



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

DIPLOMOVÁ PRÁCE

Bc. Martin Hanák

**Počítačové modelování inhibice
SARS-CoV-2 proteázy**

Fyzikální ústav UK

Vedoucí diplomové práce: RNDr. Ivan Barvík, Ph.D.

Studijní program: Biofyzika a chemická fyzika

Studijní obor: FBCHPT

Praha 2023

Prohlašuji, že jsem tuto diplomovou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V dne

Podpis autora

Rád bych poděkoval vedoucímu mojí diplomové práce, RNDr. Ivanu Barvíkovi, Ph. D., jehož odborné vedení, rady a podpora pro mě byly neocenitelné během celého procesu. S jeho pomocí jsem byl schopen nejen pochopit složitou problematiku této práce, ale také značně vylepšit její kvalitu.

Děkuji také svým rodičům a sourozencům, kteří mě podporovali během tvorby této práce a bez jejichž podpory bych ji nedokázal s úspěchem dokončit. Jsem vděčný, že mám úžasnou rodinu, která mě podporuje na mé životní cestě.

Výpočetní prostředky byly poskytnuty v rámci projektu e-INFRA CZ (ID:90140) financovaného Ministerstvem školství, mládeže a tělovýchovy České republiky.

Název práce: Počítačové modelování inhibice SARS-CoV-2 proteázy

Autor: Bc. Martin Hanák

Katedra: Fyzikální ústav UK

Vedoucí diplomové práce: RNDr. Ivan Barvík, Ph.D., Fyzikální ústav UK

Abstrakt: Práce se zabývá různými metodami výpočtu volné energie pomocí molekulárně-dynamických simulací za účelem studia inhibice biologických makromolekul a identifikace možných léčiv. Konkrétně byla prozkoumána možnost využití nerovnovážných metod na základě Jarzynského rovnosti a Crooksova fluktuačního teorému. Použité metody byly otestovány na jednoduchých systémech aminokyselin a komplexu éterové koruny s draslíkem. Následně byly zkoumané metody aplikovány na komplexy adamantánů s cyklodextriny a především na inhibici hlavní proteázy SARS-CoV-2 Mpro.

Klíčová slova: volná energie, inhibice, SARS-CoV-2, proteáza Mpro

Title: Computer modeling of SARS-CoV-2 protease inhibition

Author: Bc. Martin Hanák

Department: Institute of Physics of Charles University

Supervisor: RNDr. Ivan Barvík, Ph.D., Institute of Physics of Charles University

Abstract: This work deals with various methods of computing free energy using molecular dynamics simulations to study the inhibition of biological macromolecules and to identify potential drugs. Specifically, the possibility of using non-equilibrium methods based on Jarzynski's equality and Crooks fluctuation theorem was explored. The methods used were tested on simple systems of amino acids and a complex of ether crown with potassium. Subsequently, the tested methods were applied to complexes of adamantanes with cyclodextrins and especially to the inhibition of the main protease of SARS-CoV-2 Mpro.

Keywords: free energy, inhibition, SARS-CoV-2, protease Mpro

Obsah

Úvod	3
Cíle práce	4
1 SARS-CoV-2	5
1.1 Virionová částice	5
1.2 Životní cyklus koronaviru	6
1.3 Spike protein	7
1.4 Proteáza Mpro	9
2 Molekulární dynamika	15
2.1 Silové pole	16
2.2 Silové pole CHARMM	17
2.3 Silové pole CGenFF	18
2.4 Integrátory	19
2.5 Periodické okrajové podmínky	20
2.6 Omezení vazeb	23
2.7 Statistické soubory	24
3 Výpočty volné energie	27
3.1 Gibbsova a Helmholtzova volná energie	27
3.2 Vazebná volná energie	28
3.3 Termodynamický cyklus	29
3.4 Alchymistické transformace	32
3.5 Perturbace volné energie	33
3.6 Termodynamická integrace	34
3.7 Výpočet volné energie prostřednictvím rovnovážných MD simulací	35
3.8 Výpočet volné energie prostřednictvím nerovnovážných MD simulací	35
3.9 Dodatečné příspěvky (korekce) k vypočtené volné energii	37
4 Výsledky MD simulací	39
4.1 Solvatace aminokyselin	40
4.1.1 Příprava simulovaných systémů	42
4.1.2 Výsledky simulací	43
4.1.3 Diskuze	50
4.2 Komplex éterové koruny a draslíku	51
4.2.1 Příprava simulovaných systémů	52
4.2.2 Výsledky simulací	52
4.2.3 Diskuze	56
4.3 Komplex adamantánů a cyklodextrinů	58
4.3.1 Příprava simulovaných systémů	59
4.3.2 Výsledky simulací	62
4.3.3 Diskuze	66
4.4 Nekovalentní inhibitory proteázy Mpro	68
4.4.1 Příprava simulovaných systémů	70
4.4.2 Výsledky simulací	70

4.4.3	Diskuze	75
4.5	Kovalentní inhibitory proteázy Mpro	77
4.5.1	Příprava simulovaných systémů	78
4.5.2	Výsledky simulací	79
4.5.3	Diskuze	85
Závěr		89
Seznam použité literatury		91
A Přílohy		102
A.1	Tcl skript pro přípravu systémů	102
A.2	Tcl skript pro tvorbu topologie, solvataci a ionizaci	104
A.3	Konfigurační soubor programu NAMD	107
A.4	Tcl skript využitý k analýze trajektorií	110

Úvod

Pandemie COVID-19 zasáhla celosvětově miliony lidí a zapříčinila nevídané ekonomické a sociální otřesy. Virus zodpovědný za tuto pandemii, SARS-CoV-2, je vysoce nakažlivý a primárně napadá náš dýchací systém. Jeho rapidní rozšíření nám připomnělo urgentní nutnost vývoje efektivních vakcín a léčiv. Jeden z potenciálních přístupů k boji proti tomuto viru je inhibice jeho proteáz, které jsou nezbytné pro replikaci a šíření viru.

Hlavní proteáza viru SARS-CoV-2, také označovaná Mpro nebo 3CLpro, je slibným cílem léčiv, neboť hraje klíčovou roli v životním cyklu viru. Bez působení proteázy Mpro jsou virem tvořeny pouze neaktivní polyproteiny. Navíc struktura proteázy Mpro je značně konzervovaná (tj. bez mutací důležitých aminokyselin) pro různé koronaviry, což snižuje pravděpodobnost vzniku rezistance vůči lékům, které jsou na ni zaměřeny.

Počítačové modelování obecně představuje účinný nástroj při hledání nových léčiv, neboť umožňuje prozkoumat knihovny s velkým množstvím potenciálních inhibitorů a identifikovat slibné kandidáty. Prostřednictvím počítačového modelování je možné predikovat účinnost inhibitoru, optimalizovat jeho strukturu a zmenšit tím nároky na drahou a časově náročnou syntézu a následné experimentální testování potenciálních léků.

Molekulárně-dynamické simulace představují jednu z široce využívaných metod počítačového modelování. Umožňují simulovat pohyb molekul na atomární úrovni a studovat tak mimo jiné dynamiku proteázy SARS-CoV-2 a její interakci s různými inhibitory. S pomocí molekulární dynamiky jsme schopni predikovat vazebnou volnou energii komplexu inhibitoru a proteázy fyzikálně korektním způsobem a jsme tak schopni vybrat z velkého množství zkoumaných inhibitorů ten nejúčinnější.

Tradičně byly vazebné volné energie určovány pomocí rovnovážných molekulárně-dynamických simulací, kdy je její hodnota vypočtena na základě jedné či několika málo velmi dlouhých trajektorií. Nevýhodou toho přístupu je čekání na výsledek po dobu několika dnů či dokonce týdnů.

V posledních letech se proto začalo využívat nerovnovážných molekulárně-dynamických simulací, kdy se vychází z analýzy velkého množství nezávislých trajektorií. Pokud máme k dispozici hardwarové prostředky, které umožní masivně paralelní realizaci výpočtů, tak můžeme výsledek získat podstatně rychleji.

V této práci se zaměřuji na prozkoumání možností realizace výpočtů volné energie pomocí nerovnovážných molekulárně-dynamických simulací v prostředí tzv. MetaCentra (metavo.metacentrum.cz). Metody výpočtu volné energie jsou nejprve testovány na jednoduchých modelových systémech, jako jsou aminokyseliny, komplex éterové koruny s iontem draslíku či komplexy adamantánů s cyklo-dextriny. Následně jsou tyto přístupy aplikovány na komplexy koronavirové proteázy Mpro a jejích inhibitorů.

Cíle práce

Prvním cílem práce je ověření využití nerovnovážných metod výpočtu volné energie na jednoduchých modelových systémech, pro které jsou známé referenční hodnoty volné energie z experimentu či z předchozích rovnovážných molekulárně-dynamických simulací.

Za tímto účelem je nejprve studována hydratační volná energie postranních řetězců aminokyselin. Cílem je odladit technické detaily spojené s volbami různých simulačních parametrů či zahrnutím korekcí pro nabitě molekuly nebo oříznutí van der Waalsova potenciálu. Jde také o výběr vhodného silového pole a testování vlastní reparametrizace provedené na základě ab initio výpočtů.

Pro ověření výpočtů vazebné volné energie na základě termodynamických cyklů je nejprve studován komplex éterové koruny a draslíku. Na tomto extrémně jednoduchém modelovém systému typu hostitel-host (angl. host-guest), je zkoumán vliv délky a počtu nerovnovážných trajektorií na výslednou hodnotu volné energie, možnost zavedení dodatečných omezujících potenciálů či aplikace Jarzynského rovnosti a Crooksova fluktuačního teorému.

Takto prověřené a odladěné metody jsou pak aplikovány na složitější komplex adamantánů s cyklodextriny. Adamantány jsou mimo jiné zkoumány v souvislosti s inhibicí vazby tzv. Spike proteinu SARS-CoV-2 na receptory lidských buněk. Cyklodextriny zajišťují transport léků. Prvním dílčím cílem je zde nalezení samotných vazebných konformací komplexů. Druhým dílčím cílem je určení vazebné volné energie pro všechny kombinace šesti studovaných adamantánů a tří cyklodextrinů.

Poznatky získané při studiu jednoduchých modelových systémů jsou nakonec využity pro studium inhibice koronavirové proteázy Mpro. První sadou studovaných molekul jsou nekovalentní inhibitory, pro které jsou v literatuře dostupné referenční hodnoty vazebné volné energie získané však prostřednictvím velmi odlišného způsobu výpočtu než je ten, který je použit v rámci této práce.

Poslední sadu studovaných molekul představují kovalentní inhibitory koronavirové proteázy Mpro, mezi nimiž je i klinicky používaný Paxlovid. Opět jde o určení vazebné volné energie. Přitom je mimo jiné třeba prozkoumat vliv protonace histidinů a konformační proměnlivosti aktivního místa proteázy na konformaci ligandů a tudíž i na výsledné hodnoty vazebné volné energie.

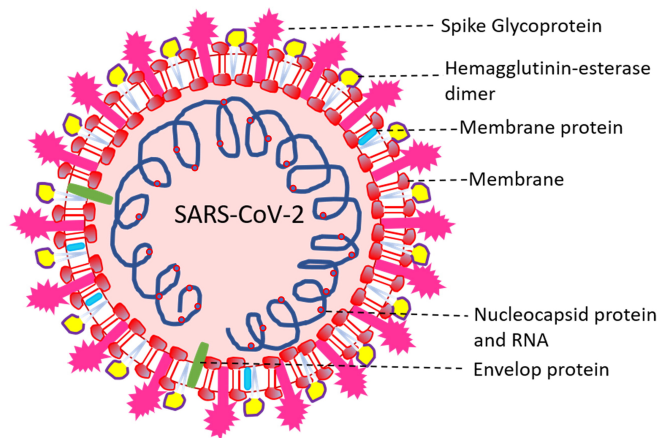
1. SARS-CoV-2

Hlavním cílem této práce je prozkoumat možnosti využití počítačových simulací pro identifikaci potenciálních inhibitorů koronaviru SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2), který způsobil pandemii onemocnění COVID-19, s jejímiž následky se vypořádáváme dodnes. Počáteční příznaky onemocnění zahrnují horečku, kašel, únavu atd. Projevují se do týdne od dne nákazy. Při závažném průběhu se do dvou týdnů objeví pneumonie, která může vést až k smrti nakaženého.

Pro využití počítačových simulací k identifikaci potenciálních inhibitorů virových proteinů, které by umožnily efektivní léčbu onemocnění COVID-19, je zapotřebí znát detailní strukturu a životní cyklus viru SARS-CoV-2. To nám umožní identifikovat klíčové kroky v replikaci viru, jejichž narušením je možné šíření viru buďto zpomalit, nebo zcela zastavit.

1.1 Virionová částice

Genom koronaviru SARS-CoV-2 je složen z přibližně 30 000 nukleotidů. Je v něm uložena informace o poměrně malém počtu proteinů. Dle funkce rozlišujeme 4 strukturální a 16 nestrukturních proteinů koronaviru. Struktura virionu je schématicky znázorněno na obrázku 1.1 [1].



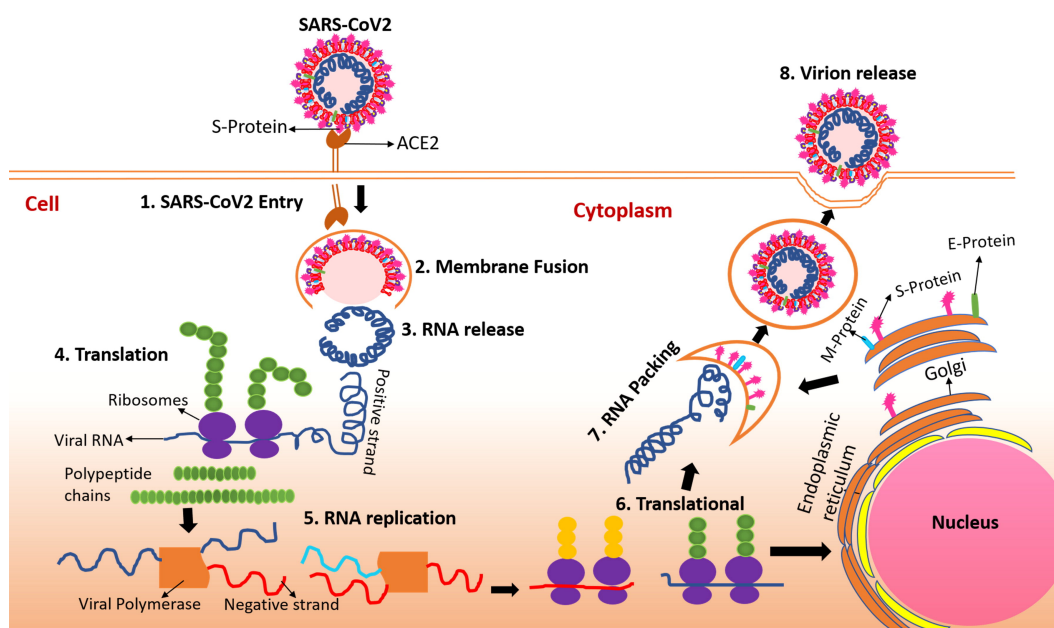
Obrázek 1.1: Schématické znázornění struktury virionu SARS-CoV-2. Strukturální proteiny se vyskytují především v membráně viru kromě N-proteinu, který tvoří komplex s virovou RNA [1].

Virionová částice je obalena membránou, ve které jsou umístěny zejména strukturální proteiny. M-protein (angl. Membrane protein) se vyskytuje především na povrchu membrány a je zásadní při tvorbě nových virionových částic. E-protein (angl. Envelope protein) je menší membránový protein zajišťující vhodnou permeabilitu membrány. S-protein (angl. Spike protein) je zabudován do virové membrány, ale z velké části z ní vyčnívá ven, což umožňuje viru rozeznat receptory hostitelských buněk. S-protein zajišťuje navázání viru na hostitele, následnou fúzi membrán a vniknutí viru do hostitelské buňky. Při tomto procesu dochází k podstatným konformačním změnám S-proteinu, který tak představuje ideální

cíl pro potenciální inhibitory, které by těmto změnám mohly zabránit. Funkce HE-dimeru (angl. Hemagglutinin-esterase dimer) není zatím přesně známa, ale předpokládá se, že hraje roli při infekci hostitelských buněk. Uvnitř virionu je ukryta virová RNA v komplexu se strukturálním N-proteinem (angl. Nuclear capsid protein), který hraje důležitou roli při virové replikaci a transkripci. Nestrukturní proteiny nejsou ve virionu přítomny, jsou vytvářeny až při translaci v hostitelské buňce.

1.2 Životní cyklus koronaviru

Pro efektivní návrh potenciálních inhibitorů je zapotřebí znát detailně mechanismy toho, jak virus napadá hostitelské buňky a jak se replikuje. Hlavní kroky životního cyklu koronaviru jsou schématicky znázorněny na obrázku 1.2 [1].



Obrázek 1.2: Schématické znázornění životního cyklu viru SARS-CoV-2. Vstup koronaviru do buňky je zajištěn S-proteinem, který rozpozná hostitelské ACE2 receptory a spustí fúzi membrán. Virová RNA je pak uvolněna do cytoplazmy. Translací virové RNA vznikají neaktivní polyproteiny. Působením virových proteáz jsou odděleny a aktivovány jednotlivé proteiny, které pak přispívají k replikaci virové RNA. Nově syntetizované strukturální proteiny a vlákno virové RNA s pozitivní polaritou nakonec vytvářejí nové virionové částice a opouštějí buňku exocytózou [1].

Proces napadení hostitele začíná navázáním Spike proteinu na receptory ACE2 (Angiotensin Converting Enzyme 2), které se nacházejí na povrchu mnoha lidských buněk. Spike protein je následně štěpen na dvou místech hostitelskými proteázami trypsin a furin, což vede k fúzi membrán. Vir se tak do hostitelských buněk dostává pomocí endocytózy.

Endozom se poté otevře a uvolní virovou RNA do hostitelské cytoplazmy. N-proteiny obalující virovou RNA jsou postupně odstraňovány proteazomy. Vlákno virové RNA s pozitivní polaritou je přeloženo ribozomy hostitele do dlouhých po-

lyproteinových řetězců. Při translaci zde dochází k posunu ribozomálního čtecího rámce (angl. ribosomal frameshift) [2], což umožní syntetizovat z jednoho vlákna virové RNA dva různě dlouhé polyproteiny, označované pp1a a pp1ab. Jim odpovídající sekvence virové RNA jsou označovány jako ORF1a a ORF1b (angl. open reading frame).

Vytvořené polyproteiny jsou následně štěpeny virovými proteázami na 16 nestrukturních proteinů (NSP, nonstructural proteins). Virové proteázy jsou přitom na počátku též zabudované v polyproteinech. Patří sem hlavní proteáza (Mpro, main protease) a PLpro (papain-like protease), což jsou 5. a 3. nestrukturní protein v polyproteinech pp1a a pp1ab.

Většina z ostatních nestrukturních proteinů se pak podílí na tvorbě virového replikačně-transkripčního komplexu (RTC), který syntetizuje komplementární RNA vlákno s negativní polaritou RNA(-), jež dále slouží jako vzor pro syntézu dalších vláken RNA s pozitivní polaritou RNA(+), což umožňuje množení viru.

Dále může dojít i k tzv. diskontinuální transkripci [3], při které vznikají různě dlouhé subgenomické mRNA, jejichž translací následně vznikají různé virové proteiny. Vir tak může regulovat množství vytvářených proteinů. Diskontinuální transkripci přitom vznikají především strukturní proteiny.

Důležitou roli přitom hraje 1. nestrukturní protein NSP1, který se váže k hostitelským ribozomům. Blokuje translaci lidské mRNA a naopak podporuje translaci virové RNA [2]. Vir tak postupně přebírá kontrolu nad veškerou translací probíhající v hostitelské buňce.

Strukturní proteiny jsou syntetizovány v cytoplazmě hostitelských buněk a následně přemístěny do endoplazmatického retikula a dále do Golgiho aparátu. V cytoplazmě také dochází k obalování nově vzniklé virové RNA N-proteinem a k jejich shlukování v blízkosti Golgiho aparátu. Postupně tak vznikají nové viriony, které pomocí exocytózy opouštějí původní hostitelskou buňku. Celý proces se pak opakuje od začátku.

1.3 Spike protein

Struktura S-proteinu je znázorněna na obrázku 1.3. Lze ji rozdělit na tři hlavní části: ektodoménu, transmembránovou kotvu a krátký ocásek uvnitř viru. Ektodoména je složena z trimeru S1 podjednotek, jehož hlavní funkcí je rozpoznání a navázání se k hostitelským ACE2 receptorům. RBD (angl. receptor binding domain) je specifická část S1 podjednotek, která se váže na ACE2 receptory. Součástí ektodomény jsou dále S2 podjednotky, které zajišťují následnou fúzi membrán.

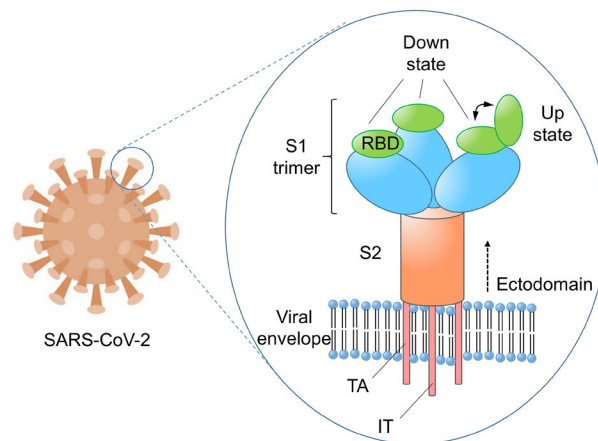
Spike protein se může vyskytovat ve dvou konformacích. V otevřené konformaci je více flexibilní a může se vázat na ACE2 receptory. V uzavřené konformaci je RBD schována uvnitř Spike proteinu a nemůže tedy ACE2 receptory efektivně rozeznávat. Tato konformační změna Spike proteinu chrání do určité míry koronavirus vůči hostitelskému imunitnímu systému [4]. Ke stejnému účelu slouží i glykosylace Spike proteinu. Navázané glykany pokrývají povrch Spike proteinu a brání vir před rozeznáním hostitelskými protilátkami.

Oproti SARS-CoV je Spike protein viru SARS-CoV-2 štěpen pomocí furinu (proteázy hostitelských buněk) mezi podjednotkami S1 a S2 [5]. Vzniká tak S1/S2 heterodimer. Podjednotka S2 je dále štěpena hostitelskými proteázami jako je

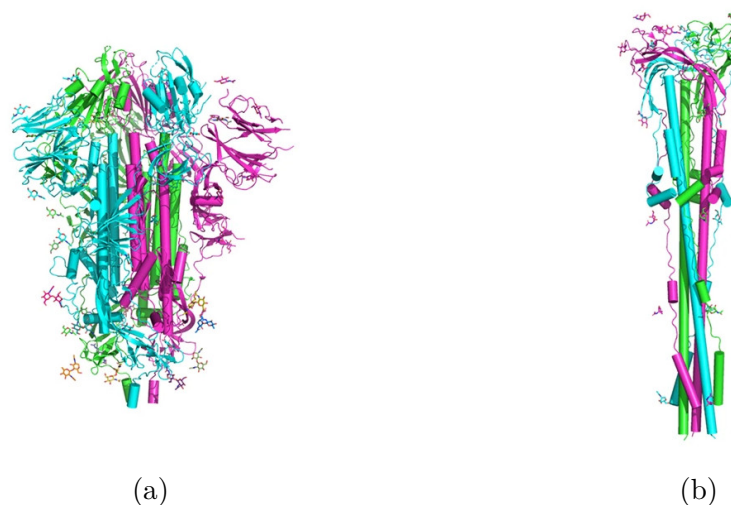
např. TMPRSS (angl. transmembrane-protease serin 2), čímž dojde k odkrytí peptidu, který zprostředkovává fúzi membrán. Po navázání S1 podjednotky na ACE2 receptor je prefúzní struktura Spike proteinu destabilizována a podjednotka S1 se disociuje. V postfúzní konformaci tvoří podjednotka S2 strukturu sestávající z šesti pevně vázaných helixů. Obě tyto struktury jsou zobrazeny na obrázku 1.4.

Spike protein viru SARS-CoV-2 se váže k ACE2 s větší afinitou než Spike protein viru SARS-CoV [6]. Vazba je zprostředkována především polárními aminokyselinami na jednom z alfa-helixů receptoru ACE2 a aminokyselinami z RBD domény Spike proteinu, jak je detailně zobrazeno na obrázku 1.5 [7].

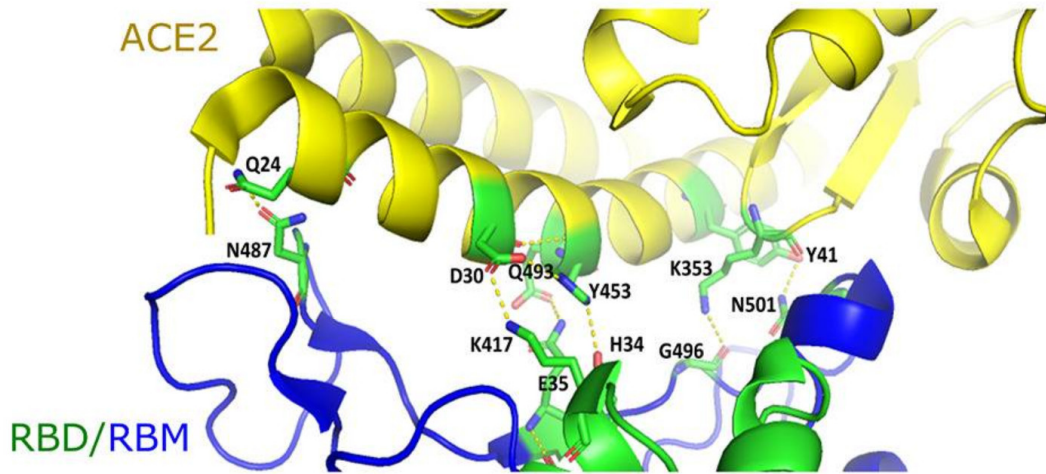
Inhibice ACE2 receptoru může zabránit koronaviru v rozpoznání hostitelské buňky. Vir pak nemůže do buňky vniknout. Jeden z možných přístupů v boji proti infekci SARS-CoV-2 je tak založený na inhibici ACE2 receptorů prostřednictvím modifikovaných molekul adamantánů [8]. V rámci této práce byla studována schopnost adamantánů vytvářet komplexy s cyklodextriny, které mohou sloužit jako nosiče léků.



Obrázek 1.3: Schématické znázornění struktury spike (S) proteinu. Rozlišení a následné navázání hostitelských ACE2 receptorů obstarává S1 podjednotka, zatímco podjednotka S2 zajišťuje následnou fúzi membrán [4].



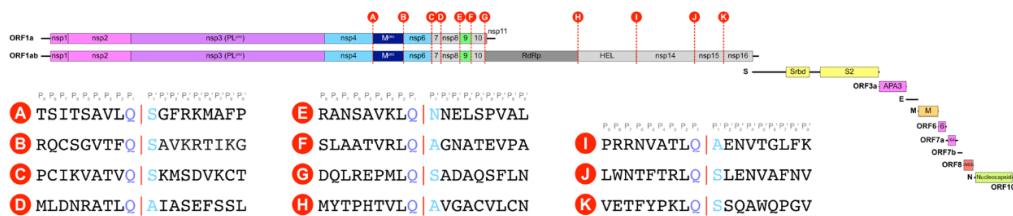
Obrázek 1.4: (a) Prefúzní struktura a (b) postfúzní struktura Spike proteinu [5].



Obrázek 1.5: Hlavní interakce zprostředkávající vazbu ACE2 receptoru a Spike proteinu [7].

1.4 Proteáza Mpro

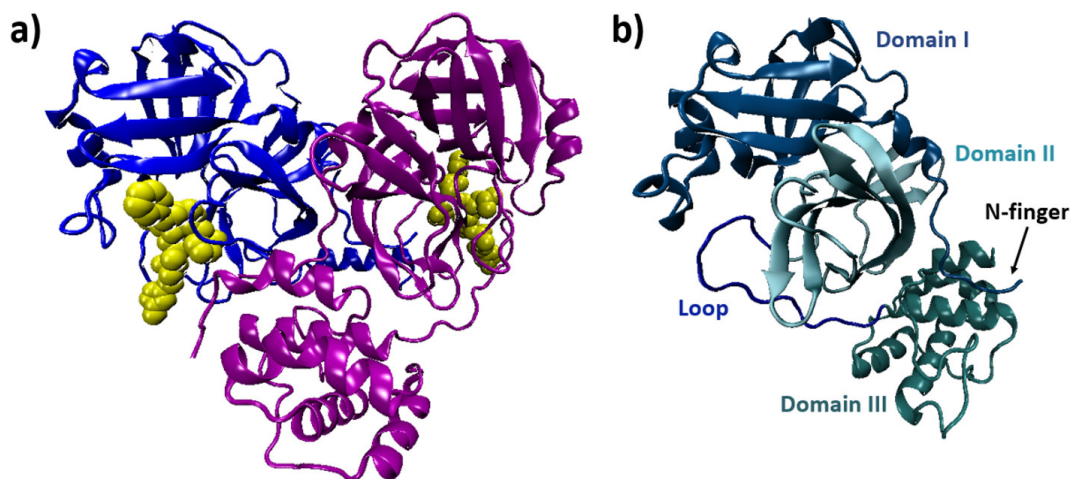
Struktura genomu SARS-CoV-2 je schematicky znázorněna na obrázku 1.6. Jsou zde rozlišeny části genomu kódující polyproteiny pp1a a pp1ab, které jsou štěpeny proteázami PLpro a Mpro. Proteáza PLpro odštěpuje z polyproteinů nestrukturní proteiny 1 až 3. Hlavní proteáza Mpro (někdy také označovaná 3CLpro, angl. 3-chymotrypsin-like protease) odštěpuje nestrukturní proteiny 4 až 16. Bez tohoto štěpení jsou virové proteiny v polyproteinu neaktivní. Virové proteázy tak představují ideální cíl inhibitorů.



Obrázek 1.6: Schématické znázornění struktury genomu SARS-CoV-2 [9]. Genom lze rozdělit na části kódující nestrukturní a strukturní proteiny. Červeně jsou označena místa působení proteázy Mpro a je pro ně uvedena i odpovídající sekvence aminokyselin.

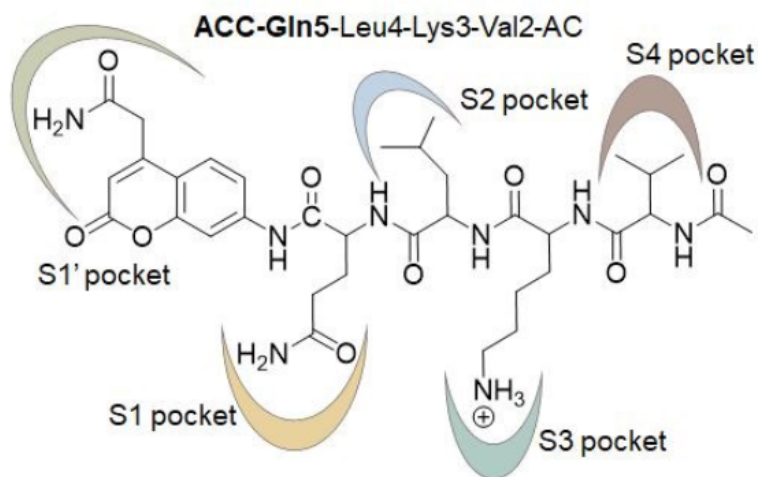
První struktura SARS-CoV-2 Mpro v komplexu s kovalentním inhibitorem N3 byla získána prostřednictvím rentgenové krystalografie (vědci mimo jiné i přímo z Wuhanu) již na počátku roku 2020. 26. ledna 2020 byla tato struktura vložena do proteinové databanky s identifikátorem 6LU7. O něco později přibyla ještě její upřesněná verze 7BQY. A již 9. února 2020 byl do časopisu Nature odeslán doprovodný manuskript [10].

Tato historicky první struktura SARS-CoV-2 Mpro je znázorněna na obrázku 1.7. Proteáza je aktivní ve formě homodimeru. Jednotlivé protomery jsou přitom na sebe orientované přibližně kolmo. Každý z protomerů lze rozdělit do 3 domén. Z hlediska sekundární struktury jsou domény I a II tvořeny převážně beta-skládanými listy, zatímco v doméně III převažují alfa-šroubovice.



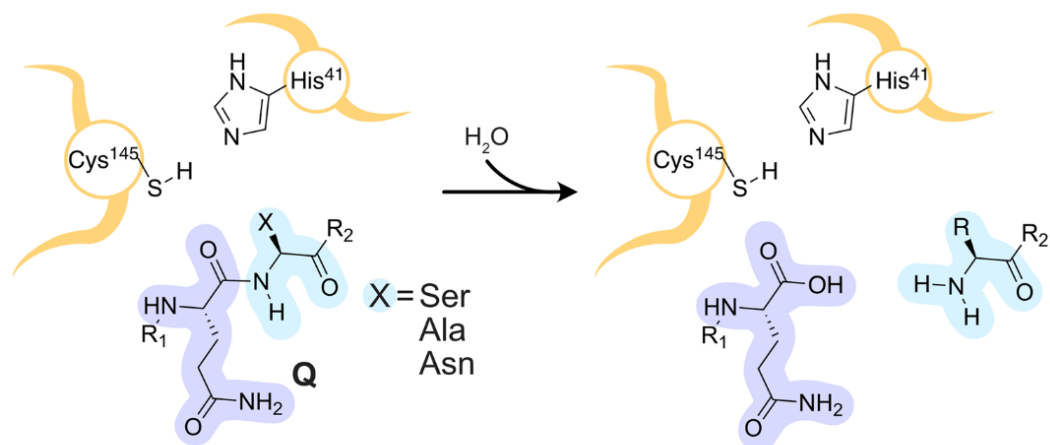
Obrázek 1.7: Krystalová struktura Mpro (PDB ID 6LU7) (a) homodimer (b) jeden protomer s vyznačenými třemi doménami [4].

Katalytické aktivní místo proteázy Mpro (obrázek 1.8) se nachází mezi doménami I a II a je tvořeno dyádou His41 a Cys145. Detail katalytické dyády je zobrazen na obrázku 1.9. Proteáza MPro štěpí polyproteiny výhradně za aminokyselinou glutaminem [11], jak je schematicky znázorněno na obrázku 1.9. Kromě toho proteáza Mpro preferuje při štěpení peptidů hydrofobní aminokyselinu na pozici před glutaminem (především leucin), a spíše malou aminokyselinu bezprostředně za místem štěpení (obvykle serin nebo alanin). Jelikož není známá žádná lidská proteáza s obdobnými vlastnostmi, tak lze předpokládat, že inhibitory Mpro budou mít minimální vedlejší účinky.



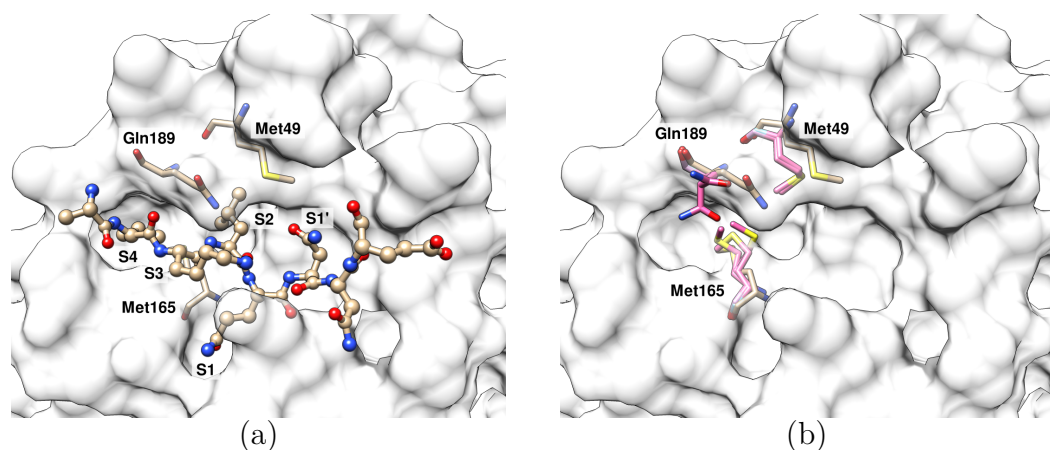
Obrázek 1.8: Schématické znázornění interakcí proteázy Mpro s peptidy v blízkosti katalytické dyády Cis-His [12].

Postupně byly získány komplexy proteázy Mpro s částmi štěpených peptidů jak pomocí krystalografie a rentgenové difrakce [13, 14] tak prostřednictvím elektronového mikroskopu [15]. Byly zachyceny stavy jak bezprostředně před tak i bezprostředně po přestípnutí peptidů [16].



Obrázek 1.9: Obecná reakce katalyzovaná proteázou Mpro [9]. R1 a R2 zde obecně značí zbylé části štěpeného peptidu. X značí možné postranní řetězce aminokyselin, u kterých dochází ke štěpení.

Struktura proteázy Mpro bez substrátu (v apo formě) se může v okolí aktivního místa značně lišit od struktur s navázaným peptidem, jak ukazuje obrázek 1.10. Postranní řetězce několika aminokyselin mají v apo formě Mpro jinou konformaci. Největší rozdíly lze pozorovat u aminokyselin Gln189, Met49 a Met165. Jejich postranní řetězce více či méně okupují místa či kavity, do kterých se obvykle váží peptidy či inhibitory. Aktivní místo proteázy Mpro je tedy poměrně mělké, konformačně plastické a do jisté míry i kryptické. To z Mpro činí poměrně obtížný cíl, pokud jde o racionální design potenciálních inhibitorů.



Obrázek 1.10: Rozdíl v konformaci aktivního místa proteázy Mpro (a) s navázaným peptidem (PDB ID 7DVP) a (b) v apo formě (PDB ID 6M03, 6Y2E).

Kovalentní inhibitor N3, který je přítomen v historicky první krystalové struktuře (6LU7) koronavirové proteázy Mpro z ledna 2020, byl vytvořen již v roce 2005 v rámci vývoje širokospektrálních inhibitorů proteáz koronaviřů [17]. Na jeho vývoji se podílel i Rolf Hilgenfeld, který se vývojem inhibitorů proteáz virů SARS a MERS zabývá již 20 let [18]. Společně se známým virologem Christianem Drostenem a dalšími vložili do proteinové databanky vlastní krystalové struktury Mpro již v únoru 2020 [19]. Zkoumali přitom potenciální inhibitory

na bázi α -ketoamidů. Vycházeli přitom z inhibitoru 11r, který v subnanomolární koncentraci účinně inhiboval vir MERS [20]. Ještě dříve byly podobné látky testovány v klinických zkouškách jako inhibitory proteázy Rhinoviru a tedy jako léky proti běžnému nachlazení [21]. V proteinové databance lze prostřednictvím identifikátoru 1CQQ snadno nalézt krystalovou strukturu rhinovirové proteázy v komplexu s inhibitorem AG7088 z roku 1999. Všechny výše zmíněné inhibitory obsahují společný strukturní motiv - analog kruciálního glutaminu [21].

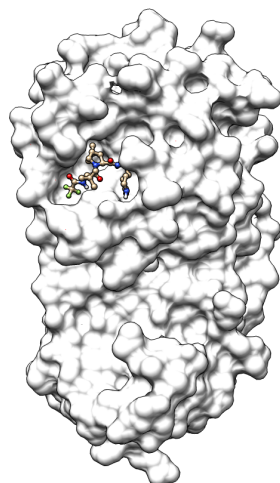
Ten obsahuje i látka PF-00835231, která byla roku 2003 zkoumána jako možný inhibitor viru SARS-CoV-1 [22]. Jelikož ale hrozba šíření SARS-CoV-1 byla ve své době velmi rychle vyřešena, tak struktura inhibitoru PF-00835231 byla firmou Pfizer publikována až v roce 2020 v souvislosti s pandemií COVID-19. Jelikož tento inhibitor nefungoval dobře při orálním podání, tak byl firmou Pfizer dále vylepšován, až se dospělo k látce PF-07321332, která je více známa pod názvy Nirmatrelvir či Paxlovid [23–25]. Paxlovid je zatím nejúspěšnější inhibitor proteázy Mpro, je hojně klinicky využíván.

Pro Paxlovid byla získána řada krystalových struktur deponovaných v proteinové databance, díky nimž lze studovat, jakým způsobem dochází k efektivní inhibici proteázy Mpro. Struktura celé proteázy Mpro s navázaným inhibitorem PF-07321332 je zobrazena na obrázku 1.11. Obrázek 1.12 pak ukazuje detailní srovnání aktivního místa Mpro jednak s peptidem a jednak s inhibitorem PF-07321332. Je zřejmé, že pro efektivní inhibici proteázy Mpro je optimální, když je inhibitor pevně navázaný v blízkosti katalytické dyády a interaguje s částmi proteázy označenými jako S1, S2, S3 a S4, kam se běžné vážou aminokyseliny štípaného peptidu. Jelikož proteáza Mpro štěpí peptidy výhradně za glutaminem, tak je pro inhibitory užitečné, když ve své struktuře obsahují analog glutaminu, který interaguje s příslušným vazebným místem S1 stejně jako samotný glutamin. Jak ukazuje obrázek 1.13, Paxlovid má ve své struktuře stejný analog glutaminu, jako výše zmíněné historické inhibitory. Zajímavé je, že kovalentní vazba inhibitoru PF-07321332 na Cys145 v aktivním místě Mpro je zřejmě vratná, neboť v krystalové struktuře 7VH8 byly zachyceny dvě různé konformace postranního řetězce tohoto cysteinu z katalytické dyády. Pouze jedna z těchto konformací je přitom vhodná pro tvorbu kovalentní vazby s inhibitorem, jak můžeme vidět na obrázku 1.14. Již nyní je zřejmé, že Paxlovid bude potřeba dále modifikovat, neboť v buněčných kulturách byly vyšlechtěny mutované varianty koronaviru, které jsou vůči Paxlovidu rezistentní. Můžeme proto očekávat, že se podobné varianty dříve či později objeví i v běžné populaci [26, 27].

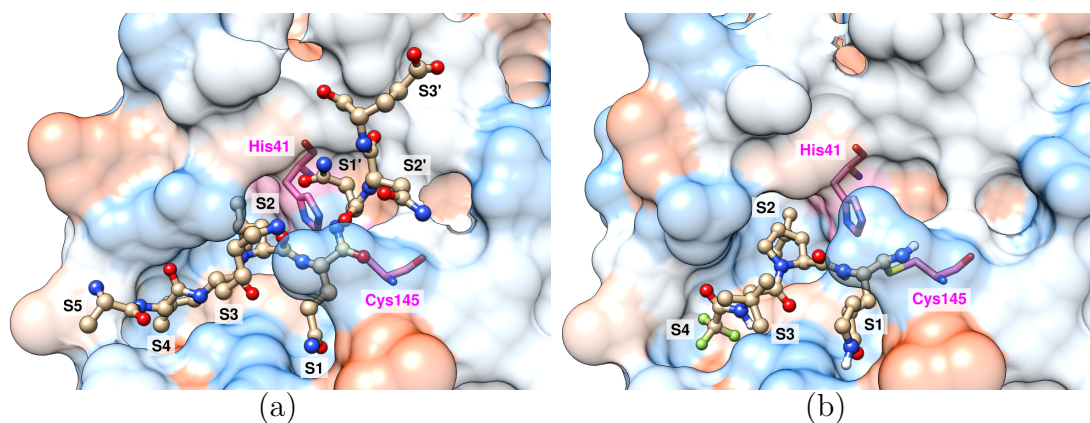
Kromě historické linky sahající až do 90. let a vedoucí k Paxlovidu byla vyzkoušena celá řada možných přístupů k inhibici proteázy Mpro [28]. Obecně lze inhibitory rozdělit do čtyř hlavních skupin: i) kovalentní inhibitory imitující peptidy, ii) kovalentní inhibitory nezaložené na analogii s peptidy, iii) nekovalentní inhibitory blokuující vstup štěpených polyproteinů do aktivního místa Mpro iv) nekovalentní inhibitory blokuující aktivitu Mpro alostericky.

O racionální design inhibitorů proteázy Mpro se od roku 2020 usiluje i prostřednictvím počítačového modelování [29–34].

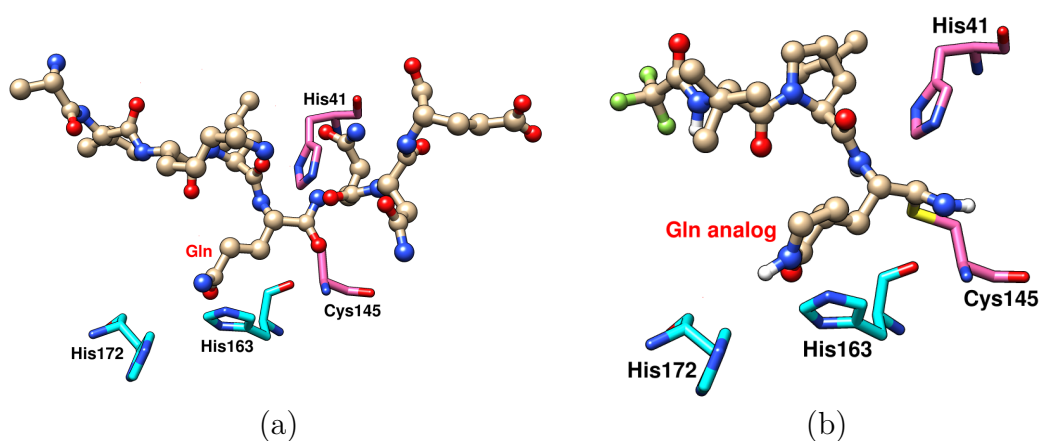
Velká část studií byla provedena s využitím dockingu, který umožňuje velmi rychle porovnat obrovské množství potenciálních inhibitorů za cenu toho, že vazebná volná energie je určována prostřednictvím jednoduchých skórovacích funkcí a výstupy z dockingu jsou tak spíše orientační, obsahují řadu falešně pozitivních



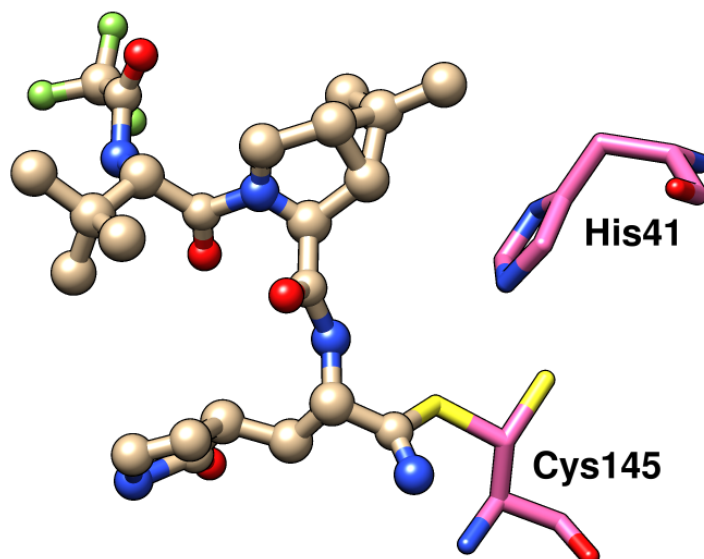
Obrázek 1.11: Vazba inhibitoru PF-07321332 na proteázu Mpro (PDB ID 7RFS).



Obrázek 1.12: Detail aktivního místa proteázy Mpro (a) s peptidem (PDB ID 7MGR) a (b) inhibitrem PF-07321332 (PDB ID 7RFS).



Obrázek 1.13: Znázornění hlavních interakcí (a) peptidu štěpeného za glutami- nem (PDB ID 7MGR) a (b) inhibitrem PF-07321332 obsahující analog glutaminu (PDB ID 7RFS).



Obrázek 1.14: Dvě konformace cysteinu z katalytické dyády pozorované v krystalové struktuře proteázy Mpro (PDB ID 7VH8).

nálezů. Ilustruje to fakt, že dvě studie, ve kterých byla prostřednictvím dockingu proseta vždy více než miliarda ligandů, nevedly k nalezení inhibitoru s experimentálně potvrzenou schopností inhibovat Mpro v dostatečně nízkých koncentracích [35–37].

Úspěšná byla studie vycházející z počátečního dockingu 235 milionů ligandů, díky čemuž byl po řadě optimalizací vytvořen inhibitor s aktivitou srovnatelnou s Paxlovidem [38].

K hledání inhibitorů Mpro byly využity i fyzikálně korektní výpočty vazebné volné energie prostřednictvím velkého množství krátkých nerovnovážných trajektorií. Tento přístup je ovšem velmi náročný na počítačový čas. Proto byla využita exa-scale kapacita gridového projektu FoldingHome, který k výpočtům využívá grafické karty počítačů stovek tisíc dobrovolníků z celého světa. Výpočty vazebné volné energie zde byly využity k prioritizaci ligandů pro syntézu v rámci megaprojektu COVID Moonshot a přispěly k nalezení nanomolárních inhibitorů Mpro srovnatelných s Paxlovidem [39].

V této práci jsme se inspirovali i články skupiny P. Procacciho, která metodiku výpočtu vazebné volné energie prostřednictvím velkého množství krátkých nerovnovážných MD simulací intenzivně zkoumá a rozvíjí posledních 10 let a též se zapojila do úsilí o nalezení efektivních inhibitorů koronavirové Mpro [40–44].

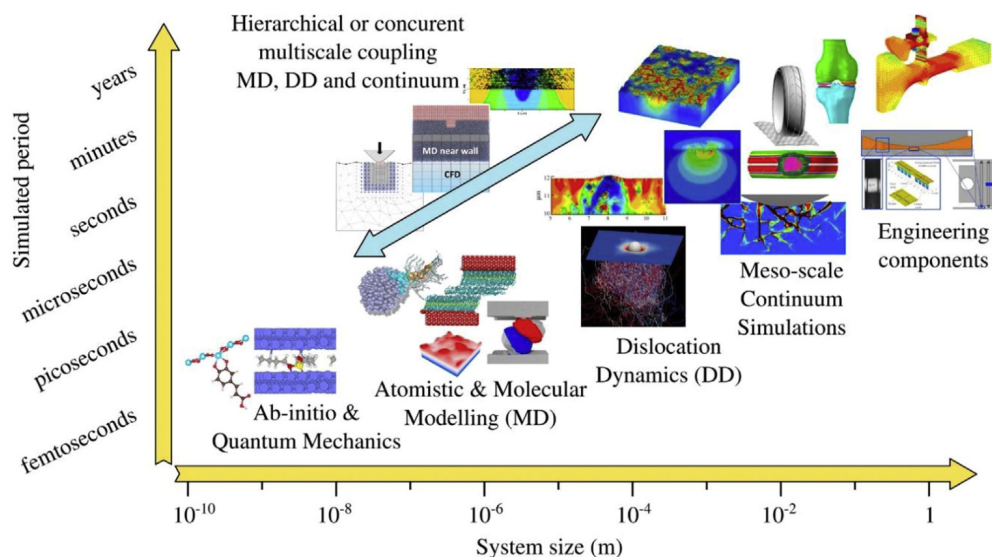
2. Molekulární dynamika

Molekulární dynamika (MD) umožňuje studovat pohyb molekul na atomární úrovni, což je mimo jiné přínosné pro pochopení toho, jak fungují velké biomolekulární systémy. MD simulace doplňují experimentální metody (rentgenovou kystalografii, NMR, cryoEM), které produkují statické struktury biomolekul často v mezních konformacích.

Základní myšlenka molekulární dynamiky je prostá. Máme-li počáteční souřadnice atomů, můžeme spočítat síly, které na atomy působí a numerickým řešením pohybových rovnic tak získat dynamický vývoj daného molekulárního systému [45].

Dle úrovně teorie, kterou při popisu molekulárního systému používáme, rozlišujeme klasickou MD, kde je pohyb jader atomů i vliv elektronů popsán klasicky, ab-initio MD, která elektrony popisuje kvantově a jádra klasicky, a nakonec kvantově-mechanickou MD, jež popisuje jádra i elektrony pomocí kvantové mechaniky.

Úroveň teorie popisu rozhoduje o tom, pro jak velké systémy a na jak dlouhých časových škálách jsme schopni MD simulace provést. Zatímco prostřednictvím klasické MD jsme schopni simulovat řádově 100 000 atomů po dobu až několika μs , tak ab-initio a kvantově-mechanické MD simulace jsou se současnými výpočetními prostředky proveditelné pouze pro o několik řádů menší molekulární systémy (desítky či stovky atomů) a na podstatně kratších časových škálách. To názorně ukazuje obrázek 2.1.



Obrázek 2.1: Velikost simulovaných systémů a časové škály, pro něž jsou se současnými výpočetními prostředky proveditelné MD simulace na různých úrovních teorie popisu molekul [46].

2.1 Silové pole

Molekulární interakce je v principu potřeba popisovat pomocí kvantové mechaniky. Vliv kvantových jevů ovšem může být aproximován. Základem je obvykle Born-Oppenheimerova aproximace, která umožňuje oddělit popis pohybu jader a elektronů. Předpokládá se, že jádra se pohybují v efektivním poli tvořeném elektrony, které získáme řešením bezčasové Schrödingerovy rovnice pro elektrony. Klasickou aproximaci pak získáme, pokud efektivní pole tvořené elektrony popíšeme jednoduchou analytickou funkcí a pohyb jader prostřednictvím klasických Newtonových pohybových rovnic.

Konkrétní tvar silového pole záleží na tom, které fenomény potřebujeme studovat. Pro získání korektních vibračních spekter molekul je potřeba volit silová pole s tzv. cross-termy popisujícími vazby mezi délkami vazeb a velikostmi valenčních či dihedrálních úhlů. Pokud naopak potřebujeme popsat velké biomolekulární systémy na dlouhých časových škálách, tak volíme co nejjednodušší (tzv. diagonální) silová pole. Příklad takového silového pole je uveden na obrázku 2.2.

$$E_{total} = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

Obrázek 2.2: Příklad tvaru silového pole [47].

Celkovou energetickou nadplochu E_{total} , po které se pohybují jádra, je možné rozdělit na dvě skupiny členů popisující vazebné a nevazebné interakce.

Vazebné členy popisují síly působící na atomy v rámci jedné molekuly. V souřadném systému dané molekuly je možné tyto členy zapsat prostřednictvím délek vazeb a velikostí valenčních či dihedrálních úhlů. Jednotlivé členy jsou přitom jednoduché funkce, které vystihují charakteristický tvar příslušného potenciálu. Pro délku vazeb je tak například běžné používat harmonický potenciál, který je adekvátní zejména v blízkosti rovnovážné hodnoty. Je ale zřejmé, že nedokáže popsat vznik či přestípnutí vazby - disociaci molekul.

Nevazebné interakce lze rozdělit do dvou skupin na elektrostatické a van der Waalsovy.

Elektrostatické interakce je v MD simulacích běžné aproximovat pomocí bodových parciálních nábojů a Coulombova zákona [48]. Každému atomovému jádru je přiřazen neceločíselný náboj. Není to skutečná fyzikální veličina. Hodnoty parciálních nábojů se v různých silových polích mohou lišit v závislosti na metodě použité při jejich parametrizaci. V silovém poli AMBER jsou voleny tak, aby byl co nejlépe aproximován elektrostatický potenciál v okolí molekuly. V silovém poli CHARMM jde o co nejlepší aproximaci interakce s okolními molekulami vody.

Van der Waalsovy interakce zahrnují celou třídu interakcí mezi dipóly a indukovanými dipóly. Zahrnují tak Londonovy, Debyeovy a další síly. V MD simulacích se obvykle popisují společně s Pauliho repulzí pomocí Lennard-Jonesova potenci-

álu [49]. Ten reprezentují první dva členy v ne vazebných interakcích na obrázku 2.2.

Pro popis molekul vody byl v této práci zvolen model TIP3P [50], který molekulu vody popisuje jako rigidní molekulu se třemi bodovými náboji a Lennard-Jonesovými parametry na polohách atomů.

2.2 Silové pole CHARMM

Pro popis koronavirové proteázy Mpro bylo v rámci této práce zvoleno silové pole CHARMM36 [51], které se hojně využívá pro MD simulace proteinů, nukleových kyselin a lipidů.

Silové pole CHARMM obsahuje následující členy [51]:

$$U(\mathbf{R}) = \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_\phi[1 + \cos(n\phi - \delta)] \quad (2.1)$$

$$+ \sum_{\text{Urey-Bradley}} K_{UB}(S - S_0)^2 + \sum_{\text{impropers}} K_\omega(\omega - \omega_0)^2 \quad (2.2)$$

$$+ \sum_{\text{residues}} U_{\text{CMAP}}(\phi, \psi) \quad (2.3)$$

$$+ \sum_{\text{non-bonded pairs}} \left\{ \epsilon_{ij}^{\min} \left[\left(\frac{R_{ij}^{\min}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^{\min}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 \epsilon r_{ij}} \right\}, \quad (2.4)$$

kde první řádek 2.1 představuje tradiční vazebné členy zahrnující délky vazeb b , valenční úhly θ a dihedrální úhly ϕ . Silové pole CHARMM neobsahuje cross-termy, ani vyšší mocinné členy. Vazby a valenční úhly jsou popsány pomocí jednoduchých harmonických funkcí a dihedrální úhly pomocí periodické funkce, kde n značí multiplicitu a δ fázový posun. Silové pole CHARMM tak představuje kompromis mezi přesností a rychlostí výpočtu, který dovoluje simulovat velké biomolekulární systémy na dlouhých časových škálách.

Druhý řádek 2.2 obsahuje vazebné členy, které se v jiných silových polích často nevyskytují. Tzv. Urey-Bradleyho potenciál spojuje terminální atomy A a C v trojici kovalentně vázaných atomů A-B-C. Nevlastní dihedrální člen (angl. improper dihedral) se používá v případě, kdy máme tři atomy hvězdicovitě vázané na jeden centrální atom a lze pomocí něj kontrolovat, zda centrální atom leží v rovině ostatních tří atomů, nebo mimo ní. Zavedením těchto členů je možné lépe popsat a ovládat prostorové uspořádání motivů vyskytující se v proteinech, jako jsou např. peptidové vazby, které jsou přibližně planární, a jejichž popis by bez zahrnutí těchto členů byl obtížný.

Třetí řádek 2.3 představuje tzv. CMAP korekci (angl. correction map) zavedenou pro další vylepšení popisu proteinů. Jedná se o numerickou korekci energie definovanou na diskrétní mříži pro tzv. Ramachandranovy úhly, což jsou dihedrální úhly popisující konformaci tzv. hlavního vlákna - páteře (angl. main chain - backbone) proteinu.

Poslední řádek 2.4 obsahuje ne vazebné členy, které jsou tradičně popsány pomocí Lennard-Jonesova potenciálu u van der Waalovy interakce a pomocí parciálních bodových nábojů v případě elektrostatické interakce.

Veličiny s indexem 0 značí rovnovážné hodnoty.

Tzv. silové konstanty K_i jsou parametry silového pole určené na základě experimentu či ab initio výpočtů a jejich hodnoty jsou specifické pro různé molekuly. Každému atomu je totiž v molekule přiřazen určitý atomový typ, který může být specifický např. i pro různé pozice stejného atomu v dané molekule. V konečném výsledku tak i atomy stejného druhu, např. atomy uhlíku ve funkční skupině ketonu, mohou být popsány různými atomovými typy. Silové konstanty K_i nabývají různých hodnot pro různé kombinace atomových typů.

Při parametrizaci silového pole CHARMM bylo využito toho, že proteiny jsou složeny z konečného množství základních stavebních kamenů - aminokyselin. Pro kompletní parametrizaci proteinových molekul tak stačí naparametrizovat jednotlivé aminokyseliny a peptidovou vazbu.

2.3 Silové pole CGenFF

V případě malých organických látek, které jsou zkoumány jako možné inhibitory proteinů a případně jako budoucí léčiva, je jejich parametrizace složitější v tom smyslu, že struktura těchto organických ligandů je obvykle poměrně unikátní. V principu je tedy ideální parametrizovat tyto ligandy jednotlivě a na základě kvantově-mechanických výpočtů v kombinaci s experimentálními daty.

Nicméně i zde lze využít toho, že jednotlivé ligandy sestávají z více či méně obvyklých dílčích strukturních motivů. Zejména pokud potřebujeme popsat a zkoumat větší množství ligandů, jejichž struktura je různorodá, tak je obvyklé použít v první fázi webových serverů, které ligandům atomové typy a silové konstanty přiřadí na základě analogie.

Pro popis ligandů v rámci této práce bylo zvoleno silové pole CGenFF (CHARMM General Force Field) [52], které bylo navrženo tak, aby bylo kompatibilní se silovým polem CHARMM a nedocházelo tak k míšení různých přístupů k parametrizaci molekul, které spolu mají v simulovaném systému koexistovat a interagovat.

Silové pole CGenFF bylo parametrizováno pro dvě třídy molekul. První zahrnovala řadu bioorganických heterocyklů. Druhá velké množství funkčních skupin. Příslušné parametry by tak měly být použitelné pro většinu malých bioorganických látek.

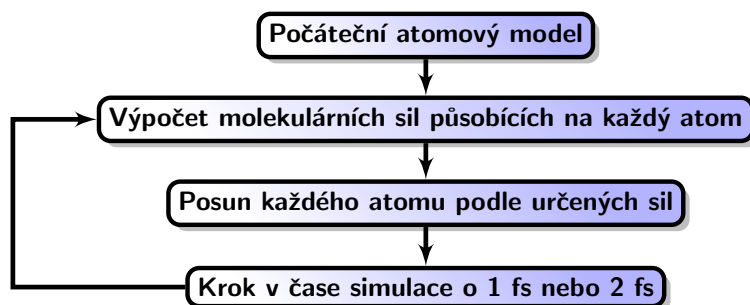
Parametrizace ligandů, které v CGenFF nejsou explicitně uvedeny, se provádí na základě analogie vhodnou kombinací dostupných parametrů. Pokud analogie není dostatečná, tak se doporučuje provést dodatečné kvantově-mechanické výpočty pro optimalizaci příslušných parametrů.

Pro silové pole CGenFF existuje algoritmus, který dokáže automaticky přiřadit atomové typy zadané molekule [53, 54] především na základě konektivity atomů. Přiřazení vazebných silových konstant K_i se provádí postupnou substitucí atomových typů do známých parametrů CGenFF. Každá substituce přidá určitou hodnotu k celkové penalizaci a kombinace substitucí s nejmenší penalizací pak určí volbu výsledné silové konstanty K_i . Parciální náboje jsou přiřazeny na základě rozšířeného “vazba-náboj” schématu, které je schopné popsat indukční a mezomerní efekty až do tří kovalentních vazeb.

2.4 Integrátory

MD simulace je možné zahájit, pokud jsou k dispozici počáteční souřadnice všech atomů simulovaného systému a pokud je přiřazeno vhodné silové pole všem jeho složkám (tj. proteinu, ligandu, vodní obálce, iontům atd.).

Řešení Newtonových pohybových rovnic pak lze provést pomocí integrátorů, které diskretizují čas na konečné časové kroky a v každém takové časovém kroku aproximují pohyb molekul, jak je schématicky znázorněno na obrázku 2.3.



Obrázek 2.3: Schématické znázornění řešení Newtonových pohybových rovnic pomocí integrátorů s konečným časovým krokem [47].

Konkrétní tvar integrátoru se může lišit v důsledku volby např. termostatu apod. Základní odvození vychází z Liouvillova operátoru iL [55] definovaného na fázovém prostoru vztahem:

$$iL = \sum_{j=1}^{3N} \left[x_j \frac{\partial}{\partial x_j} + F_j \frac{\partial}{\partial p_j} \right], \quad (2.5)$$

kde x_j a p_j značí kanonickou souřadnici a hybnost, N je rovno počtu atomů v simulaci a F_j představuje sílu působící na j -tý stupeň volnosti.

Klasický propagátor $U(t)$ pak lze vyjádřit jako exponenciálu z Liouvillova operátoru

$$U(t) = e^{iLt}.$$

Liouvillov operátor lze rozložit na součet dvou operátorů $iL = iL_1 + iL_2$, které odpovídají dvou členům v rovnici 2.5 a aplikovat Trotterův teorém [55]

$$e^{i(L_1+L_2)t} = \left[e^{iL_1 \frac{\Delta t}{2}} e^{iL_2 \Delta t} e^{iL_1 \frac{\Delta t}{2}} \right]^P + \mathcal{O}(t^3/P^2),$$

kde délka časového kroku je označena $\Delta t = t/P$ a P značí počet kroků za čas t .

Dostaneme tak aproximativní vyjádření propagátoru rozložené na součin exponenciál z jednotlivých částí původního Liouvillova operátoru, které jsou úměrné derivacím buďto pouze souřadnice, nebo pouze polohy. Vyhne se tak křížovým členům a působení propagátoru se značně zjednoduší, neboť exponenciála derivace souřadnice působící na funkci fázového prostoru způsobí pouhý posun souřadnice [55]

$$e^{c \frac{\partial}{\partial q}} f(q) = f(q + c).$$

Rozpisem působení integrátoru na počáteční souřadnice ve fázovém prostoru dostaneme konkrétní tvar integrátoru. Příkladem je rychlostní Verletův integrátor, který je často implementován v softwarových balících pro MD simulace biomolekul. Sestává ze tří dílčích podkroků při výpočtu nových souřadnic a rychlostí v rámci jednoho časového kroku:

$$\mathbf{v}\left(t + \frac{\Delta t}{2}\right) = \mathbf{v}(t) + \frac{1}{2}\mathbf{a}(t)\Delta t, \quad (2.6)$$

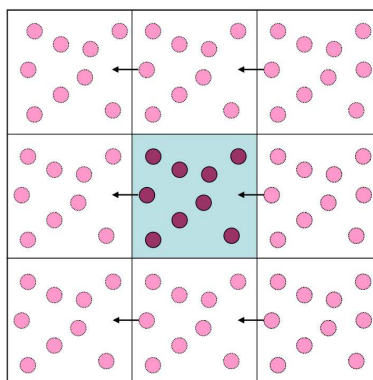
$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \mathbf{v}\left(t + \frac{\Delta t}{2}\right)\Delta t, \quad (2.7)$$

$$\mathbf{v}(t + \Delta t) = \mathbf{v}\left(t + \frac{\Delta t}{2}\right) + \frac{1}{2}\mathbf{a}(t + \Delta t)\Delta t, \quad (2.8)$$

kde výpočet sil a zrychlení v prvním a třetím dílčím podkroku závisí na zvoleném silovém poli. Kromě úplně prvního kroku MD simulace je nutné počítat síly pouze jednou za časový krok. Výpočet sil je velmi náročná operace, proto použití Verletova algoritmu umožňuje provádět MD simulace relativně rychle. Další jeho výhodou je enormní stabilita i při použití relativně velkého časového kroku. Verletův integrátor je implementován v softwarovém balíku NAMD, který byl použit pro MD simulace v rámci této práce.

2.5 Periodické okrajové podmínky

Abychom se v MD simulacích vyhnuli povrchovým efektům, zavádíme obvykle tzv. periodické okrajové podmínky [56]. Simulovaný systém sestávající z biomolekuly obklopené vodní obálkou tvoří základní box, který je díky periodickým okrajovým podmínkám nekonečně replikován ve všech směrech s periodou rovnou délkám hran boxu, jak je znázorněno na obrázku 2.4.



Obrázek 2.4: Schématické znázornění periodických okrajových podmínek [57].

Atom na hraně základního boxu, který by simulovaný systém v dalším časovém kroku opustil, je díky periodickým okrajovým podmínkám nahrazen kopií sebe sama přicházející do základního boxu z protější strany.

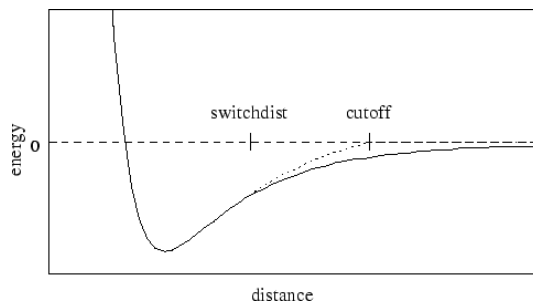
Periodické okrajové podmínky tak umožňují přibližně simulovat vlastnosti pseudonekonečného systému pomocí výpočtů dynamiky uvnitř konečného boxu.

Takto definovaný nekonečný systém je ale problematický z hlediska dalekohodových sil, které je v principu potřeba počítat mezi všemi páry atomů, tedy nejen uvnitř základního boxu. Prakticky se tento problém řeší různě pro van der Waalovy a elektrostatické interakce.

V případě van der Waalových interakcí můžeme využít toho, že jejich velikost klesá s šestou mocninou vzdálenosti. Je proto možné jednoduše zadefinovat tzv. vzdálenost oříznutí (angl. cut-off distance), od které můžeme tyto síly položit rovny nule [56].

Abychom se vyhnuli nespojitostem sil na cut-off vzdálenosti, tak se navíc definuje tzv. přepínací vzdálenost (angl. switch distance), od které van der Waalovy síly modifikujeme tak, aby spojitě klesly na nulovou hodnotu pro cut-off vzdálenost. Vliv zavedení switch vzdálenosti je zobrazen na obrázku 2.5. Běžné hodnoty switch a cut-off vzdálenosti se pohybují kolem 10, resp. 12 Å.

K urychlení výpočtů se navíc definuje vzdálenost pro generování seznamu párů atomů, které jsou od sebe vzdáleny méně než cut-off vzdálenost. Tento seznam není nutné měnit v každém výpočetním kroku, ale pouze s určitou frekvencí v závislosti na definovaných switch a cut-off vzdálenostech.



Obrázek 2.5: Modifikace Lennard-Jonesova potenciálu zavedením switch a cut-off vzdálenosti [56].

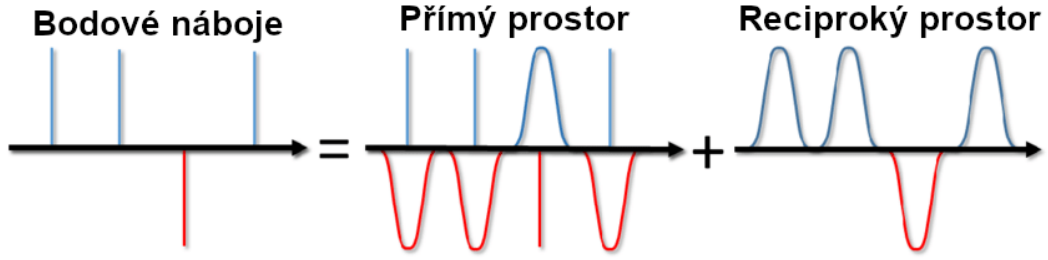
Elektrostatický potenciál klesá ve srovnání s van der Waalovými interakcemi podstatně pomaleji a nelze proto jednoduše definovat cut-off vzdálenost. Místo toho se obvykle aplikuje Ewaldova sumace [56], která využívá periodičnosti simulovaného systému, a umožňuje provést výpočet elektrostatických interakcí nejen pro základní box, ale i pro všechny jeho periodické obrazy. V softwarových balících používaných pro MD simulace biomolekul je obvykle implementována tzv. “částice-mřížka” verze Ewaldovy sumace (anglicky Particle-Mesh Ewald summation) [58].

Ewaldova sumace rozděljuje součet příspěvků od všech bodových nábojů na dvě sumy. Vychází se zde ze základního vztahu [59]

$$1 = \operatorname{erfc}(\beta r) + \operatorname{erf}(\beta r),$$

kde ke všem bodovým nábojům v systému přičteme a odečteme stínící Gaussovo rozložení nábojů, jak je znázorněno na obrázku 2.6. Parametr β zde představuje šířku použitých stínících hustot a r značí obecně vzdálenost mezi dvěma atomy. Funkce erf a erfc značí chybovou, resp. doplňkovou chybovou funkci.

Pokud má simulovaný systém nenulový celkový náboj, tak se navíc definuje uniformní hustota náboje, která celkový náboj vyruší [59]. Řešením Poissonovy



Obrázek 2.6: Stínění bodových nábojů při výpočtu elektrostatických interakcí Ewaldovou metodou [60].

rovnice pro takto definované hustoty nábojů, pak obdržíme odpovídající elektrostatické potenciály. Celkovou energii lze rozdělit na část, která rychle konverguje v přímém prostoru, a druhou část, která rychle konverguje v reciprokém prostoru. Celkově tak pro elektrostatickou interakci E_{Ewald} můžeme (bez prefaktoru $1/4\pi\epsilon$) psát [56]:

$$E_{\text{Ewald}} = E_{\text{dir}} + E_{\text{rec}} + E_{\text{self}} + E_{\text{surface}}, \quad (2.9)$$

$$E_{\text{dir}} = \frac{1}{2} \sum_{i,j=1}^N q_i q_j \sum_{\mathbf{n}_r} \frac{\text{erfc}(\beta|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}_r|)}{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}_r|} - \sum_{(i,j) \in \text{excluded}} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{v}_{ij}|}, \quad (2.10)$$

$$E_{\text{rec}} = \frac{1}{2\pi V} \sum_{\mathbf{m} \neq 0} \frac{\exp(-\pi^2|\mathbf{m}|^2/\beta^2)}{|\mathbf{m}|^2} \left| \sum_{i=1}^N q_i \exp(2\pi i \mathbf{m} \cdot \mathbf{r}_i) \right|^2, \quad (2.11)$$

$$E_{\text{self}} = -\frac{\beta}{\sqrt{\pi}} \sum_{i=1}^N q_i^2, \quad (2.12)$$

$$E_{\text{surface}} = \frac{2\pi}{(2\epsilon_s + 1)V} \left| \sum_{i=1}^N q_i \mathbf{r}_i \right|^2. \quad (2.13)$$

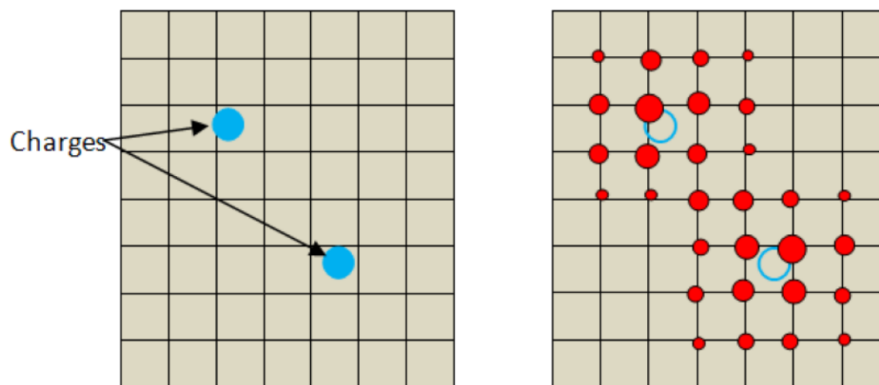
První člen E_{dir} (2.10) značí část Ewaldovy sumace vyhodnocenou v přímém prostoru. V silových polích je zvykem vynechávat působení nevazebných interakcí mezi nejbližšími atomy. Jejich působení je zahrnuto v jiných členech silového pole. Musíme zde proto odečíst příspěvek párů atomů, jejichž nevazebné interakce jsou v použitém silovém poli vynechány. Náboje jsou označeny q_i , polohy nábojů \mathbf{r}_i , mřížkový vektor \mathbf{n}_r a \mathbf{v}_{ij} představuje mřížkový vektor, který minimalizuje $|\mathbf{r}_i - \mathbf{r}_j + \mathbf{v}_{ij}|$ pro daný pár i a j .

Druhý člen E_{rec} (2.11) je roven části Ewaldovy sumace vyhodnocené v reciprokém prostoru. V je rovno objemu simulačního boxu a \mathbf{m} značí mřížkový vektor v reciprokém prostoru.

Třetí člen E_{self} (2.12) je roven triviální konstantě a čtvrtý povrchový člen E_{surface} (2.13) lze většinou zanedbat díky dielektrické konstantě vody $\epsilon_s \approx 80$ [56].

Pro vhodně definovaný parametr β určující šířku rozdělení stínících hustot můžeme pro přímou část Ewaldovy sumace E_{dir} definovat cut-off vzdálenost, od které bude příspěvek tohoto členu zanedbán. Výpočetní náročnost se tak zmenší z $O(N^2)$ na $O(N)$. Přímou část Ewaldovy sumace tak lze provést přímočaře.

Naopak pro reciprokou část Ewaldovy sumace nelze definovat cut-off vzdálenost a její výpočetní náročnost je stále $O(N^2)$. Proto se v prostoru zavádí periodická mřížka (angl. mesh). Bodové náboje, které se mohou vyskytovat kdekoli v prostoru, jsou interpolovány na body mřížky s vahami, které odpovídají vzdálenosti náboje od daného mřížového bodu. Schématicky tuto situaci vystihuje obrázek 2.7 [61].



Obrázek 2.7: Schématické znázornění interpolace nábojů na mříž metodu PME. Interpolace nábojů na pravidelnou mříž umožňuje využití algoritmu rychlé Fourierovy transformace a značně tak přispívá k urychlení výpočtů [61].

Při využití algoritmu rychlé Fourierovy transformace (angl. Fast Fourier Transform) se výpočetní náročnost snižuje na $O(N \log N)$. V reciprokém prostoru je spočten příspěvek k energii a následnou zpětnou Fourierovou transformací, opět využívající FFT algoritmu, se dopočítají síly v přímém prostoru.

2.6 Omezení vazeb

Délku časového kroku je třeba v MD simulacích volit podle nejrychlejšího pohybu ve studovaném systému. V případě bioorganických molekul jsou většinou nejrychlejší oscilace vodíků v kovalentních vazbách. Pokud bychom chtěli popsat jejich pohyb s dostatečnou přesností, museli bychom volit velmi malý časový krok, cca. 0,5 fs.

Pokud ale délku těchto vazeb zafixujeme pomocí vhodných holonomních omezení, můžeme volit delší časový krok, obvykle 2,0 fs, a výpočet se tak značně urychlí. Jedním ze způsobů zavedení holonomních omezení je algoritmus SHAKE [62].

Zavedením k holonomních vazeb $\sigma_k(\mathbf{r}_i) = 0$ zároveň přidáme dodatečné síly \mathbf{G}_{ik} do Newtonových pohybových rovnic

$$m\mathbf{a}_i = \mathbf{F}_i + \sum_k \mathbf{G}_{ik}, \quad (2.14)$$

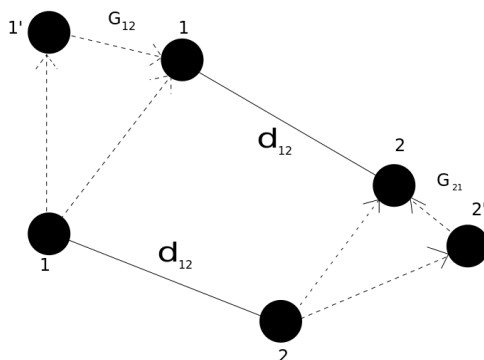
$$\mathbf{G}_{ik} = -\lambda_k \frac{\partial \sigma_k}{\partial \mathbf{r}_i}, \quad (2.15)$$

kde m značí hmotnost atomu i , celková síla je rozdělena na část plynoucí z potenciální energie silového pole \mathbf{F}_i a část z omezení vazeb \mathbf{G}_{ik} , λ_k značí Lagrangeův multiplikátor.

Přesné řešení pro Lagrangeovy multiplikátory může být v závislosti na zvolených holonomních vazbách σ_k výpočetně náročné, nebo dokonce i analyticky neřešitelné a musíme proto použít aproximativní metody.

V případě algoritmu SHAKE se postupuje iterativně [62]. Pozice atomů jsou nejdříve spočteny bez omezení. Poté se pro každou ze zavedených holonomních vazeb σ_k přibližně vyjádří odpovídající Lagrangeův multiplikátor λ_k a posunou se původně neomezené pozice atomů. Omezení se zde aplikují po sobě a zavádí se zjednodušující předpoklad, že jsou nezávislé na předchozích. Celá procedura se opakuje, dokud nejsou zadané vazby σ_k splněny se zvolenou tolerancí. SHAKE algoritmus je názorně ilustrován na obrázku 2.8.

SHAKE algoritmus byl původně odvozen pro obyčejný Verletův integrátor. Pro rychlostní Verletův integrátor existuje jeho rychlostní verze nazývaná RATTLE algoritmus [63]. Pro molekuly vody modelované pomocí modelu TIP3P existuje analytické řešení pro splnění holonomních vazeb a nazývá se SETTLE algoritmus [64].



Obrázek 2.8: Ilustrace SHAKE algoritmu. Atomy 1 a 2 si po výpočtu nových poloh obecně nezachovají svoji meziatomovou vzdálenost d_{12} . Aplikací omezení pomocí SHAKE algoritmu můžeme určit sílu G_{12} potřebnou pro zachování konstantní délky vazby [65].

2.7 Statistické soubory

Pohybové rovnice řešené pomocí integrátorů odvozených z Liouvillova operátoru 2.5 popisují izolovaný systém, ve kterém se zachovává energie, objem i počet částic. Ve výsledku tak popisujeme systém z mikrokanonického souboru.

V termodynamické limitě platí, že by si různé statistické soubory měly být ekvivalentní, ale v MD simulacích máme pouze konečný počet molekul. Může jich být tak malé množství, že se jejich chování v různých statistických souborech může značně lišit. Je proto vhodné simulovat systémy v souborech, které nejlépe odpovídají experimentálním podmínkám. V případě bioorganických molekul se

nejčastěji volí izotermicko-izobarický soubor (NpT soubor), ve kterém se udržuje konstantní střední hodnota teploty T a tlaku p .

Rovnice výše uvedeného Verletova integrátoru 2.6, 2.7 a 2.8 je tak potřeba modifikovat zavedením vhodných termostatů a barostatů.

Jako termostaty se označují algoritmy, které zaručí konstantní hodnotu střední hodnoty teploty během MD simulace. Možných přístupů je mnoho a podle konkrétního systému, který chceme studovat, je potřeba zvolit vhodný termostat. Pro některé dříve hojně používané termostaty, jako byl např. Berendsenův termostat, bylo ukázáno, že sice udržují konstantní teplotu, ale negenerují správný kanonický soubor [66], neboť některé stupně volnosti mají v konečném výsledku nižší teplotu než ostatní.

Existují tzv. deterministické termostaty, které explicitně popisují stupně volnosti lázně, která je vázána na náš systém. Příkladem takového termostatu je Nosé-Hooverův termostat [67].

Kanonický soubor lze generovat i pomocí Andersenova termostatu [68], který jednoduše periodicky změní všechny rychlosti atomů v simulaci na nové rychlosti generované podle Maxwell-Boltzmannova rozdělení. Nevýhodou tohoto přístupu je, že trajektorie jsou značně narušeny a tento termostat tak nelze použít na studium časově korelovaných jevů, nýbrž pouze k získání statistických rovnovážných hodnot.

V této práci byl v MD simulacích použit termostat založený na Langevinovské dynamice [56]. Pohybové rovnice jsou v tomto případě modifikovány dodatečnou stochastickou silou, která reprezentuje interakci s lázní a zajišťuje generování správného statistického souboru. Kromě stochastického členu je nutné rovnice doplnit o člen tření, aby byl splněn fluktučně-disipační teorém. Celkově tak pohybové rovnice nabývají podoby:

$$\dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m_i}, \quad (2.16)$$

$$\dot{\mathbf{p}}_i = \mathbf{F}_i - \gamma \mathbf{p}_i + \sqrt{2k_B T \gamma m_i} \mathbf{R}_i(t), \quad (2.17)$$

kde \mathbf{r}_i a \mathbf{p}_i označují kanonicky sdruženou souřadnici a hybnost, m_i hmotnost i -tého atomu. Síla \mathbf{F}_i značí část celkové síly odvozené ze silového pole bez zahrnutí termostatu, γ je koeficient tření, k_B Boltzmannova konstanta, T teplota a $\mathbf{R}_i(t)$ značí bílý Gaussovský šum s nulovou střední hodnotou.

Po zavedení stochastické síly se pozice a rychlosti stanou náhodnými veličinami a jejich odpovídající pravděpodobnostní rozdělení se v čase vyvíjí podle Smoluchowského rovnice [56]. Rovnice 2.16 a 2.17 jsou navrženy tak, aby časově nezávislé řešení Smoluchowského rovnice odpovídalo rozdělení požadovaného statistického souboru.

Jako barostaty se označují algoritmy, které nám zaručí konstantní střední hodnotu tlaku během MD simulace. Opět existuje více možností a je potřeba vybrat vhodný barostat pro námi studovaný systém.

V MD simulacích realizovaných v rámci této práce byl využit modifikovaný Nosé-Hooverův barostat, jehož fluktuace byly popsány Langevinovou dynamikou [69]. Objem simulačního boxu V je zde dynamická veličina, která představuje kanonickou souřadnici barostatu a její odpovídající kanonickou hybnost označíme

p_ε . Definujeme zde fiktivní hodnotu hmotnosti pístu W jako další parametr simulace a hybnost barostatu p_ε lze pak určit jako součin hmotnosti pístu W a rychlosti deformace boxu $\dot{\varepsilon}$. Celkově pak pohybové rovnice společně s termostatem a barostatem nabývají následující podoby:

$$\dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m_i} + \frac{p_\varepsilon}{W} \mathbf{r}_i, \quad (2.18)$$

$$\dot{\mathbf{p}}_i = \mathbf{F}_i - \gamma \mathbf{p}_i + \sqrt{2k_B T \gamma m_i} \mathbf{R}_i(t) - \frac{p_\varepsilon}{W} \mathbf{p}_i, \quad (2.19)$$

$$\dot{V} = \frac{3p_\varepsilon}{W} V, \quad (2.20)$$

$$\dot{p}_\varepsilon = 3V(X - P_{\text{ext}}) + \frac{3}{N_f} \sum_{i=1}^N \frac{\mathbf{p}_i^2}{m_i} - \gamma_p p_\varepsilon + \sqrt{2k_B W \gamma_p} R_p, \quad (2.21)$$

kde R_p je bílý Gaussovský šum představující stochastickou sílu působící na píst barostatu, γ_p je třecí koeficient barostatu, P_{ext} cílový tlak a skalární veličina X představuje hodnotu tlaku počítanou podle viriálového teorému bez příspěvku šumu působícího na píst. Lze ji určit podle vztahu:

$$X = \frac{1}{3V} \left[\sum_{i=1}^N \frac{\mathbf{p}_i \cdot \mathbf{p}_i}{m_i} + \sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{F}_i \right] - \frac{\partial}{\partial V} \Phi(\mathbf{r}^N, V), \quad (2.22)$$

kde $\Phi(\mathbf{r}^N, V)$ značí potenciál určený podle silového pole.

3. Výpočty volné energie

Podle druhého zákona termodynamiky může entropie pro izolovaný systém pouze růst nebo zůstat konstantní. Spontánně tak probíhají pouze procesy, které zvětšují entropii systému. Rovnovážný stav je stav s maximální entropií. Většina systémů, které pozorujeme, nebo s nimiž pracujeme v rámci MD simulací, ale izolovaná není. Pro popis pomocí principu maximální entropie bychom museli sledovat entropii jak systému tak i okolního rezervoáru.

Zavedením termodynamických potenciálů ovšem můžeme tento popis zjednodušit. V každém termodynamickém souboru totiž platí princip minimalizace odpovídajícího termodynamického potenciálu vzhledem ke všem vnitřním proměnným [70]. Spontánně tak probíhají pouze procesy, pro které termodynamický potenciál (volná energie) klesá.

Výpočet změny volné energie při navázání inhibitoru do aktivního místa enzymu pak je významným nástrojem při racionálním designu léků.

3.1 Gibbsova a Helmholtzova volná energie

V MD simulacích biomolekul se nejčastěji používá izotermicko-izobarický soubor popisující systémy, které si mohou vyměňovat teplo se svým okolím a jejichž objem může fluktuovat. Střední hodnoty teploty a tlaku jsou zde konstantní, proto se tento soubor obvykle označujeme jako NpT. Pracujeme zde s Gibbsovu volnou energií G , pro níž platí výše zmíněný princip minimalizace:

$$\left(\frac{\partial G}{\partial X}\right)_{T,P} = 0, \quad (3.1)$$

$$\left(\frac{\partial^2 G}{\partial X^2}\right)_{T,P} > 0, \quad (3.2)$$

kde X značí libovolný vnitřní parametr systému a rovnice 3.1 platí pouze za předpokladu rovnovážného stavu.

Gibbsovu volnou energii je možné vyjádřit pomocí partiční funkce $\Delta(N, P, T)$:

$$G(N, P, T) = -\frac{1}{\beta} \ln \Delta(N, P, T). \quad (3.3)$$

Z Gibbsovy volné energie můžeme odvodit další termodynamické veličiny, pokud vyjádříme její totální diferenciál

$$dG = \mu dN + V dP - S dT, \quad (3.4)$$

$$\mu = \left(\frac{\partial G}{\partial N}\right)_{P,T}, \quad (3.5)$$

$$V = \left(\frac{\partial G}{\partial P}\right)_{N,T}, \quad (3.6)$$

$$S = -\left(\frac{\partial G}{\partial T}\right)_{N,P}, \quad (3.7)$$

kde μ značí chemický potenciál a S entropii systému.

V případě kanonického souboru (NVT), který se v MD simulacích biomolekul také někdy používá, termodynamický potenciál odpovídá Helmholtzově volné energii A . Význam Gibbsovy a Helmholtzovy volné energie je analogický a v literatuře se často zaměňují.

Od kanonického souboru popsaného, Helmholtzovou volnou energií $A(N,V,T)$, můžeme přejít k izotermicko-izobarickému souboru, popsanému Gibbsovou volnou energií $G(N,P,T)$, pomocí Legendreovy transformace [55], pokud vyjádříme objem jako funkci tlaku a využijeme jednu z Maxwellových relací $P = -\partial A/\partial V$

$$G(N,P,T) = A(N,V(P),T) + PV(P). \quad (3.8)$$

Partiční funkci izotermicko-izobarického souboru $\Delta(N,P,T)$ lze odvodit na základě úvah, kdy uvažujeme kanonický systém, který pomyslně rozdělíme na dva podsystémy a jeden z nich uvažujeme jako barostat [55]. Ve výsledku můžeme psát

$$\Delta(N,P,T) = \frac{1}{V_0 N! h^{3N}} \int_0^\infty dV \int d\mathbf{x} e^{-\beta(H(\mathbf{x})+PV)}, \quad (3.9)$$

kde $H(\mathbf{x})$ značí hamiltonián systému, faktor $\beta = 1/k_B T$, h Planckova konstanta a V_0 je objemový člen zahrnutý pro správnou normalizaci.

Ze znalosti partiční funkce můžeme získat volnou energii a odvodit všechny termodynamické vlastnosti studovaného systému. Přesné určení partiční funkce je však často obtížné a pro řadu jevů nám stačí jednodušší popis pomocí změny volné energie mezi dvěma rovnovážnými stavy.

3.2 Vazebná volná energie

V případě inhibitorů enzymů se konkrétně jedná o rozdíl hodnot volné energie mezi stavy s volným a navázaným ligandem. Porovnání vazebných volných energií pro sérii inhibitorů pak umožňuje vybrat z nich ten nejúčinnější. To se hojně využívá při racionálním designu léků prostřednictvím dockingu či MD simulací. Důležité přitom je, porovnávat vypočtené hodnoty vazebné volné energie s dostupnými experimentálními daty, aby bylo jisté, že zvolený výpočetní protokol je adekvátní.

Označíme-li studovaný enzym E a potenciální inhibitor I , tak jejich vazebná reakce lze zapsat:



kde IE značí produkt, tedy komplex navázaného inhibitoru a enzymu.

V rovnováze pro tento systém platí, že poměr koncentrací reaktantů a produktů je konstantní v čase. Můžeme proto definovat rovnovážnou vazebnou konstantu K_b (nebo její převrácenou hodnotu, označovanou jako disociační konstanta K_d), následujícím způsobem [71]

$$K_b = \frac{1}{K_d} = \frac{[IE]}{[I][E]}. \quad (3.11)$$

Rovnovážná vazebná konstanta souvisí s rozdílem Gibbsovy volné energie mezi stavy s volným a navázaným ligandem prostřednictvím vztahu:

$$\Delta G_0 = -k_B T \ln(C_0 K_b), \quad (3.12)$$

kde ΔG_0 značí standardní vazebnou volnou energii a $C_0 = 1M$ odpovídá standardní koncentraci. Ze znalosti Gibbsovy vazebné volné energie tak můžeme určit vazebnou rovnovážnou konstantu a naopak. Zároveň je zřejmé, že čím více záporná bude vazebná volná energie, tím větší bude rovnovážná vazebná konstanta a vytvořený komplex tak bude termodynamicky více stabilní.

3.3 Termodynamický cyklus

V principu je možné simulovat navázání ligandu k proteinu přímo. Na počátku umístíme ligandy volně do okolní vodní obálky a čekáme (obvykle velmi dlouho), až dojde k navázání. Vazebnou volnou energii je pak v principu možné určit z podílu počtu simulačních snímků, kdy je ligand navázan a kdy je volný. Problémem tohoto postupu je, že proces navázání inhibitoru do aktivního místa enzymu trvá velmi dlouho a pokud se tak náhodou stane, pak inhibitor již aktivní místo obvykle spontánně neopustí. K výpočtu vazebné volné energie by přitom bylo potřeba získat trajektorie, kde dojde k velkému počtu realizací asociace a disociace mezi inhibitorem a enzymem. To s běžnými hardwarovými prostředky není v současné době možné. V praxi tak byl tento postup úspěšně aplikován pouze za využití specializovaného superpočítače Anton, jehož procesory jsou navrženy přímo pro algoritmus molekulární dynamiky [72]. Kromě toho se jednalo o chabé milimolární inhibitory s velmi malými vazebnými volnými energiemi $\sim 2-3$ kcal/mol. Pro potenciální léky účinkující v nano- či pikomolárních koncentracích a s vazebnými volnými energiemi většími než 10 kcal/mol je tento přístup v současné době nepoužitelný. Ve většině případů tak musíme použít specializované metody, které výpočet vazebné volné energie urychlí.

Jedna z možností je zavést dodatečné síly či potenciály, jež ligand táhnou z či naopak tlačí do vazebného místa. Patří sem metody SMD (angl. Steered Molecular Dynamics), ABF (angl. Adaptive Biasing Force), metadynamika, umbrella sampling. Hovoříme přitom o urychlení tzv. geometrické transformace - na obrázku 3.1 podél horní horizontální šipky.

Druhý možný přístup [73], který byl využit v rámci této práce, představují tzv. alchymistické transformace - na obrázku 3.1 podél vertikálních šipek. V tomto případě se využívá skutečnosti, že volná energie je stavová veličina. Její rozdíl mezi dvěma stavy tedy nezávisí na cestě, po které jsme přešli z jednoho stavu do druhého. Zajímá nás zde rozdíl volné energie mezi stavem, kdy je ligand navázaný na protein, a stavem, kdy je ligand volný v rozpouštědle. Namísto přímého výpočtu podél horní horizontální šipky na obrázku 3.1 půjdeme nyní oklikou - tj. druhou možnou cestou podél termodynamického cyklu. To bude obnášet výpočet rozdílu volné energie při anihilaci (odpoutání) ligandu z vodní obálky, od něhož bude třeba odečíst hodnotu rozdílu volné energie při anihilaci ligandu z vazebného místa proteinu. V praxi se to realizuje tak, že se postupně škálují nevazebné interakce ligandu takovým způsobem, aby se ligand postupně odpoutal od svého

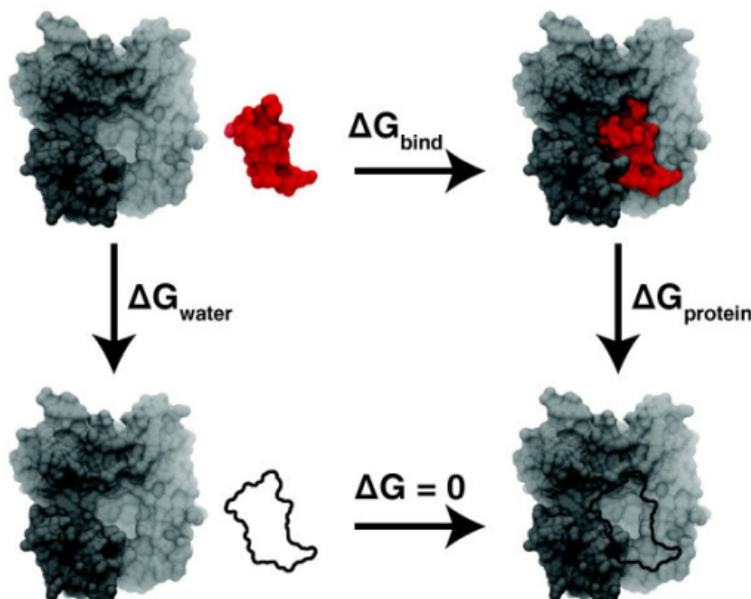
okolí (vodní obálky či vazebného místa v proteinu) a ze simulovaného systému tak efektivně zmizel.

Provádí se tedy dvě nezávislé MD simulace. V první MD simulaci uvažujeme ligand volně rozpuštěný ve vodě, protein v tomto případě vůbec není potřeba v simulaci uvažovat, díky čemuž je možné použít menší simulační box. Ligand postupně mizí ze simulovaného systému a získáme zde hodnotu ΔG_{water} . Ve druhé MD simulaci začínáme s ligandem navázaným na protein. Ligand pak v průběhu alchymistického výpočtu ze simulovaného systému mizí a určíme hodnotu $\Delta G_{\text{protein}}$. Konečné stavy ligandu v obou simulacích jsou shodné - ligand ve vakuu. Rozdíl volné energie mezi těmito koncovými stavy (tj. podél spodní horizontální šipky na obrázku 3.1) je tak nulový. Vazebnou volnou energii díky tomu určíme jako rozdíl hodnot ΔG_{water} a $\Delta G_{\text{protein}}$. Využíváme přitom faktu, že volná energie je stavová veličina, proto suma rozdílů volné energie podél celého termodynamického cyklu musí být nulová. To znamená, že musí platit:

$$0 = \Delta G_{\text{water}} - \Delta G_{\text{protein}} - \Delta G_{\text{bind}} \quad (3.13)$$

a díky tomu můžeme určit vazebnou volnou energii z rozdílu hodnot získaných v průběhu obou dílčích alchymistických transformací:

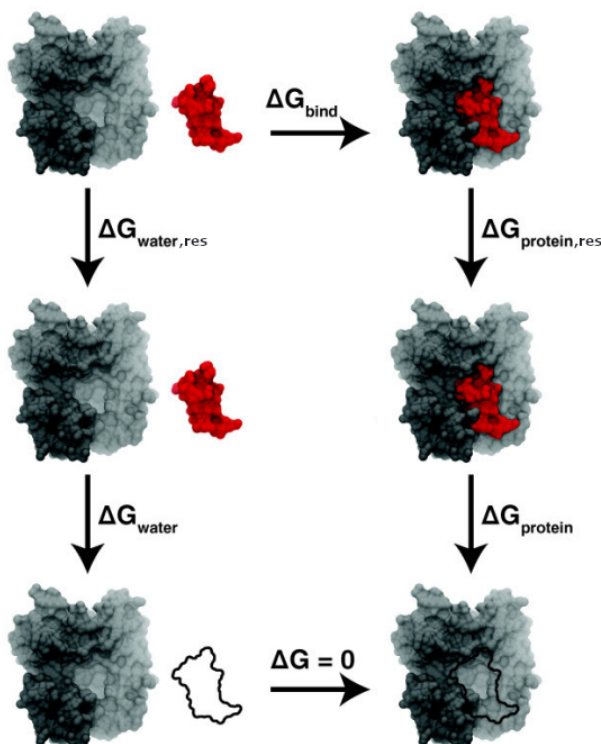
$$\Delta G_{\text{bind}} = \Delta G_{\text{water}} - \Delta G_{\text{protein}}. \quad (3.14)$$



Obrázek 3.1: Termodynamický cyklus pro výpočet vazebné volné energie. V případě geometrické transformace se postupuje nejkratší možnou cestou přímo podél horní horizontální šipky. V případě alchymistických transformací se využije delší cesta, kdy se provádí dvě dílčí simulace/transformace, během nichž jsou interakce ligandu s okolím škálovány postupně až na nulu. Vazebná volná energie je pak určena jako rozdíl hodnot vypočtených během těchto dvou dílčích simulací [47].

V praxi se ukazuje, že výpočty vazebné volné energie často konvergují velmi pomalu a je proto potřeba simulovat neúnosně dlouhé trajektorie. Jedním z možných řešení je zavést do simulovaných systémů dodatečné omezující potenciály,

kteřé budou ligand držet v určité pozici, orientaci, či konformaci [74, 75]. Ve výsledku tím omezíme konformační prostor, který musíme během MD simulace navzorkovat, a stačí tak simulovat kratší trajektorie. Nevýhodou tohoto přístupu je, že musíme zároveň určit příspěvek omezujících potenciálů k vazebné volné energii. Ten lze v některých jednoduchých případech určit analyticky. Pokud ovšem používáme komplikované omezující potenciály, tak je potřeba modifikovat použitý termodynamický cyklus a provést další nezávislé MD simulace, ve kterých příspěvek omezujících potenciálů určíme, jak je znázorněno na obrázku 3.2.



Obrázek 3.2: Upravený termodynamický cyklus využitý pro výpočet vazebné volné energie v případě zavedení dodatečných omezujících potenciálů (angl. restraints). Oproti výpočtu bez dodatečných potenciálů je potřeba provést další dvě simulace navíc, ve kterých se určí příspěvky omezujících potenciálů. [47].

Provádí se zde tedy navíc MD simulace, ve které je ligand navázaný na protein a postupně zapínáme omezující potenciály. V této MD simulaci určíme $\Delta G_{\text{protein, res}}$. Podobnou simulaci je potřeba provést i pro ligand rozpuštěný ve vodě, kdy určíme $\Delta G_{\text{water, res}}$. Zbylé dvě simulace, alchymistické transformace, probíhají podobně jako v předchozím případě, pouze zde jsou v jejich průběhu přítomny ony omezující potenciály. Výslednou vazebnou volnou energii pak určíme z výrazu:

$$\Delta G_{\text{bind}} = \Delta G_{\text{water}} + \Delta G_{\text{water, res}} - \Delta G_{\text{protein}} - \Delta G_{\text{protein, res}} \quad (3.15)$$

Značnou výhodou tohoto přístupu je to, že dílčí MD simulace je často možné provést obousměrně, protože ligand odpoutaný od proteinu zůstává díky omezujícím potenciálům ve vazebném místě. Během alchymistických transformací je tak možné simulovat nejen anihilaci ale i kreaci ligandu. Omezující potenciály ve zbývajících MD simulacích je možné zapínat i vypínat. Případný rozdíl mezi

hodnotami volné energie určenými v jednom a druhém směru přitom indikuje, že výsledky nejsou dostatečně zkonvergované a že je žádoucí provést delší MD simulace. Kreační běhy při výpočtu solvatačních volných energií obvykle konvergují podstatně rychleji než běhy anihilační. Na druhou stranu vazebná místa proteinů nemusí být vždy rigidní po odpoutání ligandu mohou v průběhu anihilačního běhu značně změnit své prostorové uspořádání, což může konvergenci navazujících kreačních běhů značně zpomalit až znemožnit. To je i případ koronavirové Mpro s mimořádně flexibilním aktivním místem, které je uzpůsobené interakci s 10 či 11 různými přirozenými substráty, které musí spolehlivě rozpoznávat. Proto byla v této práci dominantně využita metoda prosté dvojité anihilace bez dodatečných omezujících potenciálů.

3.4 Alchymistické transformace

Při alchymistických transformacích se obvykle definuje bezrozměrný parametr λ , který nabývá hodnot od 0 do 1 a popisuje přechod z počátečního referenčního do cílového stavu. Během alchymistických transformací může docházet k anihilaci, kreači ale i mutaci atomů či jejich skupin. Atomy v simulovaném systému je tak možné rozdělit do tří skupin [76]: atomy beze změny, atomy patřící referenčnímu počátečnímu stavu a a atomy cílového stavu b . Hybridní hamiltonián během tohoto procesu můžeme zapsat následovně:

$$\mathcal{H}(\mathbf{x}, \mathbf{p}_x; \lambda) = \mathcal{H}_0(\mathbf{x}, \mathbf{p}_x) + \lambda \mathcal{H}_b(\mathbf{x}, \mathbf{p}_x) + (1 - \lambda) \mathcal{H}_a(\mathbf{x}, \mathbf{p}_x), \quad (3.16)$$

kde \mathcal{H}_0 popisuje atomy beze změny, \mathcal{H}_b popisuje atomy v cílovém stavu a \mathcal{H}_a atomy referenčního stavu.

Interakce atomů referenčního stavu jsou postupně škálovány až na nulovou hodnotu a tyto atomy tak efektivně mizí ze simulovaného systému. Naopak atomy cílového stavu b se postupně v simulovaném systému objevují.

Jelikož je volná energie stavová veličina, tak hodnota jejího rozdílu mezi dvěma stavy nezávisí na procesu, kterým jsme se dostali ze stavu počátečního do stavu koncového. Můžeme tak volit i nefyzikální (alchymistické) mezistavy jakož i přesný průběh škálování interakcí v podstatě libovolně. Z praktického hlediska je ovšem dobré volit takový způsob škálování, při kterém se při malé změně parametru λ změny interakce pouze málo. Velké změny v energii pro malé změny λ lze při alchymistických simulacích pozorovat především v koncových bodech, kdy anihilované atomy ze simulovaného systému zcela mizí, nebo se v něm naopak objeví z ničeho. Mluví se zde o tzv. katastrofě koncového bodu [77]. Dochází zde k numerickým nestabilitám a je proto třeba modifikovat způsob, jakým interakce škálujeme. Možným řešením je zavést tzv. soft-core potenciál, který modifikuje nevazebné van der Waalovy a elektrostatické interakce, aby nedocházelo k příliš velkým změnám naráz.

Místo jednoho parametru λ se zde zavádí dva nezávislé parametry λ_{LJ} a λ_{elec} a škálování provádíme následovně:

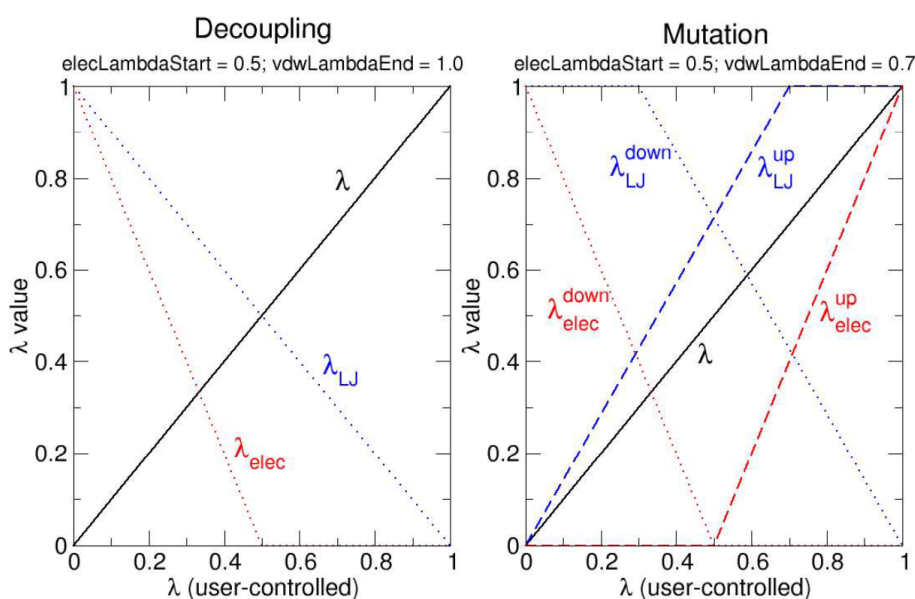
$$V_{NB}(r_{ij}) = \lambda_{LJ} \varepsilon_{ij} \left[\left(\frac{R_{ij}^{\min 2}}{r_{ij}^2 + \delta(1 - \lambda_{LJ})} \right)^6 - \left(\frac{R_{ij}^{\min 2}}{r_{ij} + \delta(1 - \lambda_{LJ})} \right)^3 \right] + \lambda_{elec} \frac{q_i q_j}{\varepsilon_1 r_{ij}},$$

kde parametr δ v soft-core potenciálu zajišťuje, že van der Waalsův potenciál nabývá všude konečných hodnot i pro malé hodnoty λ_{LJ} .

Obrázek 3.3 názorně ukazuje, jakých hodnot konkrétně nabývají parametry λ_{LJ} a λ_{elec} v typické alchymistické MD simulaci prováděné prostřednictvím softwarového balíku NAMD [56].

Pro atomy, které ze simulace mizí, vypínáme elektrostatické interakce rychleji. Pro objevující se atomy je naopak zapínáme až s určitou prodlevou. Předtím zapínáme van der Waalsovy interakce, atomy tak mají dostatek času na to, aby se přizpůsobily, nedocházelo k překryvům a v okamžiku zapnutí elektrostatické interakce objevujících se atomů již byly všechny atomy od sebe dostatečně vzdálené.

Vazebné interakce přitom zůstávají během alchymistické transformace beze změny.



Obrázek 3.3: Průběh škálování elektrostatické λ_{elec} a van der Waalsovy interakce λ_{LJ} během alchymistické transformace v programu NAMD. Pro vyhnutí se katastrofy koncového bodu je škálování obou interakcí navzájem posunuto. První graf znázorňuje situaci pro atomy, které ze simulace mizí. Druhý graf znázorňuje případ, kdy některé atomy ze simulovaného systému mizí a jiné se v něm naopak objevují [56].

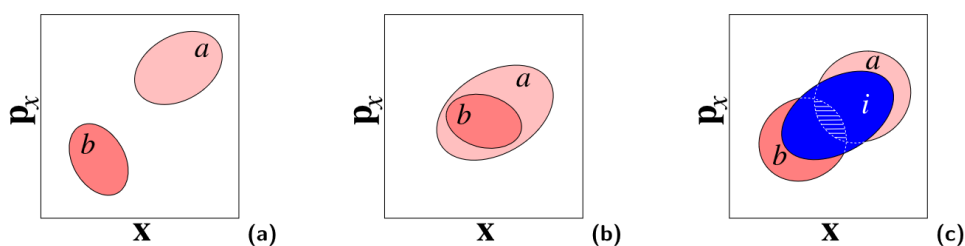
3.5 Perturbace volné energie

Jeden z možných přístupů k výpočtu volné energie během alchymistické transformace je perturbace volné energie (angl. Free Energy Perturbation) [78]. Uvažujeme-li kanonický soubor a chceme určit změnu volné energie mezi dvěma stavy a a b popsány dvěma různými potenciály U_a a U_b , tak je možné tuto změnu vyjádřit pomocí střední hodnoty na fázovém prostoru [79]:

$$\Delta A_{a \rightarrow b} = -kT \ln \left\langle e^{-\beta(U_b - U_a)} \right\rangle_a, \quad (3.17)$$

kde index a za střední hodnotou značí, že středujeme přes konfigurační prostor stavu a .

Tuto rovnici je možno interpretovat tak, že z molekulárních simulací získáme konfigurace stavu a , které pak beze změny použijeme ke vzorkování konfiguračního prostoru stavu b . Problém s tímto přístupem nastane, pokud konfigurační prostory stavů a a b nemají dostatečný překryv, jak je znázorněno na obrázku 3.4. V tomto případě konfigurace navzorkované ve stavu a nejsou konfigurace s vysokou pravděpodobností pro stav b . Rozdíl potenciálů v obou stavech je zde značný a exponenciála ve vzorci 3.17 nabývá malých hodnot. V konečném výsledku tak výpočet volné energie konverguje velmi pomalu.



Obrázek 3.4: Znáznornění fázového prostoru v jedné dimenzi při výpočtu volné energie pomocí metody perturbace volné energie. (a) Počáteční stav a nemá dostatečný překryv s koncovým stavem b , výpočet vůbec nekonverguje. (b) Dochází k překryvu, výpočet konverguje rychle. (c) Je možné zavést mezistavy i , které zaručí postupný přechod mezi stavy a a b [79].

Abychom mohli efektivně využít výpočet volné energie podle vzorce 3.17, tak stav b musí být malou poruchou pro stav a . Pokud tomu tak není, tak musíme zadefinovat mezistavy i , které zajistí postupný přechod mezi stavy a a b . Mezistavy i definujeme pomocí dříve zavedeného parametru λ , který škáluje interakce ligandu s okolím a nabývá hodnot mezi 0 a 1. Interval hodnot parametru λ se rozdělí do konečného počtu podintervalů a pro každý z nich se provede dílčí výpočet podle vzorce 3.17. Celkový rozdíl volné energie je pak roven součtu dílčích výsledků z jednotlivých podintervalů:

$$\Delta A_{a \rightarrow b} = -kT \sum_{i=1}^N \ln \left\langle e^{-\beta(U(\lambda_{i+1}) - U(\lambda_i))} \right\rangle_i. \quad (3.18)$$

3.6 Termodynamická integrace

Alternativní způsob výpočtu volné energie prostřednictvím alchymistické transformace představuje termodynamická integrace [80]. Stav a je kontinuálně převáděn do stavu b . Celý proces by měl být prováděn adiabaticky, simulovaný systém by tedy měl být v termodynamické rovnováze. Rozdíl volné energie mezi stavy a a b může být i v tomto případě vyjádřen pomocí střední hodnoty v kanonickém souboru [81]:

$$\Delta A_{a \rightarrow b} = \int_0^1 \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_{\lambda} d\lambda, \quad (3.19)$$

kde index λ za střední hodnotou opět naznačuje, že střední hodnotu počítáme pro konkrétní kanonický soubor s danou hodnotou λ .

Praktická implementace výpočtu volné energie dle vzorce 3.19 je analogická předchozí poruchové metodě. Interval hodnot λ rozdělíme na konečně mnoho podintervalů a v každém z nich určíme střední hodnotu derivace potenciálu podle λ . Volnou energii pak získáme podle vzorce 3.19 numerickou integrací.

3.7 Výpočet volné energie prostřednictvím rovnovážných MD simulací

Přesný způsob výpočtu volné energie se může lišit nejen podle použité výpočetní metody. Obecně totiž dále rozlišujeme výpočty provedené prostřednictvím rovnovážných a nerovnovážných MD simulací. V případě rovnovážných MD simulací nám k analýze výsledků v principu stačí jedna jediná velmi dlouhá trajektorie. Střední hodnoty termodynamických veličin jsou pak určeny, za předpokladu ergodického teorému, jako střední hodnoty v čase z této trajektorie.

3.8 Výpočet volné energie prostřednictvím nerovnovážných MD simulací

V případě nerovnovážných (většinou velmi krátkých) MD simulací je potřeba získat velké (stovky či tisíce) množství nezávislých trajektorií. Jejich analýzou získáme pravděpodobnostní rozdělení studované veličiny, např. práce, při daném procesu. Aplikací některého z teorémů platných pro nerovnovážné děje získáme opět rovnovážnou hodnotu volné energie.

Práce není stavová veličina, neboť její hodnota závisí na prošlé trajektorii, ale stále se jedná o termodynamickou veličinu a je jí možné zadefinovat pomocí střední hodnoty souboru nějaké funkce na fázovém prostoru. Konkrétně můžeme tvrdit, že vykonaná práce $W_{a \rightarrow b}$ musí odpovídat střední hodnotě funkce $W_{a \rightarrow b}(\mathbf{x})$, která vyjadřuje práci vykonanou v jednom členu souboru při přechodu ze stavu a do stavu b . Je zde důležité zdůraznit, že se nejedná o prosté rovnovážné střední hodnoty, neboť funkce $W_{a \rightarrow b}(\mathbf{x})$ je definována podél určité trajektorie [55].

Připravíme-li soubor mikrostavů \mathbf{x}_0 odpovídající stavu a a z každého vedeme určitou trajektorii $\mathbf{x}_t(\mathbf{x}_0)$, pak můžeme vyjádřit práci $W_{a \rightarrow b}$ jako střední hodnotu $W_{a \rightarrow b}(\mathbf{x}_0)$ přes počáteční kanonický soubor. Práci $W_{a \rightarrow b}(\mathbf{x}_0)$ je také možno chápat jako funkcionál trajektorie $W_{a \rightarrow b}[\mathbf{x}_t]$, ale jelikož předpokládáme, že trajektorie \mathbf{x}_t je jednoznačně určena počáteční podmínkou \mathbf{x}_0 , tak můžeme psát

$$W_{a \rightarrow b} = \langle W_{a \rightarrow b}(\mathbf{x}_0) \rangle_a. \quad (3.20)$$

Z termodynamiky je známo, že pokud přecházíme ze stavu a do stavu b nevratným procesem, tak práce $W_{a \rightarrow b}$ vykonaná během tohoto procesu bude větší nebo rovna změně rozdílu volné energie [55]:

$$W_{a \rightarrow b} \geq \Delta A_{a \rightarrow b}, \quad (3.21)$$

kde rovnost platí pouze pro případ vratného děje. Z rovnice 3.21 by se tedy mohlo zdát, že prostřednictvím práce můžeme získat pouze horní odhad rovnovážné

volné energie. Nicméně není tomu tak, neboť při dostatečně velkém množství realizovaných nerovnovážných MD simulací získáme z některých trajektorií hodnoty práce, které budou nižší, než je hodnota rovnovážné volné energie. Jde tedy o to, nahradit prosté průměrování vhodnou alternativou. Tou je exponenciální průměrování v Jarzynského rovnosti [82–86], poprvé zavedené v roce 1997. Pomocí ní lze získat vztah mezi hodnotami práce vykonanými během série nerovnovážných MD simulací a rovnovážnou hodnotou změny volné energie:

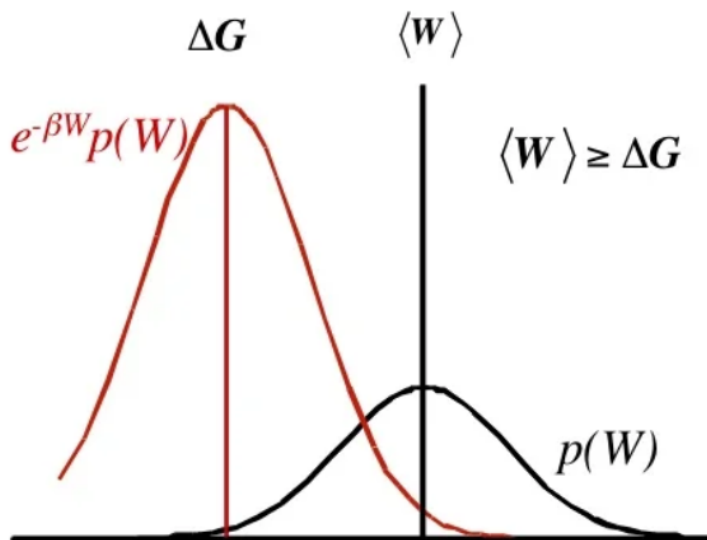
$$\Delta A_{a \rightarrow b} = -kT \ln \langle \exp(-\beta W_{a \rightarrow b}(x_0)) \rangle_a, \quad (3.22)$$

kde uvedenou střední hodnotu je nutné chápat stejně jako v případě 3.20.

Oproti rovnovážným MD simulacím, při kterých se vzorkují mikrostavy pro výpočet souborových středních hodnot, tak v případě nerovnovážných MD simulací přecházíme ke vzorkování pravděpodobnostního rozdělení $P(W)$ hodnot práce ze všech možných (či spíše reálně realizovaných) nerovnovážných trajektorií [55]. Střední hodnotu v rovnici 3.22 tak lze alternativně přepsat do tvaru:

$$\langle e^{-\beta W} \rangle = \int dW P(W) e^{-\beta W}. \quad (3.23)$$

Situace je názorně ilustrovaná na obrázku 3.5. Je zřejmé, že pravděpodobnost získání hodnoty práce blízké rovnovážné hodnotě změny volné energie může být velmi malá. Pro spolehlivý odhad volné energie je tak obvykle zapotřebí velkého množství nerovnovážných trajektorií, abychom získali dostatečný počet realizací s malými hodnotami práce W .



Obrázek 3.5: Pravděpodobnostní rozdělení $P(W)$ hodnot práce vzorkované při nerovnovážných MD simulacích. Aplikací Jarzynského identity jsme z tohoto rozdělení schopni určit hodnotu rovnovážné změny volné energie [55].

V případě, že rozdělení hodnot práce je gaussovské nebo jemu blízké, tak můžeme uvedenou exponenciální střední hodnotu 3.23 vyjádřit přibližně pomocí

omezeného rozvoje kulantů [55]:

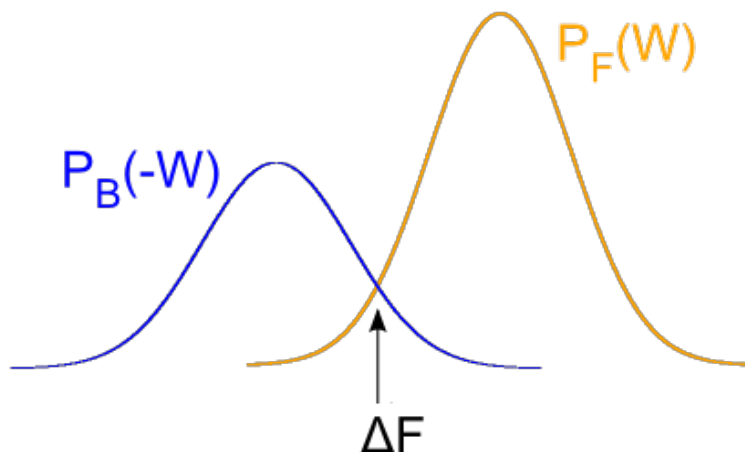
$$\ln \langle e^{-\beta W} \rangle \approx -\beta \langle W \rangle + \frac{\beta^2}{2} (\langle W^2 \rangle - \langle W \rangle^2), \quad (3.24)$$

který eliminuje problém se vzorkováním malých hodnot práce, neboť nám stačí spolehlivě určit střední hodnotu pravděpodobnostního rozdělení práce a její varianci.

Kromě Jarzynského identity můžeme k výpočtu volné energie využít Crooksův flukтуаční teorém [87, 88], který dává do souvislosti pravděpodobnostní rozdělení hodnot práce, které označíme $P_F(W)$, a rozdělení hodnot práce z opačného procesu $P_B(W)$. Crooksův flukтуаční teorém lze pak zapsat následovně [89]:

$$\frac{P_F(W)}{P_B(-W)} = e^{\beta(W-\Delta A)}. \quad (3.25)$$

Situace je graficky znázorněna na obrázku 3.6. Z rovnice 3.25 je patrné, že pokud se rozdělení hodnot práce $P_F(W)$ pro dopředný proces (např. kreaci atomu či molekuly) překrývá s rozdělením hodnot práce $P_B(-W)$ pro opačný proces (např. anihilaci atomu či molekuly), tak exponent na pravé straně rovnice musí být nulový a hodnota práce na průsečíku tak odpovídá rozdílu volné energie.



Obrázek 3.6: Využití Crooksova teorému. Pokud jsme schopni určit průsečík pravděpodobnostního rozdělení práce dopředného procesu $P_F(W)$ s analogickým rozdělením pro opačný, zpětný proces $P_B(-W)$, tak z něho lze určit rovnovážnou hodnotu rozdílu volné energie ΔF [90].

3.9 Dodatečné příspěvky (korekce) k vypočtené volné energii

Při praktickém provádění alchymistických transformací prostřednictvím MD simulací je potřeba věnovat zvláštní pozornost nabitým ligandům. Některé softwarové balíky, jako je např. i program NAMD, totiž započítávají i interakci nabitého ligandu se svými periodickými kopiemi. Periodické obrazy ligandů se většinou nachází dostatečně daleko, takže na zkoumaný ligand působí pouze elektrostatickými silami, které jsou počítány PME metodou. Pro přesné výsledky alchymistických transformací s nabitými ligandy je proto zapotřebí započíst korekci na tuto

self-interakci ligandu se svými periodickými kopiemi. Příspěvek k volné energii ze self-interakce lze pro kubické periodické systémy s periodou L vyjádřit analyticky [91]:

$$\Delta G_{\text{charge}} = \frac{\xi_{\text{Ewald}}}{L} (q_1^2 - q_0^2), \quad (3.26)$$

kde ΔG_{charge} značí korekci volné energie na vliv self-interakce, ξ_{Ewald} je Ewaldova konstanta a q_1 značí náboj molekuly na konci alchymistické transformace a q_0 na jejím počátku. Z rovnice 3.26 je patrné, že bez započítání této korekce můžeme očekávat zdánlivou závislost vypočtené volné energie na velikosti simulovaného boxu.

Při výpočtu tzv. standardní vazebné volné energie je dále zapotřebí vzít v potaz koncentraci studovaných ligandů. Většinou se totiž pro srovnání s experimentem používají hodnoty volné energie odpovídající standardní koncentraci 1 M. V MD simulacích je přitom pohyb ligandu tak či onak omezen na určitý objem V .

V případě rovnovážných MD simulací může být tento objem roven objemu celého simulačního boxu, pokud je simulace dostatečně dlouhá a ligand skutečně postupně projde celý objem simulovaného boxu. Pokud jsou ovšem při alchymistických MD simulacích zavedeny dodatečné omezující potenciály držící ligand v určitém objemu kolem vazebného místa, tak se tím také efektivně změní koncentrace ligandu.

Pro krátké nerovnovážné alchymistické MD transformace, ve kterých ligand začíná navázaný ve vazebném místě proteinu, z něhož se postupně odpoutává, je ligand efektivně omezen na objem, který je určen délkou jednotlivých trajektorií $V(\tau)$, neboť za čas τ se zkoumaný ligand může dostat pouze do omezené vzdálenosti od vazebného místa.

Korekci volné energie pro standardní stav lze souhrnně zapsat pro všechny výše zmíněné případy následovně [92]

$$\Delta G_{\text{vol}} = k_B T \ln \left(\frac{V}{V_0} \right), \quad (3.27)$$

kde k_B značí Boltzmannovu konstantu, T teplotu, objem $V_0 = 1661 \text{ \AA}^3$ odpovídá standardnímu stavu a význam objemu V je různý, podle toho, jakým způsobem byla MD simulace provedena. Obecně V značí objem, ve kterém se ligand nacházel během simulace.

4. Výsledky MD simulací

V následujících podkapitolách jsou prezentovány výsledky výpočtů volné energie pro jednoduché modelové systémy a pro komplexy koronavirové proteázy Mpro s potenciálními inhibitory. Výpočty volné energie byly prováděny prostřednictvím velkého množství krátkých nerovnovážných MD simulací. Proto byly tyto výpočty realizovány v MetaCentru (metavo.metacentrum.cz), které sdružuje univerzitní a akademická výpočetní centra po celé České republice. Jeho hardwarové prostředky jsou studentům VŠ k dispozici volně (pro studijní či vědecké účely). V rámci této práce bylo v MetaCentru spuštěno přibližně 140000 úloh a propočítáno 180000 CPU dnů.

Při výpočtech volné energie bylo potřeba běžně spouštět a následně analyzovat série 100 až 1000 MD simulací. K tomuto účelu bylo nezbytné vytvořit cca. 200 (často poměrně obsáhlých) skriptů. Několik ukázek je přiloženo v Apendixu. Po drobných úpravách je možné tyto skripty použít k MD simulacím podobných molekulárních systémů. Skripty jsou proto k dispozici na vyžádání u autora či vedoucího této práce.

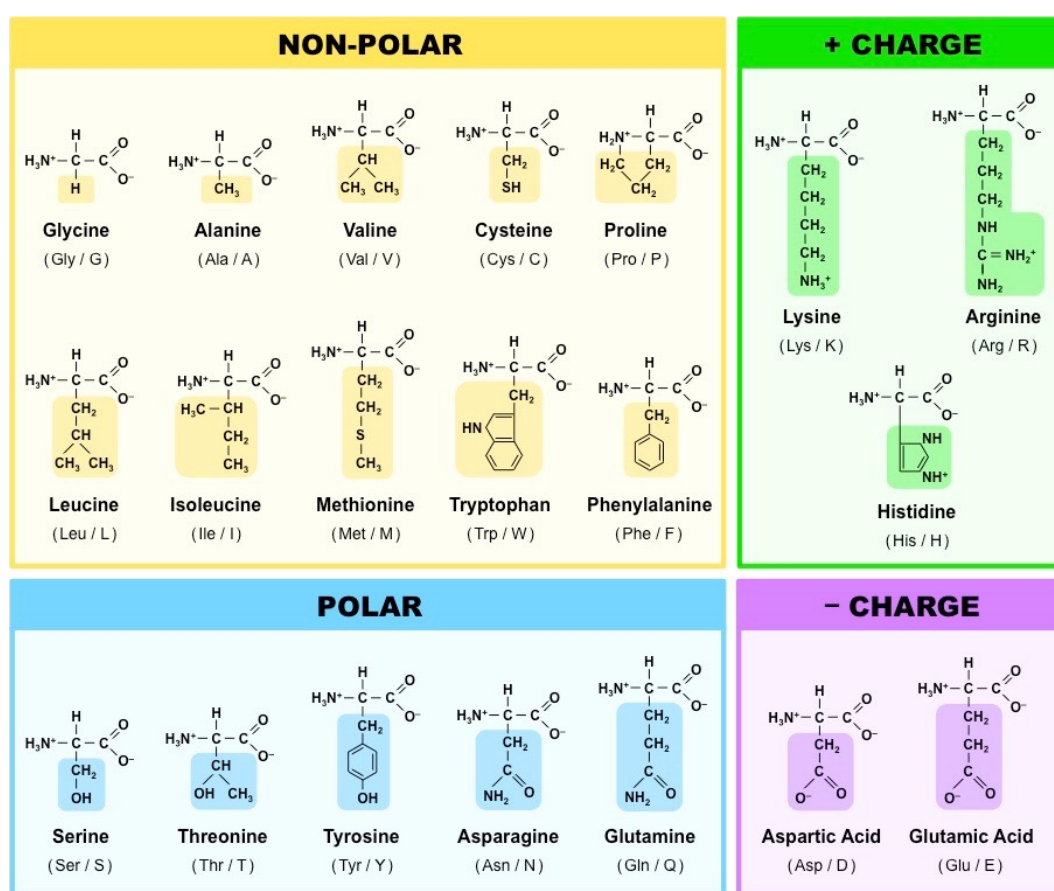
MD simulace byly realizovány prostřednictvím softwarového balíku NAMD [56]. Pro popis koronavirové proteázy Mpro jsem zvolil silové pole CHARMM [51], které je určeno pro MD simulace biomolekul - proteinů, nukleových kyselin či lipidů. Pro malé molekuly a ligandy jsem ve většině případů použil silové pole CGenFF (verze 1.0) [52]. Silová pole CHARMM a CGenFF jsou kompatibilní. Například vynechávají ne vazebné interakce mezi atomy, které jsou propojené přes méně než tři kovalentní vazby. Tyto interakce jsou zahrnuty ve vazebných členech.

Všeckrá příprava simulačních protokolů a následná analýza MD trajektorií byla prováděna pomocí programu VMD [93] a jeho programovacího rozhraní Tk konzole. Během přípravy simulačních protokolů jsem se soustředil především na optimalizaci parametrů spojených přímo s alchymistickými výpočty volné energie. Obecné parametry byly nastaveny pro všechny simulované systémy více méně stejně a obvyklým způsobem. Cut-off vzdálenost pro van der Waalsovy interakce činila 12,0 Å a zároveň byly van der Waalsovy interakce od switch vzdálenosti 10,0 Å spojitě modifikovány tak, aby ve zvolené cut-off vzdálenosti měly nulovou hodnotu. Vzdálenost párů atomů, mezi kterými byly van der Waalsovy interakce počítány, byla 14,0 Å. Krátkodosahové ne vazebné interakce byly aktualizovány v každém časovém kroku a plná elektrostatická interakce každý druhý krok. Pro výpočet elektrostatické interakce byla použita metoda PME s mřížkou o periodě 1 Å. Pomocí algoritmu SHAKE byly všechny kovalentní vazby s atomem vodíku omezeny na konstantní délku. Díky tomu jsem mohl použít časový integrační krok 2 fs. Pro kontrolu teploty byl zvolen Langevinův termostat s koeficientem tření 1 ps^{-1} . Většina MD simulací probíhala při teplotě 300 K. Pro kontrolu tlaku byl použit Nosé-Hooverův barostat v kombinaci s Langevinovou dynamikou. Cílový tlak odpovídal 1 bar.

Vizualizaci simulovaných systémů jsem prováděl pomocí programů VMD [93] a UCSF Chimera [94].

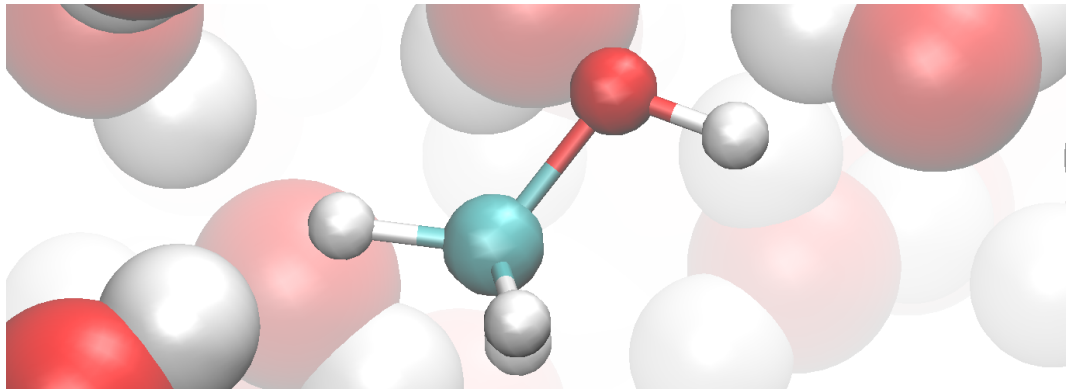
4.1 Solvatace aminokyselin

První skupina molekul, které jsem se ve své práci věnoval, byly postranní řetězce aminokyselin. Pro všechny základní aminokyseliny, zobrazené na obrázku 4.1, jsem tedy připravil molekulu odpovídající pouze jejímu postrannímu řetězci. To znamená bez alfa uhlíku a také bez amino- a karboxylové skupiny. Například pro serin je analogem jeho postranního řetězce molekula methanolu, jak ukazuje obrázek 4.2. Glycin (bez postranního řetězce) a velmi specifický prolin byly vynechány, neboť analogy jejich postranních řetězců není možné smysluplně definovat. Pro neutrální histidin byly připraveny dvě verze jeho analogu s různě umístěným vodíkem. Hid má protonovaný δ -dusík a Hie ϵ -dusík. V následujícím textu pro jednoduchost někdy označuji analogy postranních řetězců aminokyselin jako “aminokyseliny” a také pro ně používám názvy jejich mateřských aminokyselin.



Obrázek 4.1: Struktura základních aminokyselin. Pro MD simulace byly připraveny molekuly odpovídající postranním řetězcům (s výjimkou glycinu a prolinu) [95].

Pro analogy postranních řetězců aminokyselin byla následně studována jejich solvatační volná energie. Tedy volná energie spojená s přechodem těchto molekul z vakua do vodního prostředí. Jedná se o podstatně jednodušší proces, než je vazba ligandu na protein. Pro výpočet solvatační volné energie totiž nepotřebujeme uvažovat termodynamický cyklus, který byl definován rovnicí 3.15. Stačí zde uvažovat jednu alchymistickou transformaci, při které molekula mizí, nebo se naopak postupně objevuje ve vodním prostředí.



Obrázek 4.2: Příklad analogu postranního řetězce aminokyseliny serinu ve vodním prostředí. Bez alfa uhlíku, amino- a karboxylové skupiny je analogem postranního řetězce serinu molekula methanolu.

Nerovnovážený způsob výpočtu volné energie pro komplexní molekulární systémy stále není běžně používán a v literatuře se setkáváme spíše s tradičními výpočty prováděnými pomocí rovnovážných MD simulací. Bylo proto potřeba vyřešit některé technické detaily a připravit pracovní postup, který by mi umožnil najednou spustit a analyzovat velké množství trajektorií, řádově 1000, pro každou jednu studovanou molekulu. K přípravě simulovaných systémů a následné analýze výsledků MD simulací jsem využil kombinaci vlastních bash skriptů a TCL programovatelné rozhraní integrovaného v softwarovém balíku VMD. Jeden příklad takového skriptu je uveden jako příloha A.1.

Kromě toho bylo třeba ověřit shodu výsledků získaných prostřednictvím nerovnovážných MD simulací s referenčními hodnotami získanými prostřednictvím experimentu a rovnovážných MD simulací. K tomuto účelu jsou výpočty hydratační volné energie postranních řetězců aminokyselin velmi vhodné, neboť v literatuře je dostatek referenčních hodnot hydratační volné energie získaných jak z experimentů, tak z rovnovážných MD simulací.

Simulované systémy jsou v případě solvovaných analog postranních řetězců aminokyselin velmi malé a MD simulace jsou díky tomu poměrně rychlé. Tento typ modelového systému byl tedy pro počáteční odladění metodiky výpočtů volné energie prostřednictvím nerovnovážných MD simulací vhodnější než komplexnější systémy, pro které by výpočty trvaly podstatně déle.

Kromě toho jsem analoga postranních řetězců aminokyselin zvolil také kvůli ověření, zda je možné silové pole CGenFF později použít pro věrohodné výpočty vazebné volné energie ligandů na bázi analog peptidů (tedy ligandů jejichž stavebními kameny jsou analoga aminokyselin) do aktivního místa koronavirové proteázy Mpro. Mimo jiné jsem zkoumal, jak se shodují výsledky získané prostřednictvím různých verzí silového pole CGenFF a zda je možné shodu s experimentem vylepšit vlastní reparametrizací silového pole CGenFF prostřednictvím ab initio výpočtů.

4.1.1 Příprava simulovaných systémů

Výchozí struktury analog postranních řetězců aminokyselin byly připraveny v modulu Molefactory, který je součástí softwarového balíku VMD. Výstupem z tohoto programu jsou soubory se souřadnicemi atomů (pdb) a jejich konektivitou (psf), které ale neobsahují žádnou informaci o parametrech silového pole. Pro popis analog postranních řetězců aminokyselin bylo zvoleno silové pole CGenFF (cgenff.umaryland.edu) které existuje ve dvou verzích (v1.0 a v2.5) díky mírně odlišnému algoritmu pro odhad parciálních nábojů na atomech [52]. V této podkapitole jsem otestoval obě verze tohoto silového pole a následně jsem otestoval i vlastní parametrizaci.

Připravené struktury posloužily jako vstup pro webovou aplikaci CHARMM-GUI [96], která je schopná rychle generovat veškeré nezbytné vstupní soubory pro simulační program NAMD. Molekuly analog postranních řetězců aminokyselin přitom byly obklopeny molekulami vody reprezentovanými TIP3P modelem. Byl definován simulační box s hranou o délce přibližně 50 Å a byly aplikovány periodické okrajové podmínky. Do simulovaných systémů byly přidány ionty draslíku popř. chloru tak, aby simulační box byl celkově neutrální. Většina ostatních simulačních parametrů zůstala beze změny tak, jak bylo popsáno v úvodní sekci. S parametry specifickými pro alchymistický výpočet, jako je dělení intervalu u λ parametru na subintervaly či jako je počet integračních kroků realizovaných na jednotlivých subintervalech, jsem experimentoval a konkrétní hodnoty těchto parametrů jsou uvedeny u odpovídajících dílčích výsledků níže.

MD simulace pro všechny aminokyseliny začaly 200 kroky minimalizace, při které byly odstraněny blízké kontakty atomů. Navazující MD simulace tak byly stabilní. Po minimalizaci jsem spustil 20 ps ekvilibrační běh. Jednalo se o obyčejnou MD simulaci. Počáteční rychlosti byly náhodně vygenerovány programem NAMD v souladu Maxwellovým-Boltzmannovým rozdělením. Následné působení termostatu a barostatu uvedlo systém do termodynamické rovnováhy, která odpovídá statistickému souboru NpT .

Pro výpočty hydratační volné energie jsem zvolil přístup podle Crooksova teorému 3.25. Prováděl jsem alchymistické transformace, při kterých byly interakce aminokyselin s okolními molekulami škálovány postupně na nulovou hodnotu, a aminokyseliny tak efektivně postupně mizely ze simulovaného systému. Tento směr jsem označil jako dopředný. Následně jsem připravil a provedl MD simulace, ve kterých probíhal opačný proces. Aminokyseliny se postupně objevovaly z ničeho ve vodním prostředí. Tento směr simulací jsem označil jako zpětný. Z překryvu distribucí hodnot prací z dopředných a zpětných běhů jsem pak byl s pomocí Crooksova teorému schopen určit rovnovážnou hodnotu hydratační volné energie.

U alchymistických transformací je důležité správně zvolit, jak rozdělit interval hodnot 0 až 1 pro parametr λ vystupující v rovnici 3.16 na subintervaly, a jak rychle alchymistické transformace v jednotlivých λ -oknech provádět.

V případě rovnovážných MD simulací bych musel zaručit, aby v každém λ -okně měl simulovaný systém dostatek času dostat se do termodynamické rovnováhy. Až poté by bylo možné začít vzorkovat rovnovážnou střední hodnotu a určit volnou energii např. podle rovnice 3.17.

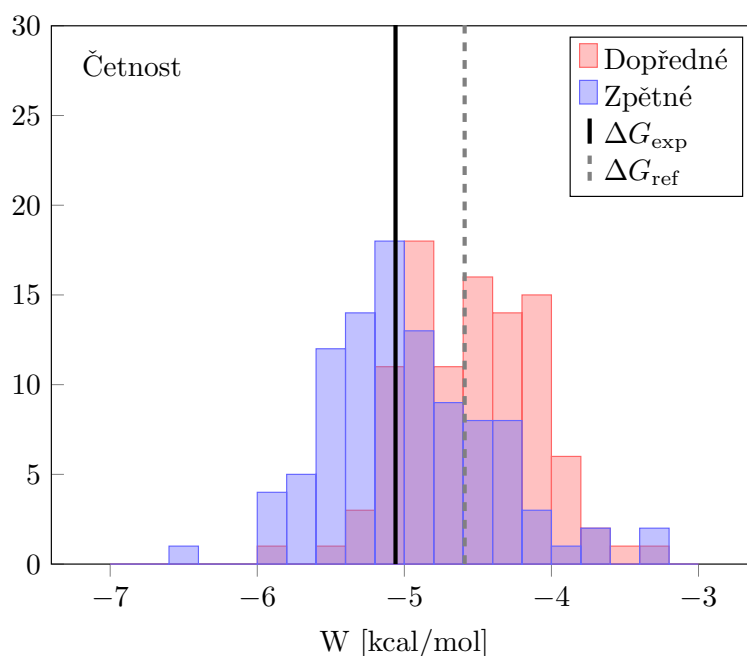
V případě nerovnovážných MD simulací, které jsem použil zde, ale stačí, aby byl simulovaný systém v termodynamické rovnováze na počátku alchymistických

transformací. V následujících λ oknech pak již není potřeba systém ekvilibrovat. Volba rychlosti nerovnovážné alchymistické transformace je do značné míry libovolná, ale pro velmi rychlé transformace lze očekávat, že simulované systémy budou daleko od termodynamické rovnováhy, a pro spolehlivou aplikaci Crooksova teorému 3.25 bude zapotřebí simulovat velmi velké množství trajektorií. Volba rychlosti alchymistických transformací tak obvykle představuje kompromis učiněný s ohledem na realizovatelnou délku jednotlivých MD běhů ve vymezeném čase a s ohledem na celkový počet MD běhů, který je možno naráz spustit na dané hardwarové platformě.

Nejprve jsem důkladně otestoval různé rychlosti alchymistických transformací. Zkoušel jsem měnit jak dělení intervalu λ , tak počet integračních kroků realizovaných v jednotlivých λ oknech. Uspokojivých výsledků jsem dosáhl v případě, kdy jsem spustil 100 dopředných a 100 zpětných běhů, λ interval byl ekvidistantně rozdělen do 10 oken s krokem $\Delta\lambda = 0,1$ a jednotlivá λ okna byla simulována po dobu 20 ps s celkovou délkou alchymistické transformace 200 ps.

4.1.2 Výsledky simulací

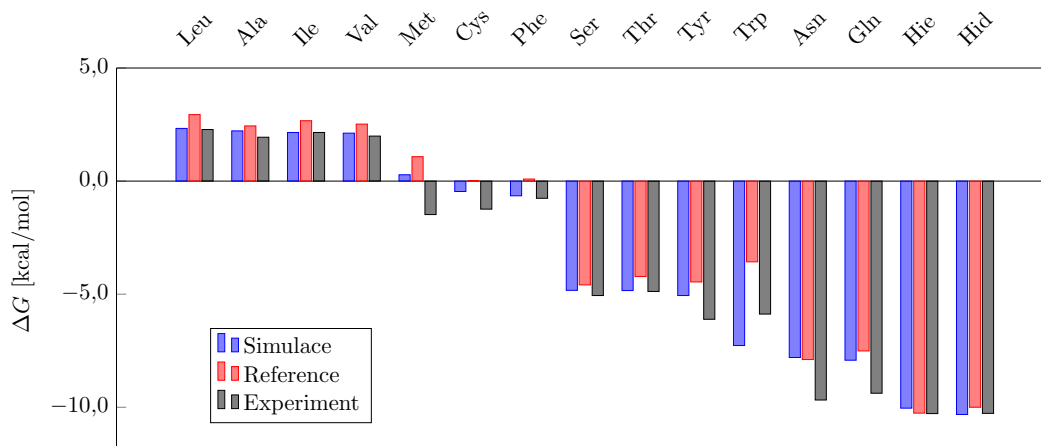
Příklad výsledku MD simulací pro postranní analog serinu a silové pole CGenFF verze 1 je znázorněn na obrázku 4.3. Výsledné hodnoty práce ze všech 2x100 trajektorií jsou znázorněny pomocí histogramů. Efektivně se tak během takto provedených MD simulací vzorkují pravděpodobnostní rozdělení vystupující v Crooksově teorému 3.25 a z překryvu histogramů hodnot z dopředného a zpětného běhu je možné určit rovnovážnou hodnotu hydratační volné energie.



Obrázek 4.3: Příklad výsledků ze 100 dopředných a 100 zpětných trajektorií pro analog postranního řetězce serinu. Překryv histogramů je zvýrazněn fialovou barvou.

Obdobně získané výsledky pro zbytek analog postranních řetězců neutrálních aminokyselin jsou shrnuty v tabulce 4.1 a graficky znázorněny na obrázku 4.4. Re-

ferenční hodnoty značí hodnoty převzaté z předchozích studií [97, 98], ve kterých studovali stejné molekuly se stejným silovým polem, ale s využitím rovnovážných MD simulací. Referenční experimentální hodnoty byly určeny dynamickým měřením tlaku páry v rovnováze s řídkými roztoky aminokyselin [99].



Obrázek 4.4: Hydratační volné energie pro analoga postranních řetězce neutrálních aminokyselin získané s využitím silového pole CGenFF verze 1.0. Výsledky z nerovnovážných MD simulací jsou zde porovnány s referenčními hodnotami rovnovážných MD simulací [97, 98] a také s experimentálními hodnotami [99].

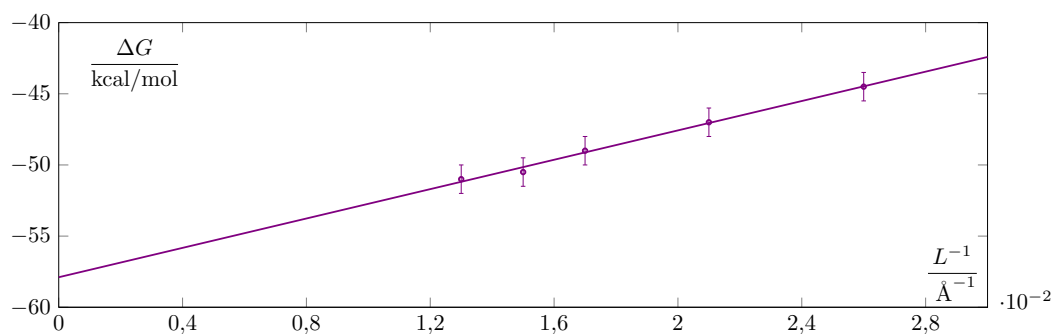
Aminokyselina	ΔG_{v1}	ΔG_{v2}	ΔG_{exp}	ΔG_{ref}
Leu	2,3	2,5	2,3	2,9
Ala	2,2	2,4	1,9	2,4
Ile	2,2	2,3	2,2	2,7
Val	2,1	2,2	2,0	2,5
Met	0,3	0,5	-1,5	1,1
Cys	-0,5	-0,4	-1,2	0,0
Phe	-0,7	-0,4	-0,8	0,1
Ser	-4,8	-4,8	-5,1	-4,6
Thr	-4,8	-4,7	-4,9	-4,2
Tyr	-5,1	-4,9	-6,1	-4,5
Trp	-7,3	-6,5	-5,9	-3,6
Asn	-7,8	-8,0	-9,7	-7,9
Gln	-7,9	-7,8	-9,4	-7,5
Hie	-10,0	-10,6	-10,3	-10,3
Hid	-10,3	-10,7	-10,3	-10,0

Tabulka 4.1: Výsledné hodnoty hydratační volné energie postranních řetězců neutrálních aminokyselin. Hodnoty jsou uvedeny v jednotkách kcal/mol. Výsledky pro silové pole CGenFF verze 1.0 jsou označeny ΔG_{v1} a pro verzi 2.5 ΔG_{v2} .

Pro nabité aminokyseliny je při výpočtu hydratační volné energie potřeba vzít do úvahy několik dalších faktorů. Především jde o interakci nabitých molekul se svými periodickými kopiemi a přechod přes vodní rozhraní.

Během alchymistické transformace mizí nabitá aminokyselina postupně ze simulovaného systému a tím se mění i jeho celkový náboj. Studovaná molekula zároveň v průběhu celé transformace interaguje se svými periodickými obrazy a to prostřednictvím elektrostatických interakcí, jejichž vliv je většinou započten prostřednictvím Ewaldovy sumace 2.12. V současné implementaci programu NAMD není ve výsledné volné energii odstraněn příspěvek z této self-interakce. Z rovnice 3.26 je patrné, že bez započítání příslušné korekce můžeme očekávat zdánlivou závislost volné energie na velikosti simulovaného boxu.

To jsem si ověřil na příkladu kladně nabitého lysinu, pro který jsem spustil MD simulace s různými velikostmi simulačního boxu. Na obrázku 4.5 jsou znázorněny výsledky z této série MD simulací. Vidíme zde, že závislost volné energie na velikosti simulačního boxu skutečně odpovídá očekávané teoretické hodnotě 3.26 a že výsledné hodnoty volné energie je tedy potřeba o tuto hodnotu korigovat. Pro většinu zde provedených MD simulací byl příspěvek této korekce roven $-8,12$ kcal/mol.



Obrázek 4.5: Výpočty hydratační volné energie pro kladně nabitou molekulu lysinu pro různé velikosti simulačního boxu. Extrapolací na nekonečnou velikost simulačního boxu lze ověřit vzorec pro analytickou korekci 3.26.

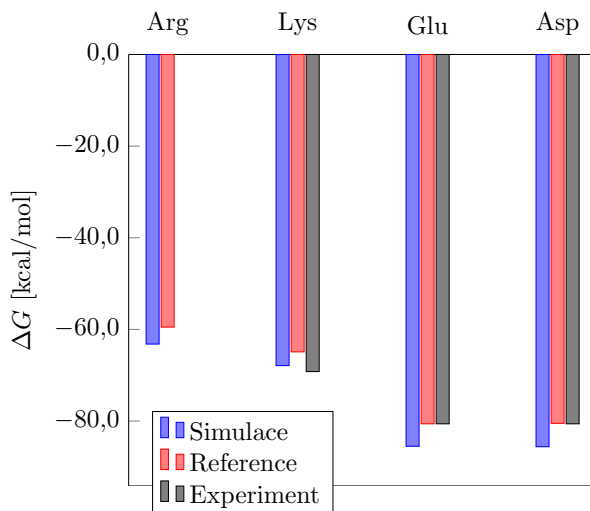
Aminokyselina	ΔG_{v1}	ΔG_{v2}	ΔG_{exp}	ΔG_{ref}
Arg	-63,2	-59,4	-	-59,5
Lys	-67,9	-69,7	-69,2	-64,9
Glu	-85,5	-86,0	-80,6	-80,6
Asp	-85,6	-85,8	-80,6	-80,5

Tabulka 4.2: Výsledné hodnoty hydratační volné energie nabitých aminokyselin. Všechny hodnoty jsou uvedené v jednotkách kcal/mol. Výsledky pro silové pole CGenFF verze 1.0 jsou označeny ΔG_{v1} a pro verzi 2.5 ΔG_{v2} .

Experimentální hodnoty hydratační volné energie v sobě navíc zahrnují příspěvek odpovídající přenosu nabité molekuly přes rozhraní voda/vzduch. Pro simulovaný systém s molekulami vody typu TIP3P je tento příspěvek roven

-11,53 kcal/mol pro kladně nabitě molekuly a +11,53 kcal/mol pro molekuly nabitě záporně.

Pro analoga postranních řetězců nabitých aminokyselin jsou výsledné hodnoty hydratační volné energie uvedeny v tabulce 4.2 a graficky jsou znázorněny v grafu 4.6.



Obrázek 4.6: Výsledné hodnoty hydratační volné energie nabitých aminokyselin pro silové pole CGenFF verze 1.0. Výsledky simulací jsou zde porovnány s referenčními hodnotami získanými z rovnovážných MD simulací [97, 98] a také s experimentálními hodnotami [99].

Výsledné hodnoty hydratační volné energie mohou být ovlivněny i některými nastaveními simulačního protokolu, ve kterých se často dopouštíte zjednodušujících předpokladů pro urychlení výpočtů. Důležitý je zde např. parametr cut-off pro van der Waalovy interakce, jelikož pro vzdálenosti větší než je hodnota tohoto parametru, tyto interakce zcela zanedbáváme.

Pro odhad korekce ΔG_{LRC} na konečnou hodnotu cut-off vzdálenosti jsem použil dva různé přístupy.

První přístup spočíval ve spuštění dodatečných MD simulací s nastavením LJcorrection na hodnotu on. Díky tomu softwarový balík NAMD aplikuje analytickou korekci k van der Waalově energii založenou na středování všech van der Waalových parametrů přítomných v simulovaném systému a předpokladu jejich homogenního rozložení za cut-off vzdáleností [100].

Druhý způsob byl založen na dodatečné rovnovážné MD simulaci, ve které jsem pro jednotlivé snímky určil rozdíl van der Waalovy energie počítaný pro původní cutoff 12 Å a pro větší cutoff 50 Å. Výsledky jsou shrnuty v tabulce 4.3.

Nakonec jsem se pokusil hodnoty hydratační volné energie postranních řetězců aminokyselin upřesnit s využitím vlastní parametrizace. Pro tento účel jsem využil modul FFTk (angl. Force Field Toolkit) [101] zabudovaný v programu VMD, který umožňuje systematickou a rychlou reparametrizaci malých molekul v souladu s filozofií silového pole CHARMM.

Vyšel jsem ze silového pole CGenFF verze 1.0. Vazebné silové konstanty jsem ponechal beze změny a zaměřil jsem se především na reparametrizaci bodových nábojů postupem, který se uplatňuje při tvorbě silového pole CHARMM.

Aminokyselina	$\Delta G_{\text{LRC},1}$	$\Delta G_{\text{LRC},2}$	$\Delta G_{\text{LRC,ref}}$
Ala	-0,01	-0,12	-0,11
Arg	-0,34	-0,64	-0,60
Asn	-0,23	-0,33	-0,30
Asp	-0,19	-0,28	-0,40
Cys	-0,16	-0,28	-0,27
Gln	-0,25	-0,42	-0,39
Glu	-0,24	-0,36	-0,40
Hid	-0,37	-0,47	-0,41
Hie	-0,26	-0,47	-0,41
Ile	-0,17	-0,39	-0,38
Leu	-0,20	-0,39	-0,36
Lys	-0,61	-0,48	-0,20
Met	-0,31	-0,47	-0,47
Phe	-0,40	-0,56	-0,53
Ser	-0,12	-0,18	-0,17
Thr	-0,19	-0,27	-0,26
Trp	-0,54	-0,76	-0,74
Tyr	-0,45	-0,62	-0,58
Val	-0,12	-0,30	-0,29

Tabulka 4.3: Odhady korekce volné energie na konečný cutoff. Výsledky využívající NAMD LJcorrection jsou označeny $\Delta G_{\text{LRC},1}$ a výsledky počítané z rozdílu energií pro menší a větší cutoff $\Delta G_{\text{LRC},2}$. Referenční hodnoty byly převzaty z [97, 98]. Veškeré hodnoty jsou uvedeny v jednotkách kcal/mol.

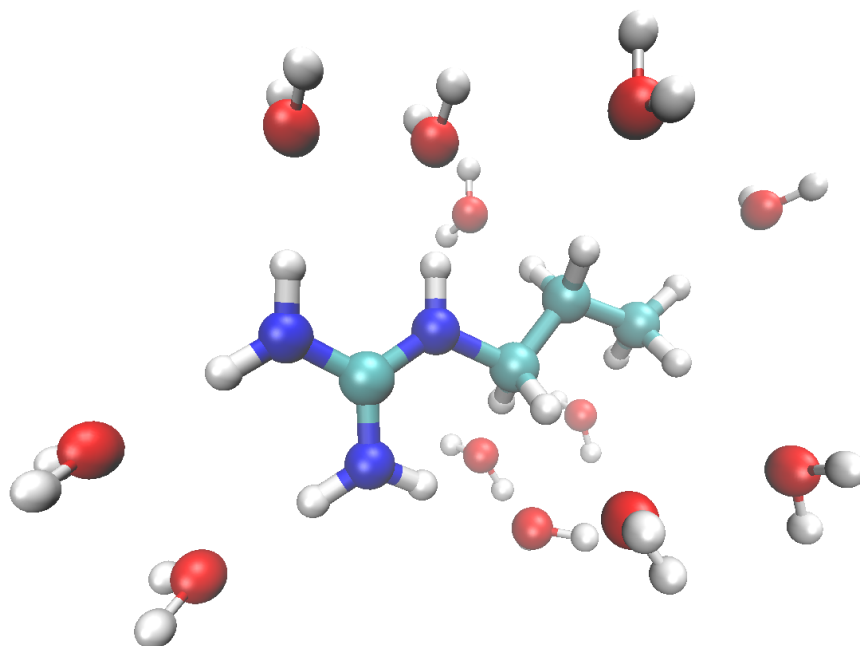
Každý vnější atom studované molekuly, který je v přímém kontaktu s vodními molekulami, je označen jako potenciální donor nebo akceptor vodíkové vazby. Pro každý z těchto atomů jsou modulem FFTk připraveny vstupní soubory pro kvantově-mechanické výpočty, ve kterých je na počátku jedna molekula vody umístěna v pokud možno optimální pozici a orientaci pro tvorbu vodíkové vazby. Následně je provedena omezená ab initio optimalizace geometrie komplexu, ve které je optimalizována pouze vzdálenost a natočení molekuly vody. Tyto kvantově-mechanické výpočty jsem provedl prostřednictvím programu Gaussian [102]. Volí se zde velmi přibližná Hartree-Fockova metoda a malá báze HF/6-31G(d), aby výsledky byly konzistentní se silovým polem CHARMM.

Výsledky takto provedených kvantově-mechanických optimalizací jsem následně využil v modulu FFTk, který určí interakční energii jako rozdíl energie komplexu a jednotlivých molekul, a extrahuje optimalizované vzdálenosti. Pro aproximaci polárního vodního prostředí jsou navíc optimalizované vzdálenosti posunuty o $-0,2 \text{ \AA}$ a interakční energie škálovány faktorem 1,16 [101].

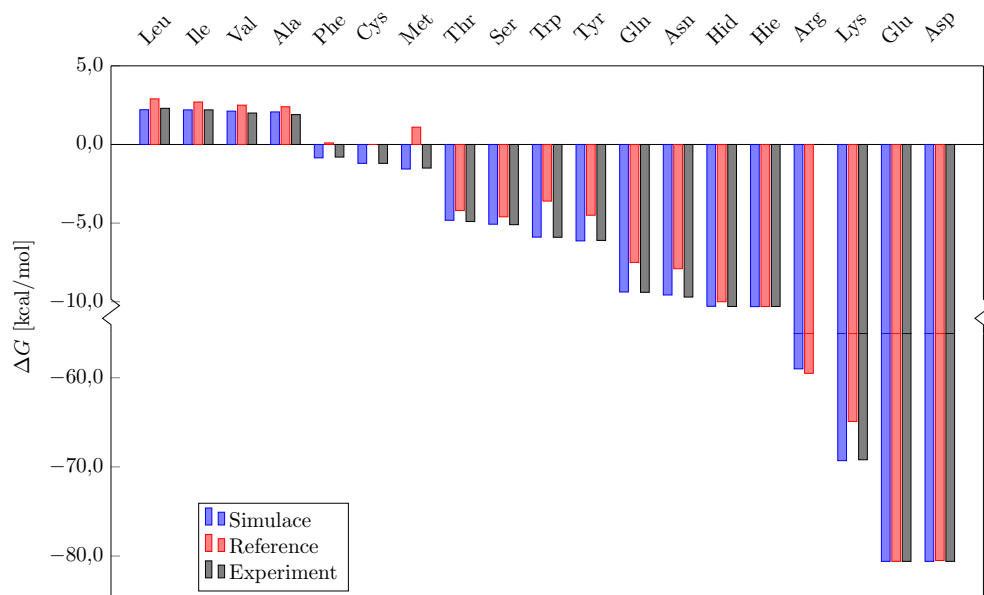
Při následné optimalizaci bodových nábojů jsou náboje chemicky ekvivalentních atomů nastaveny na stejnou hodnotu a nepolárním vodíkům ve vazbách s uhlíkem je konvenčně přiřazena fixní hodnota +0,09.

Optimalizace bodových nábojů je pak založena na reprodukování kvantově-mechanických výsledků pomocí molekulárně-mechanického popisu při využití TIP3P modelu pro molekuly vody. Při optimalizaci se minimalizuje rozdíl pro celkový dipól molekuly, energii v minimu a optimální vzdálenost molekuly vody. Optimalizace probíhá iterativně dokud není rozdíl pod zadanou tolerancí. Příklad výstupu z této optimalizace je znázorněn na obrázku 4.7 pro analog postranního řetězce argininu. Stejný postup jsem aplikoval na všechna ostatní studovaná analoga postranních řetězců aminokyselin.

Výsledky z kvantově-mechanických optimalizací jsem posléze využil v modulu FFTk pro další reparametrizaci nábojů. Zaměřil jsem se především na empirický faktor $f = 1,16$, který FFTk modul využívá k aproximaci polárního vodního prostředí a škáluje jím interakční energie. Hodnotu tohoto faktoru jsem se pokoušel postupně měnit. Dostal jsem tak nové sady parciálních nábojů studovaných molekul, které jsem využil pro nové výpočty hodnot hydratační volné energie. Výpočty volné energie pro nové sady parciálních nábojů získané prostřednictvím FFTk modulu jsem spustil se stejným nastavením jako v předchozím případě. Výsledné hodnoty shrnuje tabulka 4.4 a obrázek 4.8.



Obrázek 4.7: Výsledné vzdálenosti molekul vody získané z kvantově-mechanické optimalizace pro analog postranního řetězce argininu. Tyto vzdálenosti byly následně využity při parametrizaci bodových nábojů v modulu FFTk.



Obrázek 4.8: Výsledné hodnoty hydratační volné energie analog postranních řetězců aminokyselin parametrizovaných pomocí FFTk.

Aminokyselina	f	ΔG_{FFTk}	ΔG_{exp}	ΔG_{ref}
Leu	1,16	2,2	2,3	2,9
Ile	1,06	2,2	2,2	2,7
Val	0,90	2,1	2,0	2,5
Ala	1,36	2,1	1,9	2,4
Phe	1,43	-0,9	-0,8	0,1
Cys	1,46	-1,2	-1,2	0,0
Met	1,52	-1,6	-1,5	1,1
Thr	1,14	-4,8	-4,9	-4,2
Ser	1,25	-5,1	-5,1	-4,6
Trp	1,36	-5,9	-5,9	-3,6
Tyr	1,16	-6,1	-6,1	-4,5
Gln	1,19	-9,4	-9,4	-7,5
Asn	1,10	-9,6	-9,7	-7,9
Hid	1,05	-10,3	-10,3	-10,0
Hie	1,16	-10,3	-10,3	-10,3
Arg	0,98	-59,0	-	-59,5
Lys	1,06	-69,3	-69,2	-64,9
Glu	0,87	-80,6	-80,6	-80,6
Asp	0,89	-80,6	-80,6	-80,5

Tabulka 4.4: Výsledné hodnoty hydratační volné energie analog postranních řetězců aminokyselin parametrizovaných pomocí modulu FFTk. Faktor f zde značí empirický parametr, který škáluje kvantově-mechanické interakční energie.

4.1.3 Diskuze

Vypočtené hodnoty korekcí na konečný cutoff van der Waalsových interakcí, které jsou uvedené v tabulce 4.3 nabývají poměrně malých hodnot. Navíc při pozdějších výpočtech vazebných volných energií by byl příspěvek této korekce ještě menší, neboť její vliv se částečně odečte díky použití termodynamického cyklu 3.1. Pro další studované systémy jsem tedy tuto korekci neuvažoval.

Ve výsledných hodnotách hydratační volné energie získaných prostřednictvím nerovnovážných MD simulací s využitím Crooksova flukтуаčního teorému 4.4 lze jasně rozlišit hydrofobní aminokyseliny s kladnou hodnotou hydratační volné energie a hydrofilní aminokyseliny se zápornou hodnotou hydratační volné energie. Můžeme zde také vidět, že jsem byl schopen aplikací nerovnovážných MD simulací dosáhnout srovnatelné přesnosti, jako bylo dosaženo prostřednictvím referenčních rovnovážných MD simulací. Díky masivně paralelnímu provedení krátkých nerovnovážných MD simulací v MetaCentru byly výsledky z nich získány ve výrazně kratším čase, než by tomu bylo u dlouhé rovnovážné MD simulace.

V tabulce 4.4 můžeme pozorovat pouze nepatrný rozdíl mezi výsledky získanými se silovým polem CGenFF verze 1.0 a 2.5. Pro další studované systémy jsem proto využil verzi 1.0, která je v literatuře více využívána.

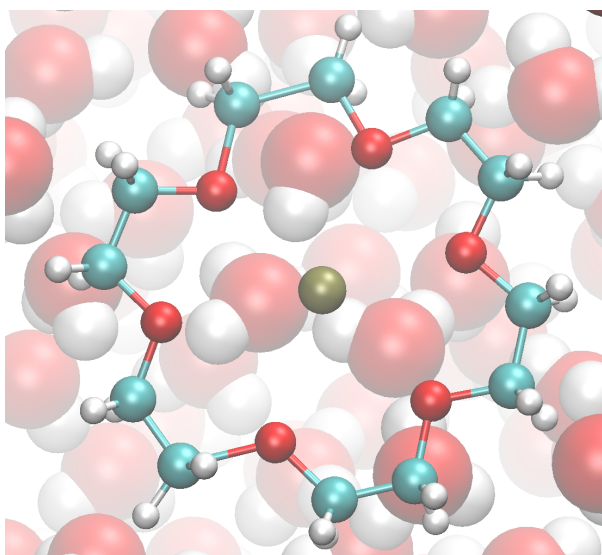
Shoda mezi experimentálními a vypočtenými hodnotami je obecně poměrně dobrá. V případě methioninu má ovšem výsledná (velmi malá) hodnota opačné znaménko oproti experimentální hodnotě. V případě tryptofanu můžeme pozorovat rozdíl v řádu 1 kcal/mol. Podobný nesoulad ale nastal i u referenčních hodnot z rovnovážných MD simulací a lze proto předpokládat, že byl způsoben systematickou chybou v silovém poli CGenFF.

Tento nesoulad jsem se pokoušel eliminovat vlastní reparametrizací bodových nábojů všech zkoumaných analog. Z grafu 4.8 je jasně vidět výrazné přiblížení nově vypočtených výsledků k referenčním experimentálním hodnotám (ve srovnání s předchozími výsledky 4.4 získanými pomocí generických CGenFF parametrů) a to včetně kritických analog postranních řetězců methioninu a tryptofanu. Z tabulky 4.4 je přitom vidět, že optimální hodnota empirického faktoru f , který byl variován, aby ke sblížení vypočtených a experimentálních hydratačních volných energií došlo, se pro jednotlivá analoga postranních řetězců aminokyselin značně liší. Není tedy možné stanovit jednu optimální univerzální hodnotu tohoto empirického parametru f , kterým imitujeme přítomnost polárního vodního prostředí.

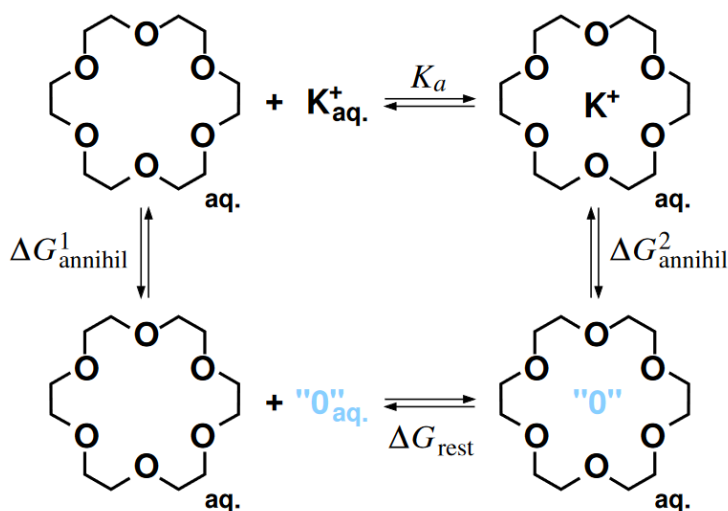
Nicméně pro potenciální inhibitory proteázy Mpro, které často mají formu analog krátkých peptidů, jistě nebude problém zkorigovat parciální náboje jejich stavebních kamenů na základě strukturní analogie s aminokyselinami a s využitím zde vypočtených nových hodnot parciálních nábojů. Kromě toho by bylo zajímavé použít tyto nové hodnoty parciálních nábojů i u aminokyselin v aktivním místě proteázy Mpro. A samozřejmě zjistit, jak to vše ovlivní vazebné volné energie.

4.2 Komplex éterové koruny a draslíku

Dalším studovaným systémem byl komplex éterové koruny a iontu draslíku, jehož struktura je znázorněna na obrázku 4.9. Jedná se o jednoduchý modelový systém, na kterém je možné otestovat výpočty vazebné volné energie. Pro tyto výpočty již bylo zapotřebí využít termodynamického cyklu znázorněného na obrázku 4.10. Bylo tedy třeba realizovat separátní alchymistické transformace jednak pro navázaný a jednak pro volný ion.



Obrázek 4.9: Struktura komplexu éterové koruny s navázaným iontem draslíku ve vodním prostředí.



Obrázek 4.10: Termodynamický cyklus použitý při výpočtu vazebné volné energie pro komplex éterové koruny a draslíku [103].

Jedná se o modelový systém, pro který je známá jak experimentální hodnota vazebné volné energie [104], tak její hodnoty z MD simulací založených na rovnovážném přístupu [103]. Proto jsem tento modelový systém využil k tomu,

abych zjistil, do jaké míry jsou nerovnovážné MD simulace vhodným nástrojem pro výpočet vazebné volné energie. Malé rozměry tohoto simulovaného systému umožňují získat výsledky relativně rychle. To dovoluje rozsáhle experimentovat se simulačními parametry, vyzkoušet různé alternativní přístupy a odladit technické detaily výpočtů vazebné volné energie prostřednictvím krátkých nerovnovážných MD simulací.

4.2.1 Příprava simulovaných systémů

Pro popis éterové koruny jsem opět využil silové pole CGenFF.

Obecné simulační parametry byly stejné jako v případě předchozích MD simulací solvatace aminokyselin. Zaměřil jsem se především na parametry přímo spojené s alchymistickými transformacemi, jako je jejich délka a celkový počet. Nastavení většiny simulačních parametrů jsem testoval pouze pro MD simulace s iontem navázaným do éterové koruny. MD simulace pro samotný ion ve vodě byly provedeny až následně čistě za účelem určení vazebné volné energie. Konkrétní hodnoty simulačních parametrů jsou uvedeny u jednotlivých výsledků níže.

Pro tento jednoduchý modelový systém jsem se zaměřil na srovnání dvou způsobů výpočtu vazebné volné energie. První přístup spočíval v aplikaci Crooksova teorému analogicky, jak jsem to prováděl při výpočtu hydratační volné energie aminokyselin. Takže bylo zapotřebí spouštět alchymistické transformace oběma směry. Druhý testovaný přístup spočíval v přímé aplikaci Jarzynského identity (3.22). Alchymistické transformace byly v tomto případě spouštěny pouze jedním směrem.

Velkou pozornost jsem věnoval využití dodatečných omezení držících iont draslíku ve vazebném místě éterové koruny během alchymistických transformací. (tj. aplikaci postupu popsaného rovnicí 3.15).

Jako dopředné simulace zde opět označuji ty alchymistické transformace, ve kterých ion postupně mizel ze simulovaného systému, a jako zpětné simulace ty, ve kterých se naopak v simulovaném systému postupně objevoval.

4.2.2 Výsledky simulací

Při testování prvního způsobu výpočtu vazebné volné energie pomocí Crooksova teorému se ukázalo, že zpětné simulace nelze vůbec využít, pokud ion není držen v blízkosti vazebného místa v éterové koruně dodatečným omezujícím potenciálem. Bez zavedení tohoto omezení docházelo k tomu, že ion na konci dopředné simulace, kdy jeho interakce s okolím byly téměř nulové, oddifundoval z vazebného místa v éterové koruny. Zpětné simulace pak začínaly s iontem mimo vazebné místo a pouze s velmi malou pravděpodobností se během nich ion vrátil do původní polohy.

V dalších MD simulacích jsem proto zavedl omezující potenciál $V(r)$ působící mezi hmotnými středy éterové koruny a iontu. Pro tento účel jsem použil modul Colvars zabudovaný v použitém simulačním programu NAMD. Tento modul umožňuje jednoduché zavedení rozmanitých omezujících potenciálů do simulovaných systémů. Konkrétně jsem zvolil omezující potenciál ve tvaru harmonické stěny, který je nulový až do definovaného poloměru r_0

$$V(r) = \begin{cases} \frac{1}{2}k \left(\frac{r - r_0}{w} \right)^2 & r > r_0, \\ 0 & r < r_0, \end{cases} \quad (4.1)$$

kde silová konstanta k byla nastavena na 100,0 kcal/mol a šířka w měla hodnotu 0,1 Å. Ion tak byl efektivně držen v objemu koule o poloměru r_0 kolem středu éterové koruny.

Zavedením omezujícího potenciálu 4.1 omezíme konfigurační prostor dostupný studovanému iontu a dojde tak ke ztrátě jeho translační entropie, což ovlivní i výslednou volnou energii. Pro určení vlivu této vazby bych obecně musel postupovat podle termodynamického cyklu z obrázku 3.2, kde se navíc provádí dodatečné MD simulace určující příspěvky omezujících potenciálů. V případě tohoto konkrétního velmi jednoduchého omezujícího potenciálu ovšem existuje analytické vyjádření korekce volné energie [92], pro které mi stačí odhadnout objem ΔV navzorkovaný iontem během MD simulace (3.27). Objem ΔV jsem jednoduše odhadl jako objem koule o poloměru zavedeného omezujícího potenciálu r_0 .

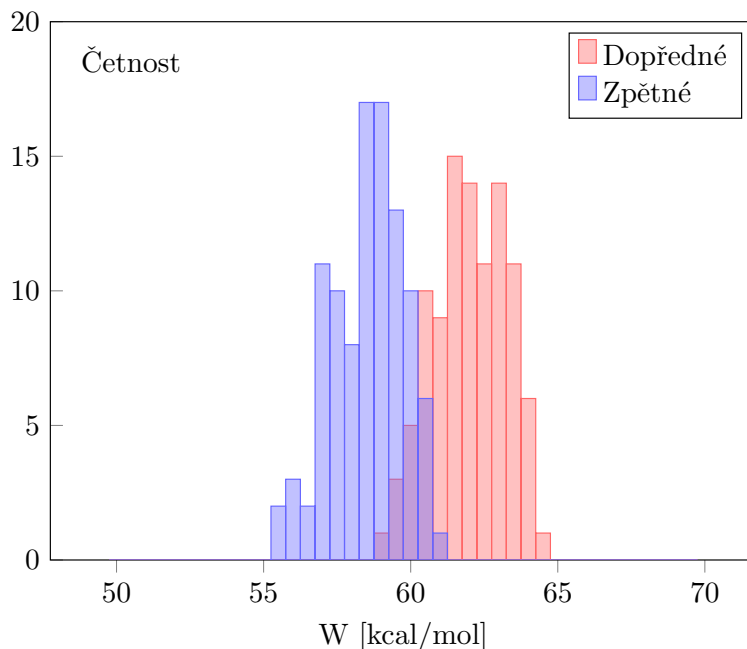
Hodnota poloměru r_0 omezujícího potenciálu $V(r)$ by měla být zvolena tak, aby vázaný ion byl schopen navzorkovat všechny možné vazebné konfigurace během MD simulace. Není vhodné volit příliš malou hodnotou poloměru, neboť pak je ion vázán v malém objemu a nemůže se dostat do vzdálenějších vazebných konformací. Pro příliš velké hodnoty poloměru ion naopak může zcela opustit vazebné místo, nevzorkuje pak již vazebné konfigurace a výpočet je neefektivní. Optimální hodnota poloměru se nachází v rozmezí těchto dvou extrémů.

V prvotních MD simulacích jsem otestoval nastavení simulačních parametrů spojených s alchymistickými transformacemi. Rozumných výsledků jsem dosáhl, když jsem spustil 100 dopředných a 100 zpětných alchymistických transformací, které byly rozděleny do 10 λ -oken a když transformace byla provedena během 200 ps. Výsledná volná energie byla určena podle Crooksova teorému.

Tabulka 4.5: Výsledné hodnoty volné energie pro vázaný ion pro různé poloměry r_0 omezujícího potenciálu. K hodnotám získaných analýzou 2x100 trajektorií ΔG_{Crooks} bylo potřeba započítat vliv omezujícího potenciálu ΔG_{res} na volnou energii. Poloměr je uveden v jednotkách Å a volné energie v kcal/mol.

r_0	ΔG_{Crooks}	ΔG_{res}	ΔG
0,5	61,8	-4,8	57,0
1,0	61,0	-3,6	57,4
2,0	59,8	-2,3	57,5
5,0	58,1	-0,7	57,4
7,4	56,9	0,0	56,9
25,0	56,4	2,2	58,6
50,0	56,5	3,4	59,9

Hodnotu poloměru r_0 jsem testoval v rozmezí od 0,5 Å do 50 Å. Korekce na omezující potenciál byla určena dle vzorce (3.27). Výsledky pro různé poloměry r_0 jsou shrnuty v tabulce 4.5 a na obrázku 4.11 je ukázka výsledných rozdělání práce pro dopředné a zpětné běhy v případě poloměru $r_0 = 2$ Å.



Obrázek 4.11: Výsledky ze 100 dopředných a 100 zpětných alchymistických transformací pro vázaný ion v případě poloměru vazby $r_0 = 2$ Å.

Druhý způsob výpočtu vazebné volné energie iontu spočíval v použití metody dvojité anihilace a v přímé aplikaci Jarzynského identity. V tomto případě bylo zapotřebí provést větší množství trajektorií, řádově 1000, pro získání dostatečně kvalitního výsledku.

Výhodou tohoto přístupu bylo, že nebylo nutné realizovat zpětné kreační transformace a nebylo zapotřebí definovat žádný omezující potenciál.

Nicméně ve výsledných hodnotách vazebné volné energie pozorujeme závislost na rychlosti provedení alchymistické transformace [105]. Problém spočívá v tom, že na konci velmi krátkých a rychlých nerovnovážných alchymistických transformací může být ligand nalezen pouze v omezeném objemu v okolí počáteční vazebné konfigurace. To je způsobeno tím, že ligand je na začátku alchymistické transformace vázaný a až ke konci této transformace je téměř zcela volný a může se difúzním pohybem dostat dále od původní polohy. Objem, ve kterém se ligand na konci alchymistické MD simulace nachází, je přitom závislý na rychlosti anihilační transformace. Je proto třeba k výsledkům přičíst korekci na standardní koncentraci ΔG_{vol} , jejíž hodnotu můžeme opět určit podle vzorce (3.27).

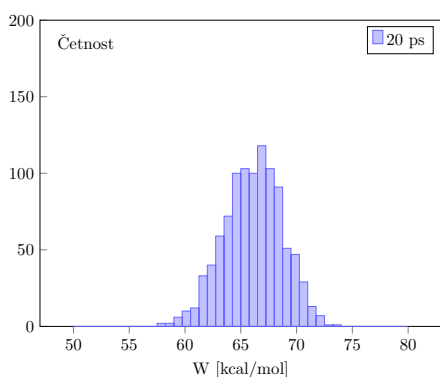
Způsob určení objemu $V(\tau)$ na němž by byl všeobecný konsenzus zatím neexistuje. Problém otevřela před téměř 10 lety skupina P. Procacciho [105] která od té doby s možnými přístupy různě experimentuje. V další sérii MD simulací jsem se proto pokusil určit nejvhodnější způsob určení tohoto objemu.

Je vhodné zdůraznit, že díky logaritmické závislosti 3.27 je i při 100% chybě určení objemu výsledná hodnota volné energie korekcí posunuta pouze o hodnotu $k_B T \ln(2) = 0,4$ kcal/mol.

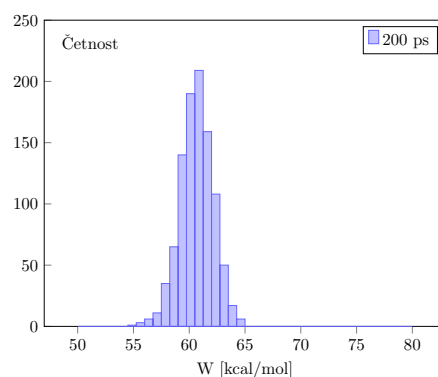
Otestoval jsem různě dlouhé trajektorie v rozmezí od 20 ps do 2 000 ps a pro každé nastavení jsem spustil 1000 alchymistických transformací. Výsledná rozdělení práce jsou zobrazena na obrázku 4.12 a výsledné hodnoty volných energií shrnuje tabulka 4.6.

Tabulka 4.6: Výsledné hodnoty volné energie pro vázaný ion počítané přímou aplikací Jarzynského identity pro různě dlouhé trajektorie. Délka trajektorie τ je udána v ps a hodnoty volných energií v kcal/mol. K výsledkům byla přičtena korekce ΔG_{vol} počítána podle vzorce (3.27).

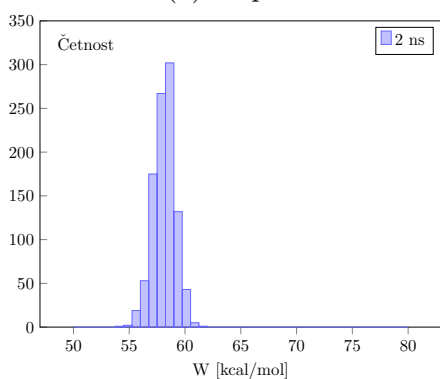
τ	$\Delta G_{\text{Jarzynski}}$	ΔG_{vol}	ΔG
20	60,4	-2,1	58,3
200	58,8	-1,9	56,9
2000	57,3	0,4	57,7
20000	56,4	1,5	57,9



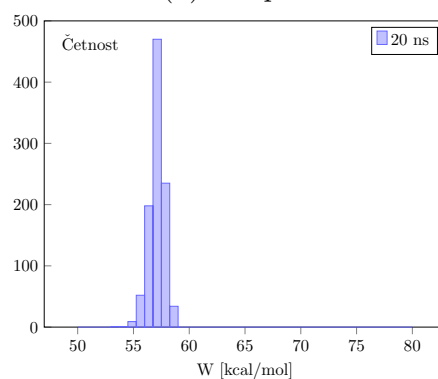
(a) 20 ps



(b) 200 ps



(c) 2 000 ps



(d) 20 000 ps

Obrázek 4.12: Výsledné rozdělení práce pro různě rychlé nerovnovážné alchymistické transformace.

Objem vystupující v rovnici (3.27) jsem nejprve zkusil určit jako objem koule s poloměrem rovným maximální vzdálenosti, do které se ion během simulace dostal. Výsledné hodnoty korekce ovšem s tímto odhadem nabývaly příliš velkých hodnot a výsledky nebyly konzistentní. Lepšího odhadu jsem dosáhl v případě,

kdy jsem uvažoval objem $V(\tau)$ rovný objemu koule s poloměrem rovným standardní odchylce vzdálenosti iontu od středu éterové koruny během MD simulace.

Výslednou vazebnou volnou energii draslíku a éterové koruny získám, když od předchozích výsledků odečtu volnou energii alchymistické transformace (anihilace) samotného draslíku ve vodě, jak popisuje termodynamický cyklus (3.15).

Pro volný ion není potřeba započítávat korekci 3.27 a výpočet je tak zcela přímočarý. Pro oba postupy (obousměrné vs. pouze anihilační transformace) jsem hodnotu volné energie solvatace iontu určil na 53,8 kcal/mol.

Je ovšem třeba zdůraznit, že MD simulace vázaného a volného iontu probíhaly nezávisle na sobě a i když počáteční velikosti jejich simulačních boxů byly nastaveny na přibližně stejnou velikost s délkou hrany 28 Å, tak v konečném výsledku byl díky vlivu barostatu simulační box komplexu o něco málo větší. To znamená, že v použitém termodynamickém cyklu se korekce na self-interakci náboje (3.26), která závisí na velikosti simulačního boxu, neodečte přesně a je potřeba ji započítat. Zde se negativně projevil problematický vliv malé počáteční velikosti boxu 28 Å, neboť pro tak malou délku hrany boxu vede i malý rozdíl na značnou hodnotu korekce (3.26). Celkový příspěvek self-interakce náboje k volné energii jsem určil jako rozdíl hodnot počítaných podle (3.26) pro délku hrany simulačního boxu s komplexem a délku hrany simulačního boxu s volným iontem.

Výsledné hodnoty vazebné volné energie pro oba použité postupy shrnuje tabulka 4.7, včetně započtení korekce (3.26). Pro výpočet jsem z výsledků 4.5 vybral případ s poloměrem vazby 2 Å a z výsledků 4.6 případ s délkou trajektorií 2 000 ps.

Tabulka 4.7: Výsledné hodnoty vazebné volné energie komplexu éterové koruny a iontu draslíku. Hodnoty byly počítány dvěma metodami: z obousměrných transformací pomocí Crooksova teorému a z jednosměrných anihilačních transformací přímou aplikací Jarzynského identity. K výsledným hodnotám byla navíc připočtena korekce na self-interakci náboje během alchymistické transformace. Všechny hodnoty jsou udané v kcal/mol.

ΔG_{Crooks}	$\Delta G_{\text{Jarzynski}}$	ΔG_{charge}	$\Delta G_{\text{Crooks,charge}}$	$\Delta G_{\text{Jarzynski,charge}}$	ΔG_{exp}
-3,7	-3,9	0,4	-3,3	-3,5	-2,9

4.2.3 Diskuze

V případě výsledků získaných prostřednictvím obousměrných transformací a s využitím Crooksova teorému můžeme z hodnot v tabulce 4.5 vidět, že výsledné hodnoty volné energie pro alchymistickou transformaci vázaného iontu jsou konzistentní v rozsahu poloměru vazeb 1,0 Å až 5,0 Å. Pro příliš omezující potenciál s poloměrem 0,5 Å pozorujeme malou odchylku v řádech 0,5 kcal/mol. Ion byl v tomto případě omezen na příliš malý objem a nebyl schopen navzorkovat veškeré vázané stavy. Také pro příliš velké poloměry od hodnoty 7,4 Å lze pozorovat

odchylku výsledných hodnot. Analýzou trajektorií pro tyto větší poloměry se ukázalo, že ion se nevrátil do vazebného místa ve všech zpětných simulacích a pro extrémně velké poloměry byl efektivně volný. Ion mimo vazebné místo nemohl navzorkovat vázané stavy, a proto pozorujeme nekonzistentní výsledky.

Pro druhý způsob výpočtu volné energie pomocí přímé aplikace Jarzynského identity je z výsledků znázorněných v grafu 4.12 patrné, že čím pomaleji provádíme alchymistické transformace, tak tím více se blížíme rovnovážným simulacím. V limitě nekonečně dlouhé trajektorie bychom čekali, že výsledná hodnota práce bude přímo rovna rovnovážné hodnotě změny volné energie. Naopak pro velmi rychlé alchymistické transformace je patrná značná variance výsledného rozdělení práce. Můžeme vidět, že hodnoty v tabulce 4.6 získané přímou aplikací Jarzynského identity jsou konzistentní s předchozími výsledky získanými pomocí Crooksova teoremu 4.5. Největší odchylku můžeme pozorovat pro kratší trajektorie a tudíž rychlejší transformace, což mohlo být způsobeno pomalou konvergencí při přímé aplikaci Jarzynského identity. Pro velmi rychlé transformace by tedy bylo vhodné napočítat ještě více než 1000 velmi krátkých trajektorií.

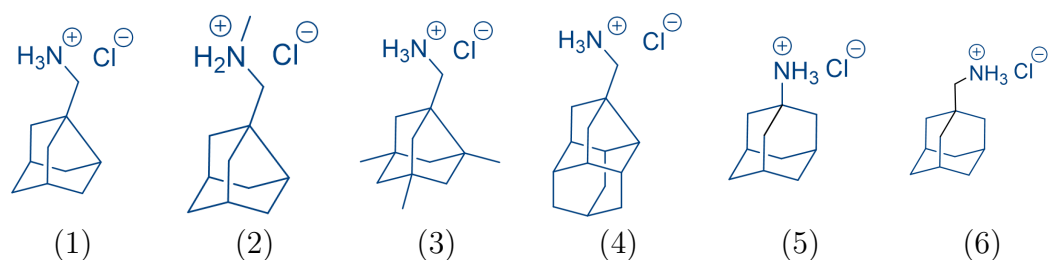
Výsledné hodnoty vazebné volné energie uvedené v tabulce 4.7 jasně ukazují, že jsem byl schopen určit vazebnou volnou energii studovaného komplexu pomocí obou použitých přístupů se srovnatelnou přesností. Je zde patrný rozdíl od experimentálních hodnot [104], což může být způsobeno mimo kvality silového pole například i tím, že vazebnou volnou energii počítám jako rozdíl dvou velkých hodnot dílčích volných energií (z alchymistických transformací provedených zvláště pro volný a vázaný iont) a přesnost výsledku tím může být ovlivněna.

4.3 Komplex adamantánů a cyklodextrinů

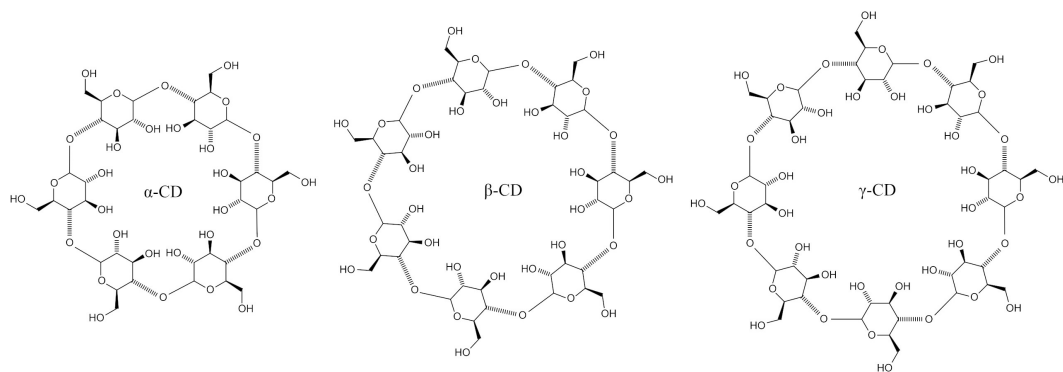
Další soubor molekul, kterými jsem se zabýval, tvořily komplexy adamantánů a cyklodextrinů. Jde o spolupráci s Přírodovědeckou fakultou Univerzity Karlovy konkrétně se skupinou pod vedením doc. RNDr. Jindřicha Jindřicha, CSc., která tyto látky syntetizuje a měří vazebné volné energie metodami izotermální kalorimetrie (ITC) a NMR titrace.

Deriváty adamantánů je možné využít mimo jiné jako inhibitory receptorů ACE2, čímž je možné zamezit pronikání koronaviru do lidských buněk. Cyklodextriny se používají jako nosiče léků apod.

Konkrétně jsem se zabýval strukturami uvedenými na obrázcích 4.13 a 4.14. Cyklodextriny jsou oligosacharidy složené z několika glukóz, které jsou spojené přes α -1,4 glykosidickou vazbu. Cyklodextriny v prostoru zaujímají toroidní tvar s větším otvorem na straně sekundárních hydroxylových skupin a s menším otvorem na opačné straně, kde jsou situované primární hydroxylové skupiny. Podle počtu glukóz v molekule cyklodextrinu rozlišujeme alfa-, beta- a gamma- cyklodextrin s 6, 7, resp. 8 jednotkami glukózy. Studované adamantány dále zjednodušeně označuji jako ligandy 1 až 6.



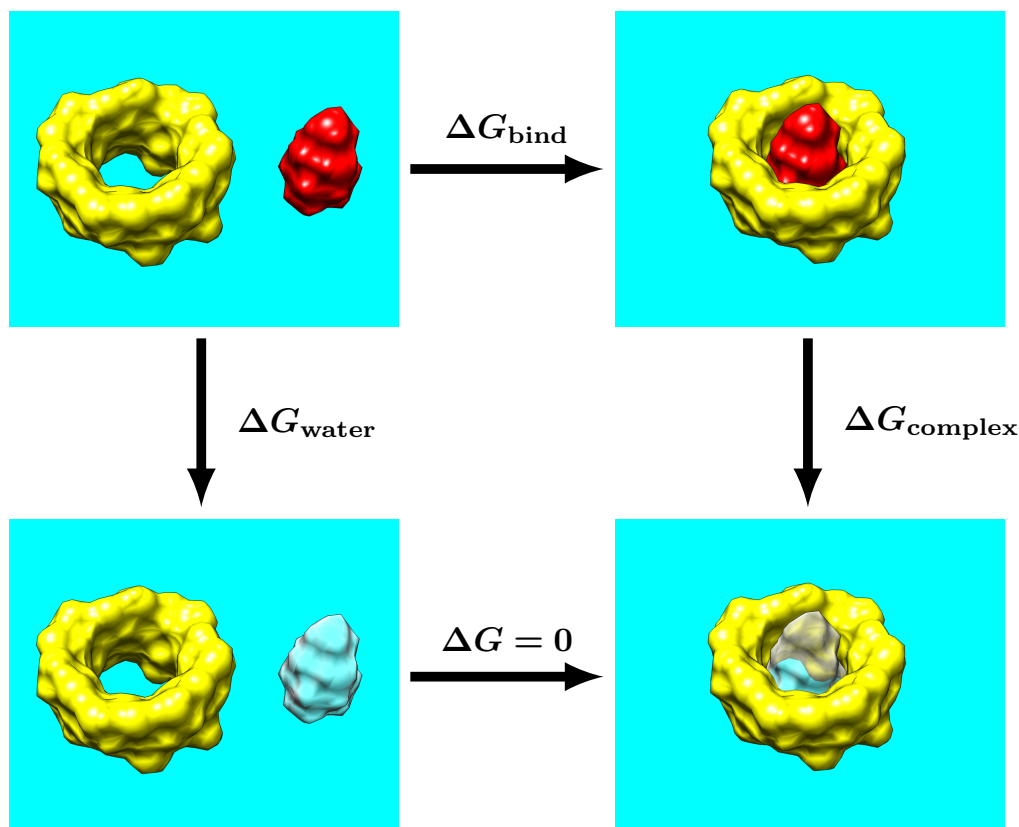
Obrázek 4.13: Struktury studovaných adamantánů a noradamantánů, které byly pro zjednodušení označeny jako ligandy 1 až 6.



Obrázek 4.14: Struktury studovaných cyklodextrinů.

Vazebné volné energie byly opět počítány pomocí termodynamického cyklu, který je zde specificky pro studované komplexy adamantánů s cyklodextrinů znázorněn na obrázku 4.15.

Jde o další jednoduchý modelový systém typu host:guest, který díky nižší výpočetní náročnosti umožňuje odladit metodiku výpočtu vazebných volných energií prostřednictvím nerovnovážných MD simulací. Atraktivní přitom je i to, že



Obrázek 4.15: Termodynamický cyklus využitý při výpočtech vazebné volné energie komplexů adamantánů a cyklodextrinů.

konformační plasticita cyklodextrinů do určité míry imituje obdobnou vlastnost aktivního místa koronavirové proteázy Mpro, a že struktura adamantánových ligandů je na rozdíl od triviálního draslíkového iontu použitého v předchozí kapitole velmi komplexní.

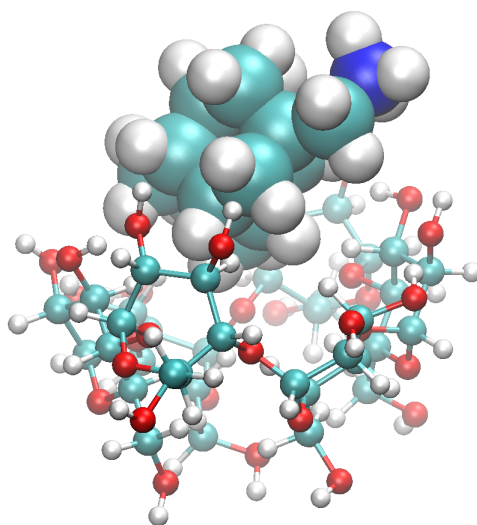
4.3.1 Příprava simulovaných systémů

Pro zkoumané komplexy jsem neměl k dispozici jejich vazebné konformace a mým prvním úkolem tedy bylo určit, jak přesně se budou v jednotlivých případech adamantány a cyklodextriny k sobě vázat. Pro tento účel jsem pro všechny možné kombinace šesti adamantánů a tří cyklodextrinů připravil výchozí konformace komplexů s různým uspořádáním. Testoval jsem obě možné orientace nabitě amino skupiny (do vodní obálky a do nitra cyklodextrinu) a zároveň jsem testoval 10 různých natočení adamantánu vůči cyklodextrinu s krokem v natočení 36° . Celkově jsem tak pro 18 kombinací 6 adamantánů a 3 cyklodextrinů, 2 možné orientace adamantánů a 10 jejich různých natočení připravil výchozí struktury pro celkem 360 rovnovážných MD simulací s různými počátečními podmínkami.

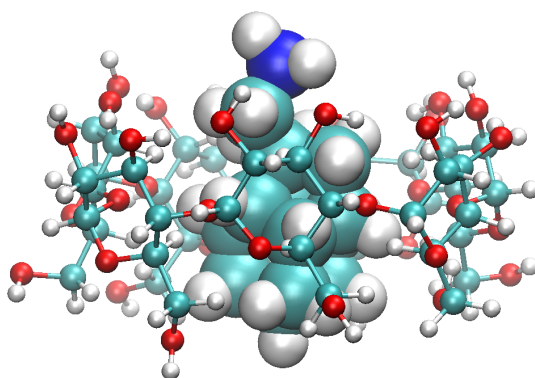
Pro všech 360 takto připravených komplexů bylo provedeno 2000 kroků minimalizace, aby došlo k odstranění případných příliš blízkých kontaktů atomů. Následně jsem spustil 1 ns ekvilibrační běhy. Nakonec jsem spustil 80 ns produkční běhy, ze kterých jsem analyzoval výsledky.

Opět jsem použil silové pole CGenFF.

Ze spuštěných rovnovážných MD simulací jsem se zaměřil na ty běhy, ve kterých adamantán zůstal pevně navázaný na cyklodextrin. Nakonec jsem došel k závěru, že konformace v komplexech budou obecně značně odlišné pro alfa-cyklodextrin oproti beta- a gamma- cyklodextrinům. Příklady jsou uvedeny na obrázcích 4.16 a 4.17. V případě alfa-cyklodextrinu došlo k přechodu do vaničkové konformace. Cyklodextrin je tak z jedné strany uzavřený a adamantány se váží spíše z opačné strany. Oproti tomu beta- a gamma- cyklodextriny zůstaly v počáteční otevřené konformaci a adamantány do nich mohly pronikat z obou stran.

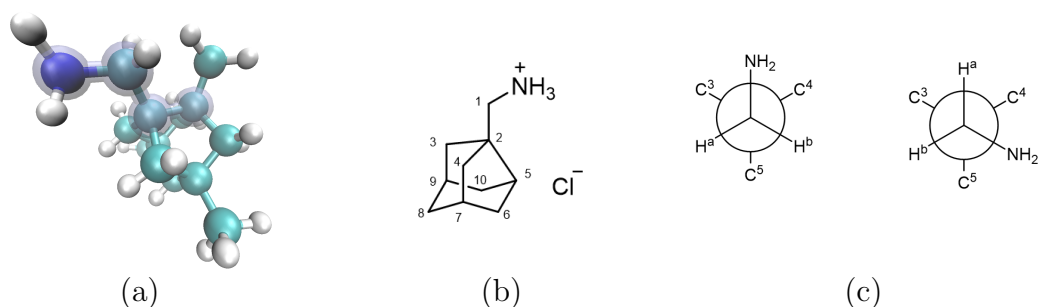


Obrázek 4.16: Vazebná konformace alfa-cyklodextrinu. Cyklodextrin je v tomto případě ve vaničkové konformaci, uzavřený na straně primárních hydroxylových skupin a otevřený na straně sekundárních hydroxylových skupin.



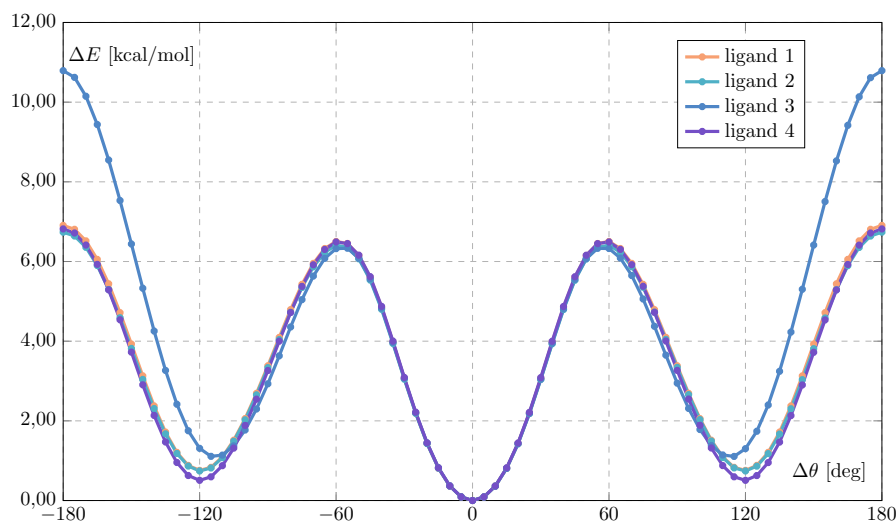
Obrázek 4.17: Vazebná konformace beta-cyklodextrinu. Cyklodextrin zůstal v tomto případě otevřený a adamantány jím mohly procházet. Pro gamma-cyklodextriny byla pozorována analogická konformace.

Z experimentálních dat se ukázal být důležitý dihedrální úhel znázorněný na obrázku 4.18, který určuje natočení nabitě aminoskupiny vůči zbytku noradamantánu. Ve výsledcích z nukleární magnetické rezonance lze rozlišit různá natočení této skupiny, neboť signál z methylenové skupiny mezi amino-skupinou a zbytkem adamantánu je v NMR spektru obvykle pozorován jako singlet. Zde se tento singletní signál v některých případech rozštěpil na dva dublety, nebo na kombinaci dvou dubletů a singletu. Výsledky experimentů lze vysvětlit rotací aminoskupiny o 120° , po které jsou vodíky v chemicky neekvivalentních polohách.



Obrázek 4.18: (a) Zkoumaný dihedrální úhel, pro který byla ověřena parametrizace CGenFF. (b) Použité číslování atomů. (c) Dvě možná natočení aminoskupiny způsobující rozdíly v NMR spektrech.

Adamantány byly popsány silovým polem CGenFF. Vzhledem k důležitosti uvedeného dihedrálního úhlu byly provedeny kontrolní kvantově-mechanické výpočty jako základ pro ověření parametrizace. Pro tuto kontrolu byl využit FFTk modul v programu VMD. Zaměřil jsem se přitom pouze na ligandy 1 až 4, protože ligandy 5 a 6 zkoumaný dihedrální úhel neobsahují.



Obrázek 4.19: Relaxovaný sken dihedrálního úhlu provedený na úrovni MP2 v bázi 6-31G(d) pomocí programu Gaussian.

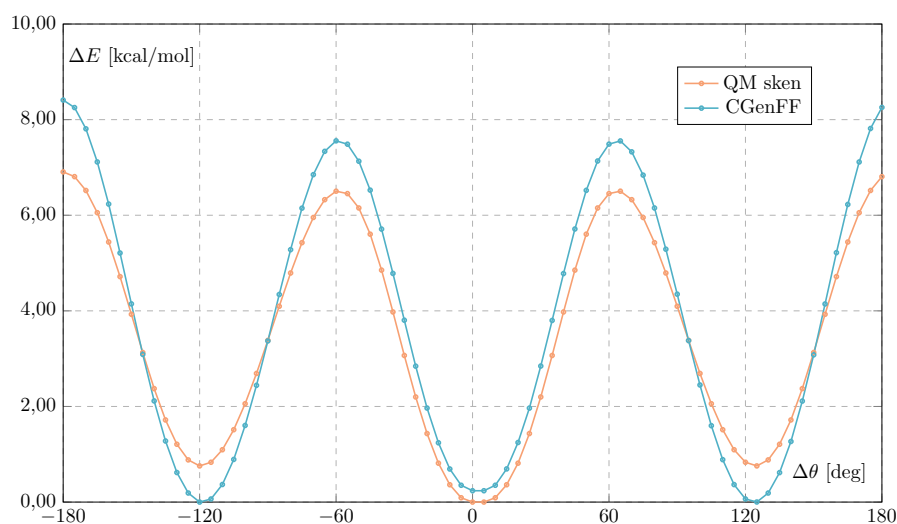
Studoval jsem struktury samotných ligandů ve vakuu. Jejich výchozí konformery jsem převzal z výsledků předchozích ekvilibračních MD simulací provedených pro komplexy. Pro všechny zkoumané ligandy jsem v programu Gaussian provedl kvantově-mechanické výpočty optimalizace geometrie na úrovni teorie

MP2 (Møller–Plessetova poruchová teorie v 2.řádu) s použitím báze 6-31G(d). Následně jsem provedl relaxovaný sken výše zmíněného dihedrálního úhlu na stejné úrovni teorie a se stejnou bazí. Sken byl proveden v celém rozmezí 360° s krokem 5°. Výsledky relaxovaných skenů shrnuje obrázek 4.19.

Kvantově-mechanické profily energie byly pomocí modulu FFTk porovnány s analogickými výpočty provedenými na molekulárně-mechanické úrovni. Příspěvek zkoumaného dihedrálního úhlu k silovému poli je v CGenFF popsán funkcí:

$$E_{\text{dih}} = A_1[1 + \cos(\theta)] + A_2[1 + \cos(2 \cdot \theta)] + A_3[1 + \cos(3 \cdot \theta)]. \quad (4.2)$$

Porovnání klasického molekulárně-mechanického a kvantově-mechanického popisu pro případ ligandu 1 ukazuje obrázek 4.20. Můžeme zde pozorovat, že parametry CGenFF kvalitativně dobře popisují celkový tvar potenciálu, ale je zde nezanedbatelný rozdíl ve výšce energetických bariér. Pozorovaný rozdíl je důsledkem toho, jakým způsobem byly odvozeny parametry v silovém poli CGenFF, které novým molekulám přiřadí parametry na základě analogie s dříve parametrizovanými molekulami [52]. Pro další výpočty jsem ale považoval původní parametrizaci CGenFF za dostačující a neprováděl jsme v tomto případě vlastní reparametrizaci.



Obrázek 4.20: Porovnání kvantově-mechanického skenu dihedrálního úhlu s výsledky získanými pomocí klasického molekulárně-mechanického popisu a silového pole CGenFF.

4.3.2 Výsledky simulací

Jakmile jsem měl k dispozici výchozí vazebné konformace komplexů a ověřenou parametrizaci zkoumaných ligandů, tak jsem mohl spustit výpočty vazebné volné energie. Nejdříve jsem zkoušel postupovat analogicky jako v případě výpočtu hydratačních volných energií aminokyselin, tj. prováděl jsem alchymistické transformace v obou směrech a využíval Crooksova flukтуаčního teorému. Zkusil jsem spustit 100 dopředných a 100 zpětných alchymistických transformací pro vázaný i pro volný adamantán. MD simulace byly rozděleny do 20 λ-oken s celkovou dobou jedné alchymistické transformace 400 ps.

Zpětné MD simulace byly opět problematické, nastal obdobný problém jako při výpočtech vazebné volné energie komplexu éterové koruny a draslíku. Ligandy se pouze s malou pravděpodobností vracely zpět do vazebné konfigurace. Pro omezení tohoto problému jsem zkoušel přidat do simulace omezující potenciál mezi hmotnými středy adamantánu a cyklodextrinu ve tvaru harmonických stěn 4.1. Tento omezující potenciál vyřešil problém se zpětnými transformacemi v případě éterové koruny. V tomto případě jsem ovšem pracoval se složitějšími ligandy, u kterých vazebné konformace adamantánů závisely i na jejich orientaci vůči cyklodextrinu a triviální omezující potenciál 4.1 se zde ukázal být nedostatečným. Ve zpětných MD simulacích se adamantány stále vracely do původní vazebné konfigurace pouze s malou pravděpodobností.

Možným řešením by bylo použít kombinaci více omezujících potenciálů, tak aby nebyla omezena pouze vzdálenost adamantanových ligandů od cyklodextrinu, ale aby byla omezena i jejich orientace a možná i konformace. V tomto případě by bylo nutné pro každý ligand provést další nezávislé simulace, které by určily příspěvek k vazebné volné energii od omezujícího potenciálu, jak popisuje termodynamický cyklus 3.2. MD simulace by se tak podstatně zkomplikovaly.

Místo toho jsem se zaměřil na výpočet vazebné volné energie metodou dvojité anihilace s přímou aplikací Jarzynského identity popř. její aproximace pomocí kumulantů. Alchymistické transformace jsem prováděl pouze dopředným směrem, kdy ligand postupně mizel ze simulovaného systému. MD simulace byla přitom rozdělena do 20 λ -oken a její celková délka byla 400 ps. Vycházel jsem přitom z výsledků ekvilibračních MD simulací, takže adamantány byly na počátku alchymistické transformace navázány k cyklodextrinům. Žádný dodatečný omezující potenciál jsem v tomto případě nedefinoval. Výsledné hodnoty pro vázané adamantány shrnuje tabulka 4.8. Pro volné adamantány ve vodním prostředí byly provedeny analogické anihilační transformace a výsledky shrnuje tabulka 4.9. Výsledné vazebné volné energie vypočítané z rozdílu výsledku obou dílčích sérií alchymistických transformací jsou uvedeny v tabulce 4.10. Jelikož MD simulace volných a vázaných ligandů probíhaly v různě velkých simulačních boxech a ligandy byly kladně nabitě, tak bylo nutné započítat opravu na self-interakci kladného náboje s jeho periodickými obrazy. Tato oprava byla vypočtena dle vzorce (3.26). Zároveň bylo nutné zahrnout korekci na standardní stav (3.27), protože zvolená délka nerovnovážných MD simulací efektivně limitovala pohyb adamantanových ligandů na omezený objem. Výsledné hodnoty jsem vyhodnotil jak pomocí Jarzynského identity (3.22), tak i pomocí její aproximace kumulanty (3.24).

Výsledné hodnoty počítané pomocí aproximace kumulanty jsou souhrnně zobrazeny v grafu 4.21 pro všechny možné kombinace adamantánů a cyklodextrinů. Graf 4.22 následně porovnává uvedené výsledky s experimentálními hodnotami izotermální kalorimetrie, které jsou zatím známé pouze pro ligandy 1 a 2, a graf 4.23 porovnává výsledky simulací s experimentálními daty NMR, které byly získány přepočtem z rovnovážných vazebných konstant určených z výsledků NMR a zatím byly proměřeny pro ligandy 1 až 4.

Tabulka 4.8: Výsledky výpočtů volné energie pro vázané adamantány. Výsledky byly počítány pomocí Jarzynského identity i pomocí rozvoje kumulantů. K výsledkům byla navíc přičtena korekce na standardní stav a pro self-interakci náboje v simulovaném systému s periodickými okrajovými podmínkami. Veškeré hodnoty jsou uvedeny v kcal/mol.

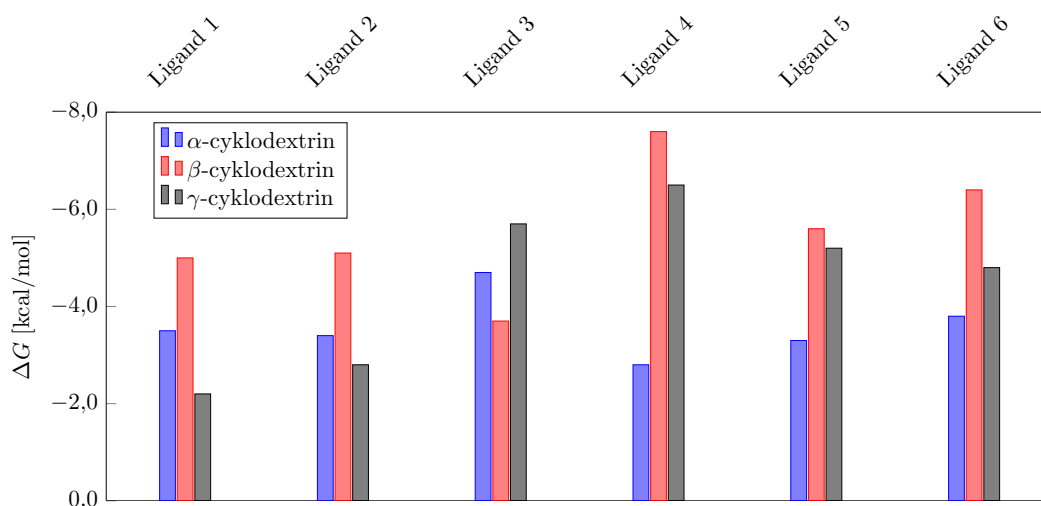
Cyklodextrin	Ligand	$\Delta G_{\text{Jarzynski}}$	$\Delta G_{\text{kumulant}}$	ΔG_{vol}	ΔG_{charge}
α -cyklodextrin	1	54,5	54,6	0,0	12,0
	2	12,8	15,3	-0,3	11,9
	3	102,9	104,1	0,0	11,9
	4	57,5	58,4	-0,9	11,9
	5	117,0	118,0	-0,1	11,5
	6	81,4	81,8	-0,2	11,5
β -cyklodextrin	1	56,1	56,4	-0,2	11,5
	2	15,7	17,7	0,2	11,5
	3	102,8	103,9	0,4	11,6
	4	63,0	64,3	0,0	11,5
	5	120,3	120,3	0,0	11,6
	6	84,8	84,6	0,0	11,5
γ -cyklodextrin	1	53,1	53,0	-1,0	11,3
	2	14,2	14,9	-0,5	11,3
	3	106,1	105,6	-0,1	11,4
	4	62,5	62,9	-0,6	11,3
	5	119,3	119,3	-0,8	11,4
	6	82,5	82,3	-0,9	11,4

Tabulka 4.9: Výsledky výpočtů volné energie pro volné adamantány ve vodním prostředí. Výsledky byly počítány pomocí Jarzynského identity i pomocí rozvoje kumulantů. K výsledkům byla navíc přičtena korekce self-interakce náboje. Veškeré hodnoty jsou uvedeny v kcal/mol.

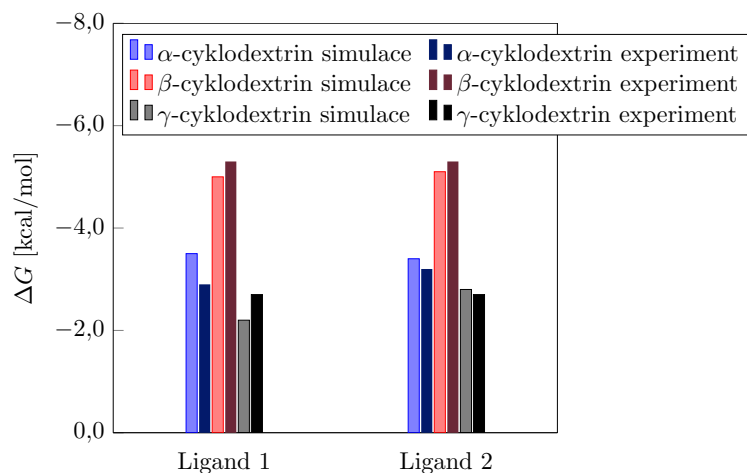
Ligand	$\Delta G_{\text{Jarzynski}}$	$\Delta G_{\text{kumulant}}$	ΔG_{charge}
1	49,1	49,1	14,0
2	10,0	10,1	13,9
3	97,7	97,7	13,6
4	54,4	54,7	13,6
5	110,7	112,3	14,0
6	75,1	75,8	13,9

Tabulka 4.10: Vazebné volné energie komplexů adamantánů s cyklodextriny. Pro srovnání byly hodnoty určeny jak pomocí Jarzynského identity, tak pomocí aproximace kumulanty. Veškeré hodnoty jsou uvedeny v kcal/mol.

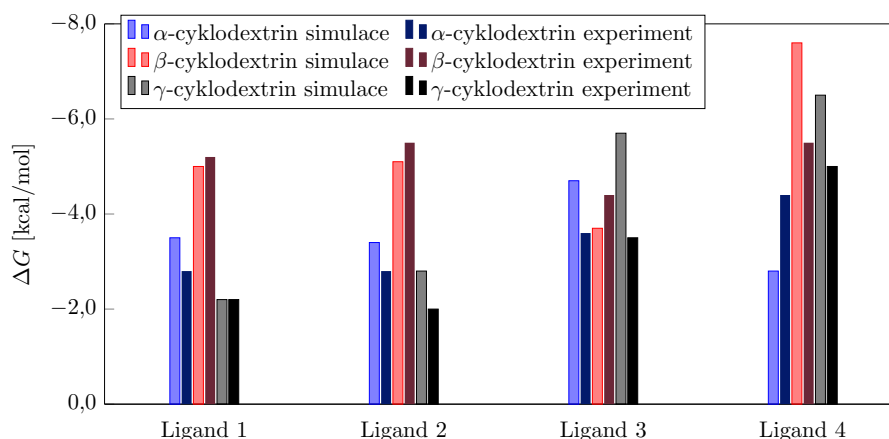
Cyklodextrin	Ligand	$\Delta G_{\text{Jarzynski}}$	$\Delta G_{\text{kumulant}}$	$\Delta G_{\text{exp,ITC}}$	$\Delta G_{\text{exp,NMR}}$
α -cyklodextrin	1	-3,3	-3,5	-2,9	-2,8
	2	-1,1	-3,4	-3,2	-2,8
	3	-3,4	-4,7	-	-3,6
	4	-2,3	-2,8	-	-4,4
	5	-3,9	-3,3	-	-
	6	-4,1	-3,8	-	-
β -cyklodextrin	1	-4,7	-5,0	-5,3	-5,2
	2	-3,2	-5,1	-5,3	-5,5
	3	-2,6	-3,7	-	-4,4
	4	-6,4	-7,4	-	-5,5
	5	-7,2	-5,6	-	-
	6	-7,3	-6,4	-	-
γ -cyklodextrin	1	-2,3	-2,2	-2,7	-2,2
	2	-2,2	-2,8	-2,7	-2,0
	3	-6,3	-5,7	-	-3,5
	4	-6,3	-6,5	-	-5,0
	5	-6,7	-5,2	-	-
	6	-5,7	-4,8	-	-



Obrázek 4.21: Výsledné vazebné volné energie počítané pomocí aproximace kumulanty pro všechny studované kombinace adamantánů a cyklodextrinů.



Obrázek 4.22: Srovnání vazebných volných energií počítaných pomocí aproximace kumulanty s experimentálními výsledky ITC.



Obrázek 4.23: Srovnání vazebných volných energií počítaných pomocí aproximace kumulanty s experimentálními výsledky NMR.

4.3.3 Diskuze

Z výsledných hodnot uvedených v tabulce 4.10 je patrný značný rozdíl mezi hodnotami počítanými přímo pomocí Jarzynského identity a hodnotami počítanými pomocí její aproximace kumulanty. Jelikož bylo vždy spuštěno pouze 100 trajektorií, tak tento rozdíl byl pravděpodobně způsoben především pomalou konvergencí výsledků získaných přímou aplikací Jarzynského identity, kdy nejvíce přispívají málo pravděpodobné trajektorie s malými hodnotami práce. Pro zmenšení rozdílu mezi výsledky získanými prostřednictvím Jarzynského identity a prostřednictvím kumulantů by bylo zapotřebí spustit větší množství MD simulací.

Výsledky počítané prostřednictvím kumulantů (3.24) nejsou zatíženy závislostí na trajektoriích s malou hodnotou práce, neboť pro jejich určení je dostatečující určit střední hodnotu a varianci pravděpodobnostního rozdělení hodnot

práce. Z tohoto důvodu dále komentuji pouze výsledky počítané prostřednictvím kumulantů. Hodnoty v grafu 4.21 jasně ukazují, že většina adamantánů se nejlépe váže k β -cyklodextrinu. Jedinou výjimku zde tvoří ligand 3, který jako jediný měl postranní methylové skupiny a nejlépe se váže ke γ -cyklodextrinu. Při porovnání dostupných experimentálních dat ITC pro ligandy 1 a 2 4.22 s výsledky MD simulací je zřejmá dobrá shoda. Podobně dobrou shodu můžeme pozorovat pro výsledky NMR 4.23, kde je ovšem patrný větší rozdíl mezi výsledky simulací a experimentem.

4.4 Nekovalentní inhibitory proteázy Mpro

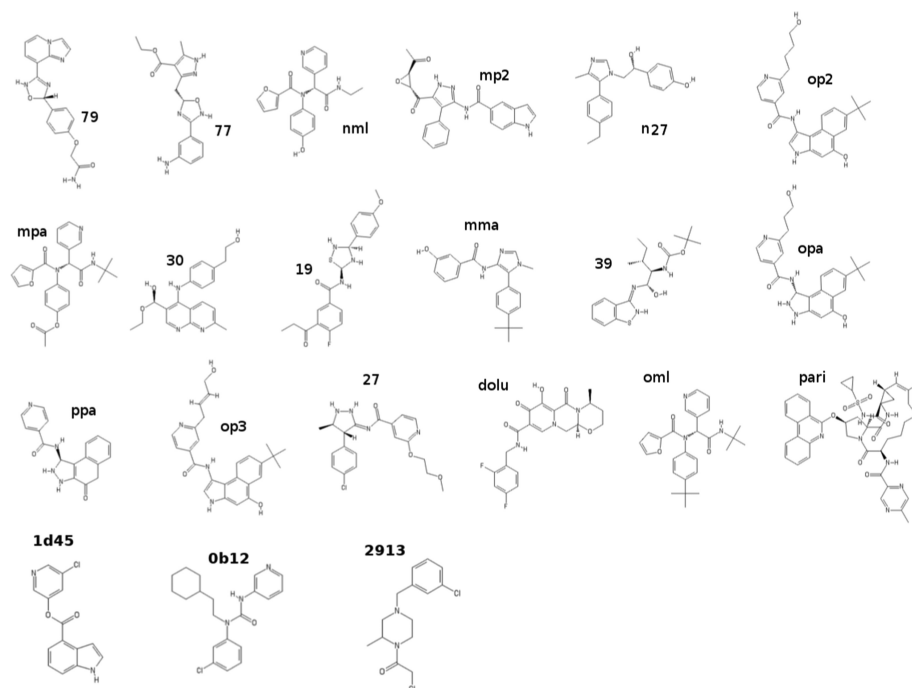
V další sérii MD simulací jsem se zaměřil na výpočet vazebné volné energie pro ligandy, které byly v předchozích studiích identifikovány jako potenciální inhibitory koronavirové proteázy Mpro [42]. Potenciální inhibitory byly identifikovány na základě empirických metod, pomocí nichž je možné odhadnout vazebné volné energie ligandů za velmi zjednodušených předpokladů. Jednou z těchto metod je molekulární docking, prostřednictvím něhož je možné určit vazebné volné energie na základě tzv. empirických skórovacích funkcí. Výpočty prováděné pomocí molekulárního dockingu jsou nenáročné na hardwarové prostředky a je možné je provádět extrémně rychle. Odhady vazebné volné energie z dockingu jsou ovšem pouze velmi přibližné. O tom svědčí i fakt, že dvě kampaně popsané v literatuře, během nichž byla prostřednictvím dockingu otestována vždy více než miliarda ligandů, nevyústily v získání použitelného inhibitoru koronavirové proteázy Mpro [35, 36].

Mým úkolem proto bylo, pokusit se určit přesnější hodnoty vazebné volné energie ligandů na základě fyzikálně korektních alchymistických transformací provedených prostřednictvím nerovnovážných MD simulací.

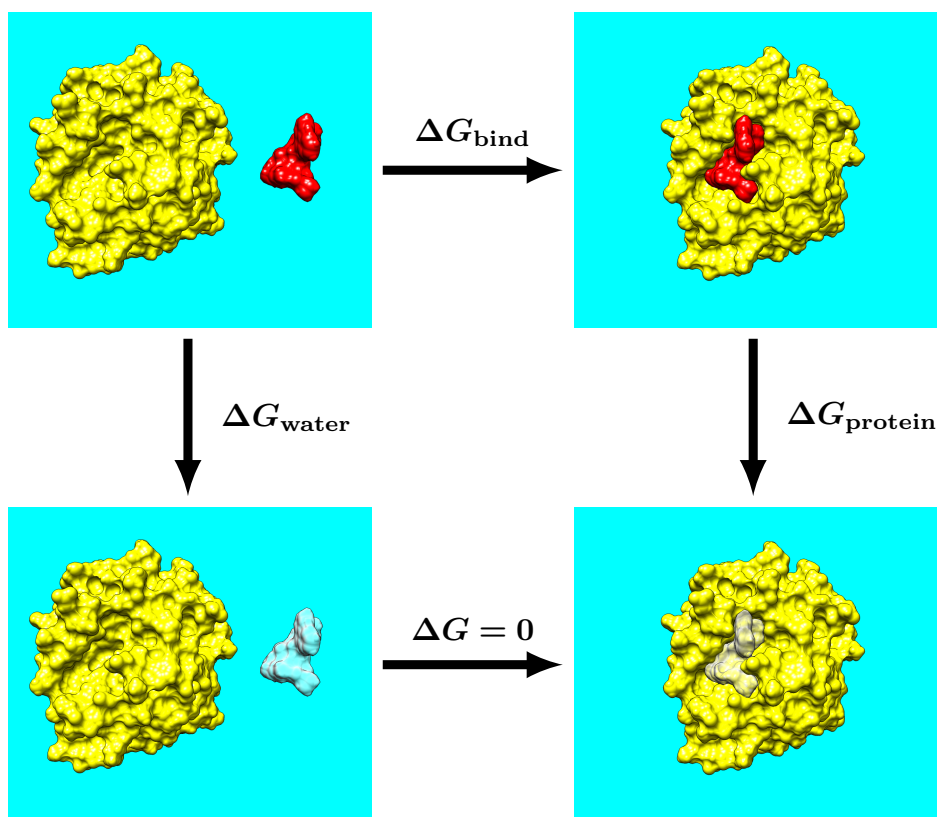
Ve struktuře studované proteázy Mpro, která je aktivní ve formě dimeru s monomery orientovanými na sebe přibližně kolmo, můžeme pozorovat, že aktivní proteolytická místa jsou poměrně mělká a vystavená tak vodnímu prostředí. Aktivní místa obou monomerů jsou přitom od sebe značně vzdálená. Pro výpočty vazebné volné energie je proto možné omezit se na studium jednoho monomeru resp. pouze jeho domén I a II, mezi kterými se aktivní místo nachází.

Struktury zkoumaných ligandů jsou znázorněny na obrázku 4.24. Z uvedených ligandů jsou pouze látky dolu, pari a oml komerčně dostupné a pouze pro 1d45, 0b12 a 2913 jsou známy experimentální hodnoty vazebné volné energie vůči koronavirové proteáze Mpro [42]. V literatuře jsou ovšem dostupné referenční hodnoty vazebných volných energií všech ligandů z obrázku 4.24 získané prostřednictvím nerovnovážných MD simulací, které ovšem byly provedeny prostřednictvím velmi komplikovaného simulačního protokolu NEW-vDSSB (angl. new virtual double-system single-box) [42], který se v mnoha ohledech liší od velmi jednoduchého simulačního protokolu, který jsem zde vytvořil a optimalizoval při studiu jednoduchých modelových systémů v předchozích podkapitolách této práce. Aplikace mého postupu na ligandy z obrázku 4.24 a proteázu Mpro a srovnání výsledků s referenčními daty z literatury [42] tedy může napomoci posoudit, nakolik je můj simulační protokol využitelný i při studiu komplexních biomolekulárních systémů.

Obrázek 4.25 znázorňuje termodynamický cyklus využitý při výpočtech vazebné volné energie v této podkapitole. Opět se jedná o klasickou dvojistou anihilaci ligandů, kde se provádí anihilace jednak z vodní obálky a jednak z aktivního místa koronavirové proteázy Mpro.



Obrázek 4.24: Struktury studovaných ligandů převzaté z předchozích studií, kde byly vyselektovány na základě výsledků molekulárního dockingu [42].



Obrázek 4.25: Termodynamický cyklus využitý při výpočtech vazebné volné energie nekovalentních inhibitorů.

4.4.1 Příprava simulovaných systémů

Počáteční souřadnice atomů ligandů navázaných na proteázu Mpro byly převzaty z předchozích studií, kde byly vyselektovány prostřednictvím molekulárního dockingu [42]. Pro všechny ligandy byly vytvořeny dva nezávislé simulované systémy. Jednak pro ligand vázaný na proteázu Mpro a jednak pro samotný ligand. V obou případech byly zkoumané systémy obklopeny molekulami vody, které byly popsány modelem TIP3P. Simulační box se samotnými ligandy ve vodním prostředí měl hrany dlouhé přibližně 30 Å. Simulační box s ligandy navázanými k proteáze Mpro přibližně 75 Å. Do simulačních boxů byly navíc přidány ionty chloru a sodíku kvůli neutralizaci a také pro doplnění koncentrace iontů na standardní hodnotu 0,15 mol/L.

Pro všechny připravené simulované systémy byla nejdříve provedena minimalizace s 2000 kroky. Následně byla provedena tzv. termalizace (angl. annealing), což je MD simulace, která simulovaný systém postupně zahřeje a přivede do termodynamické rovnováhy bez narušení struktury proteinu či komplexu. Při této MD simulaci byla stabilizována konformace hlavního vlákna či páteře proteinu (angl. main chain či backbone) pomocí harmonických omezení tak, aby setrvala v původní konformaci. Teplota byla postupně zvětšována z počáteční hodnoty 60 K až na cílovou hodnotu 300 K s krokem 1 K. Pro každou hodnotu teploty byla přitom provedena krátká dílčí simulace o délce 1,2 ps. Na tuto termalizační MD simulaci jsem navázal ekvilibračním během při teplotě 300 K a s celkovou dobou trvání 1 ns. Harmonická omezení přitom stále udržovala původní konformaci páteře proteázy Mpro. Závěrečná evilibrace už ale proběhla bez těchto omezení.

4.4.2 Výsledky simulací

Pro následující alchymistické transformace všech ligandů jsem nejprve otestoval možné alternativy v simulačním protokolu a různá nastavení simulačních parametrů na ligandu 0b12, pro který je známa experimentální hodnota vazebné volné energie vůči proteáze Mpro.

První z testovaných přístupů k výpočtu vazebné volné energie byl založen na využití obousměrných alchymistických transformací a aplikaci Crooksova teorému (3.25). To je způsob, jakým probíhaly MD simulace pro hydratační volné energie aminokyselin v úvodní podkapitole výsledkové části práce. V tomto případě jsem provedl dopředné anihilační alchymistické transformace, ve kterých se ligand ze simulovaného systému postupně odpoutával, a posléze zpětné kreační simulace, ve které se ligand v simulovaném systému naopak postupně objevoval. Série dopředných i zpětných běhů byly spuštěny pro volný i vázaný ligand.

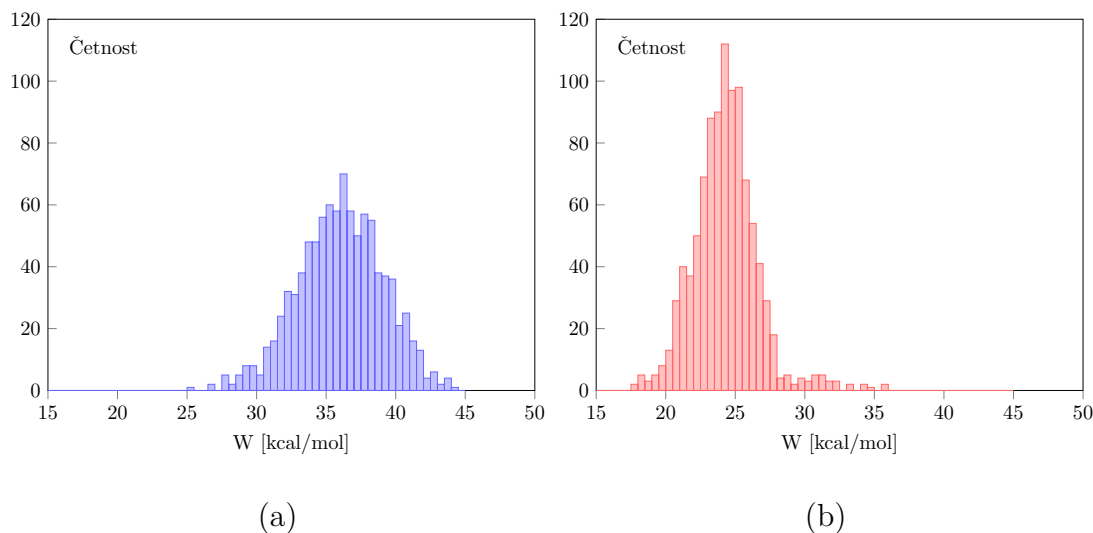
Pro volný ligand ve vodním prostředí bylo možné aplikovat Crooksův teorém bez komplikací, neboť pravděpodobnostní rozdělení práce z dopředných a zpětných simulací měla dostatečný překryv. Nicméně pro ligandy vázané na proteázu nastal problém u zpětných kreačních alchymistických transformací. Ligand se pouze s malou pravděpodobností vracel do vazebné konformace a výsledná rozdělení práce pro zpětné simulace tak byla značně rozšířena, měla velkou varianci a nebylo možné jednoznačně určit překryv s rozdělením práce z dopředných MD simulací. Nebylo tak možné aplikovat Crooksův teorém. Tento problém nastal i u obou jednoduchých modelových systémů popsanych v předchozích podkapitolách tj. v případě komplexu draslíku a éterové koruny a také u komplexů adamantánů

s cyklodextriny.

Zkusil jsem tedy opět zavést do MD simulace omezující potenciál ve tvaru (4.1), který působil mezi atomy ligandu a několika atomy aktivního místa proteázy Mpro. Snažil jsem se tak dosáhnout toho, aby se ligand během alchymistických transformací držel v blízkosti vazebného místa v proteáze Mpro. Problém se zpětnými kreačními alchymistickými transformacemi ovšem přetrval. Možným řešením by bylo zavést více složitějších omezujících potenciálů, které by kromě vzdálenosti ligandu od aktivního místa proteázy Mpro kontrolovaly i orientaci a konformaci ligandu. Pak by ovšem bylo zapotřebí určit příspěvek těchto potenciálů k vazebné volné energii v dalších nezávislých MD simulacích podle termodynamického cyklu (3.15). Výpočet by se tak značně zkomplikoval, což v případě prosévání většího počtu ligandů během hledání optimálního inhibitoru, není žádoucí.

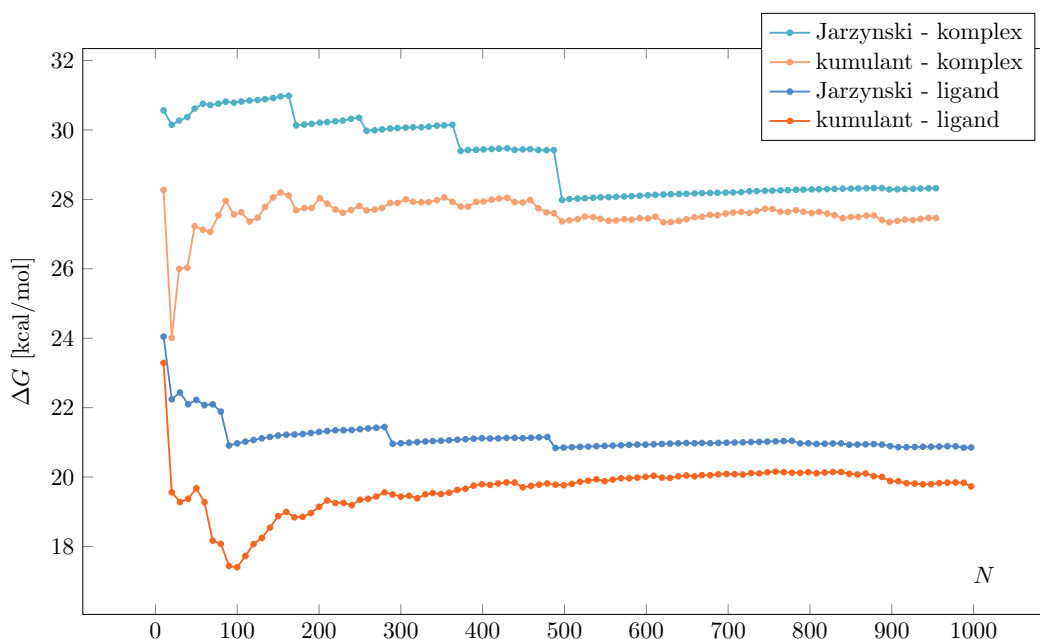
Proto jsem se opět zaměřil na výpočet vazebné volné energie prostřednictvím prosté dvojité anihilace a přímé aplikaci Jarzynského identity (3.22). To znamená, že jsem prováděl pouze dopředné anihilační alchymistické transformace, při kterých ligandy postupně mizely ze simulovaných systémů. Opět jsem tedy prováděl výpočty pro vázaný i volný ligand podle termodynamického cyklu (3.1), přitom jsem nedefinoval žádné dodatečné omezující potenciály.

Na ligandu 0b12 jsem testoval různá nastavení anihilační alchymistické transformace, především pokud jde o dělení transformace na různý počet λ -oken a celkovou dobu transformace. Příklad výsledků pro jedno takové nastavení parametrů je uveden na obrázku 4.26, kde uvádím výsledky ze série 1000 anihilačních alchymistických transformací ligandu 0b12 jednak z komplexu s proteázou Mpro a jednak z vodní obálky. Nakonec jsem na základě mnoha testů provedených s ligandem 0b12 zvolil pro výpočty s ostatními ligandy jako optimální rozdělení anihilační transformace do 20 λ -oken s krokem $\Delta\lambda = 0,05$ a s celkovou dobou během níž byla realizována jedna alchymistická transformace 400 ps.



Obrázek 4.26: Příklad výsledků z alchymistických transformací pro (a) ligand 0b12 vázaný na proteázu Mpro, (b) pro volný ligand 0b12.

Dalším důležitým parametrem u nerovnovážných MD simulací je celkový počet trajektorií potřebných k získání věrohodného výsledku. Za tímto účelem jsem provedl analýzu 1000 trajektorií naakumulovaných pro ligand 0b12, kdy jsem postupně do vyhodnocovaného souboru přidával výsledky po 10 trajektoriích. Volnou energii jsem zkoušel počítat přímo pomocí Jarzynského identity (3.22) i pomocí její aproximace kumulanty (3.24). Analýzu jsem provedl pro MD simulace vázaného i volného ligandu. Výsledky jsou znázorněny v grafu 4.27. Je zřejmé, že k získání věrohodných výsledků je optimální naakumulovat 1000 trajektorií.



Obrázek 4.27: Analýza konvergence výsledků prostřednictvím 1000 trajektorií získaných pro ligand 0b12. Výsledky byly počítány opakovaně pro postupně rostoucí počet trajektorií. Analýza byla provedena zvláště pro ligand v komplexu s proteázou a zvláště pro volný ligand ve vodním prostředí. Pro srovnání byl využit jak přímý výpočet volné energie pomocí Jarzynského identity, tak její aproximace pomocí kumulantů.

Následně jsem provedl anihilační alchymistické transformace pro všechny ostatní zkoumané ligandy. Použil jsem parametry odladěné při testování výpočtech s ligandem 0b12. Alchymistické transformace byly tedy rozděleny do 20 λ -oken. Celková doba jedné transformace přitom byla 400 ps. Pro každý ligand jsem prováděl série 1000 transformací. MD simulace byly provedeny jak pro ligandy vázané v komplexu, tak pro volné ligandy ve vodním prostředí.

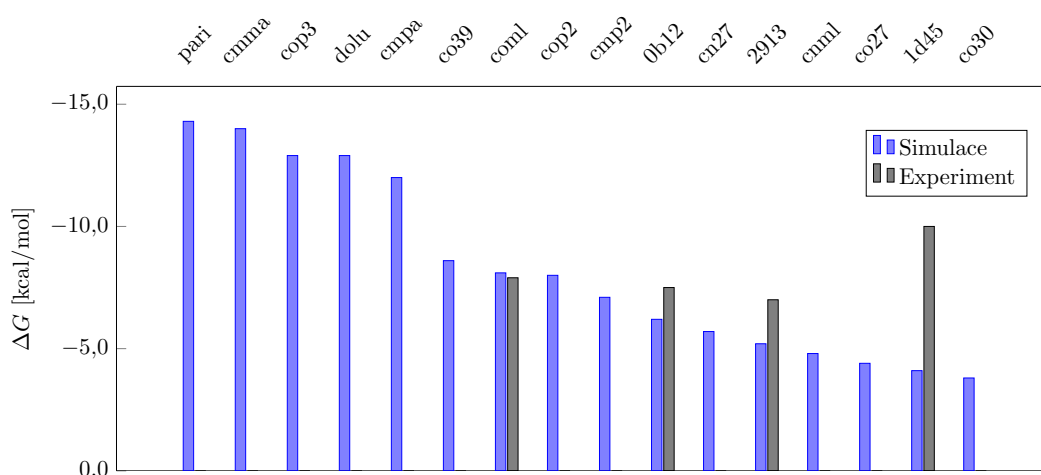
Většina ligandů byla neutrálních s výjimkou ligandu 2913, pro který bylo potřeba započít korekci volné energie na self-interakci náboje během alchymistické transformace počítané podle vzorce (3.26). Výslednou hodnotu této korekce jsem určil z rozdílu velikosti simulačních boxů pro volný a vázaný ligand na hodnotu $-11,6$ kcal/mol.

Pro všechny ligandy bylo potřeba započít korekci na standardní stav počítanou podle vzorce (3.27), neboť omezená délka trajektorií získaných z krátkých nerovnovážných MD simulací efektivně omezuje pohyb ligandu na určitý objem, i když v MD simulacích není definovaný žádný dodatečný omezující potenciál.

Výsledné hodnoty změny volné energie při anihilační transformaci pro vázané ligandy shrnuje tabulka 4.11, pro volné pak tabulka 4.12. Z rozdílu těchto hodnot a se započtením korekcí byla určena výsledná vazebná volná energie, jejíž hodnoty shrnuje tabulka 4.13.

Tabulka 4.11: Změna volné energie během anihilační alchymistické transformace pro ligandy v komplexu s proteázou Mpro. Veškeré hodnoty jsou uvedeny v kcal/mol.

Ligand	$\Delta G_{\text{kumulant}}$	$\Delta G_{\text{Jarzynski}}$	ΔG_{vol}
pari	35,2	39,7	-0,4
mma	28,1	31,6	0,2
cop3	36,8	39,2	-0,1
dolu	25,0	27,5	0,2
mpa	27,2	27,1	-0,2
co39	24,0	25,0	-0,2
coml	20,6	20,1	-0,2
cop2	33,1	34,3	0,3
mp2	37,0	39,0	0,4
ob12	21,0	21,9	0,0
cn27	23,0	23,8	0,0
2913	54,5	54,1	0,3
cnml	19,4	20,7	1,4
co27	21,7	22,8	0,5
ld45	15,1	15,3	0,8
co30	23,2	25,5	0,8



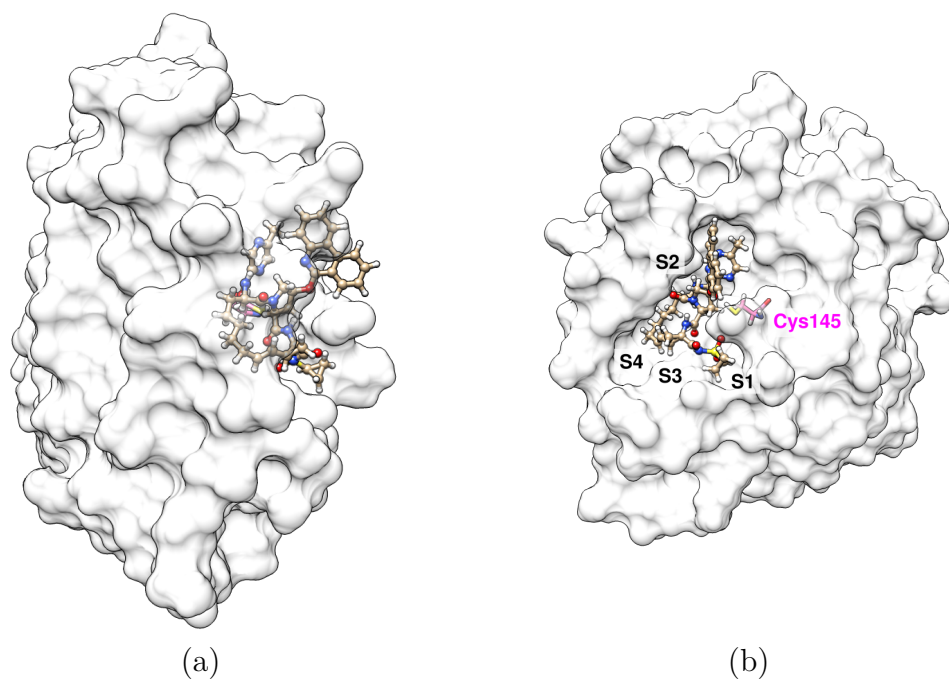
Obrázek 4.28: Výsledné vazebné volné energie počítané pomocí přímé aplikace Jarzynského identity pro všechny studované inhibitory proteázy Mpro

Hodnoty počítané přímo pomocí Jarzynského identity jsou graficky znázorněny na obrázku 4.28. Z těchto výsledků lze určit ligand pari jako nejlepší inhi-

bitor a ligand co30 jako ten nejhorší. Vazba ligandu pari na proteázu Mpro je znázorněna na obrázku 4.29.

Tabulka 4.12: Změna volné energie během anihilační alchymistické transformace pro volné ligandy ve vodním prostředí. Veškeré hodnoty jsou uvedeny v kcal/mol.

Ligand	$\Delta G_{\text{kumulant}}$	$\Delta G_{\text{Jarzynski}}$
pari	24,7	24,9
mma	15,9	17,7
cop3	25,1	26,2
dolu	14,3	14,8
mpa	14,4	14,9
co39	15,5	16,2
coml	10,8	11,8
cop2	25,5	26,7
mp2	30,6	32,3
0b12	15,0	15,1
cn27	18,0	18,0
2913	37,0	37,6
cnml	16,0	17,3
co27	18,7	18,9
1d45	11,0	12,0
co30	21,2	22,5



Obrázek 4.29: Vazba zkoumaného ligandu pari na proteázu Mpro ze dvou různých úhlů pohledu.

Tabulka 4.13: Výsledné vazebné volné energie zkoumaných inhibitorů proteázy Mpro. Veškeré hodnoty jsou uvedeny v kcal/mol.

Ligand	$\Delta G_{\text{kumulant}}$	$\Delta G_{\text{Jarzynski}}$	ΔG_{ref}	ΔG_{exp}
pari	-10,2	-14,3	-10,7	-
mma	-12,5	-14,0	-7,4	-
cop3	-11,6	-12,9	-8,8	-
dolu	-10,9	-12,9	-6,7	-
mpa	-12,6	-12,0	-4,3	-
co39	-8,3	-8,6	-7,6	-
coml	-9,6	-8,1	-9,0	-7,9
cop2	-7,9	-8,0	-7,1	-
mp2	-6,8	-7,1	-5,8	-
0b12	-6,0	-6,2	-9,3	-7,5
cn27	-5,0	-5,7	-3,9	-
2913	-6,2	-5,2	-5,8	-7,0
cnml	-4,8	-4,8	-3,1	-
co27	-3,5	-4,4	-8,3	-
1d45	-4,9	-4,1	-5,4	-10,0
co30	-2,6	-3,8	-6,1	-

4.4.3 Diskuze

Testování vývoje hodnot volné energie v závislosti na celkovém počtu trajektorií uvedené v grafu 4.27, bylo provedené zvlášť pro volný a vázaný ligand. Vazebná volná energie by byla určena rozdílem těchto dvou hodnot. Ve vývoji volné energie v závislosti na počtu trajektorií počítané podle Jarzynského identity můžeme pozorovat výrazné skoky, které byly způsobeny tím, že k výsledkům počítaným podle Jarzynského identity přispívají nejvíce málo pravděpodobné trajektorie s malou hodnotou práce. Oproti tomu ve výsledcích počítaných pomocí aproximace kumulanty žádné skoky nepozorujeme, neboť pro jejich určení stačí znát pouze střední hodnotu a varianci pravděpodobnostního rozdělení práce a výsledky tak nejsou zatíženy závislostí na málo pravděpodobných trajektoriích s malou hodnotou práce. V konečném výsledku tak hodnoty volné energie počítané pomocí aproximace kumulatnů konvergují rychleji s počtem trajektorií. Nevýhodou výpočtu pomocí kumulantů je, že tato aproximace platí přesně, pouze pokud pravděpodobnostní rozdělení práce jsou gaussovské, což jak vidíme v grafu 4.26 úplně splněno není.

Experimentální hodnoty vazebné volné energie byly známy pouze pro některé se zkoumaných ligandů. Pro většinu těchto ligandů můžeme pozorovat dobrou shodu experimentálních výsledků s výsledky simulací. Výjimkou je ligand 1d45, pro který je vidět značný rozdíl mezi simulacemi a experimentem. Tento rozdíl byl pravděpodobně způsoben tím, že ligand 1d45 zde byl simulován jako nekovalentní

inhibitor, ve skutečnosti zřejmě v aktivním místě proteázy Mpro tvoří kovalentní vazbu [42].

Výsledné hodnoty vazebné volné energie jsou v tabulce 4.13 srovnány i s referenčními hodnotami [42], které byly prostřednictvím nerovnovážných MD simulací získány nově navrženou metodou NEW-vDSSB (angl. new virtual double-system single-box). Tato metoda je ovšem poměrně komplikovaná jak z hlediska realizace MD simulací (zahrnuje dlouhé rovnovážné REMD běhy), tak z hlediska jejich analýzy (konvoluce rozdělení práce). Díky této komplikovanosti se metoda NEW-vDSSB příliš nehodí pro provádění MD simulací pro větší série ligandů v rané fázi hledání optimálního inhibitoru. V řadě ohledů se navíc metoda NEW-vDSSB principiálně odlišuje od simulačního protokolu používaného v této práci. U vázaných ligandů sice i metoda NEW-vDSSB používá anihilační transformace, u volných ligandů transformace kreační. Způsob výpočtu korekce na standardní stav je v rámci metody NEW-vDSSB zcela odlišný, je založen na odhadu objemu vazebného místa z rovnovážných REMD běhů. V rámci referenční NEW-vDSSB studie [42] bylo také použito zcela jiné silové pole (AMBER99SB-ILDN pro proteázu a GAFF2 pro ligandy) než to, které je využíváno v rámci této práce (CHARMM36 pro proteázu a CGenFF pro ligandy).

Pro většinu ligandů referenční výsledky získané metodou NEW-vDSSB [42] odpovídají výsledkům získaným v rámci této práce. Vzhledem ke všem výše zmíněným rozdílům v metodice je shoda poměrně dobrá. Pro několik ligandů je však rozdíl ve výsledných hodnotách nezanedbatelný, přičemž většinou je hodnota vazebné volné energie získaná v rámci této práce nižší (více záporná) než ta, která byla získána metodou NEW-vDSSB. Některé ligandy se zde tedy jeví jako nadějnější inhibitory, než jak je ohodnocuje metoda NEW-vDSSB.

Domnívám se, že je to důsledek skutečnosti, že nerovnovážné MD simulace provedené v rámci této práce navazovaly přímo na docking, kdežto v rámci metody NEW-vDSSB byl proveden ještě REMD mezikrok, kdy aminokyseliny ve vazebném místě proteázy byly zahřáty na velmi vysokou teplotu. Cílem bylo zahrnout do výpočtu i suboptimální konfigurace ligandu ve vazebném místě proteázy. Otázkou ovšem je, do jaké míry výsledné vazebné volné energie pak odpovídají realitě. Problém vidím v tom, že pro každou odlišnou konfiguraci ligandu ve vazebném místě proteázy by pak bylo třeba provést řádově 1000 alchymistických transformací, aby se dosáhlo konvergence a teprve výsledky z těchto dílčích sérií MD simulací by se měly zkombinovat. Výpočty takového rozsahu jsou ovšem v dnešní době obtížně realizovatelné a ani v referenční NEW-vDSSB studii takový rozsah neměli [42]. Obecně mi přijde žádoucí, aby metoda neprodukovala falešně negativní výsledky, tedy aby z výběru nevypadly potentní ligandy pouze díky artefaktům ve výpočetním protokolu. A u metody NEW-vDSSB si nejsem jistý, zda k tomu v její aktuální podobě nedochází.

Celkově tedy jednoduchý simulační protokol používaný v rámci této práce obstál v testu na komplexním biomolekulárním systému velmi dobře. Do budoucna (až budou k dispozici významně výkonnější hardwarové prostředky, než které v současné době poskytuje MetaCentrum), bude dobré ověřit, zda jednoduchý simulační protokol občas neprodukuje i falešně pozitivní výsledky díky zanedbání suboptimálních alternativních konfigurací ligandů ve vazebných místech proteinů.

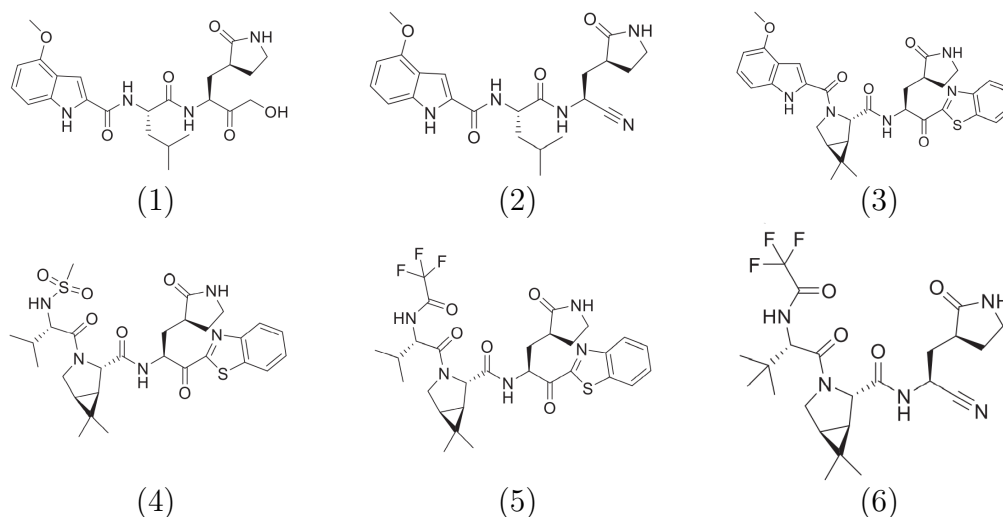
4.5 Kovalentní inhibitory proteázy Mpro

Poslední sada molekul, kterou jsem v této práci studoval, sestávala ze série šesti ligandů, jejichž struktury jsou znázorněny na obrázku 4.30 [106]. V dalším textu je označuji jednoduše jako ligandy 1 až 6.

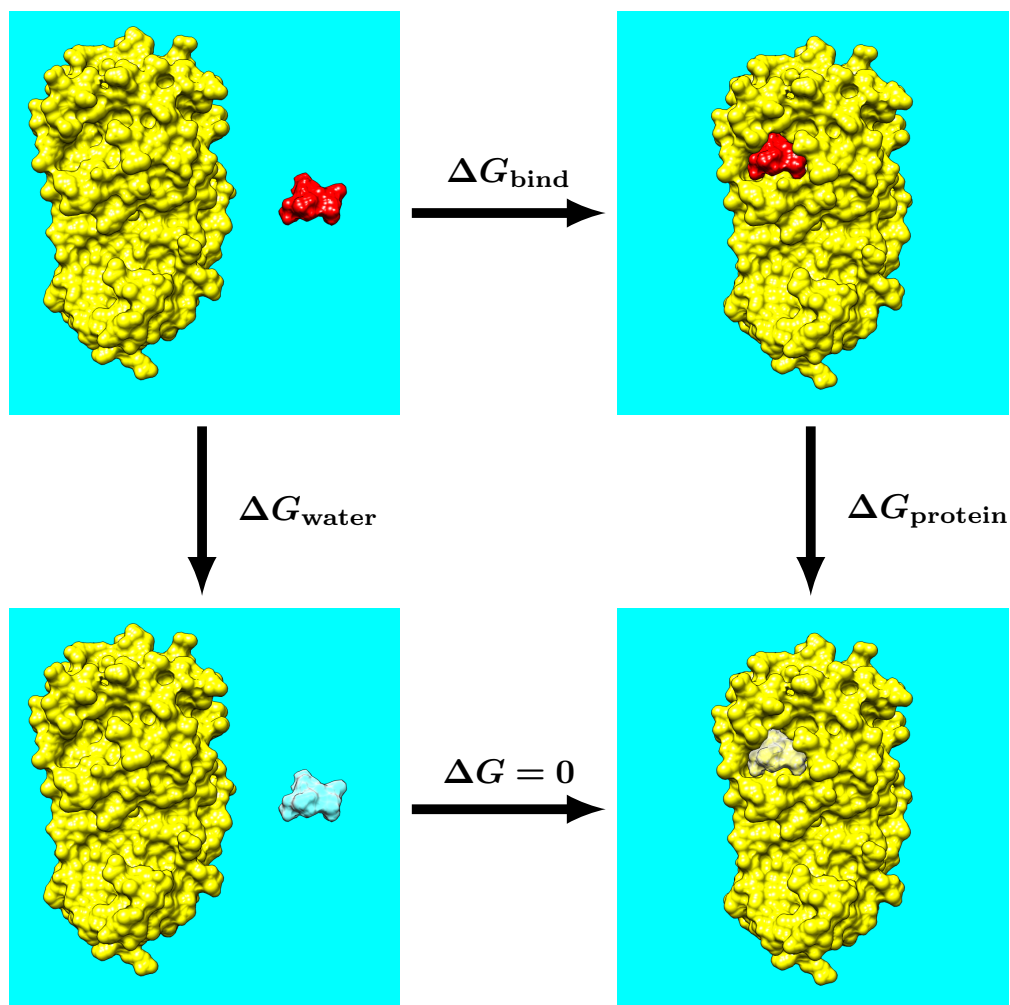
Tyto ligandy byly zkoumány firmou Pfizer jako potenciální inhibitory proteázy Mpro za účelem léčby onemocnění Covid-19. Kvůli zvýšení orální absorpce byly generovány strukturně podobné sloučeniny lišící se pouze v některých funkčních skupinách, jako je např. nitrilová skupina ligandu 2 a 6. Nakonec byl pro klinické testování vybrán ligand 6. Výběr byl založen na inhibičních schopnostech, orální absorpci a možnosti syntézy ve velkém měřítku. V současné době je ligand 6 známý pod názvem Paxlovid a běžně se používá v klinické praxi.

Pro sérii ligandů 1 až 6 jsou známé experimentální hodnoty inhibičních konstant vůči koronavirové proteáze Mpro. Jedná se vesměs o kovalentní inhibitory. Váží se přímo na cystein z katalytické dyády Cys-His proteázy Mpro. Ke kovalentní vazbě dochází buďto přes nitrilovou skupinu ligandů 2 a 6, nebo přes keto skupinu ostatních ligandů. Pro některé ligandy jsou dostupné krystalové struktury jejich komplexů s koronavirovou proteázou Mpro. Z těchto struktur mimo jiné vyplývá, že kovalentní vazba ligandů na proteázu Mpro je dost možná reverzibilní [25], protože postranní řetězec Cys145 z katalytické dyády se může nacházet ve dvou různých konformacích, z nichž jedna je s výše zmíněným kovalentním navázáním ligandů nekompatibilní.

Já jsem se zde zaměřil na klasické MD simulace ligandů v aktivním místě proteázy Mpro ve stavu těsně před kovalentním připojením ligandů k Cys145. Termodynamický cyklus, který jsem použil pro výpočty vazebné volné energie těchto kovalentních inhibitorů, je znázorněn na obrázku 4.31. Jde tedy opět o dvojitou anihilaci, tedy přístup, který se osvědčil v předchozích podkapitolách.



Obrázek 4.30: Struktury studovaných kovalentních inhibitorů proteázy Mpro. V následujícím textu jsou pro zjednodušení značeny jako ligandy 1 až 6.



Obrázek 4.31: Termodynamický cyklus využitý při výpočtech vazebné volné energie kovalentních inhibitorů.

4.5.1 Příprava simulovaných systémů

Počáteční souřadnice použité v MD simulacích byly z větší části převzaty z proteinové databanky RCSB (www.rcsb.org), ve které jsou uloženy krystalové struktury koronavirové proteázy Mpro v komplexu s kovalentně navázanými ligandy 3, 4 a 6 pod identifikátory 7RFR, 7RFU a 7RFS. Uvedené struktury byly získány prostřednictvím rentgenové difrakce, takže neobsahují atomy vodíku, které bylo třeba doplnit na základě předdefinovaných topologií aminokyselin apod. v programu VMD.

Chystal jsem se provádět klasické MD simulace, neboť mě zajímala především nekovalentní složka interakce ligandů s proteázou. Proto byla ve výchozích krystalových strukturách odstraněna kovalentní vazba ligandů na Cys145 a byly doplněny atomy vodíku tak, aby to odpovídalo stavu s kovalentně nenavázaným ligandem.

Jelikož se zkoumané ligandy lišily pouze v několika funkčních skupinách, tak i pro zbylé ligandy 1, 2 a 5 byly jako základ využity výše zmíněné krystalové

struktury, ve kterých jsem modifikoval strukturu ligandů 3, 4 a 6 pomocí modulu Molefactory ze softwarového balíku VMD.

Ze struktur komplexů ligandů s proteázou byly převzaty souřadnice ligandů i pro simulace volných ligandů ve vodním prostředí.

Proteáza Mpro byla opět popsána pomocí silového pole CHARMM a ligandy pomocí silového pole CGenFF. Před aplikací algoritmů CGenFF pro generování silových konstant pro zkoumané ligandy jsem provedl jednoduchou optimalizaci geometrie ligandů ve vakuu prostřednictvím programu Gaussian na úrovni teorie MP2-6-31G(d). Tato optimalizace odstranila možné nepřesnosti či deformace v konformacích ligandů převzatých z krystalových struktur komplexů.

Pro přípravu simulačních protokolů jsem využil modul QwikMD v programu VMD. Počáteční struktury komplexů ligandů s proteázou jakož i počáteční struktury samotných ligandů byly umístěny do simulačního boxu a obklopeny molekulami vody popsanými pomocí modelu TIP3P. Simulované systémy komplexů obsahovaly přibližně 100 000 atomů a simulované systémy samotných ligandů zhruba 10 000 atomů. Do simulovaných systémů byly doplněny ionty sodíku a chloru tak, aby bylo dosaženo standardní koncentrace 0,15 mol/l.

Pro všechny připravené simulované systémy byla spuštěna krátká minimalizace s 2000 kroky pro odstranění blízkých kontaktů atomů. Následně byl spuštěn annealing, při kterém byla teplota postupně zvětšována z počáteční hodnoty 60 K na konečnou hodnotu 300 K a konformace proteázy byla udržována harmonickými omezení v počáteční konformaci převzatá z krystalové struktury. Následoval obyčejný ekvilibrační běh s délkou 1 ns, během něhož byla stále zapnuta harmonická omezení udržující počáteční konformaci proteinu. Nakonec byl spuštěn další ekvilibrační již zcela bez dodatečných omezení a s celkovou délkou trvání 2 ns.

4.5.2 Výsledky simulací

Výpočet vazebné volné energie jsem provedl stejně jako v předchozí podkapitole věnované nekovalentním ligandům. Tedy prostřednictvím metody dvojité anihilace s přímou aplikací Jarzynského identity resp. její aproximace pomocí kumulantů. Bylo proto vždy spuštěno 1000 MD běhů a to pouze v dopředném anihilačním směru alchymistické transformace, kdy ligand postupně mizel ze simulovaného systému. Alchymistická transformace byla přitom rozdělena do 20 λ -oken s krokem $\Delta\lambda = 0,05$. Celková délka jedné trajektorie činila 400 ps. Výslednou hodnotu volné energie jsem pak určoval jak pomocí Jarzynského identity (3.22), tak pomocí aproximace kumulanty (3.24). K výsledným hodnotám vazebné volné energie byla posléze připočtena ještě korekce na standardní stav (3.27).

Výsledné hodnoty změny volné energie pro vázané ligandy shrnuje tabulka 4.14 a pro volné ligandy ve vodě tabulka 4.15. Z rozdílu těchto dílčích výsledků byla získána výsledná vazebná volná energie dle termodynamického cyklu 3.1. Konkrétní hodnoty vazebné volné energie jsou uvedeny v tabulce 4.16 a vizuálně zobrazeny na obrázku 4.32.

Z výsledků ekvilibrace a následných alchymistických trajektorií bylo zřejmé, že ligandy preferovaly odlišné konformace než jsou ty, které zachycují výchozí krystalové struktury. Především atomy, u kterých dochází k tvorbě kovalentní vazby, tj. atomy z ligandů a z postranního řetězce katalytického cysteinu, se vzdálily na cca. 5 Å. Heterocyklus odpovídající derivátu pyrrolidinu a imitující

Tabulka 4.14: Výsledky alchymistických transformací pro kovalentní inhibitory vázané k proteáze Mpro. Hodnoty jsou uvedeny v jednotkách kcal/mol.

Ligand	$\Delta G_{\text{kumulant}}$	$\Delta G_{\text{Jarzynski}}$	ΔG_{vol}
1	33,8	37,0	-0,2
2	40,4	40,7	0,0
3	40,3	42,4	0,1
4	40,9	44,2	0,2
5	34,9	35,3	0,3
6	30,0	31,7	-0,1

Tabulka 4.15: Výsledky alchymistických transformací pro volné kovalentní inhibitory ve vodním prostředí. Hodnoty jsou uvedeny v jednotkách kcal/mol.

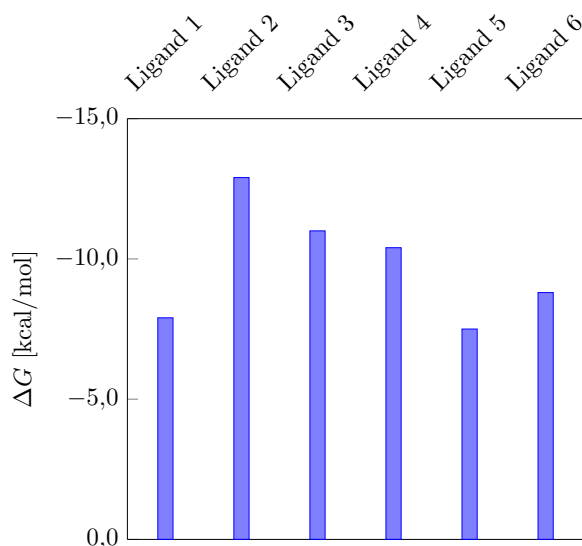
Ligand	$\Delta G_{\text{kumulant}}$	$\Delta G_{\text{Jarzynski}}$
1	26,1	28,3
2	27,5	27,5
3	29,1	31,1
4	30,3	31,9
5	27,0	28,0
6	21,3	22,2

Tabulka 4.16: Výsledné hodnoty vazebné volné energie komplexů kovalentních inhibitorů a proteázy Mpro. Hodnoty volné energie jsou uvedeny v jednotkách kcal/mol.

Ligand	$\Delta G_{\text{kumulant}}$	$\Delta G_{\text{Jarzynski}}$	$K_{i,\text{exp}}$ [nM]	ΔG_{exp}
1	-7,9	-8,9	0,3	-13,1
2	-12,9	-13,2	27,7	-10,4
3	-11,0	-11,1	230,0	-9,1
4	-10,4	-12,1	7,9	-11,1
5	-7,5	-7,0	12,1	-10,9
6	-8,8	-9,6	3,1	-11,7

přítomnost klíčového Gln, který je společný všem zkoumaným ligandům (a mnoha jejich předchůdcům se stopou sahající až do 90. let) se vzdálil z vazebné kavity S1, v níž byl původně ukotven v krystalové struktuře.

Na konformační změny ligandů mohla mít vliv zvolená protonace histidinů proteázy Mpro. Vodíky totiž nelze prostřednictvím rentgenové difrakce v krystalcích rozlišit, a proto je třeba je doplnit na základě předdefinovaných topologií aminokyselin. V případě histidinů ovšem existuje více možných protonací. V



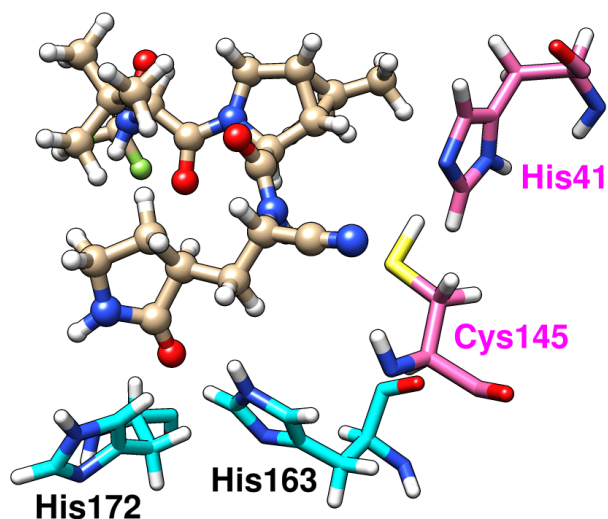
Obrázek 4.32: Vazebná volná energie kovalentních inhibitorů při jejich vazbě k proteáze Mpro. Výsledky první sady MD simulací s histidiny proteázy Mpro protonovanými na δ -dusíku.

proteáze Mpro se přitom vyskytují hned dva histidiny (His163 a His172) v bezprostřední blízkosti postranního řetězce klíčového Gln ze štípaného substrátu, či v bezprostřední blízkosti heterocyklu strukturního analogu Gln, který je tradiční součástí struktury potenciálních inhibitorů koronavirových polymeráz. Zvolená protonace obou zmíněných histidinů proteázy Mpro tak může mít podstatný vliv na konečnou konformaci ligandů.

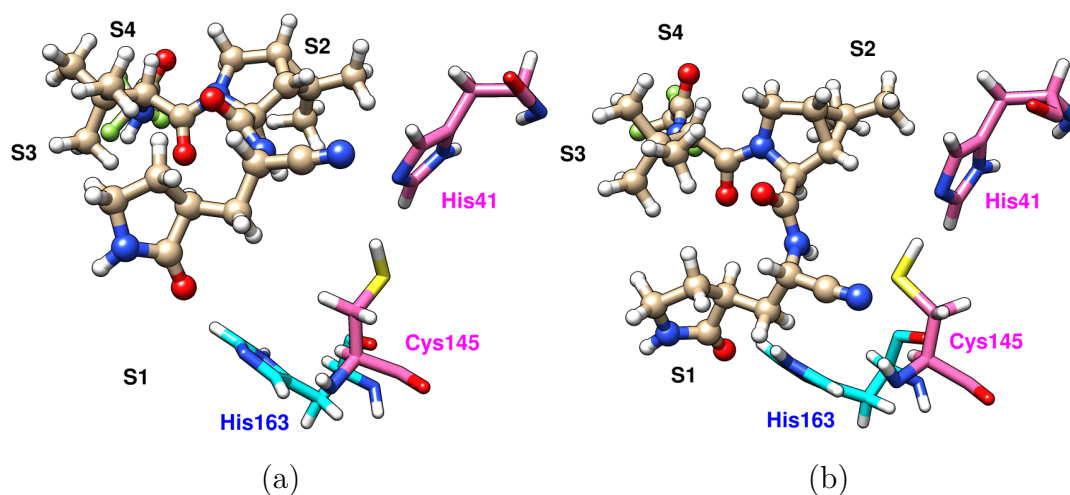
V úvodní sérii MD simulací jsem uvažoval oba tyto histidiny s protonovaným δ -dusíkem. Pro ověření vlivu protonace jsem připravil nové simulované systémy se všemi možnými kombinacemi protonací zmíněných dvou histidinů. Neutrální verze histidinů měly protonovaný buď δ -dusík nebo ε -dusík. Kladně nabitě histidiny měly protonované oba zmíněné histidiny. Pro test vlivu protonace histidinů jsem použil pouze ligand 6. Takže celkově bylo připraveno 9 nových simulovaných systémů.

Přípravné MD simulace byly realizovány stejným způsobem jako s původní protonací histidinů. Postupně jsem spustil minimalizaci, annealing, ekvilibraci s omezeními a nakonec obyčejnou ekvilibraci bez omezení páteře proteázy Mpro. V řadě případů došlo opět k markantním konformačním změnám konformace ligandu 6. Nicméně v případě, kdy byly histidiny 163 a 172 protonované na ε -dusíku, zůstala konformace ligandu stabilní a vytvořila se stabilní vodíková vazba mezi postranním řetězcem His163 a heterocyklem analogu Gln z ligandu 6, jak znázorňuje obrázek 4.33. Atomy ligandu 6, které se mají kovalentně navázat k Cys145, byly též v adekvátní vzdálenosti.

Rozdíl v konformacích ligandu 6 pro původní a pro nově zvolenou protonaci histidinů proteázy Mpro je skutečně markantní, jak znázorňuje obrázek 4.34. Proto jsem se rozhodl výpočty volné energie opakovat pro všechny ligandy s nově zvolenou protonací histidinů 163 a 172, které jsem nyní uvažoval protonované na ε -dusíku.



Obrázek 4.33: Nově zvolená protonace histidinů 163 na 172 na ϵ -dusíku, která umožňuje vytvoření stabilní vodíkové vazby mezi His163 a analogem Gln ve struktuře ligandů.



Obrázek 4.34: Srovnání konformací ligandu 6 (a) pro původní protonaci histidinů 163 a 172 proteázy Mpro na δ -dusíku a (b) pro nově zvolenou protonaci histidinů na ϵ -dusíku

Kromě toho jsem ještě doladil počáteční konformace ligandů tak, aby polohy všech jejich funkčních skupin maximálně korespondovaly s polohami odpovídajících motivů v krystalových strukturách.

Připravil jsem simulované systémy pro zbylé ligandy 1 až 5 a provedl jejich ekvilibraci stejně jako v případě ligandu 6. Poté jsem pro jednotlivé ligandy opět provedl 1000 alchymistických transformací se stejným nastavením a délkou (400 ps) jako v první sérii výpočtů.

Hodnoty rozdílu volné energie z alchymistické transformace pro ligandy vázané v komplexu s Mpro shrnuje tabulka 4.17 a pro volné ligandy tabulka 4.18. Vazebná volná energie počítaná z jejich rozdílu je pak uvedena v tabulce 4.19 a graficky znázorněna na obrázku 4.35.

Tabulka 4.17: Výsledky alchymistických transformací pro ligandy 1 až 6 vázané k Mpro s nově zvolenou protonací histidinů 163 a 172. Hodnoty jsou uvedeny v jednotkách kcal/mol.

Ligand	$\Delta G_{\text{kumulant}}$	$\Delta G_{\text{Jarzynski}}$	ΔG_{vol}
1	44,6	44,7	0,0
2	40,0	40,9	-0,3
3	52,5	53,2	0,0
4	48,9	52,0	0,2
5	43,2	44,2	0,2
6	39,5	41,2	0,2

Tabulka 4.18: Výsledky alchymistických transformací pro ligand 1 až 6 ve vodní obálce. Hodnoty jsou uvedeny v jednotkách kcal/mol.

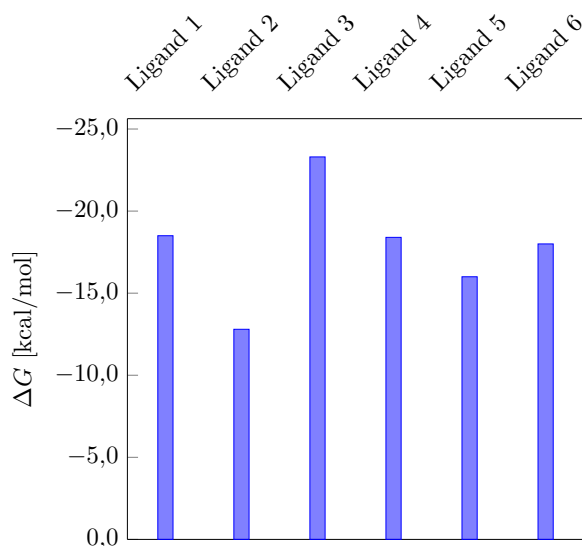
Ligand	$\Delta G_{\text{kumulant}}$	$\Delta G_{\text{Jarzynski}}$
1	26,1	28,3
2	27,5	27,5
3	29,1	31,1
4	30,3	31,9
5	27,0	28,0
6	21,3	22,2

Tabulka 4.19: Vazebná volná energie ligandů 1 až 6 s nově zvolenou protonací histidinů 163 a 172 proteázy Mpro. Hodnoty jsou uvedeny v jednotkách kcal/mol.

Ligand	$\Delta G_{\text{kumulant}}$	$\Delta G_{\text{Jarzynski}}$	$K_{i,\text{exp}}$ [nM]	ΔG_{exp}
1	-18,5	-16,4	0,3	-13,1
2	-12,8	-13,7	27,7	-10,4
3	-23,3	-22,0	230,0	-9,1
4	-18,4	-19,9	7,9	-11,1
5	-16,0	-16,0	12,1	-10,9
6	-18,0	-18,8	3,1	-11,7

Na první pohled je zřejmé, že hodnota vazebné volné energie vypočtená pro ligand 3 se poněkud vymyká a ligand 3 by na základě ní měl být zdaleka nejúčinnějším inhibítorem z celé série ligandů 1 až 6. Přitom experimentálně určená vazebná volná energie svědčí o opaku.

Ligand 3 se v experimentech popsaných v literatuře choval jinak než zbývající ligandy v řadě ohledů. Ligand 3 má velmi specifickou strukturu se dvěma



Obrázek 4.35: Vazebná volná energie ligandů 1 až 6 s nově zvolenou protonací histidinů 163 a 172 proteázy Mpro.

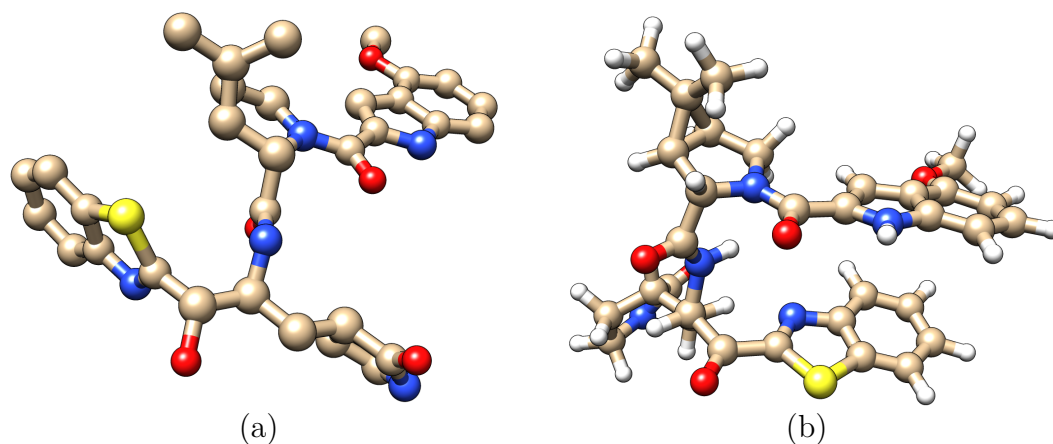
objemnými planárními motivy. Lze si představit, že tyto části ligandu 3 mohou stackovat analogicky jako báze nukleových kyselin. Taková interakce by mohla ligand 3 “uzamknout” v konformaci, která se značně liší od té, kterou ligand 3 zaujímá v aktivním místě koronavirové proteázy Mpro ve struktuře 7RFR. Lze si dále představit, že molekuly ligandu 3 v takové kompaktní konformaci by mohly dále agregovat a vytvářet komplexy vyššího řádu. To vše by mohlo zhoršit schopnost ligandu 3 pronikat do buněk apod.

Zde je ale nejpodstatnější, že by to také mohlo významně snížit počet molekul ligandu 3, které jsou volné a navíc v konformaci kompatibilní s aktivním místem proteázy Mpro. Což by se mohlo efektivně projevit zvýšením experimentální K_i , z níž byla vypočtena referenční experimentální vazebná volná energie.

Provedl jsem proto sérii dlouhých MD simulací volného ligandu 3 ve vodním prostředí, které startovaly z konformeru zachyceného v krystalové struktuře. Po jisté době jsem opakovaně pozoroval přechod ligandu 3 do předpokládané kompaktní konformace, ve které jeho planární části stackovaly 4.36, analogicky jako k tomu dochází například u dinukleotidů.

Pokud jde o prostorové uspořádání samotného aktivního místa proteázy Mpro, v průběhu provedených MD simulací bylo možné pozorovat dvě různé konformace postranního řetězce cysteinu z katalytické dyády proteázy, jak ukazuje obrázek 4.37. To není překvapení, neboť i v krystalové struktuře 7VH8 zobrazené na obrázku 1.14, je možné tyto dvě konformace postranního řetězce Cys145 rozlišit. Svědčí to o tom, že případná kovalentní vazba je nejspíš poměrně slabá a reversibilní.

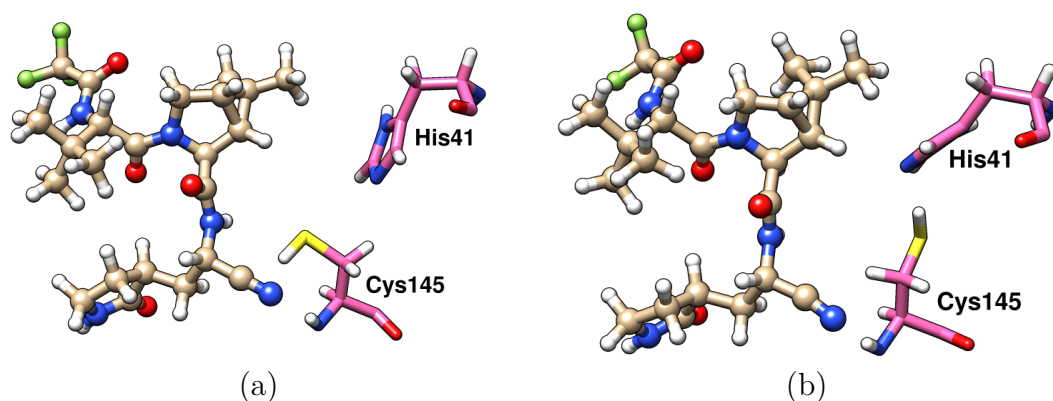
Zajímavé je i to, že ve zde provedených MD simulacích konformer postranního řetězce Cys145, který je nekompatibilní s kovalentním navázáním ligandů, obvykle dominoval. Vzniká tak otázka, zda v krystalových strukturách nemají vliv umělé krystalové kontakty (angl. crystal packing forces), které komplexy Mpro deformují a vychylují ligandy ze stavu substrátů do stavu produktů reakce. Vizuální inspekce krystalových struktur ukazuje, že například ligand 6 se



Obrázek 4.36: (a) Konformace ligandu 3 v krystalové struktuře (PDB ID 7RFR). (b) Kompaktní konformace volného ligandu 3 ve vodním prostředí se stackingem jeho planárních motivů.

artificiálních krystalových kontaktů přímo účastní.

Ve zde prezentovaných výpočtech vazebné volné energie jsem příspěvek kovalentního navázání ligandů k postrannímu řetězci Cys145 nezahrnul. Díky reverzibilitě kovalentní vazby by zřejmě nebyl příliš velký. Jeho velikost by bylo možné dopočítat pomocí QMMM výpočtů, které by bylo možné realizovat například pomocí kombinace programů NAMD a Orca.



Obrázek 4.37: Dvě konformace cysteinu 145 z katalytické dyády proteázy Mpro pozorované během rovnovážných simulací.

4.5.3 Diskuze

Pro první sérii MD simulací s histidiny 163 a 172 protonovanými na δ -dusíku, jsou výsledné hodnoty vazebné volné energie uvedené v tabulce 4.16. Pro druhou sérii MD simulací s histidiny 163 a 172 protonovanými na ε -dusíku, jsou výsledné hodnoty vazebné volné energie uvedené v tabulce 4.19. Z uvedených hodnot je zřejmé, že protonace histidinů v blízkosti vazebného místa S1 poblíž aktivního místa proteázy Mpro má značný vliv na vazebné konformace ligandů a v konečném výsledku i na vypočtené hodnoty vazebné volné energie.

Dle experimentálně určených inhibičních konstant $K_{i,\text{exp}}$ by ligandy 1, 4 a 6 měly být lepšími inhibitory než ligandy 2, 3 a 5. Výsledné hodnoty vazebné energie 4.35 ze druhé série MD simulací tomu kvalitativně odpovídají, ligandy 1, 4 a 6 z nich vycházejí jako lepší inhibitory než ligandy 2 a 5. Výjimku tvoří ligand 3, který z výsledků druhé série MD simulací vyšel jako nejlepší, z experimentu ale naopak jako nejhorší. Domnívám se, že to je důsledek toho, že volný ligand 3 spontánně přechází do velmi kompaktní konformace, která je zcela nekompatibilní s aktivním místem Mpro. Tento konformační přechod jsem pozoroval v MD simulacích volného ligandu 3 ve vodním prostředí, kdy došlo ke stackingu planárních částí ligandu 3, jak znázorňuje obrázek 4.36.

Dále je zřejmé, že hodnoty vazebné volné energie ze druhé série MD simulací jsou podstatně nižší (více záporné) než experimentální hodnoty. MD simulace tak inhibiční schopnosti ligandů do určité míry přeceňují. To je zřejmě způsobeno několika faktory.

Nerovnovážné MD simulace jsou velmi rychlé. V tomto případě byly realizovány během pouhých 400 ps. Hodnoty vazebných volných energií z anihilačních transformací obecně ke správným výsledkům konvergují zdola. Ligandy se tedy při příliš rychlých anihilacích mohou jevit jako lepší inhibitory, než jakými ve skutečnosti jsou. Proto jsem pro ligand 6 provedl ještě dodatečnou sérii 1000 anihilačních transformací, kdy každá trvala 2000 ps. Transformace byly tedy provedeny 5x pomaleji než transformace z první i druhé série MD simulací provedených pro všechny ligandy. Výsledná vazebná volná energie $-14,6$ kcal/mol pro ligand 6 se skutečně významně přiblížila k experimentální hodnotě $-11,7$ kcal/mol. V tomto duchu by bylo možné dále pokračovat, alchymistické transformace by bylo možné prodlužovat a vypočtená hodnota by se opět o něco více přiblížila experimentální hodnotě.

Otázka je, zda je k tomu nějaký důvod, tedy zda by to poskytlo dostatečnou přidanou hodnotu. Kumulovaná délka transformací pro ligand 6 z poslední série MD simulací totiž byla již $2 \mu\text{s}$, což vyžaduje poměrně značné hardwarové prostředky. Smyslem výpočtu přitom je seřazení (ranking) ligandů dle inhibičních schopností a k tomu kvantitativní shoda mezi vypočtenou a experimentální volnou energií není nezbytně nutná. V publikovaných studiích totiž autoři často cílí pouze na korelaci, kvantitativně se přitom hodnoty experimentální a vypočtené prostřednictvím různých metod (MM/PBSA ale i DFT) mohou lišit až řádově [107].

Otázka také je, zda by v tomto konkrétním případě bylo vůbec možné dosáhnout dokonalé kvantitativní shody mezi vypočtenou a experimentální hodnotou vazebné volné energie pouhým prodlužováním alchymistických transformací.

Je totiž možné, že současné krystalové struktury neposkytují úplný obrázek o tom, jak vypadá vazba ligandů do aktivního místa proteázy Mpro. Kromě výše prezentovaných výpočtů vazebné volné energie prostřednictvím anihilačních alchymistických transformací provedených pomocí nerovnovážných MD simulací jsem totiž učinil několik pokusů získat vazebnou volnou energii i prostřednictvím rovnovážných MD simulací s obousměrnými alchymistickými transformacemi. Při tom jsem využíval dodatečná omezení, což vyžaduje výpočty podle složitějšího termodynamického cyklu. Tato omezení se aplikují pouze na ligand a k vypočtení jejich příspěvku je potřeba provést dodatečné MD simulace. Během těchto poměrně dlouhých rovnovážných běhů (až 50 ns) jsem pozoroval významné kon-

formační změny proteázy Mpro v oblasti aktivního místa nejen v průběhu alchymistických transformací, ale i v průběhu dodatečných běhů, během nichž ligand po celou dobu interaguje se svým okolím, pouze se mění intenzita jeho geometrických omezení. Z těchto MD simulací jsem získal dojem, že aktivní místo Mpro je konformačně daleko plastičtější, než jak ukazují dosavadní krystalové struktury a že by bylo zajímavé provést rovnovážné MD běhy o délce v řádu μs či dokonce desítek μs , které by umožnily získat alternativní struktury komplexu Mpro a ligandů. Je nepochybné, že hodnoty vazebné volné vypočtené na základě takových struktur by byly v absolutní hodnotě nižší než ty, získané na základě “ideálních” krystalových struktur ve druhé sérii simulací a tedy také podstatně blíže experimentálním hodnotám (kterým jsou ostatně poměrně blízko výsledky z první série simulací, kdy bylo “ideální” prostorové uspořádání komplexů z krystalu narušeno suboptimální protonací histidinů).

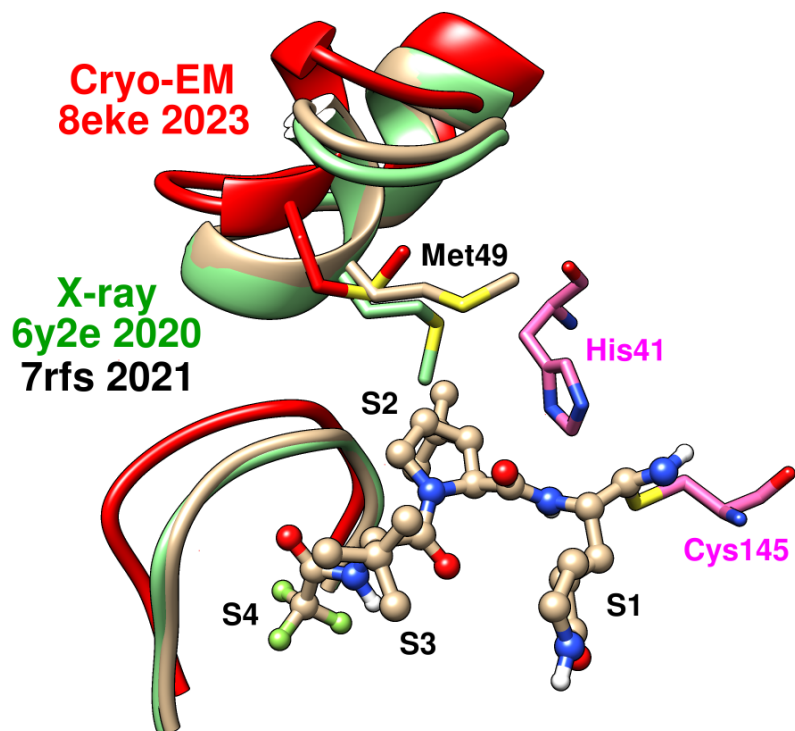
Můj dojem, že konformační plasticita Mpro v oblasti aktivního místa je podstatně výraznější, než jak ukazují stávající krystalové struktury, který jsem získal na základě dlouhých rovnovážných MD simulací, potvrzují první studie Mpro provedené prostřednictvím elektronového mikroskopu, které byly publikované během posledního měsíce [15].

Pokud provedeme strukturní alignment krystalových struktur proteázy Mpro v apo formě a Mpro s navázaným ligandem, tak je konformace páteře Mpro v obou případech téměř totožná, viz obrázek 4.38. Příspěvek umělé stabilizující síly v krystalu se stane patrný, když do strukturního alignmentu přidáme apostrukturu získanou pomocí elektronového mikroskopu [15], kde je konformace smyčky nad aktivním místem diametrálně odlišná 4.38.

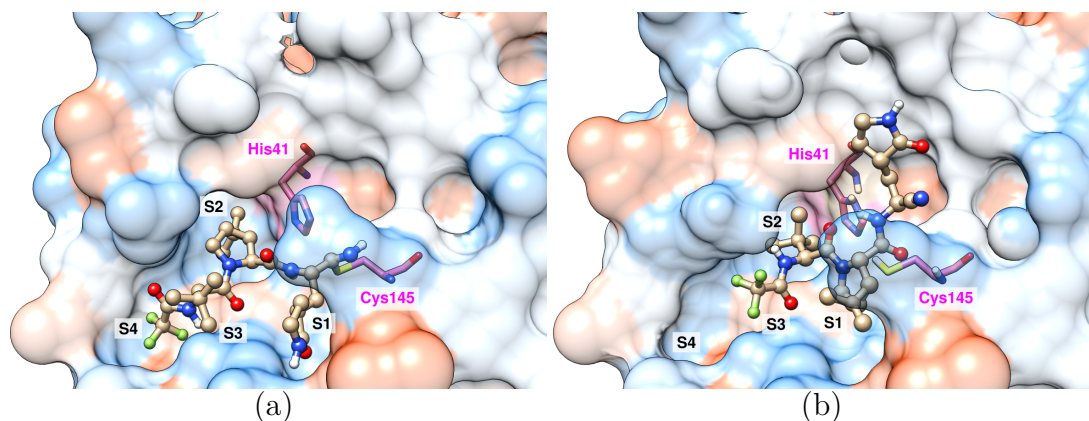
Kromě toho, že je aktivní místo proteázy Mpro mimořádně konformačně plastické, což koresponduje s jeho funkcí, neboť musí rozpoznávat 11 různých míst ve štípaném polyproteinu, tak je třeba vzít do úvahy i to, že aktivní místo je poměrně mělké. Obsahuje přitom sérii vazebných míst pro postranní řetězce aminokyselin a některé peptidy či inhibitory na bázi analog peptidů (jako jsou i ligandy 1 až 6) se mohou v roztoku do aktivního místa proteázy vázat dost možná více způsoby.

Například ligand 6 byl zkoumán prostřednictvím metody NEW-vDSSB skupinou P. Procacciho [43], přičemž na rozdíl od nás nevycházeli z krystalové struktury, ale z vlastního dockingu. Jak ukazuje obrázek 4.39, tak se jim nepodařilo ligand 6 umístit správně. Je ve vazebných místech proteázy Mpro o jednu pozici posunutý. Strukturní analog glutaminu je například mimo vazebné místo S1. Otázka je, do jaké míry je třeba takové alternativní konfigurace ligandů započíst do vazebné volné energie a do jaké míry je to u proteázy Mpro s extrémně flexibilním a mělkým aktivním místem vůbec reálné provést vyčerpávajícím způsobem.

Se současnými hardwarovými prostředky je nejspíše třeba počítat s tím, že hodnoty vazebné volné energie vypočtené prostřednictvím nerovnovážných MD simulací startujících z krystalových struktur Mpro jsou o něco nižší, než by odpovídalo realitě, necílit na dokonalou kvantitativní shodu s experimentem a pragmaticky používat vypočtené hodnoty vazebné volné energie k rankingu potenciálních inhibitorů, které mohou být navrženy například prostřednictvím softwarového balíku AutoGrow, který je v MetaCentru dostupný jako modul.



Obrázek 4.38: Rozdíl v konformaci páteře proteázy Mpro v aktivním místě. Proteáza Mpro s navázaným ligandem 6 (PDB ID 7RFS), apo forma v krystalové struktuře (PDB ID 6Y2E) a apo forma zachycená prostřednictvím elektronového mikroskopu (PDB ID 8EKE).



Obrázek 4.39: Umístění ligandu 6 (a) v krystalové struktuře (PDB ID 7RFS) a (b) v komplexu z dockingu P. Procacciho [43].

Závěr

Výpočty hydratační volné energie postranních řetězců aminokyselin posloužily k odladění postupu založeného na nerovnovážných MD simulacích. Hodnoty hydratační volné energie získané nerovnovážnými metodami byly srovnatelné s referenčními hodnotami získanými prostřednictvím rovnovážných MD simulací. Pro nabitě aminokyseliny byl prozkoumán vliv náboje na výsledné hodnoty hydratační volné energie a byla ověřena použitelnost analytické korekce. V případě vlivu oříznutí nevazebných van der Waalsových interakcí v konečné vzdálenosti jsem si ověřil, že vliv oříznutí má na výsledné hodnoty hydratační volné energie pouze zanedbatelný vliv. S využitím různých verzí silového pole CGenFF jsem dosáhl podobné přesnosti výsledných hydratačních volných energií. Pomocí vlastní reparametrizace bodových nábojů prostřednictvím ab initio výpočtů založené na variování empirického parametru umožňujícího implicitní zahrnutí polarizace v důsledku předpokládaného vlivu okolního prostředí jsem byl schopen výsledné hodnoty hydratační volné energie výrazně upřesnit. Soulad s experimentem se zlepšil řádově - z jednotek na 0,1 kcal/mol.

Na komplexu éterové koruny a draslíku jsem prostudoval alchymistický výpočet vazebné volné energie založený na termodynamickém cyklu. Zkoumal jsem využití dodatečných omezujících potenciálů pro urychlení konvergence a různé přístupy, jak započítat korekci na standardní stav. Studoval jsem také vliv délky a celkového počtu dílčích alchymistických transformací na konečné výsledky. Získal jsem srovnatelné výsledky přístupy založenými jak na Crooksově flukтуаčním teorému tak i na Jarzynského identitě.

Dále jsem provedl dlouhé rovnovážné MD simulace, na základě nichž byly určeny vazebné konformace komplexů adamantánů s cyklodextriny. Následně jsem na tyto komplexy s úspěchem aplikoval nerovnovážnou metodu výpočtu vazebné volné energie založenou na přímé aplikaci Jarzynského rovnosti. Vypočtené hodnoty volné energie odpovídaly referenčním hodnotám z experimentu.

Simulační parametry použitelné v případě složitějších biomolekulárních systémů jsem doladil na základě výpočtů vazebné volné energie pro sadu nekovalentních inhibitorů proteázy Mpro. Hodnoty vazebné volné energie získané pomocí nerovnovážných metod umožnily seřadit potenciální inhibitory dle předpokládané účinnosti. Bylo provedeno srovnání vypočtených hodnot s dostupnými referenčními daty jak z experimentu tak z výpočtů provedených ale v řadě ohledů odlišným způsobem.

V případě série kovalentních inhibitorů proteázy Mpro byla prostudována nekovalentní část jejich interakce, kdy se ligandy, ještě kovalentně nepřipojené, nacházejí v aktivním místě proteázy Mpro ve formě substrátů reakce. Na základě rovnovážných MD simulací byl určen vliv protonace histidinů v aktivním místě proteázy Mpro na vazebné konformace zkoumaných ligandů. Vazebná volná energie byla nerovnovážnými metodami určena pro dvě možné protonace těchto histidinů. Vypočtené hodnoty vazebné volné energie se lépe shodovaly s experimentem, pokud histidiny byly protonované na δ -dusíku. V tom případě se ale konformace ligandů nezanedbatelně lišily od jejich konformací ve výchozích krystalových strukturách. V případě protonace histidinů na ϵ -dusíku se ligandy jeví jako účinnější inhibitory, než jak ukazuje experiment. Konformace z výchozích

krystalových struktur přitom zůstala dokonale zachována. Bylo ověřeno, že neso-
ulad mezi vypočtenou a experimentální hodnotou je možné do jisté míry zmen-
šit pomocí delších simulací. Nicméně se zdá, že se zde musí uplatňovat ještě další
faktory.

Dlouhé rovnovážné MD simulace, které jsem prováděl při testování alterna-
tivního způsobu výpočtu volné energie, naznačovaly, že aktivní místo proteázy
Mpro může být konformačně značně bohatší, než jak ukazují dosavadní kry-
stalové struktury, kde má díky tzv. “crystal packing forces” stejnou konformaci
proteáza s navázaným ligandem i bez něj. První struktury proteázy Mpro získané
pomocí elektronového mikroskopu, které se během posledního měsíce objevily v
proteinové databance, dojem z dlouhých rovnovážných MD simulací potvrzují.
Např. v případě apo formy Mpro je konformace aktivního místa významně od-
lišná a vykazuje velké fluktuační. Konformační fluidita aktivního místa má nepo-
chybně funkční význam, neboť proteáza musí správně rozpoznávat 11 odlišných
míst ve štípaném polyproteinu. Dramaticky to ovšem komplikuje racionální de-
sign potenciálních inhibitorů. Jedním z vysvětlení, proč více méně selhaly dvě
velké kampaně, při nichž byla prostřednictvím dockingu testována vždy více než
miliarda ligandů, je to, že potenciální nekovalentní inhibitory byly dockovány
do historicky první krystalové struktury 6LU7, která ovšem původně obsahovala
kovalentní inhibitor.

S ohledem na výše uvedené se tedy zdá, že vazba ligandů do mělkého, konfor-
mačně proměnlivého aktivního místa proteázy Mpro může být labilnější, než jak
naznačují stávající krystalové struktury. Ideální tedy asi bude, provádět úvodní
docking do hned několika reprezentativních struktur Mpro získaných například na
základě dlouhých (jednotky až desítky μs) rovnovážných MD simulací. Ideálním
nástrojem pro virtuální syntézu ligandů a jejich následný docking, přitom může
být například softwarový balík AutoGrow dostupný v MetaCentru jako modul.
Navazující výpočty volné energie pomocí nerovnovážných MD simulací (v du-
chu této práce), pak mohou posloužit k fyzikálně korektnímu rankingu nejlepších
ligandů navržených programem AutoGrow.

Simulační protokol, velké množství skriptů, způsob výpočtu korekce volné
energie na standardní stav či postup při reparametrizaci silového pole, které byly
vytvořené a odladěné v rámci této práce, se mohou uplatnit i mimo kontext ko-
ronavirové proteázy Mpro. Jsou univerzálně využitelné při obdobných výpočtech
vazebné volné energie pro nejrůznější (bio)molekulární komplexy a pro racionální
design inhibitorů řady jiných medicínsky zajímavých proteinů.

Seznam použité literatury

- [1] Subramanian Boopathi, Adolfo B. Poma, and Ponmalai Kolandaivel. Novel 2019 coronavirus structure, mechanism of action, antiviral drug promises and rule out against its treatment. *Journal of Biomolecular Structure and Dynamics*, 39(9):3409–3418, 2021.
- [2] Sk Manirul Haque, Omar Ashwaq, Abdulla Sarief, and Abdul Kalam Azad John Mohamed. A comprehensive review about SARS-CoV-2. *Future Virology*, 15(9):625–648, 2020.
- [3] Isabel Sola, Fernando Almazán, Sonia Zúñiga, and Luis Enjuanes. Continuous and discontinuous RNA synthesis in coronaviruses. *Annual Review of Virology*, 2(1):265–288, 2015.
- [4] Antonio Francés-Monerris, Cécilia Hognon, Tom Miclot, Cristina García-Iriepa, Isabel Iriepa, Alessio Terenzi, Stéphanie Grandemange, Giampaolo Barone, Marco Marazzi, and Antonio Monari. Molecular basis of SARS-CoV-2 infection and rational design of potential antiviral agents: Modeling and simulation approaches. *Journal of Proteome Research*, 19(11):4291–4315, 2020.
- [5] Can-Rong Wu, Wan-Chao Yin, Yi Jiang, and H Eric Xu. Structure genomics of SARS-CoV-2 and its omicron variant: drug design templates for COVID-19. *Acta Pharmacologica Sinica*, 43(12):3021–3033, 2022.
- [6] Van A. Ngo and Ramesh K. Jha. Identifying key determinants and dynamics of SARS-CoV-2/ACE2 tight interaction. *PLOS ONE*, 16(9):1–15, 2021.
- [7] Anshumali Mittal, Kavyashree Manjunath, Rajesh Kumar Ranjan, Sandeep Kaushik, Sujeet Kumar, and Vikash Verma. COVID-19 pandemic: Insights into structure, function, and hACE2 receptor recognition by SARS-CoV-2. *PLOS Pathogens*, 16(8):1–19, 2020.
- [8] Chang-ki Oh, Tomohiro Nakamura, Nathan Beutler, Xu Zhang, Juan Piña-Crespo, Maria Talantova, Swagata Ghatak, Dorit Trudler, Lauren N. Carnevale, Scott R. McKercher, Malina A. Bakowski, Jolene K. Diedrich, Amanda J. Roberts, Ashley K. Woods, Victor Chi, Anil K. Gupta, Mia A. Rosenfeld, Fiona L. Kearns, Lorenzo Casalino, Namir Shaabani, Hejun Liu, Ian A. Wilson, Rommie E. Amaro, Dennis R. Burton, John R. Yates, Cyrus Becker, Thomas F. Rogers, Arnab K. Chatterjee, and Stuart A. Lipton. Targeted protein S-nitrosylation of ACE2 inhibits SARS-CoV-2 infection. *Nature Chemical Biology*, 19(3):275–283, 2023.
- [9] Henrique S Fernandes, Sérgio F Sousa, and Nuno M F S A Cerqueira. New insights into the catalytic mechanism of the SARS-CoV-2 main protease: an ONIOM QM/MM approach. *Molecular Diversity*, 26(3):1373–1381, 2022.
- [10] Zhenming Jin, Xiaoyu Du, Yechun Xu, Yongqiang Deng, Meiqin Liu, Yao Zhao, Bing Zhang, Xiaofeng Li, Leike Zhang, Chao Peng, Yinkai Duan,

- Jing Yu, Lin Wang, Kailin Yang, Fengjiang Liu, Rendi Jiang, Xinglou Yang, Tian You, Xiaoce Liu, Xiuna Yang, Fang Bai, Hong Liu, Xiang Liu, Luke W. Guddat, Wenqing Xu, Gengfu Xiao, Chengfeng Qin, Zhengli Shi, Hualiang Jiang, Zihe Rao, and Haitao Yang. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature*, 582(7811):289–293, 2020.
- [11] Sven Ullrich and Christoph Nitsche. The SARS-CoV-2 main protease as drug target. *Bioorganic And Medicinal Chemistry Letters*, 30(17):127377, 2020.
- [12] Katarzyna Świderek and Vicent Moliner. Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 Mpro by QM/MM computational methods. *Chemical Science*, 11(39):10626–10630, 2020.
- [13] Jaeyong Lee, Liam J. Worrall, Marija Vuckovic, Federico I. Rosell, Francesco Gentile, Anh-Tien Ton, Nathanael A. Caveney, Fuqiang Ban, Artem Cherkasov, Mark Paetzel, and Natalie C. J. Strynadka. Crystallographic structure of wild-type SARS-CoV-2 main protease acyl-enzyme intermediate with physiological C-terminal autoprocessing site. *Nature Communications*, 11(1):5877, 2020.
- [14] Elizabeth A. MacDonald, Gary Frey, Mark N. Namchuk, Stephen C. Harrison, Stephen M. Hinshaw, and Ian W. Windsor. Recognition of divergent viral substrates by the SARS-CoV-2 main protease. *ACS Infectious Diseases*, 7(9):2591–2595, 2021.
- [15] Manju Narwal, Jean-Paul Armache, Thomas J Edwards, and Katsuhiko S Murakami. SARS-CoV-2 polyprotein substrate regulates the stepwise Mpro cleavage reaction. *Journal of Biological Chemistry*, page 104697, 2023.
- [16] Yao Zhao, Yan Zhu, Xiang Liu, Zhenming Jin, Yinkai Duan, Qi Zhang, Chengyao Wu, Lu Feng, Xiaoyu Du, Jinyi Zhao, Maolin Shao, Bing Zhang, Xiuna Yang, Lijie Wu, Xiaoyun Ji, Luke W Guddat, Kailin Yang, Zihe Rao, and Haitao Yang. Structural basis for replicase polyprotein cleavage and substrate specificity of main protease from SARS-CoV-2. *Proceedings of the National Academy of Sciences of the United States of America*, 119(16):e2117142119, 2022.
- [17] Haitao Yang, Weiqing Xie, Xiaoyu Xue, Kailin Yang, Jing Ma, Wenxue Liang, Qi Zhao, Zhe Zhou, Duanqing Pei, John Ziebuhr, Rolf Hilgenfeld, Kwok Yung Yuen, Luet Wong, Guangxia Gao, Saijuan Chen, Zhu Chen, Dawei Ma, Mark Bartlam, and Zihe Rao. Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLOS Biology*, 3(10):e324, 2005.
- [18] Rolf Hilgenfeld. From SARS to MERS: crystallographic studies on coronavirus proteases enable antiviral drug design. *The FEBS Journal*, 281(18):4085–4096, 2014.
- [19] Linlin Zhang, Daizong Lin, Xinyuanyuan Sun, Ute Curth, Christian Drost, Lucie Sauerhering, Stephan Becker, Katharina Rox, and Rolf Hilgenfeld. Crystal structure of SARS-CoV-2 main protease provides a basis for

- design of improved α -ketoamide inhibitors. *Science*, 368(6489):409–412, 2020.
- [20] Linlin Zhang, Daizong Lin, Yuri Kusov, Yong Nian, Qingjun Ma, Ji-ang Wang, Albrecht von Brunn, Pieter Leyssen, Kristina Lanko, Johan Neyts, Adriaan de Wilde, Eric J. Snijder, Hong Liu, and Rolf Hilgenfeld. α -ketoamides as broad-spectrum inhibitors of coronavirus and enterovirus replication: Structure-based design, synthesis, and activity assessment. *Journal of Medicinal Chemistry*, 63(9):4562–4578, 2020.
- [21] Kanchan Anand, John Ziebuhr, Parvesh Wadhvani, Jeroen R Mesters, and Rolf Hilgenfeld. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science*, 300(5626):1763–1767, 2003.
- [22] Robert L. Hoffman, Robert S. Kania, Mary A. Brothers, Jay F. Davies, Rose A. Ferre, Ketan S. Gajiwala, Mingying He, Robert J. Hogan, Kirk Kozminski, Lilian Y. Li, Jonathan W. Lockner, Jihong Lou, Michelle T. Marra, Lennert J. Jr. Mitchell, Brion W. Murray, James A. Nieman, Stephen Noell, Simon P. Planken, Thomas Rowe, Kevin Ryan, George J. III Smith, James E. Solowiej, Claire M. Steppan, and Barbara Taggart. Discovery of ketone-based covalent inhibitors of coronavirus 3CL proteases for the potential therapeutic treatment of COVID-19. *Journal of Medicinal Chemistry*, 63(21):12725–12747, 2020.
- [23] Dafydd R Owen, Charlotte M N Allerton, Annaliesa S Anderson, Lisa Aschenbrenner, Melissa Avery, Simon Berritt, Britton Boras, Rhonda D Cardin, Anthony Carlo, Karen J Coffman, Alyssa Dantonio, Li Di, Heather Eng, Roseann Ferre, Ketan S Gajiwala, Scott A Gibson, Samantha E Greasley, Brett L Hurst, Eugene P Kadar, Amit S Kalgutkar, Jack C Lee, Jisun Lee, Wei Liu, Stephen W Mason, Stephen Noell, Jonathan J Novak, R Scott Obach, Kevin Ogilvie, Nandini C Patel, Martin Pettersson, Devendra K Rai, Matthew R Reese, Matthew F Sammons, Jean G Sathish, Ravi Shankar P Singh, Claire M Steppan, Al E Stewart, Jamison B Tuttle, Lawrence Updyke, Patrick R Verhoest, Liuqing Wei, Qingyi Yang, and Yuao Zhu. An oral SARS-CoV-2 mpro inhibitor clinical candidate for the treatment of COVID-19. *Science*, 374(6575):1586–1593, 2021.
- [24] Daniel W. Kneller, Hui Li, Gwyndalyn Phillips, Kevin L. Weiss, Qiu Zhang, Mark A. Arnould, Colleen B. Jonsson, Surekha Surendranathan, Jyothi Parvathareddy, Matthew P. Blakeley, Leighton Coates, John M. Louis, Peter V. Bonnesen, and Andrey Kovalevsky. Covalent narpaprevir- and boceprevir-derived hybrid inhibitors of SARS-CoV-2 main protease. *Nature Communications*, 13(1):2268, 2022.
- [25] Yao Zhao, Chao Fang, Qi Zhang, Ruxue Zhang, Xiangbo Zhao, Yinkai Duan, Haofeng Wang, Yan Zhu, Lu Feng, Jinyi Zhao, Maolin Shao, Xiuna Yang, Leike Zhang, Chao Peng, Kailin Yang, Dawei Ma, Zihe Rao, and Haitao Yang. Crystal structure of SARS-CoV-2 main protease in complex with protease inhibitor PF-07321332. *Protein Cell*, 13(9):689–693, 2022.

- [26] Rana Abdelnabi, Dirk Jochmans, Kim Donckers, Bettina Trüeb, Nadine Ebert, Birgit Weynand, Volker Thiel, and Johan Neyts. Nirmatrelvir-resistant SARS-CoV-2 is efficiently transmitted in female syrian hamsters and retains partial susceptibility to treatment. *Nature Communications*, 14(1):2124, 2023.
- [27] Dirk Jochmans, Cheng Liu, Kim Donckers, Antitsa Stoycheva, Sandro Bolland, Sarah K Stevens, Chloe De Vita, Bert Vanmechelen, Piet Maes, Bettina Trüeb, Nadine Ebert, Volker Thiel, Steven De Jonghe, Laura Vangeel, Dorothée Bardiot, Andreas Jekle, Lawrence M Blatt, Leonid Beigelman, Julian A Symons, Pierre Raboisson, Patrick Chaltin, Arnaud Marchand, Johan Neyts, Jerome Deval, and Koen Vandyck. The substitutions L50F, E166A, and L167F in SARS-CoV-2 3CLpro are selected by a protease inhibitor in vitro and confer resistance to nirmatrelvir. *MBio*, 14(1):e0281522, 2023.
- [28] Guangdi Li, Rolf Hilgenfeld, Richard Whitley, and Erik De Clercq. Therapeutic strategies for COVID-19: progress and lessons learned. *Nature Reviews Drug Discovery*, 2023.
- [29] John Chodera, Alpha A. Lee, Nir London, and Frank von Delft. Crowdsourcing drug discovery for pandemics. *Nature Chemistry*, 12(7):581–581, 2020.
- [30] Aayush Gupta and Huan-Xiang Zhou. Profiling SARS-CoV-2 main protease (Mpro) binding to repurposed drugs using molecular dynamics simulations in classical and neural network-trained force fields. *ACS Combinatorial Science*, 22(12):826–832, 2020.
- [31] Son Tung Ngo, Ngoc Quynh Anh Pham, Ly Thi Le, Duc-Hung Pham, and Van V. Vu. Computational determination of potential inhibitors of SARS-CoV-2 main protease. *Journal of Chemical Information and Modeling*, 60(12):5771–5780, 2020.
- [32] Son Tung Ngo, Hung Minh Nguyen, Le Thi Thuy Huong, Pham Minh Quan, Vi Khanh Truong, Nguyen Thanh Tung, and Van V Vu. Assessing potential inhibitors of SARS-CoV-2 main protease from available drugs using free energy perturbation simulations. *RSC Advances*, 10(66):40284–40290, 2020.
- [33] Son Tung Ngo, Nguyen Minh Tam, Minh Quan Pham, and Trung Hai Nguyen. Benchmark of popular free energy approaches revealing the inhibitors binding to SARS-CoV-2 Mpro. *Journal of Chemical Information and Modeling*, 61(5):2302–2312, 2021.
- [34] T Ngoc Han Pham, Trung Hai Nguyen, Nguyen Minh Tam, Thien Y Vu, Nhat Truong Pham, Nguyen Truong Huy, Binh Khanh Mai, Nguyen Thanh Tung, Minh Quan Pham, Van Vu, V, and Son Tung Ngo. Improving ligand-ranking of AutoDock Vina by changing the empirical parameters. *Journal of Computational Chemistry*, 43(3):160–169, 2022.

- [35] José P Cerón-Carrasco. When virtual screening yields inactive drugs: Dealing with false theoretical friends. *ChemMedChem*, 17(16):e202200278, 2022.
- [36] Christoph Gorgulla, Krishna M Padmanabha Das, Kendra E Leigh, Marco Cespugli, Patrick D Fischer, Zi-Fu Wang, Guilhem Tesseyre, Shreya Pandita, Alec Shnapir, Anthony Calderaio, Minko Gechev, Alexander Rose, Noam Lewis, Colin Hutcheson, Erez Yaffe, Roni Luxenburg, Henry D Herce, Vedat Durmaz, Thanos D Halazonetis, Konstantin Fackeldey, J J Patten, Alexander Chuprina, Igor Dziuba, Alla Plekhova, Yurii Moroz, Dmytro Radchenko, Olga Tarkhanova, Irina Yavnyuk, Christian Gruber, Ryan Yust, Dave Payne, Anders M Näär, Mark N Namchuk, Robert A Davey, Gerhard Wagner, Jamie Kinney, and Haribabu Arthanari. A multi-pronged approach targeting SARS-CoV-2 proteins using ultra-large virtual screening. *iScience*, 24(2):102021, 2021.
- [37] Anh-Tien Ton, Francesco Gentile, Michael Hsing, Fuqiang Ban, and Artem Cherkasov. Rapid identification of potential inhibitors of SARS-CoV-2 main protease by deep docking of 1.3 billion compounds. *Molecular Informatics*, 39(8):e2000028, 2020.
- [38] Andreas Lutten, Hjalmar Gullberg, Eldar Abdurakhmanov, Duy Duc Vo, Dario Akaberi, Vladimir O. Talibov, Natalia Nekhotiaeva, Laura Vangeel, Steven De Jonghe, Dirk Jochmans, Janina Krambrich, Ali Tas, Bo Lundgren, Ylva Gravenfors, Alexander J. Craig, Yoseph Atilaw, Anja Sandström, Lindon W. K. Moodie, Åke Lundkvist, Martijn J. van Hemert, Johan Neyts, Johan Lennerstrand, Jan Kihlberg, Kristian Sandberg, U. Helena Danielson, and Jens Carlsson. Ultralarge virtual screening identifies SARS-CoV-2 main protease inhibitors with broad-spectrum activity against coronaviruses. *Journal of the American Chemical Society*, 144(7):2905–2920, 2022.
- [39] The COVID Moonshot Consortium. Open science discovery of potent non-covalent SARS-CoV-2 main protease inhibitors. *bioRxiv*, 2023.
- [40] Piero Procacci, Marina Macchiagodena, Marco Pagliai, Guido Guarnieri, and Francesco Iannone. Interaction of hydroxychloroquine with SARS-CoV2 functional proteins using all-atoms non-equilibrium alchemical simulations. *Chemical Communications (Cambridge)*, 56(62):8854–8856, 2020.
- [41] Marina Macchiagodena, Marco Pagliai, and Piero Procacci. Identification of potential binders of the main protease 3CLpro of the COVID-19 via structure-based ligand design and molecular modeling. *Chemical Physics Letters*, 750(137489):137489, 2020.
- [42] Marina Macchiagodena, Marco Pagliai, Maurice Karrenbrock, Guido Guarnieri, Francesco Iannone, and Piero Procacci. Virtual double-system single-box: A nonequilibrium alchemical technique for absolute binding free energy calculations: Application to ligands of the SARS-CoV-2 main protease. *Journal of Chemical Theory and Computation*, 16(11):7160–7172, 2020.

- [43] Marina Macchiagodena, Maurice Karrenbrock, Marco Pagliai, and Piero Procacci. Virtual double-system single-box for absolute dissociation free energy calculations in GROMACS. *Journal of Chemical Information and Modeling*, 61(11):5320–5326, 2021.
- [44] Marina Macchiagodena, Marco Pagliai, and Piero Procacci. Characterization of the non-covalent interaction between the PF-07321332 inhibitor and the SARS-CoV-2 main protease. *Journal of Molecular Graphics and Modelling*, 110(108042):108042, 2022.
- [45] Scott A Hollingsworth and Ron O Dror. Molecular dynamics simulation for all. *Neuron*, 99(6):1129–1143, 2018.
- [46] S. Goel, M. Knaggs, G. Goel, X.W. Zhou, H.M. Upadhyaya, V.K. Thakur, V. Kumar, G. Bizarri, A. Tiwari, A. Murphy, A. Stukowski, and A. Matthews. Horizons of modern molecular dynamics simulation in digitalized solid freeform fabrication with advanced materials. *Materials Today Chemistry*, 18:100356, 2020.
- [47] Jacob D. Durrant and J. Andrew McCammon. Molecular dynamics simulations and drug discovery. *BMC Biology*, 9(1):71, 2011.
- [48] Sereina Riniker. Fixed-charge atomistic force fields for molecular dynamics simulations in the condensed phase: An overview. *Journal of Chemical Information and Modeling*, 58(3):565–578, 2018.
- [49] Chetan R. Rupakheti, Alexander D. Jr. MacKerell, and Benoît Roux. Global optimization of the Lennard-Jones parameters for the Drude polarizable force field. *Journal of Chemical Theory and Computation*, 17(11):7085–7095, 2021.
- [50] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.
- [51] B R Brooks, C L Brooks, 3rd, A D Mackerell, Jr, L Nilsson, R J Petrella, B Roux, Y Won, G Archontis, C Bartels, S Boresch, A Caffisch, L Caves, Q Cui, A R Dinner, M Feig, S Fischer, J Gao, M Hodoseck, W Im, K Kuczera, T Lazaridis, J Ma, V Ovchinnikov, E Paci, R W Pastor, C B Post, J Z Pu, M Schaefer, B Tidor, R M Venable, H L Woodcock, X Wu, W Yang, D M York, and M Karplus. CHARMM: the biomolecular simulation program. *Journal of Computational Chemistry*, 30(10):1545–1614, 2009.
- [52] K Vanommeslaeghe, E Hatcher, C Acharya, S Kundu, S Zhong, J Shim, E Darian, O Guvench, P Lopes, I Vorobyov, and A D Mackerell, Jr. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry*, 31(4):671–690, 2010.

- [53] K. Vanommeslaeghe and A. D. Jr. MacKerell. Automation of the CHARMM general force field (CGenFF) i: Bond perception and atom typing. *Journal of Chemical Information and Modeling*, 52(12):3144–3154, 2012.
- [54] K Vanommeslaeghe, E Prabhu Raman, and A D MacKerell, Jr. Automation of the CHARMM general force field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *Journal of Chemical Information and Modeling*, 52(12):3155–3168, 2012.
- [55] Mark Tuckerman. *Statistical mechanics: Theory and molecular simulation*. Oxford Graduate Texts. Oxford University Press, London, England, 2010.
- [56] James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kalé, and Klaus Schulten. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16):1781–1802, 2005.
- [57] Paraskevi Gkeka. *Molecular dynamics studies of peptide-membrane interactions: insights from coarse-grained models*. PhD thesis, 01 2010.
- [58] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, 1993.
- [59] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, and Lee G. Pedersen. A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, 103(19):8577–8593, 1995.
- [60] Jacob Chapman. *Improving the Functional Control of Ferroelectrics using Insights from Atomistic modelling*. PhD thesis, 09 2018.
- [61] Iupinov Alexei. Implementation of the particle mesh ewald method on a gpu. Master’s thesis, 2016.
- [62] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman J.C Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3):327–341, 1977.
- [63] Hans C Andersen. Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics*, 52(1):24–34, 1983.
- [64] Shuichi Miyamoto and Peter A. Kollman. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*, 13(8):952–962, 1992.
- [65] Jostein Blyverket. *Molecular dynamics modeling of clay-fuild interfaces*. PhD thesis, 2015.

- [66] Efrem Braun, Seyed Mohamad Moosavi, and Berend Smit. Anomalous effects of velocity rescaling algorithms: The flying ice cube effect revisited. *Journal of Chemical Theory and Computation*, 14(10):5262–5272, 2018.
- [67] D. J. Evans and B. L. Holian. The Nose–Hoover thermostat. *The Journal of Chemical Physics*, 83(8):4069–4074, 1985.
- [68] E. A. Koopman and C. P. Lowe. Advantages of a Lowe-Andersen thermostat in molecular dynamics simulations. *The Journal of Chemical Physics*, 124(20):204103, 2006.
- [69] D. Quigley and M. I. J. Probert. Langevin dynamics in constant pressure extended systems. *The Journal of Chemical Physics*, 120(24):11432–11441, 2004.
- [70] Herbert B Callen. *Thermodynamics and an Introduction to Thermostatistics*. John Wiley & Sons, Nashville, TN, 2 edition, 1985.
- [71] Vittorio Limongelli. Ligand binding free energy and kinetics calculation in 2020. *WIREs Computational Molecular Science*, 10, 2020.
- [72] Albert C. Pan, Huafeng Xu, Timothy Palpant, and David E. Shaw. Quantitative characterization of the binding and unbinding of millimolar drug fragments with molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 13(7):3372–3377, 2017.
- [73] William L. Jorgensen, J. Kathleen Buckner, Stephane Boudon, and Julian Tirado-Rives. Efficient computation of absolute free energies of binding by computer simulations. Application to the methane dimer in water. *The Journal of Chemical Physics*, 89(6):3742–3746, 1988.
- [74] Stefan Boresch, Franz Tettinger, Martin Leitgeb, and Martin Karplus. Absolute binding free energies: A quantitative approach for their calculation. *The Journal of Physical Chemistry B*, 107(35):9535–9551, 2003.
- [75] James C. Gumbart, Benoît Roux, and Christophe Chipot. Standard binding free energies from computer simulations: What is the best strategy? *Journal of Chemical Theory and Computation*, 9(1):794–802, 2013.
- [76] David A. Pearlman. A comparison of alternative approaches to free energy calculations. *The Journal of Physical Chemistry*, 98(5):1487–1493, 1994.
- [77] Thomas C. Beutler, Alan E. Mark, René C. van Schaik, Paul R. Gerber, and Wilfred F. van Gunsteren. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chemical Physics Letters*, 222(6):529–539, 1994.
- [78] Robert W. Zwanzig. High-Temperature Equation of State by a Perturbation Method. II. Polar Gases. *The Journal of Chemical Physics*, 23(10):1915–1922, 1955.

- [79] Surjit B. Dixit, Christophe Chipot, Floris Buelens, and Christopher Harrison. Alchemical free energy methods @ONLINE. <https://www.ks.uiuc.edu>, 2023.
- [80] John G. Kirkwood. Statistical Mechanics of Fluid Mixtures. *The Journal of Chemical Physics*, 3(5):300–313, 1935.
- [81] W. K. den Otter. Thermodynamic integration of the free energy along a reaction coordinate in cartesian coordinates. *The Journal of Chemical Physics*, 112(17):7283–7292, 2000.
- [82] C. Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78:2690–2693, 1997.
- [83] C. Jarzynski. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Physical Review E*, 56:5018–5035, 1997.
- [84] C. Jarzynski. Equilibrium free energies from nonequilibrium processes. *Acta Physica Polonica Series B*, 29:1609–1622, 1998.
- [85] D. A. Hendrix and C. Jarzynski. A fast growth method of computing free energy differences. *The Journal of Chemical Physics*, 114(14):5974–5981, 2001.
- [86] C. Jarzynski. Equilibrium and nonequilibrium foundations of free energy computational methods. In *Computational Methods for Macromolecules: Challenges and Applications*, pages 287–303. Springer Berlin Heidelberg, 2002.
- [87] Gavin E. Crooks. Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems. *Journal of Statistical Physics*, 90(5):1481–1487, 1998.
- [88] Gavin E. Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical Review E*, 60:2721–2726, 1999.
- [89] Piero Procacci and Chiara Cardelli. Fast switching alchemical transformations in molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 10(7):2813–2823, 2014.
- [90] Bert de Groot. Non-equilibrium free energy calculation @ONLINE. <https://www3.mpibpc.mpg.de>, 2023.
- [91] G Hummer, L R Pratt, and A E Garcia. Free energy of ionic hydration. *Journal of Physical Chemistry*, 100(4), 1 1996.
- [92] Jiyao Wang, Yuqing Deng, and Benoît Roux. Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophysical Journal*, 91(8):2798–2814, 2006.

- [93] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.
- [94] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Greenblatt, Elaine C Meng, and Thomas E Ferrin. UCSF chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004.
- [95] B. Cornell. Amino acids @ONLINE. <http://ib.bioninja.com.au>, 2016.
- [96] Sunhwan Jo, Taehoon Kim, Vidyashankara G Iyer, and Wonpil Im. CHARMM-GUI: a web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, 29(11):1859–1865, 2008.
- [97] Michael R. Shirts, Jed W. Pitera, William C. Swope, and Vijay S. Pande. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *The Journal of Chemical Physics*, 119(11):5740–5761, 2003.
- [98] Van A. Ngo, John Keenan Fanning, and Sergei Yu Noskov. Comparative analysis of protein hydration from MD simulations with additive and polarizable force fields. *Advanced Theory and Simulations*, 2(2):1800106, 2019.
- [99] R Wolfenden, L Andersson, P M Cullis, and C C Southgate. Affinities of amino acid side chains for solvent water. *Biochemistry*, 20(4):849–855, 1981.
- [100] Michael R. Shirts, David L. Mobley, John D. Chodera, and Vijay S. Pande. Accurate and efficient corrections for missing dispersion interactions in molecular simulations. *The Journal of Physical Chemistry B*, 111(45):13052–13063, 2007.
- [101] Christopher G Mayne, Jan Saam, Klaus Schulten, Emad Tajkhorshid, and James C Gumbart. Rapid parameterization of small molecules using the Force Field Toolkit. *Journal of Computational Chemistry*, 34(32):2757–2770, 2013.
- [102] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. Gaussian~16 Revision C.01, 2016. Gaussian Inc. Wallingford CT.

- [103] Haochuan Chen, Julio D C Maia, Brian K Radak, David J Hardy, Wensheng Cai, Christophe Chipot, and Emad Tajkhorshid. Boosting free-energy perturbation calculations with GPU-accelerated NAMD. *Journal of Chemical Information and Modeling*, 60(11):5301–5307, November 2020.
- [104] Gabriel Michaux and Jacques Reisse. Solution thermodynamic studies. part 6. enthalpy-entropy compensation for the complexation reactions of some crown ethers with alkaline cations: a quantitative interpretation of the complexing properties of 18-crown-6. *Journal of the American Chemical Society*, 104(25):6895–6899, 1982.
- [105] Robert B. Sandberg, Martina Banchelli, Carlo Guardiani, Stefano Menichetti, Gabriella Caminati, and Piero Procacci. Efficient nonequilibrium method for binding free energy calculations in molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 11(2):423–435, 2015.
- [106] Dafydd R Owen, Charlotte M N Allerton, Annaliesa S Anderson, Lisa Aschenbrenner, Melissa Avery, Simon Berritt, Britton Boras, Rhonda D Cardin, Anthony Carlo, Karen J Coffman, Alyssa Dantonio, Li Di, Heather Eng, Roseann Ferre, Ketan S Gajiwala, Scott A Gibson, Samantha E Greasley, Brett L Hurst, Eugene P Kadar, Amit S Kalgutkar, Jack C Lee, Jisun Lee, Wei Liu, Stephen W Mason, Stephen Noell, Jonathan J Novak, R Scott Obach, Kevin Ogilvie, Nandini C Patel, Martin Pettersson, Devendra K Rai, Matthew R Reese, Matthew F Sammons, Jean G Sathish, Ravi Shankar P Singh, Claire M Steppan, Al E Stewart, Jamison B Tuttle, Lawrence Updyke, Patrick R Verhoest, Liuqing Wei, Qingyi Yang, and Yuao Zhu. An oral SARS-CoV-2 Mpro inhibitor clinical candidate for the treatment of COVID-19. *Science*, 374(6575):1586–1593, 2021.
- [107] Yuhang Wang, Sruthi Murlidaran, and David A Pearlman. Quantum simulations of SARS-CoV-2 main protease Mpro enable high-quality scoring of diverse ligands. *Journal of Computer-Aided Molecular Design*, 35(9):963–971, 2021.

A. Přílohy

A.1 Tcl skript pro přípravu systémů

Příklad Tcl skriptu, který byl využit v Tk konzoli VMD programu pro přípravu počátečních souřadnic komplexů adamantánů s cyklodextriny. Vycházel jsem zde ze souřadnic samotných cyklodextrinů a samotných adamantánů a zkombinoval jsem je dohromady s různou počáteční orientací. Pro každou kombinaci cyklodextrinu a adamantánu bylo připraveno 10 různých počátečních orientací adamantánu vůči cyklodextrinu a dvě různé orientace nabitě amino-skupiny adamantánu.

```
1 package require Orient
2 namespace import Orient::orient
3 package require psfgen
4 package require autoionize
5
6 # list of all cyklodextins and ligands
7 set cyklodextrin_list { alpha beta gamma }
8 set ligand_list { 1 2 3 4 5 6 }
9 set orientation_list { 1 2 3 4 5 6 7 8 9 10 }
10 set pi 3.1415926535897931
11
12 # prepare complex structures for different initial orientations
13 foreach cyklodextrin $cyklodextrin_list {
14   foreach ligand $ligand_list {
15     foreach i $orientation_list {
16
17       cd ${cyklodextrin}_${ligand}
18       puts "pdb for $cyklodextrin, ligand $ligand and orientation $i\n\n"
19
20       # load ligand and cyklodextrin structures
21       mol new ../../01_mol2_files/${cyklodextrin}_cyklodextrin.mol2
22       mol new ../../01_mol2_files/ligand_${ligand}.mol2
23
24       set top_id [expr [molinfo top get id] ]
25       set below_top_id [expr [molinfo top get id] - 1 ]
26
27       # selections
28       set sel_glc [atomselect $below_top_id "resname GLC"]
29       set center_glc [measure center $sel_glc weight [$sel_glc get mass]]
30       set sel_lig [atomselect $top_id "resname LIG"]
31       set center_lig [measure center $sel_lig weight [$sel_lig get mass]]
32
33       # translate both to center
34       set transfo [transoffset [vecinvert $center_glc]]
35       $sel_glc move $transfo
36       set transfo [transoffset [vecinvert $center_lig]]
37       $sel_lig move $transfo
38
39       # rotate molecules
40       set unit_vec_y [expr cos($i * $pi / 5)]
41       set unit_vec_z [expr sin($i * $pi / 5)]
42
43       set I [draw principalaxes $sel_glc]
44       set A [orient $sel_glc [lindex $I 0] {1 0 0}]
```

```

45 $sel_glc move $A
46 set I [draw principalaxes $sel_glc]
47 set A [orient $sel_glc [lindex $I 1] {0 1 0}]
48 $sel_glc move $A
49 set I [draw principalaxes $sel_glc]
50
51 set I [draw principalaxes $sel_lig]
52 set A [orient $sel_lig [lindex $I 2] {1 0 0}]
53 $sel_lig move $A
54 set I [draw principalaxes $sel_lig]
55 set A [orient $sel_lig [lindex $I 1] [list 0 [expr $unit_vec_y] [expr
↪ $unit_vec_z]]]
56 $sel_lig move $A
57 set I [draw principalaxes $sel_lig]
58
59 # translate both to center
60 set sel_glc [atomselect $below_top_id "resname GLC"]
61 set center_glc [measure center $sel_glc weight [$sel_glc get mass]]
62 set sel_lig [atomselect $top_id "resname LIG"]
63 set center_lig [measure center $sel_lig weight [$sel_lig get mass]]
64 set transfo [transoffset [vecinvert $center_glc]]
65 $sel_glc move $transfo
66 set transfo [transoffset [vecinvert $center_lig]]
67 $sel_lig move $transfo
68
69 # create pdb files with complex structures
70 # ligand only
71 $sel_lig writepdb "lig_only_$i.pdb"
72 # cyclodextrin only
73 $sel_glc writepdb "glc_only_$i.pdb"
74
75 # reverse orientation
76 set reverse_matrix [transaxis z 180]
77 $sel_lig move $reverse_matrix
78
79 # translate both to center
80 set sel_glc [atomselect $below_top_id "resname GLC"]
81 set center_glc [measure center $sel_glc weight [$sel_glc get mass]]
82 set sel_lig [atomselect $top_id "resname LIG"]
83 set center_lig [measure center $sel_lig weight [$sel_lig get mass]]
84 set transfo [transoffset [vecinvert $center_glc]]
85 $sel_glc move $transfo
86 set transfo [transoffset [vecinvert $center_lig]]
87 $sel_lig move $transfo
88
89 # create pdb files with complex structures
90 # ligand only
91 $sel_lig writepdb "lig_only_reverse_$i.pdb"
92 # cyclodextrin only
93 $sel_glc writepdb "glc_only_reverse_$i.pdb"
94
95
96 mol delete $below_top_id
97 mol delete $top_id
98 cd ..
99
100 }}}

```

A.2 Tcl skript pro tvorbu topologie, solvataci a ionizaci

Ukázkový Tcl skript, který bral počáteční souřadnice komplexů adamantánů a cyklodextrinů a vytvořil k nim odpovídající topologické psf soubory, vygeneroval vodní obálku, přidal ionty pro neutralizaci systému, vytvořil .xst soubory s informacemi o periodických okrajových podmínkách a vytvořil .fep soubory, které určovaly, které atomy budou během alchymistické transformace mizet.

```
1 package require Orient
2 namespace import Orient::orient
3 package require psfgen
4 package require autoionize
5
6 # list of all cyklodextrins and ligands
7 set cyklodextrin_list { alpha beta gamma }
8 set ligand_list { 1 2 3 4 5 6 }
9 set orientation_list { 1 2 3 4 5 6 7 8 9 10 }
10 set pi 3.1415926535897931
11
12 # psf generation for complexes
13 foreach cyklodextrin $cyklodextrin_list {
14   foreach ligand $ligand_list {
15     foreach i $orientation_list {
16
17       cd ${cyklodextrin}_${ligand}
18       puts "psf for $cyklodextrin, ligand $ligand and orientation $i"
19
20       # load topology files
21       topology ../../02_str_files/CHARMM/top_all136_cgenff.rtf
22       topology ../../02_str_files/${cyklodextrin}_cyklodextrin.str
23       topology ../../02_str_files/ligand_${ligand}.str
24
25
26       # get initial position from pdb files
27       segment GLC {
28         pdb "glc_only_${i}.pdb"
29       }
30       coordpdb "glc_only_${i}.pdb" GLC
31
32       segment LIG {
33         pdb "lig_only_${i}.pdb"
34       }
35       coordpdb "lig_only_${i}.pdb" LIG
36
37       # create psf files
38       writepsf "complex_only_${i}.psf"
39       writepdb "complex_only_${i}.pdb"
40
41       resetpsf
42       psfcontext reset
43
44       # reverse orientation
45       # topology files
46       topology ../../02_str_files/CHARMM/top_all136_cgenff.rtf
47       topology ../../02_str_files/${cyklodextrin}_cyklodextrin.str
48       topology ../../02_str_files/ligand_${ligand}.str
```



```

49
50 # get initial position from pdb files
51 segment GLC {
52     pdb "glc_only_reverse_${i}.pdb"
53 }
54 coordpdb "glc_only_reverse_${i}.pdb" GLC
55
56 segment LIG {
57     pdb "lig_only_reverse_${i}.pdb"
58 }
59 coordpdb "lig_only_reverse_${i}.pdb" LIG
60
61 # create psf files
62 writepsf "complex_only_reverse_${i}.psf"
63 writepdb "complex_only_reverse_${i}.pdb"
64
65 resetpsf
66 psfcontext reset
67
68 # ligand only pdf files
69 # topology
70 topology ../../02_str_files/CHARMM/top_all136_cgenff.rtf
71 topology ../../02_str_files/ligand_${ligand}.str
72
73 # get initial position from pdb files
74 segment LIG {
75     pdb "lig_only.pdb"
76 }
77 coordpdb "lig_only.pdb" LIG
78
79 writepsf "ligand_only.psf"
80 writepdb "ligand_only.pdb"
81
82 # create psf files
83 resetpsf
84 psfcontext reset
85
86 cd ..
87
88 }}}
89
90
91 # solvation
92 foreach cyklodextrin $cyklodextrin_list {
93     foreach ligand $ligand_list {
94         foreach i $orientation_list {
95
96             cd ${cyklodextrin}_${ligand}
97             puts "solvation for $cyklodextrin, ligand $ligand and orientation $i"
98
99             # solvate complex
100            solvate "complex_only_${i}.psf" "complex_only_${i}.pdb" -o "solvate_${i}" -s WAT
            ↪ -rotate -rotsel { all } -rotinc 10 -x 15 -y 15 -z 15 +x 15 +y 15 +z 15 -b
            ↪ 2.4
101            mol delete top
102            mol delete top
103
104            # solvate complex, reverse orientation

```

```

105 solvate "complex_only_reverse_${i}.psf" "complex_only_reverse_${i}.pdb" -o
    ↪ "reverse_solvate_${i}" -s WAT -rotate -rotsel { all } -rotinc 10 -x 15 -y 15
    ↪ -z 15 +x 15 +y 15 +z 15 -b 2.4
106 mol delete top
107 mol delete top
108
109 # solvate ligand only
110 solvate "ligand_only.psf" "ligand_only.pdb" -o "solvate_ligand" -s WAT -rotate
    ↪ -rotsel { all } -rotinc 10 -x 15 -y 15 -z 15 +x 15 +y 15 +z 15 -b 2.4
111 mol delete top
112 mol delete top
113
114 cd ..
115
116 }}}
117
118 # autoionization, creation of the .fep file and .xst file
119 foreach cyklodextrin $cyklodextrin_list {
120   foreach ligand $ligand_list {
121     foreach i $orientation_list {
122
123       cd ${cyklodextrin}_${ligand}
124       puts "autoionization for $cyklodextrin, ligand $ligand and orientation $i"
125
126       # complex ionization
127       autoionize -psf "solvate_${i}.psf" -pdb "solvate_${i}.pdb" -sc 0.15 -o
        ↪ "ionized_${i}"
128       set ligand_sel [atomselect top "resname LIG"]
129       $ligand_sel set beta -1
130       set all_sel [atomselect top all]
131       $all_sel writepdb "ionized_${i}.fep"
132       pbc writextst "ionized_${i}.xst"
133       mol delete top
134
135       # complex ionization, reverse oientation
136       autoionize -psf "reverse_solvate_${i}.psf" -pdb "reverse_solvate_${i}.pdb" -sc
        ↪ 0.15 -o "reverse_ionized_${i}"
137       set ligand_sel [atomselect top "resname LIG"]
138       $ligand_sel set beta -1
139       set all_sel [atomselect top all]
140       $all_sel writepdb "reverse_ionized_${i}.fep"
141       pbc writextst "reverse_ionized_${i}.xst"
142       mol delete top
143
144       # ligand only ionization
145       autoionize -psf "solvate_ligand.psf" -pdb "solvate_ligand.pdb" -sc 0.15 -o
        ↪ "ionized_ligand"
146       set ligand_sel [atomselect top "resname LIG"]
147       $ligand_sel set beta -1
148       set all_sel [atomselect top all]
149       $all_sel writepdb "ionized_ligand.fep"
150       pbc writextst "ionized_ligand.xst"
151       mol delete top
152
153       cd ..
154
155     }}}

```

A.3 Konfigurační soubor programu NAMD

Ukázkový konfigurační soubor, který slouží jako vstup do programu NAMD ke spuštění alchymistických MD transformací kovalentních inhibitorů proteázy Mpro.

```
1 #####
2 # Free Energy Perturbation for ligand 6
3 #####
4
5 # MD SETUP
6
7 timestep                2.0
8
9 # FLEXIBLE CELL
10
11 useflexiblecell        no
12
13 # INPUT
14
15 structure                ../common/QwikMD.psf
16 coordinates              ../common/QwikMD.pdb
17 binCoordinates           ../common/MD_2.restart.coor
18 binVelocities            ../common/MD_2.restart.vel
19 extendedSystem           ../common/MD_2.restart.xsc
20
21 # Force Field Parameters
22
23 paraTypeCharmm           on
24 parameters ../common/toppar_water_ions_namd.str
25 parameters ../common/toppar_all36_carb_glycopeptide.str
26 parameters ../common/par_all36_lipid.prm
27 parameters ../common/par_all36_na.prm
28 parameters ../common/par_all36_prot.prm
29 parameters ../common/par_all36_carb.prm
30 parameters ../common/par_all36_cgenff.prm
31 parameters ../common/ligand1_qwikmd.str
32 parameters ../common/ligand3_qwikmd.str
33 parameters ../common/ligand4_qwikmd.str
34 parameters ../common/ligand5_qwikmd.str
35 parameters ../common/ligand2_qwikmd.str
36 parameters ../common/ligand6_qwikmd.str
37
38 # OUTPUT
39
40 binaryoutput             no
41 outputname               forward
42 outputenergies           2000
43 outputtiming              2000
44 outputpressure           2000
45
46 dcdfile                  forward.dcd
47 dcdfreq                  2000
48 XSTFreq                  2000
49
50 binaryrestart             yes
51 restartfreq              2000
52 restartname              forward.restart
```

```

53
54 # PME
55
56 PME                yes
57 PMEGridSpacing    1
58
59 # WRAP WATER FOR OUTPUT
60
61 wrapAll            on
62
63 # CONSTANT-T
64
65 langevin           on
66 langevinTemp       300.0
67 langevinDamping    1.0
68
69 # CONSTANT-P
70
71 LangevinPiston     on
72 LangevinPistonTarget 1.01325
73 LangevinPistonPeriod 200
74 LangevinPistonDecay 100
75 LangevinPistonTemp 300
76 useGroupPressure   yes
77
78 # SPACE PARTITIONING
79
80 splitpatch         hydrogen
81 hgroupcutoff       2.8
82 stepspercycle      10
83
84 # CUT-OFFS
85
86 switching          on
87 switchdist         10.0
88 cutoff             12.0
89 pairlistdist       14.0
90 limitDist          1.3
91
92 # RESPA
93
94 fullElectFrequency 2
95 nonbondedFreq      1
96
97 # 1-4 NON-BONDED
98
99 exclude            scaled1-4
100 1-4scaling         1.0
101
102 # COM
103
104 commotion          no
105
106 # SHAKE
107
108 rigidbonds         all
109 rigidtolerance     0.000001
110 rigiditerations    400

```

```
111
112 # COLVARS SECTION
113
114 colvars                off
115 colvarsConfig          ../common/COMCOM.in
116
117
118 # FEP PARAMETERS
119
120 source                 ../common/fep.tcl
121
122 alch                   on
123 alchType               FEP
124 alchFile               ../common/FEP.fep
125 alchCol                B
126 alchOutFreq           200
127 alchOutFile           forward.fepout
128
129 alchElecLambdaStart   0.1
130 alchVdwLambdaEnd      1.0
131 alchVdwShiftCoeff     5.0
132 alchdecouple          on
133
134 alchEquilSteps        2000
135 set numSteps          10000
136
137 set dLambda           0.05
138
139 runFEP 0.0 1.0 $dLambda $numSteps
```

A.4 Tcl skript využitý k analýze trajektorií

Příklad Tcl skriptu, který byl využit v programovacím rozhraní programu VMD pro analýzu trajektorií kovalentních inhibitorů proteázy Mpro. Skript projde veškeré adresáře pro vázané a volné ligandy vždy pro všech 1000 trajektorií a přečte výsledné hodnoty práce ze souboru pod názvem forward.fepout. Tyto hodnoty jsou následně využity k výpočtu volné energie dle Jarzynského rovnosti a aproximace kumulanty. K výsledkům je také započtena korekce na standardní stav, pro kterou bylo nutné určit objem, ve kterém se ligand během simulace nacházel. Podobné skripty byly využity v průběhu celé práce.

```
1 #
2 # CONSTANTS
3 #
4 set T 300.0
5 set k_boltzmann 1.38064852e-23
6 set NA 6.0221409e+23
7 set joule_to_kcal 0.000239006
8 set pi 3.1415926535897931
9 set R_kcal 0.00198721806
10
11 set V_0 1660.0
12 set treshold_multiplier 2.0
13
14 set subdir_num 1000
15 set upper_bound_length 200
16
17 set folder_prefix_list { ligand }
18 set ligand_number_list { 1 2 3 4 5 6 }
19 set subdir_list { complex ligand }
20
21 foreach prefix $folder_prefix_list {
22   foreach lig $ligand_number_list {
23     foreach subdir $subdir_list {
24
25       cd ./${prefix}_${lig}/${subdir}
26
27       puts "Analysis start for the folder ${prefix}_${lig}/${subdir}\n\n"
28
29       # initialization
30       set work_list {}
31       set work_num 0
32
33       set results(${prefix}_${lig},${subdir},work_avg) 0.0
34       set results(${prefix}_${lig},${subdir},work_sigma) 0.0
35       set results(${prefix}_${lig},${subdir},dG_kumulant) 0.0
36       set results(${prefix}_${lig},${subdir},dG_Jarzynski) 0.0
37
38       set results(${prefix}_${lig},${subdir},max_dist_avg) 0.0
39       set results(${prefix}_${lig},${subdir},max_dist_sigma) 0.0
40       set results(${prefix}_${lig},${subdir},V_1) 0.0
41       set results(${prefix}_${lig},${subdir},vol_corr) 0.0
42
43       set results(${prefix}_${lig},${subdir},charge_corr) 0.0
44       set results(${prefix}_${lig},${subdir},dG_kumulant_corr) 0.0
45       set results(${prefix}_${lig},${subdir},dG_Jarzynski_corr) 0.0
46       set completed_runs(${prefix}_${lig},${subdir}) {}
```

```

47
48 # look through all subdir
49 # check if run successfully finished
50 # read results from the forward.fepout file
51 for {set i 1} { $i <= $subdir_num } {incr i} {
52     cd srun$i
53
54     if { [catch {set fepout_file [open forward.fepout r]} error_text] } {
55         puts "Could not find forward.fepout file in
56             ↪ ${prefix}_${lig}/$subdir/srun$i"
57     } else {
58         catch {seek $fepout_file -$upper_bound_length end}
59         set data [read -nonewline $fepout_file]
60         set lastline [lindex [split $data "\n"] end]
61         close $fepout_file
62
63         if { [lindex $lastline 0] == "#Free"
64             && [lindex $lastline 8] == 1.0
65             && [file exists "z_cas"]} {
66             lappend work_list [lindex $lastline 18]
67             set work_num [expr $work_num + 1]
68             lappend completed_runs(${prefix}_${lig},$subdir) [expr $i]
69         } else {
70             puts "Run inside ${prefix}_${lig}/$subdir/srun$i did not finish
71                 ↪ successfully."
72         }
73     }
74     cd ..
75 }
76 #
77 # Analyze results
78 #
79 # work average
80 set work_avg 0
81 for {set i 0} { $i < [expr $work_num] } {incr i} {
82     set work_avg [expr $work_avg + [lindex $work_list $i]]
83 }
84 set work_avg [expr $work_avg / $work_num]
85
86 # work standard deviation
87 set work_dev_sum 0
88 set work_sigma 0
89 for {set i 0} { $i < [expr $work_num] } {incr i} {
90     set work_dev_sum [expr $work_dev_sum +
91         ($work_avg - [lindex $work_list $i])
92         * ($work_avg - [lindex $work_list $i]) ]
93 }
94 set work_sigma [expr sqrt($work_dev_sum/$work_num)]
95
96 #
97 # Kumulant dG
98 #
99 set dG_kumulant [expr $work_avg - $work_sigma * $work_sigma / (2 * $T *
100     ↪ $R_kcal)]
101 #

```

```

102 # Jarzynski dG
103 #
104 set exp_work_list {}
105 for {set i 0} {$i < [expr $work_num]} {incr i} {
106     lappend exp_work_list [expr exp(- [lindex $work_list $i]/ ($R_kcal * $T))]
107 }
108 set exp_work_avg 0
109 for {set i 0} {$i < [expr $work_num]} {incr i} {
110     set exp_work_avg [expr $exp_work_avg + [lindex $exp_work_list $i]]
111 }
112 set exp_work_avg [expr $exp_work_avg / $work_num]
113
114 set dG_Jarzynski [expr - $R_kcal * $T * log($exp_work_avg)]
115
116 # update contribution lists
117 puts "avg work: $work_avg +/- $work_sigma, dG_kumulant: $dG_kumulant,
↪ dG_Jarzynski: $dG_Jarzynski\n\n"
118 set results(${prefix}_${lig},${subdir,work_avg}) $work_avg
119 set results(${prefix}_${lig},${subdir,work_sigma}) $work_sigma
120 set results(${prefix}_${lig},${subdir,dG_kumulant}) $dG_kumulant
121 set results(${prefix}_${lig},${subdir,dG_Jarzynski}) $dG_Jarzynski
122
123 cd ../../
124
125 }}}
126
127 #
128 # Volume correction for the standard state
129 #
130
131 set folder_prefix_list { ligand }
132 set ligand_number_list { 1 2 3 4 5 6 }
133 set subdir_list { complex }
134
135 package require pbctools
136
137 foreach prefix $folder_prefix_list {
138     foreach lig $ligand_number_list {
139         foreach subdir $subdir_list {
140
141             cd ./${prefix}_${lig}/${subdir}
142
143             puts "Analysis start for the standard state vol corr inside the folder
↪ ${prefix}_${lig}/${subdir}\n\n"
144
145             set max_dist_list {}
146             set max_dist_num 0
147
148             for {set i 1} { $i <= $subdir_num } {incr i} {
149                 cd srun$i
150
151                 # load in trajectory and topology
152                 mol new ../common/QwikMD.psf
153                 mol addfile forward.dcd waitfor all
154
155                 set subdir_max_dist 0.0
156
157                 # selection for ligand and protein

```



```

158 set sel_ligand [atomselect top "resname L1 or resname L2 or resname L3 or
↳ resname L4 or resname L5 or resname L6"]
159 set sel_protein [atomselect top "protein"]
160 set n [molinfo top get numframes]
161
162 # COM initial
163 set sel_ligand_COM_initial [measure center $sel_ligand weight [$sel_ligand
↳ get mass]]
164 set sel_protein_COM_initial [measure center $sel_protein weight
↳ [$sel_protein get mass]]
165
166 # loop all frames
167 for {set j 0} {$j < $n } {incr j} {
168     $sel_ligand frame $j
169     $sel_protein frame $j
170
171     set a [lindex [lindex [pbc get -first $j -last $j] 0] 0]
172     set b [lindex [lindex [pbc get -first $j -last $j] 0] 1]
173     set c [lindex [lindex [pbc get -first $j -last $j] 0] 2]
174     set box_dim [list $a $b $c ]
175
176     # COM
177     set sel_ligand_COM [measure center $sel_ligand weight [$sel_ligand get
↳ mass]]
178     set sel_protein_COM [measure center $sel_protein weight [$sel_protein
↳ get mass]]
179
180     # COM COM distance modified for periodic boundary condition
181     set vec_diff_COM [vecsub
182     [vecsub $sel_ligand_COM $sel_ligand_COM_initial]
183     [vecsub $sel_protein_COM $sel_protein_COM_initial]
184     ]
185     set scaled_rounded_vec_diff_COM [list
186     [expr round([lindex $vec_diff_COM 0]/[lindex $box_dim 0])]
187     [expr round([lindex $vec_diff_COM 1]/[lindex $box_dim 1])]
188     [expr round([lindex $vec_diff_COM 2]/[lindex $box_dim 2])]
189     ]
190     set vec_diff_COM [list
191     [expr [lindex $vec_diff_COM 0]
192     - [lindex $scaled_rounded_vec_diff_COM 0]
193     * [lindex $box_dim 0]]
194     [expr [lindex $vec_diff_COM 1]
195     - [lindex $scaled_rounded_vec_diff_COM 1]
196     * [lindex $box_dim 1]]
197     [expr [lindex $vec_diff_COM 2]
198     - [lindex $scaled_rounded_vec_diff_COM 2]
199     * [lindex $box_dim 2]]
200     ]
201
202     set current_dist [veclength $vec_diff_COM]
203
204     # update if bigger than previous max distance
205     if { $current_dist > $subdir_max_dist } {
206         set subdir_max_dist $current_dist
207     }
208 }
209
210 # append max distance from one specific srin into the max_dist_list

```

```

211     lappend max_dist_list $subdir_max_dist
212     set max_dist_num [expr $max_dist_num + 1]
213
214     mol delete top
215     cd ..
216 }
217
218 #
219 # Analyze distance results
220 #
221
222 # avgerage maximum distance
223 set max_dist_avg 0.0
224 for {set i 0} {$i < [expr $max_dist_num]} {incr i} {
225     set max_dist_avg [expr $max_dist_avg + [lindex $max_dist_list $i]]
226 }
227 set max_dist_avg [expr $max_dist_avg / $max_dist_num ]
228
229 # standard deviation for maximum distance
230 set max_dist_dev_sum 0.0
231 set max_dist_sigma 0.0
232 for {set i 0} {$i < [expr $max_dist_num]} {incr i} {
233     set max_dist_dev_sum [expr $max_dist_dev_sum + ($max_dist_avg - [lindex
        ↪ $max_dist_list $i]) * ($max_dist_avg - [lindex $max_dist_list $i])]
234 }
235 set max_dist_sigma [expr sqrt($max_dist_dev_sum / $max_dist_num)]
236
237 #
238 # Treshold
239 #
240 set treshold [expr $max_dist_avg + $treshold_multiplier*$max_dist_sigma]
241
242 #
243 # Standard state correction
244 #
245 set V_1 [expr min(4 * $pi * $treshold * $treshold * $treshold / 3, $a * $b *
        ↪ $c)]
246 set vol_corr [expr - $R_kcal * $T * log($V_1 / $V_0 )]
247
248 # append results to the contribution list
249 puts "Maximum distance: $max_dist_avg +/- $max_dist_sigma, vol V1: $V_1, corr:
        ↪ $vol_corr\n\n"
250
251 set results(${prefix}_${lig},${subdir},max_dist_avg) $max_dist_avg
252 set results(${prefix}_${lig},${subdir},max_dist_sigma) $max_dist_sigma
253 set results(${prefix}_${lig},${subdir},V_1) $V_1
254 set results(${prefix}_${lig},${subdir},vol_corr) $vol_corr
255
256 cd ../../
257
258 }}}

```