

Atomistické počítačové simulace ribozomu

Habilitační práce

Michal H. Kolář

2022-10-26

Obsah

Seznam obrázků	vi
Seznam zkratk	vii
Předmluva	ix
1 Úvod	1
2 Metody	7
2.1 Struktura simulačních projektů	8
2.2 Vstupní data	9
2.3 Stavba simulovaného systému	11
2.4 Přípravné fáze neboli ekvilibrace	14
2.5 Produkční simulace / sběr dat	15
2.6 Analýzy	16
2.6.1 Základní analýzy	17
2.6.2 Analýza sekundární struktury	19
2.6.3 Analýza hlavních komponent	19
2.6.4 Analýzy šité na míru	21
3 Výsledky a diskuze	23
3.1 Translační uvěznění peptidu VemP	23
3.1.1 VemP v tunelu a mimo něj	24
3.1.2 VemP11	26
3.1.3 Metadynamika mutantů VemP11	29

3.2	Dekódování nestandardního genetického kódu	31
3.3	Allosterický přenos signálu ribozomem	34
3.4	Konformační preference fragmentů ribozomálních proteinů	38
3.5	Dynamika nejužšího místa v tunelu bakteriálního ribozomu	45
4	Závěr	49
	Reference	51

Seznam obrázků

1.1	Bakteriální ribozom	2
1.2	Schéma bakteriální translace	3
2.1	Schéma MD simulací	8
2.2	Schéma simulačního projektu	9
2.3	Výřez ribozomu	12
2.4	Ošetření periodických okrajových podmínek	17
2.5	Analýza hlavních komponent 2D	20
2.6	PCA modelového proteinu	20
3.1	Schéma ribozomálního tunelu	24
3.2	VemP a model translačního uvěznění	25
3.3	Charakteristiky VemP11	27
3.4	Rozbalování VemP11	28
3.5	Umbrella sampling simulace VemP11	28
3.6	Metadynamika VemP11	30
3.7	Dekódovací centrum ribozomu	32
3.8	Konformační soubory eRF1 a kodonu	33
3.9	Interakce eRF1 s kodonem	34
3.10	Ribozom a peptidová deformyláza	36
3.11	Allosterie skrz ribozom	37
3.12	Fragment ribozomálních proteinů	39
3.13	Počáteční konformace uL22	40

3.14 Gyrační poloměry fragmentů	41
3.15 Sekundární struktura fragmentů	42
3.16 Torzní úhel ϕ fragmentu uL2	44
3.17 Srovnání veličin z NMR a MD simulací	45
3.18 Schéma kulového výřezu	46
3.19 Šířka zúžení ribozomálního tunelu	47
3.20 Markovův stavový model dynamiky zúžení	48

Seznam zkratek

A99SB-D	silové pole Amber [1]
aa-tRNA	aminoacylovaná tRNA
fM-tRNA	tRNA nesoucí formylmetioninovou skupinu
BPTI	inhibitor trypsinu hovězího pankreatu (z angl. <i>bovine pancreatic trypsin inhibitor</i>)
C36M	silové pole CHARMM [2]
CD	cirkulární dichroismus
CMAP	kolektivní souřadnice popisující kontakty
cryo-EM	kryogenní elektronová mikroskopie
DNA	deoxyribonukleová kyselina
DSSP	definice sekundární struktury proteinu
eRF1	eukaryotický release factor 1
GIGO	odpad jako vstup, odpad jako výstup (z angl. <i>garbage in, garbage out</i>)
IDP	vnitřně nestrukturovaný protein (z angl. <i>intrinsically disordered protein</i>)
LJ	Lennard-Jones
MD	molekulová dynamika
metaD	metadynamika
mRNA	messengerová ribonukleová kyselina
MSM	Markovův stavový model
NMR	nukleární magnetická rezonance
PBC	periodické okrajové podmínky (z angl. <i>periodic boundary conditions</i>)
PC	hlavní komponenta (z angl. <i>principal component</i>)
PCA	analýza hlavních komponent (z angl. <i>principal component analysis</i>)
PCR	regrese hlavních komponent (z angl. <i>principal component regression</i>)
PDB	proteinová databanka
PDF	peptidová deformyláza
PMF	potenciál střední síly (z angl. <i>potential of mean force</i>)
RF	release faktor
RMSD	odmocnina střední kvadratické odchylky (z angl. <i>root mean square deviation</i>)
RMSF	odmocnina střední kvadratické fluktuace (z angl. <i>root mean square fluctuation</i>)
RNA	ribonukleová kyselina

rRNA ribozomální RNA

r-protein ribozomální protein

SRN Spolková republika Německo

tRNA transferová RNA

US umbrella sampling

USA Spojené státy americké (z angl. *United States of America*)

VemP Vibrio-export monitoring peptide

VemP11 VemP zkrácený na 11 reziduí

VŠCHT Vysoká škola chemicko-technologická

Předmluva

Mým oblíbeným slovním obratem, který používám v populárně naučných a někdy i odborných textech, je „život, jak jej známe“. Obvykle doplňuji, že takový život je založený na biomolekulách. Zároveň tím ale připouštím, že nelze vyloučit existenci života stojícího na jiných principech. Ten zatím neznáme. Z tohoto pohledu jsou biomolekuly pro současné vědce a vědkyně stěžejní objekty, které má smysl studovat.

Redukcionismus přistupuje k vědění na základě předpokladu, že celek lze pochopit jeho rozložením na části a jejich následným studiem. V mnoha ohledech takto funguje i současná přírodověda, tedy vědní oblasti biologie, chemie, fyziky a jejich různých překryvů a odnoží. Lidé se zaobírají atomy a molekulami, jdou-li hlouběji do mikrokosmu, tak třeba gluony nebo kvarky. Stejně i já jsem poslední roky hledal odpovědi na přírodovědné otázky studiem jednotlivostí a tato habilitační práce mé úsilí shrnuje. Život, nebo alespoň některé jeho aspekty, lze pochopit studiem atomů. Doufám. Myslel si to i R. Feynman, když prohlásil „Everything that living things do can be understood in terms of the jiggings and wiggings of atoms.“

Kromě představení zásadních výsledků a závěrů mého výzkumu, jež jsou potřebné k úspěšné habilitaci na české vysoké škole, je také mým cílem zpřístupnit bádání budoucím studentům a studentkám, kteří projeví zájem o počítačové simulace. Ty totiž reprezentují můj hlavní pracovní rámec. Tímto chci objasnit, proč jsem se do psaní pustil ve své mateřštině (a velmi pravděpodobně i mateřštině budoucích studentů a studentek) a nikoli mezinárodně, anglicky. Je to projevem mé snahy snižovat bariéry k vědění, která s cizojazyčným textem vzniká, byť pozoruji u nejmladší generace studentstva, že to je bariéra docela malá. Anglická literatura je jistě plná dostupných a možná¹ i lepších textů o počítačových simulacích (bio)molekul, než je tento.

Text popisuje pilíře mé vědecké práce posledních asi 5 let, často ale toto období přesahuje. Vždyť svou první atomistickou simulaci jsem provedl jako bakalářský student druhého ročníku. Jednalo se o komplex dvanáctimeru DNA s 4',6-diamidin-2-fenylindolem, fluorescenčním barvivem známým jako DAPI. Navzdory tomu jsem se pak zevrubněji

¹Neskromně si dovoluji pochybovat, neboť Předmluva je jedna z prvních pasáží mé habilitační práce, kterou píšu, a kvalitu budoucího textu nyní nemohu zhodnotit.

věnoval kvantové chemii a k simulacím se řádně vrátil až po ukončení doktorského studia při svých dvou stážích v Německu. Nejprve v Jülichu, kde jsem pomocí molekulárních simulací studoval interakce halogenovaných sloučenin s RNA. Studium to bylo spíše neúspěšné, neboť vědeckých cílů, kterých jsem si vytyčil, jsem nedosáhl. Příroda prostě nefungovala tak, jak jsem si představoval. I tak to byla hodnotná lekce, možná díky které jsem se posunul do Göttingenu, kde jsem se počítačovým simulacím začal věnovat, jak se říká, na plný úvazek.

Mnoho z toho, o čem zde píšu, by nebylo možné, kdybych kolem sebe neměl rodinu, přátele a kolegy, kteří mi to umožnili a kterým srdečně děkuji. Děkuji také těm studentkám a studentům VŠCHT Praha, se kterými mám potěšení spolupracovat. Přečtením rané verze textu pomohli najít a opravit mnoho překlepů a chyb.

Kapitola 1

Úvod

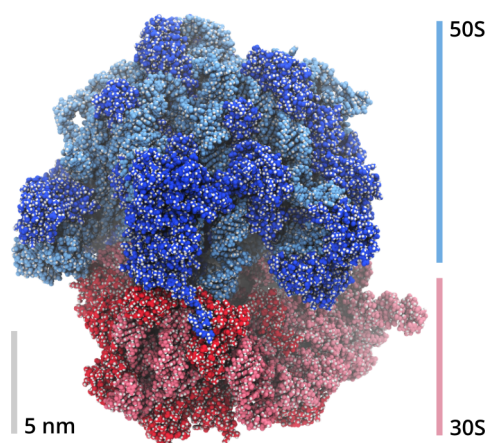
Biopolymery, jako jsou proteiny nebo nukleové kyseliny, se účastní drtivé většiny biologických dějů. Kromě biopolymerů v živých organizmech vystupují stovky typů menších molekul jako kofaktory enzymů, metabolity nebo hormony. Společně tvoří nesmírně složitou síť chemických reakcí, kterou nazýváme životem.

Z relativně malého množství stavebních jednotek (monomerů) lze postavit nesčetně polymerů. Např. ze 4 typů nukleotidů založených na adeninu, thyminu, cytosinu a guaninu lze sestavit řádově 10^6 různých DNA o délce 100 párů nukleotidů, což je skutečně astronomické množství. Lidská DNA ovšem obsahuje přibližně 3^9 párů nukleotidů. Teoretická rozmanitost proteinového světa je ještě větší, neboť stavebních kamenů – aminokyselinových zbytků – existuje přibližně 20. Proteinů o sekvenční délce 100 reziduí teoreticky existuje 10^{130} . Pouze malé množství z nich se ale v mikrosvětě živých organismů skutečně nachází. Proto existují přísně regulované metabolické dráhy, které zajišťují nejen syntézu biomolekul, ale i jejich transport, modifikaci, vzájemnou interakci nebo degradaci.

Předkládaná práce se zaměřuje na syntézu bílkovin. Ta je mezi různými organismy a napříč životními fázemi vybraného organismu značně konzervovaná. Proteiny jsou v živých organizmech syntetizovány výhradně na ribosomech, což jsou velké biomolekulové komplexy, nebo – v závislosti na tom, kdo o nich hovoří – malé organely.

Ribozomy byly objeveny pomocí elektronové mikroskopie v 50. letech 20. století Georgem E. Paladem jako *malé částicové složky cytoplazmy* [3]. Název *ribosom* se začal používat o několik let později. Že se jedná o komplexy ribonukleové kyseliny (RNA) a proteinů, bylo zřejmé velice brzy po jejich objevu, stejně jako bylo brzy jasné, že ribozomy jsou zodpovědné za proteosyntézu, tzv. translaci.

Podle centrálního dogmatu molekulární biologie je proteosyntéza posledním krokem exprese genetické informace, během které je DNA transkribována RNA polymerázou do



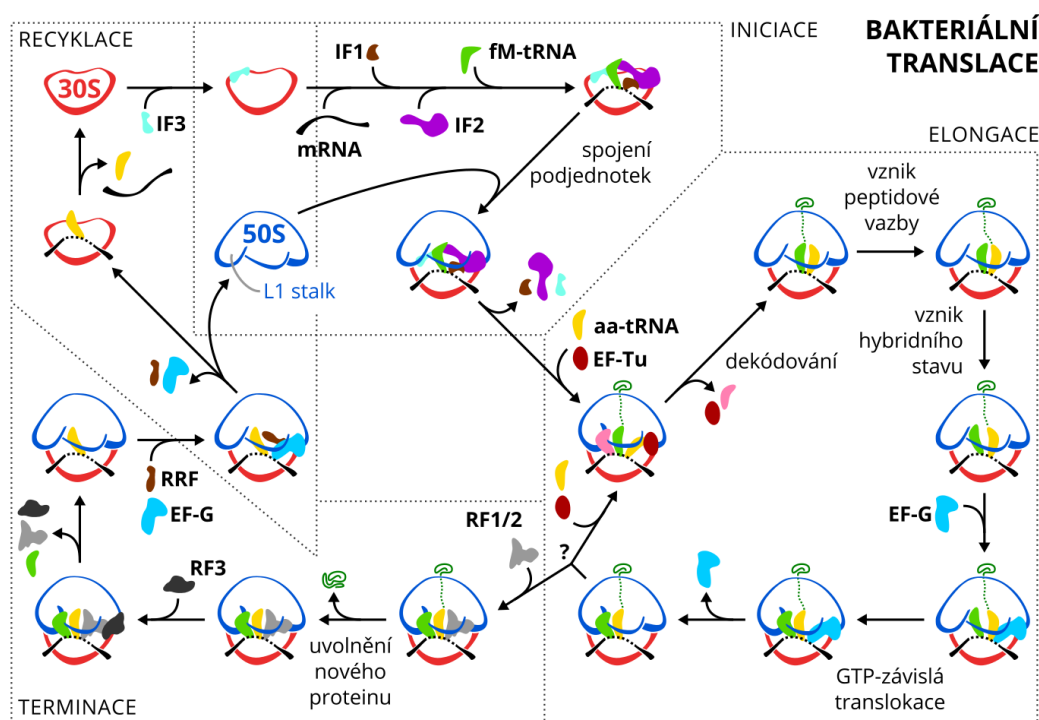
Obrázek 1.1: Bakteriální ribozom s velkou 50S podjednotkou v odstínech modré a malou 30S podjednotkou v odstínech červené. Tmavší odstíny značí ribozomální proteiny, světlejší odstíny jsou rRNA. Atomy vodíku jsou znázorněny bíle. Vytvořeno z cryo-EM struktury *E. coli* ribozomu s rozlišením 2.9 Å (PDB 5AFI [7]). Upraveno z Ref. 8.

messengerové RNA (mRNA). mRNA je na ribozomech následně (nebo už během transkripce) translatována na proteiny. Složitost translačního aparátu ale zhusta přesahuje sdělení z předešlé věty. Mnoha okolnostem vzniku proteinů již rozumíme, zároveň ale víme, že v našem chápání funkce ribozomu existuje mnoho bílých míst.

Ribozom plní mnoho rozmanitých úkolů. Na ribozomu nejde jen o prostou katalýzu tvorby peptidových vazeb mezi aminokyselinovými zbytky. Přesto i ta sama o sobě představuje spletitou soustavu dějů [4]. Ribozom hraje zásadní roli v regulaci proteosyntézy [5], podílí se na časoprostorovém cílení proteinů a skrze synchronizaci s externími neribozomálními procesy má souvislost s množstvím patofyziologických stavů [6].

Ribozom se skládá z několika vláken ribozomální RNA (rRNA) a asi 50 ribozomálních proteinů (r-proteinů), které jsou organizovány do dvou ribozomálních podjednotek (Obr. 1.1). Přesný počet molekul rRNA a r-proteinů je pro tři domény života (bakterie, archea neboli staruše a eukaryota) rozdílný. Jednodušší organizmy mají ribozomy menší a proporčně bohatší na rRNA než složitější organizmy. Specializované ribozomy např. některých plastidů mohou být sekvenčně a strukturně velmi specifické [9]. Organizace do dvou podjednotek je však konzervovaná.

Napříč různými ribozomy je malá ribozomální podjednotka zodpovědná za čtení genetické informace dočasně uložené v mRNA. Jedna po druhé se do ribozomu vážou aminoacylované transferové RNA (tRNA) a aminokyselinové zbytky se propojují peptidovými vazbami (Obr. 1.2). Malá ribozomální podjednotka zajišťuje správné řazení tRNA. Velká podjednotka katalyzuje chemické reakce rodícího se proteinu, které jsou dvě. Zaprvé jde o reakci vzniku peptidové vazby mezi aminoacylovanou tRNA v tzv. A místě ribozomu a tRNA nesoucí rodící se protein v tzv. P místě. V každém proteinu vzniká několik de-



Obrázek 1.2: Schéma bakteriální translace rozdělené do čtyř logických fází: iniciace, elongace, terminace a recyklace. Přepřacováno z Ref. 10 na základě známých ribozomálních struktur shrnutých v Ref. 4. Význam zkratk: aa-tRNA = aminoacylovaná tRNA, fM-tRNA = tRNA nesoucí formylmethioniovou skupinu, IF = iniciační faktor, EF = elongační faktor, RF = release factor, RRF = ribozomální recyklační faktor, GTP = guanosintrifosfát.

sítek až stovek peptidových vazeb. Po připojení posledního aminokyselinového zbytku do rodícího se proteinu ribozom katalyzuje rozštěpení vazby mezi tRNA v P místě a kompletním proteinem, který se tím z ribozomu uvolňuje.

Struktura bakteriálního ribozomu byla na atomární úrovni rozřešena rentgenovou krystalografií na přelomu milénia [11–13]. Vědecké skupiny vedené Adou Yonath, Venkim Ramakrishnanem a Thomasem Steitzem určily nezávisle na sobě atomární složení velké a malé ribozomální podjednotky. Určení atomární struktury ribozomu otevřelo novou éru studia ribozomu, za což byli vedoucí skupin v roce 2009 odměněni Nobelovou cenou za chemii. Zásadní bylo např. zjištění, že aktivní místo velké ribozomální podjednotky je tvořeno výhradně rRNA a katalýza vzniku peptidových vazeb tedy lokálně probíhá bez přispění r-proteinů [11].

Struktura ribozomu je, stejně jako u jiných (bio)molekul, proměnlivá. Konformační změny ribozomu spojené s translací probíhají na široké časové škále od pikosekund (např. vznik peptidové vazby) po sekundy (např. kotranslační balení proteinů). Experimentálně určené struktury ribozomu potom reprezentují konformační stavy, průměrované po dobu trvání experimentu. V poslední dekádě se stěžejní experimentální technikou pro

studium struktury ribozomu stala kryogenní elektronová mikroskopie (cryo-EM) [14], která z hlediska prostorového rozlišení aktuálně dosáhla kvalit rentgenové krystalografie: Fromm et al. ohlásili určení struktury bakteriálního ribozomu s průměrným rozlišením 1,55 Å [15]. Experimenty s explicitním časovým rozlišením jsou pro ribozom zatím málo dostupné, ale zřetelně představují směr, kterým se bude výzkum ubírat [16].

Vhodným nástrojem pro studium dynamiky ribozomu, a ostatně i jiných biomolekul a jejich komplexů, jsou počítačové simulace. V této práci se věnuji klasickým molekulárně dynamickým (MD) simulacím. Pečlivěji je popisují v kapitole 2.5 Produkční simulace / sběr dat. Shrnutí simulací ribozomu je také dostupné ve dvou souborných článcích, na kterých jsem se podílel [8, 10].

MD simulace generují statistický soubor biomolekulárních konformací. Děje se tak obvykle (ale ne výhradně) numerickou integrací Newtonových pohybových rovnic. Takové simulace poskytují *trajektorii*, tedy časovou závislost kartézských souřadnic jednotlivých atomů. Aparát statistické mechaniky potom umožňuje mikroskopické informace o atomech interpretovat v termínech makroskopického světa a vypočítat veličiny jako relativní Gibbsova energie nebo rychlostní konstanta.

Od pionýrských MD simulací lennard-jonesovských tekutin [17] prošly simulační techniky dlouhým vývojem. První simulace biomolekuly, proteinu BPTI (z angl. *bovine pancreatic trypsin inhibitor*), byla z dnešního pohledu provedena velmi aproximativními algoritmy [18]; simulační model např. nepopisoval vodné prostředí proteinu. Práce však poukázala na podstatný, v té době přehlížený, fakt, že proteiny jsou dynamické objekty. Otázka „Jak se biomolekuly pohybují,“ se stala naprosto relevantní. Tvzení, že funkce biomolekuly je dána její strukturou, musela být revidována a *dynamika* biomolekul se dostala do popředí [19, 20]. Stojí za zmínku, že mimořádné úspěchy zaznamenaly MD simulace při studiu balení proteinů [21, 22], interakcí proteinů s nízkomolekulárními látkami [23, 24] či přenosu vody a iontů membránovými kanály [25, 26].

Jak se zvyšoval výkon počítačů a efektivita simulačních programů, rostly také časové škály simulací, případně velikosti simulovaných systémů. První MD simulace kompletního ribozomu byla provedena v roce 2003 v Los Alamos National Laboratory, USA pod vedením K. Sanbonmatsu [27]. Její vědecká skupina se zaměřila na kritický krok dekodování mRNA, a to tzv. akomodaci tRNA ze stavu A/T do stavu A/A. Simulace byly provedeny na Q Machine, druhém nejvýkonnějším superpočítači světa té doby s výkonem 20 teraFLOPS. Dobový přínos pro pokrok v simulacích mělo také nobelovské určení atomární struktury obou ribozomálních podjednotek o několik let dříve.

Atomistické MD simulace ribozomu jsou i dnes výpočetně extrémně náročné, a to ze tří hlavních důvodů:

1. ribozom je velký,

2. význačné děje mají široký rozsah charakteristických časů,
3. ribozom je chemicky heterogenní.

Ribozom o velikosti 20–25 nm je pro běžné počítače stále příliš velké sousto. Pracovní stanice s moderní grafickou kartou vyprodukuje přibližně 2–5 ns trajektorie kompletního ribozomu v explicitním vodném prostředí za den. Pro některé otázky ohledně translace tato časová škála může stačit, mnohem zajímavější děje se ale na ribozomu odehrávají asi 1000× pomaleji. Ty lze studovat na superpočítačích. Je také možné „obětovat“ atomární rozlišení a simulace ribozomu tzv. zhrubit [28]. Náročné je také zajistit dostatečně kvalitní simulační parametry, zejména interakční potenciál, který musí správně popisovat RNA, proteiny i solvent.

Předkládanou habilitační práci jsem uspořádal okolo dvou hlavních kapitol. Kapitola 2 Metody je určena zejména studentkám a studentům. Popisují v ní simulační techniky pro studium ribozomu včetně některých technických záležitostí. Soustředím se na atomistické MD simulace, okrajově zmiňuji i pokročilé techniky jako např. simulace výřezů ribozomu. Snažím se tím usnadnit si práci při seznamování budoucích studentů se simulacemi a nejde mi jen o to, aby zvládli technickou stránku věci. Potřebuji je seznámit i s obecnější koncepcí. Celkově doufám, že tím zvýším užitečnost textu, do jehož psaní vkládám úsilí. Habilitačních prací jsem ve své kariéře mnoho nečetl a umím si představit, že v tom nejsem výjimkou. Hrozí tak, že se můj text přidá k těm, které nikdo nečetl. Kapitola 3 Výsledky a diskuze směřuje primárně na komisi a členy akademické obce. Popisují v ní projekty, jenž vedly k novým znalostem o fungování ribozomu a které tvoří těžiště mého vědeckého přínosu posledních let. U každého projektu jsou uvedeny spolupracovníci případně spolupracující skupiny, neboť MD simulace jsou obvykle součástí širšího spektra vědeckých metod. U již publikovaných výsledků a závěrů se snažím být stručnější a odkazuji na recenzované práce. U projektů, které zatím publikované nejsou, uvádím informací více tak, aby je komise mohla sama kriticky zhodnotit. Stručně formuluji i směr, kterým se každý z projektů dál ubírá.

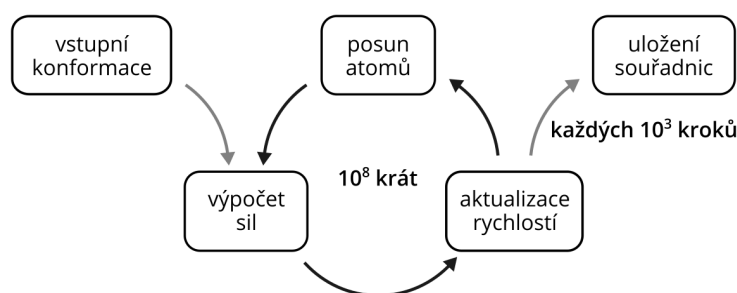
Kapitola 2

Metody

V této kapitole se věnuji obecnějšímu popisu atomistických MD simulací. Při psaní kapitoly mám na paměti především studentstvo technických, přírodních případně i lékařských věd. Věnuji se jakési struktuře simulačních projektů, proto některé pasáže mohou působit až banálně. Konkrétní algoritmy nechávám až na pár výjimek troufale stranou, neboť je o nich skvěle pojednáno např. v česky psaném skriptu I. Nezbedy, J. Kolafy a M. Kotrly *Úvod do molekulárních simulací – Metody Monte Carlo a molekulární dynamiky* (Univerzita Karlova, 2002). Publiku se zájmem o pečlivější teoretický rozbor počítačových simulací mohu otevřeně doporučit anglicky psanou knihu A. R. Leache *Molecular Modelling: Principles and Applications* (Pearson Education, 2001) nebo knihu M. E. Tuckermanna *Statistical Mechanics: Theory And Molecular Simulation* (Oxford University Press 2010). Z nedávných literárních počínů zmíním články v časopisu *Living Journal of Computational Molecular Science*. Časopis se snaží měnit stávající paradigma vědeckého publikování a umožňuje pomocí moderních softwarových technologií aktualizovat průběžně již publikované články s tím, jak se mění úroveň lidského poznání o daném tématu. Vyzdvihl bych práci Browna *et al.* o základech molekulárních simulací [29], J. A. Lemkula o simulacích v programovém balíku GROMACS [30, 31] a A. Grossfielda *et al.* o kvantifikaci chyb v simulačních projektech [32].

Simulační projekty uvedené v kapitole 3 Výsledky a diskuze byly z velké části provedeny v programovém balíku **GROMACS**. Napříč kapitolou proto někdy uvádím příklady, jak jednotlivé metodologické úkony v GROMACSu provést. Alternativami balíku GROMACS jsou např. **NAMD**, **Amber**, nebo **CHARMM**.

V této kapitole se také soustředím na obecnější simulační strategie, které jsou typické pro studium ribozomu, případně pro jiné velké simulační systémy. Takové pracovní protokoly je v literatuře obtížné dohledat a od dobré praxe pro malé systémy se mohou lišit.



Obrázek 2.1: Schéma molekulárně dynamických simulací. Vstupní konformace slouží k počátečnímu výpočtu meziatomových sil, z Newtonových zákonů jsou síly přepočteny na zrychlení a upraveny rychlosti atomů, atomy jsou posunuty ve směru sil. V pravidelných intervalech jsou aktuální souřadnice (případně i rychlosti či jiná data) uloženy pro pozdější analýzu.

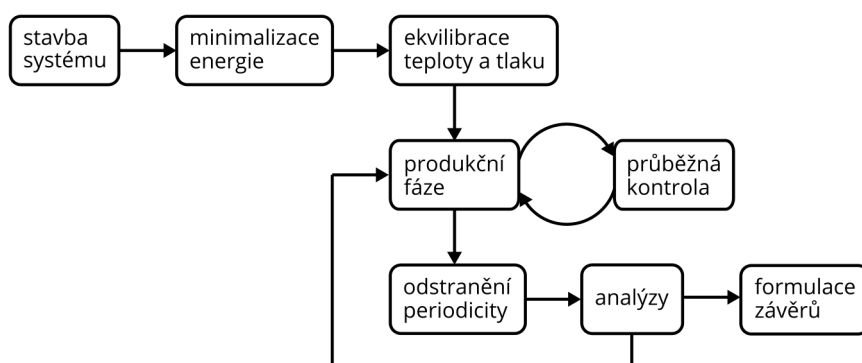
2.1 Struktura simulačních projektů

Schematické znázornění atomistických simulací, o kterých zde pojednávám, je na Obr. 2.1. MD simulacemi mám na mysli sady algoritmů, které iterativně vzorkují konformační prostor biomolekul a tím produkují trajektorii biomolekuly, tj. závislost kartézských souřadnic na čase. Vše je založeno na myšlence, že studiem trajektorií můžeme pochopit fungování biomolekul. Problém můžeme přirovnat k úkolu pochopit fungování automobilu na základě pozorování, z čeho se automobil skládá (karoserie, šrouby, lanka, hadičky, atp.) a jaké vnitřní pohyby vykonává. Rotační pohyb kol, translace válců v motoru, stahování oken, pohyb stěračů – všechny tyto pohyby automobil vykonává, avšak jejich význam pro transport osob a nákladu např. mezi Prahou a Osoblažským výběžkem se liší. I v případě biomolekul se důležitost různých konformačních změn pro funkci biomolekuly liší a MD simulace toto mohou objasnit.

V MD simulacích jde o numerické řešení Newtonových pohybových rovnic pro sadu atomů. Pro počáteční konfiguraci simulačního systému je z interakčního potenciálu spočtena síla působící na každý atom, ze síly je vypočteno zrychlení a aktualizován vektor rychlosti každého atomu. Atomy jsou následně v prostoru posunuty o kousek odpovídající časovému kroku v řádu femtosekund. Pro novou konfiguraci je opět spočtena síla, nové rychlosti a nové kartézské souřadnice. Během simulace dojde k takovému výpočtu řádově 10^8 krát, přičemž jsou výstupní data (kartézské souřadnice, rychlosti, atp.) uchovávány pro pozdější analýzu přibližně každých 10^3 až 10^5 kroků (Obr. 2.1).

Je zřejmé, že výpočet sil musí být výpočetně nenáročný, aby šlo vygenerovat dostatečně dlouhé trajektorie, které zachytí konformační přechody spojené s biochemickými ději. Interakční potenciál je pro biomolekuly obvykle založen na harmonické aproximaci vazebných sil. Nevazebné interakce jsou obvykle popsány bodovými náboji a Lennard-Jonesovým (LJ) potenciálem.

Typický simulační projekt se skládá z několika logických částí, jak je znázorněno na



Obrázek 2.2: Obecné schéma simulačního projektu.

Obr. 2.2.

1. stavba simulovaného systému
2. přípravné a ekvilibrační simulace
3. produkční simulace / sběr dat
4. analýza simulačních dat

Každé části se níže věnuji podrobněji. Řazení zhruba odpovídá časovému uspořádání, přičemž po stavbě simulovaného systému se projekt může větvit do několik nezávislých ramen. Analýzy mohou být také provázány s produkčními simulacemi a zajišťovat tak zpětnou vazbu pro adaptivní rozšiřování produkčních simulací.

2.2 Vstupní data

K praktickému provedení atomistické simulace jsou potřeba tři druhy vstupních informací: i) souřadnice studovaného systému, ii) parametry interakčního potenciálu, iii) parametry simulace.

Počáteční souřadnice studovaného systému a zejména rozpuštěné látky (biomolekuly) představují vstupní data, jež mohou značně ovlivnit kvalitu výsledných MD simulací (ve smyslu konceptu GIGO¹). Lze je získat z biofyzikálních experimentů, jako je rentgenová krystalografie nebo cryo-EM. Pro ribozom existují stovky strukturních modelů s atomárním rozlišením uložené v PDB (z angl. *Protein Data Bank*, <https://www.rcsb.org/>). K dispozici jsou modely ribozomů z různých bakterií i ribozomů vyšších organismů včetně organelárních ribozomů. Experimenty zachycují ribozom v různých funkčních a konformačních stavech, proto lze simulačnímu projektu vybrat strukturu ribozomu na míru. Případně mohou být simulace různých strukturních modelů podstatou strategie pro objasnění příslušné vědecké otázky, jako u studie translokace tRNA ribozomem [33].

¹z angl. *garbage in, garbage out*

Při výběru startovní konformace je třeba se věnovat faktu, že různé části ribozomu bývají experimentálně „viditelné“ s různou přesností. Struktury některých flexibilních r-proteinů na povrchu ribozomu nemusí být v daném modelu k dispozici vůbec. Často se jedná např. o protein bS1 nebo uL3. V této souvislosti je potřeba zmínit, že v poslední době mohou pro MD simulace úspěšně sloužit i modely získané z hlubokých neuronových sítí, jako je AlphaFold2 [34] nebo RosettaFold [35].

Balík GROMACS používá pro strukturní informace vlastní `gro` formát. Pro ribozom je výhodou, že GROMACS umí zpracovat i formát typu `pdb`, který umožňuje snadněji pracovat s mnohosložkovými systémy skrze identifikátor *chain* nebo *segid*.²

Významnou vstupní informací je interakční potenciál, neboli silové pole. Např. pro vazbu C–C popsanou harmonicky $E = \frac{1}{2}k_{CC}(d - d_{CC})^2$ jsou potřeba dva parametry: rovnovážná délka vazby d_{CC} a silová konstanta k_{CC} . Takto je třeba mít před simulací k dispozici knihovnu párových parametrů (vazeb), třístředových interakcí (úhlů), čtyřstředových interakcí (torzí) i atomárních parametrů (elektrický náboj, LJ parametry). Obvykle se parametry určují z experimentálních údajů nebo pomocí kvantově chemických výpočtů.

Silová pole biomolekul představují oblast výzkumu, která se intenzivně rozvíjí od 80. let 20. století a prošla si vzestupy i pády. Např. silová pole pro strukturované proteiny jsou dnes velmi kvalitní [36], což nelze tvrdit o silových polích pro tzv. vnitřně nestrukturované proteiny (IDPs z angl. *intrinsically disordered proteins*). Např. Rauscher *et al.* ukázali, že konformační soubor IDPs je silně závislý na použitém silovém poli [37]. Značného pokroku dosáhla i silová pole pro nukleové kyseliny, byť okolo roku 2007 musela projít významnou revizí, když bylo pozorováno, že pro simulační časy v řádu stovek ns dochází k chybnému rozplétání helikálních struktur. Revidované parametry páteře DNA a později RNA takové chování opravily [38, 39].

Ribozom je chemicky heterogenní systém, proto je potřeba sada vyvážených interakčních parametrů pro r-proteiny (ty jsou většinou strukturované), kanonické nukleotidy rRNA a solvent, tedy vodu a ionty. Ribozom a tRNA navíc obsahují asi 100 typů nekanonických nukleotidů, jako je např. pseudouridin nebo N6-methyladenosin, jejichž parametry jsou taktéž potřeba. K dispozici jsou v silových polích Amber [40] a CHARMM [41]. Pokud jsou součástí simulovaného systému další nekanonické molekuly, např. fluorescenční značky nebo antibiotika, získává se jejich silové pole parametrizací tzv. na zakázku na základě GAFF [42], CGenFF [43], či OpenMM [44].

Celkově lze konstatovat, že pro atomistické MD simulace ribozomu prakticky existují jen dvě sady interakčních parametrů: jedna je založená na silových polích z rodiny Amber [45] a druhá z rodiny CHARMM [46]. Jejich použití není obvykle striktně racionalizováno a podléhá tradicím výzkumné skupiny, případně dobré praxi doložené v literatuře

²Manuál `pdb` formátu je dostupný na <https://www.wwpdb.org/documentation/file-format-content/format33/v3.3.html>.

o předešlých projektech. Je běžné, že validace probíhá na menších simulačních modelech, případně křížmo srovnáním více silových polí.

V balíku GROMACS jsou interakční parametry systému uloženy v souborech `top` a `itp` v pracovním adresáři nebo instalačních adresářích balíku. Jedná se o rozsáhlé soubory, pro solvovaný ribozom jde o přibližně 3 miliony řádků. Protože jsou soubory v textovém lidsky čitelném formátu, jsou snadno editovatelné a dají se tak upravit pro specifický účel.

Třetím druhem vstupních informací jsou parametry MD simulace. Sem patří teplota případně i tlak systému, počet integračních kroků, délka kroku, frekvence ukládání dat pro pozdější analýzu, atp. Patří sem také parametry mnoha algoritmů, které se zavádí pro zvýšení efektivity simulace. Jsou to např. cut-off vzdálenosti, parametry tzv. hledání sousedů (neighbour search), parametry algoritmů fixující délky vazeb. Balík GROMACS tyto parametry obsahuje v souboru typu `mdp`.

2.3 Stavba simulovaného systému

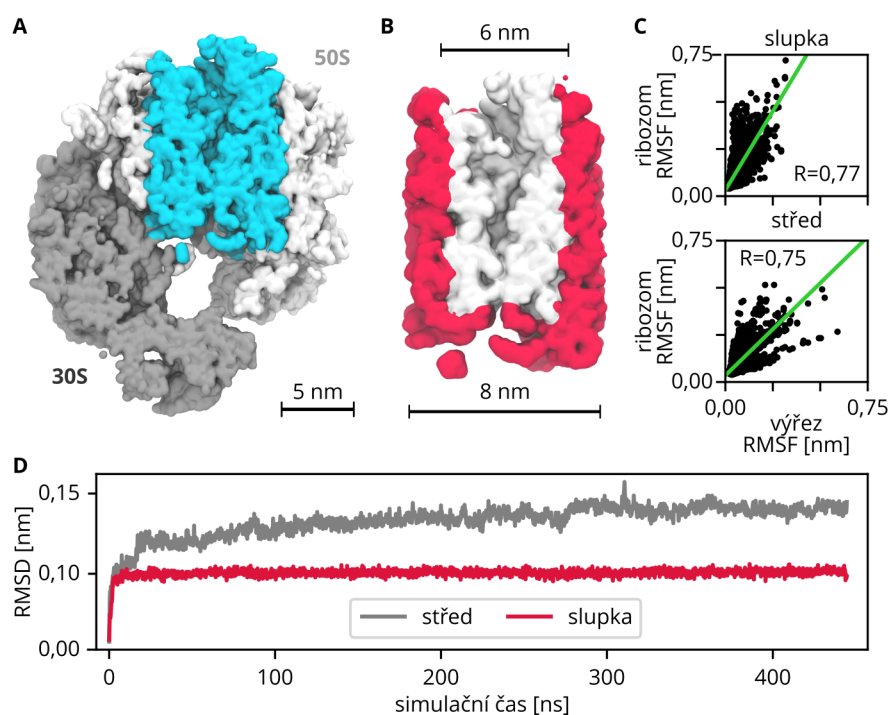
Experimentální struktury deponované v PDB nelze pro atomistické simulace použít přímo. Struktury jsou často nekompletní, chybí např. flexibilní části biomolekul nebo postranní řetězce proteinů. Krystalografické a cryo-EM struktury také obvykle neobsahují vodíkové atomy, jejichž elektronová hustota je příliš nízká, aby byla při typickém rozlišení 2–3 Å měřitelná. Experimentální data neobsahují ani strukturní informace o okolním prostředí. Proto je potřeba simulovaný systém nejprve sestavit, tedy připravit vstupní strukturní data tak, aby mohla během simulace reprezentovat fyzikální realitu.

Ke kompletaci biomolekul existují specializované programy, např. [Pymol](#), [VMD](#) nebo [Molmodeller](#). Simulační balíky obsahují programy, které zkompletovaným biomolekulám opatří vodíky podle standardizovaných knihoven. Zde se využívá faktu, že biopolymery se skládají pouze z několika typů monomerních jednotek. V balíku GROMACS k tomu slouží `gmx pdb2gmx`.

U velkých biomolekulových komplexů, jako je ribozom, existuje mnoho otázek, které se týkají prostorově omezených oblastí. Může se jednat např. o studium pohybu aminokyseliny zbytku tRNA v A místě ribozomu a jeho interakce s rodícím se peptidem, nebo o párování nukleotidů v dekodovacím centru malé ribozomální podjednotky. Simulovat celý ribozom pak může být neefektivní, lze-li oprávněně předpokládat, že vzdálené části ribozomu se studovaného místa netýkají. V takovém případě lze simulovat pouze výřez ribozomu, jak je naznačeno na Obr. 2.3A.

Při sestavení výřezu je potřeba se vypořádat s několika technickými omezeními:

- Ač je možné simulovat i nekompletní monomerní jednotky, např. pouze cukr-



Obrázek 2.3: Výřez ribozomu zahrnující ribozomální tunel. A) Vertikální průřez bakteriálního ribozomu s malou 30S podjednotkou vyznačenou tmavě, velkou 50S podjednotkou světle a výřezem modře. B) Výřez ribozomu s červeně vyznačenou částí podléhající pozičním omezením (tj. slupkou). C) Korelační diagram odmocniny středních kvadratických fluktuací (RMSF, sekce 2.6.1 Základní analýzy) nevodíkových atomů z 300 ns dlouhých MD simulací výřezu a celého ribozomu pro část podléhající pozičním omezením (slupka) a pro zbytek výřezu (střed). Referenční hodnoty ze simulací celého ribozomu jsou uvedeny na svislé ose. D) Časová závislost odmocniny střední kvadratické odchylky souřadnic (RMSD) pro slupku a střed výřezu, počítaná pro všechny těžké atomy vzhledem k souřadnicím v cryo-EM struktuře (PDB 5AFI [7]).

fosfátovou páteř bez nukleobáze, je výhodnější do výřezu zahrnout celá rezidua. Simulační program je potom rozezná ve své knihovně a uživatel/ka se tak vyhne pracné parametrizaci nestandardních reziduí.

- Odstranění vzdálených částí vede k rozdělení dlouhých řetězců (např. rRNA) na kratší podle toho, jak se ve výřezu vyskytují. Je simulačně výhodné řetězce o délce jedné monomerní jednotky vynechat, neboť se v knihovnách reziduí standardně nevyskytují.
- Odstraněním vzdálených částí dojde k narušení přirozené chemické a strukturní celistvosti biomolekulového komplexu. Oblasti na okraji výřezu je proto potřeba fixovat pomocí tzv. pozičních omezení (angl. *position restraints*). To jsou obvykle harmonické potenciály, které udržují jednotlivé atomy blízko referenčních pozic v absolutním souřadném systému. Vede to k vytvoření jakési vnější slupky (Obr. 2.3B), která výřez udržuje strukturně stabilní. Popis struktury a dynamiky vnitřní části výřezu bez pozičních omezení je potom realističtější.

Poziční omezení vyžadují referenční strukturu. Vhodné je použít experimentální strukturu, z níž byl systém postaven. Hodnoty silových konstant k_{pr} se volí v řádu stovek až tisíců $\text{kJ mol}^{-1} \text{nm}^{-2}$. Pokud je k dispozici MD simulace kompletního ribozomu, lze k_{pr} vypočítat z fluktuací atomových pozic jako

$$k_{pr} = \frac{RT}{\text{RMSF}^2}, \quad (2.1)$$

kde R je univerzální plynová konstanta, T je termodynamická teplota a RMSF je odmocnina střední kvadratické fluktuace atomu (z angl. *root-mean-square fluctuation*).

Při sestavování výřezu zahrnujícího ribozomální tunel jsme zkoumali, jaký vliv mají chybějící rezidua na strukturu a dynamiku studované části uvnitř výřezu. Závěr, že vnitřní část výřezu se chová srovnatelně jako stejná část v simulaci celého ribozomu, plyne z odmocniny střední kvadratické odchylky souřadnic (RMSD) a RMSF. RMSD nevodíkových atomů dosáhla během 400 ns dlouhé simulace stabilních hodnot okolo 0,15 nm pro část bez pozičních omezení (Obr. 2.3D). Korelační koeficient R pro RMSF části výřezu bez pozičních omezení měl hodnotu 0,75 (Obr. 2.3C), což je srovnatelné s korelačním koeficientem mezi hodnotami RMSF získanými ze dvou nezávislých MD simulací kompletního ribozomu (mezi 0,7 a 0,8). Střední kvadratická odchylka RMSF z výřezu a RMSF z celého ribozomu pro atomy bez pozičních omezení (tj. „odchylka z odchylek“) byla 0,046 nm, což je opět srovnatelné s hodnotou pro dvě nezávislé simulace celého ribozomu (okolo 0,035 nm).

Pro MD simulace biomolekul včetně ribozomu se běžně využívá tzv. periodických okrajových podmínek (PBC, z angl. *periodic boundary conditions*). PBC byly zavedeny pro

simulace homogenních kondenzovaných fází, kde nabídlý méně artefaktů, než simulace malých kapalných klastrů nebo simulace v tzv. odrazných okrajových podmínkách (angl. *bounce-back boundary conditions*). PBC byly použity i v projektech prezentovaných později v této práci.

Aby byla konformační dynamika ribozomu realistická, je nutné modelovat i vodné prostředí ribozomu. Jednoduché modely používají implicitní popis solvatace [47]. Studie v této práci používaly explicitní popis, kdy je studovaná biomolekula obalena velkým množstvím molekul vody. Model vody bývá spjatý s biomolekulovým silovým polem. Obvykle se používají tříšředové modely TIP3P [48] nebo SPC/E [49]. Solvataci v balíku GROMACS zajišťuje program `gmX solvate`, jehož vstupem je periodický box s biomolekulou.

V poslední fázi přípravy simulovaného systému je potřeba nastavit iontové podmínky simulace. V důsledku nabitých fosfátových částí rRNA a/nebo kyselých a bazických postranních řetězců některých aminokyselin je solvatovaný simulační box elektricky nabitý. Box s bakteriálním ribozomem má náboj přibližně 3500 e . Část tohoto náboje je kompenzována hořečnatými ionty, které bývají součástí experimentálních struktur a je vhodné je použít tak, jak jsou k dispozici. Neutralizace se zajišťuje přidáním K^+ iontů případně Cl^- , pokud je náboj rozpuštěné látky kladný. Pro realističtější popis rozpouštění se zároveň zavádí přebytečná (angl. *excess*) koncentrace dalších iontů, které reprezentují prostředí uvnitř buňky nebo experimentální podmínky. V případě ribozomu se používají K^+ , Na^+ a Mg^{2+} kationty doplněné Cl^- anionty. Dbá se zejména na správný popis experimentální iontové síly prostředí. Technicky se ionty do systému přidávají pomocí programu `gmX genion`, který funguje tak, že příslušný počet náhodně zvolených molekul vody nahradí ionty. Pro přesnější umístění iontů lze spočítat elektrostatický potenciál a ionty umístit do lokálních minim nebo maxim. V případě ribozomu je to ale výpočet náročný, a proto se nepraktikuje. Realistická poloha iontů se potom zajišťuje ekvilibračními simulacemi.

2.4 Přípravné fáze neboli ekvilibrace

Periodický box se solvatovaným ribozomem nelze pro MD simulace použít přímo. Během stavby simulačního systému je totiž časté, že některé meziatomové kontakty, obvykle mezi molekulami vody v 1. solvatační vrstvě a biomolekulou, jsou velmi blízko sebe. Systém se tak nachází ve stavu s vysokou potenciální energií. Může se jednat např. o coulombickou repulzi způsobenou souhlasně nabitými atomy blízko sebe. Numerická integrace Newtonových pohybových rovnic by z tohoto bodu vedla k numerické nestabilitě simulace a následné „havárii“, kdy by byl program s chybou ukončen. Proto je potřeba pomocí sady krátkých simulací připravit systém pro sběr dat (Obr. 2.2).

Energie, jako funkce kartézských souřadnic atomů, se numericky minimalizuje algoritmy založenými na sestupu ve směru gradientu energie. Nalezení globálního minima energie není pro vysokodimenzionální systém, jako je ribozom, možné. Není to ale pro potřeby MD simulace potřeba, neboť je postačující minimalizací odstranit pouze stupně volnosti s nejvyšší energií.

Vzorkování konformačního prostoru systému probíhá postupně. Nejprve jsou z Maxwellova–Boltzmannova rozdělení vygenerovány počáteční rychlosti a systém je uveden na cílovou teplotu, obvykle 300 K nebo 310 K. Může být výhodné nejprve ohřát vodu a poté až celý simulační box. Souřadnice rozpuštěné látky bývají v takovém případě udržovány pomocí harmonických pozičních omezení, aby v důsledku velkých sil nebyla počáteční konformace biomolekuly příliš narušena. Algoritmus pro udržení konstantní teploty – termostat – lze také použít odděleně pro biomolekulu a pro solvent, a udržovat teplotu biomolekuly nízko, např. na 10 K. Následně je dosaženo rovnovážného tlaku 1 bar použitím algoritmu pro konstantní tlak – barostatu.

Posiční omezení biomolekuly lze po ohřátí rozpouštědla uvolnit náhle. Pro velké systémy, jako je ribozom, je výhodné je uvolňovat postupně tím, že se silová konstanta v průběhu simulace snižuje k nule. Integrační krok ekvilibračních simulací bývá 1–2 fs. Délka simulací bývá krátká, v řádu desítek ps až ns.

2.5 Produkční simulace / sběr dat

Produkční simulací nazýváme finální simulaci, jejíž data použijeme pro ověřování hypotéz a formulování závěrů. Oproti ekvilibračním simulacím bývá produkční simulace delší (až 1000×) a některé algoritmy mohou být použity odlišně. Lze např. prodloužit integrační krok, neboť po ekvilibraci je systém více numericky stabilní. S využitím algoritmů jako *hydrogen mass repartitioning* [50] nebo *virtual sites* [51] lze Newtonovy rovnice integrovat s krokem 4–5 fs. Provázání systému s termostatem a barostatem může být méně intenzivní, neboť ekvilibrační simulace teplotu a tlak nastavují, zatímco produkční simulace je „jen“ udržuje. Výsledkem produkční simulace je trajektorie, v programu GROMACS soubor typu `xtc` a energetické veličiny v souboru typu `edr`.

Jelikož je pro simulace čas omezeným zdrojem,³ existují snahy vylepšit vzorkování konformačního prostoru. Triviálním postupem je simulovat mnoho kratších simulací namísto jedné dlouhé. Je potřeba zajistit nezávislost jednotlivých trajektorií např. použitím rozdílných počátečních konformací biomolekuly a/nebo různých počátečních rychlostí. Postup aplikoval např. projekt [Folding@home](#), který využívá spořiče obrazovek počítačů v domácnostech po celém světě [52, 53]. Omezení takové strategie si uvědomíme při představě extrémního případu mnoha simulací dlouhých jeden simulační krok. Je evi-

³A nejen pro ně! Viz [Čtyři tisíce týdnů](#) od O. Burkemana

dentní, že minimální délka simulace je dána rychlostí děje, který studujeme. Postupu bylo použito v několika projektech popsanych v kapitole 3 Výsledky a diskuze, kdy typická délka jednotlivých simulací byla několik stovek ns. Podobně bylo např. použito tisíce kratších simulací (v průměru 5 ns nebo 12,5 ns) ke studiu aktivace membránového receptoru [24]. Stojí také za zmínku, že závěry formulované na základě výsledků z jedné simulace mohou být chybné. Gapsys a de Groot nedávno zpochybnili [54] výsledky vědecké skupiny M. Karpluse, mj. nositele Nobelovy ceny za chemii za rok 2013, která popsala závislost konformačního chování hemoglobinu na velikosti simulačního boxu [55–57].

Velkou skupinou metod pro vylepšené vzorkování konformačního prostoru jsou metody založené na úpravě interakčního potenciálu, shrnuté např. v Ref. 58–60. Patří sem např. umbrella sampling simulace [61], řízené MD simulace [62], nebo metadynamika [63]. Jejich podstatou je cílené ovlivnění silového pole systému tak, aby simulace preferenčně vzorkovala oblasti konformačního prostoru s vyšší energií.

2.6 Analýzy

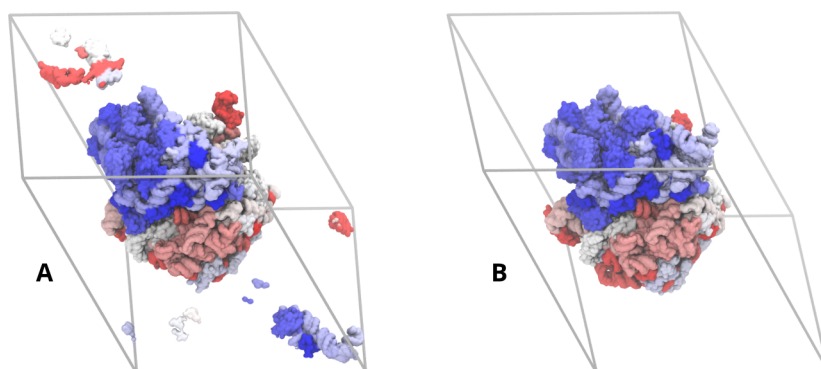
MD simulace produkují velké množství dat a souborů. V jednom z předešlých projektů jsme studovali relaxaci vody v závislosti na poloze náboje ve vlákne DNA [64]. Jelikož je relaxace rychlá, v řádu desítek femtosekund až pikosekund, uchovávali jsme souřadnice systému s vysokou frekvencí, což vedlo k vygenerování řádově terabytů dat za den.

V případě projektů popsanych v této práci sloužila voda pouze jako model prostředí, ve kterém se biomolekuly pohybují. Samotný pohyb molekul vody nebyl pro vědecké otázky podstatný.⁴ Pro pozdější analýzu jsme tedy uchovávali pouze kartézské souřadnice biomolekul.

Při použití periodických okrajových podmínek je často potřeba odstranit z trajektorie periodické přeskoky (Obr. 2.4). Jsou to artefakty vizualizace, kdy se biomolekula zdánlivě rozpadne do několika částí podle toho, ve kterém periodickém obraze simulace se právě nachází. Analýza takových trajektorií by poskytla řádné výsledky, ale implementace analytických nástrojů by se tím zbytečně komplikovala. K odstranění periodických přeskoků slouží program `gmx trjconv` s přepínačem `-pbc`.

Zjednodušeně lze říci, že z MD simulací nás zajímá několik typů výsledků: střední hodnoty, fluktuace, trendy a relaxace různých veličin. Soustředíme se na ně v závislosti na tom, zda studujeme termodynamiku rovnovážného stavu systému, nebo kinetiku určitého děje. Např. bychom mohli chtít porovnat simulační modely s experimentální strukturou určenou pomocí rentgenové krystalografie. Experimentální model reprezen-

⁴V souvislosti s ribozomem se vlastnostmi vody zabýval např. Lucent *et al.* [65].



Obrázek 2.4: A) Ribozom v periodickém boxu bez ošetřených periodických přeskoků. B) Ribozom s ošetřenými periodickými přeskoky. Barevná škála reprezentuje čísla reziduí od červené přes šedou k modré.

tuje souborový průměr konformací, neboli časový průměr přes dobu trvání experimentu. Dává proto smysl spočítat střední hodnoty kartézských souřadnic atomů v MD simulaci. Fluktuace popisují rovnovážnou dynamiku. Např. fluktuace kartézských souřadnic lze porovnat s teplotními faktory (B-faktory) krystalografických měření.

2.6.1 Základní analýzy

Pro simulace biomolekul je vhodné sledovat určitou „sadu“ veličin pokaždé, nezávisle na tom, jestli přímo souvisí se studovanou otázkou/hypotézou. Alespoň takové je subjektivní doporučení, které dávám studentkám a studentům při simulačních projektech.

Nejprve je vhodné spočítat střední hodnotu teploty, tlaku a hustoty a jejich drift. V prostředí GROMACS to lze provést jednoduše analýzou edr souboru pomocí programu `gmx energy`. NPT statistický soubor by měl vykazovat hodnoty nastavené v parametrech simulace a nulový drift. Ten je počítán lineárním fitem na data a reportovaný jako rozdíl fitu na začátku a na konci simulace.

Konformační dynamiku biomolekuly lze popsat odmocninou střední kvadratické odchylky poloh atomů, RMSD (z angl. *root-mean-square deviation*). RMSD představuje zobecněnou vzdálenost dvou sad atomů, např. referenční struktury a snímku z MD trajektorie, vypočtenou jako

$$\text{RMSD} = \sqrt{\frac{1}{N_a} \sum_i^{N_a} |\vec{r}_i - \vec{r}_{i,\text{ref}}|^2}, \quad (2.2)$$

kde r_i je polohový vektor i -tého atomu, $r_{i,\text{ref}}$ je jeho referenční hodnota a součet jde přes počet atomů N_a . Význam má RMSD až po odstranění celkové translace a rotace biomolekuly tzv. superpozicí dvou struktur. Běžně se vykresluje RMSD jako funkce simulačního

času. Jako referenční se běžně bere počáteční konformace biomolekuly, což bývá obvykle struktura pocházející z nějakého experimentu. V rovnovážné simulaci by se RMSD měla po určitém čase ustálit. Pro proteinovou páteř je typickou hodnotou, na které se RMSD ustálí, 1,5–2,5 Å, pro páteř celého ribozomu (nebo atomy C_α a P) hodnota 4–6 Å.

Informaci o dynamice různých částí biomolekuly získáme např. pomocí odmocniny středních kvadratických fluktuací atomů, RMSF (z angl. *root-mean-square fluctuation*). Jedná se o veličinu, která popisuje směrodatnou odchylku polohy i -tého atomu vypočtenou jako

$$\text{RMSF}_i = \sqrt{\frac{1}{N_f} \sum_f (\vec{r}_{i,f} - \langle \vec{r}_i \rangle_f)^2}, \quad (2.3)$$

kde $\vec{r}_{i,f}$ je polohový vektor i -tého atomu v f -tém snímku trajektorie, $\langle \cdot \rangle_f$ značí střední hodnotu přes snímky simulace a součet jde přes počet snímků trajektorie N_f . RMSF se vynáší do grafů jako funkce čísel atomů. Při výpočtu RMSF se tiše předpokládá⁵, že hustota pravděpodobnosti polohy vybraného atomu je normální (gaussovská). To ale často není splněno. Můžeme si představit smyčku určité biomolekuly, která se může vyskytovat ve dvou konformačních stavech. Hustoty pravděpodobností poloh atomů smyčky by v takovém případě byly bimodální, čímž by předpoklad normality nebyl splněn. Veličina RMSF jako kvalitativní měřítko flexibility by i v takovém případě zřejmě posloužila spolehlivě.

Velikost biomolekuly lze charakterizovat gyračním poloměrem R_g . Ten představuje vzdálenost od těžiště molekuly, ve které by musela být umístěna jedna částice o hmotnosti biomolekuly, která by měla stejný moment setrvačnosti jako biomolekula. R_g se vypočítá podle

$$R_g = \sqrt{\frac{\sum_i^{N_a} m_i |\vec{r}_i|^2}{\sum_i^{N_a} m_i}}, \quad (2.4)$$

kde i -tý atom má hmotnost m_i a $|\vec{r}_i|$ je jeho vzdálenost od těžiště biomolekuly. Součet jde přes celkový počet atomů N_a . Gyrační poloměr představuje vhodnou charakteristiku pro molekuly, u nichž se očekává proměnlivá velikost, např. IDPs.

RMSD, RMSF i R_g jsou příklady tzv. kolektivních souřadnic. Jedná se o veličiny, které jsou funkcí kartézských souřadnic atomů a snižují dimenzionalitu problému: namísto $3N_a$ -rozměrného vektoru popisuje konformaci jedno číslo (např. R_g). Nalezení vhodných kolektivních souřadnic pro popis dynamiky systému představuje jednu z rozhodujících výzev pro zodpovězení studovaných vědeckých otázek [66, 67].

⁵angl. *tacit assumption*

2.6.2 Analýza sekundární struktury

Struktura proteinů je hierarchická, rozlišujeme tedy primární, sekundární, terciární a kvarterní strukturu. Primární struktura proteinu je nutnou vstupní informací do MD simulací a není potřeba ji v simulacích analyzovat – je neměnná. Motivy sekundární struktury jsou klíčové pro základní koncepty proteinové chemie, jako jsou proteinové domény nebo rodiny foldů [68]. Sekundární struktura počáteční konformace je známá, ale může se v průběhu MD simulace v důsledku termálních pohybů a konformačních přechodů měnit.

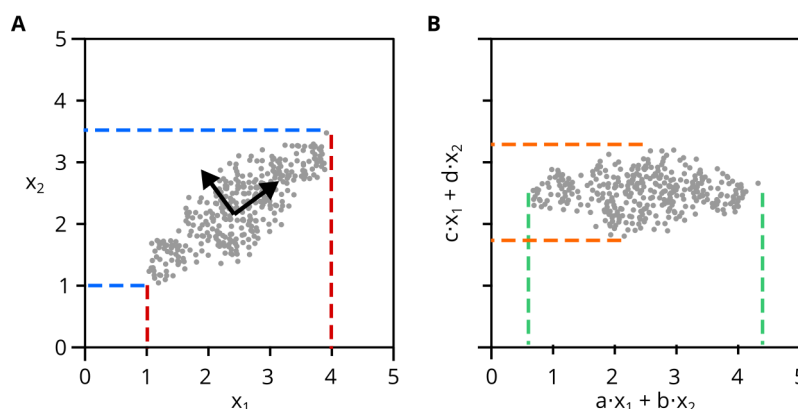
Existuje několik klasifikátorů sekundární struktury, mezi populární patří DSSP [69], STRIDE [70], nebo z novějších např. DISICL [71]. V některých projektech popsanych v této práci zaměřených na krátké peptidy jsme používali klasifikátor DSSP založený na analýze vodíkových vazeb proteinové páteře. DSSP rozlišuje 7 strukturních motivů (uvádím zde jejich anglické názvy): α -helix, 3/10-helix, π -helix, β -strand, β -bridge, turn, bend. Osmým DSSP motivem je coil, tedy část proteinu bez sekundární struktury.

Analýzu sekundární struktury DSSP lze provést např. programem `gmx do_dssp`, který se odkazuje na externí DSSP program. Ten lze získat zdarma online. Výstupem DSSP je klasifikace každé aminokyseliny do jedné z osmi kategorií pro každý snímek trajektorie. Pro typickou MD simulaci typického proteinu je DSSP klasifikace značně nepřehledná a je nutné ji dále zpracovat, např. vhodně průměrovat nebo vytvořit histogramy.

2.6.3 Analýza hlavních komponent

Atomistické MD simulace jsou mnohorozměrný problém. Smysluplně charakterizovat taková data je hlavním úkolem vědkyň a vědců v oboru. Populárním statistickým nástrojem pro zpracování mnohorozměrných dat je tzv. analýza hlavních komponent, PCA (z angl. *principal component analysis*). Tu lze s úspěchem aplikovat i na MD simulace biomolekul [72]. Detailní popis PCA lze nalézt v učebnicích statistiky. Šikovný úvod nabízí Ref. 73, aplikace na biomolekulární MD simulace shrnuje např. Stein *et al.* [74]. Zde se omezím na úsporný popis PCA. Dvourozměrný případ je vysvětlený na Obr. 2.5.

Pro jednoduchost si představme sbalený protein např. o délce 100 reziduí, jehož terciární struktura je dobře definovaná (Obr. 2.6). Fluktuace struktury v průběhu MD simulace jsou malé až na jednu výjimku – smyčku řekněme o 10 reziduích, která fluktuuje víc než zbytek proteinu. Její fluktuace jsou navíc spojené (korelované) s pohybem N-konce proteinu. Chtěli bychom objektivně identifikovat pohyb této smyčky, např. jednoznačně říci, která rezidua se pohybují více, případně pohyb kvantifikovat. Z Obr. 2.6C lze dovést, že pro (velkou) fluktuaci smyčky nebude (malá) fluktuace rezidua na opačné straně proteinu důležitá. Zároveň je také zřejmé, že celkové fluktuace proteinu budou dány především fluktuací smyčky. PCA umožňuje identifikovat kolektivní souřadnice, které



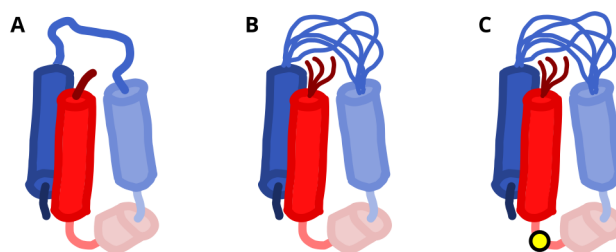
Obrázek 2.5: Uvažujme data ve dvou rozměrech x_1 a x_2 , což mohou být např. nějaké dvě přímo měřené veličiny. Jejich hodnoty jsou vyneseny v grafu A. Rozsah dat ve směru x_1 je od 1 do 4 (červeně), ve směru x_2 od 1 do 3,5 (modře). Rozsahy jsou tedy přibližně stejné. Jenže z rozmístění bodů v grafu A jde vidět, že data mají „tvar“ elipsy a daly by se tedy najít jiné dva charakteristické směry, znázorněné šipkami, které by data popisovaly vhodněji. Nové rozměry jsou lineární kombinací původních x_1 a x_2 . PCA poskytuje rozvojové koeficienty $a-d$ takové, že je rozsah dat (přesněji variance) v daných rozměrech maximalizovaný. Transformovaná data jsou v grafu B. Jde vidět, že rozsah v horizontálním směru (zeleně) je mnohem větší než ve vertikálním (oranžově). Pro každý bod bychom proto mohli vynechat informaci o hodnotě $c \cdot x_1 + d \cdot x_2$. Podstatné informace o datech by přesto zůstaly zachovány.

málo fluktuují, a odfiltrovat je z dalších analýz.

Konformaci proteinu lze popsat vektorem $\vec{k}$ (jako konformace).

$$\vec{k} = (x_1, y_1, z_1, x_2, y_2, z_2, x_3, \dots, z_{N_a}), \quad (2.5)$$

kde x_i, y_i, z_i jsou kartézské souřadnice i -tého atomu. Vektor $\vec{k}$ má $3N_a$ prvků a v průběhu simulace se mění. Každý snímek trajektorie je reprezentován jedním bodem v tomto $3N_a$ -rozměrném prostoru. Podstatou PCA je sestrojení kovarianční matice $\mathbf{C}$ a její matematické zpracování. Účelem je najít ortogonální báze vektory PCs (z angl. *principal components*) takové, aby maximalizovaly mnohorozměrnou varianci původních dat



Obrázek 2.6: Modelový protein značený barevnou škálou od červené na N-konci k modré na C-konci. A) Terciární struktura proteinu je dobře definovaná. Protein se skládá ze čtyř α -helixů. B) N-konec a smyčka spojující helix-3 a helix-4 vykazují vysoké fluktuace. C) Reziduum na opačné straně proteinu (žlutý kruh) se smyčkou a N-koncem neinteraguje.

(Obr. 2.5).

Obecná kovarianční matice se konstruuje z N -rozměrného vektoru souřadnic

$\vec{x} = (x_1, x_2, x_3, \dots, x_N)$ a její prvky c_{ij} jsou definovány jako

$$c_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle, \quad (2.6)$$

kde $\langle \cdot \rangle$ značí střední hodnotu. Diagonální členy, kde $i = j$, jsou variance jednotlivých prvků vektoru $\vec{x}$. Střední hodnoty se počítají přes statistický soubor, v případě simulací tedy přes snímky trajektorie. Diagonalizací $\mathbf{C}$ získáme N vlastních čísel a N vlastních vektorů (PC). Každému PC přísluší jedno vlastní číslo, které charakterizuje důležitost vektoru pro popis celkové variance.

PCA lze proto vnímat jako transformaci původních souřadnic do nových souřadnic, přičemž nové souřadnice jsou seřazené podle klesající důležitosti pro popis fluktuací v systému. První báze vektor PC1 reprezentuje nejvýznamnější pohyb biomolekuly. Je dobré zmínit, že biologická relevance takového pohybu není garantována. Biomolekula např. nemusí reálně vykonávat pohyb v PC1, ale ve směru, který je kombinací více PCs.

2.6.4 Analýzy šité na míru

Velká rozmanitost biomolekulárního světa neumožňuje nalézt univerzální sadu analytických nástrojů a strategií. Je to úkol vědkyň a vědců, aby našli vhodné metriky pro pochopení biomolekulární dynamiky. Řada analýz proto proto vzniká souběžně s projektem, během toho, jak se vyjasňují různá zákoutí problému, hodnoty měření nebo simulační výsledky.

Mezi aktuální přístupy k analýze simulací patří Markovovské stavové modely [75], nástroje adaptivních simulací [76, 77], techniky založené na strojovém učení [78, 79], jejich vzájemné kombinace, apod. Některé z nich byly použity i u projektů popsanych v této práci.

Kapitola 3

Výsledky a diskuze

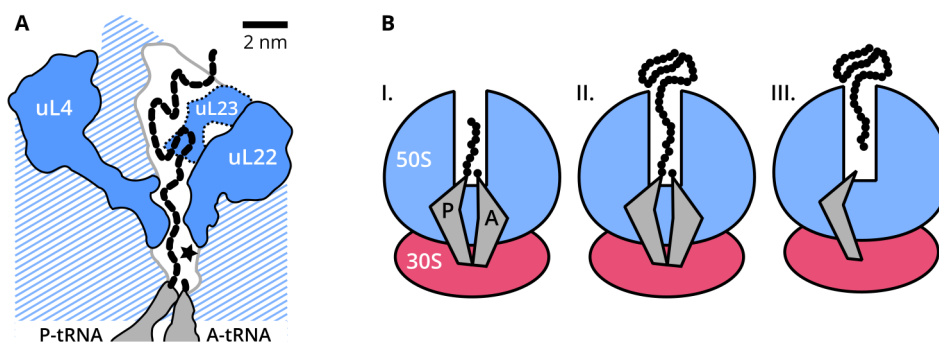
V této kapitole shrnuji několik projektů z posledních asi 5 let, v nichž jsem hrál vůdčí roli a které se týkají otázek o ribozomu. K jejich zodpovězení významně přispěly atomistické MD simulace. Studentky a studenti podílející se na výzkumu zpracovávali své bakalářské, diplomové, nebo dizertační práce pod mým vedením na VŠCHT Praha.

Několik vědeckých publikací souvisejících s projekty prošlo recenzním řízením. Jedná se o práci v Chemických listech, kde na populárně-naučné úrovni popisují translaci a zavádím koncept tří etap elongace proteinu [80], práci v Nucleic Acids Research o translačním uvěznění VemP [81], práci v recenzním řízení v Biophysical Journal o allosterickém působení peptidové deformylázy na povrchu ribozomu [82] a dvou souborných textech v Current Opinion in Structural Biology [10] a Annual Reviews in Biophysics [8]. Jelikož jsou práce dostupné online díky Open Access modelu, v případě allostérie jako [preprint](#), rozhodl jsem se jejich plné texty k habilitační práci nepřikládat jako přílohu, ač to bývá zvykem.

Projekty, které zatím recenzním řízením neprošly, popisují pečlivěji. Jedná se tedy o nové vědecké poznatky ve smyslu § 72 odst. 3 zákona o vysokých školách. Bez výjimky plánuji jejich odeslání k recenznímu řízení v některém z vědeckých časopisů o biofyzice, molekulární biologii, fyzikální chemii, apod., jakmile to situace umožní. Zejména sekce 3.2 Dekódování nestandardního genetického kódu a 3.4 Konformační preference fragmentů ribozomálních proteinů jsou úzce propojené s prací experimentálních skupin a jejich postup není plně v mé režii.

3.1 Translační uvěznění peptidu VemP

Vznik peptidových vazeb mezi aminokyselinovými zbytky probíhá v katalytickém centru uvnitř velké ribozomální podjednotky. Z katalytického centra vede na povrch ribozomu



Obrázek 3.1: A) Schéma ribozomálního tunelu ve velké ribozomální podjednotce. Šrafovaná plocha značí rRNA, modře jsou vybarveny r-proteiny, které u bakterií zasahují do tunelu a spolu-vytváří jeho stěny. Rodící se protein je znázorněn uvnitř tunelu černě. Hvězdička značí vazebné místo makrolidových antibiotik. B) Tři etapy elongace rodícího se proteinu. Modře je znázorněna velká ribozomální podjednotka 50S, červeně malá podjednotka 30S, šedě jsou tRNA v A a P místech ribozomu. Upraveno z Ref. 80.

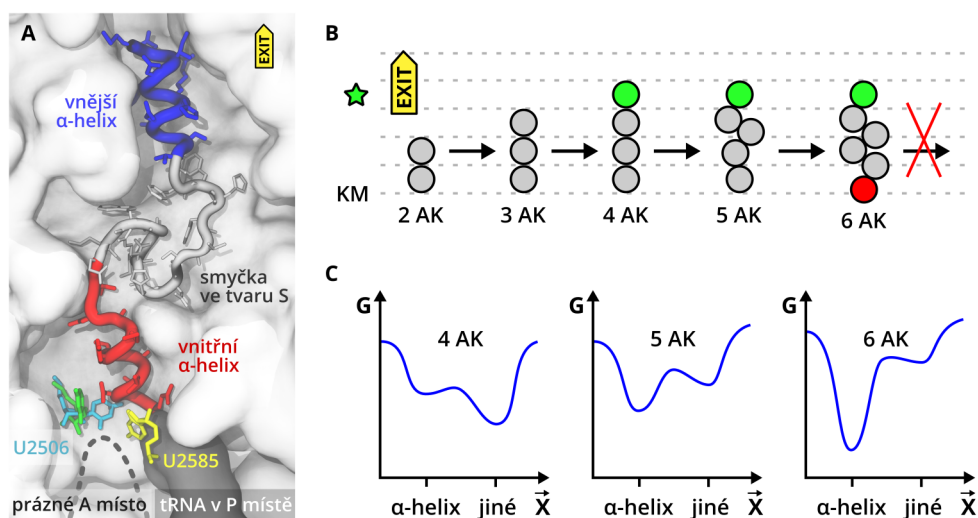
asi 10 nm dlouhý tunel, kterým rodící se protein opouští ribozom (Obr. 3.1A). Existence tunelu vyplynula v 60. letech 20. století z experimentů, které naznačovaly, že část rodícího se proteinu je chráněna před působením hydrolytických enzymů [83]. Cryo-EM struktury přítomnost tunelu ve velké podjednotce prokázaly [84].

Po objasnění atomární struktury velké podjednotky [11] se předpokládalo, že tunel funguje pouze jako pasivní prostor [85]. Ve skutečnosti ale tunel tvoří vysoce funkční prostředí, které s rodícím se proteinem specificky interaguje [86]. Např. se do ribozomálního tunelu váže významná třída klinicky používaných antibiotik [87] (Obr. 3.1A). Poměry uvnitř tunelu ovlivňují rychlost, s jakou se rodící se protein prodlužuje [88]. Za určitých okolností může dojít k tzv. translačnímu uvěznění (angl. *translational arrest* [89, 90])¹. V katalytickém místě ribozomu nemůže během uvěznění dojít k vytvoření nové peptidové vazby (důvody mohou být různé), zároveň však nemůže být ribozom recyklován, neboť rodící se protein nebyl ještě z ribozomu uvolněn, což je nutný předpoklad pro zahájení recyklace [91].

3.1.1 VemP v tunelu a mimo něj

Během své postdoktorské stáže na Ústavu Maxe Plancka pro biofyzikální chemii v Göttingenu (SRN) jsem se začal zabývat peptidem VemP (z angl. *Vibrio export-monitoring peptide*), který translační uvěznění způsobuje u druhu bakterie rodu *Vibrio* žijící v brakických vodách. Translační uvěznění slouží bakterii k regulaci exportu proteinů skrz membránu ven z buňky, změnil-li se koncentrace Na^+ iontů ve vnějším prostředí. Mechanismus působení VemP navržený Ischii *et al.* je fascinující a ukazuje, jak významné může mít translační uvěznění ribozomu fyziologické účinky [92].

¹Český ekvivalent jsem zavedl v Ref. 80



Obrázek 3.2: A) Struktura peptidu VemP uvnitř ribozomálního tunelu za stavu translačního uvěznění. α -helixy jsou znázorněny modře a červeně, nukleotidy v katalytickém centru ribozomu související s translačním uvězněním jsou žlutě a zeleně. B) Schéma prodlužování rodičního se proteinu. Protein se prodlužuje až do chvíle, než kotva (zelené kolečko) začne interagovat se částí stěny tunelu (úroveň se zelenou hvězdou). Poté je elongace zpomalena, což vede k tvorbě α -helixu, jenž pro délku modelového proteinu 6 aminokyselinových zbytků (6AK) způsobí v katalytickém místě (KM) translační uvěznění. C) Schéma možných profilů Gibbsovy energie G jako funkce zobecněné konformační souřadnice $\tilde{X}$. V průběhu prodlužování se mění profil G ve prospěch α -helikální konformace, což vede k translačnímu uvěznění. Upraveno z Ref. 81.

Cryo-EM struktura komplexu ribozom–VemP odhalila, že VemP zaujímá v ribozomálním tunelu značně kompaktní konformaci [93]. Během translačního uvěznění tvoří VemP dva α -helixy propojené smyčkou ve tvaru S (Obr. 3.2A). Přítomnost vnitřního α -helixu v katalytickém místě byla označena jako příčina translačního uvěznění [93]. Důvody, které k tomu vedou, zůstaly neobjasněny.

Se skupinami v Göttingenu (SRN) a Tempe (Arizona, USA) jsme si položili otázku, proč VemP zaujímá konformaci, která vede k translačnímu uvěznění. Preference VemP tvořit α -helixy může být zakódována v primární struktuře (tj. sekvenci reziduí), α -helixy mohou být indukovány prostředím tunelu specifickými interakcemi, nebo nespecificky (např. permitivitou prostředí), případně kombinací více faktorů.

Provedli jsme MD simulace kompletního *E. coli* ribozomu s VemP v tunelu a MD simulace VemP mimo ribozom ve vodném prostředí. Zároveň jsme využili algoritmů predikujících motivy sekundární struktury z primární struktury peptidu a spektroskopie cirkulárního dichroismu (CD) VemP ve vodě a vodných roztocích 2,2,2-trifluorethanolu, jež měly napodobovat některé vlastnosti prostředí uvnitř tunelu. Kombinací simulačních a experimentálních technik jsme ukázali, že preference tvořit α -helix mimo ribozom je pro VemP velmi nízká [81]. To podpořilo závěr, že sekundární struktura je indukovaná ribozomálním tunelem.

Atomistické MD simulace ribozomu také naznačily, že část smyčky VemP propojující oba α -helixy vykazuje nízkou strukturní dynamiku (dle RMSF jednotlivých reziduí). Navrhli jsme proto model translačního uvěznění VemP (Obr. 3.2B), ve kterém část smyčky slouží jako kotva, která specifickými interakcemi mezi VemP a stěnami tunelu zpomaluje posun VemP tunelem a umožňuje tvorbu vnitřního α -helixu, jež vede k uvěznění.

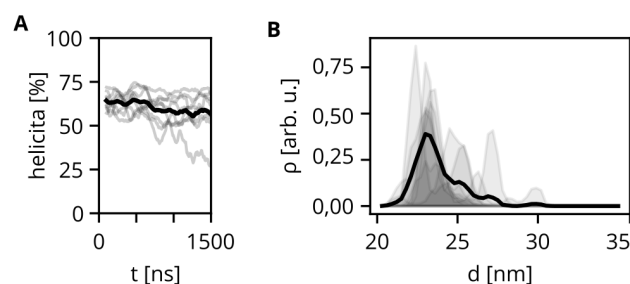
3.1.2 VemP11

Kromě simulací VemP o plné délce odpovídající cryo-EM struktuře (36 reziduí) [93], jsme provedli MD simulace zkráceného VemP o délce 11 reziduí (VemP11), tedy vnitřního α -helixu (Obr. 3.2A). Předchozí biochemické experimenty ukázaly, že zkrácené varianty VemP mají nižší schopnost způsobovat translační uvěznění – varianta o 19 reziduích vykazovala 10 % efektivitu oproti VemP o plné délce [93]. O zkoumaném VemP11 jsme proto předpokládali, že translační uvěznění nezpůsobuje. Zajímalo nás, jakým způsobem probíhá rozbalování VemP11 v tunelu a zda-li je rozbalení spojeno s uvolněním nukleotidů U2506 a U2585. Ty se nachází v katalytickém centru ribozomu a hrají hlavní roli v mechanismu translačního uvěznění. Dle Su *et al.* indukuje vnitřní α -helix konformační změnu U2506 a U2585, která je nekompatibilní s vazbou aminoacylované tRNA do A místa ribozomu.

Během osmi nezávislých 1,5 μ s dlouhých simulací komplexu ribozom–VemP11 ke kompletnímu rozbalení nedošlo. Pouze v jednom případě se helicity VemP11 snížila z 65 % na přibližně 30 % (Obr. 3.3A) [81]. Délka VemP11 d , definovaná jako Z-složka vzdálenosti těžiště N-koncového rezidua a těžiště skupiny C-koncového rezidua VemP11 a tří 3'-terminálních nukleotidů P-tRNA, vykazovala nejpravděpodobnější hodnotu okolo 2,3 Å. Delší VemP11 se v simulacích v tunelu vyskytoval velmi zřídka (Obr. 3.3B). MD simulace i CD spektroskopie jednoznačně ukázaly, že VemP11 ve vodném prostředí α -helix netvoří [81].

Ve zkoumání konformačních preferencí VemP11 jsme proto pokračovali, neboť se zdálo, že uvnitř tunelu existuje bariéra Gibbsovy energie, která zpomaluje přechod VemP11 z helikální do rozpletené konformace. Podstata takové bariéry nebyla známa.

Se studentkou I. Švecovou jsme se v rámci její bakalářské práce zabývali MD simulacemi VemP11 v ribozomálním tunelu [94]. Oproti předešlým MD simulacím bylo prostředí ribozomu modelováno redukovane: ribozomální tunel byl z ribozomu „vyříznut“ a na okrajích zafixován (viz sekce 2.3 Stavba simulovaného systému). To vedlo k dramatickému snížení výpočetních nároků a možnosti studovat rozbalování VemP10 pomocí řízených MD simulací a metodou umbrella sampling (US) [61, 95]. Řízené MD i US simulace byly provedeny v nativní implementaci balíku GROMACS za využití kolektivní souřadnice d odpovídající délce VemP11 a tří nukleotidů tRNA, přibližně v rozsahu



Obrázek 3.3: Charakteristiky VemP11 ze simulací ribozom–VemP11 [81]. Data z 8 nezávislých simulací jsou šedě, jejich průměr černě. A) Helicita jako funkce simulačního času t . B) Hustota pravděpodobnosti ρ vzdálenosti těžišť C- a N-konce VemP11.

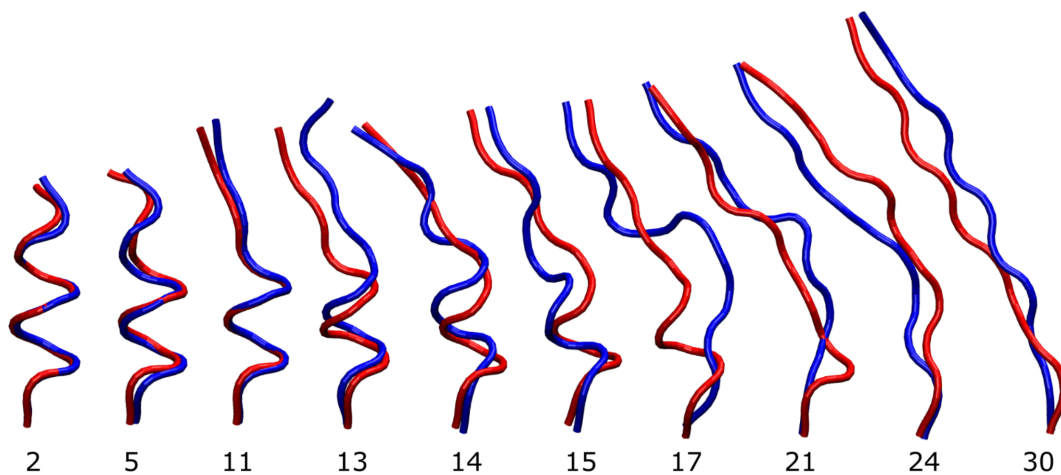
2,2–4,5 nm. Jelikož se v cryo-EM struktuře komplexu ribozom-VemP vyskytuje nukleotid U2506 v katalytickém místě ve dvou stejně populovaných konformacích, označených A a B (Obr. 3.2A), zaměřili jsme se ve studiu i na význam počáteční konformace U2506 na rozbalování VemP11.

Studentka provedla simulace 32 umbrella oken každé o délce 150 ns. Počáteční konformace VemP11 byly vybrány ze 100 ns dlouhé řízené MD simulace, během níž se VemP11 rozpletl. Silová konstanta umbrella oken byla zkusmo optimalizována vzhledem k překryvu histogramů studované souřadnice mezi jednotlivými okny. Pro produkční simulace byla použita hodnota 5000 kJ mol^{-1} . První a poslední okno bylo prodlouženo na délku 1 μs . Silové pole a nastavení simulací odpovídalo předešlému simulačnímu protokolu s tím rozdílem, že nebyl simulován celý ribozom, ale pouze jeho výřez obsahující tunel a jeho okolí.

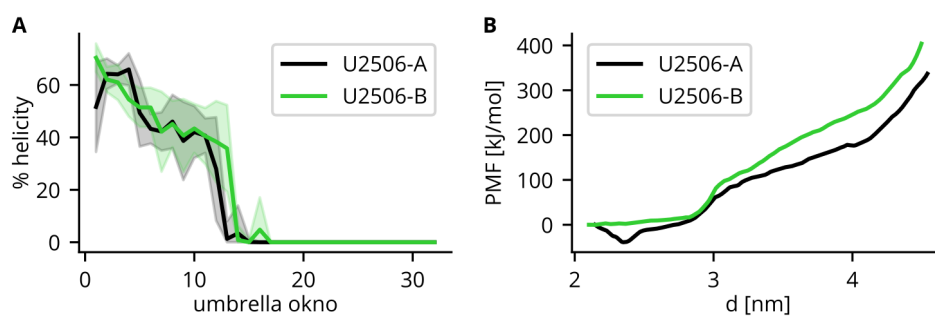
V daném simulačním modelu se VemP11 během řízené MD simulace rozbaloval postupně od N-konce směrem k C-konci, o čemž vypovídají také reprezentativní struktury pro jednotlivá umbrella okna na Obr. 3.4. Naznačuje to, že síla, která působí na N-konec VemP11 se obtížně přenáší směrem ke katalytickému centru. To je v souladu s předchozími řízenými MD simulacemi kompletního VemP [96].

Dle algoritmu DSSP se helikální charakter peptidu kompletně ztratil přibližně ve 14. umbrella okně, což odpovídá délce VemP11 asi $3,2 \text{ \AA}$. To je zřejmé ze závislosti procentuálního zastoupení α -helixu (helicity) VemP11 na umbrella okně na Obr. 3.5. Ztráta helicity nebyla v simulacích závislá na počáteční konformaci U2506. Zároveň je evidentní (Obr. 3.4), že snížení helicity není korelováno s natažením peptidu.

Potenciál střední síly (PMF, z angl. *potential of mean force*) odpovídá závislosti Gibbsovy energie (simulace poskytly izobaricko-izotermický statistický soubor) na vybrané kolektivní souřadnici. Profil PMF na Obr. 3.5B ukazuje, že uvnitř tunelu je rozpletení VemP11 charakterizované kolektivní souřadnicí d energeticky nevýhodné pro obě počáteční konformace U2506.



Obrázek 3.4: Schéma rozbalování VemP11 během umbrella sampling simulací. Číslo odpovídá oknu. Modře je znázorněn systém s U2506-A, červeně s U2506-B. Převzato z bakalářské práce I. Švecové [94].



Obrázek 3.5: Umbrella sampling MD simulace VemP11 uvnitř ribozomálního tunelu pro dvě startovní konformace nukleotidu U2506 u C-konce VemP11. A) Procentuální zastoupení α -helikální konformace jako funkce umbrella okna. Linky znázorňují střední hodnoty přes simulační čas umbrella okna (150 ns), plochy odpovídají směrodatné odchylce. B) Potenciál střední síly (PMF, z angl. *potential of mean force*) jako funkce délky VemP11 d , tedy Z-složky vzdálenosti těžišť reziduí na C- a N-konci peptidu.

Patrný je ale vliv U2506 na tvar PMF; zatímco pro systém s počáteční konformací U2506-A vykazuje PMF bariéry pro rozbalení dvě (pro d okolo 2,4 Å a 3,0 Å), u systému s U2506-B je bariéra pouze jedna (3,0 Å). Zajímavý je rozdíl pro malé délky peptidu. Pro U2506-A vykazuje PMF o 40 kJ mol^{-1} hlubší minimum oproti systému U2506-B. Systém U2506-B má křivku PMF prakticky plochou až do délky $d = 3 \text{ Å}$. Rozdíl v počátečních konformacích nukleotidu U2506 má tedy význam pro stabilizaci konformace s malou délkou VemP11.

Profil PMF pro U2506-B zároveň naznačuje, že natažení VemP11 je v souřadnici d energeticky málo náročné a za teploty systému 310 K by mělo být dostupné. To souhlasí s pozorováním v simulacích VemP11 v kompletním ribozomu (Obr. 3.3), kde VemP11 zůstává ochotně v helikální konformaci a jeho délka d se pohybuje v rozmezí do 3 Å. Další natažení VemP11 je energeticky nevýhodné nezávisle na konformaci U2506, přičemž z US simulací plyne, že bariéra Gibbsovy energie má výšku 80–100 kJ mol^{-1} (Obr. 3.5B).

Dynamika VemP11 probíhá na mnohorozměrné ploše Gibbsovy energie. Ačkoliv PMF do 3 Å pro U2506-B nevykazuje bariéru pro rozbalení v souřadnici d , není vyloučeno, že v jiné souřadnici taková bariéra existuje. Vzhledem ke konformačnímu chování VemP11 uvnitř tunelu je pravděpodobné, že bariéru je nutné popsat v jiných kolektivních souřadnicích.

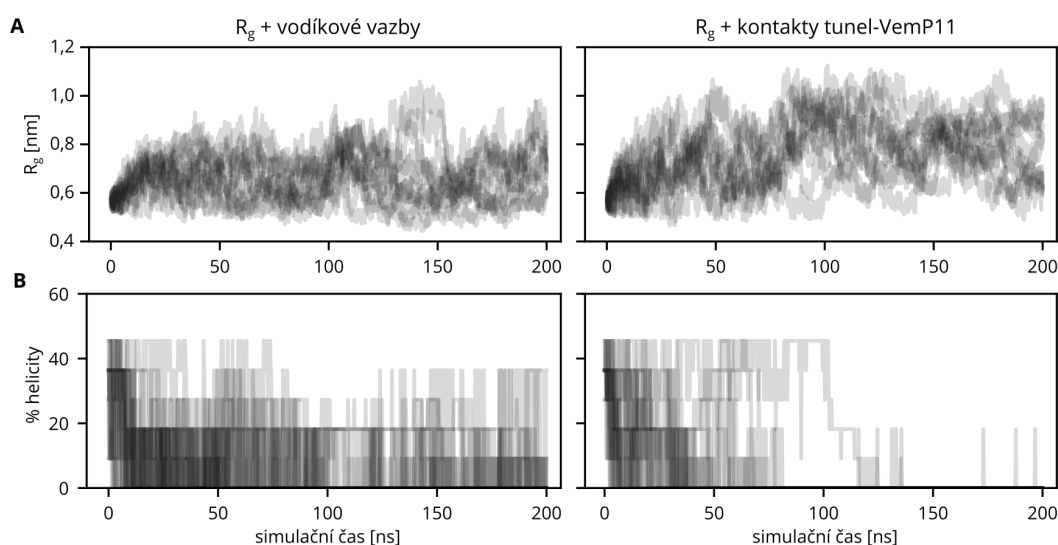
3.1.3 Metadynamika mutantů VemP11

Abychom lépe prozkoumali předpokládanou mnohorozměrnou energetickou bariéru rozpletení VemP11 v ribozomálním tunelu, provedli jsme dodatečné simulace za využití metadynamiky (metaD) [63]. Simulace byly součástí tréninku PhD studentky M. Černekové a byly provedeny se srovnatelným nastavením jako předešlé US simulace, t.j. s využitím výřezu tunelu a jeho okolí. MetaD byla provedena pomocí komunitně vyvíjené knihovny Plumed [97, 98] propojené s balíkem GROMACS.

Simulovali jsme několik systémů s VemP11. Kromě dvou A a B startovních konformací nukleotidu U2506 jsme simulovali VemP11 s kladně nabitým N-koncem a s N-koncem bez náboje, tj. nesoucím acetylovou skupinu. Pro každý systém jsme provedli tři nezávislé metaD simulace lišící se počátečními rychlostmi, přičemž ekvilibrační simulace běžely pokaždé s jinými počátečními rychlostmi. Startovní konformace VemP11 pocházela vždy z cryo-EM [93].

MetaD používaly dvě kolektivní souřadnice. Z-složku vzdálenosti těžišť z US simulací jsme nahradili obecnějším gyračním poloměrem R_g , který umožňuje dodávat energii i do stupňů volnosti, které souvisí se šířkou VemP11. R_g jsme kombinovali s tzv. CMAP souřadnicí, která charakterizuje meziatomové kontakty.

První metaD jsme provedli s CMAP popisující intramolekulární vodíkové vazby VemP11,



Obrázek 3.6: Metadynamika VemP11. Data reprezentují 12 simulací (3 nezávislé simulace pro každý ze 4 systémů). A) Gyrační poloměr R_g VemP11 jako funkce simulačního času t . B) Procentuální zastoupení reziduí VemP11 v helikální konformaci jako funkce simulačního času t .

abychom ověřili, nakolik brání rozpletení α -helixu. Vzhledem k předešlým CD experimentům a MD simulacím komplexu ribozom–VemP11 jsme očekávali, že pro stabilizaci helikální struktury velkou roli hrát nebudou. Další sadu metaD simulací jsem provedl s CMAP pro intermolekulární interakce. Vizuální inspekci a rešerší dostupné literatury jsme vybrali dvě interakce, které jsme ovlivňovali dodatečným potenciálem. Jednalo se o interakce G2505–Val151 a C2610–Tyr150.

Obr. 3.6 demonstruje rozdíl mezi MetaD s intramolekulárními a intermolekulárními CMAP souřadnicemi. Horní panely ukazují vývoj R_g ve 12 simulacích několika systémů s VemP11 (nelze je proto průměrovat). MetaD s intramolekulární CMAP (vodíkové vazby) vykazuje pomalejší nárůst R_g a v pozdějších fázích simulací také nižší vzorkované hodnoty R_g oproti metaD s intermolekulární CMAP (kontakty tunel–VemP11). Dolní panely Obr. 3.6 ukazují procentuální zastoupení helikálních reziduí (dle DSSP) jako funkci simulačního času. Narušení intramolekulárních vodíkových vazeb vedlo ve většině simulací ke snížení helicity z počátečních 50 % na úroveň 10–30 %. α -helix VemP11 se tak rozpletl pouze částečně. Oproti tomu narušení dvou intermolekulárních kontaktů mezi tunelem a VemP11 vedlo ve všech 12 simulacích ke kompletní ztrátě α -helikální konformace dle DSSP. Je potřeba zmínit, že v těchto simulacích byly intramolekulární vodíkové vazby popsány standardním způsobem, nebyly tedy narušeny dodatečným potenciálem.

MetaD simulace potvrzují hypotézu, že VemP11 je stabilizován zejména nekovalentními vazbami se stěnami tunelu, zatímco intramolekulární vodíkové vazby přispívají méně. Kvantifikace těchto příspěvků především z hlediska Gibbsovy energie je naším úkolem v následné práci. Zároveň sestavujeme systémy, kde budou interakce VemP11 narušeny

nikoli přidáním potenciálem, ale bodovými mutacemi.

Mimo to se se studentem J. Cikhartem v rámci jeho diplomové práce věnujeme studiu různě dlouhých analogů VemP, jak byly studovány v Ref. [81] pomocí CD spektroskopie. Pomocí simulací ve vodných roztocích 2,2,2-trifluorethanolu bychom chtěli vysvětlit molekulární podstatu konformačního chování peptidů v rozpouštědle.

3.2 Dekódování nestandardního genetického kódu

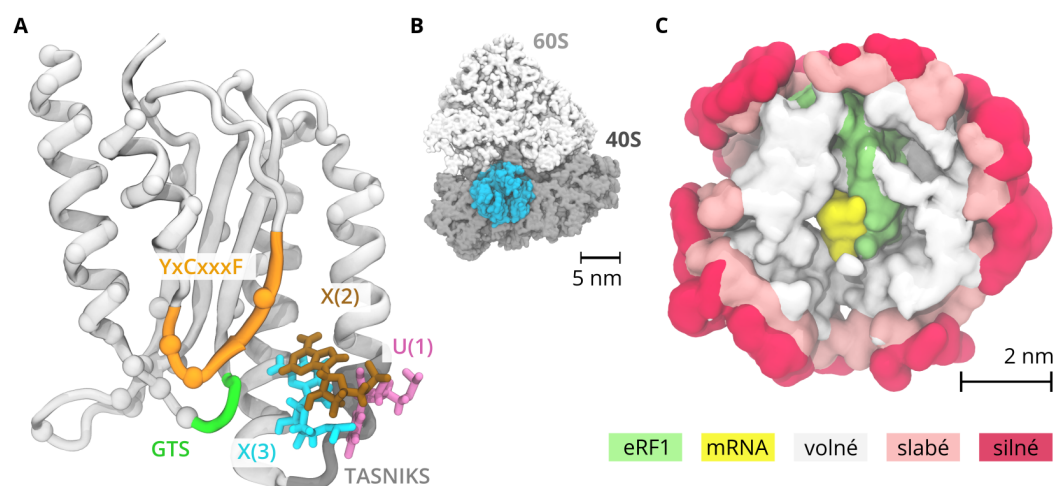
Malá ribozomální podjednotka je zodpovědná za čtení genetického kódu dočasně uloženého v mRNA. Každá trojice nukleotidů na mRNA (kodon) je rozpoznávána trojicí nukleotidů na aminoacylované tRNA (antikodon), která se váže do ribozomu. Tím je definován genetický kód, tedy sada pravidel, podle nichž se sekvence nukleotidů v mRNA překládá do sekvence rodícího se proteinu.

Některé kodony mají speciální význam. Např. kodon AUG běžně slouží jako start kodon na začátku překládané sekvence. V bakteriích jsou tři kanonické terminační kodony UAA, UAG, UGA, které naopak signalizují konec translace (označují se také jako stop kodony). Ty nejsou rozpoznávány tRNA, ale specializovanými proteiny, tzv. release faktory (RF). RF rozpoznává stop kodon v mRNA a v katalytickém místě ribozomu umožňuje hydrolýzu vazby mezi tRNA v P místě a rodícím se proteinem.

Genetický kód je napříč organizmy relativně konzervovaný, avšak existují rozdíly. Typická je např. změna kódu v translačním aparátu mitochondrií, kde kodon UGA, v jaderném genomu překládaný jako stop kodon, je v mitochondriích obratlovců překládán jako tryptofan. Jedná se o tzv. stop-to-sense záměnu, poprvé objevenou v lidských mitochondriích na konci 70. let 20. století [99]. Od té doby bylo shromážděno mnoho nestandardních přiřazení, nejen typu stop-to-sense, které jsou nejčastější, ale i sense-to-stop a sense-to-sense [100]. Variace genetického kódu jsou shromažďovány na [webu Národního centra pro biotechnologické informace v USA](#).

Ve spolupráci se skupinou M. Eliáše z Ostravské univerzity jsme se zaměřili na genetický kód nedávno objeveného jednobuněčného eukaryotického organismu ze skupiny bičíkovců. Jelikož se jedná o studii dosud nepublikovanou, nemohu sdílet úchvatné detaily ohledně biologického kontextu. Proto se zaměřím pouze na aspekt dekódování mRNA, který jsme studovali pomocí MD simulací.

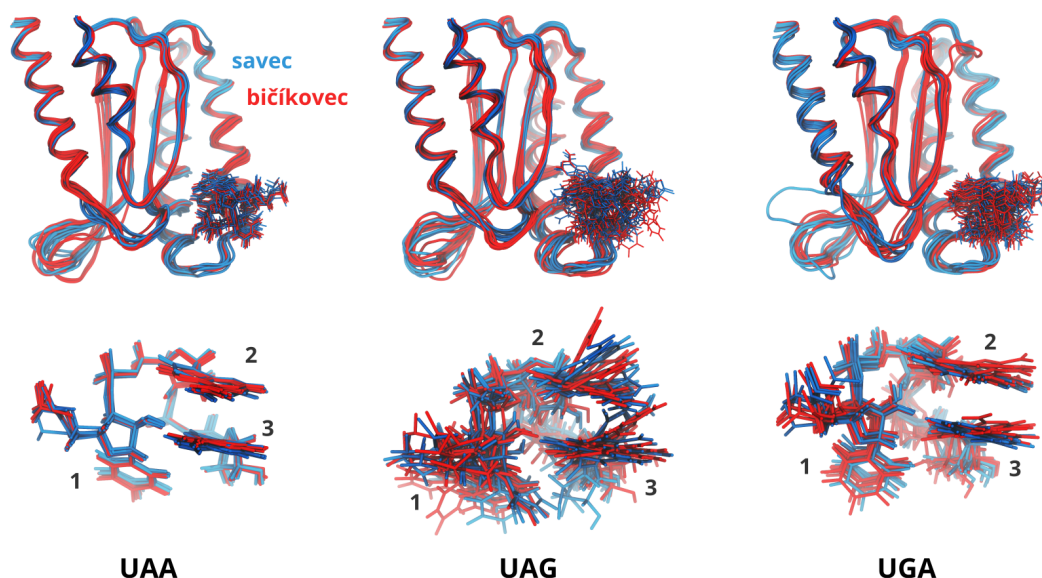
Bičíkovec pomocí kanonického stop kodonu UAG kóduje glutamin a zbylé dva stop kodony využívá standardně. Molekulární mechanismus této záměny není znám. Rozpoznávání kanonických stop kodonů v eukaryotických organizmech zajišťuje jediný RF označovaný eRF1 (na rozdíl od bakterií, kde jsou RF dva). Biochemické experimenty a cryo-EM struktury ukázaly, že za rozpoznání terminačních kodonů jsou na eRF1 zodpo-



Obrázek 3.7: A) N-terminální doména eRF1 se třemi sekvenčně konzervovanými motivy (TASNIKS šedě, GTS zeleně a YxCxxxF oranžově) zodpovědnými za správné rozpoznání kodonu (růžově, hnědě a modře). Sekvenční odchylky mezi eRF1 bičíkovce a savce jsou vyznačeny kuličkami. B) Řez eukaryotickým ribozomem s malou 40S podjednotkou tmavě a velkou 60S podjednotkou světle. Výběr atomů použitých v MD simulacích dekodovacího centra je označen modře. Připraveno z cryo-EM modelu komplexu savčího ribozomu s eRF1 (PDB 3JAG [102]). C) Průřez kulovým simulačním modelem. Červeně jsou označeny části podléhající silným pozičním omezením, růžově části se slabými pozičními omezeními. Ostatní části se pohybovaly bez pozičních omezení. Zeleně je eRF1, žlutě mRNA.

vědné tři konzervované motivy: T₅₃ASNIKS, G₂₆TF a Y₁₂₀xCxxxF [101, 102] (Obr. 3.7). U eRF1 bičíkovce jsme v těchto motivech identifikovali jedinou nekonzervativní mutaci oproti lidskému eRF1, a to v motivu TASNIKS, kde první serin je nahrazen valinem. V motivu YxCxxxF jsou tři kritická rezidua zachována, ale z ostatních reziduí vykazují tři konzervativní mutaci, konkrétně L121M, N124S a K125H. Mimo tyto tři motivy vykazují sekvence savčího eRF1 a eRF1 bičíkovce několik dalších odlišností. MD simulace jsme využili k otestování hypotézy, že eRF1 bičíkovce má díky mutacím ke kodonu UAG nižší afinitu než savčí eRF1.

Vytvořili jsme atomistické modely komplexu ribozomu a obou variant eRF1. Každý komplex obsahoval mRNA s jedním z kodonů UAA, UAG a UGA. Ke stavbě komplexu jsme využili cryo-EM model savčího ribozomu a eRF1 s rozlišením 3,5–3,8 Å (PDB 3JAG, 3JAH, 3JAI [102]). eRF1 bičíkovce jsme vytvořili bodovými mutacemi savčího eRF1 pomocí programu [Open Pymol](#) s implicitními kritérii pro orientaci postranních řetězců. Pro jednoduchost jsme nezaváděli mutace do rRNA, neboť nejbližší nekonzervovaný nukleotid se nacházel 1,2 nm od dekodovacího centra. Díky lokálnímu charakteru dekodovacího centra a studovaného procesu jsme pro snížení výpočetních nároků simulovali pouze kulový výřez eukaryotického ribozomu zahrnující rezidua velké i malé podjednotky do vzdálenosti 4,0 nm od atomu C5 z druhého nukleotidu kodonu (Obr. 3.7). Rezidua ve vzdálenosti větší než 2,8 nm od atomu C5 podléhala pozičním omezením se silovou konstantou 500 kJ mol⁻¹ nm⁻². Rezidua ve vzdálenosti 3,4 nm od atomu C5 podléhala



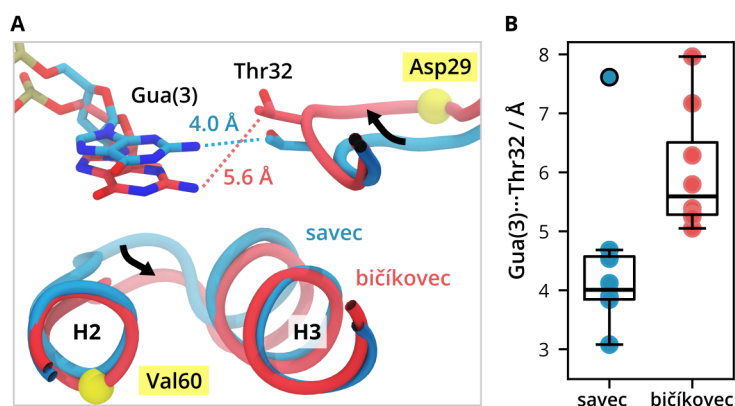
Obrázek 3.8: Konformační soubory eRF1 a kodonu. Každý model odpovídá struktuře průměrované přes posledních 150 ns každé trajektorie.

dodatečným pozičním omezením se silovou konstantou $9500 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. Včetně solventu a iontů obsahoval simulační systém asi 85000 atomů.

Použili jsme interakční potenciál z rodiny Amber, stejný, jako byl použit v ostatních simulacích ribozomu diskutovaných v této práci. Simulacemi jsme generovali NpT statistický soubor při teplotě 300 K a tlaku 1 bar s použitím v-rescale termostatu [103] a Parrinellova a Rahmannova barostatu [104]. Pro každý eRF1 (savčí a bičíkovce) jsme provedli 8 nezávislých simulací lišících se v počátečních rychlostech atomů. Pro analýzy jsme využili posledních 150 ns z celkem 300 ns dlouhých trajektorií. Průměrováním kartézských souřadnic jsme pro každou trajektorii získali strukturní model dekódovacího centra ribozomu.

Inspekce strukturních modelů z jednotlivých trajektorií ukázala, že variabilita dekódovacího centra se pro jednotlivé kodony liší a vzrůstá v řadě UAA, UGA, UAG (Obr. 3.8). Kodon UAA je strukturně nejrigidnější, ale konformační soubory eRF1 savce a bičíkovce se neliší. Kodon UAG vykazuje nejvyšší variabilitu v obou variantách eRF1.

Simulace UAG kodonu odhalily rozdíl, který podporuje hypotézu snížené afinity eRF1 bičíkovce oproti UAG kodonu oproti savčímu. Na základě cryo-EM struktury Brown *et al.* uvedli [102], že pro správné rozpoznání třetí pozice kodonu (savčího) je stěžejní vodíková vazba mezi Thr32 eRF1 a aminoskupinou guaninu v UAG. MD simulace však naznačují, že tato vodíková vazba je v dekódovacím centru s eRF1 bičíkovce narušena (Obr. 3.9). Medián vzdálenosti Thr32–Gua(3) v savčím dekódovacím centru, jak je naznačeno na Obr. 3.9A, je $4,0 \text{ Å}$, zatímco u bičíkovce je to $5,6 \text{ Å}$. Toto oslabení interakce je spojeno s konformační změnou smyčky poblíž motivu $G_{26}TS$. Z hlediska těchto změn



Obrázek 3.9: Interakce eRF1 s kodonem. A) Detail interakce Thr32 s třetím nukleotidem kodonu UAG pro eRF1 savec (modře) a bičkovce (červeně). Rozdíl v konformacích smyček obsahujících GTS a TASNIS motivy pro dva eRF1 je naznačen šipkami. Dvě nekonzervovaná rezidua jsou vyznačena žlutě. Helixy H2 a H3 jsou označeny podle jejich pořadí v N-terminální doméně eRF1. Jedná se o modely získané průměrováním přes posledních 150 ns všech 8 nezávislých trajektorií. B) Vzdálenost mezi Thr32 a Gua(3) vypočtená z 8 nezávislých trajektorií. P-hodnota pro nulovou hypotézu, že mediány jsou identické, je 0,00060 s vyloučením jednoho odlehlého výsledku pro savčí komplex.

může být významná i mutace rezidua 29 poblíž motivu, kterým je u savců nebo bičkovců se standardním genetickým kódem nenabitý Asn nebo Ser. U studovaného bičkovce se stop-to-sense změnou genetického kódu je to Asp, jehož postranní řetězec nese záporný náboj, což dle našich simulací vede k posunu smyčky asi o 2 Å.

Bodová mutace serinu na valin v TASNIS motivu je z pohledu guaninu v UAG lokalizovaná na zadní straně helixu 2 (Obr. 3.9A). Přestože přímo neovlivňuje interakci s uracilem v pozici 1 na kodonu, může mít souvislost s posunem celého motivu o 1,5 Å v dekódovacím centru bičkovce oproti savčímu, jak naznačují simulace. Z toho usuzujeme, že tyto dvě konformační změny oslabují interakci eRF1 s UAG kodonem v případě bičkovce.

Přesnější výpočty jsou prakticky proveditelné. Např. Lind *et al.* studovali rozpoznání terminačních kodonů v mitochondriích [105]. Kromě simulační strategie použité námi (porovnávání nezávislých trajektorií) vypočítali i relativní Gibbsovy energie pro různé komplexy mitochondriálních RF a kodonů. Vzhledem k rozsahu naší studie vybraného bičkovce a jejímu biologickému zaměření jsme se spokojili s kvalitativními závěry podloženými strukturními úvahami.

3.3 Allosterický přenos signálu ribozomem

Funkcí ribozomu není pouze katalyzovat vznik peptidových vazeb. Ribozom je významným činitelem v koordinaci procesů přidružených k translaci. Např. v bakteriích je trans-

lace na ribozomech úzce spjata s transkripcí DNA do mRNA. Proto je ribozom časově i prostorově synchronizován s RNA polymerázou. Takový biomolekulový komplex, který zpracovává DNA a produkuje protein, je označován jako expresom. Dle aktuálních cryo-EM modelů je v expresomu několik ribozomálních proteinů malé ribozomální podjednotky významně zapojeno do interakce s polymerázou [106, 107].

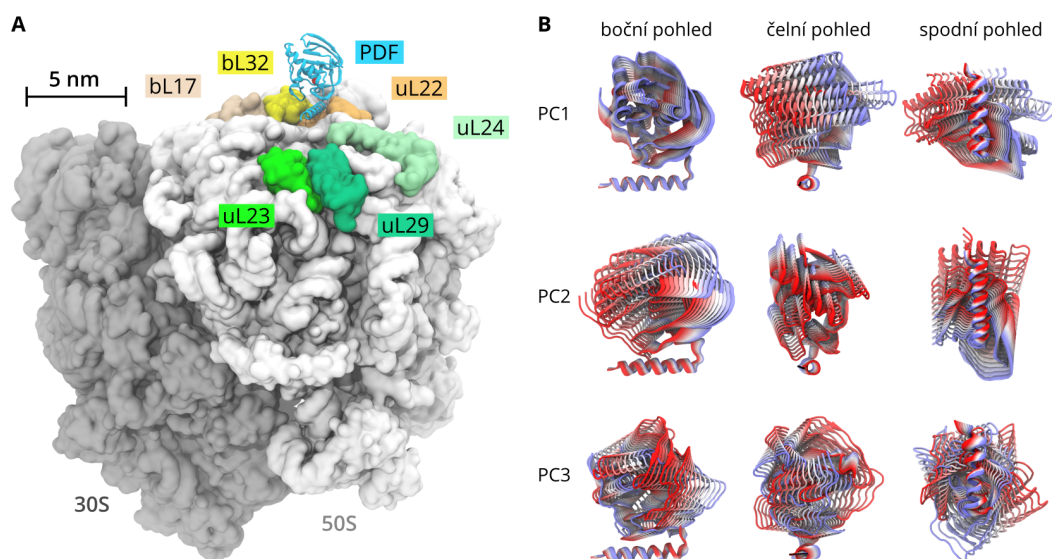
Ribozom také slouží jako logický uzel pro kotranslační a posttranslační modifikace rodících se proteinů [5]. Váže se na něj řada translačních faktorů, které ovlivňují směřování zrozeného proteinů. Patří sem např. signál rozpoznávající částice (SRP z angl. *signal recognition particle*), díky níž jsou membránové proteiny naváděny během translace přímo do membrány. Vzhledem k velikosti ribozomu přesahující 20 nm vyvstává otázka, jak jsou tyto interakce koordinovány.

Na povrch velké ribozomální podjednotky se také váže tzv. peptidová deformyláza (PDF), enzym, který v bakteriích a některých organelách eukaryot odstraňuje z N-konce rodícího se proteinu signální formylovou skupinu. Bingel-Erlenmeyer *et al.* [108] pomocí biochemických experimentů a cryo-EM ukázali, že PDF třídy I se na ribozom váže pomocí C-terminálního α -helixu PDF, který interaguje s r-proteinem uL22. Tento protein má, podobně jako i jiné r-proteiny, specifickou konstrukci. Globulární doména uL22 na povrchu ribozomu je propojena s prodlouženou smyčkou, která sahá do nitra velké podjednotky. Špička prodloužení tvoří část ribozomálního tunelu, a je tak v kontaktu s každým rodícím se proteinem, který tunelem projde. A projdou jím všechny!

Položili jsme si otázku, zda-li je informace o vazbě PDF přenášena do nitra velké ribozomální podjednotky. Nabízela se hypotéza, že by takový přenos mohl allostericky zprostředkovávat r-protein uL22. Pomocí MD simulací kompletního ribozomu jsme se pokusili hypotézu ověřit. Studie byla součástí bakalářské práce H. McGratha [109] a PhD tréninku M. Černekové. Výsledky a závěry studie jsou ve formě preprintu dostupné [online](#) [82] a práce je v recenzním řízení.

Studovaný systém je na Obr. 3.10A. Atomárně rozlišená struktura komplexu ribozom–PDF potřebná pro MD simulace nebyla dostupná. Proto jsme ji sestavili z atomistických modelů jednotlivých součástí komplexu, a to *E. coli* ribozomu (PDB 5AFI [7]), komplexu ribozomu s C-terminálním helixem PDF (PDB 4V5B [108]) a kompletní izolované PDF (PDB 2AI8 [110]). Provedli jsme 4 nezávislé atomistické MD simulace označené PDF+, každou o délce 1040 ns. Abychom určili vliv PDF na konformační chování ribozomu, provedli jsme pro srovnání také 4 nezávislé simulace ribozomu bez PDF (označené PDF–). MD simulace využívaly pro ribozom běžný aparát algoritmů uvedených v sekci 2 Metody včetně silového pole z rodiny Amber.

Cryo-EM struktura reportující o interakci PDF a ribozomu byla založena na zkráceném modelu PDF, který obsahoval pouze C-terminální α -helix. Naše simulace jsou tak prvním zdrojem informací o interakci kompletní PDF s ribozomem na atomární úrovni.



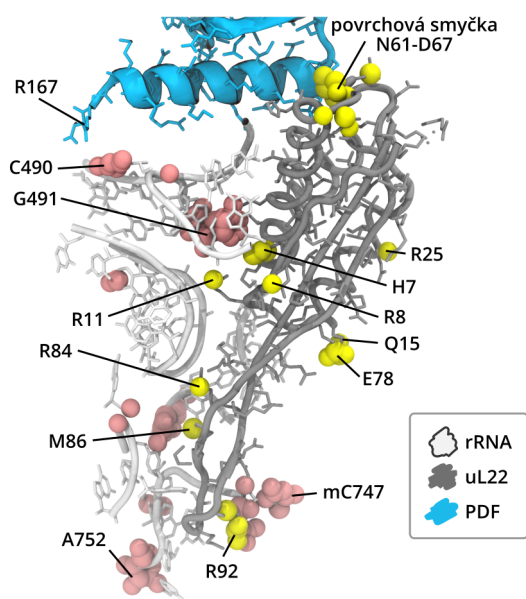
Obrázek 3.10: A) Komplex ribozomu s peptidovou deformylázou (PDF) použitý pro MD simulace. Šedě jsou označeny velká 50S podjednotka, malá 30S podjednotka, modře PDF a ostatními barvami r-proteiny v blízkosti interakčního rozhraní. B) Vizualizace prvních tří hlavních komponent (PC) ze tří ortogonálních pohledů.

Simulace ukázaly, že kontakt mezi PDF a ribozomem není pouze skrze C-terminální α -helix, nýbrž i částí katalytické domény. Analýzou souřadnic PDF jsme prokázali, že PDF vykonává na povrchu ribozomu rozsáhlý pohyb na časových škálách pod $1 \mu\text{s}$ (Obr. 3.10B). Analýza hlavních komponent odhalila, že vzájemný pohyb PDF a ribozomu lze popsat několika málo kolektivními souřadnicemi. První 3 hlavní komponenty popisují 92 % celkových fluktuací kartézských souřadnic PDF.

Analýzy experimentálních modelů uL22 a modelů získaných z MD simulací shodně naznačily, že PDF ovlivňuje strukturu dvou částí uL22: povrchové smyčky, která je součástí rozhraní ribozom–PDF, a špičky prodloužení uL22, která sahá do tunelu a od PDF je vzdálena přibližně 5 nm. Tento nelokální strukturní efekt jsme dále studovali metodami pokročilé statistické analýzy.

Pomocí regrese hlavních komponent (PCR, z angl. *principle component regression*), která je jakýmsi rozšířením PCA a v kontextu biomolekul byla dříve označena jako Functional Mode Analysis [111], jsme vytvořili klasifikátor přítomnosti PDF na povrchu ribozomu. Ke klasifikaci jsme použili různé části ribozomu s předpokladem, že části poskytující spolehlivou klasifikaci jsou peptidovou deformylázou přímo ovlivňovány.

Kromě triviálního výsledku, že přítomnost PDF dobře koreluje se strukturou povrchové smyčky ve vazebném místě PDF, umožnila PCR odhalit několik podstatných faktů. Prokázali jsme, že pro spolehlivou predikci přítomnosti PDF je nutné zahrnout postranní řetězce uL22; páteř uL22 poskytovala slabou predikci. Zároveň jsme ukázali, že dobrou predikci poskytuje i špička uL22 v tunelu a rRNA v jejím nejbližším okolí.



Obrázek 3.11: Přenos allosterického signálu mezi vazebným místem peptidové deformylázy (PDF) na povrchu ribozomu a nitrem ribozomu. Atomy, které jsou z hlediska PCR klasifikátoru nejvíce citlivé na přítomnost/absenci PDF na povrchu ribozomu, jsou vyznačeny žlutě (na uL22) nebo růžově (na rRNA).

Nejlépe přítomnost PDF predikoval rozsáhlý výběr reziduí zahrnující uL22 včetně postranních řetězců a rRNA tvořící povrch ribozomálního tunelu (Obr. 3.11). Na základě analýzy výstupů PCR jsme navrhli, že přenos allosterického signálu mezi vazebným místem PDF na povrchu a nitrem ribozomu probíhá skrze rRNA nukleotidy poblíž C-konce PDF.

MD simulace a strukturní analýzy experimentálních dat tedy naznačily, že vazba PDF ovlivňuje strukturu a dynamiku vnitřních částí ribozomu. Fyziologický význam však není zřejmý. Především biochemické experimenty ukázaly, že asociační konstanta PDF není závislá na translačním statusu ribozomu; PDF se váže stejně silně nezávisle na tom, jestli je v tunelu přítomný rodící se protein nebo ne [112]. Lze spekulovat, že nelokální změny v ribozomálním tunelu mohou ovlivňovat průchod rodícího se proteinu z katalytického místa ven z ribozomu. Jelikož r-protein uL22 tvoří nejužší místo tunelu, mohou mít i malé změny v jeho dynamice význam pro kinetiku transportu rodícího se peptidu, a to zejména v nejranější fázi elongace (Obr. 3.1B), kdy se N-konec rodícího se peptidu stále nachází uvnitř tunelu. V pozdějších fázích je totiž translokace proteinu tunelem řízena silou, kterou generuje kotranslační balení vně ribozomu [113].

Projekt rozvíjíme dvěma směry. Přišlo nám zajímavé, že PDF třídy I se váže na povrch ribozomu pomocí C-terminálního α -helixu. Jak se vážou jiné třídy PDF zůstává neobjasněno. Během Středoškolské odborné činnosti se J. Kubíček věnoval zkoumání variability peptidových deformyláz [114]. Pomocí PCA proteinové databáze definoval 5 klastrů a v

každém klastru našel vhodného reprezentanta pro další výzkum. Společně se studentkou bakalářského programu A. Hrádkovou se nyní věnují popisu konformačního chování C-terminálních částí PDF ve vodném prostředí (mimo ribozom), a to jak v přítomnosti katalytické domény PDF, tak bez ní. Od MD simulací si slibujeme lepší pochopení vazebných možností ostatních tříd PDF. Druhým směrem je rozvoj algoritmů pro popis dalokodosahové allostérie, kterému se v rámci diplomové práce věnuje H. McGrath. Chtěli bychom využít algoritmy strojového učení pro vědecké problémy podobné vazbě ribozom–PDF.

3.4 Konformační preference fragmentů ribozomálních proteinů

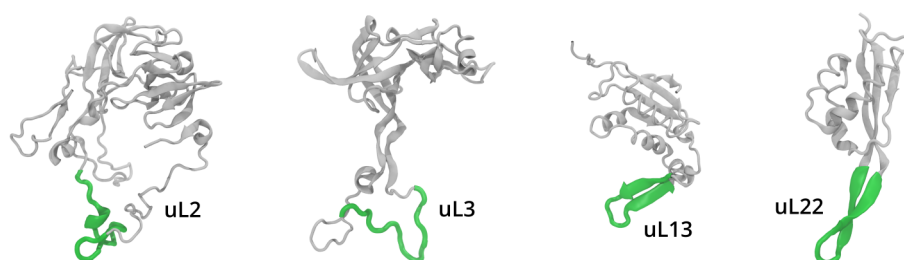
Ribozom, jako složitý nukleoproteinový komplex, formovala evoluce na základě plejády okolností. Předpokládá se, že evolučně starší ribozomy byly ještě jednodušší než současné bakteriální. Odpovídá to i trendu ve složitosti moderních ribozomů, kdy eukaryotické ribozomy jsou složitější než prokaryotické.

Bokov *et al.* formulovali model evoluce ribozomu, tzv. *Accretion model*. Skládá se ze 6 časových fází pro bakteriální ribozom a dvě další přidává pro eukaryotický ribozom [115]. Absolutní stáří fází není známo až na fázi VI, o které se předpokládá, že nastala před asi 3,9 miliardami let [116]. V prvních dvou fázích byl předchůdce moderního ribozomu proteinů prostý a první proteinové části se zabudovaly až ve fázi III [117]. I katalytické místo velké podjednotky moderního ribozomu je napříč taxonomickými jednotkami silně konzervované a skládá se pouze z rRNA [11].

Na základě euklidovské vzdálenosti fragmentů r-proteinů od nitra ribozomu lze určit jejich stáří. Kovacs *et al.* využili *Accretion model* a obohatili jej o r-proteiny. Předpokládali, že moment začlenění fragmentu r-proteinu do prapředka ribozomu se shoduje se stářím rRNA, se kterou fragment interaguje. Rozdělili fragmenty r-proteinů podle jejich stáří do fází III–VI [117].

Důsledkem toho je, že moderní r-proteiny jsou složené z několika evolučně různě starých domén. Evolučně mladá globulární doména je exponovaná na povrchu ribozomu a do nitra zasahují evolučně starší smyčky nebo výběžky. Některé výběžky mají dobře definovanou sekundární strukturu, např. výběžek uL22 tvoří β -list, uL2 má krátkou α -helikální část, jiné jsou neuspořádané (např. uL3). Principy, podle kterých probíhala počáteční selekce fragmentů r-proteinů, však nejsou známy.

Se skupinou K. Hlouchové z výzkumného ústavu BIOCEV jsme se zaměřili na otázku, jakým způsobem probíhal prvotní kontakt mezi proteinovými fragmenty s prapředkem ribozomu složeným pouze z RNA. Vybrali jsme několik r-proteinů s charakteristickými strukturními motivy a studovali jejich konformační chování. Jejich struktury z moder-



Obrázek 3.12: Vybrané r-proteiny a jejich fragmenty studované pomocí atomistických MD simulací zvýrazněné zeleně. Struktury odpovídají ribozomu bakterie *T. thermophilus*, PDB 7MD7 [118]

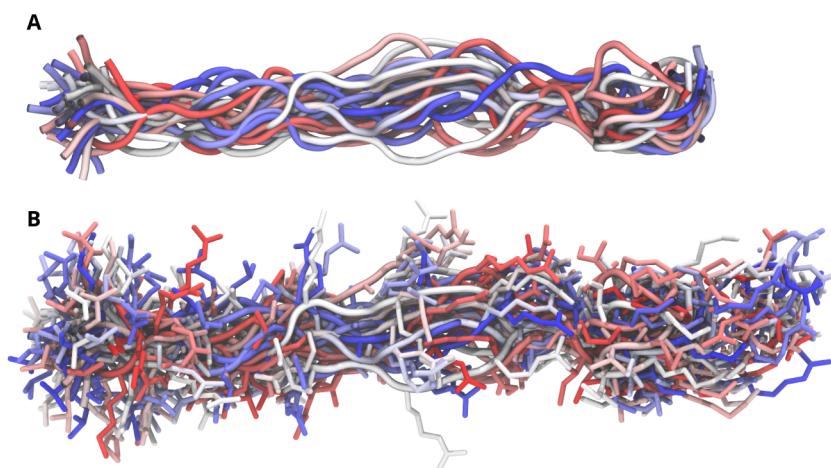
fragment	rezidua	sekvence	počet +/– ^a
uL2	219–241	PHVRG AAMNP VDHPH GGEG RAP	2/2
uL3	136–153	RHPGS IGNRK TPGRV YKG	5/0
uL13	71–86	IYTRY SGYPG GLKKI P	3/0
uL22	82–99	LKRVL PRARG RADII KKR	7/1

Tabulka 3.1: Primární struktury fragmentů vybraných r-proteinů. Číslování odpovídá bakterii *T. thermophilus*. ^a značí počet kladně a záporně nabitých postranních řetězců.

ního bakteriálního ribozomu (*T. thermophilus*, PDB 7MD7 [118]) jsou uvedené na Obr. 3.12 a sekvence v Tab. 3.1. Pro některé r-proteiny existují i experimentálně určené struktury mimo ribozom (např. pro uL22, PDB 1BXE [119]). Pomocí atomistických MD simulací jsme se pokusili objasnit, jestli evolučně staré fragmenty mají vnitřní preferenci, tj. na úrovni sekvence reziduí, vyskytovat se v konformacích, jaké zaujímají v moderním ribozomu. Nebo bychom mohli naopak tvrdit, že jejich struktura v ribozomu je plně indukovaná prostředím a interakcemi s rRNA.

Na projektu jsme pracovali se studentem P. Linhartem, který se v rámci své bakalářské práce zaměřil na fragment r-proteinu uL22 [120]. Při přípravě na diplomovou práci následně rozšířil sadu proteinů a na základě zkušenosti s uL22 aktualizoval simulační protokol. K některým fragmentům jsme měli k dispozici data z nukleární magnetické rezonance (NMR), a to chemické posuny vybraných atomů a výstupy programu TALOS-N [121, 122, 122]. Ten poskytuje odhady dihedrálních úhlů ϕ a ψ a na jejich základě klasifikuje sekundární strukturu reziduí do tří kategorií: helix, extended a coil.

Simulace jsme provedli pomocí dvou silových polí optimalizovaných pro popis rovnováhy strukturovaný–nestrukturovaný protein, konkrétně CHARMM36m ze skupiny A. D. MacKerella [2] a A99SB-disp ze skupiny D. E. Shawa [1]. Očekávali jsme, že se fragmenty budou chovat jako IDPs. Nasvědčuje tomu např. vysoké zastoupení reziduí s bazickými a kyselými postranními řetězci typické pro IDPs. Terminální rezidua byla modelována jako nabitá, tedy se skupinami $-\text{NH}_3^+$ a $-\text{COO}^-$. Modely vody byly použity v souladu s parametrizací jednotlivých silových polí. CHARMM36m jsme kombinovali



Obrázek 3.13: 32 počátečních konformací fragmentu uL22 vygenerovaných MD simulací v plynné fázi při teplotě 500 K. A) Proteinová páteř. B) Proteinová páteř a nevodíkové atomy.

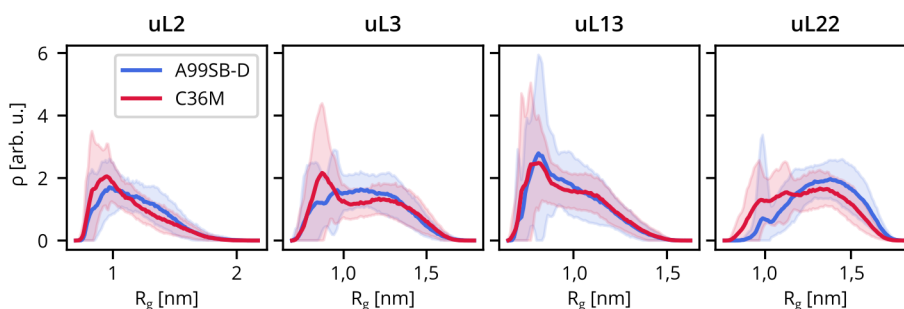
s třístředovým modelem vody TIP3P [48], zatímco A99SB-disp se čtyřstředovým modelem odvozeným od TIP4P-D [123].

Pro zlepšení vzorkování konformačního prostoru a pro dekorelaci počátečních podmínek jsme nejprve pro každý fragment vygenerovali 32 počátečních konformací (pro ilustraci uL22 na Obr. 3.13). Konformace pocházely z 65 ns dlouhé MD simulace v plynné fázi provedené při 500 K z plně natažené konformace peptidu. První nanosekunda byla využita k ekvilibraci. Následně byly konformace ukládány každé 2 ns. Každou konformaci, takto vygenerovanou v plynné fázi, jsme vložili do periodického boxu tvaru rhombického dvanáctistěnu a solvatovali explicitními molekulami vody. Iontové podmínky jsme nastavili tak, aby iontová síla roztoku NaCl odpovídala NMR měřením, což vedlo ke koncentraci NaCl 38 mmol dm^{-3} .

Produkční simulace proběhly při konstantní teplotě 300 K a tlaku 1 bar s použitím Busiho termostatu [103] a Parrinellova-Rahmanova barostatu [104]. Délka každé trajektorie byla 330 ns. Pro každý fragment jsme tedy získali konformační soubor ekvivalentní $10,56 \mu\text{s}$, který jsme analyzovali. Simulace běžely na superpočítači Karolina ve výpočetním centru IT4Innovations v Ostravě.

Sesbírané výsledky sloužily ke třem hlavním účelům. Zaprvé jsme chtěli porovnat chování fragmentů ve vodném prostředí a v ribozomu a hledat případné evoluční implikace. Zadruhé, NMR měření poskytla referenční data pro MD simulace, což umožnilo zhodnotit simulační protokol a porovnat kvalitu silových polí. Zatřetí jsme chtěli porovnat silová pole i pro veličiny, pro které nemáme referenční data. Řada studií totiž ukázala, že popis nestrukturovaných proteinů může být na silovém poli silně závislý [37, 124, 125].

Sbalování fragmentů jsme sledovali pomocí gyračního poloměru R_g . Jeho hustoty pravděpodobnosti získané z úseků trajektorií mezi 130 a 330 ns jsou na Obr. 3.14. Z hlediska



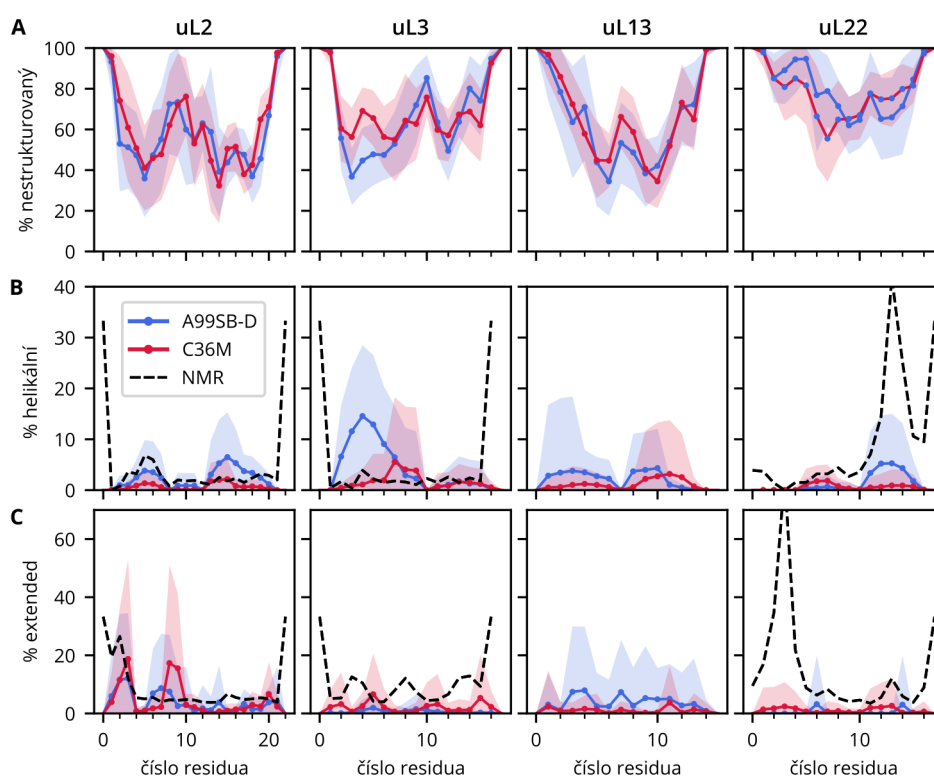
Obrázek 3.14: Hustota pravděpodobnosti ρ pro gyrační poloměr R_g počítaný pomocí silového pole A99SB-D [1] (modře) a C36M [2] (červeně). Linky reprezentují střední hodnotu přes 32 nezávislých simulací, barevná plocha reprezentuje směrodatnou odchylku. Pro konstrukci ρ byly použity úseky trajektorií v rozmezí 130–330 ns.

R_g byl nejméně sbalený uL22, což lze vysvětlit nejvyšším počtem nabitých postranních řetězců mezi sadou studovaných fragmentů. Vícestavové chování lze vypožorovat u uL3 a uL13. Pro R_g jsme pozorovali velmi dobrou shodu obou silových polí, a to zejména u fragmentů uL2 a uL13.

Sekundární strukturu jsme v MD simulacích kvantifikovali algoritmem DSSP [69]. Hlavním závěrem je, že všechny studované fragmenty r-proteinů se chovají jako nestrukturované, přičemž nejméně strukturovaný je uL22. Terminální části fragmentů vykazují méně sekundární struktury než centrální části. Je to zřejmé z Obr 3.15A, který ukazuje procento simulačního času, ve kterém byla jednotlivá rezidua klasifikována jako nestrukturovaná (DSSP třída coil). Pro konstrukce profilů byly použity úseky trajektorií mezi 130 a 330 ns. Vzhledem k silovým polím jsou tyto závěry robustní, neboť profily se pro obě silová pole shodují kvantitativně pro většinu reziduí.

Dále jsme kvantifikovali procento simulačního času v helikální a extended konformaci dle TALOS-N klasifikace (Obr. 3.15BC). Pro potřeby srovnání osmi tříd DSSP s algoritmem TALOS-N, který definuje třídy pouze tři, jsme za helikální považovali DSSP třídy α -helix a 3/10-helix, a za extended β -strand a β -bridge. Všechny fragmenty vykazovaly nízké zastoupení jak helikální, tak extended konformace, a to v NMR měřeních i MD simulacích. Výjimku tvoří uL22, u něhož bylo okolí rezidua 14 na základě NMR dat klasifikováno jako dočasně (40 %) helikální a okolí rezidua 5 dočasně (70 %) v extended konformaci. Výsledky MD simulací však pro uL22 ukazují nižší podíl sekundární struktury, obvykle méně než 20 %.

Chtěli jsme lépe porozumět konformačnímu chování fragmentů a zdánlivému nesouladu klasifikace sekundárních struktur některých reziduí mezi NMR a MD simulacemi. Proto jsme se zaměřili na dihedrální úhly ϕ a ψ proteinové páteře. V MD simulacích je lze spočítat přímo z kartézských souřadnic. TALOS-N úhly počítá z chemických posunů a následně z úhlů vyvozuje sekundární strukturu.



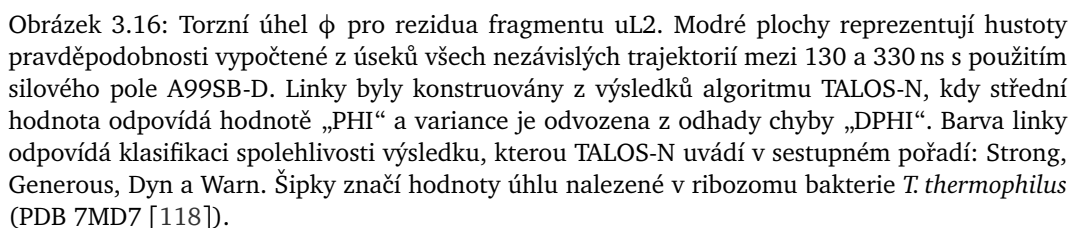
Obrázek 3.15: Profily sekundární struktury dle algoritmu DSSP, založeném na simulačních datech v silovém poli A99SB-D [1] (modře) nebo C36M [2] (červeně), a algoritmu TALOS-N z NMR experimentů (černě). Body reprezentují střední hodnotu přes 32 nezávislých simulací, barevná plocha reprezentuje směrodatnou odchylku, linky byly přidány pro přehlednost. Pro konstrukci profilů byly použity úseky trajektorií v rozmezí 130–330 ns.

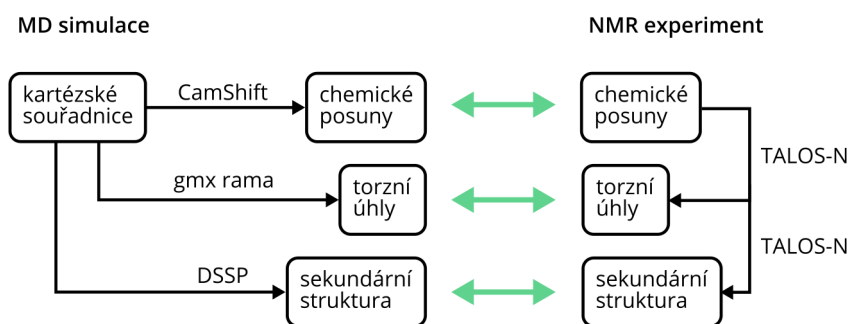
Porovnali jsme hodnoty úhlů v MD simulacích s hodnotami vypočtenými z NMR chemických posunů. Pro ilustraci, Obr. 3.16 ukazuje hodnoty dihedrálních úhlů ϕ pro jednotlivá rezidua fragmentu uL2 (mimo terminálních). Z hlediska MD simulací se jedná o hustoty pravděpodobnosti počítané z úseků 32 nezávislých trajektorií mezi 130 a 330 ns. Z hlediska NMR jde o hustoty pravděpodobnosti aproximované Gaussovou funkcí, kde maximum je v hodnotě “PHI” spočtené algoritmem TALOS-N a pološířka odpovídá odhadu chyby “DPHI,” kterou TALOS-N pro každé reziduum také uvádí. Kvalita NMR výstupů není stejná: TALOS-N rozlišuje 4 třídy spolehlivosti, sestupně Strong, Generous, Dyn a Warn. Ty jsou v Obr. 3.16 rozlišeny barevně. Obrázek také obsahuje informace o konformaci peptidu nalezené v ribozomu (PDB 7MD7 [118]).

Je zřejmé, že v uL2 jsou pro většinu reziduí MD simulace v souladu s NMR měřeními. Hodnoty ϕ pro rezidua Asn-9, Pro-10, Asp-12, His-13 a Pro-14, jejichž odhady TALOS-N klasifikoval jako spolehlivé (Strong nebo Generous), kvantitativně odpovídají hodnotám z MD simulací. Zároveň data naznačují, že hustoty pravděpodobnosti úhlů ve vodném prostředí odpovídají hodnotám nalezeným v ribozomu. Poznatek je platný zejména pro „vnitřní“ rezidua, tj. ta, která jsou daleko od konců peptidu. V tomto ohledu tedy sekvence střední část strukturně „pasuje“ do prostředí ribozomu a zdá se, že ribozom populace úhlu ϕ výrazně nemění.

Analýza jednotlivých reziduí demonstrována na fragmentu uL2 také potvrzuje nestrukturovaný charakter peptidu. Mnoho reziduí má hustoty pravděpodobnosti úhlu ϕ více-modální (např. Gly-5, Met-8, nebo Arg-21). Poukazuje to zároveň na fakt, že průměrování těchto úhlů přes simulační trajektorie je nevhodné a mohlo by vést k chybným závěrům. Např. Gly-20 vykazuje dvě maxima okolo 80° a -80° se střední hodnotou v 0° , což by naznačovalo neshodu s úhlem 70° nalezeným v prostředí ribozomu. Ve skutečnosti se úhel ϕ pohybuje v okolí této hodnoty asi 50 % simulačního času.

Je zřejmé, že porovnání provedených NMR experimentů a MD simulací není přímočaré a nesoulad klasifikace sekundárních struktur je jen zdánlivý. Techniky negenerují veličiny, které by byly přímo porovnatelné, jak je naznačeno na Obr. 3.17, proto je buď ze strany NMR, nebo MD simulací potřeba kompromisu. NMR chemické posuny jsou pomocí TALOS-N přepočteny na úhly ϕ a ψ a klasifikovány do tří motivů sekundárních struktur. MD simulace generují kartézské souřadnice, z nichž lze úhly vypočítat přímo. Ze simulací lze klasifikovat sekundární struktury analýzou vodíkových vazeb v DSSP. Bylo by v principu možné použít totožnou klasifikaci jako TALOS-N a klasifikovat dle ϕ a ψ . Z MD simulací lze také vypočítat chemické posuny některou z empirických metod. Např. skupina M. Vendruscola vytvořila algoritmus CamShift [126], který je na základě empiricky definovaných souvislostí mezi chemickými posuny a meziatomovými vzdálenostmi schopen počítat chemické posuny ze známe konformace proteinu. Jelikož je definována i derivace (zobecněná síla) vztahu chemický posun–vzdálenost, je možné





Obrázek 3.17: Srovnání veličin měřených nukleární magnetickou rezonancí (NMR) a výstupů MD simulací.

CamShift použit v průběhu MD simulace ke kalibraci silového pole [127].

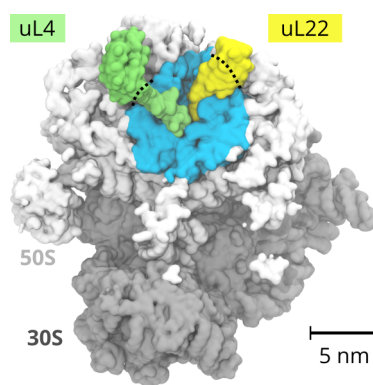
Ve studiu fragmentů r-proteinů dále pokračujeme. Prvotní porovnání chemických posunů vypočítaných algoritmem CamShift naznačuje, že budeme schopni kalibrovat konformační statistické soubory, aby lépe reprezentovaly NMR data. Jelikož je prostředí ribozomu kvůli rRNA páteři silně elektricky nabitě, nabízí se také možnost zaměřit se na vliv iontového prostředí peptidů na jejich konformační preference.

3.5 Dynamika nejužšího místa v tunelu bakteriálního ribozomu

Každý protein v živém organismu je syntetizován na ribozomu a opustí jej tunelem ve velké ribozomální podjednotce. Fyzikálně chemické a geometrické parametry tunelu jsou proměnlivé. Mění se od katalytického místa směrem k ústí, ale také v důsledku termálních fluktuací a konformačních pohybů. Např. Petrone *et al.* na základě MD simulací ukázali [128], že krátké postranní řetězce aminokyselin (jako methan pro Ala, acetat pro Asp), vykazují rozdílné Gibbsovy energie na různých místech tunelu. Predikovali bariéru Gibbsovy energie v koncové části tunelu blízko ústí pro transport aminokyselinových zbytků. Dao Duc *et al.* analyzovali struktury ribozomálního tunelu v několika organismech [129]. Ukázali, že tunely vyšších organismů jsou v průměru užší než tunely bakterií.

Nejužším místem bakteriálního ribozomu je zúžení tvořené smyčkami r-proteinů uL4 a uL22. Poloměr zúžení je asi 0,4 nm [129]. V eukaryotických ribosomech existuje druhé zúžení, asi o 2 nm vzdálenější od katalytického místa než první zúžení, které může být ještě užší než první zúžení.

Zúžení tvořené špičkami proteinů uL4 a uL22 je sekvenčně i strukturně konzervované. Přitom jeho funkce, není známa. Obsahuje mnoho bazických postranních řetězců, což by mohlo naznačovat, že bude citlivé na sekvenci rodícího se peptidu. Na základě bioche-



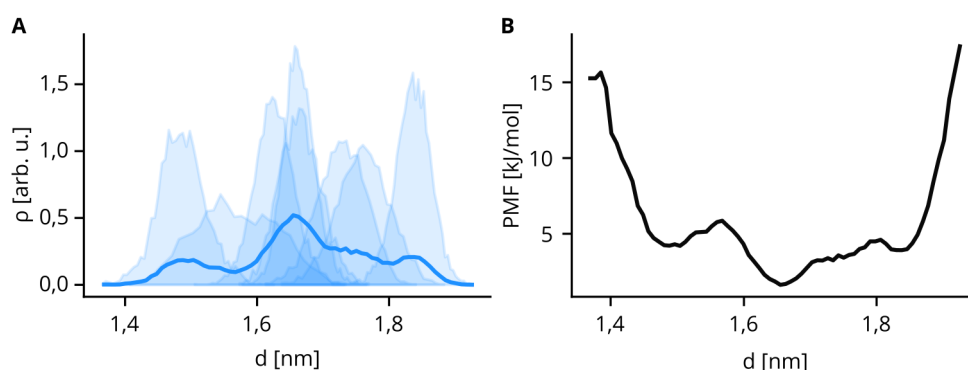
Obrázek 3.18: Schéma simulovaného systému v kontextu bakteriálního ribozomu. Malá ribozomální podjednotka 30S je tmavě, velká ribozomální podjednotka je světle, simulovaný výřez modře. R-proteiny uL4 (zeleně) a uL22 (žlutě) tvořící zúžení jsou zobrazeny pro přehlednost celé, ale v simulaci byly ořezány, jak je naznačeno tečkovaně.

mických analýz regulačního peptidu SecM, který způsobuje translační uvěznění [80], došli Nakatogawa a Ito k závěru, že zúžení slouží jako rozlišovací brána (v orig. *discriminating gate*) [130]. Dalším zkoumáním ale bylo prokázáno, i na základě MD simulací, že mechanismus translačního uvěznění SecM spočívá v sekvenčně závislých konformačních změnách katalytického centra [131, 132]. V blízkosti zúžení se také váže významná třída tzv. makrolidových antibiotik, které allostericky inhibují katalytické místo ribozomu [133, 134]. Zúžení tunelu může mít také vliv na kotranslační balení a kolaps některých domén rodičního se proteinu [135].

Se studentkou T. Svatoňovou jsme se v rámci její diplomové práce zaměřili na studium zúžení tunelu v ribozomu bakterie *E. coli* [136]. Provedli jsme MD simulace kulového výřezu okolí zúžení a sestavili Markovovský stavový model dynamiky zúžení.

K simulacím jsme vybrali rezidua ve vzdálenosti menší než 3,5 nm od uL4:Gly64 a uL22:Arg92 (Obr. 3.18). Do vzdálenosti 2,5 nm byla rezidua volně pohyblivá, za touto vzdáleností byl systém fixován pozičními omezeními se silovými konstantami $2000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. Simulační protokol byl stejný jako v případě dekódovacího centra nebo VemP (sekce 3.1 a 3.2). Stručně, použili jsme silové pole z rodiny Amber, generovali NpT statistický soubor při teplotě 300 K a tlaku 1 bar a provedli jsme 8 nezávislých simulací, každou o délce 1,5 μs . Startovní konformace systému jsme vygenerovali krátkou NVT simulací při 400 K a s rozdílnými počátečními rychlostmi. Pro účely analýz byly špičky r-proteinů definovány jako rezidua 60–69 v uL4 a 86–94 v uL22.

Nejprve jsme analyzovali šířku zúžení jako funkci simulačního času. Šířka byla definována jako vzdálenost d těžišť špiček proteinů uL4 a uL22. K analýzám jsme použili úseky trajektorií mezi 500 a 1500 ns. PMF reprezentuje řez plochou Gibbsovy energie ve vybrané souřadnici, zde vzhledem ke vzdálenosti těžišť špiček r-proteinů. Hustoty pravděpodobnosti ρ pro vzdálenosti těžišť d jsou uvedené na Obr. 3.19A. Data z jednotlivých



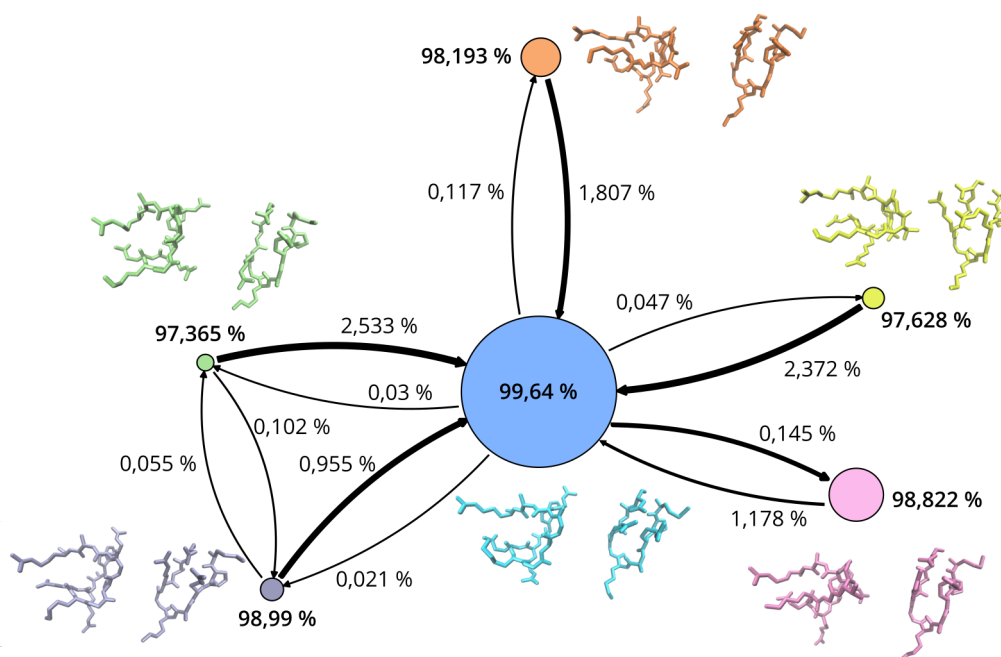
Obrázek 3.19: Šířka zúžení ribozomálního tunelu reprezentovaná vzdáleností těžišť špiček r-proteinů uL4 a uL22. A) Hustoty pravděpodobnosti ρ vzdálenosti d vypočtené z 8 nezávislých MD simulací (modré plochy) a z agregovaných dat. B) Potenciál střední síly (PMF) jako funkce vzdálenosti těžišť d .

trajektorií se sice překrývají, ale nejpravděpodobnější hodnoty jsou v širokém rozsahu od asi 1,5 nm do 1,8 nm.

PMF zkonstruovaný z histogramu dat agregovaných ze všech trajektorií tak vykazuje široké a mělké minimum s nejnižší hodnotou pro 1,66 nm (Obr. 3.19B). PMF naznačuje, že špičky uL4 a uL44 se mohou pohybovat v relativně širokém rozmezí; přiblížit jejich těžiště blíž než na 1,4 nm ale vyžaduje překonat bariéru Gibbsovu energii vyšší než 10 kJ/mol. Jedná se spíše o kvalitativní výsledek, neboť právě ve vysokoenergetických oblastech vzorkují MD simulace konformační prostor málo a PMF je zatížena vysokou chybou. Řešení nabízí metody rozšířeného vzorkování, jako je umbrella sampling nebo metadynamika. Zároveň je třeba poznamenat, že tvar PMF pro prázdný tunel nemusí být pro translaci a postup rodícího se proteinu tunelem rozhodující. S každou přidanou aminokyselinou se totiž hyperplocha Gibbsovy energie mění, a bylo by proto zajímavé se v budoucnu soustředit i na tunely s rodícím se proteinem.

Globálně nejnižší vzdálenost libovolných dvou atomů z obou špiček byla 0,26 nm. Přestože se jednalo o krátkodobý kontakt, může být zajímavé vysvětlit okolnosti, za kterých k němu došlo. Dlouhodobější přiblížení obou špiček, které by mohlo mít fyziologické důsledky, jsme však nepozorovali.

Dynamiku zúžení jsme chtěli popsat v termínech několika málo proměnných, ideálně s přímým propojením ke strukturním informacím. Vhodným rámcem pro popis takových systémů nabízí Markovovské stavové modely (MSM) [75, 137]. MSM reprezentují dynamiku systému jako přeskoky mezi diskrétními stavy. Přeskoky jsou tzv. bez paměti, což znamená, že pravděpodobnost přeskoku ze stavu A do stavu B nezávisí na stavu, který předcházel A. Pro popis biomolekul byl vytvořen rozsáhlý ekosystém teoretických konceptů, knihoven a nástrojů [138–141], což značně usnadňuje aplikaci MSM na konkrétní problémy. My jsme využili balík PyEMMA [140]. Stavy byly definovány pomocí kolek-



Obrázek 3.20: MSM dynamiky zúžení. Velikost kruhů odpovídá populaci stavů. Pravděpodobnosti přechodů jsou uvedené u šipek, tučněji je uvedena pravděpodobnost setrvání v aktuálním stavu. Molekulové modely špiček barevně souhlasí s barvou stavu. Vlevo je vždy špička uL4, vpravo uL22. Přepřacováno z diplomové práce T. Svatoňové [136].

tivních souřadnic TICA (z angl. *time-lagged independent component analysis*) [142, 143]. TICA transformují kartézské souřadnice tak, že maximalizují autokorelace, čímž reprezentují pomalé děje, které ale nemusí nutně vykazovat vysokou varianci.

MSM pro dynamiku špiček je na Obr. 3.20. Zhruběním modelu jsme identifikovali 6 stavů označených barvami světle modrá, růžová, oranžová, fialová, žlutá, zelené (v pořadí klesající populace). Výsledný model má tři hlavní vlastnosti: i) modelu dominuje jeden stav (světle modrý), ii) přechody mezi stavy jsou málo pravděpodobné a iii) až na výjimku přechodu mezi zeleným a fialovým stavem přechody vždy zahrnují centrální světle modrý stav. Z hlediska MSM je tedy konformační dynamika špiček pomalá.

Projekt dále rozvíjíme dvěma hlavními směry. Chtěli bychom kvantifikovat odchylky PMF související s výřezem. Sběr referenčních dat ze simulací celého ribozomu je výpočetně náročný a v tuto chvíli přesahuje naše technické možnosti. Proto simulujeme výřezy různých velikostí a budeme sledovat konvergenci fyzikálních případně i strukturních veličin. Dále pomocí umbrella sampling simulací lépe vzorkujeme konformace zúžení pro různé vzdálenosti těžišť a vypočteme přesnější PMF. Plánujeme to provést nejen pro prázdný tunel jako v diplomové práci T. Svatoňové, ale také pro komplexy s rodícím se peptidem.

Kapitola 4

Závěr

Činnost ribozomu představuje významný prvek v metabolismu živých organizmů. Předkládaná habilitační práce měla za úkol prezentovat vybrané projekty, ve kterých jsem v posledních asi 5 letech hrál vůdčí roli a které s ribozomem úzce souvisely.

Metodologicky byla má práce zaměřena na počítačové simulace, především ty s atomárním rozlišením, proto jim bylo v textu věnováno proporčně více místa než biologii ribozomu. Argumentoval jsem, že ribozom je pro simulace stále velkou výzvou, a to zejména kvůli své velikosti a časovým škálám dějů, kterých se účastní. Snažil jsem se ukázat, pro jaké otázky týkající se ribozomu mohou atomistické simulace najít uplatnění. Ještě širší kontext nabízí souhrnné články jiných skupin [144–146] a ty, na kterých jsem se sám podílel [8, 10, 80].

Většina zde prezentovaných projektů se nadále rozvíjí a to i díky vědkyním a vědcům, se kterými mám potěšení spolupracovat. Další simulace buď plánujeme, nebo již provádíme. Fascinující fungování ribozomu a jeho částí stále nabízí další otázky. A jak to ve vědě bývá, jedna zodpovězená otázka generuje mnoho dalších, jenž je třeba zodpovědět.

MD simulace se během dekad rozvoje staly etablovanou vědeckou technikou. Stále častěji dochází k jejich propojování s experimentálními technikami jako je cryo-EM nebo NMR, ale i s novými teoretickými přístupy v čele se strojovým učením a kvantovou chemií. Očekávám proto, že MD simulace budou schopné atakovat stále větší množství otázek, jejichž různorodost bude také růst. Nemyslím, že by mělo dojít k útlumu MD simulací z nějakých racionálních důvodů. Atomy a molekuly se totiž budou pohybovat i nadále, a proto i nadále bude potřeba teoretický rámec pro popis jejich chování, který simulace a statistická mechanika mohou nabídnout. Může se ovšem stát, že MD simulace postihne to, co se děje jinde ve společnosti; že vyjdou z módy. Tahle má obava je podložená dvěma postřehy. Móda je nevyzpytatelná a z vědy jako takové se stále více stává show pro netrpělivé.

Reference

- [1] P. Robustelli, S. Piana, D. E. Shaw, *Proc. Natl. Acad. Sci. U. S. A.* **115**, E4758 (2018). [DOI](#).
- [2] J. Huang, *et al.*, *Nat. Methods* **14**, 71 (2017). [DOI](#).
- [3] G. E. Palade, *J. Biophys. Biochem. Cytol.* **1**, 59 (1955).
- [4] T. M. Schmeing, V. Ramakrishnan, *Nature* **461**, 1234 (2009). [DOI](#).
- [5] G. Kramer, D. Boehringer, N. Ban, B. Bukau, *Nat. Struct. Mol. Biol.* **16**, 589 (2009). [DOI](#).
- [6] E. W. Mills, R. Green, *Science* **358**, eaan2755 (2017). [DOI](#).
- [7] N. Fischer, *et al.*, *Nature* **520**, 567 (2015). [DOI](#).
- [8] L. V. Bock, S. Gabrielli, M. H. Kolář, H. Grubmüller, *Ann. Rev. Biophys.* pp. 1–28 (2022).
- [9] V. Tobiasson, I. Berzina, A. Amunts, *Nat. Commun.* **13**, 6132 (2022). [DOI](#).
- [10] L. V. Bock, M. H. Kolář, H. Grubmüller, *Curr. Opin. Struc. Biol.* **49**, 27 (2018). [DOI](#).
- [11] N. Ban, P. Nissen, J. Hansen, P. B. Moore, T. A. Steitz, *Science* **289**, 905 (2000). [DOI](#).
- [12] B. T. Wimberly, *et al.*, *Nature* **407**, 327 (2000). [DOI](#).
- [13] F. Schlutzenzen, *et al.*, *Cell* **102**, 615 (2000). [DOI](#).
- [14] A. Brown, S. Shao, *Curr. Opin. Struc. Biol.* **52**, 1 (2018). [DOI](#).
- [15] S. A. Fromm, *et al.*, The translating bacterial ribosome at 1.55 Å resolution by open access cryo-EM (2022). [DOI](#).
- [16] J. Frank, *Journal of Structural Biology* **200**, 303 (2017). [DOI](#).
- [17] B. J. Alder, T. E. Wainwright, *J. Chem. Phys.* **31**, 459 (1959). [DOI](#).
- [18] J. A. McCammon, B. R. Gelin, M. Karplus, *Nature* **267**, 585 (1977). [DOI](#).
- [19] M. W. van der Kamp, *et al.*, *Structure* **18**, 423 (2010). [DOI](#).
- [20] U. Hensen, *et al.*, *PLOS ONE* **7**, e33931 (2012). [DOI](#).
- [21] K. Lindorff-Larsen, S. Piana, R. O. Dror, D. E. Shaw, *Science* **334**, 517 (2011). [DOI](#).
- [22] T. J. Lane, D. Shukla, K. A. Beauchamp, V. S. Pande, *Curr. Opin. Struc. Biol.* **23**, 58 (2013). [DOI](#).

- [23] Y. Shan, *et al.*, *J. Am. Chem. Soc.* **133**, 9181 (2011). DOI.
- [24] K. J. Kohlhoff, *et al.*, *Nature Chem.* **6**, 15 (2014). DOI.
- [25] B. L. de Groot, H. Grubmüller, *Science* (2001). DOI.
- [26] W. Kopec, *et al.*, *Nature Chem.* **10**, 813 (2018). DOI.
- [27] K. Y. Sanbonmatsu, S. Joseph, C.-S. Tung, *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15854 (2005). DOI.
- [28] M. Levi, J. K. Noel, P. C. Whitford, *Methods* **162–163**, 128 (2019). DOI.
- [29] E. Braun, *et al.*, *LiveCoMS* **1**, 5957 (2019). DOI.
- [30] J. A. Lemkul, *LiveCoMS* **1**, 5068 (2019). DOI.
- [31] M. J. Abraham, *et al.*, *SoftwareX* **1–2**, 19 (2015). DOI.
- [32] A. Grossfield, *et al.*, *LiveCoMS* **1**, 5067 (2019). DOI.
- [33] L. V. Bock, *et al.*, *Nat. Struct. Mol. Biol.* **20**, 1390 (2013). DOI.
- [34] J. Jumper, *et al.*, *Nature* **596**, 583 (2021). DOI.
- [35] M. Baek, *et al.*, *Science* **373**, 871 (2021). DOI.
- [36] K. Lindorff-Larsen, *et al.*, *PLOS ONE* **7**, e32131 (2012). DOI.
- [37] S. Rauscher, *et al.*, *J. Chem. Theory Comput.* **11**, 5513 (2015). DOI.
- [38] A. Pérez, *et al.*, *Biophys. J.* **92**, 3817 (2007). DOI.
- [39] M. Zgarbová, *et al.*, *J. Chem. Theory Comput.* **7**, 2886 (2011). DOI.
- [40] R. Aduri, *et al.*, *J. Chem. Theory Comput.* **3**, 1464 (2007). DOI.
- [41] Y. Xu, K. Vanommeslaeghe, A. Aleksandrov, A. D. MacKerell Jr., L. Nilsson, *J. Comput. Chem.* **37**, 896 (2016). DOI.
- [42] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **25**, 1157 (2004). DOI.
- [43] K. Vanommeslaeghe, E. P. Raman, A. D. MacKerell, *J. Chem. Inf. Model.* **52**, 3155 (2012). DOI.
- [44] Y. Qiu, *et al.*, *J. Chem. Theory Comput.* **17**, 6262 (2021). DOI.
- [45] W. D. Cornell, *et al.*, *J. Am. Chem. Soc.* **117**, 5179 (1995). DOI.
- [46] J. Huang, A. D. MacKerell Jr., *J. Comput. Chem.* **34**, 2135 (2013). DOI.
- [47] J. K. Noel, *et al.*, *PLoS Comp. Biol.* **12**, e1004794 (2016). DOI.
- [48] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983). DOI.
- [49] H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma, *J. Phys. Chem.* **91**, 6269 (1987). DOI.
- [50] C. W. Hopkins, S. Le Grand, R. C. Walker, A. E. Roitberg, *J. Chem. Theory Comput.* **11**, 1864 (2015). DOI.
- [51] K. A. Feenstra, B. Hess, H. J. C. Berendsen, *J. Comput. Chem.* **20**, 786 (1999). DOI.
- [52] M. Shirts, V. S. Pande, *Science* **290**, 1903 (2000). DOI.
- [53] V. S. Pande, *et al.*, *Biopolymers* **68**, 91 (2003). DOI.

-
- [54] V. Gapsys, B. L. de Groot, *eLife* **9**, e57589 (2020). [DOI](#).
- [55] K. El Hage, F. Hédin, P. K. Gupta, M. Meuwly, M. Karplus, *eLife* **7**, e35560 (2018). [DOI](#).
- [56] V. Gapsys, B. L. de Groot, *eLife* **8**, e44718 (2019). [DOI](#).
- [57] K. El Hage, F. Hédin, P. K. Gupta, M. Meuwly, M. Karplus, *eLife* **8**, e45318 (2019). [DOI](#).
- [58] R. C. Bernardi, M. C. R. Melo, K. Schulten, *Biochimica et Biophysica Acta (BBA) - General Subjects* **1850**, 872 (2015). [DOI](#).
- [59] Y. I. Yang, Q. Shao, J. Zhang, L. Yang, Y. Q. Gao, *J. Chem. Phys.* **151**, 070902 (2019). [DOI](#).
- [60] M. Invernizzi, P. M. Piaggi, M. Parrinello, *Phys. Rev. X* **10**, 041034 (2020). [DOI](#).
- [61] G. M. Torrie, J. P. Valleau, *J. Comput. Phys.* **23**, 187 (1977). [DOI](#).
- [62] H. Grubmüller, B. Heymann, P. Tavan, *Science* **271**, 997 (1996). [DOI](#).
- [63] A. Laio, M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12562 (2002). [DOI](#).
- [64] M. H. Kolář, T. Kubař, *J. Phys. Chem. B* **121**, 1520 (2017). [DOI](#).
- [65] D. Lucent, C. D. Snow, C. E. Aitken, V. S. Pande, *PLoS Comp. Biol.* **6**, e1000963 (2010). [DOI](#).
- [66] F. Noé, C. Clementi, *Curr. Opin. Struc. Biol.* **43**, 141 (2017). [DOI](#).
- [67] D. Trapl, *et al.*, *Front. Mol. Biosci.s* **6** (2019).
- [68] J. M. Thornton, C. A. Orengo, A. E. Todd, F. M. G. Pearl, *J. Mol. Biol.* **293**, 333 (1999). [DOI](#).
- [69] W. Kabsch, C. Sander, *Biopolymers* **22**, 2577 (1983). [DOI](#).
- [70] D. Frishman, P. Argos, *Proteins: Struct. Funct., Function, and Bioinformatics* **23**, 566 (1995). [DOI](#).
- [71] G. Nagy, C. Oostenbrink, *J. Chem. Inf. Model.* **54**, 266 (2014). [DOI](#).
- [72] A. Amadei, A. B. M. Linssen, H. J. C. Berendsen, *Proteins: Struct. Funct., Function, and Bioinformatics* **17**, 412 (1993). [DOI](#).
- [73] J. Shlens, A Tutorial on Principal Component Analysis (2014). [DOI](#).
- [74] S. A. M. Stein, A. E. Loccisano, S. M. Firestine, J. D. Evanseck, *Annual Reports in Computational Chemistry*, D. C. Spellmeyer, ed. (Elsevier, 2006), vol. 2, pp. 233–261. [DOI](#).
- [75] V. S. Pande, K. Beauchamp, G. R. Bowman, *Methods* **52**, 99 (2010). [DOI](#).
- [76] S. O. Nielsen, R. E. Buló, P. B. Moore, B. Ensing, *Phys. Chem. Chem. Phys.* **12**, 12401 (2010). [DOI](#).
- [77] E. Hruska, V. Balasubramanian, H. Lee, S. Jha, C. Clementi, *J. Chem. Theory Comput.* **16**, 7915 (2020). [DOI](#).
- [78] Y. Wang, J. M. Lamim Ribeiro, P. Tiwary, *Curr. Opin. Struc. Biol.* **61**, 139 (2020). [DOI](#).
- [79] F. Noé, G. De Fabritiis, C. Clementi, *Curr. Opin. Struc. Biol.* **60**, 77 (2020). [DOI](#).

- [80] M. H. Kolář, *Chemické listy* **115**, 550 (2021).
- [81] M. H. Kolář, *et al.*, *Nucleic Acids Res.* **50**, 2258 (2022). DOI.
- [82] H. McGrath, M. Černeková, M. H. Kolář, Binding of the peptide deformylase on the ribosome surface modulates the structure and dynamics of the exit tunnel interior (2022). DOI.
- [83] L. I. Malkin, A. Rich, *J. Mol. Biol.* **26**, 329 (1967). DOI.
- [84] J. Frank, *et al.*, *Nature* **376**, 441 (1995). DOI.
- [85] N. R. Voss, M. Gerstein, T. A. Steitz, P. B. Moore, *J. Mol. Biol.* **360**, 893 (2006). DOI.
- [86] D. N. Wilson, R. Beckmann, *Curr. Opin. Struc. Biol.* **21**, 274 (2011). DOI.
- [87] D. N. Wilson, *Nat. Rev. Microbiol.* **12**, 35 (2014). DOI.
- [88] E. Samatova, J. Dabberger, M. Liutkute, M. V. Rodnina, *Front. Microbiol.* **11** (2021). DOI.
- [89] K. Ito, S. Chiba, *Ann. Rev. Biochem.* **82**, 171 (2013). DOI.
- [90] A. R. Buskirk, R. Green, *Philos. Trans. R. Soc. Lon. B., Biol. Sci.* **372**, 20160183 (2017). DOI.
- [91] M. C. Kiel, H. Kaji, A. Kaji, *Biochem. Mol. Biol. Educ.* **35**, 40 (2007). DOI.
- [92] E. Ishii, *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* **112**, E5513 (2015). DOI.
- [93] T. Su, *et al.*, *eLife* **6**, e25642 (2017). DOI.
- [94] I. Švecová, Studium kotranslačního balení modelového peptidu metodami výpočetní biochemie, Bakalářská práce, VŠCHT Praha, Praha, Česká republika (2021).
- [95] J. Kästner, *WIREs Comp. Mol. Science* **1**, 932 (2011). DOI.
- [96] M. H. Zimmer, M. J. Niesen, T. F. Miller, *Biophys. J.* **120**, 2425 (2021). DOI.
- [97] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, *Comp. Phys. Commun.* **185**, 604 (2014). DOI.
- [98] M. Bonomi, *et al.*, *Nat. Methods* **16**, 670 (2019). DOI.
- [99] B. G. Barrell, A. T. Bankier, J. Drouin, *Nature* **282**, 189 (1979). DOI.
- [100] J. Ling, P. O'Donoghue, D. Söll, *Nat. Rev. Microbiol.* **13**, 707 (2015). DOI.
- [101] P. Kolosov, *et al.*, *Nucleic Acids Res.* **33**, 6418 (2005). DOI.
- [102] A. Brown, S. Shao, J. Murray, R. S. Hegde, V. Ramakrishnan, *Nature* **524**, 493 (2015). DOI.
- [103] G. Bussi, D. Donadio, M. Parrinello, *J. Chem. Phys.* **126**, 014101 (2007). DOI.
- [104] M. Parrinello, A. Rahman, *J. Appl. Phys.* **52**, 7182 (1981). DOI.
- [105] C. Lind, J. Sund, J. Åqvist, *Nat. Commun.* **4**, 1 (2013). DOI.
- [106] R. Kohler, R. A. Mooney, D. J. Mills, R. Landick, P. Cramer, *Science* **356**, 194 (2017). DOI.
- [107] M. W. Webster, *et al.*, *Science* **369**, 1355 (2020). DOI.
- [108] R. Bingel-Erlenmeyer, *et al.*, *Nature* **452**, 108 (2008). DOI.

-
- [109] H. McGrath, Studium interakce enzymů s povrchem ribozomu metodami výpočetní biochemie, Bakalářská práce, VŠCHT Praha, Praha, Česká republika (2021).
- [110] K. J. Smith, *et al.*, *Protein Sci.* **12**, 349 (2003). DOI.
- [111] J. S. Hub, B. L. de Groot, *PLoS Comp. Biol.* **5**, e1000480 (2009). DOI.
- [112] A. Sandikci, *et al.*, *Nat. Struct. Mol. Biol.* **20**, 843 (2013). DOI.
- [113] D. H. Goldman, *et al.*, *Science* **348**, 457 (2015). DOI.
- [114] J. Kubíček, Strukturní variabilita peptidových deformyláz, Středoškolská odborná činnost, Gymnázium, Nad Kavalírkou 5, Praha, Česká republika (2022).
- [115] K. Bokov, S. V. Steinberg, *Nature* **457**, 977 (2009). DOI.
- [116] G. E. Fox, *Evolution of the Protein Synthesis Machinery and Its Regulation*, G. Hernández, R. Jagus, eds. (Springer International Publishing, Cham, 2016), pp. 31–60. DOI.
- [117] N. A. Kovacs, A. S. Petrov, K. A. Lanier, L. D. Williams, *Mol. Biol. Evol.* **34**, 1252 (2017). DOI.
- [118] C.-W. Chen, *et al.*, *Antibiotics* **10**, 390 (2021). DOI.
- [119] J. Unge, *et al.*, *Structure* **6**, 1577 (1998). DOI.
- [120] P. Linhart, Computational biochemistry of disordered peptides, Bakalářská práce, VŠCHT Praha, Praha, Česká republika (2022).
- [121] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, *J. Biomol. NMR* **44**, 213 (2009). DOI.
- [122] Y. Shen, A. Bax, *J. Biomol. NMR* **56**, 227 (2013). DOI.
- [123] S. Piana, A. G. Donchev, P. Robustelli, D. E. Shaw, *J. Phys. Chem. B* **119**, 5113 (2015). DOI.
- [124] J. Henriques, C. Cragnell, M. Skepö, *J. Chem. Theory Comput.* **11**, 3420 (2015). DOI.
- [125] J. Mu, H. Liu, J. Zhang, R. Luo, H.-F. Chen, *J. Chem. Inf. Model.* **61**, 1037 (2021). DOI.
- [126] K. J. Kohlhoff, P. Robustelli, A. Cavalli, X. Salvatella, M. Vendruscolo, *J. Am. Chem. Soc.* **131**, 13894 (2009). DOI.
- [127] P. Robustelli, K. Kohlhoff, A. Cavalli, M. Vendruscolo, *Structure* **18**, 923 (2010). DOI.
- [128] P. M. Petrone, C. D. Snow, D. Lucent, V. S. Pande, *Proc. Natl. Acad. Sci. U. S. A.* **105**, 16549 (2008). DOI.
- [129] K. Dao Duc, S. S. Batra, N. Bhattacharya, J. H. D. Cate, Y. S. Song, *Nucleic Acids Res.* **47**, 4198 (2019). DOI.
- [130] H. Nakatogawa, K. Ito, *Cell* **108**, 629 (2002). DOI.
- [131] S. Bhushan, *et al.*, *PLOS Biology* **9**, e1000581 (2011). DOI.
- [132] J. Gumbart, E. Schreiner, D. N. Wilson, R. Beckmann, K. Schulten, *Biophys. J.* **103**, 331 (2012). DOI.
- [133] K. Kannan, *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15958 (2014). DOI.

- [134] S. Sothiselvam, *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* **111**, 9804 (2014). [DOI](#).
- [135] J. Lu, C. Deutsch, *Nat. Struct. Mol. Biol.* **12**, 1123 (2005). [DOI](#).
- [136] T. Svatoňová, Struktura a dynamika zúžení peptidového tunelu v ribozomu, Diplomová práce, VŠCHT Praha, Praha, Česká republika (2022).
- [137] B. E. Husic, V. S. Pande, *J. Am. Chem. Soc.* **140**, 2386 (2018). [DOI](#).
- [138] G. R. Bowman, X. Huang, V. S. Pande, *Methods* **49**, 197 (2009). [DOI](#).
- [139] G. R. Bowman, K. A. Beauchamp, G. Boxer, V. S. Pande, *J. Chem. Phys.* **131**, 124101 (2009). [DOI](#).
- [140] M. K. Scherer, *et al.*, *J. Chem. Theory Comput.* **11**, 5525 (2015). [DOI](#).
- [141] A. Mardt, L. Pasquali, H. Wu, F. Noé, *Nat. Commun.* **9**, 5 (2018). [DOI](#).
- [142] G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, F. Noé, *J. Chem. Phys.* **139**, 015102 (2013). [DOI](#).
- [143] C. R. Schwantes, V. S. Pande, *J. Chem. Theory Comput.* **9**, 2000 (2013). [DOI](#).
- [144] K. Y. Sanbonmatsu, *Curr. Opin. Struc. Biol.* **22**, 168 (2012). [DOI](#).
- [145] G. I. Makarov, T. M. Makarova, N. V. Sumbatyan, A. A. Bogdanov, *Biochem. Mosc.* **81**, 1579 (2016). [DOI](#).
- [146] F. Trovato, E. P. O'Brien, *Ann. Rev. Biophys.* **45**, 345 (2016). [DOI](#).