Vysoké učení technické v Brně

Fakulta chemická

Rovnovážné a kinetické procesy v roztocích biologicky aktivních makromolekul pod vlivem vnějšího prostředí

Habilitační práce

Obor: Fyzikální chemie

Brno 2021   RNDr. Marek Ingr, Ph.D.

Obsah

Seznam zkratek

Zkratka	Význam
B3LYP	Hybridní funkcionál používaný v teorii funkcionálu hustoty
b5	Cytochrom b5
C_{12}-DMAP	N-dodekanoyl-4-(dimethylamino)-pyridiniový kationt
CAP	Komplexní absorbující potenciál
CI	Konfigurační interakce
CMC	Kritická micelární koncentrace
CSM	Spektrální těžiště (z angl. Center of spectral mass)
CYP	Cytochrom P450
ČOV	Čistírna odpadních vod
DABCYL	4-((4-(dimethylamino)fenyl)azo)benzoová kyselina
DDM	n-dodecyl-β-D-maltopyranosid
DEXAS	Přímá exponenciální amplifikace a sekvenování (z angl. Direct Exponential Amplification and Sequencing)
DFT	Teorie funkcionálu hustoty
DHS	Dehydrošikimát (dehydrogenovaná forma kyseliny šikimové)
DMAP	4-dimethylaminopyridin
DMR	Discrete momentum representation (reprezentace diskrétního impulsu)
DNA	Deoxyribonukleová kyselina
DVR	Discrete variable representation (reprezentace diskrétní proměnné)
EDANS	5-((2-aminoethyl)amino)naftalene-1-sulfonová kyselina
FRET	Försterův rezonanční přenos energie
GCU	Kyselina glukuronová
HA	Kyselina hyaluronová
HIV-1	Virus lidské imunodeficience (angl. Human Immunodeficiency Virus) typu 1
HIV-1 PR	Proteinasa z viru HIV-1
HPE	Hyperplocha potenciální energie
hSR	Lidská serinová racemasa
MAV	Přidružený virus myeloblastózy (z angl. Myeloblastosis associated virus)

MD	Molekulová dynamika
MPMV	Masonův-Pfizerův opičí virus (z angl. Mason-Pfizer monkey virus)
MRCI	Multireferenční konfigurační interakce
mSR	Myší serinová racemasa
NADP$^+$	Nikotinamidadenindinukleotidfosfát (oxidovaná forma)
NADPH	Nikotinamidadenindinukleotidfosfát (redukovaná forma)
NAG	N-acetylglukosamin
NAMD	Nanoscale Molecular Dynamics – programový balík pro provádění simulací metodou molekulové dynamiky
NGS	Sekvenování nové generace (z angl. Next-generation sequencing)
NMR	Nukleární magnetická rezonance
PCA	Polymerasové cyklické skládání (z angl. Polymerase cycling assembly)
PCR	Polymerasová řetězová reakce (z angl. Polymerase chain reaction)
PEPICO	Fotoelektronová-fotoiontová koincidenční spektroskopie (z angl. photoelectron-photoion coincidence spectroscopy)
PVA	Polyvinylalkohol
QSAR	Quantitative structure-activity relationship (Modely kvantitativní závislosti aktivity na struktuře)
RCCSD(T)	Metoda spřažených klastrů (Coupled cluster) vycházející z omezené (restricted) Hartreeho-Fockovy metody
RNA	Ribonukleová kyselina
rRNA	Ribozomální ribonukleová kyselina
SA	Kyselina šikimová
SR	Serinová racemasa
TAMRA	Tetramethylrhodamin
TGGE	Gelová elektroforéza v teplotním gradientu (z angl. Temperature gradient gel electrophoresis)
UHF	Neomezená (unrestricted) Hartreeho-Fockova metoda
UMP2	Møllerova-Plessetova poruchová teorie 2. řádu vycházející z vlnové funkce UHF
VMD	Visual molecular dynamics – program pro vizualizaci výsledků simulačních a kvantověchemických programů

Úvod

Předkládaná habilitační práce shrnuje veškeré moje dosavadní vědecké aktivity, jejichž výsledkem byla publikace v odborných časopisech. Protože však moje dosavadní vědecká kariéra byla poněkud heterogenní, obsahuje práce publikace na různá témata, která sice vždy souvisejí s fyzikální chemií, ale mají mnohdy přesah do oblastí biofyziky a biochemie. Vedle fyzikální chemie, z níž jsem získal doktorát roku 2000, jsem totiž vystudoval ještě magisterské studium biochemie a tomuto oboru jsem se věnoval i během svého cca desetiletého působení v komerční oblasti v první dekádě tohoto století. Většina mých prací s přesahem do biochemie se však věnuje tématům fyzikálněchemickým jako je enzymová kinetika nebo rovnováhy v roztocích oligomerních proteinů.

Z důvodů uvedených výše se tedy zahrnuté publikace věnují tématům ne zcela úzce souvisejícím, protože některá z těchto témat pocházejí ještě z doby mého studia, jiná vznikla v době mého komerčního působení, kdy jsem na částečný úvazek pracoval na katedře biochemie PřF UK, a další až po mém nástupu na FT UTB ve Zlíně v roce 2011. Proto ani nebylo možno shrnout všechna dílčí témata pod jeden zastřešující název práce. Z tohoto důvodu jsem název zvolil tak, aby odpovídal tématům, jimž se aktuálně věnuji se svojí skupinou spolupracovníků a studentů. Třebaže i zde jde o dva oddělené směry zájmu, jejich společným jmenovatelem je vliv vnějšího prostředí na biologicky aktivní makromolekuly, tedy na jejich strukturu a dynamiku v roztocích. V prvním případě jde o vliv vysokého tlaku na molekuly proteinů, který je využíván jako další nezávislá termodynamická proměnná umožňující rozšířený experimentální pohled na proteinové molekuly. V případě druhého tématu se jedná o vliv složení roztoku na chování molekul kyseliny hyaluronové a jejich derivátů. Toto druhé téma, v současné době celosvětově atraktivní, je hlavním směrem mého současného badatelského zájmu. V jeho rámci se naše skupina věnuje teoretickým simulacím zmíněných molekul v roztocích vodných i nevodných metodou molekulové dynamiky, přičemž ovšem spolupracujeme i s experimentálními skupinami ve společnosti Contipro a.s. Obě tato témata jsou pojata poněkud šířeji a jsou v textu zařazena jako první.

Pro představení svých ostatních odborných zájmů jsem však doplnil i další publikované práce z předchozího i současného období, které s hlavním tématem habilitační práce souvisejí pouze okrajově, případně jsou tematicky odlišné. Tyto práce se týkají témat enzymové kinetiky jedno- i vícesubstrátových reakcí, která má přesah i do oblasti výzkumu proteinů za vysokého tlaku. Dále jde o směry zahrnující vývoj i teoretický popis aplikací metody polymerázové

řetězové reakce (PCR), zahrnujících sekvenaci DNA, aplikace kvantitativní PCR i syntézu genů. Všechny tyto biochemicky zaměřené práce v sobě obsahují rovněž metodiky exprese a purifikace rekombinantních proteinů, jíž jsem se věnoval i v rámci svého komerčního působení. Konečně jsou v habilitační práci zařazeny i moje nejstarší odborné publikace z oblasti kvantové chemie, které se zabývají především výpočty rozptylu elektronů na molekulách, ale také výpočty reakčních mechanismů radikálových reakcí a výpočty konstant charakterizujících vliv substituentů pro metodu QSAR.

Publikace zahrnuté do této habilitační práce jsou rozděleny do tematických skupin a jsou vždy uvedeny před komentářem k dané oblasti. Aby byly odlišeny od ostatních citovaných prací, jsou označeny zkratkami P1 – P20, pod nimiž figurují v textu. Kromě toho jsou však označeny i číslem společně s jinými citacemi a jsou uvedeny v přehledu literatury.

Motivace

Téma vlivu vnějšího prostředí na biologicky aktivní makromolekuly jsem pro tuto práci zvolil z důvodu mého dlouhodobého zaujetí jevy s ním souvisejícími. Zejména v rámci svého biochemicky zaměřeného působení jsem neustále narážel na problémy vyvolané i poměrně malými změnami vnějších fyzikálních podmínek i složení roztoků. Ty se projevovaly především v oblasti exprese a purifikace rekombinantních proteinů, kde často i malá změna složení roztoku způsobila výraznou změnu v rozpustnosti proteinu, čímž jeho izolaci někdy zjednodušila, jindy naopak zcela znemožnila. Změny pH, iontové síly, polarity roztoku, a v neposlední řadě teploty však ovlivňují nejen oblast rozpustnosti či srážení proteinů, ale i disociaci podjednotek oligomerních proteinů, unfolding (rozpad terciární struktury) či drobnější konformační změny ovlivňující vazbu ligandů či enzymovou aktivitu. Výrazným příkladem důsledku těchto vlivů, s nímž jsem sám přišel do styku, je nejednoznačnost výsledků měření rovnovážné konstanty disociace dimeru HIV-1 proteinasy, které různými autory uváděny odlišně v rozmezí mnoha řádů. Pravděpodobným důvodem přitom byly právě odlišné podmínky jednotlivých stanovení, nesporný vliv však jistě měla také různost metodiky měření.

Posouzení vlivu vnějšího prostředí na molekulární strukturu a funkci je sice částečně proveditelné na základě dlouhodobé chemické zkušenosti a z ní vyvozených empirických pravidel, systematický popis těchto jevů však možný není, a patrně ani v dohledné době nebude. Vzájemné interakce mezi jednotlivými molekulami totiž tvoří natolik komplexní systém, že

v něm i relativně malá změna jak na straně makromolekuly, tak na straně prostředí, může vyvolat zásadní změnu chování systému. Markantním příkladem je chování proteinů v roztocích různých solí, jak je před více než sto lety popsal pražský rodák Franz Hofmeister, které se pro jednotlivé soli významně liší, třebaže se neliší žádné makroskopické parametry roztoku včetně jeho iontové síly. Ve svých dosavadní pracích jsem se proto zaměřoval na takové systémy, u nichž je buď jednoduše popsatelný externí vliv, nebo obsahují molekuly strukturně jednodušší než molekuly proteinů. K prvnímu případu patří výzkum disociačních rovnováh oligomerních proteinů za vysokého tlaku. Tlak zde vystupuje jako jednoduše popsatelný externí vliv umožňující posun rovnováh do koncentračních oblastí umožňujících lepší experimentální odezvu, čehož jsme využili při studiu dimerizace HIV-1 proteinasy. Do druhé skupiny prací potom patří výzkum chování molekul kyseliny hyaluronové a jejích derivátů v různých vodných a nevodných prostředích metodami výpočetní chemie. Tento biopolymer, jakkoli stále dosti komplexní, je molekulou strukturně podstatně jednodušší než proteiny, proto je snazší vlivy prostředí pochopit a vzájemně odlišit. V simulačních metodách lze navíc jednoduše měnit nejen prostředí, ale i studovanou molekulu definovaným připojováním vhodných substituentů, což umožňuje systematický popis externích vlivů. Kromě samotné zajímavosti a odborné relevantnosti studia těchto jevů přispívá atraktivitě tématu i okolnost, že kyselina hyaluronová je technologicky významným polymerem, takže výsledky našeho zkoumání mohou mít vliv na návrh pokročilých materiálů typu nosičů léčiv nebo tkáňových náhrad, jakož i na vývoj syntetických metod jejich přípravy. Proto toto téma momentálně tvoří hlavní náplň výzkumného zájmu mne samotného i mojí výzkumné skupiny.

Komentář k souboru publikovaných prací

1. Struktura a dynamika molekul kyseliny hyaluronové v různých prostředích metodou molekulové dynamiky

Publikované práce

P1   M. Ingr, E. Kutálková, J. Hrnčiřík, Hyaluronan random coils in electrolyte solutions- a molecular dynamics study, Carbohydr. Polym. 170 (2017) 289–295. [1]

P2   E. Kutálková, J. Hrnčiřík, R. Witasek, M. Ingr, Effect of solvent and ions on the structure and dynamics of a hyaluronan molecule, Carbohydr. Polym. 234 (2020) 115919. [2]

P3   E. Kutálková, J. Hrnčiřík, R. Witasek, M. Ingr, G. Huerta-Ángeles, M. Hermannová, V. Velebný, The rate and evenness of the substitutions on hyaluronan grafted by dodecanoic acid influenced by the mixed-solvent composition, Int. J. Biol. Macromol. 189 (2021) 826–836. [3]

Kyselina hyaluronová (HA) je přírodní polysacharid tvořený pravidelně se opakujícími monosacharidovými zbytky kyseliny glukuronové a N-acetylglukosaminu. Monomerní jednotku tedy tvoří jejich dvojice, jejíž struktura je znázorněna na obr. 1 a lze ji popsat vzorcem [4)-β-D-GlcpA-(1→3)-β-D-GlcpNAc-(1→]$_n$. Vyskytuje se v pojivových tkáních obratlovců, kde tvoří významnou součást mezibuněčné hmoty. Vzhledem k hodnotě pH v těchto prostředích se zde vyskytuje jako sůl, proto se pro ni používá také název hyaluronan. Jde o silně hydrofilní polymer hojně využívaný v kosmetice a farmacii pro jeho příznivé vlastnosti pro regeneraci pokožky a hojení ran [4]. Pro svoji biokompatibilitu bývá také často využíván ve vývoji nosičů léčiv [5] a tkáňových náhrad [6].
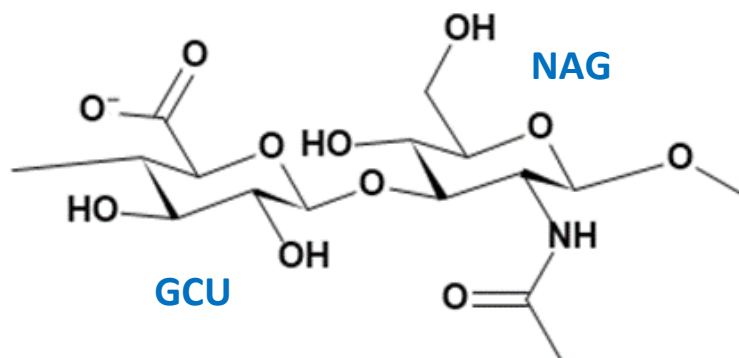
Obr. 1. Strukturní vzorec monomeru kyseliny hyaluronové tvořené monosacharidovými zbytky kyseliny glukuronové (GCU) a N-acetylglukosaminu (NAG).

Vlastnosti molekul kyseliny hyaluronové jsou určovány zejména její značnou hydrofilitou a relativně vysokou rigiditou polymerního řetězce. Hydrofilní povaha molekuly je dána polárními skupinami, mezi něž patří především karboxylová skupina kyseliny glukuronové a také hydroxylové skupiny obou monosacharidových jednotek (obr. 1). K hydrofilní povaze přispívají také kyslíkové atomy glykosidických vazeb a acetamidová skupina N-acetylglukosaminu. Všechny tyto skupiny jsou vázány v ekvatoriálních polohách vůči heterocyklům obou zbytků, zatímco v axiálních směrech jsou vázány pouze atomy vodíku. Proto tato oblast „nad" a „pod" heterocykly bývá považována za částečně hydrofobní. Uvedené polární skupiny jsou schopny vytvářet vodíkové vazby prostřednictvím atomů kyslíku, případně dusíku, s okolními molekulami vody, avšak také v rámci řetězce samotného. Tyto intramolekulární vodíkové vazby, přispívají k vyšší konformační rigiditě řetězce HA, protože omezují volné protáčení řetězců kolem vazeb glykosidických spojení.

V rámci naší výzkumné skupiny se zaměřujeme na zkoumání vlivu prostředí, v němž se molekula HA nachází, na konformaci a dynamiku polymerního řetězce molekuly HA metodami teoretické a výpočetní chemie. Základem výzkumu jsou simulace metodou molekulové dynamiky (MD), které zachycují chování řetězce HA v časovém intervalu v řádu stovek ns. Simulace probíhají v systému NpT, tedy za konstantního počtu částic, tlaku a teploty a jsou prováděny programem NAMD [7], k vizualizacím simulací a vyhodnocením některých parametrů je používán program VMD [8]. Ostatní vyhodnocení jsou prováděna pomocí námi napsaných programů. V průběhu našich výzkumných prací byla simulována různá prostředí počínaje čistou vodou (obsahující jen nutné množství protiiontů pro neutralizaci systému), přes vodné roztoky různě koncentrovaných solí až po směsná rozpouštědla tvořená vodou a organickými látkami s ní dokonale mísitelnými, konkrétně 1,4-dioxanem a terc-butanolem (2-methylpropan-2-ol).

1.1. Simulace náhodných klubek makromolekul kyseliny hyaluronové v roztocích elektrolytů

První část tohoto výzkumu, publikovaná v práci P1, byla zaměřena na simulaci náhodných klubek HA ve vodě a vodných roztocích solí. Třebaže konečným cílem bylo modelovat náhodná klubka velkých polymerních molekul o relativní molekulové hmotnosti v řádu stovek kDa až jednotek MDa, samotné simulace obsahovaly vždy jeden oligosacharid HA o délce 24 monomerních jednotek, tj. 48 monosacharidových zbytků. Simulace byly provedeny za teploty 310 K pro dvě soli, NaCl a $MgCl_2$ ve třech koncentracích, 0 M (za přítomnosti nezbytného počtu kationtů pro neutralizaci systému), 0,2 M a 1 M. Zároveň byly provedeny obdobné simulace i pro neutrální analog HA obsahující místo monosacharidové jednotky kyseliny glukuronové glukosu.

Pro každý systém byly vytvořeny modely velkých klubek HA. Tato klubka byla konstruována skládáním náhodně vybraných fragmentů simulovaných řetězců. Nejdříve byla vytvořena statistika distribuce dihedrálních úhlů jednotlivých spojení. Následně byly náhodně vybírány fragmenty řetězce HA z různých snímků simulace a tyto fragmenty byly postupně spojovány za vzniku dlouhé makromolekuly, přičemž dihedrální úhly každého spojení byly náhodně vybrány na základě výše uvedené statistiky. Touto metodou byl vytvořen celý statistický soubor náhodných klubek a byly vyhodnoceny střední hodnoty charakteristických veličin, především vzdálenosti konců řetězce a gyračního poloměru. Reprezentativní klubko je znázorněno na obr. 2.
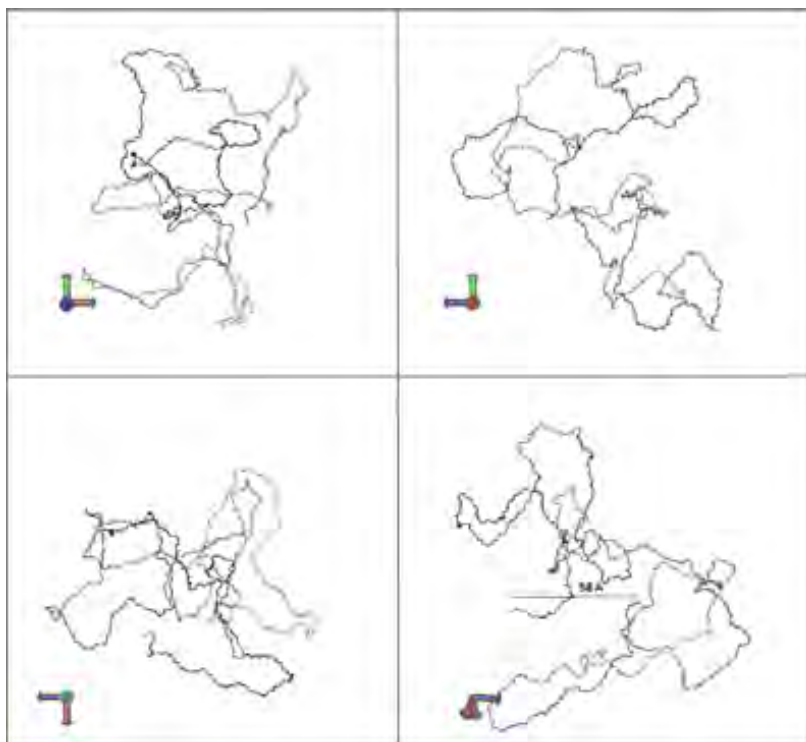
Obr. 2. Náhodné klubko makromolekuly HA vytvořené náhodným skládáním fragmentů řetězce z jednotlivých snímků simulace (každý panel obsahuje pohled z jiného směru). Simulace probíhala v 1M MgCl2, celková délka molekuly je 2000 monosacharidových zbytků. Gyrační poloměr klubka je 329 Å, vzdálenost konců 58 Å (vyznačena v pravém dolním panelu). Převzato z P1.

Analýza gyračních poloměrů náhodných klubek ukázala, že velikost klubka klesá s rostoucí koncentrací solí v roztoku (obr. 3). Tato závislost je v souladu s mnohými experimentálními studiemi, a to nejen trendově, ale i co se týče absolutní velikosti klubek vyjádřené právě gyračním poloměrem [9–13]. To potvrzuje relevanci použitého modelu a otevírá možnost simulace náhodných klubek silně botnajících polymerů pomocí metody MD.
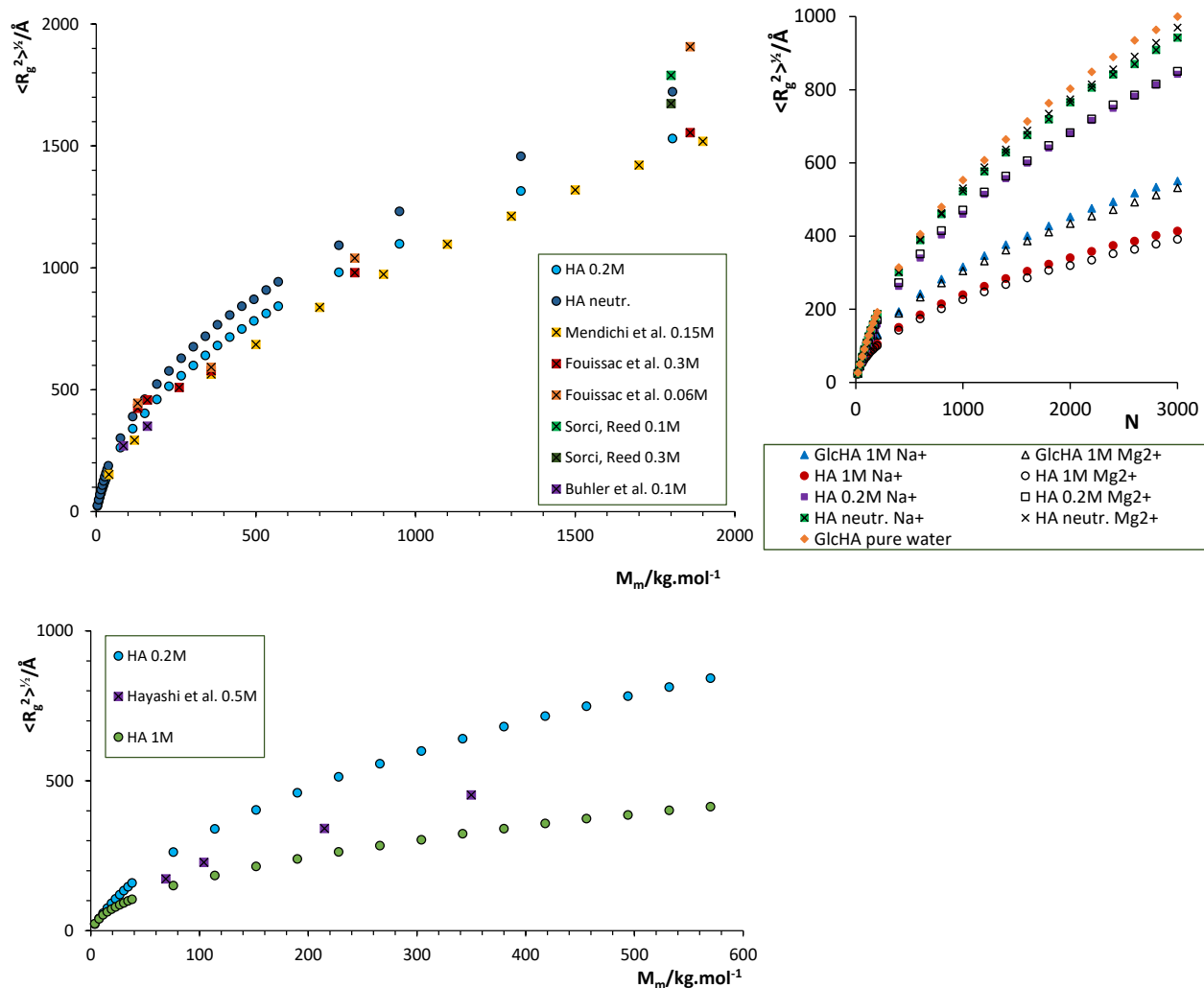
Obr. 3. Gyrační poloměr simulovaných náhodných klubek jako funkce molární hmotnosti makromolekuly. Levé panely – srovnání modelovaných klubek pro tři koncentrace NaCl – 0 M (ozn. „neutr."), 0,2 M, 1 M s různými publikovanými experimentálními výsledky [9–13]. Pravý panel – srovnánígyračních poloměrů modelovaných klubek HA a jejího neutrálního analogu (ozn. „GlcHA") jako funkce počtu monosacharidových jednotek (N) pro různé koncentrace NaCl a MgCl2. Kroužky – výsledky simulací z P1, ostatní symboly – experimentální výsledky z literatury. Převzato a upraveno z P1.

V rámci této studie byly také srovnány vlivy různých solí, NaCl a MgCl₂. U obou solí se projevovaly podobné trendy, výsledky nebyly příliš odlišné (obr. 3, vpravo). Případné zařazení dalších solí je plánováno jako možná cesta pokračování výzkumu. Bylo prokázáno, že s rostoucí koncentrací soli gyrační poloměr klubka klesá nejen pro nabitou molekulu HA, ale i pro její neutrální analog. To ukazuje, že hlavní příčinou této závislosti není pouhé stínění elektrostatické interakce mezi záporně nabitými karboxylátovými skupinami přítomnými ionty soli, naopak že tento vliv je spíše minoritní. Velikosti elektrostatické interakce mezi skupinami je totiž, vzhledem k jejich vzdálenosti cca 10 Å a vysoké relativní permitivitě vody, poměrně malá ve srovnání s ostatními interakčními energiemi mezi HA a okolním prostředím.

Detailnější analýzou vlivu iontů na velikost náhodného klubka se následně zabývala naše další studie P2.

Použitý model nahlížel na molekulu HA jako na neinteragující řetězec, nezohledňoval tedy možné vzájemné interakce monosacharidových zbytků, které se mohly náhodně ocitnout ve vzájemné blízkosti. Z tohoto důvodu roste velikost modelovaného klubka pro řetězce výrazně přesahující perzistenční délku přesně s exponentem 0,5, což je pro neinteragující klubko typické, avšak mírně odlišné od některých experimentálních prací, které se shodují na hodnotě 0,59 [9,10]. To je pravděpodobně důsledkem v modelu zanedbané interakce mezi řetězci, třebaže tato se, vzhledem k vysokému stupni botnání klubka, může projevit až u velmi dlouhých řetězců.

1.2. *Vliv iontů na konformaci a dynamiku oligosacharidů kyseliny hyaluronové v roztocích*

V další publikované práci P2 jsme se věnovali dynamice řetězce HA. Cílem bylo zejména objasnit, jakým mechanismem dochází k zmenšení gyračního poloměru molekuly HA v roztocích o zvyšující se iontové síle. Jak vyplynulo z předchozí studie P1, není možné zmenšení náhodného klubka přisoudit pouze poklesu elektrostatické repulze v důsledku zvýšení permitivity prostředí. Simulace byly prováděny pouze pro molekulu HA s plně disociovanými karboxylovými skupinami v roztocích NaCl o koncentracích 0 M, 0,2 M, 0,6 M, a 1 M za teplot 275 a 310 K.

Nejdříve jsme se zaměřili na distribuci hodnot dihedrálních (torzních) úhlů jednotlivých glykosidických spojení. Řetězec HA obsahuje spojení dvou typů, spojení 1-3 mezi uhlíkem C1 GCU a uhlíkem C3 NAG a spojení 1-4 mezi uhlíkem C1 NAG a uhlíkem C4 GCU. Konformaci každého spojení lze charakterizovat dvěma dihedrálními úhly popisujícími otočení zbytků molekuly kolem dvou jednoduchých vazeb vycházejících z atomu kyslíku příslušného glykosidického spojení. Definice úhlů je naznačena na obr. 4.
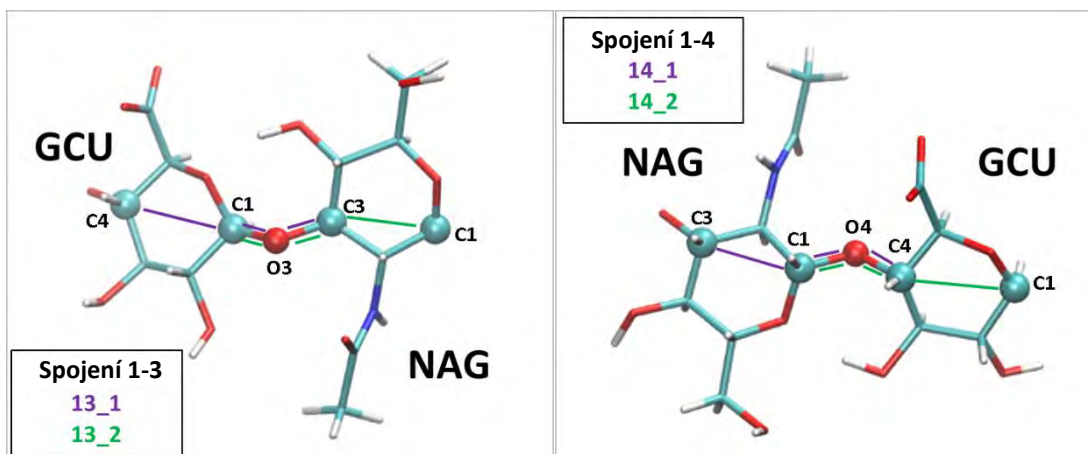
Obr. 4. Definice dihedrálních úhlů byla provedena pomocí atomů tvořících glykosidická spojení mezi monosacharidovými zbytky (takto zvolená definice se jevila jako vhodná pro konstrukci velkých klubek v práci P1). Atomy, které definují daný úhel, jsou spojeny čarami stejné barvy jako jeho označení. Např. úhel 13_1 (levý panel, fialově), který charakterizuje otočení kolem vazby mezi C1 GCU a glykosidickým kyslíkem O3 NAG je definován jako úhel svíraný rovinami C4, C1, O3 a C1, O3, C3. Převzato a upraveno z P2.

Bylo zjištěno, že každý z úhlů osciluje kolem jedné základní rovnovážné polohy s relativně úzkou distribucí hodnot s pološířkou cca 50-60° (obr. 5). Vzájemný vztah mezi dihedrálními úhly jednoho typu spojení lze znázornit pomocí Ramachandranova diagramu – tyto grafy tedy musíme pro molekulu HA zkonstruovat dva jeden pro spojení 1-3 a druhý pro spojení 1-4. Ukázalo se však, že ani samotná distribuce úhlů, ani jejich vzájemná poloha v Ramachandranově diagramu se pro jednotlivé systémy nijak výrazně neliší (Ramachandranovy diagramy proto nebyly v práci uvedeny). Bylo však pozorováno, že s koncentrací NaCl roste výskyt hodnot dihedrálních úhlů mimo oblast svých rovnovážných hodnot. To znamená, že za vyšší koncentrace soli snáze dochází k přetočení monosacharidového zbytku do jiné polohy. S nejvyšší četností k tomu dochází u dihedrálního úhlu 14_2, který je největším zdrojem flexibility molekuly HA (obr. 5). Dihedrální úhel je v přetočeném stavu stabilní jednotky až desítky ns, poté se vrací do rovnovážné polohy.
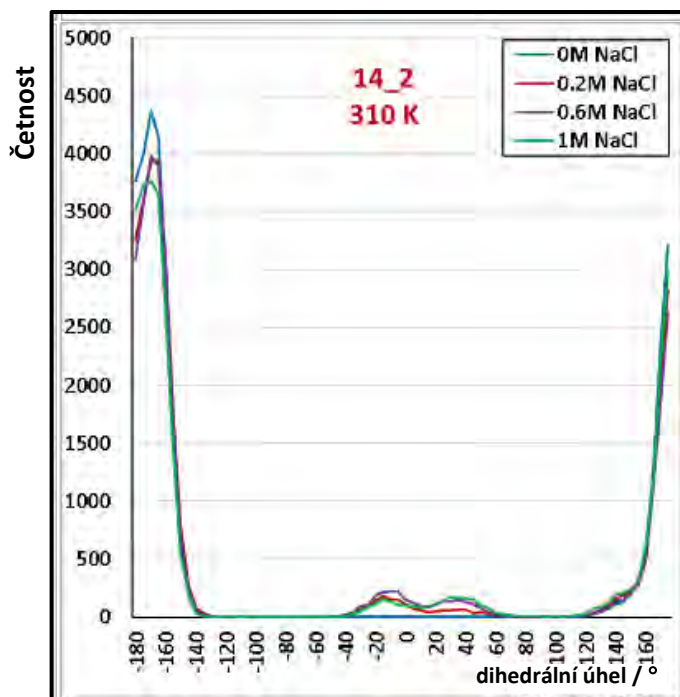
Obr. 5. Distribuce dihedrálních úhlů 14_2 za teploty 310 K pro různé koncentrace soli. Monosacharidovým zbytkům v rovnovážné poloze odpovídá majoritní pík s maximem kolem -160°, zbytkům přetočeným z rovnovážné polohy odpovídají minoritní oblasti uprostřed grafu. Jejich velikost roste s koncentrací NaCl, v čisté vodě se prakticky nevyskytuje. Převzato a upraveno z P2.

Dále jsme se proto zaměřili na pozorování, jak řetězec na takové přetočení zareaguje. Nejdříve bylo zjištěno, že protočení v jednom dihedrálním úhlu nevede k bezprostřední kompenzaci protočením jinde v řetězci, naopak všechny ostatní dihedrální úhly v celém řetězci zůstávají, až na náhodné fluktuace, beze změny. Napětí vzniklé protočením tedy disipuje do ostatních stupňů volnosti, což umožňuje protočení jednotlivého dihedrálního úhlu bez nutnosti změny jiného. Kolem vazby odpovídající protočenému dihedrálu však vznikne konformace energeticky nevýhodná, což vede k postupné relaxaci do konformace odlišné. Na rozdíl od konformace výchozí, v níž řetězec HA neobsahuje prudší zakřivení, spěje řetězec po přetočení do konformace typu vlásenky, tedy k prudkému zahnutí o téměř 180°, což vede k celkovému zkrácení řetězce. Po určitém čase, typicky v řádu jednotek až desítek ns, dojde ke zpětnému protočení uvažovaného dihedrálního úhlu do jeho rovnovážné hodnoty, v důsledku čehož se řetězec opět „narovná" do rovnovážné konformace. V některých případech proběhne konformační změna ve značném rozsahu, jindy jen zčásti, pokud ke zpětnému přetočení dojde příliš brzy. Průběh několika takových strukturních změn je zachycen na obr. 6. Je tedy pravděpodobné, že příčinou změny velikosti klubka se změnou iontové síly roztoku je větší flexibilita řetězce projevující se častějšími konformačními změnami vedoucími k jeho opakovanému zkracování.
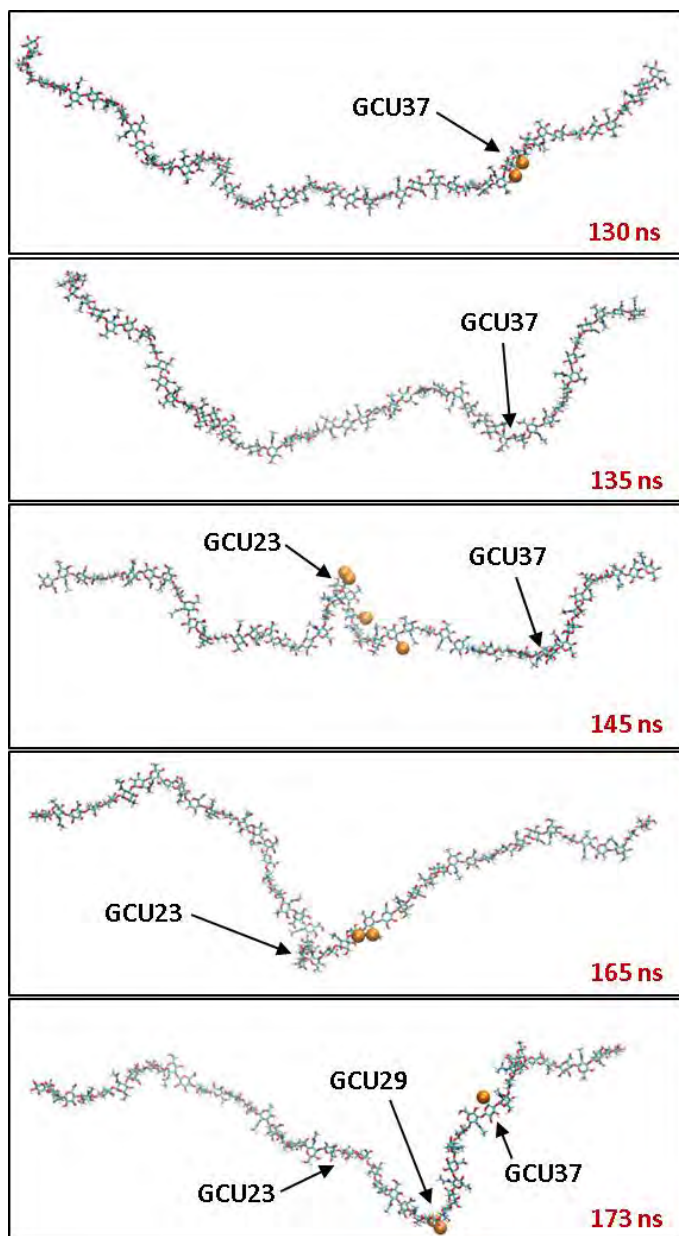
Obr. 6. Ukázka dynamiky řetězce HA. Série snímků z MD simulace ukazuje postupný vznik a zánik ohybů řetězce vzniklých v důsledku přetočení zbytků GCU č. 37, 23 a 29 (chronologicky). Žluté kuličky zobrazují iont Na+, které svojí přítomností přispěly ke vzniku přetočení. Převzato z P2.

Vysvětlení závislosti četnosti přetočení dihedrálních úhlů na koncentraci NaCl je založeno na dvou jevech. Prvním z nich je celkový pokles počtu vodíkových vazeb, a to jak intramolekulárních v rámci řetězce HA, tak mezi HA a vodou. To dává molekule HA za vyšší koncentrace NaCl větší volnost k možným konformačním změnám. Vedle toho bylo také pozorováno, že výskyt přetočených dihedrálních úhlů 14_2 pozitivně koreluje s výskytem iontů Na$^+$ v blízkosti kyslíku O5 zbytku kyseliny glukuronové. Ten nejdříve s koncentrací NaCl roste, ale nabývá maxima přibližně u 0,6M NaCl a následně mírně klesá. Toto maximum je patrné zejména za nižší teploty 275 K, kdy je výskyt iontů Na$^+$ v této poloze dokonce vyšší než v bezprostřední blízkosti karboxylátové skupiny. Toto maximum také koreluje s minimální hodnotou gyračního poloměru náhodných klubek simulovaných metodou popsanou v článku P1. Popsaný trend vazby iontů do uvedeného místa nebyl v plném rozsahu vysvětlen,

pravděpodobně souvisí se závislostmi solvatačních energií jak karboxylátové skupiny, případně dalších skupin HA, tak samotných sodných kationtů, na koncentraci NaCl.

Z prací P1 a P2 vyplývají následující závěry. Metodou molekulové dynamiky lze simulovat oligosacharidy HA, jejichž charakteristické vlastnosti, jakož i vlastnosti velkých náhodných klubek z nich vytvořených, odpovídají experimentálním hodnotám. Metoda je tedy vhodná pro studium fyzikálněchemických vlastností HA. Simulací získaný gyrační poloměr klesá s rostoucí koncentrací solí v roztoku, což je v souladu s experimentálními zjištěními. Zejména za nízkých teplot však tento pokles nemusí být monotónní, gyrační poloměr nabývá minima při koncentraci NaCl cca 0,6 M. Závislost gyračního poloměru na koncentraci solí souvisí s dynamickým chováním řetězce, především s náhodným protáčením jednotlivých monosacharidových zbytků kolem glykosidických vazeb z rovnovážné polohy, jejichž četnost roste s koncentrací solí. Tato protočení vyvolají konformační změnu vedoucí k vytvoření ohybu na řetězci HA, což vede k jeho dočasnému zkrácení, a tedy následně ke zmenšení gyračního poloměru.

1.3. Oligosacharidy kyseliny hyaluronové ve směsných rozpouštědlech

V další publikované práci (P3) jsme se věnovali chování oligosacharidů HA ve směsných rozpouštědlech tvořených vodou a organickými látkami s vodou mísitelnými, 1,4-dioxanem (dále jen dioxan) a terc-butanolem. Výzkum probíhal ve spolupráci se společností Contipro a.s. a jejím primárním cílem bylo vysvětlení různé reaktivity HA v esterifikačních reakcích na jejích hydroxylových skupinách v závislosti na použitém rozpouštědle. Tyto reakce jsou prováděny metodou směsných anhydridů za katalýzy 4-dimethylaminopyridinem (DMAP) [14].

Nejprve bylo studováno chování samotných oligosacharidů HA ve směsných rozpouštědlech obou typů připravených v různých poměrech obou složek za dvou různých teplot. Krom jiného byla pozorována především separace směsných rozpouštědel v bezprostřední blízkosti molekuly HA. Ukázalo se, že molekula HA je solvatována přednostně molekulami vody, zatímco organickou složku odpuzuje. Tento efekt je přitom výraznější ve směsi voda:terc-butanol. Následně byly simulovány molekuly HA substituované jedním dodekanoylovým zbytkem vázaným esterovou vazbou na hydroxylovou skupinu uhlíku C6 NAG. Separace směsného rozpouštědla byla v oblasti řetězce HA stejná jako u nesubstituované molekuly HA, avšak v okolí postranního alifatického řetězce byla přesně opačná – řetězec byl solvatován především organickou složkou. V tomto případě byl efekt výrazně silnější v případě

terc-butanolu, což je pravděpodobně dáno jeho amfifilní povahou, která symetričtější molekule dioxanu chybí. Rozložení molekul směsného rozpouštědla kolem substituované molekuly HA ilustruje obr. 7. Na obr. 8 jsou pak zachyceny radiální distribuční funkce jednotlivých složek rozpouštědla kolem jednotlivých částí této molekuly.
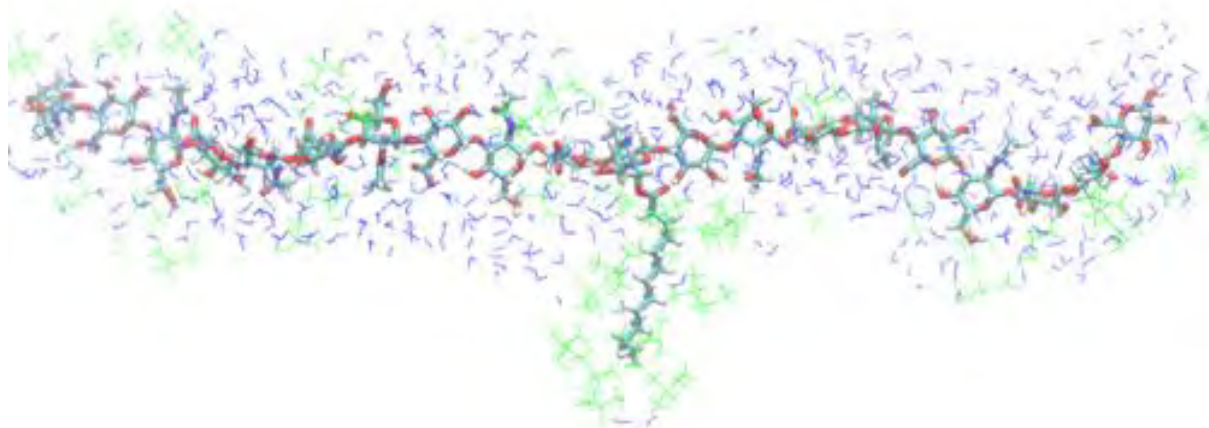
Obr. 7. Oligosacharid HA substituovaný dodekanoylovým zbytkem ve směsném rozpouštědle tvořeném vodou (modré molekuly) a terc-butanolem (zelené molekuly) v objemovém poměru 1:1. V okolí řetězce HA převládá voda, v blízkosti substituentu jednoznačně dominuje terc-butanol. Převzato a upraveno z P3.
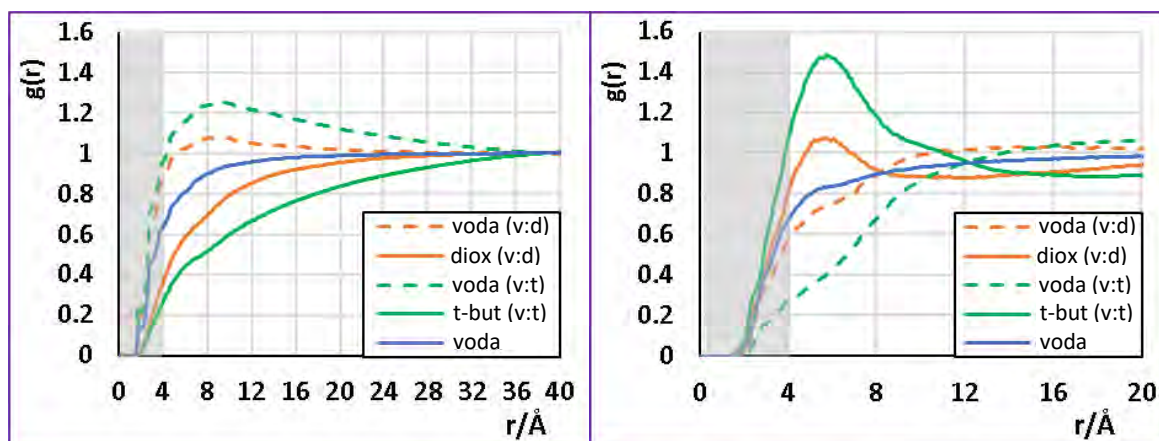
Obr. 8. Radiální distribuční funkce g(r) popisující výskyt jednotlivých komponent směsného rozpouštědla kolem jedlotlivých částí substituované molekuly HA. Vlevo: distribuce v okolí řetězce HA. Vpravo: distribuce v okolí dodekanoylového substituentu. Oranžově – směs voda:dioxan (v:d), zeleně – směs voda:terc-butanol (v:t), modře – čistá voda. Převzato a upraveno z P3.

Konečně byly simulovány systémy obsahující nesubstituovanou nebo substituovanou molekulu HA a reaktant substituční reakce N-dodekanoyl-4-(dimethylamino)-pyridiniový kationt (dále jen C_{12}-DMAP), který vzniká v reakční směsi (v reálném experimentálním provedení) vazbou dodekanoylového zbytku, původně přítomného jako součást směsného anhydridu, na katalyzátor DMAP. Samotný alifatický řetězec reakčního prekurzoru vykazuje podobnou separaci směsného rozpouštědla jako řetězec vázaný na HA. Bylo prokázáno, že

výskyt C$_{12}$-DMAP v blízkosti molekuly HA je podstatně četnější ve směsi voda:dioxan než voda:terc-butanol, přičemž pravděpodobným důvodem tohoto jevu je právě menší separace směsného rozpouštědla u první směsi. Naopak silná separace směsi voda:terc-butanol představuje energetickou bariéru pro přístup C$_{12}$-DMAP k molekule HA. Toto pozorování je v dobrém souladu s experimentálním provedením reakce, která ve směsi voda:dioxan vykazuje výrazně vyšší stupeň substituce než ve směsi voda:terc-butanol. Reakční prekurzor C$_{12}$-DMAP má navíc v roztoku voda:terc-butanol silnou afinitu k alifatickému řetězci již navázanému na HA a tato interakce mu umožňuje přiblížení k HA, které je v tomto rozpouštědle jinak méně četné. Naopak ve směsi voda:dioxan se tento efekt prakticky neprojevuje. To vysvětluje experimentální zjištění, že ve směsi voda:terc-butanol bude docházet k častějšímu shlukování substituentů, zatímco ve směsi voda:dioxan bude jejich distribuce po délce řetězce rovnoměrnější. Všechny závěry naší teoretické studie byly navíc nezávisle potvrzeny jinou experimentální prací našich spolupracovníků z Contipro a.s. a dalších, jimž se podařilo prokázat obě zásadní pozorování, tedy vyšší stupeň substituce ve směsi voda:dioxan a častější shlukování substituentů ve směsi voda:terc-butanol i jiným alifatickým substituentem, a to oleoylem, který je oproti dodekanoylu o 6 uhlíkových atomů delší a obsahuje jednu dvojnou vazbu [15]. Na základě této shody teorie s experimentem lze předpokládat obecnější platnost získaných závěrů o vlivu směsného rozpouštědla na průběh esterifikačních chemických reakcí na řetězci HA.

1.4. Další plány a perspektivy

V návaznosti na publikované výsledky pokračujeme ve výzkumu struktury a dynamiky molekul HA několika směry. První z nich představuje simulace interakcí dvou řetězců HA ve vodných roztocích. Některými experimenty provedenými v 90. letech 20. století byla naznačena možnost existence reverzibilních dvojšroubovicových struktur molekul HA ve vodném prostředí [16–18]. To však bylo později poměrně přesvědčivě vyvráceno jinými experimentálními pracemi [19–21]. Naše simulace, kterými se snažíme objasnit možnost vzniku takových struktur, ukazují, že existence pravidelných dvojšroubovicových struktur ve vodných roztocích pravděpodobně možná není, nicméně s rostoucí koncentrací NaCl roste četnost vzájemných interakcí obou řetězců, která může hrát podstatnou roli v chování velkých makromolekulárních klubek nebo koncentrovaných roztoků HA. To lze nepřímo potvrdit experimentálně pozorovaným chováním těchto roztoků metodami viskozimetrie nebo rozptylu světla. Výsledky této studie jsou připravovány k publikaci.

Vedle výzkumu interakcí molekul HA ve vodě pokračujeme zejména ve zkoumání chování molekul HA ve směsných rozpouštědlech. Naše dosavadní výsledky ukazují především

výraznější interakci HA s ionty v roztoku v důsledku celkově nižší polarity směsného rozpouštědla. To má ale za následek rovněž pevnější interakci mezi řetězci, která v určitých poměrech koncentrací směsného rozpouštědla může vést až ke vzniku stabilních duplexových struktur. Některé výsledky této části jsou momentálně připravovány k publikaci, další jsou postupně dokončovány.

Relativně nemnoho výzkumných prací se dosud zabývalo simulacemi substituovaných řetězců HA, třebaže právě chemicky modifikované molekuly HA jsou díky své nižší hydrofilitě často technologicky využívány. Vliv substitucí na strukturu oligosacharidů HA ve vodném prostředí byl studován experimentálně i teoreticky ve skupině A. Mohse [22,23], přičemž byl prokázán vliv hydrofobních substituentů na tvorbu kompaktnějších makromolekulárních klubek. V rámci dalšího pokračování našeho výzkumu proto plánujeme navázat na studie uvedeného typu a zahrnout do nich i vliv směsných rozpouštědel, protože právě v nich často probíhají substituční reakce vedoucí k navázání hydrofobních zbytků na molekulu. Rovněž plánujeme podrobnější výzkum vlivu koncentrace solí na substituované řetězce HA. Simulace těchto systémů momentálně probíhají.

2. Rovnováhy a kinetika v roztocích oligomerních proteinů za vysokého tlaku

Publikované práce

P4　　E. Kutálková, J. Hrnčiřík, M. Ingr, Pressure induced structural changes and dimer destabilization of HIV-1 protease studied by molecular dynamics simulations, Phys. Chem. Chem. Phys. 16 (2014) 25906–25915. [24]

P5　　M. Ingr, R. Lange, V. Halabalova, A. Yehya, J. Hrncirik, D. Chevalier-Lucia, L. Palmade, C. Blayo, J. Konvalinka, E. Dumay, Inhibitor and Substrate Binding Induced Stability of HIV-1 Protease against Sequential Dissociation and Unfolding Revealed by High Pressure Spectroscopy and Kinetics, Plos One. 10 (2015) e0119099. [25] .

P6　　Ingr, E. Kutálková, J. Hrnčiřík, R. Lange, Equilibria of oligomeric proteins under high pressure - A theoretical description, J. Theor. Biol. 411 (2016) 16–26. [26]

Chemické rovnováhy v roztocích jsou ovlivnitelné řadou externích faktorů, mezi něž zpravidla počítáme především teplotu a složení rozpouštědla, v němž reakce probíhá, v případě biochemických reakcí probíhajících ve vodném prostředí jsou nejvýznamnějšími faktory iontová síla a pH. Obvykle však mezi ovlivňující faktory nepočítáme tlak, jehož vliv na reakce v kondenzované fázi je vpravdě minimální. To je důsledkem nízké stlačitelnosti kondenzovaných látek, která neumožňuje velké změny objemů reakčních směsí v důsledku probíhajících chemických reakcí.

Změnu rovnovážné konstanty chemické reakce, resp. jejího logaritmického vyjádření $pK_a = -\log K_a$, s měnícím se tlakem lze vyjádřit vztahem

$$\left(\frac{\partial pK_a}{\partial p}\right)_T = -\log e \left(\frac{\partial \ln K_a}{\partial p}\right)_T = \frac{\log e}{RT}\left(\frac{\partial \Delta G_r^0}{\partial p}\right)_T = \frac{\log e}{RT}\Delta V_r^0, \tag{1}$$

kde p je tlak, T je termodynamická teplota, R je molární plynová konstanta, ΔG_r^0 je standardní molární reakční změna Gibbsovy energie a ΔV_r^0 je standardní molární reakční změna objemu. Ze vztahu je patrné, že pokud molární objem reakční směsi během reakce, případně jiného fyzikálněchemického děje, narůstá, zvyšuje se i pK_a, tedy klesá rovnovážná konstanta, a tudíž se rovnováha posune na stranu reaktantů. Naopak pokud celkový objem reakční směsi klesá, rovnováha se posouvá na stranu produktů, což je v souladu s principy chemické termodynamiky. Rovnice (1) však také ukazuje, že pokud má dojít k posunu pK_a o jednotku, tedy změnu rovnovážné konstanty o jeden řád, musí platit vztah

$$\Delta p \Delta V_r^0 = \frac{RT}{\log e}, \tag{2}$$

což za teploty 298,15 K (25 °C) znamená, že

$$\Delta p = \frac{5708 \ ml \ MPa \ mol^{-1}}{\Delta V_r^0}. \tag{3}$$

Protože např. u proteinových molekul se změna molárního objemu související s procesy typu asociace nebo disociace podjednotek pohybuje v řádu desítek až stovek $ml \ mol^{-1}$, je k posunu rovnovážné konstanty o řád zapotřebí tlaku v řádu stovek MPa. Takové tlaky, s výjimkou dolní meze tohoto intervalu, se v biosféře Země vůbec nevyskytují, jelikož odpovídají hloubkám v řádu desítek km pod vodní hladinou. Přesto byly některé výzkumy v této oblasti zaměřeny právě na vliv tlaku na proteiny hlubokomořských mikroorganismů [27].

Výzkum proteinů za vysokého tlaku tedy zpravidla nemá primárně význam pro popis jevů probíhajících v přírodě, je však zajímavý z jiného hlediska. Vzhledem k nízké stlačitelnosti vody totiž aplikace vysokého tlaku na vodný roztok nevede k dramatickému zvýšení vnitřní energie systému, proto tímto způsobem můžeme posunout rovnovážnou konstantu různých procesů, aniž bychom riskovali tepelnou denaturaci proteinových molekul. Toho lze využít k měření rovnovážných konstant procesů, jejichž rovnováha je výrazně posunuta na jednu stranu, takže není jednoduché pozorovat rovnovážné zastoupení obou stavů reagujících látek [28].

Typickým příkladem je měření disociační konstanty oligomerního proteinu složeného z několika podjednotek – pro jednoduchost předpokládejme homodimer. Jestliže je rovnováha silně posunuta na stranu dimeru, je možné pozorovat rovnovážné zastoupení monomeru a dimeru jen za velmi nízkých koncentrací, což následně znemožňuje dobrou detekci většinou použitelných experimentálních metod. Pokud ovšem disociace dimeru vede ke snížení celkového objemu systému, což je typický případ, lze aplikací vysokého tlaku rovnovážnou konstantu posunout tak, že rovnoměrné zastoupení obou forem je pozorovatelné v takové oblasti koncentrací, kde je stanovení poměru monomeru a dimeru snazší. Proto můžeme rovnovážnou konstantu stanovit za zvýšeného tlaku a následně její hodnotu přepočítat na tlak atmosférický. Vysokotlakých metod bylo v minulosti použito k výzkumu proteinů dimerních [29–31], tetramerních [32–34], hexamerních [35], vyšších oligomerů [36–38] či virových kapsid [39–41], prionů [42], nebo dokonce polymerních proteinových struktur jako mikrotubuly nebo mikrofilamenta [43–45]. Procesy, u nichž naopak vysoký tlak podporuje asociaci proteinových molekul do vyšších útvarů, jsou méně obvyklé, ale vyskytují se rovněž, jde zpravidla o agregace nespecifikovaného počtu podjednotek [46].

Použití vysokotlakých metod se neomezuje pouze na asociační rovnováhy podjednotkových proteinů, využívá se rovněž ke studiu unfoldingu proteinů [47–49] a dalších procesů [50,51]. Vysokotlaké experimenty se navíc neomezují jen na studium rovnováh, ale také na kinetiku zkoumaných procesů [48,52].

Studium proteinů za vysokého tlaku je z technického hlediska náročnější oproti experimentům za tlaku atmosférického, neboť zkoumaný vzorek musí být umístěn v prostředí, kde je vysoký tlak udržován. V praxi to znamená uzavření kyvety se vzorkem pružnou membránou a její umístění do cely naplněné vodou, v níž se pomocí vysokotlaké pumpy vytvoří požadovaný tlak. Pokud jsou zapotřebí vstupy pro světelné paprsky pro účely spektroskopických měření, jsou v cele vytvořeny otvory opatřené safírovými průzory.

Nejběžnějšími metodami používanými v kombinaci s vysokým tlakem jsou spektrofotometrie a fluorimetrie (použité ve většině výše citovaných prací), vysoký tlak však byl aplikován i v kombinaci s rozptylem světla [53] nebo neutronů [27] či gelovou elektroforézou [54,55]. Pro studium nativních konformací jednotlivých podjednotek oligomerních proteinů jsou pak obzvláště vhodné metody vysokotlaké rentgenové krystalografie a vysokotlaké NMR [56]. I přes zmiňovanou technickou komplikovanost je tedy paleta vysokotlakých metod poměrně pestrá, což umožňuje jejich aplikaci na širokou škálu výzkumných problémů.

V naší skupině jsme se věnovali zejména výzkumu rovnovážných a kinetických vlastností proteinasy z viru HIV 1 (HIV-1 PR), a to jak experimentálně, tak teoreticky. HIV-1 PR je homodimerní aspartátovou proteinasou, její katalytické centrum tedy tvoří dva zbytky kyseliny asparagové, z nichž každý je součástí jedné podjednotky. HIV-1 PR je malým proteinem o relativní molekulové hmotnosti monomeru 10,8 kDa. Aktivní místo proteinasy se nachází v relativně hydrofobní dutině, do které je před reakcí vázán substrát. Ten je v průběhu reakce uzavřen do aktivního místa pohyblivými chlopněmi (flaps), které naopak ve stavu bez substrátu mohou být otevřeny. Dimerní struktura je stabilizována zejména dimerizační doménou tvořenou vzájemně provázanými N- i C-konci obou polypeptidových řetězců. Struktura je znázorněna na obr. 9.

2.1. *Simulace vlivu vysokého tlaku na molekulu HIV-1 proteinasy metodou molekulové dynamiky*

V první práci na toto téma (P4) jsme simulovali molekulu HIV-1 PR za různých tlaků od 1 do 600 atm metodou molekulové dynamiky. Tento přístup byl dříve několikrát použit na různé proteiny, jeho aplikace shrnuje přehledový článek [57]. Cílem studie bylo popsat konformační změny molekuly a určit objemovou změnu provázející disociaci dimeru. Nejdříve byly sledovány změny konformace dimerní molekuly HIV-1 proteainsy kvantitativně charakterizované jednak vzdáleností α-uhlíků aminokyselin Pro39 a Pro39', které leží na nejvzdálenějších místech jednotlivých monomerů (byly měřeny i jiné vzdálenosti, které však vykazovaly jen nepatrné změny, viz obr. 9), jednak gyračním poloměrem molekuly. Obě tyto veličiny s rostoucím tlakem monotónně klesají, což jednoznačně prokazuje vliv tlaku na konformaci dimeru, změny jsou však relativně malé v porovnání s celkovou velikostí molekuly (obr. 9).
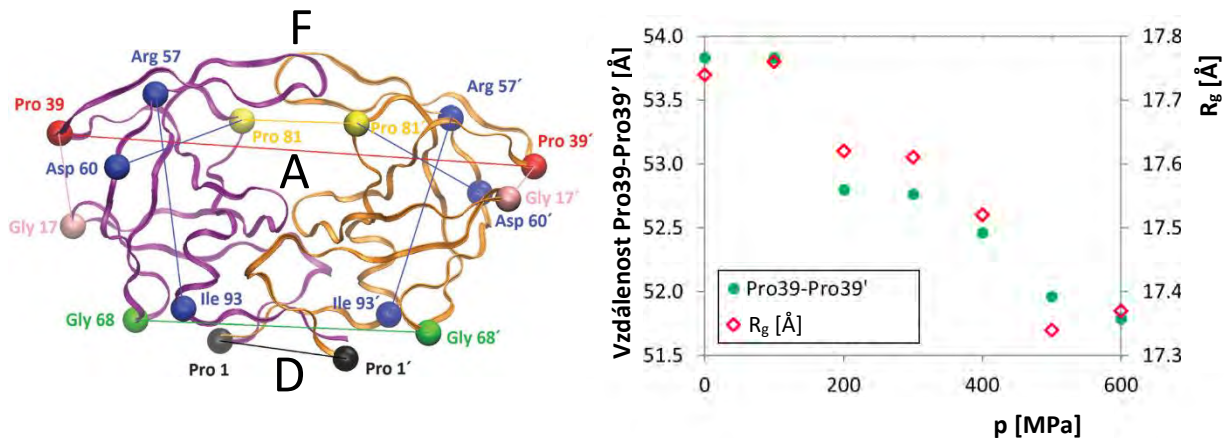
Obr. 9. Vlevo: struktura molekuly HIV-1 PR s vyznačením významných domén: A – aktivní místo obklopené dutinou pro vazbu substrátu, F – chlopně (flaps), D – dimerizační doména. Barevnými kuličkami a čarami jsou vyznačeny vzdálenosti, jejichž změna byla sledována v závislosti na tlaku v systému. Vpravo: Závislost vzdálenosti Pro39-Pro39' (která vykazovala nejvýraznější změnu) a gyračního poloměru Rg molekuly na tlaku. Převzato a upraveno z P4.

Aby bylo možno posoudit a termodynamicky kvantifikovat vliv vysokého tlaku na rovnováhu monomer-dimer, byla vypočtena změna objemu související s disociací dimeru na dva monomery. Protože celková změna objemu systému je velmi malá v porovnání s objemem simulačního boxu, byla použita aproximativní metoda zjištění objemového rozdílu založená na stanovení počtu molekul vody v dutině aktivního místa, neboť lze předpokládat, že míra naplnění této dutiny vodou ovlivní celkovou objemovou změnu systému nejvýrazněji. Vycházíme z předpokladu, že její hydrofobní vnitřní povrch odpuzuje molekuly vody, takže celková hustota těchto molekul je uvnitř dutiny za atmosférického tlaku nižší než ve volné vodě. Zvyšováním tlaku je však tato odpudivá tendence překonávána a do dutiny se dostává více molekul vody z okolí, čímž klesá celkový objem systému. Obdobný výpočet byl proveden také pro monomer, kde je efekt zřejmě menší, protože monomer obsahuje pouze polovinu dutiny aktivního místa (obr. 10).
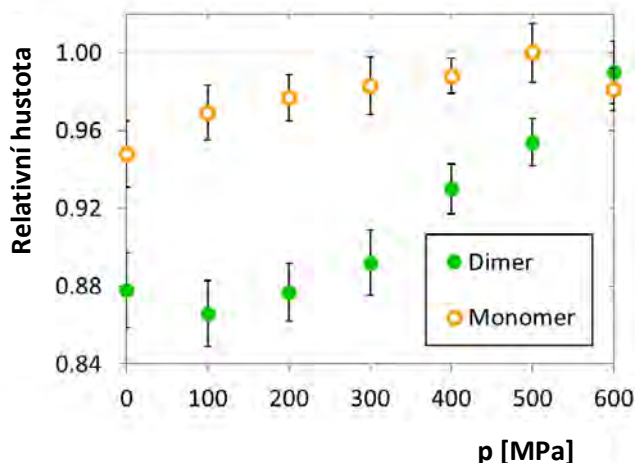
Obr. 10. Relativní hustota vody (oproti vodě volné) v hydrofobní dutině aktivního místa dimeru a v jí odpovídající oblasti v molekule monomeru. Převzato a upraveno z P4.

Bylo zjištěno, že za atmosférického tlaku až do necelých 100 MPa je objemová změna disociace dokonce kladná, tedy že dimer je za těchto podmínek tlakově stabilizován. Za tlaků vyšších ovšem objemová změna klesá do záporných hodnot, což vede k očekávané destabilizaci dimeru. Velikost objemové změny je minimální (tj. nejvíce záporná) pro tlak 300 MPa, poté se vrací zpět k nulové hodnotě. Toto chování je pravděpodobně důsledkem skutečnosti, že dvě poloviny dutiny aktivního místa obsažené ve dvou monomerech za atmosférického tlaku dokáží odpudit více vody než kompletní dutina v dimeru. Jakmile však tlak vzroste, otevřené poloviny dutiny již nedokáží zabránit přiblížení molekul vody tak efektivně, jako dutina kompletní, proto se celkový objem monomerů zmenšuje rychleji než objem dimeru. Při dalším vzrůstu tlaku však již objem monomerů dále neklesá, na rozdíl od dimeru, jehož hydrofobní dutina je již také zaplňována vodou. Tím se objemová změna vrací zpět k nule. Je tedy patrno, že objemová změna provázející disociaci dimeru HIV-1 PR není konstantní přes všechny tlaky, ale může se dosti podstatně měnit.

Aby bylo možno provést srovnání s experimentálními daty, byla vypočtena hodnota průměrné objemové změny pro interval od nuly do daného tlaku. Takto vypočtené hodnoty odpovídají objemovým změnám ovlivňující experimentální chování při daném tlaku. Průměrné hodnoty začínají opět v kladných oblastech, do záporných hodnot přecházejí u cca 160 MPa a nakonec se stabilizují kolem hodnoty $-50\ ml/mol$. Tomu potom odpovídá i průběh změny hodnoty disociační konstanty dimeru (obr. 11). Tato změna je v oblasti do 160 MPa nepatrně kladná, poté však klesá do záporných hodnot a pro tlaky nad 300 MPa se stabilizuje kolem hodnoty $\Delta pK_d = -4$, tedy rovnovážná konstanta disociace dimeru vzroste asi o 4 řády. To svědčí o výrazné tlakové destabilizaci dimeru. Vypočtené hodnoty změny objemu i disociační konstanty jsou, i přes vysokou míru aproximace, v dobrém souladu s naší následnou experimentální prací na stejném enzymu P5.

26
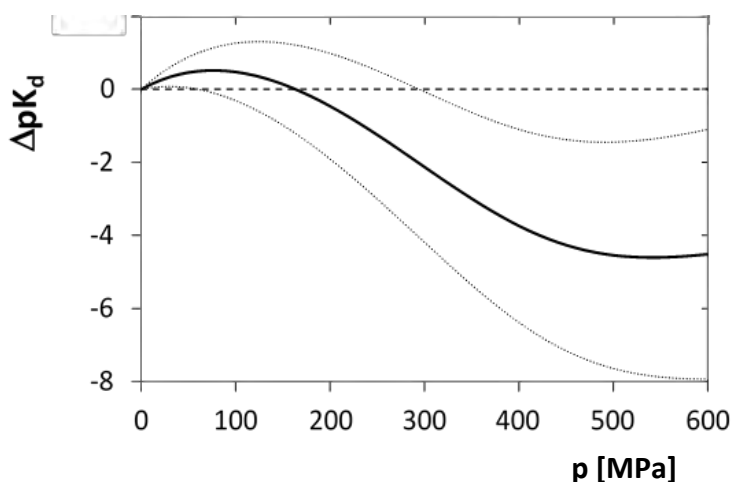
Obr. 11. Tlaková závislost změny disociační konstanty vyjádřené logaritmicky jako pK_d (plná čára). Směrodatné odchylky jsou vyznačeny čarami tečkovanými. Převzato a upraveno z P4.

Naše teoretická studie tedy ukazuje potenciál simulací metodou molekulové dynamiky pro výpočty objemových změn molekul proteinů za vysokého tlaku. Dále ukazuje, že objemová změna nemusí být konstantní v celém intervalu aplikovaných tlaků, což je v experimentálních pracích obvykle předpokládáno. Ovlivnění rovnovážné konstanty je však v dobré shodě s experimentem.

2.2. Studium rovnovážných a kinetických vlastností HIV-1 proteinasy metodou vysokotlaké fluorescenční spektroskopie

Zároveň se simulacemi jsme ve spolupráci se skupinami Dr. Reinharda Langeho a prof. Eliane Dumay z Univerzity v Montpellieru provedli analogickou experimentální studii, která je obsahem práce P5. V této studii jsme vystavili roztok HIV-1 PR postupně se zvyšujícímu tlaku a pozorovali fluorescenční odezvu v oblasti fluorescence aromatických aminokyselin, zejména tryptofanu. Pozorované změny v intenzitě a poloze těžiště emisního spektra svědčí o strukturních změnách v molekule proteinu.

Jako první byl ověřen obecně známý předpoklad, že kompetitivní inhibitory HIV-1 PR, které se vážou do aktivního místa enzymu, stabilizují jeho dimerní strukturu. Bylo zjištěno, že zatímco bez inhibitoru lze pozorovat typickou denaturační křivku, v přítomnosti inhibitoru darunaviru [58,59] se sledované spektrální charakteristiky buď nemění, nebo jen monotónně klesají. To potvrzuje hypotézu, že inhibitor nejen stabilizuje dimerní strukturu, ale že takto stabilizovaný dimer nemůže ani podléhat rozpadu terciární struktury (unfoldingu), a to ani za vysokého tlaku.

Protože vysoký tlak často vyvolává agregaci proteinů, byl měřen i tento jev, který by mohl další stanovení komplikovat. Bylo prokázáno, že agregace proteinu nastává až za tlaků vyšších než 350 MPa, což je hodnota umožňující provedení měření parametrů dimerizace i unfoldingu (popsaných dále).

Abychom odlišili procesy disociace dimeru a následného unfoldingu monomerů, sledovali jsme rovněž závislost tlaků, při nichž dochází k jednotlivým spektrálním změnám, na koncentraci. V případě unfoldingu, který je v obou směrem procesem prvního řádu, dochází k přechodu ze strukturovaného do nestrukturovaného stavu za stejného tlaku bez ohledu na koncentraci proteinu. Na druhou stranu přechod mezi dimerem a monomerem, kdy z jedné částice vznikají dvě, je na tlaku výrazně závislý, protože vyšší koncentrace stabilizuje dimerní stav. Proto za nízkých koncentrací dojde k disociaci dimeru za nižšího tlaku než v případě koncentrace vyšší. V případě HIV-1 PR byla pozorována zřejmá koncentrační závislost polohy inflexního bodu tlakové závislosti těžiště emisního spektra (obr. 12, vlevo), naopak inflexní bod závislosti intenzity spektra na tlaku se s koncentrací neměnil. Proto byla první závislost považována za odezvu disociace dimeru, zatímco druhá za odezvu unfoldingu disociovaných monomerů.

Z první závislosti byla následně vypočtena hodnota disociační konstanty i objemová změna provázející tento proces. Byla použita linearizovaná rovnice závislosti polohy inflexního bodu Δp_{inf} na logaritmu koncentrace proteinu $[M_0]$ (vyjádřené jako koncentrace formálního monomeru)

$$\Delta p_{inf}([M_0]) = -\frac{RT}{\Delta V_d}\ln[M_0] + \frac{RT}{\Delta V_d}\ln\frac{(1+\sqrt{2})K_{d,atm}}{4}, \tag{4}$$

kde $K_{d,atm}$ je disociační konstanta dimeru za atmosférického tlaku a ΔV_d je objemová změna disociace (považovaná za konstantu). Z proložení experimentálně získaných hodnot inflexních bodů touto lineární závislostí (obr. 12, vpravo) byly získány hodnoty obou parametrů, $\Delta V_d = -32,5\ ml/mol$ a $K_{d,atm} = 0,92\ \mu M$. Hodnota disociační konstanty zapadá do oblasti hodnot získaných jinými metodami nezávislými na interakci se substrátem či inhibitorem [60–62]. Naopak kinetické studie na těchto interakcích závislé [63–65], včetně naší starší práce [66], poskytují obecně hodnoty nižší, které mohou být způsobeny vlivem stabilizačního efektu substrátu či inhibitoru.
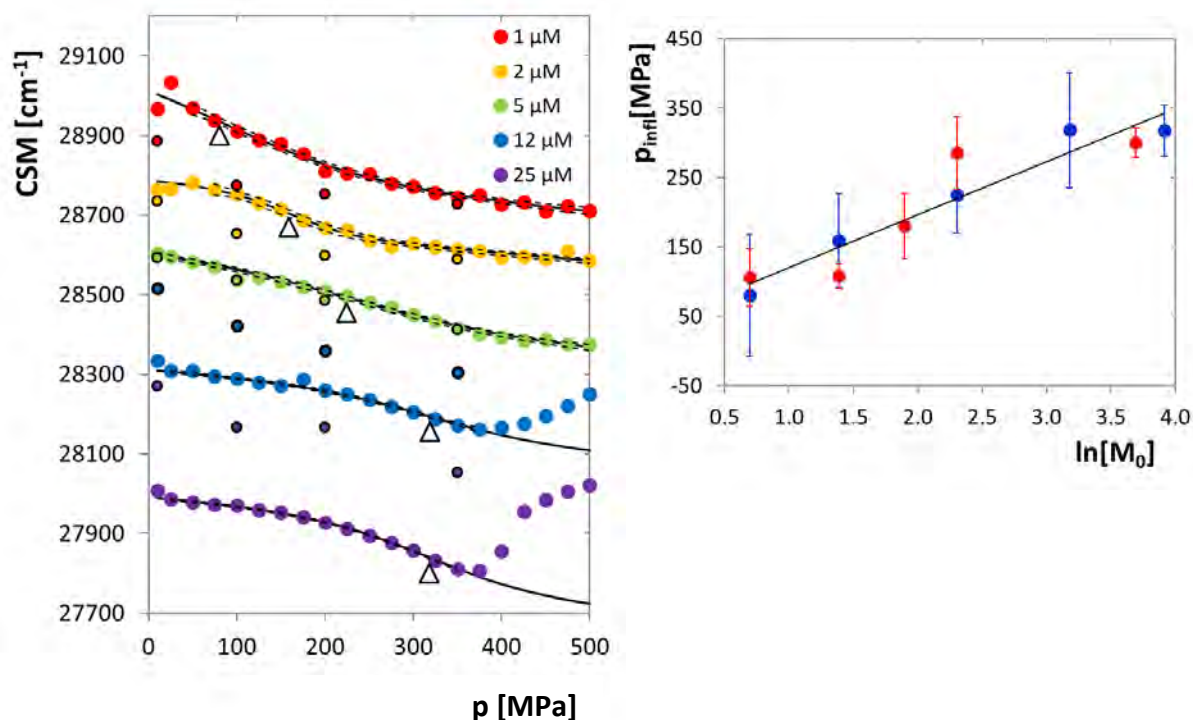
Obr. 12. Závislost inflexního bodu křivek disociace dimer-monomer na koncentraci HIV-1 PR. Vlevo: experimentálně naměřené body pro jednotlivé koncentrace proložené teoretickou závislostí (plná čára). Přerušované čáry vymezují pásy 95% spolehlivosti regresních křivek. Černě ohraničené tečky reprezentují hodnoty naměřené při zpětném snižování tlaku – tyto jsou ovlivněny předchozím působením velmi vysokého tlaku vyvolávajícího jak unfolding, tak i agregaci proteinu. Trojúhelníčky označují polohu inflexního bodu. Křivky jsou pro přehlednost vzájemně vertikálně posunuty. Vpravo: tlak v inflexním bodě jako funkce koncentrace. Body různých barev označují opakovaná měření, chybové úsečky vyznačují směrodatnou odchylku vypočtenou z chyb nelineární regrese. Plná čára vyznačuje regresní přímku proloženou všemi body dle rovnice (4). Převzato a upraveno z P5.

Denaturační křivka daná odezvou celkové intenzity fluorescence, na rozdíl od těžiště spektra, nevykazuje žádný výrazný posun s měnící se koncentrací HIV-1 PR. Proto je pravděpodobné, že tato charakteristika je spíše odezvou unfoldingu monomerů než jejich disociace. V tomto případě lze velmi snadno určit z polohy inflexního bodu objemovou změnu provázející unfolding i hodnotu rovnovážné konstanty tohoto procesu: $\Delta V_u = -104\ ml/mol$ a $K_{u,atm} = 1,5 \cdot 10^{-5}\ \mu M$. To znamená, že za teploty 25 °C, za níž měření probíhalo, je monomer za atmosférického tlaku velmi stabilní, jeho unfolding nastává až za vyšších tlaků. To je v souladu i s experimentální studií metodou NMR, v níž byly pozorovány monomery mutantních dimerizačně defektních forem HIV-1 PR [67]. Tomuto pozorování odpovídají i hodnoty rychlostních konstant foldingu a unfoldingu monomerů, které byly stanoveny na základě pozorované časové změny fluorescence. Za atmosférického tlaku je rychlostní konstanta pro unfolding o 4-5 řádů nižší než pro folding, se zvyšujícím se tlakem se však obě vyrovnávají, přičemž jejich vyrovnání nastane za tlaku, v němž má křivka tohoto přechodu

inflexní bod. Poté se vzájemný poměr rychlostních konstant obrátí a jejich hodnoty se opět vzdalují.

Nakonec byla měřena i závislost kinetických parametrů enzymového štěpení umělého peptidového substrátu. Z technických důvodů nebylo možné detailně proměřit rychlost reakce v celé oblasti koncentrací jak enzymu, tak substrátu. Proto bylo provedeno pouze srovnání hodnot pro různé tlaky za dané koncentrace obou složek s teoreticky odvozenou křivkou popisující vliv disociace dimeru na kinetiku při konstantnosti Michaelisovy konstanty K_m a čísla přeměny k_{cat}. Výsledek ukazuje, že disociace dimeru je velmi podstatným faktorem ovlivňujícím rychlost štěpení substrátu, jelikož experimentální data se modelové křivce velmi blíží, ale přesto skutečný průběh vykazuje určité odchylky. To svědčí o tlakové závislosti kinetických konstant K_m a k_{cat}, k jejichž stanovení by však bylo zapotřebí dalších experimentů. Tyto experimenty v současné době probíhají a výsledky budou publikovány v budoucnosti.

Experimentální studie HIV-1 PR za vysokého tlaku ukázala, že za teploty 25 °C je dimer proteinu velmi stabilní, jeho disociační konstanta koresponduje s jinými dříve získanými experimentálními hodnotami. Dimer stabilizovaný inhibitorem ani za vysokého tlaku nepodléhá unfoldingu, který proto může nastat jedině v případě disociace dimeru. Samotné monomery jsou za uvedené teploty za atmosférického tlaku stabilní vůči unfoldingu, jejich terciární struktura se rozpadá až za zvýšeného tlaku. Enzymová kinetika je zásadně ovlivněna disociací dimeru, ale jistý vliv na ni má i tlaková závislost samotných kinetických konstant, K_m a k_{cat}. Ten však dosud nebyl detailně prozkoumán a bude předmětem dalších měření.

2.3. *Teoretický popis rovnováh v roztocích oligomerních proteinů za vysokého tlaku*

V následující publikované práci (P6) jsme se zaměřili na teoretický popis oligomerizačních rovnováh za vysokého tlaku. Cílem práce bylo vytvořit univerzální popis těchto dějů pro různé počty podjednotek ve variantách homo- i heterooligomerů. Popis byl založen především na změnách poloh inflexních bodů tlakových závislostí stupňů konverze jednotlivých oligomerizačních stavů. Popis byl doplněn matematickými vzorci pro jednotlivé rovnováhy, které jsou sice mnohdy komplikované pro manuální počítání, ale lze je použít pro počítačové zpracování problému.

Teoretický popis byl vypracován pro několik typů rovnováh. Nejjednodušším typem je uzavřená rovnováha homooligomerního systému, tj. rovnováha mezi oligomerem o daném počtu podjednotek a volnými podjednotkami (monomery) bez předpokladu vzniku jakýchkoli mezistupňů. Křivka přechodu má v tomto případě jen jeden inflexní bod a je popsatelná

podobnou rovnicí jako v případě homodimeru. Dále bylo zpracováno několik základních uzavřených rovnováh pro heterooligomery, konkrétně heterodimer a heterotrimery typů A_2B nebo ABC. Teoretický model byl vytvořen i pro rovnováhy postupné (konsekutivní), kdy mezi monomery a výsledným oligomerem stojí ještě několik mezistavů daných postupnou tvorbou výsledného komplexu. Popis těchto rovnováh už je relativně složitý, proto byl vytvořen jen pro homotrimer a oba typy heterotrimerů.

Tato teoretická práce tedy přináší přehledné zpracování popisu rovnováh v různých typech oligomerních proteinových komplexů podléhajících disociaci vlivem vysokého tlaku. Jejím cílem je usnadnit vyhodnocení případných dalších experimentů s oligomerními proteiny různé komplexnosti.

Použití vysokého tlaku představuje jiný, méně obvyklý přístup ke studiu rovnováh i kinetiky procesů probíhajících na proteinových molekulách, jimiž jsou zejména unfolding, disociace podjednotek, agregace, enzymová kinetika, případně interakce s ligandy. Vysoký tlak představuje nezávislou termodynamickou proměnnou umožňující posun rovnováh jednotlivých dějů do koncentračních oblastí, kde je snazší jejich experimentální detekce. To umožňuje zkoumání řady jevů a měření mnoha veličin, které by za atmosférického tlaku nebylo proveditelné.

2.4. *Další plány a perspektivy*

I přes jistou logistickou komplikaci danou nutností měření na zahraničním pracovišti je tento projekt stále aktivní, byť v současné době není tolik ve středu našeho výzkumného zájmu jako projekt předchozí. Přesto stále pracujeme na výzkumu vlivu vysokého tlaku na strukturu a enzymovou aktivitu HIV-1 PR. Plánujeme zejména srovnání přírodní varianty enzymu s jeho některými mutantními variantami, jakož i s kovalentně spojeným dimerem, který nepodléhá disociaci. Dalším možným směrem výzkum chování HIV-1 PR v makromolekulárně zahuštěném prostředí (macromolecular crowding) simulujícím reálné prostředí, v němž se molekuly proteinů nacházejí.

3. Kinetika enzymově katalyzovaných reakcí

Publikované práce

P7   M. Ingr, T. Uhlíková, K. Strísovský, E. Majerová, J. Konvalinka, Kinetics of the dimerization of retroviral proteases: the "fireman's grip" and dimerization, Protein Sci. 12 (2003) 2173–2182. [66]

P8   V. Hýsková, K. Bělonožníková, I. Šmeringaiová, D. Kavan, M. Ingr, H. Ryšlavá, How is the activity of shikimate dehydrogenase from the root of Petroselinum crispum (parsley) regulated and which side reactions are catalyzed?, Phytochemistry. 190 (2021) 112881. [68]

P9   A. Tichá, S. Stanchev, J. Škerle, J. Begán, M. Ingr, K. Švehlová, L. Polovinkin, M. Růžička, L. Bednárová, R. Hadravová, E. Poláchová, P. Rampírová, J. Březinová, V. Kašička, P. Majer, K. Strisovsky, Sensitive Versatile Fluorogenic Transmembrane Peptide Substrates for Rhomboid Intramembrane Proteases, J. Biol. Chem. 292 (2017) 2703–2713. [69]

P10   H.E. Hoffman, J. Jirásková, M. Ingr, M. Zvelebil, J. Konvalinka, Recombinant human serine racemase: enzymologic characterization and comparison with its mouse ortholog, Protein Expr. Purif. 63 (2009) 62–67. [70]

Téma enzymové kinetiky částečně souvisí s výše diskutovaným tématem studia proteinů za vysokého tlaku, věnoval jsem se mu však již dříve a také v rámci jiných projektů, vzájemně velmi odlišných, přičemž všechny představovaly odborně relevantní témata zajímavá z mnoha různých úhlů pohledu. Jednalo se nejdříve o výzkum enzymové kinetiky a disociační rovnováhy dimerních retrovirových proteinas, později jsem se však podílel i na výzkumu aktivity serinové racemasy. Velmi zajímavým tématem byl také výzkum rhomboidových proteinas, což jsou enzymy membránové, takže popis jejich enzymové aktivity musí brát v potaz jejich rozmístění ve dvojrozměrném prostoru biomembrán. Konečně v rámci poslední práce na toto téma jsem se podílel na identifikaci mechanismu dvojsubstrátové zvratné reakce katalyzované šikimátdehydrogenasou, rostlinným enzymem připraveným z kořene petržele.

3.1. Kinetika dimerizace retrovirových proteinas a její ovlivnění strukturním motivem „fireman's grip".

Moje první práce z této oblasti (P7) souvisí se studiem kinetiky a dimerizační rovnováhy retrovirových proteinas, avšak na rozdíl od P5 klasickými nízkotlakými metodami. Retrovirové

proteinasy jsou dimerní proteiny, přičemž právě dimerizace je nezbytná pro jejich enzymovou aktivitu [71]. Stanovení disociační konstanty dimeru bylo v minulosti prováděno mnoha výzkumnými skupinami s relativně nejednotnými výsledky závislými na použité metodice a konkrétních podmínkách experimentu. Kromě toho bylo také zjištěno, že enzymová aktivita retrovirových proteinas závisí na tom, jestli vedle kyseliny asparagové v aktivním místě enzymu se nachází aminokyselina threonin nebo serin, což jsou jediné dvě varianty v retrovirech nalezené. Tato aminokyselina tvoří spolu s leucinem předcházejícím aspartát aktivního místa a analogickými aminokyselinami druhého monomeru strukturní motiv zvaný „fireman's grip" (hasičský chvat – název je vychází z podobnosti se stoličkou ze čtyř rukou používanou k záchraně raněného), který propojuje oba monomery sítí vodíkových vazeb, čímž přispívá ke stabilizaci dimeru. Mutační studie provedené v obou směrech prokázaly, že varianty s threoninem jsou vždy aktivnější [72–75].

V naší práci jsme si proto kladli za cíl vysvětlit vliv této aminokyseliny na stabilitu dimeru retrovirových proteinas. Vyšli jsme tedy ze dvou přirozeně se vyskytujících (wild-type) proteinas z retrovirů HIV-1 a MAV (myeloblastosis-associated virus), z nichž první obsahuje ve zmíněném motivu threonin a druhý serin, a k oběma připravili metodami rekombinantní exprese a purifikace i opačné varianty. Kromě toho jsme ještě použili přirozenou variantu viru MPMV obsahující ve fireman's gripu rovněž threonin. Pro všechny tyto enzymy jsme stanovili rovnovážnou konstantu disociace dimeru, stejně jako jednotlivé rychlostní konstanty procesů asociace a disociace, na základě měření rychlostí štěpení fluorescenčního substrátu daným enzymem. Měření probíhalo tak, že koncentrovaný zásobní roztok proteinasy byl zředěn pufrem vhodným pro průběh enzymové reakce a byl inkubován za teplotního optima daného enzymu. V určitých časech byly z tohoto roztoku odebírány vzorky, k nimž byl v kyvetě fluorimetru přidán fluorogenní peptidový substrát, který nesl na jednom konci fluorofor (donor) a na druhém zhášeč fluorescence (akceptor) poskytující Försterův rezonanční přenos energie (FRET). Aktivita enzymu byla tedy monitorována rychlostí růstu fluorescence donoru. Z naměřených dat byla vypočtena jak rovnovážná konstanta disociace dimeru, tak i rychlostní konstanty jeho disociace i asociace. Z naměřených hodnot bylo patrné, že všechny enzymové varianty obsahující threonin v motivu fireman's gripu měly disociační konstantu v průměru o řád nižší než varianty obsahující serin, byly tedy termodynamicky stabilnější. Příčina tohoto jevu není zcela jednoznačně objasněna, ale lze ji anticipovat z okolnosti, že u termodynamicky stabilnějších T-variant je zjevně vyšší rychlostní konstanta asociace dimeru, zatímco rychlostní konstanty disociace se od S-variant příliš neliší. Je tedy pravděpodobné, že struktura fireman's

gripu má význam pro počáteční stadium asociace monomerů a že methylové skupiny threoninu tuto strukturu stabilizují lépe než postranní řetězce serinu.

Naše závěry tedy přispívají k poznání vztahu mezi strukturou a funkcí retrovirových proteinas. Navíc podporují hypotézu, že volba mezi serinem a threoninem je evolučně podmíněna a souvisí s regulací aktivity příslušné proteinasy ve virové částici, protože pro infektivitu viru je škodlivá aktivita jak příliš nízká, tak příliš vysoká [71,76].

3.2. *Mechanismus dvojsubstrátové zvratné reakce katalyzované šikimátdehydrogenasou z kořene petržele*

V mé dosud poslední práci z oblasti enzymové kinetiky (P8) jsme se s kolegy z katedry biochemie PřF UK zabývali mechanismem syntézy kyseliny šikimové enzymem šikimátdehydrogenasou. Ta je klíčovým enzymem šikimátové dráhy, která je významnou cestou syntézy aromatických biologicky aktivních sloučenin. Jelikož se však tato vyskytuje jenom u bakterií, hub, rostlin a parazitů podkmene výtrusovců, ale ne u živočichů, mohou inhibitory této dráhy být používány jako antimikrobiální či herbicidní látky [77,78]. Cílem této práce proto bylo popsat reakční mechanismus reakce katalyzované šikimátdehydrogenasou z petržele a identifikovat její přírodní inhibitory.

Jedná se o dvojsubstrátovou reakci

$$NADPH\ (A) + H^+ + DHS\ (B) \rightleftarrows SA\ (P) + NADP^+\ (Q), \tag{5}$$
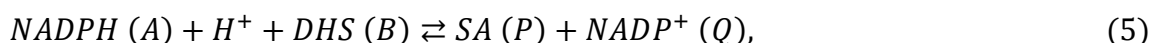
kde $NADP^+$ je oxidovaná forma koenzymu nikotinamidadenindinukleotidfosfátu, $NADPH$ jeho forma redukovaná, SA je kyselina šikimová (šikimát) a DHS její dehydrogenovaná forma (dehydrošikimát), písmena v závorkách jsou uvedena jako zjednodušující zkratky. Rovnováha reakce není výrazně posunuta ani jedním směrem, takže se tato chová jako reakce zvratná. Reakce tohoto typu může obecně probíhat mechanismy několika druhů. Typickým příkladem je mechanismus uspořádaný (sekvenční), kdy se na enzym naváže postupně nejdříve substrát A, po něm substrát B a pak teprve proběhne reakce. Jinou možností je mechanismus typu ping-pong, kdy se nejdříve naváže substrát A, je přeměněn na produkt za současné modifikace enzymu, a na takto modifikovaný enzym se naváže substrát B, jenž je následně přeměněn na druhý produkt, čímž se enzym přemění zpět do výchozího stavu. Je však známa celá řada dalších mechanismů, které se u dvojsubstrátových reakcí mohou vyskytovat [79]. V případě zvratné reakce pak musíme obecně předpokládat různost průběhu reakce z jedné a druhé strany.

K identifikaci mechanismu lze využít stanovení typu inhibicí jednotlivými reakčními produkty, a to v obou směrech. To znamená, že např. k reakci, jíž se účastní substráty A a B, přidáme jako inhibitor produkt P a tento pokus opakujeme s postupně se měnícími koncentracemi obou substrátů i inhibitoru. Přídavek produktu obecně plní roli inhibitoru, je však třeba stanovit, zda se jedná o inhibici kompetitivní či nekompetitivní, případně akompetitivní. Pokud se jedná o inhibici kompetitivní, znamená to, že produkt se navazuje na stejnou chemickou entitu (molekulu enzymu nebo jeho komplex s druhým substrátem) jako sledovaný substrát, v opačném případě se váže jiným způsobem. S využitím stanovení mechanismu inhibicí pro všechny kombinace substrátů a inhibitorů jsme poté schopni získat cenné informace o reakčním mechanismu.

V případě našeho enzymu šikimátdehydrogenasy jsme získali výsledky shrnuté v tabulce 1.

Tabulka 1. Typy inhibice pro reakce jednotlivých substrátů inhibované jednotlivými produkty. C – kompetitivní inhibice, NC – nekompetitivní inhibice.

Inhibitor	Q (NADP)		P (SA)		A (NADPH)		B (DHS)	
Substrát s proměnnou koncentrací	NADPH	DHS	NADPH	DHS	NADP	SA	NADP	SA
Typ inhibice	C	NC	C	NC	C	NC	NC	C

Toto schéma není typické pro žádný popsaný mechanismus, především proto, že vůči nukleotidovému substrátu NADPH (A) působí jako kompetitivní inhibitory oba produkty P (šikimát) a Q (NADP), což znamená, že se na volný enzym mohou navázat oba tyto produkty, čímž dojde k zamezení vazby substrátu A. Pokud naopak považujeme za substráty látky P a Q, pak proti Q působí jako kompetitivní inhibitor látka A a proti P látka B. Naopak A není kompetitivním inhibitorem proti P, z čehož plyne, že sice P se může navázat na volný enzym, ale tato vazba je tzv. mrtvým koncem (dead end) reakčního mechanismu, tedy stavem, který nevede k reakci a pro další aktivitu enzymové molekuly musí dojít k jeho zpětnému rozpadu. Proti substrátu B působí všechny produkty pouze jako nekompetitivní inhibitory, což je obvyklé i pro uspořádaný mechanismus. Na druhou stranu však proti látce P v roli substrátu působí jako kompetitivní inhibitor látka B, což znamená, že se tato může navázat i na komplex EQ. Takto může vzniknout komplex EQB, který je ale nutně zase mrtvým koncem mechanismu. S využitím poznatku o významné inhibici substrátem B z mechanismu dále vyplývá, že substrát B se může vázat i na volný enzym, čímž však brání v navázání kosubstrátu A, ale nikoli

produktu Q, což vede k vytvoření mrtvého konce EQB opačným pořadím navazování. I komplex EB lze tedy rovněž považovat za mrtvý konec.
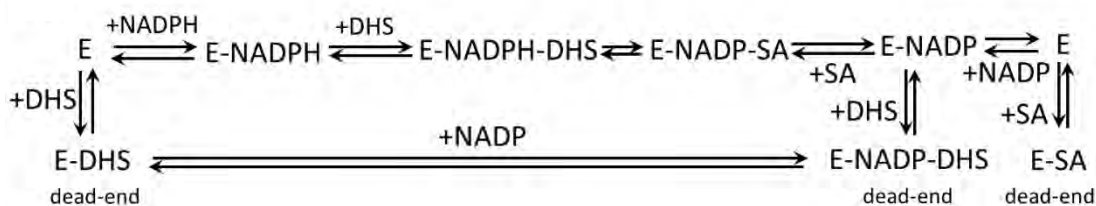
Obr. 13. Reakční mechanismus oxidace kyseliny šikimové šikimátdehydrogenasou. Převzato a upraveno z P8.

Z uvedených argumentů vyplývá reakční mechanismus uvedený na obr. 13. Jde o uspořádaný mechanismus v obou směrech, avšak s tím že v obou směrech mohou vznikat i komplexy enzymu a substrátů, z nichž reakce nemůže produktivně pokračovat, takže musí dojít k jejich zpětnému rozpadu. Tyto jsou celkem tři a tvoří mrtvé konce reakčního mechanismu.

V rámci práce byla dále měřena inhibice šikimátdehydrogenasy aromatickými látkami, které se mohou vyskytovat v rostlinné tkáni a sloužit jako regulátory její aktivity. Bylo zjištěno, že inhibiční efekt obecně roste s počtem hydroxylových skupin na těchto molekulách přítomných. Ze skupiny měřených látek byly nejlepšími inhibitory kyselina tříslová, kyselina kávová a kyselina chlorogenová.

Naše práce tedy přispívá k poznání mechanismu reakce katalyzované enzymem, který může být potenciálním cílem fungicidních a antimikrobiálních přípravků, uvedené výsledky proto mohou mít i nemalý technologický potenciál.

3.3. *Návrh a příprava fluorescenčních substrátů pro měření enzymové kinetiky rhomboidových proteinas*

Rhomboidové proteinasy (z angl. rhomboid proteinases), zkráceně též rhomboidy, jsou početnou skupinou serinových transmembránových proteinas vyskytujících se u téměř všech organismů. Od většiny ostatních proteolytických enzymů se liší tím, že jejich aktivní místo je lokalizováno v transmembránové doméně, nachází se tedy uvnitř biomembrány, na níž je enzym ukotven. Rhomboidy hrají roli především v různých signálních drahách, proto jsou často využívány jako cíle terapeutického zásahu [80–82]. Jelikož byly objeveny až v roce 2001, poznatky o nich nejsou dosud tak rozsáhlé jako v případě jiných skupin enzymů. Jejich výzkumem se zabývá mj. skupina Dr. Kvida Stříšovského na ÚOCHB AVČR, kde vznikla také práce P9, na níž jsem měl možnost se podílet.

Na rozdíl od jiných proteinas nebyly dosud pro výzkum rhomboidů navrženy vhodné nízkomolekulární peptidové substráty poskytující odezvu měřitelnou jednoduchými spektrálními metodami, tj. spektrofotometricky nebo fluorimetricky. Návrh takových substrátů byl proveden v rámci tohoto projektu na základě transmembránové sekvence běžně používaného substrátu LacYTM2. Z něj byla převzata sekvence 36 aminokyselin, přičemž štěpicí místo bylo za 11. aminokyselinou od N-konce peptidu. Na tento peptid byly potom umístěny zbytky molekul poskytujících Försterův rezonanční přenos energie (FRET), a to tak, že do pozice P5 (tj. 5. aminokyselina od štěpicího místa směrem k N-konci) byl umístěn fluorescenční donor EDANS (5-((2-aminoethyl)amino)naftalene-1-sulfonová kyselina) a do pozice P4' (4. směrem k C-konci) zhášeč fluorescence DABCYL (4-((4-(dimethylamino)fenyl)azo)benzoová kyselina). Byly změřeny kinetické parametry štěpení tohoto substrátu vybranými rhomboidy. Tato měření prokázala, že aktivita všech použitých enzymů byla dostatečná na to, aby navržený substrát byl vhodný pro kinetická měření enzymové aktivity rhomboidů. Vzhledem k tomu, že předpokládaná další kinetická měření mohou obsahovat i další látky, např. inhibitory, které by případně mohly interferovat se spektrální odezvou použitého fluoroforu (EDANS), byla zkonstruována též varianta substrátu se spektrem posunutým k vyšším vlnovým délkám. V této variantě byl použit fluorofor TAMRA (tetramethylrhodamin) a zhášeč QXL610. Zatímco pro EDANS leží excitační vlnová délka v UV oblasti spektra (335 nm) a emisní je a 495 nm, odpovídající hodnoty pro TAMRA jsou 553 nm, resp. 583 nm, což vylučuje rušení absorpcí např. isokumarinem, používaným jako inhibitor. I tento substrát byl kineticky charakterizován a vykazoval dobré vlastnosti pro další výzkumné použití.

Na základě mutačních studií štěpicího místa byl dále připraven substrát obdobný předchozí variantě, tedy také s fluorescenčními značkami TAMRA-QXL610, u něhož však byla změněna sekvence aminokyselin v místech P5-P1 z původní sekvence HISKS na RVRHA. Vůči tomuto novému substrátu byly testované rhomboidy více než 50-krát aktivnější a tedy i vhodnější pro kinetická měření.

V rámci této práce byla studována i závislost reakční rychlosti na koncentraci amfifilní látky tvořící membránové částice (lipozomy), na nichž reakce probíhají (zde DDM, n-dodecyl-β-D-maltopyranosid). Bylo totiž pozorováno, že s rostoucí koncentrací této látky v oblasti nad kritickou micelární koncentrací (CMC) enzymová aktivita klesá. Toto se ukázalo být důsledkem rozdělení molekul enzymu a substrátu na různé lipozomy, k čemuž může dojít zejména v případě nízkých koncentrací obou látek. Tyto látky jsou navíc připraveny

v zásobních roztocích rovněž na lipozomech, proto po jejich smíchání v reakční směsi musí nejdříve dojít k přechodu jejich molekul mezi lipozomy, aby byly molekuly enzymu i substrátu přítomny na společných lipozomech. Pokud je koncentrace lipozomů větší než koncentrace enzymu i substrátu, je počet lipozomů obsazených oběma partnery dán vztahem

$$n_{ESM} = \frac{n_E \cdot n_S}{n_L},$$
(6)

kde n_E, n_S a n_L jsou po řadě počty molekul enzymu, molekul substrátu a lipozomů v reakční směsi. Odtud plyne, že rychlost reakce by měla být nepřímo úměrná koncentraci lipozomů, tedy i koncentraci detergentu DDM v reakční směsi, tj. $v \sim [DDM]^{-1}$. Toto bylo skutečně pozorováno v oblasti vyšších koncentrací DDM (kolem 100 mM), zatímco se snižováním koncentrace se snižuje i absolutní hodnota exponentu závislosti – za koncentrace v řádu jednotek mM platilo $v \sim [DDM]^{-0,4}$. V této oblasti totiž už přestává platit předpoklad o nadbytku lipozomů nad biomolekulami, což vede ke zmírnění uvažované závislosti.

Uvedená zjištění vedou k následujícím doporučením pro měření enzymové kinetiky rhomboidů a membránových proteinas obecně. Zaprvé je nezbytné, aby pro srovnatelnost výsledků různých experimentů bylo vždy pracováno za konstantní koncentrace detergentu vytvářejícího lipozomy. Zadruhé, nejvhodnější substráty pro kinetická měření jsou takové, které lze skladovat v zásobních roztocích bez detergentů nebo s jejich velmi nízkým obsahem – tím se zmírní neurčitost zanesená do systému redistribucí molekul mezi lipozomy. Právě z tohoto druhého důvodu jsou velmi vhodné krátké peptidové substráty připravené v rámci této práce, neboť tyto lze skladovat v bezvodém dimethylsulfoxidu bez přítomnosti jakýchkoli amfifilních látek. To ještě zvyšuje jejich užitečnost pro výzkum rhomboidových proteinas.

3.4. *Enzymologická charakterizace rekombinantní lidské a myší serinové racemasy*

Serinová racemasa (SR) je enzym katalyzující přeměnu aminokyseliny L-serinu na D-serin [83,84]. Vzhledem k identické termodynamické stabilitě obou izomerů v achirálním prostředí může reakce snadno běžet oběma směry. Zajímavý je však především směr od L-serinu k D-serinu, protože představuje biosyntézu této neproteinogenní D-aminokyseliny, která hraje roli v řadě fyziologicky významných procesů. Kromě katalýzy racemizace optických izomerů serinu katalyzuje SR také β-eliminaci každého z nich za vzniku opticky neaktivního pyruvátu [85,86]. Vzhledem k tomu, že D-serin je významným neurotransmiterem [87,88], je SR dlouhodobě zkoumána jako potenciální terapeutický cíl v léčbě nemocí jako mozková mrtvice nebo Alzheimerova choroba.

V práci P10, která vznikala ve skupině prof. Jana Konvalinky na ÚOCHB AVČR, bylo naším cílem srovnání kinetických vlastností lidské SR (hSR) a jejího myšího ortologu (mSR), který s ní vykazuje 89 %-ní sekvenční shodu. Kromě toho byl cílem rovněž vývoj postupu exprese a purifikace rekombinantní hSR exprimované v kultuře bakterie *Escherichia coli* (*E. coli*).

Exprese proteinu byla prováděna z genu optimalizovaného podle četnosti využití kodonů (codon usage) v daném organismu. Optimalizace byla provedena pomocí webového nástroje GeMS (http://software.kosan.com/GeMS) [89]. Takto upravený gen nebyl v době provádění studie k dispozici, proto byl syntetizován složením ze synteticky připravených oligonukleotidů metodou polymerázového cyklického skládání (polymerase cycling assembly - PCA) [90,91] založeného na použití metody PCR. Nejdříve byly zakázkově syntetizovány oligonukleotidy o co největších délkách (limitujícím faktorem jsou možnosti chemické syntézy), které se překrývají střídavě svými 5' a 3' konci. Tato směs byla potom smíchána s ostatními komponentami běžně používanými v PCR (reakční pufr, směs deoxynukleosidtrifosfátů, termostabilní DNA polymeráza), přičemž použité oligonukleotidy plnily zároveň funkci templátu i primerů. V průběhu teplotních cyklů obdobných PCR dochází k postupnému propojení oligonukleotidů a vzniku celé požadované sekvence. Jelikož však je tento produkt výrazně kontaminován nekompletními fragmenty cílové sekvence, provede se nakonec její amplifikace klasickou PCR tak, že se malá část výsledné reakční směsi z PCA použije jako templát, který je následně amplifikován koncovými primery. Jelikož sekvence genu hSR má délku 1020 párů bází (bp), byla rozdělena na dvě části zvané syntony po cca 500 bp, protože pro delší úseky je metoda PCA málo účinná. Tyto syntony pak byly spojeny klasickými metodami genového inženýrství s využitím štěpení restrikčními endonukleasami a spojení pomocí DNA ligasy. (V době této práce byla syntéza genů poměrně inovativním a u nás zřídka používaným postupem, v dnešní době se však její provádění vlastními silami už nevyplatí z důvodů snadné komerční dostupnosti.) Syntetizovaný gen byl poté vložen do plasmidu pUC19 a následně do expresního plasmidu pMPM-A4-2 s arabinosou indukovatelným promotorem *ara*BAD. Samotná exprese a purifikace byla prováděna metodami převzatými z literatury.

Následně byly stanoveny kinetické konstanty K_m a k_{cat}, definované v modelu jednosubstrátové enzymové kinetiky dle Michaelise a Mentenové, pro nově připravený enzym a jejich hodnoty byly porovnány s myším ortologem SR (mSR), připraveným již dříve. Byly sledovány čtyři reakce, a to racemizace v obou směrech a β-eliminace obou optických izomerů

serinu. Oba enzymy se shodovaly prakticky ve všech kinetických parametrech, jedinou výjimkou byla β-eliminace D-serinu, u níž byla myší varianta více než třikrát aktivnější. Vysvětlení tohoto rozdílu na strukturní bázi však nebylo možno provést, jelikož v té době nebyla známa krystalová struktura ani jednoho z enzymů. Vedle konstant K_m a k_{cat} byly pro oba enzymy stanoveny také konstanty inhibice K_i pro racemizaci L-serinu třemi kompetitivními inhibitory, malonátem, L-erythro-3-hydroxyaspartátem a glycinem. Bylo zjištěno, že všechny inhibitory inhibují oba enzymy a že jejich inhibiční konstanty jsou pro oba enzymy řádově srovnatelné. Pro hSR však jejich hodnoty byly vždy 2-4-krát nižší než pro mSR, vysvětlení tohoto rozdílu na strukturní úrovni však rovněž nebylo možné.

V rámci této studie tedy byla nalezena optimalizovaná metoda přípravy hSR s kodonově optimalizovaného synteticky připraveného genu a byly porovnány kinetické vlastnosti tohoto enzymu s jeho myším ortologem mSR. Kinetické i inhibiční vlastnosti obou enzymů byly vzájemně velmi blízké, což ospravedlnilo používání mSR jako modelu hSR a jeho použití ve vyhledávání (screeningu) inhibitorů.

3.5. *Další plány a perspektivy*

Oblast enzymové kinetiky a chemické kinetiky obecně propojuje řadu témat v této práci uvedených. Proto je pravděpodobné, že se jí přímo či nepřímo budu věnovat i nadále. Některé z připravovaných nebo čerstvě zahájených projektů, jak už bylo zmíněno, se věnují výzkumu vlastností proteinů v makromolekulárně zahuštěném prostředí (macromolecular crowding). Této oblasti se chceme věnovat jak metodami vysokotlakými, tak také konvenčními metodami výzkumu proteinů. Rovněž předpokládáme, že v některých experimentech nám bude velmi ku prospěchu nově budovaná aparatura pro manipulaci mikroskopickými objekty metodou optických pastí, kterou plánujeme využít k řadě experimentů s proteinovými i neproteinovými makromolekulami. Jelikož však jsou tyto experimenty stále ve fázi příprav, podrobněji je zde nerozvádím.

4. Výpočty elektronového rozptylu na molekulách metodami kvantové chemie

Publikované práce

P11   M. Polášek, M. Juřek, M. Ingr, P. Čársky, J. Horáček, Discrete momentum representation of the Lippmann-Schwinger equation and its application to electron-molecule scattering, Phys. Rev. A. 61 (2000) 032701. [92]

P12   M. Ingr, M. Polášek, P. Čársky, J. Horáček, Discrete momentum representation method for polar molecules: Calculation of the elastic electron scattering on the H_2O molecule, Phys. Rev. A. 62 (2000) 032703. [93]

P13   M. Ingr, H.-D. Meyer, L.S. Cederbaum, Potential energy curve of the X2Sigmau$\mathplus$resonance state of F2-computed by CAP/CI, J. Phys. B: At. Mol. Opt. Phys. 32 (1999) L547–L556. [94]

Rozptyl elektronů na molekulách a površích pevných látek a krystalech je jednou z metod studia struktury molekul. Třebaže není využívaný tak často jako rentgenová difrakce nebo NMR, patří mezi metody vhodné zejména pro studium struktury molekul absorbovaných na površích pevných látek [95], v současnosti však nabývá na významu i jako metoda studia struktury krystalů [96,97]. Elektronový rozptyl může být buď elastický, kdy rozptýlený elektron opouští místo srážky s molekulou se stejnou energií, s níž přišel, a neelastický, kdy je jeho konečná energie nižší než před srážkou. Neelastický rozptyl tedy může být využíván jako spektroskopická metoda komplementární ke spektroskopii optické, jelikož přechody mezi jednotlivými stavy se řídí odlišnými výběrovými pravidly. Tato metodika je známa pod zkratkou EELS (electron energy-loss spectroscopy) [98].

4.1.   *Výpočty účinných průřezů elastického rozptylu elektronů na molekulách v bázi rovinných vln*

V našich pracích P11 a P12 jsme se ve skupině prof. Petra Čárského na ÚFCH JH AVČR zabývali vývojem kvantověchemické metody pro výpočet účinných průřezů elastického elektronového rozptylu, tj. pravděpodobností rozptylu elektronu do různých směrů v závislosti na energii rozptylovaného elektronu. Výpočet byl založen na řešení Lipmannovy-Schwingerovy rovnice v bázi rovinných vln, které jsou vhodné pro popis stavu nevázaného elektronu. Báze těchto rovinných vln byla rozmístěna v diskrétních bodech impulsového prostoru, přičemž jejich angulární pozice byly dány směrovými vektory vrcholů a dalších významných bodů pravidelných mnohostěnů ikosaedrické symetrie (pravidelný dodekaedr a ikosaedr). Distribuce v radiální dimenzi byla dána transformovanými body Gaussových integračních kvadratur. Metoda dostala název discrete momentum representation (DMR) (v analogii s běžně používanou metodou discrete variable representation - DVR).

Jelikož jsme se zabývali rozptylem na molekulách v plynné fázi, bylo kromě samotného řešení rovnice ještě třeba provést zprůměrování hodnot účinného průřezu přes všechny prostorové orientace cílové molekuly. Protože řešením dostaneme hodnoty účinného průřezu pro přechod elektronu z jakéhokoli výchozího směru do jakéhokoli směru finálního, byla vytvořena fitovací procedura, pomocí níž byl účinný průřez proložen hladkou funkcí a tato byla následně zintegrována.

Protože podle Pauliho principu jsou vlnové funkce víceelektronového systému antisymetrické vzhledem k záměně elektronů, musí toto pravidlo platit i v systému, kde dochází k rozptylu, a to nejen pro elektrony cílové molekuly, ale pro všechny elektrony včetně elektronu rozptylovaného. Tato podmínka činí řešení úlohy komplikovaným, proto byl v našich výpočtech použit jenom nejjednodušší přístup zvaný statická výměnná aproximace (static exchange approximation), která vychází z modelu nezávislých elektronů, a je tedy obdobou Hartreeho-Fockovy metody pro nevázané stavy.

V první práci P11 byl počítán elastický rozptyl na empirickém Yukawově potenciálu a na nepolárních molekulách vodíku a methanu. Tato práce potvrdila dobrou shodu s analytickým řešením pro Yukawův potenciál a s experimentálními hodnotami účinných průřezů pro obě molekuly a tedy i její dobrou aplikovatelnost na další systémy.

Na tuto práci potom navazoval analogický výpočet na molekule vody (P12). Principiální rozdíl mezi touto molekulou a molekulami předchozími spočívá v jejím nenulovém dipólovém momentu, který působí singularitu ve výpočtu elementů matice interakčního potenciálu pro rozptyl v dopředném směru (elektron odchází ve stejném směru, v jakém přišel). Bylo však prokázáno, že vzhledem k povaze této singularity a rozmístění mřížkových bodů v impulsovém prostoru stačí nahradit integrál dopředného rozptylu nulou a dále výpočet probíhá analogicky případu nepolární molekuly. Výpočet pak prokázal dobrou shodu mezi vypočtenými a experimentálními hodnotami účinných průřezů.

4.2. *Výpočty parametrů rezonančního elektronového stavu na molekule F_2 s využitím komplexních absorbujících potenciálů*

V další práci (P13) jsem se věnoval výpočtům rezonančního elektronového rozptylu na molekule F_2. Tento výzkum probíhal ve spolupráci s Dr. Hans-Dieterem Meyerem ze skupiny Prof. Lorenze S. Cederbauma z Univerzity v Heidelbergu. Rezonancemi nazýváme kvazivázané stavy, v nichž je dopadající elektron zachycen na dobu řádově delší, než je délka kolize elektronu s molekulou v běžném nerezonančním rozptylu. Tím vzniká útvar podobný

aniontu dané molekuly, který ovšem není v čase stabilní a následně se rozpadá uvolněním elektronu. Proto pro každý rezonanční stav můžeme definovat střední dobu života, jejíž určení je cílem prováděného výpočtu společně s určením energie, za níž ke vzniku rezonančního stavu dochází. Je zjevné, že účinný průřez je pro rezonanční stav výrazně větší než pro rozptyl nerezonanční. Výpočet energie a střední doby života je možné převést na problém nalezení vlastních hodnot a vlastních funkcí Hamiltoniánu systému, jako je tomu v případě stacionárních stavů. Vzhledem k tomu, že však rezonanční stav není stacionární, v čase zaniká, nemá vlnová funkce tohoto stavu v nekonečnu nulovou limitu, naopak exponenciálně roste k nekonečnu. Energie tohoto stavu je potom třeba vyjádřit jako komplexní číslo, jehož imaginární část udává právě střední dobu života rezonance.

Pokud chceme k praktickému výpočtu využít běžných metod kvantové chemie sloužících k výpočtu vázaných stavů, musíme vlnovou funkci zbavit její divergující části. To lze provést různými metodami, z nichž nejvýznamnějšími jsou metoda Feshbachovy projekce [99,100], stabilizační metoda [101] nebo metoda komplexního škálování [102–104]. Jinou možností, která byla použita v naší práci, je aplikace komplexních absorbujících potenciálů (complex absorbing potential – CAP) [105]. Jde o potenciálové funkce působící na rozptylovaný elektron, které přičteme k základnímu potenciálu pocházejícímu od samotné cílové molekuly. V blízké oblasti kolem molekuly jsou tyto funkce nulové, neovlivňují tedy strukturu samotného rezonančního stavu, od jisté vzdálenosti pak plynule rostou k nekonečnu. Tyto funkce jsou však vynásobeny imaginární jednotkou, mají tedy ryze imaginární hodnotu, což umožňuje postupnou „absorpci", tj. zánik divergující části vlnové funkce, která potom limituje k nule jako u stavu vázaného. Pokud by potenciál byl reálný, způsoboval by odrazy vlnové funkce, a tedy vznik stojaté vlny, která by potom ovlivňovala skutečnou vlnovou funkci samotného rezonančního stavu. Komplexní absorbující potenciály jsou však empiricky volenými funkcemi a není jednoduché najít jejich vhodný tvar. Proto je nutné k němu dospět iterativně, tedy tak, že provádíme sérii výpočtů komplexní energie rezonančního stavu s postupně se měnícím CAPem a hledáme takovou jeho velikost, kolem níž se hodnota energie stabilizuje. Pokud je totiž CAP příliš malý v oblasti pokryté bázovými funkcemi použitými k výpočtu, nemá na výpočet dostatečný vliv, pokud je naopak příliš silný, ovlivňuje výpočet až příliš a mění tedy vlnovou funkci v oblasti existence rezonančního stavu.

V naší studii jsme s využitím CAP počítali metodou multireferenční konfigurační energie (MRCI) – tato metoda jako celek byla označena CAP/CI – komplexní energie aniontu F_2^- v základním stavu ($X\,\Sigma_u^+$) pro řadu mezijaderných vzdáleností v oblasti, kde se energie

tohoto iontu nachází nad energií základního stavu neutrální molekuly. V této oblasti se totiž molekulový aniont stává rezonančním stavem, poněvadž může emisí jednoho elektronu přejít do energeticky nižšího stavu neutrální molekuly. V oblasti, kde je energie aniontu nižší než energie neutrální molekuly, je aniont stabilním vázaným stavem, neboť neexistuje stav neutrální molekuly, do něhož by emisí elektronu mohl přejít. Vypočítané hodnoty šířky rezonance jako funkce mezijaderné vzdálenosti byly extrapolovány k nule, čímž byl nalezen bod křížení křivek potenciální energie základních stavů aniontu a neutrální molekuly. To umožnilo správné vertikální nastavení obou křivek, které bylo možno zkontrolovat porovnáním se známou experimentálně zjištěnou hodnotou elektronové afinity atomu fluoru. Toto srovnání poskytlo velmi dobrou shodu, která prokázala správnost použitého přístupu.

Podařilo se tedy prokázat použitelnost metody CAP/CI a konceptu komplexních absorbujících potenciálů obecně pro výpočty rezonančních stavů molekulových aniontů. Kromě toho byly získány hodnoty energií a šířek rezonančního stavu $X\,\Sigma_u^+$ aniontu F_2^- pro řadu mezijaderných vzdáleností. Tato data jsou potřebná pro následný výpočet kvantové dynamiky rezonančního stavu, a tedy i simulace procesu disociativního elektronového záchytu nebo výpočty spekter vibračně excitovaného elektronového rozptylu na molekule F_2.

5. Aplikace metody polymerázové řetězové reakce (PCR) a jejich teoretický popis

Publikované práce

P14   M. Ingr, J. Dostál, T. Majerová, Enzymological description of multitemplate PCR-Shrinking amplification bias by optimizing the polymerase-template ratio, J. Theor. Biol. 382 (2015) 178–186. [106]

P15   M. Ingr, J. Konvalinka, Theoretical description of the direct exponential amplification and sequencing (DEXAS) method, Biol. Chem. 381 (2000) 439–445. [107]

P16   H. Marušincová, L. Husárová, J. Růžička, M. Ingr, V. Navrátil, L. Buňková, M. Koutny, Polyvinyl alcohol biodegradation under denitrifying conditions, International Biodeterior. Biodegrad. 84 (2013) 21–28. [108]

P17   V. Kotrbová, D. Aimová, M. Ingr, L. Borek-Dohalská, V. Martínek, M. Stiborová, Preparation of a biologically active apo-cytochrome b5 via heterologous expression in Escherichia coli, Protein Expr. Purif. 66 (2009) 203–209. [109]

Polymerasová řetězová reakce (PCR) byla objevena v roce 1983 Kary Mullisem jako metoda mnohonásobné amplifikace vybraného úseku molekuly DNA. Její mimořádný vědecký přínos nejen že vynesl jejímu objeviteli Nobelovu cenu, ale vedl k rozšíření metody do mnoha příbuzných vědních oborů od samotného základního výzkumu v oblasti molekulární biologie, genetiky a biochemie přes metody klinické diagnostiky až po forenzní analýzy typu určování otcovství či pachatelů trestných činů [110]. Třebaže princip metody je velmi jednoduchý, vyvinula se pestrá škála jejích různých variant používaných k účelům preparativním i ke kvalitativní a kvantitativní analýze. Typické provedení metody PCR obsahuje v reakční směsi pufr umožňující aktivitu DNA polymerasy, templátovou DNA, směs čtyř deoxynukleosidtrifosfátů (dNTP) dva primery vymezující konce amplifikovaného úseku a termostabilní DNA polymerasu. Reakce probíhá ve třech cyklicky se opakujících krocích, z nichž první probíhá za vysoké teploty (94-98 °C) a dochází při něm k disociaci (nazývané též „tání") řetězců tvořících dvojšroubovicovou molekulu DNA, druhý za teploty v rozmezí cca 50-68 °C, při němž dochází k nasedání (anealingu) primerů na komplementární sekvence templátu, a třetí za teplotního optima funkce zvolené DNA polymerasy (typicky 72 °C), při němž dochází k samotné polymerizaci nových vláken DNA enzymovou replikací vláken templátu přítomného ve směsi.

5.1.   Enzymologický popis multitemplátové PCR

V klasickém provedení PCR je amplifikován jeden úsek DNA pomocí dvou primerů za vzniku jednoho PCR produktu. Existují však metody, kdy je v jedné reakční směsi amplifikováno více PCR produktů. Tyto mohou vznikat přepisem několika templátů pomocí stejné dvojice primerů – tato metoda se nazývá multitemplátová PCR, nebo pomocí více dvojic templátů – potom jde o multiplexovou DNA [111]. Tyto metody slouží k detekci a kvantifikaci více sekvencí DNA ve směsi, a to včetně sekvencí málo zastoupených. Cílem naší práce bylo teoreticky popsat průběh souběžné amplifikace více templátů s ohledem na příčiny jevů narušujících správnost stanovení. Aby totiž bylo kvantitativní srovnání všech těchto sekvencí relevantní, je nutné udržet po celou dobu reakce konstantní poměr všech stanovovaných templátů. To však může být narušováno několika jevy, zejména náhodností amplifikace

v případě velmi nízkých koncentrací minoritního templátu. Daná molekula templátu totiž v jednom konkrétním cyklu buď přepsána je, nebo není, což u koncentrovaných templátů vede jen ke konstantnímu snížení celkové účinnosti, ale např. v případě jediné templátové molekuly může dojít ke ztrátě celého jednoho cyklu a tedy zdánlivému poklesu jeho koncentrace na polovinu. Další problém může nastat v situaci, kdy již koncentrace PCR produktu, který se stává templátem dalších cyklů, dosáhne stejné nebo vyšší hodnoty než koncentrace DNA polymerasy. Potom už amplifikace probíhá nikoli exponenciálně, ale pouze lineárně, což znamená, že produktu přibývá velmi pomalu. Protože však DNA polymeráza je sdílena všemi templáty, dostanou se do tohoto stavu všechny v okamžiku, kdy do něj dospěje templát majoritní. Pokud tedy chceme stanovit templát minoritně zastoupený, musíme ho dokázat detekovat ještě v oblasti, kdy amplifikace všech templátů probíhala exponenciálně.

V rámci naší práce P14 jsme odvodili teoretický model amplifikace v multitemplátové PCR s platností také pro PCR multiplexovou. Popis je založen na modelu enzymové kinetiky jednosubstrátové reakce dle Michaelise a Mentenové, kde substrátem je komplex enzym primer, či přesněji replikované vlákno DNA [112,113]. Významným vstupním parametrem je procesivita enzymu vyjádřená formou rovnovážné konstanty mezi enzymem asociovaným se substrátem a enzymem volným. Na základě této teorie byl proveden nejen deterministický výpočet amplifikačních křivek, ale i simulace amplifikace jako náhodného procesu na jednotlivých molekulách. S jeho pomocí bylo stanoveno, jak procesivita enzymu ovlivňuje možnost přesné detekce minoritních templátů a v jakém nejmenším poměru k templátu majoritnímu lze při dané procesivitě enzymu minoritní templát detekovat. Výsledky byly následně diskutovány v kontextu různých aplikací klasické i kvantitativní [114,115] PCR.

5.2. *Metoda přímé exponenciální amplifikace a sekvenování (Direct Exponential Amplification and sequencing - DEXAS)*

Metoda DEXAS byla navržena koncem 90. let 20. století Kilgerem a Pääbem [116] jako vylepšení klasické Sangerovy metody sekvenace DNA [117]. Její podstata spočívá v oboustranné amplifikaci sekvenovaného úseku DNA metodou PCR se simultánní produkcí sekvenačních fragmentů. V reakční směsi tedy musí být přítomny dva primery, z nichž každý je značen odlišnou fluorescenční značkou. Reakční směs dále obsahuje složky obvyklé pro sekvenaci DNA, tj. vhodný reakční pufr, analyzovaný vzorek DNA, termostabilní DNA polymerasu, směs čtyř deoxinukleosidtrifosfátů (dNTP) a jeden zvolený dideoxinukleosidtrifosfát (ddNTP). Reakce je provedena paralelně čtyřikrát, vždy s jedním konkrétním ddNTP (tj. ddATP, ddCTP, ddGTP, nebo ddTTP). Výhodou metody oproti

klasickému provedení je jednak amplifikace templátu a jednak možnost souběžné analýzy sekvence z obou stran daného úseku DNA.

Cílem naší práce (P15) bylo vytvoření teoretického popisu metody DEXAS, který by umožňoval snadné nalezení některých parametrů klíčových pro optimální průběh reakce. Pro nalezení optimálního průběhu je však třeba znát také poměr afinit zvolené DNA polymerasy k oběma typům nukleotidů. Účinnost produkce fragmentů DNA pro následnou elektroforetickou analýzu totiž závisí nejen na tomto parametru, ale také na počtu amplifikačních cyklů, a především na poměru dNTP a terminátorů reakce ddNTP v reakční směsi. Proto byl odvozen teoretický model, který při znalosti poměru afinit enzymu k nukleotidům a počtu cyklů umožňuje určit optimální poměr dNTP a ddNTP. Výpočet byl dále rozšířen i na pokročilejší variantu metody DEXAS, v níž mohou být použity dva různé enzymy lišící se v afinitě k ddNTP tak, že jeden tyto terminátory reakce začleňuje do DNA velmi ochotně, zatímco druhý téměř vůbec. Z popisu vyplývá, že použití takových enzymů poskytuje větší variabilitu v přípravě reakční směsi, takže je snazší zvolit takový poměr dNTP a ddNTP, aby bylo dosaženo optimálních reakčních podmínek. Vedle toho byl posouzen také vliv účinnosti enzymové amplifikace, a tedy všech výše diskutovaných parametrů, na rovnoměrnost množství fragmentů jednotlivých délek, která je dalším omezujícím předpokladem pro volbu parametrů reakce. Následně byl vyhodnocen také vliv chybové frekvence dané DNA polymerasy na možné znečištění signálu fragmenty vzniklými v důsledku vnesených mutací. Tento vliv však byl shledán relativně málo významným. Nakonec byl navržen algoritmický postup, jak správně určit parametry PCR amplifikace v metodě DEXAS, tj. zejména poměr dNTP a ddNTP, poměr použitých DNA polymeras a počet cyklů reakce, aby bylo dosaženo optimálního výtěžku sekvenačních fragmentů.

V následujících letech však došlo k velmi bouřlivému rozvoji sekvenačních metod a zejména k rozšíření kapilárních sekvenátorů využívajících čtyřbarevnou fluorescenční detekci fragmentů značených nikoli na primerech, ale reakčních terminátorech [118]. Kromě toho došlo k prudkému rozvoji sekvenačních metod „nové generace" (Next-Generation sequencing, NGS) [119,120] vhodných pro rozsáhlé sekvenace celých genomů a metagenomů. Všechny tyto metoda zefektivnily sekvenování do té míry, že výhody přinášené metodou DEXAS se staly nevýznamnými. Proto se tato metoda dále nerozvíjela a nedošlo ani k širšímu využití výsledků této práce. Přesto byl DEXAS ve své době zajímavou myšlenkou a teoretické zpracování jeho popisu se jevilo jako smysluplný příspěvek k rozvoji sekvenačních metod.

5.3. Degradace polyvinylalkoholu za denitrifikujících podmínek

Polyvinylalkohol (PVA) je ve vodě rozpustný polymer s širokým využitím ve výrobcích běžné denní potřeby, v nichž je využíván jako zahušťovadlo, emulgátor, látka tvořící filmy či adhezivum [121]. Vzhledem k jeho roční produkci přesahující 1 Mt dochází k úniku nemalého množství tohoto polymeru do životního prostředí, zejména v odpadních vodách. Naopak příznivou skutečností je, že PVA je biodegradabilní prostřednictvím některých druhů bakterií a hub [122]. Biodegradační procesy jsou vždy založeny na enzymatické oxidaci PVA, v níž je zpravidla konečným příjemcem elektronů molekula kyslíku. Tyto procesy tedy probíhají v aerobním prostředí, proto umožňují degradaci PVA např. ve vodách obsahujících rozpuštěný kyslík. Naopak v půdách, kde je kyslíku nedostatek, je možnost degradace PVA těmito mikroorganismy značně omezená [123,124]. Některé předchozí práce však ukazovaly na možnost anaerobní degradace PVA, v nichž by konečným oxidačním činidlem byla jiná látka, přičemž jedním z hlavních kandidátů jsou dusičnanové anionty [125–127].

Naše práce P16, vznikající v laboratoři prof. Marka Koutného na FT UTB ve Zlíně, byla proto zaměřena na prokázání možnosti této degradace v denitrifikačních podmínkách (oxidačním činidlem jsou anionty NO_3^- redukující se na molekulární dusík N_2) prostřednictvím bakteriálního společenstva z čistírny odpadních vod (ČOV – vzorky byly odebrány v ČOV ve Zlíně-Malenovicích) a identifikovat případně nalezený bakteriální druh, který tuto degradaci způsobuje.

Bakteriální kulturou odebranou z ČOV bylo zaočkováno minerální médium, v němž jedinou látkou obsahující uhlík byl právě PVA, a tato kultura byla kultivována za aerobních i anaerobních podmínek v přítomnosti i bez přítomnosti dusičnanů. Experimenty prokázaly, že k degradaci PVA dochází v obojích podmínkách, přičemž však v anaerobních je degradace výrazně urychlena přítomností dusičnanů v médiu. To ukazuje, že PVA je oxidován právě na úkor redukce dusičnanových aniontů. Následně jsme provedli sérii experimentů s cílem najít bakteriální druh, který degradaci PVA v denitrifikačních podmínkách způsobuje. Nejdříve byla provedena analýza metodou TGGE (temperature-gradient gel electrophoresis – gelová elektroforéza v teplotním gradientu) [128]. Z PVA degradující bakteriální kultury byla metodou PCR amplifikována hypervariabilní sekvence genu 16S rRNA, který bývá často používán pro identifikaci bakterií, jelikož je ve všech bakteriích univerzálně přítomen, ale jeho sekvence je výrazně druhově závislá. Na konec sekvence byla připojena CG-svorka, tedy sekvence o délce 40 párů bází tvořených pouze guaninem (G) a cytosinem (C), která za podmínek elektroforézy (teplota 35-50 °C, 8M močovina) nedisociuje na jednovláknové

molekuly DNA. Protože zbylá část DNA obsahuje podstatné zastoupení AT párů, dochází k její disociaci za některé teploty z uvedeného intervalu v závislosti na její konkrétní sekvenci. Vzhledem k tomu, že PCR produkty z jednotlivých bakterií se elektromigračně pohybují v gelu o daném teplotním rozmezí, dochází k disociaci každého z nich za jiné teploty, a tedy i v jiné vzdálenosti od startu. Jakmile k této disociaci dojde, změní se dsDNA na molekulu tvořenou, kromě GC svorky, ssDNA, která má v gelu výrazně nižší mobilitu, a proto se prakticky zastaví. Rozdíly mezi jednotlivými bakteriálními druhy tedy můžeme vyhodnotit na základě vzdálenosti, do níž daná DNA doputuje. Konkrétní bakteriální rod pak lze určit srovnáním se známými tabelovanými výsledky pro identicky prováděnou TGGE. Z našeho pokusu vyplynulo, že během kultivace v kultuře výrazně roste zastoupení bakterie rodu *Steroidobacter*, která proto byla označena za pravděpodobného původce degradace. Její přítomnost v kultuře byla následně prokázána sekvenací genu 16S rRNA amplifikovaného PCR z narostlé bakteriální kultury. Přibližně třetina všech sekvenovaných klonů obsahovala stejnou sekvenci odpovídající (na základě srovnání s databází GenBank) bakteriálnímu kmenu *Steroidobacter* ZUMI 37. Přímý vztah této bakterie k degradaci PVA v denitrifikačních podmínkách byl dále prokázán stanovením jejího zastoupení v bakteriální kultuře v různých dnech od začátku kultivace. Toto stanovení bylo provedeno metodou kvantitativní PCR (qPCR) [114,115], která byla provedena paralelně se dvěma dvojicemi primerů. První dvojice nasedala na konzervativní sekvence genu 16S rRNA, které jsou společné všem bakteriím, takže produktem této amplifikace bude pro jakoukoli bakterii přibližně stejně velký PCR produkt. Druhá dvojice primerů nasedala do hypervariabilní sekvence, a byla tedy specifická právě pro hledanou bakterii. Standardem byl v obou případech plazmid pUC19 nesoucí dříve zaklonovaný gen 16S rRNA bakterie *Steroidobacter* ZUMI 37, templátem byla DNA izolovaná z bakteriální kultury v různých dnech kultivace. Porovnáním množství PCR produktu amplifikací s první a druhou dvojicí primerů jsme získali zastoupení bakterie *Steroidobacter* ZUMI 37 v kultuře. Toto zastoupení bylo na začátku velmi nízké, cca 0,02 %, ale v průběhu kultivace rostlo úměrně tomu, jak ubývalo množství PVA. Na konci kultivace pak dosáhla frakce této bakterie cca 16 %. Tento výsledek jasně prokázal souvislost bakterie *Steroidobacter* ZUMI 37 s degradací PVA v denitrifikačních podmínkách, proto ji lze prohlásit za hledaný degradující mikroorganismus.

Tato studie tedy prokázala, že bakteriální degradace PVA je možná i v anaerobních podmínkách, kdy konečným oxidačním činidlem je dusičnanový aniont. Zároveň byl identifikován bakteriální kmen *Steroidobacter* ZUMI 37, který tuto degradaci způsobuje. Třebaže se tento kmen nepodařilo vypěstovat v čisté kultuře, což může být důsledkem nutné

symbiózy s jinými bakteriemi, má výsledek potenciální význam pro využití v technologiích cílících na odstraňování PVA z životního prostředí, zejména v čištění odpadních vod.

5.4. *Příprava apo-cytochromu b rekombinantní expresí v* E. coli

Cytochrom b5 (dále jen b5) je malý hemoprotein o relativní molekulové hmotnosti cca 17 kDa, který je schopen přenášet jednotlivé elektrony v enzymově katalyzovaných biochemických reakcích [129,130]. V buňkách savců je b5 přítomen ve třech isoformách vyskytujících se v endoplazmatickém retikulu, mitochondriích a cytoplazmě erytrocytů [131]. Retikulární forma se vyskytuje na membráně endoplazmatického retikula, proto bývá po destrukci buněk přítomna také v mikrozomech a bývá rovněž nazývána jako mikrosomální. Je složena z N-terminální domény obsahující místo pro vazbu hemu a menší hydrofobní C-terminální domény odpovědné za fixaci b5 na membráně [132]. Tato doména je klíčová pro interakci b5 s dalšími reakčními partnery, zejména cytochromem P450 (CYP), které jsou samy ukotveny na biomembránách. Aby mohl být b5 aktivním přenašečem elektronů v redoxních reakcích, musí obsahovat hem jako redoxní kofaktor. Takový protein se potom označuje jako holo-protein, v tomto případě holo-b5. Funkce holo-b5 spočívá ve stimulaci aktivity CYP, třebaže některé experimenty ukázaly i opačný efekt [133]. Jiné ukazují, že b5 je schopen stimulovat aktivitu CYP i jako apo-protein (apo-b5), tj. bez přítomnosti hemového kofaktoru [134–136]. Proto je pro výzkum těchto reakcí nezbytné mít k dispozici nejen holo-, ale i apo-b5 ve formě čistého proteinu. Protože v buňkách je b5 přítomen zpravidla v holo-formě, byly činěny pokusy o přípravu apo-b5 extrakcí hemu acetonem v kyselém prostředí, což ovšem vedlo k denaturaci nezanedbatelné části proteinu [137,138]. Proto jsme se v této studii P17 zaměřili na rekombinantní přípravu králičího apo-b5 heterologní expresí v bakteriích *Escherichia coli*. Práce probíhala na katedře biochemie PřF UK ve skupině prof. Marie Stiborové.

Gen proteinu byl připraven syntézou z chemicky syntetizovaných oligonukleotidů metodou polymerase cycling assembly (PCA – viz též komentář k publikaci P10) [90,91]. Jelikož je poměrně krátký, pouze 405 párů bází (bp) byl připraven celý jako jeden synton. Takto připravený gen byl předlohou pro expresi celého b5 včetně transmembránové domény. Syntetizovaný gen byl dále vložen do expresního plazmidu pET22b a byl exprimován v buňkách *E. coli* BL21(DE3) Gold. Aby byl protein exprimován v apo-formě, nebyla do kultivačního média přidána kyselina δ-aminolevulová, která je prekurzorem syntézy hemu a bez níž bude tento kofaktor vytvářen jen ve velmi malém množství. Purifikace membránového proteinu apo-b5 se lišila od běžných postupů purifikace rozpustných proteinů zejména tím, že

po ultrazvukové destrukci buněk se protein apo-b5 nacházel na fragmentech biologických membrán. Proto byl buněčný lyzát odstředěn na preparativní ultracentrifuze a membránová frakce byla shromážděna v sedimentu, který byl následně rozpuštěn v roztocích detergentů Brij 35 a cholátu sodném. Následně byly provedeny dvě chromatografie na měniči aniontů DEAE-Sepharosa v mobilních fázích lišících se koncentrací detergentů a s odlišným typem eluce – první gradientová a druhá isokratická. Preparát byl následně zbaven detergentů dialýzou a byl zahuštěn pomocí centrifugační ultrafiltrační cely Amicon.

Izolovaný protein byl dále charakterizován spektrofotometricky a kineticky aby bylo prokázáno, že jde o apo-formu a že po rekonstituci s hemem je protein biologicky aktivní. Výrazná převaha apo-formy v preparátu byla prokázána měřením absorpčního spektra v oblasti 225-600 nm. U izolovaného produktu byl pozorován výrazný pík při 280 nm charakteristický pro proteiny (způsobený absorpcí aromatických aminokyselinových zbytků) a pouze nepatrný pík při 413 nm, tedy v oblasti Soretova pásu charakteristického pro hemoproteiny. Ten odpovídal cca 1 % zastoupení holo-b5 vzniklému při expresi z hemu přirozeně syntetizovaného v bakteriích. Titrací preparátu hemin chloridem docházelo k navazování hemové skupiny na apo-b5 a tím i nárůstu píku při 413 nm až do velikosti odpovídající prakticky stoprocentní vazbě hemu na apo-formu. Tím byla prokázána schopnost rekombinantně připraveného preparátu navazovat svoji prostetickou skupinu, což je nezbytný předpoklad jeho správné biologické funkce. Dále bylo naměřeno charakteristické spektrum redukované formy hemu s maximem posunutým k 424 nm s dalšími menšími píky při 526 a 556 nm, což opět prokazovalo správnou funkci rekombinantního preparátu. Nakonec byl zkoumán vliv rekombinantního b5 na aktivaci oxidace barviva Sudan I lidským CYP3A4, a to jak v holo-, tak v apo-formě. Rekombinantní preparát v holo-formě zesiloval oxidační účinek CYP3A4 prakticky stejně jako kontrolní vzorky králičího a lidského holo-b5, což opět prokazovalo správnou funkci rekombinantně připraveného proteinu. Navíc bylo zjištěno, že i sama apo-forma zvýšila úroveň oxidace, byť významně méně než holo-b5. To je důsledkem nemožnosti apo-formy podílet se na reakci přenosem elektronu, ale potvrzuje to správnost teorie předpokládající alosterický vliv apo-b5 na funkci CYP3A4 [134–136].

V rámci této práce byl tedy syntetizován gen králičího cytochromu b5 a byl vyvinut postup rekombinantní exprese a purifikace tohoto proteinu. Byla prokázána vysoká biologická aktivita rekombinantního preparátu ve všech zkoumaných vlastnostech. To poskytlo možnost využívat tento protein k dalším výzkumům biochemických redoxních reakcí, jichž se účastní.

6. Ostatní publikované práce jinde nezařazené

6.1. Inhibice produkce bioplynu produkty degradace biomasy

P18   S. Pekařová, M. Dvořáčková, P. Stloukal, M. Ingr, J. Šerá, M. Koutny, Quantitation of the Inhibition Effect of Model Compounds Representing Plant Biomass Degradation Products on Methane Production, BioResources 12 (2017) 2421–2432. [139]

Produkce bioplynu anaerobní fermentací biomasy je ekonomicky schůdnou alternativou ke spotřebě fosilních paliv, která navíc může jako surovinu využívat odpadní produkty různých průmyslových a zemědělských procesů a tím přispívat k jejich likvidaci. Rostlinná biomasa je tvořená především celulosou, hemicelulosou a ligninem, které tvoří komplexní matrici odolnou proti biotické i abiotické degradaci [140]. Vzhledem k tomu je biomasa před samotným použitím k produkci bioplynu upravována tak, aby se co nejvíce rozvolnila její struktura. Běžnou metodou těchto úprav je metoda parní exploze [141], kdy je biomasa po stanovený čas zahřívána ve speciálním kanálu za vysokého tlaku externě dodávanou párou, aby po následném prudkém snížení tlaku na konci kanálu došlo k jejímu rozmělnění. Možným problémem tohoto postupu je vznik nízkomolekulárních vedlejších produktů, které mohou následně inhibovat proces samotné produkce methanu, jenž je klíčovou složkou bioplynu.

Práce P18, která vznikala v laboratoři prof. Marka Koutného na FT UTB ve Zlíně, se zabývala studiem vlivu některých takových látek na produkci methanu bakteriálním společenstvem pocházejícím z digestátu biomasy z bioplynového produkčního zařízení z Českých Budějovic. Jako zdroj uhlíku byly použity dva umělé substráty, celulóza a acetát. V prvním případě musel proběhnout celý proces produkce methanu jako v případě skutečné biomasy, tedy hydrolýza, acetogeneze a methanogeneze, zatímco v druhém se produkce omezila pouze na poslední fázi. Jako inhibující látky byly testovány furfural, 5-hydroxymethylfurfural, kyselina gallová a kyselina tříslová.

V případě furfuralu nebyl pozorován prakticky žádný inhibiční efekt s výjimkou vysoké koncentrace 2 g/l, která se však ve skutečném reaktoru na produkci methanu nemůže objevit. Naopak byl pozorován pozitivní vliv na produkci methanu svědčící o tom, že furfural sám může sloužit jako substrát pro fermentaci. O tom svědčí i fakt, že i za výše zmíněné vysoké koncentrace je furfural jako inhibitor nakonec inaktivován, což pouze prodlouží lagovou fázi produkce methanu, ale po ní už proces běží jako bez přítomnosti inhibující látky. Naopak 5-hydroxymethylfurfural vykazoval zjevný inhibiční efekt od koncentrace 0,2 g/l při použití

celulózy jako substrátu, v případě acetátu se inhibice posunula ke koncentraci 0,5 g/l. Snížení rychlosti produkce methanu bylo navíc stálé, inhibitor tedy nebyl postupem času bakteriální kulturou degradován, a tedy inaktivován. Z uvedeného také plyne, že kritickým krokem nebyla samotná methanogeneze, ale spíše hydrolýza celulózy nebo acetogeneze.

Fenolické sloučeniny, kyseliny gallová a tříslová, nevykazovaly žádný výrazný inhibiční, ale na druhou stranu ani stimulační, efekt na produkci methanu.

Z testovaných látek vykazoval významnější inhibiční efekt pouze 5-hydroxymethylfurfural. V koncentracích dosažitelných v bioplynovém reaktoru však ani on nepředstavuje pro proces výrazný problém. Lze proto konstatovat, že žádná ze zkoumaných látek nepředstavuje technologické riziko pro produkci methanu z biomasy po předchozí úpravě metodou parní exploze.

6.2. Ab initio *výpočty substituentových konstant pro metod QSAR*

P19   O. Exner, M. Ingr, P. Čársky, Ab initio calculations of substituent constants: A reinvestigation, Theochem-J. Mol. Struct. 397 (1997) 231–238. [142]

Konceptem QSAR (quantitative structure-activity relationship – modely kvantitativní závislosti aktivity na struktuře) se rozumí soubor pravidel a výpočetních postupů sloužících k predikci chemických a biologických vlastností sloučenin, především organických, na základě znalosti jejich struktury a veličin (deskriptorů) z ní přímo odvoditelných [143]. Je znám a rozvíjen po dobu více než 55 let a v dnešní době je tvořen širokou škálou různých výpočetních metod.

Součástí QSAR je také posouzení vlivu konkrétního substituentu na vlastnosti dané molekuly. Těmito vlastnostmi mohou být různé spektrální charakteristiky (intenzita píků v IR spektru, chemické posuny v NMR), ale také např. hodnoty disociačních konstant protolyticky aktivních skupin, případně charakteristiky rovnováhy a rychlosti konkrétních chemických reakcí. Tento vliv je do QSAR promítnut skrze odlišné hodnoty vhodně zvolených deskriptorů, v našem případě konstant charakterizujících jednotlivé dílčí vlivy, kterými substituent působí na sloučeninu, na niž je vázán. Pro kvantitativní posouzení těchto vlivů substituentů na zvolenou vlastnost y uvažované sloučeniny byla navržena rovnice [144]

$$y = y_0 + \rho_F \sigma_F + \rho_R \sigma_R + \rho_\alpha \sigma_\alpha + \rho_\chi \sigma_\chi + \varepsilon. \tag{7}$$

Zde y_0 značí hodnotu dané veličiny pro nesubstituovanou molekulu, parametry ρ jsou rozvojovými konstantami udávajícími vliv příspěvků dle jednotlivých forem působení substituentu na molekulu a konstanty σ jsou charakteristikami jednotlivých substituentů. Veličina ε zahrnuje ostatní nespecifikované vlivy. Index F označuje vliv induktivního efektu substituentu (field-inductive), R efektu rezonančního (mezomerního), α vliv polarizace molekuly v důsledku přítomnosti substituentu a χ vliv jeho elektronegativity (tento vliv je nejméně významný a také nejméně rigorózně definovatelný).

Práce P19 byla zaměřena na výpočty hodnot jednotlivých konstant *ab initio* metodami kvantové chemie na základě modelů navržených dříve Marriottem a Topsomem [145–148]. Byla realizována ve skupině prof. Petra Čárského na ÚFCH JH AVČR pod vedením prof. Otto Exnera z ÚOCHB AVČR. Vzhledem k nižším výpočetním možnostem v době vzniku práce bylo cílem jednak přepočítání dříve vypočtených konstant ve větších bázích atomových orbitalů, dále pak stanovení konstant příslušejících dosud nezkoumaným substituentům. Kromě toho byl posuzován vliv konformace na hodnotu konstanty, pokud u daného substituentu více konformací přichází v úvahu. Byly vypočteny konstanty pro 21 substituentů, pro něž již konstanty počítány byly, a pro 16 substituentů nových. Nová skupina obsahovala substituenty amidové a thioamidové, a to jednak čisté, jednak dále substituované methylovou skupinou umístěnou buď na uhlíkovém, nebo dusíkovém atomu – připojení substituentu k molekule bylo na straně opačné. Kromě toho byly počítány také substituenty, v nichž místo skupiny methylové figurovala aminoskupina, jednalo se tedy o zbytky močoviny a thiomočoviny. Tyto rozšířené amidové a thioamidové substituenty byly počítány ve dvou planárních konformacích E a Z.

Naše výpočty poskytly upřesněné hodnoty všech konstant a ukázaly, že hodnoty vypočtené se od experimentálních neliší v průměru více, než se liší jednotlivé sady experimentálních hodnot mezi sebou navzájem. Menší shoda byla nalezena jen u konstant σ_χ, které jsou však málo významné a jejich použití je obecně diskutabilní. Srovnání hodnot konstant pro E a Z konformery jednotlivých substituentů ukázalo významný vliv na hodnoty konstant, zejména na σ_F, v některých případech i σ_α. Proto lze očekávat prokazatelný vliv konformace na vlastnosti molekul. Bohužel však pro tyto výpočty nebylo k dispozici relevantní srovnání s experimentem, uváděné experimentální hodnoty jsou bližší četněji zastoupené konformaci Z. I přesto však lze konstatovat, že *ab initio* výpočty představují rozumnou cestu získání hodnot konstant i pro substituenty, pro které experimentální provedení není k dispozici nebo není snadno proveditelné.

6.3. Výpočty hyperplochy potenciální energie (HPE) radikálového kationtu $C_4H_6^+\bullet$ ab initio *metodami*

P20   V. Hrouda, P. Čársky, M. Ingr, Z. Chval, G.N. Sastry, T. Bally, The C4H6+ potential energy surface. 2. The reaction of ethylene radical cation with acetylene, J. Phys. Chem. A 102 (1998) 9297–9307. [149]

Reakce ve zkřížených molekulových paprscích jsou významnou experimentální technikou sloužící ke studiu reakční dynamiky elementárních chemických reakcí, jejíž přínos byl oceněn Nobelovou cenou za chemii roku 1986 (Herschbach, Lee). Je založena na kolizi dvou paprsků tvořených vysoce urychlenými molekulami, které v důsledku srážek reagují za vzniku různých meziproduktů a následně konečných produktů. Ty poté letí v různých směrech vzhledem ke směrům vstupujících paprsků, kde také mohou být detekovány a může být měřena jejich energie [150,151]. K detailnímu studiu vlastností vznikajících částic se používá řada spektrálních technik. Metodou umožňující analýzu iontů selektovaných podle energie je fotoelektronová-fotoiontová koincidenční spektroskopie (photoelectron-photoion coincidence spectroscopy – PEPICO) [152,153]. Použití těchto metod umožňuje mimo jiné i výzkum reakčních mechanismů reakcí radikálováchh kationtů, kterému se věnovala laboratoř prof. Zdeňka Hermana na ÚFCH JH AVČR. V návaznosti na to byly ve skupině prof. Petra Čárského, jejímž jsem byl členem, prováděny kvantověchemické výpočty kladoucí si za cíl vysvětlení vzájemných přeměn jednotlivých meziproduktů těchto reakcí. Výzkum byl prováděn ve spolupráci se skupinou prof. Thomase Ballyho z Univerzity ve Fribourgu ve Švýcarsku. V rámci tohoto projektu vznikla i naše publikace P20.

Radikálový kationt $C_4H_6^+\bullet$ vzniká jako produkt srážky molekul acetylenu a ethylenu, z nichž jedna je v ionizovaném stavu, tedy postrádá jeden elektron. K jeho vzniku dochází v experimentech ve zkřížených molekulových paprscích, vzhledem k jeho nízké stabilitě však dochází k rozkladu na různé fragmenty, které jsou následně detekovány. Mezi reaktanty a konečnými produkty však leží celá řada meziproduktů a tranzitních stavů, přes které musí komplex $C_4H_6^+\bullet$ projít a které tvoří složitou hyperplochu potenciální energie, na níž se systém v průběhu reakce pohybuje. Cílem této práce (P20) tedy bylo zmapovat hyperplochu potenciální energie (HPE) systému $C_4H_6^+\bullet$ na několika úrovních *ab initio* výpočtů a navrhnout reakční koordinátu vedoucí od reaktantů k produktům, které byly pozorovány experimentálně [154]. Byla zaměřena zejména na popis mechanismu vzniku a následné fragmentace $C_4H_6^+\bullet$,

čímž navazovala na některé předchozí studie z naší i jiných skupin popisující některé partikulární aspekty tohoto mechanismu [155–158]. Struktury ležící v minimech HPE byly navrženy intuitivně a následně optimalizovány na několika úrovních *ab initio* výpočtů, UHF (neomezená (unrestricted) Hartreeho-Fockova metoda), UMP2 (Møllerova-Plessetova poruchová teorie 2. řádu vycházející z vlnové funkce UHF) a také metodou DFT (teorie funkcionálu hustoty) využívající hybridního funkcionálu B3LYP. Následně byly nalezeny i struktury tranzitních stavů tato minima propojující. Energie ve stacionárních bodech byla následně ještě přepočítána metodou RCCSD(T) poskytující přesnější hodnoty rozdílů jednotlivých stavů.

Pomocí *ab initio* výpočtů byla nalezena reakční koordináta vedoucí od srážky ionizovaného ethylenu s acetylenem k experimentálně pozorovaným produktům, cyklopropeniovému kationtu a methylovému radikálu. Reakce začíná cykloadičním krokem za vzniku cyklobuteniového kationtu. Následně dochází k otevření tohoto cyklu za vzniku radikálového kationtu butadienu. Mechanismus jeho další přeměny není jednoznačný, proto byly navrženy dvě možné cesty vedoucí k produktům obsahujícím methylovou skupinu, obě s aktivačními bariérami kolem 45 kcal mol^{-1}. Jedna z nich vede k radikálovému kationtu methylallenu, mechanismus jeho případného rozpadu však nebyl dále studován. Druhá cesta vede k propenylkarbenovému radikálovému kationtu, jenž se následným přesmykem mění na radikálový kationt methylcyklopropanu. Ten se poté rozpadá na experimentálně pozorované produkty, cyklopropeniový kationt a methylový radikál.

Výsledek naší kvantověchemické studie se tedy shoduje s experiemntálním pozorováním, přičemž navíc propojuje reaktanty a produkty reakční koordinátou vysvětlující mechanismus studované chemické reakce.

Závěr

Tato habilitační práce je sestavena z výsledků získaných v období 1997-2021, tedy v průběhu téměř celého jednoho čtvrtstoletí. Obsahuje výsledky různých studií experimentálních i teoretických, které se vždy věnují výzkumu fyzikálních vlastností molekul, byť v rámci několika různých výzkumných směrů. Nejstarší z těchto prací se věnují vesměs molekulám izolovaným, pozdější práce se však týkají molekul v roztoku, a právě zde je vždy možné pozorovat zásadní vliv vnějšího prostředí na jejich chování. Poslední práce se proto věnují právě tomuto fenoménu na příkladu biotechnologicky atraktivní molekuly kyseliny hyaluronové. Vzhledem k značnému rozvoji výpočetní techniky v posledních desetiletích je v současné době možné provádět molekulárně-dynamické simulace relativně velkých systémů, které umožňují zkoumat nejen strukturu a konformaci hlavní studované molekuly, ale i rozložení molekul rozpouštědla, případně dalších látek v prostředí přítomných. Jde vesměs o jevy, které jsou experimentálně těžko měřitelné, proto je pro jejich pochopení přínos výpočetní chemie nezastupitelný. Výsledky zahrnuté do této práce vysvětlují, jak roztoky solí a vybraná směsná rozpouštědla ovlivňují experimentálně pozorovatelné vlastnosti molekul kyseliny hyaluronové. Tím také ukazují, že vysvětlení molekulární podstaty experimentálních pozorování na základě analýzy simulovaných systémů je perspektivní cestou zkoumání vlivu prostředí na biomakromolekuly. Proto je tento přístup hlavní náplní práce mojí výzkumné skupiny, který se svými kolegy hodlám i nadále rozvíjet.

Zkoumání vlivu externích podmínek bylo obsaženo i v dalších publikovaných pracích věnujících se vlivu vysokého tlaku na molekuly oligomerních proteinů. Vysokotlaká metodika byla úspěšně aplikována na popis rovnováhy dimerizace, unfoldingu i enzymové kinetiky HIV-1 proteinasy a naznačila možnosti aplikace na další typy systémů.

Vliv prostředí hrál zásadní roli i v dalších uvedených pracích z oblasti enzymové kinetiky, byť nebyl přímo předmětem zkoumání. V budoucnu proto plánujeme zaměřit se i experimentálně na vliv prostředí na proteinové molekuly, především na vliv makromolekulárně zahuštěného prostředí simulujícího podmínky v buněčné cytoplazmě.

Na závěr bych rád poděkoval velmi dlouhé řadě lidí, kteří mě podporovali v celém období mého dosavadního vědeckého působení, a to jak po stránce odborné, tak čistě osobní. Především chci tedy poděkovat všem mým spolupracovníkům, školitelům, konzultantům, mentorům i studentům na všech mých dosavadních pracovištích ve Zlíně, Praze, Montpellieru, Heidelbergu a Fribourgu, jakož i všem ostatním spoluautorům svých publikací, z nichž některé

Přehled literatury

[1]   M. Ingr, E. Kutálková, J. Hrnčiřík, Hyaluronan random coils in electrolyte solutions-a molecular dynamics study, Carbohydr. Polym. 170 (2017) 289–295. https://doi.org/10.1016/j.carbpol.2017.04.054.

[2]   E. Kutálková, J. Hrnčiřík, R. Witasek, M. Ingr, Effect of solvent and ions on the structure and dynamics of a hyaluronan molecule, Carbohydr. Polym. 234 (2020) 115919. https://doi.org/10.1016/j.carbpol.2020.115919.

[3]   E. Kutálková, J. Hrnčiřík, R. Witasek, M. Ingr, G. Huerta-Ángeles, M. Hermannová, V. Velebný, The rate and evenness of the substitutions on hyaluronan grafted by dodecanoic acid influenced by the mixed-solvent composition, Int. J. Biol. Macromol. 189 (2021) 826–836. https://doi.org/10.1016/j.ijbiomac.2021.08.137.

[4]   D. Jiang, J. Liang, P.W. Noble, Hyaluronan in Tissue Injury and Repair, Annu. Rev. Cell Dev. Biol. 23 (2007) 435–461. https://doi.org/10.1146/annurev.cellbio.23.090506.123337.

[5]   K.N. How, W.H. Yap, C.L.H. Lim, B.H. Goh, Z.W. Lai, Hyaluronic Acid-Mediated Drug Delivery System Targeting for Inflammatory Skin Diseases: A Mini Review, Front. Pharmacol. 11 (2020) 1105. https://doi.org/10.3389/fphar.2020.01105.

[6]   D.D. Allison, K.J. Grande-Allen, Review. Hyaluronan: a powerful tissue engineering tool, Tissue Eng. 12 (2006) 2131–2140. https://doi.org/10.1089/ten.2006.12.2131.

[7]   J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, K. Schulten, Scalable Molecular Dynamics with NAMD, J. Comput. Chem. 26 (2005) 1781–1802. https://doi.org/10.1002/jcc.20289.

[8]   W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38, 27–28. https://doi.org/10.1016/0263-7855(96)00018-5.

[9]   R. Mendichi, L. Soltés, A. Giacometti Schieroni, Evaluation of radius of gyration and intrinsic viscosity molar mass dependence and stiffness of hyaluronan, Biomacromolecules. 4 (2003) 1805–1810. https://doi.org/10.1021/bm0342178.

[10]   K. Hayashi, K. Tsutsumi, F. Nakajima, T. Norisuye, A. Teramoto, Chain-stiffness and excluded-volume effects in solutions of sodium hyaluronate at high ionic strength, Macromolecules. 28 (1995) 3824–3830. https://doi.org/10.1021/ma00115a012.

[11]   E. Fouissac, M. Milas, M. Rinaudo, R. Borsali, Influence of the ionic strength on the dimensions of sodium hyaluronate, Macromolecules. 25 (1992) 5613–5617. https://doi.org/10.1021/ma00047a009.

[12]   E. Buhler, F. Boué, Chain Persistence Length and Structure in Hyaluronan Solutions: Ionic Strength Dependence for a Model Semirigid Polyelectrolyte, Macromolecules. 37 (2004) 1600–1610. https://doi.org/10.1021/ma0215520.

[13]   G.A. Sorci, W.F. Reed, Effect of Valence and Chemical Species of Added Electrolyte on Polyelectrolyte Conformations and Interactions, Macromolecules. 37 (2004) 554–565. https://doi.org/10.1021/ma035551z.

[14]   G. Huerta-Ángeles, M. Brandejsová, K. Kopecká, F. Ondreáš, T. Medek, O. Židek, J. Kulhánek, H. Vagnerová, V. Velebný, Synthesis and Physicochemical Characterization of Undecylenic Acid Grafted to Hyaluronan for Encapsulation of Antioxidants and Chemical Crosslinking, Polymers. 12 (2020) 35. https://doi.org/10.3390/polym12010035.

[15]   O. Štrympl, J. Vohlídal, M. Hermannová, M. Maldonado-Domínguez, M. Brandejsová, K. Kopecká, V. Velebný, G. Huerta-Ángeles, Oleate-modified hyaluronan: Controlling the number and distribution of side chains by varying the reaction conditions, Carbohydr. Polym. 267 (2021) 118197. https://doi.org/10.1016/j.carbpol.2021.118197.

[16]   J.E. Scott, F. Heatley, Hyaluronan forms specific stable tertiary structures in aqueous solution: A 13C NMR study, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 4850–4855. https://doi.org/10.1073/pnas.96.9.4850.

[17]   J.E. Scott, F. Heatley, Biological Properties of Hyaluronan in Aqueous Solution Are Controlled and Sequestered by Reversible Tertiary Structures, Defined by NMR Spectroscopy, Biomacromolecules. 3 (2002) 547–553. https://doi.org/10.1021/bm010170j.

[18] E. Fischer, P.T. Callaghan, F. Heatley, J.E. Scott, Shear flow affects secondary and tertiary structures in hyaluronan solution as shown by rheo-NMR, J. Mol. Struct. 602–603 (2002) 303–311. https://doi.org/10.1016/S0022-2860(01)00733-5.

[19] P. Gribbon, B.C. Heng, T.E. Hardingham, The analysis of intermolecular interactions in concentrated hyaluronan solutions suggest no evidence for chain-chain association., Biochem. J. 350 (2000) 329–335.

[20] C.D. Blundell, P.L. DeAngelis, A. Almond, Hyaluronan: the absence of amide–carboxylate hydrogen bonds and the chain conformation in aqueous solution are incompatible with stable secondary and tertiary structure models, Biochem. J. 396 (2006) 487–498. https://doi.org/10.1042/BJ20060085.

[21] N.R. Yaffe, A. Almond, E.W. Blanch, A New Route to Carbohydrate Secondary and Tertiary Structure Using Raman Spectroscopy and Raman Optical Activity, J. Am. Chem. Soc. 132 (2010) 10654–10655. https://doi.org/10.1021/ja104077n.

[22] W.M. Payne, D. Svechkarev, A. Kyrychenko, A.M. Mohs, The role of hydrophobic modification on hyaluronic acid dynamics and self-assembly, Carbohydr. Polym. 182 (2018) 132–141. https://doi.org/10.1016/j.carbpol.2017.10.054.

[23] D. Svechkarev, A. Kyrychenko, W.M. Payne, A.M. Mohs, Probing the self-assembly dynamics and internal structure of amphiphilic hyaluronic acid conjugates by fluorescence spectroscopy and molecular dynamics simulations, Soft Matter. 14 (2018) 4762–4771. https://doi.org/10.1039/c8sm00908b.

[24] E. Kutálková, J. Hrnčiřík, M. Ingr, Pressure induced structural changes and dimer destabilization of HIV-1 protease studied by molecular dynamics simulations, Phys. Chem. Chem. Phys. 16 (2014) 25906–25915. https://doi.org/10.1039/c4cp03676j.

[25] M. Ingr, R. Lange, V. Halabalova, A. Yehya, J. Hrncirik, D. Chevalier-Lucia, L. Palmade, C. Blayo, J. Konvalinka, E. Dumay, Inhibitor and Substrate Binding Induced Stability of HIV-1 Protease against Sequential Dissociation and Unfolding Revealed by High Pressure Spectroscopy and Kinetics, Plos One. 10 (2015) e0119099. https://doi.org/10.1371/journal.pone.0119099.

[26] M. Ingr, E. Kutálková, J. Hrnčiřík, R. Lange, Equilibria of oligomeric proteins under high pressure - A theoretical description, J. Theor. Biol. 411 (2016) 16–26. https://doi.org/10.1016/j.jtbi.2016.10.001.

[27] U.R. Shrestha, D. Bhowmik, J.R.D. Copley, M. Tyagi, J.B. Leão, X.-Q. Chu, Effects of pressure on the dynamics of an oligomeric protein from deep-sea hyperthermophile, Proc. Natl. Acad. Sci. U.S.A. 112 (2015) 13886–13891. https://doi.org/10.1073/pnas.1514478112.

[28] C.A. Royer, Application of pressure to biochemical equilibria: the other thermodynamic variable, Meth. Enzymol. 259 (1995) 357–377. https://doi.org/10.1016/0076-6879(95)59052-8.

[29] J. Silva, E. Miles, G. Weber, Pressure Dissociation and Conformational Drift of the Beta-Dimer of Tryptophan Synthase, Biochemistry. 25 (1986) 5780–5786. https://doi.org/10.1021/bi00367a065.

[30] K. Ruan, G. Weber, Dissociation of Yeast Hexokinase by Hydrostatic-Pressure, Biochemistry. 27 (1988) 3295–3301. https://doi.org/10.1021/bi00409a026.

[31] M.J. Kornblatt, R. Lange, C. Balny, Use of hydrostatic pressure to produce "native" monomers of yeast enolase, Eur. J. Biochem. 271 (2004) 3897–3904. https://doi.org/10.1111/j.1432-1033.2004.04326.x.

[32] K. Ruan, G. Weber, Hysteresis and conformational drift of pressure-dissociated glyceraldehydephosphate dehydrogenase, Biochemistry. 28 (1989) 2144–2153. https://doi.org/10.1021/bi00431a028.

[33] K. Ruan, G. Weber, Physical heterogeneity of muscle glycogen phosphorylase revealed by hydrostatic pressure dissociation, Biochemistry. 32 (1993) 6295–6301. https://doi.org/10.1021/bi00075a025.

[34] S. Pin, C. Royer, E. Gratton, B. Alpert, G. Weber, Subunit Interactions in Hemoglobin Probed by Fluorescence and High-Pressure Techniques, Biochemistry. 29 (1990) 9194–9202. https://doi.org/10.1021/bi00491a013.

[35] D. Foguel, G. Weber, Pressure-induced Dissociation and Denaturation of Allophycocyanin at Subzero Temperatures, J. Biol. Chem. 270 (1995) 28759–28766. https://doi.org/10.1074/jbc.270.48.28759.

[36] J. Silva, M. Villasboas, C. Bonafe, N. Meirelles, Anomalous Pressure Dissociation of Large Protein Aggregates - Lack of Concentration-Dependence and Irreversibility at Extreme Degrees of Dissociation of Extracellular Hemoglobin, J. Biol. Chem. 264 (1989) 15863–15868.

[37] J.L. Silva, D. Foguel, A.T. DaPoian, P.E. Prevelige, The use of hydrostatic pressure as a tool to study viruses and other macromolecular assemblages, Curr. Opin. Struct. Biol. 6 (1996) 166–175. https://doi.org/10.1016/S0959-440X(96)80071-6.

[38] G. Weber, A.T. DaPoian, J.L. Silva, Concentration dependence of the subunit association of oligomers and viruses and the modification of the latter by urea binding, Biophys. J. 70 (1996) 167–173. https://doi.org/10.1016/S0006-3495(96)79557-7.

[39] J. Silva, G. Weber, Pressure-Induced Dissociation of Brome Mosaic-Virus, J. Mol. Biol. 199 (1988) 149–159. https://doi.org/10.1016/0022-2836(88)90385-3.

[40] J. Silva, P. Luan, M. Glaser, E. Voss, G. Weber, Effects of Hydrostatic-Pressure on a Membrane-Enveloped Virus - High Immunogenicity of the Pressure-Inactivated Virus, J. Virol. 66 (1992) 2111–2117.

[41] A.T. Da Poian, A.C. Oliveira, L.P. Gaspar, J.L. Silva, G. Weber, Reversible Pressure Dissociation of R17 Bacteriophage: The Physical Individuality of Virus Particles, J. Mol. Biol. 231 (1993) 999–1008. https://doi.org/10.1006/jmbi.1993.1347.

[42] J. Torrent, R. Lange, H. Rezaei, The Volumetric Diversity of Misfolded Prion Protein Oligomers Revealed by Pressure Dissociation, J. Biol. Chem. 290 (2015) 20417–20426. https://doi.org/10.1074/jbc.M115.661710.

[43] P. Messier, C. Seguin, Effects of High Hydrostatic-Pressure on Microfilaments and Microtubules, Journal of Embryology and Experimental Morphology. 44 (1978) 281–295.

[44] H. Kobori, M. Sato, A. Tameike, K. Hamada, S. Shimada, M. Osumi, Changes in microfilaments and microtubules of yeasts induced by pressure stress, in: R. Hayashi, C. Balny (Eds.), Progress in Biotechnology, Elsevier, Amsterdam, 1996: pp. 83–94.

[45] M. Nishiyama, Y. Shimoda, M. Hasumi, Y. Kimura, M. Terazima, Microtubule depolymerization at high pressure, Ann. N. Y. Acad. Sci. 1189 (2010) 86–90. https://doi.org/10.1111/j.1749-6632.2009.05411.x.

[46] R. Gebhardt, W. Doster, J. Friedrich, W. Petry, A. Schulte, Pressure-induced critical association of myoglobin, Springer-Verlag Berlin, Berlin, 2003.

[47] J. Torrent, J.P. Connelly, M.G. Coll, M. Ribó, R. Lange, M. Vilanova, Pressure versus heat-induced unfolding of ribonuclease A: the case of hydrophobic interactions within a chain-folding initiation site, Biochemistry. 38 (1999) 15952–15961. https://doi.org/10.1021/bi991460b.

[48] J. Font, J. Torrent, M. Ribó, D.V. Laurents, C. Balny, M. Vilanova, R. Lange, Pressure-jump-induced kinetics reveals a hydration dependent folding/unfolding mechanism of ribonuclease A, Biophys J. 91 (2006) 2264–2274. https://doi.org/10.1529/biophysj.106.082552.

[49] C.A. Royer, Revisiting volume changes in pressure-induced protein unfolding, Biochim. Biophys. Acta. 1595 (2002) 201–209. https://doi.org/10.1016/s0167-4838(01)00344-2.

[50] E.M. Dumay, M.T. Kalichevsky, J.C. Cheftel, High-Pressure Unfolding and Aggregation of .beta.-Lactoglobulin and the Baroprotective Effects of Sucrose, J. Agric. Food Chem. 42 (1994) 1861–1868. https://doi.org/10.1021/jf00045a006.

[51] M.B. Seefeldt, Y.-S. Kim, K.P. Tolley, J. Seely, J.F. Carpenter, T.W. Randolph, High-pressure studies of aggregation of recombinant human interleukin-1 receptor antagonist: Thermodynamics, kinetics, and application to accelerated formulation studies, Protein Sci. 14 (2005) 2258–2266. https://doi.org/10.1110/ps.051490205.

[52] J. Font, A. Benito, R. Lange, M. Ribó, M. Vilanova, The contribution of the residues from the main hydrophobic core of ribonuclease A to its pressure-folding transition state, Protein Sci. 15 (2006) 1000–1009. https://doi.org/10.1110/ps.052050306.

[53] G. Meier, H. Kriegs, A high pressure cell for dynamic light scattering up to 2 kbars with conservation of plane of polarization, Rev. Sci. Instrum. 79 (2008) 013102. https://doi.org/10.1063/1.2827137.

[54] A.A. Paladini, J.L. Silva, G. Weber, Slab gel electrophoresis of oligomeric proteins under high hydrostatic pressure, Anal. Biochem. 161 (1987) 358–364. https://doi.org/10.1016/0003-2697(87)90463-5.

[55] A. Paladini, G. Weber, L. Erijman, Analysis of Dissociation and Unfolding of Oligomeric Proteins Using a Flat Bed Gel-Electrophoresis at High-Pressure, Anal. Biochem. 218 (1994) 364–369. https://doi.org/10.1006/abio.1994.1193.

[56] M.D. Collins, C.U. Kim, S.M. Gruner, High-Pressure Protein Crystallography and NMR to Explore Protein Conformations, in: D.C. Rees, K.A. Dill, J.R. Williamson (Eds.), Annual Review of Biophysics, Vol 40, Ann. Rev. Biophys., Palo Alto, 2011: pp. 81–98.

[57] E. Paci, High pressure simulations of biomolecules, Biochim. Biophys. Acta-Protein Struct. Molec. Enzym. 1595 (2002) 185–200. https://doi.org/10.1016/S0167-4838(01)00343-0.

[58] S. De Meyer, H. Azijn, D. Surleraux, D. Jochmans, A. Tahri, R. Pauwels, P. Wigerinck, M.-P. de Béthune, TMC114, a Novel Human Immunodeficiency Virus Type 1 Protease Inhibitor Active against Protease Inhibitor-Resistant Viruses, Including a Broad Range of Clinical Isolates, Antimicrob. Agents Chemother. 49 (2005) 2314–2321. https://doi.org/10.1128/AAC.49.6.2314-2321.2005.

[59] B. Clotet, N. Bellos, J.-M. Molina, D. Cooper, J.-C. Goffard, A. Lazzarin, A. Wöhrmann, C. Katlama, T. Wilkin, R. Haubrich, C. Cohen, C. Farthing, D. Jayaweera, M. Markowitz, P. Ruane, S. Spinosa-Guzman, E. Lefebvre, Efficacy and safety of darunavir-ritonavir at week 48 in treatment-experienced patients with HIV-1 infection in POWER 1 and 2: a pooled subgroup analysis of data from two randomised trials, Lancet. 369 (2007) 1169–1178. https://doi.org/10.1016/S0140-6736(07)60497-8.

[60] E.M. Towler, S.V. Gulnik, T.N. Bhat, D. Xie, E. Gustschina, T.R. Sumpter, N. Robertson, C. Jones, M. Sauter, N. Mueller-Lantzsch, C. Debouck, J.W. Erickson, Functional characterization of the protease of human endogenous retrovirus, K10: can it complement HIV-1 protease?, Biochemistry. 37 (1998) 17137–17144. https://doi.org/10.1021/bi9818927.

[61] D. Xie, S. Gulnik, E. Gustchina, B. Yu, W. Shao, W. Qoronfleh, A. Nathan, J.W. Erickson, Drug resistance mutations can effect dimer stability of HIV-1 protease at neutral pH, Protein Sci. 8 (1999) 1702–1707. https://doi.org/10.1110/ps.8.8.1702.

[62] K. Strisovsky, U. Tessmer, J. Langner, J. Konvalinka, H.G. Kräusslich, Systematic mutational analysis of the active-site threonine of HIV-1 proteinase: rethinking the "fireman's grip" hypothesis, Protein Sci. 9 (2000) 1631–1641. https://doi.org/10.1110/ps.9.9.1631.

[63] P.L. Darke, S.P. Jordan, D.L. Hall, J.A. Zugay, J.A. Shafer, L.C. Kuo, Dissociation and association of the HIV-1 protease dimer subunits: equilibria and rates, Biochemistry. 33 (1994) 98–105. https://doi.org/10.1021/bi00167a013.

[64] S.P. Jordan, J. Zugay, P.L. Darke, L.C. Kuo, Activity and dimerization of human immunodeficiency virus protease as a function of solvent composition and enzyme concentration, J. Biol. Chem. 267 (1992) 20028–20032.

[65] C.A. Pargellis, M.M. Morelock, E.T. Graham, P. Kinkade, S. Pav, K. Lubbe, D. Lamarre, P.C. Anderson, Determination of kinetic rate constants for the binding of inhibitors to HIV-1 protease and for the association and dissociation of active homodimer, Biochemistry. 33 (1994) 12527–12534. https://doi.org/10.1021/bi00207a021.

[66] M. Ingr, T. Uhlíková, K. Strísovský, E. Majerová, J. Konvalinka, Kinetics of the dimerization of retroviral proteases: the "fireman's grip" and dimerization, Protein Sci. 12 (2003) 2173–2182. https://doi.org/10.1110/ps.03171903.

[67] R. Ishima, D.A. Torchia, J.M. Louis, Mutational and structural studies aimed at characterizing the monomer of HIV-1 protease and its precursor, J. Biol. Chem. 282 (2007) 17190–17199. https://doi.org/10.1074/jbc.M701304200.

[68] V. Hýsková, K. Bělonožníková, I. Šmeringaiová, D. Kavan, M. Ingr, H. Ryšlavá, How is the activity of shikimate dehydrogenase from the root of Petroselinum crispum (parsley) regulated and which side reactions are catalyzed?, Phytochemistry. 190 (2021) 112881. https://doi.org/10.1016/j.phytochem.2021.112881.

[69] A. Tichá, S. Stanchev, J. Škerle, J. Began, M. Ingr, K. Švehlová, L. Polovinkin, M. Růžička, L. Bednárová, R. Hadravová, E. Poláchová, P. Rampírová, J. Březinová, V. Kašička, P. Majer, K. Strisovsky, Sensitive Versatile Fluorogenic Transmembrane Peptide Substrates for Rhomboid

Intramembrane Proteases, J. Biol. Chem. 292 (2017) 2703–2713.
https://doi.org/10.1074/jbc.M116.762849.

[70] H.E. Hoffman, J. Jirásková, M. Ingr, M. Zvelebil, J. Konvalinka, Recombinant human serine racemase: enzymologic characterization and comparison with its mouse ortholog, Protein Expr. Purif. 63 (2009) 62–67. https://doi.org/10.1016/j.pep.2008.09.003.

[71] H.G. Kräusslich, Human immunodeficiency virus proteinase dimer as component of the viral polyprotein prevents particle assembly and viral infectivity., Proc. Natl. Acad. Sci. U S A. 88 (1991) 3213–3217.

[72] J. Konvalinka, M.A. Litterst, R. Welker, H. Kottler, F. Rippmann, A.M. Heuser, H.G. Kräusslich, An active-site mutation in the human immunodeficiency virus type 1 proteinase (PR) causes reduced PR activity and loss of PR-mediated cytotoxicity without apparent effect on virus maturation and infectivity, J Virol. 69 (1995) 7180–7186. https://doi.org/10.1128/JVI.69.11.7180-7186.1995.

[73] J.R. Rosé, L.M. Babé, C.S. Craik, Defining the level of human immunodeficiency virus type 1 (HIV-1) protease activity required for HIV-1 particle maturation and infectivity, J Virol. 69 (1995) 2751–2758. https://doi.org/10.1128/JVI.69.5.2751-2758.1995.

[74] T.W. Ridky, D. Bizub-Bender, C.E. Cameron, I.T. Weber, A. Wlodawer, T. Copeland, A.M. Skalka, J. Leis, Programming the Rous sarcoma virus protease to cleave new substrate sequences, J. Biol. Chem. 271 (1996) 10538–10544. https://doi.org/10.1074/jbc.271.18.10538.

[75] G. Arad, M. Chorev, A. Shtorch, A. Goldblum, M. Kotler, Point mutation in avian sarcoma leukaemia virus protease which increases its activity but impairs infectious virus production, J. Gen. Virol. 76 (Pt 8) (1995) 1917–1925. https://doi.org/10.1099/0022-1317-76-8-1917.

[76] H. Burstein, D. Bizub, A.M. Skalka, Assembly and processing of avian retroviral gag polyproteins containing linked protease dimers, J Virol. 65 (1991) 6165–6172. https://doi.org/10.1128/JVI.65.11.6165-6172.1991.

[77] Y. Carrington, J. Guo, C.H. Le, A. Fillo, J. Kwon, L.T. Tran, J. Ehlting, Evolution of a secondary metabolic pathway from primary metabolism: shikimate and quinate biosynthesis in plants, Plant J. (2018). https://doi.org/10.1111/tpj.13990.

[78] Y. Deng, S. Lu, Biosynthesis and Regulation of Phenylpropanoids in Plants, Critical Reviews in Plant Sciences. 36 (2017) 257–290. https://doi.org/10.1080/07352689.2017.1402852.

[79] D.L. Purich, Enzyme Kinetics: Catalysis and Control: A Reference of Theory and Best-Practice Methods, 1st edition, Elsevier, Amsterdam, 2010.

[80] M. Freeman, Rhomboid proteases and their biological functions, Annu. Rev. Genet. 42 (2008) 191–210. https://doi.org/10.1146/annurev.genet.42.110807.091628.

[81] S. Urban, S.W. Dickey, The rhomboid protease family: a decade of progress on function and mechanism, Genome Biol. 12 (2011) 231. https://doi.org/10.1186/gb-2011-12-10-231.

[82] K. Strisovsky, Why cells need intramembrane proteases - a mechanistic perspective, FEBS J. 283 (2016) 1837–1845. https://doi.org/10.1111/febs.13638.

[83] H. Wolosker, S. Blackshaw, S.H. Snyder, Serine racemase: a glial enzyme synthesizing D-serine to regulate glutamate-N-methyl-D-aspartate neurotransmission, Proc. Natl. Acad. Sci. U S A. 96 (1999) 13409–13414. https://doi.org/10.1073/pnas.96.23.13409.

[84] D.L. Graham, M.L. Beio, D.L. Nelson, D.B. Berkowitz, Human Serine Racemase: Key Residues/Active Site Motifs and Their Relation to Enzyme Function, Front. Mol. Biosci. 6 (2019) 8. https://doi.org/10.3389/fmolb.2019.00008.

[85] K. Strísovský, J. Jirásková, C. Barinka, P. Majer, C. Rojas, B.S. Slusher, J. Konvalinka, Mouse brain serine racemase catalyzes specific elimination of L-serine to pyruvate, FEBS Lett. 535 (2003) 44–48. https://doi.org/10.1016/s0014-5793(02)03855-3.

[86] V.N. Foltyn, I. Bendikov, J. De Miranda, R. Panizzutti, E. Dumin, M. Shleper, P. Li, M.D. Toney, E. Kartvelishvily, H. Wolosker, Serine racemase modulates intracellular D-serine levels through an alpha,beta-elimination activity, J. Biol. Chem. 280 (2005) 1754–1763. https://doi.org/10.1074/jbc.M405726200.

[87] H. Wolosker, E. Dumin, L. Balan, V.N. Foltyn, D-amino acids in the brain: D-serine in neurotransmission and neurodegeneration, FEBS J. 275 (2008) 3514–3526. https://doi.org/10.1111/j.1742-4658.2008.06515.x.

[88]   M.J. Scolari, G.B. Acosta, D-serine: a new word in the glutamatergic neuro-glial language, Amino Acids. 33 (2007) 563–574. https://doi.org/10.1007/s00726-006-0481-0.

[89]   S.J. Kodumal, K.G. Patel, R. Reid, H.G. Menzella, M. Welch, D.V. Santi, Total synthesis of long DNA sequences: Synthesis of a contiguous 32-kb polyketide synthase gene cluster, Proc. Natl. Acad. Sci. U S A. 101 (2004) 15573–15578. https://doi.org/10.1073/pnas.0406911101.

[90]   H.O. Smith, C.A. Hutchison, C. Pfannkoch, J.C. Venter, Generating a synthetic genome by whole genome assembly: φX174 bacteriophage from synthetic oligonucleotides, Proc. Natl. Acad. Sci. U S A. 100 (2003) 15440–15445. https://doi.org/10.1073/pnas.2237126100.

[91]   J.R. TerMaat, E. Pienaar, S.E. Whitney, T.G. Mamedov, A. Subramanian, Gene synthesis by integrated polymerase chain assembly and PCR amplification using a high-speed thermocycler, J. Microbiol. Methods. 79 (2009) 295–300. https://doi.org/10.1016/j.mimet.2009.09.015.

[92]   M. Polášek, M. Juřek, M. Ingr, P. Čársky, J. Horáček, Discrete momentum representation of the Lippmann-Schwinger equation and its application to electron-molecule scattering, Phys. Rev. A. 61 (2000) 032701. https://doi.org/10.1103/PhysRevA.61.032701.

[93]   M. Ingr, M. Polášek, P. Čársky, J. Horáček, Discrete momentum representation method for polar molecules: Calculation of the elastic electron scattering on the ${\mathrm{H}}_{2}\mathrm{O}$ molecule, Phys. Rev. A. 62 (2000) 032703. https://doi.org/10.1103/PhysRevA.62.032703.

[94]   M. Ingr, H.-D. Meyer, L.S. Cederbaum, Potential energy curve of the X2Sigmau$\mathplus$resonance state of F2-computed by CAP/CI, J. Phys. B: At. Mol. Opt. Phys. 32 (1999) L547–L556. https://doi.org/10.1088/0953-4075/32/19/103.

[95]   F. Jona, J.A. Strozier, W.S. Yang, Low-energy electron diffraction for surface structure analysis, Rep. Prog. Phys. 45 (1982) 527–585. https://doi.org/10.1088/0034-4885/45/5/002.

[96]   T. Gruene, E. Mugnaioli, 3D Electron Diffraction for Chemical Analysis: Instrumentation Developments and Innovative Applications, Chem. Rev. 121 (2021) 11823–11834. https://doi.org/10.1021/acs.chemrev.1c00207.

[97]   M. Gemmi, E. Mugnaioli, T.E. Gorelik, U. Kolb, L. Palatinus, P. Boullay, S. Hovmöller, J.P. Abrahams, 3D Electron Diffraction: The Nanocrystallography Revolution, ACS Cent. Sci. 5 (2019) 1315–1329. https://doi.org/10.1021/acscentsci.9b00394.

[98]   F. Hofer, F.P. Schmidt, W. Grogger, G. Kothleitner, Fundamentals of electron energy-loss spectroscopy, IOP Conf. Ser.: Mater. Sci. Eng. 109 (2016) 012007. https://doi.org/10.1088/1757-899X/109/1/012007.

[99]   H. Feshbach, Unified theory of nuclear reactions, Ann. Phys. 5 (1958) 357–390. https://doi.org/10.1016/0003-4916(58)90007-1.

[100]   H. Feshbach, A unified theory of nuclear reactions. II, Ann. Phys. 19 (1962) 287–313. https://doi.org/10.1016/0003-4916(62)90221-X.

[101]   A.U. Hazi, H.S. Taylor, Stabilization Method of Calculating Resonance Energies: Model Problem, Phys. Rev. A. 1 (1970) 1109–1120. https://doi.org/10.1103/PhysRevA.1.1109.

[102]   J. Aguilar, J.M. Combes, A class of analytic perturbations for one-body Schrödinger Hamiltonians, Commun. Math. Phys. 22 (1971) 269–279. https://doi.org/10.1007/BF01877510.

[103]   E. Balslev, J.M. Combes, Spectral properties of many-body Schrödinger operators with dilatation-analytic interactions, Commun. Math. Phys. 22 (1971) 280–294. https://doi.org/10.1007/BF01877511.

[104]   B. Simon, Resonances in n-Body Quantum Systems With Dilatation Analytic Potentials and the Foundations of Time-Dependent Perturbation Theory, Ann. Math. 97 (1973) 247–274. https://doi.org/10.2307/1970847.

[105]   U.V. Riss, H.-D. Meyer, Calculation of resonance energies and widths using the complex absorbing potential method, J. Phys. B: At. Mol. Opt. Phys. 26 (1993) 4503–4535. https://doi.org/10.1088/0953-4075/26/23/021.

[106]   M. Ingr, J. Dostál, T. Majerová, Enzymological description of multitemplate PCR-Shrinking amplification bias by optimizing the polymerase-template ratio, J. Theor. Biol. 382 (2015) 178–186. https://doi.org/10.1016/j.jtbi.2015.06.048.

[107]   M. Ingr, J. Konvalinka, Theoretical description of the direct exponential amplification and sequencing (DEXAS) method, Biol. Chem. 381 (2000) 439–445. https://doi.org/10.1515/BC.2000.057.

[108] H. Marušincová, L. Husárová, J. Růžička, M. Ingr, V. Navrátil, L. Buňková, M. Koutny, Polyvinyl alcohol biodegradation under denitrifying conditions, Int. Biodeterior. Biodegrad. 84 (2013) 21–28. https://doi.org/10.1016/j.ibiod.2013.05.023.

[109] V. Kotrbová, D. Aimová, M. Ingr, L. Borek-Dohalská, V. Martínek, M. Stiborová, Preparation of a biologically active apo-cytochrome b5 via heterologous expression in Escherichia coli, Protein Expr. Purif. 66 (2009) 203–209. https://doi.org/10.1016/j.pep.2009.03.011.

[110] L. Garibyan, N. Avashia, Research Techniques Made Simple: Polymerase Chain Reaction (PCR), J. Invest. Dermatol. 133 (2013) e6. https://doi.org/10.1038/jid.2013.1.

[111] E. Kalle, M. Kubista, C. Rensing, Multi-template polymerase chain reaction, Biomol. Detect. Quantif. 2 (2014) 11–29. https://doi.org/10.1016/j.bdq.2014.11.002.

[112] S. Schnell, C. Mendoza, Enzymological considerations for a theoretical description of the quantitative competitive polymerase chain reaction (QC-PCR), J. Theor. Biol. 184 (1997) 433–440. https://doi.org/10.1006/jtbi.1996.0283.

[113] S. Schnell, C. Mendoza, Theoretical description of the polymerase chain reaction, J. Theor. Biol. 188 (1997) 313–318. https://doi.org/10.1006/jtbi.1997.0473.

[114] H.D. VanGuilder, K.E. Vrana, W.M. Freeman, Twenty-five years of quantitative PCR for gene expression analysis, BioTechniques. 44 (2008) 619–626. https://doi.org/10.2144/000112776.

[115] S. Deepak, K. Kottapalli, R. Rakwal, G. Oros, K. Rangappa, H. Iwahashi, Y. Masuo, G. Agrawal, Real-Time PCR: Revolutionizing Detection and Expression Analysis of Genes, Curr. Genomics. 8 (2007) 234–251. https://doi.org/10.2174/138920207781386960.

[116] C. Kilger, S. Pääbo, Direct exponential amplification and sequencing (DEXAS) of genomic DNA, Biol. Chem. 378 (1997) 99–105. https://doi.org/10.1515/bchm.1997.378.2.99.

[117] F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. U S A. 74 (1977) 5463–5467. https://doi.org/doi: 10.1073/pnas.74.12.5463.

[118] M.L. Metzker, Emerging technologies in DNA sequencing, Genome Res. 15 (2005) 1767–1776. https://doi.org/10.1101/gr.3770505.

[119] T. Hu, N. Chitnis, D. Monos, A. Dinh, Next-generation sequencing technologies: An overview, Hum. Immunol. 82 (2021) 801–811. https://doi.org/10.1016/j.humimm.2021.02.012.

[120] S. Goodwin, J.D. McPherson, W.R. McCombie, Coming of age: ten years of next-generation sequencing technologies, Nat. Rev. Genet. 17 (2016) 333–351. https://doi.org/10.1038/nrg.2016.49.

[121] N.B. Halima, Poly(vinyl alcohol): review of its promising applications and insights into biodegradation, RSC Adv. 6 (2016) 39823–39832. https://doi.org/10.1039/C6RA05742J.

[122] F. Kawai, X. Hu, Biochemistry of microbial polyvinyl alcohol degradation, Appl. Microbiol. Biotechnol. 84 (2009) 227–237. https://doi.org/10.1007/s00253-009-2113-6.

[123] E. Chiellini, A. Corti, S. D'Antone, R. Solaro, Biodegradation of poly (vinyl alcohol) based materials, Prog. Polym. Sci. 28 (2003) 963–1014. https://doi.org/10.1016/S0079-6700(02)00149-1.

[124] G.A.I. Mejía, O.B.L. López, P.A. Mulet, Biodegradation of poly (vinylalcohol) with enzymatic extracts of phanerochaete chrysosporium, Macromol. Symp. 148 (1999) 131–147. https://doi.org/10.1002/masy.19991480112.

[125] S. Gartiser, M. Wallrabenstein, G. Stiene, Assessment of several test methods for the determination of the anaerobic biodegradability of polymers, J. Environ. Polym. Degrad. 6 (1998) 159–173.

[126] H. Yu, G. Gu, L. Song, Degradation of Polyvinyl Alcohol in Sequencing Batch Reactors, Environ. Technol. 17 (1996) 1261–1267. https://doi.org/10.1080/09593331708616496.

[127] J. Hrnčiřík, J. Pšeja, J. Kupec, S. Bernkopfová, Anaerobic Biodegradation of Polyvinyl Alcohol Modified by Extracellular Polysaccharides, J. Polym. Environ. 18 (2010) 98–103. https://doi.org/10.1007/s10924-010-0180-5.

[128] V. Viglasky, Polyacrylamide temperature gradient gel electrophoresis, Methods Mol. Biol. 1054 (2013) 159–171. https://doi.org/10.1007/978-1-62703-565-1_10.

[129] T.D. Porter, The roles of cytochrome b5 in cytochrome P450 reactions, J. Biochem. Mol. Toxicol. 16 (2002) 311–316. https://doi.org/10.1002/jbt.10052.

[130] S.F. Velick, P. Strittmatter, The oxidation-reduction stoichiometry and potential of microsomal cytochrome, J. Biol. Chem. 221 (1956) 265–275.

[131] G. Vergéres, L. Waskell, Cytochrome b5, its functions, structure and membrane topology, Biochimie. 77 (1995) 604–620. https://doi.org/10.1016/0300-9084(96)88176-4.

[132] T.A. Clarke, S.-C. Im, A. Bidwai, L. Waskell, The role of the length and sequence of the linker domain of cytochrome b5 in stimulating cytochrome P450 2B4 catalysis, J. Biol. Chem. 279 (2004) 36809–36818. https://doi.org/10.1074/jbc.M406055200.

[133] J.B. Schenkman, I. Jansson, The many roles of cytochrome b5, Pharmacol. Ther. 97 (2003) 139–152. https://doi.org/10.1016/s0163-7258(02)00327-3.

[134] H. Yamazaki, W.W. Johnson, Y.F. Ueng, T. Shimada, F.P. Guengerich, Lack of electron transfer from cytochrome b5 in stimulation of catalytic activities of cytochrome P450 3A4. Characterization of a reconstituted cytochrome P450 3A4/NADPH-cytochrome P450 reductase system and studies with apo-cytochrome b5, J. Biol. Chem. 271 (1996) 27438–27444. https://doi.org/10.1074/jbc.271.44.27438.

[135] H. Yamazaki, T. Shimada, M.V. Martin, F.P. Guengerich, Stimulation of Cytochrome P450 Reactions by Apo-cytochromeb 5: EVIDENCE AGAINST TRANSFER OF HEME FROM CYTOCHROME P450 3A4 TO APO-CYTOCHROME b 5 OR HEME OXYGENASE *, J. Biol. Chem. 276 (2001) 30885–30891. https://doi.org/10.1074/jbc.M105011200.

[136] H. Yamazaki, M. Nakamura, T. Komatsu, K. Ohyama, N. Hatanaka, S. Asahi, N. Shimada, F.P. Guengerich, T. Shimada, M. Nakajima, T. Yokoi, Roles of NADPH-P450 Reductase and Apo- and Holo-Cytochrome b5 on Xenobiotic Oxidations Catalyzed by 12 Recombinant Human Cytochrome P450s Expressed in Membranes of Escherichia coli, Protein Expr. Purif. 24 (2002) 329–337. https://doi.org/10.1006/prep.2001.1578.

[137] D.L. Cinti, J. Ozols, Binding of homogeneous cytochrome b5 to rat liver microsomes. Effect on N-demethylation reactions, Biochim Biophys Acta. 410 (1975) 32–44. https://doi.org/10.1016/0005-2744(75)90205-3.

[138] B. Mrazova, M. Martinkova, V. Martínek, E. Frei, M. Stiborová, Optimalization of preparation of apo-cytochrome b(5) utilizing apo-myoglobin, Interdisciplinary Toxicology. (2008). https://doi.org/10.2478/v10102-010-0037-8.

[139] S. Pekařová, M. Dvořáčková, P. Stloukal, M. Ingr, J. Šerá, M. Koutny, Quantitation of the Inhibition Effect of Model Compounds Representing Plant Biomass Degradation Products on Methane Production, BioResources. 12 (2017) 2421–2432. https://doi.org/10.15376/biores.12.2.2421-2432.

[140] W.S. Adney, C.J. Rivard, M. Shiang, M.E. Himmel, Anaerobic digestion of lignocellulosic biomass and wastes - Cellulases and related enzymes, Appl. Biochem. Biotechnol. 30 (1991) 165–183. https://doi.org/10.1007/BF02921684.

[141] F.R. Amin, H. Khalid, H. Zhang, S. u. Rahman, R. Zhang, G. Liu, C. Chen, Pretreatment methods of lignocellulosic biomass for anaerobic digestion, AMB Express. 7 (2017) 72. https://doi.org/10.1186/s13568-017-0375-4.

[142] O. Exner, M. Ingr, P. Čársky, Ab initio calculations of substituent constants: A reinvestigation, Theochem-J. Mol. Struct. 397 (1997) 231–238. https://doi.org/10.1016/S0166-1280(96)04891-9.

[143] E.N. Muratov, J. Bajorath, R.P. Sheridan, I.V. Tetko, D. Filimonov, V. Poroikov, T.I. Oprea, I.I. Baskin, A. Varnek, A. Roitberg, O. Isayev, S. Curtalolo, D. Fourches, Y. Cohen, A. Aspuru-Guzik, D.A. Winkler, D. Agrafiotis, A. Cherkasov, A. Tropsha, QSAR without borders, Chem. Soc. Rev. 49 (2020) 3525–3564. https://doi.org/10.1039/D0CS00098A.

[144] R.W. Taft, R.D. Topsom, The Nature and Analysis of Substitutent Electronic Effects, in: Progress in Physical Organic Chemistry, John Wiley & Sons, Ltd, 1987: pp. 1–83. https://doi.org/10.1002/9780470171950.ch1.

[145] R.D. Topsom, Some Theoretical Studies of Electronic Substituent Effects in Organic Chemistry, in: Progress in Physical Organic Chemistry, John Wiley & Sons, Ltd, 1987: pp. 125–191. https://doi.org/10.1002/9780470171950.ch3.

[146] S. Marriott, R.D. Topsom, Theoretical studies of the inductive effect. IV. A theoretical scale of substituent field parameters, ACS Publications. (2002). https://doi.org/10.1021/ja00313a002.

[147] S. Marriott, R.D. Topsom, A theoretical scale of substituent resonance parameters (σR°), J. Chem. Soc., Perkin Trans. 2. (1985) 1045–1047. https://doi.org/10.1039/P29850001045.

[148] S. Marriott, W.F. Reynolds, R.W. Taft, R.D. Topsom, Substituent electronegativity parameters, J. Org. Chem. 49 (1984) 959–965. https://doi.org/10.1021/jo00180a002.

[149] V. Hrouda, P. Čársky, M. Ingr, Z. Chval, G.N. Sastry, T. Bally, The C4H6+ potential energy surface. 2. The reaction of ethylene radical cation with acetylene, J. Phys. Chem. A. 102 (1998) 9297–9307. https://doi.org/10.1021/jp982590h.

[150] A. Yu, A mini review of the crossed molecular beam apparatus in molecular reaction dynamics, J. Saudi Chem. Soc. 23 (2019) 1–6. https://doi.org/10.1016/j.jscs.2018.04.004.

[151] N. Balucani, G. Capozza, F. Leonori, E. Segoloni, P. Casavecchia, Crossed molecular beam reactive scattering: from simple triatomic to multichannel polyatomic reactions, Int. Rev. Phys. Chem. 25 (2006) 109–163. https://doi.org/10.1080/01442350600641305.

[152] T. Baer, J. Booze, K.-M. Weitzel, Photoelectron Photoion Coincidence Studies of Ion Dissociation Dynamics, in: Vacuum Ultraviolet Photoionization and Photodissociation of Molecules and Clusters, WORLD SCIENTIFIC, 1991: pp. 259–296. https://doi.org/10.1142/9789812814821_0005.

[153] T. Baer, Ion dissociation dynamics and thermochemistry by photoelectron photoion coincidence (PEPICO) spectroscopy, Int. J. Mass Spectrom. 200 (2000) 443–457. https://doi.org/10.1016/S1387-3806(00)00327-4.

[154] Z. Herman, K. Birkinshaw, Reactive Scattering of Ions, Berichte Der Bunsengesellschaft Für Physikalische Chemie. 77 (1973) 566–575. https://doi.org/10.1002/bbpc.19730770805.

[155] G.N. Sastry, T. Bally, V. Hrouda, P. Cársky, The C4H6•+ Potential Energy Surface. 1. The Ring-Opening Reaction of Cyclobutene Radical Cation and Related Rearrangements, J. Am. Chem. Soc. 120 (1998) 9323–9334. https://doi.org/10.1021/ja981651m.

[156] P. Du, W.T. Borden, Ab initio calculations predict a singlet ground state for tetramethyleneethane, ACS Publications. (2002). https://doi.org/10.1021/ja00237a066.

[157] D.J. Bellville, R. Chelsky, N.L. Bauld, A nonelectrocyclic path from the cyclobutene cation radical to the 1,3-butadiene cation radical, J. Comput. Chem. 3 (1982) 548–551. https://doi.org/10.1002/jcc.540030411.

[158] O. Wiest, Ab Initio Studies of the Ring-Opening Reaction of the Cyclobutene Radical Cation, J. Am. Chem. Soc. 119 (1997) 5713–5719. https://doi.org/10.1021/ja964334t.

Publikované práce

P1. Hyaluronan random coils in electrolyte solutions-a molecular dynamics study.

Contents lists available at ScienceDirect

Carbohydrate Polymers

Hyaluronan random coils in electrolyte solutions—a molecular dynamics study

CrossMark

Marek Ingr*, Eva Kutálková, Josef Hrnčiřík

Tomas Bata University in Zlín, Faculty of Technology, Department of Physics and Materials Engineering, Nám. T. G. Masaryka 5555, 76001 Zlín, Czechia

ABSTRACT

A computational method of modeling random coils of hyaluronan was developed based on the molecular-dynamics simulations. An oligosaccharide of 48 monosaccharide units was equilibrated within a 70–100 ns simulation and randomly chosen pieces of this molecule from different simulation frames were combined to constitute a long polysaccharide chain, both for hyaluronan and its non-ionic analog containing glucose instead of glucuronic acid. The dihedral angles of the glycoside connections of the pieces obeyed the statistics deduced from the simulation. The simulations were performed at various concentrations of NaCl and MgCl$_2$. The calculated radii of gyration show a striking agreement with experimental data from the literature and indicate a key importance of the polymer-ion interactions for the random-coil conformation, but a low influence of the excluded volume of the chain and the carboxylate-groups repulsion. The method has thus the potential to become a versatile tool of modeling macromolecules of various semirigid polymers.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Hyaluronic acid is a natural polysaccharide consisting of alternating units of β-D-glucuronic (GCU) acid and β-D-N-acetylglucosamine (NAG). The monomeric units are connected by β-1,3 glycosidic bond between C1 of GCU and C3 of NAG and by β-1,4 glycosidic bond between C1 of NAG and C4 of GCU [4]-β-D-GlcpA-(1→3)-β-D-GlcpNAc-(1→]$_n$. At physiological conditions it mostly occurs in an ionized form (hyaluronan). Hyaluronic acid is a biologically active molecule occurring in connective tissues, especially the synovial fluid, vitreous fluid of eyes, umbilical cords and in chicken combs. Due to its biological functions it is an object of interest of both pharmaceutical and cosmetics industry. For more detailed information see current reviews (Allison & Grande-Allen, 2006; Jiang, Liang, & Noble, 2007; Necas, Bartosikova, Brauner, & Kolar, 2008). Hyaluronic acid is synthesized by membrane enzymes hyaluronan synthases and is catabolized by hydrolases called hyaluronidases (Stern, 2003). Industrially it is produced from animal tissues or genetically modified bacteria, but the cell-free technologies are being developed, too (Sze, Brownlie, & Love, 2016).

Hyaluronan is a hydrophilic polymer with a strong retention of water. In water environment it forms highly swollen random coils, the shape and dimensions of which are influenced by the solution composition. As hyaluronic acid is a polyelectrolyte practically fully dissociated in physiological conditions, it is conceivable that the concentration of ions is one of the key factors influencing the shape of its macromolecules. Indeed, Fouissac, Milas, Rinaudo, and Borsali (1992) studied the dependence of the radius of gyration (R$_g$) on the concentration of NaCl and confirmed a good agreement with the theory of Odijk, Skolnick and Fixman (Odijk, 1977, 1978; Odijk & Houwaart, 1978; Skolnick & Fixman, 1977), i.e. a continuous decrease of the random-coil dimensions when salt concentration increases. Later on numerous experimental studies were published presenting the measured values of different characteristics of hyaluronan random coils, especially the radius of gyration, hydrodynamic radius, diffusion coefficient, intrinsic viscosity of the solution or persistence length of the chain, often in dependence on ionic strength of the solution. Hayashi, Tsutsumi, Nakajima, Norisuye, and Teramoto (1995) studied the properties of the hyaluronan solutions in higher salt concentrations, 0.2 and 0.5 M NaCl, Mendichi, Soltés, and Giacometti Schieroni (2003) carried out a complex study of hyaluronan properties at a single concentration of 0.15 M NaCl. In both these studies the dependences of R$_g$ and intrinsic viscosity on molecular weight as well as the scaling factors of both the quantities were determined. Sorci & Reed (2004) studied hyaluronan solutions at varying concentrations of NaCl and CaCl$_2$ up to the ionic strength of 0.1 M. They showed a continuous decrease of the radius of gyration with grow-

ing ionic strength as well as the formation of smaller coils in CaCl$_2$ with respect to NaCl. Buhler & Boué (2004) showed that the persistence length of a hyaluronan molecule decreases with growing ionic strength of the solution. In addition, interactions of hyaluronan with other compounds, especially quaternary ammonium salt surfactants, were studied (Bjoerling, Hersloef-Bjoerling, & Stilbs, 1995; Grundelova, Mracek, Kasparkova, Minarik, & Smolka, 2013).

The structure of free hyaluronan molecules has also been studied by means of theoretical chemistry. One of the first attempts was carried out by Holmbeck, Petillo, and Lerner (1994) who determined average values of the dihedral angles of the glycosidic bonds in hyaluronan by molecular-mechanics approach. Kaufmann, Möhle, Hofmann, and Arnold (1998) determined the pairs of dihedral angles of both the β-1,3 and β-1,4 glycosidic bonds and constructed the respective Ramachandran plots using the molecular-dynamics approach applied to short pieces of hyaluronan, namely dimers and trimers of the basic monosaccharide units. Pereira et al. (2006) generalized this method to several different disaccharides. Similar approach was used by Almond, Brass, and Sheehan (1998), who determined the average helix of the polymer molecule, and Donati, Magnani, Bonechi, Barbucci, and Rossi (2001), who compared the MD simulations on short pieces of hyaluronan molecule with the experimental NMR data. Kirschner & Woods (2001) showed, using the quantum-mechanical approach, that explicit solvent is necessary for a good reproduction of the dihedral angles. Ivanov & Neamtu (2013) applied this technique to study the influence of dimethylsilanediol to the hyaluronan structure. Furlan, La Penna, Perico, and Cesàro (2004, 2005) used Monte Carlo simulation using all-atoms molecular potential to model the properties of hyaluronan random coils. In a different approach, Nyström et al. (2010) used Monte Carlo method to simulate larger hyaluronan random coils using the bead-and-spring model. This approach enabled simulations of molecules of real sizes, but without the use of the exact molecular potentials. Recently, Mutter et al. (2015) used molecular dynamics in combination with DFT in order to determine the vibrational spectra of hyaluronan.

In this work we apply the molecular-dynamics approach to an oligosaccharide of 48 monosaccharide units and compose large random coils connecting randomly chosen pieces of this chain from different frames of the simulation. This allows us to make all-atom models of random coils up to at least 10000 monosaccharide units and to determine some of their physico-chemical properties.

2. Methods

This study deals with the macromolecules of hyaluronan [4)-β-D-GlcpA-(1→3)-β-D-GlcpNAc-(1→]$_n$ (hereafter abbreviated as HA) and its non-charged analog in which the carboxyl groups are substituted by hydroxymetyl groups, i.e. the glucuronic acid unit is substituted by glucose [4)-β-D-Glcp-(1→3)-β-D-GlcpNAc-(1→]$_n$ (hereafter abbreviated as GlcHA). Equilibrium structures of 48 monosaccharide-units long chains of hyaluronan and its analog were generated by means of molecular-dynamics simulations in environments containing various concentrations of two salts, NaCl and MgCl$_2$. All MD simulations were performed in NAMD Version 2.10 program package (Phillips et al., 2005) using the CHARM36 carbohydrate topology and force field parameters (Guvench et al., 2008; Guvench, Hatcher, Venable, Pastor, & MacKerell, 2009). Interatomic distances and non-bonding interactions were evaluated using VMD 1.9.2 program (Humphrey, Dalke, & Schulten, 1996). High molecular weight random coils were generated by connecting randomly selected pieces of the simulated molecules by a method of selection of the glycoside-bond dihedral angles in accord with their distribution in the simulated molecules. For details see SI.
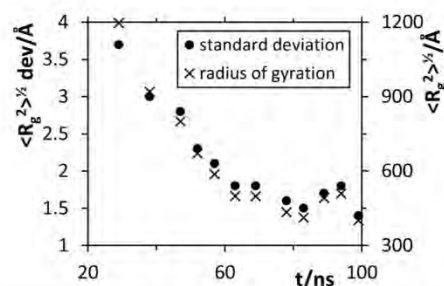
Fig. 1. Time evolution of the mean value (right axis) and the standard deviation (left axis) of the radius of gyration of a set of 5000 random coils of a 2000 monosaccharide-units long molecule of the non-charged hyaluronan analog in water with 1 M MgCl$_2$. The simulation starts from the artificial conformation of a regular helix (Fig. S1, lower panel).

3. Results and discussion

3.1. Generation of equilibrium random coils

MD simulations of polysaccharide molecules of 48 monosaccharide units were carried out for various systems. The chosen molecules were hyaluronan, simulated in 1 M, 0.2 M and 0 M (polyelectrolyte only neutralized by counterions) NaCl and MgCl$_2$, and its non-charged analog in which the glucuronic acid unit was substituted by glucose, simulated in 1 M NaCl and MgCl$_2$ and in pure water.

The simulated molecules were used as the source material for building large macromolecular coils by a procedure similar to that used by Furlan et al. (2005) or Ivanov & Neamtu (2013), but applied to much larger molecules and using a specific statistical procedure to determine the dihedral angles of the glycosidic connections of the pieces (see SI). In order to monitor the equilibration of the system under the MD simulation, the construction of the random coils was performed in every interval of approx. 5 ns. For each such interval 5000 coils of a reference length of 2000 monosaccharides (i.e. 379 kDa) were generated and the mean square value of the radius of gyration and its standard deviation of the mean were calculated. The chosen length is sufficient for the formation of a realistic random coil, but not too long for being biased by the neglect of the excluded volume of the chain (for the discussion of this point see Section 3.2). Fig. 1 shows that the radius of gyration decreases during the course of the simulation reaching a stable region, biased only by random fluctuations, after about 60 ns. This supports the hypothesis that the equilibration of individual parts of the chain leads to the equilibrium structures of the whole generated random coils. Interestingly, together with the R$_g$ value its standard deviation of the mean decreases, too, in a highly regular manner (this phenomenon is discussed in detail in SI).

The initial configuration of the simulated molecules was an artificially constructed regular helix (Fig. S1). Although the molecule immediately adopted a more physically relevant shape of an irregular helix-like chain, the full equilibration lasted for tens nanoseconds, during which it still contained some remainder of the initial, non-physical structure. Hence, the set of frames from the equilibrated region that provided the R$_g$ value closest to the average of this region was further used to model the random coils of different sizes. As the random-coil models were constructed for different electrolyte concentrations, the reversibility of the conformation changes accompanying the transfer of the molecule from one concentration to another was tested on an example of the non-ionic analog of hyaluronan in MgCl$_2$. An oligosaccharide was

initially equilibrated in 1 M $MgCl_2$ and the random coils of 2000 monosaccharide units derived from it showed R_g of (434.4 ± 1.6) Å (see also the next section). When an oligosaccharide of the same geometry was re-equilibrated in pure water, the R_g of the constructed random coils started to increase immediately approaching the value of (802.7 ± 2.7) Å obtained for the same system after 70 ns of independent MD simulation. Hence, the model is well reversible when the electrolyte concentration is being varied.

3.2. The shape and dimensions of the macromolecules

After finding the equilibrated set of frames, random coils were generated for a wide range of polymer lengths spanning from 20 to 10000 monosaccharide units. As can be seen in Figs. S3–S6, the resulting macromolecular coils are relatively rarely packed with a very low probability of chain crossing in the space. (Supplementary information (SI) contains also files in the XYZ format containing the coordinates of the random coils shown in Figs. S3-S6 that can be 3D-viewed in any relevant program.) Therefore, within this study we use only a simple model of a non-interacting coil in which the macromolecular chain can, in principle, intersect itself. To judge the applicability of the method for modeling macromolecules of various molecular weights, the comparison of the determined radii of gyration with miscellaneous experimental data was carried out. Mendichi et al. (2003) measured R_g for a wide scale of molecular weights spanning from 40 kDa to 6 MDa in 0.15 M NaCl. Figs. 2 A and S2 indicates a good agreement of our model (0.2 M NaCl) with this experiment in a substantial part of this interval. In addition, several discrete values of R_g determined by Fouissac et al. (1992), measured at 0.3 M NaCl, agree with our model also well. In that study values for lower NaCl concentrations, 0.06 M and 0.01 M, are presented, too. They indicate a slight expansion of the coils when the salt concentration decreases, which is an identical trend that can be observed in our simulation when going from 0.2 M NaCl to just neutralized hyaluronan molecules (Figs. 2B and S2). For higher salt concentrations less data are available. To our knowledge, no experiments carried out in 1 M NaCl have been published yet, therefore a direct comparison of our simulation at these conditions cannot be done. However, Hayashi et al. (1995) published the values of R_g in 0.5 M NaCl for a wide range of molecular weights. It can be seen in Figs. 2C and S2 that these values fit between our values for 0.2 M and 1.0 M NaCl. This demonstrates a good applicability of the theoretical approach for molecules of the given size. At higher molecular weights this model may be in principle biased by the neglect of the interactions of the distant monosaccharide units and allowed intersection of the polymer chain, which might cause smaller R_g values of the simulated molecules. However, the experimentally determined radius of gyration of a 1.8 MDa molecule in 0.1 M NaCl (Sorci & Reed, 2004) is about 1760 Å while the simulated values are 1530 Å for 0.2 M NaCl and 1720 Å for just neutralized system. Considering the effect of electrolyte concentration on R_g (see also Section 3.3.), this deviation is obviously small even for molecules of this size. Hence, the resulting values show a striking agreement with the experimental data of various sources for molecules up to 10000 monosaccharide units, even under considerably varying conditions. This result, in addition, indicates that the excluded volume of the chain and the interactions of the distant parts of the molecule have only a minor influence on the overall shape of the macromolecule.

3.3. Random coils in varying electrolyte concentrations

Reproducible differences of the molecular sizes can be found among the different simulated systems, especially among the molecules simulated in different electrolyte concentrations. For every system, i.e. a given polymer in a given salt, R_g was deter-
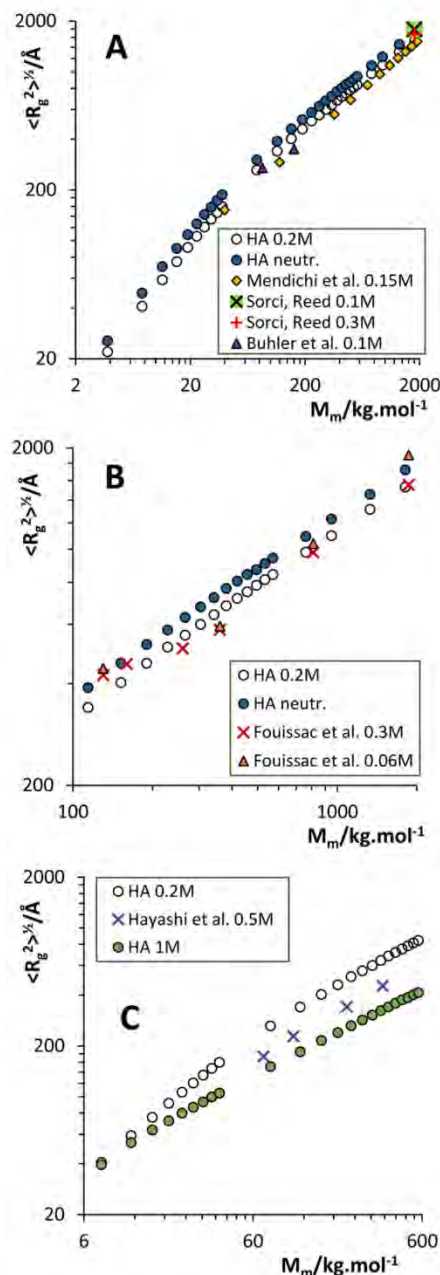

Fig. 2. Comparison of radii of gyration of the modeled hyaluronan random coils in water with various concentrations of NaCl (circles) with experimental data (other symbols) of different sources.
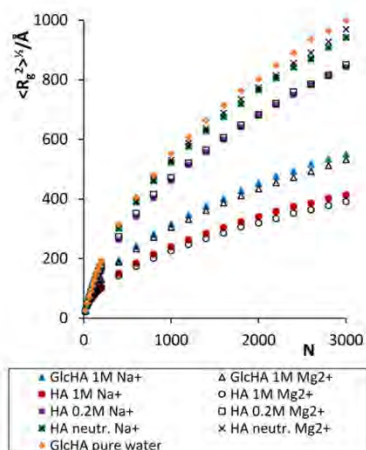
Fig. 3. Radius of gyration as a function of the number of monosaccharide units N for all the simulated systems.

Fig. 4. Scaling factor as a function of the number of monosaccharide units N for the different simulated polysaccharides in NaCl solutions (concentrations are indicated in the legend).

mined for three salt concentrations, 1.0 M, 0.2 M and 0 M (in case of hyaluronan accurate number of counterions was present). Fig. 3 exhibits the R_g dependence on the concentrations of NaCl and MgCl$_2$ and indicates an apparent decrease of this quantity with the increasing salt concentration. This result remarkably agrees with the experimental values (Figs. 2 and S2), although the experimental data are obtained from different sources.

The decrease of the radius of gyration with the growing salt concentration is reflected also by the changes of the persistence length of the polymer chain. Based on the Kratky-Porod model of a worm-like chain (Kratky & Porod, 1949) the persistence length can be calculated from the determined mean radius of gyration using the equation (Benoit & Doty, 1953)

$$\langle R_g^2 \rangle = l_p \left(\frac{L}{3} - l_p + \frac{2l_p^2}{L^2} \left[L - l_p \left(1 - \exp\left(-\frac{L}{l_p} \right) \right) \right] \right), \quad (1)$$

where l_p is the persistence length and L is the contour length of the chain. Numerical solution of Eq. (1) carried out in the Wolfram Mathematica 9 package gives the resulting persistence lengths for the different systems summarized in Table S1, all of them calculated for the reference chain length of 2000 monosaccharide units and the contour length of 10 Å per a disaccharide unit.

As the persistence length reacts very sensitively on the changes in R_g, it decreases rapidly when the salt concentration grows. Hence, in low salt concentration the polymer chain is remarkably stiffer then in higher ionic strength which explains the tendency to form bigger random coils. The determined values of the total persistence length are roughly in agreement with the experimental findings of Buhler & Boué (2004) who also report values close to 200 Å in the salt-free solutions and the decrease of them when the salt concentration grows. They attribute this effect to the screening of the coulombic repulsion of the carboxylate groups in a high salt concentration and absence of this effect in low concentration. However, our simulations indicate that this effect is of a minor importance and the hyaluronan-salt interactions are facilitated especially by the other functional groups of the chain (see also Section 3.5.).

The scaling factor of the radius of gyration, i.e. the power of M_w the radius of gyration grows with, shows a remarkably different behavior in dependence on the salt concentration (Fig. 4). In every case it tends to 1 for very low M_w which is a trivial con-

sequence of the almost rod-like shape of very short molecules. On the other hand, as the model considers only a non-interacting random coil, the scaling factor must necessarily converge to the value of 0.5 when the chain length tends to infinity. In spite of this, considerable differences can be observed for different electrolyte concentrations. In the high electrolyte concentration the scaling factor reaches the value of 0.5 at about 500 monosaccharide units (95 kDa), while in the low concentration the decrease is much slower and the scaling factor does not approach the value of 0.5 below 4000 monosaccharide units (760 kDa) – see Fig. 4. Hence, the observed decrease of the scaling factor roughly correlates with the persistence length – the longer the persistence length, the slower the scaling-factor decrease. Accordingly, the characteristic ratio of the random coil, i.e. the ratio of the mean value of squares of the chain-ends distances $\langle r^2 \rangle$ to the mean value of squares of radii of gyration $\langle R_g^2 \rangle$, decreases in all cases to the value of 6.0 characteristic for an unperturbed random coil, but the decrease is significantly more rapid for macromolecules in high electrolyte concentration (see Fig. S7). However, comparison of the scaling factor with experimental data brings mixed results. Fouissac et al. (1992) present data roughly agreeing with our model, their scaling factor decreases from 0.585 at 0.01 M NaCl towards 0.5 at 0.3 M. On the contrary, the studies of Mendichi et al. (2003) and Hayashi et al. (1995) both provide the constant value of approx. 0.59 for 0.15 M and 0.5 M NaCl, respectively. Thus, according to those works the scaling factor is independent of the salt concentration and does not fall to 0.5 even if it is rather high. Unfortunately, it is not easy to explain this discrepancy as not enough independent data are available, and we can only speculate about its reasons. First of all, the dependencies of R_g on M_w for different scaling factors within the region of 0.5 and 0.6 are all quite similar and a minor inaccuracy can come from the experimental/theoretical setup and data evaluation. Furthermore, the experimental studies a priori expected a constant scaling factor independent on the chain length which might be inappropriate, especially for the short molecules. Finally, our theoretical model ignores the excluded-volume influence, which necessarily causes the convergence of the scaling factor to 0.5, the scaling factor can thus be underestimated for longer molecules. On the other hand, as can be deduced from the images of the modeled molecules (Figs. S3-S6), the chain crossing seems to be very rare, especially for shorter molecules, no matter what the salt concentration is. Therefore, it is unlikely that the excluded-volume effect has a significant influence except for the large molecules of at least MDa molecular weight. In addition, the results of the experimental study investigating the scaling factor consistently for three different salt concentrations (Fouissac et al., 1992) support our findings. Hence, it is not possible to make a clear an unambiguous conclusion in this point.

Fig. 5. A typical picture from MD simulations of the interaction of the sodium (left) and magnesium (right) cations with the carboxylate group.

3.4. Hyaluronan molecules in different electrolytes

Although the general polyelectrolyte theories are based on the concept of ionic strength, it is known from the pioneering work of Hofmeister (1891) that the influence of ions on the polyelectrolytic biomolecules is a more complex problem based on specific characteristics of individual ions. As was shown by Jungwirth and collaborators (Lund, Vácha, & Jungwirth, 2008; Vlachy et al., 2009), the interaction of an ion with a given group of the macromolecule depends on the electronic structure of both the ion and the group in a non-trivial way. Due to this fact even the originally proposed Hofmeister series is not valid generally for all systems. To show the influence of different salts on hyaluronan, we simulated its random coils in the environment of two different salts, NaCl and MgCl$_2$. Both the salts were applied in the concentrations of 0 M (hyaluronan only neutralized by counterions), 0.2 M and 1 M for hyaluronan and 0 M and 1 M for its non-ionic analog. As can be seen in Fig. 3, both the molecules behave almost equally in solutions of both the salts, i.e. their radii of gyration are given only by the concentration but not by the identity of the electrolyte. This feature can be justified in the following way. First, Table S1 shows that the interaction of chloride ions is very small in comparison with any of the cations, their influence is, therefore, minor even though their concentration is twice higher at MgCl$_2$ in comparison with NaCl. Second, Fig. 5 indicates an apparent difference between the interactions of both the ions with the functional groups of the polymer. While Na$^+$ favors the direct interaction, more strongly solvated Mg^{2+} prefers interactions across water molecules rather firmly attached to its surface and thus shielding the electrostatic interaction partially. This may also be a plausible explanation of the partial disruption of hyaluronan-water hydrogen bonds by high concentrations of NaCl, but not MgCl$_2$ (see Table S1). Therefore, the interactions of sodium and magnesium cations with the functional groups are approximately equal which results in a similar behavior of the hyaluronan random coil in equal concentrations of the two salts. However, simulations comprising more different electrolytes might be helpful in order to formulate general conclusions regarding the interaction of hyaluronan and similar molecules with ions, which is planned as a future perspective of this project.

3.5. Interactions driving the changes of the random coil shape

The presented simulations, as well as the discussed experiments, show a strong response of the random coils of both the studied polysaccharides to the changes of the electrolyte concentration. Considering the polyelectrolytic essence of the molecule, the Coulombic interactions between the charged carboxylate groups and their shielding by the ions of the electrolyte may be considered as the most probable cause of this effect. However, as can be seen in Fig. 3, the influence of the salt concentration on the radius of gyration is approximately equal at both the charged molecule of hyaluronan and the neutral molecule of its non-charged ana-

log. It indicates that this interaction is, in fact, minor, which is also documented by Fig. S8 showing that even the pure Coulombic interaction ($\varepsilon_r = 1$) between the carboxylate groups is smaller than other relevant interactions by several orders of magnitude.

This is, actually, not surprising, considering that the average distance between these groups (the C6 atoms) is about 10 Å (Fig. S8B), at which the electrostatic force within the aqueous solution is already very week – the distance is, in fact, very close to the electrostatic cut-off of 10–12 Å applied generally in the MD simulations. However, the interactions of the simulated oligosaccharide chain with the surrounding water molecules and ions are strong and highly dependent on the electrolyte concentration. It indicates the key influence of these interactions on the polysaccharide conformation. Fig. S9 shows that the interaction with ions grows with the electrolyte concentration, while the interaction with water decreases. The non-ionic analog shows the same trend regarding the interaction with water, but in a lower extent (Fig. S10). When the interactions of the carboxylate groups (or hydroxymethyl groups of the analog) with the surrounding water molecules and ions are subtracted, the remaining interactions are roughly equal for both the systems, indicating that their influence on the conformation of the chain is the same at both the molecules (Figs. S11-S14). This finding justifies the hypothesis that the primary influence on the molecular conformation is given by the interactions of the ions and water molecules with the "backbone" of the molecule rather than the carboxylate groups, most likely via their influence on the flexible parts of the molecule, especially the glycoside bonds. Although, to our knowledge, no experimental comparison of this or another structurally similar couple of charged and neutral polymer has been reported yet, numerous studies of different neutral water-soluble polymers in salt solutions indicate the plausibility of this hypothesis. A similar effect was observed at polyethylene glycol (PEG), a polymer resembling HA by the ether bond connecting the monomeric units, although its structure is simpler and more flexible. The increasing salt concentration causes the decrease of the intrinsic viscosity of PEG solutions with the intensities determined by the salt identity (Bailey & Callard, 1959; Brunchi & Ghimici, 2013) which indicates the shrinkage of the macromolecular coils. Another experiment showed the decreased PEG swelling in high salt concentrations by a direct measurement of the mass of the solvating water molecules using a quartz microbalance (Heeb, Lee, Venkataraman, & Spencer, 2009). An analogous behavior was found also for polyvinylpyrrolidone in various salt solutions (Guner, 1996). However, other polymers, e.g. polyvinylalcohol (Bianchi, Conio, & Ciferri, 1967) and polyacrylamide (Livney et al., 2003), showed mixed responses, i.e. swelling or shrinkage, to the electrolyte-concentration growth depending on the used salts. Hence, the electrolyte influence on neutral polymers is likely a consequence of a complex system of interactions of ions both with the solvation sphere of the macromolecule and with the individual polar groups of the polymer, which prevents a simple generalization of this phenomenon. In our case, separation of the total interaction energy to the contributions of water and ions interactions with the molecule shows that the increasing salt concentration leads not only to a stronger interaction of the chain with ions, but also to a weaker interaction with water. Thus, addition of salt causes restructuring of the solvation shell of hyaluronan which results in the tiny variations of the conformation of a short piece of the chain, but – consequently – in a substantial change of the radii of gyration of the large random coils. This phenomenon has, therefore, an analogous essence as the well-known salting-out effect. However, the salt concentration has only a minor effect on the number of hydrogen bonds. Table S1 shows that the number of hydrogen bonds between the groups of the chain is almost independent of the salt concentration for both hyaluronan and its analog. The number of hydrogen bonds between the chain and the sur-

rounding water is also quite stable, although its moderate decrease of about 12.5% can be observed when NaCl concentration grows from 0 M to 1 M, which corresponds with the decrease of water molecules in the solvation shell of hyaluronan. On the other hand, in MgCl$_2$ solution this effect cannot be seen, probably due to the indirect interaction of the cation with the chain and thus smaller ability to repel water molecules. Hence, it seems that the formation or disruption of hydrogen bonds is not the key interaction forming the shape of hyaluronan and similar molecules.

4. Conclusion

A method of modeling the semirigid polysaccharide molecules was developed and applied on the macromolecules of hyaluronan and its non-charged analog in solutions of two electrolytes, NaCl and MgCl$_2$. The method is based on an equilibrium molecular-dynamics simulation of a 48 monosaccharide-units long molecule and combining randomly selected pieces of this molecule from different simulation frames in order to construct a bigger macromolecule. Using this method sets of random coils, each of a given molecular weight in a given environment, were generated and the averaged radius of gyration was determined for each set. The radius of gyration of the modeled random coils decreases and finally tends to a stable value as long as the MD-simulated oligosaccharide is drawing to equilibrium, indicating that the equilibration of the individual pieces of the macromolecule leads to the equilibration of the whole random coil. The stabilized radii of gyration show a strikingly good agreement with the experimentally observed values as well as the scaling with the molecular weight of the macromolecule. Moreover, the simulated random coils respond to the varying electrolyte concentration in an obvious accordance with the experimental observations – both the radius of gyration and the persistence length of the chain grow when the electrolyte concentration is decreasing, similarly for both hyaluronan and its non-charged analog and for both the salts under consideration. Thanks to this excellent agreement of the model and experiment we can deduce that the main force driving the conformational changes of the macromolecule is the restructuring of the solvation shell of the polymer, i.e. the interactions of water molecules and the cations with the polymer backbone. On the contrary, the mutual repulsion of the carboxylate groups as well as the excluded volume of the chain has only a minor effect on the molecular conformation. Furthermore, the model-experiment agreement supports the hypothesis that the behavior of a selected piece of the chain is equal, no matter whether it exists as a free oligosaccharide or as a part of a long macromolecule. Therefore, the developed method is able to generate all-atom models of the random coils of semirigid polysaccharide molecules on the base of the complete set of their local pair interactions within the macromolecule itself and between the macromolecule and the environment. This feature gives the method the capability to become a versatile tool for the prediction of physical properties of random coils of semirigid polysaccharides.

References

Allison, D. D., & Grande-Allen, K. J. (2006). Review. Hyaluronan: A powerful tissue engineering tool. *Tissue Engineering*, *12*(8), 2131–2140. http://dx.doi.org/10.1089/ten.2006.12.2131

Almond, A., Brass, A., & Sheehan, J. K. (1998). Deducing polymeric structure from aqueous molecular dynamics simulations of oligosaccharides: Predictions from simulations of hyaluronan tetrasaccharides compared with hydrodynamic and X-ray fibre diffraction data. *Journal of Molecular Biology*, *284*(5), 1425–1437. http://dx.doi.org/10.1006/jmbi.1998.2245

Bailey, F. E., & Callard, R. W. (1959). Some properties of poly(ethylene oxide)1 in aqueous solution. *Journal of Applied Polymer Science*, *1*(1), 56–62. http://dx.doi.org/10.1002/app.1959.070010110

Benoit, H., & Doty, P. (1953). Light scattering from non-Gaussian chains. *The Journal of Physical Chemistry*, *57*(9), 958–963. http://dx.doi.org/10.1021/j150510a025

Bianchi, E., Conio, G., & Ciferri, A. (1967). Intrinsic viscosity of poly(vinyl alcohol) in aqueous salt solutions. *The Journal of Physical Chemistry*, *71*(13), 4563–4564. http://dx.doi.org/10.1021/j100872a071

Bjoerling, M., Hersloef-Bjoerling, A., & Stilbs, P. (1995). An NMR self-diffusion study of the interaction between sodium hyaluronate and tetradecyltrimethylammonium bromide. *Macromolecules*, *28*(20), 6970–6975. http://dx.doi.org/10.1021/ma00124a038

Brunchi, C.-E., & Ghimici, L. (2013). PEG in aqueous salt solutions. viscosity and separation ability in a TiO2 suspension. *Revue Roumaine De Chimie*, *58*(2–3), 183–188.

Buhler, E., & Boué, F. (2004). Chain persistence length and structure in hyaluronan solutions: Ionic strength dependence for a model semirigid polyelectrolyte. *Macromolecules*, *37*(4), 1600–1610. http://dx.doi.org/10.1021/ma0215520

Donati, A., Magnani, A., Bonechi, C., Barbucci, R., & Rossi, C. (2001). Solution structure of hyaluronic acid oligomers by experimental and theoretical NMR, and molecular dynamics simulation. *Biopolymers*, *59*(6), 434–445. http://dx.doi.org/10.1002/1097-0282(200111)59:6<434:AID-BIP1048>3.0.CO;2-4

Fouissac, E., Milas, M., Rinaudo, M., & Borsali, R. (1992). Influence of the ionic strength on the dimensions of sodium hyaluronate. *Macromolecules*, *25*(21), 5613–5617. http://dx.doi.org/10.1021/ma00047a009

Furlan, S., La Penna, G., Perico, A., & Cesàro, A. (2004). Conformational dynamics of hyaluronan oligomers in solution. 3. molecular dynamics from monte carlo replica-Exchange simulations and mode-coupling diffusion theory. *Macromolecules*, *37*(16), 6197–6209. http://dx.doi.org/10.1021/ma049641v

Furlan, S., La Penna, G., Perico, A., & Cesàro, A. (2005). Hyaluronan chain conformation and dynamics. *Carbohydrate Research*, *340*(5), 959–970. http://dx.doi.org/10.1016/j.carres.2005.01.030

Grundelova, L., Mracek, A., Kasparkova, V., Minarik, A., & Smolka, P. (2013). The influence of quarternary salt on hyaluronan conformation and particle size in solution. *Carbohydrate Polymers*, *98*(1), 1039–1044. http://dx.doi.org/10.1016/j.carbpol.2013.06.057

Guner, A. (1996). Properties of aqueous salt solutions of polyvinylpyrrolidone.I. Viscosity characteristics. *Journal of Applied Polymer Science*, *62*(5), 785–788. http://dx.doi.org/10.1002/(SICI)1097-4628(19961031)62:5<785:AID-APP10>3.0.CO;2-S

Guvench, O., Greene, S. N., Kamath, G., Brady, J. W., Venable, R. M., Pastor, R. W., & Mackerell, A. D. (2008). Additive empirical force field for hexopyranose monosaccharides. *Journal of Computational Chemistry*, *29*(15), 2543–2564. http://dx.doi.org/10.1002/jcc.21004

Guvench, O., Hatcher, E., Venable, R. M., Pastor, R. W., & MacKerell, A. D. (2009). CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. *Journal of Chemical Theory and Computation*, *5*(9), 2353–2370. http://dx.doi.org/10.1021/ct900242e

Hayashi, K., Tsutsumi, K., Nakajima, F., Norisuye, T., & Teramoto, A. (1995). Chain-stiffness and excluded-volume effects in solutions of sodium hyaluronate at high ionic strength. *Macromolecules*, *28*(11), 3824–3830. http://dx.doi.org/10.1021/ma00115a012

Heeb, R., Lee, S., Venkataraman, N. V., & Spencer, N. D. (2009). Influence of salt on the aqueous lubrication properties of end-grafted, ethylene glycol-based self-assembled monolayers. *Acs Applied Materials & Interfaces*, *1*(5), 1105–1112. http://dx.doi.org/10.1021/am900062h

Hofmeister, F. (1891). Zur Lehre von der Wirkung der Salze. *Archiv für experimentelle Pathologie und Pharmakologie*, *28*(3–4), 210–238. http://dx.doi.org/10.1007/BF01824334

Holmbeck, S. M., Petillo, P. A., & Lerner, L. E. (1994). The solution conformation of hyaluronan: A combined NMR and molecular dynamics study. *Biochemistry*, *33*(47), 14246–14255.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, *14*(1), 33–38, 27–28.

Ivanov, D., & Neamtu, A. (2013). Molecular dynamics evaluation of hyaluronan interactions with dimethylsilanediol in aqueous solution. *Revue Roumaine De Chimie*, *58*(2–3), 229–238.

Jiang, D., Liang, J., & Noble, P. W. (2007). Hyaluronan in tissue injury and repair. *Annual Review of Cell and Developmental Biology*, *23*(1), 435–461. http://dx.doi.org/10.1146/annurev.cellbio.23.090506.123337

Kaufmann, J., Möhle, K., Hofmann, H.-J., & Arnold, K. (1998). Molecular dynamics study of hyaluronic acid in water. *Journal of Molecular Structure: THEOCHEM*, *422*(1-3), 109–121. http://dx.doi.org/10.1016/S0166-1280(97)00084-5

Kirschner, K. N., & Woods, R. J. (2001). Solvent interactions determine carbohydrate conformation. *Proceedings of the National Academy of Sciences*, *98*(19), 10541–10545. http://dx.doi.org/10.1073/pnas.191362798

Kratky, O., & Porod, G. (1949). Rontgenuntersuchung Geloster Fadenmolekule. *Recueil Des Travaux Chimiques Des Pays-Bas-Journal of the Royal Netherlands Chemical Society*, *68*(12), 1106–1122.

Livney, Y. D., Portnaya, I., Faupin, B., Ramon, O., Cohen, Y., Cogan, U., & Mizrahi, S. (2003). Interactions between inorganic salts and polyacrylamide aqueous solutions and gels. *Journal of Polymer Science Part B-Polymer Physics*, *41*(5), 508–519. http://dx.doi.org/10.1002/polb.10406

Lund, M., Vácha, R., & Jungwirth, P. (2008). Specific ion binding to macromolecules: Effects of hydrophobicity and ion pairing. *Langmuir*, *24*(7), 3387–3391. http://dx.doi.org/10.1021/la7034104

Mendichi, R., Soltés, L., & Giacometti Schieroni, A. (2003). Evaluation of radius of gyration and intrinsic viscosity molar mass dependence and stiffness of hyaluronan. *Biomacromolecules*, *4*(6), 1805–1810. http://dx.doi.org/10.1021/bm0342178

Mutter, S. T., Zielinski, F., Cheeseman, J. R., Johannessen, C., Popelier, P. L. A., & Blanch, E. W. (2015). Conformational dynamics of carbohydrates: Raman optical activity of D-glucuronic acid and N-acetyl-D-glucosamine using a combined molecular dynamics and quantum chemical approach. *Physical Chemistry Chemical Physics*, *17*(8), 6016–6027. http://dx.doi.org/10.1039/c4cp05517a

Necas, J., Bartosikova, L., Brauner, P., & Kolar, J. (2008). Hyaluronic acid (hyaluronan): A review. *Veterinarni Medicina*, *53*(8), 397–411.

Nyström, B., Kjøniksen, A.-L., Beheshti, N., Maleki, A., Zhu, K., Knudsen, K. D., . . . & García de la Torre, J. (2010). Characterization of polyelectrolyte features in polysaccharide systems and mucin. *Advances in Colloid and Interface Science*, *158*(1–2), 108–118. http://dx.doi.org/10.1016/j.cis.2009.05.003

Odijk, T., & Houwaart, A. C. (1978). On the theory of the excluded-volume effect of a polyelectrolyte in a 1-1 electrolyte solution. *Journal of Polymer Science: Polymer Physics Edition*, *16*(4), 627–639. http://dx.doi.org/10.1002/pol.1978.180160405

Odijk, T. (1977). Polyelectrolytes near the rod limit. *Journal of Polymer Science: Polymer Physics Edition*, *15*(3), 477–483. http://dx.doi.org/10.1002/pol.1977.180150307

Odijk, T. (1978). Electrostatic persistence length and its relation to a unified theory of polyelectrolytes in solution. *Polymer*, *19*(8), 989–990. http://dx.doi.org/10.1016/0032-3861(78)90212-4

Pereira, C. S., Kony, D., Baron, R., Müller, M., van Gunsteren, W. F., & Hünenberger, P. H. (2006). Conformational and dynamical properties of disaccharides in water: A molecular dynamics study. *Biophysical Journal*, *90*(12), 4337–4344. http://dx.doi.org/10.1529/biophysj.106.081539

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., . . . & Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, *26*(16), 1781–1802. http://dx.doi.org/10.1002/jcc.20289

Skolnick, J., & Fixman, M. (1977). Electrostatic persistence length of a wormlike polyelectrolyte. *Macromolecules*, *10*(5), 944–948. http://dx.doi.org/10.1021/ma60059a011

Sorci, G. A., & Reed, W. F. (2004). Effect of valence and chemical species of added electrolyte on polyelectrolyte conformations and interactions. *Macromolecules*, *37*(2), 554–565. http://dx.doi.org/10.1021/ma035551z

Stern, R. (2003). Devising a pathway for hyaluronan catabolism: Are we there yet? *Glycobiology*, *13*(12), 105R–1115R. http://dx.doi.org/10.1093/glycob/cwg112

Sze, J. H., Brownlie, J. C., & Love, C. A. (2016). Biotechnological production of hyaluronic acid: A mini review. *3 Biotech*, *6*(1) http://dx.doi.org/10.1007/s13205-016-0379-9

Vlachy, N., Jagoda-Cwiklik, B., Vácha, R., Touraud, D., Jungwirth, P., & Kunz, W. (2009). Hofmeister series and specific interactions of charged headgroups with aqueous ions. *Advances in Colloid and Interface Science*, *146*(1–2), 42–47. http://dx.doi.org/10.1016/j.cis.2008.09.010

P2. Effect of solvent and ions on the structure and dynamics of a hyaluronan molecule.
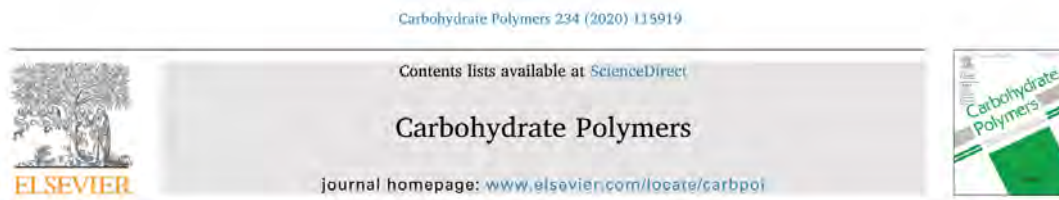
Effect of solvent and ions on the structure and dynamics of a hyaluronan molecule

Eva Kutálková[a], Josef Hrnčiřík[a], Roman Witasek[a], Marek Ingr[a,b,*]

[a] Tomas Bata University in Zlín, Faculty of Technology, Department of Physics and Materials Engineering, Nám. T.G. Masaryka 5555, 76001 Zlín, Czech Republic
[b] Charles University, Faculty of Science, Department of Biochemistry, Hlavova 8/2030, 12843 Praha 2, Czech Republic

ARTICLE INFO

ABSTRACT

Hyaluronic acid (hyaluronan, HA) is a negatively charged polysaccharide forming highly swollen random coils in aqueous solutions. Their size decreases along with growing salt concentration, but the mechanism of this phenomenon remains unclear. We carry out molecular-dynamics simulations of a 48-monosaccharide HA oligomer in varying salt concentration and temperature. They identify the interaction points of Na^+ ions with the HA chain and reveal their influence on the HA solvation-shell structure. The salt-dependent variation of the molecular size does not consist in the distribution of the dihedral angles of the glycosidic connections but is driven by the random flips of individual dihedral angles, which cause the formation of temporary hairpin-like structures effectively shortening the chain. They are induced by the frequency of cation-chain interactions that grow with the salt concentration, but is reduced by the simultaneous decrease of ions' activities. This leads to an anomalous random-coil shrinkage at 0.6 M salt concentration.

1. Introduction

Hyaluronan (HA), a natural polysaccharide of the extracellular matrix of connective tissues, is a linear alternating co-polymer of glucuronic acid (GCU) and N-acetyl glucosamine (NAG) of the formula ([4)-β-D-GlcpA-(1→3)-β-D-GlcpNAc-(1→]$_n$). As a highly hydrophilic polyelectrolyte it forms highly swollen random coils in aqueous solutions the shape of which varies with the changes of the environment. Indeed, the variations of the radius of gyration (R_g) of HA macromolecules were studied by several research groups (Fouissac, Milas, Rinaudo, & Borsali, 1992; Hayashi, Tsutsumi, Nakajima, Norisuye, & Teramoto, 1995; Mendichi, Soltés, & Giacometti Schieroni, 2003) showing that the growing salt concentration causes the decrease of R_g and thus the shrinkage of the molecule. Recently, differences in the HA random-coil size in the solutions of different Hofmeister-series ions was investigated (Musílová, Kašpárková, Mráček, Minařík, & Minařík, 2019). In addition, more detailed structural studies based on NMR were carried out aimed at the determination of the detailed 3D structure of the HA molecule and its dynamics in solution, often supplemented by the methods of computational chemistry. This approach was first used by Holmbeck et al. who determined the interglycosidic dihedral angles and predicted the properties of the helical structure of HA oligosaccharides (Holmbeck, Petillo, & Lerner, 1994). Donati et al. showed that the relative rigidity of the HA chain is supported by the formation

of direct or water-mediated intramolecular hydrogen bonds (Donati, Magnani, Bonechi, Barbucci, & Rossi, 2001). The dynamics of the chain was studied by Furlan, La Penna, Perico, & Cesàro (2005) and by the group of Blundell who used the ^{13}C- and ^{15}N-labeled oligosaccharides to evaluate the flexibility of both the sidechains and the glycosidic connections (Almond, Deangelis, & Blundell, 2006; Almond, DeAngelis, & Blundell, 2005; Blundell, DeAngelis, Day, & Almond, 2004). These studies also confirmed the left-handed 4-fold helical structure of the HA oligosaccharides previously known from the crystal structures.

Recently, HA oligosaccharides modified by aliphatic chains of different lengths were simulated by MD showing an increasing tendency of formation of particles with hydrophobic core when the chain is growing (Payne, Svechkarev, Kyrychenko, & Mohs, 2018). MD was also used to study interactions of HA oligosaccharides with phospholipids (Bełdowski, Mazurkiewicz, Topoliński, & Małek, 2019; Siódmiak et al., 2017) and biomembranes (Smith, Ziolek, Gazzarrini, Owen, & Lorenz, 2019). Using quantum-chemical methods HA oligosaccharides were studied by Pogány and Kovács (Pogany & Kovacs, 2010) who carried out a DFT study of isolated di- to decasaccharides and determined their hydrogen-bonds structure and its influence on the conformation. For a comprehensive survey of NMR and computational studies see the recent reviews (Almond, 2018; Nagarajan, Sankaranarayanan, & Desai, 2019; Pomin, 2014).

In our previous work we presented a method of calculating the

random-coil size from a simulation of a short oligosaccharide (Ingr, Kutálková, & Hrnčiřík, 2017). The values of the radius of gyration obtained this way agreed well with experiment for a wide range of HA molecular weight and for several different values of the salt concentration in the solution. However, the detailed mechanisms by which the dissolved ions influence the size and shape of the random coil remains unclear. No apparent differences in the chain structure under different salt concentration were observed, especially not the changes in dihedral angles of glycosidic connections. Moreover, the common cause of polyelectrolyte-molecule shrinkage in salty solutions – screening of the electrostatic repulsion, seems to be rather insignificant for HA due to the large distances of the charged groups. Hence, in this work we concentrate on the explanation of how the interactions of the HA chain with the solvation shell, especially the ions, are transmitted to the dynamical changes of the chain structure and, subsequently, to the changes of the shape and size of the random coils. The aim of this study is thus to verify the following hypothesis: Ions in solution influence the size and shape of hyaluronan random coils by direct interactions with specific sites on the chain via the dynamic formation and decay of hairpin-like loops.

2. Methods

2.1. Simulation of hyaluronan oligosaccharides

HA oligonucleotides of 48 monosaccharide units (24-mer, as a monomeric units consists of two residues, GCU and NAG) were simulated in 1 M, 0.6 M, 0.2 M and 0 M (polyelectrolyte only neutralized by counterions Na^+) NaCl solution at two different temperatures (275 K, 310 K). The initial dimensions of the water box were $(177 \times 103 \times 172)$ Å^3. Topology files were constructed using a NAMD automatic PSF builder, solvent molecules were represented by the explicit TIP3P solvent model, periodic boundary conditions in the isobaric-isothermal (NPT) ensemble were used. All simulations were performed using NAMD Version 2.10 program package (Phillips et al., 2005) with CHARMM36 carbohydrate topology and force field parameters (Guvench, Hatcher, Venable, Pastor, & MacKerell, 2009; Guvench et al., 2008) which is one of the standard ways used for carbohydrate simulations (Nagarajan et al., 2019). A timestep of 1 fs for bonding and 2 fs for non-bonded interactions and 10 Å cutoff of non-bonded interactions were used. Full electrostatic calculations were performed every sixth fs using the Particle Mesh Ewald method (PME), implemented in the NAMD package. Prior to each MD simulation the energy of systems was minimized for 180 fs. Subsequent simulations were carried out at the constant atmospheric pressure and one of the temperatures mentioned. The pressure was controlled using the Langevin piston Nosé-Hoover method and the temperature was controlled using Langevin dynamics. Interatomic distances and non-bonded interactions were evaluated using VMD 1.9.2 program (Humphrey, Dalke, & Schulten, 1996). In order to reach the thermodynamic equilibrium of the system, the simulation times were between 80 and 100 ns for systems at 310 K temperature about 200 ns for systems at 275 K temperature.

2.2. Cumulative solvation-shell diagrams (CSSD)

CSSD show the projection of all atoms within a certain shape around a given residue to a selected plane. The diagram is a superposition of these projections over a selected number of residues of a given kind (GCU or NAG) and time frames. It is a useful tool to investigate the distribution of individual components of the HA solvation shell. For more details see SI Methods, section SM1.

2.3. Distribution functions of solvent molecules

Distribution functions of solvent molecules show the numbers of

atoms of a given kind along a selected axis in one of the three possible directions summed over the two other coordinates. They are summed over selected residues of a given kind and selected time frames and can be limited to a special part of the space. For more details see SI Methods, section SM2.

2.4. Composition of random coils as a function of dihedral-angle distributions

Composition of random coils as a function of dihedral-angle distributions was carried out using the subsequent connecting of individual residues in accord with the distributions of the dihedral angles of the glycosidic connections obtained from the simulation of a certain system. For more details see SI Methods, section SM3.

3. Results and discussion

3.1. End-to-end distance and persistence length of the HA chain

Previously (Ingr et al., 2017), we proposed a method for calculating the radii of gyration of HA composing a long random coil from fragments of a short oligosaccharide. In that calculation the dihedral angles of the glycosidic connections were inherently present in the chain fragments. For the glycosidic bond connecting two fragments the couple of dihedral angles was chosen on random in accord with the distribution of these angles in the ensemble formed by all the glycosidic connections in all the time frames of the simulation. In this work we investigate the mechanism by which the ions of the electrolyte influence the random coil size. MD simulations of HA oligomer of 48 monosaccharide units were carried out for eight systems varying in temperature and NaCl concentration. The mean distances of residue couples separated by different numbers of monomeric units were evaluated. For every separation the distance was averaged over all possible couples throughout the chain length and the selected number of time frames. Fig. 1 shows that for both the temperatures the end-to-end distance of a given separation is the highest for 0 M NaCl. For 310 K this distance decreases continuously along with the growing salt concentration, as expected from the published results (see e.g. Fig. 1B). For the lower temperature, 275 K, the distance decreases up to the concentration of 0.6 M NaCl but above this value it shows an opposite trend.

Given the curves in Fig. 1A, the persistence length of the chain can be calculated for every studied system numerically solving the formula (Kratky & Porod, 1949)

$$\langle d^2 \rangle = 2l_p L \left[1 - \frac{l_p}{L} \left(1 - e^{-\frac{L}{l_p}} \right) \right]. \tag{1}$$

where d is the end-to-end distance, L the contour length and the l_p persistence length. As can be seen from the lower limit of these curves, the contour length per residue is approximately 4.6 Å. The persistence lengths are shown in Fig. 1B (see the legend) together with the dependencies of radii of gyration on molecular weight for systems of given persistence lengths determined by the equation (Benoit & Doty, 1953)

$$\langle R_g^2 \rangle = l_p \left\{ \frac{L}{3} - l_p + \frac{2l_p^2}{L^2} \left[L - l_p \left(1 - \exp\left(-\frac{L}{l_p}\right) \right) \right] \right\}. \tag{2}$$

Direct comparison of the calculated persistence lengths with experiment can be done for 0 M NaCl – the two experimental values obtained by two different methods at 298 K, 159 Å and 184 Å (Buhler & Boué, 2004), show a good agreement with our results. Furthermore, Fig. 1B shows a good agreement of the calculated R_g values with experiment and indicates that the properties of the simulated oligosaccharide are kept even by a long random coil. Obviously, the flexibility of the chain is strongly influenced by the ions, but the question
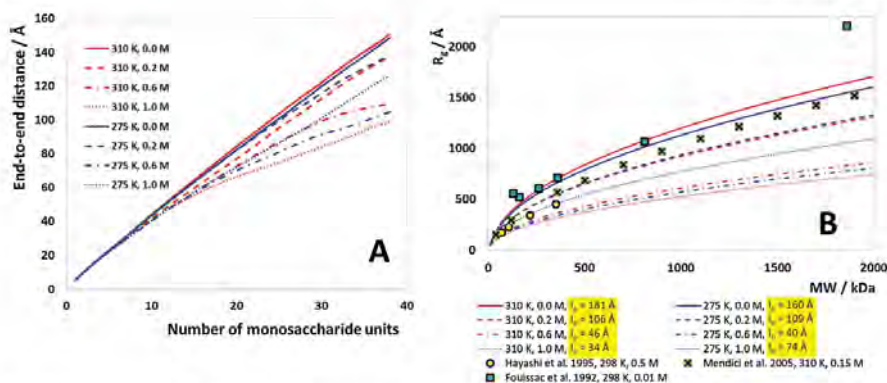
Fig. 1. A. End-to-end distance as a function of the number of monosaccharide units of the oligosaccharide for different temperatures and NaCl concentrations. Every point of each curve is given by the average distance of all possible oligosaccharides of the given length contained in the simulated chain. B. Theoretical dependencies of the radius of gyration on the molecular weight of the HA chain for different conditions (line design corresponds with panel A). R_g was calculated (using eq. 2) from the persistence length of every system (yellow emphasis in the legend) determined by the numerical solution of eq. 1. Selected experimental data are shown for comparison. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

remains by what mechanism they effect the structure and shape of the random coil.

3.2. Distribution of dihedral angles of the glycosidic connections

Every two HA residues are connected by a glycosidic bond characterized by two dihedral angles (defined in Fig. S2). As there are two types of glycosidic connection – 1–3 connection of the reducing end of GCU and non-reducing end of NAG and 1–4 connection for the opposite variant– four dihedral angles have to be considered in order to characterize the HA chain. Obviously, when fixed values of the dihedral angles are taken into account, the resulting chain adopts a shape of a regular helix the diameter and ascent of which depend on the choice of the dihedral values. When the distribution is broadened, the regular helix changes to a random coil and its radius of gyration decreases with the width of the distribution. The distribution of individual dihedral angles is shown in Fig. S4, which indicates that the widths of the major peaks are practically independent of the temperature and salt concentration. When an ensemble of random coils of a certain number of monosaccharide units is generated combining individual residues with the dihedral angles of their glycosidic connection obeying the distribution given solely by the major peaks, the radius of gyration is in a good agreement with the values for the salt-free solution of HA from our previous calculation (Ingr et al., 2017) as well as the experiment (Fig. S3). Mutual orientations of the neighboring residues vary only very slightly from the conformation plotted in Fig. 2A, hereafter referred to as "standard conformation".

A typical appearance of a longer HA molecule is shown in Fig. 3, first panel. Thus, this simple model corresponds with a real system not disturbed by the dissolved ions and shows that at these conditions no additional flexibility of the chain has to be taken into account. In addition, it also proves that the simulation provides a good approximation of the real system.

As the salt concentration increases, minor peaks can be observed in the dihedral-angles distribution which correspond with the flipped glycosidic connections, i.e. connections containing at least one dihedral angle of a value out of the peak corresponding with the standard conformation (see Fig. S4).

If these peaks are added to the dihedrals distribution, the radius of gyration decreases immediately, but the correspondence with experiment is not good – the calculated radii of gyration are generally underestimated (e.g. the R_g value for 0.2 M NaCl, 310 K and 2000
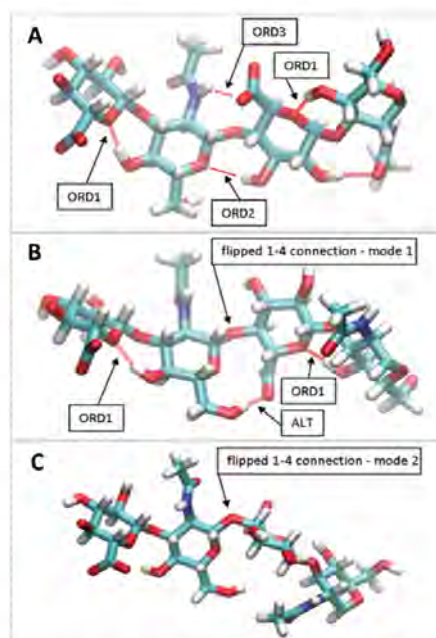
Fig. 2. Conformations of the HA chain. A. Standard conformation of both the 1–3 (left) and 1–4 (right) glycosidic connections of the monosaccharide units with all four dihedral angles in their most typical values. Ordinary hydrogen bonds are depicted by the red lines. B. The most common flipped conformation (mode 1) with the 14_2 dihedral turned to -5°. C. Another flipped conformation with the (mode 2) 14_2 dihedral turned to 42°. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

monosaccharide units, i.e. 378 kDa, is 465 Å, which is significantly lower than expected – see the experiments in Fig. 1B). On one hand, this observation indicates that the increased chain flexibility leads to a
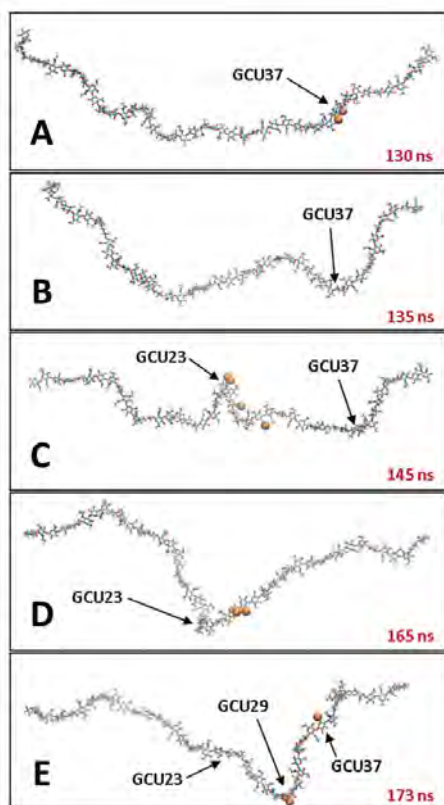
Fig. 3. The development of the HA chain in 1 M NaCl at 275 K during the time course of the simulation. Individual time frames show the rise and vanishing of three hairpin-like structures formed as a consequence of the flip of the 14_2 dihedral angle on the non-reducing end of the GCU units 37, 23 and 29. These structures arise after the disturbance of the standard conformation of the chain by Na$^+$ (orange balls). The lifetime of these structures is about $10-20$ ns.

shrinkage of random coils, but, on the other hand, the influence of the minor peaks seems to be exaggerated. The reason is that, due to the hydrodynamic friction, the rest of the molecule remains almost unchanged immediately after the dihedral flip and only slowly relaxes to the new equilibrium conformation enforced by the flip (for detailed discussion see section 3.3.). On the contrary, the model random coil composed of individual residues by the purely statistical approach takes an unnatural turning whenever a dihedral angle from the minor peak is incorporated into the chain as if the equilibrium conformation after the flip was adopted immediately. This strongly increases the disorder of the coil and thus reduces its radius of gyration. Hence, the random incorporation of the dihedral flips into the chain model does not explain the dependence of the radius of gyration on the salt concentration.

3.3. Chain dynamics

Explanation of the chain-dimension variations can be based on the dynamical changes of the chain resulting from the flips of individual dihedral angles of the glycosidic connections. First of all, Fig. 2B and C shows that an individual flip occurs typically at a single glycosidic-
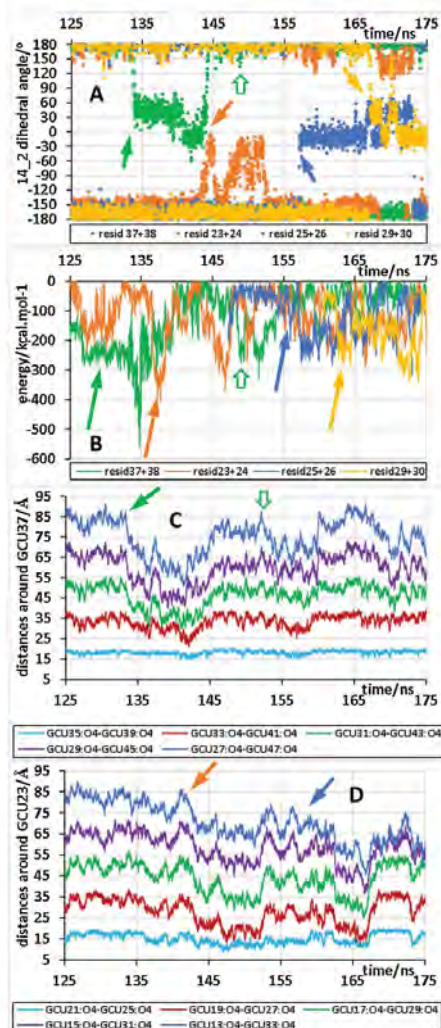
Fig. 4. Time course of the simulation of HA at 1 M NaCl at 275 K. A. Four individual 14_2 dihedral angles between the residues denoted in the legend as functions of time. B. Interaction energy of the residues surrounding the flipping dihedral with Na$^+$ ions – negative energy indicates Na$^+$ in the proximity of the residues. C and D. The distances of the residue couples (indicated in the legend) located symmetrically on both sides of the flipping glycosidic connection (indicated in the graph) with growing distance separating them. Note that the flips cause the shortening of the measured distances – formation of hairpin-like structures. The arrows of a given color indicate the related processes in different panels – the residue flip in A, the preceding Na$^+$-HA interaction in B and the onset of the shortening of the distances between the surrounding residues in C and D. This effect can be, in a lesser extent, caused also by the occurrence of a Na$^+$ ion close to the respective residues eliciting only a small deviation from the standard dihedral value instead of the complete flip (see the green curves in panels A and B and all curves in panel C from 145 ns on, indicated by the hollow green arrows). Note also that the curves in panel C reflects also the flips at the residues 25 as it is close to residue 23 (the center of measurement in this graph). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

4

connection dihedral angle and no other compensation flips occur any-where in the chain. With the exception of the flipped dihedral angle, all the geometric parameters are practically equal to the state before the flip. This can be seen comparing Fig. 2A and B, where the connecting dihedral angles are obviously different, but the conformation of the surrounding chains is conserved to such details that e.g. the in-tramolecular hydrogen bond ORD1 remains unchanged. Therefore, since Fig. 4A shows that the flip itself is a rapid process, in the first moment after its occurrence the chain gets twisted as a ribbon held firmly on one end and turned around its axis on the other end. More-over, the flip elicits a tendency of the chain to a bending given by the new orientation of the residues neighboring with the flip that assume different directions of each side of the chain compared to the un-perturbed chain. Subsequently, the tension in the chain starts to relax and the chain undergoes bending in the point of the flipped connection, therefore the residues on the different sides of this connection get closer to one another and form a loose hairpin-like loop until the chain con-formation reaches a new equilibrium.

As the flipped conformation is a thermodynamically unfavorable state, after a certain time period it undergoes a reverse flip that gen-erates oppositely oriented tension and results in the reverse relaxation to the original chain conformation followed by the growth of the dis-tance between the residues on the opposite sides of the flip. The time dependence of the separating distance of several such residue couples is shown in Fig. 4C and D. Every formation of the hairpin-like loop is preceded by a local interaction of Na$^+$ ions with the respective spot of the chain. Fig. 4B documents these interactions as a decrease of the HA-Na$^+$ interaction energy. It indicates the key role of these interactions for the formation of the loops (the mechanism is described in detail in section 3.4.2.). Fig. 3 in section 3.2. documents the subsequent for-mation of several hairpin-like loops during the course of the simulation. Fig. 3A shows the initial interaction of the chain at GCU37 with the Na$^+$ ions while Fig. 3B shows the formed loop around this residue which later decays at Fig. 3C. This process is also shown in detail in Fig. S5 which indicates that the formation of the hairpin-like loop culmi-nates at 138 ns. Fig. 3C also shows the loop formed around GCU23 with the Na$^+$ ions still present. In Fig. 3D it is already decayed but other Na$^+$ ions are present close to GCU29. These ions initiate the formation of a loop around this residue as can be seen in Fig. 3E. Figs. 3 and S5 show that an average lifetime of the loop is in lower tens of ns. In the HA chain the most often flip occurs on the 1–4 connection, especially the dihedral 14_2. The standard mean value of this dihedral is about –171° and there are two main extraordinary values, –5° and 42°. As can be seen in Fig. 4A, the flipped dihedral often jumps between these two values and the jump from one to another may lead to even deeper decrease of the distance between the surrounding residues. Obviously, the repeated process of a formation and decay of the loops leads to the shortening of the mean end-to-end distance and radius of gyration of the molecule. The described process is thus the cause of the variations in the random-coil size of the HA macromolecule in different environ-ments which support the formation of the hairpin-like loops with dif-ferent intensity.

3.4. Solvent effects on the chain dynamics

The frequency of the dihedral flips is approximately proportional to the NaCl concentration (see the secondary peaks in Fig. S4). The effect of the ions can be mediated either by the influence of the ions on its solvation shell including the formation of hydrogen bonds or by a direct interaction of an ion with the oligosaccharide.

3.4.1. Hydrogen bonds formation

Hyaluronan chain contains various functional groups serving as both donors and acceptors of hydrogen bonds. Therefore, a peculiar network of these bonds is formed within the molecule and many others interconnect the HA molecule and the surrounding solvent. A generally
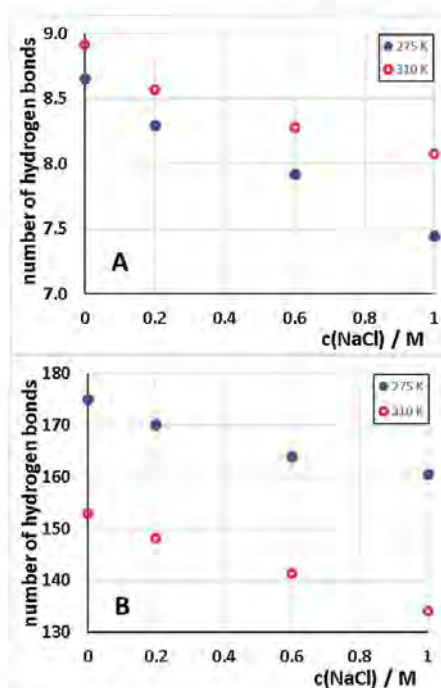
Fig. 5. Numbers of ordinary hydrogen bonds (A) within the simulated HA chain and HA-water hydrogen bonds (B). The error bars smaller than the used sym-bols.

accepted concept describes three main hydrogen bonds within the HA chain which stabilize the ordinary structure of the macromolecule (Fig. 2A, hereafter called "ordinary hydrogen bonds"). These hydrogen bonds are formed by 1. O4 (NAG) – O5 (GCU), 2. O3 (GCU) – O5 (NAG) and 3. N (NAG) – O61 (donor group is written first; for identification of the atom see Fig. S1). The hydrogen bond 1 bridges the 1–3 connection while the hydrogen bonds 2 and 3 bridge the connection 1–4. In ad-dition, several different hydrogen bonds (hereafter called "alternative hydrogen bonds") can be formed occasionally that are capable of a temporary stabilization of different extraordinary structures occurring especially as a consequence of the dihedral-angle flips (e.g. "ALT" in Fig. 2B). Using equilibrium MD simulations we have evaluated the average numbers of hydrogen bonds within the last 36 ns of the simu-lation at two different temperatures and several different NaCl con-centrations. Fig. 5 shows that the total number of intramolecular hy-drogen bonds within the simulated HA oligosaccharide increases along with the temperature.

On the contrary, the number of hydrogen bonds between the mo-lecule and the solvent shows an opposite trend. Both these hydrogen bonds likely contribute to the straightening of the HA chain, the in-tramolecular ones by fixing the standard conformation and the HA-water ones by forcing the oligosaccharide to expose the oxygen con-taining groups to water. These effects are thus counteracting, but the relative decrease of the HA-water hydrogen bonds is higher than the increase of the intramolecular ones (compare the numbers given in the two panels of Fig. 5). Temperature increase may therefore rather de-stabilize the standard conformation. However, in pure water this effect is negligible, but may contribute to the observed conformation changes in the presence of ions. When the salt concentration is increased, the
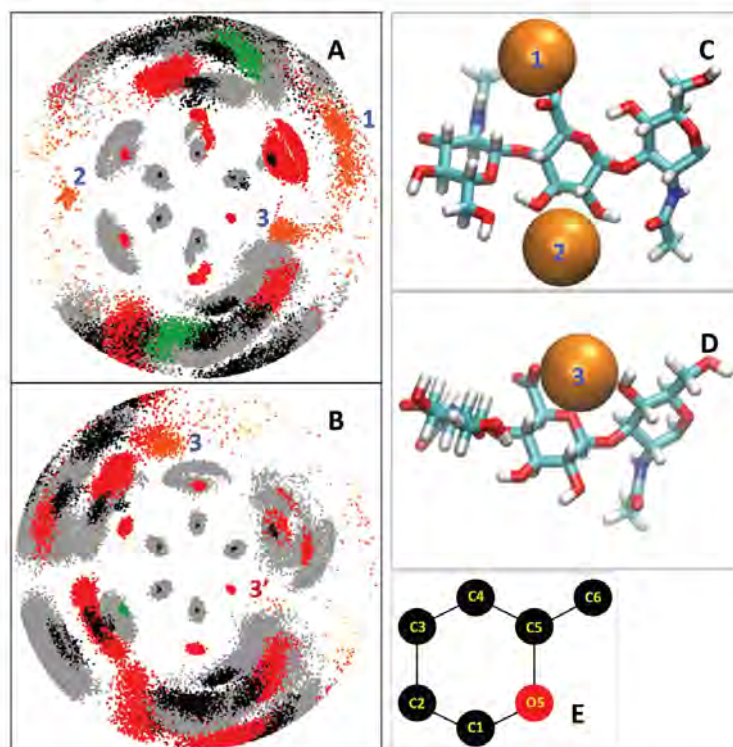
Fig. 6. A and B. Cumulative solvation-shell diagrams of the GCU and NAG residues, respectively. The monosaccharide ring is viewed in the direction of its 3-fold "axis of symmetry". Carbon atoms are shown in black, oxygen in red, hydrogen in grey, nitrogen in green, Na$^+$ in orange and Cl− in yellow. Atoms of water are not shown. The three typical locations of Na$^+$ are indicated: 1. at the carboxylate group of GCU, 2. between the oxygens O2 and O3 of GCU, 3. in the proximity of the oxygen O5 of GCU (labeled also in the NAG diagram). With NAG the Na$^+$ ions interact weakly in a position analogous to 3 – labeled as 3′. C. Structure of a HA chain with two Na$^+$ ions in positions 1 and 2. D. Structure of a HA chain with a Na$^+$ ion in the position 3. E. Orientation scheme for CSSDs. For more detailed analysis of CSSDs see Figs. S6-S8. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

number of hydrogen bonds of any kind decreases at both the temperatures. Hence, the increasing salt concentration may weaken the highly swollen standard configuration by lowering the hydrogen-bond number and thus further support the dynamic phenomena of loop formations (see section 3.3.) by the increased flexibility of the chain.

3.4.2. HA-ions interactions

The solvation shell of a HA molecule was studied using the cumulative solvation-shell diagrams showing the superposition of atoms around individual residue types in a given direction of view and a selected part of the space over the whole simulation run (see section 2.2.) and the distribution functions of individual atoms of the solvent molecules (see section 2.3.). Especially the distribution of ions around the HA chain was of a special interest. Fig. 6 shows specific positions in which Na$^+$ ions are often located. The most frequent position of these ions is in the close proximity of the carboxylate group (position 1).

Another position of Na$^+$ location is between the oxygens 2 and 3 of the GCU unit where the ion is stabilized by the partial charges of these oxygens (position 2). The position 3 is in the proximity of the hemiacetal oxygen O5 of the GCU residue. Here, the ion is attracted by the partial negative charges on this oxygen as well as the oxygen on C6 of the NAG residue and the oxygens of the carboxylate group. The analogous position on the NAG residue (denoted as 3′) is also occasionally occupied by the Na$^+$ ions, but this event is much less frequent than on GCU (compare the number of dots in these positions in Fig. 6A and B as well as in the corresponding panels of Figs. S7 and S8). If the ion gets in this position, the two participating residues turn out of the standard conformation and impose some tension to the neighboring glycosidic connections. Moreover, the ionic interaction competes with the

hydrogen bond (ordinary hydrogen bond ORD1, Fig. 2) formation, which further destabilizes the standard conformation. As a result, the occurrence of Na$^+$ in this position often induces an orientation flip of the neighboring glycosidic connection, especially the closest 14_2 dihedral angle, as shown in Fig. 2. This phenomenon can be directly observed in the simulations and is also documented by the graph in Fig. 4 which shows the interaction energy between the ion and the participating residues simultaneously with the value of the flipped dihedral angle. Obviously, the increased interaction is well correlated with the dihedral flip. As can be seen in Fig. 4, not only the flip, but, with lower intensity, even the occurrence of a Na$^+$ ion in the respective position itself leads to the formation of the hairpin-like structures discussed in section 3.3. and therefore to the shortening of the end-to-end distance. Thus, the presence of Na$^+$ in the position 3 induces the orientation flips of the neighboring residues, which is likely the essence of the mechanism by which the ions influence the random-coil structure.

An ion residing at the proximity of a GCU residue often fluctuates between the positions at the carboxylate group and at O5. In most cases the equilibrium between these two states is shifted towards the carboxylate group, i.e. the ion spends more time in this position than in the other one. However, at the NaCl concentration of 0.6 M the equilibrium seems to be shifted towards the position at O5, which is apparent especially at the lower temperature of 275 K. Figs. S7 and S9 show the distributions of Na$^+$ ions in these two positions and Fig. 7 the equilibrium ratio of these two states

$$K = \frac{n_{fr}\,(pos.\ 3)}{n_{fr}\,(pos.\ 1)}, \tag{3}$$

where n_{fr} denotes the numbers of events when an ion resides in the
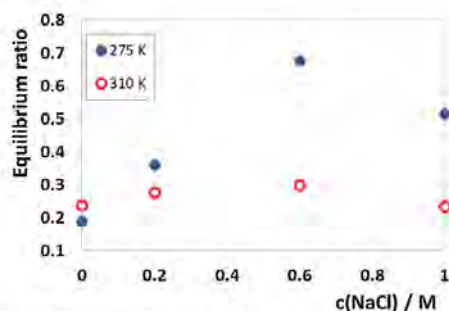
6

80

Fig. 7. Equilibrium ratio of ion binding to position 3 (at oxygen O5) vs. position 1 (at carboxylate group) as a function of NaCl concentration. Only the values for non-zero concentrations are derived from the simulation, the values for c (NaCl) = 0 is obtained by extrapolation.

respective position in a given time frame of the simulation.

Fig. S10 shows the total time spent by the ions in position 3 and the mean lifetime spent there during one event of ion binding to this position. Data of all these plots correlate with the reversed order of the end-to-end distances for different NaCl concentrations including the anomaly at 275 K, where the distance for 0.6 M is apparently lower than for 1.0 M. Similar reversed salt influence at a given concentration, known as the Hofmeister series reversal, was observed experimentally for various molecules and biomolecules (Jordan, Gibb, Wishard, Pham, & Gibb, 2018; Salis, Cugia, Parsons, Ninham, & Monduzzi, 2012; Zhang & Cremer, 2009). This behavior may be a consequence of two counteracting quantities – the growing NaCl concentration and the decreasing activity of the ions (Hamer & Wu, 1972) and charged groups along with it. Hence, the ability of constituting the carboxylate-Na$^+$ couples is likely the lowest at the NaCl concentration close to 0.6 M which results in a prolonged lifetime of the ion at O5 and an increased probability of the dihedral flip leading to the shortening of the end-to-end distance. As this shortening correlates with the occurrence of Na$^+$ in the position 3, the anomalous behavior even further supports the hypothesis of the key of the HA-Na$^+$ interaction in this position as the triggering agent of the dynamic changes leading to the random-coil shrinkage.

In summary, the increased NaCl concentration affects the conformation of a HA molecule in two ways, both of which lead to the disruption of the standard conformation. First, the number of hydrogen bonds between HA and water decreases which disturbs the standard conformation by the destabilization of the solvation shell of the molecule. Second, the increased number of Na$^+$ ions causes their more frequent occurrence in the proximity of charged and polar groups of the HA chain, especially at the proximity of the O5 oxygen (of GCU, but in a lesser extent also for NAG) that plays a key role in the formation of the temporary hairpin-like loops that are essential for the fluctuations of the macromolecular dimensions.

4. Conclusions

A hyaluronan macromolecule is a highly dynamic system the structure of which reflects the surrounding environment. In pure water the chain is more rigid which leads to larger, i.e. more swollen random coils, while in the growing concentration of NaCl the size of the coil shrinks. However, the major peaks of the distributions of the four dihedral angles of both the glycosidic connections of the alternating monosaccharide units show negligible changes in the varying salt concentration and the random coils composed in accord with them are close to the experimental values for HA in pure water irrespective of the salt concentration in the simulated system. The coil-size variations thus

do not consist in the differences of the dihedrals distributions, but rather in the structural fluctuations caused by the occasional temporary flips of individual dihedral angles to different orientations or at least partial deflections from the standard orientation. These flips are correlated with the presence of ions, especially Na$^+$, which interact with the chain in several specific positions but have also an indirect influence on the solvation sphere of the molecule resulting in a lower number of hydrogen bonds. The flip occurs in just one glycosidic connection of the residues and thus brings considerable tension to the chain. Subsequent relaxation of this tension provokes structural changes leading to hairpin-like loops. Their presence in the chain shortens its end-to-end distance as well as the radius of gyration. The lifetime of the flips is in the orders of units to tens of ns after which the glycosidic connection returns to the standard orientation and the hairpin-like loop relaxes to the standard, i.e. straightened, chain conformation. Hence, although these structural variations appear and disappear dynamically, their presence in the chain makes the random coil more compact. The trend of random-coil shrinkage with the growing salt concentration shows an anomaly at the concentration of 0.6 M where the size of the coil is below the general trend, which is especially evident at low temperature. This is likely caused by the lowered occurrence of the Na$^+$ in the proximity of the carboxylate group that might be a consequence of the optimum combination of the growing salt concentration and the simultaneous decrease of activities of both the charged entities. This destabilizes the interaction between the carboxylate group and the Na$^+$ ion which has therefore a higher tendency to interact with the HA chain in the proximity of the oxygen O5 where it can provoke the conformation changes.

Although the presented approach is able to reproduce the experimentally observed characteristics of the HA chain and to give a plausible explanation of its behavior, it might be interesting to include the interaction of distant parts of the chain in the model. Simulating the two-chains interaction will be the subject of the forthcoming study. Furthermore, studies of the HA chain in mixed water-organic solvents might be another interesting perspective of this project.

CRediT authorship contribution statement

Eva Kutálková: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Josef Hrnčiřík:** Conceptualization, Formal analysis, Methodology, Validation, Writing - original draft, Writing - review & editing. **Roman Witasek:** Formal analysis, Investigation, Methodology, Software, Validation, Visualization. **Marek Ingr:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Writing - original draft, Writing - review & editing.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.carbpol.2020.115919.

References

Almond, A. (2018). Multiscale modeling of glycosaminoglycan structure and dynamics: current methods and challenges. *Current Opinion in Structural Biology, 50*, 58–64. https://doi.org/10.1016/j.sbi.2017.11.008.

Almond, A., DeAngelis, P. L., & Blundell, C. D. (2005). Dynamics of hyaluronan oligosaccharides revealed by 15N relaxation. *Journal of the American Chemical Society, 127*(4), 1086–1087. https://doi.org/10.1021/ja043526i.

Almond, A., Deangelis, P. L., & Blundell, C. D. (2006). Hyaluronan: the local solution conformation determined by NMR and computer modeling is close to a contracted left-handed 4-fold helix. *Journal of Molecular Biology, 358*(5), 1256–1269. https://doi.org/10.1016/j.jmb.2006.02.077.

Bełdowski, P., Mazurkiewicz, A., Topoliński, T., & Małek, T. (2019). Hydrogen and water bonding between glycosaminoglycans and phospholipids in the synovial fluid: Molecular dynamics study. *Materials, 12*(13), 2060. https://doi.org/10.3390/ma12132060.

Benoit, H., & Doty, P. (1953). Light scattering from non-gaussian chains. *The Journal of Physical Chemistry, 57*(9), 958–963. https://doi.org/10.1021/j150510a025.

Blundell, C. D., DeAngelis, P. L., Day, A. J., & Almond, A. (2004). Use of 15N-NMR to resolve molecular details in isotopically-enriched carbohydrates: sequence-specific observations in hyaluronan oligomers up to decasaccharides. *Glycobiology, 14*(11), 999–1009. https://doi.org/10.1093/glycob/cwh117.

Buhler, E., & Boué, F. (2004). Chain persistence length and structure in hyaluronan solutions: Ionic strength dependence for a model semirigid polyelectrolyte. *Macromolecules, 37*(4), 1600–1610. https://doi.org/10.1021/ma0215520.

Donati, A., Magnani, A., Bonechi, C., Barbucci, R., & Rossi, C. (2001). Solution structure of hyaluronic acid oligomers by experimental and theoretical NMR, and molecular dynamics simulation. *Biopolymers, 59*(6), 434–445. https://doi.org/10.1002/1097-0282(200111)59:6<434::AID-BIP1048>3.0.CO;2-4.

Fouissac, E., Milas, M., Rinaudo, M., & Borsali, R. (1992). Influence of the ionic strength on the dimensions of sodium hyaluronate. *Macromolecules, 25*(21), 5613–5617. https://doi.org/10.1021/ma00047a009.

Furlan, S., La Penna, G., Perico, A., & Cesàro, A. (2005). Hyaluronan chain conformation and dynamics. *Carbohydrate Research, 340*(5), 959–970. https://doi.org/10.1016/j.carres.2005.01.030.

Guvench, O., Greene, S. N., Kamath, G., Brady, J. W., Venable, R. M., Pastor, R. W., et al. (2008). Additive empirical force field for hexopyranose monosaccharides. *Journal of Computational Chemistry, 29*(15), 2543–2564. https://doi.org/10.1002/jcc.21004.

Guvench, O., Hatcher, E., Venable, R. M., Pastor, R. W., & MacKerell, A. D. (2009). CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *Journal of Chemical Theory and Computation, 5*(9), 2353–2370. https://doi.org/10.1021/ct900242e.

Hamer, W. J., & Wu, Y. (1972). Osmotic coefficients and mean activity coefficients of uni-univalent electrolytes in water at 25°C. *Journal of Physical and Chemical Reference Data, 1*(4), 1047–1100. https://doi.org/10.1063/1.3253108.

Hayashi, K., Tsutsumi, K., Nakajima, F., Norisuye, T., & Teramoto, A. (1995). Chain-stiffness and excluded-volume effects in solutions of sodium hyaluronate at high ionic strength. *Macromolecules, 28*(11), 3824–3830. https://doi.org/10.1021/ma00115a012.

Holmbeck, S. M., Petillo, P. A., & Lerner, L. E. (1994). The solution conformation of hyaluronan: a combined NMR and molecular dynamics study. *Biochemistry, 33*(47), 14246–14255.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics, 14*(1), 33–38 27–28.

Ingr, M., Kutálková, E., & Hrnčiřík, J. (2017). Hyaluronan random coils in electrolyte solutions-a molecular dynamics study. *Carbohydrate Polymers, 170*, 289–295. https://doi.org/10.1016/j.carbpol.2017.04.054.

Jordan, J. H., Gibb, C. L. D., Wishard, A., Pham, T., & Gibb, B. C. (2018). Ion–hydrocarbon and/or ion–ion interactions: direct and reverse hofmeister effects in a synthetic host. *Journal of the American Chemical Society, 140*(11), 4092–4099. https://doi.org/10.1021/jacs.8b00196.

Kratky, O., & Porod, G. (1949). Rontgenuntersuchung Geloster Fadenmolekule. *Recueil Des Travaux Chimiques Des Pays-Bas-Journal of the Royal Netherlands Chemical Society, 68*(12), 1106–1122.

Mendichi, R., Soltés, L., & Giacometti Schieroni, A. (2003). Evaluation of radius of gyration and intrinsic viscosity molar mass dependence and stiffness of hyaluronan. *Biomacromolecules, 4*(6), 1805–1810. https://doi.org/10.1021/bm0342178.

Musilová, L., Kašpárková, V., Mráček, A., Minařík, A., & Minařík, M. (2019). The behaviour of hyaluronan solutions in the presence of Hofmeister ions: A light scattering, viscometry and surface tension study. *Carbohydrate Polymers, 212*, 395–402. https://doi.org/10.1016/j.carbpol.2019.02.032.

Nagarajan, B., Sankaranarayanan, N. V., & Desai, U. R. (2019). Perspective on computational simulations of glycosaminoglycans. *Wiley Interdisciplinary Reviews-Computational Molecular Science, 9*(2), e1388. https://doi.org/10.1002/wcms.1388.

Payne, W. M., Svechkarev, D., Kyrychenko, A., & Mohs, A. M. (2018). The role of hydrophobic modification on hyaluronic acid dynamics and self-assembly. *Carbohydrate Polymers, 182*, 132–141. https://doi.org/10.1016/j.carbpol.2017.10.054.

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry, 26*(16), 1781–1802. https://doi.org/10.1002/jcc.20289.

Pogany, P., & Kovacs, A. (2010). Theoretical study of hyaluronan oligosaccharides. *Structural Chemistry, 21*(6), 1185–1194. https://doi.org/10.1007/s11224-010-9658-y.

Pomin, V. H. (2014). NMR-based dynamics of free glycosaminoglycans in solution. *Analyst, 139*(15), 3656–3665. https://doi.org/10.1039/C4AN00531G.

Salis, A., Cugia, F., Parsons, D. F., Ninham, B. W., & Monduzzi, M. (2012). Hofmeister series reversal for lysozyme by change in pH and salt concentration: insights from electrophoretic mobility measurements. *Journal of the Chemical Society Faraday Transactions, 14*(13), 4343–4346. https://doi.org/10.1039/C2CP40150A.

Siódmiak, J., Bełdowski, P., Augé, W. K., Ledziński, D., Śmigiel, S., & Gadomski, A. (2017). Molecular dynamic analysis of hyaluronic acid and phospholipid interaction in tribological surgical adjuvant design for osteoarthritis. *Molecules (Basel, Switzerland), 22*(9), https://doi.org/10.3390/molecules22091436.

Smith, P., Ziolek, R. M., Gazzarrini, E., Owen, D. M., & Lorenz, C. D. (2019). On the interaction of hyaluronic acid with synovial fluid lipid membranes. *Journal of the Chemical Society Faraday Transactions, 21*(19), 9845–9857. https://doi.org/10.1039/C9CP01532A.

Zhang, Y., & Cremer, P. S. (2009). The inverse and direct Hofmeister series for lysozyme. *Proceedings of the National Academy of Sciences, 106*(36), 15249–15253. https://doi.org/10.1073/pnas.0907616106.

P3. The rate and evenness of the substitutions on hyaluronan grafted by dodecanoic acid influenced by the mixed-solvent composition.

E. Kutálková, J. Hrnčiřík, R. Witasek, M. Ingr, G. Huerta-Ángeles, M. Hermannová, V. Velebný, Int. J. Biol. Macromol. 189 (2021) 826–836.

The rate and evenness of the substitutions on hyaluronan grafted by dodecanoic acid influenced by the mixed-solvent composition

Eva Kutálková [a], Josef Hrnčiřík [a], Roman Witasek [a], Marek Ingr [a,b,*], Gloria Huerta-Ángeles [c], Martina Hermannová [c], Vladimír Velebný [c]

[a] Tomas Bata University in Zlín, Faculty of Technology, Department of Physics and Materials Engineering, Nám. T.G. Masaryka 5555, 76001 Zlín, Czech Republic
[b] Charles University, Faculty of Science, Department of Biochemistry, Hlavova 8/2030, 12843 Praha 2, Czech Republic
[c] Contipro a.s., Dolní Dobrouč 401, 561 02 Dolní Dobrouč, Czech Republic

ABSTRACT

In this work, low molecular weight (17 kDa) hyaluronan was modified by dodecanoyl substituents. The activation of dodecanoic acid was mediated by benzoyl chloride towards the preparation of a mixed anhydride, which reacts in a second step with HA in water mixed with an organic solvent. The effect of the cosolvent was studied and showed an even distribution of substituents and higher reaction rate in water: 1,4-dioxane compared to water:tert-butanol where substituents occupy adjacent positions. The chemical characterization of the prepared derivatives was elucidated by NMR, FTIR spectroscopy, thermal analyses, and gas chromatography, while the distribution of substituents was evaluated by enzymatic degradation. Molecular-dynamics simulations reveal opposite solvent separations around HA and dodecanoyl chains, that is stronger in water:tert-butanol solution. The resulting incompatibility of solvation-shells of the two entities repels the reaction intermediates from the HA chain and drives them towards the already bound substituents, explaining the observed differences in the distribution evenness. Thus, the influence of the solvent on the reaction selectivity is observed by shielding reactive sites around HA. Therefore, a control of the distribution of the substituents was obtained by defining the concentration of HA and used cosolvent.

1. Introduction

Hyaluronan or hyaluronic acid (HA) is a naturally occurring polymer consisting of D-glucuronic acid (GlcA) and N-acetyl-D-glucosamine (GlcNAc) units linked by alternating "-(1 → 4)- and "-(1 → 3)- glycosidic bonds. Its highly hydrophilic nature, caused by the Gibbs-free-energy lowering by constituting the HA-water interactions on account of interactions between different parts of the HA chain, results in strong swelling of the HA random coils in aqueous environment. Numerous experimental works [1 3] described HA random coils along with the determination of their radii of gyration R_g, which corresponded with our previous molecular-dynamics (MD) study [4]. The tertiary structures of oligo HA and higher size chains are thermodynamically unstable in aqueous solution and are considered inexistent [5,6], although some authors declared their identification [7,8]. Intramolecular hydrogen bonds stabilize the semi-rigid nature of HA chains. Still, HA-water hydrogen bonds anchor the molecule to a well-structured solvation shell, protecting the molecule from intermolecular interactions [9]. Thus, HA is soluble in water, inert to interactions with other molecules and interacts with the binding sites of protein receptors (hyaladherins) [10].

The chemical modification of HA i.e. hydrophobization increases its shelf life and decrease its solubility in water. Hydrophobized high molecular weight HA has been used in ophthalmology [11,12], while low molecular weight modified HA is processed for drug delivery [13,14]. The hydrophobization of HA is usually performed in solvents such as dimethylsulfoxide (DMSO) or formamide for grafting long alkyl chain amines [15,16] or polymers to HA [17] and uses anhydrous conditions at high temperature. However, these conditions make the process expensive and inconvenient for upscaling as HA had to be converted to either acid form or tetrabutylammonium salt (TBAHA) to become soluble in the organic solvent. This approach increases the chain fragmentation of HA, decreases the yield, and provides lower conversion due to harsh reaction conditions. For example, Wang and collaborators [18]

reported the degradation of the molecular weight from 1.33×10^5 to 0.874×10^5 after the conversion to HATBA. Besides, Khetan and Burdick [19] mentioned that the purified yield of the modified HA was only 65% (based on moles) of HA present in the TBAHA. Lower yield was reported also by Zerobin et al. [20]. To overcome these problems, the chemical modification of HA is carried out in a mixture of organic solvents miscible with water, i.e. water:DMF [21], water:isopropanol [22] or water:DMSO [23]. These solvents may also bring some advantage for HA processing [24]. Despite the results of these studies, to our knowledge no systematic MD study of HA in any mixed water-organic solvent has been done yet.

A recent study shows the chemical HA modification by oleoyl residues using the mixed-anhydride method in water:1,4-dioxane and water:tert-butanol (2-methyl-2-propanol) mixed solvents [25]. The solvent composition influences not only the degree of substitution, but also the regioselectivity of the reaction. Therefore, this work was conducted to investigate the analogous esterification reaction with dodecanoyl residues to extend the versatility of the solvent-controlled approach for medium-length fatty acids. To the best of our knowledge this is the first time MD simulations are performed with HA in mixed solvents to explain how the mixed-solvent composition controls the efficiency and evenness of the esterification reaction.

2. Materials and methods

2.1. Materials

Low molecular weight sodium hyaluronate ($M_w = 17,000$ g/mol, $M_w/M_n = 1.50$) was provided by Contipro a.s., Dolní Dobrouč, Czech Republic. The weight-average molecular weight (M_w) and polydispersity (M_w/M_n) were determined by SEC-MALLS before the chemical modification as described previously [26,27]. Dodecanoic acid (C12, $\geq$99.9%), 4-(dimethylamino)pyridine (DMAP, $\geq$99%) and tert-butanol (99.9%) were obtained from Merck. Analytical grade CHROMASOLV® solvents were used for the chromatographic analysis. Deuterium oxide (D_2O), deuterated chloroform ($CDCl_3$) and DMSO-D6 were purchased from CortecNet (France). 1,4-Dioxane (99%), isopropanol (IPA, 99%), and sodium chloride (NaCl, 98%) were obtained from Lachner (Czech Republic). Benzoyl chloride (BC $\geq$ 99.0%) was obtained from Sigma-Aldrich. Triethylamine (TEA, $\geq$99.0%) was obtained from Penta, Czech Republic.

2.2. Synthesis of sodium dodecanoyl hyaluronate (HA-C12)

In the reaction flask, sodium hyaluronate (5 g, 12.5 mmol) was dissolved in distilled water (50 mL). To that solution, 25 mL of the chosen organic solvent (tert-butanol or 1,4-dioxane) was slowly added. After the solution was homogeneous, TEA (3.8 mL, 37.5 mmol) and DMAP (0.076 g, 0.63 mmol) were added, then the mixture was stirred until a homogeneous solution was obtained. In a second flask, dodecanoic acid activation was carried out. Dodecanoic acid (3.255 g, 16.3 mmol) was dissolved in 20 mL of the organic solvent used in flask 1. After that, TEA (3.8 mL, 37.5 mmol) was added, followed by 2.28 g of benzoyl chloride (1.88 mL, 16.3 mmol). The formation of the mixed anhydride was carried out for 30 min at 25 °C. Then, this solution was slowly poured to the flask containing HA. The flask 2 was washed out with 5 mL of the chosen organic solvent. The acylation reaction proceeded for 2 h at room temperature under vigorous stirring. After 2 h, the reaction was stopped by adding a saturated solution of NaCl. In the case of lower concentration (the described solution contains 5% (w/v) HA considering only HA and the mixed solvent — this quantification is used throughout the whole text) only the total volume of the reaction was multiplied (e.g. for 0.5% 10 times), keeping the molar equivalents of the reaction constant. The reaction product was precipitated with an excess of IPA (500 mL). The product was washed several times with solutions of IPA:water (85% v/v, 4 × 500 mL). Finally, the precipitate

was washed two more times with 500 mL of IPA. The white powder was decanted, dried in an oven at 40 °C for 24 h.

2.3. Determination of the degree of substitution by gas chromatography

The degree of substitution (DS_{GC}), i.e. the content of esters moieties (acyls), was measured by GC after alkaline hydrolysis of the sample as described recently [28]. The detailed description of the method is given in the Supplementary information, Section S4. The DS_{GC} means the extent of HA modification by hydrophobic side groups and is calculated according to Eq. (1):

$$Fl_{MD} = \frac{m_{FA,} - m_{FA,free}}{m_{sample}} \frac{M_{SA}}{M_{HA}} \times \beta\Delta\Delta) \tag{1}$$

where $m_{FA,tot}$ is the amount of fatty acid determined in the derivative after alkaline hydrolysis, $m_{FA,free}$ is the amount of fatty acid unbound to HA chain, m_{sample} is the weight of derivative, M_{HA} is molar mass of HA disaccharide (401 g/mol for sodium salt), and M_{FA} is molar mass of fatty acid.

2.4. Liquid chromatography coupled to mass spectrometry

Sample (3–5 mg) was dissolved in a mixture of 0.1 M sodium phosphate buffer pH 5.0 and isopropanol (1:1, v/v) at the concentration 2 mg/mL. After complete dissolution, the sample was diluted with 0.1 M sodium phosphate buffer to a final concentration of the derivative and isopropanol 1 mg/mL and 25% (v/v), respectively. Then, to an aliquot (1 mL) a solution of hyaluronan lyase from *Streptococcus pneumoniae* (SpHyl, 25 μL, approx. 6500 IU/mL) was added and incubated at 37 °C for 2 h. The sample was diluted to 20 μg mL^{-1} with 30% methanol in water (v/v) and analyzed by LC-MS. An Acquity UPLC I-class chromatographic system connected to Synapt G2-Si mass spectrometer (Waters, Manchester, UK) was used for structural characterization of HA-C12. The separation was performed on Waters Acquity UPLC Protein BEH C4 column (1.7 μm, 300 Å, 150 × 2.1 mm) at 40 °C. The chromatographic eluent consisted of 0.1% formic acid in water and acetonitrile. Flow rate was set at 0.4 mL min^{-1} and the following gradient of acetonitrile was applied: 0 min 30%, 16.5 min 95%, 17 min 95%, 17.1 min 30%. The injection volume was 5 μL.

The mass spectrometer was equipped with an electrospray ionization source operating in negative ion mode. Sodium formate (0.5 mmol L^{-1} in water/isopropanol = 10/90, v/v) was used for instrument calibration in the m/z range 50–1850. Instrumental parameters for C12-modified oligosaccharide analysis were set as follows (unless otherwise stated): spray capillary voltage 2.7 kV, source temperature 100 °C, sampling cone voltage 140 V, source offset 80 V, desolvation temperature 500 °C, cone gas flow (N_2) 50 L h^{-1}, desolvation gas flow (N_2) 850 L h^{-1}, nebulizer gas flow (N_2) 6.0 bar, trap collision voltage 4.0 V, transfer collision voltage 2.0 V, trap gas flow (Ar) 2.0 mL min^{-1}, helium cell gas flow (He) 180.0 mL min^{-1}, ion mobility spectrometry gas (drift gas) flow (N_2) 90 mL min^{-1} and wave height 40.0 V. Ion mobility wave velocity was set at 500 m s^{-1}. Leucine enkephalin (200 ng mL^{-1}) at flow rate 5 μL min^{-1} was used as a lock mass. Mass chromatograms belonging to unsaturated di-, tri- and tetrasaccharides modified with 1, 2 and 3 acyl chains (m/z 939.3822 for ΔHA4-1 × C12; m/z 1121.5492 for ΔHA4-2 × C12; m/z 1303.7163 for ΔHA4-3 × C12; m/z 560.2707 for ΔHA2-1 × C12; m/z 736.3031 for ΔHA3-1 × C12) were extracted and integrated. The extraction window was 0.02 amu. The percentage of each modified oligosaccharide in the mixture was calculated by Eq. (2).

$$(\Delta HAx - i \times C12) = 100 \times \frac{A(\Delta HAx - i \times C12)}{\sum_{i=1}^{3} A(\Delta HAx - i \times C12)} \tag{2}$$

$w(\Delta HAx-i \times C12)$ is the percentage of peak area belonging to one specific oligosaccharide in the sum of peak areas belonging to all other

detected oligosaccharides. _HAx can be _HA4, _HA2 and _HA3. The amount of other unsaturated modified oligosaccharides (mainly _HA2-1 × C12, _HA3-1 × C12) was low (≤5%). Therefore, they are included neither in the results section nor in Eq. (2). Acquired data were processed with MassLynx 4.1 software (Waters, Milford, MA, USA).

2.5. Molecular-dynamics (MD) simulations

All MD simulations were carried out by NAMD 2.10 program package [29]. CHARMM 36 force field containing the parts for saccharide molecules [30] and fatty acids [31] was used for the both HA molecule and the saturated aliphatic chains, CGenFF topology and parameter files were used for the 1,4-dioxane and tert-butanol molecules, TIP3P model of water was applied. The force field of N-dodecanoyl-4-(dimethylamino)-pyridinium cation (C_{12}-DMAP$^+$) and triethylammonium (TEA$^+$) ions were constructed using CHARMM-GUI Ligand Reader & Modeler [32]. This instrument was also used to obtain the ester-bond parameters between HA (C6 carbon of GlcNAc) and the 12-carbon aliphatic side-chain (C_{12}), while the topology of this link was adopted from an analogously acetylated HA molecule constructed by the CHARMM-GUI Glycan Modeler [33] (none part of the CHARMM-GUI package can process the whole substituted HA molecule). Two configurations of substituted HA molecules were considered   with the substituent at the end or in the middle of the chain, i.e. on the GlcNAc residue number 1 or 11 (out of 20 residues), respectively, from the non-reducing end. Formulas of all the simulated molecules are given in Fig. S11.

The preparation of the simulated systems started by the equilibration of the boxes of pure mixed solvent. HA was then wrapped by these boxes to form the complete simulation box of about 100 Å edge. The energy of each system was minimized for 5400 fs prior to the MD simulation. Integration was performed by the Verlet-I/r-RESPA MTS method with the slow-force mollification, a timestep of 1 fs for bonding and 2 fs for non-bonding interactions and 10 Å cutoff of non-bonding interactions was used. Full electrostatic calculations were performed every 6th fs using the Particle Mesh Ewald (PME) method. Simulations were carried out in NPT system at the constant pressure of 1 atm and selected temperature for 200 ns. The pressure was controlled using the Langevin piston Nosé-Hoover method and the temperature was controlled using Langevin dynamics.

HA oligosaccharides of 20 monosaccharide units were simulated in 7 different media: pure water and three compositions of water:1,4-dioxane and water:tert-butanol solutions. In both cases, the volume to volume ratios were defined as 2:1, 1:1 and 1:2 (organic-component volume fractions 0.33, 0.50, 0.67), at 310 K and 277 K. Every simulation contains 10 Na$^+$ ions compensating the negative charge of HA. Other HA-oligosaccharides simulations were performed in the presence of 5 C_{12}-DMAP$^+$ and 5 TEA$^-$ ions in the mixed solvents of water:1,4-dioxane and water:tert-butanol of organic-component volume fraction 0.5 at 277 K and 310 K. C_{12}-substituted HA oligosaccharides were simulated in the presence of either 10 Na$^+$ or 10 C_{12}-DMAP$^-$ ions in both solvents. For composition details of the systems see Table S3.

The simulation results were visualized, and specific analyses, especially the determinations of radial distribution functions (RDFs), were carried out using the VMD program [34]. The number of atoms of a given molecule within a distance of 4 Å from a selected entity, HA or the aliphatic chain, hereinafter called number of close atoms, was evaluated as a descriptor of the closest part of the solvation shell of the entity. In case of the C_{12}-DMAP$^+$ reactants number of whole molecules having at least one atom in equally defined area was evaluated (hereinafter called number of close molecules). The output quantities were averaged within the equilibrated time interval 50 200 ns and the standard deviations of the mean were determined (for all the quantities it is lower than the symbols, therefore it is not shown in the graphs).

3. Results

3.1. Synthesis of dodecanoic acid grafted to hyaluronan (HA-C12)

The activation of dodecanoic acid is mediated by benzoyl chloride (BC) and yields a mixed aliphatic-aromatic anhydride in organic solvents miscible with water, i.e., tetrahydrofuran, 1,4-dioxane, isopropyl alcohol or tert-butanol. The mixed anhydride reacts in the next step with the hydroxyl moieties of HA, yielding esterified HA (Scheme 1). However, the solvent mixture drastically influences the selectivity of the reaction towards the substitution at the primary and/or secondary hydroxyl moieties. While HA acylation carried out in DMSO was reported to be more selective for short fatty acids [35], the selectivity for medium fatty acids is still unknown.

Several factors influence the degree of chemical modification and the distribution of substituents: polarity of the reaction media, the molar ratio of reagents, and the hydrophobicity of the fatty acid (ligand), the M_w of HA used for the chemical modification and the concentration of the reaction feed components.

Even though water-insoluble high molecular weight dodecanoyl derivatives of hyaluronan can be used to prepare thin films [36]. It is necessary to develop water-soluble derivatives for drug delivery [37]. It is worth mentioning that the structural characterization of derivatives is preferred by using low molecular weight HA due to the higher reaction efficiency and solubility of the conjugates. The results showed that the polymer concentration in the reaction feed drastically changes the conjugate's solubility probably due to self-aggregation (Table 1). The derivatives prepared at low concentration (0.5 1.5% of HA in the reaction feed) were fully water soluble (lower turbidity and higher transmittance, T ≥ 65%). Contrary, the derivatives prepared at higher concentration (2.5 5%) produced highly viscous solutions or were insoluble in water due to the uneven substitution.

Fig. 1 depicts the ^{1}H NMR spectrum of HA-C12 recorded in NaOD (see the methodical description in Section S1). NaOD was added to the sample, mixed and the spectra were acquired immediately to avoid the possible hydrolysis of HA, but the spectral resolution was strongly improved due to the in situ hydrolysis of substituents. On the contrary, the ^{1}H NMR spectrum of HA-C12 recorded in D$_2$O derivatives presented a broadening of the signals due to self-aggregation (see Figs. S1, S2). The degrees of substitution measured in NaOD and determined by NMR are in a good agreement with the ones determined by gas chromatography (Fig. 1). All the spectra show typical proton chemical shifts of HA involving signal at 2.0 ppm belonging to -NCOCH$_3$ group, skeletal signals at 3.4 3.9 and anomeric resonances at 4.4 4.6 ppm. The ^{13}C-DEPT-HSQC NMR shows the remaining signals at 0.8, 1.3, 1.6 and 2.4 ppm, which were attributed to CH$_3$, and CH$_2$ (Fig. S3). Furthermore, the signals located at 7.6 (triplet), 7.5 (triplet) and at 7.9 (doublet of doublets) ppm, respectively reveal the benzoylation of hyaluronan. The signals located at 3.7 and 3.9 corresponding to the methylene protons C6 (GlcNAc) deshielded to 4.2 and 4.5 ppm in D$_2$O, respectively due to the esterification. Interestingly, the signals corresponding to the diasterotopic methylene protons in position C6, upfield to 2.78 and 3.11 ppm in NaOD. Several signals in HA backbone become double due to the esterification, but they were overlapped. Therefore, the enzymatic fragmentation of HA-C12 reveals the substitution positions (Section 3.2).

The formation of a covalent bond between dodecanoic acid and HA was further established by diffusion ordered NMR spectroscopy (DOSY) as depicted in Fig. 2A. The signals coming from the same molecule have the same diffusion coefficient. Because of the marked difference between the diffusion coefficients of the low molecular weight dodecanoic acid and HA, the DOSY map could easily establish the presence of non-attached fatty acid groups to HA, which obviously diffuse much faster than the bound acyl groups as compared to free dodecanoic acid (Fig. 2B). DOSY experiment showed similar diffusion behaviour for all signals (except for residual isopropanol and the standard of
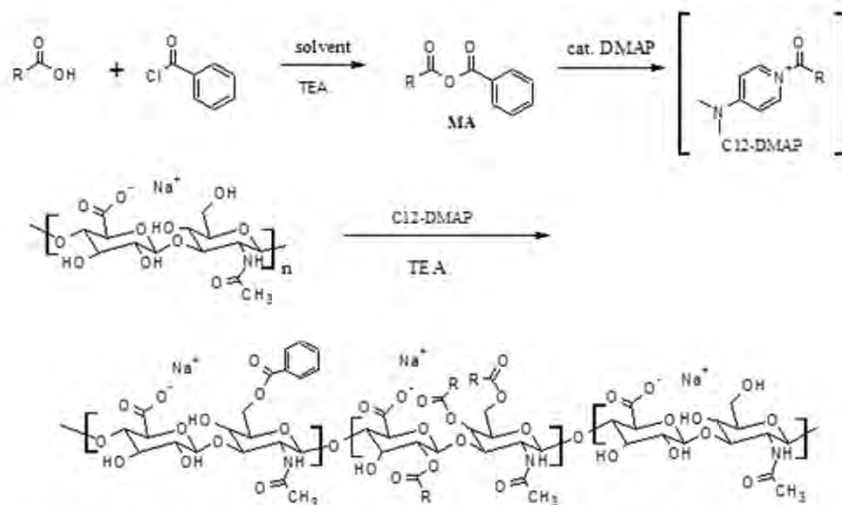
Scheme 1. The esterification of HA by the mixed anhydride method is mediated by 4-dimethylamino pyridine and carried out in a mixed water:organic solvent system. R = -C$_{11}$H$_{23}$ (undecanoyl).

Table 1

Effect of the concentration of HA on the synthesis of HA-C12 in 1,4-dioxane or tert-butanol. The esterification reaction was performed using LMW-HA (17,000 g/mol) and 1.3 eq. of activated dodecanoic acid for 2 h at 25 °C.

Entry	Conc. HA^a (g/L)	DS$_{GC}$^{b} (%)	NTU^c	T (%)^d
1,4-Dioxane				
1	5	7.1 ± 0.01	1.1 ± 0.05	99.6 ± 0.003
2	10	14.5 ± 0.02	1.3 ± 0.07	99.2 ± 0.019
3	15	20.4 ± 0.02	1.7 ± 0.52	99.4 ± 0.003
4	25	28.0 ± 0.1	20.3 ± 0.32	64.4 ± 0.015
5	30	30.7 ± 0.2	48.4 ± 0.28	37.6 ± 0.069
6	50	36.8 ± 0.4	82.1 ± 1.02	24.7 ± 0.032
tert-Butanol				
7	5	7.4 ± 0.1	0.7 ± 0.05	95.4 ± 0.003
8	25	20.9 ± 0.2	2.2 ± 0.13	93.3 ± 0.003
9	50	9.8 ± 0.1	2.7 ± 0.04	94.4 ± 0.003

^{a} Concentration of HA used for the chemical modification (w/v).

^{b} DS$_{GC}$ is the average of three independent determinations.

^{c} NTU stands for nephelometric turbidity, see Section S5.

^{d} Transmittance was determined by UV Vis at 660 nm in 1% (w/v) solution.

(trimethylsilyl)propionic-2,2,3,3-d4 acid sodium salt, which presented a higher diffusion), thus, indicated that all of the proton resonances in this region belonged to one structural complex (HA-C12).

The dodecanoyl moieties present the same diffusion behaviour as HA which confirmed the formation of a covalent bond between the fatty acid and HA. In addition, it can be well recognized that the diffusion of the methyl substituent of C12, which is located at 0.8 ppm presented even a slower diffusion (i.e. the signal is placed at larger log D values along y axis). The restricted mobility in D$_2$O is explained by self-aggregation of the conjugate.

The aggregation phenomenon was confirmed acquiring DOSY in D$_2$O with 0.9% (w/v) of NaCl. The kosmotropic salt made all the molecule appear even larger due to slower diffusion, similar to reported earlier [38]. This means that there is aggregation between the hydrophobic acyl substituents that it is leading to slower diffusion. Furthermore, the critical aggregation concentration (CAC) for the derivative (HA-C12) confirmed the self-aggregation. CAC was determined by Nile red

encapsulation and was found to be ~0.02 mg/mL (Section S2, Fig. S4).

The formation of the ester bond was further evidenced by infrared spectroscopy (Section S3). The esterification was confirmed by the peak at 1729 cm^{-1}, as reported earlier [12]. The signals at 2924 and 2835 cm^{-1} corresponded to the asymmetric and symmetric vibrations of C H due to the presence of the alkyl moiety as substituent, which were more prominent compared to HA used for chemical modification (Fig. S5).

Furthermore, the purity of the samples was determined by GC (Section S4) and thermal analyses (Table S2). The purity was given by the absence of free dodecanoic and benzoic acids (Fig. S6), and determination of ash and dry matter (Figs. S7 S9).

Even though it is expected that the esterification reactions tend not to reach completion, Table 1 shows that 1,4-dioxane was able to produce ~30% conversion at high concentration of HA. Contrary, the conversion of HA to HA-C12 in alcohols (tert-butanol) was not efficient. Esterifying multiple hydroxyls on a disaccharide provides an additional challenge due to steric hindrance between neighboring alkyl moieties.

3.2. Enzymatic degradation for the determination of the distribution of substituents

The solubility of the derivative can be explained in terms of even distribution of the synthesized derivatives. Unsaturated C12-modified tetrasaccharides (– HA4-1 × C12, – HA4-2 × C12, – HA4-3 × C12) are the main SpHyl degradation products of HA-C12 prepared in both 1,4-dioxane and tert-butanol. Extracted ion chromatograms of stated species present in HA-C12 prepared in 1,4-dioxane or tert-butanol are shown in Fig. S10. The amount of – HA4-1 × C12 decreases while – HA4-2 × C12 and – HA4-3 × C12 increase with increasing DS (Fig. 3). At similar DS, the reactions carried out in tert-butanol showed a high percentage of di- and trisubstituted tetrasaccharides. Interestingly, the substitution predominantly takes place at C4 and C6 of one GlcNAc, which can be only observed after enzymatic degradation. The derivative containing a higher amount of – HA4-1 × C12 had an even distribution of acyl chains and is more soluble in water. Notably, the combination of 1,4-dioxane and diluted conditions produced these derivatives.
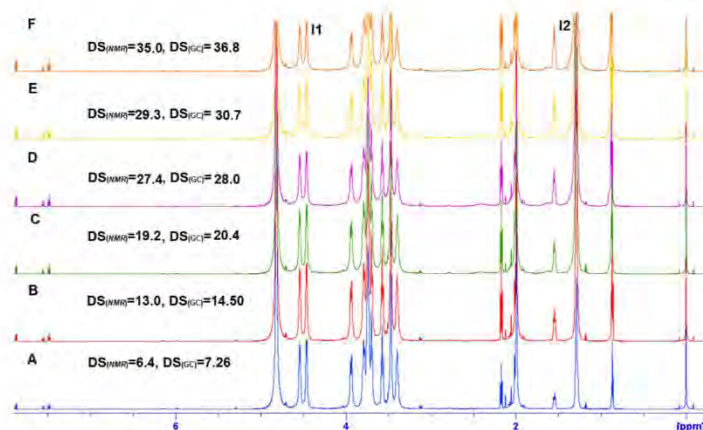
Fig. 1. ^{1}H NMR of HA-C12 measured in NaOD. DS values are in %.
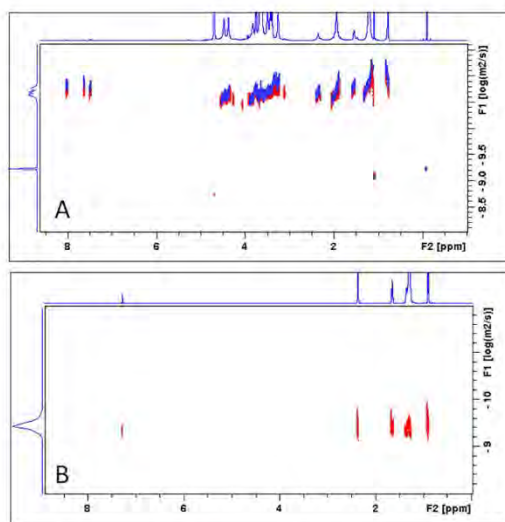
Fig. 2. A. DOSY NMR of HA-C12 measured in D$_2$O (red) and blue (0.9% NaCl) corresponding to sample Table 1, entry 6. B. Lauric acid measured in CDCl$_3$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.3. Separation of solvent components around the free and substituted HA oligosaccharides

In order to explain the different properties of the substitution reaction in the different mixed solvents, 200 ns MD simulations of both free and C12-monosubstituted HA chains of 20 monosaccharide units were simulated in both kinds of the mixed solvents of the organic-component volume fraction of 0.5. Equilibration of these simulations is demonstrated by the stable values of the numbers of close atoms of individual solvent components (for definition see Section 2.5) from the whole molecules and from the substituents (Fig. S12) and also of the interaction energies between the aliphatic sidechain and the solvent

components (Fig. S13). Simulations of non-substituted HA molecules were carried out also at the temperatures of 277 K and 310 K and for the organic-component volume fractions of 0.33, 0.5 and 0.66. The RDFs for the volume fraction of 0.5 are shown in Fig. 4 and show a strong solvent separation in the vicinity of the HA chain reaching the distance of about 25 Å.

In both the mixed solvents the HA solvation shell is enriched by water and deprived of the organic component relative to the bulk solvent. Obviously, the separation is remarkably stronger in the water:tert-butanol mixture (see a simulation snapshot in Fig. S14), where the maximum water concentration, occurring at the distance of 8 Å, is about 1.2 higher than the bulk concentration compared to the value of about 1.06 for water:1,4-dioxane. Simultaneously, analogous ratios for the organic component in the same distance are 0.52 and 0.68, respectively. Interestingly, the temperature dependence of the solvent separation is low and is slightly stronger at 277 K. A comparison of the close-atoms numbers of the solvent components (Fig. S15) shows that the solvent separations are consistent throughout the studied composition range. Moreover, the strong attractive interaction between HA and water is also supported by the quadratic dependence, with higher slope for water:tert-butanol, of relative enrichment of the close solvation shell by water calculated as the ratio of the close-atoms number and its projected bulk value (Fig. S16).

As the solvent separation around the free HA molecules showed a small temperature dependence, the simulation of substituted HA was carried out at 298 K. The radial distribution functions (RDFs) of the individual solvent components around the HA chain are shown in Fig. 5 for the substituent located in the middle of the HA chain.

Obviously, the solvent distribution around HA is very similar to that at the free HA chains. The numbers of close atoms in Fig. 5B indicate a negligible dependence of the HA solvation-shell structure on the substituent position, at the end or in the middle of the chain.

Analogously, RDFs of the solvent components around the substituent, i.e. the C$_{12}$ aliphatic sidechain, were evaluated (Fig. 5C, position in the middle of HA chain, 298 K). They show a completely opposite solvent separation than at the HA chain. The solvation shell of the sidechain is strongly deprived of water and enriched by the organic component whose concentration reaches its maximum at about 6 Å from the sidechain and then relaxes towards its bulk value reaching it approx. at 12 Å.

In water:tert-butanol mixture the solvent separation is remarkably stronger than in water:1,4-dioxane. The ratios of the water and organic-component close atoms numbers at the sidechain (Fig. 5D) are 3.3 and 1.2, respectively (bulk solvent value 0.86), again almost independently
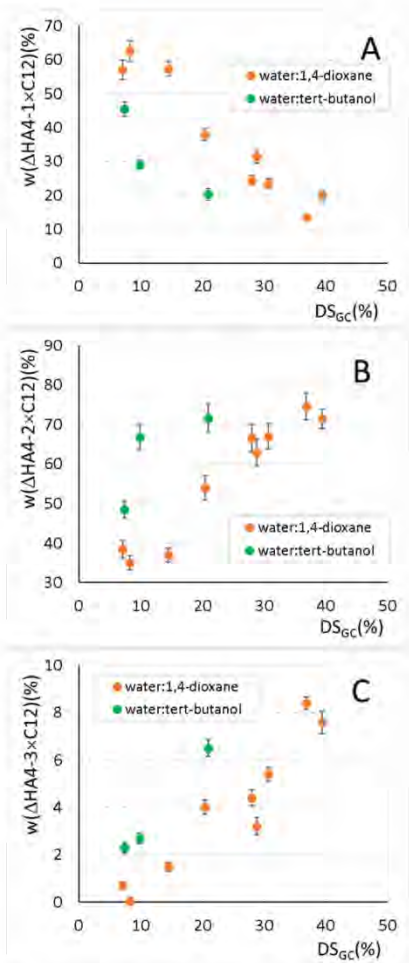
Fig. 3. The amounts of mono- (A), di- (B) or trisubstituted (C) unsaturated tetrasaccharides plotted against DSGC. Derivatives prepared in 1,4-dioxane and tert-butanol are showed in orange and green, respectively. The esterification reaction was performed using HA (17,000 g/mol) and 1,3 eq of activated C12 for 2 h at 25 °C. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of the substituent position. In addition, the RDFs in Fig. S17B show that the concentration maximum of tert-butanol methyl-carbon atoms is closer to the aliphatic sidechain than that of the oxygen atoms by approx. 1.5 Å which agrees with the distance of these atoms in the tert-butanol molecule. The amphiphilic character of tert-butanol molecules thus causes their strong preferential orientation by the methyl groups towards the aliphatic sidechain, while at the HA chain no such orientation takes place (Fig. S17A). On the contrary, RDFs for 1,4-dioxane carbon and oxygen atoms show significant difference neither at HA nor at the aliphatic sidechain (Fig. S17) indicating the random orientation of these molecules.

Similar solvent separation was observed also around the aliphatic chain of the C_{12}-DMAP$^+$ reactant — see the comparison with the

sidechain bound on HA in Fig. S18. The small differences of the distribution functions are caused rather by the different surroundings of the C_{12} chain in the two cases that necessarily influences the RDFs at the longer distances.

3.4. Separation of solvent components around the free and substituted HA oligosaccharides

Free and C_{12}-substituted HA oligosaccharides were simulated in the presence of the reaction intermediates (C_{12}-DMAP$^-$) to evaluate the influence of the mixed solvents on the course of the esterification. The free HA chains were simulated together with 5 C_{12}-DMAP$^+$ molecules. Furthermore, 5 TEA$^+$ molecules were added to make the environment closer to the experimental conditions. The simulation was carried out in 50% (v/v) mixed solvents at 277 K and 310 K.

RDFs in Fig. 6 show an essential difference between the two mixed solvents. In water:1,4-dioxane the distributions are almost independent of temperature, only in the closest area the number of C_{12}-DMAP$^+$ atoms is somewhat higher (Fig. 6, inset) at the lower temperature indicating the disturbance of the attractive interaction by the thermal motion. This indicates an essentially enthalpic stabilization of C_{12}-DMAP$^+$ at HA by the electrostatic interaction between the positive and negative charges of C_{12}-DMAP$^+$ and HA, respectively. In water:tert-butanol, on the contrary, the occurrence of C_{12}-DMAP$^+$ in the HA vicinity is very rare at 277 K while at 310 K it grows substantially indicating an entropy-based stabilization of this state. It indicates that in this solvent the approach of C_{12}-DMAP$^+$ to HA requires a certain restructuring of the solvation shells of HA and C_{12}-DMAP$^+$ supported by the higher temperature. Its deeper understanding, however, needs more targeted MD simulations and will be a matter of further research.

3.5. Interactions of the reaction intermediates with substituted HA oligosaccharides

Substituted HA chains were simulated in both 50% (v/v) mixed solutions in the presence of 10 C_{12}-DMAP$^+$ molecules at 298 K. The C_{12}-DMAP$^+$ distributions around HA and the sidechain are well equilibrated within the 200 ns simulations — see the time development of the numbers of close C_{12}-DMAP$^+$ molecules in Fig. S19.

RDFs of the C_{12}-DMAP$^+$ molecules were calculated separately around individual thirds of the HA chain containing first 6, middle 8 and last 6 monosaccharide residues (Fig. 7) as well as around the substituent itself. In water:1,4-dioxane the situation at individual parts of HA is practically equal and independent of the substituent position. The C_{12}-DMAP$^+$ molecules are rather strongly electrostatically attracted to the HA chain which not only increases their concentration at HA, but also causes the preferential orientation with the positively charged head (DMAP$^+$ moiety) towards this molecule — see the RDFs of head and tail (last 5 carbons of the aliphatic chain) atoms in Fig. 7. On the contrary, in water:tert-butanol the C_{12}-DMAP$^+$ molecules are far less attracted to HA in the substituent-free parts of the molecule, consistently with the non-substituted HA chains. Close to the substituent, however, the number of C_{12}-DMAP$^+$ is remarkably higher. It indicates that the aliphatic sidechain disturbs the water-rich solvation shell of HA, being itself wrapped by tert-butanol molecules, and thus allows C_{12}-DMAP$^+$ to approach the HA chain. This phenomenon is more pronounced for the substituent located in the middle of the HA chain, since at the end (Fig. S20) the aliphatic chain may stick aside of the HA chain diminishing thus its positive effect on the reactants approach.

4. Discussion

The promising potential of the mixed solvents to control the degree of substitution of the esterification of HA, as well as the distribution of the substituents along the chain, was shown recently on the case of oleoyl residues [25]. To investigate the universality of these findings,
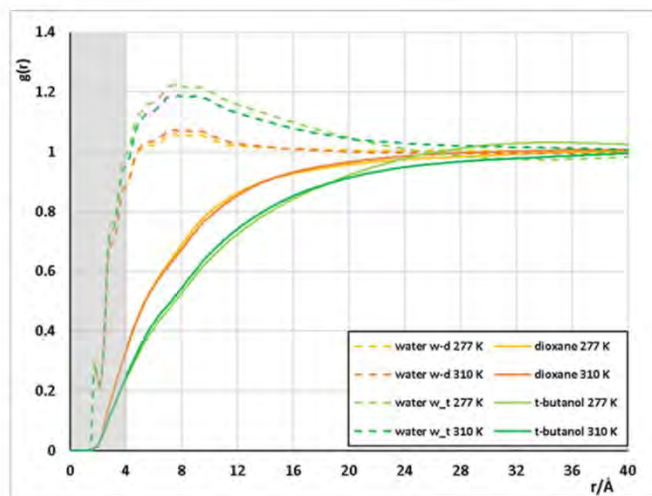
Fig. 4. RDFs of the solvent components of the 50% (v/v) mixed solvents water:1,4-dioxane (denoted w:d) and water:tert-butanol (w:t) for 277 K and 310 K.
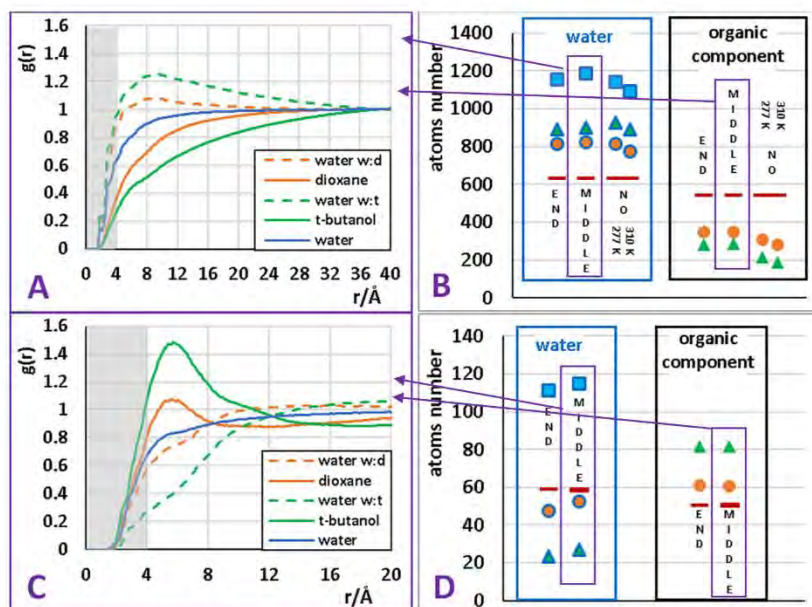
Fig. 5. RDFs of the solvent components in the 50% (v/v) mixed solvents water:1,4-dioxane (orange) and water:tert-butanol (green) and pure water (blue) around the C_{12}-substituted HA molecules. A: RDFs at the HA chain itself for 298 K and substituent in the middle. B: Numbers of close atoms (within the 4 Å distance from HA, indicated by the grey background in panel A) of water (left) and organic component (right) for both substituent positions and free HA molecules (simulated at different temperatures). Violet frames and arrows indicate the system shown in panel A. C and D: Same as A and B, but for the substituent instead of HA chain. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

esterification of HA by dodecanoyl residues was carried out by the same methodology in the same mixed solvents, water:1,4-dioxane and water:tert-butanol in the volume ratio 1:1. The observed trends reproduced those of the previous study. Remarkably higher degree of substitution was reached in water:1,4-dioxane as well as its growth along with the concentration of the reactants in the reaction feed. While in water:1,4-

89
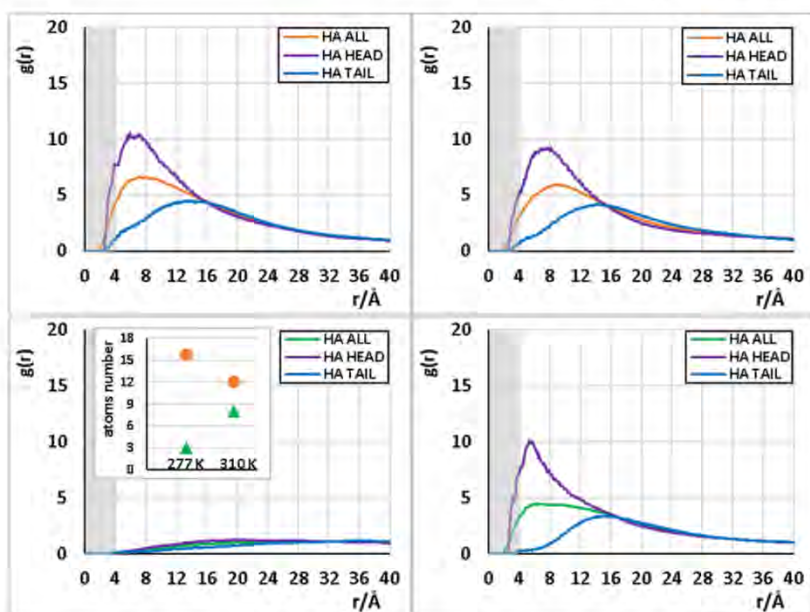
Fig. 6. RDFs of C_{12}-DMAP$^+$ around HA molecules at 277 K (left) and 310 K (right) for water:1,4-dioxane (up) and water:*tert*:butanol (down) for the whole molecule, the head (carbon atoms of the DMAP heterocycle) and tail (last five carbon atoms of the sidechain). Inset   numbers of close atoms of C_{12}-DMAP$^+$ at HA within the region indicated in grey.

dioxane the distribution of the substituents was relatively even, in water: tert-butanol clustering of the substituents, resulting in more frequent occurrence of multiply substituted oligosaccharides after cleaving HA enzymatically, was observed. The agreement of both these studies indicates that the influence of the mixed solvents on the reaction course is general for various kinds of non-polar aliphatic substituents. The solvent composition affects especially the approach of the reactants to the HA chain. An investigation of this phenomenon was based on MD simulations of non-substituted and substituted HA molecules both in pure mixed solvents and in the presence of the reaction intermediates.

MD simulations were carried out for 200 ns and this time was sufficient for the equilibration of all the distributions of molecules as well as the interaction energies (Figs. S12, S13, S19) since the systems only contain small molecules without internal motions with high energetic barriers. Compared to water:1,4-dioxane solution, water:tert-butanol mixture is more prone to solvent separation at both the hydrophilic HA molecule and the aliphatic sidechain (Figs. 4, 5). In the vicinity of HA both the organic components are repelled from the solvation shell, but they differ in the intensity of the effect and in the radius of the zone where the solvent composition differs from the bulk value. For instance, if the bulk ratio the organic to water atoms is 1, in the closest area it is only about 0,30 for tert-butanol and 0,43 for 1,4-dioxane and in the point of maximum water concentration 0,44 and 0,70, respectively. Hence, the solvation shell in water:tert-butanol is enriched by water remarkably more than in water:1,4-dioxane. Moreover, the radius of the water-enriched zone is much wider for tert-butanol reaching the bulk value at approx. 34 Å in contrast to 24 Å. If the solvation shell is considered as a cylinder wrapping the HA molecule, the simulated HA molecule repels approx. 170 tert-butanol- and attracts 700 water molecules, while for the water:1,4-dioxane solution these numbers are only 50 and 240, respectively. On the contrary, the solvation shell of the aliphatic sidechain of the substituted HA molecule is enriched by the less

polar organic component (Fig. 5). Although the long-distance parts of the distribution functions are biased by the reverse solvent separation around HA, the ratios of the numbers of close atoms of the organic component and water   3.3 for water:tert-butanol, 1.2 for water:1,4-dioxane, 0.86 for bulk   support the apparent fact of stronger solvent separation in water:tert-butanol. Amphiphilic tert-butanol molecules are uniquely oriented by their methyl groups towards the aliphatic chain by hydrophobic interactions. Accordingly, this mixture is also more prone to components micro-separation (Fig. S21) reported previously by [39].

As shown in Table 1, the reaction rate, and thus also the DS, initially grows with the concentration of both the reactants, HA and the activated dodecanoic acid, in accord with the standard kinetic control. It was, however, shown that it is only valid below the coil overlap point of the polymer. This point was reported to be at concentration ~ 0.86 wt% for LMW-HA (MW ≥ 37,000 g/mol) in water [40,41]. As the HA chains fill the solution, the reaction rate is affected by steric effect of the coils (at concentration ≥ 0.86 wt%). Moreover, the viscosity of the reaction mixture influences the rate of reaction, as well as the solubility of the mixed anhydride. In the mixed solvents the separation of their components may further contribute to the decrease of the reaction efficiency observed especially in water:tert-butanol above the HA concentration of 25 g/L in the reaction feed. Considering the solvation shell as a cylinder of a mean radius of 35 Å inferred from MD, its formally calculated volume at the HA concentration of 25 g/L is about 1.4-times higher than the whole volume of the solution. This situation likely leads to concentration of HA in water-rich zones while C_{12}-DMAP$^+$ remains in the tert-butanol enriched areas having thus limited access to HA. In water:1,4-dioxane this effect plays only a minor role since the solvation shell is narrower and the solvent separation weaker, which makes the rate decrease rather small.

MD simulations show that for the C_{12}-DMAP$^+$ reactants, even at low
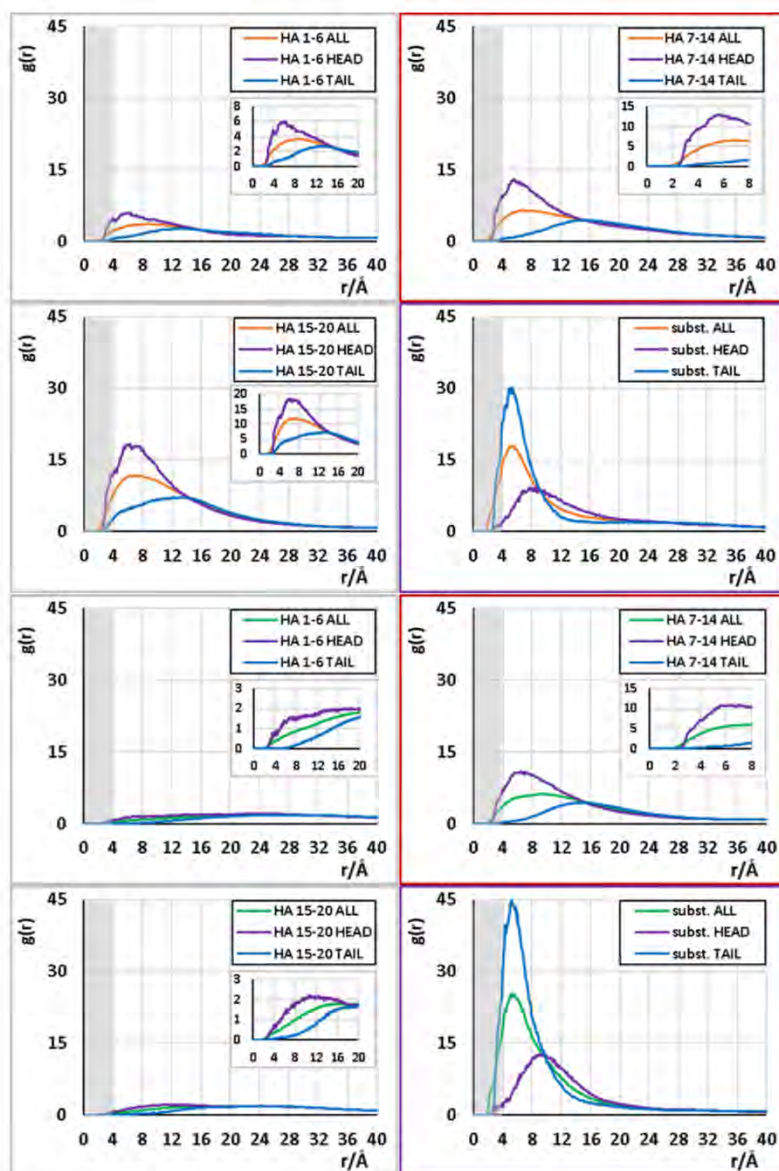
Fig. 7. RDFs of the C_{12}-DMAP$^+$ molecules, 50% (v/v) water:1,4-dioxane (upper quartet) and water:tert-butanol (lower quartet), 298 K, for individual parts of the HA molecule substituted in the middle of the chain. In every quartet, grey-framed plots are related to the end parts of the HA chain (residues 1 6 and 15 20), red-framed to the central part of HA (residues 7 14) bearing the substituent and violet-framed to the substituent itself. For better comparability, all graphs have equal scales, details of the closer parts of the RDFs are shown in the insets. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

HA concentration, it is more difficult to approach HA in water:tert-butanol solution. It is, therefore, not surprising that the efficiency of the reaction measured by DS is lower here than in water:1,4-dioxane. At 298 K, the occurrence of the C_{12}-DMAP$^+$ head in the close area at HA in the latter solvent is, in average, about 3 4 times lower which roughly corresponds with the observed difference in DS. Although the temperature dependence of the reaction efficiency was not investigated experimentally, it can be expected that in water:1,4-dioxane the temperature effect will not be strong. With the temperature growth the number of reactant molecules attacking HA slightly decreases which may, on the other hand, be compensated by a higher rate of the reaction step itself. In water:tert-butanol, however, the number of reactants approaching HA grows quite rapidly with the temperature (Fig. 6) which could lead to the overall reaction-rate increase.

The evenness of the substituent distribution on the HA chain was investigated by enzymatic cleavage of HA-C12 by SpHyl. The fraction of

monosubstituted fragments is higher in water:1,4-dioxane (Fig. 3), while the fraction of bi- and trisubstituted oligomers is significantly higher in the water:tert-butanol mixture. This observation agrees with the MD simulations showing that in the water:tert-butanol solution the C_{12}-$DMAP^-$ ions are very hardly able to approach HA except for the close neighborhood of the already bound substituent. This preference is particularly strong at the substituent bound in the middle of the HA chain, a location dominating when longer chains are used. Therefore, the aliphatic sidechain probably serves as an entering gate for C_{12}-$DMAP^-$ in water:tert-butanol while in water:1,4-dioxane its influence on the subsequent substitutions is negligible.

5. Conclusions

The composition of mixed media of water and organic solvent have been shown to be a promising factor determining the efficiency of the mixed-anhydride based esterification reaction on hyaluronic acid and the evenness of the distribution of the substituents along the chain. In water:1,4-dioxane solution the substitution of HA by dodecanoyl residues provided higher degree of substitution as well as more even substituents distribution than in water:tert-butanol, both in the volume ratio 1:1. MD simulations of non-substituted and substituted HA molecules in the mixed solvents showed stronger water-enriched solvation shell of HA in water:tert-butanol forming a higher barrier for the approach of the reactant C_{12}-$DMAP^+$ to HA and thus causing a lower reaction rate in this solvent and even its rapid decrease at higher HA concentration in the reaction feed. However, an opposite solvent separation around the aliphatic chain enables the approach of C_{12}-$DMAP^+$ to HA in the proximity of an already bound substituent which may cause the clustering of substituents observed in water:tert-butanol. The MD results thus explain the experimental observations, consistent also with the previous study with oleoyl substituents [25], on the molecular level and open a way to the systematic investigation and utilization of the mixed solvents on the substitution reactions on polysaccharides.

CRediT authorship contribution statement

Eva Kutálková Methodology, Formal analysis, Investigation, Writing - Original draft, Writing - Review & editing
Josef Hrnčiřík Methodology, Investigation, Writing - Original draft, Writing - Review & editing
Roman Witasek Methodology, Formal analysis, Investigation
Marek Ingr Conceptualization, Methodology, Formal analysis, Investigation, Writing - Original draft, Writing - Review & editing, Supervision, Project administration
Gloria Huerta-Ángeles Conceptualization, Methodology, Formal analysis, Investigation, Writing - Original draft, Writing - Review & editing, Supervision
Martina Hermannová Methodology, Formal analysis, Investigation, Writing - Original draft, Writing - Review & editing

Vladimír Velebný Conceptualization, Methodology, Investigation, Supervision, Project administration, Funding acquisition.

References

[1] E. Fouissac, M. Milas, M. Rinaudo, R. Borsali, Influence of the ionic strength on the dimensions of sodium hyaluronate, Macromolecules 25 (1992) 5613 5617, https://doi.org/10.1021/ma00047a009.

[2] K. Hayashi, K. Tsutsumi, F. Nakajima, T. Norisuye, A. Teramoto, Chain-stiffness and excluded-volume effects in solutions of sodium hyaluronate at high ionic strength, Macromolecules 28 (1995) 3824 3830, https://doi.org/10.1021/ma00115a012.

[3] R. Mendichi, L. Soltés, A. Giacometti Schieroni, Evaluation of radius of gyration and intrinsic viscosity molar mass dependence and stiffness of hyaluronan, Biomacromolecules 4 (2003) 1805 1810, https://doi.org/10.1021/bm0342178.

[4] M. Ingr, E. Kutálková, J. Hrnciřík, Hyaluronan random coils in electrolyte solutions-a molecular dynamics study, Carbohydr. Polym. 170 (2017) 289 295, https://doi.org/10.1016/j.carbpol.2017.04.054.

[5] C.D. Blundell, P.L. DeAngelis, A. Almond, Hyaluronan: the absence of amide carboxylate hydrogen bonds and the chain conformation in aqueous solution are incompatible with stable secondary and tertiary structure models, Biochem. J. 396 (2006) 487 498, https://doi.org/10.1042/BJ20060085.

[6] P. Gribbon, B.C. Heng, T.E. Hardingham, The analysis of intermolecular interactions in concentrated hyaluronan solutions suggest no evidence for chain-chain association, Biochem. J. 350 (2000) 329 335.

[7] J.E. Scott, F. Heatley, Hyaluronan forms specific stable tertiary structures in aqueous solution: a 13C NMR study, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 4850 4855, https://doi.org/10.1073/pnas.96.9.4850.

[8] J.E. Scott, F. Heatley, in: Biological Properties of Hyaluronan in Aqueous Solution Are Controlled and Sequestered by Reversible Tertiary Structures, Defined by NMR Spectroscopy, Biomacromolecules 3, 2002, pp. 547 553, https://doi.org/10.1021/bm010170j.

[9] E. Kutálková, J. Hrnciřík, R. Witasek, M. Ingr, Effect of solvent and ions on the structure and dynamics of a hyaluronan molecule, Carbohydr. Polym. 234 (2020), 115919, https://doi.org/10.1016/j.carbpol.2020.115919.

[10] F. Bano, M.I. Tammi, D.W. Kang, E.N. Harris, R.P. Richter, Single-molecule unbinding forces between the polysaccharide hyaluronan and its binding proteins, Biophys. J. 114 (2018) 2910 2922, https://doi.org/10.1016/j.bpj.2018.05.014.

[11] F. Bongiovi, C. Fiorica, F.S. Palumbo, G. Di Prima, G. Giammona, G. Pitarresi, Imatinib-loaded micelles of hyaluronic acid derivatives for potential treatment of neovascular ocular diseases, Mol. Pharm. 15 (2018) 5031 5045, https://doi.org/10.1021/acs.molpharmaceut.8b00620.

[12] G. Huerta-Ángeles, F. Ondreáš, M. Brandejsová, K. Kopecká, H. Vagnerová, J. Kulhánek, T. Drmota, Formulation of hyaluronan grafted with dodecanoic acid as a potential ophthalmic treatment, Carbohydr. Polym. 246 (2020), 116578, https://doi.org/10.1016/j.carbpol.2020.116578.

[13] G. Huang, H. Huang, Application of hyaluronic acid as carriers in drug delivery, Drug Deliv. 25 (2018) 766 772, https://doi.org/10.1080/10717544.2018.1450910.

[14] D. Šmejkalová, T. Muthný, K. Nešporová, M. Hermannová, E. Achbergerová, G. Huerta-Angeles, M. Svoboda, M. Cepa, V. Machalová, D. Luptáková, V. Velebný, Hyaluronan polymeric micelles for topical drug delivery, Carbohydr. Polym. 156 (2017) 86 96, https://doi.org/10.1016/j.carbpol.2016.09.013.

[15] A. Cadete, A. Olivera, M. Besev, P.K. Dhal, L. Gonçalves, A.J. Almeida, G. Bastiat, J.-P. Benoit, M. de la Fuente, M. Garcia-Fuentes, M.J. Alonso, D. Torres, Self-assembled hyaluronan nanocapsules for the intracellular delivery of anticancer drugs, Sci. Rep. 9 (2019) 11565, https://doi.org/10.1038/s41598-019-47995-8.

[16] P. Walvekar, R. Gannimani, M. Salih, S. Makhathini, C. Mocktar, T. Govender, Self-assembled oleylamine grafted hyaluronic acid polymersomes for delivery of vancomycin against methicillin resistant Staphylococcus aureus (MRSA), Colloids Surf. B: Biointerfaces 182 (2019) 110388, https://doi.org/10.1016/j.colsurfb.2019.110388.

[17] G. Huerta-Angeles, M. Brandejsová, R. Nigmatullin, K. Kopecká, H. Vágnerová, D. Šmejkalová, I. Roy, V. Velebný, Synthesis of graft copolymers based on hyaluronan and poly(3-hydroxyalkanoates), Carbohydr. Polym. 171 (2017) 220 228, https://doi.org/10.1016/j.carbpol.2017.05.011.

[18] X. Wang, J. Messman, J.W. Mays, D. Baskaran, Polypeptide grafted hyaluronan: synthesis and characterization, Biomacromolecules 11 (2010) 2313 2320, https://doi.org/10.1021/bm1004146.

[19] S. Khetan, J.A. Burdick, Patterning network structure to spatially control cellular remodeling and stem cell fate within 3-dimensional hydrogels, Biomaterials 31 (2010) 8228 8234, https://doi.org/10.1016/j.biomaterials.2010.07.035.

[20] E. Zerobin, M. Markovic, Z. Tomášiková, X.-H. Qin, D. Ret, P. Steinbauer, J. Kitzmüller, W. Steiger, P. Gruber, A. Ovsianikov, R. Liska, S. Baudis, Hyaluronic acid vinyl esters: a toolbox toward controlling mechanical properties of hydrogels for 3D microfabrication, J. Polym. Sci. 58 (2020) 1288 1298, https://doi.org/10.1002/pol.20200073.

[21] W.M. Payne, D. Svechkarev, A. Kyrychenko, A.M. Mohs, The role of hydrophobic modification on hyaluronic acid dynamics and self-assembly, Carbohydr. Polym. 182 (2018) 132 141, https://doi.org/10.1016/j.carbpol.2017.10.054.

[22] G. Huerta-Angeles, M. Bobek, E. Príkopová, D. Šmejkalová, V. Velebný, Novel synthetic method for the preparation of amphiphilic hyaluronan by means of aliphatic aromatic anhydrides, Carbohydr. Polym. 111 (2014) 883 891, https://doi.org/10.1016/j.carbpol.2014.05.035.

[23] Y. Xi, T. Jiang, Y. Yu, J. Yu, M. Xue, N. Xu, J. Wen, W. Wang, H. He, Y. Shen, D. Chen, X. Ye, T.J. Webster, Dual targeting curcumin loaded alendronate-hyaluronan- octadecanoic acid micelles for improving osteosarcoma therapy, Int. J. Nanomedicine 14 (2019) 6425 6437, https://doi.org/10.2147/IJN.S211981.

[24] L. Vítková, I. Musilová, E. Achbergerová, A. Minařík, P. Smolka, E. Wrzecionko, A. Mráček, Electrospinning of hyaluronan using polymer coelectrospinning and intermediate solvent, Polymers 11 (2019), https://doi.org/10.3390/polym11091517.

[25] O. Štrympl, J. Vohlídal, M. Hermannová, M. Maldonado-Domínguez, M. Brandejsová, K. Kopecká, V. Velebný, G. Huerta-Ángeles, Oleate-modified hyaluronan: controlling the number and distribution of side chains by varying the reaction conditions, Carbohydr. Polym. 267 (2021), 118197, https://doi.org/10.1016/j.carbpol.2021.118197.

[26] S. Podzimek, M. Hermannova, H. Bilerova, Z. Bezakova, V. Velebny, Solution properties of hyaluronic acid and comparison of SEC-MALS-VIS data with off-line capillary viscometry, J. Appl. Polym. Sci. 116 (2010) 3013 3020, https://doi.org/10.1002/app.31834.

[27] D. Cožíková, T. Šílová, V. Moravcová, D. Šmejkalová, S. Pepeliaev, V. Velebný, M. Hermannová, Preparation and extensive characterization of hyaluronan with narrow molecular weight distribution, Carbohydr. Polym. 160 (2017) 134 142, https://doi.org/10.1016/j.carbpol.2016.12.045.

[28] J. Chmelar, A. Kotzianová, M. Hermannová, R. Šuláková, D. Šmejkalová, J. Kulhánek, V. Velebný, Evaluating the degree of substitution of water-insoluble acyl derivatives of hyaluronan using raman spectroscopy: method development and comparison with gas chromatography and 1H NMR, Anal. Methods 9 (2017) 232 239, https://doi.org/10.1039/C6AY03067J.

[29] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, Scalable molecular dynamics with NAMD, J. Comput. Chem. 26 (2005) 1781 1802, https://doi.org/10.1002/jcc.20289.

[30] O. Guvench, S.S. Mallajosyula, E.P. Raman, E. Hatcher, K. Vanommeslaeghe, T. J. Foster, F.W. Jamison, A.D. MacKerell, CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein

[31] J.B. Klauda, R.M. Venable, J.A. Freites, J.W. O Connor, D.J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A.D. MacKerell, R.W. Pastor, Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types, J. Phys. Chem. B 114 (2010) 7830 7843, https://doi.org/10.1021/jp101759q.

[32] J. Lee, X. Cheng, J.M. Swails, M.S. Yeom, P.K. Eastman, J.A. Lemkul, S. Wei, J. Buckner, J.C. Jeong, Y. Qi, S. Jo, V.S. Pande, D.A. Case, C.L. Brooks, A. D. MacKerell, J.B. Klauda, W. Im, CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field, J. Chem. Theory Comput. 12 (2016) 405 413, https://doi.org/10.1021/acs.jctc.5b00935.

[33] S.-J. Park, J. Lee, Y. Qi, N.R. Kern, H.S. Lee, S. Jo, I. Joung, K. Joo, J. Lee, W. Im, CHARMM-GUI glycan modeler for modeling and simulation of carbohydrates and glycoconjugates, Glycobiology 29 (2019) 320 331, https://doi.org/10.1093/glycob/cwz003.

[34] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (33 38) (1996) 27 28.

[35] D. Šmejkalová, M. Hermannová, R. Šuláková, A. Prušová, J. Kucerík, V. Velebný, Structural and conformational differences of acylated hyaluronan modified in protic and aprotic solvent system, Carbohydr. Polym. 87 (2012) 1460 1466, https://doi.org/10.1016/j.carbpol.2011.09.057.

[36] J. Chmelar, J. Mrázek, M. Hermannová, L. Kubala, G. Ambrožová, A. Kocurková, T. Drmota, K. Nešporová, L. Grusová, V. Velebný, Biodegradable free-standing films from lauroyl derivatives of hyaluronan, Carbohydr. Polym. 224 (2019), 115162, https://doi.org/10.1016/j.carbpol.2019.115162.

[37] Y.K. Sung, S.W. Kim, Recent advances in polymeric drug delivery systems, Biomater. Res. 24 (2020) 12, https://doi.org/10.1186/s40824-020-00190-7.

[38] D. Ret, G. Steiner, S. Gentilini, S. Knaus, Exact determination of the degree of substitution of high molar mass hyaluronan by controlling the conformation in solution, Carbohydr. Polym. 204 (2019) 124 130, https://doi.org/10.1016/j.carbpol.2018.10.003.

[39] S.D. Overduin, A. Perera, G.N. Patey, Structural behavior of aqueous t-butanol solutions from large-scale molecular dynamics simulations, J. Chem. Phys. 150 (2019), 184504, https://doi.org/10.1063/1.5097011.

[40] M.K. Cowman, T.A. Schmidt, P. Raghavan, A. Stecco, Viscoelastic properties of hyaluronan in physiological conditions, F1000Res. 4 (2015) 622, https://doi.org/10.12688/f1000research.6885.1.

[41] M. Cowman, S. Matsuoka, The intrinsic viscosity of hyaluronan, Hyaluronan. 1 (2002) 75 78, https://doi.org/10.1533/9781845693121.75.

modeling, J. Chem. Theory Comput. 7 (2011) 3162 3180, https://doi.org/10.1021/ct200328p.

93

P4. Pressure induced structural changes and dimer destabilization of HIV-1 protease studied by molecular dynamics simulations.

PCCP

PAPER

Pressure induced structural changes and dimer destabilization of HIV-1 protease studied by molecular dynamics simulations†

Eva Kutálková, Josef Hrnčiřík and Marek Ingr*

High-pressure methods have become attractive tools for investigation of the structural stability of proteins. Besides protein unfolding, dimerization can be studied this way, too. HIV-1 protease is a convenient target of experimental and theoretical high-pressure studies. In this study molecular-dynamics simulations are used to predict the response of HIV-1 protease to the pressure of 0.1 to 600 MPa. The protease conformation of both the monomer and the dimer is highly rigid changing insignificantly with growing pressure. Hydrophobicity of the protease decreases with increasing pressure. Water density inside the active-site cavity grows from 87% to 100% of the bulk water density within the pressure range. The dimer-dissociation volume change is negative for most of the pressure ranges with the minimum of −105 ml mol^{-1}, except for a short interval of positive values at low pressures. The dimer is thus slightly stabilized up to 160 MPa, but strongly destabilized by higher pressures.

1. Introduction

The protease of type-1 human immunodeficiency virus (HIV-1 PR) is one of the most studied enzymes in general. As an inevitable part of the life cycle of HIV virus, the causative agent of AIDS disease, this protein is an important therapeutic target in AIDS treatment. Therefore, a huge number of experimental studies in the fields of molecular biology, biochemistry and biophysics have been performed since the AIDS pandemics burst out which contributed to revealing the structure and enzymological properties of the enzyme. As a result of this effort, nine inhibitors have been brought to the stage of commercially produced and therapeutically used drugs during this period.[1,2]

In addition to the experimental research, theoretical studies based on the methods of molecular dynamics and quantum chemistry have been carried out, which also helped us to understand better the structure–function relationship within the enzyme and provided a valuable contribution to the drug development by means of rational drug design.[3]

Structural stability is one of the thoroughly studied features of HIV-1 PR. The first study of this phenomenon was performed by Todd *et al.*[4] by differential scanning calorimetry, who presumed a tight connection between the dimer dissociation and unfolding. However, recent NMR experiments showed that the monomers of retroviral proteases of M-PMV,[5] HIV-1 PR[6] and HIV-2 PR[7] can maintain their structures even without dimerization partners. The hypothesis of a stable HIV-1 PR monomer was also supported by the molecular-dynamics simulations by Levy *et al.*[8] and Yan *et al.*[9] Unfolding of the HIV-1 PR monomer facilitated by denaturing agents was studied by Noel *et al.*[10] who showed that certain mutations in the protease sequence increase its tendency to unfold. Molecular-dynamics simulations were carried out in order to explain the mechanism of monomer folding and characterize its intermediates.[11,12]

As the enzyme is active as a homodimer, it was believed that an efficient inhibitor can be designed that disrupts dimer formation and several experimental and theoretical studies were carried out with this aim.[13] A number of research teams made an effort to measure the dimerization equilibrium constant $K_d = [M]^2/[D]$, where $[M]$ is the concentration of monomer and $[D]$ is the concentration of dimer.[14–18] Unfortunately, the results of these measurements differ by several orders of magnitude among the different studies depending on the method used, therefore the overall experimental evidence is so far inconclusive. An obvious cause of the experimental troubles is the high stability of the dimer and the subsequent necessity to do the experiments at very low concentrations in which the experimental sensitivity is limited by the low signal to noise ratio.

High-pressure methods represent a suitable way of studying thermodynamic transitions of biomolecules that cannot be easily observed at atmospheric pressure. Although the high-pressure conditions are non-physiological, they are often used

Tomas Bata University in Zlín, Faculty of Technology, Department of Physics and Materials Engineering, Nám. T.G. Masaryka 5555, 76001 Zlín, Czech Republic. E-mail: ingr@ft.utb.cz; Fax: +420 576035141; Tel: +420 576031417
† Electronic supplementary information (ESI) available. See DOI: 10.1039/c4cp03676j

in combination with several spectroscopic and other techniques in order to study the structural stability of proteins with respect to unfolding. Besides this, these techniques are attractive tools to study the equilibrium behavior of the oligomeric proteins or the active site structure and the mechanism of enzyme catalysis can be studied this way as well.[19,20] The necessary condition of applicability of the high-pressure methods is a negative volume change ΔV of the studied process in the direction that should be enforced by pressure increase.

Molecular dynamics simulations of proteins under high pressure have been carried out for various proteins, especially in order to study protein unfolding.[21–26] For the theoretical background overview see the review by Paci[27] discussing various aspects of high-pressure molecular-dynamics simulations including the usability of the standard force fields parameterized at atmospheric pressure. A reliable assessment of this problem is difficult since not many parallel experiments and calculations have been carried out and the accuracy of the experimental data is also limited.[27] In general, all the intermolecular interactions are supposed to be well parametrized, especially in the attractive regions of the potential curves, in order to reproduce the atmospheric-pressure properties of the molecules correctly. Application of high pressure causes small changes in the mean intermolecular distances (at 600 MPa the linear intermolecular distances decrease approx. by 5%) and only an insignificant internal-energy increase (pressure increase by 600 MPa increases the energy similar to the temperature increase by 10 K), therefore the molecules are not shifted out of the reliable potential range. Recently Lerbret et al.[28] compared the vibrational density of states of lysozyme and bulk water obtained both experimentally (inelastic neutron scattering) and theoretically (molecular dynamics), which resulted in a remarkably good agreement for both the systems. Calculations of ΔV have been reported for the process of unfolding of several proteins.[21,24] They are in general based on the calculation of the difference in partial molar volumes (PMV) of the unfolded and folded systems. A troublesome point of these calculations is that the volume change is often about 2 orders of magnitude lower than the volume fluctuations of the simulated system, which results in errors of ΔV comparable to the value itself even when long sets of data are used. To our knowledge, this approach has not been used for the evaluation of ΔV of pure dimer or oligomer dissociation yet.

Considering the HIV-1 PR dimerization, application of high pressure can help to enforce dimer dissociation, which may facilitate the determination of K_d and enable further experiments with free monomers. However, to our knowledge no experimental and only one theoretical study[29] describing the behavior of HIV-1 PR under the conditions of high pressure have been published to date. In their work Meher et al. compare the HIV-1 PR dimer at atmospheric pressure and 300 MPa as regards the geometrical properties of the molecule.

In this study we use molecular dynamics simulations to get deeper insight into the high pressure response of HIV-1 PR in both dimeric and monomeric states. Simulations are carried out in the pressure range from atmospheric pressure up to 600 MPa. The changes in geometry of the molecules are

evaluated as well as the variability in interactions of the protein molecules with water. A special interest is dedicated to the water density in the active-site cavity, which is also used as a base for the estimate of the volume change ΔV of the dimer dissociation as a function of pressure.

2. Results and discussion

2.1. Conformational changes of the dimer and monomer of HIV-1 PR

Unrestrained, all-atom molecular dynamics simulations with the smaller water box initiated from both the closed and open conformations of the protease dimer at atmospheric pressure suggested that closed, semi-open and open conformations are in dynamical equilibrium. The average life time of the individual conformation is in the order of nanoseconds. Semi-open form prevails in the statistical ensemble which agrees with the previously published data of Hornak et al.[30] The same situation was observed also for the high pressure end of the studied interval, i.e. pressures of 500 and 600 MPa.

Simulations with the bigger water box were carried out for all the pressures under study (0.1, 100, 200, 300, 400, 500, 600 MPa) for the purpose of the comparative study of the different physical properties of the HIV-1 PR dimer. In order to select sufficiently equilibrated states of the protein at different pressures, conformational changes of the dimer and the backbone root mean square deviation (rmsd) were analyzed. Closed conformation states were chosen as starting points of the comparative study of all the investigated parameters as they maintain the best conformational similarity over the whole pressure range. In addition, their ability of water repulsion is stronger in comparison with the semi-open and open conformations, which makes them more suitable for the calculations regarding hydrophobicity.

Fig. 1 shows a typical closed conformation of the protein dimer and Fig. 2 the evolution of a typical backbone rmsd of the dimer. As in the case of the smaller water box, the individual conformations are stable for several nanoseconds and are separated by relatively sharp transitions.

Pressure induced structural changes of the HIV-1 protease dimer were investigated by the equilibrium MD simulations at an interval of 0.1 to 600 MPa. Evaluation of the characteristic
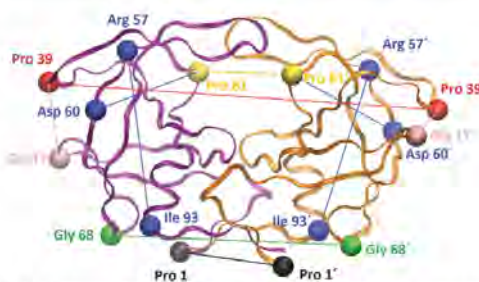
Fig. 1 Distances of α-carbons of selected amino acids in the molecule of the HIV-1 protease dimer.

95
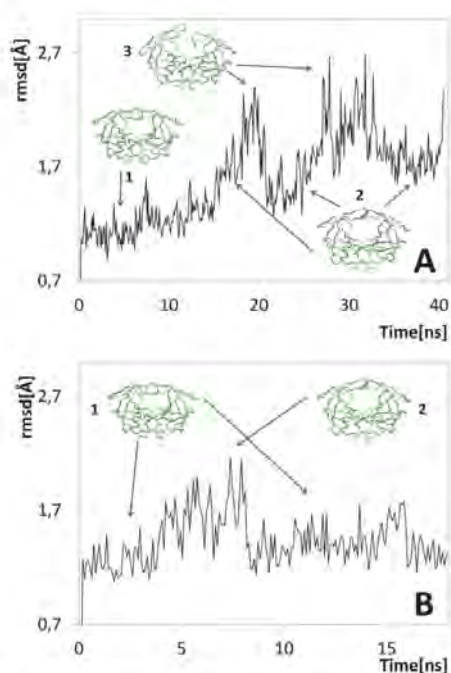
Fig. 2 Root mean square deviation (rmsd) of a typical dimer simulation at atmospheric pressure. The stable regions of individual conformations as well as the transitions between them can be seen. (A) Equilibrium of all three conformations, closed (1), semi-open (2) and open (3), during a 40 ns run. (B) Equilibrium between closed (1) and semi-open (2) conformations in a different simulation. After the return of the system to the closed conformation (in approx. 8 ns), Na⁺ ion was bound by Asp25 and 25' (in approx. 12 ns), which stabilizes the closed conformation.

Fig. 3 Distance of Pro39–Pro39' α-carbons and the radius of gyration of the dimer. Standard deviations of the mean of the individual points do not exceed 0.1% of the absolute values as a consequence of averaging over large data sets. Minor inhomogeneities in the trends are caused by small differences in the protein conformations in different simulations.

quantities were performed on well equilibrated parts of the simulation runs of 4.5 ns for the dimer (closed conformation) and 9 ns for the monomer. Fig. S1 and S2 (ESI†) show the rmsd profile of these parts of the simulations. It can be seen that the dimer structure maintains the rmsd values generally below 1.5, which indicates a considerably rigid structure. In contrast, monomer structure seems to be somewhat looser with higher rmsd fluctuations but without a significant rmsd drift. Volume, pressure and temperature fluctuations during the course of the simulation are shown in Fig. S3 (ESI†). The volume and pressure fluctuations are considerably lower than the differences of these quantities calculated for an arbitrary pair of the selected pressures. The amplitude of the volume fluctuations decreases with growing pressure, while the fluctuations of pressure and temperature are pressure independent.

In accord with the previous study[29] a tiny shrinkage of the structure can be observed at high pressure, but no dramatic change can be seen within the typical simulation time. The general change in the structure can be characterized by the pressure dependence of the radius of gyration (Fig. 3). It decreases

slowly and monotonously across the whole pressure range by approx. 2.0%. In order to find out which parts of the molecule are mostly responsible for this change, several linear dimensions between topologically significant atoms of the molecule were measured. The distances are shown in Fig. 1 and their values are listed in Table 1. It can be seen that the distance of the α-carbons of prolines 39 and 39' (the non-primed and primed numbers are used to distinguish between the two monomers) situated at the outermost points of the structure in a turn familiarly called "ear" by Perryman et al.,[31] decreases in a remarkable accordance with the radius of gyration (Fig. 3). The distance shrinks from the original 53.8 Å at atmospheric pressure to 51.8 Å at 600 MPa, i.e. by approx. 3.7%. Distance shortening is observed also for the "ear" and "cheek", i.e. Gly17 and Pro39 of the same subunit, but the data are rather scattered and the trend is not so clear. In contrast to this, other distances between significant points of the structure do not show any remarkable change (Table 1). This is, for instance, the case of the distance between Gly68 and 68' situated at the turn close to the dimerization domain, and the distance between Pro1 and 1' in the dimerization domain at the "whiskers". Both do not show any noticeable trend and only randomly oscillate around their mean values. Even the distance between prolines 81 and 81' located directly at the inner surface of the active-site cavity does not express a monotonous trend, but rather big oscillations during the time course of the simulation. Hence, the molecule seems to be a relatively rigid structure changing its conformation only in the less firmly fixed parts of the surface. The overall structure of the molecule resists even very high pressures without significant changes. This is particularly true for the dimerization domain composed of the intertwined ends of both monomer subunits. Even the active-site cavity is not much sensitive to the pressure, but its dimensions change with the conformation oscillations that occur during the time at any selected pressure.

Two characteristic distances were also determined for the monomer (blue lines in Fig. 1). One of them, the distance between the α-carbons of Arg57 and Ile93, runs from the flap to the dimerization domain, while the other one, Asp60 to Pro81

25908 | Phys. Chem. Chem. Phys., 2014, 16, 25906–25915

This journal is © the Owner Societies 2014

96

Table 1 Distances of α-carbons of selected amino acids in the molecule of the HIV-1 protease dimer and monomer

Pressure [MPa]	Distance [Å]								
	Dimer							Monomer	
	Pro39–Pro39' ("ears")	Pro81–Pro81' (cavity)	Gly68–Gly68'	Gly17–Pro39 ("cheek"–"ear")	Pro1–Pro1' ("whiskers")	Asp60–Pro81	Arg57–Ile93	Asp60–Pro81	Arg57–Ile93
0.1	53.83	21.15	35.55	12.37	17.51	21.8	23.9	21.8	23.8
100	53.83	21.50	36.63	11.95	16.47	21.7	23.7	21.6	23.4
200	52.80	21.05	34.62	11.77	17.38	21.7	23.5	21.8	23.3
300	52.76	20.87	34.88	11.97	17.80	21.7	23.5	21.7	23.0
400	52.46	20.51	34.33	11.41	17.57	21.7	23.4	21.7	23.5
500	51.96	19.37	34.88	11.66	16.91	21.6	23.0	21.3	23.3
600	51.79	22.15	34.79	11.58	17.38	21.5	23.1	21.5	23.7

α-carbon distance, is approximately perpendicular to it. These two distances (Table 1) are almost pressure independent with only tiny fluctuations around their mean values. Moreover, there is no change between these distances in the dimer and monomer structures. At 600 MPa partial distortion of the geometry can be observed which may indicate initiation of the pressure induced unfolding of monomers. The monomer molecule thus seems to be a well-defined structure keeping its conformation irrespective of the dimerization state. This observation, therefore, is in agreement with the NMR study of Ishima et al.[6] that confirmed the existence of stable free HIV-1 PR monomers, and even extends this conclusion to the high-pressure conditions. However, at some simulation runs small structural distortions in the monomer structure were observed for 600 MPa indicating the possible initiation of the pressure induced unfolding of the monomers.

Water-accessible surface area of the monomer and dimer molecules was calculated as an auxiliary geometrical indicator. The result is shown in Fig. 4. For the dimer the decrease of this area is monotonous within the whole pressure range, while for the monomer the decrease is much lower and biased by fluctuations. At 600 MPa it decreases more significantly, probably as a consequence of the partial structural collapse described previously as a possible initiation of a pressure-induced unfolding of a monomer. The size of water-accessible surface area reflects the slight shrinkage of the dimer molecule indicated by the radius of gyration and the Pro39–Pro39' distance, but decreases only insignificantly for the monomer. This is in accord with the

other geometrical parameters and confirms the low adaptability of both the structures to high pressure.

2.2. Interaction of HIV-1 PR with water and water density in the active-site cavity

Interaction of the protein molecule with water may play the principal role in the dimer dissociation and structural changes in general. To investigate this effect, calculations of water density in the closed cavities (defined in Methods section) of a dimer and monomer were carried out for the chosen pressures. For every pressure, a set of calculations was carried out with varying radius of the boundary sphere from 6 to 20 Å for the dimer and from 6 to 16 Å for the monomer. In these calculations the number of water molecules in the closed cavity was determined as well as the number of water molecules in the same closed cavity projected to a region of bulk water. The situation is depicted in Fig. 5. The ratio of these numbers represents the relative density of water in the closed cavity and is shown in Fig. 6 as a function of the boundary sphere radius. For small radii the cavity-water density is higher than that of free water but with growing radius it decreases quickly and the ratio gets reversed. At a certain radius, roughly corresponding with the real dimensions of the cavity, the density reaches its minimum value and then slowly grows to the bulk-water value as a consequence of the decreasing influence of the protein inside the relatively big sphere. The initial decrease can be

Fig. 4 Water accessible surface area (WASA) of the monomer and dimer. The value for the monomer is multiplied by 2. Standard deviations of the mean of the individual points do not exceed 0.1%.
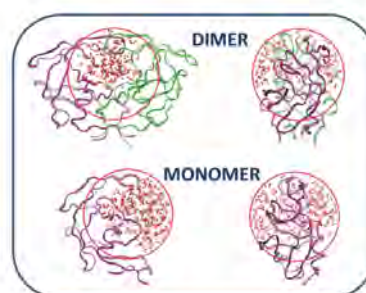
Fig. 5 Water molecules (red) inside the closed cavity of a radius 14 Å in the molecules of the HIV-1 PR dimer and monomer. Left images – front view, right images – side view.

This journal is © the Owner Societies 2014

Phys. Chem. Chem. Phys. 2014, **16**, 25906–25915 | 25909

97
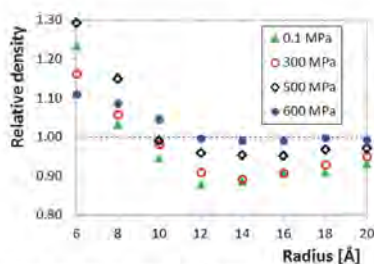
Fig. 6 Relative density of water as a function of the boundary-sphere radius for selected pressures. Standard deviations of the mean of the individual points do not exceed 1.2% for the biggest sphere and 4.5% for the smallest sphere.

Fig. 7 Relative density of water in the closed cavity of the dimer and monomer. Relative density is calculated as a ratio of the number of water molecules in the respective closed cavity and the same closed cavity projected to the bulk-water region. Standard deviations of the data are indicated by error bars.

Table 2 Relative density of water in the closed cavity of the dimer with a boundary-sphere radius of 6 Å.

Pressure [MPa]	Relative density
0.1	1.23
100	1.17
200	1.13
300	1.16
400	1.20
500	1.29
600	1.11

attributed to the electrostriction effect that concentrates water molecules around the charged aspartate residues. Indeed, the water density in the close proximity of the active-site aspartate residues exceeds significantly the bulk-water density and does not show any clear trend with changing pressure. Table 2 shows the relative density of water in a small closed cavity of a boundary-sphere radius of 6 Å. In contrast, repulsion of water from the hydrophobic amino-acid residues constituting the inner surface of the cavity lowers the density of water at its walls. The minimum value was, therefore, taken as the average water density inside the cavity. Fig. 7 shows that for the pressure up to about 200 MPa the density of water remains approximately constant at about 87% of the bulk water density. When the pressure is continuously increased up to 600 MPa, the density grows rapidly towards the bulk-water value.

An analogous calculation for the HIV-1 PR monomer revealed a similar trend, but the curve is shifted to much lower values of pressure. At atmospheric pressure the density is at about 96% of the bulk-water value and grows rapidly with increasing pressure approaching the bulk-water density closely between 100 and 200 MPa (Fig. 7).

These results indicate that the strongly curved hydrophobic inner surface of the active-site cavity repels some water from the interior. This effect remains almost constant up to the pressure of 200 MPa showing that the cohesive forces among the water molecules are strong enough to oppose the pressure tending to push them into the hydrophobic regions of the cavity. However, when the pressure is increased even higher, the cohesion of water is no more able to stand the external force
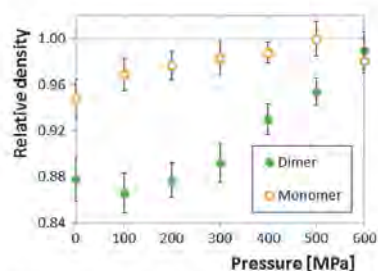
and water begins to fill the cavity continuously. At the highest pressure considered in the simulation, 600 MPa, the external forces are so strong that there is no difference between the water inside and outside the cavity. For the monomer the dependence is much weaker and the steepest increase of the density occurs at a much lower pressure. The curvature of the concave shape remaining after splitting the active-site cavity to two parts is thus insufficient to facilitate an equally strong effect as the dimer cavity, but is still capable of weaker repulsion of water molecules, which is almost negligible above 200 MPa.

An interesting comparison of monomer's and dimer's inter-action with water follows from relating the water-accessible surface area to the number of water molecules located up to 3 Å from the surface. At lower pressures the dimer value is slightly higher while at higher pressures the values for the dimer and monomer approach one another (data not shown). It can be explained by the stronger repulsion of water from the hydrophobic cavity of a dimer than from the monomer surface, which is in agreement with the calculation of average water density at both the molecular forms.

Additional calculations were carried out in order to evaluate the ratio of the numbers of water molecules located in the proximity of the hydrophilic and hydrophobic parts of the surface. Fig. 8 presents this quantity for both the dimer and the monomer. Although it decreases almost equally for monomer and dimer up to 100 MPa, probably due to hydration of analogous surface regions, a principle difference occurs at this pressure. For the monomer the rapid decrease continues reaching its minimum value at 200 MPa and then remains constant; for the dimer the decrease is rather slow with the steepest descent between 300 and 400 MPa. The observed behavior corresponds well with the calculated pressure profiles of monomer- and dimer-associated water densities. Hence, it can be assumed that the hydrophobic cavity has the most significant influence on the redistribution of water in the surroundings of the dimer and is mostly responsible for the total-volume change of the dimer to monomer dissociation.

2.3. Volume change of dimer dissociation

A principal point regarding the relationship between computer simulations and the experimental approach of dimer–monomer

25910 | Phys. Chem. Chem. Phys. 2014, 16, 25906–25915

This journal is © the Owner Societies 2014

98
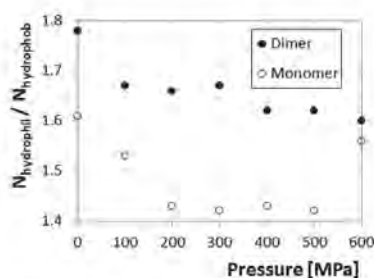
Fig. 8 Ratio of water molecules at the hydrophilic and hydrophobic parts of the HIV-1 PR dimer and monomer surfaces. Standard deviations of the mean of the individual points do not exceed 0.3%.

equilibrium under high pressure is the sign and magnitude of the volume change ΔV accompanying the dissociation of the dimer to monomers, because it is a measure of the pressure dependence of the equilibrium constant of this process. Considering the rigidity of both the dimer and monomer structures, this volume change is presumably facilitated especially by interaction of the protein molecule with water. As discussed previously, the water density inside the active-site cavity can strongly depend on pressure forcing the molecules inside against the repelling hydrophobic force.

Direct calculation of the volume change is rather troublesome as it is too small in comparison with the volumes of both the solvated dimer and monomers and even the volume fluctuations of the system. Nevertheless, an attempt was made to calculate it as a difference of PMVs of the dimer and two monomers. The resulting partial molecular volumes of both the systems are shown in Fig. S4 (ESI†). For both the dimer and the monomer it decreases along with the growing pressure in a similar way in the order of hundreds to thousands Å^3. This decrease is likely caused by the reorganization of the solvation sphere of the protein molecules, especially at the hydrophobic parts of their surfaces. Fig. S3 (ESI†) shows the time development of the volume of the water box of the HIV-1 PR dimer at various pressures. Obviously, the obtained partial molecular volumes of both the systems (dimer and two monomers) are significantly higher than the volume fluctuations of the water box, which allows us to calculate the partial molecular volumes with sufficient accuracy. However, the resulting molar volume changes ΔV_{PMV} are biased by rather high standard deviations which make the trends of this quantity unclear (Fig. S5, ESI†). Nevertheless, their values are in the order of tens of ml mol^{-1} for all the pressures under study, which is in agreement with the analogous experimental data for different proteins related to their molecular weight.[32,33]

To circumvent this problem of high volume fluctuations, a different approach was used that is focused on the local surroundings of the active-site cavity. The calculation is based on the assumption of different numbers of water molecules in the real closed cavity and in the closed cavity projected to free water region. The latter state simulates a hypothetical ideal

solution in which the interactions of a water molecule with its surroundings are equal at any point of the system. In such a state water surrounds the protein uniformly, therefore the total volume of a system composed of an HIV-1 PR molecule and a certain number of water molecules is equal in both dimeric and monomeric states provided that the monomer shape is unchanged in both the states. When some molecules are repelled from the protein's closed cavity by the real intermolecular interactions, each of them increases the volume of the surrounding water by the volume it occupies in the bulk water. In order to eliminate the influence of the outer surface of the molecules that is assumed to be equal for the dimer and free monomers, the closed cavities of both the monomer and the dimer were composed of a boundary sphere of equal radius of 14 Å, which corresponds to the minimum of the relative water density in the dimer cavity for most of the pressures. Table 3 summarizes the input data of the calculation of ΔV and its resulting pressure dependence is shown in Fig. 9A. The volume change is positive (i.e. the dissociation is accompanied by volume increase) in the lower part of the pressure interval but decreases rapidly with growing pressure and turns to negative values at about 80 MPa. It reaches a minimum at 300 MPa and then grows back to zero or even small positive values at the highest pressure considered.

The described behavior of ΔV can be explained as follows. The curved inner surface of the dimer cavity has a higher ability to repel water than only its half contained by the monomer. The monomer curvature is obviously insufficient to counteract pressure above 200 MPa, but still it can repel water at lower pressures. Dimer cavity, on the other hand, is almost equally hydrophobic up to 200 MPa above which the water-repelling ability decreases, too. The minimum of ΔV occurs in a point where the monomer surface is almost completely filled with water, while the cavity of the dimer is still highly hydrophobic. At low pressures the number of molecules repelled by the dimer is approximately equal to or even slightly lower than the number of molecules repelled by two monomers. This might

Table 3 Input parameters of the calculation of the volume change of HIV-1 PR dimer dissociation as a function of pressure. The data were calculated for the boundary sphere radius of 14 Å where the water density in the dimer closed cavity reached its minimum

Pressure [MPa]	$V_{\text{m,H}_2\text{O,bulk}}{}^{a}$ [ml mol^{-1}]	$\Delta N_{\text{H}_2\text{O,dim}}{}^{b}$	$\Delta N_{\text{H}_2\text{O,mon}}{}^{c}$
0.1	18.08 (18.14)	14.7	11.0
100	17.21 (17.46)	18.7	7.3
200	16.67 (16.92)	15.7	5.3
300	16.19 (16.49)	14.7	4.0
400	15.93 (16.13)	9.7	2.7
500	15.64 (15.81)	6.0	0.0
600	15.42 (15.52)	1.3	4.3

^{a} The molar volume of free water. The calculated values agree well with the experimental data[34] (in parentheses) and thus confirm the reliability of the used model of water. ^{b} The number of water molecules inside the closed cavity of the dimer projected to the bulk-water region minus the number of water molecules in the real closed cavity of the dimer. ^{c} The number of water molecules inside the closed cavity of the monomer projected to the bulk-water region minus the number of water molecules in the real closed cavity of the monomer.

99
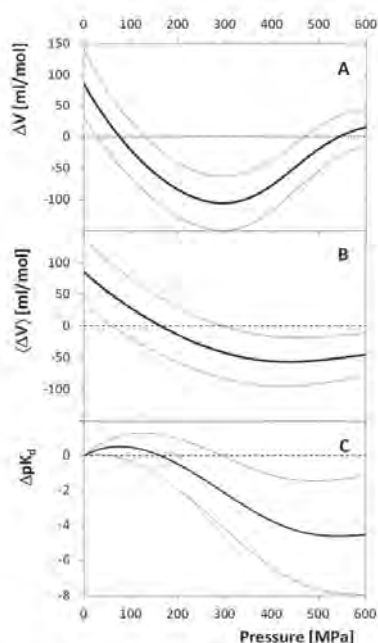
Fig. 9 Volume change of the dimer dissociation and the shift of equilibrium constant K_d as a function of pressure. Standard deviations are indicated by dotted lines. (A) Volume change of the dimer dissociation. (B) Average volume change in the interval from atmospheric to the given pressure. (C) The shift of the equilibrium constant expressed in the logarithmic scale as ΔpK_d

be a consequence of strong attraction of water by the two active-site aspartate residues bearing two closely located negative charges. Therefore, ΔV can reach even positive values. In contrast, the positive ΔV at the upper end of the pressure range might be rather a consequence of a partial structural distortion of the monomer that occurs at 600 MPa.

A common practice in the high-pressure experimental studies is to consider the volume change to be constant over the whole pressure range, often due to the limited sensitivity of the experimental devices. In order to enable the comparison of the current simulations with experiment, mean values of ΔV, defined as

$$\langle \Delta V \rangle = \frac{\int_{p_{atm}}^{p} \Delta V(p')\mathrm{d}p'}{p - p_{atm}}. \tag{1}$$

were calculated. Fig. 9B shows that the dependence of $\langle \Delta V \rangle$ on pressure is smoother than that of ΔV itself. The transition from positive to negative values occurs at 160 MPa. Above this pressure $\langle \Delta V \rangle$ stays in the order of tens of ml mol^{-1}, mostly between 20 and 60 ml mol^{-1}. This value corresponds well with our so far unpublished experimental data indicating ΔV of the same order of magnitude.

A principal experimental quantity that depends on ΔV is the equilibrium constant K_d. If ΔV is positive, high pressure shifts K_d to lower values, i.e. stabilizes the dimeric state, while with negative ΔV the situation is opposite. The shift of the equilibrium constant expressed as $\Delta pK_d(p)$ was evaluated according to eqn (14) and is plotted in Fig. 9C. At lower pressures up to 160 MPa $\Delta pK_d(p)$ is slightly positive, the increased pressure therefore stabilizes the dimer in this range. However, at higher pressures this quantity is negative with significantly higher absolute values indicating a strong destabilization of the dimer at high pressure. High pressure can therefore help the experimental techniques to observe monomeric HIV-1 PR at higher concentrations and thus with a better signal to noise ratio.

Standard errors shown in Fig. 9 indicate some inaccuracy in the determination of ΔV and ΔpK_d, which is a consequence of random fluctuations in the number of water molecules in the individual simulation frames. However, in spite of this the trends remain clear and the resulting curves provide a reasonable estimate of the high-pressure behavior of HIV-1 PR.

2.4. Protein–ion interactions

An interesting phenomenon is the stabilization of the active site by interaction of the two aspartate residues with a sodium cation. It was shown previously[35] and was also confirmed by our calculations that the interaction is highly favourable at atmospheric pressure as the interaction partially compensates the strong repulsion of the negatively charged aspartate residues. Thus, even if the simulation starts from a state without a sodium cation bound in the active site, it is often trapped by the aspartates and the system reaches equilibrium in this state. In contrast, for the high pressure of 500 or 600 MPa the tendency of binding the cation is much weaker – in some runs of the simulation it is not trapped at all, while in others the cation sits to one of the residues, stays there for some time and then leaves. Only rarely the firm complex of the sodium cation with both Asp residues is formed. Moreover, in some cases the escape of the cation from the bound state was observed. It indicates that the equilibrium between the states of the free active site and the active site bound with the cation is not strongly shifted to one or the other side at high pressure. This can be explained in the following way. The free Asp residues of the active site strongly attract the polar water molecules in order to compensate the mutual repulsion. This leads to the formation of a dense solvation sphere around these two residues, as can be seen in Table 2. When the ion is trapped by the two residues, it strongly contributes to the charge compensation and some water molecules leave the solvation sphere. At low pressure both the enthalpic contribution and the entropy increase of the released water molecules strongly favour the ion trapping by the aspartates. However, at high pressure this stabilizing effect can be overwhelmed by the influence of the volume growth when the water molecules pass from the highly organized solvation sphere to the bulk water of lower density. Therefore, formation of the Asp25–Asp25'–Na$^+$ complex can be less favourable at high pressures. At intermediate pressures no apparent trend of the tendency of the active site to bind the

25912 | Phys. Chem. Chem. Phys. 2014, 16, 25906–25915

This journal is © the Owner Societies 2014

cation was observed within the time scale of the simulation, but this phenomenon deserves further investigation.

3. Methods

3.1. Model construction and MD equilibrium simulations

The structural stability of a HIV-1 PR dimer and monomer was studied by all-atom molecular dynamics (MD) simulations, where the solvent molecules were represented by an explicit solvent model. Simulations were performed in the pressure series of 0.1, 100, 200, 300, 400, 500, and 600 MPa at a constant temperature of 310 K. The model was based on the X-ray crystal structure obtained from the RCSB protein data bank (PDB ID: 1OHR).[36] The inhibitor and the crystallographic water molecules were deleted from the source PDB file and hydrogens were added. A topology file has been constructed using an NAMD automatic PSF builder. The dimer model was solvated with TIP3P[37] water and neutralized by adding 4 Cl⁻ ions. NaCl concentration was set to 0.20 mol l⁻¹. Initial simulations of the dimer aimed at demonstration of the equilibrium properties of the system were performed with a water box of the dimensions 74.2 × 57.2 × 57.5 Å³ ("smaller water box") for 54 ns and the pressures of 0.1 and 500 MPa. In the case of the following simulations used for the calculation of the studied molecular properties the water box of the dimensions of 115.0 × 97.8 × 98.1 Å³ ("bigger water box") contained 3128 atoms of the protein, 35 664 water molecules, 134 Na⁺ ions and 138 Cl⁻, *i.e.* 110 392 atoms in total. The monomer model was solvated by TIP3P water molecules and NaCl concentration was set to 0.20 mol l⁻¹, *i.e.* equally as in the dimer case. The water box of the dimensions of 96.9 × 84.3 × 97.8 Å³ contained 1564 atoms of the protein, 26 169 water molecules, 99 Na⁺ ions and 101 Cl⁻ ions, *i.e.* 80 271 atoms in total.

In accord with previous high-pressure molecular-dynamics studies,[24–26] all simulations were performed using NAMD Version 2.9 program package[38] with the CHARMM22/CMAP topology and force-field parameters.[39,40] A timestep of 1 fs, 10 Å cutoff of the non-bonding interactions and a switching function starting at a distance of 8 Å were used. Full electrostatic calculations were performed using the Particle Mesh Ewald method (PME) implemented in the NAMD package. Native contacts between pairs of atoms (i,j) with $|i - j| < 4$ were discarded from the native contact list (1–4 scaling). The energy of the system was minimized for 180 fs prior to each MD simulation, which was subsequently carried out at the constant temperature of 310 K and the selected pressure using periodic boundary conditions in the isobaric-isothermal (*NPT*) ensemble. The temperature was controlled using Langevin dynamics and the pressure was controlled using the Langevin piston Nosé–Hoover method. The piston oscillation period was 100 fs, the barostat damping decay was 50 fs. The length of the simulation runs was between 18 and 40 ns for different systems.

Interatomic distances and other geometric properties were evaluated using the VMD 1.9.1 program[41] as average values over all frames of the selected part of the simulation where the system stayed in the closed conformation for a sufficient time period.

3.2. Water density and dimer-dissociation volume change calculation

In order to calculate the ratio of water density inside the active-site cavity and the density of bulk water located out of the reach of the protein–water interactions (hereafter called "relative density"), a boundary sphere of a selected radius centered in the geometric center of the α-carbons of Asp25, Ala28 and Pro79 of both chains was constructed in order to define the active-site cavity borders. Analogously, a boundary sphere centered in the geometric center of the α-carbons of Asp25, Ala28, Ile50 and Pro79 was constructed around the concave shaped half-cavity of a monomer. The boundary of the region of interest (hereafter called "closed cavity") is then formed by either the van der Waals surface of the protein atoms or the boundary-sphere surface depending on what is closer to the sphere centre in a given radial direction. The atoms of the water molecules inside the closed cavity are counted. In case that a molecule sticks out of the cavity through the boundary sphere, only the inner atoms are taken into account. In a simultaneous calculation the closed cavity was projected geometrically to an arbitrarily chosen location in which the bulk water was unperturbed by the interactions with the protein. The atoms of water-molecules located inside the projected closed cavity were counted. In this case, molecules protruding through the boundary-sphere surface were treated as previously, but those protruding through the van der Waals surface of the protein atoms were excluded from the count completely. The number of water molecules in the projected and real closed cavities N_{proj} and N_{real}, respectively, were calculated as 1/3 of the water atoms present in the respective regions and their ratio, N_{real}/N_{proj}, which is equal to the relative density of water in the closed cavity, was used as a measure of the hydrophobicity of the cavity. Furthermore, the difference

$$\Delta N_{H_2O,dim} = N_{proj,dim} - N_{real,dim} \qquad (2)$$

was defined, which is equal to the number of water molecules repelled by the hydrophobic surface of the cavity of the dimer with respect to a hypothetic reference state of no hydrophobic interaction, *i.e.* of equal density of water in every point of the system except the space occupied by the protein. Analogous quantities were defined for the monomer:

$$\Delta N_{H_2O,mon} = N_{proj,mon} - N_{real,mon}. \qquad (3)$$

Finally, the number of water molecules repelled by a dimer was subtracted from the number of molecules repelled by two monomers giving the change in the number of repelled molecules during the dimer-dissociation process:

$$\Delta N_{H_2O,dissoc} = 2\Delta N_{H_2O,mon} - \Delta N_{H_2O,dim} \qquad (4)$$

These quantities were evaluated by home-made program and were averaged over 10 frames selected from the well equilibrated parts of the simulation.

This journal is © the Owner Societies 2014

Phys. Chem. Chem. Phys., 2014, **16**, 25906–25915 | **25913**

101

The molar volume $V_{m,H_2O,bulk}$ of unperturbed bulk water at the respective pressure was calculated as an Avogadro-number multiplied ratio of the volume of an arbitrarily chosen sphere positioned in the unperturbed bulk-water area and the number of water molecules located inside it.

In order to evaluate ΔV as a function of pressure in the whole range of 0.1 to 600 MPa the functions $\Delta N_{H_2O,dim}(p)$ and $\Delta N_{H_2O,mon}(p)$ were fitted by suitable smooth functions

$$\Delta N_{H_2O,dim}(p) = Q\frac{e^{-a(p-b)}}{1+e^{-a(p-b)}} \qquad (5)$$

$$\Delta N_{H_2O,mon}(p) = Re^{-cp}, \qquad (6)$$

where the parameters a, b, c, Q, and R were obtained by means of non-linear regression. Accordingly,

$$\Delta N_{H_2O,dissoc}(p) = 2\Delta N_{H_2O,mon}(p) - \Delta N_{H_2O,dim}(p). \qquad (7)$$

In addition, the molar volume of bulk water was fitted by a cubic polynomial

$$V_{m,H_2O,bulk}(p) = A + Bp + Cp^2 + Dp^3. \qquad (8)$$

Non-linear regressions were performed using Wolfram Mathematica 9 program.

The resulting volume change of dimer dissociation $\Delta V(p)$ was calculated as

$$\Delta V(p) = \Delta N_{H_2O,dissoc}(p)V_{m,H_2O,bulk}(p). \qquad (9)$$

Given the volume change, pressure dependence of the equilibrium constant of dimer dissociation

$$K_d = \frac{M^2}{D} \qquad (10)$$

can be determined. As

$$K_d(p) = \exp\left(-\frac{\Delta G_r^0(p)}{RT}\right), \qquad (11)$$

where $\Delta G_r^0(p)$ is the standard reaction change of the Gibbs energy at pressure p, and

$$\Delta G_r^0(p) = \Delta G_r^0(p_{atm}) + \int_{p_{atm}}^{p} \Delta V(p')dp'$$
$$= \Delta G_r^0(p_{atm}) + \Delta G_{r,HP}^0(p), \qquad (12)$$

the pressure dependence of K_d is

$$K_d(p) = \exp\left(-\frac{\Delta G_r^0(p_{atm}) + \Delta G_{r,HP}^0(p)}{RT}\right)$$
$$= K_{d,atm}\exp\left(-\frac{\Delta G_{r,HP}^0(p)}{RT}\right). \qquad (13)$$

Here $\Delta G_{r,HP}^0(p) = \int_{p_{atm}}^{p} \Delta V(p')dp'$. It is convenient to present the equilibrium-constant shift in the logarithmic scale, i.e.

$$pK_d(p) = pK_{d,atm} + \frac{\Delta G_{r,HP}^0(p)}{2.303\,RT} = pK_{d,atm} + \frac{\int_{p_{atm}}^{p}\Delta V(p')dp'}{2.303\,RT}, \qquad (14)$$

where the pK quantities are conventionally defined as $-\log_{10}K$. The difference in pK_d for atmospheric pressure and given pressure p denoted $\Delta pK_d(p)$ is thus

$$\Delta pK_d(p) = \frac{\int_{p_{atm}}^{p}\Delta V(p')dp'}{2.303\,RT}. \qquad (15)$$

3.3. Calculation of partial molar volumes

Total volumes of water boxes of the dimer, the monomer and free water were calculated averaging the value over 5000 frames for the dimer and monomer and 1000 frames for water. The water box of free water contained equal number of water molecules and ions as the water box of the dimer except the ions compensating the protein charge. The average volume of one water molecule was evaluated. Partial molar volumes of the dimer and monomer were calculated as a difference in the total volume of the water box of the respective system and the volume of the number of free-water molecules equal to the number of water molecules in this water box. The volume change of dimer dissociation ΔV_{PMV} (the index "PMV" was used to distinguish this quantity from ΔV defined in Section 3.2.) was evaluated as

$$\Delta V_{PMV} = 2PMV(monomer) - PMV(dimer). \qquad (16)$$

4. Conclusions

High-pressure responses of the HIV-1 PR dimer and monomer were investigated by molecular dynamics methods. Within the simulation time scale the conformations of both the forms are stable up to 500 MPa with only minor changes in the outermost surface domains around Pro39. Partial structural distortion of the monomer can be seen at 600 MPa. With growing pressure the water molecules approach the protein surface even at the hydrophobic regions. An interesting consequence of this phenomenon is observed at the active-site cavity. There, the density of water is approximately at 87% of the bulk-water value up to 200 MPa, after which it starts to grow up reaching the bulk-water density at about 600 MPa. A similar, but weaker effect can be seen also at the monomer molecule. The difference in both these effects gives rise to the volume change of dimer dissociation, which is positive at low pressures but changes the sign at 80 MPa, reaches minimum at 300 MPa and then returns to zero between 500 and 600 MPa. The dimer is thus slightly stabilized at the lower part of the pressure range but strongly destabilized at higher pressures. The equilibrium constant of dimer dissociation is thus shifted to higher values by high pressure, which can enable experimental studies of HIV-1 PR dimerization at higher concentrations and thus with

25914 | Phys. Chem. Chem. Phys. 2014, 16, 25906–25915

This journal is © the Owner Societies 2014

improved accuracy of the detection methods. Generally, high-pressure methods are helpful tools for the investigation of oligomerization equilibria of proteins possessing hydrophobic cavities at the interfaces of their subunits.

Acknowledgements

References

1 M. Kožíšek, J. Bray, P. Řezáčová, K. Šašková, J. Brynda, J. Pokorná, F. Mammano, L. Rulíšek and J. Konvalinka, *J. Mol. Biol.*, 2007, **374**, 1005–1016.

2 M. Kožíšek, K. G. Šašková, P. Řezáčová, J. Brynda, N. M. van Maarseveen, D. De Jong, C. A. Boucher, R. M. Kagan, M. Nijhuis and J. Konvalinka, *J. Virol.*, 2008, **82**, 5869–5878.

3 A. Wlodawer and J. Vondrasek, *Annu. Rev. Biophys. Biomol. Struct.*, 1998, **27**, 249–284.

4 M. J. Todd, N. Semo and E. Freire, *J. Mol. Biol.*, 1998, **283**, 475–488.

5 V. Veverka, H. Bauerova, A. Zabransky, J. Lang, T. Ruml, I. Pichova and R. Hrabal, *J. Mol. Biol.*, 2003, **333**, 771–780.

6 R. Ishima, D. A. Torchia and J. M. Louis, *J. Biol. Chem.*, 2007, **282**, 17190–17199.

7 J. M. Louis, R. Ishima, A. Aniana and J. M. Sayer, *Protein Sci.*, 2009, **18**, 2442–2453.

8 Y. Levy, A. Caflisch, J. N. Onuchic and P. G. Wolynes, *J. Mol. Biol.*, 2004, **340**, 67–79.

9 M. C. Yan, Y. Sha, J. Wang, X. Q. Xiong, J. H. Ren and M. S. Cheng, *Proteins*, 2008, **70**, 731–738.

10 A. F. Noel, O. Bilsel, A. Kundu, Y. Wu, J. A. Zitzewitz and C. R. Matthews, *J. Mol. Biol.*, 2009, **387**, 1002–1016.

11 M. Cavallari, C. Ghio, S. Monti, M. Ferrario, A. Maritan and P. Carloni, *THEOCHEM*, 2006, **769**, 111–121.

12 M. Bonomi, A. Barducci, F. L. Gervasio and M. Parrinello, *PLoS One*, 2010, **5**, e13208.

13 H. J. Schramm, J. Boetzel, J. Buttner, E. Fritsche, W. Gohring, E. Jaeger, S. Konig, O. Thumfart, T. Wenger, N. E. Nagel and W. Schramm, *Antiviral Res.*, 1996, **30**, 155–170.

14 T. F. Holzman, W. E. Kohlbrenner, D. Weigl, J. Rittenhouse, D. Kempf and J. Erickson, *J. Biol. Chem.*, 1991, **266**, 19217–19220.

15 E. M. Towler, S. V. Gulnik, T. N. Bhat, D. Xie, E. Gustschina, T. R. Sumpter, N. Robertson, C. Jones, M. Sauter, N. Mueller-Lantzsch, C. Debouck and J. W. Erickson, *Biochemistry*, 1998, **37**, 17137–17144.

16 C. A. Pargellis, M. N. Morelock, E. T. Graham, P. Kinkade, S. Pav, K. Lubbe, D. Lamarre and P. C. Anderson, *Biochemistry*, 1994, **33**, 12527–12534.

17 P. L. Darke, S. P. Jordan, D. L. Hall, J. A. Zugay, J. A. Shafer and L. C. Kuo, *Biochemistry*, 1994, **33**, 98–105.

18 M. Ingr, T. Uhlikova, K. Strisovsky, E. Majerova and J. Konvalinka, *Protein Sci.*, 2003, **12**, 2173–2182.

19 E. Girard, S. Marchal, J. Perez, S. Finet, R. Kahn, R. Fourme, G. Marassio, A. C. Dhaussy, T. Prange, M. Giffard, F. Dulin, F. Bonnete, R. Lange, J. H. Abraini, M. Mezouar and N. Colloc'h, *Biophys. J.*, 2010, **98**, 2365–2373.

20 S. Marchal, A. Marabotti, M. Staiano, A. Varriale, T. Domaschke, R. Lange and S. D'Auria, *PLoS One*, 2012, **7**, e50489.

21 S. Sarupria, T. Ghosh, A. E. García and S. Garde, *Proteins*, 2010, **78**, 1641–1651.

22 D. Paschek and A. E. Garcia, *Phys. Rev. Lett.*, 2004, **93**, 238105.

23 N. Smolin and R. Winter, *BBA-Proteins and Proteomics.*, 2006, **1764**, 522–534.

24 T. Yamazaki, T. Imai, F. Hirata and A. Kovalenko, *J. Phys. Chem. B*, 2007, **111**, 1206–1212.

25 A. Paliwal, D. Asthagiri, D. P. Bossev and M. E. Paulaitis, *Biophys. J.*, 2004, **87**, 3479–3492.

26 J. A. Yancey, N. A. Vellore, G. Collier, S. J. Stuart and R. A. Latour, *Biointerphases*, 2010, **5**, 85–95.

27 E. Paci, *Biochim. Biophys. Acta*, 2002, **1595**, 185–200.

28 A. Lerbret, A. Hédoux, B. Annighöfer and M.-C. Bellissent-Funel, *Proteins*, 2013, **81**, 326–340.

29 B. R. Meher, M. V. Satish Kumar and K. Sen, in *ICIT 2008 Proceedings of the 11th international konference on information technology*, ed. B. S. Panda and A. Nayak, 2008, pp. 118–122.

30 V. Hornak, A. Okur, R. C. Rizzo and C. Simmerling, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 915–920.

31 A. L. Perryman, J.-H. Lin and A. McCammon, *Protein Sci.*, 2004, **13**, 1108–1123.

32 M. J. Kornblatt, R. Lange and C. Balny, *Eur. J. Biochem.*, 2004, **271**, 3897–3904.

33 K. C. Ruan and G. Weber, *Biochemistry*, 1988, **27**, 3295–3301.

34 C. W. Burnham, J. R. Holloway and N. F. Davis, *Am. J. Sci.*, 1969, **267-A**, 70–95.

35 D. Kovalskyy, V. Dubyna, A. E. Mark and A. Kornelyuk, *Proteins*, 2005, **58**, 450–458.

36 S. W. Kaldor, V. J. Kalish, J. F. Davies 2nd, B. V. Shetty, J. E. Fritz, K. Appelt, J. A. Burgess, K. M. Campanale, N. Y. Chirgadze and D. K. Clawson, *et al.*, *J. Med. Chem.*, 1997, **40**, 3979–3985.

37 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

38 J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781–1802.

39 A. D. MacKerell Jr., D. Bashford, M. Bellott, R. L. Dunbrack Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo and S. Ha, *et al.*, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.

40 A. D. MacKerell, M. Feig and C. L. Brooks, *J. Comput. Chem.*, 2004, **25**, 1400–1415.

41 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.

This journal is © the Owner Societies 2014

Phys. Chem. Chem. Phys., 2014, **16**, 25906–25915 | **25915**

P5. Inhibitor and Substrate Binding Induced Stability of HIV-1 Protease against Sequential Dissociation and Unfolding Revealed by High Pressure Spectroscopy and Kinetics.

M. Ingr, R. Lange, V. Halabalova, A. Yehya, J. Hrncirik, D. Chevalier-Lucia, L. Palmade, C. Blayo, J. Konvalinka, E. Dumay, Plos One 10 (2015) e0119099.

PLOS ONE

RESEARCH ARTICLE

Inhibitor and Substrate Binding Induced Stability of HIV-1 Protease against Sequential Dissociation and Unfolding Revealed by High Pressure Spectroscopy and Kinetics

Marek Ingr[1,2]*, Reinhard Lange[3], Věra Halabalová[1], Alaa Yehya[4], Josef Hrnčiřík[1], Dominique Chevalier-Lucia[5], Laetitia Palmade[5], Claire Blayo[5], Jan Konvalinka[2], Eliane Dumay[5]

1 Tomas Bata University in Zlín, Faculty of Technology, Department of Physics and Material Engineering, nám. T.G. Masaryka 5555, 76001 Zlín, Czech Republic, 2 Charles University in Prague, Department of Biochemistry, Hlavova 2030, 128 43 Prague 2, Czech Republic, 3 Université Montpellier 2, INRA UMR IATE, Biochimie et Technologie Alimentaires, cc023, Place Eugene Bataillon, 34095 Montpellier cedex 05, France, 4 INSERM U710, Université Montpellier 2, Place Eugene Bataillon, 34095 Montpellier cedex 05, France, 5 Université Montpellier 2, UMR IATE, Biochimie et Technologie Alimentaires, cc023, 2 Place Eugene Bataillon, 34095 Montpellier cedex 05, France

* ingr@ft.utb.cz

Abstract

High-pressure methods have become an interesting tool of investigation of structural stability of proteins. They are used to study protein unfolding, but dissociation of oligomeric proteins can be addressed this way, too. HIV-1 protease, although an interesting object of biophysical experiments, has not been studied at high pressure yet. In this study HIV-1 protease is investigated by high pressure (up to 600 MPa) fluorescence spectroscopy of either the inherent tryptophan residues or external 8-anilino-1-naphtalenesulfonic acid at 25°C. A fast concentration-dependent structural transition is detected that corresponds to the dimer-monomer equilibrium. This transition is followed by a slow concentration independent transition that can be assigned to the monomer unfolding. In the presence of a tight-binding inhibitor none of these transitions are observed, which confirms the stabilizing effect of inhibitor. High-pressure enzyme kinetics (up to 350 MPa) also reveals the stabilizing effect of substrate. Unfolding of the protease can thus proceed only from the monomeric state after dimer dissociation and is unfavourable at atmospheric pressure. Dimer-destabilizing effect of high pressure is caused by negative volume change of dimer dissociation of −32.5 mL/mol. It helps us to determine the atmospheric pressure dimerization constant of 0.92 μM. High-pressure methods thus enable the investigation of structural phenomena that are difficult or impossible to measure at atmospheric pressure.

Introduction

HIV-1 protease is a small aspartic protease active as a homodimer of the molecular weight of 21.6 kDa. The dimer is stabilized by a β-sheet domain formed by the intertwined N- and C- termini of the monomers [1]. It is expressed as a part of Gag-Pol polyprotein and is autocatalytically excised from this precursor. After this release the protease plays an important role in the process of maturation of the viral particle as it cleaves the viral polyprotein precursors Gag at five sites and Gag-Pol at eleven sites. This proteolytic processing produces the structural and enzymatically active viral proteins and enables the formation of mature infectious viral particle. Inhibition of HIV-1 protease prevents viral maturation which makes the enzyme an important therapeutic target in HIV/AIDS treatment [2,3].

Stability of the HIV-1 protease dimer is an important property determining the enzyme's activity, since only the dimeric form is enzymatically active. For quantifying its stability, many research groups have therefore conducted experiments for determining the equilibrium constant of dimerization (K_d) of HIV-1 protease [4–7] as well as of some other retroviral proteases and HIV-1-protease mutants [7–9]. However, the obtained values varied considerably depending on the method of determination. While some groups reported extremely stable dimer characterized by subnanomolar dissociation constant K_d [4], other teams determined values several orders of magnitude higher [9,10].

A probable explanation of these discrepancies is the high dimer stability which leads to the necessity of measurements in very low protease concentrations at the detection limit of most analytical methods. In addition, it is not possible to observe the association of monomers directly, because an isolated monomer is impossible to prepare. Thus, when kinetics of the processes of dissociation and association of the dimer is investigated, only the first process can be directly observed, which makes the results less reliable.

High pressure methods [11–14] represent a possible way how to circumvent the experimental troubles arising from the high dimer stability. When high pressure is applied to a physicochemical system, the equilibrium is shifted towards the state with lower total volume. As there is a general experience that the volume decreases when oligomeric proteins are dissociated, it can be expected that the HIV-1 protease dimer will dissociate under high pressure. Thus, the dimer dissociation will be able to be observed at considerably higher concentration. Moreover, if the process is reversible, it should also be possible to observe the association of monomers directly after pressure release.

High-pressure techniques are increasingly used to study unfolding and oligomerization equilibrium of many proteins [15–24]. The structural changes are most often studied by variety of spectroscopic methods. However, structural distortions can result also in different set of antibodies obtained from an experimental animal after immunization by antigen treated by high pressure [25]. High pressure methods can be used also for preparative purposes, e.g. the extraction of proteins for LC-MS analysis from formaldehyde-fixed tissue samples [26]. In some cases the transition curves of dimer or oligomer dissociation were used to evaluate the volume change of this process and the atmospheric-pressure equilibrium constant [20–22]. Recently, we applied this approach to study bovine odorant-binding protein (bOBD) of the lipocalin family, which appears to share some structure-function features with HIV-1 protease. Although not enzymatically active, it is a small homodimer able of binding its ligand only in dimeric state. High-pressure methods in the range up to 600 MPa have revealed that the dissociated monomers unfold at lower pressure than a dimer without the bound ligand, while ligand binding shifts the unfolding threshold even higher [23]. Thus, dimerization stabilizes the structure of the protein which is even more stabilized by ligand binding.

105

Similar behaviour can be expected also in the case of HIV-1 protease. To our knowledge, however, high pressure studies with retroviral proteases have not been carried out yet. Thus, it is not known whether a reversible equilibrium of dimer and folded monomers occurs at high pressure or the dimer dissociation is directly coupled with unfolding. Recent NMR studies with mutants of HIV-1 and HIV-2 proteases with defects in dimerization domains [27,28] indicate that folded monomers can exist, but they do not allow the observation of dimerization. Another question is whether the dimer's interaction with substrate or inhibitor stabilizes the molecule with respect to the high-pressure dissociation and unfolding. Furthermore, dimerization equilibrium constant K_d and the rate constants of dimer dissociation and association under high pressure have not been measured yet. Our intention is therefore to study HIV-1 protease under varying pressure in order to observe the monomer-dimer equilibrium and unfolding and to contribute to understanding its equilibrium and kinetics properties.

Methods and Material

HIV-1 protease expression and purification

HIV-1 protease purification procedure was modified according to [10]. It was produced by heterologous expression in *Escherichia coli* strain BL21(DE3)RIL from a gene cloned in a plasmid pET11c. Four 2L flasks with 0.5 L LB medium containing 0.1 mg/mL ampicillin were inoculated by the cells harbouring the plasmid and the cultivation was carried out until $OD_{600} = 1.0$. The protein expression was then induced by the addition of 0.25 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and cultivated for 3 h. Cells were harvested by centrifugation, washed by PBS buffer, resuspended in buffer A (50 mM Tris, 50 mM NaCl, 1 mM EDTA, pH 8.0) and subjected to three freeze-thaw cycles. Cells were further disintegrated by sonication in five 1-min cycles on ice. Cell debris was centrifuged for 30 min, 15000 g, 4°C, the supernatant was removed and the peleted inclusion bodies were resuspended in buffer SA (A + 1 M NaCl) and sonicated in three 1-min cycles on ice. After centrifugation for 20 min, 15000 g, 4°C the previous washing procedure was repeated in TA buffer (A + 1% Triton X-100). One half of the inclusion bodies were resuspended in 2 mL of water and dissolved in 4 mL of 99% acetic acid, then it was diluted 30 times by 0.05% 2-mercaptoethanol in water and dialyzed against 2 L of the same solution. The dialysis continued against three 2L pools of buffer 20 mM Tris, 20 mM MES, 10% glycerol, 0.05% 2-mercaptoethanol, 1 mM EDTA, pH 6.7. Precipitate was removed by centrifugation for 10 min, 15000 g, 4°C and the supernatant was purified by cation-exchange chromatography on Capto S column (GE healthcare) with the binding buffer CA equal to the dialysis buffer and the elution buffer CB (CA + 1 M NaCl). Fractions eluted by the 30-min gradient of 0–100% CB containing the protease were pooled and further purified by gel permeation chromatography on Sephacryl S-100 16/60 (GE Healthcare) in buffer 20 mM CH$_3$COONa, 200 mM NaCl, 10% glycerol, 0.05% 2-mercaptoethanol, 1 mM EDTA, pH 5.0 (hereafter called "assay buffer"). Fractions containing the protease were pooled and concentrated on Amicon centrifugation cell with cut-off 10 kDa to the final value of about 0.5 mg/mL.

Theoretical description of the structural transitions

Monomer-dimer equilibrium at high pressure. At given temperature and pressure the equilibrium between dimer (D) and monomer (M)

$$D \rightleftarrows 2M \tag{1}$$

106

is described by an equilibrium constant K_d

$$K_d(p) = \frac{[M]^2}{[D]}. \tag{2}$$

Where $[M]$ and $[D]$ are the relative molar concentrations of monomer and dimer, respectively. If the total concentration of the protein expressed as monomer is M_0, the equilibrium monomer concentration is

$$[M] = \frac{-K_d + \sqrt{K_d^2 + 8K_d[M_0]}}{4}. \tag{3}$$

Under the varying pressure the equilibrium constant takes the form

$$K_d(p) = e^{\frac{-\Delta G_d^0(p)}{RT}} = e^{\frac{-\Delta G_{d,atm}^0 + \Delta V_d \Delta p}{RT}} = K_{d,atm} e^{\frac{-\Delta V_d \Delta p}{RT}}, \tag{4}$$

where $\Delta G_d^0(p)$ is the standard molar reaction change of Gibbs energy at certain pressure p and $\Delta G_{d,atm}^0$ is the same quantity at atmospheric pressure, $K_{d,atm}$ is the atmospheric pressure equilibrium constant and ΔV_d is the volume change of dimer dissociation which is considered to be constant within the whole pressure range. The symbol Δp stands for the difference of a certain pressure and atmospheric pressure.

Combining Eqs. (3) and (4) the expression for monomer concentration at given $[M_0]$ and Δp is obtained

$$[M] = \frac{K_{d,atm} e^{\frac{-\Delta V_d \Delta p}{RT}}}{4} \left(-1 + \sqrt{1 + \frac{8[M_0]}{K_{d,atm} e^{\frac{-\Delta V_d \Delta p}{RT}}}} \right). \tag{5}$$

Introducing the degree of dimer dissociation, or rather a monomer fraction, $\alpha_d = [M]/[M_0]$ Eq. (3) can be expressed as

$$\alpha_d(\Delta p) = \frac{K_{d,atm} e^{\frac{-\Delta V_d \Delta p}{RT}}}{4[M_0]} \left(-1 + \sqrt{1 + \frac{8[M_0]}{K_{d,atm} e^{\frac{-\Delta V_d \Delta p}{RT}}}} \right), \tag{6}$$

which defines a sigmoid-like curve with the limits of 0 and 1 for $[M_0]$ tending to infinity and zero, respectively. After some algebraic rearrangements the position of inflex point of this curve can be derived:

$$\Delta p_{inf} = \frac{-RT}{\Delta V_d} \ln \frac{4[M_0]}{K_{d,atm}(1 + \sqrt{2})} = \frac{-RT}{\Delta V_d} \ln \frac{4[M_0]}{(1 + \sqrt{2})} - \frac{\Delta G_{d,atm}^0}{\Delta V_d} \tag{7}$$

Thus, Δp_{inf} can be used to determine the atmospheric-pressure equilibrium constant of dimerization $K_{d,atm}$ (or equivalently $\Delta G_{d,atm}^0$) and the molar change of volume ΔV_d, provided that the dependence of Δp_{inf} on $[M_0]$ described by Eq. (7) is known at least for two different total enzyme concentrations $[M_0]$. Measuring this curve for a series of concentrations, $K_{d,atm}$ and ΔV_d can be determined either from non-linear regression of this equation or with the aid

107

of the linearized formula

$$\Delta p_{\text{inf}}([M_0]) = \frac{-RT}{\Delta V_d} \ln[M_0] + \frac{RT}{\Delta V_d} \ln \frac{(1+\sqrt{2})K_{d,\text{atm}}}{4}, \tag{8}$$

Where $\frac{-RT}{\Delta V_d}$ is the slope and $\frac{RT}{\Delta V_d} \ln \frac{(1+\sqrt{2})K_{d,\text{atm}}}{4}$ is the Δp_{inf}-axis intercept of the resulting line. The slope of the curve in the inflex point is then given by

$$\frac{\partial \alpha_d}{\partial p}\Big|_{\Delta p_{\text{inf},d}} = \frac{\Delta V_d}{RT}(2\sqrt{2} - 3). \tag{9}$$

Monomer folding-unfolding equilibrium and kinetics. For this process described by an equation

$$M \underset{k_f}{\overset{k_u}{\rightleftarrows}} U. \tag{10}$$

Where U denotes the unfolded monomer and k_u and k_f are the rate constants of unfolding and folding processes, respectively, the inflex point of the dependence of degree of unfolding $\alpha_u = (M_0 - M)/M_0$ is given by the expression

$$\Delta p_{\text{inf},u} = -\frac{\Delta G_{u,\text{atm}}}{\Delta V_u} = \frac{RT}{\Delta V_u} \ln K_{u,\text{at}} \tag{11}$$

and the slope at the inflex point is

$$\frac{d\alpha_u}{d\Delta p}\Big|_{\Delta p_{\text{inf},u}} = -\frac{\Delta V_u}{4RT}. \tag{12}$$

Where ΔV_u is the volume change of this process considered as pressure independent, $\Delta G_{u,\text{atm}}$ is the change of the Gibbs energy of unfolding at atmospheric pressure and $K_{u,\text{atm}}$ is the atmospheric pressure equilibrium constant. The rate constants k_u and k_f were calculated based on the determined equilibrium parameters and the observed relaxation rate constant of the unfolding

$$k_{\text{obs}} = k_f + k_u \tag{13}$$

$$k_u(p) = k_{\text{obs}} \frac{K_u(p)}{K_u(p) + 1} \tag{14}$$

$$k_f(p) = k_{\text{obs}} \frac{1}{K_u(p) + 1}, \tag{15}$$

where

$$K_u(p) = K_{u,\text{atm}} e^{-\frac{\Delta V_u \Delta p}{RT}}. \tag{16}$$

108

High pressure fluorescence experiments

Fluorescence measurements were carried out at 25°C using an SLM Series 2 spectrofluorimeter (Aminco Bowman, Foster City, CA), modified to accommodate a thermostated high pressure optical cell. HIV-1 protease was diluted to various concentrations (1 to 25 μM dimer) in the assay buffer. For some experiments, inhibitor darunavir (120 μM) or 8-anilino-1-naphthalene-sulfonate (ANS) (700 μM) was added to HIV-1 protease solutions. The sample was placed in a 5 mm diameter quartz cuvette, closed at the top with a flexible polyethylene film that was attached by a rubber O-ring.

Pressure dependence of tryptophan fluorescence spectra. Emission spectra of both inhibited and non-inhibited protease were measured at the excitation wavelength of 280 nm (slit of 4 nm) in the emission-wavelength range of 300–400 nm (slit of 8 nm). A sample of HIV-1 PR was prepared diluting the stock solution by the assay buffer to the desired concentration from the range of 1 to 25 μM dimer, pipetted into the quartz cuvette and placed in the instrument. Initial pressure of 10 MPa was set, the sample was incubated for 3 min and then the fluorescence spectrum was recorded in triplicate and averaged. Afterwards the pressure was set to the next value and the procedure was repeated. A typical pressure series consisted of pressures of 10, 25 and 50 MPa and further pressures increasing by 25 MPa up to 500 MPa, in some cases 600 MPa, excitation wavelength was 280 nm and the emission wavelength range was 300–400 nm. Centre of spectral mass (CSM) and spectral integral (SI) defined as

$$SI = \int_{\lambda_i}^{\lambda_f} F(\lambda)d\lambda \tag{17}$$

$$CSM = \frac{\int_{\bar{v}_i}^{\bar{v}_f} v F(v)dv}{\int_{\bar{v}_i}^{\bar{v}_f} F(v)dv}. \tag{18}$$

were determined for every pressure point giving the dependencies of these quantities as functions of pressure. $F(\lambda)$ is the fluorescence intensity at the wavelength λ while λ_i and λ_f denote the lower and upper boundary of the spectral range and $\bar{v}$, $\bar{v}_i$ and $\bar{v}_f$ are the respective wavenumbers corresponding to the wavelengths by the relations $\bar{v} = c/\lambda$ and analogous. In addition, a set of differential spectra was calculated for every concentration. A spectrum for a given pressure was normalized to the spectral integral of the spectrum at 10 MPa and the 10-MPa spectrum was subtracted from it.

Determination of dimerization equilibrium constant and volume change. Experimentally determined curves of CSM vs. pressure were fitted by model functions of a sigmoidal shape in the form of

$$F(\Delta p) = I_0 - q\Delta p + \frac{I_f - I_0 + (q - m)\Delta p}{1 + e^{\frac{\Delta G - \Delta p \Delta V}{RT}}} \tag{19}$$

Where I_0, I_f, ΔG and ΔV are variable parameters of the fit and q and m were set firmly as a limit of the slope at p → −∞ or +∞, respectively. Eq. (19) was used as a suitable fitting function allowing the determination of inflex points, although in case of second-order transition the parameters ΔG and ΔV do not have the meaning of Gibbs energy and volume changes. Direct non-linear regression by a function based of Eq. (6) was avoided as it is unstable due to its complexity and many unknown parameters. An inflex point of the model functions was determined

109

as

$$\Delta p_{int} = -\frac{\Delta G}{\Delta V} \qquad (20)$$

and plotted vs. logarithm of concentration providing a linear dependence given by Eq. (8). The values of ΔV_d and $K_{d,atm}$ were obtained by linear regression of the experimental points.

Pressure-jump kinetics. A sample of a selected concentration was prepared and the pressure was set to 10 MPa. After 3 min incubation the pressure was set to 50 MPa by a sudden increase of pressure and the fluorescence was recorded for hundreds to thousands of seconds. The dead time of the pressure jump was less than 5 ms. The excitation and emission wavelengths were 280 nm and 350 nm, respectively. This procedure was repeated for a series of pressures increasing by 50 MPa up to 450 MPa. The whole sequence was performed for the dimer concentrations of 1, 2, 5, and 20 μM. The obtained curves were fitted by the single-exponential decay function and the equilibrium values of fluorescence as well as the apparent rate constants were determined. For every concentration the pressure dependence of the equilibrium fluorescence was fitted by the sigmoidal function given by Eq. (19) with the parameters ΔG and ΔV renamed to $\Delta G^0_{u,atm}$ and ΔV_u denoting the atmospheric pressure standard change of Gibbs energy and the volume change of the unfolding process, respectively. Furthermore, the values of rate constants of unfolding and folding of the protease monomers were calculated for every pressure point using Eqs. (14, 15).

Reversibility assay. A sample of 5 μM (considered as dimer) HIV-1 PR in the assay buffer was prepared, incubated for 3 min at 10 MPa in the cuvette space of the fluorometer and the tryptophan-emission spectrum was recorded as previously. Then the pressure was set to the selected target pressure, the sample was incubated for 50 min and another spectrum was recorded. The pressure was then released back to 10 MPa and the final spectrum was measured after 3 min incubation. CSM of each spectrum was evaluated. The experiment was repeated for several target pressures from the interval of 100–350 MPa. In a parallel experiment the incubation time was shortened to 3 min.

ANS-indicated transition. 700 μM 8-anilino-1-naphtalenesulfonic acid (ANS) was added to the sample of HIV-1 PR of a given dimer concentration (2, 5, 12, and 20 μM) in the assay buffer. The sample was pre-incubated in the fluorometer pressure cell at 10 MPa for 5 min. Emission spectrum was recorded in triplicate, excitation wavelength was 350 nm and emission-wavelength range 400–600 nm. The procedure was repeated for a series of pressures up to 500 MPa with typical increase of 40–50 MPa. At the pressure points where the spectrum was obviously changing the measurement was repeated several times with 4-min incubation periods between the two subsequent measurements.

Kinetic experiments with ANS. In the first experiment, the sample was prepared identically as in the previous paragraph and the pressure-jump experiment was carried out in the same way as in case of the tryptophan spectra with the excitation and emission wavelengths of 350 nm (slit of 2 nm) and 472 nm (slit of 8 nm), respectively. In the second experiment whole spectra were measured in a series of time points after each pressure jump and the CSM was evaluated for every spectrum.

High-pressure enzyme-kinetics experiments

Absorbance kinetic experiments were carried out using a Cary 3 spectrophotometer (Foster City, CA) modified to accommodate a high pressure cell. Chromogenic peptide substrate of the sequence KARVNle*NphEANle-NH$_2$ (Nph stands for p-nitrophenylalanine, Nle for norleucine, asterisk indicates the cleavage site) with known kinetics parameters (Km = 15.1 μM; kcat

= 30.0 s^{-1} [29,30]) was used. The volume of 2900 μL of the reaction buffer were mixed with 60 μL of substrate at 3.95 μM. The mixture was equilibrated in a water bath at 25°C for 17 min before adding the enzyme. 40 μL of enzyme (diluted by reaction buffer to the final concentration of 1 μM considered as dimer) were promptly introduced at the mixture and the sample was transferred immediately to the high-pressure cell of the instrument. As this manipulation requires 2–4 min, the concentrations were chosen in order that the linear decrease of substrate concentration takes considerably longer time. The excitation wavelength was set to 305 nm (2 nm slit width) and the absorbance was recorded for 20 min under the varying pressure in the region of 10–350 MPa at 25°C. The measured reaction rate was then normalised to the initial value at 10 MPa and plotted vs. pressure.

Results

High pressure treatment of HIV-1 protease complex with inhibitor darunavir

HIV-1 protease solution of 12 μM dimer was mixed with 120 μM darunavir and the pressure was increased stepwise from the lowest value of 10 MPa to the maximum of 500 MPa recording the fluorescence spectrum in every pressure point. Fig. 1 shows the comparison of pressure dependences of CSM and spectral integral for inhibited and non-inhibited enzyme. It can be seen that the spectral integral change of the inhibited enzyme is relatively small in comparison with the strong sigmoid-like decrease of the free-enzyme curve. Accordingly, the CSM curve of the inhibited-enzyme decreases almost linearly, while the free-enzyme curve follows this trend only at the beginning of the pressure range up to approximately 200 MPa and then diverges from the inhibited curve to lower values. This trend is then reversed at 375 MPa, from where the curve goes steeply up. The continuous CSM decrease common for both the curves is likely an expression of a direct influence of the high pressure to the fluorophores as was demonstrated by [31] on free tryptophan and its derivatives. Its slope is equal for both the curves in the pressure range where the free enzyme can be expected to stay in the unperturbed dimeric form. Hence, the "inhibited" curve shows neither the characteristic sigmoidal shape indicating a conformation transition nor any other substantial variations of the trend, while the "free" curve does. Thus, both CSM and spectral intensity indicate that the structure of the enzyme is stabilized by the inhibitor within the chosen pressure range. This conclusion is supported also by the differential spectra shown in S2 Fig. (see the next section).

Similar experiments were carried out also for the dimer concentration of 2 μM with two different concentrations of the inhibitor, 120 μM and 6.7 μM. In accord with the previous experiment, obviously changing trend of the free-enzyme CSM curve can be seen while the inhibited curve is linear (data not shown).

Dimerization equilibrium studied by high-pressure fluorescence spectroscopy

Spectral response of non-inhibited HIV-1 protease to high pressure was determined for a dimer concentration range of 1 to 25 μM. The structure of this enzyme with the location of the fluorescently active amino acids is shown in Fig. 2. For every concentration point tryptophan emission spectra were measured in the range of 10 to 500 MPa and the pressure dependences of CSM were evaluated. A set of emission spectra for selected pressures is shown in S1 Fig. The experiment was carried out in two series, the resulting curves of one of them are shown in Fig. 3. As can be seen, the curves show the typical sigmoidal profile characteristic for structural transitions.

111
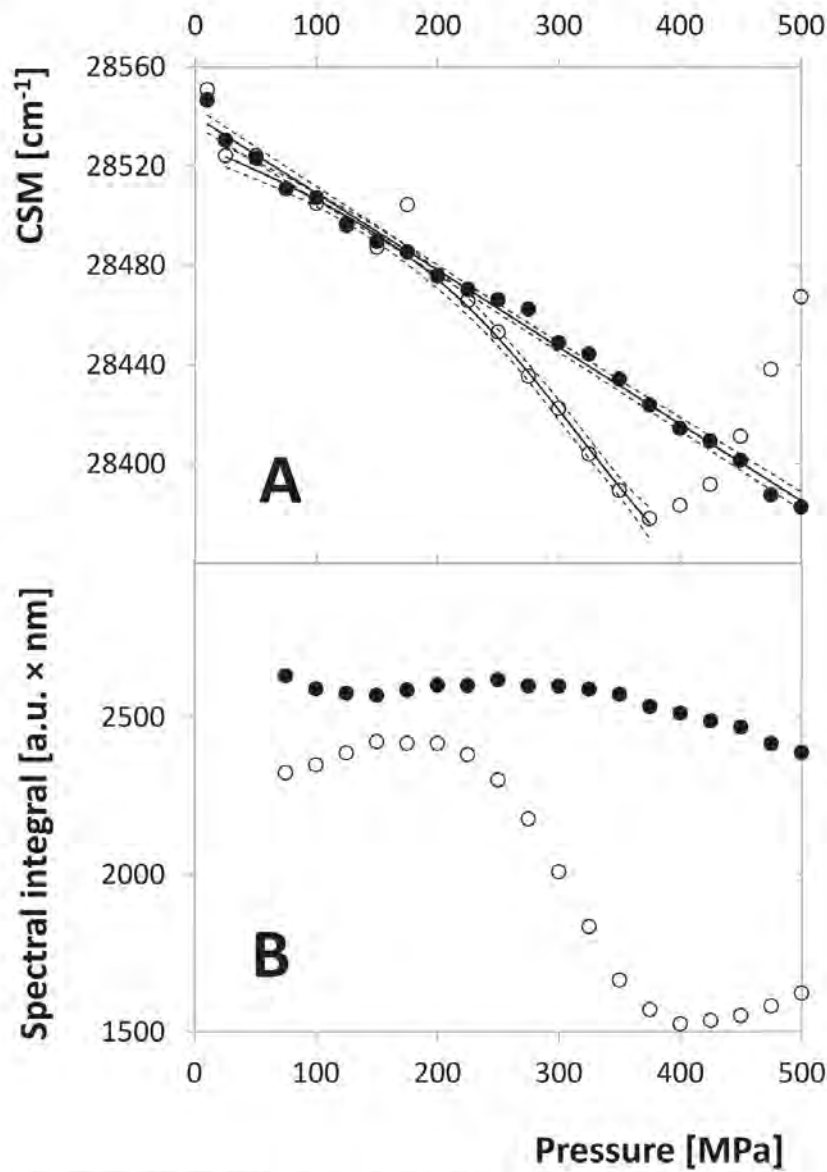
Fig 1. Spectral properties of inhibited (solid circles) and non-inhibited (open circles) HIV-1 PR as functions of pressure. A. CSM of the tryptophan fluorescence spectrum, $\lambda_{ex} = 280$ nm, $\lambda_{em} = 300$–400 nm. Curves are mutually shifted by an additive constant in order to allow the comparison of their shape details. Solid lines represent the regression curves of the data series (linear for inhibited and sigmoidal according to Eq.19 for non-inhibited curve) and dashed lines the corresponding confidence bands. As in Fig. 3, points of the non-inhibited series above 375 MPa, which are influenced by aggregation, are excluded from the regression procedure. B. Spectral integral of the same spectra.

112

Differential-spectra sets calculated for different protease concentrations are shown in S2 Fig. In the presence of the inhibitor all the curves show equal shape and differ only by their amplitude, likely as a consequence of the direct pressure influence on the fluorophores. Without the inhibitor the positions of the extremes of the spectra move with the varying pressure indicating more complex structural changes which is in agreement with Fig. 1A.

To confirm that the CSM at a given pressure does not change with time, two different kinetics experiments were carried out. In the first one a sample of the enzyme was put into the pressure cell and the pressure was set manually to the desired value from the range of 200 to 450 MPa. Then the tryptophan emission spectra were taken in a series of time points and the CSMs were determined. Fig. 4A shows that no significant CSM change can be observed at any of these curves. Up to 360 MPa slightly decreasing tendency can eventually be seen the magnitude of which is considerably lower than the CSM changes induced by the pressure increase. More significant increase can be observed at very high pressures above 400 MPa which can be associated with aggregation of the protein (see the description of this phenomenon below).

In addition, a pressure-jump kinetic experiment was carried out providing the time dependence of the emission spectrum for every pressure point. In accord with the previous experiment the CSM position is stable and does not vary within the time course (Fig. 4B). Both these experiments demonstrate that the pressure dependence of CSM refers to a fast structural transition reaching its equilibrium in the time scale shorter than the experimental response, i.e. the order of magnitude of the half-times of the observed processes is surely lower than a second.

The inflex points of the CSM curves apparently shift to higher pressures with increasing concentration, which is characteristic for transitions where the reaction volume decreases along with dissociation of multisubunit complexes, e.g. dimer dissociation. In addition, the limits of the curves at plus and minus infinity have non-zero, but practically equal slope corresponding to that of the inhibited curves. The curves of the two highest concentrations turn steeply up above 350 MPa, likely due to the protein aggregation (see below). This part of the curves was therefore excluded from the evaluation of the dimerization parameters. Inflex points of all the curves were determined and plotted vs. logarithm of monomer concentration, as shown in Fig. 5. Dimer-dissociation volume change $\Delta V_d = (-32.5 \pm 4.1) ml\ mol^{-1}$ and the atmospheric-pressure equilibrium constant $K_{d,atm} = (0.92 \pm 0.17) \mu M$ were evaluated by means of linear regression in accord with Eq. (8). The ratio of the volume change to molar weight of HIV-1 PR (-1.50 ml/kg) is in a reasonable agreement with previously studied proteins yeast enolase (-1.28 ml/kg) [20] or yeast hexokinase (from -1.07 to -1.49 ml/kg) [21].

In order to confirm the reversibility of the dimer dissociation, reversibility assay was carried out. Here, spectra are recorded at 10 MPa, then after 50 min incubation at the target pressure and again at 10 MPa after the pressure release. As shown in Fig. 6, CSM values follow approximately the sigmoid-like curves shown in Fig. 3 for the target-pressure spectra and return to the original value after the pressure release. Similar experiment, but with missing the 50 min incubation period, is shown in S3 Fig. At 300 MPa the final value after the incubation period is a little lower than the initial one, which is not observed without incubation. This might be a consequence of the folding kinetics, but the difference is quite small and should be taken with caution. However, at 350 MPa a partial contribution of the protein aggregation shifts up both the high pressure value and the value after releasing the pressure in the experiment with the 50 min incubation period, which is apparently different from the incubation-free. Thus, in accord with the transition curves shown in Fig. 3, the dimerization reflected by the CSM transition seems to be a reversible process in the pressure interval of 10–350 MPa.

113

114
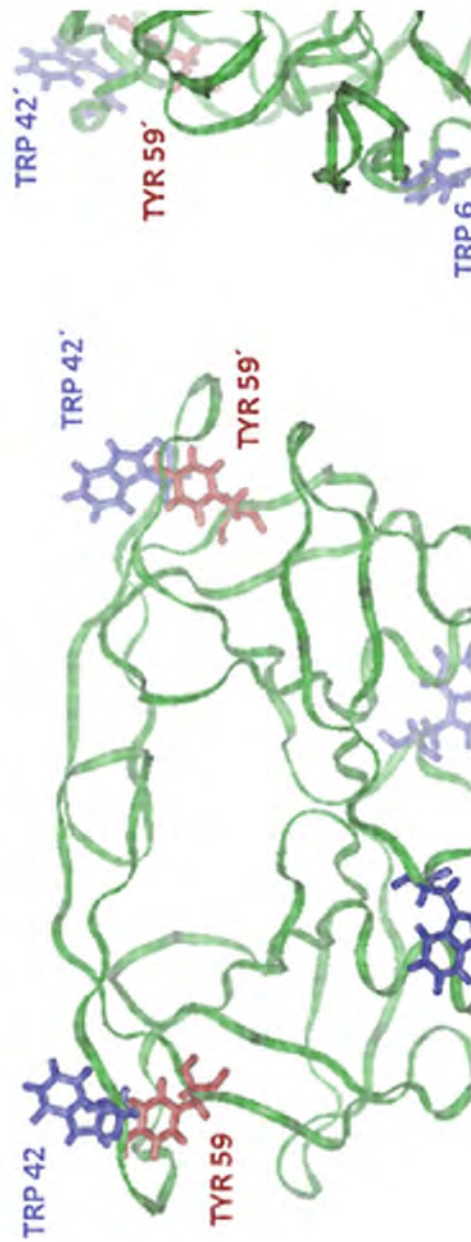
Fig 2. Model of the HIV-1 PR backbone in front (left) and side (right) view (structure taken from RSCB protein data bank, PDB ID: 1OHR). The residues of fluorescent amino acids, Trp6, Trp42 and Tyr59, are indicated. Trp6 is located close to the dimerization interface, therefore its spectroscopic properties are probably influenced by dimer dissociation. On the contrary, Trp42 and Tyr59 are located at the most distant site from the dimerization interface, therefore their fluorescence is probably unaffected by dimer dissociation, provided that the monomer keeps its conformation. Trp42 and Tyr59 can eventually partially influence this effect due to the change of their mutual configuration because the conformation of this part of the molecule undergoes partial changes with growing pressure, as was shown by the molecular-dynamics simulation [32]. All the fluorophores can contribute to the spectral changes induced by unfolding of monomers or protein aggregation when they undergo more extensive changes of their environment.

doi:10.1371/journal.pone.0119099.g002

Unfolding of monomers influences the fluorescence intensity

In addition to CSM, spectral intensity of the measured spectra was studied, too. The curves representing the pressure dependence of the spectral integral look qualitatively similar to the CSM curves (see the non-inhibited curves in Fig. 1), but they vary for quite a long time period. Therefore, a series of kinetic experiments was carried out in order to determine the time dependence of the spectral intensity at different pressures and concentrations. The measured curves (Fig. 7) show an obvious decrease within the time course, the magnitude of which depends on the pressure. A series of pressure-jump experiments with varying enzyme concentrations was performed. A typical set of the measured intensity curves including the fitting single-exponential functions is presented in Fig. 8A. The pressure dependent fluorescence intensity changes were not fully reversible. However, at the end of each kinetics, no further time dependent changes occurred, allowing the determination of apparent thermodynamic parameters. The apparent equilibrium values of fluorescence were evaluated for each curve and plotted as functions of pressure (an example is given in Fig. 8B). These functions have a sigmoid-like profile typical for the structural transitions. The values of the apparent volume change ΔV_u and atmospheric pressure Gibbs-energy change $\Delta G_{u,atm}^{i}$ of this transition were determined by means of non-linear regression. They are listed in Table 1 and show relatively high stability with respect to the varying concentration, which indicates a transition of first-order kinetics in both directions. Thus, this transition represents presumably an unfolding of the protease monomers. $\Delta G_{u,atm}^{i}$ of the series with the highest concentration is somewhat higher in comparison with the other concentrations which might be a consequence of preventing the unfolding by dimer formation.

The average values of the unfolding characteristics are $\Delta V_u = (-104 \pm 6) ml\, mol^{-1}$ and $\Delta G_{u,atm}^{i} = (27.6 \pm 1.7) kJ\, mol^{-1}$ resulting in $K_{u,atm} = (1.5 \pm 1.0) \times 10^{-5}$. Thus, the unfolding process is highly unfavourable at atmospheric pressure but can be induced by the high pressure of more than 200 MPa. It agrees with the NMR study by Ishima et. al. [27] who observed the existence of folded monomers of HIV-1 PR, as well as with observations of folded monomers of different retroviral proteases [28,33].

The values of the relaxation rate constant $k_{obs}(p)$ were determined for all the pressures and the rate constants $k_f(p)$ and $k_u(p)$ were calculated using Eqs. (13–15). Fig. 9 shows linear dependences of the folding and unfolding rate constants, while the relaxation constant $k_{obs}(p)$ has a minimum in the crossing point of the former lines where the equilibrium constant equals 1. This feature indicates microreversibility of the transition. Finally, the apparent activation volume change and the atmospheric pressure rate constants $k_{f,atm}$ and $k_{u,atm}$ were determined using linear regression of the data by the functions

$$\ln k_i(p) = \ln k_{i,atm} - \frac{\Delta p \Delta V_i^{\neq}}{RT}, \qquad (21)$$

115
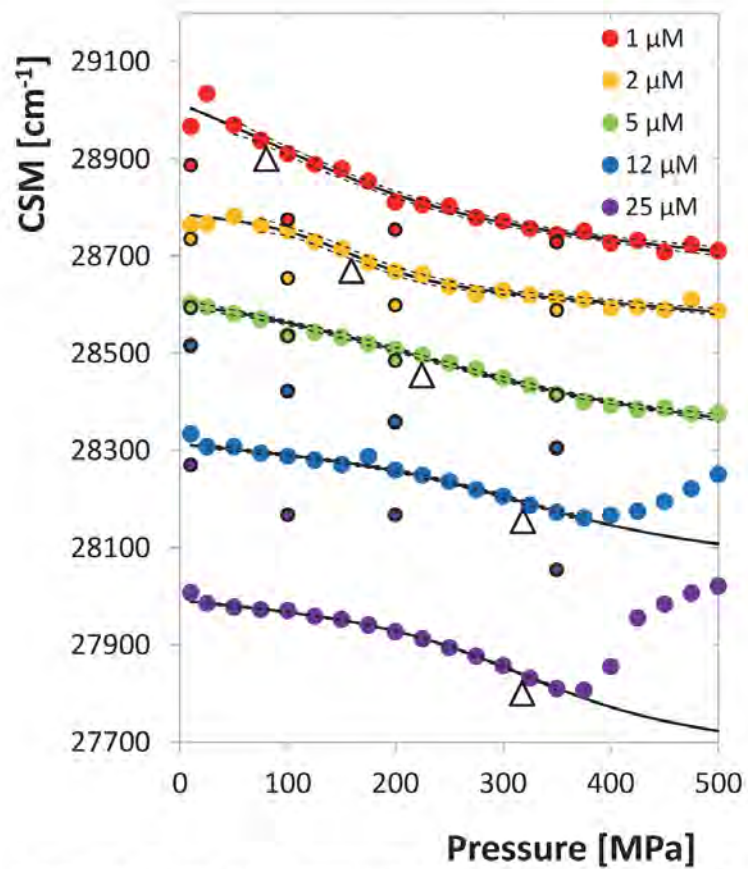
Fig 3. Centre of spectral mass of the emission spectra as a function of pressure for different HIV-1 protease concentrations. The curves are shifted up or down by artificial additive constants for the sake of lucidity. Solid lines express the fitting curves, empty triangles indicate the inflex points. Dashed lines represent the 95% confidence bands of the regression curves. The high pressure regions of the curves for 12 and 25 μM dimer concentration are excluded from fitting as they are influenced by protein aggregation. The smaller circles with black bordering line indicate the CSM values for the subsequent release of pressure from the highest to the lowest value.

doi:10.1371/journal.pone.0119099.g003

where "*I*" stands for "u" or "f" and $V_I^{\neq}$ is the activation volume of the respective reaction. The values for 2 and 5 μM dimer are shown in Table 1. It is obvious that these values, especially the rate constants, are strongly pressure dependent which might be a consequence of the dimerization influence on the unfolding process. Thus, these parameters cannot be considered as reliable characteristic of the unfolding itself, but as an evidence of the mutual interplay of unfolding and dimerization.

116
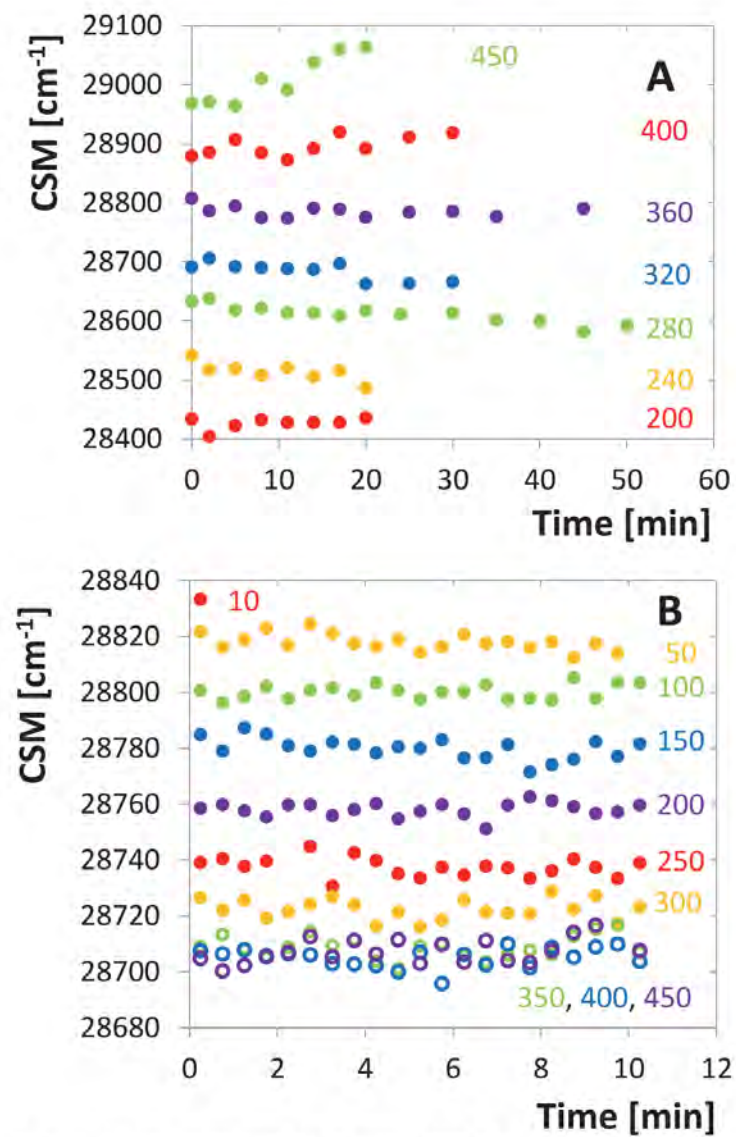
Fig 4. Time development of the tryptophan-fluorescence CSM. A. Time series for different pressures, each started with a new enzyme sample, concentration of 10 μM dimer. The individual curves are shifted up or down by artificially chosen constant for the sake of better orientation in the graph. B. Time series measured with the same sample of 2 μM concentration, pressure setting is facilitated by pressure jumps. The CSM scale is genuine for all the series.

doi:10.1371/journal.pone.0119099.g004

117
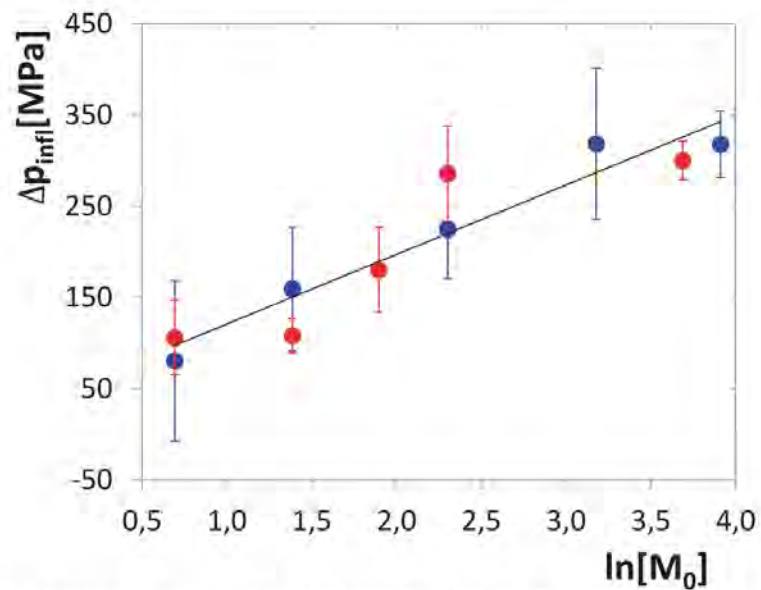
Fig 5. Concentration dependence of the inflex point of the CSM curves. The data consist of two independent experimental series (blue and red circles). Linear-regression line of the slope of 76.3 MPa and intercept of 45 MPa is shown which is used to determine the constants $K_{d,atm} = 0.92 \, \mu M$ and $\Delta V_r = 32.5 \, ml \, mol^{-1}$. Regression lines of the individual series are not shown as they are almost identical; their slopes and intercepts are 77.3 MPa and 43 MPa, respectively (blue), and 74.7 MPa and 47 MPa, respectively (red), which corresponds with $K_{d,atm} = 0.95 \, \mu M$, $\Delta V_r = 32.1 \, ml \, mol^{-1}$ (red) and $K_{d,atm} = 0.88 \, \mu M$, $\Delta V_r = 33.2 \, ml \, mol^{-1}$ (blue). The error bars of the individual measurements are calculated from the errors of the regression parameters of Eq. 19 using Eq. 20 and the error-transition law. $[M_0]$ is considered to be a dimensionless quantity related to the units of µM ($[M_0] \rightarrow [M_0]/\mu M$).

doi:10.1371/journal.pone.0119099.g005

Unfolding of monomers influences also the reversibility of the CSM transition assigned to the dimer dissociation. Reversed transition curves measured along with the release of the pressure from the highest value down to atmospheric pressure are shown in Fig. 6 together with the original transition curves. They approximately follow the trend of the original curves, but they are variously shifted for different concentrations. The protease monomers unfold at about 264 MPa in a reversible, but slow process, which can start only after dissociation of the dimer. Therefore, for the lower concentrations of 1–5 µM the recovery of CSM is delayed by folding of the unfolded monomers that are not able of dimerization. The weakening of this effect from 1 to 5 µM might be explained by the longer time the protein spent in the monomeric form for the lower concentrations at which the dimer dissociates at lower pressure. At the higher concentrations of 12 and 25 µM the recovery curve has the expected shape but is shifted up, most likely due to the contribution of protein aggregation occurring at this concentration range (see section "Protein aggregation" below).

118
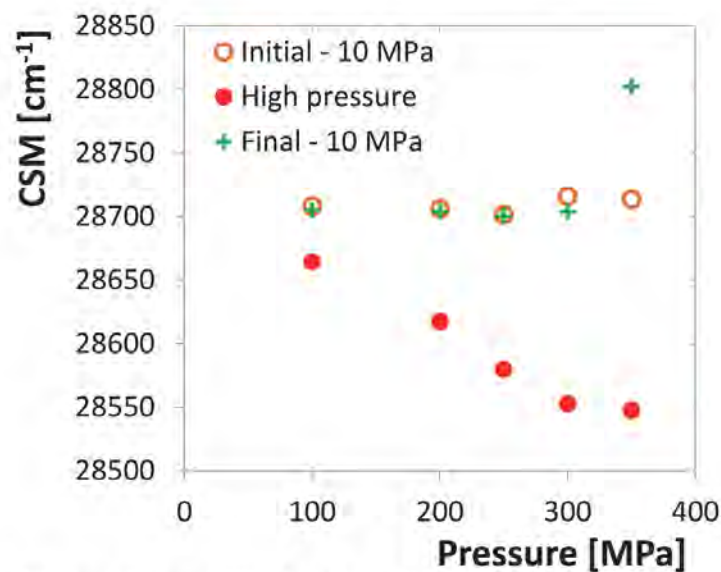
Fig 6. Reversibility assay of HIV-1 PR of 5 µM concentration (expressed as dimer). For every enzyme sample three CSM values were determined: at 10 MPa before pressurizing, at the selected target pressure from the interval of 100 to 350 MPa after 50 min incubation and again at 10 MPa after releasing the pressure. The experiment indicates good reversibility of CSM up to 300 MPa, above this value the reversibility is perturbed by protein aggregation.

doi:10.1371/journal.pone.0119099.g006

High-pressure behaviour of HIV-1 protease indicated by ANS fluorescence

8-anilino-1-naphthalenesulfonic acid (ANS) is a fluorescent indicator that binds to proteins, mainly by hydrophobic interactions. This interaction is accompanied with a fluorescence increase and a blue shift of ANS emission spectrum, which can be used to indicate the structural changes of proteins, especially the unfolding [34]. A series of fluorescence spectra under various pressures was measured as described in Methods and Material. Up to 240 MPa only a continuous decrease of the CSM with growing pressure occurs without any sign of a transition. Above this pressure, a remarkable change in the spectral shape takes place due to the increased emission at 472 nm which shifts CSM steeply up. The time course of this transition was investigated by pressure-jump experiments. A time series of emission spectra was taken at each pressure point and the pressure dependence of CSM was fitted by Eq. (19). The apparent equilibrium values were plotted vs. pressure (Fig. 10) and the transition parameters were determined as in the previous section. The values of apparent $\Delta G_{ANS,atm} = (43.5 \pm 5.8) kJ\ mol^{-1}$ and $\Delta V_{ANS} = (-160 \pm 21) ml\ mol^{-1}$ confirm a transition strongly unfavourable at pressures up to 240 MPa with atmospheric-pressure equilibrium constant $K_{ANS,atm} = 2.4 \times 10^{-8}$, but highly favourable above 320 MPa. Above this pressure the shift of equilibrium CSM value is small indicating that the transition runs almost to completeness. Stepwise release of the pressure did not lead to the return of the fluorescence intensity to the value before the transition. Hence, this transition is either irreversible or the kinetics of the reverse process is very slow.

119
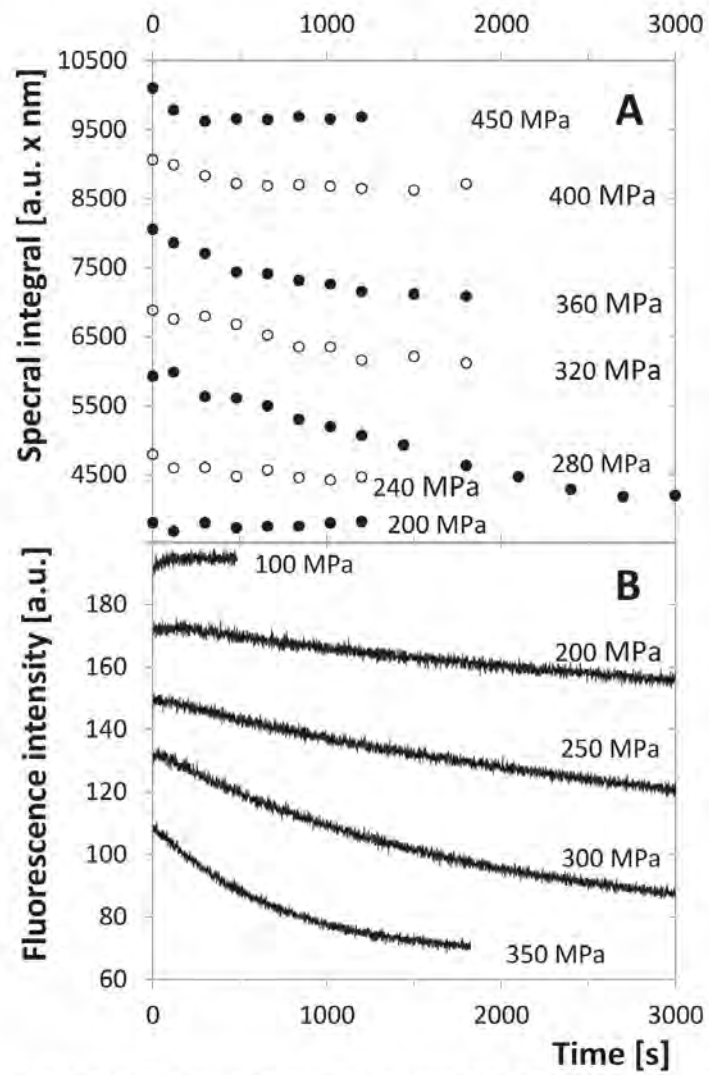
Fig 7. Time development of the emission intensity of the tryptophan fluorescence at λ_{ex} = 280 nm. In both panels the curves are shifted up or down by artificially chosen constant for better orientation in the graph. A. Spectral integral of spectra taken in time series for different pressures, each series with new protease sample, λ_{em} = 300–400 nm, dimer concentration 10 µM. B. Emission intensity at a single wavelength of λ_{em} = 350 nm for the dimer concentration of 5 µM.

doi:10.1371/journal.pone.0119099.g007

120
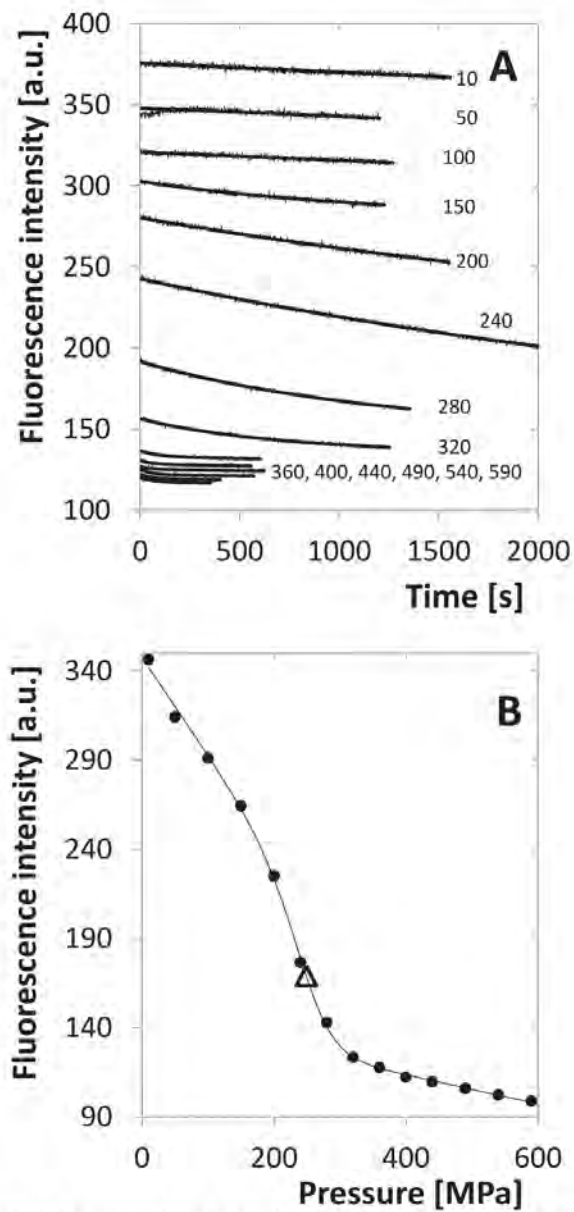
Fig 8. Fluorescence-intensity indicated transition for 5 µM dimer. A. Time dependence of the intensity for different pressures. Each series is plotted together with the fitting single-exponential decay curve. The

121

same sample was used for the whole set of measurements, pressure setting was facilitated by pressure jumps. The numbers indicate pressure in MPa. B. Equilibrium values of fluorescence for the same experimental series determined from the limit of the fitting curves for time tending to infinity. Inflex point is indicated by the open triangle.

The observed relaxation rate of this transition is almost pressure independent above the inflex point at about 270 MPa. At 320 MPa, where the apparent equilibrium is strongly shifted to the unfolded state, the observed relaxation rate constant $k_{obs,ANS} = (0.040 \pm 0.003)min^{-1}$ which gives an unfolding half time $\tau_{1/2,ANS} = (17.5 \pm 1.23)min$.

The spectral properties of this transition show no significant concentration dependence which indicates a first-order transition. For low concentration of 2 μM dimer almost no transition-like shape was observed (data not shown), most likely due to the low spectral response of ANS at low protein concentration.

Protein aggregation

At protein (dimer) concentrations above 10 μM a dramatic increase of both the CSM and the spectral intensity was observed at pressures above 300 MPa. It is demonstrated on both the raw and differential spectra in S1 Fig. and S2 Fig. The steepness of the respective part of the transition curves grows dramatically with protein concentration, while the point where this effect starts tends to lower pressure. Figs. 1 and 3 show this effect especially for 12,5 and 25 μM dimer and the difference in steepness and position shift is obvious. It indicates that, in contrast to the monomer-dimer equilibrium, in this process the total volume decreases along with the decreasing number of particles. Most probably this transition represents pressure-induced protein aggregation. Formation of higher aggregates agrees with the CSM increase (i.e. shift to lower wavelengths) as the spectrum is partially contaminated by the elastically scattered excitation light. Indeed, at the highest concentration the aggregates were visible after the experiment in the sample cuvette as white opacity.

Enzyme kinetics of HIV-1 protease

Enzyme kinetics was investigated under the varying pressure in the region of 10–350 MPa. The substrate was mixed in a spectrophotometric cuvette with the reaction buffer and enzyme and this sample was transferred immediately to the high-pressure cell of the instrument. As this manipulation requires time in the order of seconds to minutes, the concentrations were chosen in order that the linear decrease of substrate concentration takes considerably longer time. The measured reaction rate was then normalised to the initial value at 10 MPa and plotted vs. pressure (see Fig. 11). The resulting curve is decreasing steeply but the activity remains detectable up to 350 MPa. The identification of the individual phenomena contributing to this decrease is

Table 1. Equilibrium properties of monomer unfolding indicated by intensity of the tryptophan fluorescence spectrum.

Enzyme concentration(expressed as dimer) [μM]	Δp_{infl} [MPa]	ΔG_u^0 [kJ. mol^{-1}]	ΔV_u [mL. mol^{-1}]	$\Delta V_u^{\ddagger}$ [mL. mol^{-1}]	$\Delta V_f^{\ddagger}$ [mL. mol^{-1}]	$k_{u,atm}$ [s^{-1}]	$k_{f,atm}$ [s^{-1}]
1	252 ± 15	21.3 ± 1.2	−85 ± 4				
2	267 ± 51	29.6 ± 3.8	−111 ± 14	−47.5	63.2	0.014	2153
5	249 ± 54	23.0 ± 4.0	−92 ± 15	−74.9	23.5	8.4×10^{-4}	8.9
20	288 ± 45	36.4 ± 3.4	−126 ± 12				
Average	265 ± 22	27.6 ± 6.9	−104 ± 19				

122
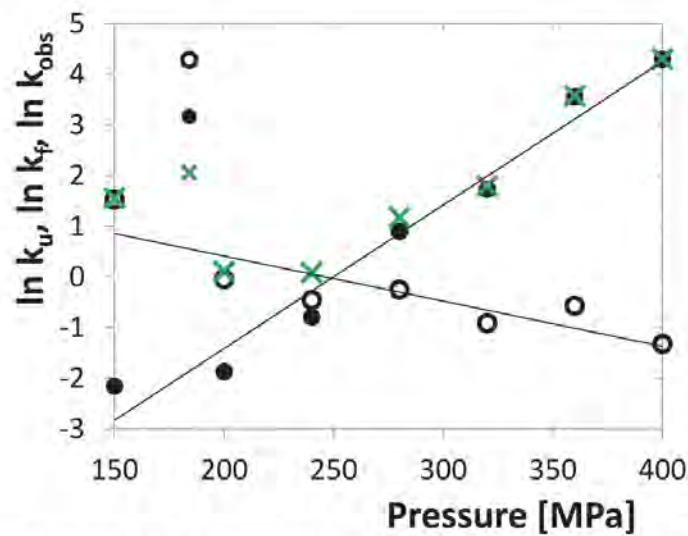
Fig 9. Individual rates of folding and unfolding of the protease monomers and the observed relaxation rate for the concentration of 5 μM dimer. Both rates are fitted by linear functions. The crossing point roughly corresponds with the inflex point of the unfolding transition. Rate constants are considered as dimensionless quantities related to the units of h^{-1}.

doi:10.1371/journal.pone.0119099.g009

rather difficult. The general function describing the rate of the enzyme reaction, which can be derived from the ordinary Michaelis-Menten kinetics and the monomer-dimer equilibrium (Eq. 3), is

$$v = k_{cat} \frac{S}{K_m + S} \left[D_0 + \frac{K_m K_d}{8(K_m + S)} \left(1 - \sqrt{1 + \frac{16 D_0 (K_m + S)}{K_m K_d}} \right) \right]. \quad (22)$$

However, in principle all the three parameters, K_m, K_d and k_{cat} are pressure dependent, but the dependence is known to us only for K_d. Determination of this function for the other two quantities would require measurements at different enzyme concentration, which was impossible on the used equipment. In addition, the volume changes that determine the pressure dependencies of the parameters need not be constant. Finally, considering reasonable values of volume changes, all the parameters may influence the reaction rate in a quantitatively comparable way. Nevertheless, considering the pressure-independent behaviour of the inhibited enzyme, it is conceivable that the interaction of enzyme and substrate will behave similarly and that K_m and k_{cat} will be less pressure dependent. The solid curve in Fig. 11 shows the model situation if only K_d is pressure dependent with the previously determined values of ΔV_d and $K_{d,atm}$. The curve obviously follows the trend of experimental points, but there are some deviations, especially at the high-pressure region, which may be attributed to the pressure dependence of K_m and k_{cat}.

123
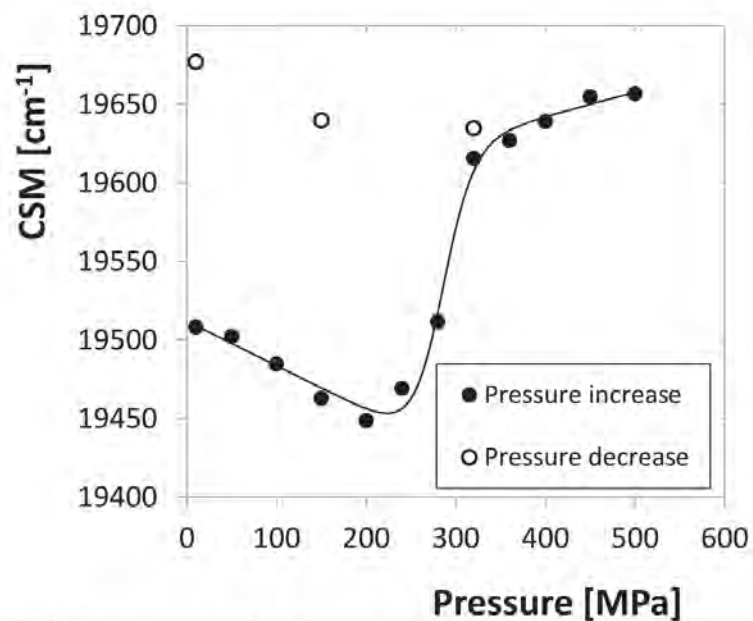
Fig 10. Equilibrium values of CSM of ANS emission spectrum for 10 µM dimer fitted by a model function. Values for the reverse course of pressure indicate irreversibility of the transition.

Discussion

HIV-1 protease structure changes have been investigated under high-pressure ranging from 10 to 600 MPa. The transitions between structural states of the enzyme have been monitored with the aid of tryptophan and ANS fluorescence spectra.

A pressure-dependent transition that moves to higher pressures with growing concentration was identified with the dissociation of the protease dimer. It is followed by another, concentration independent, transition which can be associated with the unfolding of monomers. The former process seems to be very fast keeping the system in permanent monomer-dimer equilibrium, while the latter one is rather slow with the half-time in the order of tens of minutes.

The former process is reflected by changes in the shape of the tryptophan emission spectrum which can be observed as a difference in the CSM profile for inhibited and non-inhibited enzyme (Fig. 1A). Obviously, the total spectral change consists of a structure-independent contribution of the direct influence of high pressure to the fluorophores that shifts the CSM monotonously to lower values [31] and a contribution depending on dimer dissociation which is missing at the inhibited enzyme. This is apparent also in the differential spectra shown in S2 Fig., especially on the branch where the minima are located. For the higher concentrations of 12 and 25 µM, where the dimer dissociation occurs only at relatively high pressures, the minima of the curves are initially stable at about 320–330 nm, similarly to the inhibited enzyme. (The small difference in their position the inhibited and non-inhibited enzyme is likely caused

124
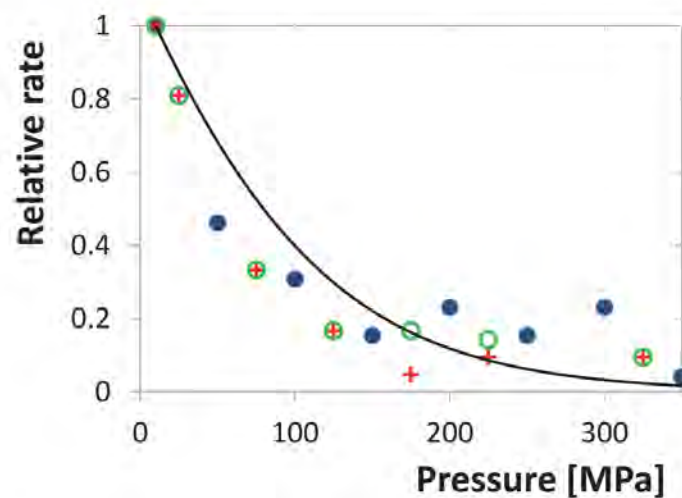
Fig 11. Initial rate of substrate cleavage as a function of pressure related to the value for 10 MPa.
Different symbols denote three experimental series. The solid curve shows the relative rate for the ideal case
of pressure dependent K_d but pressure independent K_m and k_{cat}.

doi:10.1371/journal.pone.0119099.g011

by the presence of darunavir, which is itself an aromatic molecule.) With growing pressure the
position of the minima moves up. On the contrary, for the low concentration of 2 μM, where
the dimer dissociation progresses already at the lower pressures, the minimum is initially locat-
ed approximately at 350 nm and moves down when the pressure grows. It is thus probable that
the structure-independent contribution causes an intensity decrease at 320–330 nm while the
dimer dissociation at higher wavelength. At high concentration the structure-independent con-
tribution prevails at lower pressures and is supplemented by the dimer-dissociation contribu-
tion only in the high-pressure region, while at lower concentrations the dimer-dissociation
contribution dominates at low pressures and is overwhelmed by the structure-independent
contribution when the dimer dissociation draws to completeness. Thus, for high concentra-
tions the minimum position moves up with increasing trend while for low concentrations they
move down with decreasing trend, in all cases tending to approximately equal point between
330 and 340 nm. For the higher concentrations of 12 and 25 μM the high-pressure spectra are
probably influenced by protein aggregation which causes the intensity increase at the low-
wavelength range as a consequence of the elastic scattering of the excitation light. Obviously, it
also changes the position of the differential-spectrum minimum dramatically.

The unfolding process is reflected by the intensity of both tryptophan and ANS fluorescence
spectra. The ANS spectral intensity shows a transition at equal pressure as the tryptophan fluo-
rescence, but considerably steeper and with a rather sudden onset above 240 MPa. This may be
explained by co-operative action of the monomer unfolding and stabilization of the unfolded
state by binding of ANS. Given the values of $K_{u,atm}$ and ΔV_u, the degree of unfolding starts to
grow rapidly from the point where $K_u(p) = 1$. As a consequence, the rate of unfolding increases
dramatically in this region. Thus, the unfolding process becomes strongly favoured above the
inflex point both thermodynamically and kinetically. As previously discussed, ANS fluores-
cence provides a sufficient signal only at protein concentration higher than 2 μM. Unfolded

125

protein reaches this concentration closely after the pressure crosses the inflex point. The interaction with ANS can probably further stabilize the unfolded state or change the direction of the process to another unfolding pathway, which may contribute to the irreversibility of this process. A similar process of destabilization of the native structure of human serum albumin by interaction with ciprofloxacin was reported recently [35]. The common base of both the tryptophan- and ANS-indicated transitions can be further supported also by the rough agreement of the relaxation rate constants.

When an inhibitor is added to the system, none of the transitions occurs within the whole pressure range. It can be deduced that the inhibitor stabilizes the dimer structure, as is generally assumed even at atmospheric pressure. These results indicate that the dimer structure is remarkably firm and does not undergo unfolding even at very high pressure. Unfolding transitions have been observed only at the non-inhibited enzyme above about 265 MPa where a substantial part of the dimer was already dissociated, which indicates that unfolding can be initiated only in the monomeric state of the protease.

The high-pressure destabilization of HIV-1 PR dimer allowed us to determine the equilibrium constant of dimer dissociation $K_{d,atm}$, which is a challenging experimental problem for decades. At atmospheric pressure determination of this constant is complicated by the high stability of dimer and requires rather low enzyme concentrations. In contrast with this, high pressure, together with negative ΔV_d, destabilizes the dimer and shifts the transition to higher concentrations. This fact can be helpful also for other experimental assays where free monomers are needed, e.g. to confirm their structure by NMR. Furthermore, compounds interacting directly with monomer, especially potential inhibitors of dimerization [36] or modular inhibitors inhibiting both the catalytic activity and dimer formation [37] studied by some research groups in the past, can be designed with the aid of these techniques.

The method of determination of $K_{d,atm}$ presumes that ΔV_d is constant within the whole pressure range. This assumption is supported by the linearity of the dependence of the CSM inflex point vs. pressure depicted in Fig. 5. Eventual deviations from linearity might occur at the lowest part of the pressure range where no experimentally detected inflex points are located, but their influence on the equilibrium constant would be minor, because the reaction Gibbs energy $\Delta G^0_{d,atm}$ is an integral of ΔV_d along the pressure axis. Thus, the resulting $K_{d,atm}$ value can be considered to be correct at least as regards the order of magnitude.

The previously reported values of equilibrium constant $K_{d,atm}$ differ by several orders of magnitude among each other. In general, the values obtained by kinetic assays [4–7], including our study [6], are lower than those obtained by substrate/inhibitor-independent assays [9,10,38], which might be caused by an influence of the substrate/inhibitor stabilization. Our current value is higher that the kinetically determined values, but belongs to the lower end of the set of values determined by the substrate/inhibitor-independent methods.

The present results show that the dimer dissociation and association are both fast processes that keep the system in permanent equilibrium. Together with the observed substrate or inhibitor stabilization it means that the dimerization equilibrium is no hindrance of the enzyme activity since the active dimers re-associate even at low concentrations immediately after the addition of the substrate. Previously, rather mixed results have been reported regarding the question of fast or slow dimerization equilibrium. The method of sedimentation equilibrium [9,10,38] could not resolve this question, while the kinetic methods usually reported a slow dimerization process [4,5,7,37,39–41]. However, at atmospheric pressure dimerization can be easily confused with some unfolding process, because the missing pressure coordinate prevents the distinguishing between the first and second order transitions. Although at the current

126

experiment no unfolding processes running at atmospheric pressures have been found, at different conditions, e.g. temperature of 37°C, they may be enabled.

Enzyme-kinetics experiments showed that the enzyme keeps a part of its activity up to 350 MPa. It indicates that also the substrate provides a similar structure stabilizing effect as an inhibitor, although it is continuously cleaved to products. There is, however, a strong decrease of the reaction rate in the pressure interval of 10 to 150 MPa which might be caused by pressure dependences of at least three different parameters, K_m, K_d and k_{cat}. It is plausible that the enzyme with bound substrate behaves similarly as with inhibitor and is thus rather inert with respect to pressure. In such a case K_d can be considered as a major source of pressure dependence. Indeed, the experimental results follow the theoretical model with reasonable accuracy, but deviations from it can be observed at the high pressure region where the measured rate exceeds the theoretical assumption. However, to explain this effect, that can be caused by several partial phenomena, further experiments are required that allow us to vary the substrate concentrations in a broad range. Thus, high pressure enzyme kinetics remains an interesting experimental challenge for the future.

Conclusion

Inhibitor and substrate binding protect HIV-1 PR against dissociation and unfolding. Unfolding of the protein is initiated only from the monomeric state. In the absence of substrate, the protein is not extremely stable, however, the $K_{d,atm}$ in the submicromolar range protects the protein from dissociation and unfolding at atmospheric pressure and 25°C. Unfolding processes occur only at pressures above 250 MPa. High pressure destabilizes the dimer structure, which opens an interesting way for experiments requiring higher population of monomers in the system. It can facilitate studies of the process of dimerization, structure of monomers and their interaction with different compounds which can be used for design of potential dimerization inhibitors.

Supporting Information

S1 Fig. Selected tryptophan emission spectra of HIV-1 PR in the concentration of 25 μM dimer.
(PDF)

S2 Fig. Sets of differential tryptophan-emission spectra for different concentrations of HIV-1 PR dimer. For 12 μM spectra for inhibited and non-inhibited enzyme are shown. Each spectrum is fitted by 4th-order polynomial in order to identify the changes in the shape of the spectrum and positions of its extremes. For the method of calculation of these spectra see Methods and materials, section "High pressure fluorescence experiments".
(PDF)

S3 Fig. Reversibility assay of HIV-1 PR of 5 μM concentration (expressed as dimer). For every enzyme sample three CSM values were determined: at 10 MPa before pressurizing, at the selected target pressure from the interval of 100 to 350 MPa 3 min after setting the pressure (in contrast to an analogous experiment shown in Fig. 6 where the incubation time was 50 min) and again at 10 MPa after releasing the pressure.
(EPS)

127

Acknowledgments

Author Contributions

Conceived and designed the experiments: MI RL JH. Performed the experiments: MI RL VH AY DCL LP CB. Analyzed the data: MI RL AY DCL LP CB JH JK ED. Contributed reagents/materials/analysis tools: MI RL JK ED. Wrote the paper: MI RL JH JK ED.

References

1. Wlodawer A, Miller M, Jaskolski M, Sathyanarayana BK, Baldwin E, Weber IT, et al. Conserved folding in retroviral proteases: Crystal structure of a synthetic HIV-1 protease. Science. 1989; 245: 616–621. PMID: 2548279

2. Turner BG, Summers MF (1999) Structural biology of HIV. J Mol Biol. 285(1): 1–32. PMID: 9878383

3. Pokorná J, Machala L, ezáčová P, Konvalinka J. Current and Novel Inhibitors of HIV Protease. Viruses. 2009; 1: 1209–1239. doi: 10.3390/v1031209 PMID: 21994591

4. Jordan SP, Zugay J, Darke PL, and Kuo LC. Activity and dimerization of human immunodeficiency virus protease as a function of solvent composition and enzyme concentration. J Biol Chem. 1992; 267: 20028–20032. PMID: 1400318

5. Pargellis CA, Morelock MM, Graham ET, Kinkade P, Pav S, Lubbe K, et al. Determination of kinetic rate constants for the binding of inhibitors to HIV-1 protease and for the association and dissociation of active homodimer. Biochemistry. 1994; 33: 12527–12534. PMID: 7918476

6. Darke PL, Jordan SP, Hall DL, Zugay JA, Shafer JA, Kuo LC Dissociation and association of the HIV-1 protease dimer subunits—equilibria and rates. Biochemistry. 1994; 33: 98–105. PMID: 8286367

7. Ingr M, Uhlikova T, Strisovsky K, Majerova E,Konvalinka J. Kinetics of the dimerization of retroviral proteases: The "fireman's grip" and dimerization. Protein Sci. 2003; 12: 2173–2182. PMID: 14500875

8. Holzman TF, Kohlbrenner WE, Weigl D, Rittenhouse J, Kempf D, Erickson J. Inhibitor stabilization of Human-immunodeficiency-virus type-2 proteinase dimer formation. J Biol Chem. 1991; 266: 19217–19220. PMID: 1918040

9. Towler EM, Gulnik SV, Bhat TN, Xie D, Gustchina E, Sumpter TR, et al. Functional characterization of the protease of human endogenous retrovirus, K10: Can it complement HIV-1 protease? Biochemistry. 1998; 37: 17137–17144. PMID: 9860826

10. Stříšovský K., Tessmer U, Langner J, Konvalinka J, Kräusslich HG. Systematic mutational analysis of the active-site threonine of HIV-1 proteinase: Re-thinking the "fireman's grip" hypothesis. Protein Sci. 2000; 9: 1631–1641. PMID: 11045610

11. Prehoda KE, Mooberry ES, Markley JL. High pressure effects on protein structure. In: Protein dynamics, function, and design, Springer US, Jardetzky O, Lefévre JF, Holbrook RE, NATO ADV Sci I A-Lif 1998; 301: 59–86.

12. Silva JL, Foguel D, Royer CA. Pressure provides new insights into protein folding, dynamics and structure. Trends Biochem Sci. 2001; 26: 612–618. PMID: 11590014

13. Gross M, Jaenicke R. Proteins under pressure. The influence of high hydrostatic pressure on structure, function and assembly of proteins and protein complexes. Eur J Biochem. 1994; 221: 617–630. PMID: 8174542

14. Marchal S, Torrent J, Masson P, Kornblatt JM, Tortora P, Fusi P, et al. The powerful high pressure tool for protein conformational studies. Braz J Med Biol Res. 2005; 38: 1175–1183. PMID: 16082457

15. Torrent J,Connelly JP, Coll MG, Ribo M, Lange R, Vilanova M. Pressure versus Heat-Induced Unfolding of Ribonuclease A: The Case of Hydrophobic Interactions within a Chain-Folding Initiation Site. Biochemistry. 1999; 38: 15952–15961. PMID: 10625462

16. Font J, Benito A, Lange R, Ribo M, Vilanova M. The contribution of the residues from the main hydrophobic core of ribonuclease A to its pressure-folding transition state. Protein Sci. 2006; 15: 1000–1009. PMID: 16597833

17. Font J, Torrent J, Ribo M, Laurents DV, Balny C, Vilanova M, et al. Pressure-jump-induced kinetics reveals a hydration dependent folding/unfolding mechanism of ribonuclease A. Biophys J. 2006; 91: 2264–2274. PMID: 16798802

128

18. El Moustaine D, Perrier V, Smeller L, Lange R, Torrent J. Full-length prion protein aggregates to amyloid fibrils and spherical particles by distinct pathways. FEBS J. 2008; 275: 2021–2031. doi: 10.1111/j.1742-4658.2008.06356.x PMID: 18355314

19. Girard E, Marchal S, Perez J, Finet S, Kahn R, Fourme R, et al. Structure-Function Perturbation and Dissociation of Tetrameric Urate Oxidase by High Hydrostatic Pressure. Biophys J. 2010; 98: 2365–2373. doi: 10.1016/j.bpj.2010.01.058 PMID: 20483346

20. Kornblatt MJ, Lange R, Balny C. Use of hydrostatic pressure to produce 'native' monomers of yeast enolase. Eur J Biochem. 2004; 271: 3897–3904. PMID: 15373835

21. Ruan KC, Weber G. Dissociation of yeast hexokinase by hydrostatic-pressure. Biochemistry. 1988; 27: 3295–3301. PMID: 3291947

22. Deville-Bone D, Else AJ. Revisible high hydrostatic-pressure inactivation of phosphofructokinase from Escherichia-coli. Eur J Biochem. 1991; 200: 747–750. PMID: 1833191

23. Marchal S, Marabotti A, Staiano M, Varriale A, Domaschke T, Lange R, et al. Under Pressure That Splits a Family in Two. The Case of Lipocalin Family. PLOS ONE. 2012; 7: e50489. doi: 10.1371/journal.pone.0050489 PMID: 23209756

24. Cioni P, Gabellieri E, Marchal S, Lange R. Temperature and pressure effects on C112S azurin: Volume, expansivity, and flexibility changes. Proteins. 2014; 82: 1787–1798. doi: 10.1002/prot.24532 PMID: 24652750

25. Ferreira de Lima Neto D, Bonafe CFS, Arns CW. Influence of High Hydrostatic Pressure on Epitope Mapping of Tobacco Mosaic Virus Coat Protein. Viral Immunol. 2014; 27: 60–74. doi: 10.1089/vim.2013.0088 PMID: 24605789

26. Fowler CB, O'Leary TJ, Mason JT. Improving the Proteomic Analysis of Archival Tissue by Using Pressure-Assisted Protein Extraction: A Mechanistic Approach. J Proteomics Bioinform. 2014; 7: 151–157. PMID: 25049470

27. Ishima R, Torchia DA, Louis JM. Mutational and structural studies aimed at characterizing the monomer of HIV-1 protease and its precursor. J Biol Chem. 2007; 282: 17190–17199. PMID: 17412697

28. Louis JM, Ishima R, Aniana A, Sayer JM. Revealing the dimer dissociation and existence of a folded monomer of the mature HIV-2 protease. Protein Sci. 2009; 18: 2442–2453. doi: 10.1002/pro.261 PMID: 19798742

29. Kožíšek M, Bray J, ezáčová P, Šašková K, Brynda J, Pokorna J, et al. Molecular analysis of the HIV-1 resistance development: Enzymatic activities, crystal structures, and thermodynamics of nelfinavir-resistant HIVprotease mutants. J Mol Biol 2007; 374: 1005–1016. PMID: 17977555

30. Kožíšek M, Šašková KG, ezáčová P, Brynda J, van Maarseveen NM, De Jong D, et al. Ninety-nines not enough: Molecular characterization of inhibitor-resistant human immunodeficiency virus type 1 protease mutants with insertions in the flap region. J Virol. 2008; 82: 5869–5878. doi: 10.1128/JVI.02325-07 PMID: 18400858

31. Ruan K, Tian S, Lange R, Balny C. Pressure Effects on Tryptophan and its Derivatives. Biochem. Biophys Res Commun. 2000; 269(3): 681–686. PMID: 10720476

32. Meher BR, Satish Kumar MV, Kausik Sen (2008) Pressure induced conformational dynamics of HIV-1 protease: A Molecular Dynamics simulation study. In ICIT 2008: Proceedings of the 11th international conference on information technology; Panda, B.S., Nayak, A., eds., 118–122.

33. Veverka V, Bauerova H, Zabransky A, Lang J, Ruml T, Pichova I, et al. (2003) Three-dimensional Structure of a Monomeric Form of a Retroviral Protease. J Mol Biol 333(4): 771–780. PMID: 14568536

34. Gasymov OK, Glasgow BJ (2007) ANS Fluorescence: Potential to augment the identification of the external binding sites of proteins. Biochim Biophys Acta 1774(3): 403–411. PMID: 17321809

35. Iranfar H, Rajabi O, Salari R, Chamani J (2012) Probing the Interaction of Human Serum Albumin with Ciprofloxacin in the Presence of Silver Nanoparticles of Three Sizes: Multispectroscopic and ζ Potential Investigation. J. Phys. Chem. B 116(6), 1951–1964. doi: 10.1021/jp210685q PMID: 22224861

36. Schramm HJ, Boetzel J, Büttner J, Fritsche E, Göhring W, Jaeger E, et al.(1996) The inhibition of human immunodeficiency virus proteases by 'interface peptides'. Antiviral Res. 30(2–3): 155–170. PMID: 8783807

37. Uhlíková T, Konvalinka J, Pichová I, Souček M, Kräusslich HG, Vondrášek J (1996) A modular approach to HIV-1 proteinase inhibitor design. Biochem. Biophys. Res. Commun. 222(1): 38–43. PMID: 8630071

38. Xie D, Gulnik S, Gustchina E, Yu B., Shao W, Qoronfleh W, et al. Drug resistance mutations can effect dimer stability of HIV-1 protease at neutral pH. Protein Sci. 1999; 8: 1702–1707. PMID: 10452615

39. Cheng YS, Yin FH, Foundling S, Blomstrom D, Kettner CA. Stability and activity of human immunodeficiency virus protease: Comparison of the natural dimer with a homologous, single-chain tethered dimer. Proc Natl Acad Sci USA 1999; 87: 9660–9664.

129

40. Zhang ZY, Poorman RA, Maggiora LL, Heinrikson RL, Kézdy FJ. Dissociative inhibition of dimeric enzymes. Kinetic characterization of the inhibition of HIV-1 protease by its COOH-terminal tetrapeptide. J Biol Chem. 1991; 266: 15591–15594. PMID: 1874717

41. Kuzmic P. Kinetic assay for HIV proteinase subunit dissociation. Biochem Biophys Res Commun. 1993; 191: 998–1003. PMID: 8466539

130

P6. Equilibria of oligomeric proteins under high pressure - A theoretical description.

Equilibria of oligomeric proteins under high pressure – A theoretical description

CrossMark
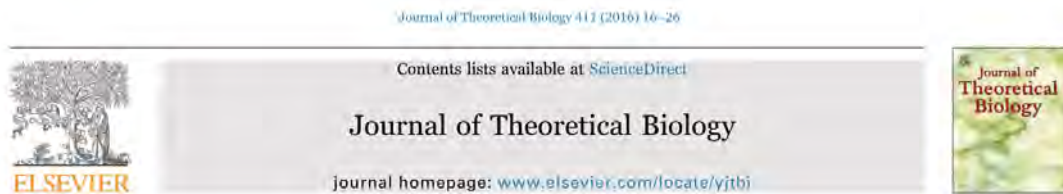
Marek Ingr[a,b,*], Eva Kutálková[a], Josef Hrnčiřík[a], Reinhard Lange[c]

[a] Tomas Bata University in Zlín, Faculty of Technology, Department of Physics and Materials Engineering, nám. T. G. Masaryka 5555, 76001 Zlín, Czechia
[b] Charles University in Prague, Faculty of Science, Department of Biochemistry, Hlavova 2030, 12843 Prague 2, Czechia
[c] Université Montpellier, INRA UMR IATE, Biochimie et Technologie Alimentaires, cc023, Place Eugene Bataillon, 34095 Montpellier Cedex 05, France

ARTICLE INFO

ABSTRACT

High pressure methods have become a useful tool for studying protein structure and stability. Using them, various physico-chemical processes including protein unfolding, aggregation, oligomer dissociation or enzyme-activity decrease were studied on many different proteins. Oligomeric protein dissociation is a process that can perfectly utilize the potential of high-pressure techniques, as the high pressure shifts the equilibria to higher concentrations making them better observable by spectroscopic methods. This can be especially useful when the oligomeric form is highly stable at atmospheric pressure. These applications may be, however, hindered by less intensive experimental response as well as interference of the oligomerization equilibria with unfolding or aggregation of the subunits, but also by more complex theoretical description. In this study we develop mathematical models describing different kinds of oligomerization equilibria, both closed (equilibrium of monomer and the highest possible oligomer without any intermediates) and consecutive. Closed homooligomer equilibria are discussed for any oligomerization degree, while the more complex heterooligomer equilibria and the consecutive equilibria in both homo- and heterooligomers are taken into account only for dimers and trimers. In all the cases, fractions of all the relevant forms are evaluated as functions of pressure and concentration. Significant points (inflection points and extremes) of the resulting transition curves, that can be determined experimentally, are evaluated as functions of pressure and/or concentration. These functions can be further used in order to evaluate the thermodynamic parameters of the system, i.e. atmospheric-pressure equilibrium constants and volume changes of the individual steps of the oligomer-dissociation processes.

1. Introduction

High-pressure methods became a common tool of investigation of structure and function of proteins during the last two decades (Gross and Jaenicke, 1994; Royer, 1995; Mozhaev et al., 1996; Silva et al., 2001; Marchal et al., 2005; Rivalain et al., 2010; Silva et al., 2014). In some cases they are used to study the properties of proteins from marine organisms living deeply under the sea level (Shrestha et al., 2015), but vast majority of these studies is aimed at elucidation of the structure-function relationships of proteins from common organisms extrapolating their high-pressure behavior to the atmospheric pressure. High-pressure methods are used to investigate protein denaturation, unfolding, conformational changes, enzyme kinetics, etc., but they also have valuable application in studying quaternary structure and equilibria of oligomeric proteins. Many oligomeric proteins have been investigated by high-pressure methods, including those of low number of subunits, mainly dimers (Paladini and Weber, 1981; Silva et al.,

1986; Ruan and Weber, 1988; Erijman et al., 1993; Kornblatt et al., 2004; Marchal et al., 2012; Ingr et al., 2015) and tetramers (Jaenicke and Koberstein, 1971; Royer et al., 1986; Ruan and Weber, 1989; Pin et al., 1990; Devillebonne and Else, 1991; Ruan and Weber, 1993; Girard et al., 2010), hexamers (Foguel and Weber, 1995), higher oligomers and viral capsids (Silva and Weber, 1988; Silva et al., 1989, 1992; Da Poian et al., 1993; Silva et al., 1996; Weber et al., 1996) or prion oligomers (Torrent et al., 2015), protein aggregates with less organized structure like casein micelles (Gebhardt et al., 2005, 2006, 2011) and even polymeric structures, e.g. TMV-virus (Bonafe et al., 1998) or microtubules and microfilaments (Messier and Seguin, 1978; Kobori et al., 1996; Nishiyama et al., 2010). Structural changes of oligomeric proteins prior to subunit dissociation were studied, too (Cioni and Strambini, 1996). These studies were concerned with different structural and functional features and in some cases the key thermodynamic parameters, especially the volume change of the oligomer dissociation ΔV and the atmospheric pressure equilibrium

* Corresponding author at: Tomas Bata University in Zlín, Faculty of Technology, Department of Physics and Materials Engineering, nám. T. G. Masaryka 5555, 76001 Zlín, Czechia.
E-mail address: ingr@ft.utb.cz (M. Ingr).

constant K_{am}, were determined (Silva et al., 1986; Ruan and Weber, 1988, 1989; Pin et al., 1990; Da Poian et al., 1993; Erijman et al., 1993; Ruan and Weber, 1993; Foguel and Weber, 1995; Kornblatt et al., 2004; Ingr et al., 2015). These studies exploit the fact that the high pressure favors a process accompanied with a negative change of the total volume of the system. They show that the oligomeric form is destabilized by high pressure, i.e. the total volume of the monomers is lower than that of the oligomer, which is considered as a general rule for the oligomer dissociation processes supported by whole the experimental evidence. This fact allows us to study the dissociation equilibrium even for oligomers highly stable at atmospheric pressure and dissociating only at very low concentration where the signal of the detection methods is insufficient (Royer, 1995). Application of high pressure is most often coupled with different spectroscopic and fluorometric detection techniques, but methods of light (Meier and Kriegs, 2008) or neutron (Shrestha et al., 2015) scattering, as well as optical microscopy (Nishiyama et al., 2006, 2010) or gel electrophoresis (Paladini et al., 1987, 1994), can be used, too. In addition, properties of the monomeric forms of highly stable oligomers as well as intermediate structures of protein unfolding can be studied by high-pressure X-ray crystallography and NMR (Collins et al., 2011).

Application of high pressure can induce various structural changes from oligomer dissociation via reversible unfolding to an irreversible aggregation, sometimes observable at a single protein (Dumay et al., 1994; Seefeldt et al., 2005; Ingr et al., 2015). It is, therefore, necessary to be able to distinguish among these processes, especially unfolding and oligomer dissociation. The processes can be identified according to the concentration dependence of their transition curves. As was shown in numerous experimental studies (Lange et al., 1996; Mozhaev et al., 1996; Ruan et al., 2001; Royer, 2002; Rouget et al., 2010, 2011; Cioni et al., 2014), the transition curve of a reversible folding-unfolding equilibrium, i.e. the dependence of the fraction of one of the forms on pressure, is a concentration independent sigmoid with the inflection point pressure

$$p_{inf} = \frac{RT}{\Delta V} \ln K_{am}$$

(1)

and the fraction of the unfolded form α_n is ½ (for the proof follow the derivation given by Eqs. (3)–(14) below for n=1). The slope of the transition curve in the inflection point is

$$\frac{d\alpha_u}{dp} = -\frac{\Delta V}{4RT}.$$

(2)

These two quantities allow us to determine the thermodynamic characteristics ΔV and K_{am} of the process. On the contrary, the transition curve of the oligomer-monomer equilibrium moves towards higher pressures when the concentration grows. This shift can be used to determine the volume change of the process ΔV and the equilibrium constant K for any pressure, including the atmospheric pressure equilibrium constant K_{am}, as was previously shown for several oligomeric proteins (Ruan and Weber, 1988, 1989, 1993; Foguel and Weber, 1995; Kornblatt et al., 2004; Ingr et al., 2015). In addition, processes with negative ΔV in the direction of association of monomeric subunits are also known. They are usually aggregations with high and not precisely defined number of monomeric units, as was demonstrated on the case of myoglobin (Gebhardt et al., 2003).

Besides many experimental studies, some theoretical works dealing with the thermodynamics of oligomeric-proteins dissociation under high pressure were published as well (Weber, 1986, 1993). In this paper we provide a contribution to the theoretical analysis of some of these equilibria with the stress on the detailed description of the transition curves, especially their significant points, i.e. inflection points and extremes, which may be used as a versatile tool for evaluation of eventual future experiments with oligomeric proteins.

2. Results and discussion

2.1. General assumptions

In this work we describe closed equilibria (i.e. equilibria between the highest oligomer and the monomer without any intermediate states) of homooligomers of any degree, heterodimer and heterotrimers, and consecutive equilibria (containing intermediate oligomers of lower degree than the highest one) of homo- and heterotrimer. In all the cases only pressure-independent negative volume changes of oligomer dissociations will be considered as it seems to be a good approximation supported by the overall experimental evidence as well as our recent theoretical simulation (Kutalkova et al., 2014).

The parameters that should be determined using the proposed theoretical background are especially the volume changes ΔV accompanying the individual oligomer-dissociation steps and the atmospheric-pressure equilibrium constants of these processes K_{am}. Their determination is based on the analysis of the transition curves – i.e. responses of the experimental device to the system under changing pressure at a given concentration or, in a reciprocal approach, changing concentration at a given pressure.

For simplicity, the concentrations of individual chemical entities are denoted by simple capital letters with intuitive meaning (M – monomer, A – subunit A, D – dimer, etc.) equal with those used in the respective chemical equations. All the concentrations are considered as dimensionless relative quantities related to the standard concentration of 1 mol dm^{-3}. Accordingly, the equilibrium constants are dimensionless, too. As many of the mathematical derivations are rather lengthy, they are presented in Electronic Supplementary Information, hereafter referred to as ESI.

2.2. Closed equilibria systems

2.2.1. Homooligomers

Consider the equilibrium between a homooligomeric protein M_n, consisting of n monomeric subunits, and its subunits M described by a chemical equation

$$M_n \;\rightleftarrows\; n\,M.$$

(3)

The equilibrium constant of this process for a given pressure p is

$$K(p) = \frac{(M)^n}{M_n}.$$

(4)

Denoting the total protein concentration related to the monomeric form M_0 and considering the balance equation

$$M_0 = M + n\,M_n,$$

(5)

the relation between M_0 and the monomer fraction $\alpha = M/M_0$ is given by the equation

$$n\,\alpha^n + \frac{K(p)}{M_0^{n-1}}\alpha - \frac{K(p)}{M_0^{n-1}} = 0.$$

(6)

The reaction change of the Gibbs energy is

$$\Delta G = \Delta G_{am} + \Delta p \Delta V$$

(7)

where ΔG_{am} is the same quantity at atmospheric pressure and $\Delta p = p - p_{am}$ is the difference between the current and atmospheric pressures. For simplicity, we approximate Δp by p because p_{am} is negligible in comparison with p, which is in the order of tens to hundreds MPa in all relevant experiments. As

$$K(p) = \exp\left(-\frac{\Delta G}{RT}\right) = \exp\left(-\frac{\Delta G_{am} + p\Delta V}{RT}\right) = K_{am}\exp\left(-\frac{p\Delta V}{RT}\right)$$

(8)

where R is the molar gas constant and T is the thermodynamic temperature, Eq. (6) becomes

$$n \ \alpha^n + \frac{K_{atm} \exp\left(-\frac{p\Delta V}{RT}\right)}{M_0^{n-1}} (\alpha - 1) = 0. \tag{9}$$

This equation can be solved analytically for α as a function of pressure only for $n \leq 4$ and only for $n = 2$ it is relatively simple. In this case

$$\alpha(p) = \frac{K_{atm} e^{-\frac{p\Delta V}{RT}}}{4M_0} \left(-1 + \sqrt{1 + \frac{8M_0}{K_{atm} e^{-\frac{p\Delta V}{RT}}}} \right) \tag{10}$$

(Fig. 1). However, it is generally possible to express K as a function of α.

$$K(p) = K_{atm} \exp\left(\frac{p\Delta V}{RT}\right) = \frac{n \ \alpha(p)^n}{(1 - \alpha(p))} M_0^{n-1}. \tag{11}$$

Based on this equation K_{atm} and ΔV can be determined provided that $\alpha(p)$ is known for a series of pressure points and $K(p)$ can be determined for each of them, as was shown previously (Ruan and Weber, 1988, 1989, 1993; Foguel and Weber, 1995; Kornblatt et al., 2004; Ingr et al., 2015). In this case

$$\ln K(p) = \ln K_{atm} - \frac{p\Delta V}{RT}. \tag{12}$$

The expression $-\Delta V/RT$ is thus the slope of the linear dependence of $\ln K$ on pressure while $\ln K_{atm}$ is its intercept on y-axis. A drawback of the experimental method based on Eq. (12) may arise from the necessity to measure the signal for the whole pressure range in order to reach the extreme cases, i.e. practically pure monomer on one side and oligomer on the other side. This may be difficult for several reasons. First, the "pure" oligomer region may be experimentally inaccessible if the oligomer is not highly stable. Second, the "pure" monomer region may be biased by interference of the oligomer dissociation with other structural transitions, unfolding or aggregation, that often take place at high pressure. Finally, the spectral response can drift with pressure in a manner independent of structural transitions, often differently for the monomer and oligomer cases.

A more robust evaluation method based on the inflection points of the transition curves can be derived from Eq. (9) expressing p as a function of α:

$$p(\alpha) = \frac{RT}{\Delta V} \ln \frac{(1 - \alpha)}{n \ \alpha^n} + \frac{RT}{\Delta V} \ln K_{atm} - (n - 1)\frac{RT}{\Delta V} \ln M_0. \tag{13}$$

It can be shown (see ESI, Section S-1) that its inflection point and the corresponding value of α are given by the expressions

$$p_{infl} = \frac{RT}{\Delta V}\left[(n - 1)\ln(1 + \sqrt{n}) - \left(\frac{n}{2} + 1\right)\ln n \right] + \frac{RT}{\Delta V} \ln K_{atm}$$
$$- (n - 1)\frac{RT}{\Delta V} \ln M_0 \tag{14}$$

$$\alpha_{infl} = \frac{n - \sqrt{n}}{n - 1}. \tag{15}$$

Obviously, for a given n the inflection point occurs at constant value of α independently of other parameters – this value moves towards 1 with growing number of subunit (Fig. S-2). This effect can be observed e.g. in (Bonafe et al., 1994) – see Fig. 4 and solid circles in Fig. 5 thereof – where a protein consisting of 20 identical subunits is studied.

Eq. (14) represents the inflection point pressure as a linear function of the logarithm of the total protein concentration (expressed as monomer). The thermodynamic parameters ΔV and K_{atm}, eventually $\Delta G_{atm} = -RT \ln K_{atm}$, can be obtained by a linear regression of this function. Using previously published works it can be shown that this approach gives results in a good agreement with those reported by the authors (Paladini and Weber, 1981; Ruan and Weber, 1988; Da Poian et al., 1993; Foguel and Weber, 1995) in spite of the fact that the published transition curves were usually measured only at two different

concentrations. Certainly, if the transition curves can be reliably detected in the whole range of α, then a single curve measured at any concentration is, in principle, sufficient to evaluate ΔV and K_{atm} in accord with Eq. (12). However, if the limits of the transition curve for α tending to 0 and 1 cannot be well detected, but the central part containing the inflection point is still observable, then the evaluation method based on Eq. (14) is an optimal choice, although the curve should be measured for more concentrations. As an example, consider the data by Ruan and Weber (Ruan and Weber, 1988) who determined, according to their Fig. 2, the thermodynamic parameters for dissociation of hexokinase dimer of $\Delta V = -150 \ ml \ mol^{-1}$ and $K_{atm} = 6.1 \times 10^{-10}$. Using the same plot, we can estimate the inflection-point pressures as 140 MPa for 1.7 µM dimer and 100 MPa for 0.17 µM dimer. Linear regression according to Eq. (14) then gives $\Delta V = -131 \ ml \ mol^{-1}$ and $K_{atm} = 8.9 \times 10^{-10}$ showing a good agreement of both the methods. (For more detailed analysis including the reproduced original plot see ESI, Section S-15.) However, this comparison must be taken with caution since only two points were taken into account.

Actually, when ΔG_{atm} is not too small in absolute value, i.e. K_{atm} is substantially different from 1, the first term in Eq. (14) is negligible since the expression in brackets is within the range $(-1; 1)$ for all values of n up to 59 (for $n = 8$ it is 1.00003) and even further it does not grow dramatically. Thus, in these cases the y-axis intercept can be in a good approximation identified with $RT \ln K_{atm}/\Delta V$. (For instance, when $K_{atm} = 10^{-3}$, i.e. $\Delta G_{atm} = 17,1 \ kJ \ mol^{-1}$ at 25 °C, the value of $\ln K_{atm} = -6.9$ and thus the error of this quantity is at maximum 17% and is decreasing when the oligomer is more stable at atmospheric pressure.) It may be useful, especially for oligomers with higher number of subunits n, to define a volume change related to one monomer as

$$\Delta V_{mon} = \Delta V/n. \tag{16}$$

The slope of Eq. (14) then takes the form $[-(n - 1)/n]RT/\Delta V_{mon}$, which tends to $-RT/\Delta V_{mon}$ for higher n. Hence, it is possible to determine the volume change per monomer unit for equilibria of higher oligomers even when the number of subunits n is not known. It should, however, be known in order to determine ΔV Eq. (16) and K_{atm}.

Although for most of proteins n is usually known, its determination can be desirable for many-subunits, e.g. micelles or viral capsids. It can be estimated from Eq. (15), but it is necessary to measure the transition curve in a wide pressure range in order to see both the limit cases, complete oligomerization and complete monomerization of the system. The former state can be usually assumed at the atmospheric pressure. The latter state may be approximated by the point of maximum curvature of the transition curve which is always very close to $\alpha = 1$,
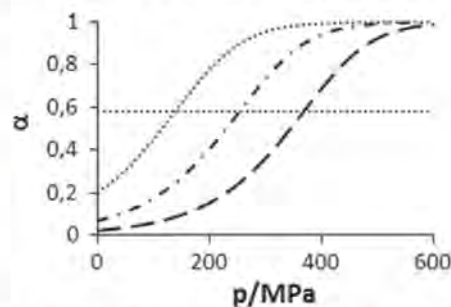
Fig. 1. Monomer-dimer equilibrium at high pressure. Fraction of monomer α is plotted as a function of pressure for different total concentrations of monomer. $K_{atm}=10^{-6}$, $\Delta V = -50$ ml mol^{-1}, $R=8.314$ J mol^{-1} K^{-1}, $T=300$ K; —— $M_0=10^{-3}$; ---- $M_0=10^{-4}$; ···· $M_0=10^{-5}$. Inflection points of all the curves lie on the horizontal dotted line – Eq. (15). Concentrations are given in mol l^{-1}.
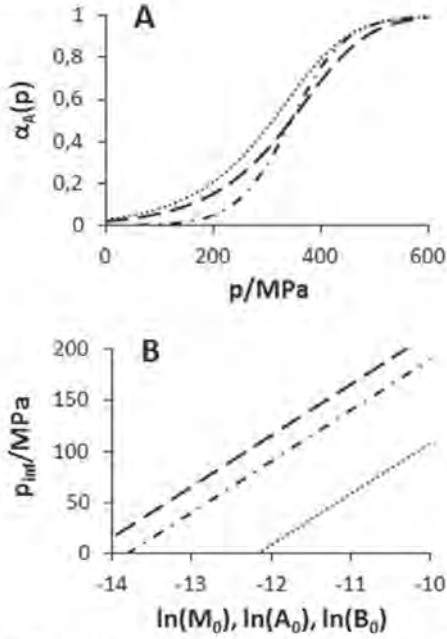
Fig. 2. A: Fraction of the monomeric subunit A as a function of pressure for heterodimer at stoichiometric (·····) and non-stoichiometric (-----) ratio of concentrations of subunits A and B and for a homodimer (———). $K_{d, atm}=10^{-6}$; $\Delta V=-50$ ml mol^{-1}; $R=8.314$ J mol^{-1} K^{-1}; $T=300$ K; $B_0=10^{-3}$ mol l^{-1}; $M_0=10^{-3}$ mol l^{-1}. In case of non-stoichiometric ration of A and B the concentration $A_0=10^{-8}$ mol l^{-1}. B: Linear dependences of inflection points of transition curves plotted in panel A on the monomer concentration.

especially for high values of n (see Fig. S-2). The region behind this point can be used to identify the biasing trend of the curve. This determination should be, however, considered only as a rough estimate, due to the eventual limited sensitivity of the experimental methods.

2.2.2. Heterodimer

Heterodimer is the only system consisting of non-equal subunits that necessarily undergoes only one-step oligomerization equilibrium

$$D \;\rightleftharpoons\; A + B. \tag{17}$$

It is, therefore, interesting to compare it with homodimer. Considering the dissociation constant definition

$$K_d = \frac{AB}{D} \tag{18}$$

and balance equation

$$A_0 = A + D \tag{19}$$

(in case that the total concentrations of subunits A and B are different, we consider, without the loss of generality, A to be the lower concentrated one), the fraction of monomeric form of A defined as

$$\alpha_A = \frac{A}{A_0} \tag{20}$$

is obtained as a solution of the quadratic equation

$$A_0 \alpha_A^2 + \alpha_A (B_0 - A_0 + K_d(p)) - K_d(p) = 0. \tag{21}$$

This equation gives, after introducing the pressure dependence of

K_d given by Eq. (8), a solution

$$\alpha_A(p) = \frac{A_0 - B_0 - K_{d,atm} e^{\frac{-p\Delta V}{RT}}}{2A_0} \left(1 + \sqrt{1 + \frac{4 K_{d,atm} e^{\frac{-p\Delta V}{RT}} A_0}{\left(A_0 - B_0 - K_{d,atm} e^{\frac{-p\Delta V}{RT}}\right)^2}} \right) \tag{22}$$

which is plotted in Fig. 2.

In analogy with the previous section, an equation can be derived that gives the value of the inflection point α_A as a function of A_0 and B_0:

$$\alpha_A^4 - 4\alpha_A^3 + 2\alpha_A^2 \left(3 - 2\frac{B_0}{A_0}\right) - 2\alpha_A \left(+\frac{B_0^2}{A_0^2} - 3\frac{B_0}{A_0} + 2\right) + \frac{B_0^2}{A_0^2} - 2\frac{B_0}{A_0} + 1 = 0. \tag{23}$$

This equation is of the fourth order and thus difficult to solve in general. However, the equilibrium can be studied in two significant special cases, for $A_0 = B_0$ and for $B_0 >> A_0$. In the former case, which corresponds e.g. to the work with isolated natural heterodimeric protein, the linear and absolute terms in α_A vanish and the equation gets, after some algebra, the form

$$\alpha_A^2 - 4\alpha_A + 2 = 0. \tag{24}$$

with the solution $\alpha_{A,inf} = 2 - \sqrt{2} = 0.586$, equally as in the case of a homodimer. The corresponding inflection point pressure is

$$p_{inf} = \frac{RT}{\Delta V} \ln\left(\frac{1 + \sqrt{2}}{2}\right) + \frac{RT}{\Delta V} \ln K_{d,atm} - \frac{RT}{\Delta V} \ln A_0. \tag{25}$$

Thus, this expression is completely analogous with Eq. (14) and can be analyzed in the same way in order to obtain $K_{d,atm}$ and ΔV. Its use can be verified e.g. on the published data by Foguel and Weber (Foguel and Weber, 1995), see ESI, Section S-15. The only difference is in the logarithmic expression in the first term which has the value of $\ln(1.207) = 0.188$ for heterodimer, while for homodimer it is equal to $\ln[(1 + \sqrt{2})/4] = -0.504$.

The latter case, in which $B_0 >> A_0$, can be helpful especially at systems of less stable heterodimers where the individual subunits can be either purified or prepared by recombinant expression as independent proteins. In this case the concentration of subunit B can be considered as a constant of the value B_0. Then

$$\alpha_A = \frac{1}{\frac{B_0}{K_{d,atm}} e^{\frac{p\Delta V}{RT}} + 1} \tag{26}$$

and the inflection point pressure of this transition curve is

$$p_{inf} = \frac{RT}{\Delta V} \ln K_{d,atm} - \frac{RT}{\Delta V} \ln B_0 \tag{27}$$

(for derivation see ESI, Section S-2) which allows us to determine ΔV and $K_{d,atm}$ by means of a linear regression. Thus, the analysis of a heterodimer is formally equal to that of a homodimer, but the subunit difference gives us two ways how to easily determine the parameters $K_{d,atm}$ and ΔV, one suitable especially for highly stable heterodimers and the other one for heterodimers of lower stability.

2.2.3. Closed equilibria of heterotrimers

As the subunits of heterotrimers are not equal, it is generally reasonable to consider the equilibria in heterotrimers as consecutive. For instance, the dissociation of a trimer consisting of two subunits A and one subunit B (A_2B heterotrimer) may consist of the trimer dissociation to A_2 and B followed by the dissociation of the A_2 dimer. However, the consecutive equilibrium changes to the closed equilibrium when the dimer is unstable in comparison with the trimer, and therefore the dimeric subsystem A_2 exists in substantial amount only in a complex with the subunit B. In this case the following equilibrium can be expected

19

134

$$T \rightleftharpoons 2A + B, \tag{28}$$

which is characterized by the dissociation constant

$$K = \frac{A^2 B}{T} \tag{29}$$

and the balance equations

$$A_0 = A + 2T; \qquad B_0 = B + T. \tag{30}$$

Let us define the fractions of monomer and trimer as fractions of the subunit A present in one or another state, i.e.

$$a_M = \frac{A}{A_0}; \qquad a_t = \frac{2T}{A_0}. \tag{31}$$

At the stoichiometric ratio of the subunits A and B the position of the inflection point of the trimer-monomer transition curve is given by a linear function

$$p_{\text{inf}} = \frac{RT}{\Delta V} \ln\left(\frac{6 + 4\sqrt{3}}{9} K_{\text{atm}}\right) - 2\frac{RT}{\Delta V} \ln A_0 \tag{32}$$

which can be used to determine the thermodynamic parameters K_{atm} and ΔV.

Analogously, a similar derivation can be carried out for the heterotrimer composed of three different subunits A, B and C (ABC heterotrimer). Considering the stoichiometric ratios of all three subunits, we arrive at a solution identical to the A$_2$B heterotrimer, the inflection point pressure is thus given by Eq (32).

2.3. Consecutive-equilibria systems

2.3.1. Homotrimer

If an oligomer consists of more than two subunits, its association or dissociation can run consecutively, i.e. via one or several intermediate states. The situation was partially described by Ruan and Weber (Ruan and Weber, 1989) for the system homotetramer-dimer-monomer and later even studied experimentally by Foguel and Weber on a hexamer-dimer-monomer system (Foguel and Weber, 1995). In both the cases, however, the dimer was very unstable and thus the equilibrium was reduced to the closed one, as discussed below. Even simpler example is a protein consisting of three identical subunits (homotrimer) which can consecutively dissociate from trimer (T) to dimer (D) plus monomer (M) and the dimer can further dissociate to monomers:

$$T \rightleftharpoons D + M; \qquad D \rightleftharpoons 2M. \tag{33}$$

At a given pressure the equilibrium of the system is described by the trimer and dimer dissociation constants

$$K_T = \frac{DM}{T}; \qquad K_D = \frac{M^2}{D}, \tag{34}$$

If the total concentration of the protein expressed in terms of the monomeric units is M_0, the balance equation

$$M_0 = M + 2D + 3T \tag{35}$$

is valid. The mathematical model derivation is complicated by the presence of the two pressure-dependent equilibrium constants. It is, therefore, convenient to solve the problem for a fixed pressure and then introduce the pressure dependence to the obtained solutions.

Let us define the fractions of monomeric units present in the forms of monomer, dimer and trimer

$$a_M = \frac{M}{M_0}; \qquad a_D = \frac{2D}{M_0}; \qquad a_T = \frac{3T}{M_0}, \tag{36}$$

respectively. Considering this, the following equation is obtained (for derivation see ESI, Section S-4):

$$\sqrt{\frac{K_D a_D}{2M_0}} + a_D + 3\sqrt{\frac{K_D a_D^2 M_0}{8K_T^2}} = 1. \tag{37}$$

Fig. 3. Consecutive equilibrium of homotrimer – concentration dependence. Fractions of trimer, dimer and monomer as functions of protein concentration (expressed as a monomer) for different sets of thermodynamic parameters. A: $K_{D,\text{atm}} = 10^{-8}$; $K_{T,\text{atm}} = 10^{-4}$; $\Delta V_D = -20$ ml mol^{-1}; $\Delta V_T = -50$ ml mol^{-1}; B: $K_{D,\text{atm}} = K_{T,\text{atm}} = 10^{-4}$; $\Delta V_D = -50$ ml mol^{-1}; $\Delta V_T = -20$ ml mol^{-1}. In both cases $T = 298,15$ K; $p = 150$ MPa; ····· monomer; ---- dimer; —— trimer.

This equation can be solved for a_D as a function of M_0 and, consequently, a_M and a_T can be determined, too. However, as the equation is of the third degree in a_D, the solution is rather cumbersome and complicated to understand. Nevertheless, the equation was solved by the Wolfram Mathematica 9 package and the result is shown in Fig. 3 for several sets of the thermodynamic parameters. The complete expression is given in ESI, Appendix S-1.

If the pressure dependence of K_D and K_T is introduced in analogy with Eq. (8) with $K_{D,\text{atm}}$ and $K_{T,\text{atm}}$ being their atmospheric-pressure values and V_D and V_T the respective volume changes, fractions of the individual states can be plotted also as functions of pressure. Fig. 4 presents this plot for two different sets of parameters.

A more complex view of this situation can be obtained if a 3D plot is constructed of the fraction of a dimer (or, eventually, the other states) as 2D functions of concentration and pressure (Fig. 5). Here, it can be seen that the shape of the function differs considerably depending on the thermodynamic parameters and the positions of the significant points of the standard 1D functions can be deduced from it.

As can be seen in Fig. 4, the nature of the equilibrium at given pressure differs in dependence on the values of equilibrium constants. If $K_D << K_T$, i.e. if the dimer is remarkably more stable than the trimer, there is a wide range of dimer prevalence and practically no region of co-existence of all three forms. On the other hand, when the trimer's stability highly exceeds that of the dimer, i.e. $K_T << K_D$, the dimer is almost absent at any concentration and the trimer directly decays to the monomer – the equilibrium thus changes to the closed one and can be described by Eq. (6). Depending on the pressure, the system can change its nature between these two cases in several different ways, as is shown in Fig. 4A and B. Regarding the experi-
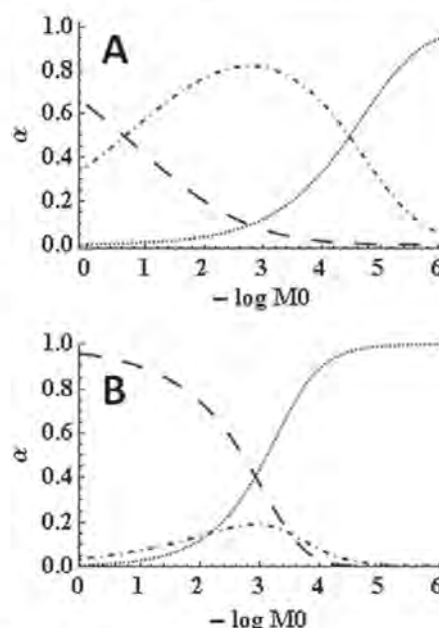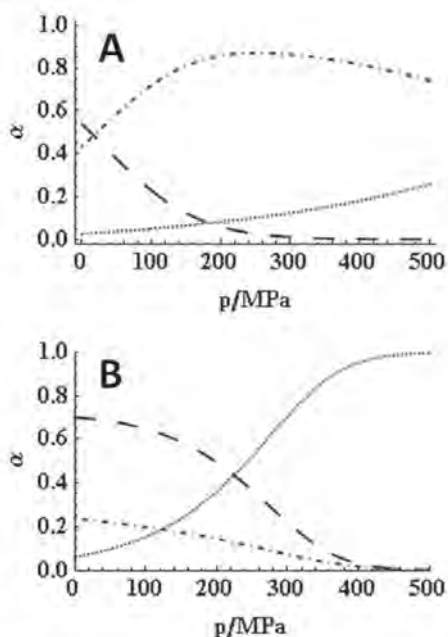
135

Fig. 4. Consecutive equilibrium of homotrimer – pressure dependence. Fraction of trimer, dimer and monomer as functions of pressure for different sets of thermodynamic parameters. A: $K_{D,atm}=10^{-5}$; $K_{T,atm}=10^{-4}$; $\Delta V_D=-20$ ml mol^{-1}; $\Delta V_T=-50$ ml mol^{-1}; B: $K_{D,atm}=K_{T,atm}=10^{-4}$; $\Delta V_D=-50$ ml mol^{-1}; $\Delta V_T=-20$ ml mol^{-1}; In both cases $T=298,15$ K; $-\log M_0=2.5$: ····· monomer; ---- dimer; ——— trimer.

mental approach, it is, therefore, necessary to measure the experimental response of the system as a function of concentration and pressure on a 2D area given by a product of suitably long intervals of both these quantities. Depending on the method of choice, the experimental response can be measured either as a function of concentration at a fixed pressure, repeating the measurement for different pressure values, or vice versa.

It should be noted that at the consecutive equilibria of trimers or higher oligomers the sensitivity of the experimental methods to the individual oligomerization states is a crucial issue. Contrary to the closed equilibria, where only two states have to be distinguished, consecutive systems require at least three-state resolution. Depending on the mutual ratios of the detected signals of the individual states, the overall transition curve of a given system may look differently for different detection methods. Determination of the relative contributions of each state to the total signal may be possible in case that the states can be isolated chemically (Foguel and Weber, 1995). Nevertheless, it is especially valuable if a method sensitive exclusively to one of the states can be used. Possible suggestions of such techniques based on the Förster resonant energy transfer (FRET) (Sun et al., 2010) and neutron scattering (Jacrot, 1976, Svergun and Nierhaus, 2000) are given in ESI, Section S-7.

As the nature of the equilibrium may change with the varying pressure and concentration, it is worthwhile to identify regions where the complex model can be approximated by linear equations. The simplest way is to choose a range of pressure (or concentration in a complementary approach) with a significant area of dimer prevalence when the concentration (pressure) changes from the lower to the upper border of its interval. If this range exists and has a sufficient extent, the two inflection points corresponding to the transitions from trimer to

dimer and from dimer to monomer are well separated and are characterized by a practical absence of the third form of the system, i.e. monomer in the former and trimer in the latter case. In this case the last or the first term in Eq. (37) can be neglected for the trimer-dimer (TD) and the monomer-dimer (DM) transitions, respectively. The solutions of the resulting equations lead to linear dependences of the inflection point pressures on $\ln(M_0)$ (for derivation see ESI, Section S-6)

$$P_{inf,TD} = \frac{RT}{2\Delta V_T - \Delta V_D} \ln\left(\frac{K_{T,atm}^2}{K_{D,atm}} \frac{8(6+2\sqrt{6})}{81}\right) - \frac{RT}{2\Delta V_T - \Delta V_D} \ln M_0 \tag{38}$$

$$P_{inf,DM} = \frac{RT}{\Delta V_D} \ln\left(K_{D,atm} \frac{1+\sqrt{2}}{4}\right) - \frac{RT}{\Delta V_D} \ln M_0 \tag{39}$$

which allows us to determine the thermodynamic parameters $K_{D,atm}$, $K_{T,atm}$, V_D, and V_T by means of a linear regression.

Similarly, the thermodynamic parameters can be estimated from the inflection points of α_D as a function of pressure provided that both the inflection points exist and are well separated in the experimentally accessible region. As can be proved (see ESI, Section S-6), the inflection points are identical with those identified in the previous way, therefore Eqs. (38) and (39) remain unchanged. It is, therefore, only necessary to decide which way is more convenient for the specific experimental data.

In case that no region of dimer prevalence exists in the experimentally accessible region of pressure and concentration, it is probable that an area exists where, on the contrary, $K_T << K_D$. In this situation the dimer is practically absent from the system and the equilibrium can be considered as closed and analyzed in accord with Eq. (14). The equilibrium constant K of Eq. (14) is then equal to the product $K_T K_D$, and, accordingly, $K_{atm} \rightarrow K_{T,atm} K_{D,atm}$ and $\Delta V \rightarrow \Delta V_T + \Delta V_D$. In case that $K_T << K_D$ everywhere in the experimentally accessible region, the individual equilibrium constants and volume changes for the processes of dimerization and trimerization cannot be determined and their separation has no meaning.

In some cases, however, none of the discussed simplified cases takes place in a sufficiently large region. For instance, if $\Delta V_D = \Delta V_T$ (Fig. 5C), the inflection points run along parallel lines and the pressure changes have no effect on their separation. If their separation corresponds with none of the previously discussed cases, determination of the thermodynamic parameters from simple linear functions is impossible. In such cases it may be convenient to start from an estimate based on the distance between the inflection points of the experimentally determined transition curve. These points do not have to correspond with the inflection points of α_D, but rather of α_M (for the DM transition – low concentration side) and α_T (for the TD transition – high concentration side). Solving Eq. (37) for α_M and determining the inflection point of α_M (as a function of concentration) numerically in Wolfram Mathematica 9 it can be demonstrated that this inflection point occurs within the interval $\alpha_{M,inf} \in (0.56; 0.64)$ for any values of K_D and K_T. The corresponding variation of $\ln M_0$ is small, therefore an arbitrary value from this interval can be taken to estimate the thermodynamic parameters. Analogously, the inflection point of $\alpha_{T,inf}$ occurs within the interval $(0.33; 0.45)$. If it is possible to estimate the nature of the equilibrium, both the values can be chosen with higher accuracy as they can be determined exactly for some special cases (Table 1).

The inflection point concentrations (in the logarithmic form) are then given by

$$\ln M_{T,inf} = \ln 3 - \ln \alpha_{T,inf} + 2\ln K_D - \ln K_T$$
$$- 3\ln\left(\sqrt{1 + 3\frac{K_D(1-\alpha_{T,inf})}{K_T \alpha_{T,inf}}} - 1\right) \tag{40}$$
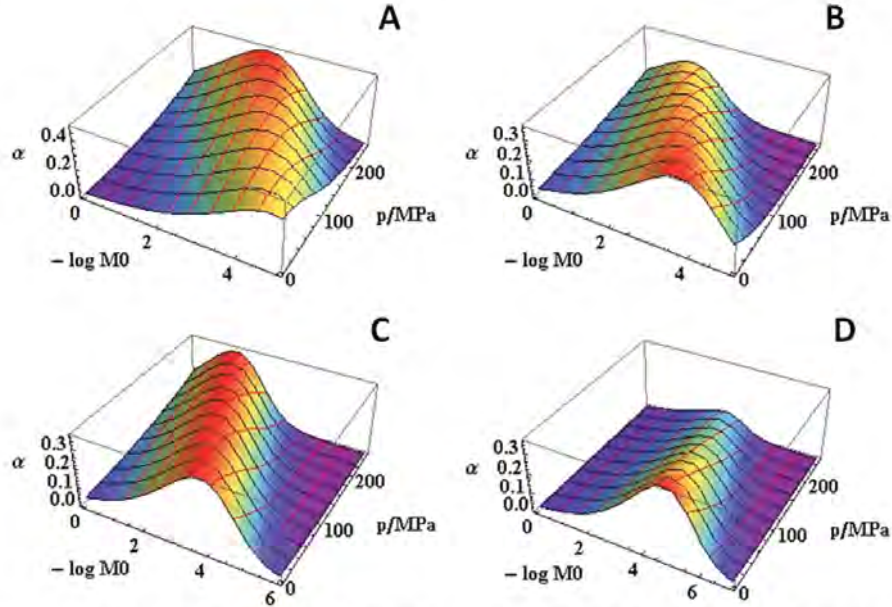
Fig. 5. Dimer fraction as a function of pressure and concentration for different sets of thermodynamic parameters. A: Volume change (in absolute value) of the trimer-dimer (TD) transition is bigger than that of the dimer monomer (DM) transition – the separation of the TD and DM transitions is growing with increasing pressure. $K_{D,atm}=K_{T,atm}=10^{-5}$; $\Delta V_D=-50$ ml mol^{-1}; $\Delta V_T=-60$ ml mol^{-1}; B: opposite situation than in panel A – the separation of the transitions is decreasing with growing pressure $K_{D,atm}=K_{T,atm}=10^{-4}$; $\Delta V_D=-60$ ml mol^{-1}; $\Delta V_T=-50$ ml mol^{-1}; C: volume changes of both the transitions are equal, the separation of TD and DM transitions is pressure-independent. $K_{D,atm}=K_{T,atm}=10^{-4}$; $\Delta V_D=\Delta V_T=-50$ ml mol^{-1}; D: absolute value of ΔV_D is more than twice bigger than absolute value of ΔV_T – no cross sections of the 3D plot of α_D at constant concentration M_0 has maximum – it only decreases with growing pressure. $K_{D,atm}=K_{T,atm}=10^{-5}$; $\Delta V_D=-50$ ml mol^{-1}; $\Delta V_T=-20$ ml mol^{-1}. In all cases $T=298,15$ K.

Table 1

Inflection-point values of monomer and trimer fractions of a homotrimer for different natures of the equilibrium.

Nature of equilibrium	$\alpha_{M,inf}$	$\alpha_{T,inf}$
General	0.56–0.64	0.33–0.45
Closed	0.634	0.366
Consecutive – well separated transitions	0.586	0.450
Consecutive – $K_T = K_D$	0.562	0.334

$$\ln M_{M,inf} = -\ln 3 - \ln \alpha_{M,inf} + \ln K_T + \ln\left(\sqrt{1 + 3\frac{K_D(1 - \alpha_{M,inf})}{K_T \alpha_{M,inf}}} - 1\right). \tag{41}$$

These lines, together with other significant-point dependences on concentration and pressure, are depicted in Fig. 6.

It should be pointed out that the difference of these expressions, i.e. the approximate distance between the inflection points, is only a function of the ratio K_D/K_T:

$$\ln M_{M,1/2} - \ln M_{T,1/2} = -\ln 9 + \ln \frac{\alpha_{T,inf}}{\alpha_{M,inf}} - 2\ln\frac{K_D}{K_T}$$
$$+ + \ln\left(\sqrt{1 + 3\frac{K_D(1 - \alpha_{M,inf})}{K_T \alpha_{M,inf}}} - 1\right)$$
$$+ 3\ln\left(\sqrt{1 + 3\frac{K_D(1 - \alpha_{T,inf})}{K_T \alpha_{T,inf}}} - 1\right). \tag{42}$$

This equation, together with Eqs. (40) and (41), can thus be used to estimate the desired thermodynamic parameters using methods based on approximate model functions (see ESI, Sections S-8 and S-9). If the

experimental methods enable reliable determination of the functions α_M and α_T, an iterative method can be used to determine the parameters with higher accuracy (see ESI, Section S-10).

In some cases the experimental methods may allow us to determine also the maximum of α_D, either as a function of concentration at a given pressure, or vice versa. If α_D is considered as a function of concentration, its maximum also corresponds with the crossing point of α_M and
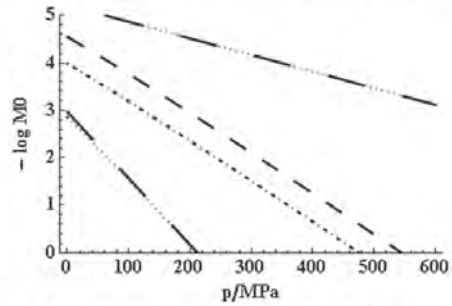
Fig. 6. Significant points of the homotrimer equilibrium. Individual significant points as functions of pressure and concentration lie on curves not much different from linear functions. Inflection point of TD transition is represented by the two lowest lines, dotted and dashed, inflection point of DM transition by the two uppermost lines (dotted – Eq. (40) with $\alpha_{T,inf} = 0.450$ and Eq. (41) with $\alpha_{M,inf} = 0.586$, respectively; dashed – estimate according to Eqs. (38) and (39), respectively. The curves of maxima in the concentration (dot-dashed – Eq. (43)) and pressure (short-dashed – Eq. (44)) domains are close to parallel lines. $K_{D,atm} = 10^{-5}$, $K_{T,atm} = 10^{-4}$, $\Delta V_D = -20$ml mol^{-1} and $\Delta V_T = -50$ ml mol^{-1}.

137

α_T and occurs at the concentration given by the formula

$$\ln[M_{0,\max}(p)] = \ln\left(\frac{2}{3}K_{T,atm}\right) - \frac{p\Delta V_T}{RT}$$
$$+ \ln\left(1 + \sqrt{\frac{3K_{D,atm}}{K_{T,atm}}\exp\left[\frac{p(\Delta V_T - \Delta V_D)}{RT}\right]}\right) \quad (43)$$

Its derivation is given in ESI, Section S-4, together with its application in an iterative method of determination of the thermodynamic parameters (Section S-10).

In many cases it can be more convenient to measure the signal at a given concentration as a function of pressure. The inflection points of the transition curve can be again approximated by the inflection points of α_M and α_T. Using all the previous approximations, these points are given by Eqs. (40) and (41). A set of linear equations can be derived that allow us to estimate the thermodynamic parameters, and, if the functions α_M and α_T can be determined reliably, an iterative procedure can be used to determine the parameters accurately (see ESI, Section S-9 and S-10). It is also possible to find the maximum of α_D at a given concentration (as a function of p) which is, in general, a different point than the maximum at constant pressure (see Fig. 6). As can be seen in ESI, Section S-5, its position is given by the formula

$$\ln M_0(p_{max}) = \ln\left(\frac{2}{3}K_{T,atm}\right) + \ln\left(\frac{\Delta V_D}{2\Delta V_T - \Delta V_D}\right) - \frac{p_{max}\Delta V_T}{RT} +$$
$$+ \ln\left(1 - \exp\left[\frac{p_{max}(\Delta V_T - \Delta V_D)}{2RT}\right]\Delta V_T\sqrt{\frac{3K_{D,atm}}{K_{T,atm}\Delta V_D(2\Delta V_T - \Delta V_D)}}\right), \quad (44)$$

This equation is rather complicated for direct determination of the thermodynamic parameters, but can be used in an iterative way analogously to Eq. (43) (see ESI, section S-10). It should be pointed out that this maximum does not exist for all possible values of ΔV_D and ΔV_T. As follows from the derivation given in ESI, Section S-5, the condition $|\Delta V_T| > |\Delta V_D|/2$ has to be fulfilled. If it is not, the fraction of dimer α_D decreases with growing pressure within the whole range irrespective of the concentration, because the dimer is destabilized much more rapidly than trimer and the DM transition has always larger extent than the TD transition (Fig. 5D). In this case the analysis based on Eq. (44) cannot be carried out. However, the thermodynamic parameters can be evaluated by analyzing the pressure-dependent transition curve at two different concentrations, the first one at which the system at atmospheric pressure is predominantly trimeric and the second one at which it is mostly dimeric. In the first situation the system reduces to the closed trimer-monomer equilibrium, in the second one to the dimer-monomer equilibrium, both described by Eq. (13).

It is, unfortunately, impossible to demonstrate the validity of the theory on real experimental data, since, to our knowledge, no experiment directly coherent with this model was published so far. However, it was reported several times that for equilibria considered as closed there was an incoherence between the ΔV values determined from the individual transition curves (in accord with Eq. (12), often denoted as ΔV_β in the literature) and from the shift of transition curves (in analogy with Eq. (14), denoted as ΔV_C), the latter method providing higher values (King and Weber, 1986; Silva et al., 1986; Ruan and Weber, 1993). This discrepancy was always observed only for higher oligomers than dimer, even for oligomers with high number of subunits like viral capsids (Weber et al., 1996). It was explained by the heterogeneity of the oligomer population and the deterministic equilibrium (Erijman and Weber, 1991). However, another plausible explanation, at least in some cases, can be that the equilibria might behave as consecutive. If the TD and DM transitions are separated, but the separation is not very large, the resulting transition curve may still appear as that of the closed equilibrium due to the low sensitivity and noise of the experimental methods. However, the value of ΔV determined by the two methods is inconsistent. Consider an example (in detail discussed in

ESI, Section S-11) of a homotrimer equilibrium of the parameters ΔV_T, ΔV_D, $K_{T,atm}$, and $K_{D,atm}$. Depending on their values the equilibrium can behave either as closed or as consecutive with more or less separated TD and DM transitions. For determination of the apparent ΔV (i.e. for the putatively closed equilibrium) often the pressure difference between the points of $\alpha = 0.1$ and $\alpha = 0.9$ is used. These points can be, in a good approximation, identified with those of $\alpha_T = 0.9$ and $\alpha_M = 0.9$, respectively. Using Eq. (37) the pressure difference between these points can be determined. Considering, for simplicity of the example, that $\Delta V_T = \Delta V_D$, the difference is

$$\Delta p = p_{T90} - p_{M90} = \frac{RT}{\Delta V_D + \Delta V_T}\left(4\ln\frac{K_{T,atm}}{K_{D,atm}} + 18.72\right) \quad (45)$$

For the genuine closed equilibrium, where $\Delta V = \Delta V_T + \Delta V_D$, the corresponding difference determined by Eq. (14) is

$$\Delta p = p(\alpha = 0.1) - p(\alpha = 0.9) = 8.79\frac{RT}{\Delta V_T + \Delta V_D}. \quad (46)$$

As the dimer has to be rather stable in case of a consecutive equilibrium, it is probable that $K_{D,atm}$ is not considerably higher than $K_{T,atm}$ (ten or more times). In this case, however, the expression in Eq. (45) is higher than that in Eq. (46). Thus, when the experimental quantity corresponding with Eq. (45) is substituted to Eq. (46) and the apparent volume change ΔV_P, is determined, the resulting value is lower than in the case of authentic closed equilibrium. On the other hand, the other method based on the concentration shift of the transition curves provides rather different results. Here, the apparent inflection point of the curve can be identified with the maximum of α_D and determined with the aid of Eq. (44). If we consider, for simplicity, only the limit case of the consecutive equilibrium with well separated transitions, i.e. when the expression in the parenthesis of Eq. (44) reduces to 1, the maximum of α_D occurs at

$$p_{max} = \frac{RT}{\Delta V_T}\ln\left(\frac{K_{T,atm}}{3}\frac{2\Delta V_D}{2\Delta V_D - \Delta V_T}\right) - \frac{RT}{\Delta V_T}\ln M_0. \quad (47)$$

Comparing this expression with Eq. (14) describing the closed equilibrium, it is obvious that the apparent volume change $\Delta V_C/2$ is equal to ΔV_T. Thus, if $\Delta V_T \approx \Delta V_D$, then $\Delta V_C \approx \Delta V_D + \Delta V_T$, which is a result expected for the closed equilibrium. Hence, although some approximations were taken into account, it can be seen that the apparent inflection points move with the concentration changes in a similar way for both closed and consecutive equilibrium and the apparent volume change ΔV_C is not much dependent on the character of the equilibrium. On the contrary, the apparent volume change ΔV_β depends on it considerably, which might be the reason of the observed incoherence between the two quantities. Thus, although it is not possible to reevaluate the previously published data, they may indicate that the studied equilibria were in fact consecutive but the used experimental methods did not enable uncovering this feature.

It should be, however, noted that in some cases the inconsistence of ΔV_P and ΔV_C might have still another reason. The dissociation process can be irreversible or only partially reversible, especially due to the simultaneously running unfolding of the monomers. This phenomenon, which then governs the observed dissociation process, is of the first order and thus concentration independent. In case of partial reversibility the concentration dependence of the transition curve can be seen, but smaller than in the fully reversible case. This can be demonstrated on the published examples of higher oligomers in (Silva et al., 1989; Bonafe et al., 1994) – especially in the latter case the difference between the irreversible and partially reversible system is apparent. Therefore, the mutual agreement of ΔV_P and ΔV_C can be used as an indicator of reversibility of the dissociation-association process.

2.3.2. Heterotrimer A₂B

Many oligomeric proteins form a heterotrimer consisting of two

138

identical subunits A and a different subunit B. The most common mechanism of its association – dissociation equilibrium is the following:

$$T \rightleftharpoons D + B \tag{48}$$

$$D \rightleftharpoons 2A \tag{49}$$

Here, T denotes the trimer A_2B and D stands for the dimer A_2. The principal difference with respect to the homotrimer is that, using recombinant subunits A and B, the dimerization process (Eq. (49)) can be studied separately in accord with Eq. (14). Therefore, the dimerization parameters $K_{D,atm}$ and ΔV_D can be determined prior to studying trimer formation. In addition, the difference in the subunits allows us to use non-stoichiometric ratios of A and B which can further help the experimental procedures.

Heterotrimer equilibrium is described by the dissociation constants of dimer

$$K_D = \frac{A^2}{D}; \qquad K_V = \frac{BD}{T}. \tag{50}$$

In addition, the following balance equations have to be obeyed:

$$A_0 = A + 2D + 2T; \qquad B_0 = B + T, \tag{51}$$

where A_0 and B_0 are the total concentrations of the subunits A and B. As the subunit A is present in all three forms of the system, it is useful to define the fractions of monomer, dimer and trimer, respectively, as follows:

$$\alpha_M = \frac{A}{A_0}; \qquad \alpha_D = \frac{2D}{A_0}; \qquad \alpha_T = \frac{2T}{A_0} \tag{52}$$

Restricting our interest immediately on the stoichiometric case, in which

$$T_0 \equiv B_0 = \frac{A_0}{2}, \tag{53}$$

where T_0 denotes the total concentration of trimer, i.e. the concentration of trimer provided that all the subunits are present in the trimeric form, we arrive at the following equation

$$\sqrt{\frac{K_D \alpha_D}{4 T_0}} + \alpha_D + \frac{\alpha_D T_0}{K_T + \alpha_D T_0} = 1. \tag{54}$$

This equation can be rearranged to the form of an algebraic equation of the fourth degree in α_D. More detailed discussion is given in ESI, Section S-12, together with its analytical solution carried out in Wolfram Mathematica 9 (Appendix S-2).

It is again useful to discuss the individual special cases of Eq. (54). Let us start with the case presuming a highly stable dimer and less stable trimer which is characterized by the relation $K_T \gg K_D$. As in the homotrimer case, a region of prevailing dimer exists and the two transitions, TD and DM, can be analyzed separately. Eq. (54) can thus be simplified by the neglect of the first or the last term, respectively. Both these simplified equations depend on only one dissociation constant. The TD transition can be considered as a heterodimer equilibrium of subunits A_2 and B, while the DM transition as a homodimer equilibrium of two A subunits. These transitions are, therefore, described by Eqs. (25) and (14), respectively. (For the justification see ESI, Section S-12.) Hence, the two transitions can be considered separately and the parameters $K_{D,atm}$, ΔV_D, $K_{T,atm}$, and ΔV_T can then be determined straightforwardly.

If, on the other hand, the trimer is substantially more stable than the dimer, i.e. $K_T \ll K_D$, the system undergoes the closed equilibrium described by Eq. (32) with the constants K_{atm} and ΔV given by the expressions $K_{atm} = K_{T,atm} K_{D,atm}$ and $\Delta V = \Delta V_T + \Delta V_D$, analogously with the homotrimer case.

If the separation of the two transitions is not so striking and it does not allow us to use these approximations at any pressure, it is valuable when an experimental method sensitive exclusively to either the trimer

or the free subunit B is available. If this is the case, the fraction of free B denoted α_B obeys the fourth-degree equation

$$4T_0^2 \alpha_B^4 + 8K_T T_0 \alpha_B^3 + K_T \alpha_B^2 (K_T + 4K_D - T_0) - K_T \alpha_B (8K_T + K_D) + 4K_T^2 = 0. \tag{55}$$

However, Eq. (55) is only quadratic equation in T_0 and can be solved for T_0 as a function of α_B:

$$T_0 = \frac{K_T (1 - \alpha_B)}{\alpha_B^2} \left(1 + \sqrt{\frac{K_D}{4K_T} \frac{\alpha_B}{1 - \alpha_B}} \right). \tag{56}$$

Although the inflection point of $\ln T_0$ cannot be determined analytically, the numerical solution in Wolfram Mathematica 9 shows that the inflection point value of $\alpha_{B,inf}$ is in a very narrow interval (0.58; 0.64) for any values of K_D and K_T. For K_D/K_T ratio between 0.1 and 10, i.e. when none of the previously discussed special cases applies, the interval reduces to (0.60; 0.64). Thus, the inflection point value of T_0 can be obtained from Eq. (56) substituting $\alpha_{B,inf} \approx 0.62$ for α_B. Introducing the pressure dependence of K_T and K_D and exploiting the fact that $K_{D,atm}$ and ΔV_D can be determined independently, Eq. (56) becomes (in the logarithmic form)

$$\ln T_0 = \ln K_{T,atm} + \ln \frac{1 - \alpha_{B,inf}}{\alpha_{B,inf}^2} - \frac{p \Delta V_T}{RT}$$
$$+ \ln \left[1 + \sqrt{\frac{K_{D,atm}}{4K_{T,atm}} \frac{\alpha_B}{1 - \alpha_B} \exp\left(\frac{p}{RT} (\Delta V_T - \Delta V_D) \right)} \right]. \tag{57}$$

This equation can be solved iteratively in a similar manner as in the case of homotrimer (see ESI, Section S-13).

2.3.3. Heterotrimer ABC

Another kind of a heterotrimer is composed of three different subunits A, B and C. As the solution of this system is formally analogous to the A_2B heterotrimer, it is given in ESI, Section S-14.

3. Conclusions

A theoretical description of the equilibria among the subunits of oligomeric proteins under high pressure is provided in this work. The dependences of significant points of the transition curves – inflection points and extremes – on pressure and concentration have been derived, in most cases in the form of linear functions. Analysis of these functions allows us to determine the thermodynamic characteristics of the systems, i.e. the volume changes and atmospheric-pressure equilibrium constants of the individual transitions. The theory can be developed for the closed equilibrium of homooligomers of any number of subunits in a close analytical form. For these systems a rough estimate of the number of subunits is possible in case that it is not known a priori. In addition, an analogous description was developed also for the closed equilibria of a heterodimer and different forms of heterotrimers. Systems undergoing the consecutive equilibria represent far more complex problem. A homotrimer, which was chosen as the simplest example, was analyzed in detail. Here, it is difficult to develop a universal theory valid in the whole pressure and concentration ranges because the description is based on algebraic equations of higher than the second degree. Although these solutions were developed using Wolfram Mathematica 9 software and are shown in several figures, they are too complicated for the detailed analysis. Therefore, special cases that can be reached by a specific choice of pressure and concentration ranges were studied in which the solution is simplified and can be turned to linear functions. Finally, a heterotrimeric system undergoing consecutive equilibria was analyzed in a similar way.

Although we have recently verified the theory on the model of dimerization of HIV-1 protease (Ingr et al., 2015), more extensive experimental studies including also higher oligomers are needed to demonstrate its utility. In spite of possible experimental difficulties we

139

hope that this work may encourage further high-pressure experimental studies with oligomeric proteins.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jtbi.2016.10.001.

References

Bonafe, C.F., Araujo, J.R., Silva, J.L., 1994. Intermediate states of assembly in the dissociation of gastropod hemocyanin by hydrostatic pressure. Biochemistry 33, 2651–2660.

Bonafe, C.F.S., Vital, C.M.R., Telles, R.C.B., Gonçalves, M.C., Matsuura, M.S.A., Pessine, F.B.T., Freitas, D.R.C., Vega, J., 1998. Tobacco mosaic virus disassembly by high hydrostatic pressure in combination with urea and low temperature. Biochemistry 37, 11097–11105. http://dx.doi.org/10.1021/bi980349n.

Cioni, P., Gabellieri, E., Marchal, S., Lange, R., 2014. Temperature and pressure effects on C112S azurin: volume, expansivity, and flexibility changes. Proteins-Struct. Funct. Bioinform. 82, 1787–1798. http://dx.doi.org/10.1002/prot.24532.

Cioni, P., Strambini, G.B., 1996. Pressure effects on the structure of oligomeric proteins prior to subunit dissociation. J. Mol. Biol. 263, 789–799. http://dx.doi.org/10.1006/jmbi.1996.0616.

Collins, M.D., Kim, C.U., Gruner, S.M., 2011. High-pressure protein crystallography and NMR to explore protein conformations. In: Rees, D.C., Dill, K.A., Williamson, J.R. (Eds.), Annual Review of Biophysics 40. Annual Reviews, Palo Alto, 81–98.

Da Poian, A.T., Oliveira, A.C., Gaspar, L.P., Silva, J.L., Weber, G., 1993. Reversible pressure dissociation of R17 bacteriophage: the physical individuality of virus particles. J. Mol. Biol. 231, 999–1008. http://dx.doi.org/10.1006/jmbi.1993.1347.

Devillebonne, D., Else, A., 1991. Reversible high hydrostatic-pressure inactivation of phosphofructokinase from Escherichia-coli. Eur. J. Biochem 200, 747–750. http://dx.doi.org/10.1111/j.1432-1033.1991.tb16240.x.

Dumay, E.M., Kalichevsky, M.T., Cheftel, J.C., 1994. High-pressure unfolding and aggregation of beta -lactoglobulin and the baroprotective effects of sucrose. J. Agric. Food Chem. 42, 1861–1868. http://dx.doi.org/10.1021/jf00045a006.

Erijman, L., Lorimer, G., Weber, G., 1993. Reversible dissociation and conformational stability of dimeric ribulose bisphosphate carboxylase. Biochemistry 32, 5187–5195. http://dx.doi.org/10.1021/bi00070a030.

Erijman, L., Weber, G., 1991. Oligomeric protein associations - transition from stochastic to deterministic equilibrium. Biochemistry 30, 1595–1599. http://dx.doi.org/10.1021/bi00220a022.

Foguel, D., Weber, G., 1995. Pressure-induced dissociation and denaturation of allophycocyanin at subzero temperatures. J. Biol. Chem. 270, 28759–28766. http://dx.doi.org/10.1074/jbc.270.48.28759.

Gebhardt, R., Doster, W., Friedrich, J., Kulozik, U., 2006. Size distribution of pressure-decomposed casein micelles studied by dynamic light scattering and AFM. Eur. Biophys. J. Biophys. Lett. 35, 503–509. http://dx.doi.org/10.1007/s00249-006-0058-6.

Gebhardt, R., Doster, W., Friedrich, J., Petry, W., Schulte, A., 2003. Pressure-Induced Critical Association of Myoglobin. Springer-Verlag, Berlin, Berlin.

Gebhardt, R., Doster, W., Kulozik, U., 2005. Pressure-induced dissociation of casein micelles: size distribution and effect of temperature. Braz. J. Med. Biol. Res. 38, 1209–1214. http://dx.doi.org/10.1590/S0100-879X2005000800008.

Gebhardt, R., Takeda, N., Kulozik, U., Doster, W., 2011. Structure and stabilizing interactions of casein micelles probed by high-pressure light scattering and FTIR. J. Phys. Chem. B 115, 2349–2359. http://dx.doi.org/10.1021/jp107622d.

Girard, E., Marchal, S., Perez, J., Finet, S., Kahn, R., Fourme, R., Marassio, G., Dhaussy, A.-C., Prange, T., Giffard, M., Dulin, F., Bonnete, F., Lange, R., Abraini, J.H., Mezouar, M., Colloc'h, N., 2010. Structure-function perturbation and dissociation of tetrameric urate oxidase by high hydrostatic pressure. Biophys. J. 98, 2365–2373. http://dx.doi.org/10.1016/j.bpj.2010.01.05.

Gross, M., Jaenicke, R., 1994. Proteins under pressure. The influence of high hydrostatic pressure on structure, function and assembly of proteins and protein complexes. Eur. J. Biochem. FEBS 221, 617–630.

Ingr, M., Lange, R., Halabalova, V., Yehya, A., Hrncirik, J., Chevalier-Lucia, D., Palmade, L., Blayo, C., Konvalinka, J., Dumay, E., 2015. Inhibitor and substrate binding induced stability of HIV-1 protease against sequential dissociation and unfolding revealed by high pressure spectroscopy and kinetics. PLoS One 10, e0119099. http://dx.doi.org/10.1371/journal.pone.0119099.

Jacrot, B., 1976. Study of biological structures by neutron-scattering from solution. Rep. Prog. Phys. 39, 911–953. http://dx.doi.org/10.1088/0034-4885/39/10/001.

Jaenicke, R., Koberstein, R., 1971. High pressure dissociation of lactic dehydrogenase. FEBS Lett. 17, 351–354. http://dx.doi.org/10.1016/0014-5793(71)80185-0.

King, L., Weber, G., 1986. Conformational drift of dissociated lactate dehydrogenases. Biochemistry 25, 3632–3637.

Kobori, H., Sato, M., Tameike, A., Hamada, K., Shimada, S., Osumi, M., 1996. Changes in microfilaments and microtubules of yeasts induced by pressure stress. In: Hayashi, R., Balny, C. (Eds.), Progress in Biotechnology 13, High Pressure Bioscience and Biotechnology, Proceedings of the International Conference on High Pressure Bioscience and Biotechnology. Elsevier, Kyoto, 83–94. http://dx.doi.org/10.1016/S0921-0423(06)80015-8.

Kornblatt, M.J., Lange, R., Balny, C., 2004. Use of hydrostatic pressure to produce "native" monomers of yeast enolase. Eur. J. Biochem. 271, 3897–3904. http://dx.doi.org/10.1111/j.1432-1033.2004.04326.x.

Kutalkova, E., Hrncirik, J., Ingr, M., 2014. Pressure induced structural changes and dimer destabilization of HIV-1 protease studied by molecular dynamics simulations. Phys. Chem. Chem. Phys. 16, 25906–25915. http://dx.doi.org/10.1039/c4cp03676j.

Lange, R., Frank, J., Saldana, J.-L., Balny, C., 1996. Fourth derivative UV-spectroscopy of proteins under high pressure I. Factors affecting the fourth derivative spectrum of the aromatic amino acids. Eur. Biophys. J. 24, 277–283. http://dx.doi.org/10.1007/BF00180368.

Marchal, S., Marabotti, A., Staiano, M., Varriale, A., Domaschke, T., Lange, R., D'Auria, S., 2012. Under pressure that splits a family in two. The case of lipocalin family. PLoS One 7, e50489. http://dx.doi.org/10.1371/journal.pone.0050489.

Marchal, S., Torrent, J., Masson, P., Kornblatt, J.M., Tortora, P., Fusi, P., Lange, R., Balny, C., 2005. The powerful high pressure tool for protein conformational studies. Braz. J. Med. Biol. Res. Rev. Bras. Pesqui. Médicas E Biológicas Soc. Bras. Biofisica Al 38, 1175–1183, [doi:/S0100-879X2005000800004].

Meier, G., Kriegs, H., 2008. A high pressure cell for dynamic light scattering up to 2 kbars with conservation of plane of polarization. Rev. Sci. Instrum. 79, 13102. http://dx.doi.org/10.1063/1.2827137.

Messier, P., Seguin, C., 1978. Effects of high hydrostatic-pressure on microfilaments and microtubules. J. Embryol. Exp. Morphol. 44, 281–295.

Mozhaev, V.V., Heremans, K., Frank, J., Masson, P., Balny, C., 1996. High pressure effects on protein structure and function. Proteins Struct. Funct. Bioinform. 24, 81–91. http://dx.doi.org/10.1002/(SICI)1097-0134(199601)24:1 < 81 ::AID-PROT6 > 3.0.CO;2-R.

Nishiyama, M., Kimura, Y., Nishiyama, Y., Terazima, M., 2006. Development of a high-pressure microscope and its application to biological systems, in: Proceedings of the 2006 International Symposium on Micro-NanoMechatronics and Human Science. pp. 1–5. http://dx.doi.org/10.1109/MHS.2006.320278.

Nishiyama, M., Shimoda, Y., Hasumi, M., Kimura, Y., Terazima, M., 2010. Microtubule depolymerization at high pressure. Ann. N. Y. Acad. Sci. 1189, 86–90. http://dx.doi.org/10.1111/j.1749-6632.2009.05411.x.

Paladini, A., Weber, G., Erijman, L., 1994. Analysis of dissociation and unfolding of oligomeric proteins using a flat bed gel-electrophoresis at high-pressure. Anal. Biochem. 218, 364–369. http://dx.doi.org/10.1006/abio.1994.1193.

Paladini, A.A., Silva, J.L., Weber, G., 1987. Slab gel electrophoresis of oligomeric proteins under high hydrostatic pressure. Anal. Biochem. 161, 358–364. http://dx.doi.org/10.1016/0003-2697(87)90463-5.

Paladini, A.A., Weber, G., 1981. Pressure-induced reversible dissociation of enolase. Biochemistry 20, 2587–2593. http://dx.doi.org/10.1021/bi00512a034.

Pin, S., Royer, C., Gratton, E., Alpert, B., Weber, G., 1990. Subunit interactions in hemoglobin probed by fluorescence and high-pressure techniques. Biochemistry 29, 9194–9202. http://dx.doi.org/10.1021/bi00491a013.

Rivalain, N., Roquain, J., Demazeau, G., 2010. Development of high hydrostatic pressure in biosciences: pressure effect on biological structures and potential applications in biotechnologies. Biotechnol. Adv. 28, 659–672. http://dx.doi.org/10.1016/j.biotechadv.2010.04.001.

Rouget, J.-B., Aksel, T., Roche, J., Saldana, J.-L., Garcia, A.E., Barrick, D., Royer, C.A., 2011. Size and sequence and the volume change of protein folding. J. Am. Chem. Soc. 133, 6020–6027. http://dx.doi.org/10.1021/ja200228w.

Rouget, J.-B., Schroer, M.A., Jeworrek, C., Pühse, M., Saldana, J.-L., Bessin, Y., Tolan, M., Barrick, D., Winter, R., Royer, C.A., 2010. Unique features of the folding landscape of a repeat protein revealed by pressure perturbation. Biophys. J. 98, 2712–2721. http://dx.doi.org/10.1016/j.bpj.2010.02.044.

Royer, C.A., 2002. Revisiting volume changes in pressure-induced protein unfolding. Biochim. Biophys. Acta 1595, 201–209.

Royer, C.A., 1995. Application of pressure to biochemical equilibria: the other thermodynamic variable. Methods Enzym. 259, 357–377.

Royer, C.A., Weber, G., Daly, T.J., Matthews, K.S., 1986. Dissociation of the lactose repressor protein tetramer using high hydrostatic pressure. Biochemistry 25, 8308–8315.

Ruan, K., Weber, G., 1993. Physical heterogeneity of muscle glycogen phosphorylase revealed by hydrostatic pressure dissociation. Biochemistry 32, 6295–6301.

Ruan, K., Weber, G., 1989. Hysteresis and conformational drift of pressure-dissociated glyceraldehydephosphate dehydrogenase. Biochemistry 28, 2144–2153.

Ruan, K., Weber, G., 1988. Dissociation of yeast hexokinase by hydrostatic-pressure. Biochemistry 27, 3295–3301. http://dx.doi.org/10.1021/bi00409a026.

Ruan, K., Xu, C., Yu, Y., Li, J., Lange, R., Bec, N., Balny, C., 2001. Pressure-exploration of the 33-kDa protein from the spinach photosystem II particle. Eur. J. Biochem. FEBS 268, 2742–2750.

Seefeldt, M.B., Kim, Y.-S., Tolley, K.P., Seely, J., Carpenter, J.F., Randolph, T.W., 2005. High-pressure studies of aggregation of recombinant human interleukin-1 receptor antagonist: thermodynamics, kinetics, and application to accelerated formulation studies. Protein Sci. 14, 2258–2266. http://dx.doi.org/10.1110/ps.051490205.

Shrestha, U.R., Bhowmik, D., Copley, J.R.D., Tyagi, M., Leão, J.B., Chu, X.-Q., 2015. Effects of pressure on the dynamics of an oligomeric protein from deep-sea hyperthermophile. Proc. Natl. Acad. Sci. USA 112, 13886–13891. http://dx.doi.org/10.1073/pnas.1514478112.

Silva, J., Luan, P., Glaser, M., Voss, E., Weber, G., 1992. Effects of hydrostatic-pressure

on a membrane-enveloped virus - high immunogenicity of the pressure-inactivated virus. J. Virol. 66, 2111–2117.

Silva, J., Miles, E., Weber, G., 1986. Pressure dissociation and conformational drift of the beta-dimer of tryptophan synthase. Biochemistry 25, 5780–5786. http://dx.doi.org/10.1021/bi00367a065.

Silva, J., Villas-Boas, M., Bonafe, C., Meirelles, N., 1989. Anomalous pressure dissociation of large protein aggregates - lack of concentration-dependence and irreversibility at extreme degrees of dissociation of extracellular hemoglobin. J. Biol. Chem. 264, 15863–15868.

Silva, J.L., Foguel, D., DaPoian, A.T., Prevelige, P.E., 1996. The use of hydrostatic pressure as a tool to study viruses and other macromolecular assemblages. Curr. Opin. Struct. Biol. 6, 166–175. http://dx.doi.org/10.1016/S0959-440X(96)80071-6.

Silva, J.L., Foguel, D., Royer, C.A., 2001. Pressure provides new insights into protein folding, dynamics and structure. Trends Biochem. Sci. 26, 612–618.

Silva, J.L., Oliveira, A.C., Vieira, T.C.R.G., de Oliveira, G.A.P., Suarez, M.C., Foguel, D., 2014. High-pressure chemical biology and biotechnology. Chem. Rev. 114, 7239–7267. http://dx.doi.org/10.1021/cr400204z.

Silva, J.L., Weber, G., 1988. Pressure-induced dissociation of brome mosaic virus. J. Mol. Biol. 199, 149–159. http://dx.doi.org/10.1016/0022-2836(88)90385-3.

Sun, Y., Wallrabe, H., Booker, C.F., Day, R.N., Periasamy, A., 2010. Three-color spectral FRET microscopy localizes three interacting proteins in living cells. Biophys. J. 99, 1274–1283. http://dx.doi.org/10.1016/j.bpj.2010.06.004.

Svergun, D.I., Nierhaus, K.H., 2000. A map of protein-rRNA distribution in the 70 SEscherichia coli Ribosome. J. Biol. Chem. 275, 14432–14439. http://dx.doi.org/10.1074/jbc.275.19.14432.

Torrent, J., Lange, R., Rezaei, H., 2015. The volumetric diversity of misfolded prion protein oligomers revealed by pressure dissociation. J. Biol. Chem. 290, 20417–20426. http://dx.doi.org/10.1074/jbc.M115.661710.

Weber, G., 1993. Thermodynamics of the association and the pressure dissociation of oligomeric proteins. J. Phys. Chem. 97, 7108–7115. http://dx.doi.org/10.1021/j100129a031.

Weber, G., 1986. Phenomenological description of the association of protein subunits subjected to conformational drift. Effects of dilution and of hydrostatic pressure. Biochemistry 25, 3626–3631.

Weber, G., DaPoian, A.T., Silva, J.L., 1996. Concentration dependence of the subunit association of oligomers and viruses and the modification of the latter by urea binding. Biophys. J. 70, 167–173.

141

P7. Kinetics of the dimerization of retroviral proteases: the "fireman's grip" and dimerization

M. Ingr, T. Uhlíková, K. Strísovský, E. Majerová, J. Konvalinka, Protein Sci. 12 (2003) 2173–2182.

Kinetics of the dimerization of retroviral proteases: The "fireman's grip" and dimerization

MAREK INGR,[1,2] TAT'ÁNA UHLÍKOVÁ,[1] KVIDO STŘÍŠOVSKÝ,[1,2] EVA MAJEROVÁ,[3]
AND JAN KONVALINKA[1,2]

[1]Institute of Organic Chemistry and Biochemistry, Academy of Science of the Czech Republic, 166 10 Praha 6,
Czech Republic
[2]Department of Biochemistry, School of Science, Charles University, 128 43 Praha 2, Czech Republic
[3]Biological Testing Branch, Science Applications International Corp. (SAIC), National Cancer Institute-Frederick
Cancer Research Development Center (NCI-FCRDC), Frederick, Maryland 21784, USA

Abstract

All retroviral proteases belong to the family of aspartic proteases. They are active as homodimers, each unit contributing one catalytic aspartate to the active site dyad. An important feature of all aspartic proteases is a conserved complex scaffold of hydrogen bonds supporting the active site, called the "fireman's grip," which involves the hydroxyl groups of two threonine (serine) residues in the active site Asp-Thr(Ser)-Gly triplets. It was shown previously that the fireman's grip is indispensable for the dimer stability of HIV protease. The retroviral proteases harboring Ser in their active site triplet are less active and, under natural conditions, are expressed in higher enzyme/substrate ratio than those having Asp-Thr-Gly triplet. To analyze whether this observation can be attributed to the different influence of Thr or Ser on dimerization, we prepared two pairs of the wild-type and mutant proteases from HIV and myeloblastosis-associated virus harboring either Ser or Thr in their Asp-Thr(Ser)-Gly triplet. The equilibrium dimerization constants differed by an order of magnitude within the relevant pairs. The proteases with Thr in their active site triplets were found to be ~10 times more thermodynamically stable. The dimer association contributes to this difference more than does the dissociation. We propose that the fireman's grip might be important in the initial phases of dimer formation to help properly orientate the two subunits of a retroviral protease. The methyl group of threonine might contribute significantly to fixing such an intermediate conformation.

Keywords: Retroviral protease; dimerization; HIV protease; fireman's grip; kinetic assay; fluorescence; pressure

Retroviral proteases (PRs) belong to the family of retropepsins (Dunn 1998), which are dimeric aspartic proteases. The

two monomers are interconnected by their C and N termini, which form an antiparallel β-sheet that is primarily responsible for the dimer stability (Wlodawer et al. 1989). The proteases are expressed as parts of the viral polyproteins Gag or Gag-Pol, from which they cleave themselves before cleaving the rest of the polyproteins to yield the structural proteins and replication enzymes. The Gag-Pol polyprotein is a product of ribosomal frame-shifting or readthrough suppression, which takes place with 5% to 10% frequency. In contrast, in the case of alpharetroviruses (e.g., myeloblastosis-associated virus [MAV]), the protease is a part of the Gag polyprotein and is thus expressed in equimolar ratio with its Gag substrate. In the case of betaretroviruses (Mason-Pfizer monkey virus [MPMV]) the protease also is expressed by frame-shifting event. It was observed previously that the activities of the alpharetroviral proteases are con-

siderably lower than those of lentiviral proteases, which was interpreted to be a compensation for the difference in their relative levels of synthesis and thus concentration in the virion (Arad et al. 1995).

Because the PRs are active in the form of noncovalently bound homodimers, chemical compounds preventing dimerization became possible antiviral agents. Moreover, the dimerization seems to be the key event in the activation of the polyprotein processing in retroviruses (Kräusslich 1991). Hence, many analyses of dimerization equilibrium and rate constants have been performed, especially for HIV-1 PR. The experimental approaches can be divided into two groups: substrate-independent methods, which involve techniques such as ultracentrifugation (Holzman et al. 1991; Grant et al. 1992; Towler et al. 1998; Xie et al. 1999; Stříšovský et al. 2000), and substrate (inhibitor)-dependent methods, based on kinetic measurements, most often fluorometric ones (Cheng et al. 1990; Zhang et al. 1991; Jordan et al. 1992; Kuzmič 1993; Darke et al. 1994; Pargellis et al. 1994; Uhlíková et al. 1996).

The method of sedimentation equilibrium is attractive due to its independence of substrate. The result is therefore less influenced by stabilization of the dimeric form by substrate binding. On the other hand, the experiments require relatively high concentrations of pure protein. The K_d values for HIV-1 and HIV-2 proteases determined by this method are summarized in Table 1. The majority of these values are in the micromolar range. These results, however, seem to be inconsistent with the fact that most activity assays routinely use nanomolar concentrations of HIV-PRs for activity determinations.

Although the kinetic measurements are more sensitive, require a lower concentration of protein, and are easier to perform, their interpretation can be obscured by phenomena such as the stabilization of the dimer by a substrate or an inhibitor. This effect is difficult to separate from genuine dimer stability, and therefore, the result may depend on the ligand used. On the other hand, the advantage of kinetic studies is that both of the rate constants, k_1 and k_2, describing the association and dissociation processes can be determined.

As shown in Table 1, results of both K_d and rate constants determined kinetically with fluorogenic substrate or inhibitor have been reported by several groups for both HIV-1 and HIV-2 proteases (Cheng et al. 1990; Zhang et al. 1991; Jordan et al. 1992; Kuzmič 1993; Darke et al. 1994; Pargellis et al. 1994; Uhlíková et al. 1996. For summary, see Darke 1994). The reported K_d values from kinetic studies are typically of the order of 10^{-7} to 10^{-9} M, that is, on average, three orders of magnitude lower than the values derived from sedimentation analysis. The largest difference between two HIV-1 K_d values is five orders of magnitude. This major disagreement may be partially explained by the incorrect assumption of very fast association and dissociation processes in the publication by Jordan et al. (1992), in which the lowest values were determined. Another reason for the observed differences may be the inconsistency of experimental conditions in different experiments, especially pH varying from 4.5 to 7.0. However, there still remain discrepancies in the reported K_d values that are difficult to explain.

There are two major structural features believed to be responsible for the dimer stability of PRs: the N and C

Table 1. *Values of K_d from previous studies*

Reference	Protease	pH	K_d	$\tau_{1/2}$ (sec)	Remark
Holzman et al. 1991	HIV-2	4.5	87 μM		
		7.5	28 μM		
Grant et al. 1992	HIV-1	5.0	<10 nM		
Towler et al. 1998	HIV-1	7.0	9 μM		
	HERV-K10	7.0	52.5 μM		
Xie et al. 1999	HIV-1	7.0	52.7 μM		Autolysis products interferred, K_d taken as upper limit
Stříšovský et al. 2000	HIV-1	6.0	77 μM		
Cheng et al. 1990	HIV-1	7.0	50 nM	198	
Zhang et al. 1991	HIV-1		3.6 nM		Very fast processes assumed
Jordan et al. 1992	HIV-1		<1 nM		
	HIV-2		<1 nM		
Darke et al. 1994	HIV-1	6.5	16 nM		Fluorogenic inhibitor used
		7.0	39 nM		
Pargellis et al. 1994	HIV-1	5.5	5.2 nM	533	0 M NaCl
			4.0 nM	745	1.5 M NaCl
Kuzmič et al. 1993	HIV-1	6.4	440 nM	23.9	
Uhlíková et al. 1996	HIV-1	6.4	112 nM		Integral kinetics
		4.5	19 nM		

The top part of the table shows the results of sedimentation studies; the bottom part the results of kinetic assays. $\tau_{1/2}$ is the halftime of dimer dissociation.

termini intertwined in an antiparallel β-sheet, and a hydrogen bond scaffold supporting the active site called the fireman's grip. It is formed by the Asp-Thr(Ser)-Gly triplet in which the oxygen of side chain of the active site Thr (Thr26 in the case of HIV-1 PR) forms a hydrogen bond with the main-chain amide of the active site Thr (Thr26′) of the other molecule in the dimer. Similar interactions take place between the side chain of this Thr (Thr26) and the main-chain carbonyl of the preceding Leu (Leu24′ in the case of HIV-1 PR) of the other molecule in the dimer. Both hydrogen bonds are dependent on the side-chain hydroxyls of the Thr residues. An X-ray crystallographic model of the three-dimensional structure of HIV-1 PR and a detailed view of the active site region depicting the described hydrogen bond network are shown in Figure 1. This complex structure maintains the productive conformation of the active site (Pearl and Blundell 1984; Pearl and Taylor 1987).

Previous studies showed that the Thr26Ser mutation of HIV-1-PR decreases the enzyme activity fivefold to 10-fold (Konvalinka et al. 1995; Rose et al. 1995). The opposite phenomenon was observed for the Ser38Thr mutation of MAV and Rous sarcoma virus (RSV) proteases (Arad et al. 1995; Ridky et al. 1996). A recent study (Stříšovský et al. 2000) provides evidence that the fireman's grip contributes to dimerization, which could thus be responsible for these differences in activity. Stříšovský et al. (2000) showed, by using sedimentation analysis and activity assays, that mutation of Thr26 of wild-type HIV-PR to Ala and, in lesser extent, also to Ser and Cys decreases the dimer stability considerably. This finding makes it important to address whether this is a general phenomenon, that is, whether the

Figure 1. (*A*) Structure of the dimer of HIV-1 protease. The individual subunits are depicted by different colors. Dimerization domain (D), flaps (F), and active site (A) are indicated. (*B*) Detail of the active site of HIV-1 protease with fireman's grip. Hydrogen bridges are denoted by dashed lines. Source is Protein Data Bank entry 3 PHV.

presence of either Thr or Ser in the active site triplet Asp-Thr(Ser)-Gly indeed regulates the dimer stability of PRs. Therefore, we undertook a comparative analysis of the equilibrium and rate constants of dimerization of three prototype PRs, from HIV, MAV, and MPMV. We examined the hypothesis that the presence of either Thr or Ser in the active site triplet of PRs plays a regulatory role in the PR dimerization and therefore in its activation. We also asked whether Ser-Thr changes affect the K_d of the dimer by affecting the association or the dissociation rate constants. Answering this question may give clues to the folding pathway that leads to the active PR dimer and thus triggers polyprotein processing in retroviruses.

Results

Design, fluorometric, and kinetic characterization of fluorogenic substrates

New fluorogenic substrates were designed for the purpose of the present study. Our aim was to prepare substrates with a good fluorometric response and good kinetic properties. The design was based on the previously known peptide substrates with convenient values of K_m and k_{cat} (Konvalinka et al. 1991, 1992). Donor and acceptor of energy for nonradiative energy transfer were attached to the termini of the substrates, 5-[(2′aminoethyl)-amino]naphtalenesulfonic acid (EDANS) as a donor and 4-{[4′-(dimethylamino) phenyl]azo}-benzoic acid (DABCYL) as an acceptor. Emission and excitation spectra were recorded for both the uncleaved and cleaved substrates (data not shown). The fluorescence of the cleaved form is 11 times higher than that of the noncleaved form. The wavelengths of maxima for excitation and emission, 347 and 500 nm, respectively, roughly correspond with those tabulated for EDANS. The substrates were also characterized kinetically. Determined values of Michaelis constant K_m and the turnover number k_{cat} are presented in Table 2. Both substrates are suitable for wild-type MAV-PR (MAV-PR[S]). For the other enzymes, however, only peptide FS1 can be used because the cleavage of FS2 is too slow or does not occur at all.

Equilibrium constant of the monomer–dimer process

The equilibrium between monomeric and dimeric forms of a PR is described by the chemical equation

$$2M \underset{k_2}{\overset{k_1}{\rightleftarrows}} D \tag{1}$$

The equilibrium concentration of dimer in solution is given by the equation

$$[D] = \frac{1}{8}(K_d + 4E - \sqrt{K_d(K_d + 8E)}), \tag{2}$$

Table 2. *Values of K_m and k_{cat} for the fluorogenic substrates FS1 (Glu[EDANS]ThrHisGlnValTry↓PheValArgLysAlaLys [DABCYL]-NH$_2$) and FS2 (Glu[EDANS]ThrProGlnValTyr↓ PheValArgLysAlaLys [DABCYL]-NH$_2$) and their chromogenic analogues CS1 (AlaThrHisGlnValTyr(p-nitro-Phe)ValArgLysAla) and CS2 (AlaThrProGlnValTyr(p-nitro-Phe)ValArgLysAla)*

Enzyme	Substrate	K_m (μM)	k_{cat} (sec^{-1})
MAV-PR(S)	FS1	2.2 ± 0.6	0.7 ± 0.1
MAV-PR(S)	FS2	3.5 ± 0.5	8.6 ± 1.4
MAV-PR(S)	CS1	75[a]	5[a]
MAV-PR(S)	CS2	9[a]	5[a]
HIV-1-PR(T)	FS1	3.0 ± 0.4	13.3 ± 1.4
HIV-1-PR(T)	CS1	15[b]	14[b]
HIV-1-PR(S)	FS1	4.2 ± 1.2	4.8 ± 0.7
MPMV-PR(T)	FS1	0.9 ± 0.1	1.9 ± 0.2
MPMV-PR(T)	CS1	60[c]	0.17[c]

For experimental details see Materials and Methods. Several values were adopted from [a]Konvalinka et al. (1991); [b]Konvalinka et al. (1990); and [c]Zábranský (1999).

where K_d is the equilibrium constant, [D] the concentration of dimer, and E the overall (analytical) concentration of the enzyme considered as monomer. We introduce the constants α and β (Kuzmič 1993):

$$\alpha = \sqrt{K_d(K_d + 8E)} \qquad (3a)$$

$$\beta = K_d + 4E. \qquad (3b)$$

Equation 2 now has the form

$$[D] = \frac{1}{8}(\beta - \alpha). \qquad (4)$$

Thus, the equilibrium constant has to be calculated from the ratio of the initial and equilibrium dimer concentration with the aid of equation 4. Let us denote Q_{eq} as the ratio [D$_{eq}$]/[D$_0$], where [D$_{eq}$] is the equilibrium dimer concentration and [D$_0$] is the initial dimer concentration, that is, the equilibrium concentration of the stock solution multiplied by the dilution factor. Q_{eq} fulfills the following equation:

$$Q_{eq} = \frac{E_0}{E} \frac{K_d + 4E - \sqrt{K_d(K_d + 8E)}}{K_d + 4E_0 - \sqrt{K_d(K_d + 8E_0)}} \qquad (5)$$

After some rearrangements, the quadratic equation for K_d is obtained, which has two roots, one of them physically meaningful:

$$K_d = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}, \qquad (6)$$

$$A = Q_{eq}(Q_{eq}[E + E_0]^2 - EE_0[Q_{eq} + 1]^2), \qquad (7a)$$

$$B = -4EE_0(E + E_0)Q_{eq}(Q_{eq} - 1)^2, \qquad (7b)$$

$$C = 4E^2E_0^2(Q_{eq} - 1)^4. \qquad (7c)$$

Here E is the cuvette concentration, and E_0 is the concentration of the stock solution.

Equation 6 was used for the practical determination of the dimerization equilibrium constant K_d. The parameter Q_{eq} was calculated from the initial (t$_{preincubation}$ = 0) and limit (t$_{preincubation}$ → ∞) values of reaction rates. Final enzyme concentrations after the dilution were chosen to cover the significant range properly, that is, to reach different values of the degree of dimer dissociation from low to high. Substrate concentration was kept as low as possible, generally below the value of K_m, to prevent possible substrate stabilization of the dimer.

A typical dependence of the initial rate on the preincubation time is shown in Figure 2, together with the inferred equilibrium rate. The value of K_d was calculated for every series of measurements, and the final result was then obtained as an arithmetic average of these values. These data, together with number of determinations and the initial (stock) and final concentrations, are presented in Table 3. A schematic plot of these values for all the studied enzymes is given in Figure 3. The wild-type forms of HIV-PR and MPMV-PR prefer the dimer form more than the wild type of MAV-PR, with the difference being approximately an order of magnitude. The two enzymes that can be compared to their fireman's grip mutants indicate that the dimers of T-variants of PRs are about an order of magnitude more stable than those of the S-variants. Wild-type MPMV-PR (MPMV-PR[T]) is not compared with its mutant, but the value of K_d for this enzyme is close to that of the other T-variants.

Rate constant of dimer association and dissociation

The kinetics of the dimerization–monomerization process is described by the following differential equation:
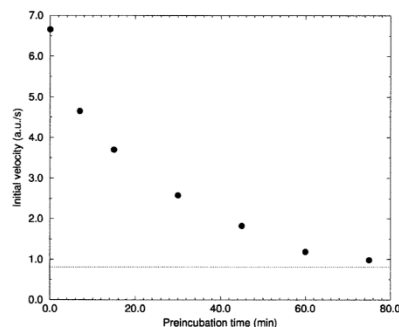
Figure 2. Initial reaction rate as a function of a preincubation time (circles). Dashed line indicates the equilibrium rate.

145

Table 3. *Values of equilibrium constants of dimerization (K_d) and kinetic parameters (k_2) of dimer dissociation and association*

Enzyme	Number of experiments	Stock enzyme concentration (μM)	Assay enzyme concentration (nM)	K_d (nM)	K_2 (10^{-4} sec^{-1})	$\tau_{1/2}$ (min)	k_1 (10^3 M^{-1} sec^{-1})
MAV-PR(S)	6	18.4	12–62	412 ± 122	7.47 ± 1.07	15.5 ± 2.2	1.81 ± 0.79
MAV-PR(*T*)	5	13–27	21–51	31.5 ± 3.3	7.81 ± 1.49	14.8 ± 2.8	24.8 ± 18.3
HIV-1-PR(T)	7	3–21	1–7	15.1 ± 4.3	6.78 ± 1.97	17.0 ± 5.0	44.9 ± 25.8
HIV-1-PR(*S*)	4	0.75–10	2–10	122 ± 28	12.1 ± 1.09	9.5 ± 0.8	10.0 ± 3.19
MPMV-PR(T)	4	2.78	4–19	13.5 ± 8.3	28.1 ± 10.8	4.1 ± 1.6	208.1 ± 207.9

Specific experimental conditions for each enzyme are indicated.

$$\frac{d[D]}{dt} = k_1[M]^2 - k_2[D] \qquad (8)$$

Using $[M] = E - 2[D]$ ($[M]$ is the monomer concentration), this equation can be solved yielding the time dependence of dimer concentration, $[D]$

$$[D] = \frac{1}{8}\frac{(\beta - \alpha)\gamma - (\beta + \alpha)e^{-k_1\alpha t}}{\gamma - e^{-k_1\alpha t}}. \qquad (9)$$

The symbol γ is defined as

$$\gamma = \frac{E(\beta_0 - \alpha_0) - E_0(\beta + \alpha)}{E(\beta_0 - \alpha_0) - E_0(\beta - \alpha)} \qquad (10)$$

and α_0 and β_0 are defined analogously to α and β in equation 3, a and b, where E is substituted by E_0.

To determine the kinetic constants k_1 and k_2, we make use of the approximate, but generally accepted, relation among these constants and K_d:

$$K_d = \frac{k_2}{k_1} \qquad (11)$$

By rearranging equation 9 with the aid of equation 11, we obtain the equation

$$W \equiv \frac{K_d}{\alpha}\ln\left(\frac{EQ(\beta_0 - \alpha_0) - E_0(\beta + \alpha)}{EQ(\beta_0 - \alpha_0) - E_0(\beta - \alpha)}\frac{1}{\gamma}\right) = k_2t, \qquad (12)$$

where Q, the ratio of current and initial dimer concentrations, is a generalization of the experimental quantity Q_{eq} from equation 5 and is determined as a ratio of the reaction rate at a certain time and at the initial time. The experimental quantity W depends linearly on time through the proportionality constant k_2. Thus, this constant can be determined by linear regression. By using equation 11, we obtain the other rate constant k_1.

The determination of the rate constant k_2 makes use of equation 12. In contrast to K_d, the values of initial reaction rate for all preincubation times are necessary. For every experimental point, the quantity W is calculated and linear regression of its dependence on time is performed. The constant k_2 is obtained as a slope of this dependence. The calculated values of the rate constant k_2 are shown in Table 3. In addition, values of the rate constant of dissociation k_1 are presented there. The last quantity presented in this table is the halftime of dimer dissociation provided that the dissociation is an independent process and no association takes place. Thus, this quantity is calculated in the following way:

$$\tau_{1/2} = \frac{\log 2}{k_2}. \qquad (13)$$

It is, in fact, only a different expression of k_2 with a more transparent meaning.

The final averaged values of kinetic parameters are presented in Table 3. It can be seen that MPMV-PR(T) is the most quickly dissociating enzyme, whereas the other two wild-type enzymes dissociate more slowly, MAV-PR(S) being slightly faster than HIV-PR(T). Comparison of wild-type enzymes and fireman's grip mutants of HIV-PR and MAV-PR shows that the dissociation rate constants are less affected by the mutations than the are K_d values. Thus, although HIV-PR*(S)* dissociates approximately twice as quickly as HIV-PR(T), there is no difference between the rate of dissociation of MAV-PR(S) and MAV-PR harboring Ser38Thr mutation (MAV-PR[*T*]). In contrast, the association rate constant k_1, calculated from the values of K_d and k_2
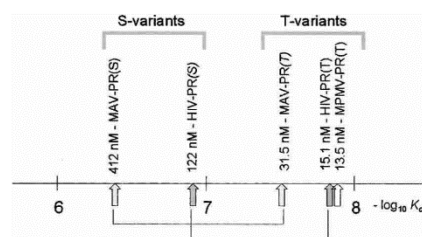
Figure 3. Values of K_d for the five studied enzymes: logarithmic plot. Each pair of the wild-type enzyme and corresponding mutant is connected for clarity.

146

(see Table 3), is strongly affected by the mutations. In summary, despite the higher experimental error of k_1 determination, the results show that the Thr→Ser mutation in fireman's grip leads to the decrease of the rate of association rather than to the increase of the rate of dissociation.

Discussion

In this article, we present a study of thermodynamic and kinetic characteristics of dimerization of three PRs and their mutant forms. These quantities have not been previously reported for these enzymes, with the exception of the wild-type HIV-1-PR. Even for this protease, which is one of the most thoroughly studied enzymes, the values provided by different experimenters and different methods are in a considerable disagreement with each other. The problems with K_d determination arise especially from the very low value of this constant, because the protein concentrations in enzymatic assays cannot differ much from K_d. The reason is that the total enzyme concentration is divided equally between monomer and dimer when it is equal to K_d (cf. equation 2). Moreover, the concentration range in which neither of the two forms prevails is rather limited, approximately one order of magnitude on each side. For example, when the enzyme concentration $E = 10K_d$, 80% of the protein is present as dimer, whereas for $E = 0.1K_d$, the dimer forms only 15% of the total enzyme. For the determination of low values of K_d (10^{-7} to 10^{-9}), the experimental enzyme concentration has to be kept close to these values, which consequently complicates the detection of the response, and thus, it allows only fluorescent detection in kinetic studies. Methods such as NMR or light scattering, which require high protein concentrations, are almost excluded from these studies unless the protease has a dimerization defect leading to a much higher K_d (Ishima et al. 2001; Schatz et al. 2001; Louis et al. 2003).

The results presented in this article were obtained under optimal conditions of pH and ionic strength for each wild-type enzyme, and the substrate concentrations were kept below the K_d of the dimer. Two different substrates for MAV PR gave similar results. Other artifacts that might interfere with the interpretation also are unlikely. Autoprocessing, precipitation, or denaturation of the enzyme or its adsorption on the walls of preincubation tubes are excluded by those experiments performed with higher enzyme concentrations, because they run to a well-pronounced equilibrium.

In the experimental results presented, there are two principal sources of the experimental error, the error of enzyme titration and the error of determination of the ratio of equilibrium and initial rates. The error of K_d was estimated using the equation

$$\Delta K_d = \left|\frac{\partial K_d}{\partial E_0}\right|\Delta E_0 + \left|\frac{\partial K_d}{\partial Q}\right|\Delta Q, \qquad (14)$$

where ΔK_d means the error of the equilibrium constant, ΔE_0 is the error of the initial concentration obtained from the nonlinear regression procedure, and ΔQ is the error estimate of the equilibrium and initial rates ratio. The error of enzyme titration depends on the availability of a tight-binding inhibitor of the corresponding enzyme. In the case of MAV and MPMV enzymes, for which only less tightly binding inhibitors are available, the error is higher. Furthermore, the fluorometric detection of enzyme activity, especially the signal-to-noise ratio, is supported either by high enzyme activity or lower dimer stability, because the experiment then can be carried out with more enzyme. Accordingly, the lowest experimental error was obtained for the enzymes with high activity and low K_d value (such as HIV-1-PR[T]) and the highest for MAV-PR*(T)* and MPMV-PR(T).

As regards the rate constant determination, we decided to determine experimentally the dissociation constant k_2 and then calculate k_1 by using the value of of K_d. The reason is that dimer dissociation is the prevalent process immediately after the dilution of the enzyme solution, and k_2 is thus less affected by the experimental error. Moreover, dimer dissociation obeys first-order kinetics, which are concentration independent. This fact precludes the influence of enzyme-titration error on k_2, when only short-time measurements are taken into account. The values of k_1, on the other hand, are influenced by all the errors of the both previous determinations, which makes them less reliable for enzymes in which both the measurements are troublesome, as, for example, MPMV-PR(T).

Because of the complexity of the mathematical procedure used to determine k_2 and its weak dependence on the enzyme concentration, the experimental error of this constant was estimated as a standard deviation of the mean of a set of independent experimental results. The errors of the dependent quantities, $\tau_{1/2}$ and k_1, were evaluated by using the following expressions:

$$\Delta\tau_{1/2} = \frac{\Delta k_2 \ln 2}{k_2^2}, \qquad (15a)$$

$$\Delta k_1 = \frac{K_d \Delta k_2 + k_2 \Delta K_d}{K_d^2}. \qquad (15b)$$

The determined values of K_d (see Fig. 3) differ considerably among the studied enzymes. Physically meaningful dependencies can, however, be observed. First, the K_d of wild-type HIV-1-PR (HIV-1-PR[T]) agrees well with the other kinetically determined values of this constant in similar conditions, especially as published by Pargellis et al. (1994) and

147

Uhlíková et al. (1996). However, the majority of the sedimentation experiments provide values that are higher by three to four orders of magnitude. We hypothesize that this disagreement is caused by the pressure dependence of the equilibrium constant, as found, for example, in the case of LexA represor (Mohana-Borges et al. 2000). High pressure in the cuvette forces proteins to decrease their volume, leading to unfolding or dissociation of multimeric structures (Silva et al. 2002). The equilibrium constant K_d depends on pressure via the standard molar Gibbs reaction energy $\Delta_r G^0$, $K_d = \exp(-\Delta_r G^0/RT)$, where R is the molar gas constant ($R = 8.314$ J mole^{-1}K^{-1}) and T is thermodynamic temperature. The partial derivative of $\Delta_r G^0$ with respect to the pressure is $\partial \Delta_r G^0/\partial p = \Delta_r V^0$, where $\Delta_r V^0$ is the molar reaction change of volume. Thus, if an approximation is made that this quantity is independent of pressure, the pressure dependence of $\Delta_r G^0$ is as follows: $\Delta_r G^0(p) = \Delta_r G^0(p_0) + \Delta_r V^0 (p-p_0)$. Now it is possible to compare results of kinetic and sedimentation studies, and estimate the value of $\Delta_r V^0$ that might cause the observed difference. If an example of the kinetic result of the present study ($K_d = 15$ nM) and the sedimentation result of Stříšovský et al. (2000; $K_d = 77$ μM) are taken, the former gives $\Delta_r G^0 = 45$kJ/mole; the latter, $\Delta_r G^0 = 24$ kJ/mole. The pressure during ultracentrifugation is in the order of 5 MPa. Thus, the corresponding $\Delta_r V^0$ is ~−4.3 × 10^{-3} m^3/mole, which is −7.1 × 10^{-27} m^3 per a reaction of one dimer molecule. According to the calculation method presented by Zamyatnin (1972), the volume of the dimer is about 2.7 × 10^{-26} m^3. Similar volume estimates can be deduced also from the Protein Data Bank crystal structures. Hence, the calculated volume change after dissociation of a single PR dimer decreases by one fourth. Although this estimate of the volume change is higher than the values measured for other proteins (Foguel et al. 1998; Mohana-Borges et al. 2000; Suarez et al. 2001), the volume change might account for the experimentally observed high K_d values, when sedimentation analysis is used for K_d determination.

The dimer of HIV-1-PR(T) is by more than an order of magnitude more stable than that of MAV-PR(S), but is comparable with MPMV-PR(T) dimer. This observation means that in the concentration range between ~10^{-6} and 10^{-9} M, the HIV-1 and MPMV PRs are considerably more active as a direct consequence of the process of dimerization. Dimerization could play an important role in regulation of activity. The concentration of the protease in the viral particle, estimated to be ~0.1 mM (Konvalinka et al. 1995), at first glance seems too high for such regulation. However, dimerization might be important for activity regulation in the cell before the formation of an immature viral particle, although the quantification of this process is difficult. It was reported that N-terminal extension of the protease may change both its activity and dimerization properties (Zybarth and Carter 1995; Schatz et al. 2001). The requirement for dimerization

prevents the cleavage of cellular proteins and the premature processing of viral polyproteins by viral PR. It was shown (Burstein et al. 1991; Kräusslich 1991) that introduction of a tethered PR dimer into a retroviral provirus lead to the loss of infectivity and particle formation, clearly due to the premature activation of the polyprotein processing.

Our experiments show that the dimers of the T-version of MAV and HIV PRs are more stable than are their S-analogs, with the difference in K_d being about an order of magnitude for both enzymes. The structural reason for this phenomenon is far from clear, but the observation itself is in correspondence with the previous virological expectations. Thus, the expression strategy of a retrovirus (PR in frame with its substrate or not) correlates well with the amino acid after the active-site aspartate (serine or threonine), as well as with the K_d value. Hence, the change in this amino acid might be a direct cause of the different dimer stability, which in turn helps to regulate the replication of the virus. Our results thus strongly support the hypothesis that Thr→Ser mutation in fireman's grip destabilizes the dimer of the protease and vice versa, and this mechanism is used in the nature for regulation of proteolytic activity of PRs.

The determined values of the kinetic constant k_2 (and also $\tau_{½}$) summarized in Table 3 show that the increased dimer stability of the T-version of PRs is caused by a higher rate of the dimer association rather than slower dissociation. Thus, the respective amino acid does not probably stabilize the dimer considerably but plays an important role in the mechanism of dimer association. It is conceivable that association and dissociation are not strictly reciprocal but proceed by different pathways. Provided that the dimer is kept together mainly by the interactions in the N- and C-terminal dimerization domain and the flap region (see Fig. 1A), it is not surprising that the influence of the fireman's grip mutation to the dissociation process is small. However, these two domains are probably insufficient to induce the dimer formation on their own. We hypothesize that the fireman's grip provides an aid to dimerization, mediating the initial contact of the two monomer molecules and adjusting them to the proper conformation and/or orientation. Several possible mechanisms for this phenomenon can be proposed. One of them can be inferred from a detail of this structure (Fig. 1B). This domain is capable of binding the two chains quite firmly by a network of hydrogen bridges and forms some predimeric intermediate. Recent structural analyses (Ishima et al. 2001; Louis et al. 2003) have demonstrated the existence of a folded monomer conformationally very similar to the subunits of a dimer. Thus, the conformation change of the monomer subunit is likely to be small and the initial intermediate may play a role in orientation of the particles to the proper positions. The fireman's grip domain is situated in an optimal position for this purpose (as opposed to, e.g., dimerization domain at the PR termini), because it is close to the mass center of the dimer. Hence, once

the subunits are connected in this region and even oriented properly, the dimerization process might be facilitated.

The remarkable difference between threonine and serine in the fireman's grip motif also can be anticipated from Figure 1B. Methyl groups of threonines may come into a tight contact during the dimer formation. An interaction of these groups may restrict the rotation around the joint connecting the two monomers and, therefore, stabilize the nascent dimer. The conformation of the threonine side chains need not change in a major way during this process, because even in the final dimer structure, they are still close to one of the favorable rotamers, as can be deduced from the rotamer library (http://www.fccc.edu/research/labs/dunbrack/bbdep.html; see also Dunbrack Jr. and Karplus 1993; Dunbrack Jr. and Cohen 1997). If threonine is substituted by serine, the methyl groups are missing. The nascent intermediate is lacking their interaction, and its geometry may therefore be less favorable for establishing the contacts with the other domains important for dimerization.

Materials and methods

Expression and purification of recombinant enzymes

DNA coding for MAV-PR was a kind gift of Jiří Hejnar (Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Prague, Czech Republic), and the DNA coding for the Ser38Thr mutant of MAV-PR was kindly provided by Moshe Kotler (Department of Molecular Genetics, Hebrew University, Hadassah Medical School, Jerusalem, Israel). The DNA coding regions of MAV-PR(S) and MAV-PR(T) were PCR-amplified and cloned into the expression vector pET 22b (Novagen) by using 5′-AGAAGCTTCTCGAGCTATAAATTTGTCAAGCG-3′ as a downstream primer and 5′-TAGGTACCATATGGGATCCCC-GGGACATTAT-3′ as the upstream primers for MAV-PR, cloned with 65 codon extension at the 5′-end, and 5′-TAGGTACCATAT GTTAGCGATGACAATGGAG-3′ for the mutated MAV-PR(T) (without extension). Restriction endonucleases *Xho*I and *Nde*I were used, respectively. Coding areas of the constructs were checked by DNA sequencing. The expression and purification protocol follows the procedure of Pichová et al. (1992) with minor modifications. Bacterial cells BL21(DE3) were transformed by the expression plasmids and grown to the optical density of 1.0. The expression was induced by 1 mM IPTG (isopropyl-β-D-galacto-pyranoside); after 3 h of expression cells were harvested, washed by PBS buffer (10 mM PO_4^{3-}, 100 mM NaCl at pH 7.2), and resuspended in buffer A (50 mM Tris-HCl, 50 mM NaCl, 5 mM EDTA at pH 8.0) with the addition of phenylmethylsulfonyl fluoride (PMSF, 10 μg/mL). Cells were disrupted by two cycles of freezing/thawing (−70°C), 30-min incubation with chicken egg lysozyme (0.5 mg/mL), and another 30-min incubation with 1% sodium deoxycholate (0.05 mL per 1 mL of the cell suspension) followed by three 1-min sonication cycles. Suspension was centrifuged for 10 min at 1000*g* at 4°C; pellet was resuspended in 3 mL buffer B (50 mM Tris-HCl, 5 mM PO_4^{3-}, 30 mM NaCl, 2 mM EDTA, 0.1% [v/v] mercaptoethanol at pH 7.5) and then dissolved in buffer B containing 9 M urea. The solution was dialyzed against buffer C (20 mM Tris-HCl, 10 mM PO_4^{3-}, 20mM NaCl, 1 mM EDTA, 0.05% [v/v] mercaptoethanol, 5% [v/v] glycerol at

pH 7.0) and centrifuged (20,000*g*, 4°C, 30 min). The refolding was repeated three times. Supernatants were collected and purified by batchwise anion-exchange chromatography on DEAE-Sephadex in buffer C, and the unbound material was further dialyzed into the buffer D (10 mM sodium acetate, 1 mM EDTA, 0.05% 2-mercaptoethanol at pH 5.0), loaded on the column of SP-Sepharose and eluted by NaCl gradient (0 to 0.4M) in buffer D.

HIV-1-PR(T) and HIV-1-PR harboring Thr26Ser mutation (HIV-1-PR*[S]*) were prepared as described in Střišovský et al. (2000). MPMV-PR (the 12-kD form; Zábranský et al. 1998) was a kind gift of Iva Pichová (Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Prague, Czech Republic).

Titration of enzyme by tight-binding inhibitor

One milliliter of assay buffer AB (0.1 M sodium acetate, *x* M NaCl, 1 mM EDTA, 10%[v/v] glycerol at pH 5.0; x = 2.0 for both wild-type and mutated MAV-PR[S] and *[T]* and 0.3 for the other enzymes) was mixed with a constant amount (typically 10 to 20 μL of 3 mM stock solution) of the peptide substrate AlaThrHis-GlnValTyr (p-nitro-Phe) ValArgLysAla (5 mg/mL) and a variable amount (0 to 100 μL) of a tight binding competitive inhibitor (Majer et al. 1993). Enzyme-specific peptide inhibitors were used for this purpose: QF34, 1 μM (Konvalinka et al. 1997) for HIV-1-PR(T) and HIV-1-PR *(S)*; ProProCysVal (PheSta) AlaMetThrMet, 13 μM for MAV-PR(S) and MAV-PR *(T)*; and ProTyrVal (PheSta) AlaMetThr, 36 μM for MPMV-PR(*T*). The reaction was started by the addition of 10 μL of the enzyme and was monitored spectrophotometrically. Typically, the reaction was performed for several different values of inhibitor concentration, and the initial velocity was determined for each of them. The enzyme concentration was determined by nonlinear regression of the dependence of the initial velocity on the inhibitor concentration.

Fluorogenic substrates for kinetic studies

Two peptide substrates were designed by using an internally quenching pair of fluorescent groups EDANS (donor) and DABCYL (acceptor) as reported (Matayoshi et al. 1990). The sequences of the two substrates are as follows: FS1, (Glu[EDANS]ThrHisGln ValTyr↓PheValArgLysAlaLys[DABCYL]-NH₂); FS2, (Glu[EDANS] ThrProGlnValTyr↓Phe ValArgLysAlaLys [DABCYL]-NH₂; the arrow denotes the cleavage site). The fluorogenic substrates were synthesized on the solid phase by using standard *O*-(benzotriazol-1-yl)-N,N,N′,N′-tetramethyluronium tetrafluoroborate (TBTU) mediated 9-fluorenylmethoxycarbonyl (Fmoc) chemistry on poly-styrene support with peptide amide linker (PAL), 5-(4-Fmoc-aminomethyl-3,5-dimethoxyphenoxy)valeric acid. Both the fluoro-phore (EDANS) and the quencher (DABCYL) were introduced onto the corresponding amino acids (Glu and Lys) prior to their attachment; hence, FmocGlu(EDANS)-OH and FmocLys (DABCYL)-OH were used for the coupling. The peptides were cleaved off the resin and simultaneously deprotected by 5% thio-anisole, 3% ethandithiole, 2% anisole, and 90% trifluoroacetic acid (TFA) and were finally purified by high-performance liquid chromography (Vydac C18 column) using water/acetonitrile/0.1% TFA as mobile phase in gradient elution (Gulnik et al. 1997).

Kinetic characterization of fluorogenic substrates

A series of reactions was performed in which increasing amounts of the substrate (typically 3 to 15 μL of 2 mM stock solution) was

149

mixed with 3 mL of assay buffer AB (see above), and the reaction was started by addition of 3 μL of the enzyme. The fluorescence increase was monitored for 300 sec, and the initial reaction rate was determined. The constants K_m and k_{cat} were calculated by fitting the data to the Michaelis-Menten equation by nonlinear regression.

Determination of dimerization parameters

A stock concentration of the enzyme was dialyzed into the assay buffer AB (see above) and titrated by tight-binding inhibitor. To determine the dimerization parameters K_d and k_2, several series of individual reaction runs were performed, each of them for different enzyme concentration. Every experiment consisted of several reactions differing in the preincubation time. The reaction for the zero preincubation time: 3 mL of the assay buffer was mixed with known amount (1 or 2 μL) of fluorogenic substrate (~2 mM), and the reaction was started by the addition of the corresponding amount of enzyme. The course of the reaction was monitored fluorometrically for 300 sec. The reaction mixtures for the other reactions were prepared mixing 3 mL of the assay buffer with the corresponding amount of the enzyme in plastic tubes. These samples were preincubated in 37°C, each one for a different time (2 to 75 min). After preincubation, the content of the tube was transferred into the cuvette, and substrate was added in the same amount as in the first reaction; the reaction course was monitored. Initial reaction rate was determined graphically as the negative value of the slope of this dependence for $t_{reaction} = 0$. After completing the whole series, the limit reaction rate for $t_{preincubation} \rightarrow \infty$ was estimated. The dimerization parameters were derived from the set of initial reaction velocities for all these series by a mathematical procedure described in Results.

Acknowledgments

References

Arad, G., Chorev, M., Shtorch, A., Goldblum, A., and Kotler, M. 1995. Point mutation in avian sarcoma leukaemia virus protease which increases its activity but impairs infectious virus production. *J. Gen. Virol.* **76:** 1917–1925.

Burstein, H., Bizub, D., and Skalka, A.M. 1991. Assembly and processing of avian retroviral Gag polyproteins containing linked protease dimers. *J. Virol.* **65:** 6165–6172.

Cheng, Y.S., Yin, F.H., Foundling, S., Blomstrom, D., and Kettner, C.A. 1990. Stability and activity of human immunodeficiency virus protease: Comparison of the natural dimer with a homologous, single-chain tethered dimer. *Proc. Natl. Acad. Sci.* **87:** 9660–9664.

Darke, P.L. 1994. Stability of dimeric retroviral proteases. *Methods Enzymol.* **241:** 104–127.

Darke, P.L., Jordan, S.P., Hall, D.L., Zugay, J.A., Shafer, J.A., and Kuo, L.C. 1994. Dissociation and association of the HIV-1 protease dimer subunits: Equilibria and rates. *Biochemistry* **33:** 98–105.

Dunbrack Jr., R.L. and Cohen, F.E. 1997. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* **6:** 1661–1681.

Dunbrack Jr., R.L. and Karplus, M. 1993. Backbone-dependent rotamer library for proteins: Application to side-chain prediction. *J. Mol. Biol.* **230:** 543–574.

Dunn, B.M. 1998. Human immunodeficiency virus 1 retropepsin. In *Handbook of proteolytic enzymes* (eds. A.J. Barret et al.), pp. 919–928. Academic Press, London.

Foguel, D., Silva, J.L., and de Prat-Gay, G. 1998. Characterization of a partially folded monomer of the DNA-binding domain of human papilloma virus E2 protein obtained at high pressure. *J. Biol. Chem.* **273:** 9050–9057.

Grant, S.K., Deckman, I.C., Culp, J.S., Minnich, M.D., Brooks, I.S., Hensley, P., Debouck C., and Meek, T.D. 1992. Use of protein unfolding studies to determine the conformational and dimeric stabilities of HIV-1 and SIV proteases. *Biochemistry* **31:** 9491–9501.

Gulnik, S.V., Suvorov, L.I., Majer, P., Collins, J., Kane, B.P., Johnson, D.G., and Erickson, J.W. 1997. Design of sensitive fluorogenic substrates for human cathepsin D. *FEBS Lett.* **413:** 379–384.

Holzman, T.F., Kohlbrenner, W.E., Weigl, D., Rittenhouse, J., Kempf, D., and Erickson, J. 1991. Inhibitor stabilization of human immunodeficiency virus type-2 proteinase dimer formation. *J. Biol. Chem.* **266:** 19217–19220.

Ishima, R., Ghirlando, R., Tozser, J., Gronenborn, A.M., Torchia, D.A., and Louis, J.M. 2001. Folded monomer of HIV-1 protease. *J. Biol. Chem.* **276:** 49110–49116.

Jordan, S.P., Zugay, J., Darke, P.L., and Kuo, L.C. 1992. Activity and dimerization of human immunodeficiency virus protease as a function of solvent composition and enzyme concentration. *J. Biol. Chem.* **267:** 20028–20032.

Konvalinka, J., Strop, P., Velek, J., Cerna, V., Kostka, V., Phylip, L.H., Richards, A.D., Dunn, B.M., and Kay, J. 1990. Sub-site preferences of the aspartic proteinase from the human immunodeficiency virus, HIV-1. *FEBS Lett.* **268:** 35–38.

Konvalinka, J., Bláha, I., Škrabana, R., Sedláček, J., Pichová, I., Kaprálek, F., Kostka, V., and Štrop, P. 1991. Subsite specificity of the proteinase from myeloblastosis associated virus. *FEBS Lett.* **282:** 73–76.

Konvalinka, J., Hořejší, M., Andreánský, M., Novek, P., Pichová, I., Bláha, I., Fábry, M., Sedláček, J., Foundling, S., and Štrop, P. 1992. An engineered retroviral proteinase from myeloblastosis-associated virus acquires pH dependence and substrate specificity of the HIV-1 proteinase. *EMBO J.* **11:** 1141–1144.

Konvalinka, J., Litterst, M.A., Welker, R., Kottler, H., Rippmann, F., Heuser, A.M., and Krausslich, H.G. 1995. An active-site mutation in the human immunodeficiency virus type 1 proteinase (PR) causes reduced PR activity and loss of PR-mediated cytotoxicity without apparent effect on virus maturation and infectivity. *J. Virol.* **69:** 7180–7186.

Konvalinka, J., Litera, J., Weber, J., Vondrášek, J., Hradilek, M., Souček, M., Pichová, I., Majer, P., Štrop, P., Sedláček, J., et al. 1997. Configurations of diastereomeric hydroxyethylene isosteres strongly affect biological activities of a series of specific inhibitors of human-immunodeficiency-virus proteinase. *Eur. J. Biochem.* **250:** 559–566.

Kräusslich, H.G. 1991. Human immunodeficiency virus proteinase dimer as component of the viral polyprotein prevents particle assembly and viral infectivity. *Proc. Natl. Acad. Sci.* **88:** 3213–3217.

Kuzmič, P. 1993. Kinetic assay for HIV proteinase subunit dissociation. *Biochem. Biophys. Res. Commun.* **191:** 998–1003.

Louis, J.M., Ishima, R., Nesheiwat, I., Pannell, L.K., Lynch, S.M., Torchia, D.A., and Gronenborn, A.M. 2003. Revisiting monomeric HIV-1 protease: Characterization and redesign for improved properties. *J. Biol. Chem.* **278:** 6085–6092.

Majer, P., Urban, J., Gregorova, E., Konvalinka, J., Novek, P., Stehlikova, J., Andreansky, M., Sedlacek, J., and Strop, P. 1993. Specificity mapping of HIV-1 protease by reduced bond inhibitors. *Arch. Biochem. Biophys.* **304:** 1–8.

Matayoshi, E.D., Wang, G.T., Krafft, G.A., and Erickson, J. 1990. Novel fluorogenic substrates for assaying retroviral proteases by resonance energy transfer. *Science* **247:** 954–958.

Mohana-Borges, R., Pacheco, A.B., Sousa, F.J., Foguel, D., Almeida, D.F., and Silva, J.L. 2000. LexA repressor forms stable dimers in solution: The role of specific DNA in tightening protein–protein interactions. *J. Biol. Chem.* **275:** 4708–4712.

Pargellis, C.A., Morelock, M.M., Graham, E.T., Kinkade, P., Pav, S., Lubbe, K., Lamarre, D., and Anderson, P.C. 1994. Determination of kinetic rate con-

150

stants for the binding of inhibitors to HIV-1 protease and for the association and dissociation of active homodimer. *Biochemistry* **33:** 12527–12534.

Pearl, L. and Blundell, T. 1984. The active site of aspartic proteinases. *FEBS Lett.* **174:** 96–101.

Pearl, L.H. and Taylor, W.R. 1987. A structural model for the retroviral proteases. *Nature* **329:** 351–354.

Pichová, I., Štrop, P., Sedláček, J., Kaprálek, F., Beneš, V., Trávníček, M., Pavlíčková, L., Souček, M., Kostka, V., and Foundling, S. 1992. Isolation, biochemical characterization and crystallization of the p15gag proteinase of myeloblastosis associated virus expressed in *E. coli. Int. J. Biochem.* **24:** 235–242.

Ridky, T.W., Bizub-Bender, D., Cameron, C.E., Weber, I.T., Wlodawer, A., Copeland, T., Skalka, A.M., and Leis, J. 1996. Programming the Rous sarcoma virus protease to cleave new substrate sequences. *J. Biol. Chem.* **271:** 10538–10544.

Rose, J.R., Babe, L.M., and Craik, C.S. 1995. Defining the level of human immunodeficiency virus type 1 (HIV-1) protease activity required for HIV-1 particle maturation and infectivity. *J. Virol.* **69:** 2751–2758.

Schatz, G.W., Reinking, J., Zippin, J., Nicholson, L.K., and Vogt, V.M. 2001. Importance of the N terminus of Rous sarcoma virus protease for structure and enzymatic function. *J. Virol.* **75:** 4761–4770.

Silva, J.L., Oliveira, A.C., Gomes, A.M., Lima, L.M., Mohana-Borges, R., Pacheco, A.B., and Foguel, D. 2002. Pressure induces folding intermediates that are crucial for protein-DNA recognition and virus assembly. *Biochim. Biophys. Acta* **1595:** 250–265.

Stříšovský, K., Tessmer, U., Langner, J., Konvalinka, J., and Kräusslich, H.G. 2000. Systematic mutational analysis of the active-site threonine of HIV-1 proteinase: Re-thinking the "fireman's grip" hypothesis. *Protein Sci.* **9:** 1631–1641.

Suarez, M.C., Lehrer, S.S., and Silva, J.L. 2001. Local heterogeneity in the pressure denaturation of the coiled-coil tropomyosin because of subdomain folding units. *Biochemistry* **40:** 1300–1307.

Towler, E.M., Gulnik, S.V., Bhat, T.N., Xie, D., Gustschina, E., Sumpter, T.R., Robertson, N., Jones, C., Sauter, M., Mueller-Lantzsch, N., et al. 1998. Functional characterization of the protease of human endogenous retrovirus, K10: Can it complement HIV-1 protease? *Biochemistry* **37:** 17137–17144.

Uhlíková, T., Konvalinka, J., Pichová, I., Souček, M., Kraüsslich, H.G., and Vondrášek, J. 1996. A modular approach to HIV-1 proteinase inhibitor design. *Biochem. Biophys. Res. Commun.* **222:** 38–43.

Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, L.M., Clawson, L., Schneider, J., and Kent, S.B. 1989. Conserved folding in retroviral proteases: Crystal structure of a synthetic HIV-1 protease. *Science* **245:** 616–621.

Xie, D., Gulnik, S., Gustchina, E., Yu, B., Shao, W., Qoronfleh, W., Nathan, A., and Erickson, J.W. 1999. Drug resistance mutations can effect dimer stability of HIV-1 protease at neutral pH. *Protein Sci.* **8:** 1702–1707.

Zábranský, A. 1999. "Analysis of the protease from Mason-Pfizer monkey retrovirus." PhD thesis, Institute of Organic Chemistry and Biochemistry, Prague, Czech Republic.

Zábranský, A., Andreánsky, M., Hrušková-Heidingsfeldová, O., Havlíček, V., Hunter, E., Ruml, T., and Pichová, I. 1998. Three active forms of aspartic proteinase from Mason-Pfizer monkey virus. *Virology* **245:** 250–256.

Zamyatnin, A.A. 1972. Protein volume in solution. *Prog. Biophys. Mol. Biol.* **24:** 107–123.

Zhang, Z.Y., Poorman, R.A., Maggiora, L.L., Heinrikson, R.L., and Kezdy, F.J. 1991. Dissociative inhibition of dimeric enzymes: Kinetic characterization of the inhibition of HIV-1 protease by its COOH-terminal tetrapeptide. *J. Biol. Chem.* **266:** 15591–15594.

Zybarth, G. and Carter, C. 1995. Domains upstream of the protease (PR) in human immunodeficiency virus type 1 Gag-Pol influence PR autoprocessing. *J. Virol.* **69:** 3878–3884.

151

P8. How is the activity of shikimate dehydrogenase from the root of Petroselinum crispum (parsley) regulated and which side reactions are catalyzed?

How is the activity of shikimate dehydrogenase from the root of *Petroselinum crispum* (parsley) regulated and which side reactions are catalyzed?

Veronika Hýsková [a,*], Kateřina Bělonožníková [a], Ingrida Šmeringaiová [a], Daniel Kavan [a], Marek Ingr [a,b], Helena Ryšlavá [a]

[a] Department of Biochemistry, Faculty of Science, Charles University, Hlavova 2030, Prague 2, 128 40, Czech Republic
[b] Tomas Bata University in Zlín, Faculty of Technology, Department of Physics and Materials Engineering, Nám. T.G. Masaryka 5555, 760 01, Zlín, Czech Republic

ARTICLE INFO

ABSTRACT

Inhibitors of the shikimate pathway are widely used as herbicides, antibiotics, and anti-infectious drugs. However, the regulation of the shikimic pathway is complex, and little is known about the feedback regulation of the shikimate dehydrogenase (SDH, EC 1.1.1.25) in plants. Thus, the aim of this study was to elucidate the kinetic mechanism of SDH purified from the root of *Petroselinum crispum* (parsley), to determine all possible reaction products and to identify phenylpropanoid compounds that affect its activity. Our results showed that the bisubstrate reaction catalyzed by *P. crispum* SDH follows a sequential ordered mechanism, except for three dead-end complexes. The main and lateral reactions of SDH were monitored by mass spectrometry, thereby detecting protocatechuic acid as a byproduct. Gallic acid was formed non-enzymatically, whereas quinate was not detected. Several polyphenolic compounds inhibited SDH activity, especially tannic, caffeic and chlorogenic acids, with IC_{50} 0.014 mM, 0.15 mM, and 0.19 mM, respectively. The number of hydroxyl groups influenced their inhibition effect on SDH, and *p*-coumaric, *t*-ferulic, sinapic, syringic and salicylic acids were less effective SDH inhibitors. Nevertheless, one branch of the phenylpropanoid pathway may affect SDH activity through feedback regulation.

1. Introduction

The shikimate pathway is well known for the synthesis of aromatic amino acids in bacteria, fungi, apicomplexan parasites, and plants. In turn, this pathway is absent in animals. For this reason, inhibitors of individual enzymes were identified as antimicrobial agents and herbicides (Carrington et al., 2018; Deng and Lu, 2017; Tzin and Galili, 2010). The key enzyme of the shikimate pathway is shikimate dehydrogenase (SDH, EC 1.1.1.25). In plants, SDH is a bifunctional enzyme that catalyzes steps three and four. The N-terminal domain functions as 3-dehydroquinate dehydratase (DHD, EC 4.2.1.10), the C-terminal domain catalyzes the reduction of 3-dehydroshikimate (DHS) to shikimate (SA) in the presence of NADPH. The regulation of the shikimate pathway in plants is more complex than in bacteria and yeast (Carrington et al.,

2018; Deng and Lu, 2017; Heldt et al., 2011; Tzin and Galili, 2010). The products of the shikimate pathway include not only phenylalanine, tyrosine, and tryptophan but also a wide range of polyphenolic compounds, such as phenolic acids, lignans, lignin, flavonoids, stilbenes and tannins. These compounds can protect plants against oxidative stress and UV radiation, impregnate cell walls and have antimicrobial properties. Thus, specialized metabolites are often synthetized and accumulated in plants in response to various stress conditions and can be associated with increased SDH activity (Bělonožníková et al., 2020; Cabane et al., 2004; Hyskova et al., 2017; Kovacik et al., 2009; Moura et al., 2010).

Another reaction catalyzed by SDH was proposed in walnut (*Juglans regia*) and *Escherichia coli*, in which 3-DHS was identified as the substrate of SDH in the presence of NADP. In the proposed SDH reaction

mechanism, 3-DHS is oxidized to 3,5-didehydroshikimate, and gallic acid is formed after enolization, thereby continuously generating both shikimate and gallic acid (Muir et al., 2011). Therefore, this reaction is important for plants such as grapevine (*Vitis vinifera*) that produce derivates of gallic acid, epicatechin 3-gallate and →glucogallin, (Bontpart et al., 2016).

Guo et al. (2014) and Tahara et al. (2021) also described the synthesis of quinate (QA) from 3-dehydroquinate (DHQ) by DHD/SDH from poplar (*Populus trichocarpa*), and from *Eucalyptus camaldulensis*, respectively. However, only some members of the gene family catalyzed this reaction in which NADH is used for reduction instead of NADPH (Guo et al., 2014; Tahara et al., 2021). These reactions were described in plants with more than one gene for DHD/SDH, which could be separated into groups with classical SDH activity and groups that produce quinic acids (Bontpart et al., 2016; Guo et al., 2014; Muir et al., 2011). Moreover, another SDH side reaction DHS dehydration to protocatechuic acid (PCA) is also considered (Bontpart et al., 2016; Muir et al., 2011).

As with other enzymes of the shikimate and phenylpropanoid pathways, SDH is localized in chloroplasts where the reducing equivalents of NADPH and ATP from photosynthesis are available. On the other hand, in non-photosynthetic plastids, the reducing equivalents of NADPH required for synthetic pathways are provided by the oxidative pentose phosphate cycle, i.e., glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase (Esposito er al., 2003). Although the shikimate pathway is a very important metabolic flow in plants, the regulation of its key enzyme (SDH) by phenylpropanoid compounds and possible SDH products are less studied, especially in non-photosynthetic tissues.

Considering the above, this study aimed to find a plant source with naturally high SDH activity and to study the kinetic properties of SDH from non-photosynthetic tissue. Root tissue of *Petroselinum crispum* (Mill.) Fuss (Apiaceae) (parsley) was identified as suited to this purpose. The mechanism of reaction was characterized in both directions, thus gaining insights into dead-end complexes involved in the regulation of parsley SDH activity. The effect of phenolic compounds as potential inhibitors was tested, and the results show that not only tannic acid but also caffeic acid and chlorogenic acid function as SDH strong inhibitors.

2. Results

2.1. Characterization of P. crispum root SDH: typical high pH optimum and 61.5 kDa molecular weight

SDH activity was screened in crude extracts from 8 different vegetables. Onion (*Allium cepa* L., bulb) and broccoli (*Brassica oleracea* var. italica, florets) crude extracts had low SDH activity, whereas crude extracts from *P. crispum* root and zucchini (*Cucurbita pepo* L. var. cylindrical, pulp) were identified as the richest sources of SDH among all plants tested. A high SDH activity per gram of fresh weight correlates with a low total phenolics content and *vice versa* (Fig. A.1). SDH was purified by ion exchange and gel chromatography from *P. crispum* (*Petroselinum crispum*) root to a final specific activity of 470 ± 18 nmol. $min^{-1}.mg^{-1}$.

SDH activity was pH-dependent, with a pH optimum between 7-8.5 and 9.5-10 in the physiological and reverse reactions, respectively (Table 1). The molecular weight was 63 kDa when determined by red native electrophoresis (Fig. A.2A) and 60 kDa when assessed by gel chromatography (Fig. A.2B). The isoelectric point of *P. crispum* SDH was 4.5 (Fig. A.2C). Only one protein band with SDH activity was detected after native red electrophoresis and isoelectric focusing. During gel chromatography, SDH was eluted as a single peak. Thus, only one isoform of SDH is present in *P. crispum* root.

Table 1

Kinetic parameters of *P. crispum* SDH. V_{max} - maximal reaction rate of enzyme reaction, K_m- Michaelis constant of the specific substrate, K'_m -apparent Michaelis constant at a given concentration of the second substrate), K_{SS} - inhibition constant at an excess substrate concentration, K_A - dissociation constant of the enzyme substrate complex, for technical reasons (absorbance above 3 or excess substrate inhibition), this parameter could not be determined. Kinetic constants were calculated as an average of values from at least 3 experiments performed in doublets, S. D. are shown.

	Direction of the reaction catalyzed by SDH	
	SA DHS	DHS SA
pH optimum	9.5 10	7 8.5
V_{max} [µmol.min^{-1}.mg^{-1}] for SA (DHS)	1.34	0.29
V_{max} [µmol.min^{-1}.mg^{-1}] for NADP(H)	1.32	0.28
K_m for SA (DHS) [mM]	0.71 ± 0.13	
K_m for NADP(H) [mM]	0.47 ± 0.16	
K'_m for SA (with 0.2 mM NADP) [mM]	0.16 ± 0.01	
K'_m for DHS (with 0.4 mM NADPH) [mM]		0.23 ± 0.03
K'_m for NADP (with 3 mM SA) [mM]	0.054 ± 0.011	
K'_m for NADPH (with 2 mM DHS) [mM]		0.016 ± 0.03
K_{SS} for SA (DHS) [mM]		0.12 ± 0.07
K_A [mM]	0.030 ± 0.008	0.25 ± 0.13

2.2. The kinetic properties of P. crispum SDH differ from those of other dehydrogenases

The kinetic parameters of the reaction catalyzed by *P. crispum* root SDH were studied in both directions, i.e., in the physiological (shikimate pathway) direction: NADPH + DHS SA + NADP, and in the reverse direction: SA + NADP NADPH + DHS (Fig. 1). The Michaelis constants and maximal velocities are summarized in Table 1. The maximal reaction rate for the reverse reaction from SA to DHS was 4.6-fold higher than in the shikimate pathway direction (DHS reduction). Because the saturating concentrations of NADP and SA were high, we also determined the apparent Michaelis constant at approximate conditions (Table 1). Furthermore, excess substrate DHS inhibited the reaction (Fig. A.3) with a substrate inhibition constant of $K_{SS} = 0.12 \pm 0.07$ mM (Table 1).

2.3. Product inhibition analysis confirmed the ordered mechanism of the SDH-catalyzed reaction in both directions

The kinetic mechanism of the bisubstrate SDH reaction was analyzed in both directions. To identify the type of mechanism, we constructed Lineweaver-Burk diagnostical plots (Fig. 1), a Hanes plot and an Eadie-Hofstee plot (data not shown) and performed product inhibition assays (Fig. 2, Table 2). All kinetic parameters including V_{max}, K_m, K_A, K_{ic}, K_{iu}, K_i, V^*_{max}, K_{SS}, $S_{0.5}$ were calculated from non-linear regression using particular equations (Bisswanger, 2002; Marangoni, 2003).

The initial rate of the reaction in the physiological direction was measured using several concentrations of DHS and NADPH, showing the typical Lineweaver-Burk plot of a sequential mechanism: straight lines with an intercept left to the ordinate (Fig. 1) and K_A 0.25 ± 0.13 mM. In this direction, the free enzyme binds to NADPH, which allows DHS, but not SA, binding in an ordered mechanism, thus partly explaining the mutual competition between NADPH and SA. NADPH binding apparently prevents NADP binding, and *vice versa*, leading to bilateral mutual competitive inhibition (Fig. 3, Table 2).

The direction of SA oxidation confirmed the strong affinity between the free enzyme and NADP. The mechanism was sequential but with a very low, almost immeasurable, dissociation constant for the complex enzyme substrate (K_A), thus it seems to look like a ping-pong. Obviously, the complex enzyme-NADP is very thermodynamically stable (Fig. 1).

NADP binding enables both DHS and SA binding. The latter follows a classical ordered mechanism in the direction of SA + NADP NADPH +
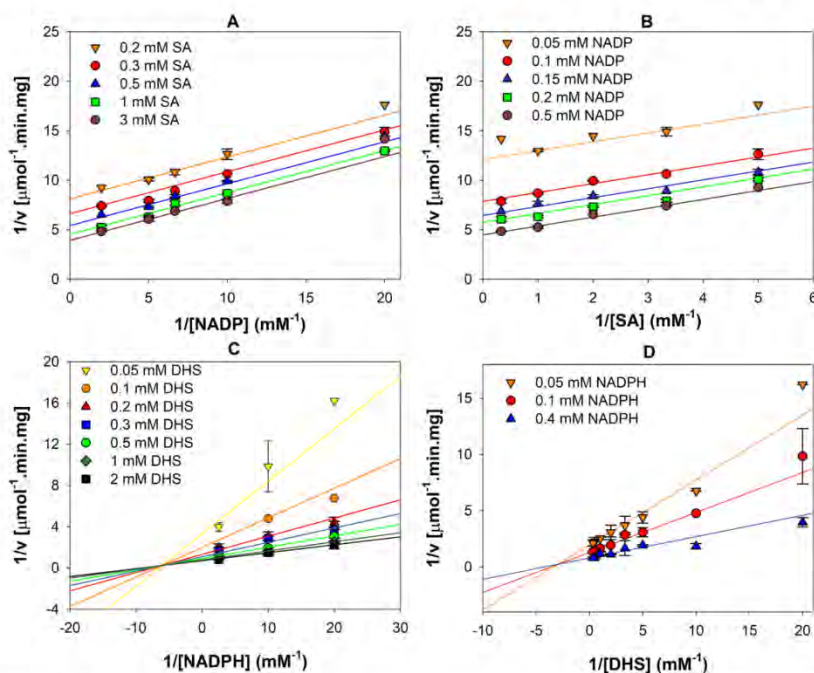
Fig. 1. Determination of the type of SDH reaction mechanism based on Lineweaver-Burk diagnostical plots for the direction from SA to DHS (A,B) and from DHS to SA (C,D) Double reciprocal plots are fitted to an equation corresponding to a sequential mechanism.

Table 2
Results of the product inhibition study for the determination of the bisubstrate reaction mechanism of *P. crispum* SDH. Inhibition constants were calculated as an average of values from 2-4 experiments carried out in doublets, S.D. are shown. NC - non-competitive C - competitive, and N.I. - no inhibition. Patterns and values in brackets mean that the reaction is saturated with the 2nd substrate.

	Product inhibitor				Product inhibitor			
	Q (NADP) Varied substrate		P (SA)		A (NADPH) Varied substrate		B (DHS)	
i.e.	1/NADPH	1/DHS	1/NADPH	1/DHS	1/NADP	1/SA	1/NADP	1/SA
2nd substrate	subsaturated				subsaturated (saturated)			
Pattern	C	NC	C	NC	C (C)	NC (N.I.)	NC (N.I.)	C (C)
Ki [mM]	0.22 ± 0.02	0.4 ± 0.1	0.7 ± 0.1	0.7 ± 0.3	0.12 ± 0.07 (0.49 ± 0.03)	0.36 ± 0.08	1.4 ± 0.2	0.13 ± 0.03 (0.33 ± 0.03)
Mechanism	steady state sequential order				steady state sequential order			

DHS; in contrast, the former results in the formation of a dead-end complex. Its origin explains the competitive inhibition of DHS against the substrate SA. Thus, these findings further confirm the competition between NADP and NADPH (Fig. 3, Table 2).

DHS binding to the free enzyme prevents NADPH binding by forming a dead-end complex (enzyme-DHS) that precludes a disordered mechanism in the direction of NADPH + DHS   SA + NADP. In addition, another dead-end complex (enzyme-DHS-NADP) is also formed when NADP binds to the enzyme-DHS (Fig. 3).

SA binding to the free enzyme prevents both NADP and NADPH binding also by forming a dead-end complex, which, in the case of NADP, prevents a disordered mechanism in the direction of SA + NADP   NADPH + DHS. In the case of NADPH, the formation of the dead-end complex completes the explanation for the competitive inhibition by the SA inhibitor against the NADPH substrate. Thus, SA competes with

NADPH to bind to free enzyme, but not the other way around, that is, NADPH does not compete with SA since SA binding to the enzyme leads to a dead-end complex. Therefore, we proposed an ordered mechanism of the bi-substrate reaction catalyzed by SDH in both directions with three dead-end complexes (enzyme-DHS, enzyme-SA, and enzyme-NADP-DHS) (Fig. 3).

2.4. Identification of all SDH products and cofactor specificity

The activity of SDH was monitored by LC-MS (Table 3), focusing on side reactions. GA was formed non-enzymatically since it was also detected without SDH in the mixtures. PCA was produced from DHS in the *P. crispum* shikimate pathway. No products were detected for DHQ with NADH and for QA with NAD as substrates. The *P. crispum* SDH active routes are shown in Fig. 4.
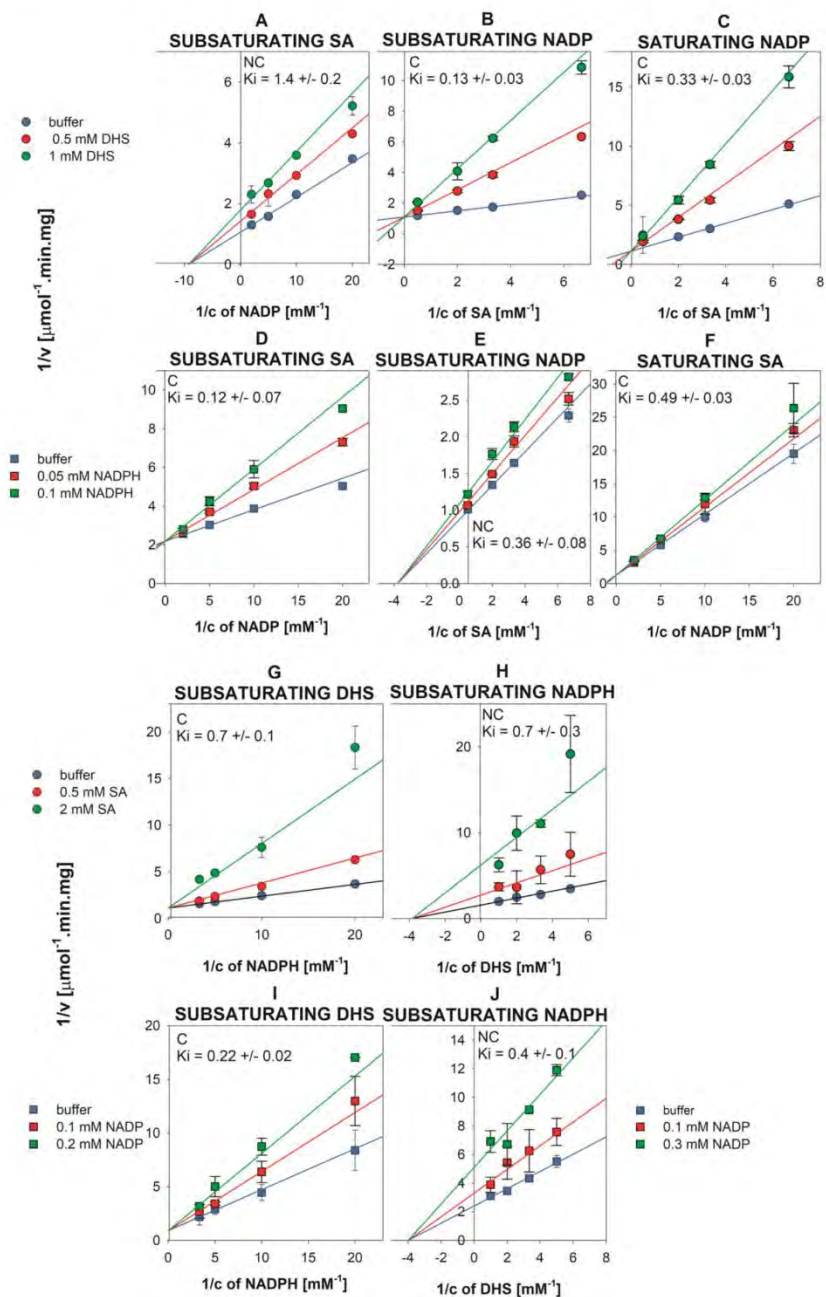
(caption on next page)

Fig. 2. Product inhibition analysis for the determination of the mechanism of bisubstrate reaction in the direction from SA to DHS (A-F) and in the direction from DHS to SA (G-J). Experimental data are fitted with calculated values determined by non-linear regression using equations charactering competitive and non-competitive inhibition, respectively. SDH products DHS (A,B,C) and NADPH (D,E,F) served as competitive (B,C,D,F) and non-competitive (A,E) inhibitors. The saturating (C,F) and subsaturating (A,B,D,E) concentrations of SA and NADP were 20 mM and 2 mM, and 0.3 mM and 0.5 mM, respectively. Saturating concentrations of NADP and variable concentrations of SA caused no inhibition (N.I.) of NADPH and saturating concentrations of SA and variable concentrations of NADP caused N.I. of DHS (data not shown). SDH products SA (G,H) and NADP (I,J) served as competitive (G,I) and non-competitive (H,J) inhibitors at 0.75 mM (subsaturating concentration) DHS (G,I) and 0.2 mM (subsaturating concentration) NADPH (H,J) as a second substrate. K_i indicates inhibition constants in mM. NC - non-competitive and C - competitive inhibition. Measurements were performed in doublets (S.D. are shown) and at least 2-4 times (enzyme preparations from different isolations).
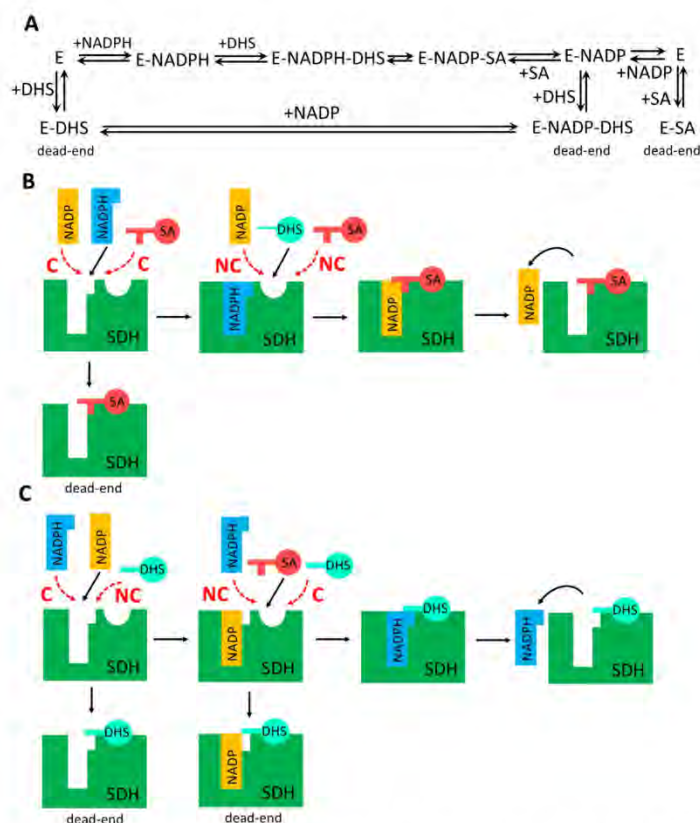
Fig. 3. Proposed kinetic mechanism of the bisubstrate reaction catalyzed by *P. crispum* SDH from non-photosynthetic tissue. SDH reaction scheme (A), graphical representation of the product inhibition study in DHS reduction direction (B) and in SA oxidation direction (C) with a formation of dead-end complexes. Non-Competitive (NC) and competitive (C) product inhibitors are indicated.

2.5. Tannic acid and some phenylpropanoids, as well as Zn^{2+} and Cu^{2+} ions, inhibit SDH activity

Various compounds that participate in phenylpropanoid metabolism and/or are related to SDH were studied as potential modulators of its activity in a screening performed at inhibitor concentrations ranging from 0.006 to 3.5 mM. Aromatic amino acids (Phe, Tyr, Trp), which are synthesized in the shikimate pathway, had no effect on SDH activity. Concurrently, the first product of the phenylpropanoid pathway (*t*-cinnamic acid) and the potential SDH product (quinic acid) did not affect SDH activity, similarly to umbelliferone (7-hydroxycoumarine) and resveratrol (trans 3,4',5 trihydroxystilbene) (data not shown). One-

way ANOVA confirmed that the effect of these compounds was nonsignificant (data not shown). In turn, salicylic, *p*-coumaric, *t*-ferulic, sinapic, syringic, caffeic, chlorogenic (caffeic acid and quinic acid ester) acids and the polyphenolic compound tannic acid had an inhibitory effect on *P. crispum* SDH activity, with tannic acid showing the lowest IC_{50} (Fig. 5). Caffeic acid and chlorogenic acid were also effective inhibitors. Hydroxycinnamic acids, such as *p*-coumaric, *t*-ferulic, sinapic, syringic and salicylic acids, showed much higher IC_{50} values (Fig. 5). In the irreversible inhibition test, the activity of the enzyme incubated with individual inhibitors did not decrease over time (data not shown). Therefore, all inhibitors mentioned above are reversible.

Although gallic acid was also tested, it strongly interfered with the

Table 3
Identification of SDH products by LC-MS analysis. List of products detected after 30 min incubation with (and without) SDH enzyme at 20 °C. DHS, 3-dehydroshikimate; DHQ, 3-dehydroquinate; PCA, protocatechuic acid; SA, shikimic acid; QA, quinate. *the incubation time was 60 min at 20 °C. SDH from *P. crispum* root (10 μl with specific activity 470 nmol.min^{-1}.mg^{-1}) was incubated in the reaction mixture (280 μl) containing 2 mM substrate (SA, DHS, DHQ, QA, GA, or PCA) and 0.14 mM cofactor (NADP, NADPH, NAD, NADH) in 100 mM Tris-HCl pH 8 (physiological direction) or 100 mM AMP-NaOH pH 9 at 20°C for 30 min. The reaction was initiated by adding SDH and stopped by adding 10 μl of 99 % (v/v) formic acid. Blank experiments were prepared similarly, albeit adding formic acid before the enzyme. Each determination was done at least 3-times, the average values and S. D. are shown.

reaction	products [nmol]
SA+NADP	DHS: 5.7 ± 0.7
	PCA: 0.1 ± 0.0
*DHS+NADPH	SA: 4.9 ± 0.2
	GA: 0.5 ± 0.0
	PCA: 0.7 ± 0.1
	DHQ: 1.1 ± 0.0
DHS+NADP	GA: 0.1 ± 0.0
	PCA: 0.3 ± 0.2
DHS+NADP	GA: 2.1 ± 0.4
without SDH	PCA: 2.7 ± 0.6
DHQ+NADP	DHS: 17.4 ± 1.4
	PCA: 4.9 ± 0.4
*DHQ+NADPH	DHS: 3.4 ± 0.8
	SA: 1.4 ± 0.2
	GA: 1.1 ± 0.5
	PCA: 1.3 ± 0.2
DHQ+NADH	No products detected.
QA+NAD	No products detected.
PCA+NADPH	No products detected.

Fig. 4. Identification of *P. crispum* SDH products by reversed-phase liquid chromatography coupled to electrospray mass spectrometry. The symbol Ⓞ indicates that no potential *P. crispum* SDH byproduct was identified in the reaction mixtures by mass spectrometry. QDH, quinate dehydrogenase; QD, quinate dehydratase; DHSD, dehydroshikimate dehydratase.

formation of color complexes, making it impossible to determine its effect. Furthermore, the color of flavanone or any representative of the flavonoids interfered with the colorimetric method, also precluding the analysis of their inhibitory activity.

Table 4 summarizes the effects on SDH activity when adding different ions to the reaction mixture. Generally, among the ions tested in this study (Ca^{2+}, Na$^+$, K$^+$, Mg^{2+}, Mn^{2+}, Zn^{2+}, and Cu^{2+}), only Zn^{2+} and Cu^{2+} had a significant effect on SDH activity. At 0.1 mM and 0.01 mM, Zn^{2+} and Cu^{2+} decreased SDH activity to 10 and 21%, respectively.

3. Discussion

The kinetic study of SDH purified from non-photosynthetic tissue *P. crispum* root containing evaluation of product inhibitors, Michaelis constants and pH optima of both reversible reactions resulted in establishing an ordered mechanism and shift of the equilibrium in favor of physiological shikimate pathway direction in the cell. Furthermore, *P. crispum* SDH exclusively used NADP(H) and SA (DHS) as a coenzyme and a substrate, respectively, while forming PCA as a by-product but not being involved in GA production. Furthermore, phenylpropanoids of one branch were found responsible for feedback regulation of *P. crispum* SDH.

3.1. P. crispum root extracts show high SDH activity

One of the reasons why SDH activity is low in plant crude extracts is the presence of phenolics. In general, when exposed to air, plant phenolics are readily oxidized, generating products that form complexes with proteins and inhibit enzyme activity (Buchanan et al., 2000), as shown in Fig. A.1. Since the total phenolic content is very high in plants, especially in medicinal herbs (Tupec et al., 2017), a group of vegetables was chosen to identify a source of high SDH activity from
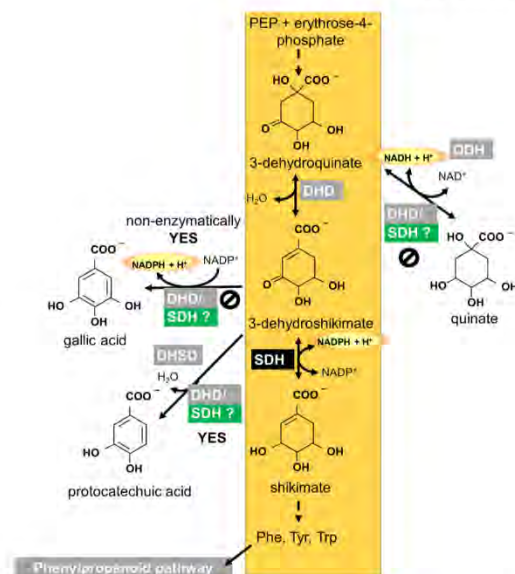
non-photosynthetic tissue. From the group of 8 vegetable sources, the *P. crispum* root exhibited the highest SDH activity. After 3-step purification procedure specific SDH activity in *P. crispum* was 470 nmol.min^{-1}.mg^{-1}. This activity was lower than SDH from recombinant sources e.g. (Tahara et al., 2021) but comparable with SDH from etiolated pea epicotyls (Balinsky and Davies, 1961), or tomato (*Solanum lycopersicum* L.) fruit (Lourenco and Neves, 1984) or *Cucumis sativus* L. pulp (Lourenco et al., 1991).

3.2. The directions of SDH-catalyzed SA oxidation and DHS reduction is controlled by the pH optimum and differ in maximal reaction rate

SDHs catalyze a reversible reaction, both DHS reduction and SA oxidation. For *P. crispum* SDH, the optimal pH was 9.5 10 for SA oxidation and 7 8.5 for DHS reduction (i.e., shikimate pathway direction) (Table 1). Accordingly, pH likely participates in the regulation of SDH activity *in vivo*. In photosynthetic tissues, the optimal pH 8 of the SDH reaction in the shikimate pathway direction matches the pH of the illuminated chloroplast, which ensures that the reactions of the shikimate pathway coincide with photosynthesis, and thus NADPH and ATP are readily available. Although SDH purified from *P. crispum* root is not a chloroplastic enzyme, SDH regulation by pH in non-photosynthesizing plastids is likely similar. NADPH is provided by the oxidative pentose phosphate pathway, and the optimal pH of glucose-6-phosphate dehydrogenase from barley root plastids is also approximately pH 8 (Esposito et al., 2001). Conversely, SDHs have a non-physiologically high optimal pH in the SA oxidation direction (Avitia-Dominguez et al., 2014; Diaz and Merino, 1997; Guo et al., 2014; Lourenco and Neves, 1984; Lourenco et al., 1991). The unusually high optimal pH of the reverse reaction may be a mechanism of physiological protection against metabolite flowback through the shikimate pathway (Table 1). The pH profiles of SDH suggest that catalysis and substrate binding involve acid/base

157
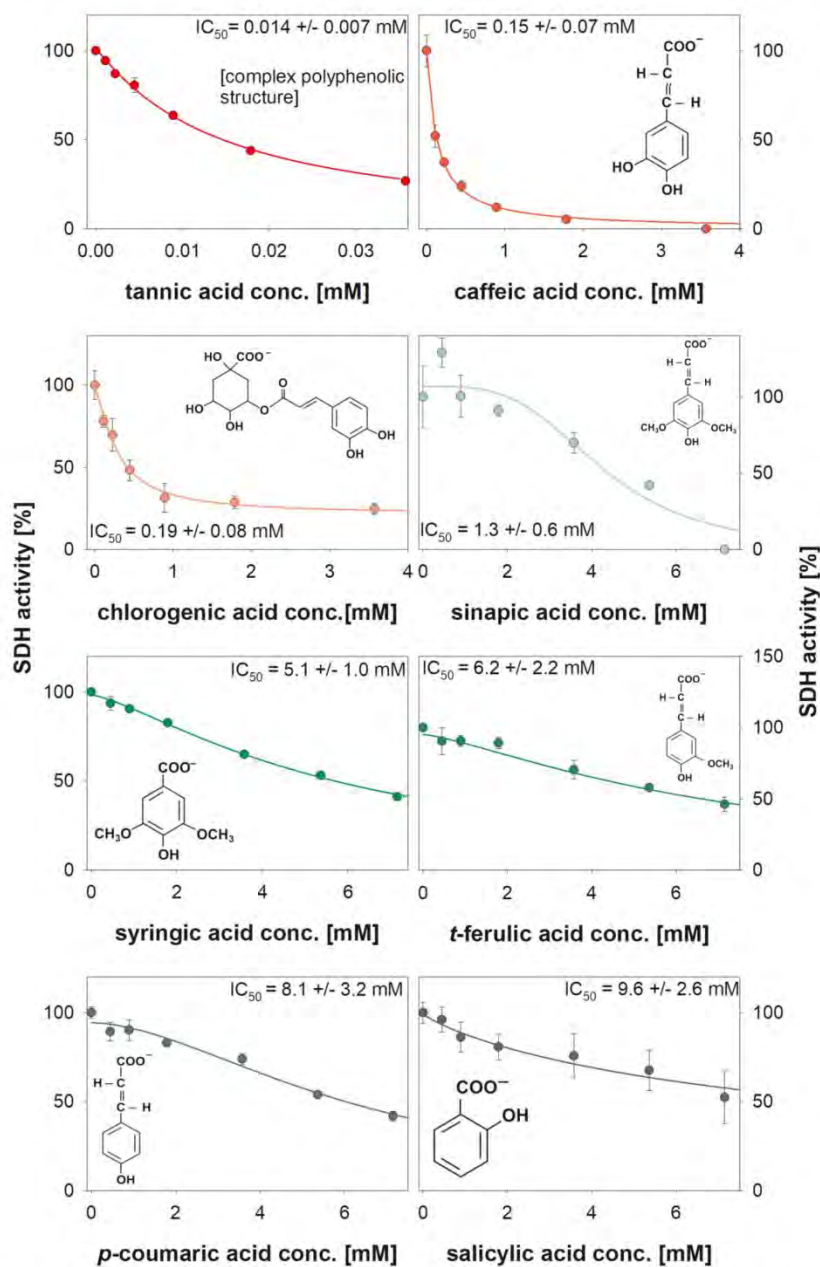
Fig. 5. Inhibition effect of various phenylpropanoids on the activity of *P. crispum* SDH. The specific activity of the enzyme preparations was 0.17 ± 0.07 μmol. min^{-1}mg^{-1}. Controls in 96 and 50 % ethanol were only slightly different from the distilled water control, with specific activities of 0.18 ± 0.06 and 0.19 ± 0.11 μmol. min^{-1}mg^{-1}, respectively. IC$_{50}$ was calculated from nonlinear regression with Eq. (1). Each determination was done at least 3-times, the average values and standard deviations are shown.

Table 4

The effect of various ions on *P. crispum* SDH activity. Screening was performed spectrophotometrically at 340 nm in the presence of 3 mM SA, 0.2 mM NADP and 5 mM*, 0.1 mM**, or 0.01 mM*** concentration of potential modulator. Each determination was done at least 3-times, the average values and S.D. are shown.

Compound	% of activity
$(NH_4)_2SO_4$	115.6 ± 9.8 *
$CaCl_2$	113.2 ± 8.2 *
NaCl	105.2 ± 3.7 *
KCl	107.0 ± 7.3 *
$MgCl_2$	99.3 ± 15.0 *
$MnCl_2$	77.6 ± 23.1 *
$ZnSO_4$	10.1 ± 4.8 **
$CuSO_4$	21.2 ± 8.3 ***

chemistry (Fonseca et al., 2007). The amino acids residues Lys385 and Asp423 (numbered with respect to the *Arabidopsis* protein), located in the SDH active site, which are conserved in DHD/SHD enzymes and which have been proposed to be involved in proton transfer during catalysis, are most likely responsible for such a high optimal pH (Singh and Christendat, 2006).

Furthermore, other factors favor the direction of shikimate pathway: metabolite channeling, which prevents the accumulation of substrates or products (Singh and Christendat, 2006), high Michaelis constants in the direction of SA oxidation (Table 1), product competitive inhibition against NADP and SA (Figs. 2 and 3) and the formation of dead-end complexes (Fig. 3)

The comparison of the maximal reaction rate in both directions shows that the reaction is 4.6 times higher in the non-physiological direction. These data were obtained at the pH optimum of the respective reactions, so they are not likely relevant to cellular context. However, this reaction is often used to study the properties of SDH from various sources.

For many SDHs, the bisubstrate reaction follows a typical sequential mechanism, as found for SDH from other sources, e. g., *Pisum sativum* (Balinsky et al., 1971; Dowsett et al., 1972) or *Capsicum annuum* (Diaz and Merino, 1997) or from *Mycobacterium tuberculosis* (Fonseca et al., 2007). In pea epicotyls, the sequential mechanism of the bisubstrate reaction was further specified to be ordered, with NADPH binding followed by DHS (Balinsky et al., 1971; Dowsett et al., 1972). The mechanism of mycobacterial SDH was also classified as steady-state ordered, albeit with DHS binding first, followed by NADPH (Fonseca et al., 2007).

Bacterial and plant SDHs also differ in their genetic framework in that each enzyme of the shikimate pathway is encoded by monofunctional genes in bacteria. In turn, DHD and SDH are fused to a bifunctional enzyme complex in plants, and even 5 enzymes of the shikimate pathway are fused to a penta-functional complex in fungi (Derrer et al., 2013). *P. crispum* SDH has the typical molecular weight of plant SDHs (Fig. A.2) (Bontpart et al., 2016; Diaz and Merino, 1997; Fiedler and Schultz, 1985; Koshiba, 1978; Lourenco and Neves, 1984; Lourenco et al., 1991; Muir et al., 2011) and is likely a monomer with both domains, DHD and SDH. The initial reaction rate (Fig. 1) and product inhibition pattern (Fig. 2) are in line with an ordered bi-bi mechanism with NADPH or NADP binding to the enzyme first except that SA is a competitive inhibitor with respect to NADPH and DHS is a competitive product inhibitor with respect to SA (Figs. 2 and 3, Table 2). These findings indicate that not only NADPH and NADP but also SA may interact with the free enzyme, most likely in the form of a dead-end complex (enzyme-SA). Both NADP and SA prevent the first substrate, NADPH, from binding to SDH, but DHS has no competitive inhibitor. Therefore, in accordance with an ordered mechanism, DHS is the second substrate. Nevertheless, the enzyme-DHS complex acts as a dead-end complex, similarly to enzyme-NADP-DHS (Fig. 3). Hence, the

formation of dead-end complexes with SA and DHS explains the key role that the NADP/NADPH ratio plays in establishing an equilibrium between the two directions. In *P. crispum*, the shikimate pathway *via* NADPH continues with the phenylpropanoid pathway, specifically with the synthesis of coumarins, where psoralen, bergapten, isoimperatorin, oxypeucedanin, xanthoxin, trioxalen, and angelicin are the most important coumarins in *P. crispum*, or with the synthesis of flavonoids (apiin and luteolin) (Kolarovic et al., 2010). In addition to the shikimate pathway, NADPH is also needed for the antioxidant system and for nitrate assimilation.

3.3. SDH also catalyzes the formation of PCA but not GA and QA

Other SDH-catalyzed reactions that may provide phenolic compounds or intermediates have been reported. Muir et al. (2011) demonstrated that SDH from walnut (*Juglans regia*) catalyzed not only the NADP-dependent dehydrogenation of SA to DHS but also the dehydrogenation of DHS to 3,5-diDHS, which provides GA after enolization. Recombinant EcDQD/SDH2 and 3 enzymes also catalyze NADP-dependent oxidation of 3-DHS to produce gallate, which is in some Eucalyptus species essential for the biosynthesis of the aluminum-detoxifying metabolite (Tahara et al., 2021). Two SDH isoforms showed a similar activity in grape wine berries (Bontpart et al., 2016). However, when analyzing *P. crispum* SDH, we also detected GA in the mixtures without the enzyme (Table 3). Therefore, we confirmed that GA could form non-enzymatically, as found in an in vitro study of recombinant SDH isoforms of Camellia sinensis (Huang et al., 2019). Also, non-enzymatic formation of GA was previously described (Kambourakis and Frost, 2000). Moreover, the reaction mechanism of *P. crispum* SDH showed that the binding of NADP and DHS to the SDH leads to the dead-end complex (Fig. 3). Consequently, enzymatic GA formation is not possible.

We also studied the pathway from DHQ to QA and back, but no products were detected in both directions. Thus, *P. crispum* SDH also belongs to the *bona fide* SDH group, using exclusively NADP as a coenzyme and SA as a substrate (Garcia-Guevara et al., 2017), whereas, e.g., some poplar and *Eucalyptus camaldulensis* SDH isoforms prefer NAD and also use QA as a substrate (Guo et al., 2014; Tahara et al., 2021). In-depth sequencing and *in vitro* biochemical assays showed that 3 poplar enzymes from 5 originally annotated as DHD/SDHs should be rather classified as QA dehydrogenases (Guo et al., 2014). In turn, we found that PCA is a byproduct in the *P. crispum* SDH pathway (Table 3, Fig. 4). As previously hypothesized (Guo et al., 2014), this reaction is ensured by SDH in *P. crispum*

3.4. *P. crispum* root SDH is regulated by metabolites of the phenylpropanoid pathway

Because the shikimate pathway is absent in mammals, searching and designing inhibitors against enzymes of this pathway may lead to the development of antimicrobials (such as the bacterial *Mycobacterium tuberculosis* and *Helicobacter pylori* SDH) and antiparasitic (malaria parasite SDH) and herbicidal (plant SDH) agents, which are harmless to humans (Diaz-Quiroz et al., 2018). There are 3 strategies for identifying compounds with an inhibitory effect on a particular enzyme: i) analyzing substrate structural analogs (Baillie et al., 1972; Diaz and Merino, 1997; Fiedler and Schultz, 1985; Koshiba, 1978; Lemos Silva et al., 1985; Lourenco and Neves, 1984; Lourenco et al., 1991; Rothe, 1974), ii) screening thousands of compounds (Avitia-Dominguez et al., 2014; Han et al., 2006; Peek et al., 2014), and iii) searching for feedback inhibitors among products of the whole pathway. The first strategy has led to the discovery of the herbicide 2,4-dichlorphenoxy acetic acid (2, 4-D) (Diaz and Merino, 1997). Concurrently, several studies have demonstrated that PCA (possible byproduct of SDH) inhibits plant SDH (Diaz and Merino, 1997; Koshiba, 1978; Lemos Silva et al., 1985; Lourenco and Neves, 1984; Lourenco et al., 1991). In this study, we have

shown that *P. crispum* SDH forms PCA in the irreversible reaction (Fig. 4). Using a screening strategy, different research groups have identified SDH inhibitors, for example, 5 novel *Helicobacter pylori* SDH inhibitors, including the natural product curcumin (Han et al., 2006), and polyphenolic inhibitors (epigallocatechin gallate and epicatechin gallate) of *Pseudomonas putida* and *Arabidopsis thaliana* SDH (Peek et al., 2014). A limited number of inhibition/activation studies have identified dihydroxybenzoic acid and its derivatives as SDH inhibitors (Fiedler and Schultz, 1985; Koshiba, 1978; Nandy and Ganguli, 1961), thus showing that SDH inhibitors are not limited to herbicides and organic reagents.

In this study, we chose the third strategy to identify plant SDH inhibitors among the products of the phenylpropanoid pathway (representative compounds of simple phenols, flavonoid, stilbene, and polyphenols). The strongest *P. crispum* SDH inhibitor was tannic acid (Fig. 5). Tannins have strong astringent properties, which may induce complexation with enzymes and substrates (Tintino et al., 2016). They bind to proteins (by hydrophobic, hydrophilic, non-specific, and specific interactions), pigments, low-molecular-weight compounds, and metallic ions (Kato et al., 2017). In microorganisms, interactions between tannic acid and the cell membrane can affect its permeability through the inhibition of the efflux pump, which may be associated with an antimicrobial effect (Tintino et al., 2016). Furthermore, the potentially extracellular localization of tannic acid may contribute to this effect because leaf mesophyll cell walls are the typical site of origin and deposition of hydrolysable tannins in oak leaves (Grundhofer et al., 2001). Furthermore, in the outer peels of pomegranate (*Punica granatum* L.), SDHs play a role in controlling the biosynthesis of hydrolysable tannins (Habashi et al., 2019).

Our results also showed that *P. crispum* SDH is inhibited at 0.15 and 0.19 mM IC_{50} by caffeic acid and chlorogenic acid (with 2 and 5 hydroxyl groups in the structure), respectively. Chlorogenic acids are esters formed between caffeic acid and quinic acid, which are strong antioxidants found in many vegetable species and coffee beans (Colon and Nerin, 2016; Guo et al., 2014; Liang and Kitts, 2015; Niggeweg et al., 2004). In plants, chlorogenic acids serve as protecting compounds against stress, e.g., viral infection (Spoustova et al., 2015), or as feeding deterrents (Ikonen et al., 2001). The *p*-coumaric, *t*-ferulic, sinapic, syringic, and salicylic acids, all with only one hydroxyl group, were milder SDH inhibitors, with IC_{50} above 5 mM, and they are not involved in regulation under physiological conditions. On the other hand, our preliminary results indicate the presence of 0.1 µM *p*-coumaric, 1.7 µM *t*-ferulic, and 0.5 µM chlorogenic acid in *P. crispum* roots (data not shown) and recently Derouich et al. (2020) published a wide scale of phenolic compounds (including chlorogenic acid as the most abundant and then *p*-coumaric, caffeic, gallic, ferulic, vanillic, and syringic acid) in aerial parts of *P. crispum* plants and discussed their high antioxidant power (Derouich et al., 2020). Considering the localization, simple phenolic compounds are probably not stored in plastids in huge amounts, modifications of cinnamic acid to *p*-coumaric, *t*-ferulic, and sinapic acid take place at the membranes of the endoplasmic reticulum, flavonoids are believed to be synthetized in the cytosol and stored in vacuoles (Kitamura, 2006) together with monolignols derivatives, which are synthesized in the cytosol with some enzymes exhibiting membrane attachment and the bulk of the monolignol pool is targeted to the apoplast for polymerization to lignin (Dixon and Barros, 2019). Salicylic acid is an important signal molecule; however, its concentration does not reach the value of the experimentally determined IC_{50}, even during stress (Belonoznikova et al., 2020). In their study, Belinsky and Davies (1961) concluded that the both carbonyl group at the C1 position and a hydroxyl group at the 4-OH position are significant determinants of ligand binding. This is true for syringic acid with IC_{50} 5.1 ± 1.0 mM. Tannic acid contains several hydroxyl groups on phenyl rings; thus, their hydroxyl groups may interact with the amino acid residue in the enzyme active center.

Under non-stress conditions, plant SDH may be inhibited by some phenylpropanoid compounds. In our previous study, we found

significant chlorogenic and quinic acid depletion in tobacco plants exposed to potyviral stress and heat shock (Hyskova et al., 2021). Such a depletion could in turn favor the shikimate pathway, producing precursors of defense compounds by enhancing SDH activity.

Plant SDH inhibition by divalent metal ions, particularly Zn^{2+} and Cu^{2+}, is known and correlated with the inactivation of functional sulfhydryl groups of SDH and also confirmed with the inhibition of plant SDH by *p*-chloromercuribenzoate which could be reversed by cysteine (Balinsky and Davies, 1961; Koshiba, 1978; Lourenco and Neves, 1984). SDH from *P. crispum* root was also inhibited by Zn^{2+} and Cu^{2+}, particularly by Cu^{2+} ions (Table 4).

4. Conclusions

P. crispum root SDH follows an ordered reaction mechanism with three dead-end complexes in both directions. PCA was identified as a side product of SDH, whereas GA was formed non-enzymatically and quinate was not detected. As such, the phenylpropanoid pathway leading to the synthesis of precursors of monolignols (initiating with cinnamic acid and continuing with *p*-coumaric acid, caffeic acid, ferulic acid, 5-hydroxyferulic and sinapic acid) affects SDH activity through feedback regulation. Moreover, their number of hydroxyl groups increases their inhibition effect on SDH. Accordingly, tannic acid, chlorogenic acid, and caffeic acid are the most effective inhibitors of *P. crispum* root SDH, which may be feedback-regulated by phenylpropanoids under stress conditions.

5. Material and methods

5.1. Extraction and purification of SDH

The *P. crispum* convar. radicosum Hill roots were purchased from the farmer company Bramko s.r.o. in the Czech Republic. All the plants were grown in the fields. SDH was purified from 100 g of *P. crispum* root by ammonium sulfate precipitation, ion exchange chromatography on DEAE-cellulose, and gel filtration on Sephacryl S-300, similarly as with NADP-malic enzyme as described previously (Ryšlavá et al., 2007).

5.2. SDH activity assay

SDH activity was monitored spectrophotometrically at 20 °C, following the formation of NADPH during the oxidation of SA into DHS by the increase of absorbance at 340 nm for 5 minutes. The SDH assay mixture (total volume of 1 ml) contained 100 mM AMP-NaOH buffer (pH 9.0), 0.2 mM NADP and 3 mM SA. The reaction was initiated by adding the enzyme (50 µl).

Potential regulatory effects of Zn^{2+} (0.1 mM), Cu^{2+} (0.01 mM); Na^+, K^+, Mg^{2+}, Ca^{2+}, Mn^{2+} and NH_4^+ (5 mM) ions on SDH activity were tested.

5.3. Determination of kinetic parameters and mechanism of the bisubstrate reaction catalyzed by SDH

The reaction rate was measured by the change of absorbance at 340 nm. In 500 mM Tris-HCl buffer (pH 8), the concentration of one of the substrates ranged from 0.005 to 0.7 mM NADPH and from 0.05 to 2 mM DHS, whereas the saturating concentration of the second substrate DHS and NADPH were 2 mM and 0.7 mM, respectively. The reverse reaction was determined similarly but in 500 mM AMP-NaOH buffer (pH 9.0) with 0.05 10 mM NADP and 0.2 10 mM SA, whereas the concentration of the second substrate was at saturation, i.e., 10 mM SA and 3 mM NADP, respectively. Kinetic parameters were determined by fitting the data to the Michaelis-Menten or to an equation characterizing substrate inhibition (Suppl. Eq. A1) and the Michaelis constant (K_m), the maximum reaction rate (V_{max}), the hypothetical maximal reaction rate corresponding to the rate that the enzyme reached in the absence of

9

inhibition by an excess of substrate (V^*_{max}) and the substrate inhibition constant (K_{ss}) were calculated by non-linear regression. All measurements were performed at 20 °C.

The kinetic mechanism of SDH was studied in both directions by varying the concentration of NADP (0.05–0.5 mM) or NADPH (0.05–0.4 mM) and the concentration of SA (0.2–3 mM) or DHS (0.05–2 mM).

The experimental data of the reaction rate as a function of the substrate (SA, NADP) concentration were fit into general reaction rate equations for the bisubstrate reactions, classified as ping—pong and sequential and particular constants were calculated by non-linear regression. Linearization methods, as the double reciprocal plot, Hanes plot or Eadie-Hofstee plot were also used for diagnostic confirmation (Bisswanger, 2002; Marangoni, 2003). V_{max} is the maximal reaction rate of the enzyme reaction, K_{mA} and K_{mB} are the Michaelis constants for the substrate A and B, respectively, and K_A is the dissociation constant for the complex enzyme-substrate. Equations derived for the determination of dead-end complexes are summarized in Appendix A. Supplementary Data as Eq. A.2- A.32.

5.4. Product inhibition study

In the direction from DHS to SA, the reaction mixture contained: 500 mM Tris-HCl buffer (pH 8) and a variable concentration of DHS (0.2; 0.3, 0.5, and 1 mM) or NADPH (0.05; 0.1, 0.2, and 0.3 mM) with a subsaturating concentration of the second substrate, i.e., 0.2 mM NADPH or 0.75 mM DHS. The concentrations of the inhibitors, NADP and SA, were in the 0.1–0.3 mM and 0.5–2 mM ranges, respectively. In this direction, saturating product concentrations could not be measured for technical reasons (NADPH absorbance above 3).

In the direction from SA to DHS, the reaction mixture contained: 375 mM AMP-NaOH buffer (pH 9.0) and a variable concentration of SA and NADP. The final variable concentrations of SA (2; 0.5; 0.3; and 0.15 mM) and NADP (0.5; 0.2; 0.1; and 0.05 mM) were tested at saturating and subsaturating concentrations of the second substrate. The saturating concentrations were 20 mM and 2 mM, and the subsaturating concentrations were 0.3 mM and 0.1 mM for SA and NADP, respectively. The concentrations of the inhibitors, NADPH and DHS, were in the 0.05–0.1 and 0.5–1 mM ranges, respectively.

The data derived from the product inhibition studies were fitted to the equation describing the relevant type of inhibition: competitive and non-competitive (Bisswanger, 2002; Marangoni, 2003), where [S] is the substrate concentration, [I] is the inhibitor concentration, and K_{ic} and K_{iu} are the inhibition constants for the inhibitor derived by slope and intercept, respectively, in the Lineweaver-Burk plot.

5.5. Identification of all SDH products

SDH isolated from *P. crispum* root (10 µl with specific activity 470 nmol.min^{-1}.mg^{-1}) was incubated in the 280 µL reaction mixture containing either 100 mM Tris-HCl pH 8 (in the physiological direction, see Fig. 4) or 100 mM AMP-NaOH pH 9 buffers, 0.14 mM NADP(H)/NAD(H) and 2 mM SA/DHS/DHQ/QA/GA/PCA for 30 min at 20 °C. The reaction was initiated by adding the enzyme and stopped by adding 10 µl of 99% (v/v) formic acid. Blank experiments were prepared similarly, albeit adding formic acid before the enzyme. The calibration standards were purchased from Sigma-Aldrich (USA) and Carl Roth (Germany).

The reaction mixtures were assayed by reversed-phase liquid chromatography coupled to electrospray mass spectrometry (LC-MS). The injection volume was 5 µL (Dionex UltiMate 3000, Thermo Fisher Scientific, USA). A Zorbax C18 reversed-phase silica-based column was used for all separations (150 × 2.1 mm, 3.5 µm, Agilent, USA). The mobile phase consisted of 0.5% (v/v) formic acid in 10% (v/v) acetonitrile in water (eluent A); and 0.3% (v/v) formic acid in 100% (v/v) acetonitrile (eluent B). The flow rate was set to 0.250 ml/min at 37 °C. The following elution program was used: isocratic A for 7 min, 0–100% B for 1 min, isocratic B for 2 min, isocratic A for 3 min. MS analysis was

performed on a ESI-Q-TOF maXis II (Bruker, USA). The MS-conditions were: nitrogen as the nebulizing (2.5 Bar) and drying gas (3.0 l.min^{-1}), dry temperature 220 °C, capillary voltage 3000 V, end plate offset 500 V. All mass spectrometric data were acquired in negative ionization mode.

5.6. Effect of specialized metabolites on the SDH

Specialized metabolites were tested for their potential regulatory effect on SDH activity. The reversibility of inhibition by these compounds was tested simultaneously (Appendix A Supplementary material). Tannic acid was dissolved in distilled water, caffeic acid, chlorogenic acid, *t*-ferulic acid, *p*-coumaric acid, and salicylic acid were dissolved in 96% ethanol, and sinapic and syringic acids were dissolved in 50% ethanol. The final concentration of ethanol did not affect SDH activity. Since phenylpropanoids interfered with the spectrophotometric assay (increasing absorbance at 340 nm), this assay was replaced with the colorimetric microplate method. A reaction mixture (280 µl) consisting of 180 mM AMP-NaOH buffer (pH 9.0), 0.14 mM NADP, 2.1 mM SA, 5 mg/ml iodonitrotetrazolium with 5 µg/ml phenazine methosulfate, and 20 µl of potential inhibitor compound was incubated with 10 µl of enzyme for 15 min. The reaction was terminated with 10 µl of 99% formic acid, measuring the absorbance at 500 nm.

Simultaneously, blank reactions were prepared in which 10 µl of enzyme was first treated with 10 µl of 99% (v/v) formic acid and then with the reaction mixture. Another blank was prepared as the individual reaction (including incubation time), albeit replacing the enzyme with distilled water.

The half maximal inhibitory concentration IC$_{50}$ was calculated from Eq. (1). [I] is the inhibitor concentration, a and b are the maximal and minimal values of reaction rate, respectively, and c corresponds to the slope of the sigmoid function.

$$\mu = a + \frac{c - b}{5 + \left(\frac{[I]}{IC_{50}}\right)} \qquad \text{Eq. (1)}$$

5.7. Determination of protein and total phenolic content

Soluble proteins were determined according to Bradford (1976), and total phenolics compounds were quantified according to Tupec et al. (2017).

5.8. Statistics

SDH isolation was performed in 20 independent biological replicates, repeating each measurement at least three times. Data were analyzed and processed in SigmaPlot v. 12 (Systat Software, Inc.) and Microsoft Excel 2019 (Microsoft Corp.).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.phytochem.2021.112881.

Author contributions

Veronika Hýsková; , Conceptualization, Data curation, Formal analysis, Investigation, Resources, Writing original draft, Kateřina Bělonožníková; , Investigation, Data curation, Methodology, Validation, Formal analysis, Funding acquisition, Writing original draft, Ingrida Šmeringaiová; , Investigation, Daniel Kavan, Investigation, Data curation, and Methodology, Marek Ingr, Software, Validation, Writing review & editing, Helena Ryšlavá; , Supervision, Project administration, Funding acquisition, Resources, Writing review & editing.

References

Avitia-Dominguez, C., Sierra-Campos, E., Salas-Pacheco, J.M., Najera, H., Rojo-Dominguez, A., Cisneros-Martinez, J., Tellez-Valencia, A., 2014. Inhibition and biochemical characterization of methicillin-resistant *Staphylococcus aureus* shikimate dehydrogenase: an in silico and kinetic study. Molecules 19, 4491 4509. https://doi.org/10.3390/molecules19044491.

Baillie, A.C., Corbett, J.R., Dowsett, J.R., McCloskey, P., 1972. Inhibitors of shikimate dehydrogenase as potential herbicides. Pestid. Sci. 3, 113 120. https://doi.org/10.1002/ps.2780030202.

Balinsky, D., Davies, D.D., 1961. Aromatic biosynthesis in higher plants. 1. Preparation and properties of dehydroshikimic reductase. Biochem. J. 80, 292 296. https://doi.org/10.1042/bj0800292.

Balinsky, D., Dennis, A.W., Cleland, W.W., 1971. Kinetic and isotope-exchange studies on shikimate dehydrogenase from *Pisum sativum*. Biochemistry 10, 1947 1952. https://doi.org/10.1021/bi00786a032.

Belonoznikova, K., Vaverova, K., Vanek, T., Kolarik, M., Hyskova, V., Vankova, R., Dobrev, P., Krizek, T., Hodek, O., Cokrtova, K., Stipek, A., Ryslava, H., 2020. Novel insights into the effect of Pythium strains on rapeseed metabolism. Microorganisms 8. https://doi.org/10.3390/microorganisms8101472.

Bisswanger, H., 2002. Enzyme Kinetics. Principles and methods. WILEY-VCH Verlag GmbH, Weinheim.

Bontpart, T., Marlin, T., Vialet, S., Guiraud, J.L., Pinasseau, L., Meudec, E., Sommerer, N., Cheynier, V., Terrier, N., 2016. Two shikimate dehydrogenases, VvSDH3 and VvSDH4, are involved in gallic acid biosynthesis in grapevine. J. Exp. Bot. 67, 3537 3550. https://doi.org/10.1093/jxb/erw184.

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248 254. https://doi.org/10.1016/0003-2697(76)90527-3.

Buchanan, B.B., Gruissem, W., Jones, R.L., 2000. Biochemistry & Molecular Biology of Plants. American Society of Plant Physiologists, Rockville, Maryland.

Cabane, M., Pireaux, J.C., Leger, E., Weber, E., Dizengremel, P., Pollet, B., Lapierre, C., 2004. Condensed lignins are synthesized in poplar leaves exposed to ozone. Plant Physiol. 134, 586 594. https://doi.org/10.1104/pp.103.031765.

Carrington, Y., Guo, J., Le, C.H., Fillo, A., Kwon, J., Tran, L.T., Ehlting, J., 2018. Evolution of a secondary metabolic pathway from primary metabolism: shikimate and quinate biosynthesis in plants. Plant J. 95, 823 833. https://doi.org/10.1111/tpj.13990.

Colon, M., Nerin, C., 2016. Synergistic, antagonistic and additive interactions of green tea polyphenols. Eur. Food Res. Technol. 242, 211 220. https://doi.org/10.1007/s00217-015-2532-9.

Deng, Y., Lu, S., 2017. Biosynthesis and regulation of phenylpropanoids in plants. Crit. Rev. Plant Sci. 36, 257 290. https://doi.org/10.1080/07352689.2017.1402852.

Derouich, M., Bouhlali, E.D.T., Bammou, M., Hmidani, A., Sellam, K., Alem, C, 2020. Bioactive compounds and antioxidant, antiperoxidative, and antihemolytic properties investigation of three *Apiaceae* species grown in the Southeast of Morocco. Hindawi Scientifica. https://doi.org/10.1155/2020/3971041.

Derrer, B., Macheroux, P., Kappes, B., 2013. The shikimate pathway in apicomplexan parasites: implications for drug development. Front. Biosci. 18, 944 969. https://doi.org/10.2741/4155.

Diaz-Quiroz, D.C., Cardona-Felix, C.S., Viveros-Ceballos, J.L., Reyes-Gonzalez, M.A., Bolivar, F., Ordonez, M., Escalante, A., 2018. Synthesis, biological activity and molecular modelling studies of shikimic acid derivatives as inhibitors of the shikimate dehydrogenase enzyme of *Escherichia coli*. J. Enzym. Inhib. Med. Chem. 33, 397 404. https://doi.org/10.1080/14756366.2017.1422125.

Diaz, J., Merino, F., 1997. Shikimate dehydrogenase from pepper (*Capsicum annuum*) seedlings. Purification and properties. Physiol. Plantarum 100, 147 152. https://doi.org/10.1111/j.1399-3054.1997.tb03465.x.

Dixon, R.A., Barros, J., 2019. Lignin biosynthesis: old roads revisited and new roads explored. Open Biol. 9 https://doi.org/10.1098/rsob.190215.

Dowsett, J.R., Middleton, B., Corbett, J.R., Tubbs, P.K., 1972. The anomalous inhibition of shikimate dehydrogenase by analogues of dehydroshikimate. Biochim. Biophys. Acta 276, 344 349. https://doi.org/10.1016/0005-2744(72)90994-1.

Esposito, S., Carfagna, S., Massaro, G., Vona, V., Di Martino Rigano, V., 2001. Glucose-6-phosphate dehydrogenase in barley roots: kinetic properties and localisation of the isoforms. Planta 212, 627 634. https://doi.org/10.1007/s004250000443.

Esposito, S., Massaro, G., Vona, V., Di Martino Rigano, V., Carfagna, S., 2003. Glutamate synthesis in barley roots: the role of the plastidic glucose-6-phosphate dehydrogenase. Planta 216, 639 647. https://doi.org/10.1007/s00425-002-0892-4.

Fiedler, E., Schultz, G., 1985. Localization, purification, and characterization of shikimate oxidoreductase-dehydroquinate hydrolyase from stroma of spinach chloroplasts. Plant Physiol. 79, 212 218. https://doi.org/10.1104/pp.79.1.212.

Fonseca, I.O., Silva, R.G., Fernandes, C.L., de Souza, O.N., Basso, L.A., Santos, D.S., 2007. Kinetic and chemical mechanisms of shikimate dehydrogenase from *Mycobacterium tuberculosis*. Arch. Biochem. Biophys. 457, 123 133. https://doi.org/10.1016/j.abb.2006.11.015.

Garcia-Guevara, F., Bravo, I., Martinez-Anaya, C., Segovia, L., 2017. Cofactor specificity switch in shikimate dehydrogenase by rational design and consensus engineering. Protein Eng. Des. Sel. 30, 533 541. https://doi.org/10.1093/protein/gzx031.

Grundhofer, P., Niemetz, R., Schilling, G., Gross, G.G., 2001. Biosynthesis and subcellular distribution of hydrolyzable tannins. Phytochemistry 57, 915 927. https://doi.org/10.1016/s0031-9422(01)00099-1.

Guo, J., Carrington, Y., Alber, A., Ehlting, J., 2014. Molecular characterization of quinate and shikimate metabolism in *Populus trichocarpa*. J. Biol. Chem. 289, 23846 23858. https://doi.org/10.1074/jbc.M114.558536.

Habashi, R., Hacham, Y., Dhakarey, R., Matityahu, I., Holland, D., Tian, L., Amir, R., 2019. Elucidating the role of shikimate dehydrogenase in controlling the production of anthocyanins and hydrolysable tannins in the outer peels of pomegranate. BMC Plant Biol. 19, 476 491. https://doi.org/10.1186/s12870-019-2042-1.

Han, C., Wang, L., Yu, K., Chen, L., Hu, L., Chen, K., Jiang, H., Shen, X., 2006. Biochemical characterization and inhibitor discovery of shikimate dehydrogenase from *Helicobacter pylori*. FEBS J. 273, 4682 4692. https://doi.org/10.1111/j.1742-4658.2006.05469.x.

Heldt, H.-W., Piechulla, B., Heldt, F., 2011. Plant Biochemistry-Translation of the, 4th German edition. Elsevier Academic Press, Amsterdam, Boston, Heildelberg, London, New York, Oxford, Paris, San Diego, San Francisco, Singapore, Sydney, Tokyo.

Huang, K., Li, M., Liu, Y., Zhu, M., Zhao, G., Zhou, Y., Zhang, L., Wu, Y., Dai, X., Xia, T., Gao, L., 2019. Functional analysis of 3-dehydroquinate dehydratase/shikimate dehydrogenases involved in shikimate pathway in Camellia sinensis. Front. Plant Sci. 10 https://doi.org/10.3389/fpls.2019.01268.

Hyskova, V., Belonoznikova, K., Doricova, V., Kavan, D., Gillarova, S., Henke, S., Synkova, H., Ryslava, H., Cerovska, N., 2021. Effects of heat treatment on metabolism of tobacco plants infected with Potato virus. Y. Plant Biol. https://doi.org/10.1111/plb.13234.

Hyskova, V., Pliskova, V., Cerveny, V., Ryslava, H., 2017. NADP-dependent enzymes are involved in response to salt and hypoosmotic stress in cucumber plants. Gen. Physiol. Biophys. 36, 247 258. https://doi.org/10.4149/gpb_2016053.

Ikonen, A., Tahvanainen, J., Roininen, H., 2001. Chlorogenic acid as an antiherbivore defence of willows against leaf beetles. Entomol. Exp. Appl. 99, 47 54. https://doi.org/10.1046/j.1570-7458.2001.00800.x.

Kambourakis, S., Frost, J.W., 2000. Synthesis of gallic acid: Cu(2+)-mediated oxidation of 3-dehydroshikimic acid. J. Org. Chem. 65, 6904 6909. https://doi.org/10.1021/jo000335z.

Kato, C.G., Goncalves, G.A., Peralta, R.A., Seixas, F.A.V., de Sa-Nakanishi, A.B., Bracht, L., Comar, J.F., Bracht, A., Peralta, R.M., 2017. Inhibition of alpha-amylases by condensed and hydrolysable tannins: focus on kinetics and hypoglycemic actions, 2017 Enzym. Res.. https://doi.org/10.1155/2017/5724902.

Kitamura, S., 2006. Transport of flavonoids from cytosolic synthesis to vacuolar accumulation. In: Grotewold, E. (Ed.), The Science of Flavonoids. Springer, New York, pp. 123 146.

Kolarovic, J., Popovic, M., Zlinska, J., Trivic, S., Vojnovic, M., 2010. Antioxidant activities of celery and parsley juices in rats treated with doxorubicin. Molecules 15, 6193 6204. https://doi.org/10.3390/molecules15096193.

Koshiba, T., 1978. Purification of two forms of the associated 3-dehydroquinate hydro-lyase and shikimate:NADP+ oxidoreductase in *Phaseolus mungo* seedlings. Biochim. Biophys. Acta 522, 10 18. https://doi.org/10.1016/0005-2744(78)90317-0.

Kovacik, J., Klejdus, B., Backor, M., 2009. Phenolic metabolism of *Matricaria chamomilla* plants exposed to nickel. J. Plant Physiol. 166, 1460 1464. https://doi.org/10.1016/j.jplph.2009.03.002.

Lemos Silva, G.M., Lourenco, E.J., Neves, V.A., 1985. Inhibition of shikimate dehydrogenase from heart-of-palm (*Euterpe oleracea* Mart.). J. Food Biochem. 9, 105 116. https://doi.org/10.1111/j.1745-4514.1985.tb00342.x.

Liang, N., Kitts, D.D., 2015. Role of chlorogenic acids in controlling oxidative and inflammatory stress conditions. Nutrients 8. https://doi.org/10.3390/nu8010016.

Lourenco, E.J., Neves, V.A., 1984. Partial purification and some properties of shikimate dehydrogenase from tomatoes. Phytochemistry 23, 497 499. https://doi.org/10.1016/S0031-9422(00)80366-0.

Lourenco, E.J., Silva, G.M., Neves, V.A., 1991. Purification and properties of shikimate dehydrogenase from cucumber (*Cucumis sativus* L.). J. Agric. Food Chem. 39, 458 462. https://doi.org/10.1021/jf00003a006.

Marangoni, A.G., 2003. Enzyme kinetics. A modern approach. John Wiley & Sons, Ltd, Chichester.

Moura, J.C., Bonine, C.A., de Oliveira Fernandes Viana, J., Dornelas, M.C., Mazzafera, P., 2010. Abiotic and biotic stresses and changes in the lignin content and composition in plants. J. Integr. Plant Biol. 52, 360 376. https://doi.org/10.1111/j.1744-7909.2010.00892.x.

Muir, R.M., Ibanez, A.M., Uratsu, S.L., Ingham, E.S., Leslie, C.A., McGranahan, G.H., Batra, N., Goyal, S., Joseph, J., Jemmis, E.D., Dandekar, A.M., 2011. Mechanism of gallic acid biosynthesis in bacteria (*Escherichia coli*) and walnut (*Juglans regia*). Plant Mol. Biol. 75, 555 565. https://doi.org/10.1007/s11103-011-9739-3.

Nandy, M., Ganguli, N.C., 1961. Studies on 5-dehydroshikimic reductase from mung bean seedlings (*Phaseolus aureus*). Arch. Biochem. Biophys. 92, 399 408. https://doi.org/10.1016/0003-9861(61)90378-2.

Niggeweg, R., Michael, A.J., Martin, C., 2004. Engineering plants with increased levels of the antioxidant chlorogenic acid. Nat. Biotechnol. 22, 746 754. https://doi.org/10.1038/nbt966.

Peek, J., Shi, T., Christendat, D., 2014. Identification of novel polyphenolic inhibitors of shikimate dehydrogenase (AroE). J. Biomol. Screen 19, 1090 1098. https://doi.org/10.1177/1087057114527127.

Rothe, G.M., 1974. Intracellular compartmentation and regulation of two shikimate dehydrogenase isoenzymes in *Pisum sativum*. Z. Pflanzenphysiol. Bd. 74, 152 159. https://doi.org/10.1016/S0044-328X(74)80168-6.

Ryšlavá, H., Doubnerová, V., Müller, K., Baťková, P., Schnablová, R., Liberda, J., Synková, H., Čeřovská, N., 2007. The enzyme kinetics of the NADP-malic enzyme from tobacco leaves. Czech. Chem. Commun. 72, 1420 1434. https://doi.org/10.1135/cccc20071420.

Singh, S.A., Christendat, D., 2006. Structure of Arabidopsis dehydroquinate dehydratase-shikimate dehydrogenase and implications for metabolic channeling in the shikimate pathway. Biochemistry 45, 7787 7796. https://doi.org/10.1016/S0044-328X(74)80168-6.

Spoustova, P., Hyskova, V., Muller, K., Schnablova, R., Ryslava, H., Cerovska, N., Malbeck, J., Cvikrova, M., Synkova, H., 2015. Tobacco susceptibility to *Potato virus Y*-NTN infection is affected by grafting and endogenous cytokinin content. Plant Sci. 235, 25 36. https://doi.org/10.1016/j.plantsci.2015.02.017.

Tahara, K., Nishiguchi, M., Funke, E., Miyazawa, S.-I., Miyama, T., Milkowski, C., 2021. Dehydroquinate dehydratase/shikimate dehydrogenases involved in gallate biosynthesis of the aluminum-tolerant tree species *Eucalyptus camaldulensis*. Planta 253. https://doi.org/10.1007/s00425-020-03516-w.

Tintino, S.R., Oliveira-Tintino, C.D., Campina, F.F., Silva, R.L., Costa Mdo, S., Menezes, I. R., Calixto-Junior, J.T., Siqueira-Junior, J.P., Coutinho, H.D., Leal-Balbino, T.C., Balbino, V.Q., 2016. Evaluation of the tannic acid inhibitory effect against the NorA efflux pump of *Staphylococcus aureus*. Microb. Pathog. 97, 9 13. https://doi.org/10.1016/j.micpath.2016.04.003.

Tupec, M., Hyskova, V., Belonoznikova, K., Hranicek, J., Cerveny, V., Ryslava, H., 2017. Characterization of some potential medicinal plants from Central Europe by their antioxidant capacity and the presence of metal elements. Food Biosci 20, 43 50. https://doi.org/10.1016/j.fbio.2017.08.001.

Tzin, V., Galili, G., 2010. New insights into the shikimate and aromatic amino acids biosynthesis pathways in plants. Mol. Plant 3, 956 972. https://doi.org/10.1093/mp/ssq048.

163

P9. Sensitive Versatile Fluorogenic Transmembrane Peptide Substrates for Rhomboid Intramembrane Proteases.

A. Tichá, S. Stanchev, J. Škerle, J. Began, M. Ingr, K. Švehlová, L. Polovinkin, M. Růžička, L. Bednárová, R. Hadravová, E. Poláchová, P. Rampírová, J. Březinová, V. Kašička, P. Majer, K. Strisovsky, J. Biol. Chem. 292 (2017) 2703–2713.

Sensitive Versatile Fluorogenic Transmembrane Peptide Substrates for Rhomboid Intramembrane Proteases*[S]

Anežka Tichá[‡§1], Stancho Stanchev[‡1], Jan Škerle[‡¶2], Jakub Began[‡||3], Marek Ingr[¶**4], Kateřina Švehlová[‡], Lucie Polovinkin[‡¶5], Martin Růžička[‡¶6], Lucie Bednárová[‡], Romana Hadravová[‡], Edita Poláchová[‡], Petra Rampírová[‡], Jana Březinová[‡], Václav Kašička[‡6], Pavel Majer[‡], and Kvido Strisovsky[‡7]

From the ‡Institute of Organic Chemistry and Biochemistry of the Czech Academy of Science, Flemingovo n. 2, Prague 166 10, the ¶Department of Biochemistry, Faculty of Science, Charles University, Hlavova 2030/8, Prague 128 43, the ||Department of Genetics and Microbiology, Faculty of Science, Charles University, Viničná 5, Prague 128 44, the §First Faculty of Medicine, Charles University, Kateřinská 32, Prague 121 08, and the **Department of Physics and Materials Engineering, Tomas Bata University in Zlín, Faculty of Technology, nám. T.G. Masaryka 5555, 76001, Zlín, Czech Republic

Edited by George N. DeMartino

Rhomboid proteases are increasingly being explored as potential drug targets, but their potent and specific inhibitors are not available, and strategies for inhibitor development are hampered by the lack of widely usable and easily modifiable *in vitro* activity assays. Here we address this bottleneck and report on the development of new fluorogenic transmembrane peptide substrates, which are cleaved by several unrelated rhomboid proteases, can be used both in detergent micelles and in liposomes, and contain red-shifted fluorophores that are suitable for high-throughput screening of compound libraries. We show that nearly the entire transmembrane domain of the substrate is important for efficient cleavage, implying that it extensively interacts with the enzyme. Importantly, we demonstrate that in the detergent micelle system, commonly used for the enzymatic analyses of intramembrane proteolysis, the cleavage rate strongly depends on detergent concentration, because the reaction proceeds only in the micelles. Furthermore, we show that the catalytic efficiency and selectivity toward a rhomboid substrate can be dramatically improved by targeted modification of the sequence of its P5 to P1 region. The fluorogenic substrates that we describe and their sequence variants should find wide use in the detection of activity and development of inhibitors of rhomboid proteases.

Rhomboid proteases are increasingly being explored as potential drug targets, but their potent and specific inhibitors are not available, and strategies for inhibitor development are hampered by the lack of widely usable and easily modifiable *in vitro* activity assays.

Rhomboid intramembrane proteases are evolutionarily widespread and regulate important biological processes including growth factor secretion (1, 2), mitochondrial dynamics (3), invasion of the malaria parasite (4), and membrane protein quality control (5). Rhomboid proteases are increasingly being explored as potential drug targets (6–9), but their selective and potent inhibitors are lacking (reviewed in Ref. 10). Rhomboid inhibitor discovery and development are complicated by the lack of widely usable and easily modifiable *in vitro* activity assays.

Rhomboid activity assays have traditionally relied on recombinant transmembrane protein substrates and gel-based readouts, but such assays are unsuitable for high-throughput screening. A fluorogenic substrate for the *Providencia stuartii* rhomboid protease AarA lacking most of the transmembrane domain of the parent substrate Gurken is cleaved very poorly by other rhomboids including the main model rhomboid protease GlpG of *Escherichia coli* (11). Other published variants of fluorogenic substrates can be used only in liposomes (12) or involve large fluorescent protein moieties making them dependent on expression in a biological system and photochemically less variable (13), which may be important for high-throughput screening of compound libraries where bright red-shifted fluorophores are preferred (14). Moreover, each of the described rhomboid substrates has been used only with one or two related rhomboid proteases, and a strategy to design widely usable or specific substrates has been lacking. Other types of activity assays employing MALDI mass spectrometry (15) and fluorescence polarization (16) have been reported, but MALDI is a low-throughput method that requires sophisticated instrumentation, and fluorescence polarization assays are based on competition of small molecular activity probes with inhibitors and are prone to detergent artifacts (16), making both of these methods unfit for routine kinetics measurements or high-throughput screening.

In view of these limitations, we have sought to develop a robust fluorogenic transmembrane peptide substrate platform for continuous activity assays that would capture all the native enzyme-substrate interactions, be applicable to both the detergent micelle system and liposomes, and would be easily adapt-

164
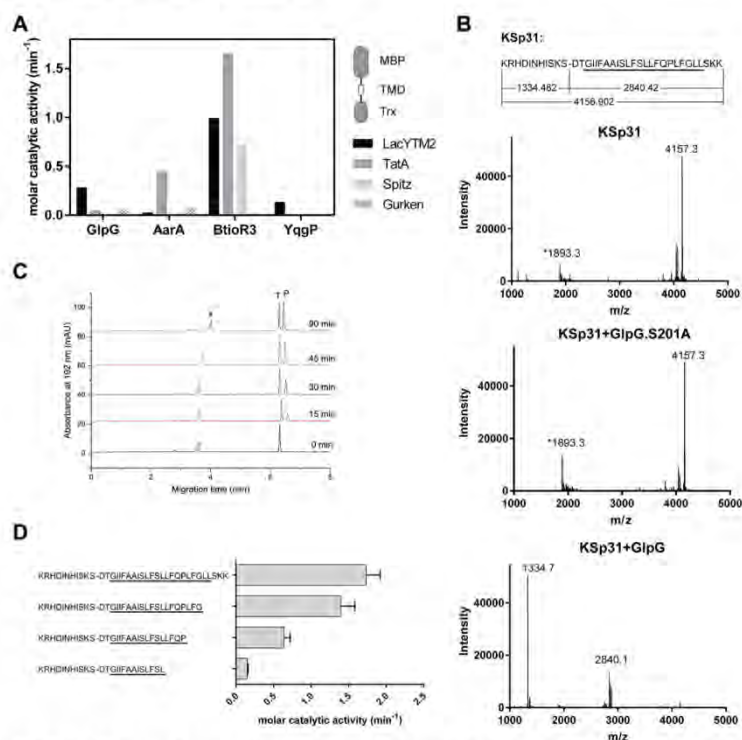
FIGURE 1. **Identification of a widely accepted transmembrane substrate for rhomboid proteases.** *A,* comparison of cleavage efficiency of model substrates LacYTM2, Gurken, TatA, and Spitz by bacterial rhomboid proteases GlpG (*E. coli*), AarA (*P. stuartii*), YqgP (*B. subtilis*), and BtioR3 (*B. thetaiotaomicron*) *in vitro.* Equal concentrations of purified recombinant substrates were exposed to purified recombinant rhomboid proteases. Cleavage products were separated by SDS-PAGE, stained, and quantified densitometrically to determine initial reaction rates, which were converted to molar catalytic activities to allow comparisons. Displayed values are representative of two independent experiments. *B,* cleavage of synthetic LacYTM2 transmembrane peptide KSp31 by GlpG. Purified synthetic peptide KSp31 was incubated with purified recombinant GlpG or its inactive mutant S201T in the presence of 0.05% (w/v) DDM, and the reaction mixtures were analyzed by MALDI mass spectrometry. The theoretical molecular masses of the expected cleavage products at the native cleavage site are denoted *below* the peptide sequence, and unambiguously match those experimentally determined and displayed in the mass spectra. The *star*-marked peak with molecular mass of 1893.3 is an unidentified minor contaminant in the preparation of KSp31. *C,* monitoring of cleavage of peptide substrate KSp31 by rhomboid protease GlpG using CE. The N-terminal cleavage product (P) of KSp31 was separated by free-flow CE in the background electrolyte composed of 100 mM H$_3$PO$_4$ and 69 mM Tris, pH 2.5, in bare fused silica capillary at separation voltage +25 kV. Samples for CE were prepared by mixing 20 μl of reaction mixture at selected reaction times (0–90 min) with 2 μl of 2.2 mM tyramine (T) as an internal standard. Samples were injected into the capillary by 20 mbar pressure for 10 s. Quantitative analysis was based on the ratio of corrected (migration time normalized) peak areas of peptides of interest and the internal standard. Analyses were performed in triplicate. *P,* cleaved N-terminal peptide; *X,* system peak. *D,* the importance of the transmembrane domain of the substrate for its recognition and cleavage by rhomboid. A series of synthetic peptides covering LacYTM2 with progressive truncations of its transmembrane domain from the C terminus was exposed to GlpG and initial rates of cleavage were quantified by capillary electrophoresis as denoted in *panel C.*

able to diverse rhomboid proteases. Because solid phase synthesis of transmembrane peptides and their purification are non-trivial, and their solution behavior often unpredictable, we place emphasis on choosing a robust system and characterizing it thoroughly, and present a generalizable framework for rhomboid substrate design.

Results and Discussion

LacYTM2 Is a Widely Accepted Rhomboid Substrate—To identify a substrate widely accepted by diverse rhomboid proteases, we have measured the efficiency of cleavage of four common model rhomboid substrate transmembrane domains (*P. stuartii* TatA, *Drosophila melanogaster* Gurken and Spitz, and *E. coli* LacYTM2) embedded in a chimeric construct by

four unrelated rhomboid proteases (*E. coli* GlpG, *Bacillus subtilis* YqgP, *P. stuartii* AarA, and *Bacteroides thetaiotaomicron* rhomboid 3 (BtioR3)) (Fig. 1*A*). Comparison of the efficiencies of cleavage (molar catalytic activities) revealed that the substrate containing the second transmembrane (TM)[8] helix of *E. coli* LacY protein (LacYTM2) (17) was the most "promiscuous" substrate.

Although it is well accepted that the region around the scissile bond, mainly P4 to P2', is key for the turnover efficiency of

---

[8] The abbreviations used are: TM, transmembrane; DDM, *n*-dodecyl-β-D-maltopyranoside; DM, *n*-decyl-β-D-maltopyranoside; CMC, critical micellar concentration; LUV, large unilamellar vesicles ; MBP, maltose-binding protein; TAMRA, tetramethylrhodamine; CE, capillary electrophoresis; BGE, background electrolyte.
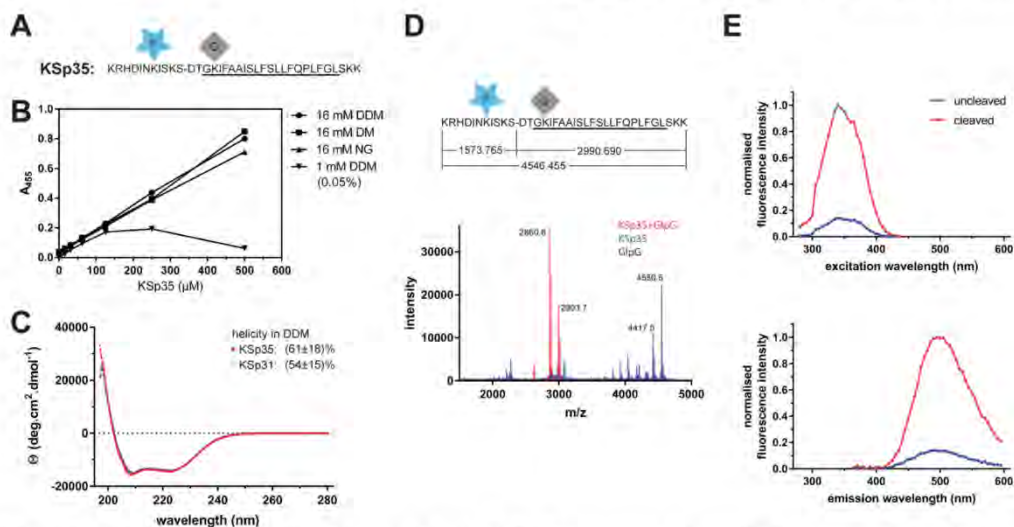
FIGURE 2. **Fluorogenic transmembrane peptide substrate based on LacYTM2.** *A*, fluorogenic variant of the LacYTM2 transmembrane helix-derived peptide (KSp31) with the P5 and P4' positions replaced by Glu-EDANS and Lys-DABCYL, respectively, yielding fluorogenic substrate KSp35. *B*, solubility of KSp35 in 16 mM detergents DDM, DM, and nonyl glucoside (*NG*) and at 1 mM DDM. Note that the concentration of DDM micelles is about 100 μM at 16 mM DDM and about 10 μM at 1 mM DDM. The peptide was dissolved to the indicated concentration by dilution from a 10 mM stock solution in DMSO, and after a 2-h incubation at 37 °C the solution was centrifuged at 21,130 $\times$ *g* for 20 min. The absorbance of the supernatant at 455 nm indicated the concentration of the chromophore in solution. *C*, circular dichroism spectra of LacYTM2-derived transmembrane peptide KSp31 and its fluorogenic variant KSp35 in detergent micelles. Peptides were reconstituted into 0.5% (w/v) DDM to 135 μM (KSp31) and 82 μM (KSp35) concentrations. The spectra show similarly significant helical content for both peptides. *D*, identification of the cleavage site in KSp35 by GlpG. Purified 95 μM KSp35 was incubated with 26 μM GlpG for 20 h and analyzed by MALDI. The *red peak* of the mass of 2993.7 corresponds well to the expected size of the C-terminal cleavage product of 2990.690. The second peak lower by 130 Da is visible in both the *blue* and *red traces* is probably a deletion product of chemical synthesis lacking a C-terminal lysine. This variant has proven difficult to purify away, but it is cleaved by GlpG and probably does not influence the kinetics properties of the substrate significantly (see Fig. 1*D*). *E*, excitation and emission spectra of KSp35 and their change upon cleavage by rhomboid GlpG measured in detergent micelles. The spectra of 10 μM KSp35 substrate in reaction buffer (20 mM HEPES, pH 7.4, 150 mM NaCl, 0.05% (w/v) DDM, 10% (v/v) DMSO) were measured at 37 °C. Excitation wavelengths ranged from 235 to 435 nm with a 10-nm increment and the emission was measured at 493 nm. The emission wavelengths ranged from 365 to 595 nm with a 10-nm increment and excitation at 335 nm.

rhomboid substrates (12, 18), the role of the TM domain of the substrate for recognition and catalysis by rhomboid is less well understood. We have thus next evaluated the importance of the transmembrane region of LacYTM2 for the recognition by *E. coli* GlpG, the main model rhomboid protease, by synthesizing a peptide covering the whole transmembrane region and adjacent juxtamembrane segments of LacYTM2, and a series of its C terminally truncated variants. The full-length LacYTM2 transmembrane peptide KSp31 was cleaved by GlpG efficiently and highly specifically at the expected Ser-Asp cleavage site (Fig. 1*B*). The kinetics of cleavage were monitored by capillary electrophoresis (Fig. 1*C*). The cleavage rate decreased significantly upon truncating the TM helix of LacYTM2 peptide by more than 5 amino acids from the C terminus (Fig. 1*D*), suggesting that most of the TM domain of the substrate is important for the interaction with and recognition by rhomboid. Thus, to develop a widely accepted fluorogenic substrate that would faithfully mimic all the relevant enzyme-substrate interactions including the intramembrane ones, we have used the full-length LacYTM2 transmembrane domain peptide KSp31 as a starting point.

Fluorogenic Transmembrane Peptide Substrate Based on LacYTM2, Basic Properties—To generate a fluorogenic variant of the LacYTM2 peptide, we have replaced the P5 and P4' positions in KSp31 by Glu-EDANS and Lys-DABCYL to yield

KSp35 (Fig. 2*A*). Previously published mutagenic analyses show that these positions are not critical for recognition by rhomboid (18, 19), and they are sufficiently close for Förster resonance energy transfer (FRET) to occur. The KSp35 peptide was soluble up to 500 μM (Fig. 2*B*) in frequently used detergents at 16 mM decyl maltoside (DM), nonyl glucoside (NG), and dodecyl maltoside (DDM). At a total DDM concentration of 16 mM (0.82%(w/v)), the concentration of micelles is about 110 μM, suggesting a partitioning ratio of more than 1 molecule of the substrate per micelle. When DDM was kept at only 1 mM (0.05% (w/v)) total concentration, which yields about 6–10 μM micelles, the solubility of KSp35 became limited to about 100 μM (Fig. 2*B*), indicating that the upper limit of the partitioning ratio is about 10–20 molecules of KSp35 per DDM micelle. The solubility of KSp35 in the absence of detergent was negligible (not shown). Circular dichroism of KSp35 in 0.5% (w/v) DDM (Fig. 2*C*) showed a significant content of α-helical structure (61 $\pm$ 18%), which is consistent with the transmembrane character of the peptide and comparable with the helical content of the parent peptide KSp31 (54 $\pm$ 15%). Cleavage of KSp35 by GlpG occurred at the expected cleavage site (Fig. 2*D*), and was accompanied by an increase in fluorescence at 495 nm (Fig. 2*E*), demonstrating that FRET between the donor and acceptor is occurring in the uncleaved peptide. Collectively, the above results show that KSp35 is a realistic model reflecting all the

166

important interactions between a rhomboid protease and its transmembrane substrate.

Kinetic Characterization of the LacYTM2-based Substrate KSp35 in Detergent Micelle System—In the detergent-solubilized state, most commonly used to study the biochemistry of intramembrane proteolysis, the reaction catalyzed by rhomboid protease occurs in detergent micelles due to the hydrophobicity of both enzyme and substrate. The system is thus microheterogeneous, the effective concentrations of the reactants depend on the volume of the micellar milieu and on the partitioning of reaction components between free solution and the micelles. To characterize the kinetic behavior of the new fluorogenic transmembrane substrates in light of these features of the micellar system, steady-state kinetics was measured with 10 μM substrate, 0.4 μM enzyme, and 0.05% (w/v) DDM, always keeping the concentrations of two components constant and varying the third one around the stated values. At 0.05% (w/v) DDM, the concentration of detergent monomers is 980 μM and micelle concentration about 6–10 μM, calculated assuming critical micellar concentration (CMC) of 0.17 mM (20) and aggregation number between 78 and 149 (20). The molar ratio of enzyme:substrate:micelles is thus 4:100:60–100. In these conditions, assuming that all the reaction partners are evenly distributed among micelles, the average number of substrate molecules per micelle is about 1.5, and only up to 4% of micelles carry an enzyme molecule (micelles containing more than one enzyme molecule are strongly improbable).

The cleavage reactions were started by either mixing two preheated solutions containing substrate or enzyme preincubated with detergent, or adding the DMSO-dissolved substrate into the rest of the preheated reaction mixture. In either case, progress curves are linear from the beginning, which implies that the redistribution of the adsorbed molecules among the micelles is significantly faster than substrate cleavage itself. In accordance with this, the reaction rate is proportional to enzyme concentration within the 0–0.6 μM range (Fig. 3A). Within this concentration range, few enzyme molecules are randomly distributed among many more micelles, providing in principle equal conditions for each enzyme molecule. A similar principle can also explain the observation that the dependence of the reaction rate on substrate concentration is linear in the 0–4 μM range (Fig. 3B). At the upper limit of 4 μM substrate, all micelles can be populated by one (or less likely more) substrate molecule, the linear dependence, furthermore, suggests that this substrate concentration is still below the apparent Michaelis constant of this process.

An important phenomenon is observed when the dependence of the initial rate on detergent concentration is measured. At concentrations above the CMC, the reaction rate rapidly decreases as DDM concentration grows (Fig. 3C), without an obvious impact on the secondary structure content of GlpG (Fig. 3D), suggesting that the effect is caused primarily by the increase in the volume of the micellar phase and consequent decrease of the effective concentrations of both substrate and enzyme. Indeed, mathematical consideration suggests that when substrate and enzyme concentrations are significantly lower than the concentration of micelles (*i.e.* at high DDM concentrations), the probability of location of a substrate molecule

on the same micelle as the enzyme molecule is inversely proportional to the concentration of DDM. Under these conditions, the fraction of substrate-occupied micelles, f_{SM}, is equal to the ratio of the numbers of substrate molecules, $n(S)$, and micelles $n(M)$.

$$f_{SM} = n(S)/n(M) \qquad \text{(Eq. 1)}$$

The mean number of micelles occupied by both the enzyme and substrate molecules, $n(ESM)$, is then given by this fraction multiplied by the number of enzyme molecules $n(E)$.

$$n(ESM) = f_{SM} \times n(E) = n(S) \times n(E)/n(M) \qquad \text{(Eq. 2)}$$

Hence, when the DDM concentration is increased at constant $n(S)$ and $n(E)$, then $n(ESM)$ reflecting the reaction rate decreases in accord with the growing value of $n(M)$. This causes the proportional decrease of the reaction rate (in other words, the reaction rate is proportional to $[DDM]^{-1}$). To inspect whether this model is correct, one can conveniently determine the power of the measured rate dependence on DDM concentration by taking a logarithm of the data from Fig. 3C (log $a^n = n \times$ log a). The logarithmic plot (Fig. 3C, *open circles, right* and *upper axes*) can be satisfyingly ($R^2 = 0.9974$) fitted by a second-order polynomial, yielding equation: $y = -0.1436x^2 - 0.3906x + 2.8852$, whose derivative $y' = -0.2872x - 0.3906$ indicates the power of DDM concentration on which the reaction rate depends. This analysis shows that for high DDM concentrations the derivative indeed tends to -1 (for $x = 2$, $y' = -0.965$; thus rate $\sim[DDM]^{-1}$), which is in accordance with the above assumption, whereas for the lower end of DDM concentrations the absolute value of the power decreases (for $x = 0$, $y' = -0.3906$; thus rate $\sim[DDM]^{-0.4}$). This is consistent with a model that upon decreasing the detergent concentration (while still being above the CMC), the density of the adsorbed molecules in the micellar phase increases, whereas total concentration of micelles decreases, which leads to less frequent collisions between them and thus less effective redistribution of the adsorbed molecules among the micelles. Possibly, the redistribution efficiency might also be insufficient because of the higher reaction rate caused by the higher reactant concentrations.

Although the reaction kinetics of intramembrane proteases in liposomes has been described in terms of interfacial kinetics (12, 21), that is, expressing the kinetic constants in relationship to the volume or molar fraction of the lipidic phase, (22, 23), the kinetic effects related to the reaction occurring in detergent micelles have surprisingly not yet been considered in enzyme kinetics studies on rhomboid proteases (12, 13) nor other intramembrane proteases, yet they are evidently important for the interpretation of kinetics measurements. Our data show that for reliable and meaningful measurement of apparent Michaelis-Menten kinetics parameters, the micelle concentration must not be limiting the solubility of the substrate, and the detergent concentration must be kept constant. The latter point also means that having a stock solution of the substrate dissolved in detergent (at a higher concentration than intended in the reaction mixture, which frequently can occur during purification and concentration) may lead to underestimation of reaction rates at high substrate concentrations due to a possibly

167
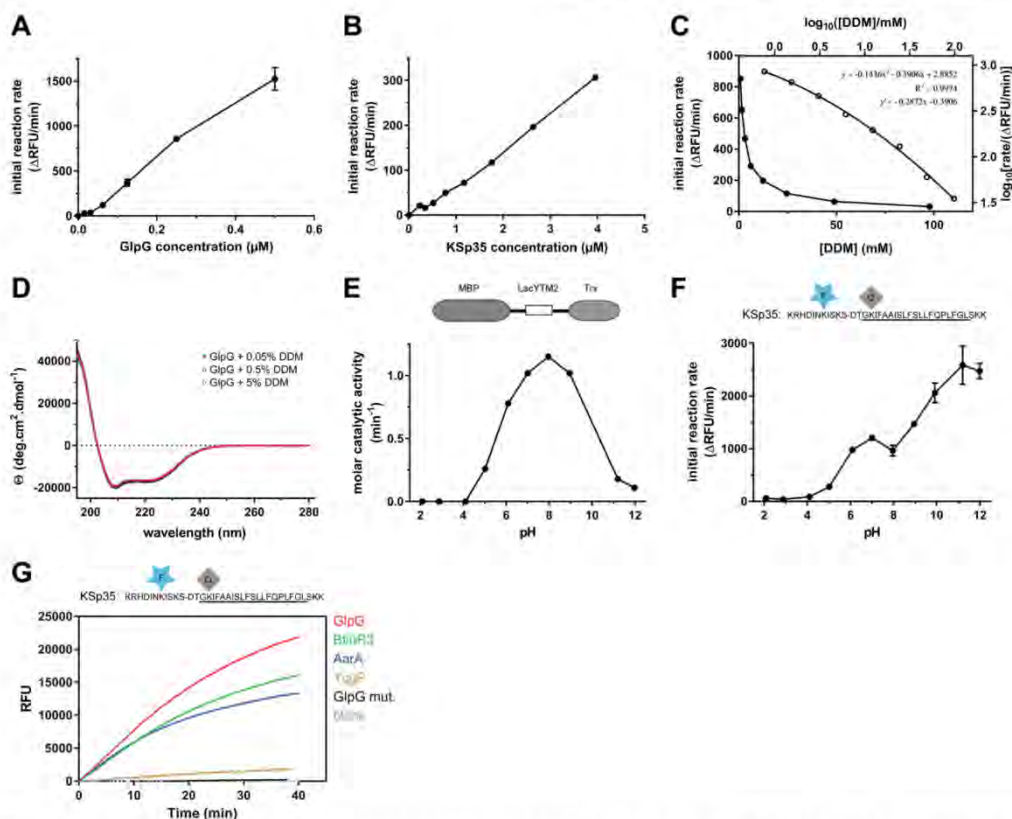
FIGURE 3. **Kinetic characterization of fluorogenic transmembrane peptide substrate KSp35 in the detergent micelle system.** *A*, dependence of the initial reaction rate on enzyme concentration. The fluorogenic substrate KSp35 (10 μM) was incubated with varying concentrations of GlpG in a reaction buffer composed of 20 mM HEPES, pH 7.4, 150 mM NaCl, 0.05% (w/v) DDM, and 10% (v/v) DMSO, and initial reaction rates were measured by following fluorescence at 493 nm. The displayed values are means from duplicate measurements with 2 × S.D. *B*, dependence of the initial reaction rate on substrate concentration. The rhomboid protease GlpG (0.4 μM) was incubated with varying concentrations of the fluorogenic substrate KSp35 in a reaction buffer composed of 20 mM HEPES, pH 7.4, 150 mM NaCl, 0.05% (w/v) DDM, 10% (v/v) DMSO, and the initial reaction rates were measured by following fluorescence at 493 nm. Representative values from one of three independent experiments are shown. *C*, dependence of the initial reaction rate on detergent concentration (*solid circles, left* and *lower axes*). The fluorogenic substrate KSp35 (10 μM) was incubated with 0.4 μM GlpG at varying concentrations of DDM in a reaction buffer composed of 20 mM HEPES, pH 7.4, 150 mM NaCl, 10% (v/v) DMSO, and initial reaction rates were measured by following fluorescence at 493 nm. Representative values from one of three independent experiments are shown. The *open circles* (*right* and *upper axes*) represent the same plot at the logarithmic scale. When this plot is fitted by second-order polynomial, the equation $y = -0.1436x^2 - 0.3906x + 2.8852$ is obtained, the derivative of which, $y' = -0.2872x - 0.3906$, is equal to the power of DDM concentration with which the reaction rate decreases. For high DDM concentrations the derivative tends to -1 (for $x = 2, y' = -0.965$), whereas for lower DDM concentrations the absolute value of the power decreases (for $x = 0, y' = -0.3906$). *D*, overall secondary structure of GlpG is not affected by high concentrations of DDM. CD spectra of GlpG at 0.05, 0.5, and 5% (w/v) (98 mM) DDM were recorded and show no variation in the secondary structure content of GlpG depending on DDM concentration. *E*, the pH dependence of GlpG activity on the LacYTM2-derived chimeric substrate MBP-LacYTM2-Trx. The substrate (2 μM) was incubated with 0.1 μM GlpG in a broad pH range buffer (38) composed of 40 mM H$_3$PO$_4$, 40 mM CH$_3$COOH, and 40 mM H$_3$BO$_3$ adjusted to pH values between 2 and 12, and initial reaction rates were measured by SDS-PAGE and densitometry as described under "Experimental Procedures." *F*, the pH dependence of cleavage of the fluorogenic LacYTM2-derived substrate KSp35 by GlpG. The substrate (10 μM) was incubated with 0.4 μM GlpG in a broad pH range buffer (38) composed of 40 mM H$_3$PO$_4$, 40 mM CH$_3$COOH, and 40 mM H$_3$BO$_3$ adjusted to pH values between 2 and 12, and initial reaction rates were measured by recording fluorescence at 493 nm. *G*, selectivity of the fluorogenic substrate KSp35 for diverse bacterial rhomboid proteases. The purified recombinant rhomboid proteases GlpG, AarA, YqgP (all at 0.4 μM), and BtioR3 (at 0.04 μM) were incubated with 10 μM KSp35 in a reaction buffer composed of 20 mM HEPES, pH 7.4, 150 mM NaCl, 0.05% (w/v) DDM, and 10% (v/v) DMSO, and progress curves were measured by recording the increase in fluorescence at 493 nm.

significant increase of detergent concentration in the final reaction mixture, as shown in Fig. 3*C*. This could result in pseudo-Michaelis kinetics and yield falsely low K_m values. Practical implications are that 1) exact detergent concentrations must be known in any kinetics measurements, and 2) it is advantageous to have the substrate stock solution dissolved in a detergent-free medium or at a detergent concentration lower or equal to that used in the final assay buffer. The transmembrane substrates presented in this article, generated by chemical synthesis, are in principle avoiding this problem, because their stock solutions are detergent-free dissolved in anhydrous dimethyl sulfoxide. Alternatively, they can be reconstituted into a detergent of choice via disaggregation in hexafluoroisopropanol, as described by Deber *et al.* (24).

168
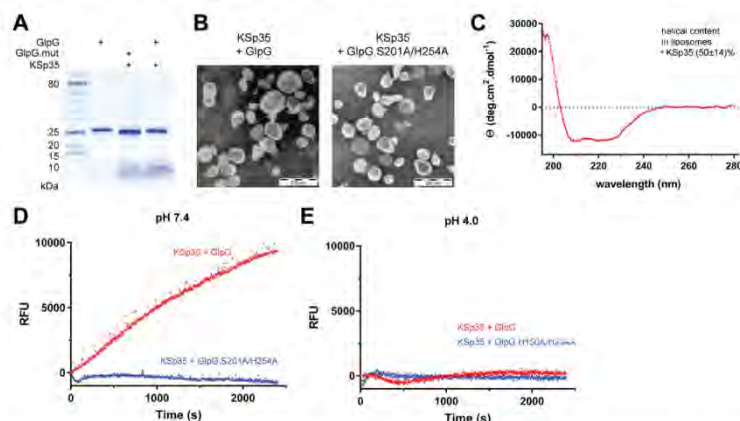
FIGURE 4. **The use of the transmembrane peptide substrate in liposomes.** *A*, KSp35 was reconstituted into liposomes (LUVs) formed from *E. coli* polar lipid extract in the presence of GlpG or its inactive mutant S201A at pH 4.0. The resulting large unilamellar vesicles were analyzed by SDS-PAGE. *B*, the shape, lamellarity, and approximate size distribution of the KSp35 + GlpG containing proteoliposomes formed at pH 4.0 were characterized by transmission electron microscopy. *C*, the integration of KSp35 into liposomes and its secondary structure content were analyzed by electronic CD. The substrate KSp35 (3 μM) was reconstituted with 2 mg/ml of *E. coli* polar lipid extract yielding an approximate peptide:lipid weight ratio of 1:500. *D*, activity of GlpG in liposomes detected by the KSp35 fluorogenic substrate. The substrate was co-reconstituted with wild type GlpG or its S201A/H254A mutant in a 30:1 molar ratio into LUVs made of *E. coli* polar lipid extract at pH 4.0, proteoliposomes were collected by ultracentrifugation and resuspended in 10 mM HEPES, 150 mM NaCl, pH 7.4, to start the cleavage reaction, which was then followed by measuring fluorescence at 493 nm. *E*, wild type GlpG or its H150A/H254A mutant were co-reconstituted with the substrate KSp35 in a 30:1 molar ratio into LUVs made of *E. coli* polar lipid extract at pH 4.0, proteoliposomes were collected by ultracentrifugation, resuspended in 50 mM sodium acetate, 150 mM NaCl, pH 4.0, and fluorescence was followed at 493 nm.

The pH dependence of cleavage rate of the unmodified LacYTM2 transmembrane segment in the context of an MBP-thioredoxin fusion protein shows a relatively broad maximum around pH 9, with substantial activity of GlpG between pH 6 and 11 and negligible activity below pH 4 and at pH 12 (Fig. 3*E*), which is largely in agreement with previous studies (12, 13). The dependence of the cleavage rate of KSp35 on pH also shows that GlpG is completely inactive at pH values below and up to 4, but the initial reaction rate of KSp35 cleavage then appears to grow up to pH 12 (Fig. 3*F*). This effect cannot be ascribed to the pH-dependent change of EDANS fluorescence (data not shown), and could possibly be due to effects of pH on the conformational dynamics of KSp35. However, this is not a concern because in most cases measurements are performed at a physiologicaly relevant pH near neutral. The apparent catalytic efficiency k_{cat}/K_m of GlpG against KSp35 measured at pH 7.4 and 0.05%(w/v) DDM is $(2.0 \pm 0.5) \times 10^{-3}$ min^{-1} μM^{-1}, which is comparable with the values reported for the TatA substrate by Dickey *et al.* (12) and Arutyunova *et al.* (13) obtained in similar conditions. Importantly, the LacYTM2-derived fluorogenic peptide substrate KSp35 is cleaved efficiently by unrelated recombinantly purified bacterial rhomboids GlpG, AarA, and BtioR3, and modestly by YqgP (Fig. 3*G*), which demonstrates its wide usability, surpassing any other currently available rhomboid substrates.

Use of the Transmembrane Peptide Substrate in Liposomes— Because the natural environment of rhomboid proteases is the lipid membrane, we next tested whether the fluorogenic peptide substrate KSp35 can also be used in liposomes. We co-reconstituted KSp35 with GlpG or its inactive mutant S201A/H254A at pH 4 into large unilamellar vesicles (LUVs) formed from *E. coli* polar lipid extract, and confirmed the composition of the resulting proteoliposomes by SDS-PAGE (Fig. 4*A*). Neg-

ative stain transmission electron microscopy showed that both empty LUVs and proteoliposomes containing KSp35 in the presence or absence of GlpG or its inactive mutant S201A/H254A had similar morphology and size distribution both at pH 7 and 4 (Fig. 4*B*). The CD spectrum of LUV-reconstituted KSp35 showed helicity of $50 \pm 14\%$ (Fig. 4*C*), which is consistent with its transmembrane helix prediction. GlpG is inactive at pH 4 (Fig. 3, *E* and *F*), and, consistently, fluorescence of proteoliposomes containing KSp35 and GlpG at pH 4 was at a constant background level (Fig. 4*E*). Upon neutralization to pH 7.4, time-dependent increase of fluorescence at 495 nm was observed in the presence of wild type GlpG but not in the presence of its active-site mutant S201A/H254A (Fig. 4*D*). These results collectively demonstrate that the LacYTM2-based fluorogenic transmembrane substrate KSp35 is widely usable both in detergent micelles or liposomes and with diverse rhomboid proteases.

*A Red-shifted Variant of the Fluorogenic Transmembrane Substrate for Rhomboids—*Large compound libraries for high-throughput screening can often contain compounds that absorb in the UV region (14), and fluorogenic substrates operating at red-shifted wavelengths are less affected by such compound interference. Because EDANS is excited in the UV region, and is thus prone to interference in library screening, we have modified the LacYTM2 peptide backbone by instead attaching the red-shifted TAMRA fluorophore to a Lys introduced into the P5 position and a compatible dark quencher QXL610 to a Cys introduced into the P4' position (Fig. 5*A*) to yield KSp76. This red-shifted fluorogenic substrate is cleaved by several bacterial rhomboid proteases with efficiencies similar to its UV variant KSp35. The apparent catalytic efficiency k_{cat}/K_m of GlpG cleaving KSp76 is $(1.6 \pm 0.5) \times 10^{-3}$ min^{-1}

*S*ASBMB   VOLUME 292 · NUMBER 7 · FEBRUARY 17, 2017

169
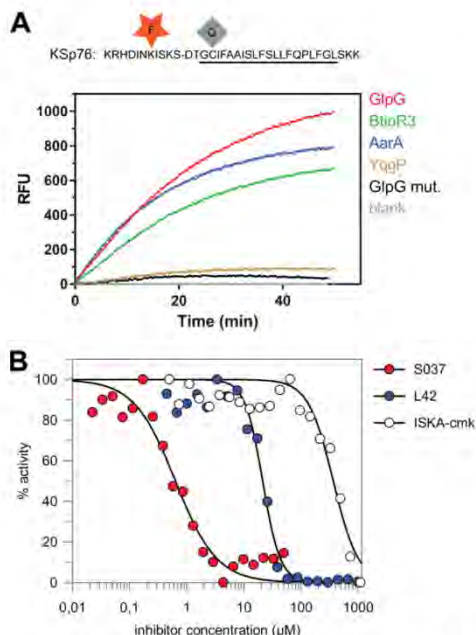
FIGURE 5. **Red-shifted variant of the LacYTM2-based fluorogenic substrate.** *A,* modification of Lys in the P5 position of KSp31 by the red-shifted TAMRA fluorophore and P4' Cys by a dark quencher QXL610 yields highly fluorogenic substrate KSp76 that is efficiently cleaved by rhomboid proteases GlpG, AarA, YqgP, and BtioR3 at identical concentrations to those used in Fig. 3G. Excitation wavelength was 553 nm, and emission was followed at 583 nm. *B,* the red-shifted fluorogenic substrate KSp76 allows measurement of inhibition by compounds that absorb in the UV region, such as isocoumarin, and is thus suitable for high-throughput screening. The dose-response curves of the chloromethylketone ISKAcmk, β-lactam L42, and isocoumarin S037 were measured after a 60-min preincubation of enzyme with inhibitor. The curves were fitted in GraFit 7 to yield apparent IC_{50} values.

μM^{-1}, which is similar to the EDANS variant KSp35 ((2.0 ± 0.5) × 10^{-3} min^{-1} μM^{-1}) under identical reaction conditions within experimental error (Fig. 6C). The utility of this red-shifted variant of the LacYTM2 substrate is demonstrated by measuring the inhibition curves of chloromethylketone ISKA-cmk (19), β-lactam L42 (11), and isocoumarin S037 (25, 26). Using a 60-min enzyme + inhibitor preincubation time, the measurements yielded apparent IC_{50} values of 370 ± 38, 12.4 ± 1.6, and 0.64 ± 0.08 μM, respectively (Fig. 5B), which are largely in agreement with published values measured in other assay systems and otherwise comparable conditions (11, 15, 19).

Efficiency and Selectivity of the Substrates Can Be Tuned by Varying Their Non-prime Side Amino Acid Sequence—One of the problems with current rhomboid protease assays is that there has been little rationale about how to modify the substrates to improve their kinetic properties and adapt them for different rhomboid proteases. Recent enzymatic analyses (12, 18) have shown that the region between the P4 and P2' residues determines the k_{cat} of the cleavage reaction, suggesting that selective substrates for rhomboids could be designed by modi-

fying the P4 to P2' region appropriately. A recent mutagenic study of the TatA substrate and structural analysis of a derived rhomboid·substrate-peptide complex revealed amino acids at the P5 to P1 positions of TatA that are preferred by GlpG (19). We tested the impact of these substitutions in the context of the LacYTM2 substrate.

Although single mutations of the P5 amino acid to the preferred Arg, P4 amino acid to Val, and P2 amino acid to His did not improve the cleavage of the purified recombinant MBP-LacYTM2-Trx substrate *in vitro*, mutation of the P1 amino acid to Ala improved the cleavage of mutant 7-fold, and mutation of the P3 residue to Arg improved the cleavage of mutant 16-fold (Fig. 6A). Combining all five mutations yielded a mutant substrate (RVRHA) that was cleaved 64-fold better than the wild type substrate (Fig. 6A), which shows that the effects of the preferred substitutions are additive. When analyzed for cleavage *in vivo*, it turns out that already the wild type MBP-LacYTM2-Trx substrate is such a good substrate of GlpG that it is turned over from 94% (Fig. 6B). The effects of the preferred P5 to P1 mutations thus cannot be assessed in this context as they all exhibit similarly high steady-state turnover (Fig. 6B).

To test this effect in our fluorogenic substrates, we have modified the TAMRA-based LacYTM2-derived fluorogenic substrate by changing the P5 to P1 segment from HISKS to RVRHA to yield KSp64, and compared the kinetic properties of both substrates. The analysis revealed that catalytic efficiency k_{cat}/K_m of GlpG cleaving KSp64 is (3.7 ± 0.4) × 10^{-2} min^{-1} μM^{-1}, which is 23-fold higher than that of the original red-shifted LacYTM2 substrate KSp76 ((1.6 ± 0.5) × 10^{-3} min^{-1} μM^{-1}) (Fig. 6C). The impact of the modifications of the P5 to P1 region on selectivity against other bacterial rhomboid proteases is particularly striking (Fig. 6D), with the initial reaction rate of KSp64 cleavage by GlpG being about 50-fold higher than that of AarA (measured from data displayed in Fig. 6D) and even higher for the other tested rhomboid proteases, revealing a straightforward strategy for designing selective rhomboid substrates.

In summary, we report novel sensitive versatile fluorogenic transmembrane peptide substrates for rhomboid intramembrane proteases that are usable both in detergent micelles and liposomes, are cleaved by diverse rhomboid proteases, and contain a red-shifted fluorophore suitable for high-throughput screening assays. Furthermore, we provide a strategy how to adapt these substrates to individual rhomboid proteases by modifying their P5 to P1 residues, and we demonstrate that controlling the detergent concentration is important for obtaining accurate kinetic data. We expect that the substrates we describe and sequence variants thereof will enable facile detection of activity and development of inhibitors of rhomboid proteases.

Experimental Procedures

General Biochemicals—Lipids were from Avanti Polar Lipids, detergents from Anatrace, buffers and other biochemicals were from Sigma or other suppliers as specified below.

DNA Constructs and Cloning—The expression constructs for rhomboid proteases GlpG, YqgP, and AarA and chimeric MBP-TMD-Trx substrate constructs where TMD = LacYTM2, Gur-

170
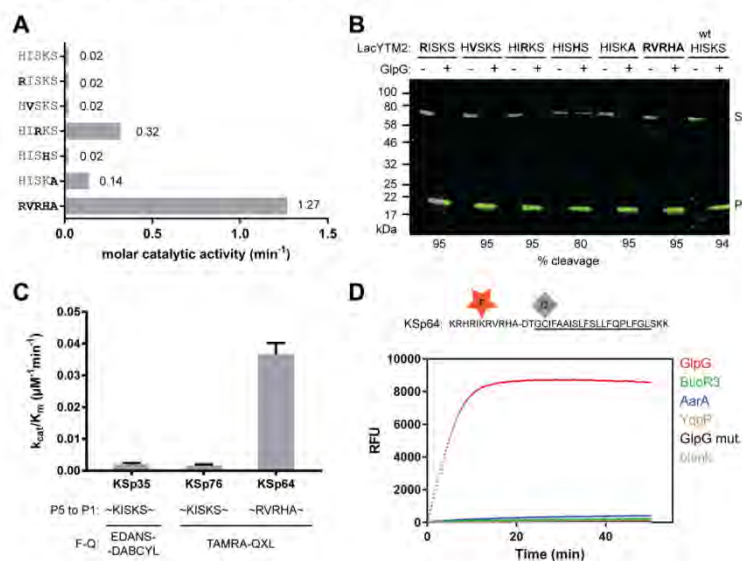
FIGURE 6. **The effect of non-prime side substitutions on the catalytic parameters and selectivity of rhomboid substrates.** *A,* preferred amino acids in the P5 to P1 positions of the LacYTM2 transmembrane substrate improve its cleavage by GlpG. The LacYTM2 embedded in the MBP-thioredoxin chimera (18) was point-mutated in the P5 to P1 positions according to the sequence preferences of *E. coli* GlpG (19). The recombinant substrates were expressed in *E. coli* ΔglpG, purified, and molar catalytic activity of GlpG in cleaving each of the substrates was determined using gel-based assay (see "Experimental Procedures" for details). The concentration of substrate was always 1.47 μM, concentration of DDM was 0.5%(w/v), the concentration of GlpG was 0.8 μM for wild type substrate (HISKS), and for the RISKS, HVSKS, and HISHS mutants the concentration was 0.08 μM for the HISKA mutant and 0.016 μM for the HIRKS and RVRHA variants (to ensure reliable measurement of the initial reaction rate). Representative values from one of three independent experiments are shown. *B,* the effects of the preferred amino acids in the P5 to P1 region of LacYTM2 on the steady-state level of cleavage by GlpG in biological membranes *in vivo.* Plasmids encoding individual mutant versions of the chimeric mutant LacYTM2 substrates described above were transformed into *E. coli* MC4100 expressing endogenous GlpG, and 2 h after induction of expression of the substrates, the cell lysates were analyzed by immunoblotting using antibody against His tag, located at the C terminus of the constructs. Detection by near-infrared laser scanning, exhibiting linearity over 6 orders of magnitude, enabled reliable quantitation. Integration of product and substrate band intensities yielded steady-state substrate conversion values that are listed *below* the image. A representative experiment is displayed. *C,* apparent kinetic parameters of fluorogenic rhomboid substrates derived from LacYTM2. Initial reaction rates at very low substrate concentrations were used to calculate catalytic efficiency values (k_{cat}/K_m) of substrates KSp35, KSp64, and Ksp76 cleaved by GlpG at 0.5% (w/v) DDM. The reaction buffer was 20 mM HEPES, pH 7.4, 150 mM NaCl, 10% (v/v) DMSO, enzyme concentration was 0.4 μM, and substrate concentration ranged from 0.5 to 20 μM. Note that a mere optimization of the P5 to P1 region of the substrate increases the catalytic efficiency (k_{cat}/K_m) of its cleavage by GlpG by 23-fold. *D,* influence of the optimization of the P5 to P1 region on the selectivity of a transmembrane substrate for rhomboids. KSp76 underwent cleavage by rhomboid proteases GlpG, AarA, YqgP, and BtioR3 at the same concentrations as described in the legends to Figs. 3*G* and 5*A.* Note that optimization of the P5 to P1 region of the substrate increases the selectivity for GlpG dramatically.

ken, TatA, or Spitz as described previously (18). The expression construct for rhomboid protease BtioR3 was generated by PCR amplification of the entire ORF encoding the Q8A3X2 (Uniprot ID) protein from *B. thetaiotaomicron* genomic DNA (purchased from ATCC), and its cloning as a C terminally His-tagged construct into pET25b+M as described previously (27). Mutations of the TatA and LacYTM2 recognition motif in the MBP-TMD-Trx construct were generated by overlap assembly PCR (28) and isothermal assembly (29). All constructs were verified by DNA sequencing.

Chemical Synthesis—All reagents were acquired from commercial sources and used without purification. Protected amino acids and amino acid derivatives were purchased from Iris Biotech (Marktredwitz, Germany). Trimellitic anhydride and 3-dimethylaminophenol were from Sigma, QXL610 vinyl-sulfone was from AnaSpec (Fremont, CA), and *N*-(9-fluorenyl) methoxycarbonyl (Fmoc)-Glu(EDANS)-OH from Merck KGaA (Darmstadt, Germany). The detailed synthetic procedures, analytical methods, and compound characterization data are included in the supporting information.

Protein Expression and Purification—Bacterial rhomboid proteases AarA, GlpG, BtioR3, and YqgP and the active site mutant GlpG.S201A were overexpressed in *E. coli* C41(DE3) (30) as full-length, C terminally His-tagged proteins from a modified pET25b+ vector (27). The cultures were grown at 37 °C in LB medium to A_{600} of 0.4 and induced by 1 mM isopropyl 1-thio-β-D-galactopyranoside. The expression was continued overnight at 20 °C. Cells were harvested, resuspended in buffer A (25 mM HEPES, pH 7.4, 100 mM NaCl, 10%(v/v) glycerol, 1 mM PMSF), and lysed by 2 to 3 passes through Avestin EmulsiFlex-C3. Cell debris was removed by a low-speed centrifugation. Cellular membranes were isolated by a 2-h centrifugation at 100,000 × g and were solubilized in 1.5%(w/v) DDM (solubilization grade, Anatrace) in Buffer B (25 mM HEPES, pH 7.4, 300 mM NaCl, 10%(v/v) glycerol, 10 mM imidazole, EDTA-free Complete Protease Inhibitor mixture (Roche Applied Science)) at room temperature for 1 h. Solubilized proteins were isolated by centrifugation at 100,000 × g for 30 min and loaded onto nickel-nitrilotriacetic acid HiTrap IMAC HP 1-ml columns (GE Healthcare). Nonspecifically bound proteins were

171

washed off with Buffer C (25 mM HEPES, pH 7.4, 300 mM NaCl, 10% (v/v) glycerol, 0.05% (w/v) DDM) containing 10, 50, and 125 mM imidazole. The protein was eluted with Buffer C containing 250 to 500 mM imidazole. The peak fractions were buffer exchanged into 25 mM HEPES, pH 7.4, 150 mM NaCl, 10% (v/v) glycerol, and 0.05% (w/v) DDM on a HiPrep 26/10 desalting column (GE Healthcare). If needed, proteins were concentrated using Vivaspin ultrafiltration spin cells with 30-kDa MWCO. Protein concentration was determined from absorbance at 280 nm, and the final concentration of DDM was determined as described (31).

Capillary Electrophoresis (CE)—Analyses of standard peptides and enzymatically cleaved peptide substrates were performed on an Agilent CE 7100 instrument (Agilent, Waldbronn, Germany) equipped with photodiode array UV-visible detector operating in the 190–600 nm range. Electropherograms were acquired at 192, 205, and 214 nm and absorbance data at 192 nm were selected for quantitative evaluation due to the highest signal to noise ratio. CE analyses were carried out in a bare fused silica capillary with polyimide outer coating (internal diameter 50 μm, outer diameter 375 μm, effective length to the detector 40 cm, total length 48.5 cm, supplied by Polymicro Technologies, Phoenix, AZ). Peptides were analyzed as cations in acidic background electrolyte (BGE) composed of 100 mM H_3PO_4, 69 mM Tris, pH 2.5. For highly hydrophobic peptides, this BGE was modified by the addition of 0.05% (w/v) DDM. The temperature of the air-cooled capillary was set to 20 °C and the sample carousel was kept at the same temperature using a circulating water bath. Prior to each CE run, the capillary was successively washed with 100 mM sodium dodecyl sulfate, ethanol, 1 M NaOH, water, 1 M HCl, and the BGE, to remove any possible carryover of hydrophobic peptides and detergents from the previous run. All washes were done at 8 bar pressure for 30 s. Peptide standards used for identification of cleavage products were solubilized in DMSO at 1 mM concentration and mixed with 50 mM HEPES buffer containing 0.05% (w/v) DDM, resulting in 50 μM peptide concentration.

The enzymatic cleavage reactions were carried out in 20 mM HEPES, pH 7.4, with 0.05%(w/v) DDM and 10% (v/v) DMSO, with 250 μM peptide substrate and 2.6 μM full-length GlpG at 37 °C. To measure the initial reaction rates, fractions were collected every 15 min for up to 2 h and the reaction was terminated by the addition of 10 mM HCl. Samples for CE were prepared by mixing 20 μl of peptide solutions with 2 μl of 2.2 mM tyramine (internal standard for quantitative analysis). Sample solutions were injected into the capillary by 20 mbar pressure for 10 s. Separations were performed at +25 kV (anode at the capillary injection end). The electrode vessels were replenished with fresh BGE after each run. All analyses were performed in triplicate. Quantitative analysis was based on the ratio of corrected (migration time normalized) peak areas of peptides of interest and the internal standard (tyramine) (32).

Mass Spectrometry—The analysis of enzymatic cleavage products of transmembrane peptides was carried out using MALDI-TOF mass spectrometry on an UltrafleXtreme™ MALDI-TOF/TOF mass spectrometer (Bruker Daltonics, Germany) with α-cyano-4-hydroxycinnamic acid matrix using a thin-layer method (33). For routine quality control during pep-

tide synthesis, mass spectra were acquired on a Waters Micromass ZQ ESCi multimode ionization mass spectrometer, and LTQ Orbitrap XL (Thermo Fisher Scientific) for HR-MS experiments, in both cases using ESI(+) ionization.

Gel-based Assay for Rhomboid Activity—For gel-based assays used in Fig. 1, the purified recombinant full-length maltose-binding protein thioredoxin fusion proteins harboring the transmembrane domains of TatA, LacYTM2, Gurken, and Spitz (18) were used as substrates. The reaction was carried out in 50 mM Tris, pH 7.4, 100 mM NaCl, 10% (v/v) glycerol, 0.05% (w/v) DDM, and 5 μM substrate. Enzyme concentrations varied to ensure adequate conditions for measurement of initial reaction rates for each enzyme-substrate combination. Time courses were measured by withdrawing 10-μl aliquots from the reaction mixture after 10, 20, 30, 40, 50, 60, and 120 min from the start of the reaction, and stopping the reaction by the addition of SDS-PAGE sample buffer. The reaction mixtures were analyzed by SDS-PAGE, Coomassie staining (InstantBlue, Expedeon, UK), and densitometry as described (19), and initial reaction rates were converted to molar catalytic activities defined as the number of substrate molecules converted by a molecule of the enzyme per unit of time (consistent with the definition by IUPAC (34, 35)). Variations in conditions used for measurements in Fig. 6 are denoted in the figure legend.

The *in vivo* assay of rhomboid activity was carried out essentially as described (19). Cleavage products were detected by SDS-PAGE and Western blotting using primary anti-penta-His mouse monoclonal antibody (Thermo) and IRDye 800CW goat anti-mouse fluorescent secondary antibody (LiCor). Densitometry was done in ImageStudio software (LiCor) and substrate conversion (α) was calculated from band intensities as $α_τ = [P]/[S] + [P]$, where [P] and [S] are product and substrate concentrations at time τ, which are proportional to the fluorescence intensity of the product and substrate bands at time τ, because the monoclonal antibody binds to the substrate or product in a constant molar ratio irrespective of their molecular weights.

Fluorescence Assay for Rhomboid Activity—The fluorescence assay of rhomboid activity was performed at 37 °C in 96-well black HTS plates (Greiner Bio-One). The reaction conditions were typically as follows: 20 mM HEPES, pH 7.4, 150 mM NaCl, 0.05% (w/v) DDM, 12% (v/v) DMSO, and 10 μM fluorogenic peptide substrate in a final volume of 50 μl, unless noted otherwise. Concentrations of stock solutions of peptide substrates and inhibitors (if applicable) were determined by quantitative amino acid analysis. Fluorescence was read continuously in a plate reader (Tecan Infinite M1000). Excitation and emission wavelengths were 335 and 493 nm, respectively, for the EDANS-DABCYL substrate, and 553 and 583 nm for the TAMRA-QXL610 substrates. Data were evaluated in i-Control (Tecan), Excel (Microsoft), GraphPad Prism 7 (GraphPad Software, Inc.), and GraFit 7 (Erithacus Software, Ltd.) software.

Inhibition Assays—The inhibition assay was carried out in 20 mM HEPES, pH 7.4, 150 mM NaCl, 12% (v/v) DMSO, 0.05% (w/v) DDM at 37 °C in 96-well black HTS plates (Greiner Bio-one). Purified recombinant full-length GlpG (0.4 μM) was pre-

172

incubated with each inhibitor at different concentrations for 1 h at 37 °C. The cleavage reaction was started by adding 10 μM KSp76 and fluorescence was read continuously to measure initial reaction rates as described above.

Reconstitution into Liposomes—*E. coli* polar lipids (20 mg), with optionally 0.1 mg of Lissamine Rhodamine B-labeled phosphatidylethanolamine (16:0) (Avanti Polar Lipids) added for visibility, were dried in a glass test tube by manual rotation under a nitrogen stream. Residual traces of solvent were removed by overnight incubation in a vacuum chamber (Binder). The resulting lipid film was hydrated in 5 ml of 50 mM acetate, 150 mM NaCl, pH 4.0, by 2 min vortexing followed by a 1-h incubation in a horizontal shaker at 200 rpm and 37 °C, and 3 cycles of freezing in liquid nitrogen and thawing in a 37 °C water bath. The lipid suspension was then extruded through a 200-nm pore membrane by 19 strokes in an Avanti Mini Extruder (Avanti Polar Lipids).

For reconstitution of proteins and peptides into liposomes, these unilamellar LUVs were solubilized in DM to a final ratio of 1.5:1 detergent:lipid, and incubated for 1 h at room temperature under gentle rotation. This mixture was diluted to a final lipid concentration of 2 mg/ml in 50 mM acetate, 150 mM NaCl, pH 4.0, and protein (GlpG or its inactive mutant) dissolved in detergent was added to a final concentration of 8 μg/ml; alternatively, the stock solution of substrate peptide KSp35 in DMSO was diluted to 10 μM. The resulting mixture was incubated at room temperature for 1 h under gentle mixing by inversion. Detergent was removed by overnight dialysis against 500-fold excess of 50 mM acetate, 150 mM NaCl, pH 4, followed by 5 h dialysis against 500-fold excess the same buffer, using 10-kDa MWCO dialysis membranes, which allowed reconstitution of proteoliposomes. These were extruded through 200-nm pore filters 9 times to ensure reproducible size distribution and lamellarity. These final proteoliposomes were harvested by ultracentrifugation (250,000 × g for 1 h at 4 °C), and resuspended in 10 mM HEPES, pH 7.4, 150 mM NaCl to a concentration of about 33 mg/ml of lipids. The morphology and size distribution of proteoliposomes was analyzed by electron microscopy.

Transmission Electron Microscopy—Liposome samples were negatively stained with 2% phosphotungstic acid on carbon-coated electron microscopy grids and analyzed with a JEOL JEM-1011 device at 80 kV beam acceleration voltage.

CD Spectroscopy—Protein and peptide samples were dissolved in 50 mM phosphate buffer at the indicated concentrations and in the presence of detergent as indicated, or reconstituted in LUVs made of *E. coli* polar lipids and extruded by 100-nm filters to minimize light scattering. Electronic circular dichroism spectra were collected by a Jasco 815 spectrometer (Tokyo, Japan) in the spectral 195–280 nm range using a cylindrical 0.02-cm quartz cell with 0.1-nm step resolution, 5 nm/min scanning speed, 16 s response time, and 1 nm spectral band. After baseline correction, the spectra were expressed as molar ellipticity per residue θ (deg cm^2 dmol^{-1}). Numerical analysis of the secondary structure and secondary structure assignment were performed using a CDPro software package and CONTIN program (36, 37).

References

1. Urban, S., Lee, J. R., and Freeman, M. (2002) A family of rhomboid intramembrane proteases activates all *Drosophila* membrane-tethered EGF ligands. *EMBO J.* **21**, 4277–4286
2. Lee, J. R., Urban, S., Garvey, C. F., and Freeman, M. (2001) Regulated intracellular ligand transport and proteolysis control EGF signal activation in *Drosophila*. *Cell* **107**, 161–171
3. McQuibban, G. A., Saurya, S., and Freeman, M. (2003) Mitochondrial membrane remodelling regulated by a conserved rhomboid protease. *Nature* **423**, 537–541
4. O'Donnell, R. A., Hackett, F., Howell, S. A., Treeck, M., Struck, N., Krnajski, Z., Withers-Martinez, C., Gilberger, T. W., and Blackman, M. J. (2006) Intramembrane proteolysis mediates shedding of a key adhesin during erythrocyte invasion by the malaria parasite. *J. Cell Biol.* **174**, 1023–1033
5. Fleig, L., Bergbold, N., Sahasrabudhe, P., Geiger, B., Kaltak, L., and Lemberg, M. K. (2012) Ubiquitin-dependent intramembrane rhomboid protease promotes ERAD of membrane proteins. *Mol. Cell* **47**, 558–569
6. Riestra, A. M., Gandhi, S., Sweredoski, M. J., Moradian, A., Hess, S., Urban, S., and Johnson, P. J. (2015) A *Trichomonas vaginalis* rhomboid protease and its substrate modulate parasite attachment and cytolysis of host cells. *PLoS Pathog.* **11**, e1005294
7. Etheridge, S. L., Brooke, M. A., Kelsell, D. P., and Blaydon, D. C. (2013) Rhomboid proteins: a role in keratinocyte proliferation and cancer. *Cell Tissue Res.* **351**, 301–307
8. Chan, E. Y., and McQuibban, G. A. (2013) The mitochondrial rhomboid protease: its rise from obscurity to the pinnacle of disease-relevant genes. *Biochim. Biophys. Acta* **1828**, 2916–2925
9. Song, W., Liu, W., Zhao, H., Li, S., Guan, X., Ying, J., Zhang, Y., Miao, F., Zhang, M., Ren, X., Li, X., Wu, F., Zhao, Y., Tian, Y., Wu, W., *et al.* (2015) Rhomboid domain containing 1 promotes colorectal cancer growth through activation of the EGFR signalling pathway. *Nat. Commun.* **6**, 8022
10. Strisovsky, K. (2016) Why cells need intramembrane proteases: a mechanistic perspective. *FEBS J.* **283**, 1837–1845
11. Pierrat, O. A., Strisovsky, K., Christova, Y., Large, J., Ansell, K., Bouloc, N., Smiljanic, E., and Freeman, M. (2011) Monocyclic β-lactams are selective, mechanism-based inhibitors of rhomboid intramembrane proteases. *ACS Chem. Biol.* **6**, 325–335
12. Dickey, S. W., Baker, R. P., Cho, S., and Urban, S. (2013) Proteolysis inside the membrane is a rate-governed reaction not driven by substrate affinity. *Cell* **155**, 1270–1281

173

13. Arutyunova, E., Panwar, P., Skiba, P. M., Gale, N., Mak, M. W., and Lemieux, M. J. (2014) Allosteric regulation of rhomboid intramembrane proteolysis. *EMBO J.* **33**, 1869–1881

14. Simeonov, A., Jadhav, A., Thomas, C. J., Wang, Y., Huang, R., Southall, N. T., Shinn, P., Smith, J., Austin, C. P., Auld, D. S., and Inglese, J. (2008) Fluorescence spectroscopic profiling of compound libraries. *J. Med. Chem.* **51**, 2363–2371

15. Vosyka, O., Vinothkumar, K. R., Wolf, E. V., Brouwer, A. J., Liskamp, R. M., and Verhelst, S. H. (2013) Activity-based probes for rhomboid proteases discovered in a mass spectrometry-based assay. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2472–2477

16. Wolf, E. V., Zeißler, A., Vosyka, O., Zeiler, E., Sieber, S., and Verhelst, S. H. (2013) A new class of rhomboid protease inhibitors discovered by activity-based fluorescence polarization. *PLoS ONE* **8**, e72307

17. Maegawa, S., Ito, K., and Akiyama, Y. (2005) Proteolytic action of GlpG, a rhomboid protease in the *Escherichia coli* cytoplasmic membrane. *Biochemistry* **44**, 13543–13552

18. Strisovsky, K., Sharpe, H. J., and Freeman, M. (2009) Sequence-specific intramembrane proteolysis: identification of a recognition motif in rhomboid substrates. *Mol. Cell* **36**, 1048–1059

19. Zoll, S., Stanchev, S., Began, J., Skerle, J., Lepšík, M., Peclinovská, L., Majer, P., and Strisovsky, K. (2014) Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate-peptide complex structures. *EMBO J.* **33**, 2408–2421

20. VanAken, T., Foxall-VanAken, S., Castleman, S., and Ferguson-Miller, S. (1986) Alkyl glycoside detergents: synthesis and applications to the study of membrane proteins. *Methods Enzymol.* **125**, 27–35

21. Kamp, F., Winkler, E., Trambauer, J., Ebke, A., Fluhrer, R., and Steiner, H. (2015) Intramembrane proteolysis of β-amyloid precursor protein by γ-secretase is an unusually slow process. *Biophys. J.* **108**, 1229–1237

22. Scheel, G., Acevedo, E., Conzelmann, E., Nehrkorn, H., and Sandhoff, K. (1982) Model for the interaction of membrane-bound substrates and enzymes: hydrolysis of ganglioside GD1a by sialidase of neuronal membranes isolated from calf brain. *Eur. J. Biochem.* **127**, 245–253

23. Parry, G., Palmer, D. N., and Williams, D. J. (1976) Ligand partitioning into membranes: its significance in determining K_m and K_s values for cytochrome P-450 and other membrane bound receptors and enzymes. *FEBS Lett.* **67**, 123–129

24. Rath, A., and Deber, C. M. (2013) Design of transmembrane peptides: coping with sticky situations. *Methods Mol. Biol.* **1063**, 197–210

25. Wolf, E. V., Zeissler, A., and Verhelst, S. H. (2015) Inhibitor fingerprinting of rhomboid proteases by activity-based protein profiling reveals inhibitor selectivity and rhomboid autoprocessing. *ACS Chem. Biol.* **10**, 2325–2333

26. Haedke, U., Küttler, E. V., Vosyka, O., Yang, Y., and Verhelst, S. H. (2013) Tuning probe selectivity for chemical proteomics applications. *Curr. Opin. Chem. Biol.* **17**, 102–109

27. Lemberg, M. K., Menendez, J., Misik, A., Garcia, M., Koth, C. M., and Freeman, M. (2005) Mechanism of intramembrane proteolysis investigated with purified rhomboid proteases. *EMBO J.* **24**, 464–472

28. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51–59

29. Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison C. A., 3rd, and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345

30. Miroux, B., and Walker, J. E. (1996) Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289–298

31. Urbani, A., and Warne, T. (2005) A colorimetric determination for glycosidic and bile salt-based detergents: applications in membrane protein research. *Anal. Biochem.* **336**, 117–124

32. Solínová, V., Kasicka, V., Koval, D., Barth, T., Ciencialová, A., and Záková, L. (2004) Analysis of synthetic derivatives of peptide hormones by capillary zone electrophoresis and micellar electrokinetic chromatography with ultraviolet-absorption and laser-induced fluorescence detection. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **808**, 75–82

33. Fenyo, D., Wang, Q., DeGrasse, J. A., Padovan, J. C., Cadene, M., and Chait, B. T. (2007) MALDI sample preparation: the ultra thin layer method. *J. Vis. Exp.* **192**, 10.3791/192

34. Nomenclature Committee of the International Union of Biochemistry (1979) Units of enzyme-activity, recommendations 1978. *Eur. J. Biochem.* **97**, 319–320

35. Nomenclature Committee of the International Union of Biochemistry (1983) Symbolism and terminology in enzyme-kinetics, recommendations 1981. *Biochem. J.* **213**, 561–571

36. Sreerama, N., and Woody, R. W. (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* **287**, 252–260

37. Provencher, S. W., and Glöckner, J. (1981) Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **20**, 33–37

38. Mongay, C., and Cerda, V. (1974) Britton-Robinson buffer of known ionic-strength. *Anal. Chim.* **64**, 409–412

174

P10. Recombinant human serine racemase: enzymologic characterization and comparison with its mouse ortholog.

Recombinant human serine racemase: Enzymologic characterization and comparison with its mouse ortholog

Hillary E. Hoffman [a,b], Jana Jirásková [a,b], Marek Ingr [b], Marketa Zvelebil [c], Jan Konvalinka [a,b,*]

[a] Gilead Sciences and IOCB Research Center, Institute of Organic Chemistry and Biochemistry of the Academy of Sciences of the Czech Republic, v.v.i., Flemingovo n. 2, 166 10 Prague 6, Czech Republic
[b] Department of Biochemistry, Faculty of Science, Charles University, Hlavova 8, Prague 2, Czech Republic
[c] Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK

ARTICLE INFO

ABSTRACT

D-Serine plays a key role in glutamatergic neurotransmission in mammalian brain as a co-agonist of N-methyl-D-aspartate receptors. The enzyme responsible for D-serine biosynthesis, serine racemase (SR), is therefore a promising target for treatment of neuropathologies related to glutamate receptor excitotoxicity, such as stroke or Alzheimer's disease. Much of the experimental work to date has been performed on mouse serine racemase, which shares a high level of sequence identity with its human ortholog. In this work, we report the synthesis of a human SR gene variant optimized for heterologous expression in Escherichia coli and describe the expression and purification of active recombinant human SR. This strategy may be of general interest to researchers wishing to express mammalian proteins in a bacterial system. Furthermore, we conduct a thorough analysis of the kinetics and inhibitor-sensitivity of the recombinant enzyme, and we provide the first direct comparison of human and mouse SR based on our kinetic data. The orthologs behave similarly overall and exhibit identical inhibition profiles, validating the use of mouse models in SR research.

Introduction

Complex roles for D-amino acids in mammalian physiology have been revealed in recent years. In particular, D-serine has emerged as a subject of research interest due to its role as a neurotransmitter and co-agonist at the "glycine" site of the N-methyl-D-aspartate receptor (NMDAR)[1] complex [1,2] (for a comprehensive review see [3,4]). The NMDAR is believed to play a crucial role in synaptic plasticity, the suggested neurochemical foundation for learning and memory [5–7]. Overactivation of NMDARs has been linked to post-ischemic brain damage such as stroke-related cell death, and blockers of the NMDAR glycine site have proven neuroprotective in animal models of stroke [8]. Aberrant regulation of D-serine levels has also been implicated in neuropathologies including schizophrenia [9], Alzheimer's disease [10], and amyotrophic lateral sclerosis [11].

The biosynthesis of D-serine is believed to be catalyzed by the pyridoxal-5′-phosphate (PLP)-dependent enzyme serine racemase (SR), which converts L-serine to its enantiomer. SR is therefore a promising target enzyme for investigation and treatment of neuropathologies linked to abnormal D-serine levels. The connection between SR and schizophrenia is one of the best-studied examples, although the data remain inconsistent. Verral et al. recently observed an increase in SR expression in schizophrenia in the dorsolateral prefrontal cortex of human brain [12], whereas with the same antibody, Steffek et al. detected increased expression in schizophrenic hippocampus but not in the prefrontal cortex [13]. Both sets of results contrast with those of Bendikov et al., who reported that SR expression levels are reduced in both the frontal cortex and hippocampus of schizophrenic brains [14]. Clearly, much remains to be understood about this enigmatic enzyme and its biological regulation in schizophrenia (and other pathologies).

Toward this end, recombinant forms of both mouse and human SR have been purified and studied by several groups [15–18], and both SR orthologs are widely and often interchangeably used in SR research. However, no direct comparison of the activities and inhibitor susceptibilities of these two enzymes has yet been detailed. Therefore, we synthesized a codon-optimized variant of the gene encoding human SR (hSR) and developed an Escherichia coli based expression and purification system that enables us to

* Corresponding author. Address: Gilead Sciences and IOCB Research Center, Institute of Organic Chemistry and Biochemistry of the Academy of Sciences of the Czech Republic, v.v.i., Flemingovo n. 2, 166 10 Prague 6, Czech Republic. Fax: +420 220183578.
E-mail address: konval@uochb.cas.cz (J. Konvalinka).
[1] Abbreviations used: NMDAR, N-methyl-D-aspartate receptor; PLP, pyridoxal-5′-phosphate; SR, serine racemase; CD, circular dichroism; LSOS, L-serine-O-sulfate.

obtain highly pure, active recombinant hSR. Furthermore, we provide the first detailed biochemical characterization of recombinant hSR, and we compare its behavior to that of recombinant mouse serine racemase (mSR) previously purified and characterized in our laboratory [17].

Results and discussion

Mouse and human SR sequence alignment and hSR gene synthesis

The gene encoding mSR was isolated from cDNA nearly a decade ago [19]. Soon after, the gene encoding its human ortholog was cloned from human brain [20], and in 2004 an hSR-encoding gene was cloned by a second group of researchers from cDNA derived from differentiated human teratocarcinoma cells [21]. The hSR gene sequences obtained independently by the two groups proved identical. The human serine racemase gene encodes a 340-amino-acid enzyme that boasts 89% sequence identity to its 339-amino-acid mouse ortholog (see Fig. 1). Both enzymes contain a consensus sequence for pyridoxal-5′-phosphate (PLP) binding and exhibit substantial sequence similarity to mammalian serine/threonine dehydratases.

Expression and purification of recombinant mSR and hSR from a bacterial source have been reported [15–18]. Bacterially expressed mammalian SR is active, indicating that post-translational modifications are not essential for racemase activity and that *E. coli* is an appropriate organism for recombinant expression. In order to obtain high-quality recombinant protein, it is often advantageous to match codon usage in the gene of interest to that of the heterologous expression host [22,23]. Comparison of two previous reports [16,18] suggests that this is the case for hSR. Nagayoshi et al. described expression of N-terminal polyhistidine-tagged hSR in *E. coli* BL21(DE3), which resulted in high expression levels with His-hSR localized in the bacterial inclusion bodies [16]. Roughly 50% of the protein was able to be refolded. Dixon et al., on the other hand, addressed the problem of codon bias by expressing His-hSR in BL21-CodonPlus(DE3)-RIL cells, which are supplemented with tRNAs rare in *E. coli*, and they report purification of His-hSR from the soluble fraction [18].

Another way to remedy codon bias is to synthesize a codon-optimized variant of the gene of interest. Recent advances in gene synthesis technology have rendered this process relatively straightforward and inexpensive. Furthermore, the gene synthesis strategy offers researchers several possibilities in addition to codon optimization, such as the ability to adjust the GC content of the gene and the option of adding or removing restriction enzyme recognition sites to facilitate downstream applications. Since the nucleotide sequence encoding hSR is well-known and relatively small (1020 base pairs), we decided to synthesize a codon-optimized variant of the hSR gene using polymerase cycling assembly (PCA) [24].

The nucleotide sequence of the hSR gene was optimized for recombinant expression in *E. coli* using the GeMS web application (http://software.kosan.com/GeMS [25]). Since the optimal length of a synthesized product is about 500–600 bp, the hSR gene sequence was split into two parts, referred to as synthons A and B, of approximately 550 bp each. In order to access two different synthetic strategies, we designed oligonucleotides with complete overlaps for synthon A (40-mers with 20 bp overlaps, 40/20 strategy) and oligonucleotides with single-chain gaps for synthon B (49-mers with 18 bp overlaps, 49/18 strategy). The strategies are outlined in Fig. 2a (see Supplementary Information data for complete oligonucleotide sequences). After two PCR steps (see Materials and methods and Fig. 2a for details), correctly sized products were obtained for both synthons. Two clones of synthon A were sequenced, and both were error-free. Five clones of synthon B were sequenced, and all contained one or two misincorporated base pairs. Site-directed mutagenesis was employed to yield the correct gene sequence.

While we did not sequence nearly enough clones to warrant a statistical analysis, it seems that the 40/20 strategy is less error-prone than the 49/18 strategy. However, the 49/18 strategy holds a key advantage over the 40/20 strategy in that fewer component oligonucleotides need to be synthesized, saving time and money. Addition of an error-cleaning step such as incubation of the product with T7 endonuclease, which cleaves DNA at mispaired bases, to the 49/18 strategy may increase the probability of identifying a correct clone.

```
hSR   MCAQYCISFADVEKAHINIRDSIHLTPVLTSSILNQLTGRNLFFKCELFQKTGSFKIRGA  60
mSR   MCAQYCISFADVEKAHINIQDSIHLTPVLTSSILNQHAGRNLFFKCELFQKTGSFKIRGA  60

hSR   LNAVRSLVPDALERKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPQTAPDCKKLAIQA  120
mSR   LNAIRGLIPDTPEEKPKAVVTHSSGNHGQALTYAAKLEGIPAYIVVPQTAPNCKKLAIQA  120

hSR   YGASIVYCEPSDESRENVAKRVTEETEGIMVHPNQEPAVIAGQGTIALEVLNQVPLVDAL  180
mSR   YGASIVYCDPSDESREKVTQRIMQETEGILVHPNQEPAVIAGQGTIALEVLNQVPLVDAL  180

hSR   VVPVGGGMLAGIAITVKALKPSVKVYAAEPSNADDCYQSKLKGKLMPNLYPPETIADGV  240
mSR   VVPVGGGMVAGIAITIKALKPSVKVYAAEPSNADDCYQSKLKGELTPNLHPPETIADGV  240

hSR   KSSIGLNTWPIIRDLVDDIFTVTEDEIKCATQLVWERMKLLIEPTAGVVAAVLSQHFQT  300
mSR   KSSIGLNTWPIIRDLVDDVFTVTEDEIKYATQLVWGRMKLLIEPTAGVALAAVLSQHFQT  300

hSR   VSPEVKNICIVLSGGNVDLTSSLINVKQAERPASYQSVSV  340
mSR   VSPEVKNVCIVLSGGNVDLTS-LNWVGQAERPAPYQTVSV  339
```

Fig. 1. Sequence alignment of mSR and hSR generated by ClustalW. Identical residues are shown against a black background, and similar residues are shown against a grey background. The PLP-binding consensus sequence (Prosite Accession No. PS00165) is underlined.
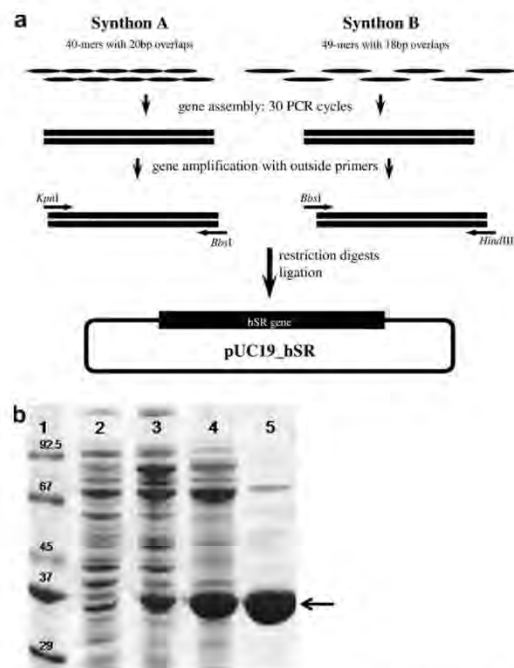
Fig. 2. Preparation and characterization of recombinant hSR. (a) Overview of hSR gene synthesis strategy; (b) 12% SDS–PAGE stained with Coomassie brilliant blue. The total amount of protein loaded into each lane was 10 μg (sample concentrations determined by Bio-Rad Protein Assay). Lane 1, molecular weight standards in kDa; lane 2, crude *E. coli* cell lysate; lane 3, recovered hSR protein following ammonium sulfate precipitation; lane 4, pooled hSR-rich fractions following Q-sepharose anion-exchange chromatography; lane 5, recombinant hSR after ATP-affinity chromatography. The arrow indicates the band corresponding to hSR.

Bacterial expression and purification of recombinant hSR

In our experience, in order to obtain soluble SR, it is necessary to keep expression levels low (unpublished observations). Therefore, we cloned the hSR gene into pMPM-A4-2 [17]. This vector exploits the arabinose-inducible *ara*BAD promoter, which allows induction to be controlled in a regulated, dose-dependent manner. While the overall expression level of hSR from this construct was rather low, most of the hSR obtained was located in the soluble fraction, allowing the enzyme to be purified under native conditions.

We observed that hSR readily precipitates from *E. coli* lysate at relatively low concentrations of ammonium sulfate. Addition of ammonium sulfate to a final concentration of 20% (w/v) allowed for almost complete recovery of hSR while minimizing the precipitation of *E. coli* proteins, resulting in approximately 10% pure hSR (see Fig. 2b, lane 3). Anion-exchange chromatography on Q-sepharose, a step that proved highly successful for purification of recombinant mSR [17], yielded protein fractions significantly enriched in hSR (see Fig. 2b, lane 4). Only fractions with the highest levels of purity were pooled; thus, most of the loss in final yield occurred during this step. The protein mixture was further purified by affinity chromatography. ATP is a known cofactor of SR, and hSR was isolated by binding to commercially available C-8 linked ATP-agarose [17] followed by elution with ATP. The resulting hSR preparation was >90% pure as determined by Coomassie-stained

SDS–PAGE, with a final yield of 2 mg purified hSR per liter of bacterial culture. The purity of this hSR preparation is comparable to that of our previously reported preparation of mSR [17], and we therefore deemed the preparation suitable for comparative kinetic and inhibition analysis, although minor contaminant bands are visible in the overloaded gel (Fig. 2b, lane 5). The purification procedure is summarized in Table 1.

Purification of bacterially expressed hSR has been described by two groups [16,18], and both groups exploited a polyhistidine tag to facilitate purification. The His-hSR preparations were reported to be active, although neither group provided a detailed catalytic characterization. We also attempted to express and purify hSR with a His-tag (data not shown), and we found that the three-step hSR preparation illustrated in Fig. 2b holds two major advantages over the His-tag-assisted purification. First, addition of a His-tag at the N- or C-terminus of the gene sequence resulted in a marked decrease in soluble expression levels along with a concomitant increase in expression in inclusion bodies, which may indicate that the His-tag impairs proper folding of the enzyme. The second and key advantage of the purification protocol described here is the ATP-affinity step, which ensures that only biologically active hSR is isolated.

To check the folding and secondary structure composition of our recombinant SR, we analyzed both enzyme preparations (mSR [17] and hSR) by circular dichroism (CD) spectroscopy. The following secondary structure composition was obtained for hSR: 40.3% α-helix, 6.9% antiparallel β-sheet, 7.0% parallel β-sheet, 15.8% β-turn, and 27.1% random coil. The composition of mSR secondary structure is analogous: 43.9% α-helix, 6.5% antiparallel β-sheet, 6.3% parallel β-sheet, 15.4% β-turn, and 24.3% random coil. These results are comparable to those previously obtained for recombinant murine SR by Cook et al. [15] and to the only available crystallographic data on a eukaryotic serine racemase (SR from the fission yeast *Schizosaccharomyces pombe*; PDB codes 1v71 and 1wtc).

hSR and mSR display similar activity and inhibition profiles

The ability to interconvert L- and D-serine gives serine racemase its name and is believed to be the primary physiological activity of the enzyme. However, SR is also known to act as a β-eliminase, converting L-serine [26] and D-serine [27] to pyruvate and various other amino acid substrates to their corresponding oxoacids [28].

The recombinant hSR described here displays both racemization and elimination activities (see Table 2). Both activities were evaluated by HPLC-based endpoint assay as previously described [17,26]. Recombinant hSR interconverts L-serine and D-serine, although the catalytic efficiency (k_{cat}/K_M) in the L-serine to D-serine direction is marginally higher than that of the reverse reaction. Human SR also catalyzes the β-elimination of L-serine, and to a far lesser extent D-serine, to pyruvate. It is somewhat surprising that D-serine is a poor β-elimination substrate while L-serine readily undergoes elimination, since all SR-catalyzed reactions are believed to proceed through a common quinonoid intermediate. However, our results are consistent with an earlier analysis of the β-elimination activity of mSR by Foltyn et al. [27] and with our previous analysis of mSR [17]. Furthermore, they support the recent hypothesis of Yoshimura and Goto [29], who proposed based on their crystal structure of SR from *S. pombe* that SR utilizes a two-base reaction mechanism. Specifically, they propose that the PLP-binding lysine residue (K56 in mSR and hSR) catalyzes the abstraction and addition of the α-hydrogen of L-serine, while the abstraction and addition of the α-hydrogen of D-serine are catalyzed by a different residue, namely S84 (in mSR/hSR numbering). Mutation of the corresponding serine residue in SR from the slime mold *Dictyostelium discoideum* resulted in loss of racemase and

Table 1
Overview of hSR purification

	Total protein (mg)	Specific activity (μmol/mg h)	Estimated purity (%)	Estimated hSR (mg)	Overall yield
E. coli lysate	1840	b.l.	<2	n.d.	n.d.
(NH₄)₂SO₄ precipitate	336	2.5	~10	34	100%
Q-sepharose pooled fractions	30	6.0	~40	13.5	40%
ATP affinity pooled fractions	13	16.2	>90	13	38%

Human SR was purified from 200 mL of *E. coli* lysate produced from 7 L of culture. The total amount of protein in each sample was determined by Bio-Rad Protein Assay. Specific activity measurements refer to the conversion of L-serine to D-serine (b.l. = below limit of detection). The purity of each sample was estimated by analysis of Coomassie-stained SDS–PAGE with ImageJ software (http://rsb.info.nih.gov/ij, n.d. = not determined). The amount of hSR in each sample was calculated based on the estimated purity, except in the case of the pooled fractions after ATP affinity chromatography. In this case, the amount of hSR was determined by quantitative amino acid analysis.

Table 2
Comparison of hSR and mSR reaction kinetics

Substrate	K_M (mM) hSR	K_M (mM) mSR[*]	k_{cat} (min⁻¹) hSR	k_{cat} (min⁻¹) mSR[*]	k_{cat}/K_M (min⁻¹ mM⁻¹) hSR	k_{cat}/K_M (min⁻¹ mM⁻¹) mSR[*]
L-Serine Racemization	4.1 ± 0.2	3.8 ± 0.1	41.5 ± 1.5	45.5 ± 0.5	10.0 ± 0.8	12.0 ± 0.4
L-Serine Elimination	4.7 ± 0.4	4.0 ± 0.5	100 ± 5.1	81.3 ± 2.8	21.2 ± 2.7	20.3 ± 2.6
D-Serine Racemization	10.8 ± 0.8	14.5 ± 1.1	84.9 ± 4.2	113 ± 3	7.9 ± 0.9	7.8 ± 0.6
D-Serine Elimination	9.5 ± 0.6	3.2 ± 0.3	7.7 ± 0.3	8.8 ± 0.2	0.81 ± 0.09	2.7 ± 0.4
L-Serine-O-sulfate Elimination	0.43 ± 0.05	0.49 ± 0.05	807 ± 49	967 ± 17	1879 ± 352	1973 ± 205

Measurements were carried out at pH 8.0 as described in Materials and methods. The kinetic parameters were determined by nonlinear regression fitting of initial reaction rates into the Michaelis–Menten equation. Values are given as means ± SEM.
[*] Kinetic parameters for recombinant mSR are taken from [17].

D-serine dehydratase activities; however, the mutant was able to catalyze the β-elimination of L-serine [29]. Our kinetic data align well with this two-base mechanism hypothesis; L-serine is processed slightly differently than D-serine as a racemization substrate, and the differences are substantially more pronounced in the β-elimination reaction.
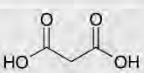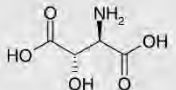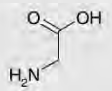
Finally, we tested the activity of hSR toward the non-physiological substrate L-serine-O-sulfate (LSOS). SR effectively catalyzed the elimination of LSOS to pyruvate but did not racemize LSOS. The LSOS-elimination activity of hSR is roughly two orders of magnitude higher than any of its serine racemization or β-elimination activities, and the affinity of LSOS for the enzyme is an order of magnitude higher than that of serine. The overall activity profile of hSR is analogous to that previously obtained by our group for recombinant mSR [17] (see Table 2).

The only significant difference between the orthologs is their apparent K_M for D-serine elimination; hSR has a three-fold higher K_M than mSR, which might suggest that the orthologs have slightly different ways of processing D-serine. While this is intriguing from a biochemical perspective, the physiological relevance remains unclear. In both mouse and human, D-amino acid oxidase (DAAO) is believed to be the enzyme primarily responsible for D-serine breakdown. However, a recent report [12] shows different localization patterns of SR and DAAO, which could suggest a role for SR-catalyzed D-serine elimination in DAAO-poor brain regions.

We next asked if both enzymes are similarly sensitive to specific inhibitors, and we evaluated the inhibition of hSR by three known mSR inhibitors: malonate, L-erythro-3-hydroxyaspartate, and glycine. Malonate and L-erythro-3-hydroxyaspartate have been identified by our group as competitive mSR inhibitors with inhibition constants (K_i) in the low micromolar range [17]. We and others [30] have shown that glycine acts as a competitive inhibitor of SR, with K_i in the high micromolar or low millimolar range. The inhibition of SR by glycine may be physiologically significant, as the concentration of glycine in neuronal tissue has been reported to be 3–6 mM [31,32].

All three compounds proved to be competitive inhibitors of the hSR-catalyzed conversion of L- to D-serine (see Table 3). The overall inhibition profile is the same for mSR and hSR. The most potent inhibitor is L-erythro-3-hydroxyaspartate with K_i in the low micromolar range (43 μM for mSR and 11 μM for hSR). Malonate is only slightly less potent with K_i of 71 μM for mSR and 33 μM for hSR. Glycine is the weakest inhibitor in both cases; it is an order of magnitude less potent than either of the other inhibitors. All three inhibitors exhibit 2- to 4-fold lower K_i values for hSR compared to mSR. In the absence of concrete structural information, it is

Table 3
Competitive inhibition of L-serine racemization by recombinant mSR and hSR

Inhibitor	Structure	K_i (μM) hSR	K_i (μM) mSR[*]
Malonate		33 ± 2.6	71 ± 16
L-erythro-3-hydroxyaspartate		11 ± 0.7	43 ± 7
Glycine		366 ± 30	1640 ± 30

Inhibitors are identified by name and basic chemical structure. Measurements were carried out at pH 8.0. Inhibition constants are given as means ± SEM.
[*] Inhibition constants for recombinant mSR have been previously reported by our group [17] and are reprinted here.

difficult to speculate about the reason for this difference. We have prepared homology models of mSR and hSR based on the crystallographic structure of *S. pombe* SR, which shares ~40% sequence identity with the mouse and human orthologs (unpublished results). All of the residues predicted to engage in substrate/inhibitor binding are conserved in both mSR and hSR; the majority of the residue differences occur at the surface of the molecule. We did, however, observe one difference in the channel leading into the active site; the side chain of tyrosine 231 of hSR points out of the channel, whereas the side chain of histidine 231 of mSR points into the channel, potentially rendering the mSR active site less inhibitor-accessible than the hSR site. To test this hypothesis, we mutated Y231 of hSR to histidine. The mutant displayed similar kinetics to the wild-type in the L-serine to D-serine conversion reaction (K_M = 5.2 ± 0.6 mM, k_{cat} = 42.2 ± 2.1 min^{-1}). Parallel analysis of wild-type hSR and the Y231H mutant showed that both enzymes are similarly sensitive to all three inhibitors (data not shown). High resolution crystal structures of both mSR and hSR will likely need to be solved before the subtle differences between the orthologs can be understood.

In conclusion, we have synthesized a variant of the human serine racemase gene optimized for *E. coli* expression, a strategy that might be of general interest to researchers wishing to express mammalian proteins in bacteria. Furthermore, we report a purification strategy that results in good yields of active hSR. Importantly, we provide the first thorough kinetic characterization of recombinant hSR from a bacterial source and compare its activity to that of mSR. Our activity and inhibition data validate, at least in part, the use of mouse models in SR research and suggest that the mouse ortholog can reasonably be used for the screening of potential hSR inhibitors.

Materials and methods

Synthesis of the hSR gene

Oligonucleotides were ordered from Generi Biotech (Hradec Kralove, Czech Republic) with the minimal amount of purification. The gene was divided into two synthons of approximately 550 bp each, and each synthon was assembled in two steps. Oligonucleotide sequences are listed in Supplementary Information. In the first step, all of the component oligonucleotides except for the first and last were included in the reaction mixture. Thermal cycles were as follows: 30 cycles of 30 s at 94 °C, 30 s at X °C (X = 50 for synthon A, 45 for synthon B), and 120 s at 73 °C. Correctly constructed synthons were amplified in a second PCR step using primers containing the following restriction endonuclease sites: KpnI and NdeI at the 5′ end and XhoI and HindIII at the 3′ end of both synthons. A BbsI restriction site was located at the 3′ end of synthon A before the HindIII site and at the 5′ end of synthon B behind the KpnI site. Synthon A was cloned into pUC19 *via* KpnI and HindIII sites, and its sequence was verified. Synthon B was then cloned into the pUC19_synthon_A construct *via* BbsI and HindIII sites to yield pUC19_hSR. Five clones were sequenced, and all of them contained at least one misincorporated base pair. Errors were fixed by site-directed mutagenesis, using the component oligonucleotides as primers, to yield the correct sequence.

Cloning, expression, and purification of recombinant human serine racemase

The assembled human serine racemase gene was subcloned from pUC19_hSR into pMPM-A4-2 [17] *via* NdeI and XhoI sites to yield pMPMhSR. Clones were verified by restriction digest and sequencing. The plasmid pMPMhSR was transformed into *E. coli*

MC1061, and cells were grown overnight on LB/agar/ampicillin plates. Each plate was then washed into 700 mL of TurboBroth (Athena Enzyme Systems, Baltimore, MD, USA) containing 100 μg/mL ampicillin. The cells were grown at 37 °C to an OD$_{600}$ of approximately 1.0. Protein production was induced by the addition of L-arabinose to a final concentration of 1 mM. The cells were harvested after 6 h at 37 °C and stored at −70 °C until further use.

After thawing, the cells were resuspended in buffer QA (20 mM triethanolamine–HCl, 1 mM MgCl$_2$, 0.1 mM DTT, 0.02 mM PLP, 0.02% sodium azide, pH 7.4); 3 mL QA were used per gram of cells. The cells were broken by lysozyme and sonication. Ammonium sulfate was added to the lysate to a final concentration of 20% (w/v), and the ammonium sulfate mixture was incubated at 4 °C for one hour with continuous stirring. The hSR-containing precipitate was separated by centrifugation at 20,000g for 20 min and subsequently resuspended in 60 mL QA. The sample was dialyzed extensively against QA to remove ammonium sulfate, and then loaded onto 25 mL of Q-sepharose fast flow resin (Amersham Biosciences, Uppsala, Sweden). After washing with several column volumes of QA, hSR was eluted in a linear gradient of 0–500 mM NaCl in buffer QA. hSR-enriched fractions were pooled, concentrated, and dialyzed against buffer QA to remove NaCl. The sample was then loaded onto C8-linked ATP-agarose resin (Sigma, Schnelldorf, Germany) and washed with 5 column volumes of QA. Pure hSR was eluted in a linear gradient of 0–2 mM ATP in buffer QA.

Addition of His-tag and cloning, expression, and purification of Y231H hSR

For convenience, a 6-histidine tag was added to pMPMhSR at the C-terminus of the hSR gene using the insertion mutagenesis protocol according to Wang and Malcolm [33] and the primers 5′-GCGAGCTATCAGTCAGTGTCTCACCACCACCACCACCACTAGAATTC TCGAGGGGGGGGCAGC-3′ and 5′-GCTGCCCCCCCTCGAGAATTCTAG TGGTGGTGGTGGTGGTGAACAGACACTGACTGATAGCTCGC-3′. The Y231H mutation was introduced with the primers 5′-CAACTCATGC CAAACCTG̲CACCCGCCTGAAACAATTGC-3′ and 5′-GCAATTGTTT CAGGCGGGT̲GCAGGTTTGGCATGAGTTTG-3′ (mutagenic bases are underlined) according to the QuikChange site-directed mutagenesis protocol (Stratagene). All DNA modifications were confirmed by sequencing over the entire hSR gene region. The Y231H mutant was expressed under the same conditions as the wild-type and was purified by immobilized metal affinity chromatography. Briefly, cells were lysed in buffer A (50 mM sodium phosphate, 300 mM NaCl, 1 mM MgCl$_2$, 20 μM PLP, 5 mM 2-mercaptoethanol, pH 8.0) containing 10 mM imidazole. The lysate was loaded onto Ni–NTA Superflow resin (Qiagen, Hilden, Germany). The column was washed with 10 volumes of buffer A containing 20 mM imidazole, and SR was eluted in a linear gradient of 20–250 mM imidazole over 15 column volumes. Fractions containing pure SR were pooled and dialyzed against buffer QA.

Activity and inhibition assays

SR activity was analyzed at 37 °C in a pH 8.0 reaction buffer containing 100 mM Hepes–NaOH, 10 μM PLP, 1 mM MgCl$_2$, 5 mM DTT, and 1 mM ATP. 470 nM enzyme was added to the buffer for analysis of racemization and elimination of L- and D-serine; a 10-fold lower enzyme concentration was used to assess elimination of LSOS to pyruvate. Enzyme concentrations were determined by Bio-Rad Protein Assay (Bio-Rad Laboratories, California, USA) and amino acid analysis. Reactions were started by the addition of substrate (concentrations ranging from 0.2 to 10K_M) and were further treated as described in [17].

For analysis of amino acid racemization, the amino acid mixture was derivatized by 1-fluoro-2,4-dinitrophenyl-5-L-alanine amide

(Marfey's reagent, Pierce, USA) and resolved using reversed-phase HPLC as described [26]. Formation of pyruvate was determined by derivatization of reaction mixtures with 2,4-dinitrophenylhydrazine as previously described [17].

For determination of mechanism of inhibition and K_i, the solution of enzyme in reaction buffer was pre-incubated with inhibitor for 15 min before addition of L-serine, and reactions were allowed to proceed for 25–60 min. The reaction mixtures were then analyzed for D-serine formation. The observed K_M and V_{max} were calculated at three different inhibitor concentrations.

All kinetic and inhibition data were analyzed with the Grafit software package (version 5.0.11, Erithacus Software, Surrey, UK).

Circular dichroism measurements

Samples for CD were prepared in 20 mM triethanolamine-HCl, pH 7.5, containing 1 mM MgCl$_2$, 1 mM DTT, and 10 μM PLP. ATP was excluded from the buffer due to its strong absorbance in the UV region. Far-UV/VIS CD spectra were recorded at room temperature on a JASCO-810 dichrometer with time constant 1 s, step size 0.5 nm, and scanning speed 100 nm/s. The optical path length was 0.1 mm. The protein concentrations used were 0.55 mg/ml for mSR and 0.5 mg/ml for hSR. The spectra are expressed as molar ellipticity per residue and are the average of four computer-controlled scans. The secondary structure percentages were calculated using spectral deconvolution according to [34–37].

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.pep.2008.09.003.

References

[1] T. Matsui, M. Sekiguchi, A. Hashimoto, U. Tomita, T. Nishikawa, K. Wada, Functional comparison of D-serine and glycine in rodents: the effect on cloned NMDA receptors and the extracellular concentration, J. Neurochem. 65 (1995) 454–458.
[2] J.P. Mothet, A.T. Parent, H. Wolosker, R.O. Brady Jr., D.J. Linden, C.D. Ferris, M.A. Rogawski, S.H. Snyder, D-Serine is an endogenous ligand for the glycine site of the N-methyl-D-aspartate receptor, Proc. Natl. Acad. Sci. USA 97 (2000) 4926–4931.
[3] M.J. Scolari, G.B. Acosta, D-Serine: a new word in the glutamatergic neuro-glial language, Amino Acids 33 (2007) 563–574.
[4] H. Wolosker, E. Dumin, L. Balan, V.N. Foltyn, D-Amino acids in the brain: D-serine in neurotransmission and neurodegeneration, FEBS J. 275 (2008) 3514–3526.
[5] M.J. Schell, The N-methyl D-aspartate receptor glycine site and D-serine metabolism: an evolutionary perspective, Philos. Trans. R. Soc. Lond. B. Biol. Sci. 359 (2004) 943–964.
[6] W. Danysz, A.C. Parsons, Glycine and N-methyl-D-aspartate receptors: physiological significance and possible therapeutic applications, Pharmacol. Rev. 50 (1998) 597–664.
[7] Y. Yang, W. Ge, Y. Chen, Z. Zhang, W. Shen, C. Wu, M. Poo, S. Duan, Contribution of astrocytes to hippocampal long-term potentiation through release of D-serine, Proc. Natl. Acad. Sci. USA 100 (2003) 15194–15199.
[8] R.A. Swanson, W. Ying, T.M. Kauppinen, Astrocyte influences on ischemic neuronal death, Curr. Mol. Med. 4 (2004) 193–205.
[9] J.H. Krystal, D.C. D'Souza, D-Serine and the therapeutic challenge posed by the N-methyl-D-aspartate antagonist model of schizophrenia, Biol. Psychiatry 44 (1998) 1075–1076.
[10] K. Hashimoto, T. Fukushima, E. Shimizu, S. Okada, N. Komatsu, N. Okamura, K. Koike, H. Koizumi, C. Kumakiri, K. Imai, M. Iyo, Possible role of D-serine in the pathophysiology of Alzheimer's disease, Prog. Neuropsychopharmacol. Biol. Psychiatry 28 (2004) 385–388.
[11] J. Sasabe, T. Chiba, M. Yamada, K. Okamoto, I. Nishimoto, M. Matsuoka, S. Aiso, D-Serine is a key determinant of glutamate toxicity in amyotrophic lateral sclerosis, EMBO J. 26 (2007) 4149–4159.
[12] L. Verrall, M. Walker, N. Rawlings, I. Benzel, J.N. Kew, P.J. Harrison, P.W. Burnet, D-Amino acid oxidase and serine racemase in human brain: normal distribution and altered expression in schizophrenia, Eur. J. Neurosci. 26 (2007) 1657–1669.
[13] A.E. Steffek, V. Haroutunian, J.H. Meador-Woodruff, Serine racemase protein expression in cortex and hippocampus in schizophrenia, Neuroreport 17 (2006) 1181–1185.
[14] I. Bendikov, C. Nadri, S. Amar, R. Panizzutti, J. De Miranda, H. Wolosker, G. Agam, A CSF and postmortem brain study of D-serine metabolic parameters in schizophrenia, Schizophr. Res. 90 (2007) 41–51.
[15] S.P. Cook, I. Galve-Roperh, A. Martínez del Pozo, I. Rodríguez-Crespo, Direct calcium binding results in activation of brain serine racemase, J. Biol. Chem. 277 (2002) 27782–27792.
[16] C. Nagayoshi, M. Ishibashi, Y. Kita, M. Matsuoka, I. Nishimoto, M. Tokunaga, Expression, refolding and characterization of human brain serine racemase in Escherichia coli with N-terminal His-tag, Protein Pept. Lett. 12 (2005) 487–490.
[17] K. Strísovský, J. Jirásková, A. Mikulová, L. Rulísek, J. Konvalinka, Dual substrate and reaction specificity in mouse serine racemase: identification of high-affinity dicarboxylate substrate and inhibitors and analysis of the beta-eliminase activity, Biochemistry 44 (2005) 13091–13100.
[18] S.M. Dixon, P. Li, R. Liu, H. Wolosker, K.S. Lam, M.J. Kurth, M.D. Toney, Slow-binding human serine racemase inhibitors from high-throughput screening of combinatorial libraries, J. Med. Chem. 49 (2006) 2388–2397.
[19] H. Wolosker, S. Blackshaw, S.H. Snyder, Serine racemase: a glial enzyme synthesizing D-serine to regulate glutamate-N-methyl-D-aspartate neurotransmission, Proc. Natl. Acad. Sci. USA 96 (1999) 13409–13414.
[20] J. De Miranda, A. Santoro, S. Engelender, H. Wolosker, Human serine racemase: molecular cloning, genomic organization and functional analysis, Gene 256 (2000) 183–188.
[21] M. Xia, Y. Liu, D.J. Figueroa, C.S. Chiu, N. Wei, A.M. Lawlor, P. Lu, C. Sur, K.S. Koblan, T.M. Connolly, Characterization and localization of a human serine racemase, Mol. Brain Res. 125 (2004) 96–104.
[22] J.F. Kane, Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli, Curr. Opin. Biotechnol. 6 (1995) 494–500.
[23] C. Gustafsson, S. Govindarajan, J. Minshull, Codon bias and heterologous protein expression, Trends Biotechnol. 22 (2004) 346–353.
[24] H.O. Smith, C.A. Hutchison III, C. Pfannkoch, J.C. Venter, Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides, Proc. Natl. Acad. Sci. USA 100 (2003) 15440–15445.
[25] S.J. Kodumal, K.G. Patel, R. Reid, H.G. Menzella, M. Welch, D.V. Santi, Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster, Proc. Natl. Acad. Sci. USA 101 (2004) 15573–15578.
[26] K. Strísovský, J. Jirásková, C. Barinka, P. Majer, C. Rojas, B.S. Slusher, J. Konvalinka, Mouse brain serine racemase catalyzes specific elimination of L-serine to pyruvate, FEBS Lett. 535 (2003) 4–48.
[27] V.N. Foltyn, I. Bendikov, J. De Miranda, R. Panizzutti, E. Dumin, M. Shleper, P. Li, M.D. Toney, E. Kartvelishvily, H. Wolosker, Serine racemase modulates intracellular D-serine levels through an alpha,beta-elimination activity, J. Biol. Chem. 280 (2005) 1754–1763.
[28] R. Panizzutti, J. De Miranda, C.S. Ribeiro, S. Engelender, H. Wolosker, A new strategy to decrease N-methyl-D-aspartate (NMDA) receptor coactivation: inhibition of D-serine synthesis by converting serine racemase into an eliminase, Proc. Natl. Acad. Sci. USA 98 (2001) 5294–5299.
[29] T. Yoshimura, M. Goto, D-Amino acids in the brain: structure and function of pyridoxal phosphate-dependent amino acid racemases, FEBS J. 275 (2008) 3527–3537.
[30] D.S. Dunlop, A. Neidle, Regulation of serine racemase activity by amino acids, Brain Res. Mol. Brain Res. 133 (2005) 208–214.
[31] S. Verleysdonk, H. Martin, W. Willker, D. Leibfritz, B. Hamprecht, Rapid uptake and degradation of glycine by astroglial cells in culture: synthesis and release of serine and lactate, Glia 27 (1999) 239–248.
[32] R. Dringen, S. Verleysdonk, B. Hamprecht, W. Willker, D. Leibfritz, A. Brand, Metabolism of glycine in primary astroglial cells: synthesis of creatine, serine, and glutathione, J. Neurochem. 70 (1998) 835–840.
[33] W. Wang, B.A. Malcolm, Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange site-directed mutagenesis, Biotechniques 26 (1999) 680–682.
[34] N. Greenfield, G.D. Fasman, Computed circular dichroism spectra for the evaluation of protein conformation, Biochemistry 8 (1969) 4108–4116.
[35] J.P. Hennessey Jr., W.C. Johnson Jr., Information content in the circular dichroism of proteins, Biochemistry 20 (1981) 1085–1094.
[36] S.W. Provencher, J. Glockner, Estimation of globular protein secondary structure from circular dichroism, Biochemistry 20 (1981) 33–37.
[37] G. Bohm, R. Muhr, R. Jaenicke, Quantitative analysis of protein far UV circular dichroism spectra by neural networks, Protein Eng. 5 (1992) 191–195.

180

P11. Discrete momentum representation of the Lippmann-Schwinger equation and its application to electron-molecule scattering.

M. Polášek, M. Juřek, M. Ingr, P. Čársky, J. Horáček, Phys. Rev. A 61 (2000) 032701.

Discrete momentum representation of the Lippmann-Schwinger equation and its application to electron-molecule scattering

Martin Polášek

Faculty of Philosophy and Science, Silesian University at Opava, Bezručovo nám. 13, 746 01 Opava, Czech Republic

Michael Juřek, Marek Ingr, and Petr Čársky

J. Heyrovský Institute of Physical Chemistry, Academy of Sciences of the Czech Republic, 18223 Prague 8, Czech Republic

Jiří Horáček

Faculty of Mathematics and Physics, Charles University, V Holešovičkách 2, 18000 Prague 8, Czech Republic

We present a way of using numerical quadrature in momentum space for solving the three-dimensional Lippmann-Schwinger equation. The integration is performed in spherical coordinates. Test calculations show that the quadrature is well suited for the electron-molecule scattering problems. Sample results of elastic scattering of electrons are presented for the empirical Yukawa potential and *ab initio* Hartree-Fock potential of the hydrogen and methane molecules.

PACS number(s): 34.80.Bm

I. INTRODUCTION

Progress in calculations of electron-molecule collisions has been recently reviewed in several books and review articles (see, for example, Refs. [1] and [2]). Recent variational treatments of electron-molecule collisions (see, for example, Refs. [3–5]) usually use Gaussian-type functions as the variational basis set. Although the use of Gaussians in bound-state calculations has become a routine task, their utilization in scattering problems is not so simple. One needs a large set of diffuse functions to represent properly all the operators appearing in the variational functional. Moreover, the S-matrix Kohn method requires additional continuum functions with correct asymptotic behavior. The choice of the resulting set is connected with some uncertainty and may lead to linear dependency. For these reasons, it is desirable to separate basis functions used for the construction of the Hartree-Fock potential and those appearing in the solution of the scattering equations. We attempted this by developing the cubic grid Gaussian basis sets (CGGBS) [6–10] consisting of s-type Gaussians centered at the points of a regular cubic lattice. The basic construction principle of the CGGBS—the best fit of the plane wave—allows the expression of the Green's function in a separable form and the acquisition of T elements by simple inversion of the Lippmann-Schwinger (LS) equation. Unfortunately, this method suffers from two main disadvantages: without semiempirical adjustment it yields infinite diagonal elements of the Green's function, and the interaction potential matrix expressed in CGGBS becomes nearly singular at lower energies.

To overcome these problems we decided to modify the method originally proposed by Walters [11] who used a numerical quadrature in **k** space to solve the Lippmann-Schwinger equation. To point out the formal analogy with the discrete variable representation, we refer to the proposed method as the discrete momentum representation (DMR) method. The method leads to a matrix equation for scattering amplitudes similar to that in the T-matrix expansion [12].

The discrete set of quadrature vectors in the **k** space may be considered as a basis set which allows to represent the T-operator matrix once the molecular potential is available in any standard basis set.

II. THEORY

A. Standard numerical solution of the Lippmann-Schwinger equation

The problem of solving the Lippmann-Schwinger equation [13]

$$T(E) = U + UG_0(E)T(E),\qquad(1)$$

where U stands for double of the potential, is encountered in many branches of physics and chemistry and its efficient solution is of a vital importance. The standard way of tackling this problem consists in carrying out the partial-wave expansion of all the quantities and solving a coupled set of one-dimensional partial wave LS equations of the following type

$$t_l(p,p';E) = u_l(p,p') + \frac{1}{\pi}\int_0^\infty \frac{u_l(p,k)t_l(k,p';E)k^2dk}{k_2^2 - k^2 + i\epsilon}\qquad(2)$$

in momentum space [14]. This equation is singular (singular integrand and infinite integration range) and special attention must be paid to the correct treatment of the singularity. There exist several more or less satisfactory ways of treating the singularity of the integral kernel (see, for example, Refs. [15–19]). Here we quote a widely used approach [19] consisting of the following. The infinite integration range $(0;\infty)$ is reduced to $(-1;1)$ by means of the transformation

$$x = \frac{k - k_0}{k + k_0},\qquad(3)$$

where

$$k_0 = \sqrt{\frac{2\mu E}{\hbar^2}}, \qquad (4)$$

which sets the kernel singularity to the middle of the integration range, $x = 0$. For computation of the resulting integrals which are of the type

$$P \int_{-1}^{1} \frac{f(x)dx}{x}, \qquad (5)$$

the use of $2n$-point Gaussian rules is recommended, since not only do they have maximum degree of exactness, but because of the symmetry they also integrate exactly the function $1/x$ [20]. It is also possible to subtract the singularity [14] by writing

$$P \int_{-1}^{1} \frac{f(x)dx}{x} = \int_{-1}^{1} \frac{f(x) - f(0)}{x} dx. \qquad (6)$$

Then also $(2n+1)$-point Gaussian rules may be used. The symmetric distribution of quadrature nodes with respect to zero permits us to neglect the subtraction term. The partial wave LS equation is then solved for a series of angular momentum quantum number l and the cross sections obtained by summing contributions from all partial waves. This approach is very efficient provided the number of contributing partial waves is low. This is true for very low-energy electrons and spherical targets. At higher energies and for heavier particles the number of partial waves increases rapidly and soon the method becomes impractical.

B. Essence of the discrete momentum representation

We propose a different method which avoids the partial-wave expansion and which is directly applicable to non-spherical targets. The idea is to solve the LS equation in the full three-dimensional momentum space:

$$\langle \mathbf{k}_1 | T | \mathbf{k}_2 \rangle = \langle \mathbf{k}_1 | U | \mathbf{k}_2 \rangle + \int \frac{\langle \mathbf{k}_1 | U | \mathbf{k} \rangle \langle \mathbf{k} | T | \mathbf{k}_2 \rangle d\mathbf{k}}{k_0^2 - k^2 + i\epsilon}. \qquad (7)$$

To do so, we express the LS equation in spherical coordinates in the $\mathbf{k}$ space and treat separately the radial and angular integrations. Let us write the T-matrix element as

$$\langle \mathbf{k}_1 | T | \mathbf{k}_2 \rangle = \langle \mathbf{k}_1 | U | \mathbf{k}_2 \rangle + \int_0^\infty \frac{k^2 f(\mathbf{k}_1, \mathbf{k}_2; k) dk}{k_0^2 - k^2 + i\epsilon}, \qquad (8)$$

where the function $f(\mathbf{k}_1, \mathbf{k}_2; k)$ is defined as

$$f(\mathbf{k}_1, \mathbf{k}_2; k) = \int d\Omega \langle \mathbf{k}_1 | U | k\mathbf{n}(\Omega) \rangle \langle k\mathbf{n}(\Omega) | T | \mathbf{k}_2 \rangle, \quad (9)$$

and the symbol $\mathbf{n}(\Omega)$ stands for the unit vector in the direction Ω. The solution of Eq. (8) may be in principle performed by the numerical technique described in the previous subsection. An efficient method for angular numerical quadrature was developed by Lebedev [21] and has been applied to problems in electronic-structure [22] and electron-scattering theories [23].

Using this numerical quadrature, we solve a set of linear algebraic equations and we obtain the T matrix on a three-dimensional grid in momentum space. For this approach, to be efficient, the total number of mesh points must be low. It can be achieved by restricting the radial integration region by using the following substitution:

$$k = k_0 \frac{a + bx}{a - bx}. \qquad (10)$$

With respect to Eq. (3), the transformation (10) diminishes considerably the extent of the numerical quadrature by cutting off the high-energy region. Without this cutoff it is difficult to reach a convergence of the numerical procedure. The transformation (10) also cuts off the low-energy region, though this is less justifiable than the cutoff of the high-energy region. However, the parameters a and b may be set so that the disregarded region is small and its exclusion does not affect the result. By numerical tests we chose $a = 10.3333$ and $b = 9.6667$ as default values since they represent a good compromise between the computational effort (given by the number of quadrature points) and accuracy (see Sec. II D for details). The transformation (10) gives a symmetric distribution of the quadrature nodes and the subtraction term in Eq. (6) vanishes.

Application of both the radial and angular quadratures to Eq. (7) converts it to the matrix form

$$\mathbf{T} = \mathbf{U} + \mathbf{U}\mathbf{G}_0^+ \mathbf{T}, \qquad (11)$$

in which the matrix elements are defined as

$$T_{qj,pi} = \langle k_q \mathbf{n}_j | T | k_p \mathbf{n}_i \rangle, \qquad (12)$$

$$V_{qj,pi} = \langle k_q \mathbf{n}_j | V | k_p \mathbf{n}_i \rangle, \qquad (13)$$

$$(G_0^+)_{qj,pi} = \begin{cases} -\delta_{qp}\delta_{ji} w_i k_0 i/2 & \text{if } p = 0, \\ \delta_{qp}\delta_{ji} \dfrac{2abw_p w_i k_p^2 k_0}{(a - bx_p)^2(k_0^2 - k_p^2)} & \text{if } p \neq 0. \end{cases} \qquad (14)$$

Subscripts p, q and i, j are indices of nodes in the radial and angular quadratures, respectively, w is the weight, and x_p is defined by Eq. (10). V is the interaction potential and $U = 2V$. The matrix $\mathbf{T}$ may be obtained by the matrix inversion

$$\mathbf{T} = (\mathbf{1} - \mathbf{U}\mathbf{G}_0^+)^{-1} \mathbf{U}. \qquad (15)$$

A peculiar feature of the DMR approach is that we do not obtain directly the scattering amplitude for an *a priori* selected scattering angle, but for a set of discrete angular coordinates given by the numerical quadrature. The scattering amplitude for a particular scattering angle can be obtained by an additional computational procedure (see Sec. II E).

C. Scattering of electrons on molecular targets

Calculations of electron-molecule collisions are usually performed with the use of the optical potential in static-exchange approximation

TABLE I. Dependence of the differential (DCS) and integral (ICS) cross sections of the elastic e-CH_4 scattering on the extent of the integration range. a, b—parameters of Eq. (10); k_{min},k_{max}—lower and upper limit of the integration range; N_{rad},N_{ang}—number of radial and angular quadrature points necessary for the convergence.

E (eV)	a	b	k_{min} (a.u.)	k_{max} (a.u.)	N_{rad}	N_{ang}	DCS [Å²/sr] 0°	90°	180°	ICS (Å²)
5	10.67	9.33	0.040	9.09	21	110	2.73	0.882	2.67	17.16
	10.5	9.5	0.030	12.12	21	110	2.64	0.892	2.4	16.6
	10.4	9.6	0.024	15.15	21	110	2.51	0.89	2.2	16.0
	10.33	9.67	0.020	18.19	25	110	2.48	0.88	2.15	15.7
20	11.00	9.00	0.121	12.12	21	146	10.50	0.495	1.436	16.49
	10.67	9.33	0.081	18.19	23	194	10.56	0.40	1.44	16.6
	10.50	9.50	0.061	24.25	25	194	10.59	0.40	1.42	16.7
	10.40	9.60	0.048	30.31	25	194	10.58	0.40	1.41	16.7
	10.33	9.70	0.040	36.37	29	194	10.59	0.41	1.40	16.7

$$V = V_s + V_{ex}. \qquad (16)$$

The static part includes electrostatic interaction of the scattered electron with the charge density

$$\rho(\mathbf{r}) = -\sum_A Z_A \delta(\mathbf{r} - \mathbf{R}_A) + 2\sum_i |\phi_i(\mathbf{r})|^2, \qquad (17)$$

where Z_A, $\mathbf{R}_A$, and ϕ_i have their usual meaning of nuclear charges, atomic positions, and occupied molecular orbitals. The corresponding matrix element is expressed as

$$\langle \mathbf{k}_1 | V_s | \mathbf{k}_2 \rangle = \frac{4\pi}{K^2} \int e^{i\mathbf{k}\cdot\mathbf{r}} \rho(\mathbf{r}) d\mathbf{r}, \qquad (18)$$

where $\mathbf{K} = \mathbf{k}_2 - \mathbf{k}_1$. Using second-order Taylor expansion for the exponential in Eq. (18) one can show that the last expression diverges for the targets with nonzero dipole moment at $K = 0$, whereas for systems with zero dipole moment the term becomes undefined. Since the forward V_S element is used only for integration and represents a zero-measure set, we can replace it by a point very close to $K = 0$. In the limit $K \rightarrow 0$ the forward Coulombic term yields the following formula,

$$\langle \mathbf{k} | V_s | \mathbf{k} \rangle = -\frac{2\pi}{k^2} \mathbf{k}^T \mathbf{M} \mathbf{k}, \qquad (19)$$

where $\mathbf{M}$ is the molecular second moment.

Finally, let us substitute for charge density of a closed shell target molecule from Eq. (17) and apply the linear combination of atomic orbitals expansion of molecular orbitals. The Coulombic matrix elements (18) in a nonforward direction become

$$\langle \mathbf{k}_1 | V_s | \mathbf{k}_2 \rangle = \frac{4\pi}{K^2} \left(-\sum_A Z_A e^{i\mathbf{k}\cdot\mathbf{R}_A} \right.$$
$$\left. + \sum_\alpha \sum_\beta P_{\alpha\beta} \left\langle \mathbf{k}_1\alpha \left| \frac{1}{r_{12}} \right| \mathbf{k}_2\beta \right\rangle \right). \qquad (20)$$

The elements of density matrix $P_{\alpha\beta}$ can be used also to calculate second moments. The exchange integrals may be evaluated in a similar manner

$$\langle \mathbf{k}_1 | V_{ex} | \mathbf{k}_2 \rangle = -\frac{1}{2} \sum_\alpha \sum_\beta P_{\alpha\beta} \left\langle \mathbf{k}_1\alpha \left| \frac{1}{r_{12}} \right| \beta\mathbf{k}_2 \right\rangle, \qquad (21)$$

and no special treatment is required for diagonal elements.

Packages for Hartree-Fock calculations use almost exclusively Gaussian basis sets. The formulas for hybrid Coulombic and exchange integrals are available in the literature [12,24–26].

D. Test of the truncated radial integration

In this section we comment on setting the constants a and b in Eq. (10). A pair of the constants a and b was chosen, which defines the integration range. For this integration range the calculations of the differential cross section were performed with increasing number of both the radial and angular quadrature mesh points, until the convergence was reached. Then the integration range was extended and a converged result was calculated again. This procedure was repeated so many times, until the result was stable. The constants a and b so obtained were accepted as standard values for all other calculations. These tests were done on the methane molecule with the molecular valence double ζ basis set [27]. The results for 5 and 20 eV are summarized in Table I.

E. Averaging over the molecular orientation

The results for electron scattering on molecules in the gas phase should account for random orientation of molecular targets. This fact requires an additional procedure to obtain differential cross sections averaged over molecular coordinates. As is usual [23], we keep the molecular target fixed and integrate the differential cross section over the pairs of $\mathbf{k}$ vectors with a fixed scattering angle. The procedure is performed in the following manner. Around each vector from the angular set we assume a circle on the sphere with the

TABLE II. Differential cross sections (in Å^2/sr) of elastic scattering of 1-eV electrons on the Yukawa potential. The results for different numbers of radial quadrature nodes (N) are compared with those obtained by partial-wave expansion (PWE). The number of angular quadrature nodes has been set to 38.

Scatt. angle (deg)	$N=5$	$N=7$	$N=9$	$N=11$	$N=15$	PWE
0.00	0.3969	0.3946	0.3939	0.3938	0.3937	0.3936
35.26	0.3658	0.3639	0.3632	0.3631	0.3630	0.3629
54.74	0.3296	0.3281	0.3274	0.3273	0.3273	0.3271
70.53	0.2971	0.2960	0.2953	0.2952	0.2951	0.2950
90.00	0.2587	0.2580	0.2573	0.2573	0.2572	0.2571
109.47	0.2262	0.2258	0.2251	0.2251	0.2250	0.2249
125.26	0.2054	0.2053	0.2047	0.2046	0.2046	0.2045
144.74	0.1873	0.1874	0.1867	0.1867	0.1866	0.1866
180.00	0.1747	0.1749	0.1743	0.1742	0.1742	0.1741

radius k_0. The radius of the circle corresponds to the scattering angle. By numerical integration along this circle we obtain the averaged differential cross section for a particular **k** vector. This is performed consecutively for all vectors from the angular set and the final value of the differential cross section is obtained by averaging the values calculated for individual **k** vectors. The procedure is general and may be used for any set of scattering angles. However, it is profitable to select scattering angles given by the angular quadrature, because such a selection reduces the number of the

FIG. 2. Angular dependence of the differential cross section of elastic electron scattering on H_2 at 10 eV. See Fig. 1 for details. The crosses represent results of static-exchange calculations of Ref. [32].

U-matrix elements needed in the numerical integration on the circles. Presently we are working on a more efficient averaging procedure.

III. RESULTS AND DISCUSSION

As the first test of the radial and angular quadratures described in this paper we calculated angular dependence of the

FIG. 1. Angular dependence of the differential cross section of elastic electron scattering on H_2 at 5 eV. The calculated results were obtained with 11 radial and 38 angular points. The lines representing results obtained for 13, 15, and 17 radial points and higher numbers of angular points are indistinguishable from the presented line. The dashed line represents the static result without inclusion of the exchange term. The experimental data are taken from Refs. [28] (squares) and [29] (triangles).

FIG. 3. Angular dependence of the differential cross section of elastic electron scattering on H_2 at 15 eV. See Fig. 1 for details.
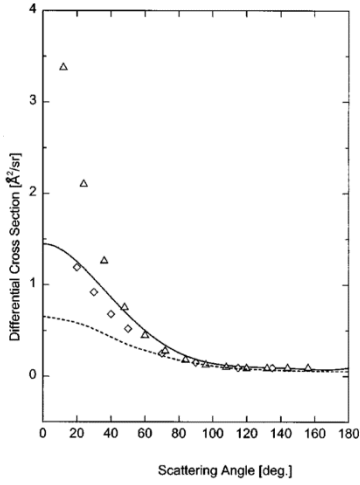
FIG. 4. Angular dependence of the differential cross section of elastic electron scattering on H$_2$ at 20 eV. See Fig. 1 for details.

differential cross section for elastic electron scattering on the Yukawa potential

$$V = \frac{e^{-r}}{r}. \tag{22}$$

The calculated elastic cross sections at 1 eV compared with those obtained by partial wave expansion are shown in Table II. We present results for five sets of radial quadrature nodes. Because of spherical symmetry of the potential, the calculated cross sections depend only very little on the number of
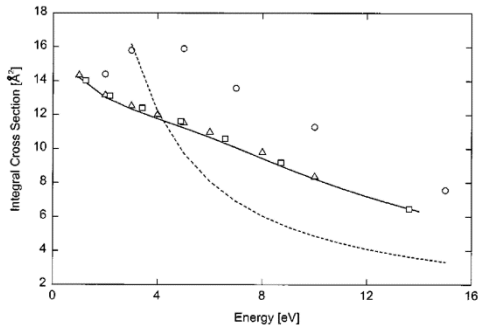
FIG. 5. Elastic integral cross section for e-H$_2$ scattering. The solid line represents results obtained by use of the quadrature with 11 radial and 38 angular points. Higher quadratures give lines that are indistinguishable from the presented line. The dashed line represents the static result without inclusion of the exchange term. The experimental data (circles) are taken from Ref. [29]. The triangles and squares represent the results of static-exchange calculations of Refs. [31] and [32], respectively.
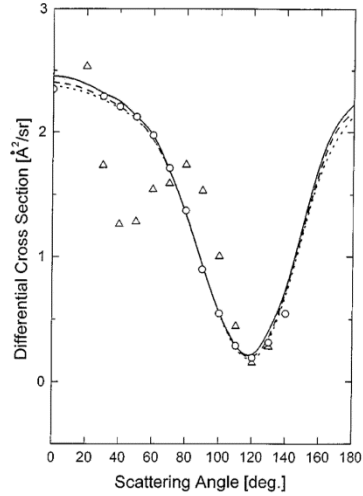
FIG. 6. Angular dependence of the differential cross section of elastic electron scattering on the methane molecule at 5 eV. The calculated results obtained by different numbers of angular points are represented as follows: 86, dotted line; 110, long dashes; 146, solid line. The experimental data are taken from Ref. [30] (triangles). The circles represent the results of static-exchange calculations of Ref. [33]. The static results are out of scale in this figure.
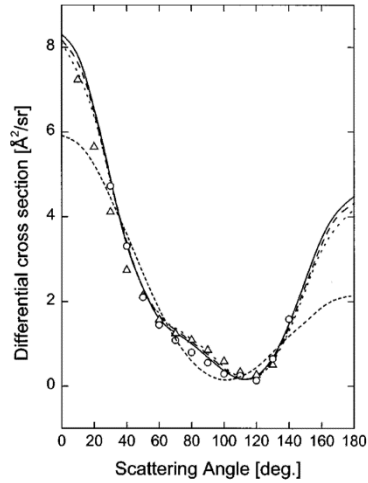
FIG. 7. Angular dependence of the differential cross section of elastic electron scattering on the methane molecule at 10 eV. See Fig. 6 for details. The dashed line represents the static result without inclusion of the exchange term.

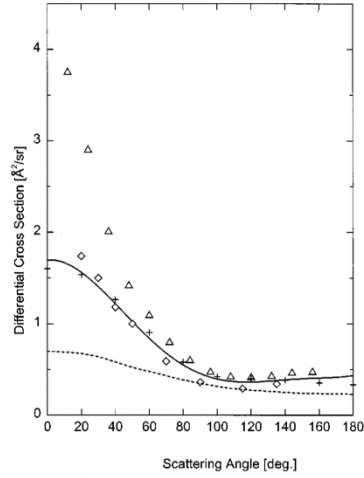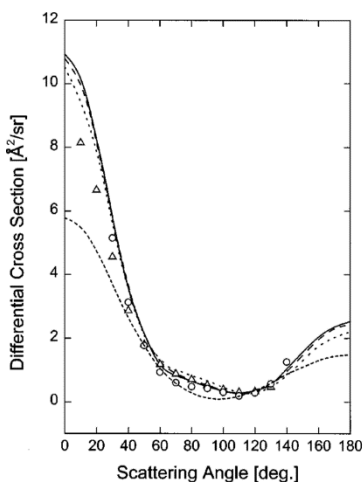FIG. 8. Angular dependence of the differential cross section of elastic electron scattering on the methane molecule at 15 eV. See Figs. 6 and 7 for details.

FIG. 9. Angular dependence of the differential cross section of elastic electron scattering on the methane molecule at 20 eV. See Figs. 6 and 7 for details.

angular nodes, and it is not necessary to apply Lebedev quadratures with more than 38 points. As can be seen in Table II, the DMR approach yields results exact to four digits.

In order to show the applicability of the DMR approach to real molecular systems, we calculated the cross sections of elastic electron scattering on hydrogen and methane molecules. Since the present version of our computer code allows only calculations in the static-exchange approximation, the polarization effects are not taken into account. For both H_2 and CH_4 the standard valence double ζ basis set [27] has been used for construction of the Hartree-Fock potential (16). The calculated cross sections for the elastic electron scattering on H_2 are presented in Figs. 1–4 along with two sets of experimental data. The angular quadrature set with 38 points is large enough in this case, and the extension of the angular set does not change the results. The curves in each one of Figs. 1–4 represent results obtained with different numbers of radial points ranging from 11 to 17. The minimal radial set with 11 points is large enough to represent the converged results. The absence of polarization contributions to the potential causes well-known underestimation of the calculated cross section near the forward direction. As can be seen in Fig. 5, the calculated cross section of elastic e-H_2 scattering follows closely the results of two other approaches [31,32]. Since the resulting integral cross section is mostly affected by near-forward contributions and no polarization effects have been introduced in the calculations, the calculated values differ significantly from the experimental ones. In all figures we also present the static results (i.e., cross sections obtained without the exchange potential), suggested for future comparison with other techniques.

The second example, dependence of the cross section of the elastic electron scattering on the methane molecule,
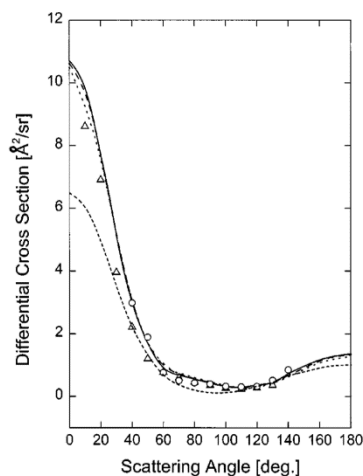
shows good convergence of the results with the increasing size of the angular set. The calculated differential cross sections for three angular sets are plotted in Figs. 6–9. The number of radial nodes has been set to 21 in all three cases. As it is seen, DMR calculations provide results very close to experimental data. The agreement is worse for 5 eV because of the neglect of polarization effects.

IV. CONCLUSION

We presented a method for the calculation of the cross sections of electron-molecule collisions based on a numerical quadrature applied to the Lippmann-Schwinger equation in momentum space. The suggested distribution of quadrature nodes maintains a relatively high precision of the numerical procedure. The method seems to be feasible for calculations of electron scattering on polyatomic molecules. Since the size of the resulting set of linear equations is not related to the number of atoms of the molecule, the presented treatment could be suitable for economical calculations of electron-molecule scattering on personal computers.

At the present time the method is restricted to calculations of elastic electron or positron scattering cross sections on nonpolar targets. Further extension to polar systems as well as the treatment of various types of inelastic collisions and inclusion of polarization effects will be the subject of subsequent papers.

[1] *Computational Methods for Electron-Molecule Collisions*, edited by W. M. Huo and F. A. Gianturco (Plenum, New York, 1995).

[2] *Novel Aspects of Electron-Molecule Collisions*, edited by K. H. Becker (World Scientific, Singapore, 1998).

[3] T. N. Rescigno, B. H. Lengsfield III, and C. W. McCurdy, in *Modern Electronic Structure Theory*, edited by D. R. Yarkony (World Scientific, Singapore, 1995), Pt. I.

[4] C. Winstead and V. McKoy, in *Modern Electronic Structure Theory* (Ref. [3]), Pt. II.

[5] M.-T. Lee, L. M. Brescansin, and L. E. Machado, Phys. Rev. A **59**, 1208 (1999).

[6] P. Čársky, V. Hrouda, and J. Michl, Int. J. Quantum Chem. **53**, 419 (1995).

[7] P. Čársky, V. Hrouda, and J. Michl, Int. J. Quantum Chem. **53**, 431 (1995).

[8] P. Čársky, V. Hrouda, J. Michl, and D. Antic, Int. J. Quantum Chem. **53**, 437 (1995).

[9] V. Hrouda, M. Polášek, P. Čársky, and J. Michl, Theor. Chim. Acta **89**, 401 (1994).

[10] P. Čársky, V. Hrouda, M. Polášek, D. E. David, D. Antic, and J. Michl, J. Phys. Chem. A **101**, 3754 (1997).

[11] H. R. J. Walters, J. Phys. B **4**, 437 (1971).

[12] T. N. Rescigno, C. W. McCurdy, and V. McKoy, Phys. Rev. A **11**, 825 (1975).

[13] R. G. Newton, *Scattering Theory of Waves and Particles* (Springer, New York, 1982).

[14] S. K. Adhikari and K. L. Kowalski, *Dynamical Collision Theory and Its Applications* (Academic, Boston, 1991).

[15] M. Harada, R. Tamagaki, and H. Tanaka, Prog. Theor. Phys. **33**, 1003 (1966).

[16] M. Harada, Prog. Theor. Phys. **38**, 353 (1967).

[17] M. I. Haftel and F. Tabakin, Nucl. Phys. **A158**, 1 (1970).

[18] W. Glöckle, *The Quantum Mechanical Few-Body Problem* (Springer, New York, 1983).

[19] I. H. Sloan, J. Comput. Phys. **3**, 332 (1968).

[20] G. Monegato, Computing **29**, 337 (1982).

[21] V. I. Lebedev, Zh. Vychisl. Mat. Mat. Fiz. **16**, 293 (1976).

[22] C. W. Murray, N. C. Handy, and G. J. Laming, Mol. Phys. **78**, 997 (1993).

[23] C. Winstead and V. McKoy, Adv. Chem. Phys. **96**, 103 (1996).

[24] D. K. Watson and V. McKoy, Phys. Rev. A **20**, 1474 (1979).

[25] P. Čársky and M. Polášek, J. Comput. Phys. **143**, 266 (1998).

[26] P. Čársky and T. Reschel, Collect. Czech. Chem. Commun. **63**, 1264 (1998).

[27] T. H. Dunning, Jr. and P. J. Hay, in *Modern Theoretical Chemistry*, edited by H. F. Schaefer III (Plenum, New York, 1977), Vol. 3.

[28] S. K. Srivastava, A. Chutjian, and S. Trajmar, J. Chem. Phys. **63**, 2659 (1975).

[29] T. W. Shyn and W. E. Sharp, Phys. Rev. A **24**, 1734 (1981).

[30] L. Boesten and H. Tanaka, J. Phys. B **24**, 821 (1991).

[31] L. A. Collins, W. D. Robb, and M. A. Morrison, Phys. Rev. A **21**, 488 (1980).

[32] T. L. Gibson, M. A. P. Lima, K. Takatsuka, and V. McKoy, Phys. Rev. A **30**, 3005 (1984).

[33] M. A. P. Lima, T. L. Gibson, W. M. Huo, and V. McKoy, Phys. Rev. A **32**, 2696 (1985).

P12. Discrete momentum representation method for polar molecules: Calculation of the elastic electron scattering on the H₂O molecule.

M. Ingr, M. Polášek, P. Čársky, J. Horáček, Phys. Rev. A 62 (2000) 032703.

Discrete momentum representation method for polar molecules: Calculation of the elastic electron scattering on the H₂O molecule

Marek Ingr,[1] Martin Polášek,[2] Petr Čársky,[1,*] and Jiří Horáček[3]

[1]*J. Heyrovský Institute of Physical Chemistry, Academy of Sciences of the Czech Republic, Dolejškova 3, 182 23 Prague 8, The Czech Republic*
[2]*Faculty of Philosophy and Science, Silesian University at Opava, Bezručovo nám. 13, 746 01 Opava, The Czech Republic*
[3]*Faculty of Mathematics and Physics, Charles University, V Holešovičkách 2, 180 00 Prague 8, The Czech Republic*

We present an extension of the method of discrete momentum representation to the calculation of elastic electron scattering from polar molecules. The essence of this method is the subtraction of the first Born term from the Lippmann-Schwinger equation, which is solved for a modified interaction potential from which the long-range part is removed. Forward scattering is described in the first Born approximation. The use of this method is demonstrated for electron scattering from the H₂O molecule in the static-exchange approximation. The results are in good agreement with former calculations as well as with experiments.

PACS number(s): 34.80.Bm

I. INTRODUCTION

The method of discrete momentum representation (DMR) was originally developed by Polášek *et al.* [1] for calculating the differential and integral cross sections of the nonresonant elastic electron-molecule scattering. It belongs to the group of methods based on the solution of the Lippmann-Schwinger equation in momentum space. A similar, but numerically different approach was used by McCarthy and Stelbovics [2] and Bray and Stelbovics [3] for atoms and by McCarthy and Rossi [4,5] for diatomic molecules. In Ref. [1], as well as in this paper, larger molecules are treated. The Hartree-Fock wave function was used for the molecule, and the scattered electron wave function was composed of plane waves, which are the natural basis of scattering processes. The static-exchange approximation was employed to evaluate the **T** matrix. Although this method is quite simple, it provided good results for the elastic electron scattering on hydrogen and methane molecules. However, the straightforward extension of this theory to polar targets is not possible because the dipole potential influences the incoming electron even at a long distance from the target. This long-range potential causes infinite values for the diagonal matrix elements of the Coulomb terms, i.e., those corresponding to the zero-momentum-transfer vector, making any further treatment of the interaction potential matrix impossible. Moreover, the problem with the dipole moment complicates not only the calculation of the elastic scattering of polar molecules, but also the calculation of inelastic scattering processes. Electronically excited states are often polar, even for molecules with a nonpolar ground state, and some vibrational modes give rise to polar molecular structures even when the target molecule has a zero dipole moment at its equilibrium geometry. This means that the original version of DMR cannot be applied to inelastic-scattering calculations.

Several papers describe methods for calculating elastic

scattering by dipole potentials analytically [6,7]. Unfortunately, analytical methods cannot be employed in calculations of scattering on molecular potentials; their short-range parts are too complicated. Similarly, even the two-potential formalism (see, e.g., Joachain [8]), where the dipole scattering is solved analytically and the rest is calculated in the basis of dipole-scattering eigenfunctions, seems to be too complicated for our case. To avoid these problems, we have developed a simpler formalism, one that is easier to apply computationally. It is based on the fact that differential cross sections as well as scattering amplitudes tend to their first Born approximations in the limit of forward scattering. This limit was recognized by Mittleman and von Holdt [6] and later used by other authors in the development of Born closure methods [9–15]. In these methods, the partial-wave expansion is truncated at some certain quantum number, and the contribution of higher partial waves is expressed in terms of the first Born approximation. In the older studies [9–13], this approach was applied directly to the differential cross section, while in the more recent ones [14,15] it is the scattering amplitude that is expressed in this way. Similarly, in our method this approximation is also applied to the scattering amplitude, but in a different way. Since no partial-wave expansion is used in this method, the scattering amplitude for low angles can be directly replaced by its first Born counterpart, and we can concentrate on calculating large-angle scattering only. Since large-angle scattering is controlled by the short-range part of the interaction potential, we may use a truncated potential, one in which the long-range part is cut off by a rapidly decreasing function (e.g., by an exponential function with a negative exponent). Thus the truncation of the partial-wave expansion is replaced by the truncation of the interaction potential. Altogether, it means that we can calculate the scattering amplitude with the subtracted first Born term and the truncated potential, and then we can simply add back the first Born amplitude.

The theoretical background of the proposed method is given in Sec. II, and Sec. III describes computational details. Section IV continues with results for our calculations of the differential and momentum-transfer cross sections for H₂O,

*Corresponding author. Email address: carsky@jh-inst.cas.cz

62 032703-1

188

followed by a comparison of these results with some other theoretical results and with experimental data.

II. THEORY

A. Overview of the DMR method

The DMR method is discussed in detail in Ref. [1] and only a brief summary is given here. An extension of the treatment to polar molecules is described below. As in Ref. [1], all the quantities in the following equations are in atomic units.

The basic equation of the scattering problem is the Lippmann-Schwinger (LS) equation:

$$|\Psi^{(+)}\rangle = |\mathbf{k}_0\rangle + \int d^3k \frac{|\mathbf{k}\rangle\langle\mathbf{k}|}{k_0^2 - k^2 + i\varepsilon} U |\Psi^{(+)}\rangle. \qquad (1)$$

In the DMR method, Eq. (1) is solved to obtain matrix elements of the T operator for all combinations of wave vectors of the incoming and outgoing waves with magnitude

$$k_0 = \sqrt{2E}, \qquad (2)$$

where E is the energy of the incident electron. To perform this, we transform the equation to the form

$$\langle\mathbf{k}'|T|\mathbf{k}_0\rangle = \langle\mathbf{k}'|U|\mathbf{k}_0\rangle + \int d^3k \frac{\langle\mathbf{k}'|U|\mathbf{k}\rangle\langle\mathbf{k}|T|\mathbf{k}_0\rangle}{k_0^2 - k^2 + i\varepsilon}. \qquad (3)$$

The integral on the right-hand side of Eq. (3) can be formally evaluated in the form of its principal value and a residue:

$$\int d^3k \frac{\langle\mathbf{k}'|U|\mathbf{k}\rangle\langle\mathbf{k}|T|\mathbf{k}_0\rangle}{k_0^2 - k^2 + i\varepsilon}$$

$$= P \int_0^\infty dk \int_{\Omega_k} d\Omega_k \frac{k^2\langle\mathbf{k}'|U|\mathbf{k}\rangle\langle\mathbf{k}|T|\mathbf{k}_0\rangle}{k_0^2 - k^2 + i\varepsilon}$$

$$- \frac{i\pi k_0}{2} \int_{\Omega_k} d\Omega_k \langle\mathbf{k}'|U|\mathbf{k}_0\rangle\langle\mathbf{k}_0|T|\mathbf{k}_0\rangle. \qquad (4)$$

In this equation, Ω_k is a solid angle in $\mathbf{k}$ space. From Eq. (4) we can obtain the $\mathbf{T}$-matrix element for any pair of incoming and outgoing wave vectors from the continuous momentum space, but we must know all the corresponding U elements, defined as twice the interaction potential $V (U = 2V)$ elements. The analytical solution of Eq. (4) is not accessible. Therefore we transformed it into matrix form, discretizing the integral by means of quadrature in both the radial and angular subspaces. The former is composed of the abscissas of Gauss-Legendre integration quadrature after transformation of the integration interval (k_{min}, k_{max}) onto the interval $(-1, 1)$ (k_{min} is a number close to zero and k_{max} is a cutoff value; see Ref. [1]) by the following substitution

$$x = \frac{a(k - k_0)}{b(k + k_0)}, \qquad (5)$$

where a and b are constants. The latter is a Lebedev integration quadrature [16] of a given order. Radial quadrature also contains the residue point, which corresponds to the energy of the incident electron.

This discretization provides the LS equation in the matrix form

$$\mathbf{T} = \mathbf{U} + \mathbf{U}\mathbf{G}_0^+\mathbf{T}, \qquad (6)$$

where $\mathbf{G}_0^+$ is the matrix of the Green's function in the discrete-momentum representation. Because of a typographical error in Ref. [1], we repeat the definition of its elements in this paper:

$$(\mathbf{G}_0^+)_{qj,pi} = -\delta_{qp}\delta_{ji}i\pi w_i k_0/2 \quad \text{for} \quad p = 0, \qquad (7a)$$

$$(\mathbf{G}_0^+)_{qj,pi} = \delta_{qp}\delta_{ji}\frac{2abw_i k_p^2 k_0}{(a - bx_p)^2(k_0^2 - k_p^2)} \quad \text{for} \quad p \neq 0. \qquad (7b)$$

In these equations, $p = 0$ when $k_p = k_0$, and x_p is defined by Eq. (5). Equation (6) can be easily solved for the $\mathbf{T}$ matrix

$$\mathbf{T} = (\mathbf{1} - \mathbf{U}\mathbf{G}_0^+)^{-1}\mathbf{U}. \qquad (8)$$

Once the $\mathbf{T}$ matrix is known, we can calculate the scattering amplitude and differential cross section for those pairs of $\mathbf{k}$ vectors that are present in the angular quadrature,

$$A_{fi} = -\frac{1}{2\pi}\langle\mathbf{k}_f|T|\mathbf{k}_i\rangle, \qquad (9)$$

$$\sigma_{fi} = \frac{|\mathbf{k}_f|}{|\mathbf{k}_i|}|A_{fi}|^2. \qquad (10)$$

The determination of the scattering amplitudes for different pairs of vectors as well as the averaging over the molecular orientation is described in detail in Sec. II D.

Once the differential cross section is calculated, we can calculate the integral and momentum-transfer cross sections:

$$\sigma_{inf} = \int_\Omega \sigma(\Omega)d\Omega, \qquad (11)$$

$$\sigma_{mt} = \int_\Omega \sigma(\Omega)[1 - \cos(\theta)]d\Omega, \qquad (12)$$

where θ is the scattering angle and Ω represents the integration angles.

B. Construction of the optical potential

The optical potential represents the interaction between the incident electron and the target molecule. Its construction allows us to describe the electron-molecule collision as a one-electron process. Its form for the DMR method [1] in the static-exchange approximation consists of the static part (in

Ref. [1], the factor $4\pi/K^2$ is also erroneously applied to the Coulombic term, but the computer code is free of this error)

$$\langle \mathbf{k}_1 | V_S | \mathbf{k}_2 \rangle = -\frac{4\pi}{K^2} \sum_A Z_A \exp(i\mathbf{K}\cdot\mathbf{R}_A)$$

$$+ \sum_{\alpha,\beta} P_{\alpha\beta} \left\langle \mathbf{k}_1 \alpha \left| \frac{1}{r_{12}} \right| \mathbf{k}_2 \beta \right\rangle, \qquad (13)$$

and the exchange part

$$\langle \mathbf{k}_1 | V_X | \mathbf{k}_2 \rangle = -\frac{1}{2} \sum_{\alpha,\beta} P_{\alpha\beta} \left\langle \mathbf{k}_1 \alpha \left| \frac{1}{r_{12}} \right| \beta \mathbf{k}_2 \right\rangle. \qquad (14)$$

Here α and β denote the basis-set functions, $P_{\alpha\beta}$ are the density-matrix elements, $\mathbf{K}=\mathbf{k}_2-\mathbf{k}_1$ is the momentum-transfer vector, Z_A is the nuclear charge of the nucleus A, and $\mathbf{R}_A$ is its position vector. Methods for calculating the Coulomb and exchange integrals that occur in Eqs. (13) and (14) are described in detail in [17].

C. Modification of DMR for polar targets

The problem of the dipole singularity originates from the static potential. As we showed in Ref. [1], by integration over $\mathbf{r}_1$, the matrix element can be expressed as

$$\langle \mathbf{k}_1 | V_S | \mathbf{k}_2 \rangle = \frac{4\pi}{K^2} \int e^{i\mathbf{K}\cdot\mathbf{r}} \rho(\mathbf{r}) d\mathbf{r}, \qquad (15)$$

where the total charge density (nuclear plus electronic) is defined as

$$\rho(\mathbf{r}) = -\sum_A Z_A \delta(\mathbf{r}-\mathbf{R}_A) + 2\sum_i |\varphi_i(\mathbf{r})|^2. \qquad (16)$$

The symbol φ_i stands for all occupied spatial orbitals. Using the Taylor expansion for the exponential in Eq. (15) we obtain

$$\langle \mathbf{k}_1 | V_S | \mathbf{k}_2 \rangle = \frac{4\pi}{K^2} \int \left[1 + i\mathbf{K}\cdot\mathbf{r} - \frac{(\mathbf{K}\cdot\mathbf{r})^2}{2} \right.$$

$$\left. -i\frac{(\mathbf{K}\cdot\mathbf{r})^3}{6} + \cdots \right] \rho(\mathbf{r}) d^3\mathbf{r}. \qquad (17)$$

The first term on the right-hand side of Eq. (17) represents integration over the total charge density, and it therefore vanishes for neutral molecules. The second term corresponds to the electron-permanent dipole interaction. It is obvious that this term tends to infinity in the forward direction, i.e., when $\mathbf{K}\rightarrow\mathbf{0}$. All other terms of the static part of the U element in Eq. (17) are finite. Therefore the dipole term is a leading term in the forward direction. As we mentioned above, this singularity is caused by the long-range part of the interaction potential and does not influence significantly scattering to larger angles. Moreover, the scattering amplitude tends to its first Born approximation in the forward direction (see, e.g.,

Ref. [6]). For these reasons, we introduce an approximate theory involving a modified potential from which the long-range part is removed.

Consequently, we propose a method for the solution of the LS equation in which the first Born term is subtracted. It is derived from Eq. (6) in the following simple way:

$$\mathbf{T} = \mathbf{U} + \mathbf{U}\mathbf{G}_0^+ \mathbf{T},$$

$$\mathbf{T}-\mathbf{U} = \mathbf{U}\mathbf{G}_0^+ \mathbf{T} = \mathbf{U}\mathbf{G}_0^+(\mathbf{T}-\mathbf{U}+\mathbf{U}) = \mathbf{U}\mathbf{G}_0^+ \mathbf{U} + \mathbf{U}\mathbf{G}_0^+(\mathbf{T}-\mathbf{U}),$$

$$\mathbf{T}-\mathbf{U} = (1-\mathbf{U}\mathbf{G}_0^+)^{-1}\mathbf{U}\mathbf{G}_0^+ \mathbf{U}. \qquad (18)$$

In general, we can derive a whole class of such equations, subtracting more than one Born term:

$$\mathbf{T} - \mathbf{B}_1 - \cdots - \mathbf{B}_n = (1-\mathbf{U}\mathbf{G}_0^+)^{-1}\mathbf{B}_{n+1}, \qquad (19)$$

where $\mathbf{B}_n$ denotes the nth Born term. For our practical purpose, however, Eq. (18) is the most suitable. The potential of a polar molecule consists of the dipole part U_D and the short-range part. For the purpose of the solution of Eq. (18), we construct a modified dipole potential U_m of the form

$$U_m(\mathbf{r}) = U_D(\mathbf{r})e^{-cr} = 2\frac{\mathbf{D}\cdot\mathbf{r}}{r^3}e^{-cr}, \qquad (20)$$

and use it instead of the dipole potential U_D ($\mathbf{D}$ is the dipole moment). The factor 2 arises from the difference between U and V potentials ($U=2V$). Whereas the U_D element of the unmodified dipole potential is

$$\langle \mathbf{k}_1 | U_D | \mathbf{k}_2 \rangle = \frac{8\pi i \mathbf{K}\cdot\mathbf{D}}{K^2}, \qquad (21)$$

the element of U_m is

$$\langle \mathbf{k}_1 | U_m | \mathbf{k}_2 \rangle = \frac{8\pi i \mathbf{K}\cdot\mathbf{D}}{K^2} \left[1 - \frac{c}{K}\arctan\left(\frac{K}{c}\right) \right]. \qquad (22)$$

Obviously, this element is analytical over the whole momentum space, including the origin of the coordinate system, for which its limit as $\mathbf{K}\rightarrow\mathbf{0}$ is vanishing. Having obtained U_m, we substitute it for the dipole potential U_D, which may be identified with the second term on the right-hand side of Eq. (17). The U potential so prepared is then used for the "subtracted" LS equation (18). Its solution may be viewed as integration of singular functions ignoring the singularity [18]. Although this method must be used with caution in some cases, it can provide reasonable results. Then, for the elements with nonzero $\mathbf{K}$, we evaluate the first Born term as it is done for nonpolar targets [1], and we add them to the solution of Eq. (18) to obtain the complete $\mathbf{T}$ matrix.

This approach is similar to the method of Rescigno *et al.* [14] and Gianturco and Scialla [19]. They also divided the scattering amplitude into the first Born term and the rest. The latter part is expressed in the form of a truncated partial-wave expansion in which the short-range potential plays the dominant role. This part is especially important in the larger-angle directions. The long-range part, on the other hand, is

described by the first Born term and is dominant in the low-angle directions. We do not truncate the partial-wave series, but the potential itself.

Next, we comment on the form of the U_m potential. The smaller the constant c, the more accurate the result we can obtain because the modified potential U_m approximates better the dipole potential. However, our grid consists of a finite number of $\mathbf{k}$ vectors. Hence for a sufficiently small value of the constant c, the U_m elements are practically indistinguishable from the U_D elements for all pairs of $\mathbf{k}$ vectors present in our grid that have a nonzero difference. Differences between U_m and U_D elements occur only for $\mathbf{K}$ vectors too small to be resolved by our grid vectors. Therefore we use U_D elements instead of U_m elements for all pairs of grid vectors with $\mathbf{K}$ different from zero, and for $\mathbf{K}=0$ we set the element to zero. Potential elements evaluated in this way correspond to elements of the U_m potential with an extremely small c. (It should be mentioned that the function e^{-cr} is not the only possibility for the cutoff function. Many functions that decrease rapidly at large distances and are equal to 1 at the origin might be used. The present choice was done mainly for mathematical convenience.)

When the conclusions of the preceding paragraph are applied, the computational procedure can be simplified considerably. The U_S elements are evaluated by means of Eq. (17) for all pairs of $\mathbf{k}$ vectors with $\mathbf{K} \neq 0$, just as they are done for nonpolar targets. Thus the second (dipole) term in Eq. (17) is kept without change. For $\mathbf{K}=0$, however, only the third (quadrupole) term in Eq. (17) is kept, since the dipole term is omitted (as explained in the preceding paragraph) and the other terms are equal to zero. Thus the singularity in the forward scattering is eliminated by this construction of the potential elements. Equation (18) can be rearranged back to the form of Eq. (8), as it is no longer necessary to subtract the singular part of the potential elements. Physically meaningful scattering amplitudes are obtained for nonzero scattering angles. The results for the water molecule were obtained in this way.

D. Averaging over the molecular orientation

A basic method of molecular-orientation averaging was described in [1]. In this technique, many pairs of $\mathbf{k}$ vectors are constructed for every scattering angle, distributing their endpoints on a circle of a given radius around each quadrature point. Although this method provides reliable results, its disadvantage is that it requires a large computational effort and therefore enormous demands on computer time. To cope with this problem we suggest an alternative method based on the same principle but using the calculated matrix elements more efficiently.

In this alternative method, a set of randomly oriented vectors of magnitude $|\mathbf{k}_0|$ is generated using a random-number generator. New T elements are calculated for every pair of the vectors using the following equations:

$$T_{i,n} = \sum_j \left[(\mathbf{1} - \mathbf{U}G_0^+)^{-1} \right]_{i,j} U_{j,n}, \qquad (23)$$

$$T_{m,n} = U_{m,n} + \sum_j U_{mj} (G_0^+)_j T_{j,n}. \qquad (24)$$

In these equations, the indices i,j correspond to the original quadrature vectors, while m,n denote the newly generated random vectors. Hence we need to evaluate the rectangle matrix $\mathbf{U}_{j,n}$ (and its complex conjugate $\mathbf{U}_{m,j}$) and the square matrix $\mathbf{U}_{m,n}$. After the evaluation of the additional T elements, all the pairs of the new vectors are sorted according to the angle between them into the intervals of a given length, e.g., 1°. Scattering amplitudes and corresponding differential cross sections are determined for these pairs and averaged within every interval. The averaged value is then considered as the value of the differential cross section at the middle point of the given interval.

Thus the values of the differential cross sections are obtained for many points over the interval [0°, 180°]. The more random vectors we generate, the smoother is the curve formed by the calculated points. Its smoothness can be further improved by expanding the individual averaging intervals, since more vectors are averaged in each interval. However, when the differential cross section changes rapidly over a short range of angles, structure can be lost in the averaging procedure if the intervals are too long. Although it is possible in principle to use a large number of random vectors and so produce very smooth curves, this technique would be very costly. Therefore it is more efficient to use only some of the vectors, e.g., one in ten, and to construct the curve by fitting with cubic splines.

Another problem with this method arises from the fact that the number of pairs of random vectors in every interval is proportional to the sine of the angle corresponding to the interval (e.g., its middle point). Due to this proportionality, the highest number of pairs occurs in the intervals around 90°, while in the intervals around 0° and 180° the number of pairs is markedly smaller. Therefore in these sparse areas the cross section must be extrapolated from the adjoining intervals. Nevertheless, when calculating polar molecules there are no problems with the region close to 0°, because here the cross section diverges, and we do not calculate values in this region at all. A different and more efficient averaging method based on the decomposition of the T element in the basis of spherical harmonic functions is presently being developed in our group.

III. COMPUTATIONAL DETAILS

All calculations for the water molecule were performed in its optimized geometry at the Hartree-Fock level using Sadlej's TZV basis set [20]. The value of the dipole moment for the SCF-optimized geometry in this basis set is 0.8060 a.u., which is slightly overestimated in comparison with the experimental value of 0.724 a.u. [21]. The basis set and the obtained molecular-orbital density matrix were transferred into the DMR program to compute Coulomb and exchange integrals by means of Eqs. (13) and (14). The numerical quadrature used for the integration of the Lippmann-Schwinger equation consisted of a radial part formed by a Gauss-Legendre quadrature transformed to the integration
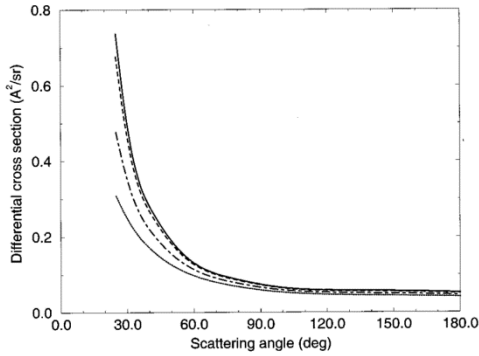
FIG. 1. Differential cross sections for the pure dipole potential (1 a.u.) for different values of the constant c. Dotted line, $c = 0.1$; dot-dashed line, $c = 0.05$; dashed line, $c = 0.01$; solid line, $c = 0$.

interval (see Sec. II A) and an angular part consisting of the points of Lebedev quadrature [16]. Radial quadratures with 19, 21, and 23 points and angular quadratures with 50, 86, and 110 points were used in individual calculations to demonstrate the convergence in both dimensions. In the averaging procedure, 1000 random vectors were generated and the averaging-interval length was set to 1°.

IV. RESULTS AND DISCUSSION

As the results of our method depend on the constant c [see Eqs. (20) and (22)], we first tested the convergence of the cross section with respect to this constant. For this purpose, the elastic electron scattering on the dipole potential (dipole moment of 1 a.u.) was calculated. The differential cross sections for several values of c are presented in Fig. 1. The

values apparently converge to the limiting value $c = 0$, as was explained in Sec. II C. Hence in practice it is not necessary to restrict the long-range potential by the exponential function since it is done automatically by the discrete integration grid in the **k** space.

Making use of this fact, we developed a remarkably simple method of using the modified dipole potential. For each pair of **k** vectors with $\mathbf{K} \neq \mathbf{0}$ the value of the unmodified interaction potential is used, and for $\mathbf{K} = \mathbf{0}$ its dipole part in Eq. (17) is set to zero. We used this method to calculate differential and momentum-transfer cross sections of elastic electron scattering from the H_2O molecule in the static-

FIG. 3. Differential cross sections of the elastic electron scattering on H_2O for 20 eV (convergence with respect to the angular quadrature, the number of radial points in all calculations is 23). Long-dashed line, 50 angular points; dashed line, 86 angular points; solid line, 110 angular points.

FIG. 2. Differential cross sections of the elastic electron scattering on H_2O for 6 eV (convergence with respect to the angular quadrature, the number of radial points in all calculations is 23). Long-dashed line, 50 angular points; dashed line, 86 angular points; solid line, 110 angular points.

FIG. 4. Differential cross sections of the elastic electron scattering on H_2O for 6 eV. Solid line, present results; long-dashed line, calculations by Brescansin *et al.* [22]; dashed line, calculations by Gianturco and Scialla [19]; +, experiment by Shyn and Cho [23]; ×, experiment by Danjo and Nishimura [24].

032703-5

192
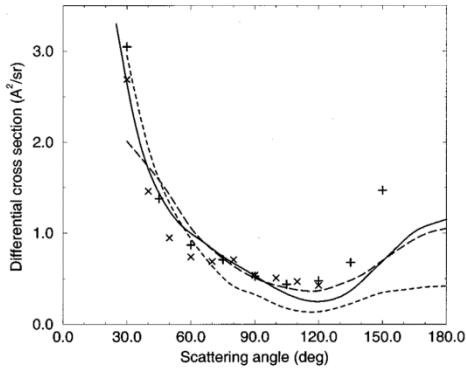
FIG. 5. Differential cross sections of the elastic electron scattering on H_2O for 10 eV. For the description of the individual curves see Fig. 4.

FIG. 7. Differential cross sections of the elastic electron scattering on H_2O for 20 eV. For the description of the individual curves see Fig. 4.

exchange approximation for incident electron energies from 2 to 20 eV. The convergence studies were performed for all combinations of three radial (19, 21, 23 points) and three angular Lebedev quadratures (50, 86, 110 points). The calculation is remarkably stable with respect to the change in the radial quadrature. The corresponding curves are, in fact, identical, and therefore not shown. The angular convergence is demonstrated in Figs. 2 and 3 for 6 and 20 eV, respectively. The angular quadrature with 50 points is apparently insufficient, but the two higher ones provide good convergence. Identical convergence tests were also performed for energies of 10 and 15 eV with analogous results (not shown). Differential cross sections for 6, 10, 15, and 20 eV are shown in Figs. 4–7 together with the results of Brescansin *et al.* [22] and Gianturco and Scialla [19], also performed in the static-exchange approximation, and with two sets of experimental values by Shyn and Cho [23] and Danjo and Nish-

imura [24]. These results are also tabulated in Table I. The agreement with previous theoretical results seems to be satisfactory. Our results are closer to those of Brescansin *et al.* [22], which may be explained by the rather different approach that was used in the calculations of Gianturco and Scialla [19]. The experimental data are also reproduced relatively well, although only the static-exchange approximation is used. Calculations for energies lower than 2 eV were not attempted since the use of the static-exchange approximation is not justifiable in this region.

Momentum-transfer cross sections were calculated using Eq. (12). Only the values of differential cross sections from

FIG. 6. Differential cross sections of the elastic electron scattering on H_2O for 15 eV. For the description of the individual curves see Fig. 4.

TABLE I. Differential cross sections for elastic electron scattering on H_2O for electron energies of 6, 10, 15, and 20 eV.

Scattering angle (deg)	DCS ($Å^2$/sr)			
	6 eV	10 eV	15 eV	20 eV
25	4.05	3.30	3.32	3.16
30	3.14	2.64	2.73	2.61
40	1.91	1.70	1.75	1.65
50	1.40	1.24	1.16	1.04
60	1.11	1.00	0.85	0.72
70	0.91	0.83	0.65	0.52
80	0.72	0.67	0.53	0.41
90	0.56	0.53	0.43	0.34
100	0.42	0.40	0.35	0.29
110	0.30	0.29	0.30	0.26
120	0.23	0.25	0.31	0.28
130	0.22	0.30	0.40	0.37
140	0.27	0.47	0.60	0.55
150	0.35	0.69	0.86	0.78
160	0.44	0.92	1.13	1.02
170	0.51	1.07	1.31	1.18
180	0.54	1.15	1.41	1.28
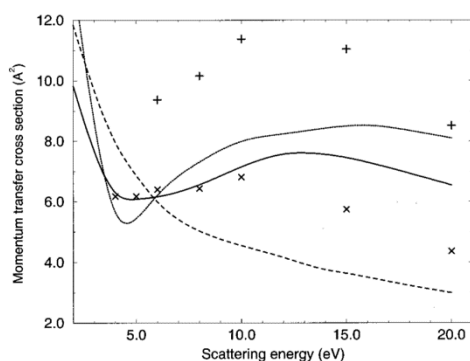
FIG. 8. Integral cross sections for the elastic electron scattering on H_2O. Solid line, present results; dotted line, calculation by Brescansin *et al.* [22]; dashed line, calculation by Gianturco and Scialla [19]; +, experiment by Shyn and Cho [23]; ×, experiment by Danjo and Nishimura [24].

25° to 180° are taken into account for two reasons. The data for lower angles are not reliable, and they contribute negligibly to the integral. The result is plotted in Fig. 8 together with other theoretical [19,22] and experimental [23,24] results, and it is tabulated in Table II. Our curve agrees better with the one from Ref. [22] than with that from Ref. [19] for the reasons discussed above. Both experimental results [23,24] show the maximum in the energy range of 10–15 eV. A more precise statement cannot be given because only a few experimental points are available in both studies (see Fig. 8). However, this feature is qualitatively reproduced by our calculation; the absolute values lie between the two experiments. We can conclude that both the calculations of differential and integral cross sections provide reliable results in the investigated energy region.

TABLE II. Integral cross sections for elastic electron scattering on H_2O.

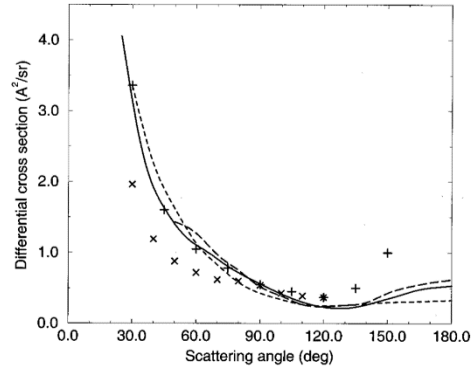| Energy (eV)	ICS ($Å^2$)	Energy (eV)	ICS ($Å^2$)
2	9.82	12	7.57
3	7.68	13	7.61
4	6.30	14	7.56
5	6.08	15	7.45
6	6.16	16	7.30
7	6.32	17	7.12
8	6.56	18	6.93
9	6.85	19	6.74
10	7.15	20	6.55
11	7.40		

V. CONCLUSIONS

We have extended the original DMR method [1] to the calculation of elastic electron-molecule-scattering cross sections for polar targets. Although the method is remarkably simple, it provides good results for intermediate scattering energies, as was demonstrated by the example of the H_2O molecule. Moreover, this method is not excessively time consuming, so it may be applied to scattering calculations for larger molecules. An alternative algorithm for averaging the result over the molecular orientation was introduced, which is based on the same principle as the former one [1] but is more efficient. With the planned addition of polarization terms to the optical potential, the DMR method seems to have the potential to be very promising as a powerful tool in the field of scattering calculations.

[1] M. Polášek, M. Jůrek, M. Ingr, P. Čársky, and J. Horáček, Phys. Rev. A **61**, 032701 (2000).
[2] I. E. McCarthy and A. T. Stelbovics, Phys. Rev. A **28**, 2693 (1983).
[3] I. Bray and A. T. Stelbovics, Phys. Rev. A **46**, 6995 (1992).
[4] I. E. McCarthy and A. M. Rossi, Phys. Rev. A **49**, 4645 (1994).
[5] A. M. Rossi and I. E. McCarthy, J. Phys. B **28**, 3593 (1995).
[6] M. H. Mittleman and R. E. von Holdt, Phys. Rev. **140**, A726 (1965).
[7] S. Altshuler, Phys. Rev. **107**, 114 (1957).
[8] C. J. Joachain, *Quantum Collision Theory* (North-Holland, Amsterdam, 1983).
[9] L. A. Collins and D. W. Norcross, Phys. Rev. A **18**, 467 (1978).
[10] D. W. Norcross and N. T. Padial, Phys. Rev. A **25**, 226 (1982).
[11] N. T. Padial, Phys. Rev. A **32**, 1379 (1985).
[12] N. T. Padial and D. W. Norcross, Phys. Rev. A **29**, 1742 (1984).
[13] A. Jain and D. W. Norcross, Phys. Rev. A **32**, 134 (1985).
[14] T. N. Rescigno, B. H. Lengsfield, C. W. McCurdy, and S. D. Parker, Phys. Rev. A **45**, 7800 (1992).
[15] W. A. Isaacs and M. A. Morrison, Phys. Rev. A **53**, 4215 (1996).
[16] V. I. Lebedev, Zh. Vychisl. Mat. Mat. Fiz. **16**, 293 (1976).
[17] P. Čársky and M. Polášek, J. Comput. Phys. **143**, 266 (1998).
[18] P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration* (Academic, New York, 1975).
[19] F. A. Gianturco and S. Scialla, J. Chem. Phys. **87**, 6468 (1987).
[20] A. J. Sadlej, Collect. Czech. Chem. Commun. **53**, 1988 (1995).
[21] J. B. Hasted, in *Water: A Comprehensive Treatise*, edited by F. Franks (Plenum, New York, 1972), Vol. 1, p. 255.
[22] L. M. Brescansin, M. A. P. Lima, T. L. Gibson, V. McKoy, and W. M. Huo, J. Chem. Phys. **85**, 1854 (1986).
[23] T. W. Shyn and S. Y. Cho, Phys. Rev. A **36**, 5138 (1987).
[24] A. Danjo and H. Nishimura, J. Phys. Soc. Jpn. **54**, 1224 (1985).

P13. Potential energy curve of the X $^2\Sigma_u^+$ resonance state of F2 computed by CAP/CI. M. Ingr, H.-D. Meyer, L.S. Cederbaum.

J. Phys. B: At. Mol. Opt. Phys. 32 (1999) L547–L556.

LETTER TO THE EDITOR

Potential energy curve of the X $^2\Sigma_u^+$ resonance state of F$_2^-$ computed by CAP/CI

M Ingr†‡, H-D Meyer† and L S Cederbaum†

† Theoretische Chemie, Physikalisch-Chemisches Institut, INF 229, D-69120 Heidelberg, Germany
‡ J Heyrovský Institute of Physical Chemistry, Academy of Sciences of the Czech Republic, Dolejškova 3, 182 23 Prague 8, The Czech Republic

Abstract. Calculations of resonance energies and widths of the X $^2\Sigma_u^+$ state of F$_2^-$ were carried out. Configuration interaction was used to solve the Schrödinger equation for a molecular electronic Hamiltonian augmented by a complex absorbing potential (CAP/CI). Several technical aspects of this method are briefly discussed. The crossing point of the anionic curve of this state with the ground state of F$_2$ is determined as the point where the resonance width vanishes. Based on this, the electron affinity of atomic fluorine is reproduced with satisfactory precision. Finally, our results are compared to those of several preceding studies and an outlook to further continuation of this project is given.

1. Introduction

The use of complex absorbing potentials (CAPs) for calculating resonance states was introduced by Jolicard and Austin [1, 2], who applied it to the Hazi–Taylor model. Riss and Meyer [3] provided a detailed mathematical investigation on the performance of the CAP method and applied it to electron–molecule resonances. In the following the method of complex absorbing potentials is combined with configuration interaction (CI). The method, denoted CAP/CI, has been used previously to calculate several resonance states of different molecular anions, such as C$_2^{2-}$ [4] and N$_2^-$ [5], and for a description of auto-ionizing (Auger) decay of (HF)$_2^+$ clusters [6].

In the present work we use the CAP/CI method to calculate the resonance energy and width for a broad range of internuclear distances of the F$_2^-$ anion. There are several motivations for this work. First, we want to demonstrate the capability of the CAP/CI method to provide a reliable dependence of the resonance energy and width on the internuclear coordinate. Secondly, all the preceding calculations on the F$_2^-$ $^2\Sigma_u^+$ state are relatively old, and the methods applied could probably be insufficient for describing the correlation effects on this system [7–9]. Moreover, these results differ markedly from each other and provide rather unreliable predictions of the crossing point of the potential energy curve of this resonance state with the curve of the ground state of a neutral F$_2$ molecule, which is the point where the resonance becomes a bound state. Our aim is to provide more accurate results. Thirdly, in order to perform dynamical studies on the system one needs the resonance parameters as input for the calculations. Finally, we intend to carry out a more extensive study on the F$_2^-$ system including some of its excited states. The present results serve as a test for these calculations.

2. Computational details

Using the CAP method one solves the Schrödinger equation for a modified Hamiltonian $H(\eta)$ which is the sum of the system Hamiltonian H and the CAP,

$$H(\eta) = H - i\eta W. \tag{1}$$

Here η is a positive parameter and W a (in general positive) potential function which vanishes in some region close to the molecule but increases in the outer region. It was shown in [3] that each eigenvalue that stabilizes in the complex plane with respect to variations of the η parameter corresponds to a complex energy E_{res} of a Siegert state. These energies are characterized by their resonance position E and width Γ,

$$E_{res} = E - i\Gamma/2. \tag{2}$$

(Note that the resonance lifetime τ is related to the width through $\tau = \hbar/\Gamma$.) Stabilization of the complex eigenvalue with respect to η is obtained if there is a pronounced minimum of $\eta|\partial E_{res}/\partial\eta|$. In the present study the same kind of CAP is used as that introduced in the work of Santra *et al* [6], who investigated the resonant decay of molecular cluster cations. The CAP is constructed as a sum of one-electron potentials, which are further decomposed into three one-dimensional terms depending on only one Cartesian coordinate of a given electron:

$$W = \sum_{i=1}^{n} [W(x_i) + W(y_i) + W(z_i)] \tag{3}$$

where n is the number of electrons. The one-dimensional potential has the following form:

$$W(x_i) = (x_i + c_x)^2 \theta(x_i + c_x) + (x_i - c_x)^2 \theta(x_i - c_x) \tag{4}$$

and analogously for y_i and z_i. Here θ denotes Heaviside's step function. The one-electron potential thus vanishes inside a box defined by constants c_x, c_y, c_z and outside this box it grows quadratically. Since F_2^- is a cylindrically symmetrical system, we have defined the CAP box by $c_x = c_y = c$ and $c_z = R/2 + c$, where R is the internuclear distance. The CAP is thus specified by two parameters, η and c. We consider an eigenvalue to correspond to a genuine resonance state if it is stable with respect to variations of both η and c.

The wavefunction of our system is constructed as a multi-reference configuration-interaction (MRCI) wavefunction. The spin-orbitals for the construction of the configuration state functions (CSF) are taken from a CASSCF calculation on the neutral fluorine molecule. The configuration space is equivalent to that used by Blomberg and Siegbahn in a calculation denoted as MC 178 [11]. A similar technique was used by Sommerfeld *et al* for N_2^- [5]. The reference configurations are chosen in the following manner: in every reference configuration the nine lowest orbitals are doubly occupied and the additional 19th electron occupies one of the other orbitals of σ_u symmetry. The 1s σ orbitals are frozen, i.e. none of the four core electrons can be excited. Only singly and doubly excited CSFs with respect to the reference configurations are used in the expansion and, additionally, we require that at least 14 active electrons occupy a specified set of lowest orbitals. This set is called the target set, whereas the remaining orbitals form the so-called continuum set. Hence at most one electron can be excited to an orbital of the continuum set. This selection procedure makes our configuration space conveniently small (see [5] for more details).

The atomic orbital basis set used in our calculation consists of two parts. One part, which describes the inner region of the wavefunction, is the standard Dunning basis set (10s6p/5s4p) for a fluorine atom [10] augmented by two uncontracted d functions and one

Table 1. List of the diffuse parts of employed basis sets. The inner part is the same for all the calculations (Dunning basis, not listed, see text). Basis sets denoted 'B' are larger by two s and one p functions than those denoted 'A'.

Basis set	Types and exponents of basis functions				Region of use
A1	s:	0.264 16	p:	0.091 00	2.28–2.30 au
		0.108 04		0.063 00	
		0.058 37		0.038 00	
		0.038 00			
		0.030 00			
		0.023 00			
		0.018 00			
		0.014 00			
A2	s:	0.264 16	p:	0.089 89	2.30–2.32 au
		0.108 04		0.060 00	
		0.058 37		0.035 00	
		0.036 480			
		0.027 000			
		0.020 000			
		0.014 000			
		0.011 000			
A3	s:	0.264 16	p:	0.089 89	2.32–2.38 au
		0.108 04		0.055 61	
		0.058 37		0.030 00	
		0.036 480			
		0.024 938			
		0.018 119			
		0.012 000			
		0.008 000			
A4	s:	0.237 75	p:	0.080 90	2.38–2.40 au
		0.097 236		0.050 050	
		0.052 535		0.027 000	
		0.032 832			
		0.022 444			
		0.016 307			
		0.009 000			
		0.005 400			

uncontracted f function, with exponents equivalent to those of basis B in [11] (i.e. 3.14 and 0.81 au for the d-function and 1.8 au for the f-function). The other part, describing the continuum component of the wavefunction, is composed of several uncontracted diffuse functions of s- and p-type. All functions are centred on the fluorine atoms. The diffuse functions must be adapted to the particular wavefunction, which is why they are different for different internuclear distances. Table 1 contains a list of the diffuse parts of the basis sets applied in our calculations.

We have also performed calculations including one additional diffuse d-function on each centre. The results were very insensitive to the inclusion of the additional function, and all the calculations reported below are without it.

As was demonstrated in [3], the unwanted influence of the CAP on the wavefunction can be partially removed using the first-order correction. In fact, this allows us to use a narrower CAP box, which is especially important in cases where the resonance width and the energy of the outgoing electron is small. These cases require larger CAP boxes, and it is more difficult

Table 1. Continued.

Basis set	Types and exponents of basis functions				Region of use
B1	s:	0.275 00	p:	0.080 90	2.40–2.45 au
		0.237 75		0.050 050	
		0.097 236		0.027 000	
		0.052 535		0.015 000	
		0.032 832			
		0.022 444			
		0.016 307			
		0.009 000			
		0.005 400			
		0.003 000			
B2	s:	0.275 00	p:	0.080 90	2.45–2.57 au
		0.237 75		0.048 500	
		0.097 236		0.024 000	
		0.052 535		0.012 000	
		0.032 832			
		0.021 700			
		0.015 100			
		0.008 250			
		0.005 000			
		0.002 600			

to reach a stabilization of the complex energy without correction, often it is not possible at all. We used the correction for internuclear distances larger than 2.45 au.

One must be aware of numerical artefacts which may occur in CAP/CI calculations. They may cause the η-trajectories (i.e. the trajectories, along which the complex eigenvalues move in the complex plane when η is increased) to exhibit spurious stabilizations. Hence, one must carefully check each stabilization point to ensure its correctness. Here we describe two typical numerical artefacts.

(a) The spectrum of the CAP-augmented Hamiltonian consists of the resonance eigenvalues and of states describing a discretized continuum. In an exact calculation (i.e. using a complete basis set) the continuum states do not stabilize, but move monotonically deeper into the complex plane when η is increased. However, in a finite basis calculation the situation is different. The eigenvalues of the discretized continuum follow the supposed trend only for those values of the η-parameter for which the basis set is sufficient for the description of the corresponding wavefunction. The basis set truncation error may compel the η-trajectories to change their direction quite rapidly, falsely indicating a stabilization. Fortunately, these artificial stabilizations are unstable with respect to the CAP box size and can be rather easily distinguished from genuine resonances.

(b) Another problem can appear when a state from the discretized continuum spectrum lies in the complex plane close to the resonance state. This again may happen if the single-particle basis set is inappropriate. Provided that both states have the same symmetry, they may influence one another in a manner similar to the well known scenario of an avoided crossing of two eigenstates of a real Hamiltonian. Due to this phenomenon the resonance eigenvalue is shifted from its proper location, with deterioration mainly in the more sensitive imaginary part. Recognition of this artefact is more troublesome than in the former case. We observed the following behaviour of eigenvalues influenced by this artefact. The η-trajectory follows a rather complicated path in the complex plane and

shows more than one point of stabilization. The quality of stabilization is generally worse than that of resonances which are not influenced by a nearby continuum state. This effect is basis set dependent and, therefore, it can be identified by performing the calculation in another basis set.

3. Results and discussion

Figure 1(*a*) shows a sketch of the potential curves of the F_2^- $X\,^2\Sigma_u^+$ state (full curve) and the neutral $^1\Sigma_g^+$ ground state (dotted curve). F_2^- is electronically stable for internuclear distances larger than 2.6 au. At smaller distances, however, the energy of the ionic state is above the energy of the neutral species and an electronic decay becomes possible. There, the ionic state is thus turned into a resonance.
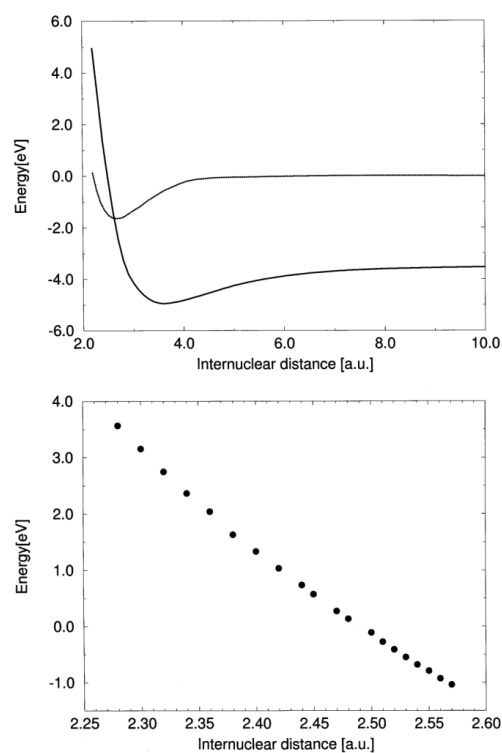
Figure 1. (*a*) Potential energy curves of the $X\,^2\Sigma_u^+$ state of F_2^- (full curve) computed in this work and the $X\,^1\Sigma_g^+$ state of F_2 (dotted curve) taken from [11]. (*b*) Individually computed resonance energies (real parts) of the $X\,^2\Sigma_u^+$ state of F_2^-. The zero point of the energy scale is the energy of two separated (neutral) F atoms. The absolute energies are listed in table 3.

Using the CAP/CI method we have calculated resonance energies and widths for several internuclear distances between 2.28 and 2.57 au. For the smaller of these internuclear distances the energy of the ionic state is about 5 eV above the crossing point of the F_2 and F_2^- ground states. This is thought to be sufficient for the description of the nuclear dynamics of the system and, therefore, we did not continue our calculations below this limit. The upper limit originates from the computational difficulties occurring above this value of internuclear distance. There, both the kinetic energy of the outgoing electron and the width of the resonance become very small and a pronounced stabilization is not obtained within the present basis sets. However, the crossing point itself may be determined by extrapolation (see figure 2(b)). and the discussion below). For internuclear distances larger than the crossing point inclusion of the CAP is, of course, not necessary.
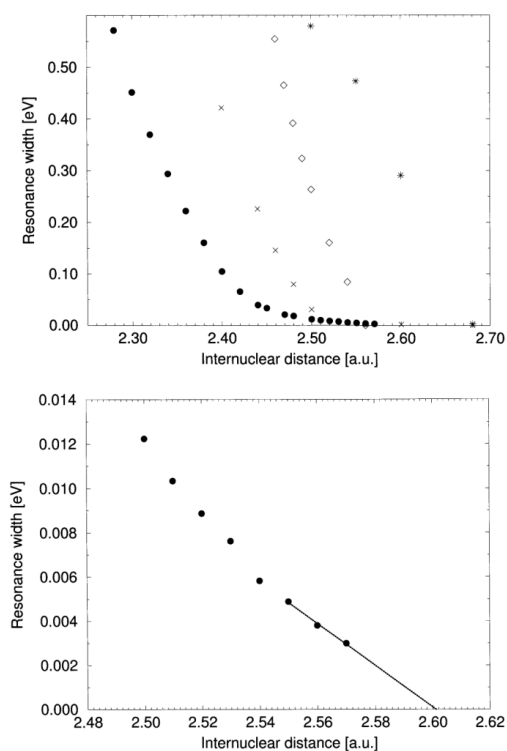
Figure 2. (a) Resonance width of the X $^2\Sigma_u^+$ state of F_2^-. Full circles, present results; crosses, Lauderdale *et al* [8]; stars, Hazi *et al* [7]; diamonds, SEP results by Morgan and Noble [9]. (b) Resonance width Γ of the X $^2\Sigma_u^+$ state of F_2^-. The vicinity of the crossing point is detailed. Full circles, calculated values; dotted line, extrapolation to $\Gamma = 0$ via a linear fit to the last three points.

As was already indicated above, it is in general impossible to perform the calculations over a wide range of internuclear distances using a single one-particle basis set (unless this basis set is very large). The exact resonance wavefunction diverges with increasing distance and it does so more quickly for broader resonances. Since we are removing the divergent part of the wavefunction by absorbing it with a CAP, it is obvious that for broader resonances a stronger CAP (i.e. with smaller c and larger η) is needed. On the other hand, the wavefunction, which is absorbed by a stronger CAP, is localized in a smaller area around the molecule. Therefore, for broader resonances, which must be absorbed at smaller internuclear distances, a less diffuse basis set is needed when compared to calculations on narrower resonances. To keep the basis set size manageably small one is thus forced to use different basis sets for different internuclear distances, each basis set being used on its own interval. The basis sets differ only in the diffuse parts and, therefore, the influence on the energy should not be large. Unfortunately, this impact of the basis set is not completely negligible, making the dependence of the resonance energy on internuclear distance discontinuous. Therefore, we calculated the resonance energy in both basis sets at all the border points of the above-mentioned intervals. Then we simply shifted all the energy values in a given interval by the difference at the border point. As the basic reference we took the basis set in which the (electronically) bound part of the potential energy curve was calculated. After these shifts the dissociation curve became smooth, as is shown in figure 1(a). The resonance energy of the $^2\Sigma_u^+$ state of F_2^- is also depicted in figure 1(b), showing the individually calculated points.

Table 2. Resonance widths at those internuclear distances where the calculation is performed in two different basis sets. The internuclear distances are given in atomic units and the energies in eV. For the basis-set notation see table 1.

Internuclear distance	A1	A2	A3	A4	B1	B2
2.45					0.0331	0.0310
2.40				0.1099	0.1055	
2.38			0.1637	0.1632		
2.32		0.3917	0.3732			
2.30	0.4564	0.4733				

Within the different basis sets used, the resonance widths Γ turned out to be surprisingly independent of the basis set employed. This can be observed at the points where the calculation was performed using two different basis sets. These points are listed in table 2. Making allowances for the rather complex situation of correlating 15 active electrons, the differences are considered as negligibly small. The dependence of the resonance width on the internuclear distance is plotted in figure 2(a) where we compare our data with those of several other investigations. As one can see, our resonance widths are smaller than those of the other studies. However, a well pronounced tendency can be observed: the resonance widths decrease systematically with the amount of correlation accounted for. The values of our calculated resonance energies and widths are listed in table 3.

An important datum, which may indicate how accurate our results are, is the point where the resonance is turned into a bound state. This point corresponds to the crossing point of the potential energy curve of the ionic state with that of the ground state of the neutral F_2 molecule, because beyond this point there is no final state to which the resonance could decay. We have determined this point in two ways. The first way is to extrapolate the resonance width to its zero value. A linear extrapolation performed using the last three points (see figure 2(b)) indicates a crossing point at an internuclear distance of about 2.60 au. Because of the neglect of the

Table 3. List of the calculated resonance energies and widths. The resonance energy (fourth column) is the energy difference between those of the X $^2\Sigma_u^+$ F$_2^-$ and X $^1\Sigma_g^+$ F$_2$ states.

Internuclear distance (au)	Energy (au) + 198 au	Width (eV)	Resonance energy (eV)
2.57	−0.9322	0.002 99	0.5757
2.56	−0.9278	0.003 81	0.6844
2.55	−0.9233	0.004 89	0.7945
2.54	−0.9187	0.005 82	0.9060
2.53	−0.9140	0.007 62	1.0189
2.52	−0.9091	0.008 87	1.1356
2.51	−0.9042	0.010 34	1.2509
2.50	−0.8982	0.012 24	1.3945
2.48	−0.8888	0.018 50	1.6061
2.47	−0.8835	0.021 22	1.7256
2.45	−0.8726	0.033 73	1.9674
2.44	−0.8670	0.039 71	2.0898
2.42	−0.8557	0.065 82	2.3319
2.40	−0.8445	0.105 54	2.6011
2.38	−0.8338	0.163 20	2.7744
2.36	−0.8193	0.221 95	3.0790
2.34	−0.8065	0.293 76	3.3263
2.32	−0.7928	0.373 18	3.5855
2.30	−0.7783	0.456 42	3.8518
2.28	−0.7631	0.571 20	4.1216

curvature of this curve there is an extrapolation error and the true crossing point is expected to lie at a slightly larger internuclear distance. Therefore, we can consider this value as a lower limit for the location of the crossing point.

The second, more accurate determination of the crossing point makes use of the Wigner threshold law, which states that

$$\Gamma(R(E)) =: \Gamma(E) \sim E^{l+1/2} \tag{5}$$

holds for low energies. Here E denotes the energy difference between the ionic and neutral states, and $R(E)$ is the inverse function of the dependence of this difference on the internuclear distance. The symbol l denotes the orbital angular momentum of the outgoing electron. Since in the present case the outgoing electron is of p-symmetry, the exponent is $\frac{3}{2}$. The form of the thus evaluated curve $\Gamma(E)$ varies considerably when the potential energy curves of the ionic and neutral states are shifted vertically with respect to each other. Shifting the curves such that the expected $E^{3/2}$ behaviour is obeyed for low energies (see figure 3) allows us to determine the crossing point very sensitively. The thus predicted crossing point is located at 2.623 au.

We did not recompute the potential energy curve of the ground state F$_2$, but rather took the curve (MC 178/basis B) computed by Blomberg and Siegbahn [11], which corresponds well with the experimental values for the equilibrium distance and dissociation energy.

Knowing the location of the crossing point, one can immediately determine an important experimentally observable property, the electron affinity of atomic fluorine. The calculation of this well known quantity provides a very useful check on the correctness of our method. Since the location of the crossing point also determines the relative vertical position of the two potential energy curves one can determine the electron affinity of atomic fluorine as the energy difference for large R where the molecule and molecular ion can be considered as being dissociated. The thus determined electron affinity of a fluorine atom is 3.54 eV, whereas
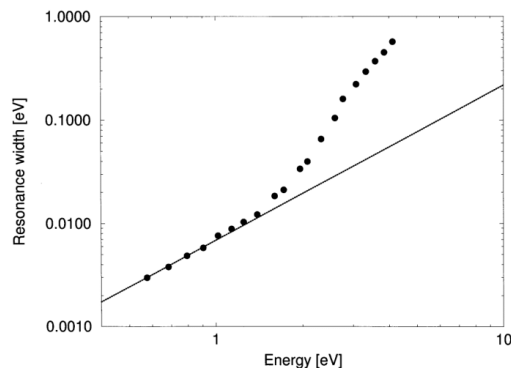
Figure 3. The full circles depict the resonance width Γ versus the resonance energy (the latter with respect to the neutral state). Dotted curve represents a curve $\Gamma \sim E^{3/2}$.

the experimental value is 3.4012 eV [12]. This means that the error of our calculation is 4%, which is a striking level of precision for this kind of calculation. It is well known that it is very difficult to achieve the same level of correlation in independent calculations on F_2 and F_2^-. Therefore, it is almost impossible to localize the potential energy curves correctly with respect to each other when one has at ones disposal only the two curves obtained from independent calculations. This indeed makes the information on the crossing point very valuable.

The very reasonable agreement of our predicted electron affinity with its experimental value is a strong argument for the accuracy and reliability of the CAP/CI method. Moreover, the described procedure provides the location of a crossing point in a systematic way, and does not rely on coincidental cancellations of correlation errors. This procedure should also be applicable to other systems.

4. Conclusions

In this paper we demonstrate the use of the CAP/CI method for calculating resonance parameters over a relatively large region of molecular geometries. The crossing point of the potential energy curves of F_2 and F_2^- is determined by investigating the resonance width, and the two computed curves are adjusted vertically such that the Wigner threshold law is obeyed. This procedure determines the relative position of the potential energy curves and yields the electron affinity. The latter is in very reasonable agreement with the experimental value. The computed resonance parameters can serve as input parameters for the dynamical calculation of processes such as dissociative attachment and vibrationally excited resonant electron–molecule scattering. Finally, this work encourages further use of the CAP/CI method for calculating other resonance states of the fluorine molecule (i.e. $^2\Pi_g$, $^2\Pi_u$, $^2\Sigma_g^+$). This is planned as a future continuation of the project.

"Schwellenverhalten Resonanzen und nichtlokale Wechselwirkungen bei niederenergetischen Elektronenstreuprozessen" an der Universität Kaiserslautern'.

References

[1] Jolicard G and Austin E J 1985 *Chem. Phys. Lett.* **121** 106
[2] Jolicard G and Austin E J 1986 *Chem. Phys.* **103** 295
[3] Riss U V and Meyer H-D 1993 *J. Phys. B: At. Mol. Opt. Phys.* **26** 4503
[4] Sommerfeld T, Riss U V, Meyer H-D and Cederbaum L S 1997 *Phys. Rev. Lett.* **79** 1237
[5] Sommerfeld T, Riss U V, Meyer H-D, Cederbaum L S, Engels B and Sutter H U 1998 *J. Phys. B: At. Mol. Opt. Phys.* **31** 4107
[6] Santra R, Cederbaum L S and Meyer H-D 1999 *Chem. Phys. Lett.* **303** 413
[7] Hazi A U, Orel A E and Rescigno T N 1981 *Phys. Rev. Lett.* **46** 918
[8] Lauderdale J G, McCurdy C W and Hazi A U 1983 *J. Chem. Phys.* **79** 2200
[9] Morgan L A and Noble C J 1984 *J. Phys. B: At. Mol. Phys.* **17** L369
[10] Dunning T H 1971 *J. Chem. Phys.* **55** 716
[11] Blomberg M R A and Siegbahn P E M 1981 *Chem. Phys. Lett.* **81** 4
[12] Blondel C, Cacciani P, Delsart C and Trainham R 1989 *Phys. Rev.* A **40** 3698

204

P14. Enzymological description of multitemplate PCR-Shrinking amplification bias by optimizing the polymerase-template ratio.

M. Ingr, J. Dostál, T. Majerová, J. Theor. Biol. 382 (2015) 178–186.

Enzymological description of multitemplate PCR—Shrinking amplification bias by optimizing the polymerase–template ratio

CrossMark

Marek Ingr [a,b], Jiří Dostál [c], Taťána Majerová [c,*]

[a] Tomas Bata University in Zlín, Faculty of Technology, Department of Physics and Materials Engineering, Nám. T.G. Masaryka 5555, 76001 Zlín, Czech Republic
[b] Department of Biochemistry, Faculty of Science, Charles University, Hlavova 2030, 128 43 Prague 2, Czech Republic
[c] Gilead Sciences and IOCB Research Center, Institute of Organic Chemistry and Biochemistry of Academy of Sciences of the Czech Republic, v.v.i., Flemingovo nam. 2, 166 10 Prague 6, Czech Republic

HIGHLIGHTS

- We use enzymological approach to describe multitemplate DNA synthesis (PCR, qPCR, SNuPE).
- We elucidate the importance of polymerase–template ratios.
- We discuss the risk of rare template loss or products ratio alterations.

GRAPHICAL ABSTRACT
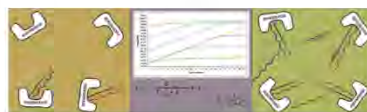
ABSTRACT

Multitemplate polymerase chain reaction (PCR) is used for preparative and analytical applications in diagnostics and research. Classical PCR and qPCR are two basic setups with many possible experimental modifications. Classical PCR is a method of choice to obtain enough material for subsequent sophisticated applications such as construction of libraries for next-generation sequencing or high-throughput screening. Sequencing and Single Nucleotide Primer Extension (SNuPE) employ one-strand synthesis and represent a distinct variant of analytical DNA synthesis.

In all these applications, maintaining the initial ratio of templates and avoiding underestimation of minority templates is desired.

Here, we demonstrate that different templates can amplify independently at low template concentrations (typical in qPCR setups, in which the polymerase concentration is usually several orders of magnitude higher than the template concentration). However, rare templates can be diluted in an effort to keep DNA amplification in the exponential phase, or template concentration can be biased by differences in amplification efficiency. Moreover, amplification of templates present in low concentrations is more vulnerable to stochastic events that lead to proportional changes in the product ratio, as well as by incomplete amplification leading to chimera formation. These undesired effects can be compensated for by using highly processive polymerases with high and equal affinity to different primer–template complexes. Novel enhanced polymerases are desired.

With increasing concentration of a primer–template of interest, the system becomes more deterministic.

Nevertheless, marked deviation from independent exponential amplification occurs when the total template concentration starts to approach the polymerase concentration. The primer–template complexes compete for enzyme molecules, and the amount of products grows arithmetically—the system starts to obey Michaelis–Menten kinetics. Synthesis of rare products in a multitemplate mixture can run more easily under the detection limit in such conditions, although it would be unequivocally detectable in a single template assay.

* Corresponding author. Tel.: +420 220183 218; fax: +420 220183 203.
E-mail address: tatana.majerova@uochb.cas.cz (T. Majerová).

When fishing out rare template variants, the best processive polymerases should be used to decrease both amplification and detection limits. The possibility of stochastic events, should be taken into account to correctly interpret the obtained data.

1. Introduction

The characterization of the first DNA polymerase (Lehman et al., 1958) and pioneering works in oligonucleotide-directed DNA synthesis (Kleppe et al., 1971; Panet and Khorana, 1974) preceded the development of polymerase chain reaction (PCR) (Saiki et al., 1985; Mullis, 1987). PCR has become a widely used technique in research, biotechnological, forensic, and clinical laboratories, and the method has undergone many improvements and modifications since its initial development.

The basic PCR setup includes heat denaturation of the double-stranded DNA template, followed by cooling to a temperature that allows annealing of each strand with one of the primers, which are present at higher concentrations than the template. The cycle is completed by setting the temperature to an optimum at which the annealed primers are extended by DNA polymerase. These three steps are typically repeated 20 to 40 times.

Multitemplate PCR involves single-tube amplification of two or more homologous templates, typically with one primer pair, whereas single-tube amplification of several non-homologous sequences with distinct primer pairs is termed multiplex PCR (Kalle et al., 2014). Our conclusions are relevant for both cases. Only formation of hetero-duplexes and possible synthesis of chimeric molecules predominantly threat co-amplification of homologous templates.

Different setups are useful in many applications, including accurate quantification of template variants by quantitative PCR (qPCR) (Stahlberg and Kubista, 2014).

Classical PCR is a method of choice to obtain a sufficient amount of DNA for subsequent analyses, including drug-resistance evaluation (Jia et al., 2014) or construction of libraries (e.g., for next-generation sequencing van Dijk et al., 2014), and it is involved in many applications including species diversity studies (Shokralla et al., 2012).

Sequencing-by-synthesis methods rely on detection of primer-extension products. Typically, only one primer is used (Buchan and Ledeboer, 2014). Single nucleotide primer extension (SNuPE) is a variant in which a primer anneals to the target sequence with its 3' end immediately adjacent to the polymorphic site. The primer is extended by DNA polymerase using fluorescent dideoxynucleotides, and the identity of the polymorphic site is revealed by identification of the incorporated fluorescent dideoxynucleotide (Nikolausz et al., 2009).

In all multitemplate PCR applications, maintaining the initial templates ratio is necessary, and knowledge of the minimal detection limit of rare template variants is essential.

Several causes of bias in multitemplate PCR setups have been reported: competition between templates (Sykes et al., 1992), competition of complementary template strands with primers (Suzuki and Giovannoni, 1996), insufficient denaturation efficiency (Booth et al., 2010), primer mismatch, different primer annealing temperatures (Booth et al., 2010), different primer elongation efficiencies (Booth et al., 2010) (usually due to different amplicon length and different GC content: Baskaran et al., 1996; Arezi et al., 2003), polymerase error rates, and high cycle numbers (Patin et al., 2013).

These factors can lead to changes in template ratios during multitemplate DNA synthesis and to underestimation of rare template quantities.

Here, we use an enzymological approach to describe different PCR setups and applications. Enzyme kinetics has been already applied to Quantitative Competitive PCR (Schnell and Mendoza, 1997a), where high template concentrations allow description using Michaelis–Menten equation. We acknowledge Schnell and Mendoza (1997a,b) for the concept of reaction efficiency.

We discuss advantages as well as pitfalls connected to different ratios of templates and polymerase. In initial cycles, the template concentration is typically several orders of magnitude lower than the enzyme concentration. In this range, templates amplify exponentially and independently, and the efficiency of amplification is not influenced by template concentration. Nevertheless, rare templates can be significantly influenced by stochastic events as well as by chimera formation. Use of highly processive polymerases with increased affinity to primer-template complexes could reduce these complications.

With increasing template concentration, the system becomes deterministic and more robust. However, when the total template concentration reaches the polymerase concentration, the templates start to compete with each other and the amount of products increases arithmetically only. Amplification of rare templates can run under the detection limit, although it would be easily detected in a single-template PCR. On the other hand, Michaelis–Menten kinetics can be applied.

2. Results and discussion

From an enzymological point of view, PCR amplification can be considered as a series of enzymatic reactions. During each cycle, a template-primer complex is converted into a product *via* enzymatic catalysis. The product of one cycle becomes the substrate for the subsequent cycle.

2.1. One substrate system

Under ideal PCR conditions, one template molecule gives rise to two product molecules. The exponential increase in copy number of a template after n cycles (T_n) can be described as

$$T_n = 2^n T_0. \tag{1}$$

In reality, the amplification efficiency is generally lower than 1; therefore, the base of the exponential function in Eq. (1) falls below 2. Saiki et al. (1985) described the amplification efficiency in the nth cycle as

$$T_n = (1+\eta)^n T_0, \tag{2}$$

where η is the amplification efficiency.

A single step of DNA synthesis in PCR can be described by the following mechanism. First, the enzyme–substrate complex (ES) is formed, in which the primer-template complex is considered the substrate (S) and DNA polymerase the enzyme (E). Second, the primer contained in ES is elongated by the enzyme in order to synthesize the new DNA strand. In this model, the second step is considered to be 100% efficient.

The formation of ES can be described as follows:

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES. \tag{3}$$

This process runs at the annealing temperature, at which no elongation is considered to take place. Therefore, this reaction can be considered to be in equilibrium described by the equilibrium constant of *ES* dissociation

$$K_d = \frac{S.E}{Q} = \frac{k_{-1}}{k_1}. \tag{4}$$

For the brevity of notation, S and E stand for the concentrations of substrate and enzyme, respectively, while Q denotes the concentration of *ES*. In this model, concentrations of primers and dNTPs are considered to be present in large excess during the amplification process. Taking into account the balance equations of substrate and enzyme

$$S_{tot} = S + Q \tag{5}$$

$$E_{tot} = E + Q, \tag{6}$$

where S_{tot} and E_{tot} are the total substrate and enzyme concentrations, respectively, the following expression is obtained for the concentration of the enzyme–substrate complex Q (for the detailed derivation, see Appendix A):

$$Q = \frac{E_{tot} + S_{tot} + K_d - \sqrt{(E_{tot} + S_{tot} + K_d)^2 - 4E_{tot}S_{tot}}}{2}. \tag{7}$$

The situation can be in principle more complex if the primer annealing-dissociation equilibrium is also considered at higher than the annealing temperature, especially the polymerization temperature. In such a case interplay of this equilibrium with the processes of polymerase association and dissociation with the template-primer complex and with the polymerization of the DNA chain should be taken into account. However, as shown by Marimuthu et al. (2014), the rate contant of primer dissociation grows rapidly with the temperature, while the association rate constant varies very weakly. Therefore, the equilibrium constant of primer dissociation, i.e., the instability of the primer-template complex, grows strongly with temperature. If we restrict our interest to "well-designed" PCR reactions, in which the primers are designed to dissociate already at temperatures slightly higher than the annealing temperature set on the instrument (in order to decrease non-specific annealing), we can assume that at the polymerization temperature of 72 °C the primers are no more annealed to the template unless they are stabilized by the associated molecule of DNA polymerase. This assumption allows us to separate the processes of annealing and polymerization. However, when the polymerase dissociates after elongation of only a part of the desired sequence, it is possible to assume that this partial product remains associated with the template since it is considerably longer, and elongation can continue within the same cycle after binding of another polymerase molecule.

Consider 10 μl of a reaction mixture containing 1 copy of a template corresponding to a concentration of 1.66×10^{-19} mol/l and 5 U of DNA polymerase corresponding to $E_{tot} = 250$ nM (TopBio 2014. Private information) with an assumed K_d of 1.5×10^{-9} mol/l (Sambrook J. 2001).

Fig. 1 shows amounts of PCR products after each amplification cycle, starting from various initial template concentrations. The non-linear curves were obtained using Eq. (7). The corresponding lines were obtained using Eq. (1).

Ideal conditions – excess dNTPs, 100% pairing of primers with template, 100% amplification efficiency – were assumed in both cases.

Fig. 1 demonstrates that, at low substrate concentrations, the amount of product doubles after each cycle. As the template concentration approaches the value $E_{tot} + K_d$, a deviation from linearity is observed.
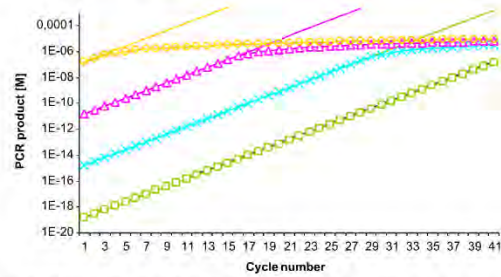
Fig. 1. Amount of PCR product after each cycle of PCR amplification for a single substrate, starting from different initial template concentrations: 1 template molecule (squares), 10^4 template molecules (crosses), 10^8 molecules (triangles), 10^{12} molecules (circles). The final volume of the reaction mixture was 10 μl; 1 copy of a template per 10 μl corresponds to 1.66×10^{-19} mol/l. K_d corresponds to 1.5×10^{-9} mol/l, E_{tot} was 5 U/10 μl, corresponding to 250 nM. Ideal conditions were assumed—excess dNTPs, 100% pairing of primers with the template, 100% amplification efficiency. The non-linear curves were obtained using Eq. (7). The corresponding linear curves were obtained using Eq. (1).

At low template concentrations (typically in the initial cycles), at which the total enzyme concentration greatly exceeds the total concentration of the substrate in a given step (S_n), the PCR product grows exponentially and assumptions for application of Michalis–Menten kinetics are not fulfilled. However, the steady state can be applied inversely. Eq. (7) can be reduced to the form (for derivation, see Appendix A)

$$Q_n = \frac{E_{tot}S_n}{E_{tot} + K_d}. \tag{8}$$

The efficiency η_n of the nth PCR cycle can be interpreted as the ratio of the concentration of *ES* complex at this step (Q_n) and the total substrate concentration in the same step (S_n)

$$\eta_n = \frac{Q_n}{S_n}. \tag{9}$$

Therefore,

$$\eta_n = \frac{E_{tot}}{E_{tot} + K_d} \equiv \eta_0. \tag{10}$$

η_0 symbolizes the independence of the reaction efficiency from the cycle number within the limits of the approximation $E_{tot} \gg S_n$. Within this range, amplification efficiency in not influenced by template concentration. The product concentration after the nth step, which is equal to the substrate concentration before the $(n+1)$th step, is

$$S_{n+1} = (1 + \eta_0)^n S_1. \tag{11}$$

Taking the natural logarithm of this expression, the following equation is obtained:

$$\ln S_{n+1} = n \ln(1 + \eta_0) + \ln S_1. \tag{12}$$

Plotting this function *vs.* n provides a line with a slope of $\ln(1 + \eta_0)$ and a y-axis intercept at $\ln S_1$, which is shown in Fig. 1.

The substrate concentration grows exponentially until it is comparable to the total enzyme concentration. Eq. (7) without simplifications is relevant in these cases.

After several subsequent cycles of PCR, the substrate concentration exceeds the enzyme concentration ($S_n \gg E_{tot}$). Therefore, the amplification is no longer exponential, as the maximum amount of PCR product obtained in a single cycle cannot exceed the amount of active enzyme. The PCR product grows arithmetically. The concentration of the *ES* complex is then given by the

expression (see Appendix A.2)

$$Q_n = \frac{E_{tot}S_n}{S_n + K_d}. \tag{13}$$

This equation is formally analogous to Michaelis–Menten enzyme kinetics, as pointed out by Schnell and Mendoza (1997a, 1997b). They described competitive PCR (cPCR) amplification in terms of initial velocities for target and competitor reactions in the $(n+1)$th cycle using classical Michaelis–Menten kinetics. In these cases the use of the Michaelis–Menten equation is valid only, since the total primer–template concentration exceeds the enzyme concentration.

Apparently, if the ES complex is highly stable, i.e., K_d is small compared to S_n, the expression is reduced to $Q_n \approx E_{tot}$, which means that every enzyme molecule produces a molecule of PCR product. If the formation of the ES complex is less efficient, the PCR product synthesis depends on substrate concentration and grows to the asymptote $Q_n = E_{tot}$. The efficiency of the amplification

$$\eta_n = \frac{E_{tot}}{S_n + K_d} \tag{14}$$

decreases continuously with growing S_n (increasing with the cycle number). Although other conditions are ideal and primers and dNTPs are in excess. Therefore, the growth of the PCR product slows down during the course of the reaction, as can be seen in Fig. 1.

Fig. 2 clarifies the role of K_d. The initial exponential part of the amplification lengthens with increasing K_d. However, both curves start to move away from exponential amplification when the polymerase becomes saturated. Several polymerases with artificially improved processivity (i.e., the number of incorporated nucleotides per one polymerase binding event) have been reported (Acevedo et al., 2014; Ishino and Ishino, 2014; Yamagami et al., 2014). Although processivity is a complex phenomenon, one of the successful strategies to obtain a more processive polymerase employed fusion of a polymerase with a DNA binding domain (Sso7d) (Wang et al., 2004), resulting in the commercial product "Phusion DNA polymerase." This enzyme is widely used for amplification of long fragments. Obviously, fusion of a polymerase with a DNA binding domain leads to a decrease in K_d compared to that of the original enzyme. Such engineered polymerases can also be suitable for amplification of extremely low-concentrated samples, as well as for multitemplate PCR, to preserve the correct product ratio for rare template variants, as will be further discussed.
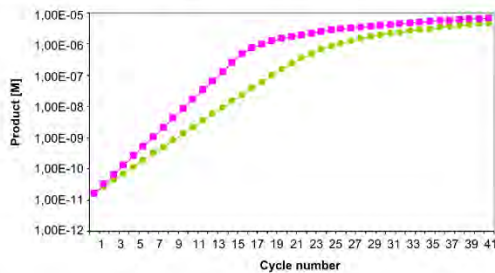
Fig. 2. Increasing K_d prolongs the linear part of the amplification curve. The data were obtained using Eq. (7), starting from 10^8 initial template molecules. K_d was 1.5×10^{-9} mol/l (squares) or 1.5×10^{-7} mol/l (circles). The final volume of the reaction mixture was 10 μl; 1 copy of a template per 10 μl corresponds to 1.66×10^{-19} mol/l. E_{tot} was 5 U/10 μl, corresponding to 250 nM. Ideal conditions were assumed—excess dNTPs, 100% pairing of primers with the template, 100% amplification efficiency.

Taken together, these observations suggest that when PCR runs with excess dNTPs and primers, the product concentration initially grows exponentially with constant efficiency η_0, depending only on polymerase concentration and its properties. Highly processive polymerases with a lowered K_d can reach the ideal case, in which each template doubles after each cycle. When product concentration approaches the total enzyme concentration, the curve representing the product concentration deflects from the original trend. At high cycle numbers, the growth is linear, as the number of product molecules synthesized in one cycle is equal to the enzyme concentration (in an ideal case). The amplification efficiency decreases in each subsequent cycle as the amount of template increases. As Fig. 1 illustrates, the position of the deflection point depends on the initial template concentration. Hence, when template concentration is very low, a considerable part of the amplification takes place exponentially. On the other hand, if the initial template concentration is rather high, the exponential amplification range may be very short or even vanish completely.

2.2. Multisubstrate system

The situation with several templates and corresponding primer pairs (irrespective of whether one primer pair has several binding sites or there are several primer pairs in the system) can be described as

$$E + \sum_i S_i \rightleftarrows \sum_i ES_i \tag{15}$$

$$E_{tot} = E + \sum_i Q_i. \tag{16}$$

Assuming the same K_d for all the templates,

$$K_d = \frac{S_i.E}{Q_i} = \frac{k_{-1}}{k_1}. \tag{17}$$

The total enzyme concentration can be written as

$$E_{tot} = Q_1 + \frac{K_d Q_1}{S_{1,tot} - Q_1} + Q_1 \sum_{i \neq 1} \frac{S_{i,tot}}{S_{1,tot}}. \tag{18}$$

The concentration of the ES_1 complex is given by the expression

$$Q_1 = \frac{S_{1,tot}}{2} \frac{\sum_i S_{i,tot} + K_d + E_{tot} - \sqrt{\left(\sum_i S_{i,tot} + K_d + E_{tot}\right)^2 - 4E_{tot}\sum_i S_{i,tot}}}{\sum_i S_{i,tot}} \tag{19}$$

and analogously for Q_2, Q_3, etc. (for detailed derivation, see Appendix A.1).

Fig. 3 shows data for amplification of two different templates obtained using Eq. (19). The sum of template concentrations in relation to enzyme concentration and K_d is the determining factor. When

$$\sum_i S_i < E + K_d \tag{20}$$

each template molecule gives rise to two product molecules (linear part of the amplification curves in Fig. 3) and all the substrates amplify independently. A sufficient amount of polymerase is present to bind each primer-template complex. Assuming $K_d = 1.5 \times 10^{-9}$ mol/l and $E_{tot} = 5$ U/10 μl, corresponding to 250 nM, the total template concentration must be below 10^{-7} mol/l. This corresponds to approximately 10^{12} template molecules (10 μl of a reaction mixture containing 1 copy of a template corresponds to a concentration of 1.66×10^{-19} mol/l). This amount is reached after more than 40 cycles when starting from 1 template molecule, after 27 cycles when starting from 10^4 template molecules, and after 13 cycles when starting from 10^8
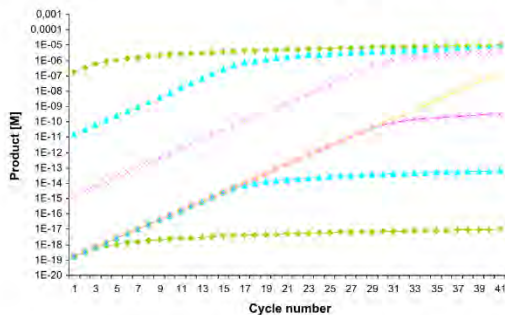
Fig. 3. Two substrate system. Amount of PCR product corresponding to template 1 after each cycle of PCR amplification, calculated using Eq. (19), starting from one template 1 molecule (solid lines) in the presence of 1 (circles), 10^4 (crosses), 10^8 (triangles), or 10^{12} (diamonds) molecules of template 2. The corresponding dashed lines represent the amount of template 2 present after each amplification cycle. The final volume of the reaction mixture was 10 µl; 1 copy of a template per 10 µl corresponds to 1.66×10^{-19} mol/l. K_d corresponds to 1.5×10^{-9} mol/l, E_{tot} was 5 U/ 10 µl, corresponding to 250 nM. Ideal conditions were assumed—excess dNTPs, 100% pairing of primers with the template, 100% amplification efficiency. The non-linear curves were obtained using Eq. (19).

template molecules. Thus, an initial template amount of approximately 10^8 molecules seems to be the upper limit at which amplification of different templates is independent, at least in the initial cycles.

In an ideal case, a single template molecule could be proportionally amplified from a mixture of 10^8 different template molecules. However, real amplification of rare templates is influenced by the stochastic nature of polymerase binding. Fig. 4 shows results of a computer simulation of amplification of a single molecule in the presence of 10^8 molecules of other template variants. Details of the simulation are given in Supplement B. Each graph shows results of 20 independent simulations under one set of conditions. Overall, the results show that a decrease in K_d leads to less scattered data closely related to an average value. The PCR drift (Wagner et al., 1994) is lower. Decreased processivity (incorporated into this simulation as the probability that an extended primer will polymerize to the full-length product) leads to more scattered data. The amount of undesired chimeras arising from annealing of partially extended primers to a different template variant is comparable with the amount of desired product.

As the copy number of individual template variants increases, the system becomes more deterministic and robust. When the concentration of templates exceeds the polymerase concentration, the templates start to compete for polymerase, the amount of products decreases after each cycle, and Michaelis–Menten kinetics can be applied. This can be used for competitive PCR (cPCR) amplification in terms of initial velocities for target and competitor reactions in the $(n+1)$th cycle using classical Michaelis–Menten kinetics (Schnell and Mendoza, 1997a).

At these higher template concentrations, formation of chimeras by template switch becomes more probable (Odelberg et al., 1995).

A tendency toward a 1:1 ratio of all products represents another phenomenon in which more abundant templates can become less available for amplification because of competing complementary template strands with primers for templates (Suzuki and Giovannoni, 1996).

2.3. Polymerase chain reaction (PCR)

Classical PCR is a powerful technique to obtain DNA pools of interest in sufficient amount for a wide range of subsequent applications including preparation of libraries for high-throughput screening and next-generation sequencing. However, it also represents a bias-prone step.

Biases caused by different template properties can lead to changes in product ratios. Changes in amplification efficiencies due to differences in amplicon lengths and AT/GC ratio can be reduced by fragment size selection, by inclusion of additives that equilibrate the melting temperature between AT and GC pairs, and by selection of a polymerase able to amplify sequences with different GC content as equally as possible (Kappa HiFi) (Quail et al., 2012).

To avoid amplification bias, the total capability of a reaction to amplify all the primer-templates independently also should be considered in the design of PCR experiments.

For a typical PCR experimental setup, a starting amount of 10^4 template molecules (1 pg–1 ng of plasmid or viral templates, 1 ng– 1 µg of genomic templates) is a common recommendation (crosses in Fig. 3). In this setup, 25 cycles seem to be reliable with a safety margin for independent amplification of different templates. Starting from 10^8 template molecules (triangles in Fig. 3), the first 11 cycles seem to be reliable. In subsequent cycles, the amplification starts to follow Eq. (19) with no negligible terms, and primer-templates compete for polymerase molecules.

Rare primer-templates are more influenced by the stochastic nature of polymerase binding in the first cycles. Thus, polymerases with higher affinity to primer-templates (with a very low K_d) can give better results. Such polymerases have been engineered in order to increase their processivity and make them more suitable for amplification of long templates. However, their potential to reduce stochastic amplification of very low-concentrated template variants makes them also suitable for multitemplate amplifications, in which preservation of the initial ratio of templates and/or fishing out of rare variants is desired, even for short amplification products. Due to the discontinuous nature of PCR amplification – at least one molecule must be present to be amplified and no lower dilutions are possible – the principal detection limit of a rare template variant is one molecule in a mixture (i.e., 1:10^4 starting from 10^4 template molecules or 1:10^8 starting from 10^8 template molecules).

However, templates present in only several copies, are influenced by the stochastic nature of amplification. Thus, for deterministic amplification 100–1000 molecules of a rare template seem to be sufficient. To keep the PCR in the exponential phase, the total concentration of templates should be under 10^8 template molecules. This moves the theoretical lower-amplification limit of a rare variant to 1: 10^5 template molecules, even when a polymerase a very low K_d would be used.

When a polymerase has higher K_d the amplification efficiency decreases and very rare templates amplify randomly. The exponential phase is prolonged. Nevertheless, the amplification curve bends down at the same template concentration as in the ideal case (Fig. 2). More cycles are necessary to reach the limit before the exponential phase of amplification ends. The same situation can occur with some types of polymerase-inhibiting impurities (working as competitive inhibitors). Other types of impurities, working irreversibly or noncompetitively, simply decrease the effective polymerase concentration. Thus, the amplification stops being exponential at a lower template concentration, meaning that fewer total templates are present to be amplified independently. In both the situations, the limit of deterministic amplification of a rare variant moves to higher concentrations.

Separation of multitemplate amplification into multiple individual reactions can prevent from changes in products ratio and chimeras production.

Emulsion PCR relies on compartmentalization of genes in a water-in-oil emulsion. This method is suitable for preparation of genomic libraries, cDNA libraries and other complex samples (Williams et al., 2006).
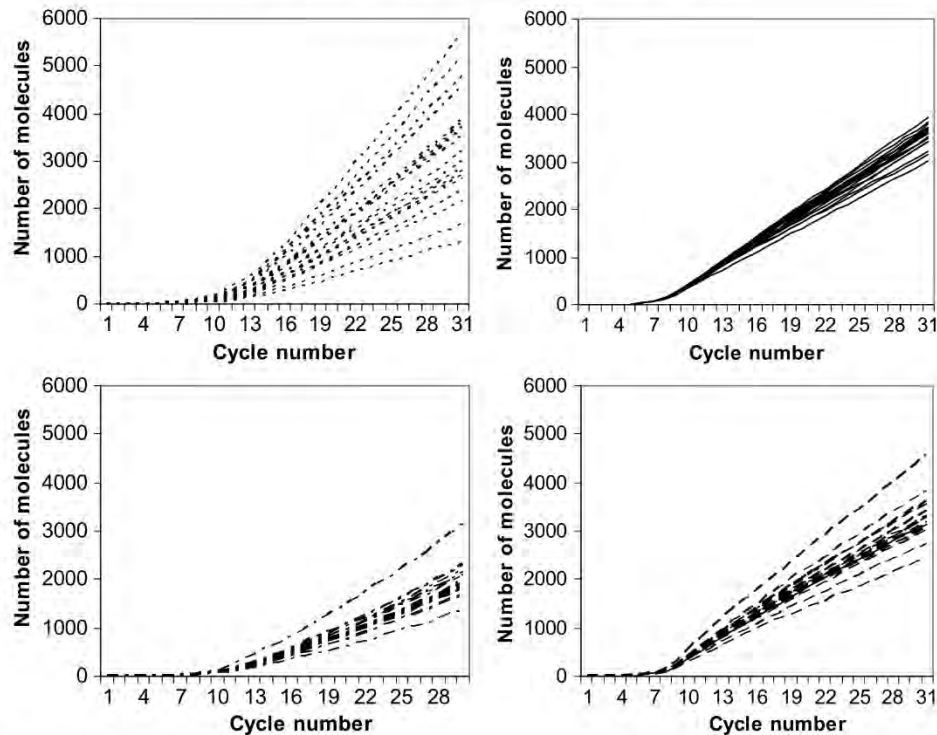
Fig. 4. Simulation of stochastic amplification: One molecule of a template of interest in presence of 10^8 molecules of competing template(s). 20 independent runs for a set of conditions: K_d 1.5 nM, processivity 100% (dotted lines), K_d 15 pM, processivity 100%, (full lines), K_d 15 pM, processivity 90% (dot-and-dash lines lines), K_d 15 pM, processivity 90%, chimeric molecules (dashed lines).

During partition enrichment of nucleotide sequences (PINS) (Kvist et al., 2014), in which a sample is divided into several tubes, the product of interest is detected after amplification. Tubes containing positive hits can be subdivided into another set of tubes until the enrichment is sufficient for subsequent analyses. This procedure can be used in classical and qPCR setups and can help obtain rare variants from a complex mixture.

For example, PINS can be employed to search for drug-resistant mutants of infectious pathogens in samples from infected individuals undergoing therapy. Drug-resistant mutants can rapidly emerge under selection pressure from a drug. Early detection of possible rare drug-resistant variants can help physicians select the most potent drugs to optimize individual drug therapy.

2.4. Quantitative PCR (qPCR)

The requirement for a low template concentration is usually fulfilled in qPCR, in which very low amounts of product(s) can be monitored with high sensitivity after each cycle. qPCR can detect minor changes in product concentrations as well as provide information about whether the reaction is running within the exponential range. In principle, it can detect even a single template molecule. When only one variant of a primer-template complex is present in a reaction mixture, increasing the number of cycles can help detect a template present at a very low concentration. However, when a rare variant is present in a mixture of other templates, increasing the number of cycles could bring the system into the non-exponential amplification phase. Due to competition with other templates, the amplification curve can bends down before it reaches the real detection limit of the method used. In such cases, the exponential amplification of the rare variant can dip below the detection limit and go unnoticed (see Fig. 3, solid lines). For detection of rare variants in mixtures of templates, use of a qPCR setup with high detection sensitivity is much more important than in singleplex assays. Highly processive polymerases give a higher yield of PCR amplification, they could increase the signal change per initial amount of a templates. They also shorten the initial random phase, suppressing stochastic events due to their higher affinity to primer-templates. Such polymerases could make qPCR more sensitive and reliable.

Moreover, the resolution of a qPCR machine, as well as the assay, is important. The difference between the amount of a corresponding PCR product in the first and late cycles of the plateau phase decreases with increasing total amount of templates. This difference is the same for all the amplification products in an individual run (see pairs of solid and dashed lines in Fig. 3). High initial primer-template concentrations can decrease the resolution of all the variants in the mixture.

In multitemplate setups, all primer–template complexes present in a reaction mixture must be taken into account as substrates for a polymerase, even if they are not amplified (e.g., 3′-mispairing primers for allele-specific discrimination) or not detected (e.g., TaqMan probes that detect only the DNA variant of interest), because they are able to compete with the primer–template of interest for polymerase binding.

Assays blocking the synthesis of the strand of interest can be advantageous in some cases and can increase the detection limit of rare template variants, if the initial amount of competing templates is below the limit for exponential amplification. Minor groove binders (MGB) can specifically block synthesis of a competing template. MGB blockers are usually used to suppress the wild type allele (Dufresne et al., 2014). This approach increases the sensitivity of detection of the rare allele by eliminating amplification of competing template, ensuring that the total primer-template concentrations are below the limit of possible competition. In such a situation, it does not matter whether the polymerase also can bind to the template-oligonucleotide blocker. In cases with extremely high initial amounts of majority templates that are under the concentration needed for exponential amplification before the reaction starts, specific degradation of competing variants could be advantageous.

2.5. Digital PCR

Digital PCR exploits a strikingly similar principle as PINS. Digital PCR is based on limited template dilution (Sykes et al., 1992) and can be a method of choice to identify very rare template variants in mixtures. In digital PCR, the sample is split into a set of individual compartments; each compartment undergoes PCR amplification. The number of compartments containing the template of interest corresponds to the proportion of this template in the initial mixture.

This method can be used in an end-point setup or can be performed in a qPCR machine (Mojtahedi et al., 2014). In these applications, the need for low template concentration can be fulfilled more easily, providing more reliable data for rare templates. Droplet digital PCR (ddPCR) splits PCR reactions into thousands of droplets, and detects the amplification in each of them. This method can give more precise results than qPCR (Doi et al., 2015).

2.6. Single primer applications

In methods using only one primer (or several primers annealing to the same DNA strand), no new template molecules are synthesized. The template concentration *vs.* polymerase concentration is constant over the course of the reaction (assuming 100% enzymatic stability) and the reaction products sum over all the cycles (sequencing, SNuPE).

Many of these applications are preceded by classical PCR to obtain enough material for subsequent analyses. This fact should be considered along with the data interpretation, because the sample heterogeneity may be reduced due to loss of rare variants during preparation of the pool (Gonzalez et al., 2012). These factors should be considered in evaluation of heterogeneity of microbial communities as well as in applications such as evaluation of drug resistance.

In some situations, high template concentration can be advantageous, since the stochastic effects are suppressed.

At high template-primer concentrations, the amount of polymerase becomes significantly lower than that of its substrates, and the system follows Michaelis–Menten kinetics as described by Schnell and Mendoza (Schnell and Mendoza, 1997a, 1997b). This situation can occur in applications such as SNuPE, which uses a single primer that anneals immediately adjacent to the nucleotide position to be analyzed and is extended by a single fluorescent nucleotide.

Eq. (13) can be rewritten for multiple substrates as follows:

$$Q_1 = E_{tot} \frac{S_1}{K_d + \sum_i S_i}. \tag{21}$$

When

$$\sum_i S_i \gg E_{tot} + K_d \tag{22}$$

and thus

$$\sum_i S_i \gg K_d, \tag{23}$$

it can be written

$$Q_1 = E_{tot} \frac{S_1}{\sum_i S_i} \tag{25}$$

$$\frac{Q_1}{S_1} = \frac{Q_i}{S_i} \tag{26}$$

$$\frac{S_1}{S_i} = \frac{Q_1}{Q_i}. \tag{27}$$

The primer–template ratios correspond to the product ratios when a polymerase is saturated and if the reactions have the same efficiency, as also has been deduced using initial reaction rates (Schnell and Mendoza, 1997a).

Moreover, the amount of competing templates can be estimated by detecting a defined amount of template in a single reaction and its same amount in a mixture of competing templates. The difference between the yield of the reaction with and without a competing template, would correspond to the amount of this competing template. When the absolute concentration of template variant 1 is known, the absolute amount of the competing DNAs ($\sum_i S_i$) can be deduced. However, the standard procedure employs detection both the competing products present in a single reaction (Wang et al., 1989).

2.7. Two limiting cases—Two substrates, one enzyme, and different K_d in low and high total substrate ranges

A polymerase may prefer certain primer sequence motifs, leading to a different ratio of products compared to the starting sample (Pan et al., 2014).

Case 1. When an enzyme is present in excess over the initial total substrate concentration, the amount of enzyme–substrate complexes is mutually independent and follows the same equations as single substrate reactions:

$$Q_1 = \frac{S_{10}E}{K_1 + E} \tag{28}$$

$$Q_2 = \frac{S_{20}E}{K_2 + E}. \tag{29}$$

Case 2. When the total substrate concentration is in excess over the enzyme concentration, then

$$Q_1 = \frac{S_1 E_0}{K_1 + S_1 + S_2 \frac{K_1}{K_2}} \tag{30}$$

$$Q_2 = \frac{S_2 E_0}{K_2 + S_2 + S_1 \frac{K_2}{K_1}}. \tag{31}$$

If the equilibrium constants differ significantly (e.g., $K_2 \gg K_1$), S_1 binds considerably better to the enzyme than S_2. In Case 1, the resulting product ratio will be influenced by the ratio of equilibrium constants and by the total enzyme concentration, while in Case 2 the substrate concentrations can influence the product ratio significantly. Modified primers or templates can be used to favor fishing out of minority templates.

3. Conclusion

PCR can be misconducted by primer mismatch, different or unspecific primer annealing, by competition of the complementary strand with a primer, insufficient denaturation efficiency, and errors introduced by a polymerase.

In addition, multitemplate PCR can be biased mainly by differences in primer elongation efficiency (usually, different amplicon lengths or different GC content play a role), by templates competition and by formation of heteroduplexes and chimeras.

Use of low template concentrations can lead to loss of rare templates by dilution or by uneven amplification due to stochastic events. Polymerases with increased affinity to primer-template can give more reproducible results for rare templates.

Our model can help to guess the lowest detection or quantification limits of rare templates in complex mixtures. The principal qPCR quantification limit of a rare variant can be considered as the possible lowest concentration for deterministic amplification of this rare template vs. the highest total templates amount, which enables exponential amplification over several cycles, i.e. does not exceed the enzyme concentration during them.

At high template concentrations, the amplification of one template is influenced by the sum of all template concentrations. Higher template concentrations can artificially increase the sample homogeneity by signal loss of rare templates due to template competition.

Subsequent rounds of PCR can be a reasonable control: An aliquot of the previous reaction mixture is taken as a template for the subsequent reaction. If the ratio of different templates remains constant over several rounds of PCR amplification (or primer extension) the sample composition can be considered representative. However loss of very low templates cannot be excluded.

Acknowledgements

Appendix A

A.1 Derivation of equation for the amount of enzyme–substrate complex in a one-substrate system

The symbols S and E stand for the concentrations of substrate and enzyme, respectively, while Q denotes the concentration of the enzyme–substrate complex ES. In this model, concentrations of primers and dNTPs are considered to be present in large excess during the amplification process. S_{tot} and E_{tot} are the total substrate and enzyme concentrations, respectively. The equation describing the concentration of ES can be derived as follows:

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES. \tag{A1}$$

$$K_d = \frac{SE}{Q} = \frac{k_{-1}}{k_1}. \tag{A2}$$

$$S_{tot} = S + Q \tag{A3}$$

$$E_{tot} = E + Q, \tag{A4}$$

$$K_d = \frac{(S_{tot} - Q)(E_{tot} - Q)}{Q}. \tag{A5}$$

After rearrangement, the following quadratic equation is obtained

$$Q^2 - Q(S_{tot} + E_{tot} + K_d) + S_{tot}E_{tot} = 0, \tag{A6}$$

which can be solved as (see also Supplement A)

$$Q = E_{tot} + S_{tot} + K_d - \frac{\sqrt{(E_{tot} + S_{tot} + K_d)^2 - 4E_{tot}S_{tot}}}{2}. \tag{A7}$$

A.2 Simplification of Eq. (7) at high concentrations of substrates or polymerase

Eq. (A.7) is symmetric with respect to E_{tot} and S_{tot}.

Thus, both the extreme cases of excess enzyme or substrate can be considered together.

Consider that $E_{tot} \gg S_{tot}$. The limit solution can be obtained with the aid of the binomial expansion of the square root according to the formula

$$(a + b)^n = \sum_{i=0}^{n} \binom{n}{i} a^{n-i} b^i. \tag{A8}$$

In this case $n = 1/2$ and $a = 1$; therefore, the formula takes the form

$$(1 + x)^{1/2} \approx 1 + \frac{1}{2}x + \cdots \quad . \tag{A9}$$

Neglecting all higher terms than the second one, Eq. (A.7) can be rearranged in the following way:

$$Q = \frac{E_{tot} + S_{tot} + K_{tot}}{2}\left(1 - \sqrt{1 - \frac{4E_{tot}S_{tot}}{(E_{tot} + S_{tot} + K_d)^2}}\right) \approx \frac{E_{tot} + S_{tot} + K_d}{2}$$

$$\times \left[1 - \left(1 - \frac{2E_{tot}S_{tot}}{(E_{tot} + S_{tot} + K_d)^2}\right)\right] = \frac{E_{tot}S_{tot}}{E_{tot} + S_{tot} + K_d}. \tag{A10}$$

Provided that the total enzyme concentration greatly exceeds the total concentration of substrate, the term S_{tot} can be neglected from the denominator of the last expression, giving the final solution

$$Q \approx \frac{E_{tot}S_{tot}}{E_{tot} + K_d}. \tag{A11}$$

Analogously, in the case that $S_{tot} \gg E_{tot}$

$$Q \approx \frac{E_{tot}S_{tot}}{S_{tot} + K_d}. \tag{A12}$$

A.3 Derivation of equation for the amount of enzyme–substrate complex in a two-substrate system

Consider a PCR reaction with several different substrates that form different enzyme–substrate complexes. The equilibrium constant K_d is assumed to be equal for all the substrates. In such a case, the following set of equations should be solved:

$$K_d = \frac{S_i E}{Q_i} \tag{A13}$$

$$S_{i,tot} = S_i + Q_i \tag{A14}$$

$$E_{tot} = E + \sum_i Q_i. \tag{A15}$$

The index i runs over all the templates; therefore, the set contains i equations of type (A.13), i equations of type (A.14), and one equation of type (A.15). Combining Eqs. (A.13) and (A.14) for

$i=1$, the following expression for E is obtained:

$$E = \frac{K_d Q_1}{S_{1,\text{tot}} - Q_1}. \tag{A16}$$

Combination of this result with the rest of Eqs. (A.14) and (A.16) for $i \neq 1$ gives

$$Q_i = Q_1 \frac{S_{i,\text{tot}}}{S_{1,\text{tot}}}. \tag{A17}$$

Eq. (A.15) then takes the form

$$E_{\text{tot}} = \frac{K_d Q_1}{S_{1,\text{tot}} - Q_1} + Q_1 + Q_1 \sum_{i \neq 1} \frac{S_{i,\text{tot}}}{S_{1,\text{tot}}} \tag{A18}$$

which can be rearranged to a quadratic equation

$$Q_1^2 \left(1 + \sum_{i \neq 1} \frac{S_{i,\text{tot}}}{S_{1,\text{tot}}}\right) - Q_1 (E_{\text{tot}} + K_d + \sum_i S_{i,\text{tot}}) + E_{\text{tot}} S_{1,\text{tot}} = 0, \tag{A19}$$

the solution of which is

$$Q_1 = E_{\text{tot}} + K_d + \sum_i S_{i,\text{tot}} - \frac{\sqrt{(E_{\text{tot}} + K_d + \sum_i S_{i,\text{tot}})^2 - 4E_{\text{tot}} \sum_i S_{i,\text{tot}}}}{2 \left(1 + \sum_{i \neq 1} \frac{S_{i,\text{tot}}}{S_{1,\text{tot}}}\right)}, \tag{A20}$$

which after a simple rearrangement, can be presented in the form

$$Q_1 = \frac{S_{1,\text{tot}}}{2} \frac{E_{\text{tot}} + K_d + \sum_i S_{i,\text{tot}} - \sqrt{(E_{\text{tot}} + K_d + \sum_i S_{i,\text{tot}})^2 - 4E_{\text{tot}} \sum_i S_{i,\text{tot}}}}{\sum_i S_{i,\text{tot}}}, \tag{A21}$$

The enzyme–substrate complex concentrations for the other substrates, i.e., Q_j for $j \neq i$, can be easily derived by rotating the indices:

$$Q_j = \frac{S_{j,\text{tot}}}{2} \frac{E_{\text{tot}} + K_d + \sum_i S_{i,\text{tot}} - \sqrt{(E_{\text{tot}} + K_d + \sum_i S_{i,\text{tot}})^2 - 4E_{\text{tot}} \sum_i S_{i,\text{tot}}}}{\sum_i S_{i,\text{tot}}}, \tag{A22}$$

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.jtbi.2015.06.048.

References

Acevedo, A., Brodsky, L., Andino, R., 2014. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. Nature 505, 686–690.
Arezi, B., Xing, W.M., Sorge, J.A., Hogrefe, H.H., 2003. Amplification efficiency of thermostable DNA polymerases. Anal. Biochem. 321, 226–235.
Baskaran, N., Kandpal, R.P., Bhargava, A.K., Glynn, M.W., Bale, A., Weissman, S.M., 1996. Uniform amplification of a mixture of deoxyribonucleic acids with varying GC content. Genome Res. 6, 633–638.
Booth, C.S., Pienaar, E., Termaat, J.R., Whitney, S.E., Louw, T.M., Viljoen, H.J., 2010. Efficiency of the polymerase chain reaction. Chem. Eng. Sci. 65, 4996–5006.
Buchan, B.W., Ledeboer, N.A., 2014. Emerging technologies for the clinical microbiology laboratory. Clin. Microbiol. Rev. 27, 783–822.
Doi, H., Uchii, K., Takahara, T., Matsuhashi, S., Yamanaka, H., Minamoto, T., 2015. Use of droplet digital PCR for estimation of fish abundance and biomass in environmental DNA surveys. PLoS One, 10.
Dufresne, A., Alberti, L., Brahmi, M., Kabani, S., Philippon, H., Perol, D., Blay, J., 2014. Impact of KIT exon 10 M541L allelic variant on the response to imatinib in aggressive fibromatosis: analysis of the desminib series by competitive allele specific Taqman PCR technology. BMC Cancer 14, 632.

Gonzalez, J.M., Portillo, M.C., Belda-Ferre, P., Mira, A., 2012. Amplification by PCR artificially reduces the proportion of the rare biosphere in microbial communities. PLoS One, 7.
Ishino, S., Ishino, Y., 2014. DNA polymerases as useful reagents for biotechnology—the history of developmental research in the field. Front. Microbiol., 5.
Jia, S., Wang, F., Li, F., Chang, K., Yang, S., Zhang, K., Jiang, W., Shang, Y., Deng, S., Chen, M., 2014. Rapid detection of hepatitis B virus variants associated with lamivudine and adefovir resistance by multiplex ligation-dependent probe amplification combined with real-time PCR. J. Clin. Microbiol. 52, 460–466.
Kalle, E., Kubista, M., Rensing, C., 2014. Multi-template polymerase chain reaction. Biomol. Detect. Quantification 2, 11–29.
Kleppe, K., Ohtsuka, E., Kleppe, R., Molineux, I., Khorana, H.G., 1971. Studies on polynucleotides.96. repair replication of short synthetic DNAs as catalyzed by DNA polymerases. J. Mol. Biol. 56 341-.
Kvist, T., Sondt-Marcussen, L., Mikkelsen, M.J., 2014. Partition enrichment of nucleotide sequences (PINS)—a generally applicable, sequence based method for enrichment of complex DNA samples. PLoS One, 9.
Lehman, I.R., Bessman, M.J., Simms, E.S., Kornberg, A., 1958. Enzymatic synthesis of deoxyribonucleic acid. 1. Preparation of substrates and partial purification of an enzyme from Escherichia-coli. J. Biol. Chem. 233, 163–170.
Marimuthu, K., Jing, C., Chakrabarti, R., 2014. Sequence-dependent biophysical modeling of DNA amplification. Biophys. J. 107, 1731–1743.
Mojtahedi, M., Fouquier d'Herouel, A., Huang, S., 2014. Direct elicitation of template concentration from quantification cycle (Cq) distributions in digital PCR. Nucleic Acids Res. 42, e126–e126.
Mullis, K.B. 1987. Process for amplifying nucleic acid sequences. U.S. Patent 4,683,202.
Nikolausz, M., Chatzinotas, A., Tancsics, A., Imfeld, G., Kaestner, M., 2009. The single-nucleotide primer extension (SNuPE) method for the multiplex detection of various DNA sequences: from detection of point mutations to microbial ecology. Biochem. Soc. Trans. 37, 454–459.
Odelberg, S.J., Weiss, R.B., Hata, A., White, R., 1995. Template-switching during DNA synthesis by Thermus aquaticus DNA polymerase I. Nucleic Acids Res. 23, 2049–2057.
Pan, W., Byrne-Steele, M., Wang, C., Lu, S., Clemmons, S., Zahorchak, R.J., Han, J., 2014. DNA polymerase preference determines PCR priming efficiency. BMC Biotechnol., 14.
Panet, A., Khorana, H.G., 1974. Studies on polynucleotides.127. Linkage of deoxyribopolynucleotide templates to cellulose and its use in their replication. Int. J. Biol. Chem. 249, 5213–5221.
Patin, N.V., Kunin, V., Lidstrom, U., Ashby, M.N., 2013. Effects of OTU clustering and PCR artifacts on microbial diversity estimates. Microb. Ecol. 65, 709–719.
Quail, M.A., Otto, T.D., Gu, Y., Harris, S.R., Skelly, T.F., McQuillan, J.A., Swerdlow, H.P., Oyola, S.O., 2012. Optimal enzymes for amplifying sequencing libraries. Nat. Methods 9, 10–11.
Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., Arnheim, N., 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle-cell anemia. Science 230, 1350–1354.
Sambrook J R.W, 2001. Molecular Cloning: A Laboratory Manual vol. 1, p. 535.
Shokralla, S., Spall, J.L., Gibson, J.F., Hajibabaei, M., 2012. Next-generation sequencing technologies for environmental DNA research. Mol. Ecol. 21, 1794–1805.
Schnell, S., Mendoza, C., 1997a. Enzymological considerations for a theoretical description of the quantitative competitive polymerase chain reaction (QC-PCR). J. Theor. Biol. 184, 433–440.
Schnell, S., Mendoza, C., 1997b. Theoretical description of the polymerase chain reaction. J. Theor. Biol. 188, 313–318.
Stahlberg, A., Kubista, M., 2014. The workflow of single-cell expression profiling using quantitative real-time PCR. Expert Rev. Mol. Diagnost. 14, 323–331.
Suzuki, M.T., Giovannoni, S.J., 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. 62, 625–630.
Sykes, P.J., Neoh, S.H., Brisco, M.J., Hughes, E., Condon, J., Morley, A.A., 1992. Quantitation of targets for PCR by use of limiting dilution. Biotechniques 13, 444–449.
van Dijk, E.L., Jaszczyszyn, Y., Thermes, C., 2014. Library preparation methods for next-generation sequencing: Tone down the bias. Exp. Cell Res. 322, 12–20.
Wagner, A., Blackstone, N., Cartwright, P., Dick, M., Misof, B., Snow, P., Wagner, G.P., Bartels, J., Murtha, M., Pendleton, J., 1994. Surveys of gene families using polymerase chain-reaction—PCR selection and PCR drift. Syst. Biol. 43, 250–261.
Wang, A.M., Doyle, M.V., Mark, D.F., 1989. Quantitation of mRNA by the polymerase chain reaction. Proc. Natl. Acad. Sci. U.S.A. 86, 9717–9721.
Wang, Y., Prosen, D.E., Mei, L., Sullivan, J.C., Finney, M., Vander Horn, P.B., 2004. A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. Nucleic Acids Res. 32, 1197–1207.
Williams, R., Peisajovich, S.G., Miller, O.J., Magdassi, S., Tawfik, D.S., Griffiths, A.D., 2006. Amplification of complex gene libraries by emulsion PCR. Nat. Meth. 3, 545–550.
Yamagami, T., Ishino, S., Kawarabayasi, Y., Ishino, Y., 2014. Mutant Taq DNA polymerases with improved elongation ability as a useful reagent for genetic engineering. Front. Microbiol. 5, 461.

P15. Theoretical description of the direct exponential amplification and sequencing (DEXAS) method.

M. Ingr, J. Konvalinka, Biol. Chem. 381 (2000) 439–445.

Theoretical Description of the Direct Exponential Amplification and Sequencing (DEXAS) Method

Marek Ingr* and Jan Konvalinka

Institute of Organic Chemistry and Biochemistry,
Academy of Sciences of the Czech Republic,
Flemingovo nám. 2, CZ-166 10 Praha 6,
The Czech Republic

* Corresponding author

We present a theoretical description of the method of DNA sequencing with simultaneous exponential PCR amplification of the template (DEXAS). Based on the theory of probability, the formula determining the optimal ratio of concentrations of deoxy- and dideoxynucleotides in the reaction mixture is derived, as well as the length distribution of sequenced DNA fragments. The prediction of the number of mutations is given and the theoretically determined aspects of DEXAS are compared with the corresponding quantities of classical sequencing methods. Some other experimentally observed effects are also discussed.
Key words: Calculation / Cycle sequencing / DNA / PCR amplification / Theory of probability.

Introduction

The Sanger dideoxy-sequencing method with laser fluorescent detection of the electrophoretic bands (Smith *et al.*, 1985) is presently the most common experimental technique for determining the primary structure of DNA. Although this method has been routinely used for a long time, it is still an attractive field for further refinement, aiming at higher throughput and sensitivity of the method. The throughput of sequencing has been improved by the simultaneous performance of the sequencing reactions on both DNA strands, distinguishing the fragments by different fluorescent labels attached to the primers (Wiemann *et al.*, 1995). This technique increases the efficiency of double-stranded DNA sequencing by the factor of two. On the other hand, this method requires an independent detection system for each strand, and thus all the primers must be fluorescently labeled, which makes the method rather expensive. The latter problem was solved by the method of *in situ* labeling of originally unlabeled primers directly in the reaction mixture (Wiemann *et al.*, 1996). An alternative way of increasing the efficacy of the method is the four-dye detection of the fragments with labeled dideoxynucleotides. This method is, however, known for its lower reliability because of the spectral overlap of the emission spectra of the four fluorescent labels (Smith *et al.*, 1985,

1986, 1987). Methods based on the 4 × 4 transformation matrix from detector space to dye space were suggested to overcome this problem (Smith *et al.*, 1987; Yin *et al.*, 1996). An alternative way of sensitive detection of four different labels consists in the time-resolved fluorescent detection (Lieberwirth *et al.*, 1998).

With respect to the sensitivity, methods of PCR amplification are often used to yield more template DNA (see *e. g.* Voss *et al.*, 1997). A method that couples PCR amplification and DNA sequencing was originally proposed by Ruano and Kidd (1991). In this method, the sample is amplified by PCR and the following sequencing reaction is performed as a further PCR amplification in the presence of polymerization terminators. Based on a similar principle, but with two differently labeled primers present in the reaction mixture, a novel method was introduced recently, that combines the PCR amplification with the DNA sequencing from both ends. As this method produces the sequencing fragments simultaneously with the template molecules, the authors called it the method of direct exponential amplification and sequencing (DEXAS) (Kilger and Pääbo, 1997a). In this method four reaction mixtures are prepared containing template DNA molecules, all four deoxynucleotides (dN), one selected dideoxynucleotide (ddN), thermostable DNA polymerase, and a pair of fluorescently labeled primers. Subsequent PCR yields:
a) a new template molecule (extension product reaching the opposite primer site), and
b) a sequencing ladder (extension products terminated by the corresponding ddNs).
The ratio of amplification and sequencing (*i. e.* the specificity and yield of the method) can be controlled by the ratio of dNs and ddNs. It is obvious that the labeled primers are necessary for this method, because the *in situ* labeling cannot be used since there is not enough template at the beginning. In spite of this disadvantage, some recent applications demonstrate the fact that DEXAS can become a powerful tool for sequencing of small amounts of DNA (Plaschke *et al.*,1998).

So far, the crucial parameters of DEXAS (ratio of dNs and ddNs, number of PCR cycles) have been only estimated without the support of any underlying theory. Therefore, a theoretical description of this method might help to reach the optimal reaction conditions. Several articles were published in the past which provide theoretical analyses of PCR-based processes. Most of them concentrate on the effect of random mutations during PCR. In one of the first works of this sort (Krawczak *et al.*, 1989), the percentage of correctly synthesized PCR product is deter-

mined by means of the theory of probability as a function of the enzyme error rate and the number of PCR cycles. Several more recent reports deal with the task of determining the error rate of the enzyme from an analysis of the PCR product (Weiss and von Haeseler, 1995; Sun, 1995). Computer simulations were used to determine the error rate of the enzyme as well as the homogeneity of the original template (Weiss and von Haeseler, 1997).

In the following we provide a detailed theoretical analysis of the DEXAS method and determine the dependence of the output of the DEXAS reaction on the initial reactant ratio. We also compare this method with the standard techniques and discuss its advantages and disadvantages.

Results

The DEXAS method is based on the simultaneous presence of dNs and ddNs in the reaction mixture. Both of them can be incorporated in the growing DNA strand and the polymerization is either continued or stopped. Provided that both dNs and ddNs are present for the particular base (which is of course only one of four in each reaction), the probability of incorporating the terminator (ddN) into a given position in the growing DNA chain is

$$P_t = Ra, \tag{1}$$

where R is defined as $c(ddN)/(c(dN) + c(ddN))$, c stands for the molar concentration, and a is the relative affinity of the polymerase to ddNs, $i.\,e.$, the probability of ddN incorporation into the synthesized DNA strand provided that the concentrations of dNs and ddNs in the reaction mixture are equal. Note that both R and a are obviously numbers between 0 and 1, therefore also P_t belongs to this interval. Taq DNA polymerase, the most common PCR enzyme, suppresses considerably the incorporation of modified nucleotides (including ddNs). To increase the affinity of the enzyme to modified nucleotides, several mutated analogs of Taq DNA polymerase were designed. In general they possess the mutation F667Y and/or several different modifications, especially deletions at the N-terminus. Their affinity to modified nucleotides was tested by Voss $et\ al.$ (1997). This affinity varies from one enzyme to another but in some cases it is almost as high as for the normal dNs. The DNA polymerase with only the F667Y mutation was found to be especially suitable for incorporating ddNs (Kilger and Pääbo, 1997b). The value of R can be adjusted by the experimentator, and our aim is to determine its optimum value; a is a constant that has to be determined experimentally.

For the sake of clarity of all the following considerations we will suppose an even distribution of all four bases in the template molecule. The probability of incorporation of a terminating nucleotide into the extending DNA strand is

$$P_{term} = \frac{1}{4} P_t, \tag{2}$$

where P_t is defined by eq. (1). The factor $\frac{1}{4}$ refers to the

fact that only one ddN is present in a given reaction mixture. The probability of chain extension at this site is then

$$P_{cont} = 1 - P_{term} = 1 - \frac{1}{4} P_t. \tag{3}$$

The probability of synthesis of a chain which is just N nucleotides long is

$$P_{chain}(N) = P_{cont}^{N-1} P_{term}. \tag{4}$$

This expression means that the chain must be $(N-1)$ times extended and then terminated at the end. A similar expression can be derived for the probability of synthesis of a chain, which is at least N nucleotides long:

$$P_{chain+}(N) = P_{cont}^{N-1}. \tag{5}$$

After $N - 1$ elongations the chain may be terminated or further extended.

To provide a new template molecule, the polymerization must proceed to the distant end of the other primer. At least N_t nucleotides have to be incorporated without termination of the reaction. (N_t denotes the number of nucleotides between the 3' end of the upper primer and the 5' end of the lower one.) Therefore we introduce the fundamental quantity P_{amp} (probability of amplification), which is the probability of synthesis of a new template DNA molecule:

$$P_{amp} = P_{chain+}(N_t) = \left(1 - \frac{1}{4} P_t\right)^{N_t-1}. \tag{6}$$

Once we have expressed the probability of amplification, we can determine the dependence of the amount of template molecules on the number of PCR cycles and thus describe the amplification process. Assuming that the initial number of template molecules is T_0, the number of these molecules after one cycle of PCR is

$$T_1 = T_0 + T_0 P_{amp} = T_0(1 + P_{amp}). \tag{7}$$

The corresponding equation for the number of template molecules after n PCR cycles can be derived in an analogous way

$$T_n = T_0(1 + P_{amp})^n. \tag{8}$$

Equation (8) has a straightforward physical meaning. The factor $(1 + P_{amp})$ is the quotient of the geometrical sequence describing the amplification process. In an ordinary PCR amplification, where no terminators are present, $P_{amp} = 1$, therefore $T_n = T_0\,2^n$. This means that the number of template molecules is doubled in every cycle, which is generally supposed in PCR. However, when the terminators are present, the amplification quotient equals a number between 1 and 2 depending on the values of R and a in eq. (1) as a consequence of the formation of sequencing fragments.

After determining the amount of template after every PCR cycle we can derive the number of sequencing fragments as a function of the number of cycles. We count all the fragments together irrespective of their lengths. The distribution function for the fragment lengths will be derived later. It is obvious that the initial number of fragments

S_0 (after 0 cycles) is equal to zero. The number of ddN terminated fragments after 1 PCR cycle is equal to the initial number of template molecules multiplied by the complementary probability to the probability of amplification:

$$S_1 = T_0(1 - P_{amp}). \tag{9}$$

It can be easily seen from eq. (9), that in a PCR experiment without terminators $S_1 = 0$, because $P_{amp} = 1$. Summing the contributions from n subsequent cycles results in the formula

$$S_n = T_0(1 - P_{amp})[1 + (1 + P_{amp}) + (1 + P_{amp})^2 + \ldots + (1 + P_{amp})^{n-1} \tag{10}$$

The right side of eq. (10) is the sum of n terms of a geometrical series with the quotient $(1 + P_{amp})$. This equation can be thus simplified to the form

$$S_n = T_0(1 - P_{amp})\frac{(1 + P_{amp})^n - 1}{(1 + P_{amp}) - 1} = T_0\frac{(1 - P_{amp})}{P_{amp}}[(1 + P_{amp})^n - 1] \tag{11}$$

Eq. (11) expresses the number of sequencing fragments present in the reaction mixture after n cycles of the PCR
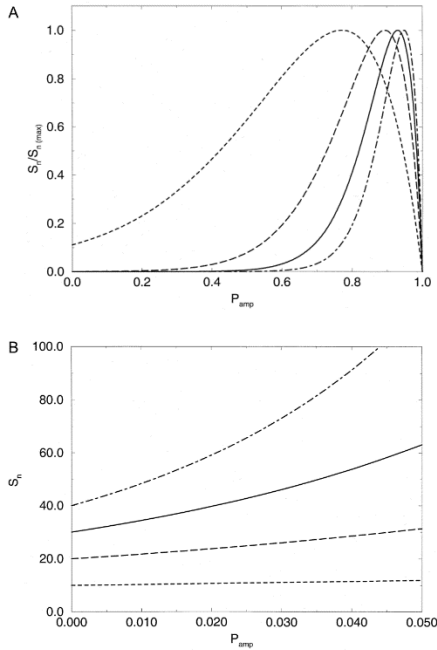
Fig. 1   Influence of P_{amp} on the Number of Sequencing Fragments.
(A) Number of sequencing fragments S_n as a function of P_{amp}. Each curve is divided by its maximum value. Dashed line: ten PCR cycles; long-dashed line: twenty PCR cycles; solid line: thirty PCR cycles; dot-dashed line: fourty PCR cycles. The maximum values of S_n/T_0 are 8.99×10^2, 4.18×10^4, 2.77×10^7 and 2.10×10^{10}, respectively. (B) Limit of S_n for $P_{amp} \to 0$. Notation of the curves is the same as in (A).

amplification. Several examples of this dependence for different numbers of PCR cycles are given in Figure 1A.

Equation 11 can be used for the determination of the optimal value of R as a function of the template length and the number of PCR cycles. Hence, we are searching for the maximum of S_n as a function of P_{amp}. For this purpose we differentiate S_n with repect to P_{amp} and put the resulting expression equal to zero. After some rearrangements we obtain the following equation

$$(1 + P_{amp})^{n-1}[nP_{amp}^2 - (n-1)P_{amp} + 1] = 1 \tag{12}$$

Due to practical reasons P_{amp} should not fall below about 0.4 and the number of PCR cycles below 20. In this region the expression $(1 + P_{amp})^{n-1}$ is very high, therefore the polynomial on the left side is considerably steep at the point where its value equals 1. Hence, this point is very close to the point where the value of the polynomial equals 0. Therefore, the constant 1 on the right side can be substituted by zero, which results in an approximate equation

$$nP_{amp}^2 - (n-1)P_{amp} + 1 = 0 \tag{13}$$

which is a quadratic equation for P_{amp} that has two roots

$$P_{amp\,1,2}^{opt} = \frac{1}{2n}[n-1 \pm \sqrt{n^2 - 6n + 1}] \tag{14}$$

The root with minus is obviously non-physical and represents an artifact of the above mentioned approximation. It is obviously located very close to zero, where we do not assume the validity of this approximation. On the contrary, the 'plus' root is a good approximation of the location of the maximum of S_n as a function of P_{amp} and, for $n \geq 20$, it is located between 0.89 and 1. It can be shown that in this region of n the error of the approximation is much less than 0.001. We can thus determine an optimal ratio of dNs and ddNs in the reaction mixture for a given number of PCR cycles making use of the eqs. (1) and (6).

$$R_{opt} = \frac{4}{a}(1 - \sqrt[N_t-1]{P_{amp}^{opt}}) \tag{15}$$

This formula answers the basic question of DEXAS, *i. e.* what R to choose for an optimal yield of sequencing fragments.

Kilger and Pääbo (1997b) suggested to use two different DNA polymerases in the same PCR, one of them strongly preferring dNs and the other with a relatively higher affinity to ddNs (the latter possessing the F667Y mutation). The authors state that the sensitivity of this modification is considerably higher than in the case discussed above. They propose its application for sequencing of a single-copy DNA molecule.

For the theoretical analysis of this approach we have to introduce two quantities a_1 and a_2 [analogous to a in eq. (1)] specific for the two polymerases. Therefore, we also obtain two different probabilities, P_{t1} and P_{t2}. We assume that the affinity to substrate (complex template-primer) is equivalent for both enzymes. Let us denote the fractions of the two polymerases with respect to the whole enzyme pool f_1 and f_2. Now we can determine the probability of am-

plification, now denoted P_{ampd}, in an analogous way. For simplicity, we will exclude all other causes of termination of the polymerization than incorporation of a terminator or reaching the end of the template molecule. This means that only one enzyme molecule can bind to one template strand in one PCR cycle. In such a case the probability P_{ampd} is simply equal to the weighed average of the probabilities of amplification of the two individual enzymes, where the fractions f_1 and f_2 represent the weights:

$$P_{ampd} = f_1 P_{amp1} + f_2 P_{amp2} \tag{16}$$

Hence, in the case of two different enzymes there is one more degree of freedom in the dependence of R_{opt} on P_{ampd}. Therefore, R_{opt} has no more only a single value, but there is a whole set of its values depending on the fractions of the two polymerases. This set is represented by a curve in a two-dimensional space described by the equation

$$\frac{P_{ampd} - \left(1 - \frac{1}{4} R_{opt} a_2\right)^{N-1}}{\left(1 - \frac{1}{4} R_{opt} a_1\right)^{N-1} - \left(1 - \frac{1}{4} R_{opt} a_2\right)^{N-1}} = f_1 \tag{17}$$

Several functions of this kind are plotted in Figure 2. Note that not all values of R are allowed for a given P_{ampd}. Depending on the values of a_1 and a_2 two forbidden regions can occur, where f_1 is (formally) either negative or higher than 1, which is of course a non-physical situation. These regions are obviously the larger the closer the values of a_1 and a_2 are. In the trivial case of their equality the curve in Figure 2 reduces to a point given by eq. (15). The borders of the allowed region are given by the expressions

$$f_1 = 0 : R = \frac{4}{a_2} (1 - \sqrt[N-1]{P_{ampd}}) \tag{18}$$

and

$$f_1 = 1 : R = \frac{4}{a_1} (1 - \sqrt[N-1]{P_{ampd}}) \tag{19}$$

Another important quantity of DEXAS is the distribution of the sequencing fragments as a function of their length. It is obvious that the procedure generates more shorter fragments than longer ones. Therefore it is important to determine the distribution of the lengths of fragments in the sequencing product. After putting eqs. (2,3) into eq. (4) and normalizing it (the fragments of the lengths between 1 and N_t are synthesized with the probability equal to one), we obtain a final expression for the distribution of fragment lengths

$$D(N) = P_{chain}(N) = \frac{\left(1 - \frac{1}{4} P_t\right)^{N-1} \frac{1}{4} P_t}{1 - \left(1 - \frac{1}{4} P_t\right)^{N_t}} = \frac{\left(1 - \frac{1}{4} P_t\right)^{N-1} \frac{1}{4} P_t}{1 - \left(1 - \frac{1}{4} P_t\right) P_{amp}} \tag{20}$$

The last expression results from the substitution of eq. (6) into the original expression for the distribution function. Examples of several such distributions are given in Figure 3.

Every PCR-based method may suffer from the problems with random mutations introduced due to the infideli-
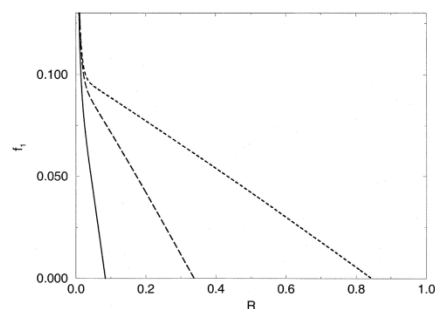
Fig. 2 Dependence of f_1 on R for Given Parameters P_{ampd}, a_1 and a_2.

In all cases $P_{ampd} = 0.9$ and $a_1 = 1$. Solid line: $a_2 = 0.01$; long-dashed line: $a_2 = 0.0025$; dashed line: $a_2 = 0.001$.
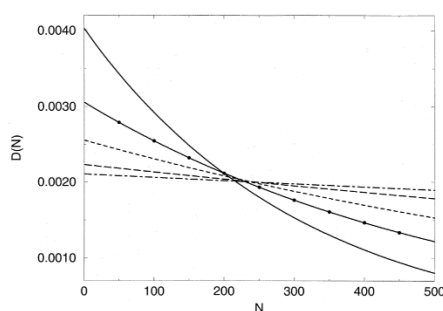
Fig. 3 Normalized Distribution Function of the Fragment Lengths.

$N_t = 500$. Solid line: $P_{amp} = 0.2$; solid line with dots: $P_{amp} = 0.4$; dashed line: $P_{amp} = 0.6$; long-dashed line: $P_{amp} = 0.8$; dot-dashed line: $P_{amp} = 0.9$.

ty of the DNA polymerase. This is certainly the case also for DEXAS. The amount of mutated DNA present in the reaction mixture after n cycles of PCR is a function of the error probability P_{err} of the enzyme. This is the probability of incorporating an incorrect nucleotide at a given position in the growing DNA chain, and it is a characteristic constant of the enzyme that can be determined experimentally.

The probability of a correct incorporation of a nucleotide is $1 - P_{err}$. If there are no terminators in the reaction, then the probability of a correct transcription is simply $(1 - P_{err})^{N_t}$, where N_t means the length of the template. In the presence of terminators the probability of synthesis of a correct full-size product molecule is equal to

$$P_{corr} = (1 - P_{err})^{N_t} P_{amp}, \tag{21}$$

and the number of correct 'full-size' products after n cycles is [note the analogy with eq. (8)]

$$T_n^{corr} = T_0 (1 + P_{corr})^n. \tag{22}$$

The total amount of synthesized template (irrespective of its fidelity) is still given by eq. (8). From eqs. (8) and (22) we can derive the contributions of correctly copied and mutated template molecules to the whole pool of full-length DNA strands. It is

$$C_{corr} = \frac{T_n^{corr}}{T_n} = \left(\frac{1 + P_{corr}}{1 + P_{amp}} \right)^n \qquad (23)$$

and analogously

$$C_{err} = \frac{T_n - T_n^{corr}}{T_n} = \frac{(1 + P_{amp})^n - (1 + P_{corr})^n}{(1 + P_{amp})^n}. \qquad (24)$$

Discussion

In the previous section we have derived a theoretical description of the DEXAS method. Now we shall discuss the consequences for its practical use. We will point out both its advantages and disadvantages and give some hints for the optimization of the experimental procedure.

The key parameter of the DEXAS procedure is certainly the fraction R of ddNs, determining the efficiency of the production of sequencing fragments. Figure 1A shows that the amount of the fragments is quite sensitive to the value of R, especially in the vicinity of the optimum value. Therefore, the precise knowledge of the parameter a for a given enzyme is crucial for a good performance of this method. It follows from Figure 1A that it is better to overestimate the fraction of ddNs rather than that of dNs, $i.\,e.$ to decrease the probability P_{amp} below the optimum value rather than to increase it. The drop of the yield of sequencing fragments is much more rapid for $R > R_{opt}$ than for $R < R_{opt}$. Hence, when the precise value of a is not available, it is better to increase the ddN fraction. The value of R can be, however, optimized for a given enzyme.

It is also possible to perform the amplification/sequencing procedure with two enzymes with different affinities to ddNs as suggested by Kilger and Pääbo (1997b). Their idea was to use one enzyme strongly suppressing the incorporation of ddNs to produce mainly the full-length products, and another enzyme that does not suppress ddNs for the synthesis of sequencing fragments. The advantage of this technique is that the optimal experimental conditions are more easily accessible, especially in the case when the trial-and-error method is used.

Theoretically this means only a different formula for the probability of amplification given by eq. (16). The equation for its optimum value (14), as well as Figure 1A, are not influenced by this modification. The only advantage of this technique is that one has a larger flexibility in adjusting the reaction mixture to provide the optimal P_{ampd}, because it is possible to vary not only the ratio of dNs and ddNs, but also the ratio of the concentrations of the two enzymes. There is only a marginal problem in the evaluation of the optimum value of R, because it is not possible to solve the corresponding algebraic equation for this quantity. However, it is possible to solve this task graphically using the

plot of f_1 (or f_2) as a function of R with given parameters P_{ampd}, a_1, and a_2 (see Figure 2), or to numerically solving eq. (17) with the aid of a computer. Although it is in principle possible to set R to any value, $i.\,e.$ to reach any value of P_{amp}, even in the single-enzyme arrangement this double-enzyme strategy may be particularly advantagous in situations where the optimal concentrations of dNs and mainly ddNs are of an 'impractical' order of magnitude (too high or too low).

As can be seen, the character of the function in Figure 1A strongly depends on the number of PCR cycles. The maximum moves to higher values of P_{amp} and becomes sharper with the increasing number of cycles. This phenomenon can be explained as follows. When only few PCR cycles are performed, it is more advantageous to synthesize mainly the sequencing fragments (lower P_{amp}), since the full-length products cannot be exploited as template molecules in too many further cycles. However, when the number of cycles increases the benefit of higher level of template production becomes fundamental and hence the higher value of P_{amp} is needed. The explanation why the maximum is getting sharper with the increasing number of cycles follows straightforward from the analysis of eq. (11).

A remark should be made to the limit case when P_{amp} approaches zero. In this case, as can be derived from eq. (11), S_n tends to n (see Figure 1B). This result has a straightforward physical meaning. When P_{amp} equals zero, then no amplification takes place. The number of sequencing fragments is then equal to the initial number of template molecules multiplied by the number of PCR steps. The amplification of sequencing fragments runs thus according to an arithmetical sequence. The situation is similar to the commonly used PCR sequencing where one primer is used for the linear amplification of the sequenced template (Murray, 1989). The only difference is that we generate differently labeled sequencing fragments synthesized from both ends.

The distribution function of the fragment lengths is plotted in Figure 3 for several different values of P_t and a given N_t. It is quite obvious that this function must decrease with the length of the fragment. However, it is possible to arrange the experimental conditions so that the distribution is almost uniform. For a given N_t, the distribution uniformity grows when the value of P_t decreases, or, equivalently, P_{amp} increases. This is a favourable situation, since we need relatively large values of P_{amp} for the higher yield of sequencing fragments.

Finally, we derived expressions for the number of template molecules mutated by random infidelity of the DNA polymerase used in PCR. Although this phenomenon is quite important when the product is used for cloning, it is less relevant for the sequencing fidelity. Consider the fact that the ratio of 'correct' and mutated template is constant during the course of amplification, because both are amplified with an equal amplification quotient $(1 + P_{amp})$. Another important presumption is that we can see an ambiguous band in the electrophoretic separation only in the

case when both fragments are present in the mixture in sufficient amount. When the amount of fragments is not 'overamplified', we can see only those false bands whose fragments are present at least in the amount of *e. g.* one hundredth or one thousandth of the amount of the non-mutated fragments. Thus, when this visibility limit is *e. g.* 1/1000 of the number of fragments in a given ordinary band and the starting number of template molecules was higher than 1000, then no mutation ambiguities should occur in the sequence, because none of them will be present in a higher ratio to the correct template. We do not consider the ambiguities caused by low electrophoretical resolution. If we start with a lower number of template molecules, then we can observe the mutations in the template molecules that were created in the amplification cycles before 1000 template molecules were present. If we denote the number of molecules at the visibility limit T_{vis} and the number of cycles necessary to amplify the template from T_0 to T_{vis} n_v, then we can easily derive from eq. (8) that

$$n_v = \frac{ln\frac{T_{vis}}{T_0}}{ln(1+P_{amp})}. \tag{25}$$

With the aid of eq. (24) we can calculate the total number of mutations in the template at the visibility limit as

$$N_{mut} = T_{vis}C_{err}. \tag{26}$$

It is of some interest to understand how the number of mutations depends on the probability of amplification P_{amp}. For this purpose we can rewrite eq. (24) into the following form

$$C_{err} = 1 - \left(\frac{1+(1-P_{err})^{N_l}P_{amp}}{1+P_{amp}}\right)^{n_v} \tag{27}$$

After analyzing this function it can be seen that the number of mutations C_{err} decreases with increasing P_{amp}. This result can be interpreted in a way that more PCR cycles are needed for the same amplification, which means that in average more cycles precede a creation of a randomly chosen template molecule. Hence, there is a higher probability of mutations. Therefore, in this point DEXAS seems to be inferior to the classical method where the template DNA is amplified by the ordinary PCR and subsequently sequenced. However, in practical cases of DEXAS performance this phenomenon does not play a significant role.

As the last point we should comment on another experimental finding of Kilger and Pääbo (1997a). They state that the unbalanced ratio of the upstream and downstream primer increases the signal-to-noise ratio and therefore improve the sensitivity of the method. This effect, however, cannot be explained by our theory, because it is most probably caused by imperfect annealing of the primers or even some more complicated processes. They are not taken into account in our simple model which might be the cause of this disagreement.

Taken together, these considerations lead to a set of instructions how to design an optimal DEXAS procedure.

We suggest to start from the distribution function and choose a value of P_{amp}, in order to obtain the distribution as uniform as possible. Using this value of P_{amp}, the number of PCR cycles necessary for the sufficient level of amplification can be chosen. Determining these parameters is always a compromise, because the maximum value of S_n has to be reached. Having optimized n and P_{amp}, we can determine the optimum value of the parameter R, or, in the case of two polymerases, an optimal pair of values of R and f_1. For this purpose we have to know the experimental constant a of the polymerase (a_1 and a_2 in the latter case).

We explain the use of this procedure on the following example. Suppose that the template DNA is 500 base pairs long and the simpler arrangement with only one enzyme is used. With the aid of Figure 3 we find that $P_{amp} = 0.8$ is sufficient for enough homogeneous mixture of the sequencing fragments. Thus, this or a higher value can be used, depending on the required amplification. From eqs. (13, 14) and Figure 1 it can be seen that for $P_{amp} = 0.8$ the maximum number of sequencing fragments is produced at about 11 PCR cycles, *i. e.*, such a procedure provides very low amplification (642 times). Hence if the initial amount of template is so small that it requires much higher amplification, a higher value of P_{amp} has to be used together with a higer number of cycles. If, for example, 30 cycles, for which the maximum number of fragments is generated at $P_{amp} = 0.93$ and the corresponding amplification is 3.7×10^8, are sufficient, we use these values to determine the fraction of ddNs R. Suppose, for simplicity, that the polymerase does not discriminate between dNs and ddNs, *i. e.* the constant a is equal to 1/2 [see eq. (1) and the following text for the definition of R and a]. Thus, making use of eq. (15) we calculate the optimal value of R, in this case $R_{opt} = 1.15 \times 10^{-3}$. It is conceivable that with the parameters determined in this way the efficiency of the DEXAS procedure will be much higher then only with their intuitive estimates.

Let us summarize the contents of this study in several concluding remarks. We provide a detailed description of the method of direct exponential amplification and sequencing. Our aim was to derive mathematical formulas which help the experimentators to find the optimal reaction conditions in a rational and systematic way. Since we know that the experimental performance of DEXAS is not as satisfactory as desired, we hope that such a description may be useful. We also discussed the influence of random mutations of the template on the sequencing result and compared it to a classical method consisting of a PCR amplification and subsequent sequencing reaction. Our conclusion is that the difference is negligible in all practical cases. At the end we summarized our findings in a set of instructions for the design of an optimal experimental procedure.

Acknowledgements

References

Kilger, C., and Pääbo, S. (1997a). Direct exponential amplification and sequencing (DEXAS) of genomic DNA. Biol. Chem. *378*, 99–105.

Kilger, C., and Pääbo, S. (1997b). Direct DNA sequence determination from total genomic DNA. Nucleic Acids. Res. *25*, 2032–2034.

Krawczak, M., Reiss, J., Schmidtke, J., and Rösler, U. (1989). Polymerase chain reaction: replication errors and reliability of gene diagnosis. Nucleic Acids Res. *17*, 2197–2201.

Lieberwirth, U., Arden-Jacob, J., Drexhage, K.H., Herten, D.P., Müller, R., Schulz, A., Siebert, S., Sagner, G., Klingel, S.,Sauer, M., and Wolfrum, J. (1998). Multiplex dye DNA sequencing in capillary gel electrophoresis by diode laser-based time-resolved fluorescence detection. Anal. Chem. *70*, 4771–4779.

Murray, V. (1989). Improved double-stranded DNA sequencing using the linear polymerase chain reaction. Nucleic Acids Res. *17*, 8889.

Plaschke, J., Voss, H., Hahn, M., Ansorge, W., and Schackert, H.K. (1998). Doublex sequencing in molecular diagnosis of hereditary diseases. BioTechniques *24*, 838–841.

Ruano, G., and Kidd, K.K. (1991). Coupled amplification and sequencing of genomic DNA. Proc. Natl. Acad. Sci. USA, *88*, 2815–2819.

Smith, L.M., Fung, S., Hunkapiller, M.W., and Hood, L.E. (1985). The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis. Nucleic Acids Res. *13*, 2399–2412.

Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B.H., and Hood, L.E. (1986). Fluorescence detection in automated DNA sequence analysis. Nature *321*, 674–679.

Smith, L.M., Kaiser, R.J., Sanders, J.Z., and Hood, L.E. (1987). The synthesis and use of fluorescent oligonucleotides in DNA sequence analysis. Meth. Enzymol. *155*, 260–301.

Sun, F. (1995). The polymerase chain reaction and branching processes. J. Comput. Biol. *2*, 63–68.

Voss, H., Nentwich, U., Duthie, S., Wiemann, S., Beneš, V., Zimmermann, J., and Ansorge, W. (1997). Automated cycle sequencing with Taquenase™: protocols for internal labeling, dye primer and "doublex" simultaneous sequencing. BioTechniques *23*, 312–318.

Wiemann, S., Stegemann, J., Grothues, D., Bosch, A., Estivill, X., Schwanger, C., Zimmermann, J., Voss, H., and Ansorge, W. (1995). Simultaneous on-line DNA sequencing on both strands with fluorescent dyes. Anal. Biochem. *224*, 117–121.

Wiemann, S., Stegemann, J., Zimmermann, J., Voss, H., Beneš, V., and Ansorge, W. (1996). "Doublex" fluorescent DNA sequencing. Two independent sequences obtained simultaneously in one reaction with internal labeling and unlabeled primers. Anal. Biochem. *234*, 166–174.

Weiss, G., and von Haeseler, A. (1995). Modeling the polymerase chain reaction. J. Comput. Biol. *2*, 49–61.

Weiss, G., and von Haeseler, A. (1997). A coalescent approach to the polymerase chain reaction. Nucleic Acids Res. *25*, 3082–3087.

Yin, Z., Severin, J., Giddings, M.C., Huang, W., Westphall, M.S., and Smith, L.M. (1996). Automatic matrix determination in four dye fluorescence-based DNA sequencing. Electrophoresis *17*, 1143–1150.

P16. Polyvinyl alcohol biodegradation under denitrifying conditions.

Polyvinyl alcohol biodegradation under denitrifying conditions

CrossMark

Hana Marušincová [a], Lucie Husárová [a,b], Jan Růžička [a,b], Marek Ingr [c,d], Václav Navrátil [d,e], Leona Buňková [a], Marek Koutny [a,b,*]

[a] Tomas Bata University in Zlín, Faculty of Technology, Department of Environmental Protection Engineering, nám. T.G.Masaryka 5555, 760 01 Zlín, Czech Republic
[b] Tomas Bata University in Zlín, Centre of Polymer Systems, nám. T.G.Masaryka 5555, 760 01 Zlín, Czech Republic
[c] Tomas Bata University in Zlín, Faculty of Technology, Department of Physics and Material Engineering, nám. T.G.Masaryka 5555, 760 01 Zlín, Czech Republic
[d] Charles University in Prague, Department of Biochemistry, Hlavova 2030, 128 43 Prague 2, Czech Republic
[e] Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo Sqr. 2, 16610 Praha 6, Czech Republic

ABSTRACT

Polyvinyl alcohol was biodegraded under denitrifying conditions with a microbial community originated from a municipal wastewater treatment plant. The derived microbial consortium was capable of poly-vinyl alcohol degradation under both denitrifying and aerobic conditions. The community dynamics was monitored by temperature gradient gel electrophoresis, and a principal utilizing organism was identified and assigned as *Steroidobacter* sp. PD. The possible role of *Steroidobacter* sp. PD was also investigated by sequencing the 16S rDNA clone library prepared from the degrading community. qPCR analysis showed that the fraction of the microorganism in the community was very low initially (0.02%) and had reached to about 16% by the end of the biodegradation experiment. The study revealed that polyvinyl alcohol can be biodegraded in a water environment not only under aerobic but also under denitrifying conditions.
© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Polyvinyl alcohol (PVA) is a water soluble polymer produced on a mass scale. Its annual production exceeds 1 megaton and a steady rise in this figure (IHS Handbook, 2007) is discernible. PVA is used as a thickening, emulsifying or film-forming agent or as an adhesive in many household and industrial applications, especially in the paper, textile and chemical industries. Due to such mass production and utilization a considerable amount is expected to leak from the afore-mentioned processes into the environment, especially into waste-water.

Fortunately, PVA was found to be biodegradable despite its full carbon backbone. It has, however, also gradually become apparent that PVA-degradation capacity is not ubiquitous, and that it is rather sparsely distributed among some bacterial and fungal taxa. Most such bacterial degraders were classified among pseudomonads but especially sphingomonads (Kawai, 1999). Some other bacterial degraders comprise Gram-negative species like *Alcaligenes faecalis* (Matsumura et al., 1994) but also Gram-positive ones, e.g. *Bacillus megaterium* (Mori et al., 1996). Examples of some fungal degraders identified so far include *Penicillium* (Qian et al., 2004), and *Aspergilus* (Jecu et al., 2010; Stoica-Guzun et al., 2011). From the literature it is obvious that substantial biodegradation can mainly be expected in an aerobic aquatic environment, while PVA removal from soils is very limited and can be mostly a result of the non-specific action of lignolytic enzymes (Mejia et al., 1999; Chiellini et al., 2003). There are some indications describing partial anaerobic removal of rela-tively low molecular weight PVA (Matsumura et al., 1993; Gartiser et al., 1998) or partial anaerobic removal of PVA as a component of a polymer blend with other easily biodegradable constituents (Hrnčiřík et al., 2010). Yu et al. (Yu et al., 1996) have suggested that the addition of nitrates could enhance PVA removal in a sequential anaerobic–aerobic bioreactor. However, none of these reports identified the relevant microorganisms or judged at least whether a specific PVA degradation was observed.

Two principal enzyme systems have been described. The first one, which relies on secondary alcohol oxidase (Shimao et al., 1983) with oxygen as an electron acceptor, can in principle be functional under

* Corresponding author. Tomas Bata University in Zlín, Faculty of Technology, Department of Environmental Protection Engineering, nám. T.G.Masaryka 5555, 760 01 Zlín, Czech Republic. Tel.: +420 576031409.
E-mail address: mkoutny@ft.utb.cz (M. Koutny).

aerobic conditions only. The second, constitutes a PVA degradation pathway beginning with the action of PQQ containing periplasmic dehydrogenase (PVADH), from which electrons are probably conducted to soluble cytochrome c and further toward a terminal acceptor; oxygen, most likely under aerobic conditions (Shimao et al., 1996). The process should be feasible under anaerobic conditions in the presence of an alternative electron acceptor, e.g. nitrogen containing electron acceptors (Zumft, 1997; Reimann et al., 2007). Subsequently, β-hydroxyl ketone products of PVA dehydrogenase can be hydrolyzed in an aldolase like reaction catalyzed by a related hydrolase (Hirota-Mamoto et al., 2006; Kawai and Xiaoping, 2009), and shorter fragments of the polymer can then probably be assimilated.

The authors sought to investigate such a possibility in a series of experiments. The study reports on PVA specific biodegradation under denitrifying conditions and brings to light evidences of PVA consumption related to denitrification, in addition to identifying microorganisms relevant to the process.

2. Materials and methods

2.1. Sampling of the microbial community

Waste-water sludge from the denitrifying compartment of the Zlin-Malenovice municipal waste-water treatment plant (Czech Republic) was sampled in March 2010, and then on two more occasions at about two-month intervals. The sludge was immediately purged with nitrogen, filtered through a screen (4 mm^2) to remove macroscopic particles, washed twice with mineral medium of pH 7.2 (Muchova et al., 2009) by centrifugation (5000 g, 12 min), and purged with nitrogen again.

2.2. Biodegradation experiments

Incubations were conducted in glass bottles (total volume, 320 ml) containing 280 ml of the mineral medium (in g l^{-1}: KH$_2$PO$_4$ 0.09, Na$_2$HPO$_4$·12H$_2$O 0.96, NH$_4$Cl 1.0, MgSO$_4$·7H$_2$O 0.1, Fe(NH$_4$)$_2$(-SO$_4$)$_2$.6H$_2$O 0.03, CaCl$_2$ 0.01, KNO$_3$ 0.86, trace element solution (Muchova et al., 2009) 1.0 ml), 100 mg l^{-1} of PVA (POVAL 205, 87–89% of hydrolysis; Kuraray Co. Ltd., Japan), 500 mg l^{-1} of NO$_3^-$ (as KNO$_3$), 1 g l^{-1} dry weight of the preprocessed sludge from the first instance of sludge sampling. The cultures were purged with nitrogen and sealed with stoppers equipped with a rubber septa and incubated anaerobically on a magnetic stirrer (each bottle contained a magnetic bar; 250 rpm) at 25 °C. The experiments were carried out in three replicates. As an alternative, the PVA biodegradation experiment was also conducted under an aerobic condition in 500 ml flasks equipped with gas permeable stoppers filled with 100 ml of identical media under vigorous shaking (180 RPM).

The PVA concentrations were assayed in microplates (Joshi et al., 1979; Vaclavkova et al., 2007). After removing biomass by centrifugation (15,000 × g, 15 min), 20 μl of a sample, 42 μl of boric acid solution (40 g per liter) and 10 μl of I$_2$/KI solution (12.7 g I$_2$ and 40 g KI per liter) were pipetted into a well, and after 5 min an absorbance reading of 660 nm was made. Actual PVA concentration was deduced from a calibration curve.

Nitrates were assayed potentiometrically with an ion-selective electrode (Perfection NO$_3$ combination, Mettler Toledo) according to the manufacturer's instructions. KNO$_3$ solutions in above described mineral medium were used for calibration.

2.3. Isolation procedures

2.3.1. Preparation of enriched PVA degrading consortia
Fresh media of the described composition were inoculated with preprocessed sludge from the second instance of sludge sampling

and cultivated for 60 days under identical conditions. PVA consumption was verified and the culture was used to inoculate another fresh medium, and this cultivation was held for 30 days. Then the microorganisms were collected by centrifugation (5000 × g, 12 min) and used for inoculating fresh medium with increased concentrations of nutrients (PVA, 150 mg l^{-1}; NO$_3^-$, 1000 mg l^{-1}); the initial dry weight of the inoculum was set to 100 mg l^{-1}. After four weeks of cultivation the culture was diluted ten times with fresh medium of the latter composition and cultivated again for another four weeks. Following this, the culture was supplemented three times in three week intervals in a manner that the PVA concentrations were set to 200 mg l^{-1} and nitrate concentrations to 1000 mg l^{-1}; this culture was designated "B". In an identical way the sludge sample from the third instance of sampling was processed and the final culture designated "D". After every manipulation, cultures were always purged with nitrogen and sealed.

2.3.2. Isolation attempts
After the enrichment procedure, microorganisms from "Culture B" were collected by centrifugation, carefully resuspended, and a series of dilutions were prepared and transferred onto agar plates containing mineral agar with nitrate (BMA, control), mineral agar with nitrate and 500 mg l^{-1} PVA (BMA-PVA), and mineral agar with nitrate, succinate and ethanol (BMA-SE, 500 mg l^{-1} of each carbon source). Variants of the described solid media containing 20 μg l^{-1} of pyrroloquinolino quinone (PQQ) were also used. The plates were incubated at 25 °C anaerobically (anaerostat MERCK, Anaerocult A) or under aerobic condition.

2.4. DNA isolation, PCR and TGGE conditions

DNA from denitrifying cultures was extracted using a commercial DNA extraction kit (PowerSoil, MoBio) according to the manufacturer's instructions, involving an initial bead-beating step. The primers fD1 and rD1 (AGAGT TTGAT CCTGG CTCAG and AAGGA GGTGA TCCAG CC, respectively) were used to amplify nearly a full-length 16S rRNA gene (Weisburg et al., 1991). Each 25 μl PCR reaction contained 12.5 μl of GoTaq Green hot start master mix (Promega), 1 μl of each primer solution (12.5 pmol), 9.5 μl of water for molecular biology, and 1 μl (5–10 ng) of bacterial DNA. All amplifications were carried out in Piko Thermal Cycler (Finnzymes); the temperature profile was as follows: initial denaturation at 94 °C for 5 min; 30 cycles at 94 °C for 1 min, 55 °C for 1 min, and 72 °C for 1 min; and a final extension at 72 °C for 10 min. Subsequently, 1 μl of the first PCR product was used as a template for nested PCR amplifying the V3-hypervariable region of the 16S rRNA gene with the primer pair 341fGC and 518r (ATTAC CGCGG CTGCT GG and CCTAC GGGAG GCAGC AG, respectively), where the GC clamp (CGCCC GCCGC GCGCG GCGGG CGGGG CGGGG GCACG GGGGG) was covalently attached to the 5' end of the forward primer (Muyzer et al., 1993). The temperature program consisted of 1 min at 94 °C and 30 cycles of 1 min at 94 °C, 1 min at 55 °C, 1 min at 72 °C, and a final extension at 72 °C for 10 min. TGGE separation of the amplified PCR products was achieved on a TGGE Maxi system (Whatman-Biometra) in an isocratic denaturing gel (8% acrylamide; 20% deionized formamide, 1× TAE, 2% glycerol and 8 M urea). Two to five micro liters of PCR products (60–100 ng DNA) were loaded into each well. A 100-bp DNA ladder (NEB) was loaded as a marker (Das et al., 2007). Electrophoresis was performed at the constant voltage of 130 V for 18 h and the optimal gradient was found to be from 35 °C to 55 °C. Afterward, the electrophoresis gels were stained with GelStar (Cambrex), according to the manufacturer's instructions, and documented. The whole procedure was repeated twice with identical results.

2.5. Phylogenetic analysis

Phylogenetic analysis of relationships between sequences retrieved from TGGE bands and the most closely related GenBank sequences were conducted in MEGA 5.05 (Nei and Kumar, 2000; Tamura et al., 2011). The distance was computed using the Neighbor-joining p-distance method and bootstrap test (1000 replicates).

2.6. Clone library and sequencing

PCR fragments containing almost full-length 16S rRNA genes from the investigated samples were amplified with the help of fD1, rD1 primers equipped with cohesive adaptors (ATGGG ATCCA GAGTT TGATC CTGGC TCAG, CAGCT GCAGA AGGAG GTGAT CCAGC C) under the above-described PCR conditions, purified and subsequently ligated into pUC19 vector between the restriction sites EcoRI and HindIII with the aid of an In-Fusion HD Cloning Kit (ClonTech) following the manufacturer's instructions. Then CaCl₂-competent Escherichia coli DH5α cells were transformed with the ligation mixture, and transformants were selected on ampicilin containing LB (Sigma–Aldrich) agar plates. About 60 clones were selected randomly, and plasmid DNAs were isolated via a High-Speed Plasmid Mini Kit (GeneAid). The clones were checked for correctness of the ligation by restriction analysis using EcoRI and HindIII restriction endonucleases. The sequences of correct-length inserts were determined (JCU, Czech Republic).

2.7. Quantification of the PVA degrading strain by qPCR

A quantitative polymerase chain reaction (qPCR) was used in order to determine the relative amount of the PVA-degrading strain in the bacterial culture. The total number of bacteria in the culture was determined using the primers 341f and 518r described in (Muyzer et al., 1993), which anneal to the conserved sequences spanning the V3 hypervariable region of 16S rRNA gene and should, therefore, amplify the genomic sequence of any bacterial strain. Another primer pair was used to quantitate the PVA-degrading strain, which was designed to selectively amplify the 16S rRNA of the degrader with respect to other 16S rRNA sequences obtained by the cloning method described above. More specifically, these primers amplify a 220 bps portion of the V5 and V6 hypervariable regions of the gene (sequences: AGGGT CTGCC TCTCG GTG and CGCAT CTCTG CAGGA TTCC). All PCR reactions were carried out in triplicate in LightCycler 96-well plates sealed with adhesive foil (Roche) using a LightCycler 480 II instrument (Roche) in a total volume of 10 μl. Each reaction consisted of 5 μl of LightCycler 480 SYBR I Green Master (Roche), 1 μl of template DNA (isolated DNA from bacterial culture or plasmid standard) and 2 μM of final concentrations of both forward and reverse primers. In order to eliminate contamination by ambient bacteria, which might cause a false positive signal when using universal bacterial primers, stock solutions of primers, water and the 96-well plate were irradiated with UVC light (λ = 254 nm) for 30 min prior to mixing the reaction. Initial denaturation for 1 min at 95 °C was followed by 45 cycles of 10 s at 95 °C, 30 s at 56 °C, and 30 s at 72 °C. The threshold cycle numbers were then determined from fluorescence intensities acquired during the PCR runs by the 2nd derivative maximum method using LightCycler 480 software (Roche).

Serial tenfold dilution of the 16S rRNA gene of the degrader cloned in pUC19 plasmid was in parallel amplified with both primer pairs and served thus as standard for absolute quantification of both targets. Initial concentration of plasmid DNA prior to dilution was determined spectrophotometrically at 260 nm (Nanodrop ND-1000, Thermo Scientific). The concentration ranges of these standards covered such ranges actually measured in the samples. The relative amount of the degrader was then calculated as the ratio of the 16S rRNA gene copy number, as determined by amplification with specific-primers, to the 16S rRNA gene copy number determined by amplification with universal primers.

2.8. DNA sequences

Sequences from the clone library were deposited in a GenBank; the accession numbers running from JQ726646–JQ726696. Almost a full 16S rDNA sequence of Steroidobacter sp. PD possesses the GenBank accession number JQ726645.

3. Results

3.1. PVA biodegradation experiment

Minimal mineral medium with nitrate content was supplemented with PVA as the sole carbon substrate, and inoculated with sludge from the denitrifying compartment of the municipal wastewater treatment plant. These cultures were flushed with nitrogen, sealed and incubated anaerobically. At regular intervals samples were taken and the PVA concentration monitored. Initially, the polymer concentration remained stable, until about day 12 when the apparent lag-phase seemed to be over and a relatively fast decrease in PVA concentration commenced (Fig. 1). This experiment was repeated several times with the same sludge sample and the sludge samples drawn from the same location, but about two months later and then again four months later, with very similar results. If the same medium was inoculated with the microbial community sampled at the end of the previously described experiment, in other words with the PVA-acclimated microbial community, no lag-phase was observed and PVA was quickly consumed (Fig. 1). Nitrate concentration decreased along with PVA biodegradation and, on the contrary, PVA concentration remained almost unchanged if no nitrate was present in the medium (Fig. 1), so PVA degradation in the absence of oxygen was apparently dependent on the presence of nitrate. These observations could indicate that the original sludge contained a relatively low concentration of a microorganism or a microbial consortium able to
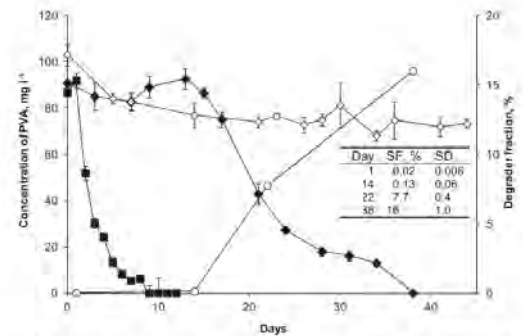
Fig. 1. PVA concentration in biodegradation experiments under various conditions. ♦, mineral medium inoculated with a PVA-non-adapted sludge microbial community with 500 mg L⁻¹ nitrate; ■, mineral medium inoculated with a PVA-adapted sludge microbial community with 500 mg L⁻¹ nitrate; ○, mineral medium inoculated with a PVA-non-adapted microbial community without nitrate; ⊙, Steroidobacter fraction in the bacterial community. The table inserted shows degrader fractions (SF) at respective points in time. Values are given as means ± 1 standard deviation (n = 3).

degrade and assimilate PVA under conditions supporting denitrification.

Another series of experiments was conducted with the enriched PVA degrading consortium "D" obtained by the procedure described. Again, under anaerobic conditions, PVA was consumed in the presence of nitrates whereas in their absence only a slight decrease in PVA concentrations at the beginning of the observation period was observed (Fig. 2A). Simultaneously, nitrate concentrations were monitored (Fig. 2B), showing that PVA degradation was accompanied by said nitrate consumption. About 150 mg of the nitrates per liter was consumed, which corresponds to about 60% of the theoretical amount necessary for the complete mineralization of the PVA inserted (about 240 mg of nitrates per 100 mg of PVA). Some degree of nitrate consumption in the incubation without PVA can be explained by the denitrification of carbon substances introduced with the sludge biomass (flocs) used as the inoculum.

An identical series of experiments was repeated under aerobic conditions. Here (Fig. 2C), both cultures, with and without nitrates, exhibited a rapid decrease in PVA concentrations, witnessing the ability of the consortia to degrade the polymer aerobically and suggesting that the key organisms of the consortium probably are facultative anaerobes. No nitrate consumption was detected in any variant of the media (i.e. with or without PVA, Fig. 2D), which is well consistent with the general idea of aerobic-anaerobic

metabolisms switching and the preference for oxygen in facultative anaerobic denitrifiers (Zumft, 1997).

3.2. Isolation attempts

The active microbial communities from biodegradation experiments were subjected to the described series of dilutions and reinoculations with the aim to enrich the key microorganisms and to dilute out the others. Dilutions of the enriched culture designated "B" were transferred onto the described selective agar media. After two months of cultivation no differences were observed between control BMA and BMA-PVA plates, distinct colonies were, however, found on BMA-SE plates, and finally 21 pure strains were isolated. Subsequently these strains were individually tested in liquid media for PVA biodegradation under denitrifying conditions, with and without PQQ addition, and for denitrification with succinate and ethanol. Despite the majority of the strains being capable of denitrification with the standard organic substrates, neither of them was able of PVA biodegradation under denitrifying conditions. Consortia composed from all isolated strains or their combinations did not prove active, either. Later the series of isolation attempts were repeated, in addition to which aerobic conditions were tested as well, but again without a positive result. After that, several further attempts for degrader's isolation were
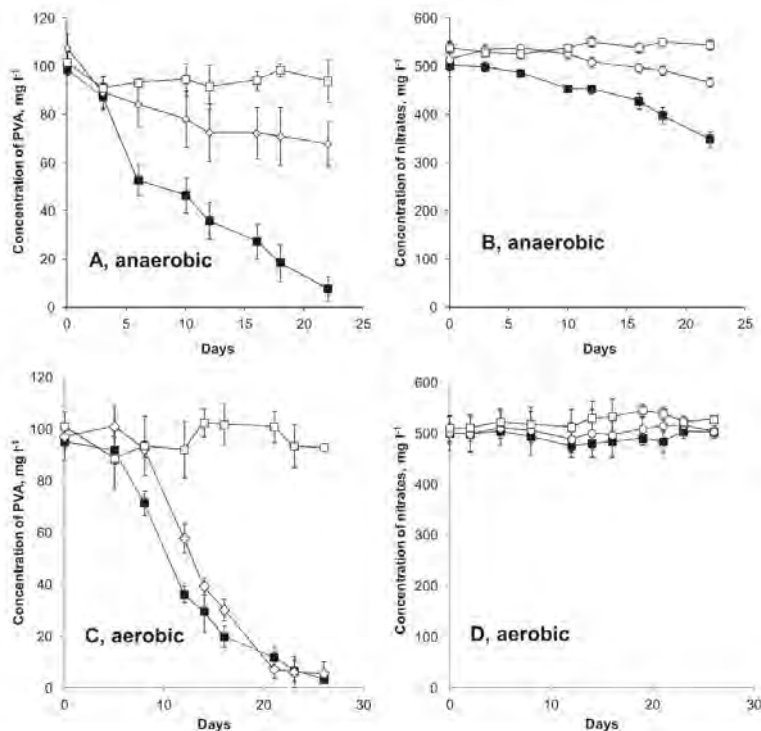
Fig. 2. Biodegradation experiment with the enriched microbial consortium "D" under anaerobic and aerobic conditions. Panel A (anaerobic) and panel C (aerobic), concentrations of PVA: ■, inoculated mineral medium with 500 mg L^{-1} of nitrates; ◇, inoculated mineral medium without nitrates; □, inoculated mineral medium with 500 mg L^{-1} of nitrates, after preparation the culture was autoclaved. Panel B (anaerobic) and panel D (aerobic), concentrations of nitrates: ■, inoculated mineral medium with 100 mg L^{-1} of PVA; ○, inoculated mineral medium without PVA; □, inoculated mineral medium with 100 mg L^{-1} of PVA; after preparation the culture was autoclaved. Values are given as means ± 1 standard deviation (n = 3).

made. Briefly, on the base of Fahrbach's description of Steroidobacter (Fahrbach et al., 2008), BMA plates amended with glutamate or acetate were inoculated with different dilutions of acclimated culture "B" and later the same was done with BMA plates with PVA amended with 5% of autoclaved acclimated culture "B" and BMA plates with partially oxidized PVA. Despite such effort and identification of many colonies found on above mentioned agars, none of them was identified as Steroidobacter or proved to be the degrader.

3.3. Investigation of PVA degradation by cultivation-independent techniques

Simultaneously, the PVA degradation experiment depicted in Fig. 1 was followed by temperature gradient gel electrophoresis (TGGE) based on amplification of the 16S rRNA gene V3 hypervariable region (Fig. 3). The microbial community in the experiment inoculated with the original sludge was sampled at time intervals, the DNA was isolated, and TGGE profiles compared with and examined against the PVA degradation curve. Time evolution of the

Fig. 3. TGGE profiles of the denitrifying PVA degrading microbial community. M, DNA marker; numbers on the top of the gel denote days during degradation or the control experiment (Fig. 1, ◆); B and D, microbial communities from enrichment cultures, where concentrations of the degraders were increased by the sequence of dilutions and transfers to fresh media. Sequences were successively obtained from the numbered bands and phylogenetic assignation made (Fig. 4).

TGGE profiles was also monitored in a blank experiment with an identical initial bacterial community but without PVA presence in the media.

It is quite obvious that a new band (Band 2) appeared between day 18 and day 22, which gradually increased in intensity. The same band was apparently missing in the control incubation. Comparing the time evolution of the investigated band with the curve in Fig. 1, one could see a clear concurrence of the onset of PVA consumption with the appearance of the band. Another interesting band (designated Band 1) exhibited a brief increase in intensity around day 22, which could also be related to the initiation of PVA consumption. The above-mentioned bands, along with other prominent bands from the last day of incubation, were excised and sequenced. The authors failed to obtain a quality sequence for some of them, which is often the case with TGGE. For the bands where sequences were successfully retrieved, their phylogenetic affiliations were sought out and are depicted in Fig. 4. Band 2 was assigned as a putative PVA utilizing organism closely related to the Steroidobacter strain found in the database.

The DNA from cultures used for isolation attempts, where enrichment for specific PVA degrading denitrifying microorganisms was induced, was processed in an identical way and the TGGE patterns visualized. Starting material for these incubations was sludge from the second and third instances of sampling at the same location. Both the obtained bacterial consortia "B" and "D" exhibited rapid consumption of PVA under denitrifying conditions, and both had highly intense bands at exactly the same position as the Band 2 previously identified. Later sequencing proved that these bands are identical with Band 2. It is worth mentioning that other significant bands found in patterns from the biodegradation experiment are apparently not present in patterns from enrichment cultures "B" and "D" and vice versa. These findings strongly supported the opinion that Band 2 represents a signal of a specific and dominant PVA utilizing organism in the described microbial community.

To obtain further independent information on the PVA degrading bacterial community, the DNA at day 38 of the same experiment was amplified using universal bacterial primers spanning almost the full length of the 16S rRNA gene, following which the fragments obtained were cloned and about 60 randomly selected correct clones were sequenced (Table 1). About 29% of the sequences from 51 correctly sequenced clones belonged to the bacterium previously identified from the TGGE band. Almost the complete length of its 16S rRNA gene was sequenced (GenBank JQ726645), the result revealing that the organism is closely related (97% identical) to the database strain Steroidobacter ZUMI 37 (GenBank AB548216.1); which, according to the note submitted with the sequence, was described as an aerobic PVA degrader isolated from soil.

Another dominant sequence found was closely related to the Dokdonella strain. Here assignment of the corresponding TGGE bands was not straightforward, as several relatively intense bands (Bands 4, 5, 9, and 10) could probably be related to this sequence. However, all these bands possessed constant intensity during the entire biodegradation experiment, and could also be found in the control incubation without PVA. Likewise, these bands are apparently missing or at least exhibit very low intensity in both enrichment cultures.

3.4. Monitoring the Steroidobacter fraction in the course of PVA biodegradation by qPCR

Absolute quantification of both Steroidobacter specific 16S rRNA and total 16S rRNA by qPCR was used to monitor the changes of the putative degrader fraction in the culture during the course of the

Fig. 4. Phylogenetic relationships of TGCE retrieved sequences and closely related sequences (GenBank). The distance was computed using the Neighbor-Joining p-distance method and bootstrap test (1000 replicates). Only bootstrap values over 50% are shown. PB, proteobacteria.

PVA biodegradation experiment. DNA isolated from the bacterial culture samples taken on days 1, 14, 22, and 38 of the cultivation was used as template for qPCR reactions, each sample with both *Steroidobacter*-specific and universal primer pairs. The 16S rRNA gene of the organism cloned in pUC19 plasmid was used as a calibration standard. Day 1 corresponded with the inoculation of the medium, while by day 38 almost all the PVA present in the medium had already been degraded. The fractions of *Steroidobacter* in the culture on the individual days of cultivation were calculated and the results were compared with the PVA degradation curve (Fig. 1).

Table 1
Frequency and affiliations of retrieved sequences.

Clones[a]	Closest related organism, accession number	Identity	Taxonomic group
16/51	*Dokdonella* sp. LM 2–5, FJ455531.1	98%	Gamma-PB, *Xanthomonadales*
15/51	*Steroidobacter* sp. ZUMI 37, AB518216.1	97%	Gamma-PB, *Xanthomonadales*
7/51	Bacterium Ellin6095 (*Burkholderia*), AY234747.1	91%	Beta-PB, *Burkholderia*
2/51	*Caldilinea aerophila* DSM 14535, AB067647.1	86%	Chloroflexi
2/51	*Phenylobacterium koreense* MKC7, EF173338.1	99%	Alpha-PB, *Caulobacter*
2/51	*Candidatus Microthrix parvicella* M2, FJ638889.1	99%	Actinobacteria
1/51	Alpha proteobacterium ZLJ-0, GU247516.1	99%	Alfa-PB
1/51	Chromatiaceae bacterium MTPP2IF163, FN293071.1	98%	Gamma-PB
1/51	Comamonadaceae bacterium PIV-20-1, AJ505862.1	99%	Beta-PB, *Burkholderia*
1/51	*Paucibacter toxinivorans* 2B5, AY515391.1	89%	Beta-PB, *Burkholderia*
1/51	*Clostridium* sp. SH-C52, FJ424481.1	98%	Firmicutes
1/51	*Longilinea arvoryzae*, AB243673.1	84%	Chloroflexi
1/51	Alpha proteobacterium WD28, HQ341736.1	97%	Alfa-PB

[a] Number of given sequences versus total number of obtained sequences; PB, *Proteobacteria*.

This experiment shows that the PVA utilizing strain is present in the original sludge in a negligible fraction with respect to the other bacteria ($\sim 0.02\%$). During cultivation its fraction grew continuously up to the value of approximately 16%, which is in general accord with the result obtained from sequencing individual 16S rRNA clones amplified from the same bacterial culture. Moreover, the trend of the data corresponds with the degree of degradation of PVA in the culture, further supporting the findings obtained from the biodegradation experiment.

4. Discussion

The biodegradation experiment proved that PVA degradation is possible under anoxic conditions using nitrate, as the terminal electron acceptor. These results indicated the presence of PVA degrading microorganisms that were probably initially present in a rather low concentration. To obtain further information, the microbiology of the process was investigated with both cultivation and cultivation-independent techniques. While attempts to isolate degrading strains on agar media, despite considerable effort, failed for the time being, TGGE revealed a distinct band which appeared to increase its intensity in correlation with decrease in PVA concentration. This band was sequenced, and later a corresponding almost full 16S rDNA sequence was obtained, showing high similarity with the 16S rDNA sequence of the previously identified aerobic PVA degrading strain *Steroidobacter* sp. ZUMI 37. The strain is also related to *Steroidobacter denitrificans* (GenBank EF605262.1), isolated as a denitrifying degrader of some steroid compounds (Fährbach et al., 2008). This supports the idea that PVA utilization and denitrification could be connected to one organism. On the other hand, the inability to isolate the pure degrading strain on agar media could suggest that the PVA degrader requires another microbial partner(s) or some environmental factor(s) either for its growth or PVA utilization.

Other bands present in TGGE profiles seemed constant during incubation but also constantly present in the control incubation without PVA, therefore, they should not be connected with PVA biodegradation. The only other exception was Band 1, briefly increasing its intensity at the end of the lag phase. The sequence from the band was retrieved and found to be related to sequences from *Brevundimonas* strains. Interestingly, *Brevundimonas vesicularis* (formerly described as *Pseudomonas vesicularis*) was earlier identified as an aerobic PVA degrader (Watanabe et al., 1976). Although the band demonstrates low intensity in later phases of PVA degradation, sequencing the 16S rRNA gene library made at the end of the degradation period provided about 4% sequences which were affiliated with *Phenylobacterium koreense* MKC7 (Table 1), a member of the same taxonomic group *Caulobacteraceae* like *Brevundimonas*. Thus, it is possible that the organism represented by Band 1 could play a role at the beginning of degradation, while it is later overgrown by a *Steroidobacter* related strain (denoted further as *Steroidobacter* sp. DP), representing at the end about 29% of the bacterial community (according to the clone library sequencing).

Another suggestion about the role of *Steroidobacter* sp. DP comes from the TGGE patterns of cultures "B" and "D", originating from the second and third instances of sludge sampling, respectively, exhibiting both strongly dominant bands of the same organism. Microscopic observation of the enrichment cultures revealed the prevalence of Gram negative rods, which was consistent with the expected *Steroidobacter* phenotype. The relatively high number of other minor bands in the enrichment culture might result from the presence of the specified amount of acetate substituents on the PVA backbone that can be easily hydrolyzed and utilized by bacteria.

The correlation of PVA consumption and *Steroidobacter* sp. DP concentration was further confirmed by the qPCR experiment. Fractions of *Steroidobacter* sp. DP were calculated for several days selected during cultivation, and a clear trend was observed showing that this species grows in accord with PVA consumption. Moreover, the determined fraction at day 38 of the degradation experiment, which was 16%, agreed reasonably with the fraction of *Steroidobacter* sp. DP sequences retrieved from the clone library (29%). This small difference might be attributed to slightly different primer affinities in both methods, varying ligation efficiency for different sequences, the small set of sequenced clones, or a combination of these. Thus, a close correlation is observed between monitoring PVA consumption and growth of the *Steroidobacter* sp. DP fraction determined by qPCR, as well as between the final fraction of *Steroidobacter* sp. determined by sequencing clones and qPCR. All this evidence supports the hypothesis that PVA can be efficiently degraded during denitrification.

5. Conclusion

In the described experiments polyvinyl alcohol was biodegraded under denitrifying conditions with a microbial community originated from a municipal waste-water treatment plant. *Steroidobacter* sp. DP was the principle PVA utilizing microorganism, and a number of indirect evidences suggest that the organism had an important role in PVA biodegradation under denitrifying conditions. Subsequent experiments also proved that the consortium dominated by *Steroidobacter* sp. DP is capable of aerobic PVA degradation. As stated in the introduction PVA is produced on a massive scale and in some cases could leak into waste-water or directly into the environment in considerable quantities. The findings presented here broaden our knowledge regarding the biodegradation of PVA and full carbon synthetic polymers in general. They also show that such processes can proceed in anaerobic environments as well, for example, in anaerobic compartments of waste-water treatment plants, but also in natural sediments, and raise the confidence in the relative environmental safety of PVA.

References

Chiellini, E., Corti, A., D'Antone, S., Solaro, R., 2003. Biodegradation of poly(vinyl alcohol) based materials. Progress in Polymer Science 28, 963–1014.
Das, M., Royer, T.V., Leff, L.G., 2007. Diversity of Fungi, Bacteria, and actinomycetes on leaves decomposing in a stream. Applied and Environmental Microbiology 73, 756–767.
Fährbach, M., Kuever, J., Remesch, M., Huber, B.E., Kämpfer, P., Dott, W., Hollender, J., 2008. *Steroidobacter denitrificans* gen. nov., sp. nov., a steroidal hormone-degrading gammaproteobacterium. International Journal of Systematic and Evolutionary Microbiology 58, 2215–2223.
Gartiser, S., Wallrabenstein, M., Stiene, G., 1998. Assessment of several test methods for the determination of the anaerobic biodegradability of polymers. Journal of Environmental Polymer Degradation 6, 159–173.
Hirota-Mamoto, R., Nagai, R., Tachibana, S., Yasuda, M., Tani, A., Kimbara, K., Kakai, F., 2006. Cloning and expression of the gene for periplasmic poly(vinyl

alcohol) dehydrogenase from *Sphingomonas* sp. strain 113P3, a novel-type quinohaemoprotein alcohol dehydrogenace. Mikrobiology 152, 1941–1949.

Hrnčiřík, J., Pšeja, J., Kupec, J., Bernkopfová, S., 2010. Anaerobic biodegradation of polyvinyl alcohol modified by extracellular polysaccharides. Journal of Polymers and Environment 18, 98–103.

IHS Chemical Economics Handbook, 2007. (Polyvinyl Alcohol). IHS, Englewood, USA.

Jecu, L., Gheorghe, A., Rosu, A., Raut, I., Grosu, E., Ghiurea, M., 2010. Ability of fungal strains to degrade PVA based materials. Journal of Polymers and Environment 18, 284–290.

Joshi, D.P., Lan-Chun-Fung, Y.L., Pritchard, J.G., 1979. Determination of poly(vinyl alcohol) via its complex with boric acid and iodine. Analytica Chimica Acta 104, 153–160.

Kawai, F., 1999. Sphingomonads involved in the biodegradation of xenobiotic polymers. Journal of Indian Microbiology and Biotechnology 23, 400–407.

Kawai, F., Xiaoping, H., 2009. Biochemistry of microbial polyvinyl alcohol degradation. Applied Microbiology and Biotechnology 84, 227–237.

Matsumura, S., Kurita, H., Shimokobe, H., 1993. Anaerobic biodegradability of polyvinyl alcohol. Biotechnology Letters 15, 749–754.

Matsumura, S., Shimura, Y., Terayama, K., Kiyohara, T., 1994. Effects of molecular weight and stereoregularity on biodegradation of poly (vinyl alcohol) by *Alcaligenes faecalis*. Biotechnology Letters 16, 1205–1210.

Mejia, A.I., Lucy Lopez, B.L., Mulet, A., 1999. Biodegradation of poly (vinyl alcohol) with enzymatic extracts of *Phaenerochaete chrososporium*. Macromolecular Symposia 148, 131–147.

Mori, T., Sakimoto, M., Kagi, T., Sakai, T., 1996. Isolation and characterization of a strain of *Bacillus megaterium* that degrades poly(vinyl alcohol). Bioscience, Bitechnology and Biochemistry 60, 330–332.

Muchova, M., Ruzicka, J., Julinova, M., Dolezalova, M., Houser, J., Koutny, M., Bunkova, L., 2009. Xanthan and gellan degradation by bacteria of activated sludge. Water Science and Technology 60, 965–973.

Muyzer, G., De Waal, E.C., Uitterlinden, A.G., 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Applied and Environmental Microbiology 59, 695–700.

Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Qian, D., Du, G., Chen, J., 2004. Isolation and culture characterization of a new polyvinyl alcohol-degrading strain: *Penicillium* sp. WSH02–21. World Journal of Microbiology Biotechnology 20, 587–591.

Reimann, J., Flock, U., Lepp, H., Honigmann, A., Adelroth, P., 2007. A pathway for protons in nitric oxide reductase from *Paracoccus denitrificans*. Biochimica Biophysica Acta 1767, 362–373.

Shimao, M., Tsuda, T., Takahashi, M., Kato, N., Sakazawa, C., 1983. Purification of membrane-bound polyvinyl alcohol oxidase in *Pseudomonas* sp. VM15C. FEMS Microbiology Letters 20, 429–433.

Shimao, M., Tamogami, T., Nishi, K., Harayama, S., 1996. Cloning and characterization of the gene encoding pyrroloquinoline quinonedependent poly(vinyl alcohol) dehydrogenase of *Pseudomonas* sp. strain VM15C. Bioscience, Biotechnology and Biochemistry 60, 1056–1062.

Stoica-Guzun, A., Jecu, L., Raut, A.G.I., Stroescu, M., Danila, M.G.M., Jipa, I., Fruth, V., 2011. Biodegradation of Poly(vinyl alcohol) and bacterial cellulose composites by *Aspergillus niger*. Journal of Polymers and Environment 19, 69–79.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28, 2731–2739.

Vaclavkova, T., Ruzicka, J., Julinova, M., Vicha, R., Koutny, M., 2007. Novel aspects of symbiotic (polyvinyl alcohol) biodegradation. Applied Microbiology and Biotechnology 76, 911–917.

Watanabe, Y., Hamada, N., Morita, M., Tsujisaka, Y., 1976. Purification and properties of a polyvinyl alcohol-degrading enzyme produced by a strain of *Pseudomonas*. Archives of Biochemistry and Biophysics 174, 575–581.

Weisburg, W., Barns, S.M., Pelletier, D.A., Lane, D.J., 1991. 16S ribosomal DNA amplification for phylogenetic study. Journal of Bacteriology 173, 697–703.

Yu, H., Gu, G., Song, L., 1996. Degradation of polyvinyl alcohol in sequencing batch reactors. Environmental Technology 17, 1261–1267.

Zumft, W.G., 1997. Cell biology and molecular basis of denitrification. Microbiology and Molecular Biology Reviews 61, 533–616.

228

P17. Preparation of a biologically active apo-cytochrome b5 via heterologous expression in *Escherichia coli*.

Preparation of a biologically active apo-cytochrome b₅ *via* heterologous expression in *Escherichia coli*

Věra Kotrbová, Dagmar Aimová, Marek Ingr, Lucie Bořek-Dohalská, Václav Martínek, Marie Stiborová [*]

Department of Biochemistry, Faculty of Science, Charles University, Albertov 2030, 128 43 Prague 2, Czech Republic

ARTICLE INFO

ABSTRACT

Cytochrome b₅ (b₅) has been shown to modulate many cytochrome P450 (CYP)-dependent reactions. In order to elucidate the mechanism of such modulations, it is necessary to evaluate not only the effect of native b₅ on CYP-catalyzed reactions, but also that of the apo-cytochrome b₅ (apo-b₅). Therefore, the apo-b₅ protein was prepared using a heterologous expression in *Escherichia coli*. The gene for rabbit b₅ was constructed from synthetic oligonucleotides using polymerase chain reaction (PCR), cloned into pUC19 plasmid and amplified in DH5α cells. The gene sequence was verified by DNA sequencing. The sequence coding b₅ was cleaved from pUC19 by NdeI and XhoI restriction endonucleases and subcloned to the expression vector pET22b. This vector was used to transform *E. coli* BL-21 (DE3) Gold cells by heat shock. Expression of b₅ was induced with isopropyl β-D-1-thiogalactopyranoside (IPTG). The b₅ protein, produced predominantly in its apo-form, was purified from isolated membranes of *E. coli* cells by chromatography on a column of DEAE–Sepharose. Using such procedures, the homogenous preparation of apo-b₅ protein was obtained. Oxidized and reduced forms of the apo-b₅ reconstituted with heme exhibit the same absorbance spectra as native b₅. The prepared recombinant apo-b₅ reconstituted with heme can be reduced by NADPH:CYP reductase. The reconstituted apo-b₅ is also fully biologically active, exhibiting the comparable stimulation effect on the CYP3A4 enzymatic activity towards oxidation of 1-phenylazo-2-hydroxynaphthalene (Sudan I) as native rabbit and human b₅.

Cytochrome b₅ (b₅) is a heme protein with molecular mass of about 17,000 [1,2], which is capable of accepting and transferring a single electron [3]. It appeared very early in evolution and is found in a wide range of phyla. Its primary structure is highly conserved. Different mammalian species show over 80% identity in the sequence of the cytosolic heme-binding domain, but reveal higher heterogeneity in the carboxy terminal domain [4,5].

In mammals, three b₅ isoforms were found in different cell compartments such as the endoplasmic reticulum, mitochondria and cytoplasm of matured erythrocytes [6]. The endoplasmic reticulum b₅ is an integral membrane protein located on the outer surface of this cell compartment, whereas its mitochondrial counterpart is located on the outer mitochondrial membrane [7]. The erythrocyte isoform is a soluble protein thought to originate from post-translational proteolysis of the endoplasmic reticulum form [8]. Interestingly, the endoplasmic reticulum b₅ is considered to be synthesized on free ribosomes and to be inserted post-translationally into the membrane of the endoplasmic reticulum, without participation of a signal recognition particle. The last 10 amino acid residues of b₅ carry information necessary for the cytochrome to be targeted to the endoplasmic reticulum membrane [9–12].

Microsomal b₅ is involved in fatty acid desaturation [6,13,14], cholesterol [15] and plasmalogen biosyntheses [6,16] as well as in various hydroxylation reactions catalyzed by mixed function oxidase system [17]. It can accept an electron from either NADH:cytochrome b₅ reductase [18,19] or NADPH:cytochrome P450 (CYP) reductase [20,21] and then reduced b₅ transfers this electron to CYPs and other enzymes (Fig. 1).

Microsomal b₅ consists of two domains, the larger soluble N-terminal heme-binding core and the smaller hydrophobic C-terminal tail, which anchors the protein to the membrane. A 15-amino acid flexible linker connects these two domains, providing a heme domain with a sufficient mobility to bind different redox partners, whereas the protein remains in the membrane. It was postulated that the linker consisting of at least 7 amino acids is necessary for productive interaction with enzymes of the mixed function oxidase system [22].

The structure of full-length protein has not been resolved yet, because the hydrophobic domain is an obstacle for crystallization and X-ray diffraction. Most of the structural studies were therefore made with truncated protein, where the membrane anchor is cleaved off. Nevertheless, the membrane anchor was observed to be essential for some biological activities of b₅ [23]. Namely, the soluble form of b₅ is able to interact with soluble partners such as cytochrome c [6,22], but does not bind to CYP and does not affect the CYP-mediated reactions at all [24,25].

[*] Corresponding author. Fax: +420 221951283.
E-mail address: stiborov@natur.cuni.cz (M. Stiborová).
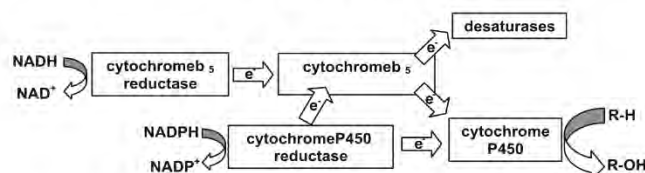
Fig. 1. Schematic demonstration of electron transfer between CYP, b_5 and their reductases (adapted from reference [32]).

The structure of soluble domain of rabbit b_5 is illustrated in Fig. 2A. The holoprotein structure contains two hydrophobic cores. Heme binds in core 1 and is coordinated by two invariant residues His 39 and His 63. These amino acid residues prevent iron from direct interactions with molecular oxygen [5]. Whereas the core 1 has a clear functional role, the highly conserved core 2 appears to play only a structural role [26–28]. The b_5 protein lacking heme consists of a well-folded hydrophobic core 2 and 42-residue loop, which is substantially disordered in the solution. Association with the heme cofactor causes the loop to organize into a well-folded core 1 [26]. The interaction of b_5 with CYPs and reductases is driven by electrostatic interactions between groups of charged amino acid residues located on the protein surface in the vicinity of heme [29,30].

Two hypothetical conformations of the anchor are shown in Fig. 2B [4,12]. In the first one, the carboxyl terminus of the hydrophobic membrane-binding domain of b_5 spans the lipid bilayer of the endoplasmic reticulum [12,25]. In the second conformation, the membrane anchor forms a hairpin loop structure with a carboxy terminal residue exposed to the cytoplasm [4,12,31]. Even though the exact topology of the membrane-binding domain is still not quite clear, some authors prefer the first hypothetical conformation [12,32].

The role of microsomal b_5 in catalytic function of CYPs has not been fully understood yet. Cytochrome b_5 has been shown to be able to stimulate, inhibit or have no effect on CYP mediated reactions [1,5,33–38]. Its effect is strongly dependent on individual CYP isoforms, their substrates and experimental conditions [5]. Several hypotheses trying to explain the influence of b_5 on CYP reactions have been proposed. One of them suggests a role of b_5 in a direct transfer of the second electron to the CYP enzyme, which is considered to be the rate limiting step in the catalytic cycle of the CYP monooxygenase reaction [39]. The electron transfer from reduced b_5 to CYP is faster than the input of electron from NADPH:CYP reductase [17,40]. Due to this stimulation effect, CYPs are prevented from uncoupling of partially activated oxygen. The

release of superoxide and hydrogen peroxide appears to various extents in all monooxygenase reactions, but also depends both on a CYP isoform and a substrate. Moreover, the addition of b_5 enhances the stability of the oxy–CYP complex, therefore, higher amount of activated oxygen is available to form hydroxylated product [5].

Another possible mechanism of the b_5 action is the formation of a complex between b_5 and CYP, which can receive two electrons from NADPH:CYP reductase in a single step, one for reduction of CYP and another for that of b_5 [5]. While CYP without b_5 has to undergo two separate interactions with NADPH:CYP reductase to complete one catalytic cycle, in the case of the presence of b_5, only one single interaction of complex of CYP and b_5 with NADPH:CYP reductase is sufficient; b_5 provides the second electron to CYP promptly after oxygen binding. Interaction of b_5 with CYP may also induce conformational changes in CYP proteins leading to breakdown of oxygenated hemoprotein complex with substrates to products. This hypothesis is based on findings showing that not only holoprotein of b_5, but also its apo-form (devoid of heme), which is not capable of electron transfer, can contribute to stimulation effects [38,41–43].

It is clear from such investigations that studies utilizing apo-cytochrome b_5 (apo-b_5) are necessary to explain the mechanisms of b_5 effects on CYP-catalyzed reactions. Preparation of this apoprotein in sufficient amounts and quality is hence crucial for further studies in this field. Several different approaches to prepare apo-b_5 were utilized. One of them, efficiently removing a heme part of this cytochrome, was the extraction of heme by acid acetone treatment [38,44,45]. Nevertheless, during this procedure, a part of such apo-b_5 protein preparations seemed to become denatured; irreversible precipitation of the apo-b_5 protein was found [45]. Therefore, other procedures, which gently remove the heme cofactor to prepare the pure apo-b_5 without altering the native protein conformation, should be developed. Here, we used a novel procedure, utilizing a heterologous expression of apo-b_5 in *Escherichia coli*. In contrast to preparation of holoprotein of b_5 by a het-
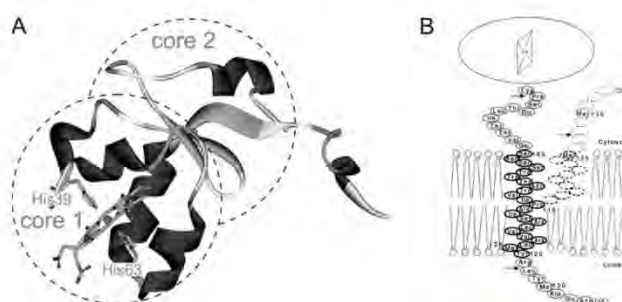
Fig. 2. (A) 3D structure of soluble domain of oxidized rabbit microsomal b_5 (based on PDB coordinates 1DO9 [50]). Two major regions (core 1 and core 2) are indicated. (B) Model for the possible topology of b_5 in membrane (adapted from [12]).

erologous expression [46], the precursor of heme biosynthesis, the δ-aminolevulinic acid, was not added to bacterial culture. Here, we demonstrate that such a procedure leads to expression of about 99% of b_5 protein in its apo-form, and that this apo-b_5 is, after reconstitution with heme, fully active to stimulate the CYP3A4-catalyzed reactions.

Materials and methods

Materials

Oligonucleotides were synthesised and dNTPs purchased from East Port (Czech Republic), *Pfu* polymerase was from Fermentas (Canada). *E. coli* BL-21 (DE3) Gold cells were obtained from Stratagene (USA). Restriction endonucleases (HindIII, EcoRI, NdeI, XhoI), plasmid pUC19 and T4 DNA ligase were purchased from New England Biolabs (USA). Plasmid pET-22b was from Novagen (USA); Luria–Bertani (LB) medium, LB agar, isopropyl β-D-1-thiogalactopyranoside (IPTG), L-α-dilauryl-sn-glycero-3-phosphocholine (DLPC), NADPH, Brij 35, glucose 6-phosphate dehydrogenase and sodium cholate were obtained from Sigma–Aldrich (USA); sodium dodecyl sulfate (SDS) was from Fluka (Switzerland); ampicillin (Amp)[1] was from Duchefa Biochemie (Netherlands). Commercial kits for purification of DNA from agarose gels and plasmid purification (Jet quick Gel Extraction Spin Kit, Jet quick Plasmid Mini and Maxiprep Spin Kit) were purchased from Genomed (Germany). Complete mini protease inhibitors tablets were obtained from Roche Diagnostic (Germany). DEAE–Sepharose CL6B was from Amersham Biosciences (Sweden). Rabbit liver NADPH:CYP reductase was purified as described earlier [47]. 1-Phenylazo-2-hydroxynaphthalene (Sudan I, Solvent Yellow 14) (>99% based on HPLC) was from British Drug Houses (UK). The derivatives 1-(4-hydroxyphenylazo)-2-naphthol (4'-OH-Sudan I), 1-(phenylazo)-naphthalene-2,6-diol (6-OH-Sudan I) and 1-(4-hydroxyphenylazo)-naphthalene-2,6-diol [4',6-di(OH)-SudanI] were synthesized as described [35]. Supersomes®, microsomes isolated from insect cells transfected with Baculovirus constructs containing cDNA of CYP3A4 with or without cytochrome b_5 and expressing NADPH:CYP reductase were from Gentest corp. (Woburn, MA, USA).

Design and construction of the complete gene of rabbit cytochrome b_5

De novo synthesis of the gene from commercially prepared oligonucleotides was carried out by the procedure similar to that described by Beck von Bodman et al. [48]. The gene coding for the whole amphipathic rabbit b_5 was optimized for *E. coli* codon usage and restriction endonucleases sites in DNA 2.0 program (available at http://www.dna20.com). Each oligonucleotide (Fig. 3) consists of approximately 80 base pairs with at least 18 nucleotides long overlapping area whose melting point is higher than 50 °C. In the first step, all oligonucleotides except for the first and last one were linked together by polymerase chain reaction (PCR) using the thermostable *Pfu* DNA polymerase. Amplification was performed after 10 min of denaturation at 94 °C in 30 cycles at 94 °C (30 s), 50 °C (30 s) and 73 °C (2 min) ending with an 8 min extension at 73 °C. The sequence of b_5 was completed by another PCR run carried out at the same conditions, where the remaining primers (first and last) were used while 1 μl of the previous PCR product was used as a template. The resulting PCR product was cleaved by EcoRI

and HindIII endonucleases and ligated to pUC19 plasmid treated with the same enzymes. DH5α cells were transformed by the resulting construct and plated on LB-agar containing 100 μg/ml Amp as a selective marker. Several colonies were picked and cultivated in 5 ml of LB medium with the same selective marker. The plasmid DNA was purified by a Jet quick Plasmid Miniprep Spin Kit (Genomed) and its correctness verified by DNA sequencing.

Protein expression and membrane isolation

The plasmid pET-22b with T7-promoter was used as an expression vector. The b_5 gene was obtained from pUC19 plasmid by digestion with NdeI/XhoI and then ligated to pET-22b vector digested with the same endonucleases (NdeI/XhoI). This expression vector was used to transform *E. coli* BL-21 (DE3) Gold cells. Transformed cells were plated on LB agar plates with 100 μg/ml Amp and grown overnight at 37 °C. The next day, 500 ml of LB medium with 100 μg/ml Amp was inoculated by the colonies wiped out of this plate. Culture was grown in an orbital shaker at 37 °C and 200 rpm. When the optical density of culture at 600 nm reached a value of approximately 1, IPTG was added to a final concentration of 0.05 mM in order to induce the protein expression. The cells were cultivated for next 4 h and then harvested by centrifugation at 3000g for 20 min and stored at −20 °C. After thawing, the pellet was resuspended in 25 ml of cold phosphate buffer A (10 mM KH_2PO_4, pH 7.7, 1 mM EDTA, supplemented with two tablets of complete mini protease inhibitor (Roche Diagnostic, Germany). Cells were disrupted by sonication using a 6 mm diameter probe in eight sonication cycles with 2.5 min duration at 40% of power, using 100 W maximal setting. Cell suspension was kept on ice during sonication and after each cycle left to cool down at 4 °C for 15 min. Insoluble proteins and cells debris was removed by centrifugation at 3000g for 15 min at 4 °C. Supernatant containing the membrane fraction was subsequently centrifuged at 105,000g for 70 min at 4 °C. The sediment was homogenized by brief 30 s sonication pulses in 25 ml of buffer B (10 mM KH_2PO_4, pH 7.7, 1 mM EDTA and 20% glycerol) and left at 4 °C till a protein concentration was determined. Protein content was determined using bicinchoninic acid (BCA) assay with bovine serum albumin as a standard [36]. The sample was then diluted with buffer B to yield protein concentration of 4 mg/ml. To solubilize b_5 from membranes, detergents Brij 35 and sodium cholate were added in amounts of 1 mg/mg of proteins and the mixture was stirred for 3 h at 4 °C. Solubilized membranes were centrifuged at 105,000g for 70 min at 4 °C.

Protein purification

All procedures were performed at 4 °C. The supernatant, obtained by centrifugation of solubilized membranes (see above), was applied on a DEAE–Sepharose CL6B column (2.5 × 20 cm) equilibrated with buffer C (20 mM KH_2PO_4, pH 7.7, 1 mM EDTA, 20% glycerol and 0.6% Brij) using a flow rate of 1 ml/min and washed with approximately 200 ml of buffer C until the absorbance of the eluent at 280 nm decreased to zero. The b_5 protein was eluted using linear gradient of KCl (0–400 mM) in buffer C. The elution profile was monitored spectrophotometrically at 280 nm. The b_5 content and purity of collected fractions were assessed by sodium dodecyl sulfate (SDS)–polyacrylamide gel electrophoresis (PAGE) [35], using 15% separating gel and stained with Coomassie brilliant blue. Pooled fractions were concentrated by ultrafiltration using PM 10 membrane (Millipore), dialyzed overnight against 2000 ml of buffer C and loaded onto a second DEAE–Sepharose CL6B column equilibrated with buffer D (20 mM KH_2PO_4, pH 7.7, 1 mM EDTA, 20% glycerol and 0.1% sodium cholate). The DEAE–Sepharose column was washed with 3-fold volume of buffer D and b_5 was eluted isocratically with buffer E

---

[1] *Abbreviations used*: Amp, ampicillin; b_5, cytochrome b_5; BCA, bicinchoninic acid; CYP, cytochrome P450; dNTPs, deoxynucleotidetriphosphates; DLPC, L-α-dilauryl-sn-glycero-3-phosphocholine; EDTA, ethylenediaminetetraacetic acid; IPTG, isopropyl β-D-1-thiogalactopyranoside; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; SDS, sodium dodecyl sulfate; 4'-OH-Sudan I, 1-(4-hydroxyphenylazo)-2-naphthol; 6-OH-Sudan I, 1-(phenylazo)-naphthalene-2,6-diol; 4',6-di(OH)-Sudan I, 1-(4-hydroxyphenylazo)-naphthalene-2,6-diol.
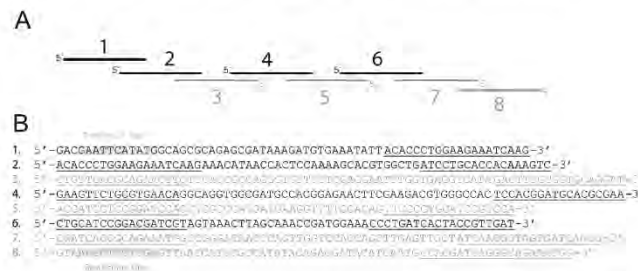
A

B

1.　5'-GACGAATTCATATGGCAGCGCAGAGCGATAAAGATGTGAAATATTACACCCTGGAAGAAATCAAG-3'
2.　5'-ACACCCTGGAAGAAATCAAGAAACATAACCACTCCAAAAGCACGTGGCTGATCCTGCACCACAAAGTC-3'
3.　5'-CTGTGGTGCAGGATCAGCCACGTGCTTTTGGAGTGGTTATGTTTCTTGATTTCTTCCAGGGTGT-3'
4.　5'-GAAGTTCTGGGTGAACAGGCAGGTGCGGATGCCACGGAGAACTTCAAGACGTGGGCCACTCCACGGATGCACGCGAA-3'
5.　5'-ACGATTCTTCGGCATCCGCACCTGCCTGTTCACCCAGAACTTCTTCACAGTTTCACGCGTGCATCCGT-3'
6.　5'-CTGCATCCGGACGATCGTAGTAAACTTAGCAAACCGATGGAAACCCTGATCACTACCGTTGAT-3'
7.　5'-GAATCAGCCAGAAACTGCCCGGAGAAGTCCCCGGTGGTAGTGATCAGGGTTTCCATCGGTTTGCTAAGTTTAC-3'
8.　5'-CTCGAGTTAAGCAAATCAGGCTCAGTTCATGTACAAGAAGTTCTGGGTGAACAGGCAGGTGCGGATGC-3'

Fig. 3. (A) Overview of rabbit microsomal b_5 gene synthesis strategy; (B) Proposal of oligonucleotides used for b_5 gene synthesis. At the 5'-end of the first oligonucleotide there are restriction sites for endonucleases EcoRI (GAATTC) and NdeI (CATATG), which contains an initializing codon ATG for protein translation. Similarly at the 5'-end of the last oligonucleotide there are restriction sites for endonucleases HindIII (AAGCTT) and XhoI (CTCGAG) followed by a stop codon TAA. Underlined areas highlight overlapping segments.

(200 mM KH_2PO_4, 200 mM KCl, pH 7.7, 1 mM EDTA, 20% glycerol and 0.1% sodium cholate). Using this chromatography step, Brij 35 detergent was removed from the b_5 preparation. An ionic detergent, sodium cholate, was additionally removed from the b_5 preparation by dialysis against 2000 ml of buffer F (50 mM KH_2PO_4, pH 7.7, 20% glycerol). After dialysis, the b_5 preparation was concentrated using an Amicon centricon apparatus and stored at −80 °C.

Incorporation of heme into apo-cytochrome b_5

The preparation of hemin chloride solution and its incorporation into apo-b_5 were performed by the procedure described elsewhere [46]. Briefly, solutions of heme were prepared by adding 2.6 mg of hemin chloride to a solution of 50% ethanol in water (4 ml) to give a final concentration of 1 mM. A small increment (10 μl) of 1 M NaOH was added and mixed to dissolve the hemin. The solution was then allowed to stand for several minutes so particulates could settle. A 10-μl aliquot was removed and diluted into 990 μl of 20 mM Tris, 1 mM EDTA, pH 8.0, at 25 °C and the absorbance of the Tris-liganded heme was measured at 385 nm. The process was repeated five times, always by adding 10 μl of 1 M NaOH, until NaOH addition caused no further increase in absorbance at 385 nm. The hemin solution was further filtered through 0.2 μm filter. Purified apo-b_5 was diluted with 20 mM Tris, pH 8.0, containing 1 mM EDTA and 0.4% sodium cholate, to yield the protein concentration of 0.25 mg/ml. Aliquots of hemin chloride were added into the apo-b_5 sample (1 ml) and the reconstitution of apo-b_5 with heme was monitored by absorbance spectroscopy. Absorbance spectra (from 350 to 500 nm) were recorded on Hewlett Packard 8453 UV spectrophotometer. The reconstitution was considered to be complete when the Soret peak of b_5 shifted from 413 to 409 nm and the increase in absorbance at 385, caused by excess of free Tris-liganded hemin, was observed in the spectrum.

Determination of reconstituted cytochrome b_5 content

The concentration of apo-b_5 reconstituted with heme was determined spectrophotometrically (the absolute absorbance spectrum) using molar extinction coefficient $\varepsilon_{413} = 117$ mM^{-1} cm^{-1} [2,49] or from the difference spectrum of reduced minus oxidized form, using molar extinction coefficient $\varepsilon_{424-409} = 185$ mM^{-1} cm^{-1}, respectively [49]. The reconstituted b_5 was reduced by the addition of approximately 1 mg of solid sodium dithionite.

Reduction of reconstituted cytochrome b_5 with NADPH:CYP reductase

The reduction of apo-b_5 reconstituted with heme using NADPH:CYP reductase was carried out in liposomes as described

[20]. NADPH:CYP reductase and b_5 were incorporated into liposomes prepared from DLPC as described previously [35]. Briefly, 20 μl of a solution of DLPC in chloroform (5 mg/ml) was evaporated to dryness under nitrogen, 50 μl of 50 mM potassium phosphate buffer (pH 7.7) was added and the mixture sonicated twice for 3 min. The mixture was diluted with the same buffer, and incubated after adding 3 μM apo-b_5 and/or 3 μM heme and 0.1 μM NADPH:CYP reductase (total volume of the mixture was 495 μl) for 10 min at room temperature and shaking at 350 rpm. The reduction reaction was initiated by adding 5 μl of 50 mM NADPH and the difference spectrum of reduced minus oxidized forms of protein was monitored on Hewlett Packard 8453 UV spectrophotometer (from 400 to 500 nm).

The effect of cytochrome b_5, apo-cytochrome b_5 and apo-cytochrome b_5 reconstituted with heme on oxidation of Sudan I by human CYP3A4

Incubation mixtures used to investigate the effects of apo-b_5 and its reconstituted form contained the following in a final volume of 500 μl: 100 mM potassium phosphate buffer (pH 7.4), 1 mM NADPH, 0.1 μM human recombinant CYP3A4 (expressed in Supersomes™ without or with b_5). In individual experiments, native rabbit liver b_5, recombinant apo-b_5, prepared in this work and its form reconstituted with heme were present in final concentrations of 0.5 μM (CYP:b_5 ratio was 1:5). After 5 min pre-incubation at 37 °C, the reaction was started by adding 5 μl of 5 mM Sudan I dissolved in methanol. After incubation (37 °C, 20 min), the mixtures were extracted twice with ethyl acetate (2 × 1 ml). Oxidation of Sudan I with CYP3A4 was linear until 30 min [35,37]. The extracts were evaporated, dissolved in methanol and the products were analyzed by HPLC on a MN Nucleosil 100-5 C18 column (Macherey-Nagel, 4.0 × 250 mm). An isocratic flow of methanol: 0.1 M NH_4HCO_3 (pH 8.5) (9:1, v/v) with flow rate of 0.7 ml/min, was used to elute the metabolites, detection was at 480 nm. The Sudan I metabolites were quantified via external calibration of synthetic standards of Sudan I metabolites as described in [35,37].

Results and discussion

Expression of apo-cytochrome b_5 in E. coli

The gene of full-length membrane rabbit microsomal b_5 was prepared from eight synthetic oligodeoxyribonucleotides, which are in length about 80 base pairs. They were designed as shown in Fig. 3. This total synthetic approach allows us to optimize the codon usage for E. coli and introduce various unique restriction sites allowing easy mutagenesis of a gene. The prepared gene was inserted into pUC19 plasmid after treating both the gene and the

Table 1
Purification of membrane bound apo-b_5.

Purification step	Volume (ml)	Total protein (mg)	Protein yield (%)
E. coli homogenate[a]	25	281	100
Membrane fraction	27	153	54
Supernatant after membrane solubilization	67	96	34
Pooled fractions after chromatography on DEAE–Sepharose	25	69	25
Final apo-b_5 preparation	3	28	10

[a] Prepared from 5.4 g of E. coli cells obtained from 0.5 l of IPTG induced culture medium.

plasmid with EcoRI and HindIII restriction endonucleases. E. coli DH5α cells were transformed by this construct. The amplified plasmid (from a single cell colony resistant to Amp) was verified by complete DNA sequencing. The expression vector based on pET-22b plasmid was prepared and used to transform BL-21 (DE3) Gold cells. Since the aim of this study was to prepare apo-b_5, no δ-aminolevulinic acid, a precursor of biosynthesis of heme was present in cell culture. The growth conditions were as described (see Materials and methods). The protein production was induced with 0.05 mM IPTG. The cells accumulated a large amount of recombinant protein, but did not show a red color in the culture. Such a color is an indicator of the presence of heme. Therefore, a negligible level of heme was present in the recombinant b_5 produced by E. coli under conditions used for its expression. This assumption was confirmed by measuring the absorbance spectra of purified protein (see below).

Purification of apo-cytochrome b_5

Cells were disrupted by sonication, and apo-b_5 was found in the cell membrane fraction. Therefore, this fraction was isolated from cell homogenate by ultracentrifugation, and further utilized for b_5 purification (Table 1). The homogenous apo-b_5 protein was prepared by a purification procedure consisting of solubilization of proteins from a membrane with detergents, and chromatography on a DEAE–Sepharose column (Table 1 and Fig. 4). Even though solubilization of apo-b_5 from membranes was not complete (the part of apo-b_5 was retained in a sediment after centrifugation), it led to a significant increase in purity of this apo-b_5 protein; only a small amount of other "ballast" proteins were solubilized together with apo-b_5 (Fig. 4). A negligible part of the solubilized sample did not

bind to DEAE–Sepharose, probably due to a high content of detergent, which can form micelles under the conditions used (4 °C). Apo-b_5 was eluted with 100 mM KCl, while other "ballast" proteins retained on a DEAE–Sepharose column.

Absorbance spectra shown in Fig. 5 confirmed that b_5 isolated by this procedure is predominantly in its apo-form (without heme). Only about 1% of a heme b_5 protein was present in this apo-b_5 preparation (Fig. 5). During the titration of the prepared apo-b_5 with heme, this cofactor was readily incorporated into the apo-form of the protein to produce a holoprotein (reconstituted recombinant b_5). The reconstitution of apo-b_5 with heme was accompanied by color change from brown color of hemin to red bright color of the holoprotein. During formation of b_5 by this procedure, an increase in absorbance at 413 nm originating from a typical Soret peak was found, which indicates that apo-b_5 reconstituted with heme. The heme-reconstituted b_5 protein exhibited the same physico-chemical properties such as the electrophoretic mobility on SDS–PAGE (data not shown) and absorbance spectra of its oxidized and reduced forms as native b_5 purified from rabbit livers. The Soret band at 413 nm of the oxidized form and maxima at 424, 526 and 556 nm of the dithionite-reduced form were found (Fig. 6). In liposome vesicles that simulate the environment of an endoplasmic reticulum membrane, purified recombinant b_5 reconstituted with heme can be reduced with NADPH:CYP reductase (see an increase in absorbance at 424 nm and a decrease in that at 409 nm in the spectrum shown in Fig. 7B). The same spectra were found also for native b_5 purified from rabbit livers reduced with NADPH:CYP reductase (Fig. 7A), but not for hemin (Fig. 7C). These findings suggest that recombinant apo-b_5 was prepared in its native structure, without any damages in its protein conformation.

The effect of cytochrome b_5, apo-cytochrome b_5 and apo-cytochrome b_5 reconstituted with heme on oxidation of Sudan I by human CYP3A4

To confirm the functional integrity of the prepared recombinant apo-b_5, we investigated whether this protein, both in its apo- and holoprotein forms, is capable of modulating the CYP3A4-mediated reactions and compared it with the isolated native b_5. The CYP3A4 enzyme was utilized in this study, because oxidation of many substrates of this enzyme is stimulated by b_5 [35,36,38,42,43]. This is also true for CYP3A4-catalyzed oxidation of a carcinogenic azo dye, 1-phenylazo-2-hydroxynaphthalene (Sudan I) [35]. Supersomes™,

Fig. 4. SDS–PAGE of apo-b_5 samples during apo-b_5 purification (15% separating gel). Lane 1, molecular weight standard (Sigma wide range); lane 2, homogenate from E. coli cells expressing apo-b_5; lane 3, supernatant of the E. coli cells after sonication; lane 4, membrane fraction; lane 5, supernatant after solubilization; lane 6, purified apo-b_5 after DEAE–Sepharose CL6B chromatography, lane 7, the final apo-b_5 preparation; lane 8, native rabbit hepatic b_5.

Fig. 5. Absorbance spectra of purified recombinant apo-b_5 (*solid line*) and the same protein reconstituted with heme (*dotted line*).
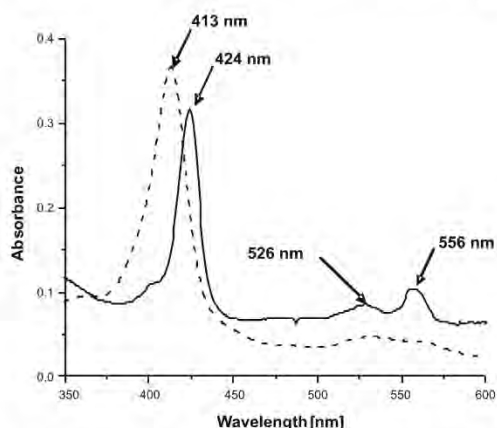
Fig. 6. Absorbance spectra of purified recombinant apo-b$_5$ reconstituted with heme. Oxidized form is represented by *dashed line* and sodium dithionite-reduced form by *solid line*.

Fig. 8. The effect of various b$_5$ proteins on the CYP3A4-mediated oxidation of Sudan I. 50 pmol human recombinant CYP3A4 expressed in Supersomes™ (without or with human recombinant b$_5$), 0 or 250 pmol of native rabbit b$_5$, recombinant rabbit apo-b$_5$, heme-reconstituted recombinant rabbit b$_5$ or recombinant human b$_5$ and 25 nmol Sudan I were incubated at 37 °C for 20 min. Grey bars correspond to amounts of individual hydroxy-derivatives of Sudan I produced during 20 min at 37 °C. Values are averages of two independent determinations.

microsomes isolated from insect cells transfected with Baculovirus constructs containing cDNA of human CYP3A4 with or without human b$_5$ and expressing human NADPH:CYP reductase, were used in these experiments. As shown in Fig. 8, the CYP3A4-mediated oxidation of Sudan I was stimulated by native rabbit hepatic b$_5$; levels of three major Sudan I metabolites, 4'-OH-Sudan I, 6-OH-Sudan I and 4',6-di(OH)-Sudan I, formed in the presence of this b$_5$ were 4-fold higher than without this protein. Similar efficiency to stimulate this reaction was also mediated by recombinant b$_5$ reconstituted with heme (3.5-fold), while apo-b$_5$ was less effective; only about a 1.8-fold increase in Sudan I oxidation was found (Fig. 8). In addition, human CYP3A4 in Supersomes™, expressing besides NADPH:CYP reductase also human b$_5$, exhibited similar efficiency to stimulate oxidation of Sudan I (3.7-fold) as systems containing native rabbit b$_5$ or recombinant apo-b$_5$ reconstituted with heme (Fig. 8). All these data demonstrate that prepared recombinant apo-b$_5$ reconstituted with heme functions analogously to native rabbit and human b$_5$ proteins and is therefore suitable for further studies that we plan to investigate the mechanisms of biological roles of b$_5$. These results also indicate that at least in the case of Sudan I as a substrate of CYP3A4, the presence of heme in b$_5$ seems not to be implicitly required for the stimulation of CYP3A4-medi-

ated oxidation of this compound. Even though stimulation effect of apo-b$_5$ was more than 2-fold lower than that of native b$_5$, apo-b$_5$ still increased Sudan I oxidation by CYP3A4. The heme lacking apo-b$_5$ is unable to directly donate the electrons to CYP3A4; therefore, this result could be considered to be consistent with the allosteric modulation theory [38,41–43]. Nevertheless, the question whether this finding (higher stimulation effect of b$_5$ *versus* apo-b$_5$) is valid also for other CYP3A4 substrates or for other CYP enzymes and their substrates awaits further investigations.

Conclusion

The results of this work describe for the first time a novel procedure of heterologous expression of apo-b$_5$, its purification and reconstitution with heme into the holoprotein form, which is fully active to modulate the CYP3A4-catalyzed reactions. Studies with other CYP enzymes and their substrates that are in progress in our laboratory will help us to shad more light on mechanisms of b$_5$ biological functions.
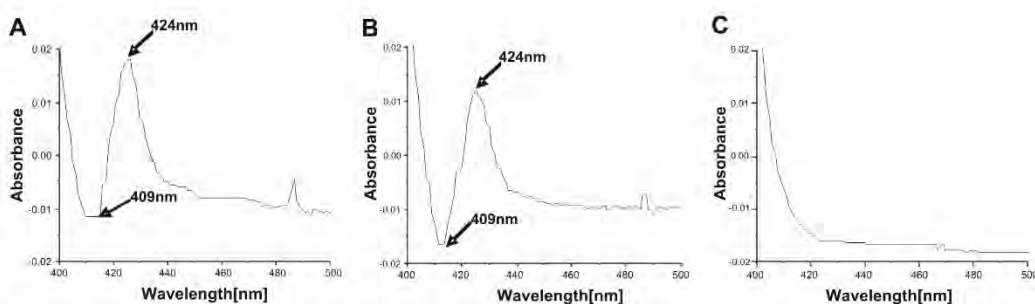
Fig. 7. Difference spectra (reduced minus oxidized) of (A) purified recombinant apo-b$_5$ reconstituted with heme, (B) purified rabbit b$_5$ and (C) hemin solution. Samples in A–C were reduced with rabbit NADPH:CYP reductase.

Acknowledgments

References

[1] T.D. Porter, The roles of cytochrome b_5 in cytochrome P450 reactions, J. Biochem. Mol. Toxicol. 16 (2002) 311–316.

[2] P. Strittmatter, S.F. Velick, The isolation and properties of microsomal cytochrome, J. Biol. Chem. 221 (1956) 253–264.

[3] S.F. Velick, P. Strittmatter, The oxidation–reduction stoichiometry and potential of microsomal cytochrome, J. Biol. Chem. 221 (1956) 265–275.

[4] M.R. Hanlon, R.R. Begum, R.J. Newbold, D. Whitford, B.A. Wallace, In vitro membrane-inserted conformation of the cytochrome b_5 tail, Biochem. J. 352 (2000) 117–124.

[5] J.B. Schenkman, I. Jansson, The many roles of cytochrome b_5, Pharmacol. Ther. 97 (2003) 139–152.

[6] G. Vergeres, L. Waskell, Cytochrome b_5, its functions, structure and membrane topology, Biochimie 77 (1995) 604–620.

[7] P. Sobrado, M.A. Goren, D. James, C.K. Amundson, B.G. Fox, A Protein Structure Initiative approach to expression, purification, and in situ delivery of human cytochrome b_5 to membrane vesicles, Protein. Expr. Purif. 58 (2008) 229–241.

[8] S.R. Slaughter, C.H. Williams Jr., D.E. Hultquist, Demonstration that bovine erythrocyte cytochrome b_5 is the hydrophilic segment of liver microsomal cytochrome b_5, Biochim. Biophys. Acta 705 (1982) 228–237.

[9] J. Mitoma, A. Ito, The carboxy-terminal 10 amino acid residues of cytochrome b_5 are necessary for its targeting to the endoplasmic reticulum, EMBO J. 11 (1992) 4197–4203.

[10] Y. Okada, A.B. Frey, T.M. Guenthner, F. Oesch, D.D. Sabatini, G. Kreibich, Studies on the biosynthesis of microsomal membrane proteins. Site of synthesis and mode of insertion of cytochrome b_5, cytochrome b_5 reductase, cytochrome P-450 reductase and epoxide hydrolase, Eur. J. Biochem. 122 (1982) 393–402.

[11] R.A. Rachubinski, D.P. Verma, J.J. Bergeron, Synthesis of rat liver microsomal cytochrome b_5 by free ribosomes, J. Cell Biol. 84 (1980) 705–716.

[12] G. Vergeres, J. Ramsden, L. Waskell, The carboxyl terminus of the membrane-binding domain of cytochrome b_5 spans the bilayer of the endoplasmic reticulum, J. Biol. Chem. 270 (1995) 3414–3422.

[13] T.C. Lee, R.C. Baker, N. Stephens, F. Snyder, Evidence for participation of cytochrome b_5 in microsomal delta-6 desaturation of fatty acids, Biochim. Biophys. Acta 489 (1977) 25–31.

[14] P. Strittmatter, L. Spatz, D. Corcoran, M.J. Rogers, B. Setlow, R. Redline, Purification and properties of rat liver microsomal stearoyl coenzyme A desaturase, Proc. Natl. Acad. Sci. USA 71 (1974) 4565–4569.

[15] G.F. Grinstead, J.L. Gaylor, Total enzymic synthesis of cholesterol from 4,4,14 alpha-trimethyl-5-alpha-cholesta-8,24-dien-3-beta-ol. Solubilization, resolution, and reconstitution of delta 7-sterol 5-desaturase, J. Biol. Chem. 257 (1982) 13937–13944.

[16] F. Paltauf, R. Prough, B. Masters, J. Johnson, Evidence for participation of cytochrome b_5 in plasmalogen biosynthesis, J. Biol. Chem. 249 (1974) 2661–2662.

[17] A. Hildebrandt, R.W. Estabrook, Evidence for the participation of cytochrome b_5 in hepatic microsomal mixed-function oxidation reactions, Arch. Biochem. Biophys. 143 (1971) 66–79.

[18] T. Iyanagi, Redox properties of microsomal reduced nicotinamide adenine dinucleotide–cytochrome b_5 reductase and cytochrome b_5, Biochemistry 16 (1977) 2725–2730.

[19] P. Strittmatter, The reaction sequence in electron transfer in the reduced nicotinamide adenine dinucleotide–cytochrome b_5 reductase system, J. Biol. Chem. 240 (1965) 4481–4487.

[20] H.G. Enoch, P. Strittmatter, Cytochrome b_5 reduction by NADPH-cytochrome P-450 reductase, J. Biol. Chem. 254 (1979) 8976–8981.

[21] F.P. Guengerich, Reduction of cytochrome b_5 by NADPH–cytochrome P450 reductase, Arch. Biochem. Biophys. 440 (2005) 204–211.

[22] T.A. Clarke, S.C. Im, A. Bidwai, L. Waskell, The role of the length and sequence of the linker domain of cytochrome b_5 in stimulating cytochrome P450 2B4 catalysis, J. Biol. Chem. 279 (2004) 36809–36818.

[23] S.B. Mulrooney, D.R. Meinhardt, L. Waskell, The [alpha]-helical membrane spanning domain of cytochrome b_5 interacts with cytochrome P450 via nonspecific interactions, Biochim. Biophys. Acta 1674 (2004) 319–326.

[24] J.Y. Chiang, Interaction of purified microsomal cytochrome P-450 with cytochrome b_5, Arch. Biochem. Biophys. 211 (1981) 662–673.

[25] G. Vergeres, L. Waskell, Expression of cytochrome b_5 in yeast and characterization of mutants of the membrane-anchoring domain, J. Biol. Chem. 267 (1992) 12583–12591.

[26] R.B. Davis Jr., J.T. Lecomte, Structural propensities in the heme binding region of apocytochrome b_5. I. Free peptides, Biopolymers 90 (2008) 544–555.

[27] C.J. Falzone, M.R. Mayer, E.L. Whiteman, C.D. Moore, J.T. Lecomte, Design challenges for hemoproteins: the solution structure of apocytochrome b_5, Biochemistry 35 (1996) 6519–6526.

[28] K. Mukhopadhyay, J.T. Lecomte, A relationship between heme binding and protein stability in cytochrome b_5, Biochemistry 43 (2004) 12227–12236.

[29] P. Strittmatter, C.S. Hackett, G. Korza, J. Ozols, Characterization of the covalent cross-links of the active sites of amidinated cytochrome b_5 and NADH:cytochrome b_5 reductase, J. Biol. Chem. 265 (1990) 21709–21713.

[30] P.P. Tamburini, R.E. White, J.B. Schenkman, Chemical characterization of protein–protein interactions between cytochrome P-450 and cytochrome b_5, J. Biol. Chem. 260 (1985) 4007–4015.

[31] J. Ozols, Structure of cytochrome b_5 and its topology in the microsomal membrane, Biochim. Biophys. Acta 997 (1989) 121–130.

[32] U.H. Durr, L. Waskell, A. Ramamoorthy, The cytochromes P450 and b_5 and their reductases-promising targets for structural studies by advanced solid-state NMR spectroscopy, Biochim. Biophys. Acta 1768 (2007) 3235–3259.

[33] E. Canova-Davis, L. Waskell, The identification of the heat-stable microsomal protein required for methoxyflurane metabolism as cytochrome b_5, J. Biol. Chem. 259 (1984) 2541–2546.

[34] E.T. Morgan, M.J. Coon, Effects of cytochrome b_5 on cytochrome P-450-catalyzed reactions. Studies with manganese-substituted cytochrome b_5, Drug Metab. Dispos. 12 (1984) 358–364.

[35] M. Stiborová, V. Martínek, H. Rýdlová, P. Hodek, E. Frei, Sudan I is a potential carcinogen for humans: evidence for its metabolic activation and detoxication by human recombinant cytochrome P450 1A1 and liver microsomes, Cancer Res. 62 (2002) 5678–5684.

[36] M. Stiborová, J. Sejbal, L. Bořek-Dohalská, D. Aimová, J. Poljaková, K. Forsterová, M. Rupertová, J. Wiesner, J. Hudeček, M. Wiessler, E. Frei, The anticancer drug ellipticine forms covalent DNA adducts, mediated by human cytochromes P450, through metabolism to 13-hydroxyellipticine and ellipticine N^2-oxide, Cancer Res. 64 (2004) 8374–8380.

[37] Y. Okada, A.B. Frey, T.M. Guenthner, F. Oesch, D.D. Sabatini, G. Kreibich, Studies on the biosynthesis of microsomal membrane proteins. Site of synthesis and mode of insertion of cytochrome b_5, cytochrome b_5 reductase, cytochrome P-450 reductase and epoxide hydrolase, Eur. J. Biochem. 122 (1982) 393–402.

[38] H. Yamazaki, W.W. Johnson, Y.F. Ueng, T. Shimada, F.P. Guengerich, Lack of electron transfer from cytochrome b_5 in stimulation of catalytic activities of cytochrome P450 3A4. Characterization of a reconstituted cytochrome P450 3A4/NADPH-cytochrome P450 reductase system and studies with apo-cytochrome b_5, J. Biol. Chem. 271 (1996) 27438–27444.

[39] F.P. Guengerich, D.P. Ballou, M.J. Coon, Spectral intermediates in the reaction of oxygen with purified liver microsomal cytochrome P-450, Biochem. Biophys. Res. Commun. 70 (1976) 951–956.

[40] J.B. Schenkman, I. Jansson, Interactions between cytochrome P450 and cytochrome b_5, Drug Metab. Rev. 31 (1999) 351–364.

[41] R.J. Auchus, T.C. Lee, W.L. Miller, Cytochrome b_5 augments the 17,20-lyase activity of human P450c17 without direct electron transfer, J. Biol. Chem. 273 (1998) 3158–3165.

[42] H. Yamazaki, M. Nakamura, T. Komatsu, K. Ohyama, N. Hatanaka, S. Asahi, N. Shimada, F.P. Guengerich, T. Shimada, M. Nakajima, T. Yokoi, Roles of NADPH-P450 reductase and apo- and holo-cytochrome b5 on xenobiotic oxidations catalyzed by 12 recombinant human cytochrome P450s expressed in membranes of *Escherichia coli*, Protein Expr. Purif. 24 (2002) 329–337.

[43] H. Yamazaki, T. Shimada, M.V. Martin, F.P. Guengerich, Stimulation of cytochrome P450 reactions by apo-cytochrome b_5: evidence against transfer of heme from cytochrome P450 3A4 to apo-cytochrome b_5 or heme oxygenase, J. Biol. Chem. 276 (2001) 30885–30891.

[44] D.L. Cinti, J. Ozols, Binding of homogeneous cytochrome b_5 to rat liver microsomes. Effect on N-demethylation reactions, Biochim. Biophys. Acta 410 (1975) 32–44.

[45] B. Mrázová, M. Martínková, V. Martínek, E. Frei, M. Stiborová, Optimization of preparation of apo-cytochrome b_5 utilizing apo-myoglobin, Interdisc. Toxicol. 1 (2008) 190–192.

[46] S.B. Mulrooney, L. Waskell, High-level expression in *Escherichia coli* and purification of the membrane-bound form of cytochrome b_5, Protein Expr. Purif. 19 (2000) 173–178.

[47] JD. Dignam, HW. Strobel, NADPH–cytochrome P-450 reductase from rat liver: purification by affinity chromatography and characterization, Biochemistry 16 (1977) 1116–1123.

[48] S. Beck von Bodman, M.A. Schuler, D.R. Jollie, S.G. Sligar, Synthesis, bacterial expression, and mutagenesis of the gene coding for mammalian cytochrome b_5, Proc. Natl. Acad. Sci. USA 83 (1986) 9443–9447.

[49] R.W. Estabrook, J. Werringloer, The measurement of difference spectra: application to the cytochromes of microsomes, Methods Enzymol. 52 (1978) 212–220.

[50] L. Banci, I. Bertini, A. Rosato, S. Scacchieri, Solution structure of oxidized microsomal rabbit cytochrome b_5. Factors determining the heterogeneous binding of the heme, Eur. J. Biochem. 267 (2000) 755–766.

P18. Quantitation of the Inhibition Effect of Model Compounds Representing Plant Biomass Degradation Products on Methane Production.

Quantitation of the Inhibition Effect of Model Compounds Representing Plant Biomass Degradation Products on Methane Production

Silvie Pekařová,[a] Marie Dvořáčková,[a] Petr Stloukal,[a,b] Marek Ingr,[c] Jana Šerá,[a] and Marek Koutny [a,*]

During the steam explosion pretreatment of plant biomass, degradation products are generated, and some of these have inhibitory activity against biogas production. The aim of this study was to investigate and quantify the effect of selected model inhibitory compounds on methane production. The results showed no significant inhibition of methane production by furfural at concentrations below 1 g/L. In addition, the microbial community was able to restore biogas production inhibited by this compound after a certain time. 5-hydroxymethylfurfural was evaluated as a more potent inhibitor, with a significant effect above 0.2 g/L. Both compounds were more effective inhibitors with cellulose as the carbon substrate, probably reflecting higher sensitivity of the cellulolytic step in biogas production. No significant inhibition was observed for the phenolic compounds tested, gallic and tannic acids, at concentrations of up to 2 g/L. Thus, the compounds investigated should not represent a problem for the biogas production involving steam explosion preprocessed plant biomass.

Keywords: Biogas; Furfural; 5-Hydroxymethylfurfural; Gallic acid; Tannic acid

*Contact information: a: Department of Environmental Protection Engineering, Faculty of Technology, Tomas Bata University in Zlín, nám. T. G. Masaryka 5555, 76001 Zlín, Czech Republic; b: Centre of Polymer Systems, University Institute, Tomas Bata University in Zlín, Trida Tomase Bati 5678, 760 01 Zlín, Czech Republic; c: Department of Physics and Materials Engineering, Faculty of Technology, Tomas Bata University in Zlín, nám. T. G. Masaryka 5555, 76001 Zlín, Czech Republic;
* Corresponding author: mkoutny@ft.utb.cz*

INTRODUCTION

Biogas production by anaerobic fermentation is an economically viable alternative to fossil fuel resources. The principal feed or substrate for industrial biogas units is plant biomass, most often wheat or corn silage. Alternative materials include manure and waste from food or wood processing industries, albeit usually in combination with already mentioned types of plant biomass, which, in any case, represents at least about 50% of the feed available for processing in biogas plants. The main reasons for the dominance of plant biomass for biogas production include high unit yields of biogas, verified technologies for the plant biomass production, good conservation and processing, and, of course, favorable cost and general availability.

Plant biomass is predominantly composed of cellulose, hemicellulose, and lignin. These natural polymers interact closely together and create a complex matrix that is rather resistant to biotic and abiotic degradation, especially under anaerobic conditions (Adney *et al.* 1991; Prochazka *et al.* 2012). Cellulose constitutes a major component of plant biomass, consisting of β(1→4) glycosidic bonds connected D-glucose units. Cellulose chains are

assembled into fibrils that possess a semi-crystalline structure, and these are interconnected with hemicelluloses and lignin at the higher level of structural hierarchy (Hendriks and Zeeman 2009). Hemicelluloses are complex polysaccharides that usually comprise pentoses, but hexoses and sugar acids are also present. The third major component of plant biomass lignin has a rather complex structure, composed of various interconnected phenol propane units. Of these three components, lignin is the most resistant to the enzymatic attack (Paulova *et al*. 2015). Indeed, this very resistance exhibited by the lignocellulose complex constitutes a major obstacle to exploit the potential of plant biomass in biotechnology, as it prolongs processes and simultaneously decreases yields.

Due to these reasons, physico-chemical pretreatment of the biomass is usually included prior to the main biotechnological process to disrupt the organized structure of the lignocellulose complex (Fan *et al*. 2006). The steam explosion method is the most often used pretreatment (Galbe and Zacchi 2007; Bruni *et al*. 2010). In this process, the biomass slurry is fed to a channel, where it is heated and pressurized with external hot steam. Upon leaving the channel, a pressure drop causes disruption to the structure of the lignocellulose complex. The three principle parameters that have to be controlled are temperature, pressure, and retention time. The mentioned procedure was used with positive results for the preprocessing of swine slurry (Ortega-Martineza *et al*. 2016), vinegar residue (Feng *et al*. 2016), or corn stover (Ji *et al*. 2016), where a significant increase of methane production was observed. Other authors reported improvements in the process kinetics (Theuretzbacher *et al*. 2015; Rincón *et al*. 2016). A more complex and somewhat skeptical view was brought by Dereix *et al*. (2006) showing that despite the increase in methane yield by about 50% the energy requirements of the steam-explosion process were substantially higher than the additional energy produced.

Bauer *et al.* (2009) studied the effect of steam explosion pretreatment on the methane yield from wheat straw, finding that at 10 min at 160 °C or 15 min at 180 °C, the treatment increased the specific methane yield by up to 14%, or 20%, respectively. Surprisingly, at 20 min at 200 °C, the positive effect on methane production was no longer observed. Treatment of plant biomass involving relatively high temperatures under elevated pressure can give rise to toxic degradation products derived from plant biomass constituents; indeed, such compounds may subsequently inhibit the key enzymes of the process. Examples of potentially inhibiting substances include phenolic compounds, particularly vanillin, gallic acid, and tannic acid, which arise through lignin decomposition, as well as the dehydration products of pentoses and hexoses, *i.e.* furfural and 5-hydroxymethylfurfural, respectively, from corresponding polysaccharides (Palmqvist and Hahn-Hagerdal 2000). The effects of furfural and 5-hydroxymethylfurfural have been studied, to some extent, in connection with biotechnological production of bioethanol or hydrogen from lignocellulose biomass (Oliva *et al*. 2006; Lu *et al*. 2007; Bellido *et al*. 2011; Paulova *et al*. 2012). Inhibition of at least three enzymes of the central carbon metabolism has been proven (Modig *et al*. 2002). The concentrations of these compounds in the pretreated biomass were over 1 g/L in some cases, and in an extreme instance even exceeded 3 g/L. Nevertheless, it was shown that for the production of hydrogen the threshold of the inhibition effect was about1 g/L. As a representative of the group of phenolic inhibition compounds derived from lignin, gallic acid was tested for the inhibition effect on methane production by Mousa and Forster (1999). The compound did not exhibit any significant effect in concentrations of up to 20 mg/L, whereas 50 mg/L caused about a 15% decrease in methane content of the biogas. The same study also showed

that supplementation with an additional, easily metabolizable carbon substrate (glucose) effectively restored methane production, likely through promoting degradation of the inhibiting compound. In another study (Hernandez and Edyvean 2008), caffeic acid and gallic acid were selected as model phenolic degradation products of lignin, and their inhibition effect on the production of methane was significant at concentrations of about 1.0 g/L.

In this study, the inhibition effect of furfural and 5-hydroxymethylfurfural, as model cellulose and hemicellulose degradation products, and gallic and tannic acids as model lignin degradation products were investigated in an experimental system mimicking general conditions at an actual industrial biogas plant. The aim was to evaluate the potential risks of the steam explosion pretreatment of plant biomass on methane production, with respect to the possible formation of inhibiting degradation products and their concentrations.

EXPERIMENTAL

Microbial Inoculum

Digestate containing the microbial community was sampled at the biogas plant (České Budějovice, Czech Republic), which mainly utilizes corn straw and grass silage as the feedstock for anaerobic fermentation at 39 °C. The sludge was sieved (1 mm^2) to remove corn grains and other solid particles and was suspended in mineral salt medium (MgSO$_4$·7H$_2$O, 22.5 g/L; CaCl$_2$, 27.5 g/L; FeCl$_3$·6H$_2$O, 0.25 g/L; (NH$_4$)$_2$SO$_4$, 10.0 g/L; KH$_2$PO$_4$, 8.2 g/L; K$_2$HPO$_4$, 21.8 g/L; Na$_2$HPO$_4$·12H$_2$O, 44.7 g/L; H$_3$BO$_3$, 0.75 g/L; FeSO$_4$·7H$_2$O, 3.00 g/L; ZnSO$_4$·7H$_2$O, 0.10 g/L; MnSO$_4$·4H$_2$O, 0.50 g/L) to reach the dry-weight concentration of 1 g/L. After dilution, the sludge was acclimated 7 days at 39 °C.

The basic characteristics of the digestate after sampling and after acclimation are summarized in Table 1 to document that the diluted and acclimated media from all three sampling instants had similar properties at the beginning of the experiments.

Table 1. Physicochemical Properties of Initial and Acclimated Digestate

	Sampling I	Sampling II	Sampling III
Initial digestate			
ORP (mV)	-289	-383	-219
pH	7.8	8.9	8.2
DW (g/L)	5.70	3.00	28.5
Acclimated digestate			
ORP (mV)	-208	-253	-225
pH	7.8	8.0	8.0
DW (g/L)	1.14	0.98	0.98
%CH$_4$(acetate)	62%	68%	-
%CH$_4$(cellulose)	-	49%	30%

ORP, oxidation-reduction potential; DW, dry weight; %CH$_4$, percentage of methane in biogas with the given substrate in the initial phase of the experiment.

Production of Methane and Carbon Dioxide during Anaerobic Digestion

Biogas production experiments were performed in 300 mL flasks filled with 100 mL of the acclimated digestate, the appropriate carbon substrate (sodium acetate or cellulose) was added at the final concentration of 10 g/L, and the contents were purged with nitrogen. Afterwards, the tested inhibiting substances were added at given concentrations. At the beginning of the tests, the pH was 7.8 ± 0.5. After repeated purging with nitrogen, all flasks were sealed with stoppers equipped with gas-tight sampling valves and incubated at 39 °C. The contents of bottles were continuously stirred, and the gas phase of each flask was sampled several times a week. The gas pressure in the flasks was controlled by a digital manometer (ISO 11734 (1995)). All experiments, including blank and control bottles, contained all the components except the tested compound or the carbon substrate, respectively, and were carried out in triplicate.

CH_4 and CO_2 production were determined by gas chromatography (Agilent GC 7890A equipped with a PORAPAK Q column and TCD detector, Santa Clara, USA; with helium as a carrier gas at 50 mL/min, $T_{injector}$ = 200 °C, T_{owen} = 50 °C, T_{det} = 220 °C) and expressed as the amount of carbon in the form methane produced per gram of carbon introduced m(CH_4), the calculation were based on the ideal gas state equation (Drimal *et al.* 2006; Hubáčková *et al.* 2013). The concentrations of CH_4 and CO_2 were derived from the calibration curve obtained using the calibration gas mixture with certified composition (0.8% CO_2, 4% CH_4, 95.2% N_2, Linde 2016). At the end of the incubation dissolved inorganic carbon content was determined in all flasks (5000A TOC analyzer, Shimadzu, Tokyo, Japan). The pH and oxidation-reduction potential (ORP) were also determined to verify the validity of experiments.

The percentage of the net mineralization of the organic carbon (D_t) was estimated from the relation of the amount of carbon released in the form of CH_4 and CO_2 (m_g, mg) and dissolved innorganic carbon (m_i, mg), corrected by a blank, to a theoretical amount of the carbon introduced as the substrate (cellulose or acetate) and the inhibiting compound eventually (m_v, mg), and expressed in terms of the percentage of anaerobic biodegradation, as follows:

$$D_t = \frac{m_g + m_i}{m_v} \times 100\ \% \qquad (1)$$

Mathematical Model

Methane production during the growth of microbial biomass was expressed according to a previously published model of inhibited biomass growth (Rial *et al.* 2011), where the amount of the produced methane is proportional to the biomass and the initial methane concentration is zero (Eq. 2),

$$m(CH_4) = k \left[\frac{X_0}{1 + \exp(c - \mu_m t)} - X_0 \right] \qquad (2)$$

$$\text{where } c = \ln\left(\frac{X_m}{X_0} - 1 \right);$$

m(CH_4) is the mass of the methane produced; k is the proportionality constant between biomass and methane production; X_0 is the biomass at the beginning of the experiment; X_m

is the asymptotic maximum of biomass; and μ_m is the the biomass increase per biomass unit and time unit (dimensions t^{-1}).

After re-parametrization (Rial *et al.* 2011), the model equation for the methane production is as follows,

$$m(CH_4) = \frac{m(CH_4)_{MAX}}{1+\exp\left[2+\frac{4v_{MAX}}{m(CH_4)_{MAX}}(\lambda-t)\right]} - \frac{m(CH_4)_{MAX}}{1+\exp\left[2+\frac{4v_{MAX}}{m(CH_4)_{MAX}}\lambda\right]} \quad (3)$$

where $m(CH_4)_{MAX}$ is the asymptotic maximum of methane production; v_{MAX} is the maximal rate of methane production; and λ is the length of lag phase. Fitting and estimation of parameters from the experimental results were performed by sum of square minimisation nonlinear method (quasi-Newton) using the macro 'Solver' of Microsoft Excel.

RESULTS AND DISCUSSION

The inhibition effect of selected model compounds was studied in experiments simulating biogas production from corn biomass, as is often the case at industrial scale biogas production plants. Semiliquid digestate from such a process was utilized as the microbial inoculum. Two groups of experiments were carried out with two different carbon substrates. The first one, with cellulose, comprised all phases of the process, *i.e.* hydrolysis, acetogenesis, and methanogenesis, whereas the other set of experiments with sodium acetate studied solely the methanogenesis. Furfural and 5-hydroxymethylfurfural were selected as model inhibition compounds formed as dehydration products during thermal treatment of carbohydrate-based plant biomass constituents (Barakat *et al.* 2012; Chiaramonti *et al.* 2012); gallic and tannic acids are phenolic model inhibition compounds derived during the degradation of plant lignin (Hernandez and Edyvean 2008; Barakat *et al.* 2012).

Inhibition Effect of Furfural and 5-Hydroxymethylfurfural on Biogas Production

The inhibition effects of furfural and 5-hydroxymethylfurfural were studied in concentrations of up to 2 g/L, which lie at the higher limit of expected concentrations in a real process (Bellido *et al.* 2011). The data obtained on methane production were fitted with the described mathematical model and the parameters of the model, *i.e.* the length of the lag phase, maximal methane production rate, and asymptotic maximal methane production level calculated (Table 2). Subsequently, these parameters were used to compare inhibition effects among the individual experiments. During the experiment with sodium acetate as the carbon substrate, furfural concentrations under 1 g/L exhibited a stimulating effect on methane production; hence, the compound could probably be utilized as an additional carbon substrate under such conditions (Fig. 1A). Even at 1.0 g/L, it was possible to discern a noticeable inhibition effect, and the effect was proportionally more pronounced at 2.0 g/L. Inhibition by furfural was manifested primarily through an increase in lag phase prior to the onset of methane production (Fig. 2). However, it seemed that even at the highest furfural concertation tested, the compound was metabolized, *i.e.* detoxified. After a certain time, methane production was substantially restored with a comparable maximal production rate to the control experiment without the inhibiting compound. The

240

same was generally applicable for experiments with cellulose as the carbon substrate (Fig. 1C), except that methane production inhibition was exhibited as early as at 0.5 g/L and then strongly at 1.0 g/L. Thus, it is likely that some processes preceding methanogenesis in the path of cellulose utilization are slightly more sensitive to furfural inhibition than methanogenesis itself.

5-hydroxymethylfurfural was found to have a slightly stronger inhibition effect than furfural, showing a noticeable influence even at 0.5 g/L (Fig. 1B). In contrast to furfural, hydroxylmethylfurfural decreased the maximal methane production rate and the level of maximum methane production (Fig. 2), presenting no evident tendency to restore methane production after a certain period. This suggested that hydroxymethylfurfural was not readily removed from the system and caused permanent inhibition of the process. As in the case of furfural, methane production from cellulose was noticeably more inhibitor-sensitive than acetate as the carbon substrate, with an inhibition appearing even at 0.2 g/L 5-hyroxymethylfurfural (Fig. 1 D). The latter finding confirms that methanogenesis was not a critical step in biogas production with respect to sensitivity to these inhibitory compounds. Rather, the critical step was cellulolysis or acetogenesis.

Table 2. Optimized Parameters of the Mathematical Model

Acetate as the carbon substrate								
c_{FUR} (g/L)	λ (d)	$m(CH_4)_{MAX}$ (mg/g)	v_{MAX} (mg/(g•d))	c_{HMF} (g/L)	λ (d)	$m(CH_4)_{MAX}$ (mg/g)	v_{MAX} (mg/(g•d))	
0.0	5.0	213	11	0.0	15	280	4.94	
0.1	0	230	12	0.1	ND	ND	ND	
0.2	0.9	226	18	0.2	7.7	324	6.39	
0.5	2.7	200	24	0.5	11	280	3.38	
1.0	9.9	180	16	1.0	4.4	103	0.91	
2.0	23	126	18	2.0	ND	ND	ND	
Cellulose as the carbon substrate								
c_{FUR} (g/L)	λ (d)	$m(CH_4)_{MAX}$ (mg/g)	v_{MAX} (mg/(g•d))	c_{HMF} (g/L)	λ (d)	$m(CH_4)_{MAX}$ (mg/g)	v_{MAX} (mg/(g•d))	
0.0	1.9	121	4.58	0.0	0.0	28.2	0.68	
0.1	0.1	132	5.01	0.1	ND	ND	ND	
0.2	3.1	114	4.28	0.2	0.0	12.9	18.0	
0.5	2.7	106	4.35	0.5	3.0	4.42	17.6	
1.0	23	140	4.95	1.0	20	2.57	ND	
2.0	ND	0	ND	2.0	ND	ND	ND	

λ, length of lag phase; $m(CH_4)_{MAX}$, asymptotic maximum of methane production; v_{MAX}, maximal rate of methane production; d, day; c_{FUR}, concentration of furfural; c_{HMF}, concentration of 5-hydroxymethylfurfural.

The reactions were characterized at the end of the incubation period, as summarized in Table 3. In all experiments, the pH and concentration of dissolved inorganic carbon indicated that adding an inhibiting compound did not cause a shift in metabolic processes from methanogenesis. The amounts of dissolved inorganic carbon were comparable in every incubation, which confirmed that the contents of gaseous-endmetabolites, which were determined in the head space of the incubation vessels, are applicable for interpreting the development of metabolic processes.

241

Table 3. Selected Properties of the Individual Incubations at the End of the Experiments

Acetate as the carbon substrate									
c_{FUR} (g/L)	pH	IC (g/L)	D_t (%)	CH_4 (%)	c_{HMF} (g/L)	pH	IC (g/L)	D_t (%)	CH_4 (%)
0.0	8.38	1.73 ± 0.05	75	93	0.0	8.36	1.74 ± 0.10	89	97
0.1	8.31	1.88 ± 0.04	78	93	0.2	8.25	1.75 ± 0.09	87	94
0.2	8.29	1.78 ± 0.08	76	92	0.5	8.05	1.69 ± 0.11	80	89
0.5	8.11	1.71 ± 0.12	70	89	1.0	8.42	1.65 ± 0.01	56	24
1.0	7.97	1.83 ± 0.16	66	87	2.0	8.18	0.0	0	0
2.0	7.21	1.41 ± 0.62	46	79	4.0	8.21	0.0	0	0
Cellulose as the carbon substrate									
c_{FUR} (g/L)	pH	IC (g/L)	D_t (%)	CH_4 (%)	c_{HMF} (g/L)	pH	IC (g/L)	D_t (%)	CH_4 (%)
0.0	7.99	1.13 ± 0.02	42	57	0.0	8.15	1.99 ± 0.07	51	39
0.1	8.15	1.10 ± 0.01	42	60	0.2	8.09	2.12 ± 0.05	50	26
0.2	8.32	1.12 ± 0.02	40	57	0.5	8.32	2.06 ± 0.05	47	8
0.5	8.13	1.10 ± 0.03	39	56	1.0	8.24	2.35 ± 0.08	47	5
1.0	8.21	1.07 ± 0.36	37	67	2.0	8.40	0	0	0
2.0	8.09	0.00	0	9	4.0	8.15	0	0	0

c_{FUR}, concentration of furfural; c_{HMF}, concentration of 5-hydroxymethylfurfural; IC, concentration of dissolved inorganic carbon in aqueous phase; D_t, % of carbon mineralization; CH_4, the percentage of CH_4 in biogas

As the simpler of the two substrates, acetate exhibited a considerably high level of mineralization, which resulted in a higher content of methane in the biogas formed. Furthermore, the effective concentrations of inhibiting compounds gradually decreased the content of methane in the biogas produced, although a dramatic effect was only seen at concentrations strongly inhibiting methane production. The last performed experiments with 5-hydroxymethylfurfural and cellulose as substrate were performed with the inoculum from the third sampling instant. In these experiments the production rate of CH_4 and its content in biogas (Table 1) was somewhat lower witnessing probably a lower initial content of methanogens on the beginning of the experiment.

242

■ 0 g L⁻¹; ■ 0.1 g L⁻¹; ■ 0.2 g L⁻¹; ■ 0.5 g L⁻¹; ■ 1.0 g L⁻¹; ■ 2.0 g L⁻¹.
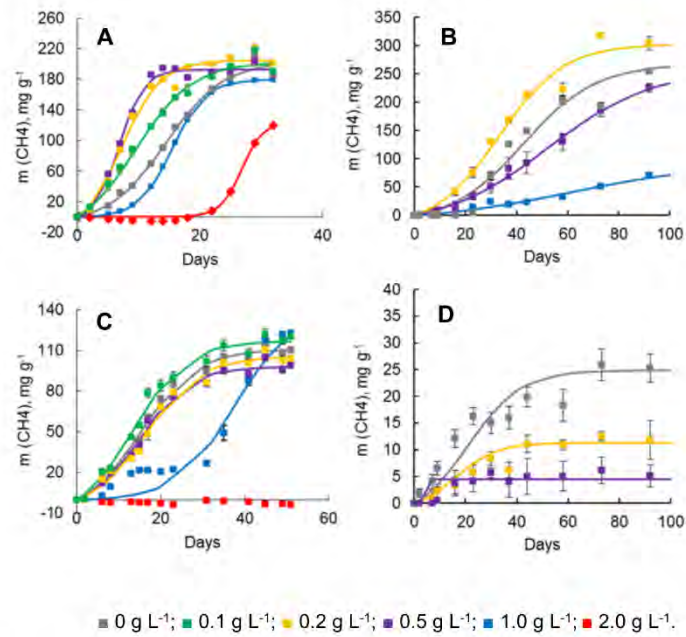
Fig. 1. Production of methane in the presence of various concentrations of furfural (A, C) and 5-hydroxymethylfurfural (B, D). The methane production is expressed per gram of carbon substrate *i.e.* sodium acetate (A, B) or cellulose (C, D). Solid lines represent described theoretical models fitted to the experimental data.
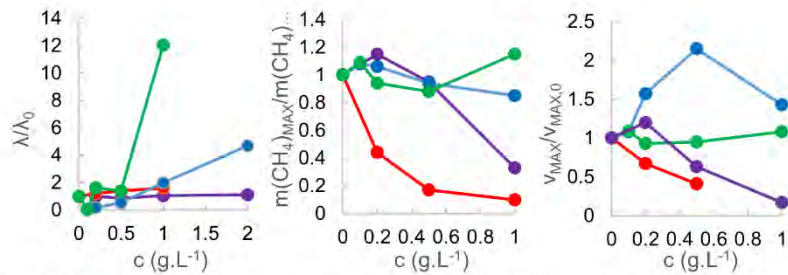
Fig. 2. Changes of model parameters: length of lag phase (λ), asymptotic maximum of methane production ($m(CH_4)_{MAX}$), and maximal rate of methane production (v_{MAX}) with the concentration of inhibiting compounds relative to the parameter value at zero concentration of the inhibiting compound (λ_0; $m(CH_4)_{MAX,0}$; $v_{MAX,0}$). ●, furfural +sodium acetate as substrate; ●,furfural+cellulose as substrate; ●, 5-hydroxymethylfurfural +sodium acetate as substrate; ●, 5-hydroxymethylfurfural +cellulose as substrate.

243

Inhibition Effect of Tannic Acid and Gallic Acid

The inhibitory effects of the phenolic compounds, as model degradation products of plant lignin, were studied up to a concentration of 2.0 g/L. As shown in Fig. 3, even at the highest concentrations they were substantially incapable of inhibiting methane production from either of the two tested carbon substrates. However, in the experiments with acetate, the maximal methane production level was slightly lower for the two highest concentrations of gallic and tannic acid. Nevertheless, such high concentrations are not expected in real conditions. Indeed, the presence of phenol compounds also failed to demonstrate any apparent stimulating effect.

The selected parameters determined at the end of the incubation period (Table 2) confirmed that the basic chemical conditions in the processes were comparable regardless of the concentration of the inhibiting compounds. The moderate decrease in the level of mineralization at the highest concentration of gallic acid in the presence of both substrates could signal the onset of inhibition again at this rather extreme concentration of the compound.
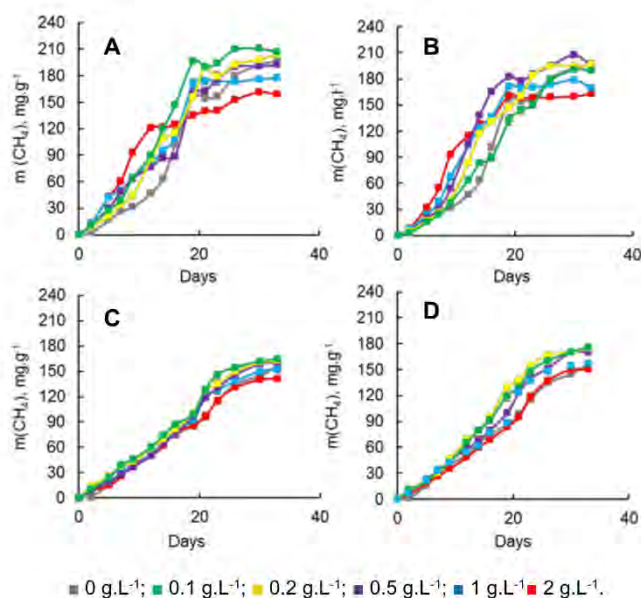
Fig. 3. Production of methane in the presence of various concentrations of gallic acid (A, C) and tanic acid (B, D). The methane production is expressed per gram of carbon substrate, *i.e.* sodium acetate (A, B) or cellulose (C, D).

CONCLUSIONS

1. The study showed no noticeable inhibition of methane production by furfural at concentrations below 1 g/L

244

2. 5-hydroxymethylfurfural was a stronger inhibitor with a noticeable effect above 0.2 g/L.

3. Both compounds were more effective inhibitors with cellulose as the carbon substrate, potentially reflecting the higher sensitivity of the cellulolytic step in the process of biogas production.

4. No inhibition was observed for the tested phenolic compounds (gallic and tannic acid) at concentrations of up to 2 g/L.

5. The compounds investigated do not raise an issue for biogas production involving plant biomass preprocessed by the steam explosion method.

ACKNOWLEDGMENTS

REFERENCES CITED

Adney, W. S., Rivard, C. J., Shiang, M., and Himmel, M. E. (1991). "Anaerobic-digestion of lignocellulosic biomass and wastes - Cellulases and related enzymes," *Applied Biochemistry and Biotechnology* 30(2), 165-183. DOI: 10.1007/BF02921684

Barakat, A., Monlau, F., Steyer, J. P., and Carrere, H. (2012). "Effect of lignin-derived and furan compounds found in lignocellulosic hydrolysates on biomethane production," *Bioresource Technology* 104, 90-99. DOI: 10.1016/j.biortech.2011.10.060

Bauer, A., Bosch, P., Friedl, A., and Amon, T. (2009). "Analysis of methane potentials of steam-exploded wheat straw and estimation of energy yields of combined ethanol and methane production," *Journal of Biotechnology* 142(1), 50-55. DOI: 10.1016/j.jbiotec.2009.01.017

Bellido, C., Bolado, S., Coca, M., Lucas, S., Gonzalez-Benito, G., and Garcia-Cubero, M. T. (2011). "Effect of inhibitors formed during wheat straw pretreatment on ethanol fermentation by *Pichia stipitis*," *Bioresource Technology* 102(23), 10868-10874. DOI: 10.1016/j.biortech.2011.08.128

Bruni, E., Jensen, A.P., and Angelidaki, I. (2010). "Steam treatment of digested biofibers for increasing biogas production," *Bioresource Technology* 101(19), 7668-7671. DOI: 10.1016/j.biortech.2010.04.064

Chiaramonti, D., Prussi, M., Ferrero, S., Oriani, L., Ottonello, P., Torre, P., and Cherchi, F. (2012). "Review of pretreatment processes for lignocellulosic ethanol production, and development of an innovative method," *Biomass & Bioenergy* 46, 25-35. DOI: 10.1016/j.biombioe.2012.04.020

Drimal, P., Hrncirik, J., and Hoffmann, J. (2006). "Assessing aerobic biodegradability of plastics in aqueous environment by GC-analyzing composition of equilibrium gaseous phase," *Journal of Polymers and the Environment* 14(3), 309-316. DOI: 10.1007/s10924-006-0024-5

Fan, Y. T., Zhang, Y. H., Zhang, S. F., Hou, H. W., and Ren, B. Z. (2006). "Efficient conversion of wheat straw wastes into biohydrogen gas by cow dung compost," *Bioresource Technology* 97(3), 500-505. DOI: 10.1016/j.biortech.2005.02.049

Feng, J., Zhang, J., Zhang, J., He, Y., Zhang, R., Liu, G., and Chen, C. (2016). "Influence of steam explosion pretreatment on the anaerobic digestion of vinegar residue," *Waste Management & Research* 34(7), 630-637. DOI: 10.1177/0734242X16644681

Galbe, M., and Zacchi, G. (2007). "Pretreatment of lignocellulosic materials for efficient bioethanol production," *Biofuels* 108, 41-65. DOI: 10.1007/10_2007_070

Hendriks, A. T. W. M., and Zeeman, G. (2009). "Pretreatments to enhance the digestibility of lignocellulosic biomass," *Bioresource Technology* 100(1), 10-18. DOI: 10.1016/j.biortech.2008.05.027

Hernandez, J. E., and Edyvean, R. G. J. (2008). "Inhibition of biogas production and biodegradability by substituted phenolic compounds in anaerobic sludge," *Journal of Hazardous Materials* 160(1), 20-28. DOI: 10.1016/j.jhazmat.2008.02.075

Hubáčková, J., Dvořáčková M., Svoboda P., Mokrejš P., Kupec J., Požárová I., Alexy P., Bugaj P., Machovský M., and Koutný M. (2013), "Influence of various starch types on PCL/starch blends anaerobic biodegradation," *Polymer Testing* 32(6), 1011-1019. DOI: 10.1016/j.polymertesting.2013.05.008

ISO 11734 (1995). "Water quality - Evaluation of the 'ultimate' anaerobic biodegradability of organic compounds in digested sludge - Method by measurement of the biogas production," International Organization for Standardization, Geneva, Switzerland.

Ji, J., Zhang, J., Yang, L., He, Y., Zhang, R., Liu, G., and Chen, C. (2016). "Impact of co-pretreatment of calcium hydroxide and steam explosion on anaerobic digestion efficiency with corn stover," *Environmental Technology* online, DOI: 10.1080/09593330.2016.1234001

Lu, P., Chen, L. J., Li, G. X., Shen, S. H., Wang, L. L., Jiang, Q. Y., and Zhang, J. F. (2007). "Influence of furfural concentration on growth and ethanol yield of *Saccharomyces kluyveri*," *Journal of Environmental Sciences* 19(12), 1528-1532. DOI: 10.1016/S1001-0742(07)60249-6

Modig, T., Liden, G., and Taherzadeh, M. J. (2002). "Inhibition effects of furfural on alcohol dehydrogenase, aldehyde dehydrogenase and pyruvate dehydrogenase," *Biochemical Journal* 363, 769-776. DOI: 10.1042/0264-6021:3630769

Mousa, L., and Forster, C. F. (1999). "The use of glucose as a growth factor to counteract inhibition in anaerobic digestion," *Process Safety and Environmental Protection* 77(B4), 193-198. DOI: 10.1205/095758299530062

Oliva, J. M., Negro, M. J., Saez, F., Ballesteros, I., Manzanares, P., Gonzalez, A., and Ballesteros, M. (2006). "Effects of acetic acid, furfural and catechol combinations on ethanol fermentation of *Kluyveromyces marxianus*," *Process Biochemistry* 41(5), 1223-1228. DOI: 10.1016/j.procbio.2005.12.003

Ortega-Martineza, E., Zaldivara, C., Phillippia, J., Carrerec, H., and Donoso-Bravoa, A. (2016). "Improvement of anaerobic digestion of swine slurry by steam explosion and chemical pretreatment application. Assessment based on kinetic analysis," *Journal of Environmental Chemical Engineering* 4(2), 2033-2039. DOI: 10.1016/j.jece.2016.03.035

Palmqvist, E., and Hahn-Hagerdal, B. (2000). "Fermentation of lignocellulosic hydrolysates. II: Inhibitors and mechanisms of inhibition," *Bioresource Technology*

246

74(1), 25-33. DOI: 10.1016/S0960-8524(99)00161-3

Paulova, L., Patakova, P., Branska, B., Rychtera, M., and Melzoch, K. (2015). "Lignocellulosic ethanol: Technology design and its impact on process efficiency," *Biotechnology Advances* 33(6 Pt 2), 1091-107. DOI: 10.1016/j.biotechadv.2014.12.002

Paulova, L., Patakova, P., Jaisamut, K., Rychtera, M., and Melzoch, K. (2012). "Cellulose ethanol - Influence of inhibitors on production strains," *Listy Cukrovarnicke a Reparske* 128(7-8), 229-232.

Prochazka, J., Mrazek, J., Strosova, L., Fliegerova, K., Zabranska, J., and Dohanyos, M. (2012). "Enhanced biogas yield from energy crops with rumen anaerobic fungi," *Engineering in Life Sciences* 12(3), 343-351. DOI: 10.1002/elsc.201100076

Rial, D., Vazquez, J. A., and Murado, M. A. (2011). "Effects of three heavy metals on the bacteria growth kinetics: A bivariate model for toxicological assessment," *Applied Microbiology and Biotechnol.* 90(3), 1095-1109. DOI: 10.1007/s00253-011-3138-1

Rincón, B., Rodríguez-Gutiérrez, G., Bujalance, L., Fernández-Bolaños, J., and Borja, R. (2016). "Influence of a steam-explosion pre-treatment on the methane yield and kinetics of anaerobic digestion of two-phase olive mil solid waste or alperujo," *Process Safety and Environmental Protection* 102, 361-369. DOI: 10.1016/j.psep.2016.04.010

Theuretzbacher, F., Lizasoain, J., Lefever, C., Saylor, M. K., Enguidanos, R., Weran, N., Gronauer, A., and Bauer, A. (2015). "Steam explosion pretreatment of wheat straw to improve methane yields: Investigation of the degradation kinetics of structural compounds during anaerobic digestion," *Bioresource Technology* 179, 299-305. DOI: 10.1016/j.biortech.2014.12.008

P19. Ab initio calculations of substituent constants: A reinvestigation.

O. Exner, M. Ingr, P. Čársky, Theochem-J. Mol. Struct. 397 (1997) 231–238.

Ab initio calculations of substituent constants: a reinvestigation

Otto Exner[a,*], Marek Ingr[1,b], Petr Čársky[b]

[a]*Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 166 10 Prague 6, The Czech Republic*
[b]*J. Heyrovský Institute of Physical Chemistry, Academy of Sciences of the Czech Republic, Dolejškova 3, 182 23 Prague 8, The Czech Republic*

Abstract

A set of various substituent constants σ_F, σ_R, σ_α, and σ_χ for 21 common substituents was recalculated according to the models of Marriott and Topsom which are based on ab initio calculations. Some inconsistencies in the original methods as to the use of molecular geometries and basis sets were eliminated and a few values of the original work were corrected. The set was extended by an additional 10 substituents derived from the amido and thioamido groups, with particular interest for QSAR. Comparison with standard experimental data confirmed the physical meaning and significance of the computational model for σ_F, σ_R, σ_α, but not for σ_χ. © 1997 Elsevier Science B.V.

Keywords: Substituent effects; Ab initio calculations; Inductive effect; Resonance; Polarizability; Electronegativity

1. Introduction

The substituent effects on any physical property are commonly described in terms of substituent constants σ [1]. In the most developed form of the empirical theory [2], four such constants are distinguished: σ_F, σ_R, σ_α, and σ_χ, corresponding to four assumed interaction mechanisms called field-inductive, resonance, polarizability, and electronegativity. A particular property is expressed as

$$y = y_o + \rho_F\sigma_F + \rho_R\sigma_R + \rho_\alpha\sigma_\alpha + \rho_\chi\sigma_\chi + \varepsilon \tag{1}$$

In Eq. (1) y and y_o means the observed quantity for the substituted and unsubstituted compound, respectively,

ρ's are proportionality constants and ε is considered to be a random variable (fitting error). Some of the terms in Eq. (1) are often insignificant—in particular, ρ_χ is different from zero only for immediately adjoining substituents and ρ_α only for some reactions in the gas phase. On the other hand, applications restricted to the terms $\rho_F\sigma_F$ and $\rho_R\sigma_R$ are innumerable and were even criticized for being used where they are not appropriate [3]. The constants σ were originally defined [1] on model processes involving experimental quantities (pK, NMR shifts, IR intensities) and are merely intuitively connected with the conception of resonance, inductive effect etc. Their actual verification is only by the fitting of many experimental data in Eq. (1). Constants determined in this way are always loaded with experimental errors, often of unknown magnitude; sometimes even the definition procedure is not suitable for a given substituent due

---

* Corresponding author.
[1] Undergraduate at the Faculty of Science, Charles University, Prague.

Table 1
Calculated substituent constants for common substituents

Substituent	σ_F	σ_R	σ_α	σ_χ
CH_3	0.02	−0.11	−0.35	0.00
C_2H_5	0.02	−0.10	−0.49	−0.01
$CH(CH_3)_2$	0.01	−0.08	−0.62	−0.02
$C(CH_3)_3$	0.02	−0.07	−0.75	−0.02
CH_2F	0.14	−0.10	−0.33	−0.02
CF_3	0.34	0.04	−0.25	−0.02
CHO	0.23	0.19	−0.46	−0.05
$COCH_3$	0.18	0.14	−0.55	−0.04
$COOCH_3$	0.17	0.15	−0.48	0.02
COF	0.37	0.20	a	0.03
CN	0.49	0.07	−0.52	0.29
NH_2	0.17	−0.56	−0.17	0.33
$N(CH_3)_2$	0.17	−0.57	−0.44	0.34
NO_2	0.58	0.17	−0.26	0.46
OH	0.31	−0.40	−0.03	0.54
OCH_3	0.30	−0.40	−0.17	0.54
F	0.46	−0.30	0.13	0.70
SH	0.21	−0.28	−0.53	−0.11
SCH_3	0.15	−0.29	−0.68	−0.15
SO_2CH_3	0.38	0.02	−0.63	−0.24
Cl	0.36	−0.23	−0.44	0.15

[a] The σ_α constant for COF is missing. 3-21G is a basis set too small for a meaningful evaluation of this constant (see the text).

to technical difficulties such as solubility, side reactions, or interference of spectral bands. For this reason, methods were advanced to predict the constants σ by quantum chemical ab initio calculations [4–8]. Within the framework of theoretical chemistry, the substituent effect may be defined as ΔH^o of the reaction (2), usually isodesmic, involving the change of a reaction center Y into Y' in dependence on the substituent X,

$$X-Y+H-Y'=X-Y'+H-Y \qquad (2)$$
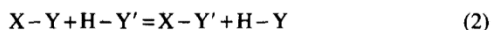
However, theoretical chemistry does not allow the calculated effect to be separated into the terms inductive, resonance, etc. Marriott and Topsom (thereafter M.T.) [4–8] have advanced specific model systems, molecules or even only conglomerates of atoms, intuitively related to these terms. The physical meaning of the calculated constants σ_F, σ_R, σ_α, and σ_χ was verified in fact only by some a posteriori correlations with experimental σ. The merit of the approach is in any case their reproducibility and applicability to all possible substituents.

Our interest in this area was primarily connected with the thioamido group which is important [9] in

pharmacology and QSAR (Quantitative Structure-Activity Relationships) but has been less characterized as substituent [10–14]. First it was necessary to reproduce the procedures described [4–8] on standard substituents which are listed in Table 1. Then we applied them to addtional substituents, not included in the original sets. These are listed in Table 2. A point of interest was also the effect of conformation in the case of substituents with a rotational freedom. Such an effect can be easily calculated but has been only exceptionally observed experimentally on two conformers of one compound [15], although the significance of substituent orientation was many times proven by comparing two different compounds [16,17]. Variable conformations were considered here in the case of eight substituents of Table 2.

2. Computational details

We intended to follow closely the methods developed by M.T. [4–8,18] who treated altogether 21 substituents. Typically, however, they studied in a single paper the basis set effect on a particular σ constant

Table 2
Calculated σ constants for substituents

Substituent	σ_F	σ_F exp.[b]	σ_R	σ_R exp.[b]	σ_α	σ_x
CONH$_2$	0.21	0.23, 0.29[f]	0.07	0.11, 0.12,[f] 0.13[g]	−0.46	−0.04
(Z)-CONHCH$_3$	0.19	0.35[c]	0.06	−0.01[c]	−0.52	−0.04
(E)-CONHCH$_3$	0.18		0.09		0.05	−0.05
CON(CH$_3$)$_2$	0.31	0.23[i]	0.08		−0.69	−0.04
(Z)-NHCOCH$_3$	0.17	0.24,[h] 0.34	−0.29	−0.22,[h] −0.21, −0.42[g]	−0.28	0.46
(E)-NHCOCH$_3$	0.43		−0.38		−0.47	0.44
(Z)-NHCONH$_2$	0.13	0.09[d]	−0.33	−0.33[d]	−0.25	0.47
(E)-NHCONH$_2$	0.49		−0.39		a	0.43
CSNH$_2$	0.28	0.24,[e] 0.29[f]	0.11	0.11,[f] 0.06[e]	−0.65	0.11
(Z)-CSNHCH$_3$	0.24	0.29[i]	0.07	0.05[i]	−0.71	0.10
(E)-CSNHCH$_3$	0.27		0.09		−0.80	0.10
CSN(CH$_3$)$_2$	0.24	0.23[j]	0.11		−0.88	0.10
(Z)-NHCSCH$_3$	0.18	0.30[c]	−0.21	−0.18[c]	−0.48	0.51
(E)-NHCSCH$_3$	0.54		−0.29		−0.68	0.47
(Z)-NHCSNH$_2$	0.12	0.26, 0.29[h]	−0.27	−0.10, −0.13[h]	a	0.52
(E)-NHCSNH$_2$	0.63		−0.32		−0.55	0.45

[a] σ_α constants for (E)-NHCONH$_2$ and (Z)-NHCSNH$_2$ are not listed since the 3-21G and 3-21G* basis sets are too small for a meaningful evaluation (see the text).

[b] Ref. [22] unless otherwise noted.

[c] Estimates derived indirectly from the constants σ_m and σ_p of Ref. [28].

[d] Derived in the same way from Ref. [30].

[e] Derived in the same way from Ref. [11].

[f] Ref. [27].

[g] Ref. [25].

[h] Ref. [29].

[i] Ref. [28].

[j] Ref. [31].

with a smaller subset of substituents. We found it therefore difficult to reproduce their complete set for all four σ constants by a uniform treatment of all 21 substituents. With some molecules we were not sure about the geometries they used (optimized or standard) and some of our calculated σ constants differed considerably from those given in the final review [18]. For this reason we considered it useful to present here Topsom's formulas [18], to specify all necessary technical details of calculations.

A field-effect constant σ_F for a substituent X is calculated by the formula

$$\sigma_F = -35.5 \Delta q_{H_\alpha} \qquad (3)$$

where Δq_{H_α} is the difference in charge,

$$\Delta q_{H_\alpha} = q_{H_\alpha(H_2 \cdots HX)} - q_{H_\alpha(H_2 \cdots H_2)} \qquad (4)$$

between the charge on hydrogen H_α in the model system 1 [4] and the charge on hydrogen H_α in an analogous system in which $X = H$. First, the geometries

$$H\text{-}H_\alpha \cdots \cdots \cdots H\text{-}X$$

$$4\text{Å}$$

$$1$$

of the H$_2$ and HX molecules were optimized at the 6-31G* level. Then a single-point 6-31G* calculation was performed for the model 1 and the calculated Mulliken populations are used in Eq. (4) for the evaluation of Δq_{H_α}. M.T. tested several variants of their calculations of σ_F [4,18] and did not give quite clearly which was used for obtaining the final data. After several attempts, we selected the 6-31G* approach with optimal geometry but some disagreements between M.T. and our values may originate at this point (see Section 3).

The σ_R constants, representing the resonance effect of substituents, may be obtained by routine 4-31G//4-31G calculations. Only the substituent was optimized.

The ethylene skeleton was kept frozen. For evaluation of the expression

$$\sigma_R = 4.167 \sum \Delta q_\pi - 0.06083 \tag{5}$$

we need to calculate Mulliken population analysis for a substituted ethylene $CH_2 = CHX$ and the parent ethylene. The overall π-electron transfer $(\sum \Delta q_\pi)$ between X and the ethylene π system is calculated as a difference between the total π-electron populations on the two carbon atoms in the substituted ethylene and the parent ethylene. From several variants of calculation given by M.T. [5,6,19] we tried to pick up their 4-31G//4-31G data [19] and our calculated constants differed only slightly from the published ones (see Section 3).

In contrast to constants σ_F and σ_R, that may be obtained by common ab initio programs, the constant σ_α, representing the polarizability effect of the substituent X, requires an extra code for the evaluation of the polarization potential [20]

$$PP = \sum_{a \neq 0} \frac{\langle \Psi_a | H' | \Psi_0 \rangle \langle \Psi_0 | H' | \Psi_a \rangle}{E_0 - E_a} \tag{6}$$

In this second-order term of the Møller–Plesset theory the summation is over all singly excited configuration state functions and the Hamiltonian H' stands for Coulombic interactions between a positive charge and the molecules CH_4 and CH_3X. Model **2** [20] defines the location of the positive charge
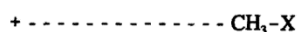
$$+ \text{- - - - - - - - - - - - - } CH_3\text{--}X$$

3 Å

2

The first step in the evaluation of σ_α is the 3-21G geometry optimization of CH_4 and CH_3X. If X contains atoms from higher than the first row of the periodic system, the 3-21G* method is applied. Using the Hartree–Fock molecular orbitals φ and orbital energies ε at the optimum geometry, Eq. (6) may be rewritten as

$$PP = 2 \sum_i \sum_j \frac{\langle \varphi_i | H' | \varphi_j \rangle \langle \varphi_j | H' | \varphi_i \rangle}{\varepsilon_i - \varepsilon_j} \tag{7}$$

The pair of indices i and j represents a single electron excitation $i \rightarrow j$ and the H' matrix elements are

evaluated for structure **2**. The σ_α constant is then expressed as the difference

$$\sigma_\alpha = PP_{CH_3X} - PP_{CH_4} \tag{8}$$

Calculation of σ_α constants met with some problems, presumably because of the use of the 3-21G and 3-21G* basis sets. It is well known that meaningful calculation of polarizability and related properties needs larger basis sets. Although the 3-21G and 3-21G* polarization potentials are reasonable for most of the substituents [18], the 3-21G calculated polarization potential for –COF is completely unsatisfactory and its value was not listed by M.T. [7,18]. The 3-21G and 3-21G* calculations of the polarization potential also failed with the –NHCO(NH$_2$) substituent in the conformation E and with –NHCS(NH$_2$) in the conformation Z. Recalculation with the 6-31G* basis set improves the results dramatically in all three instances. We plan to propose a new set of σ_α constants based on the Sadlej's basis set which was developed especially for calculation of polarization.

The constant σ_χ, representing the electronegativity effect of the substituent X, is obtained from 6-31G*// 6-31G* calculations on the CH_4 and HX molecules. The respective formula is

$$\sigma_\chi = 2(q_{H(HX)} - q_{H(CH_4)}) \tag{9}$$

where the electron populations q_H on hydrogen atoms are obtained from the Mulliken population analysis for the optimum geometries of HX and CH_4. With these constants there were no problems with selecting the right procedure since it was given by M.T. in an unambiguous way [7,18].

3. Results and discussion

The calculated constants σ_F, σ_R, σ_α and σ_χ are listed in Table 1 (standard substituents) and Table 2 (amide and thioamide substituents). In the following discussion we shall deal with two problems. The first question is to what extent are the calculations of constants σ reproducible and well described: it requires comparison of our results with those of M.T. The second question is whether the models used and the constants obtained have a clear physical meaning: it requires comparison with some standard

Table 3
Correlation between calculated and experimental constants σ

Constant	Explanatory variable	Response function	Standard deviation (σ units)	Correlation coefficient	Slope	Number of data
σ_F	calc. [4]	calc. this work	0.050	0.9530	0.793	25
	exp. [21][a]	calc. this work	0.049	0.9607	0.765	20
	exp. [21][a]	calc. [4]	0.040	0.9614	0.939	21
	exp. [21][a]	exp. [22][b]	0.042	0.9770	0.908	22
	exp. [21][a]	exp. [23][c]	0.068	0.9413	0.898	22
	exp. [21][a]	exp. [24][d]	0.173[d]	0.9889	5.221[d]	20
σ_R	calc. [5]	calc. this work	0.031	0.9911	1.003	18
	exp. [25][e]	calc. this work	0.045	0.9814	0.952	17
	exp. [25][e]	calc. [5]	0.035	0.9890	0.978	18
	exp. [23][f]	calc. this work	0.058	0.9696	0.991	15
	exp. [23][f]	calc. [5]	0.054	0.9733	0.989	19
	exp. [23][f]	exp. [25][e]	0.045	0.9851	0.990	18
σ_α	calc. [7]	calc. this work	0.016	0.9977	0.999	21
σ_χ	calc. [8]	calc. this work	0.010	0.9993	0.997	22
	exp. [8,26][g]	σ_{χ^2} this work	0.0423[g]	0.9577	0.00316[g]	18
	exp. [8,26][g]	σ_{χ^2}[8]	0.0424[g]	0.9579	0.00318[g]	18

[a] Defined by dissociation constants of substituted acetic acids in water.

[b] Derived in an indirect way from *meta* and *para* substituent effects in the aromatic series: these effects themselves have been taken in the literature [22] rather uncritically from divergent sources.

[c] Derived from ^{19}F NMR substituent induced shifts in the *meta* position.

[d] Defined by dissociation constants of 4-substituted quinuclidines in water; scaling different from the others.

[e] Derived from IR intensities of the ν_{16a} and ν_{16b} bands in monosubstituted benzenes, the sign may be uncertain.

[f] Derived from ^{19}F NMR shifts in the *meta* and *para* positions.

[g] Coupling constants $^1J_{CC}$ between the positions *ipso* and *ortho* in monosubstituted benzenes, unscaled; standard deviations in σ units; substituents Cl and SX excluded.

experimental quantities. The answers to these uestions differ sharply for the individual types of constants. For this reason, these types will be discussed separately.

3.1. Inductive constants σ_F

These constants, denoted originally σ_1, were recently renamed [4] σ_F without any actual reason. However, the reasoning in favour of "field" effect against the previous "inductive" effect was questioned and an opinion offered that the problem has no significance [16]. Since the numerical values remained unchanged, the nomenclature itself is of no importance for our discussion: we shall use here the symbol σ_F without preferring just the "field" theory.

Generally, our values agree reasonably with those of M.T. [4], see Table 3, first line. Some differences were encountered with more polar substituents $COOCH_3$, CF_3 and NO_2 but the greatest was for SO_2CH_3 (0.22 σ units) which in fact is controlling

the overall fit. In all these cases our σ_F values were smaller than those of M.T. and the difference can be attributed mainly to the more sophisticated procedure in our work (optimized geometry, 6-31G* basis). Particularly, the geometry optimization is known to diminish the molecular dipole moment and this could perhaps result in lowering the calculated σ_F constant.

For comparison with experiments, several large and systematic data sets are available. It has been also pointed out that they essentially agree with each other [22,32]. In Table 3, we refer to four such series in which all items have been obtained in the same simple way from experimental data: we have paid particular care to not include any estimated values. Of these series, the σ_F of Charton [21] have been selected most carefully and are used here as reference. The result of Table 3 is that calculated σ_F constants correspond to experimental data as closely as these experimental data correspond to each other. The deviations are only a little larger than the purely

experimental error, given for instance by Charton [21] between 0.01 and 0.02. A greater deviation was found only for the ^{19}F NMR shifts [23], the pecularities of which have been already mentioned [32]. Comparing our calculations with M.T. [4], it seems that ours fit the experimental data a little worse but this is due again mainly to the substituent SO_2CH_3. If there is no misprint in [4], one must conclude that their calculations are better in practice, i.e. more closely related to the common σ_F determined from measurements in solution. A few measurements in the gas phase [33] may suggest that in the isolated molecule the SO_2CH_3 group behaves as a weaker acceptor than in solution, in agreement with σ_F from Table 1. In our opinion, further experimental evidence is necessary.

In conclusion, the procedure advanced by M.T. [4] may be recommended for calculating missing constants σ_F and a reliability can be attained comparable to the standard experimental approach. The merit of this quantum chemical model is that calculations are easy and the result is not sensitive to computational details. There is no doubt that the resulting σ_F are of importance from the empirical point of view: they can be applied generally to reactivities of all molecules without multiple bonds and without short-range interactions between the substituent and reaction center [1,23]. The only shortcoming is that the physical meaning of the effect is not well understood [16,17] and the quantum chemical model makes use of aggregations of atoms which are not molecules. Its justification is thus only in the empirical agreement with certain experiments.

3.2. Resonance constants σ_R

Contrary to σ_F, the constants σ_R are connected with a definite theoretical concept. The quantum chemical model is immediately related to this theory and based on real molecules of substituted ethylenes [5] or benzenes [5,6]. On the other hand, it is generally accepted that the resonance effect is not constant, not even proportional in different series: it also depends always on the partner group conjugated with the substituent. Empirical relationships have been developed in such a way, that several scales of resonance constants (σ_{R^0}, σ_{R^+}, σ_{R^-}, R^+, R^-) were defined [1,2,22], to be used according to the electron demand or electron supply of the reaction center. (Unfortunately, the

nomenclature has been confused so that the symbols σ_{R^+} and R^+ for the acceptor groups mean either their strengthened conjugation with a donor reaction center [1,22] or weakened conjugation with another acceptor [2].)

We shall deal only with the scale σ_{R^0} which corresponds to conjugation with a mere phenyl group as a standard reference [1,2,22] while the M.T. quantum chemical model [4] uses the vinyl group. The agreement of our results with M.T. is better than for σ_F (Table 3). Small deviations can be explained by geometry optimization in our calculations, the only large deviation was found for the substituent NH_2 (0.09 σ units). Of the experimental data sets, we know only one set based on a single experimental property, namely the intensity of the ν_{16a} and ν_{16b} vibrational bands in monosubstituted benzenes [25]. Their correlation with $(\sigma_{R^0})^2$ is purely empirical: in fact there is no reason why the intensity should be controlled only by resonance and not by the inductive effect. In addition, the sign of σ_{R^0} must be estimated and sometimes can be uncertain. Nevertheless, the close correlation of calculated and experimental values is remarkable (Table 3) and suggests any physical meaning of the two quantities. All other scales of σ_{R^0} are based on two experimental quantities and the resonance constant is obtained by a subtraction eliminating the inductive effect. The simplest such definition is based on ^{19}F NMR shifts of *meta* and *para* substituted fluorobenzenes [23]: the values obtained have been also examined in Table 3 and found to be clearly less correlated both with the calculated σ_{R^0} and the former experimental scale.

The result is that theoretically calculated σ_{R^0} are of a similar reliability as those based on experiments. While this statement is equal as in the case of σ_F, we must stress that σ_{R^0} are much more restricted in their scope. They are applicable only in conjugation with a partner which is neither a donor nor acceptor. In addition, there are strong doubts whether σ_{R^0} for acceptors and donors have the same meaning or the same scaling: in many examples σ_{R^0} of acceptors appear to be essentially zero [2,33].

3.3. Polarizability constants σ_α

The quantum chemical model for evaluating these constants [7,18] is most sophisticated and not

completely satisfactory from the theoretical point of view (see Section 2). Nevertheless, we followed it closely and obtained a very good fit with the literature values (Table 3). The only deviation, 0.06 σ units for the substituent CN, could be even due to an error. For the substituent COF, not given in Ref. [7], the procedure failed: the apparent value obtained was far off the acceptable range.

Any efficient comparison with experimental values is not possible. Several rather convincing correlations with the gas-phase basicities and acidities were presented but only for alkyl substituents [7]. In the case of polar substituents, the polarizability effect is accompanied or even overweighted by the inductive effect: their separation was attempted [34] with the assumption of equal inductive effects in anions and cations but this is evidently only an approximation and was criticized [35]. More general are linear regressions with σ_F and σ_α as explanatory variables, successful for several gas-phase reactions [2], but they can be hardly used for checking the σ_α values.

In conclusion, the model for calculating the σ_α values is well designed in principle but not satisfactory as to the computational technique which would deserve a reinvestigation. In any case, calculations are for the time being the only source of σ_α values: direct comparison to experiments is possible only for alkyl groups.

3.4. Electronegativity constants σ_χ

The concept of electronegativity is one of the most controversial [1,8,18], even its general utility can be doubted. The quantum chemical model originally suggested [8] was later improved [18] with an ad hoc correction which uses different scaling for all carbon substituents. When we repeated the calculations with this correction, we obtained identical results (Table 3). The only deviation for the substituent CF_3 is probably due to a misprint (0.02 instead of –0.02) and is immaterial in the whole context. Comparison with experiments is a difficult problem. The only quantity claimed [8] to correlate directly is the NMR coupling constant $^1J_{CC}$ between the positions *ipso* and *ortho* in monosubstituted benzenes [8,28] which should be proportional to $(\sigma_\chi)^2$. Substituents bonded with second-row elements must be excluded. The results in Table 3 might appear sufficient according to the correlation coefficient but the standard deviations

given for $(\sigma_\chi)^2$ are unacceptable for small values: some predicted values would be even imaginary.

We conclude that M.T. calculations [8,18] represent the only reliable source of σ_χ constants, but their possible application is not evident.

3.5. Dependence of σ constants on conformation

Certain more complex substituents can exist in different conformations, see for instance Table 2. When their σ constants are determined in the usual way, they represent a weighted mean over all conformations present: separate constants for individual conformations may be accessible in special cases by means of fast spectroscopic methods. In ^{19}F NMR spectra of *meta*- and *para*-fluoroformanilides the two conformers were distinguished and constants σ_F and σ_R calculated [15]. Necessary conditions for such proceeding are separation of the signals, close dependence on σ, and reasonable abundance of rotamers. This is all seldom fulfilled. For instance, the NH frequencies in the conformers of substituted thiobenzanilides were distinguished and assigned [36], but the sensitivity to substitution was too small in relation to the experimental accuracy.

Data in Table 2 give six examples of substituents which are each in two planar conformations. Differences in σ_F and σ_α are on average 0.3, in σ_R and in σ_χ, less than 0.1. A rough comparison with the above experimental data [15] is possible for the substituents $-NHCXCH_3$ and $-NHCXNH_2$ ($X = O, S$). Calculations reproduce correctly the difference between rotamers with the right sign but overestimate its absolute value (with σ_F by a factor of six, with σ_R, say, by three, but these values are too small). The comparison is only speculative since the substituents are not identical. Remarkable is the great dependence of σ_α on conformation (one value is even positive) but an experimental verification is not possible to date.

In real molecules containing the substituents from Table 2, the conformation Z predominates. Experimental constants σ, determined by common procedures, can be compared to the Z values with a good approximation. Experimental data given in Table 2 are not numerous but reveal a general agreement with the calculations. For practical purposes, say for QSAR, the σ constants of these substituents can be calculated with the same reliability as for the others.

4. Conclusions

By recalculation at a larger basis or with geometry optimization, the resulting σ have changed very little with few exceptions. From comparison with experimental series it follows that the M.T. models [4,5] for σ_F and σ_R are well sounded and the calculated σ constants are generally as reliable as the experimental values. Particularly for the purpose of QSAR, the calculated σ constants can be recommended as sufficient. This applies particularly to the thioamide substituents of Table 2. In the case of σ_α the calculation [7] represents the only reliable source, but the calculation model should be improved. In the case of σ_χ, the calculations [8] are well reproducible, but for the time being we do not see any good reason why these constants should be calculated.

Acknowledgements

References

[1] O. Exner, Correlation Analysis of Chemical Data, Plenum, New York, 1988, Chapter 5.

[2] R.W. Taft and R.D. Topsom, Prog. Phys. Org. Chem., 16 (1987) 1.

[3] M. Ludwig, S. Wold and O. Exner, Acta Chem. Scand., 46 (1992) 549.

[4] S. Marriott and R.D. Topsom, J. Am. Chem. Soc., 106 (1984) 7.

[5] S. Marriott and R.D. Topsom, J. Chem. Soc., Perkin Trans. 2, (1985) 1045.

[6] S. Marriott, A. Silvestro and R.D. Topsom, J. Chem. Soc., Perkin Trans. 2, (1988) 457.

[7] W.J. Hehre, C.-F. Pau, A.D. Headley, R.W. Taft and R.D. Topsom, J. Am. Chem. Soc., 108 (1986) 1711.

[8] S. Marriott, W.F. Reynolds, R.W. Taft and R.D. Topsom, J. Org. Chem., 49 (1984) 959.

[9] K. Waisser, J. Dršata, F. Bartoš and K. Kosař, in M. Tichý (Ed.), QSAR in Toxicology and Xenobiochemistry, Elsevier, Amsterdam, 1985, p. 91.

[10] K. Waisser, J. Kuneš, M. Roman, M. Buděšinskyand O, Exner, Phosphorus, Sulfur, Silicon, 97 (1994) 71.

[11] K. Waisser, M. Macháček, J. Lebvowa, J. Hrbata and J. Dršata, Colln. Czech. Chem. Commun., 53 (1988) 2957.

[12] J. Gabriel, I. Němec, N. Zimová and K. Waisser, Chem. Papers, 43 (1989) 389.

[13] K. Waisser and M. Polášek, Folia Pharm. Univ. Carol., 18 (1995) 45.

[14] M. Polášek, K. Waisser and T. Bouček, Colln. Czech. Chem. Commun., 56 (1991) 2964.

[15] Z. Arnold, V. Krchňák and P. Trška, Tetrahedron Lett., (1975) 347.

[16] O. Exner and Z. Friedl, Prog. Phys. Org. Chem., 19 (1993) 259.

[17] K. Bowden and E.J. Grubbs, Prog. Phys. Org. Chem., 19 (1993) 183.

[18] R.D. Topsom, Prog. Phys. Org. Chem., 16 (1987) 125.

[19] S. Marriott and R.D. Topsom, J. Mol. Struct. (Theochem), 106 (1984), 227.

[20] S.D. Kahn, C.F. Pau, L.E. Overman and W.J. Hehre, J. Am. Chem. Soc., 108 (1986) 7381.

[21] M. Charton, Prog. Phys. Org. Chem., 13 (1981) 119.

[22] C. Hansch, A. Leo and R.W. Taft, Chem. Rev., 91 (1991) 165.

[23] R.W. Taft, E. Price, I.R. Fox, I.C. Lewis, K.K. Andersen and G.T. Davis, J. Am. Chem. Soc., 85 (1963) 709, 3146.

[24] C.A. Grob and M.G. Schlageter, Helv. Chim. Acta, 59 (1976) 264.

[25] R.T.C. Brownlee, R.E.J. Hutchinson, A.R. Katritzky, T.T. Tidwell and R.D. Topsom, J. Am. Chem. Soc., 90 (1968) 1757.

[26] V. Wray, L. Ernst, T. Lund and H.J. Jakobsen, J. Magn. Reson., 40 (1980) 55.

[27] R.T.C. Brownlee and M. Sadek, Aust. J. Chem., 34 (1981) 1593.

[28] T. Nishiguchi and Y. Iwakura, J. Org. Chem., 35 (1970) 1591.

[29] J.C. Kauer and W.A. Sheppard, J. Org. Chem., 32 (1967) 3580.

[30] O. Exner and J. Lakomý, Colln. Czech. Chem. Commun., 35 (1970) 1371.

[31] X. Creary and T. Aldridge, J. Org. Chem., 56 (1991) 4280.

[32] O. Exner, Org. Reactivity, 21 (1984) 3.

[33] M. Decouzon, O. Exner, J.-F. Gal and P.-C. Maria, J. Phys. Org. Chem., 7 (1994) 615.

[34] R.W. Taft, M. Taagepera, J.L.M. Abboud, J.F. Wolf, D.J. DeFrees, W.J. Hehre, J.E. Bartmess and R.T. Melver, J. Am. Chem. Soc., 100 (1978) 7765.

[35] M. Charton, in R.I. Zalewski, T.M. Krygowski and J. Shorter (Eds.), Similarity Models in Organic Chemistry, Biochemistry and Related Fields, Elsevier, Amsterdam, 1991, p. 629.

[36] W. Walter and Kubersky, J. Mol. Struct., 11 (1972) 207.

P20. The C4H6+ potential energy surface. 2. The reaction of ethylene radical cation with acetylene

V. Hrouda, P. Čársky, M. Ingr, Z. Chval, G.N. Sastry, T. Bally, Journal of Physical Chemistry A. 102 (1998) 9297–9307.

The $C_4H_6^{\bullet+}$ Potential Energy Surface. 2. The Reaction of Ethylene Radical Cation with Acetylene

Vojtěch Hrouda, Petr Čársky,* Marek Ingr, and Zdenek Chval[†]

J. Heyrovský Institute of Physical Chemistry, Academy of Sciences of the Czech Republic, Dolejškova 3, 182 23 Prague 8, Czech Republic

G. Narahari Sastry[‡] and Thomas Bally*

*Institut de Chimie Physique, Université de Fribourg, Pérolles, 1700 Fribourg, Switzerland*

Received: June 11, 1998; In Final Form: August 24, 1998

The reaction of the ethylene radical cation (Et[•+]) with acetylene (Ac) to form stable $C_4H_6^{\bullet+}$ intermediates and the subsequent fragmentation of these to $C_3H_3^+ + CH_3^\bullet$ or to $C_4H_5^+ + H^\bullet$ have been studied by the UMP2, RMP2, and B3LYP methods with the 6-31G* basis set, as well as by single-point calculations at the RCCSD-(T)/cc-pVTZ level of theory. The aim of this study was to identify all stationary points that might be relevant to explain the course of the observed reactions. According to their stability to dissociation, we distinguish three classes of $C_4H_6^{\bullet+}$ structures: weakly bonded complexes, structures of medium stability, and tightly bonded complexes. Methylcyclopropene radical cation seems to be the most likely ultimate precursor for the formation of the observed fragmentation products.

1. Introduction

The knowledge of potential energy surfaces (PESs) of polyatomic ions is of basic importance for interpreting experiments both on the bimolecular reaction dynamics of ion−molecule processes and on the dynamics of unimolecular decomposition of polyatomic state-selected ions. In the former, scattering studies of bimolecular ion−molecule processes often reveal the formation of intermediate complexes. Angular distributions of products may be related to the structures and dynamics of the dissociating intermediate. In the latter case, the energetics and the decomposition of ions prepared from parent molecules are investigated by the photoelectron−photoion coincidence (PEPICO) method; "experimental PESs" (energetics of minima and barriers) obtained from the resulting data can be greatly refined by theoretical calculations. C_4H_6 has been subjected to both types of experimental investigations.[1]

This study has been stimulated by crossed-beam scattering experiments on the dynamics of formation of various products in the system

$$C_2H_2^{\bullet+} + C_2H_4 \rightarrow [C_4H_6^{\bullet+}] \rightarrow$$
$$C_3H_3^+, C_4H_5^+, C_2H_4^{\bullet+}, C_2H_2^{\bullet+} \quad (1)$$
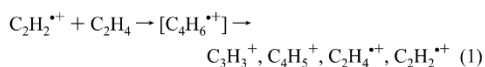
studied by Herman and his collaborators.[2] Analogous products were observed for the "charge inversed" reaction

$$C_2H_4^{\bullet+} + C_2H_2 \rightarrow [C_4H_6^{\bullet+}] \rightarrow C_3H_3^+, C_4H_5^+ \quad (2)$$
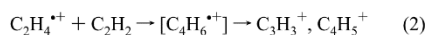
The reactions in both systems appear to proceed through common intermediates on the way to form the products. From scattering diagrams and from the angular and product transla-tional energy distributions, the formation of an intermediate

---

[†] Present address: Faculty of Biology, University of South Bohemia, Česke Budejovice, Czech Republic.
[‡] Present address: Department of Chemistry, University of Pondicherry, 605014 Pondicherry, India.

$C_4H_6^{\bullet+}$ with a mean lifetime > 5 $\times$ 10^{-12} was inferred.[2,3] For the reaction channels leading to $C_3H_3^+ + CH_3^\bullet$ and to $C_4H_5^+ + H^\bullet$, angular distributions indicated[2] that the critical config-uration of the dissociating intermediate is consistent with the structure of methylcyclopropene radical cation (MCPE).

Obviously, ab initio calculations may be very helpful in this case by yielding additional evidence on the structure and energy of intermediates and transition states, especially on the way to various dissociation products. By its size, the $C_4H_6^{\bullet+}$ system is amenable to treatments by levels of the theory that provide results of sufficient accuracy for this purpose. We decided, therefore, to undertake an ab initio study on the $C_4H_6^{\bullet+}$ potential energy surface.

Most previous studies on $C_4H_6^{\bullet+}$ focused on the ring opening of the methylenecyclopropane radical cation (MCPA)[4] or the cyclobutene radical cation (CB).[5,6] Our own results on the latter reaction were presented separately,[7] whereas the present paper is intended to focus on the pathways for $C_4H_6^{\bullet+}$ association and fragmentation processes. Before the completion of this study, another paper with a similar scope was published by Keister et al.[8] They reported on the kinetics and mechanism of methyl loss from 1,3-butadiene (BD) and MCPE studied by threshold photoelectron−photoion coincidence time-of-flight mass spectrometry (TPEPICO), ab initio UMP2/6-311G** calculations, and RRKM statistical theory. In earlier papers, Baer and collaborators[1,9,10] had described photoionization studies involving other C_4H_6 isomers, 1,2-butadiene (methylallene, MA), 1-butyne (BTY), 2-butyne (dimethylacetylene, DA), and cyclobutene (CB). They concluded that the corresponding radical cations formed by photoionizaton rearranged to a common precursor prior to dissociation to $C_3H_3^+ + CH_3^\bullet$ and that the transition states for isomerizations among the different $C_4H_6^{\bullet+}$ species lie below the dissociation limit. Preuninger and Farrar[11] arrived at the same conclusion in their photofragmen-tation study of BD, MA, CB, and MCPE. The absence of photodissociation products from BTY and DA, reported by Preuninger and Farrar,[11] was explained later by Bunn and Baer.[10]

Dissociation of the BD radical cation has also been investigated by Dannacher et al.[12,13] and by Russell et al.[14,15]

The most comprehensive study on the $C_4H_6^{\bullet+}$ PES by ab initio calculations is contained in the above-mentioned recent paper by Keister et al.[8] (other work concerned mainly the ring opening of $CB^{\cdot+}$, which is treated in detail in our other paper[7]). However, this study was based entirely on the UMP2 methodology, which is known to be prone to artifacts due to UHF spin contamination,[16] especially also in the case of weakly bound ion–molecule complexes.[17] We therefore decided to pursue our investigation of the $C_4H_6^{\bullet+}$ PES using a wider range of computational methods, hoping that this would uncover and eliminate possible artifacts of the previously used methods, such as they arise by spin contamination or symmetry breaking. Although we strived to arrive at a comprehensive picture of the $C_4H_6^{\bullet+}$ PES (insofar as it is relevant with regard to the mechanism of processes 1 and 2 above), no claim of completeness is made.

2. Computational Methods

All structures were optimized by the standard UHF, UMP2, and B3LYP methods, as implemented in the Gaussian 94 program.[18] In cases of ROHF convergence problems, optimizations were carried out with Gamess/US.[19] Stationary points were characterized by harmonic frequency calculations at all the above levels. From the available variants of the restricted open-shell MP2 method,[20] we selected that[21,22] implemented, along with analytical gradients, in the Cadpac[23] and Aces2 codes.[24] All geometry optimizations were carried out with the standard 6-31G* basis set,[25] and in the MP2 calculations, all electrons were correlated. At the optimized UMP2 and B3LYP geometries, we also performed single-point RCCSD(T) calculations, by the method of Knowles et al.[26] as implemented in the Molpro program package.[27] These were done with Dunning's correlation-consistent triple-ζ (cc-pVTZ) basis set,[28] which can be expected to give accurate energies at this level of theory.

In addition to Hessian calculations, all transition-state structures were tested by intrinsic reaction coordinate (IRC) calculations[29] to identify the minima they interconnect. Differences in zero-point energies (ΔZPE) for the evaluation of $\Delta E_0 = \Delta E + \Delta$ZPE were taken from frequencies calculated at the level used for geometry optimization, except for the cases of RMP2 and RCCSD(T)/cc-pVTZ energies, which were corrected on the basis of B3LYP frequencies, if the respective stationary point existed, and by UMP2 frequencies in other cases.

As with any calculation on processes of the type A + B → AB, where AB is a loosely bonded complex, incompleteness of the basis set forces one to consider the effects of basis set superposition error (BSSE). For some relevant cases, we calculated this error at the ROHF and RMP2 levels by a variant of the Boys–Bernardi method,[30] which takes into account the bond relaxation energy.[31] For the distonic $^+CHCHCH_2CH_2^{\bullet}$ radical cation, which may also be viewed as a $(C_2H_2\cdots C_2H_4)^{\bullet+}$ complex cation, the BSSE is 1.2 and 5.5 kcal/mol at the ROHF and RMP2 levels, respectively. For the $(C_3H_3^+\cdots CH_3^{\bullet})$ complex, the corresponding numbers are 0.6 and 1.0 kcal/mol. We decided, therefore, not to correct the energies of loosely bonded complexes for BSSE, because it brings about only a minor change in the overall energy diagram and does not alter any of our conclusions with regard to the reaction path of processes 1 and 2.

To show that the single-determinant reference wave functions are adequate to provide a correct description of the molecular characteristics, CASSCF single-point calculations using seven active electrons in eight orbitals were carried out on all UMP2 stationary points. In all cases, the ground-state reference determinant contributed $\geq$ 90% (typically by 93%) and no individual excited state contributed more than 3.9% (typically $\leq$ 2.0%) to the CASSCF wave function. The expectation value $\langle S^2 \rangle$ in the unrestricted Hartree–Fock (UHF) determinants was always less than 0.78 in the B3LYP calculations, whereas in the UHF reference wave function used for UMP2 it showed significant deviations from 0.75, especially at transition states, where it sometimes exceeded 0.95.

3. Results and Discussion

Before going into details below, we should mention that the energetics of all our calculations will refer to the reaction of the ethylene radical cation ($Et^{\bullet+}$) with acetylene (Ac), although the original experiments[2] were done with $Ac^{\bullet+} + Et$, which lie 21 kcal/mol higher in energy.[32] However, this only affects the total energy of the $C_4H_6^{\bullet+}$ system in these experiments and not the shape of the underlying PES. Under conditions of excess energy dissipation, charge transfer from $Ac^{\bullet+}$ to Et would occur before anything else, thus bringing the system to the starting point of our calculations.

3.1. Weakly Bonded Complexes. Building on our previous experience with the Et + $Et^{\bullet+}$ reaction[17] and the Ac + $Ac^{\bullet+}$ reaction,[33] we engaged in a systematic characterization of four possible types of weakly bonded complexes. At first sight, a planar 2s + 2s type complex of C_{2v} symmetry (PlC) would seem to constitute a favorable bonding arrangement as it provides optimal overlap between the π MOs of the two constituent species. Although this results in a bonding of 17 kcal/mol (see Table 1), it turns out not to be optimal, as in the previous cases of $(Et\cdots Et)^{\bullet+}$ and $(Ac\cdots Ac)^{\bullet+}$, presumably because steric repulsion prohibits a sufficiently close approach.

The PlC structure turns out to be a transition state[34] interconnecting two automeric perpendicular complexes of C_{2v} symmetry (PpC), similar to the situation found in $(Ac)_2^{\bullet+}$.[33] The perpendicular conformation seems to allow for better bonding, as PpC lies about 3 kcal/mol below PlC (see Table 1). At the SCF levels, PpC is a saddle point (-115 cm^{-1} at ROHF, -127 cm^{-1} at UHF) which, according to IRC calculations, represents a transition state for the interconversion of two three-membered-ring complexes of C_s symmetry (TC) where one end of the ethylene moiety bonds to the π system of acetylene.

At B3LYP, this stationary point has *two* imaginary modes, one of which (-96 cm^{-1}) is also associated with TC automerization, whereas the other one (-95 cm^{-1}) leads to an alternative three-membered-ring complex where one acetylene terminus binds to the ethylene π system. However, bonding in this latter species, which may be viewed as a cyclopropylcarbene radical cation (CC), is much stronger, and we will, therefore, return to this pivotal intermediate in the following section on tightly bound structures.

By UMP2, PpC is a minimum, but this is so shallow (separated from TC by a barrier of only 0.02 kcal/mol, associated with a transition state TS1 with an imaginary frequency of -88 cm^{-1}) that this should probably be regarded as an artifact. The vibrational mode connecting PpC with CC is also positive at UMP2 and thus gives rise to another transition state, TS2. This lies, however, 5 kcal/mol above PpC and turns out to be a second-order saddle point[35] whose other imaginary mode (-77 cm^{-1}) leads to the "linear complex" discussed below.

TC lies 0.5–1 kcal/mol lower in energy than PpC at all levels. By ROHF and UMP2, TC is a minimum, whereas UHF and

C$_4$H$_6$$^{\bullet+}$ Potential Energy Surface

J. Phys. Chem. A, Vol. 102, No. 46, 1998 **9299**

TABLE 1: Energies and 0 K Enthalpies (kcal/mol) Relative to Ac + Et$^{\bullet+}$ of Weakly Bonded C$_4$H$_6$$^{\bullet+}$ Complexes and Associated Transition Structures

structure[a]	B3LYP/6-31G*		RMP2/6-31G*[b]		UMP2/6-31G*		RCCSD(T)/cc-pVTZ[c]	
	ΔE	ΔE_0	ΔE	ΔE_0	ΔE	ΔE_0	ΔE	ΔE_0
PlC[d]	−25.32	−23.46	−19.05	−15.28	−18.46	−14.69	−16.67	−14.81
PpC[d]	−27.21	−25.19	−22.22	−18.49	−21.58	−17.85	−19.55	−17.53
TC[e]	−27.54	−25.32	−22.54	−19.09	−21.94	−18.49	−20.53	−18.30
LC	f		−29.58	−23.54	−21.15	−15.11	−23.17	−17.13
TS1: PpC → TC	g		g		−21.56	−17.98	−20.57	−16.99
TS2: PpC → CC	g		g		−16.34	−11.83	−21.88	−17.37
TS3: LC → CC	g		g		−20.68	−14.38	−24.07	−17.77
TS4: TC → LC	g		g		−20.57	−16.64	−21.30	−17.37
CC[j]	−41.06	−36.15	−40.69	−34.08	−39.36	−32.75	−37.04	−32.13
Ac$^{\bullet+}$ + Et	19.93	20.87	22.31	23.57	22.32	23.58	19.22	20.16
Ac + Et$^{\bullet+}$								
E^h	−155.541 20		−155.002 72		−155.002 72		−155.242 15	
ZPE^i	47.55		47.73		47.73		47.55	

^{a} For UMP2 and B3LYP optimized structures, see Figure 2; structures TSx are transition states at UMP2 and B3LYP; all other species are minima, except where indicated otherwise. ^{b} For RMP2, vibrational analysis was not performed; ΔZPE taken from vibrational analysis at UMP2 level. ^{c} Single-point calculations at B3LYP (when these are available) or UMP2 geometries; ΔZPE taken from B3LYP and UMP2 vibrational analysis, respectively. ^{d} Transition state at UMP2, second-order saddle point at B3LYP. ^{e} Transition state at B3LYP. ^{f} Not found at B3LYP. ^{g} Not found at B3LYP and RMP2. ^{h} Total energies in hartrees. ^{i} Zero point energies in kcal/mol.
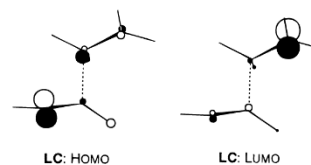
SCHEME 1: Connection of Loosely Bound Complexes, PpC and TC, to CC on the B3LYP Potential Energy Surface

SCHEME 2: Connection of Loosely Bound Complexes, PpC, TC, and LC, to CC on the UMP2 Potential Energy Surface^a
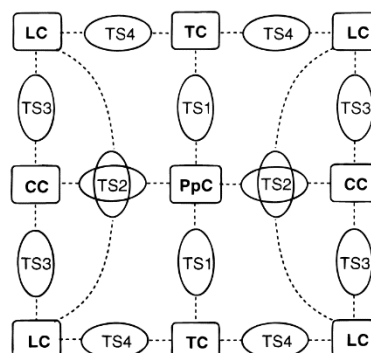
^{a} Structures, see Figure 2. Squares represent potential energy minima, ovals saddle points (crossed ovals: second-order saddle points).

Figure 1. Frontier molecular orbitals of complex LC.

B3LYP IRC calculations show it to be a transition state for CC automerization. This completes the picture at the B3LYP level (cf. Scheme 1), but all other methods predict the existence of another loosely bound species on the C$_4$H$_6$$^{\bullet+}$ PES, the above-mentioned linear complex, $^{\bullet}$CH$_2$CH$_2$CHCH$^+$ (LC),[36] which is of the type which corresponds to the most stable complexes in (Et)$_2$$^{\bullet+}$[17] and (Ac)$_2$$^{\bullet+}$.[33]

LC is more strongly bound than TC at the RMP2[37] and RCCSD(T) levels, but at UMP2, this species is penalized by the comparatively high degree of spin contamination of the UHF wave function ($\mathscr{S}^2 = 0.805$). Nevertheless, UMP2 predicts transition states TS3 (−476 cm^{-1}) and TS4 (−169 cm^{-1}), which separate LC from CC and TC, respectively (cf. Scheme 2).[38] However, on inclusion of zero-point energies, TS1 disappears, and at the RCCSD(T) level, the activation energies associated with all four transition states discussed above disappear. Thus, the minima predicted by UMP2 seem to be artifacts created by the tendency of transition states to have higher UHF spin

contamination than equilibrium structures. Consequently, Scheme 1 probably represents a PES that is closer to reality than Scheme 2.

As in (Et)$_2$$^{\bullet+}$ and (Ac)$_2$$^{\bullet+}$, LC adopts a twisted geometry (dihedral angle of 116°),[39] but in contrast to the above symmetric complexes, the inherent dissymmetry of the (Et⋯Ac)$^{\bullet+}$ complex results in a localization of spin and charge on opposite ends of the molecule. This feature expresses itself clearly in the shape of the HOMO and LUMO shown in Figure 1, which demonstrates also that the HOMO no longer has an antibonding component along the central bond so that this can assume a

shorter length. This situation is reminiscent of the species LC2 in (Ac)$_2$$^{•+}$, which lies, however, higher in energy than the corresponding delocalized state, LC1.[33]

Interestingly, the charge in LC is localized on the acetylene moiety, although the ethyl radical (IP = 8.12 eV)[40] is easier to ionize that the vinyl radical (IP = 8.25 eV).[41] The reason for this apparent contradiction is that the ethyl cation is substantially stabilized by hyperconjugation with the methyl group (CH$_3$$^•$ is 1.7 eV harder to ionize[42] than H$_3$CCH$_2$$^•$); whereas the FMOs in Figure 1 indicate that this stabilizing interaction is virtually absent in LC. In view of this, one should compare the ionization energy of CH$_3$$^•$ to that of the vinyl radical (1.6 eV in favor of the latter) to explain the distribution of spin and charge in LC. In fact, CASSCF(7,8) calculations (at the UMP2 geometry) predict that the state with exchanged spin and charge lies 54 kcal/mol above the ground state of LC. However, the separation of spin and charge in the excited state is less distinct than in the ground state.

A methodologically interesting observation on LC is, that it does not exist as a stationary point on the B3LYP surface (any attempt to locate a stationary point corresponding to LC leads to spontaneous collapse to CC). We had observed previously the reluctance of DFT models to separate spin and charge in radical cations,[33,43] and the same feature appears to be responsible for the inability of B3LYP to model LC: inspection of the spin and charge distribution from a B3LYP calculation (at the UMP2 geometry) shows that both are nearly fully delocalized, a fact which expresses itself in the shape of the B3LYP Kohn−Sham HOMO of LC. Thus, the localization of spin and charge which seems to be required in the present case to form a minimum on the PES cannot be modeled with DFT.

Summing up our results on weakly bonded (Et···Ac)$^{•+}$ complexes in Figure 2 and Table 1, we note that none of them seems to represent a stable species. In (Et)$_2$$^{•+}$ and (Ac)$_2$$^{•+}$, the linear complexes were potential energy minima, but the separation of spin and charge in the corresponding (Et···Ac)$^{•+}$ species leads to the disappearance of the barriers separating it from more tightly bound species.

3.2. Tightly Bound Structures. *3.2.1. Spontaneous Relaxation to CB.* As we have shown above, in the vicinity of the weakly bonded complexes, the surface of the potential energy is relatively flat and only feebly structured with shallow minima and low-lying transition states. Consequently, the number and nature of these stationary points varies for different methods. Of the various channels connecting the weak complexes with tightly bound C$_4$H$_6$$^{•+}$ species, the one leading to CC turns out to represent by far the most favorable pathway, with an activation barrier of less than 1 kcal/mol (see Table 1 and Scheme 1). This situation is closely analogous to that encountered in C$_4$H$_4$$^{•+}$,[33] where we also found a nearly activationless collapse of a linear complex to a CC-type structure.

As in C$_4$H$_4$$^{•+}$, CC represents a shallow minimum of C_s symmetry with spin and charge located at the formally divalent carbon atom at UMP2 and B3LYP. At the SCF levels, the C_s form of CC is a very flat saddle point, with negative frequencies of −34 and −134 cm^{-1}, respectively, connecting two slightly distorted automeric minima of C_1 symmetry. However, this distortion probably represents a case of artificial symmetry breaking that disappears when correlation is included and is therefore not relevant for practical reasons.[44] However, CC must also be considered as a fleeting intermediate, because it collapses to CB via transition state TS5 (Figure 3) with a UMP2 barrier of only 0.6 kcal/mol, which furthermore disappears at the
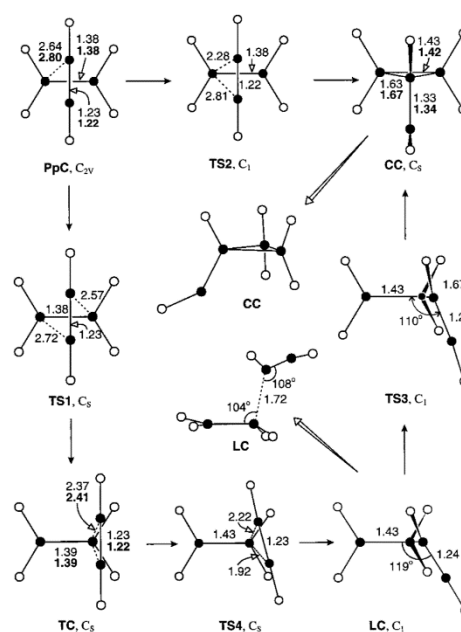

Figure 2. Interconnection and structures of stationary points in the loosely bound region of the [C$_2$H$_4$···C$_2$H$_2$]$^{•+}$ potential energy surface by UMP2 (normal face) and B3LYP (bold face). Bond lengths in Å, angles in deg. Center: alternate views of LC and CC.
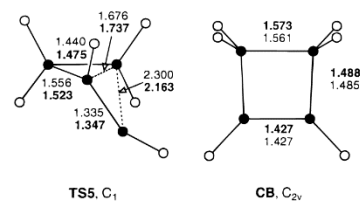

Figure 3. Structures of CB and the transition state, TS5, connecting it to CC (normal, UMP2; bold, B3LYP; bond lengths in Å, angles in deg).

CCSD(T)/cc-pVTZ level (cf. Table 2). Hence, as in C$_4$H$_4$$^{•+}$, there is no relevant obstacle for the activationless collapse of Et$^{•+}$ + Ac to the four-membered-ring structure, CB.

The rearrangements among the stable CB valence isomers are described in some detail in a separate publication.[7] However, to put the subsequently discussed H shifts into perspective, we recall three important transition states, TS6 and TS7, for the conversion of CB to the radical cations of *cis*- and *trans*-BD, respectively, and TS8 for interconversion of the two BD rotamers (for pictorial representations, see ref 7). Finally, we should also mention that for stable C$_4$H$_6$$^{•+}$ isomers as well as for some fragmentation products discussed in this study, the experimental enthalpy differences to Et$^{•+}$ + Ac are available, and they are in very good agreement with the calculated RCCSD(T) ΔH values (see Table 3).

3.2.2. Alternative Rearrangement Paths of Weakly Bonded Complexes. We have found two other reaction paths leading from the weakly bonded complexes to stable C$_4$H$_6$$^{•+}$ isomers. The first of these represents a [1,3] hydrogen shift in LC, which

259

C$_4$H$_6$$^{\bullet+}$ Potential Energy Surface

J. Phys. Chem. A, Vol. 102, No. 46, 1998 **9301**

TABLE 2: Energies and 0 K Enthalpies (in kcal/mol) Relative to Ac + Et$^{\bullet+}$ of Tightly Bonded C$_4$H$_6$$^{\bullet+}$ Structures

structure[a]	B3LYP/6-31G*		RMP2/6-31G*[b]		UMP2/6-31G*		RCCSD(T)/cc-pVTZ[c]	
	ΔE	ΔE_0	ΔE	ΔE_0	ΔE	ΔE_0	ΔE	ΔE_0
BU1	−39.02	−37.52	−29.09	−26.87	−28.83	−26.61	−30.61	−29.10
MCPE	−41.16	−38.22	−40.12	−35.89	−37.64	−33.42	−34.83	−31.89
BTY	−44.38	−40.12	−37.93	−31.67	−37.26	−31.00	−37.22	−32.96
CC	−41.06	−36.15	−40.69	−34.08	−39.36	−32.75	−37.04	−32.13
PRC	−50.96	−46.66	−44.26	−38.70	−39.98	−34.42	−43.05	−38.74
BU2	−49.68	−46.25	−44.14	−38.64	−41.60	−36.09	−43.22	−39.79
MCPA	−55.17	−51.00	−47.40	−42.14	−47.36	−42.11	−45.68	−41.50
CB	−63.08	−58.16	−57.76	−51.76	−57.60	−51.60	−55.49	−50.58
MA	−69.97	−66.14	−60.24	−54.34	−56.90	−51.00	−59.52	−55.69
cis-BD	−84.05	−78.33	−78.57	−71.84	−72.45	−65.72	−74.14	−68.41
trans-BD	−87.72	−81.98	−82.59	−75.86	−76.67	−69.94	−77.62	−71.88
TS5: CC → CB	−39.99	−34.90	−40.95	−34.43	−38.77	−32.25	−37.23	−32.15
TS6: CB → cis-BD	−44.20	−40.08	d		−35.86	−30.16	−37.24	−33.12
TS7: CB → trans-BD	−40.89	−37.36	d		−27.82	−16.30	−36.30	−32.76
TS8: cis-BD → trans-BD	e		−49.55	−41.72	−49.06	−41.23	−49.28	−41.45
TS9: LC → trans-BD	e		−11.67	−9.13	−11.44	−8.90	−12.78	−10.24
TS10: BU1 → LC	e		−11.43	−9.21	−11.39	−9.16	−15.22	−12.99
TS11: BU1 → cis-BD	−36.96	−35.50	−28.75	−25.47	−27.90	−24.62	−31.73	−30.27
TS12: BU1 → BU2	−34.99	−34.03	d		−27.70	−25.93	−29.29	−28.32
TS13: BU2 → MCPA	−49.65	−46.34	d		−41.39	−36.11	−43.15	−39.84
TS14: BU1 → MA	−34.89	−33.33	d		−19.87	−16.49	−25.62	−24.06
TS15: MA → trans-BD	−26.20	−24.04	d		−14.39	−10.75	−19.57	−17.41
TS16: CC → BTY	−8.97	−6.04	−6.20	−1.22	−1.00	−3.98	−7.63	−4.69
TS17: CC → MCPE	−14.92	−12.58	−10.62	−7.00	−10.55	−6.94	−10.18	−7.84
TS18: cis-BD → PRC	−31.87	−30.38	−24.44	−21.07	−19.01	−15.65	−25.83	−24.00
TS19: PRC → MCPE	−40.15	−37.30	−38.79	−34.33	−31.46	−27.00	−34.88	−32.03

[a] For UMP2 and B3LYP optimized structures, see Figures 3−8; structures TS*x* are transition states; all other species are minima, unless indicated otherwise. [b,c] See footnotes in Table 1. [d] Not found at RMP2. [e] Not found at B3LYP.

TABLE 3: Experimental and Calculated 298 K Enthalpies (kcal/mol) of C$_4$H$_6$$^{\bullet+}$ Isomers Relative to Et$^{\bullet+}$ + Ac[a]

	trans-BD	CB	MCPA	MCPE	C$_4$H$_5$$^+$ + H$^\bullet$	cyclopropenium + CH$_3$$^\bullet$	propargyl + CH$_3$$^\bullet$
expt	−73.7 ± 0.3[b]	−54.1 ± 1.0[c]	−40.0 ± 2.5[d]	−33.5 ± 1.7[e]	−20.0[f]	−18.9[g]	+7.0 ± 1.7[h]
calcd[i]	−73.3	−52.3	−43.0	−33.1	−21.0	−17.9	+9.4

[a] Experimental ΔH_f for isolated fragments (Ac and Et$^{\bullet+}$) is 309.1 kcal/mol.[45] [b] From ΔH_f = 26.0 ± 0.2 kcal/mol[46] and $I_v = I_a = 9.082 ± 0.004$ eV.[47] [c] From ΔH_f = 37.5 ± 0.4 kcal/mol[48] and $I_{v,1}$ = 9.43 ± 0.02 eV.[49] [d] From ΔH_f = 48.0 ± 0.4 kcal/mol[48] and $I_{v,1}$ = 9.6 ± 0.1 eV.[49] [e] From ΔH_f = 61.4 ± 0.5 kcal/mol[8] and I_a = 9.28 ± 0.05 eV.[8] [f] Assuming C$_4$H$_5$$^+$ = methylcyclopropenium; ΔH_f = 237 kcal/mol from the apperance energy of C$_4$H$_5$$^+$ from 1-butyne radical cation (no error limits given).[50] [g] From ΔH_f (cyclopropenium) = 257 kcal/mol (no error limits given)[45] and ΔH_f (CH$_3$$^\bullet$) = 35.1 ± 0.2 kcal/mol.[51] [h] From ΔH_f (propargyl) = 81 ± 1 kcal/mol,[51] $I_{a,1}$(propargyl) = 8.67 ± 0.02 eV,[52] and ΔH_f (CH$_3$$^\bullet$) = 35.1 ± 0.2 kcal/mol.[51] [i] RCCSD(T)/cc-pVTZ single-point energies, corrected for differences of ZPE as well as integrated heat capacities to 298 K on the basis of B3LYP/6-31G* structures and vibrational data.
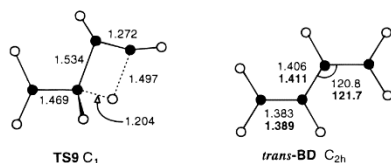

Figure 4. Structures of *trans*-BD and the transition state, TS9, connecting it to LC, which only exists on the UMP2 surface (normal, UMP2; bold, B3LYP; bond lengths in Å, angles in deg).

leads directly to *trans*-BD, thus bypassing CC. However, the barrier for this process (cf. TS9 in Figure 4) is over 10 kcal/mol, in contrast to the decay of LC to CC, which appears to be activationless (see above). Therefore, this hydrogen shift does not represent a competitive reaction, in contrast to the case of the (Et)$_2$$^{\bullet+}$ linear complex where it represents the lowest energy deactivation pathway.[17]

The other reaction passes through an intermediate, 3-buten-3-ylium-1-yl (BU1, see Figure 5), which appears in the form of two automeric C_1 minima, slightly distorted from C_s and interconnected by a very low-lying symmetric transition state at all levels. BU1 is reached from LC by a [1,2] hydrogen transfer in the acetylenic part, which results in a stabilization

by about 7.5 kcal/mol. This reaction involves transition state TS10, which lies 2.4 kcal/mol below TS9 for the above-described [1,3] H transfer, but it is still nearly 8 kcal/mol above LC at the RCCSD(T) level.

BU1 is a shallow minimum at UMP2 and B3LYP, but it is easily transformed to *cis*-BD by another [1,2] H shift via transition state TS11, which lies 1−2 kcal/mol above BU1 at these levels. RCCSD(T) single-point calculations even place TS11 below BU1, so it may well be that this intermediate does not exist. Since there is no a priori reason why BU1 should decay preferentially to *cis*-BD (as found by IRC calculations), we searched for another transition state leading to *trans*-BD. However, we failed to find this, so we assume that there must be a bifurcation point close to TS11 which effects distribution of the molecules into the two channels leading to *cis*- and *trans*-BD.

By accident, we found another, much more stable 3-buten-3-ylium-1-yl type structure, BU2, which arises when both pairs of hydrogens at the terminal carbons of BU1 are rotated by about 90°, whereupon the C1−C2−C3 angle diminishes by about 40° (see Figure 5). BU2 is a minimum at all levels but—unlike BU1—does not show any distortion from C_s symmetry. The spin and charge distributions are surprisingly similar in the two BU species, but they differ, of course, in the orientation of
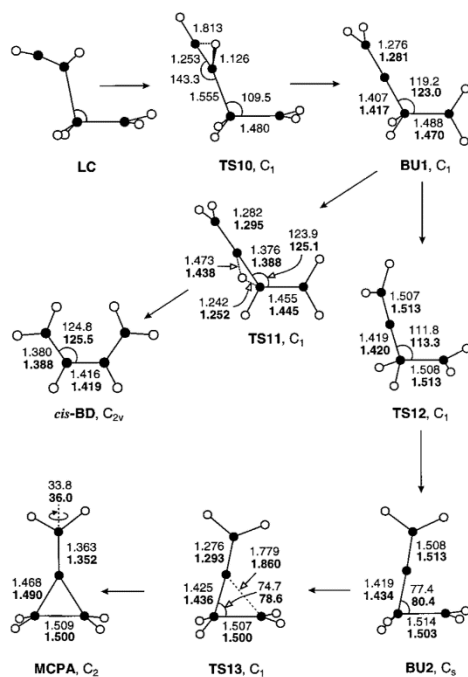
Figure 5. Interconnection and structures of stationary points involved in the transformation of LC to *cis*-BD or MCPA, respectively (normal, UMP2; bold, B3LYP; bond lengths in Å, angles in deg).

Figure 6. Interconnection and structures of stationary points involved in the transformation of BU1 to MA and MA to *trans*-BD (normal, UMP2; bold, B3LYP; bond lengths in Å, angles in deg).

the p-AOs which constitute the SOMO. Despite the close kinship of the two geometric isomers, all standard methods for locating the transition state connecting them failed. After countless trials, a grid of points obtained in the space of two judiciously chosen internal coordinates led us to a saddle point (TS12, imaginary modes at -630 and -281 cm^{-1} by B3LYP and UMP2, respectively) which looks very much like the expected elusive transition state, but all attempts to relate it to reactants or products by IRC calculations failed.

On the basis of TS12, the activation enthalpy for the transofrmation of BU1 to BU2 is, however, less than 1 kcal/mol at the UMP2 and RCCSD(T) levels, so BU1 appears to be another of these fleeting intermediates on the way to more stable $C_4H_6^{\bullet+}$ isomers. On the other hand, BU2 is poised, both from its C–H connectivities and from the close proximity of the CH$_2$ group to the central allenic carbon, to collapse to MCPA. Indeed, a slight shortening of this distance leads to a very low-lying transition state for this process, TS13 (imaginary modes at -194 and -417 cm^{-1} at B3LYP and UMP2, respectively),[53] which disappears on inclusion of zero-point energies.[53] MCPA itself is a stable product, but, due to the vibronic interactions which are typical of olefinic radical cations, its symmetry is reduced from C_2 in the neutrals by an ~35° torsion around the double bond.

The significance of the sequence of reactions described above and illustrated in Figure 5 is that it represents the lowest energy pathway leading to MCPA. Although nearly activationless processes were shown to lead from the weakly bound complexes to CB and *cis*-BD, the intermediate formation of MCPA cannot be completely discounted on the basis of the present results.
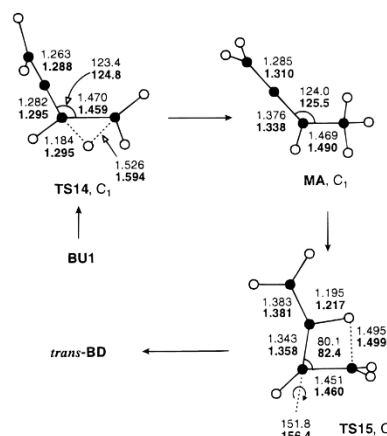
Another stable species to which BU1 can rearrange by a simple 1,2-hydrogen shift is the 1,2-butadiene (methylallene, MA) radical cation, which might serve as a direct precursor in the ultimate fragmentation to $C_3H_3^+ + CH_3^{\bullet}$. Indeed, a corresponding transition state, TS14 (see Figure 6, imaginary modes of -860 and -1545 cm^{-1} at B3LYP and UMP2, respectively), was found about 5 kcal/mol above BU1. MA has no symmetry at all levels because of the rotation in the ionized allene moiety, due to the vibronic interactions which express themselves in the Jahn–Teller distortion of the parent allene radical cation. Transition state TS15 (see Figure 6, imaginary modes of -736 and -774 cm^{-1} at B3LYP and UMP2, respectively) for the 1,3-H shift leading from MA to *trans*-BD was also located, but the barrier for this process is over 40 kcal/mol.

3.2.3. Alternative Rearrangements of CC. As mentioned above, CC represents a metastable intermediate on the way from Et$^{\bullet+}$ + Ac to CB. As in the other cases, we also explored the reaction channels corresponding to hydrogen shifts in CC. The first one is for a transfer of a H atom in the three-membered ring from the CH group to a CH$_2$ group. This is accompanied by ring opening to eventually form BTY.[54] The concerted nature of the H shift and ring opening becomes evident from transition state TS16 (see Figure 7, imaginary modes of -394 and -661 cm^{-1} at B3LYP and UMP2, respectively), which shows that the formation of the new C–H bond occurs simultaneously to the weakening of the C–C bond that is eventually broken. However, the activation barrier for that process is almost 30 kcal/mol, so it is unlikely to be competitive with the other rearrangements described above. Hence, the formation of BTY, i.e., a possible precursor to the propargyl cation, in the Et$^{\bullet+}$ + Ac reaction is very improbable.

The same hydrogen atom can be shifted to the exocyclic, formally divalent carbon to yield MCPA. The transition state for this reaction (TS17, imaginary modes of -394 and -661 cm^{-1} at B3LYP and UMP2, respectively) is very early on the reaction coordinate, and the only change in the C–C bonds is the shortening of the exocyclic one, which gives the double bond in the product. But once again, the activation barrier for this process (27 kcal/mol at the RCCSD(T) level) is much too
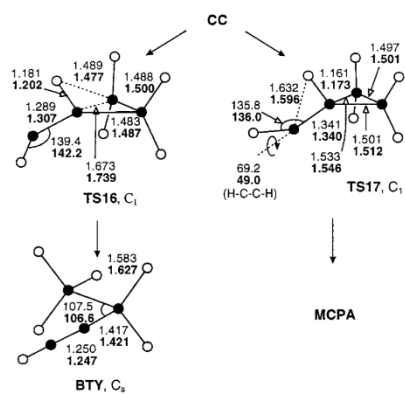
C$_4$H$_6^{•+}$ Potential Energy Surface

J. Phys. Chem. A, Vol. 102, No. 46, 1998 **9303**

Figure 7. Interconnection and structures of stationary points involved in the transformation of CC to MCPA and to BTY, respectively (normal, UMP2; bold, B3LYP; bond lengths in Å, angles in deg).

high to allow this process to compete with the formation of MCPE via BU1 (see above).

3.2.4. Rearrangements of cis-BD to MCPE. Unless the observed ultimate fragmentation to C$_3$H$_3^+$ + CH$_3^•$ involves a concerted loss of CH$_2$ and H (which appears unlikely), it requires a preceding rearrangement of the incipient radical cations (CH$_2$)$_2$(CH)$_2$ to those with a (CH)$_3$(CH$_3$) pattern of connectivities. Therefore, we wanted to find the transition states for all possible H shifts for such rearrangements. We have already described above two such processes (the one leading from BU1 via TS14 to MA and that leading from CC via TS16 to BTY), but both are associated with rather high activation barriers and are unlikely to be competitive with the nearly activationless decays to CB or BD.

The most obvious such rearrangement involves a 1,4-hydrogen shift in *cis*-BD to give a propenylcarbene radical cation (PRC) that would subsequently collapse to yield MCPE, certainly a very attractive direct precursor of C$_3$H$_3^+$ + CH$_3^•$. This pathway has been recognized and explored at the UMP2 level by Keister et al.[8] We confirmed their findings and went on to apply our standard methodology to this process. The results, shown in Figure 8 and Table 2, indicate that it is associated with a barrier of almost 50 kcal/mol. About 60% of this is due to the endothermicity of the BD → PRC rearrangement, and the 17 kcal/mol barrier for the reverse process is mostly due to the fact that the H transfer occurs over such a large distance that a good part of the CH bond dissociation energy must be invested before any bonding interaction to the formally divalent accepting C atom begins to develop.

Transition state TS18 (imaginary modes of −1396 and −1811 cm^{-1} at B3LYP and UMP2, respectively) for this rearrangement shows the expected features of being close to the product, MVC, and strongly reduced C−C−C angles to maximize the bonding interactions to the migrating hydrogen atom. As expected (and found by Keister et al.), PRC is poised for collapse to MCPE, but due to the strong ring strain which prevails in the latter compound, this process is endothermic by about 8 kcal/mol. Transition state TS19 (Figure 8, imaginary modes of −170 and −379 cm^{-1} at B3LYP and UMP2, respectively) lies very close to MCPE and only 0.05 kcal/mol above it by RCCSD(T)//UMP2.

Our RCCSD(T)/cc-pVTZ prediction for $\Delta_f H$ of MCPE (from the data in Table 3) is 276 kcal/mol, in excellent accord with
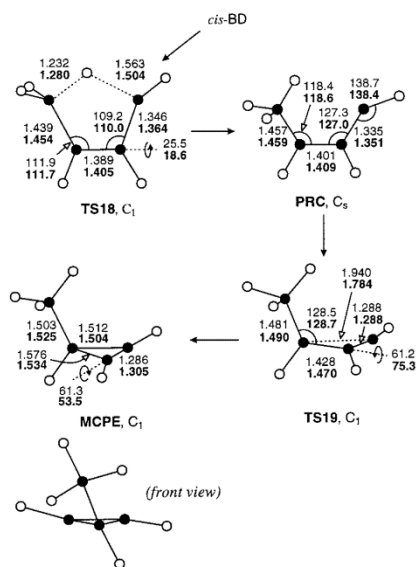

Figure 8. Interconnection and structures of stationary points involved in the transformation of *cis*-BD to MCPE (normal, UMP2; bold, B3LYP; bond lengths in Å, angles in deg). Bottom: alternate view of MCPE.

the experimental value of 275.6 ± 1.2 kcal/mol given by Keister et al.[8] To compare our result with those obtained by UMP2 with a similar basis set,[8] we should consider ΔE_0 for the *trans*-BD → MCPE reaction, for which we obtain ΔE_0 = 40.0 kcal/mol (experimental ΔH_{298} = 40.5 ± 1.7 kcal/mol), as compared to 35.4 kcal/mol by UMP2. Interstingly, the RMP2/6-31G* prediction for the same reaction is nearly identical with that obtained at our reference level of theory. Although this agreement is perhaps a bit fortuitous, it clearly demonstrates the problems of UMP2 with rearrangements involving open-shell species where the UHF wave function shows different degrees of spin contamination: S^2 is nearly 1 in *trans*-BD but only 0.78 in MCPE, which clearly puts the former species at an energetic disadvantage relative to the latter, thus resulting in a too small isomerization energy. On the other hand, B3LYP overestimates ΔE_0 for this process by about 4 kcal/mol.

We were surprised to find that the equilibrium structure of MCPE is strongly distorted from the (average) C_s symmetry, which prevails in the neutral (see separate view at the bottom of Figure 8). As it turns out, the C_s structure is a transiton state for interconversion of two C_1 mimima (over a barrier of 3.5 kcal/mol by UMP2). The reason for this distortion can be traced back to a vibronic interaction between an A′ state and an A′′ state, where the unpaired electron occupies the singly occupied π-MO (HOMO) or the subjacent antisymmetric σ_{C-C} MO (HOMO-1), respectively (see Figure 9). Distortion from C_s symmetry allows these two states to mix. This expresses itself very clearly in the HOMO at the equilibrium geometry of MCPE, which really looks like a linear combination of the two highest occupied MOs in C_s. As we have shown in the example of the CB ring opening, such vibronic interactions are quite common in radical cations with their low-lying excited states, and they may have a profound influence on the shapes of the potential energy surfaces.[33]
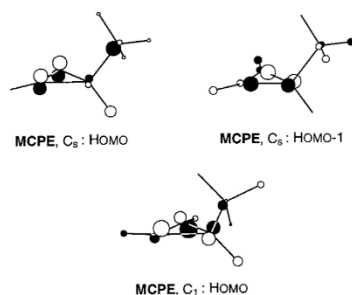
Figure 9. HOMO and HOMO-1 of MCPE in C_s symmetry and HOMO of the same compound at its nonsymmetric equilibrium geometry (cf. Figure 8).

3.3. Fragmentation Channels.

3.3.1. Cyclopropenium Cation (CPR) + $CH_3^{\bullet}$. Apart from the ring opening of CB, the $C_4H_6^{\bullet+}$ PES received the most attention in the literature in connection with the methyl loss reactions of $C_4H_6^{\bullet+}$ ions. There is every reason to believe that the precursor for this fragmentation is a $C_4H_6^{\bullet+}$ structure carrying a methyl group, because it is hardly conceivable that a concerted reaction involving C—C bond breaking accompanied by a hydrogen atom transfer would be kinetically competitive with simple C—CH_3 bond cleavage.

Since CPR is the more stable of the two possible $C_3H_3^+$ structures, MCPE appears to be the most plausible precursor for the methyl loss reaction. Indeed, it was concluded many years ago from scattering diagrams for the process $Ac^{\bullet+} + Et \rightarrow [C_4H_6^{\bullet+}] \rightarrow C_3H_3^+ + CH_3^{\bullet}$ that the last intermediate on the pathway to the ultimate fragments should have a structure close to MCPE.[2,3] This is in accordance with the experimental evidence (summarized in ref 8) that $C_4H_6^{\bullet+}$ ions of different structures loose $CH_3^{\bullet}$ by prior isomerization to a higher energy structure, which is usually assumed to be MCPE. Before the completion of the present paper, Keister et al.[8] supported this hypothesis by UMP2/6-311G** calculations. Although we basically arrived at similar conclusions, we would like to present our results because our calculations were carried out on a higher level and we can provide some interesting additional details.

As noted already by Keister et al.,[8] elongation of the C—C bond in the course of dissociation of MCPE is not associated with a monotonic increase in energy. On the way to the fragments, they encountered a [CPR⋯$CH_3^{\bullet}$] ion—molecule complex whose structure they did, however, not specify. We found this to correspond to the species C1 in Figure 10, where the unpaired electron resides in the bonding combination between one component of the degenerate cyclopropenium π-MO and the $CH_3^{\bullet}$ SOMO. It has a methyl-ring C—C distance of $d = 2.9$ Å, while the angle α between the plane of the CPR ring and the methyl carbon is 111°. It is a minimum at all levels except B3LYP, where it does not exist as a stationary point. Transition state TS20, which connects MCPE to C1, lies less than 1 kcal/mol above C1 (slightly more if the ZPE correction is based on UMP2 frequencies), so this complex is barely protected from collapse to MCPE.

In addition to C1, we found another [CPR⋯$CH_3^{\bullet}$] complex in which the methyl radical is positioned *in the plane of* CPR ($d = 3.4$ Å, $\alpha = 180°$). The bonding in this complex, which is about 1 kcal/mol more stable than C1, is entirely different from that which prevails in C1 and corresponds to a hydrogen bridge between CPR and the p-AO of the methyl radical, similar to that found in the acetylene dimer cation.[33] C2 exists as a
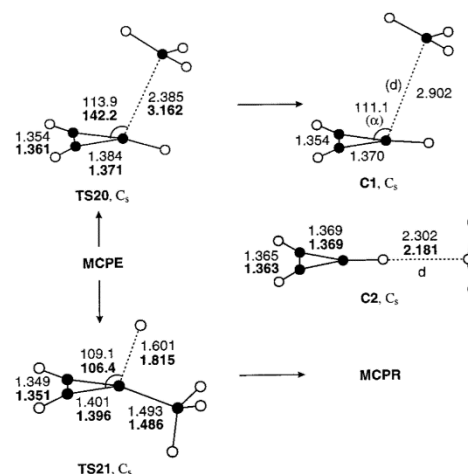

Figure 10. Structures of [$C_3H_3^+$⋯$CH_3^{\bullet}$] ion—molecule complexes C1 and C2 and the transition state, TS20, connecting them to MCPE (normal, UMP2; bold, B3LYP; bond lengths in Å, angles in deg).

minimum at all levels including B3LYP, where we also found a transition state (imaginary mode at -94 cm^{-1}) which connects it to MCPE. All attempts to locate this transition state at other levels failed, so we conclude that C2 is connected to MCPE via C1. However, we did not consider it expedient to locate the transition state connecting the two complexes, a task which would have been quite arduous in view of the flatness of the potential energy surface and did not promise to provide much additional insight.

Starting from C1 or C2, the energy rises smoothly to the dissociation products, CPR + $CH_3^{\bullet}$. Compared to the UMP2 calculations of Keister et al.,[8] we also achieved a considerable improvement in the prediction of the threshold for methyl loss from BD. Relative to *trans*-BD, ΔE_0 of the dissociation products is now 53.9 kcal/mol (experimental $\Delta H_{298} = 55.3 \pm 0.3$ kcal/mol), compared to 48 kcal/mol by UMP2, which testifies once more to the problems of UMP2 with spin-contaminated species. If we take ΔE_0 relative to the starting $Et^{\bullet+} + Ac$, then the discrepancy between the RCCSD(T) and the UMP2 results is much less pronounced (cf. Table 4).

3.3.2. Propargyl Cation (PRG) + $CH_3^{\bullet}$. Although CPR is over 25 kcal/mol more stable than its open-chain isomer, propargyl cation, PRG, the latter (combined with $CH_3^{\bullet}$) still lies below $Ac^{\bullet+} + Et$, the compounds used in the crossed-beam experiments of Herman et al.[2,3] (note from Table 4 that PRG + $CH_3^{\bullet}$ lie *above* $Et^{\bullet+} + Ac$, but Table 1 gives an energy of +20 kcal/mol for the charge-exchange process). Two $C_4H_6^{\bullet+}$ isomers must be taken into account as precursors for this fragmentation, BTY and MA. However, the transition states for methyl loss from both ions have a quasilinear structure which is incompatible with the experimental findings, so we did not pursue these channels in any more detail. However, we note that the energy difference between the two $C_3H_3^+$ isomers (experimental $\Delta H_{298} = 25.9 \pm 2$ kcal/mol) is rather well reproduced by RCCSD(T) ($\Delta E_0 = 26.8$ kcal/mol), whereas MP2 ($\Delta E_0 = 31.4$ kcal/mol) is once more off the mark, in contrast to B3LYP ($\Delta E_0 = 25.3$ kcal/mol).

3.3.3. $C_4H_5^+ + H^{\bullet}$. The second fragmentation channel observed in the crossed-beam experiments of Herman et al.[2,3]

C$_4$H$_6$$^{\bullet+}$ Potential Energy Surface

J. Phys. Chem. A, Vol. 102, No. 46, 1998 **9305**

TABLE 4: Energies and 0 K Enthalpies (kcal/mol) Relative to Ac + Et$^{\bullet+}$ of Structures Involved in the Fragmentation of C$_4$H$_6$$^{\bullet+}$

structure[a]	B3LYP/6-31G*		(R)MP2/6-31G*		UMP2/6-31G*		(R)CCSD(T)/cc-pVTZ	
	ΔE	ΔE_0	ΔE	ΔE_0[b]	ΔE	ΔE_0	ΔE	ΔE_0[c]
MCPE	−41.16	−38.22	−40.12	−35.89	−37.64	−33.42	−34.83	−31.89
TS20: MCPE → C1	−25.15[d]	−24.40[d]	−25.53	−22.84	−25.28	−22.58	−21.75	−21.00
C1	e	e	−26.00	−24.14	−26.01	−24.15	−22.32	−20.46
C2 (H-bonded)	−25.76	−24.82	−26.92	−25.09	−26.97	−25.15	−22.55	−21.61
CPR + CH$_3$$^{\bullet}$	−19.41	−19.81	−21.20	−20.96	−21.31	−21.06	−17.58	−17.97
PRG + CH$_3$$^{\bullet}$	7.30	5.46	11.65	10.49	11.54	10.38	10.70	8.86
TS21: MCPE → MCPR	−18.35	−18.37	−21.86	−20.73	−21.30	−20.17	−13.18	−13.19
MCPR + H$^{\bullet}$[f]	−36.78	−38.19	−23.35	−24.76			−19.57	−20.98
CBY + H$^{\bullet}$[f]	−16.10	−16.31	−12.35	−12.56			−10.98	−10.58
MEA + H$^{\bullet}$[f]	−3.70	−5.98	−10.41	−12.70			−4.08	−6.36

[a] For UMP2 and B3LYP optimized structures, see Figures 8−11; structures TSx are transition states. [b] ZPE differences from UMP2 vibrations. [c] ZPE differences from B3LYP. [d] TS between MCPE and C2, discussion see text. [e] Not found at B3LYP. [f] With E(H$^{\bullet}$) = 0.5 hartrees, (hence no entry for UMP2).
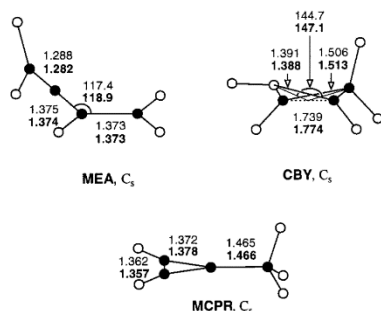

Figure 11. Structures of the three most stable C$_4$H$_5$$^+$ isomers, MEA, CBY, and MCPR (normal, UMP2; bold, B3LYP; bond lengths in Å, angles in deg).

corresponded to loss of atomic hydrogen. Exploratory B3LYP calculations indicated that only three C$_4$H$_5$$^+$ isomers are energetically accessible from Ac$^{\bullet+}$ + Et. They correspond to loss of H$^{\bullet}$ from MCPE, CB, or MA, which yields methylcyclopropenium cation (MCPR), cyclobutenium cation (CBY), or methyleneallene cation (MEA), respectively (see Table 4 and Figure 11). The last of these structures can be excluded on the basis of the observed scattering diagrams which can, however, not distinguish between the former two possibilities.

Our calculated energy of the most stable of these fragments, MCPR + H$^{\bullet}$ relative to Et$^{\bullet+}$ + Ac, is in excellent accord with experiment (cf. Table 3). However, in contrast to the methyl loss channels, no ion−molecule complex is involved in hydrogen loss, which leads to the replacement of the corresponding minimum by a transition state (TS21) which lies about 8 kcal/mol above the fragments, thus implying a corresponding activation barrier for the reverse process. In view of the nearly activationless addition of CH$_3$$^{\bullet}$ to CPR and the fact that the exothermicities of both processes are quite similar (14 vs 11 kcal/mol), this result is rather surprising. We found that addition of H$^{\bullet}$ to CPR, or to the unsubstituted carbon of MCPR, is also nearly activationless;[55] therefore, the activation barrier encountered by H$^{\bullet}$ in the attack at the methyl-substitution site may be traced to the disruption of the hyperconjugative stabilization of CPR by the methyl group.

Another way to view this phenomenon is by reference to the MOs depicted in Figure 9: TS20 and TS21 connect adiabatically to the ground state of MCPE, where the SOMO is much more strongly bonding along the CPR−CH$_3$ bond. This expresses itself also in the elongated exocyclic C(sp^2)- -C(sp^3) bond in

MCPE. Thus, the transition state for loss of CH$_3$$^{\bullet}$ is closer to the reactant than that for loss of H$^{\bullet}$, and hence, the activation barrier for the former process is smaller.

In view of the fact that TS21 for H$^{\bullet}$ loss from MCPE lies below the energy of the other possible C$_4$H$_5$$^+$ isomer, CBY (+ H$^{\bullet}$) indicates that the formation of MCPR represents the lowest energy channel for H$^{\bullet}$ loss. Therefore, we did not pursue the higher lying pathway leading to CBY + H$^{\bullet}$.

4. Conclusions

This is the last of a series of three papers on the formation, rearrangement, and subsequent redissociation of ion−molecule complexes involving ethylene and acetylene.[17,33] The three cases we have treated show some similarities, but also some surprising differences, which demonstrates that general rules governing the reactivity of radical cations are not as easy to obtain as in the case of neutral closed-shell molecules. Rather close analogies can be seen between the behavior of (Ac)$_2$$^{\bullet+}$ and the present (Et···Ac)$^{\bullet+}$ complexes. In both cases, we see a nearly activationless collapse of the incipient ion−molecule complex to stable four-membered-ring structures. Although these arise formally by a $2\pi + 2\pi$ cycloaddition of the reactants, the actual course of the reaction is, in both cases, a highly nonconcerted one, leading to several highly metastable intermediates. The last of these is a species which may be regarded as a cycloprop(en)yl carbene cation (CC).

In the present case, several alternative pathways for the rearrangement of some of these intermediates were explored, but they invariably involve rather high-lying transition states and thus do not appear to compete with the rearrangements starting from the primary stable product, i.e., the cyclobutene radical cation (CB). The ring opening of CB to butadiene radical cation (BD) occurs quite readily (E_a 18 kcal/mol[7]), but any formation of other stable C$_4$H$_6$$^{\bullet+}$ isomers with different C−H connectivities involves much higher activation barriers. Three such processes were found to lead from *cis*-BD to tautomers poised for fragmentation to C$_3$H$_3$$^+$ + CH$_3$$^{\bullet}$, i.e., the fragments observed in the crossed-beam studies of Herman et al.[2,3] (see Figure 12).

Two of these processes involve similar barriers of about 45 kcal/mol. The first is a direct 1,4-hydrogen transfer to give an intermediate best regarded as a propenylcarbene cation (PRC). The second follows a sequence of two 1,2-H transfers, the first one to give a distonic radical cation CH$_2$−C$^+$−CH$_2$−CH$_2$$^{\bullet}$, BU1, followed by another one yielding methylallene, MA. Instead of undergoing the second of these H transfers, the intermediate BU1 can cyclize via a very low-lying transition state to the
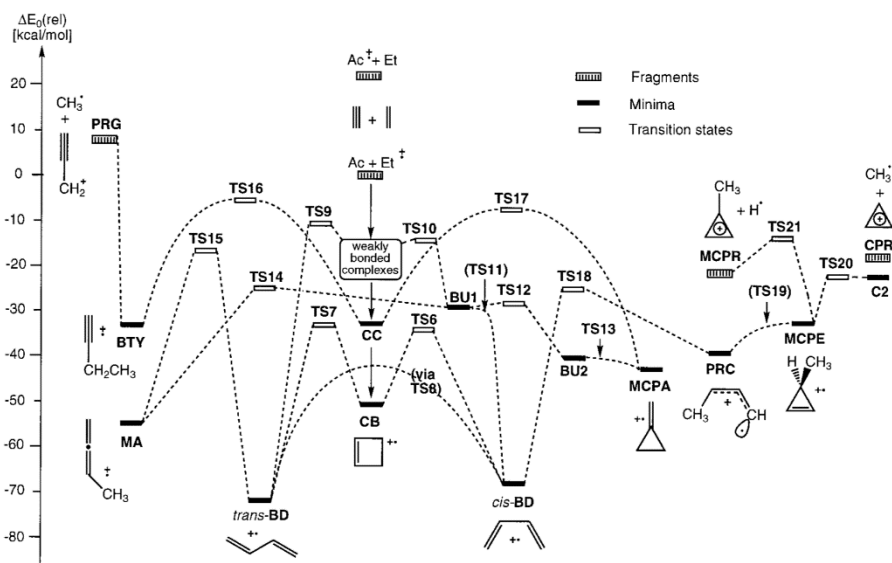
Figure 12. Schematic representation of the potential energy surface on which the Et$^{\bullet+}$ + Ac reaction and the subsequent rearrangements and fragmentations discussed in this study take place. Energies relative to Et$^{\bullet+}$ + Ac are from single-point RCCSD(T)/cc-pVTZ calculations at B3LYP/6-31G* geometries, corrected for ZPE differences on the basis of frequencies calculated at the same level (cf. Tables 1−4). Transition states which lie at lower energies than reactants at this level are indicated by vertical arrows. For details of the region labeled "weakly bonded complexes", cf. Schemes 1 and 2 and Table 1.

methylenecyclopropane radical cation, MCPA, which can, however, not lose CH$_3$ in a simple fragmentation process.

Formation of the third C$_4$H$_6$$^{\bullet+}$ isomer carrying a methyl group, 1-butyne radical cation, BTY, requires much higher activation and was not considered in detail. Instead, we concentrated on the methyl loss reaction from PRC which occurs after intermediate cyclization to the methylcyclopropene radical cation, MCPE, to yield the most stable C$_3$H$_3$$^+$ isomer, the cyclopropenium cation, CPR. This mechanism is in accord with experimental observations,[2,3,8] and it involves formation of a [CPR···CH$_3$$^{\bullet}$] ion−molecule complex which comes in two forms, C1 and C2, of entirely different nature but very similer energy. The fragmentation of MA leads to the much less stable propargyl cation (PRG), and its mechanism was not considered in detail because it is incompatible with the observed scattering diagrams.[2,3]

Finally, we considered different cleavages leading to C$_4$H$_5$$^+$ + H$^{\bullet}$, the other pair of fragments observed in the crossed-beam studies.[2,3] The most stable C$_4$H$_5$$^+$ isomer is the methylcyclopropenium cation (MCPR), which may be formed readily from MCPE in a process which does, however, not involve an ion−molcule complex of any sort. Two other C$_4$H$_5$$^+$ species, the cyclobutenium cation, CBY, and the methyleneallene cation, MEA, lie 10−12 kcal/mol higher in energy. They require as precursors CB or MA, respectively, both of which are accessible from Et$^{\bullet+}$ + Ac below the dissociation threshold. A reaction path via MEA may again be excluded, because it is not compatible with scattering diagrams.

Supporting Information Available: Tables containing the UMP2 and B3LYP total energies, thermal corrections, and Cartesian coordinates of all stationary points (minima and transition states) discussed in this study, in ASCII format, are available through the Internet. See any current masthead page for Web access instructions.

References and Notes

(1) Baer, T. In *Gas-Phase Ion Chemistry*; Bowers, M. T., Ed.; Academic: New York, 1979; Vol. 1, p 153.
(2) Herman, Z. Unpublished results.
(3) Herman, Z.; Birkinshaw, K. *Ber. Bunsen-Ges. Phys. Chem.* **1973**, *77*, 566.
(4) Du, P.; Borden, W. T. *J. Am. Chem. Soc.* **1987**, *109*, 5330.
(5) Belville, D. J.; Chelsky, R.; Bauld, N. L. *J. Comput. Chem.* **1982**, *3*, 548.
(6) Wiest, O. *J. Am. Chem. Soc.* **1997**, *119*, 5713.
(7) Sastry, G. N.; Bally, T.; Hrouda, V.; Carsky, P. *J. Am. Chem. Soc.* **1998**, *120*, 9323.
(8) Keister, J. W.; Baer, T.; Evans, M.; Ng, C. Y.; Hsu, C. W. *J. Phys. Chem. A* **1997**, *101*, 1866.
(9) Werner, A. S.; Baer, T. *J. Chem. Phys.* **1978**, *62*, 2900.
(10) Bunn, T. L.; Baer, T. *J. Chem. Phys.* **1986**, *85*, 6361.
(11) Preuninger, F. N.; Farrar, J. M. *J. Chem. Phys.* **1982**, *77*, 263.
(12) Dannacher, J.; Flamme, J. P.; Stadelmann, J.-P.; Vogt, J. *J. Chem. Phys.* **1980**, *51*, 189.

C$_4$H$_6$$^{\bullet+}$ Potential Energy Surface

J. Phys. Chem. A, Vol. 102, No. 46, 1998 **9307**

(13) Bombach, R.; Dannacher, J.; Stadelmann, J.-P. *J. Am. Chem. Soc.* **1983**, *105*, 1824.

(14) Sellers-Hann, L.; Kreiller, R. E.; Russell, D. H. *J. Chem. Phys.* **1988**, *89*, 889.

(15) Russell, D. H.; Gross, M. L.; Greef, J. V. d.; Nibbering, N. N. M. *J. Am. Chem. Soc.* **1979**, *101*, 2086.

(16) Bally, T.; Borden, W. T. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1999; Vol. 13 (in print).

(17) Jungwirth, P.; Bally, T. *J. Am. Chem. Soc.* **1993**, *115*, 5783.

(18) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Repogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; DeFrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzales, M. C.; Pople, J. A.; Gaussian 94, Version B1 and D4; Gaussian, Inc.: Pittsburgh, PA, 1995.

(19) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Kosecki, S.; Matsanuga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montogomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.

(20) Lee, T. J.; Rendell, A. P.; Dyall, K. G.; Jayatilaka, D. *J. Chem. Phys.* **1994**, *100*, 7400.

(21) Lauderdale, W. J.; Stanton, J. F.; Gauss, J.; Watts, J. D.; Bartlett, R. J. *Chem. Phys. Lett.* **1991**, *187*, 21.

(22) Amos, R. D.; Andrews, J. S.; Handy, N. C.; Knowles, P. J. *Chem. Phys. Lett.* **1991**, *186*, 130.

(23) Amos, R. D.; Alberts, I. L.; Andrews, J. S.; Colwell, S. M.; Handy, N. C.; Jayatilaka, D.; Knowles, P. J.; Kobayashi, R.; Laidig, K. E.; Laming, G.; Lee, A. M.; Maslen, P. E.; Murray, C. W.; Rice, J. E.; Simandiras, E. D.; Stone, A. J.; Su, M. D.; Tozer, D. J. Cadpac 5; Cambridge University: Cambridge, MA, 1995.

(24) Stanton, J. F.; Gauss, J.; Watts, J. D.; Lauderdale, W. J.; Bartlett, R. J. *Int. J. Quantum Chem. Symp.* **1992**, *26*, 879.

(25) Hariharan, P. C.; Pople, J. A. *Chem. Phys. Lett.* **1972**, *16*, 217.

(26) Knowles, P. J.; Hampel, C.; Werner, H.-J. *J. Chem. Phys.* **1993**, *99*, 5219.

(27) Werner, J.-J.; Knowles, P. J.; Almlöf, J.; Amos, R. D.; Deegan, M. J. O.; Elbert, S. T.; Hampel, C.; Meyer, W.; Peterson, K.; Pitzer, R.; Stone, A. J.; Taylor, P. R.; Lindh, R. MOLPRO 96.1, 1996.

(28) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.

(29) Gonzales, C.; Schlegel, H. B. *J. Chem. Phys.* **1989**, *90*, 2154.

(30) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

(31) Xantheas, S. S. *J. Chem. Phys.* **1996**, *104*, 8821.

(32) *NIST Standard Reference Database*, *69*, Aug 1997 Release.

(33) Hrouda, V.; Roeselova, M.; Bally, T. *J. Phys. Chem. A* **1997**, *101*, 3925.

(34) This transition state exhibits one imaginary mode interconnecting two automeric PpC structures at the UMP2 level (−79 cm^{-1}). At ROHF (−128, −65 cm^{-1}), UHF (−140, −67 cm^{-1}), and B3LYP (−119, −58 cm^{-1}), a second, very soft imaginary mode appears.

(35) TS2 is a true transition state with one imaginary mode at the SCF level (frequencies of −606 cm^{-1} at ROHF, −150 cm^{-1} at UHF), leading formally to PC (which is a transition state at these levels) but connecting in practice CC with TC.

(36) This structure which may be viewed as a distonic radical cation is called but-3-en-4-ylium-1-yl according to the standard chemical nomenclature.

(37) UHF/6-31G*, ROHF/6-31G*, and UMP2/6-31G* predict clearly that the twist structure LC is a minimum. In contrast, RMP2/6-31G* gives an imaginary vibrational frequency, which varies greatly with a selected step size in the numerical two-sided differentiation of the RMP2 gradient (−1227 cm^{-1} for 0.001 au, −619 cm^{-1} for 0.005 au, −427 cm^{-1} for 0.01 au, −578 cm^{-1} for 0.02 au, and −443 cm^{-1} for 0.05 au). Also, the respective normal coordinate varies greatly with differing step sizes. We tried to move downhill along the normal coordinate given by the force constant matrix obtained with a step size of 0.01 au. However, no matter how small we chose the distortion, the energy was invariably higher than that of the stationary point. Therefore, we cannot decide whether LC is a minimum or not on the RMP2 surface.

(38) By ROHF, we managed to find only TS4 (−297 cm^{-1}) but failed to locate TS3 because of the convergency problems, whereas at UHF, we located only TS3 (−512 cm^{-1}).

(39) The trans conformation (dihedral angle = 180) is a transition state for the interconversion of two twisted ones, at the UMP2 level about 4.2 kcal/mol higher than the latter. Surprisingly, the SCF methods also predict the cis structure (dihedral angle = 0) to be a minimum, but this feature is not reproduced at the correlated levels and must hence be regarded as an artifact.

(40) Ruscic, B.; Berkowitz, J.; Curtiss, L. A. *J. Chem. Phys.* **1989**, *91*, 114.

(41) Blush, J. A.; Chen, P. *J. Phys. Chem.* **1992**, *96*, 4138.

(42) Houle, F. A.; Beauchamp, J. A. *J. Am. Chem. Soc.* **1979**, *101*, 4067.

(43) Bally, T.; Sastry, G. N. *J. Phys. Chem. A* **1997**, *101*, 7923.

(44) Note that CC comes in two conformers that differ in the dihedral angle to the hydrogen atom at the divalent carbon. At all levels, the trans from is more stable than the cis form by about 1 kcal/mol. An estimate of barrier for interconvesion of the two forms (about 3 kcal/mol) may be taken from on the B3LYP value calculated for CC in C$_4$H$_4$$^{\bullet+}$.

(45) Lias, S. G.; Bartmess, J. E.; Liebman, J. F.; Holmes, J. L.; Levin, R. D.; Mallard, W. G. *J. Phys. Chem. Ref. Data Suppl. 1* **1988**, *17*, 1.

(46) Prosen, E. J.; Maron, F. W.; Rossini, R. D. *J. Res. NBS* **1951**, *46*, 106.

(47) Mallard, W. G.; Miller, J. H.; Smyth, K. C. *J. Chem. Phys.* **1983**, *79*, 5900.

(48) Wiberg, K. B.; Fenoglio, R. A. *J. Am. Chem. Soc.* **1968**, *90*, 3395.

(49) Bieri, G.; Burger, F.; Heilbronner, E.; Maier, J. P. *Helv. Chim. Acta* **1977**, *60*, 2213.

(50) Lossing, F. P.; Holmes, J. L. *J. Am. Chem. Soc.* **1984**, *106*, 6917.

(51) Tsang, W. In *Energetics of Organic Free Radicals*; Simoes, J. A. M., Greenberg, A., Liebman, J. F., Eds.; Blackie Academic & Professional: London, 1996; p 22.

(52) Minsek, D. W.; Chen, P. *J. Phys. Chem.* **1990**, *94*, 8399.

(53) At UMP2, TS12 leads to another minimum which is separated by a second transition state from MCPA. This minimum resembles MCPA, but the three-membered ring is distorted in the plane (see Supporting Information for the structure called MCPA-distorted).

(54) The orthogonal π orbitals in the triple bond of BTY give rise to two nearly degenerate electronic states of A′ and A′ symmetry. The SCF procedures tend to converge to the A′ state instead of to the A′ state, which is lower in energy at all levels (about 0.4 kcal/mol at UMP2).

(55) Sastry, G. N.; unpublished results of B3LYP calculations.