

# Nova metoda analize simulacija molekulske dinamike toplinskim mapama

---

**Kožić, Matej**

**Master's thesis / Diplomski rad**

**2024**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:217:320068>

*Rights / Prava:* [In copyright](#)/[Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2024-07-20**



*Repository / Repozitorij:*

[Repository of the Faculty of Science - University of Zagreb](#)





Sveučilište u Zagrebu  
PRIRODOSLOVNO-MATEMATIČKI FAKULTET  
Kemijski odsjek

Matej Kožić

# **NOVA METODA ANALIZE SIMULACIJA MOLEKULSKE DINAMIKE TOPLINSKIM KARTAMA**

**Diplomski rad**

predložen Kemijskom odsjeku

Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu

radi stjecanja akademskog zvanja

magistra kemije

Zagreb, 2024.



Ovaj rad je izrađen na Zavodu za fizikalnu kemiju Kemijskog odsjeka Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu pod mentorstvom prof. dr. sc. Branimira Bertoše u okviru projekta Hrvatske zaklade za znanost pod nazivom *Manganovi metalosenzori* čiji je voditelj B. Bertoša (IP-2020-02-3446). Rad je predan na ocjenu Kemijskom odsjeku Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu radi stjecanja zvanja magistra kemije.

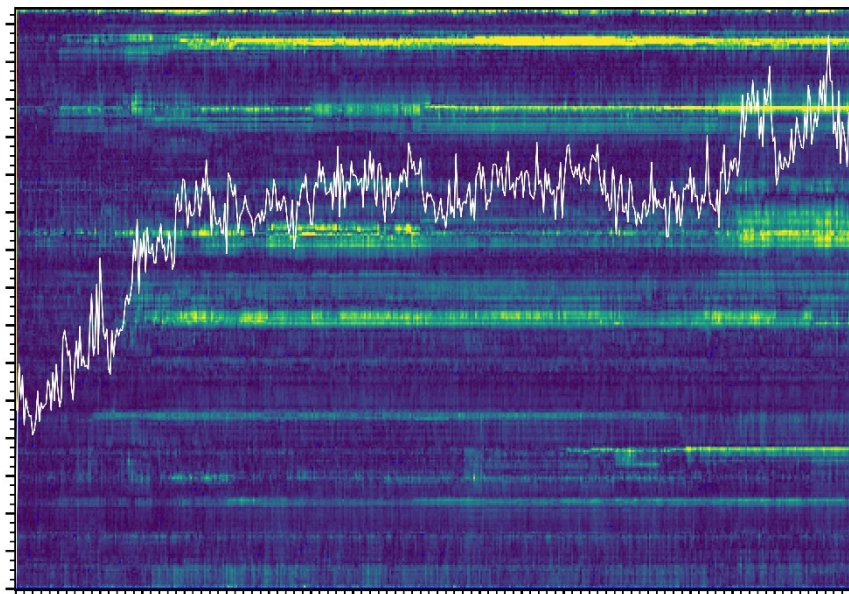


## Zahvale

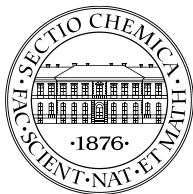
Prvenstveno, autor se zahvaljuje mentoru prof. dr. sc. Bertoši na vodstvu i strpljenju, a najviše na prepoznavanju sirove ideje kao vrijedne proganjanja, te omogućavanju toga. Posebne zahvale upućene su dr. sc. Tandarić na testiranju ranih verzija koda, dijeljenju vrijednih uvida, ustupanju simulacija za interno testiranje, i svega ostaloga povezanog uključujući ali ne i ograničeno na tehničku (i emocionalno-psihičku haha) podršku. Nakon toga upućene su zahvale mag. biol. mol. Matošević na vrlo vrijednoj tehničkoj pomoći, s posebnim naglaskom na ustupanje simulacija za testiranje i validaciju te pomoći oko toga. I na kraju, autor se zahvaljuje svima ostalima koji su prepoznali možebitnu korisnost pristupa i vjerovali u pothvat provedbe od tek difuzne ideje, do razrađenog i praktično korisnog koncepta, i funkcionalnog samodostatnog *easy-to-use software-a*.

*E da, i hvala mami tati dedi baki prijateljima bratu i dr. Pola ovog diplomskog i cijeli kod za ovu novu metodu napisan je nakon 23h. To ne bi mogao bez mira svoje sobe, stoga bi mi vjerojatno diplomski (a i cijela buduća karijera) bili znatno drukčiji bez svog „studentskog smještaja“, Lh22, gdje sam imao mira nesmetano raditi do dva ujutro i igrati se s idejama koje bi „možda mogle raditi“. Hvala.*

*I hvala na čitanju ovoga. Kao zahvala, izvoli ovu lijepu moderno-umjetničku sliku:*







Sveučilište u Zagrebu

Prirodoslovno-matematički fakultet

Kemijski odsjek

Diplomski rad

## Sažetak

# NOVA METODA ANALIZE SIMULACIJA MOLEKULSKE DINAMIKE TOPLINSKIM KARTAMA

Matej Kožić

Simulacije molekulske dinamike generiraju trajektorije koje opisuju evoluciju sustava kroz vrijeme. One se analiziraju vizualno (kvalitativno) te statističkim tehnikama (kvantitativno). Među najčešće korištenim analitičkim metodama spadaju *RMSD*, *RMSF*, i *R<sub>gyr</sub>*. Spomenute metode ograničene su svojom striktno statističkom prirodom, te ne reprezentiraju konkluzivno tijekom simulacije. Zbog kompleksnosti istraživanih sustava, robusne, brze i intuitivne metode analize od velikog su značaja i istraživačima koji provode istraživanja, kao i korisnicima tih rezultata. Jednoznačne i intuitivne kvantitativne analize pomažu razlučivanju važnih podataka od šuma, umanjuju ljudsku pogrešku, potiču efektivnije prenošenje informacija u svim stadijima istraživanja i pojednostavljuju komunikaciju znanstvenih pronalazaka. U sklopu ovog diplomskog rada razvijena je nova metoda analize simulacija temeljena na prikazivanju trajektorija toplinskim mapama nazvana „Mape trajektorija“ (engl. *Trajectory Maps*). Ovim pristupom nudi se intuitivna vizualizacija cijelog tijeka simulacije, izravna i konkluzivna usporedba više setova simulacija, intuitivna i efektivna vizualizacija pomicanja regija proteina tijekom simulacije, i dr. Metoda je razvijena sa svrhom da nadopunjuje postojeće kvantitativne i kvalitativne analize i s njima se kombinira, što ju čini posebice korisnom.

(53 stranica, 9 slika, 46 literaturnih navoda, jezik izvornika: Hrvatski)

Rad je pohranjen u Središnjoj kemijskoj knjižnici Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu, Horvatovac 102a, Zagreb i Repozitoriju Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu

**Ključne riječi:** mapa trajektorija, metodologija analize trajektorija, molekulska dinamika, računalna biokemija, računalne simulacije proteina, toplinska mapa

**Mentor:** prof. dr. sc. Branimir Bertoša

**Ocjenitelji:**

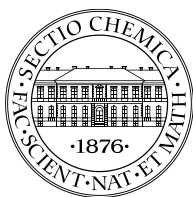
1. prof. dr. sc. Branimir Bertoša
2. izv. prof. dr. sc. Aleksandra Maršavelski
3. izv. prof. dr. sc. Ivica Đilović

Zamjena: prof. dr. sc. Tomica Hrenar

**Datum diplomskog ispita:** 29. 2. 2024.







University of Zagreb

Faculty of Science

Department of Chemistry

Master Thesis

## Abstract

# A NOVEL METHOD OF MOLECULAR DYNAMICS SIMULATIONS ANALYSIS USING HEATMAPS

Matej Kožić

Molecular dynamics simulations generate trajectories which depict the system's evolution through time. The results are analyzed both visually (qualitatively) and statistically (quantitatively). Among the most common analytical methods are *RMSD*, *RMSF*, and  $R_{gyr}$ , which due to their strictly statistical character inconclusively represent simulation courses. Owing to the complexity of analyzed systems, robust, fast, and intuitive analysis methods are invaluable to both researchers conducting the studies and wider audiences using those results. Clear and intuitive quantitative analyses aid in distinguishing important data from noise, minimize human error, and facilitate the transfer of information in all stages of research, simplifying communication of scientific findings. In this master's thesis a novel method of molecular dynamics simulation analysis was developed, based on representing trajectories with heatmaps called "Trajectory maps". This approach offers an intuitive visualization of the entire simulation course, a direct and conclusive comparison of multiple simulation sets, an intuitive and an effective visualization of protein region movements during simulation, etc. The method is designed to complement existing quantitative and qualitative analysis approaches and to combine with them, making it particularly useful to both experienced and new researchers.

(53 pages, 9 figures, 46 references, original in Croatian)

Thesis deposited in Central Chemical Library, Faculty of Science, University of Zagreb, Horvatovac 102a, Zagreb, Croatia and in Repository of the Faculty of Science, University of Zagreb

**Keywords:** Computational Biochemistry, Computational Biophysics, Heatmaps, Molecular Dynamics, Protein Simulations, Trajectory Analysis Methodology, Trajectory Map

**Mentor:** Dr. Branimir Bertoša, Professor

**Reviewers:**

1. Dr. Branimir Bertoša, Professor
2. Dr. Aleksandra Maršavelski, Associate Professor
3. Dr. Ivica Đilović, Associate Professor

Substitute: Dr. Tomica Hrenar, Professor

**Thesis defense date:** 29. 2. 2024



# Sadržaj

<b>1. Uvod</b> .....	<b>1</b>
<b>2. Literaturni pregled</b> .....	<b>2</b>
2.1. Računalne metode molekuskog modeliranja.....	2
2.2. Simulacije korištene za testiranje novo razvijene metode.....	11
<b>3. Metode</b> .....	<b>16</b>
3.1. Konceptualni razvoj mapa trajektorija.....	16
3.2. Primjena i validacija metode.....	20
<b>4. Rezultati</b> .....	<b>24</b>
4.1. Analiza simulacija TAL proteina: diferencijacija stabilnosti simuliranih sustava.....	24
4.2. Analiza simulacija MntR proteina: usporedba strukturne dinamike više simulacija.....	26
4.3. Analiza simulacija peroksidaze hrena: kvantifikacija konformacijskih promjena.....	28
<b>5. Rasprava</b> .....	<b>32</b>
<b>6. Zaključak</b> .....	<b>32</b>
<b>7. Literaturni izvori</b> .....	<b>34</b>
<b>8. Popis oznaka, kratica i simbola</b> .....	<b>37</b>
<b>9. Izjava nekorištenja alata umjetne inteligencije</b> .....	<b>39</b>
<b>10. Životopis</b> .....	<b>40</b>



## 1. Uvod

U suvremenoj znanosti teorijski i računalni pristupi neizostavni su i nezamjenjivi dijelovi znanstvenih istraživanja. Među raznim računalnim metodama koje nalaze svoju primjenu u području istraživanja bioloških sustava (engl. *life sciences*), posebno se ističe molekulska dinamika (MD). Simulacijama MD dolazi se do uvida vrijednih cijelom spektru znanstvenih disciplina: od fizike i biofizike, kroz biokemiju i farmaceutiku, do strukturne i molekulske biologije.[1-6] Stoga, nije iznenađujuće da simulacije MD svoju opsežnu uporabu nalaze kako u industriji, tako i u akademiji. Neke od istaknutih primjena su: pronalaženje novih lijekova u farmaceutskoj industriji [4], istraživanje i priprava poboljšanih proteinskih mutanata[5], dobivanje izravnog uvida u biološke, biokemijske i markomolekulske sustave[6] i dr.[1]

Simulacijama MD dobivaju se trajektorije koje opisuju evoluciju sustava kroz vrijeme, a svi uvidi dolaze upravo iz analize spomenutog. Trajektorije su primarni produkti simulacija koje je potrebno obraditi i iz njih izdvojiti posebno vrijedne informacije. Primjerice, trajektorije za protein sastavljen od 300 aminokiselina u solvacijskoj kutiji odgovarajućih dimenzija tijekom simulacije trajanja od nekoliko stotina nanosekundi, mogu iznositi i po stotine GB. Od tih stotina GB, moguće je da tek par MB sadrži korisne informacije. Od kretanja svih prisutnih molekula vode i svih aminokiselina u svim vremenskim koracima simulacije, moguće je da pomicanje tek nekoliko aminokiselina u nekom vremenskom periodu ima izrazit značaj za razumijevanje zbivanja u sustavu. Iz tog razloga, neophodno je imati efektivne i robusne metode analize kojima se mogu efikasno i konkluzivno razlučiti najbitniji podatci od manje značajnih, te podatci od šuma.[1, 2]

## 2. Literaturni pregled

### 2.1. Računalne metode molekulskog modeliranja

Cilj računalnih metoda je modelirati molekule i simulirati ponašanje sustava u realističnim uvjetima. Glavna podjela računalnih metoda zasniva se na tretiranju elektrona iz čega dalje proizlazi način modeliranja promatranog sustava. Klasična molekulska mehanika (MM) temeljena je na klasičnoj (Newtonovoj) mehanici, a kvantno-mehaničke (QM) metode još se nazivaju i *ab initio* (lat. "od početka") metode jer „od početka“ proizlaze iz fizičko-matematičkog modela koji ne koristi empirijske parametre.[2, 3]

Suštinska razlika ova dva skupa metoda je u načinu modeliranja elektrona. Klasična mehanika elektrone ne tretira eksplicitno, već su efekti koje elektroni stvaraju u molekuli uračunati implicitno. S druge strane, kvantno-mehaničke metode zasnivaju se na rješavanju pripadne vremenski ovisne Schrödingerove jednadžbe što rezultira opisom ponašanja promatranog sustava u vremenu. Kako je Schrödingerova jednadžba analitički rješiva samo za jednočestične sustave, a općenito samo za određene iznose energije, za rješavanje više-elektronskih sustava potrebno je uvesti aproksimacije. Temeljna aproksimacija je Born-Oppenheimerova aproksimacija koja pretpostavlja da se zbog razlike u masi elektrona i jezgre njihova gibanja mogu odvojiti. Time se ukupna Schrödingerova jednadžba razdvaja na dvije međusobno neovisne funkcije: **I**) elektronska valna funkcija i **II**) jezgrena valna funkcija. Tako se (uz još dodatnih aproksimacija) omogućava aproksimativno rješavanje Schrödingerove jednadžbe za više-elektronske sustave.[2, 3]

Glavne prednosti QM metoda su što one konvergiraju u egzaktno rješenje (pod uvjetom da su korištene aproksimacije prihvatljivo malenog utjecaja). Pored toga, velika prednost QM metode (za razliku od MM) je što eksplicitno modeliraju elektronske interakcije. Zbog toga, s QM metodama je moguće prikazati i istraživati nastajanje i pucanje kemijskih veza, te druge interakcije elektronskih oblaka. To ih čini nezamjenjivima u slučajevima gdje se eksplicitno istražuju navedene elektronske interakcije (npr. modeliranje kemijskih reakcija). Nažalost, te prednosti dolaze uz veliku računalnu zahtjevnost te ograničenje s obzirom na veličinu sustava. To su glavni nedostaci QM metoda. Kompleksnost računa, a samim time i njegovo trajanje, eksponencijalno raste s veličinom sustava. To QM metode čini neizvedivima za promatranje velikih sustava (kako je slučaj u biokemijskim i biofizičkim primjenama gdje se promatraju npr. proteini solvatirani u vodi). U takvim se slučajevima umjesto QM pristupa koriste manje precizne metode prihvatljivih duljina trajanja računa. Kako se simulacijama

proteina generalno ne promatraju eksplicitne elektronske interakcije, klasične metode predstavljaju prihvatljivo preciznu aproksimaciju takvih sustava.[2, 3]

Uvođenjem dodatnih aproksimacija kroz empirijski određene parametre smanjuje se kompleksnost QM računa, a samim time i njegovo trajanje. Među najpopularnijim takvim metodama je teorija funkcionala gustoće (engl. *Density Functional Theory*, DFT). Metoda DFT zasniva se na aproksimaciji da svojstva molekula proizlaze iz njihove raspodjele elektronskih oblaka. Na temelju toga se definira funkcional gustoće (koji je empirijski određen), što onda izbjegava najzahtjevniji dio računa klasičnih QM metoda.[2, 3]

Kod velikih sustava gdje je poznata lokacija gdje se događa elektronska interakcija od posebnog interesa poput pucanja ili nastanka kovalentne veze (npr. aktivno mjesto enzima), koristi se QM/MM metoda. Ona predstavlja spoj klasične i kvantne mehanike kroz promatranje samo manjeg djela sustava od interesa kvantno-mehaničkim računom, a ostatak sustava tretira se klasično molekulsko-mehaničkim računom. To pruža prednosti egzaktnijih (ili egzaktnih) računa u promatranoj regiji (npr. 1 nm oko reakcijskog centra enzima) koji eksplicitno opisuju elektronske interakcije, dok ostatak sustava (poput solvacijske kutije i dr.) biva simuliran vremenski efikasnijim pristupom klasične mehanike.[2, 3]

Dok QM metode promatraju sustav preko valnih funkcija, klasična molekulska mehanika sustav opisuje (aproksimira) kao sklop kuglica koje predstavljaju atome, povezane oprugama koje predstavljaju kovalentne veze. Svojstva kuglica i opruga (poput mase, naboja, radijusa, konstante opruge i dr.) određena su na temelju eksperimentalnih podataka. Kuglice (atomi) se nalaze u trodimenzionalnom prostoru i svaka je opisana s tri koordinate, a interakcije između njih opisane su izrazima klasične mehanike. Oni uključuju potencijal harmonijskog oscilatora (za istezanje kovalentnih veza ili savijanje valentnih kuteva), Lennard-Jonesov potencijal za opis van der Waalsovih interakcija, Coulombov potencijal za opis elektrostatskih interakcija, i dr. Svi ti parametri koji opisuju sustav opisani su poljem sila. Ovaj pristup pogodan je za istraživanje fizikalnih svojstava kod kojih nije neophodno eksplicitno uzeti elektrone u obzir, već je dovoljno samo implicitno. Kroz uvođenje parametara koji u stvarnosti ovise o elektronima, uzeti su u obzir efekti koje elektroni stvaraju u sustavu. Takvi uključuju konformacije koje sustav (protein, DNA/RNA i dr.) poprima, dostupnost određenih skupina otapalu (najčešće vodi), interakcije između molekula u sustavu, i dr.[2, 3]

Zbog veličine sustava i karaktera svojstava koja se istražuju, simulacije proteina najčešće se vrše klasičnom mehanikom. Konkretno, najkorištenija metoda je simulacijama molekulske dinamike



(MD) koja je iskazana kao pouzdana za istraživanja navedenog tipa. [2, 3] Ista je korištena u sklopu ovog istraživanja za stvaranje sustava koji su onda analizirani mapama trajektorija.

### 2.1.1. Simulacije molekulske dinamike

Molekulska dinamika (MD) je među najkorištenijim MM metodama. Metoda MD počiva na uvođenju kinetičke energije u sustav, što se čini preko temperature sustava. Svaka čestica dobiva vektor brzine određen na temelju njene mase, i prema raspodjeli poput Boltzmannove ili dr. Informacije potrebne za računanje interakcija u sustavu zapisane su u tzv. polju sila, kojim se računaju sve sile koje djeluju na svaku česticu sustava. Tako se poljem sila računaju sile koje svaka čestica osjeća. Preko resultantnog gradijenta se računaju nove koordinate svake čestice na koje se one „pomiču“, kao rezultat sumarnog djelovanja sile u prethodno određenom vremenskom koraku. Stoga, polje sila je (uz uvođenje brzina u sustav kroz temperaturu) svojevrsna konceptualna okosnica simulacija molekulske dinamike.[2, 3]

U sklopu aproksimacije kemijskog sustava Newtonovom mehanikom, poljem sila se modeliraju vezne i ne-vezne interakcije kao klasične sile s pripadnim parametrima. Parametri polja sila, koji su specifični svakom individualnom polju sila, mogu biti dobiveni eksperimentalno ili kvantno-kemijskim računima[7]. Ukupna sila koju čestica osjeća dobiva se kao suma svih sila koje djeluju na česticu. Stoga, polje sila treba opisivati sve vezne i ne-vezne interakcije između molekula. One uključuju potencijale: rastezanja kemijskih veza, savijanja veza, torzijskih kutova, elektrostatskih interakcija, i van der Waals interakcija. Pored toga, razna polja sila mogu imati specifične doprinose poput miješanih doprinosa,  $\pi$ - $\pi$  *stacking* interakcija i dr. Svaki od tih doprinosa modelira se zasebnom matematičkom funkcijom koja je najbolja aproksimacija za isti, s najmanjom računalnom zahtjevnosti izračuna. Kao primjer, rastezanje kemijskih veza može se aproksimirati harmoničkim oscilatorom, a elektrostatske interakcije modeliraju se Coulombovim zakonom. Različita polja sila različito pristupaju modeliranju tih interakcija kako bi bila specifično prilagođena određenoj svrsi. Iz tog razloga, potrebno je odabrati polje sila prigodno za simulirani sustav. Polja sila mogu biti optimizirana za male organske molekule, za tekućine, za biokemijske sustave, i dr., s razlikama u rukovanju s parametrima polja sila na način da su rezultati najbolje prilagođeni pripadajućoj svrsi. Neka od najčešće korištenih za istraživanje biokemijskih (bioloških) sustava su CHARMM[8], GROMOS[9], OPLS[10] i AMBER[11]. Polja sile CHARMM i AMBER optimizirana su za biološke makromolekule poput proteina, RNA i DNA. Polja OPLS i GROMOS više su prilagođena za manje biološke molekule. U simulacijama MD proteina najčešće korištena polja su varijante i inačice CHARMM i/ili AMBER polja sila. Budući da su polja sila empirijski određena, s vremenom se ažuriraju i postaju točnija kako eksperimentalna mjerenja postaju preciznija ili kako se

provode novi kvantno-mehanički izračuni. Stoga, polja sila se često ažuriraju novim verzijama i prilikom provođenja istraživanja vrijedi provjeriti da se koristi recentna inačica.[2, 3]

Svrha simulacija MD je vjerno reprezentirati realni sustav *in silico*. Da bi to bilo moguće, potrebno je simulaciju MD uskladiti sa fizičkim i kemijski zakonima koji opisuju realne sustave. Konkretno, sa zakonima statističke termodinamike i mehanike. Time se definira odnos makroskopskih i mikroskopskih svojstava sustava. Konkretno, odnos opservabli poput tlaka ( $p$ ), volumena ( $V$ ), broja čestica ( $N$ ) i temperature ( $T$ ) s trodimenzionalnim koordinatama atoma u prostoru s pripadnim količinama gibanja. Različiti ansambli u kojima se sustav može nalaziti definiraju se kao višedimenzijski prostori koji sadrže sva mikrostanja dostupna sustavu pri definiranim opservablama  $N$ ,  $p$ ,  $T$  i  $V$  (koje predstavljaju mjerive termodinamičke parametre). Tako se dolazi do različitih makrostanja sustava, koja se dobivaju fiksiranjem određenih opservabli. Izobarno-izotermni  $NPT$  ansambl postiže se održavanjem konstantnog tlaka i temperature (a obzirom da je sustav zatvoren, konstantna je i brojnost čestica  $N$ ). U  $NPT$  ansamblu sustavu se uravnotežuje (ekvilibrira) volumen. Održavanjem konstantnog volumena i temperature sustav biva u kanonskom  $NVT$  ansamblu, gdje se uravnotežuje tlak. Ansambli  $NPT$  i  $NVT$  najčešće su korištena dva jer odgovaraju uvjetima koji se koriste u eksperimentima. Pored navedenih, mogući su i razni drugi ansambli poput velekanonskog  $\mu VT$  gdje je održavan konstantnim kemijski potencijal i dr. Održavanje konstantnim (fiksiranje) tih parametara postiže se matematičkim algoritmičkim barostatom i/ili termostatom koji korigira brzine atoma kako bi mikroskopska slika sustava odgovarala makroskopskoj slici opservabli. Taj korak (ili više njih) sačinjava ekvilibraciju sustava. [2, 3]

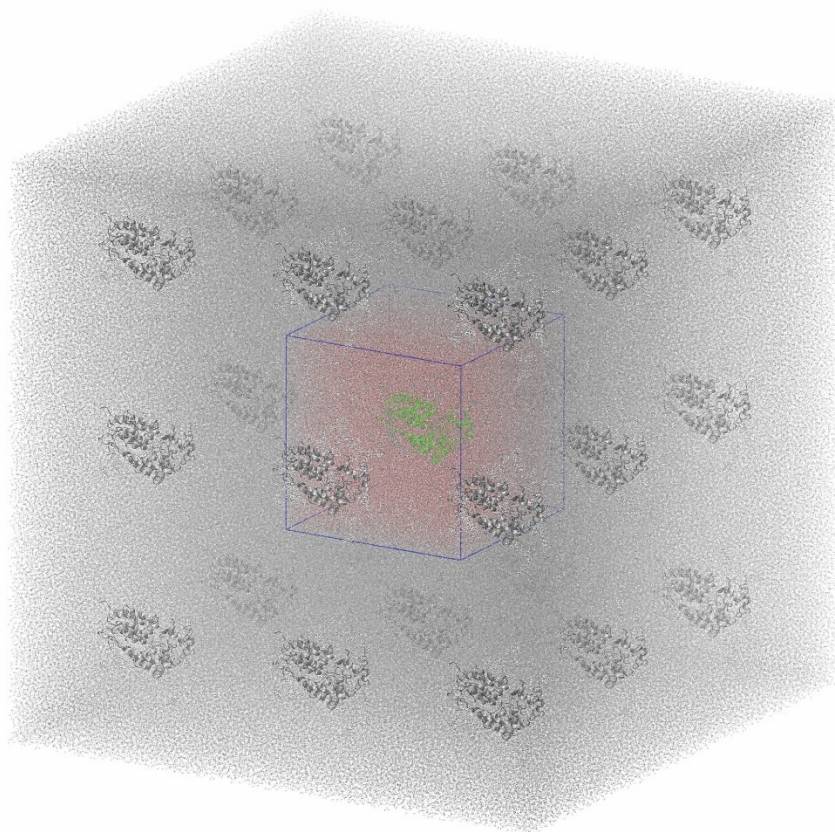
Stoga, cjelokupno gledano, na početku se atomima dodjeli nasumična brzina na temelju distribucije poput Boltzmanove, i onda im se računa nova vrijednost pozicije i brzine na temelju sila koje osjeća u vremenskom koraku koji je prethodno definiran, s opservablama ansambla definiranim kako jesu. Time se prvo uravnoteži i uskladi sa stvarnosti odnos brzina i volumena ( $NPT$  ansambl), pa odnos brzina i tlaka ( $NVT$  ansambl).[2,3]

S ekvilibriranim sustavom kreće se u produkcijsku fazu simulacija MD. U prvom koraku se računaju nove vrijednosti pozicija, brzina i ubrzanja koje su dobivene djelovanjem sila (definiranih poljem sila) u definiranom vremenskom koraku. Tako se simulira „pomak“ atoma u djeliću vremena (koje čini jedan vremenski korak). Taj vremenski korak prethodno se definira tako da iznosi manje nego najbrži mogući pomak tog sustava (u realnim uvjetima). To najčešće predstavlja period vibracije C-H veze, što iznosi cca. 10 fs. Stoga, da bi vremenski korak bio kraći, obično se uzima kao 1 fs. Što je vremenski korak kraći, to je potrebno više računalnog vremena za računanje simulacije određene

duljine trajanja iz očitog razloga. Poželjno je imati što dulji vremenski korak (bez da se žrtvuje točnost simulacija), kako bi simulacije bile manje vremenski zahtjevne. Korištenjem posebnih algoritama koji fiksiraju C-H veze, poput LINCS[12] i dr. algoritama, moguće je koristiti veće vremenske korake, najčešće oko 2 fs.[2, 3]

Simulirani pomak atoma, koji predstavlja „kretanje“ atoma, provodi se matematičkim algoritmima integracije od kojih su najpopularniji Verletov algoritam, „*leap-frog*“ algoritam i sl. Iz izračunatih novih brzina i ubrzanja računaju se nove vrijednosti sila u sustavu, te se primjenjuju periodički uvjeti uz korekcije temperature i tlaka (koristeći matematički termostat i/ili barostat). Zatim se računaju opservable, koje se s trajektorijama sustava, zapisuju u datoteku koja je produkt simulacije. To predstavlja jedan korak, kakav se iterira tijekom vremena trajanja simulacije. Na samom kraju se na temelju trajektorija i opservabli ustanovi jeli sustav postigao ravnotežu, i ako nije, simulacije se (po potrebi) produže ili ponove. Ovisno o tipu istraživanja te veličini i karakteru sustava od interesa, duljine simulacija MD u današnjim istraživanjima iznose od par desetaka nanosekundi do par mikrosekundi. Pored karaktera sustava i istraživanja, bitno je uzeti u obzir i računalne resurse na raspolaganju, koji (svojom manjkavosti) mogu biti usko grlo.[2, 3]

Što se tiče sustava simuliranih s MD, oni su najčešće biomakromolekule poput proteina ili RNA/DNA okružena molekulama vode, a sustav je ograničen kutijom periodičnih rubnih uvjeta (engl. *Periodic Boundary Conditions box*, PBC kutija, slika 1). Njena svojstva su takva da ako čestica izađe iz jedne stranice (npr.  $x$ ), identična čestica ulazi u kutiju kroz suprotnu stranicu ( $-x$ ) s istim parametrima vektora brzine i ubrzanja. Kao rezultat, efektivno se simulira beskonačno veliki (periodični) sustav gdje se čestice slobodno gibaju bez uvođenja nerealističnih granica. To ide u korist točnosti simulacije u reprezentaciji realnih sustava. Kutije PBC najčešće su kocke ili pravokutnici, no koriste se i drugi pravilni i nepravilni poliedri, ovisno o prostornim potrebama proučavanog sustava. Aproksimativno sferno simetrične molekule ne trebaju poseban tretman, dok molekule s npr. jednom vrlo dugom stranicom (npr. DNA/RNA molekule, štapićasti proteini i sl.) mogu koristiti od kvadarske PBC kutije, ili pak nekog drugog ezoteričnog oblika. Još jedan razmjerno često korišteni oblik je dodekaedar, a razlog toga je ušteda računalne snage jer je potrebno simulirati manje molekula otapala koje bi inače završile u kutovima kockaste PBC kutije. Prednost dodekaedra može biti u uštedi vremena potrebnog za simulaciju čak i kod aproksimativno sferno simetričnih sustava koji bi inače bili simulirani u kockastoj PBC kutiji.[2, 3]



**Slika 1.** Kockasta kutija periodičnih rubnih uvjeta.[2]

Sumarno, koraci izvođenja simulacije molekulske dinamike su sljedeći: **I)** izgradnja sustava, **II)** minimizacija energije (optimizacija geometrije), **III)** ekvilibracija *NVT*, **IV)** ekvilibracija *NPT* i na kraju **V)** produkcija simulacije MD. Za početak je potrebno izgraditi sustav, što uključuje korištenje 3D strukture (makro)molekule od interesa. Inicijalna 3D struktura može se dobiti dohvaćanjem kristalne ili NMR strukture molekule s mrežno dostupne baze podataka, među kojima je najčešće korištena engl. *Protein Dana Bank*, PDB, ili izgradnjom računalnim programima (npr. *AlphaFold*). S obzirom da je u fokusu diplomskog rada alat za analizu trajektorija simulacija proteinskih sustava, dalje u tekstu uzima se da se za simulaciju MD koristi proteinska struktura. Takvu strukturu često je potrebno urediti modeliranjem nedostajućih aminokiselina, određivanjem protonacijskih stanja aminokiselina, te definiranjem cisteinskih mostova, eventualne glikozilacije i dr., nakon čega se struktura protonira. Pripremljena 3D struktura proteina okružuje se molekulama vode i definiraju se periodični rubni uvjeti. Navedeno se vrši računalnim programima među kojima su popularni AMBER, CHARMM-GUI, GROMACS i sl. S pripremljenim sustavom ulazi se u korak minimizacije energije, gdje je cilj metodom MM spustiti sustav u energetski minimum. To se provodi kako bi se optimizirala geometrija sustava i otklonile eventualne steričke smetnje stvorene tijekom izgradnje istoga. Minimizirani sustav ekvilibrira

se u *NVT* ansamblu, za čime slijedi ekvibracija u *NPT* ansamblu. Tek nakon toga pokreće se simulacija MD, što se još naziva i produkcija. Nakon produkcijske faze slijedi analiza, o kojoj će se posebno diskutirati u slijedećem potpoglavlju.[2, 3]

### 2.1.2. Analize simulacija

Konceptualno, često je cilj analiza simulacija destilirati veliku količinu informacija sadržanih u trajektorijama u samo korisne i bitne uvide o simuliranom sustavu. Možemo o tome razmišljati na način da se analizom stotine GB informacija prevode u tek par bitova koji sadrže samo informacije relevantne pitanju na koje se nastojalo odgovoriti simulacijom (npr. set grafova i metrika koji ukazuju koji je od dva simulirana sustava stabilniji). U pozadini istraživanja uvijek postoji pitanje čiji se odgovor traži, a ovisno o njemu biraju se postupci kojima se analiziraju simulacije.

Univerzalan i često prvi korak u analizi simulacije je kvalitativna analiza vizualizacijom. Ona uključuje doslovno promatranje centriranog sustava koristeći neki od vizualizacijskih programa (npr. VMD [13], Chimera [14] i dr.). Više o centriranju sustava biti će rečeno u nastavku ovog poglavlja. Simulacija je četverodimenzionalni sustav, gdje se protein s tri prostorne dimenzije mijenja kroz četvrtu, vremensku, dimenziju. Stoga, vizualizacija se vrši promatranjem ponašanja trodimenzionalnog sustava kao serije kadrova koji predstavljaju vremensku evoluciju, kao interaktivni film. Na temelju promatranja i interpretiranja toga, koristeći vizualizacijske tehnike za izolirati, istaknuti ili drukčije prikazati određene dijelove proteina od interesa, istraživači donose zaključke i kontekstualiziraju ih temeljem kemijske/biološke/fizičke intuicije i drugih primjenjivih znanja. No, činjenica da se vizualnom analizom ne može cjelovito promatrati simulacija već se gleda kao serija trodimenzionalnih kadrova sa sobom nosi dva problema: teško je dobiti „kompletan pregled“ simulacije i teško je pratiti odnose kroz vrijeme (npr. ponašanje određene aminokiseline kroz vrijeme). Oboma problemima donekle se doskače kreativnim vizualizacijskim tehnikama (npr. prikazivanjem položaja određene aminokiseline u svim vremenima obojeno bojevnim skalom koja predstavlja lokaciju u vremenu, od crvene boje koja predstavlja lokaciju na početku simulacije do plave boje koja predstavlja lokaciju na kraju simulacije). No, u takvim slučajevima se pribjegne daljnjim kvantitativnim analizama koristeći nekakvu reprezentaciju kako bi se kvantificirala kvalitativna opservacija. Kao primjer, pomicanje zavojnice u proteinu lako je vidljivo tijekom vizualizacije, a može se kvantificirati mjerenjem vremenske ovisnosti udaljenosti odabrane aminokiseline u zavojnici od neke aminokiseline van zavojnice. Tako se grafički može prikazati i kvantificirati promjena određenog djela strukture proteina, kroz reprezentaciju međusobnih udaljenosti.

Svrha vizualne (kvalitativne) analize je usmjeriti istraživača koje slijedeće kvantitativne analitičke tehnike treba koristiti i na koji način, a svrha rezultata kvalitativne analize (rezultata koji su opservacije prikazane dvodimenzionalno grafički) je na jasan i razumljiv način prezentirati odabrane opservacije. Stoga, jasno da sve rezultate dobivene kvalitativnom vizualnom analizom treba kvantificirati, što se postiže prikazivanjem parametara i reprezentacija (kao što su npr. međusobne udaljenosti aminokiselina) grafovima i/ili tablicama i sl. Rezultati kvalitativne analize komplementiraju rezultate kvantitativne analize: grafike i vizualizacije pružaju kontekst i intuiciju, a grafovi i tablice dokazuju i kvantificiraju opservaciju. Simbioza kvalitativne analize i kvantitativne analize neophodna je za adekvatnu analizu. Komplementarni odnos jasnog i intuitivnog grafa koji kvantificira promjenu koja je čitko i intuitivno vizualizirana grafikom nezamjenjiv je bez ijednog od ta dva neizostavna djela.

Za kvantifikaciju se koriste matematički pristupi nerijetko temeljeni u statistici. Najčešće korištene tri statističke matematičke kvantitativne analize (koje su primjenjive u efektivno svim slučajevima) su: *RMSD* (engl. *Root Mean Square Deviation*), *RMSF* (engl. *Root Mean Square Fluctuation*), i *R<sub>gyr</sub>* (engl. *Radius of Gyration*). Metrik *RMSD*, korijen srednjeg kvadrata udaljenosti, reprezentacija je koja daje informaciju o stabilnosti sustava kroz vrijeme simulacije. Matematički izraz za računanje *RMSD* vrijednosti prikazan je jednadžbom 1. U jednadžbi *N* predstavlja broj čestica po kojima se sumira, to jest broj aminokiselina u slučaju proteina. Nadalje,  $r_i(t)$  predstavlja koordinate C $\alpha$  ugljikovog atoma aminokiseline *i* u vremenu *t*. [2, 15]

$$\rho^{RMSD}(t) = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_i(t) - r_{i,ref})^2} \quad (1)$$

Kako je rečeno, *RMSD* se koristi kao mjera stabilnosti proteina kroz simulaciju, a uspoređivanjem više simulacija vidljivo je koja je stabilnija (pod pretpostavkom da su sustavi usporedivi, npr. simulacije divljeg tipa i rekombinantnog tipa proteina). U načelu, *RMSD* se može računati za proizvoljne atome, ili npr. centre masa aminokiselina, no najčešće je računat upravo za C $\alpha$  ugljikov atom.

Tu vrijedi i napomenuti da je prije vizualne analize potrebno centrirati protein u trajektorijama. Tijekom simulacije, protein se nasumično rotira i translacija (giba) kroz kutiju periodičkih rubnih uvjeta, i promatranje toga bilo bi nemoguće. Zato se protein centrira, kako bi se uklonila translacija i rotacija, i dobio prikaz samo relativnog gibanja unutar proteina. Kao rezultat, protein je stacionarni „vibrirajući“ „objekt“ u centru kutije ispunjene molekulama vode koja se miče oko njega. Centriranje proteina

upravo se postiže minimizacijom *RMSD* vrijednosti, i posljedične promjene koordinata sustava na način da protein uvijek bude fiksiran u centru kutije.

Metrik *RMSF*, korijen prosjeka kvadriranih vrijednosti fluktuacija, mjera je količine fluktuacija individualnih čestica (u kontekstu proteina, najčešće aminokiselinskih ostataka). Izraz kojim se računa prikazan je jednadžbom **2**, tako da  $r_i$  predstavlja trodimenzionalne koordinate čestice  $i$ . Iz grafa *RMSF* vrijednosti dobiva se informacija o fluktuiranju i fleksibilnosti aminokiselinskih ostataka i regija u peptidnom lancu.[2]

$$\rho_i^{RMSF} = \sqrt{\langle (r_i - \langle r_i \rangle)^2 \rangle} \quad (2)$$

Treća najčešće korištena metrika je radijus vrtnje (ili radijus giracije). Računa se kako je prikazano jednadžbom **3**, tako da  $M$  predstavlja sveukupnu masu proteina sastavljenog od  $N$  atoma, s koordinatama centra mase  $R$ . On se također koristi kao mjera stabilnosti proteina tijekom simulacije, a sadrži i informaciju o eventualnim konformacijskim promjenama koje se očituju mijenjanjem kompaktnosti, volumena ili oblika proteina.[2]

$$R_g = \sqrt{\frac{1}{M} \sum_{i=1}^N m_i (r_i - R)^2} \quad (3)$$

Nakon provedene vizualne i osnovne kvantitativne analize (gore opisane metode), opažanja se potvrde i/ili kvantificiraju dodatnim analizama. One su većinom specifične slučaju, i ovise o tome koja se informacija pokušava ekstrahirati i/ili prenijeti. Kao primjer, ako je cilj kvantificirati promjenu u proteinu npr. „otvaranje“ aktivnog mjesta, mjeri se udaljenost određenih aminokiselina kroz vrijeme simulacije. To se zatim prikaže grafički, te se kroz porast međusobne udaljenosti određenih aminokiselina kvantitativno opiše određena promjena konformacije proteina. Razne druge korištene analize mogu prikazivati promjene površinskih naboja, volumen koji određeni dijelovi proteina zauzimaju, i dr.

Toplinske mape (toplinske karte) također se koriste u analizama, no rjeđe. Među najčešćim uporabama ističe se toplinska karta sekundarnih struktura, koja prikazuje koje aminokiseline tvore koje sekundarne strukture u kojim vremenskim periodima. Nadalje, slično se toplinskim kartama prikazuju vodikove veze (koja aminokiselina tvori vodikovu vezu u kojem vremenskom periodu), ionski mostovi, vrijednosti raznih kutova i dr. Nekada je potrebno reprezentirati protein dvodimenzionalno, što se onda

radi s tzv. mapom kontakta. Izračuna se matrica međusobnih udaljenosti svake aminokiseline sa svakom, i prikaže se toplinskom kartom.

I konačno, metrik *RMSD* može se prikazati toplinskom kartom na dva načina. Prvi je engl. *pairwise RMSD* toplinska karta gdje se izračuna *RMSD* vrijednost proteina u svakom kadru u odnosu na svaki ostali kadar. Drugim riječima, napravi se matrica *RMSD* vrijednosti gdje je svaka ćelija *RMSD* vrijednost kadra *i* od kadra *j*, za svaku kombinaciju *i* i *j* kadrova. Sasvim drugi pristup je prikazivanje *RMSD* vrijednosti za svaku individualnu aminokiselinu, u ovisnosti o vremenu. Taj engl. *per-residue RMSD* pristup onda prikazuje *RMSD* vrijednost za svaku aminokiselinu u svakom vremenskom koraku, što predstavlja cjelokupan tijek simulacije prikazan dvodimenzionalno. Zbog otvorene prirode analize simulacija postoji mnogo raznih metoda specifičnih točno određenom slučaju. A postoji i mnogo „rubnih“ metoda koje su (iako vrlo moćne) rjeđe (ili rijetko) korištene, između ostalih konkretno *per-residue RMSD* pristup. Kao primjer, navedene reference koristile su *per-residue RMSD* toplinske karte[16-18]. Njih je moguće generirati koristeći dodatke (engl. *plugins*) programa VMD, konkretno dodatak engl. *RMSD Visualizer Tool*[19], i engl. *Heatmapper*[20]. Kako je proces analize otvorenog tipa, analiza u velikom djelu ovisi o znanjima i sposobnostima istraživača, te alat koji su njemu dostupni u obliku vještina koje ima.

## 2.2. Simulacije korištene za testiranje novo razvijene metode

### 2.2.1. Struktura i simulacija TAL efektor protein

Protein TAL biljni je transkripcijski regulator koji se veže na DNA. Sastavljen je od 1123 aminokiselina koje tvore zavojnice koje se zavijaju u jednu veliku superzavojnicu. Ta superzavojnica se onda omata oko DNA lanca.[21] Autori rada *CATANA: an online modelling environment for proteins and nucleic acid nanostructures*[22] simulirali su dvije varijante sustava TAL-DNA kompleksa. Cilj istraživanja bio je predstaviti novi alat CATANA kroz usporedbu simulacija molekulske dinamike provedenih sa sustavima pripremljenim s CATANA alatom, i sustavima pripremljenim dostupnim metodama. Jedna varijanta sustava bila je s DNA iz kristalne strukture[21], dok je druga varijanta imala DNA lanac sastavljen koristeći CATANA alat. Autori su pripremili sustave za simulaciju kroz više koraka, na standardan način. Sustavi su izgrađeni, i provedena je minimizacija energije, zatim ekvilibracija u *NVT* ansamblu, pa ekvilibracija u *NPT* ansamblu, nakon kojih slijedi produkcijska faza. U svim simulacijama korišten je TIP3P eksplicitni modelom molekula vode.



### 2.2.1. Struktura i simulacije MntR proteina

Transkripcijski faktor MntR je metalosenzor koji u bakterijama *Bacillus subtilis* regulira homeostazu  $Mn^{2+}$  iona.[23] Između ostaloga, negativno regulira ekspresiju proteina odgovornih za unos iona mangana.[24] To je homodimerni protein od 284 aminokiseline, s dva mjesta za vezanje iona mangana, domenom za dimerizaciju podjedinica, i N-terminalnom domenom za vezanje na DNA.[25, 26] Iz literature se zna da aktivacija MntR s ionom mangana(II) nije popraćena velikom konformacijskom promjenom (koja bi se očekivala), s *RMSD* vrijednosti između *apo* i *holo* strukture tek 1,65 Å. [23, 27].

Autori rada[23] korištenog u sklopu ovog diplomskog rada nastojali su dublje razumjeti alosteričke efekte vezanja iona mangana na protein MntR koristeći simulacije MD, kroz promatranje i usporedbu strukturnih i dinamičkih svojstava *apo* i *holo* struktura proteina. Sustavi koji su korišteni u sklopu ovog rada autori su pripremili počevši od dvije kristalne strukture *apo* proteina (PDB kodovi: 2HYF, 2HYG [26]) i dvije kristalne strukture *holo* proteina (PDB kodovi: 2F5F, 2F5C [27]), sveukupno za četiri sustava. Nedostajući aminokiselinski ostatci su dodani, te su strukture protonirane. Budući da ion mangana nije dio standardne aminokiseline, njegove interakcije su parametrizirane koristeći DFT (engl. *Density Functional Theory*) metodu. Za svaku od četiri struktura izgrađen je jedan zasebni sustav. U sklopu istoga korišten je TIP3P model molekula vode, u kubičnoj kutiji periodičnih uvjeta tako da je minimalna udaljenost proteina od ruba kutije iznosila 20 Å, a za postizanje elektroneutralnosti dodani su ioni natrija.[23]

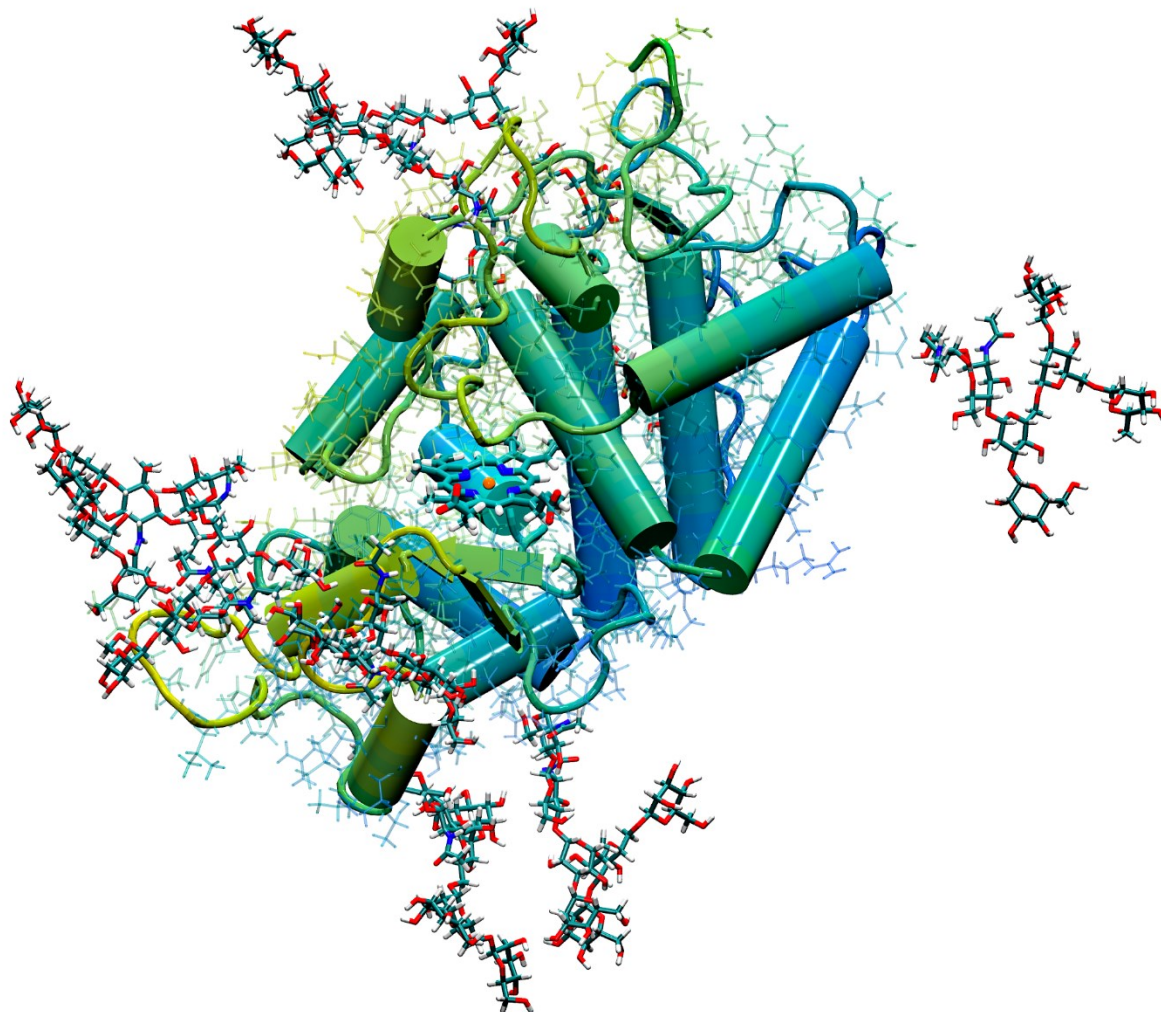
Sustavi su dalje pripremljeni za simulacije molekulske dinamike kroz više koraka. U prvom koraku provedena je minimizacija energije u pet diskretnih ciklusa tijekom kojih su se ograničavala gibanja prvo proteina i iona mangana, zatim proteina, mangana i ostalih teških atoma, i tako do zadnje ciklusa bez ikakvih ograničenja gibanja. Nakon toga slijedile su ekvilibracije pri varijacijama konstantnog tlaka, volumena, i temperature, i na kraju sama produkcijska faza simulacije MD. Sveukupno trajanje simulacija MD iznosilo je 500 ns.[23]

### 2.2.3 Peroksidaza hrena

Peroksidaza hrena (engl. *Horseradish Peroxidase*, HRP) posebno je proučavani metaloenzim podrijetlom iz biljke hren. U biljkama, on provodi reakciju oksidacije elektron donora s vodikova peroksida, pri čemu se molekula peroksida prevodi u dvije molekule vode, uz redukciju dva supstrata oblika AH (gdje H predstavlja proton, a A predstavlja ostatak molekule na kojoj HRP djeluje). Pored

karaktera reakcije koju katalizira, enzim HRP posjeduje biotehnoški atraktivna svojstva koja uključuju: šest lizina dostupnih za reakcije konjugiranja, lako dostupni terminalni krajevi peptidnog lanca, monomernost i sl. Iz tog razloga, HRP svoju primjenu nalazi u sklopu mnogih biotehnoških uporaba. Glavne uporabe koje se mogu istaknuti su kao katalizator u organskoj sintezi, kao biosenzor, i dr.[2, 28]

Sami divlji tip enzima HRP (slika 2) je monomer s 308 aminokiselina, četiri disulfidna mosta, dva kationa kalcija, prostetičkom skupinom hem (feriprotopofirin IX), i devet glikana. Glikani su na protein spojeni N-glikozilacijom preko asparaginskih bočnih ogranaka, a navedeni hem odgovoran je za reakciju koju HRP katalizira.[2, 28]



**Slika 2.** Enzim HRP (struktura PDB oznake 1H5A) [29], prikazan modelom koji ističe sekundarne strukture. Glikani i hem su prikazani štapićastim modelom i obojani po kemijskim elementima. U centru proteina jasno je vidljiv otvor aktivnog mjesta preko kojega reaktanti stupaju u interakciju s hemom, i ion željeza (II) skupine hem prikazan kao narančasta kuglica.

U prirodi, divlji tip HRP enzima je glikozilirana na devet mjesta preko bočnih ogranaka aminokiseline Asn. Uzorak koji slijedi glikozilacija je Asn-X-Thr/Ser, pri čemu X predstavlja bilo koju aminokiselinu (osim prolina ili aspartata).[28, 30] Glikani stabiliziraju enzim, što je pokazano eksperimentalno. No, pokazano je da glikozilacija nije preduvjet za smatanje enzima. Stoga, istraživane su i umjetno neglikozilirane varijante HRP enzima.[2]

Jedna od takvih varijanti pripremljena enzimskim inženjerstvom je HRP mutant koji su pripravili Martell i suradnici[31]. Kako bi pripravili enzim prilagođen uporabi biosenzora za detekciju protein-protein interakcija, koristili su metodu usmjerene evolucije u kvascu. Produkt toga je šesterostruki mutant HRP-a (mHRP). U sklopu mutacija, dolazi do gubitka jednog od glikozilacijskih mjesta, tako da mHRP ima 8 glikana umjesto 9 u divljeg tipa HRP.[2]

U sklopu istraživanja nagrađenog rektorovom nagradom godine 2022. proučavana su strukturalna i dinamička svojstva mHRP enzima.[2] Proučavano je više varijanti enzima uključujući divlji glikozilirani, divlji neglikozilirani, rekombinantni glikozilirani, i rekombinantni neglikozilirani. Nadalje, literaturno je znano da je eksperimentalno utvrđeno da se termička denaturacija enzima HRP odvija kroz dva koraka, preko intermedijernog stanja. [2, 32, 33] Kako bi se računalno istražilo postojanje takvog stanja, u sklopu referenciranog rada provedeno je još šest simulacija pri rastućim temperaturama, ukupno osam simulacija. Sustav HRP proteina pri četiri rastuće temperature, i sustav mHRP proteina pri također četiri rastuće temperature.[2]

Za izgradnju sustava korištena je kristalna struktura HRP enzima izozima C1A iz biljke *Armoracia rusticana*, PDB koda 1H5A[28]. Kao što je navedeno, izgrađene su dvije varijante sustava, divlji tip HRP, i rekombinantni mHRP s uvedenim mutacijama[31]. Sustavi su izgrađeni koristeći CHARMM-GUI mrežni poslužitelj [34-36]. U strukturu su uvedena četiri cisteinska mosta: Cys11-Cys91, Cys44-Cys49, Cys177-Cys209 i Cys97-Cys301; glikani: Man<sub>5</sub>GINAc<sub>2</sub>; nedostajuće aminokiseline C-terminusa: 307 i 308; te u slučaju rekombinantnog mHRP mutacije: Leu299Arg, Asn255Asp, Asn175Ser, Arg93Gly, Pro78Ser, i Thr21Ile.[31] Struktura je protonirana tako da su lizini i arginini pozitivnog naboja, cisteinin neutralni, glutamati i aspartati negativno nabijeni, a histidini protoniranim na delta dušikovom atomu. Kovalentna veza između iona željeza (II) hema i His170 je definirana, te je izgrađena PBC kutija udaljenosti ruba od proteina 20 Å. Sustav je solvatiran TIP3P modelom molekula vode, i neutraliziran ionima klorida. Korišteno je CHARMM36m[8] polje sila.[2]

Izgrađeni sustavi dalje su pripremljeni za simulaciju molekulske dinamike kroz faze minimizacije i ekvibracije. Minimizacija energije je provedena metodom najstrmijeg spusta, s kriterijem konvergencije kao vrijednost energije od 1000 kJ mol<sup>-1</sup> nm<sup>-1</sup>. [15] Kako bi simulacije bile provedene pri rastućim temperaturama: 300 K (26,85 °C), 318 K (44,85 °C), 333 K (59,85 °C) i 353 K (79,85 °C); koje su odabrane prema literaturnom eksperimentalno određenom intervalu u kojemu je opaženo intermedijarno stanje termičkog raspada enzima, sukladno je provedena i ekvibracija. Za svaku od osam simulacija (divlji tip i rekombinantni tip enzima pri četiri naveden temperature) provedeno je 0,6 ns ekvibracije u *NVT* ansamblu. Korišten je „*leap-frog*“ integrator s vremenskim korakom od 1 fs.

Početne brzine atoma dodijeljene su sukladno Boltzmannovoj distribuciji za temperaturu od 10 K, uz otklanjanje translacije centra mase sustava. Prvih 0,5 ns ekvibracije temperatura je kontinuirano dizana od 10 K do ciljane temperature, a zadnjih 0,1 ns temperatura je držana konstantnom, uz regulaciju s Berendsen algoritmom termostata [37]. Nakon toga slijedila je ekvibracija u *NPT* ansamblu, također s „*leap-frog*“ algoritmom integratora, u trajanju od 0,5 ns i vremenskim korakom od 0,1 fs, i algoritmom termostata Berendsen za temperaturno i Berendset izotropski barostat za tlačno sprezanje pri tlaku od 1 bar.[2]

Na to je nastavljena produkcijska faza simulacije molekulske dinamike u trajanju od 500 ns s vremenskim korakom od 2 fs. Korišten je Nose-Hoover algoritam termostata[38], i Parinello-Rahman algoritam barostata[39] izotropski pri tlaku od 1 bar. Trajektorije su ispisivane svakih 10 ps, a dalje u istraživanju su analizirane s koracima od 1 ns. Vremenski korak od 2 fs postignut je uz korištenje LINCS[12] algoritma koji ograničava vibracije veza atoma vodika. Minimizacija, ekvibracija, i produkcijska faza MD simulacije provedena je koristeći GROMACS 2020.6 programski paket.[2,15,40]

## 3. Metode

### 3.1. Konceptualni razvoj mapa trajektorija

Istraživači su trodimenzionalna (3D) bića koja žive u četverodimenzionalnom (4D) vremenu i prostoru, i proučavaju četverodimenzionalni sustav (simulaciju proteina s tri prostorne i jednom vremenskom dimenzijom). Vizualizacijom simulacije dobiva se tek serija trodimenzionalnih kadrova evolucije sustava kroz četvrtu dimenziju, a analizama poput *RMSD*, *RMSF*,  $R_{\text{gyr}}$  i sl. dobije se dvodimenzionalni prikaz evolucije nekog svojstva kroz tu četvrtu dimenziju. Oba pristupa, iako nezamjenjiva, pate od svojih nedostataka. Vizualizacijom kompleksnih sustava od stotina ili tisuća aminokiselina teško je i vremenski zahtjevno konkluzivno izvući bitne uvide te razlučiti biološki ili fizikalno značajne pomake u proteinu od onih koji nemaju taj značaj. To pogotovo vrijedi za rane stadije istraživanja. Nadalje, kompleksnost vizualne analize raste s kompleksnosti simuliranog sustava, duljinom trajanja simulacije i brojem simulacija, te je teško biti siguran da je konkluzivno ekstrahirana svaka značajna informacija i da nije previđena niti jedna. Pored toga, izravno uspoređivanje simulacija vizualnim putem zahtjevno je i nezgrapno, a problemi se još dodatno amplificiraju u komunikaciji rezultata vizualnim putem jer su istraživači zbog korištenja 2D prikaza zakinuti za još jednu dimenziju (sveukupno dvije, jedna vremenska i jedna prostorna). Na drugu ruku, statističke metode ograničene

su dvodimenzionalnim prikazivanjem samo djelića cjelokupnog 4D tijeka simulacije, što ih čini nekonkluzivnim i nejednoznačnim. Kao primjer, skok u *RMSF* grafu može označavati konformacijsku promjenu koja se dogodila jednom i rezultirala stabilnim stanjem (npr. kao da je odvijen jedan prst iz šake prema van), ili povećanje u nestabilnosti koja se očituje pojačanim fluktuacijama (npr. kao da taj prst pojačano fluktuiru u konformaciji šake). Stoga, idealno bi bilo pronaći prikaz koji će na neki način konkluzivno prikazati sve četiri dimenzije, a da je intuitivno za interpretirati dimenzionalno zakinutim bićima poput ljudi. Nažalost, prikazivanje četiri dimenzije nije trivijalno; no postoji rješenje.

Cilj je osmisлити reprezentaciju simulacije proteina koja koristi sve ljudima dostupne dimenzije. Potrebno je 4D  $\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{t}$  prostor prikazati 3D reprezentacijom novih koordinata, proizvoljno uzete:  $\mathbf{t}, \mathbf{a}, \mathbf{p}$ ; koje će sačuvati maksimalno značajnih informacija. Stoga, treba pronaći koje su to nove koordinate  $\mathbf{t}, \mathbf{a}, \mathbf{p}$ , i njihov odnos s izvornim koordinatama  $\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{t}$ . Budući da se prikazuje cjelokupni tijek simulacije, sačuvati će se vremenska os  $\mathbf{t}$ . Sada je potrebno pronaći preostale dvije koordinate. Istraživači su primarno zainteresirani za promjene u strukturi proteina (konformacijske promjene), jer funkcija proteina proizlazi iz njegovog oblika. Protein, kompleksna trodimenzionalna struktura, dugačak je polimerni lanac (ili skup više lanaca) smotan u tri prostorne dimenzije. Ako se aproksimira da se značajne konformacijske promjene očituju u promjeni konformacije okosnice, ta jednodimenzionalna struktura okosnice prirodna je 1D reprezentacija 3D strukture. Svaka aminokiselina prostornih koordinata  $\mathbf{x}, \mathbf{y}, \mathbf{z}$  ekvivalentno je položajno okarakterizirana svojim jednodimenzionalnim rednim brojem (neka bude:  $\mathbf{a}$ , od „aminokiselina“) koji označava njezin položaj u peptidnom lancu.\* Nadalje, promjene u konformaciji očituju se kao različite  $\mathbf{x}, \mathbf{y}, \mathbf{z}$  koordinate atoma za različite vrijednosti vremenske koordinate  $\mathbf{t}$ . Relativni pomak atoma iz referentnog kadra  $\mathbf{t} = \mathbf{t}_{referentan} = \mathbf{t}_r$  u kadar  $\mathbf{t} = \mathbf{t}_{r+1}$  trodimenzionalni je vektor  $(\mathbf{x}_{r+1} - \mathbf{x}_r, \mathbf{y}_{r+1} - \mathbf{y}_r, \mathbf{z}_{r+1} - \mathbf{z}_r)$ . Taj pomak može se reprezentirati jednodimenzionalno preko njegove norme, koja će odgovarati Euklidskoj udaljenosti atoma u poziciji  $\mathbf{t} = \mathbf{t}_{r+1}$  od  $\mathbf{t} = \mathbf{t}_r$  (jednadžba 4). Time je 3D pomak reprezentiran preko svoje 1D magnitude (dalje u tekstu samo „pomak“,  $\mathbf{p}$ ). Ako se iskoristi prethodno iznesena aproksimacija gdje se gleda isključivo okosnicu proteina, može se računati pomak centra mase okosnice svake aminokiseline. Nadalje, definirati će se referentna točka pomaka kao prvi (početni) kadar simulacije  $\mathbf{t} = 0$ , te će se svi pomaci računati u odnosu na  $\mathbf{t}_r = 0$ .\*\* Time se dolazi do konačnog oblika jednadžbe kojom se računa pomak  $\mathbf{p}$  centra mase okosnice aminokiseline  $\mathbf{a}$  iz vremena  $\mathbf{t}_r = 0$  u vrijeme  $\mathbf{t}$  (jednadžba za Euklidsku udaljenost u prostoru s tri dimenzije):

$$p(\mathbf{a}, \mathbf{t}; \mathbf{t}_{ref}) = \sqrt{(x_{\mathbf{a}, \mathbf{t}} - x_{\mathbf{a}, \mathbf{t}_r})^2 + (y_{\mathbf{a}, \mathbf{t}} - y_{\mathbf{a}, \mathbf{t}_r})^2 + (z_{\mathbf{a}, \mathbf{t}} - z_{\mathbf{a}, \mathbf{t}_r})^2} \quad (4)$$

Jednadžbom **4** pomak se izračuna za svaku aminokiselinu u svakom vremenskom koraku. Shodno tome, broj pomaka (koji predstavljaju točke u  $\mathbf{t}$ ,  $\mathbf{a}$ ,  $\mathbf{p}$  prostoru) iznosi broj aminokiselina pomnožen s brojem vremenskih koraka ( $\mathbf{a} \cdot \mathbf{t}$ ). Na taj način, korištenjem svojstva da su proteini uređeni 1D lanci, te prikazivanjem pomaka u 3D prostoru preko njihovih Euklidskih udaljenosti, reprezentiran je 4D prostor  $\mathbf{x}$ ,  $\mathbf{y}$ ,  $\mathbf{z}$ ,  $\mathbf{t}$  koordinata kao 3D prostor aminokiselinskog indeksa  $\mathbf{a}$  i pomaka  $\mathbf{p}$ , kroz vrijeme  $\mathbf{t}$ :  $\mathbf{t}$ ,  $\mathbf{a}$ ,  $\mathbf{p}$ ; jednadžba **5**.

$$\begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix} \rightarrow \begin{bmatrix} t \\ a \\ p \end{bmatrix} \quad (5)$$

Sada je taj novi prostor reprezentacija potrebno vizualizirati. Prikazati će se magnitudu pomaka  $\mathbf{p}$  kao funkciju aminokiselinskog indeksa  $\mathbf{a}$ , i vremena  $\mathbf{t}$ . Prikazivanje  $\mathbf{p}(\mathbf{a}, \mathbf{t})$  grafom rezultiralo bi nezgrapnom 3D površinom, čime ne bi bio postignut cilj intuitivnog reprezentiranja simulacije. Umjesto toga,  $\mathbf{p}(\mathbf{a}, \mathbf{t})$  prikazati će se toplinskom kartom. Os  $\mathbf{x}$  predstavljati će vrijeme  $\mathbf{t}$ , os  $\mathbf{y}$  predstavljati će aminokiselinski indeks  $\mathbf{a}$ , a os  $\mathbf{z}$  predstavljati će magnitudu pomaka  $\mathbf{p}$  vizualizirana skalom u boji. Time se na intuitivan i čitak način prikazuje vrijeme, lokacija i magnituda svakog pomaka centra mase okosnice proteina. Pažljivim odabirom novih koordinata, aproksimacija i načina vizualizacije efektivno se 2D grafikom\*\*\* prikazala vremenska evolucija 3D objekta (što odgovara cjelovitom prikazivanju tijekom 4D simulacije proteina). Uspješno je reprezentiran 4D objekt s 2D objektom, pri čemu su sačuvane najbitnije informacije.

\* Reprezentacijom aminokiseline u prostoru preko njenog rednog broja u potpunosti se gubi informacija o međusobnim prostornim odnosima te aminokiseline s drugima, no to je nešto lako vidljivo vizualnom analizom (jer o vremenu ovisi isključivo parametarski, svaki skup  $(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{t})$  može se promatrati u fiksnom  $\mathbf{t}$ ). Tu je potrebno ponovo naglasiti da mape trajektorija, kao niti jedna druga analiza, nisu osmišljene da zamjene vizualnu analizu, već da ju komplementiraju i nadopune.

\*\* U načelu, bilo koji kadar može biti referentna vrijednost, pa tako i npr. prethodni kadar  $\mathbf{t}_r = \mathbf{t} - 1$  za vremenski korak  $\mathbf{t}$ . O odabiru referentnog kadra ovise pomaci, a samim time i oblik površine koju oni razapinju (stoga parametarska ovisnost pomaka o referentnom kadru u jednadžbi **4**). Ako je referentni pomak prvi, dobiva se „statička“ slika evolucije proteina kroz vrijeme, što je procijenjeno kao najkorisnija varijanta (neovisno o tome što početna konformacija ne mora biti najstabilnija ili optimalna, bitno je da je vidljiva lokacija i magnituda svake promjene). Uzimanjem prethodnog kadra

kao referentne vrijednosti moguće je prikazati „dinamično“ magnitude pomaka, te vidjeti vrijednosti koje odskoču i mogu prikazivati velike konformacijske promjene. No, o odabiru prethodnog kadra ovisi i broj koraka u simulaciji, budući da što je više vremenskih koraka to su respektivni pomaci sličniji (npr. pomak od 1 nm po 1 ns vremenske rezolucije može izgledati kao deset 0,1 nm pomaka u 0,1 ns vremenskoj rezoluciji, što može biti teško razlučiti od šuma nasumičnih fluktuacija). Iz tog razloga, u sklopu ovog rada mape trajektorija isključivo kao referentni kadar koriste prvi kadar simulacije.

\*\*\*Toplinske karte reprezentiraju 3D podatke tako da boja (ili intenzitet, ili svjetlina, ili dr.) stvaraju iluziju koja (vjerno) reprezentira treću dimenziju. No, budući obitavaju u prostorno dvije dimenzije (ekran, papir, ploča...) toplinske karte smatraju se 2D objektima.

Kako bi se dobio relativan pomak bez translacije i rotacije proteina, potrebno ga je računati za centrirani sustav. Kako je već navedeno, centriranje sustava vrši se minimizacijom *RMSD* vrijednosti proteina. Kada bi se *RMSD* vrijednost računala umjesto za cijeli protein za svaki aminokiselinski ostatak individualno (ili npr. za okosnicu svake aminokiseline), u prethodno centriranom sustavu (što i je pravilo), tada taj *per-residue RMSD* upravo odgovara pomaku (kao što je definiran jednadžbom 4). Drugim riječima, *per-residue RMSD* fizikalno predstavlja relativno pomicanje proteina od referentne vrijednosti, koje se u ovom kontekstu u svrhu intuicije naziva pomak. U svrhu intuicije, jednostavnosti, izravne korespondencije s fizikalnim procesom koji se opisuje i konceptualno različitog izvoda koji je vodio do matematički ekvivalentnog (istog) izraza, dalje u tekstu pričat će se isključivo o „pomaku“ kao što je definiran jednadžbom 4; no vrijedi imati na umu da je on matematički ekvivalentan *RMSD* vrijednosti individualne aminokiseline (*per-residue RMSD*).

Pomaci okosnice proteina (relativni samom proteinu) na mapi trajektorija vidljivi su kao horizontalne pruge na odgovarajućem djelu proteinskog lanca i vremena simulacije, s pripadnom magnitudom. Razmatranjem vremena početaka raznih pruga (koje odgovaraju konformacijskim događajima okosnice) može pripomoći interpretaciji *RMSD* grafa na način da se razluči koja individualna promjena uzrokuje koji skok u *RMSD* grafu. Oblik i tip pruge može pripomoći interpretaciji *RMSF* grafa na način da je moguće razlučiti „nestabilne“ pruge koje odgovaraju povećanju fluktuacija, od „stabilnih“ pruga koje svjedoče promjeni konformacije bez značajne promjene u fluktuiranju. Promatranjem vremena početaka pruga, mapama trajektorija moguće je zaključiti o redosljedu konformacijskih događaja. To izravno doprinosi i olakšava karakterizaciji mehanizma iza konformacijskih promjena.[1]



Prikazivanjem prosječnog pomaka regije proteina (od jedne aminokiseline do cijelih podjedinica) kao dvodimenzionalni graf moguće je prikazati promjenu regije kroz vrijeme simulacije. To može koristiti kvantificiranju iste, kao dokaz npr. mehanizma, ili sl. Tako dobiveni graf analogan je gledanju tog djela mape trajektorija kroz  $y$  os toplinske karte, tako da je  $z$  os toplinske karte (inače prikazana bojom na toplinskog karti) postala  $y$  os novog grafa.[1]

Toplinske karte su matrice čije su vrijednosti prikazane bojom. Stoga, moguće ih je zbrajati i oduzimati, te provoditi druge matematičke operacije s njima i njihovim vrijednostima. Više mapa trajektorija može se uprosječiti, kako bi se dobila mapa trajektorija koja prikazuje prosječni tijek npr. triplikata simulacije. Oduzimanjem dvaju mapa trajektorija ističu se sve njihove razlike, te se razlike u konformacijskim događajima proteinske okosnice jasno i čitko vide u obliku pruga. Prikazom takve mape razlika divergentnom bojevnom skalom (npr. crveno-bijelo-plavo) ovisno o boji pruge lako je očitati jeli promjena konformacije izraženija u simulacijama s kojima se oduzima, ili simulacijama od kojih je oduzeto. Drugim riječima, mapom razlika se ističu razlike u promjenama konformacije te njihovo podrijetlo. Kao primjer, ako se uspoređuju dvije varijante simulacija istog sustava, npr. simulacija A i simulacije B. Oduzimanjem mape trajektorija B od mape trajektorija A dobiva se mapa razlike (A-B). Pomaci intenzivniji u simulaciji A biti će pozitivnih vrijednosti, dok će pomaci intenzivniji u B biti negativne vrijednosti. Identični pomaci pokratiti će se do nule. Shodno tome, divergentna mapa boje prikazati će pomake jače u A (pozitivne) kao crvene pruge, dok će pomaci jači u B (koji su negativne vrijednosti) biti plave pruge. Pomaci blizu nule biti će bijeli. Tako se bojom ističe i magnituda razlike, te podrijetlo nje (potječe li pomak iz simulacije A ili simulacije B). Isto vrijedi i za skupove simulacije npr. triplikati, gdje onda A i B predstavljaju mape prosjeka pripadnih multiplikata. [1, 2]

### 3.2. Primjena i validacija metode

Kako bi se analizirale simulacije molekulske dinamike novom metodom mapa trajektorija, Python skripta za provođenje toga je napisana: TrajMap.py (dalje u tekstu, TM.py). Skripta TM.py napisana je u obliku programa za terminal, koji se pokrene i unose se naredbe (slika 3). Alternativno, korištenje Bash skripti koje automatiziraju proces unošenja komandi je preporučen, no dalje o tome nizvodno u tekstu. Skripta TM.py temeljena je na četiri javno dostupne biblioteke za Python: Numpy[41], Pandas[42], Matplotlib[43], i MDTraj[44]. Na Linux sustavima njih je moguće instalirati s *pip* naredbama. Na Windows operativnim sustavima korištenje skripte preporučeno je preko interaktivnog Python editora poput Spyder, od Anaconda distribucije Python programskog jezika. Na Linux operativnim sustavima skripta se može koristiti izravno u terminalu, ili kao što je preporučeno preko Bash skripta. Korištenje preko skripta je preporučeno iz razloga jednostavnosti i brzine, te

minimizacije ljudskih grešaka tijekom unosa. Dodatno, korištenje preko Bash skripti (koje su već napravljene) ne zahtjeva ikakvo poznavanje programskog jezika Python, ili programiranja generalno.[1]

**TrajMap.py**  
Protein MD simulation visualization and analysis

A Python library with premade Bash scripts for easy usage through Linux terminal.

- Visualize the full course of your simulation as a heatmap.
- Plot shifts of a residue or a protein region as a graph.
- Subtract maps to compare two simulations or simulation batches.
- Accelerate and streamline your workflow.

Easy and intuitive usage!

Zero Python knowledge needed! Simply rename Bash variables!

Install dependencies<sup>1</sup> with `pip`, download `TM_kit` from the repository, adjust bash script parameters, and execute!

Comes with comprehensive documentation and an easy-to-follow guide!

<sup>1</sup> numpy, pandas, matplotlib

Fully open source!  
Use TrajMap.py as a Python library for full control.

**TrajMap.py workflow**

**PREPROCESSING**

```

gro → traj2pdb → .pdb → pdb2csv → .csv
xtc

```

**VISUALIZING**

```

matrix_subtract → matrix2map → Difference map
matrix2shifts → shifts2graph → Shift graph
matrix2map → Trajectory map

```

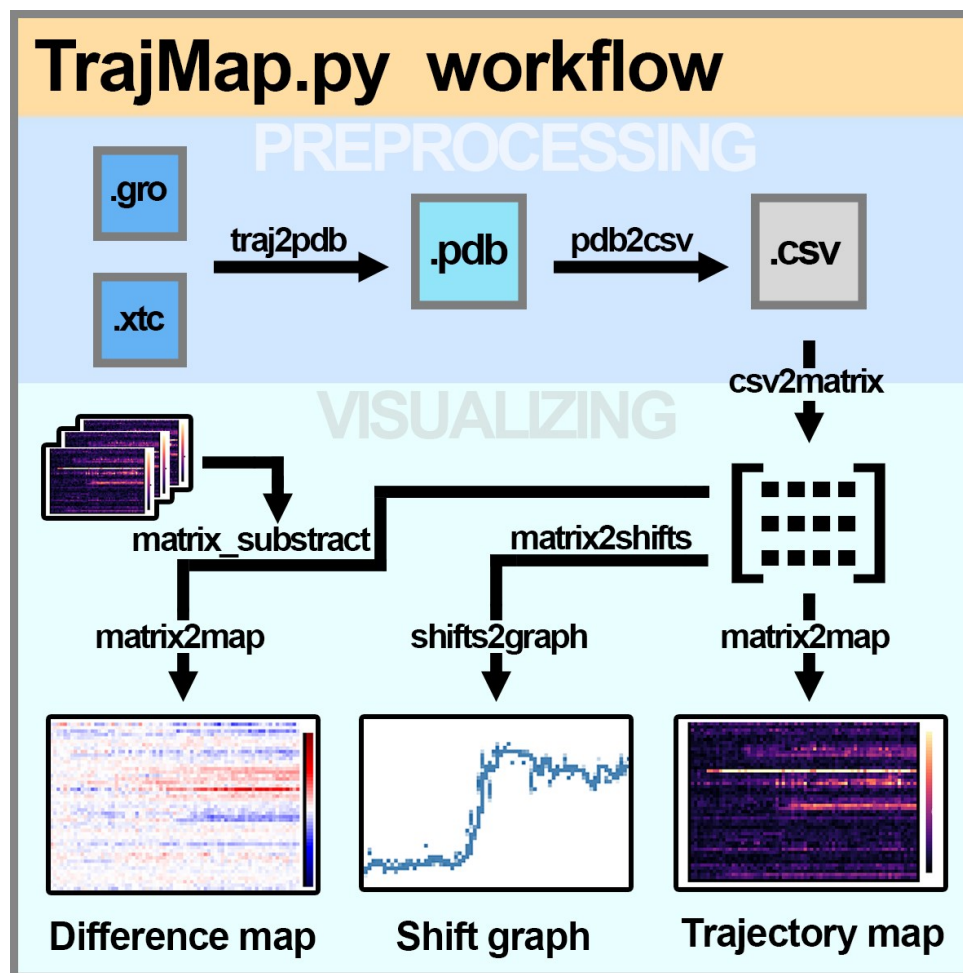
Author: Matej Kožić | Email: [mkozic@chem.pmf.hr](mailto:mkozic@chem.pmf.hr) | Repository: [github.com/matkozic/TrajMap](https://github.com/matkozic/TrajMap) | Release: 2023.02

**Slika 3.** Poster TrajMap.py GitHub repozitorija[45,46], na izvornom engleskom jeziku.[1]

Stavke dostupne s TM.py uključuju: **I)** izradu mape trajektorija simulacije, **II)** izradu grafa pomaka definirane regije simulacije, **III)** računanje prosječne mape trajektorija dvije ili tri simulacije i **IV)** izradu mape razlika simulacija ili skupova simulacija. [1]

U prvom koraku skripte trajektorije i topologije (datoteke oblika `.xtc` + `.gro`, `.xtc` + `.pdb`, `.nc` + `.prmtop`, itd.) se obrađuju na način da se protein centrira, te se eksportiraju trajektorije s korakom namještenim da bude između par stotina do tisuću kadrova. Kao primjer, ako su sirove trajektorije od 500 ns u obliku 5000 koraka, spremanjem svakog desetog kadra dobivaju se trajektorije od 500 ns u 500 koraka, što je veličina prigodnija za mapu trajektorija. U drugom koraku se iz obrađenih trajektorija računaj pomaci centara masa okosnice aminokiselina, i spremaju u matricu koja se sprema u obliku `.csv` (engl. *Comma-separated value*). Nakon toga, `.csv` matrica se učitava i iz nje se generira mapa

trajektorija. Razlog pristupa u više koraka je što je izračun matrice pomaka vremenski najzahtjevniji korak. Okvirno, za 300 aminokiselina u 500 kadrova simulacije na prosječno jakom računalu može se očekivati cca. 5 minuta računanja. Jednom spremljena .csv matrica može se brzo učitati za izradu mape trajektorija, što je proces koji je potrebno ponoviti da bi se ugodili parametri rezolucija osi i sl. Shematska reprezentacija algoritma TM.py vidljiva je na slici 4.[1]



**Slika 4.** Shematski prikaz rada Python programa TM.py, na izvornom engleskom jeziku. U prvom setu koraka radi se preprocesuiranje sirovih trajektorija i topologije do poravnate centrirane trajektorije, iz koje se računa matrica pomaka i sprema u .csv obliku. Zatim se .csv matrica vizualizira u željenom obliku. Na strelicama su prikazana imena komandi iz TM.py koje provode pripadne korake. Detaljan opis dostupan je u dokumentaciji TM.py.[1, 45, 46]

Skripta TM.py dostupna je kao folder TM\_kit. U njemu se nalazi glavna skripta TM.py, prednapravljene Bash skripte (za preprocesuiranje, vizualizaciju, i dr.), i testni folder. U testnom folderu nalaze se male testne trajektorije, i skripta za testiranje svih funkcije kako bi se provjerilo da je

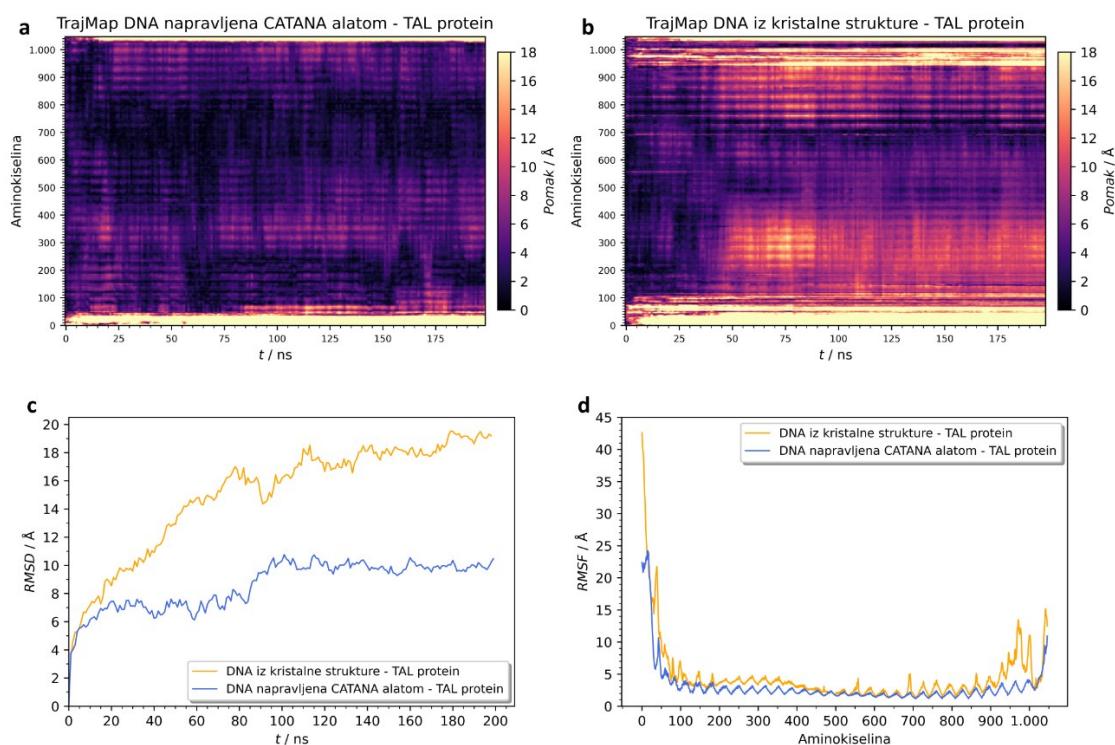
instalacija svih preduvjetnih Python biblioteka uspješno provedena. Dostupna je i dokumentacija s detaljno objašnjenim procesom instalacije preduvjeta, te korištenja same TM.py skripte i Bash skripta. Modificiranjem TM.py skripte moguće je mijenjati sve i prilagoditi skriptu u potpunosti vlastitim potrebama, od veličine osi, fonta, boja, referentnog vremena za izračun pomaka, i dr.[1]

Mape trajektorija skriptom TM.py primijenjene su na sustave simulacija molekulske dinamike proteina iz literature. Rezultati dobiveni mapama trajektorija uspoređeni su s rezultatima dobivenim uobičajenim metodama koje su koristili istraživači u izvornim radovima. Time je korištenje mapa trajektorija prezentirano na praktičnim primjerima, te je pristup validiran usporedbom s postojećim i utemeljenim metodama.

## 4. Rezultati

### 4.1. Analiza simulacija TAL proteina: diferencijacija stabilnosti simuliranih sustava

Simulacija TAL[21] proteinske strukture u kompleksu s DNA dobivena mrežno dostupnim programom CATANA [22] uspoređivana je sa simulacijom TAL proteinske strukture u kompleksu s DNA koja je izgrađena iz kristalne strukture. Za svaku od dvije simulacije napravljena je mapa trajektorija (slike **5a** i **5b**). Autori originalnog rada[22] usporedbu su izvršili *RMSD* i *RMSF* analizama, pa su također i one provedene i prikazane pored mapa trajektorija (slike **5c** i **5d**). Usporedbom mapa trajektorija, prema manjem intenzitetu pruga koje predstavljaju pomake određene regije, može se iščitati kako je TAL protein stabilniji u kompleksu s DNA strukturom dobivenom programom CATANA. Mapa trajektorija simulacije TAL proteina u kompleksu s DNA uzetom iz kristalne strukture pokazuje znatno veću nestabilnost koja se očituje jačim intenzitetom pruga što odgovara daljem pomicanju regije od početne strukture. Ova zapažanja odgovaraju zaključcima dobivenim analizom *RMSD* i *RMSF* grafova. Graf *RMSD* potvrđuju da je struktura pripravljena programom CATANA stabilnija tijekom simulacije, a graf *RMSF* također to potvrđuje i daje informaciju o regijama koje fluktuiraju više.[1]



**Slika 5.** Analiza simulacija MD proteina TAL u kompleksu s DNA mapama trajektorija (paneli **a** i **b**), uspoređena s analizama *RMSF* i *RMSD* istih simulacija (paneli **c** i **d**). Analize kompleksa TAL-DNA napravljenog s nativnom DNA koristeći program CATANA (panel **a**, i plave linija na grafovima) uspoređuje se u odnosu na TAL-DNA kompleks s DNA iz kristalne strukture (panel **b**, i žute linije na grafovima).[1]

Usporedbom *RMSF* grafa i mape trajektorija za simulaciju TAL proteina s kristalnom strukturom DNA, vidljivo je da skokovi u *RMSF* grafu odgovaraju intenzivnim prugama i „mrljama“ na mapi trajektorija ( „mrlje“ na mapi trajektorija mogu se definirati kao određene površine jačeg ili manjeg intenziteta boje). Graf fluktuacija (*RMSF*) daje informaciju o jačoj fluktuaciji regija proteina omeđenih aminokiselinama cca. 200 do 400 i cca. 900 do 1000 (jače fluktuacije u slučaju simulacije s kristalnom strukturom, u odnosu na CATANA strukturu). Iz mapa trajektorija to je također vidljivo, uz dodatne informacije o ponašanju tih regija u vremenu. U simulaciji s kristalnom strukturom DNA, regija cca. 200 do 400 prođe kroz određenu promjenu u konformacije tek u cca. 40-te nanosekunde, što traje do cca. 80-te nanosekunde. Ta velika mrlja praćena je malenim smanjenjem pomaka do pozicije u kojoj regija ostaje do kraja simulacije.[1]

Ove informacije o tijekovima simulacija i karakteru promjena opaženih *RMSF* i *RMSD* grafovima nisu vidljive iz klasičnih analiza (poput navedenih *RMSD* i *RMSF*). Sumarno, mapama trajektorija zaključuje se da je stabilnija simulacija TAL proteina u kompleksu s CATANA strukturom DNA, u odnosu

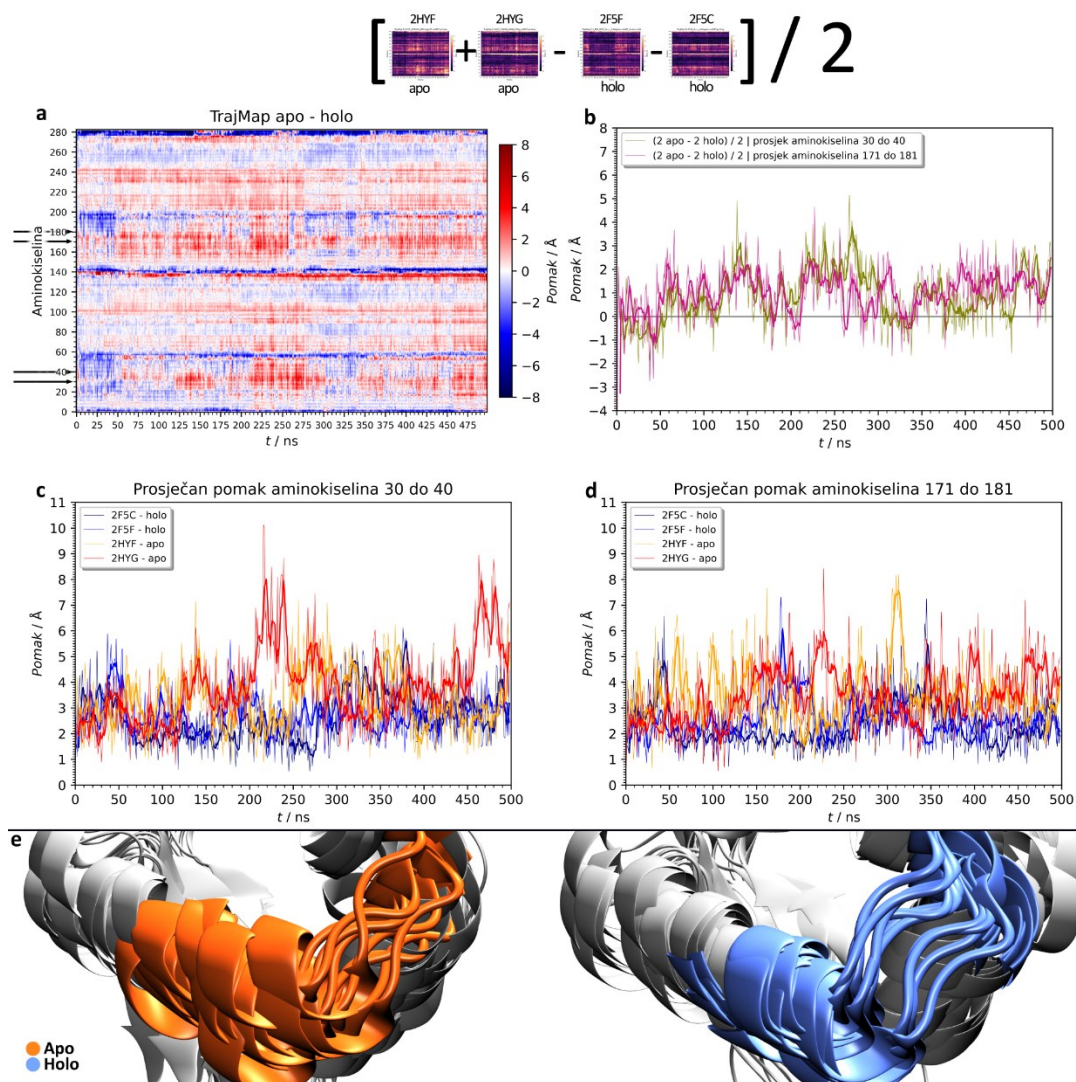
na simulaciju TAL proteina u kompleksu s kristalnom strukturom DNA. Ovaj zaključak u skladu je sa zaključcima autora donesenih na temelju klasičnih analiza poput *RMSD* i *RMSF*. Dodatno, mape trajektorija otkrivaju magnitude i vremena događanja promjena koje se očituju *RMSD* i *RMSF* grafovima, što pomaže u interpretaciji istih. Ovako dobivene informacije o tijeku simulacije mogu uvelike pomoći znanstveniku u daljnjim koracima istraživanja, npr. vizualnoj analizi, jer je sada odmah jasno koje dijelove i u kojim vremenima je potrebno pažljivo pregledati. Mape trajektorija pružaju cjeloviti prikaz tijeka simulacije koji daje širu sliku o usporedbi dvaju simulacija. Stoga, korištenjem mapa trajektorija na ovom primjeru daje iste zaključke kao klasične analize, uz otkrivanje dodatnih informacija i olakšavanja sljedećih analiza ako bi bile potrebne.[1]

#### 4.2. Analiza simulacija MntR proteina: usporedba strukturne dinamike više simulacija

U drugom istraživanju, Z. J. Matošević i suradnici istraživali su ulogu manganovih iona u strukturnoj dinamici MntR homodimernog proteina[23]. Mape trajektorija izrađene su za sveukupno četiri simulacije tog istraživanja: dvije *holo* varijante (u kompleksu s  $Mn^{2+}$  ionima, u izvornom članku strukture PDB kodova 2F5F i 2F5C)[27] i dvije *apo* varijante (bez vezanih  $Mn^{2+}$  iona, u izvornom članku strukture PDB kodova 2HYG i 2HYF)[26]. Da bi se usporedili tijekovi svih simulacija, za početak su napravljene mape prosjeka trajektorija za *holo* i *apo* varijantu. Zatim, mapa prosjeka *holo* varijante je oduzeta od mape prosjeka *apo* varijante. Tako je dobivena mapa razlika trajektorija (slika 6a) koja konkluzivno prikazuje razlike tijekom simulacija *apo* i *holo* sustava. Iz tog prikaza, svakoj pruži može se asignirati podrijetlo, *apo* ili *holo* sustav. Regije gdje su pomaci jednaki rezultiraju razlikom blizu nuli, što odgovara bijeloj boji mape. Plave regije su negativne, što odgovara dijelovima proteina gdje su pomaci bili izraženiji su slučaju *holo* sustava (budući je oduzimano s *holo*; *apo* – *holo*). Crvene, pozitivne, regije odgovaraju dijelovima proteina gdje su pomaci bili veći u *apo* sustavu.[1]

Autori izvorne studije istraživali su regiju proteina koja veže DNA (aminokiseline 30 do 40 i 171 do 181, crne strelice na *y* osi slike 6a) (obje regije odgovaraju istom djelu peptidnog lanca jer se radi o homodimernom proteinu, samo što su dva peptidna lanca numerirana sekvencijalno 1 do 141 i 142 do 282). Prosječna razlika pomaka regija 30 do 40 i 171 do 181 tijekom vremena prikazan je grafom (slika 6b). Prosječni pomaci spomenutih regija individualnih simulacija prikazani su također grafički (slika 6c i 6d).[1]





**Slika 6.** Opsežna analiza sustava od četiri simulacije. Mapama trajektorija set od dvije *apo* simulacije uspoređen je sa setom od dvije *holo* simulacije. Mapa razlike (panel **a**) napravljena je prema shemi prikazanoj na vrhu slike. **a)** Crveni, pozitivni, pikseli predstavljaju lokacije gdje su pomaci bili veće magnitude u *apo* simulacijama. Plave, negativne, regije predstavljaju lokacije gdje su pomaci bili veće magnitude u *holo* simulacijama. Bijele regije su blizu nule i predstavljaju lokacije gdje su pomaci bili vrlo slični i u *apo* i u *holo* simulacijama. **b)** Graf prosječne vrijednosti pomaka mape razlike aminokiselinskih ostataka 30 do 40 (zelena linija) i 171 do 181 (ljubičasta linija). Navedene regije označene su na mapi razlike crnim strelicama na *y* osi. **c)** Graf prosječnog pomaka regija 30 do 40 individualnih simulacije, dvije *apo* (crvena i narančasta linija) i dvije *holo* (plava i modra). **d)** Graf prosječnog pomaka regija 171 do 181 individualnih simulacije, dvije *apo* (crvena i narančasta linija) i dvije *holo* (plava i modra). Na grafovima (**b,c** i **d**) superponiran je pokretni prosjek od 10 kadrova. **e)** Regija 30 do 40 za jednu *apo* (narančasto) i jednu *holo* (plavo) simulaciju prikazana je kao superpozicija kadrova svakih 50 ns kroz 500 ns simulacije.[1]



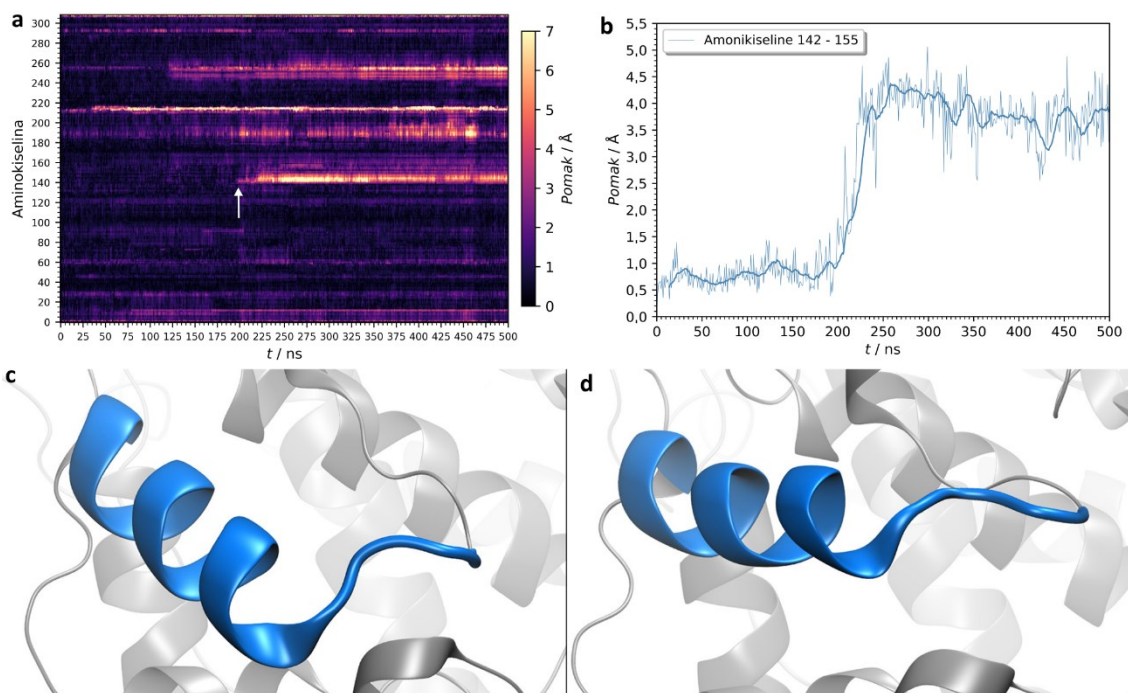
Individualne pomake dvaju homodimernih regija tih četiri simulacije teško je usporediti i interpretirati zasebno, te nije odmah jasno da obje homodimerne regije imaju veće pomake u setu *apo* simulacije. Mapom razlika tih setova od dvije simulacije moguće je svesti osam grafova pomaka dvaju homodimernih regija četiri simulacije (slike **6c**, **6d**) u samo dva grafa razlike pomaka (slika **6b**). Iz grafa pomaka mape razlika jasno je vidljivo da su pomaci dvaju homodimernih regija veći u setu *apo* simulacija. Time je kvantificirana razlika simulacija u tim regijama, i potvrđeno da su regije stabilnije u *holo* simulacijama. Stoga, zaključuje se da su *holo* forme (s vezanim ionom  $Mn^{2+}$ ) stabilnije u navedenim regijama, i imaju manje fluktuacija od *apo* forme. Ovaj rezultat sukladan je zaključku autora originalnog rada[23], te ga dodatno potvrđuje i kvantificira. Vizualno, razlika u stabilnosti je vidljiva, no suptilna (slika **6e**). Vidljivo je da *apo* forma ima „neurednije“ fluktuacije u odnosu na *holo* formu, no to je puno jasnije vidljivo iz grafova razlike (slika **6e**). Autori su do ekvivalentnih zaključaka došli koristeći druge metode, uključujući metodu analize glavnih komponenti i razne vizualizacije. U usporedbi s time, mape trajektorija su dale ekvivalentne rezultate na jednostavniji, te izravniji način, koje je jednostavnije interpretirati. Izravno prikazujući pomake regija od interesa odmah je vidljivo da dolazi do promjene između *apo* i *holo* simulacija, te je ona odmah i kvantificirana.[1]

Nadalje, grafovi pomaka svjedoče i karakteru promjene koja je u pitanju. Iz nestabilnost grafova pomaka vidljivo je da se u *apo* slučaju radi o pojačanju fluktuacija, a ne o „jednokratnoj“ promjeni konformacije. Ovo ne bi bilo zaključivo samo iz *RMSF* grafa ili drugih sličnih najčešće korištenih preliminarnih i bazičnih metoda analize. U slučaju „jednokratne“ promjene koja rezultira promjenom pozicije regije proteina očekivalo bi se dobiti diskretan skok grafa. To će biti pokazano na primjeru u sljedećem potpoglavlju (slika **7b**).[1]

### 4.3. Analiza simulacija peroksidaze hrena: kvantifikacija konformacijskih promjena

Enzim HRP istraživao je simulacijama molekulske dinamike, i u jednoj simulaciji vizualnom analizom je opažena značajna konformacijska promjena pomicanja zavojnice aminokiselinskih ostataka 142 do 155. Kako bi se kvantificirala ta konformacijska promjena napravljena je mapa trajektorija te simulacije (slika **7a**). Iz takvog prikaza spomenuta konformacijska promjena vrlo je očita i lako uočljiva kao intenzivna pruga na odgovarajućoj lokaciji proteinskog lanca (slika **7a**, bijela strelica). Nadalje, za izravnu kvantifikaciju napravljen je graf prosjeka pomaka regija 142 do 155 (koja odgovara spomenutoj zavojnici) (slika **7b**). Jasno je čitljiva magnituda navedene konformacijske promjene, a iz velikog skoka pri cca. 200 ns koji rezultira izmijenjenom konformacijom s usporedivom magnitudom fluktuacija jasan

je i karakter promjene. Drugim riječima, radi se o „jednokratnom“ pomaku zavojnice koji rezultira stabilnom konformacijom, a ne samo o pojačanju (magnituda) fluktuacija. Ovakvu informaciju nije moguće dobiti iz *RMSF* grafa kojim nije razlučivo radi li se o povećanju fluktuacija ili „jednokratnom“ pomaku.[1, 2]

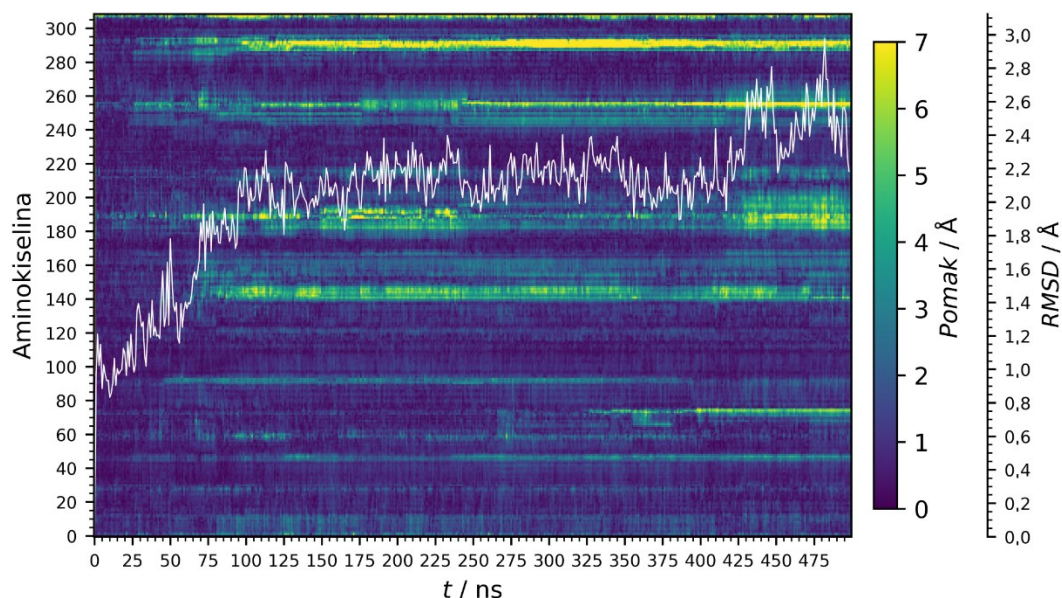


**Slika 7.** Analiza simulacije rekombinante varijante peroksidaze hrena tijekom koje je zapažena značajna promjena konformacije. **a)** Mapa trajektorije simulacije na kojoj je bijelom strelicom označena pruga koja odgovara zavojnici 142-155 koja se pomiče tijekom simulacije. **b)** Graf pomaka zavojnice 142-155, sa superponiranim pokretnim prosjekom od 15 kadrova. Zavojnica 142-155, obojana plavo, prikazana je prije konformacijske promjene u panelu **c** i nakon u panelu **d**. [1,2]

Nadalje, iz mape trajektorija vidljive su pruge u regiji cca. 250, koje su prostorno bliske zavojnici 142-155. U izvornom istraživanju, postojanje tih suptilnih konformacijskih promjena otkriveno je upravo koristeći mape trajektorija, i bilo je ključno za karakterizaciju mehanizma odgovornog za pomicanje navedene zavojnice.[1, 2]

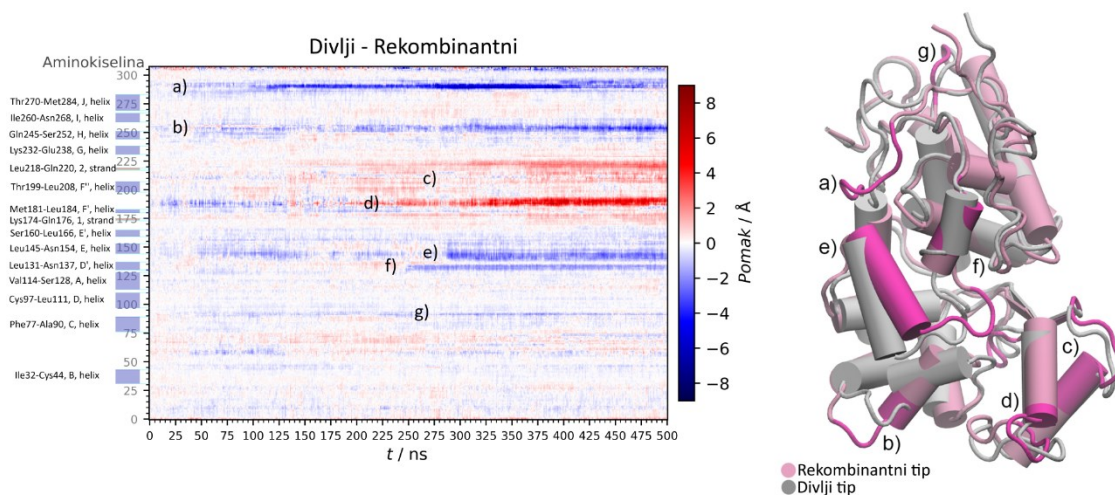
Takva promjena uočena je i u drugoj simulaciji mHRP proteina. Da bi se procijenio utjecaj te konformacijske promjene na čitavu strukturu tijekom simulacije, na mapu trajektorija superponiran je *RMSD* graf (slika **8**). Skok u *RMSD* grafu koji vremenski odgovara početku pruga koja odgovara pomicanju zavojnice 144-155 potvrđuje utjecaj promjene na cjelokupnu stabilnost proteina tijekom simulacije. Ovaj primjer pokazuje da s *RMSD* grafom tako superponiranim na mapu trajektorija moguće

je procijeniti i kvantificirati utjecaj individualnih konformacijskih promjena na sveukupnu stabilnost proteina kroz vrijeme simulacije.[1]



**Slika 8.** Mapa trajektorija mHRP simulacije sa superponiranim *RMSD* grafom ( $C\alpha$  atoma okosnice proteina) prikazanim bijelom bojom.[1]

Kako bi se istražile razlike termičkog raspada divljeg i rekombinantnog tipa HRP enzima, dva seta od četiri simulacije (za četiri različite temperature) analizirano je mapama trajektorija. Za početak je za svaki od dva seta četiri simulacije napravljena prosječna mapa trajektorija (mapa trajektorija u matričnom obliku svake simulacije se zbroji, i podjeli s brojem zbrojenih simulacija). Zatim je mapa trajektorija prosjeka četiri simulacije rekombinantnog mHRP oduzeta od mape trajektorija prosjeka četiri simulacije divljeg HRP, čime je dobivena mapa razlike. Time je sveukupno osam simulacija prikazano jednim prikazom, koji prikazuje razliku mHRP od HRP kontrolirano za rastuće temperature (slika 9).[1, 2]



**Slika 9.** Mapa razlika prosjeka četiri simulacije mHRP i prosjeka četiri simulacije HRP pri rastućim temperaturama. Pozitivne, crvene, regije odgovaraju pomacima većim u simulacijama divljeg tipa. Negativne, plave, regije predstavljaju pomake veće u simulacijama rekombinantnog tipa. Bijele regije odgovaraju lokacijama gdje su se pomaci slični u oba skupa simulacija, tj. nema razlike između simulacija divljeg i rekombinantnog tipa. Na y osi su označene lokacije sekundarnih struktura proteina. Na desnoj strani slike su strukture HRP (sivo) i mHRP (ružičasto) iz jednog kadra pripadnih simulacija, superponirane jedna na drugu. Modelom cilindra reprezentirane su zavojnice. Na mapi razlika slovima od a do g su označene individualne pruge koje odgovaraju razlikama među simulacijama. Te iste lokacije istaknute su na proteinu intenzivno ružičastom bojom, s pripadnim slovom.[1, 2]

Na takvom prikazu sve razlike u setovima simulacija jasno su vidljive, i lako je asignirati individualne događaje tijekom trajektorija. Pruge reprezentiraju razlike u simulacijama, i ovisno o koji pripadni dio proteina se više pomicao ili u setu simulacija s kojim se oduzima, ili u setu simulacija od kojeg je oduzimana. Negativne pruge odgovaraju konformacijskim događajima intenzivnijim u simulacijama rekombinantnog tipa, dok pozitivne odgovaraju konformacijskim događajima intenzivnijim u simulacijama divljeg tipa. Značajne pruge su na slici označene slovima, te dalje istražene individualno. Na ovaj način su se konkluzivno prikazale sve razlike tijekom dva seta od četiri simulacija, za sveukupno osam simulacija prikazanim jednim prikazom.[1,2]

## 5. Rasprava

Mape trajektorija omogućavaju: **I)** elegantno i intuitivno vizualiziranje cijelog tijeka simulacije, **II)** izravno uspoređivanje više (skupina) simulacija, **III)** vizualizaciju ponašanja regije proteina ili jedne aminokiseline tijekom vremena simulacije, **IV)** razlučivanje karaktera konformacijskih događaja okosnice i **V)** komplementaciju postojećih metoda analize simulacija poput *RMSD* i *RMSF* kroz olakšavanje interpretacije istih u kontekstu promatrane simulacije. Elegantan i intuitivan način kojim mape trajektorija cjelovito prikazuju tijek simulacije, uspoređuju bilo koji broj simulacija, i izravno vizualiziraju relativne pomake proteinskih regija od velike su koristi istraživačima koji se bave simulacijama MD proteina, te pogotovo početnicima s manje iskustva.[1]

Iako u literaturi postoje primjeri pristupa usporedivih ovom, konkretno spominjanja i korištenja *per-residue RMSD* toplinskih karti sljedećih referenci [16-18], te dodaci (engl. *plugins*) za program VMD[19,20] koji između ostalog imaju opcije za stvoriti *per-residue RMSD* toplinske karte, autori navode da je sve navedeno u sklopu ovog rada plod originalne ideje razvijene isključivo od strane autora nezavisno za sve navedeno u ovom paragrafu, osim ako specifično nije navedeno suprotno. Prema najboljem saznanju autora, u dostupnoj literaturi nema primjera izvođenja i/ili korištenja ovdje iznesenih koncepata i pristupa na način i/ili do mjere predstavljene ovim radom. Autor vjeruju da je znanje predstavljeno u ovom radu originalan i konkretan doprinos koji se gradi na postojećim konceptima i matematičkim alatima, te ih vodi do rezultata s konkretnom i korisnom interpretacijom .[1]

## 6. Zaključak

Mape trajektorija, nova metoda analize i vizualizacije simulacija MD proteina, detaljno je opisana te je njena primjena prezentirana na tri primjera. Mapama trajektorija moguće je elegantno i intuitivno vizualizirati cijeli tijek simulacije, čime se dobiva konkluzivan i čitak dvodimenzionalni pregled vremenske evolucije proteina tijekom simulacije.

Proces uspoređivanja simulacija je pojednostavljen jer je moguće izravno usporediti tijekove simulacija i istaknuti njihove razlike. Način na koji mape trajektorija barataju s višestrukim varijantama simulacije može znatno olakšati proces analize setova više simulacija, što može imati značajnu primjenu u području analize simulacija MD. Nadalje, mogućnost prikazivanja ponašanja određene

aminokiseline ili određene regije proteina tijekom simulacijskog vremena omogućava komunikaciju rezultata na način koji kvantitativno nadopunjuje često korištene kvalitativne pristupe vizualnog prikazivanja. Na primjerima je pokazano da mape trajektorija komplementiraju postojeće pristupe analize. Nadopunjuju vizualnu analizu na način da omogućavaju prikaz događaja u vremenu i prostoru, čime pomažu identifikaciji biološki ili fizikalno značajnih promjena u strukturi proteina tijekom simulacije. Dodatno, omogućavaju pregled cijele simulacije jednim prikazom što ukazuje na vremenske periode i proteinske regije na koje je potrebno usmjeriti pažnju tijekom vizualne analize simulacije.

Mape trajektorija komplementiraju uobičajenim metodama kvantitativne analize, poput *RMSD* i *RMSF*, tako da pomažu interpretaciji rezultata čime olakšavaju proces njihove analize. Usporedbom *RMSD* grafova s mapom trajektorija moguće je asignirati skokove u *RMSD* grafu s individualnim promjenama u strukturi proteina koje su uzrokovale taj skok. Usporedbom *RMSF* grafa s mapom trajektorija moguće je izvući informacije o karakteru fluktuacije opisane skokom u *RMSF* grafu. Konkretno, može se razlučiti radi li se o „jednokratnoj“ promjeni koja rezultira stabilnim stanjem ili povećanju nestabilnosti.

Sumarno, mape trajektorija su elegantna i intuitivna vizualizacijska metoda koja efikasno i čitko omogućava pregled cijelog tijeka simulacije proteina; olakšavajući i ubrzavajući analizu i komunikaciju rezultata. Kao obećavajuća nova metoda, one nadopunjuju postojeće pristupe i nude nove alate. Time unaprjeđuju proces prikupljanja uvida i komuniciranja rezultata, što ih čini korisnim poljima prirodnih znanosti koja se bave simulacijama MD proteina.

Mape trajektorija dostupne su za korištenje preko pripadnog GitHub repozitorija[1,45,46], kao Python i Bash skripte. Izvorni kod dolazi kao paket koji sadrži sve potrebno za korištenje, dokumentaciju i upute. Stoga, implementacija mape trajektorija u vlastita istraživanja vrlo je jednostavna i brza, te ne zahtjeva dublje poznavanje Python ili Bash programskih jezika.

## 7. Literaturni izvori

- [1] M. Kožić, B. Bertoša, *NAR genom. bioinform.* **6** (2024) 1–8.
- [2] M. Kožić, *In silico* istraživanje utjecaja glikana i mutacija na rekombinantni enzim peroksidaza hrena, rad nagrađen Rektorovom nagradom, Prirodoslovno-Matematički fakultet, Sveučilište u Zagrebu, 2022.
- [3] A. Barišić, Termodinamika procesa na međupovršini inertna tvar/vodena otopina elektrolita, Doktorska disertacija, Prirodoslovno-Matematički Fakultet, Sveučilište u Zagrebu, 2020.
- [4] L. Pipito, R. M. Rujan, C. A. Reynolds, G. Deganutti, *Bioessays* **44** (2022) 1–14.
- [5] M. Rouhani, F. Khodabakhsh, D. Norouzian, R. A. Cohan, V. Valizadeh, *J. Mol. Graph. Model.* **84** (2018) 43–53.
- [6] Y. Zhu, J. Ewalt, H. L. Ng, *Int. J. Mol. Sci.* **20** (2019) 4237.
- [7] J. Wildman, P. Repišćak, M. J. Paterson, I. Galbraith, *J. Chem. Theory Comput.* **12** (2016) 3813–3824.
- [8] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmuller, A. D. MacKerell Jr., *Nature methods* **14** (2017) 71–73
- [9] – W. F. van Gunsteren, X. Daura, A. E. Mark, *GROMOS forcefield, Encyclopedia of Computational Chemistry* (2002).
- [10] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **118** (1996) 11225–11236.
- [11] J. W. Ponder, D. A. Case, *Adv. Prot. Chem.* **66** (2003) 27–85.
- [12] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, *J. Comput. Chem.* **18** (1997) 1463–1472.
- [13] W. Humphrey, A. Dalke, K. Schulten, *J. Molec. Graphics* **14** (1996) 33–38.
- [14] E. F. Pettersen, T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris, T. E. Ferrin, *Protein Sci.* **30** (2020) 70–82.
- [15] P. Bauer, B. Hess, E. Lindahl, (2022) GROMACS 2022.1 Manual.
- [16] S. K. Farney, M. B. Dolinska, Y. V. Sergeev, *J. Anal. Pharm. Res.* **7** (2018) 621–632.
- [17] M. Minguez-Toral, B. Cuevas-Zuviria, M. Garrido-Arandia, L. F. Pacios, *Scientific Reports* **10** (2020) 13424.
- [18] A. Marce-Grau, J. Dalton, J. Lopez-Pison, M. C. Garcia-Jimenez, L. Monge-Galindo, E. Cuenca-Leon, J. Giraldo, A. Macaya, *Orphanet J. Rare Dis.* **11:38** (2016) 1–9.
- [19] A. Nurminen, S. Kukkurainen, L. S. Kaguni, V. P. Hyotnen, *RMSD Visualizer Tool v1.0 VMD extension documentation* (2012) <https://www.ks.uiuc.edu/Research/vmd/plugins/rmsdvt/> (pristupljeno 21.01.2024).
- [20] A. Nurminen, S. Kukkurainen, L. S. Kaguni, V. P. Hyotnen, *Heatmapper v1.0 VMD extension documentation* (2012) <https://www.ks.uiuc.edu/Research/vmd/plugins/heatmapper/> (pristupljeno 21.01.2024).

- [21] A. Nge-Sze Mal, P. Bradley, R. A. Cernadas, A. J. Bogdanove, B. L. Stoddard, *Science* **335** (2012) 716–719.
- [22] D. Kutak, L. Melo, F. Schroeder, Z. Jelić-Matošević, N. Mutter, B. Bertoša, I. Barišić, *Nucleic Acids Research* **50** (2022) 152–158.
- [23] Z. Jelić-Matošević, K. Radman, J. Loubser, I. Crnolatac, I. Piantanida, I. Cukrowski, I. L. Ašler, B. Bertoša, *Int. J. Mol. Sci.* **24** (2023) 1–21.
- [24] X. Huang, J. H. Shin, A. Pinochet-Barros, T. T. Su, J. D. Helmann, *Molecular Microbiology* **103** (2016) 253–268.
- [25] D. A. Capdevila, K. A. Edmonds, D. P. Giedroc, *Essays Biochem.* **9** (2017) 177–200.
- [26] M. A. DeWitt, J. I. Kliegman, J. D. Helmann, R. G. Brennan, D. L. Farrens, A. Glasfeld, *J. Mol. Biol.* **365** (2007) 1257–1265.
- [27] J. I. Kliegman, S. L. Griner, J. D. Helmann, R. G. Brennan, A. Glasfeld, *Biochemistry* **45** (2006) 3493–3505.
- [28] A. M. Azevedo, V. C. Martins, D. M. Prazers, V. Vojinović, J. M. Cabral, L. P. Fonseca, *Biotechnology Annual Review* **9** (2003) 199–247.
- [29] G. I. Berglund, G. H. Carlsson, A. T. Smith, H. Szoke, A. Henriksen, J. Hajdu, *Nature* **417** (2002) 463–468.
- [30] S. Škulj, A. Barišić, N. Mutter, O. Spadiut, I. Barišić, B. Bertoša, *CSBJ* **20** (2022) 3096–3105.
- [31] J. D. Martell, M. Yamagata, T. J. Deerinck, S. Phan, C. G. Kwa, M. H. Ellisman, J. R. Sanes, A. Y. Ting, *Nature Biotechnology* **34** (2016) 774–780.
- [32] K. Chattopadhyay, S. Mazumdar, *Biochemistry* **39** (2000) 263–270.
- [33] K. Bamdad, B. Ranjbar, H. Naderi-Manesh, M. Sadeghi, *EXCLI Journal* **13** (2014) 611–622.
- [34] S. Jo, T. Kim, V. G. Iyer, W. Im, *J. Comput. Chem.* **29** (2008) 1859–1865.
- [35] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, *J. Comput. Chem.* **30** (2009) 1545–1614.
- [36] J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, S. Jo, V. S. Pande, D. A. Case, C. L. Brooks, A. D. Mackerell Jr. J. B. Klauda, W. Im, *J. Chem. Theory Comput.* **12** (2016) 405–413.
- [37] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gusteren, A. DiNola, J. R. Haak, *J. Chem. Phys.* **81** (1984) 3684–3690.
- [38] S. Nose, *Molecular Physics* **52** (1984) 255–268.
- [39] M. Parrinello, A. Rahman, *Journal of Applied Physics* **52** (1981) 7182–7190.
- [40] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. C. Berendsen, *J. Comput. Chem.* **26** (2005) 1701–1718.



- [41] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Rio, M. Wiebe, P. Peterson, P. Garard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, T. E. Oliphant, *Nature* **585** (2020) 475–362.
- [42] W. McKinney, *Proceedings of the 9th Python in Science Conference*, Austin, 2010, 56–61.
- [43] J. D. Hunter, *Comput. Sci. Eng.* **9** (2007) 90–95.
- [44] R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernandez, C. R. Schwantes, L. P. Wang, T. J. Lane, V. S. Pande, *Biophys. J.* **109** (2015) 1528–1532.
- [45] M. Kožić, *TrajMap.py*, <https://github.com/matkozic/TrajMap> (pristupljeno 21.01.2024).
- [46] M. Kožić, *TrajMap.py publication release* (2023), <https://zenodo.org/records/10428489> (pristupljeno 21.01.2024).

## 8. Popis oznaka, kratica i simbola

TM – engl. *Trajectory Map*, Mapa trajektorija.

MM – Molekulska mehanika, računalna metoda molekuskog modeliranja temeljena na klasičnoj mehanici.

MD – Molekulska dinamika, metoda temeljena na molekulskoj mehanici kojom se provode simulacije.

QM – Kvantna mehanika.

QM/MM – Metoda gdje se dio sustava modelira QM pristupom, a ostatak s klasičnim MM pristupom.

DFT – engl. *Density Functional Theory*, Teorija funkcionala gustoće, popularno korištena računalna metoda.

$p$  – Tlak, termodinamički parametar, opservabla.

$V$  – Volumen, termodinamički parametar, opservabla.

$N$  – Brojnost, broj objekata u promatranom sustavu (atom, molekula, u generalnom smislu bilo što, npr. svileni mravojed, komet...), termodinamički parametar, opservabla.

$T$  – Temperatura, termodinamički parametar, opservabla.

$\mu$  – Kemijski potencijal, termodinamički parametar, opservabla.

$NPT$  – Termodinamički ansambl s konstantnim parametrima  $N$ ,  $P$  i  $T$ . Još znan kao izotermno-izobarni ansambl.

$NVT$  – Termodinamički ansambl s konstantnim parametrima  $N$ ,  $V$  i  $T$ . Još znan kao kanonski ansambl.

$\mu VT$  – Termodinamički ansambl s konstantnim parametrima  $\mu$ ,  $V$  i  $T$ . Još znan kao veleanonski ansambl.

VMD – engl. *Visualize Molecular Dynamics*, često korišteni program za vizualizaciju simulacija molekulske dinamike (pogotovo biomakromolekula).

$RMSD$  – engl. *Root Mean Squared Deviation*, mjera koja daje informaciju o stabilnosti sustava kroz vrijeme simulacije.

$RMSF$  – engl. *Root Mean Squared Fluctuation*, mjera je količine fluktuacija individualnih čestica.

$R_{gyr}$  – engl. *Radius of gyration*, radijus vrtnje, statistička metrika (između ostalog i) prosječnog volumena makromolekule.

1D – Jedna dimenzija.

2D – Dvije dimenzije.

3D – Tri dimenzije.

4D – Četiri dimenzije.

$x$  – Jedna od tri prostorne koordinate s arbitrarnom pripadnom jednoslovnim oznakom.

$y$  – Jedna od tri prostorne koordinate s arbitrarnom pripadnom jednoslovnim oznakom.

$z$  – Jedna od tri prostorne koordinate s arbitrarnom pripadnom jednoslovnim oznakom.

$t$  – (U kontekstu reprezentacije simulacija mapama trajektorija) Vremenski indeks simulacije, redni broj kadra.

$t_r$  – (U kontekstu reprezentacije simulacija mapama trajektorija) Referentno vrijeme kadra pomoću kojega se određuju pomaci.

$\alpha$  – (U kontekstu reprezentacije simulacija mapama trajektorija) Aminokiselinski indeks, redni broj aminokiseline u lancu (počevši od N-terminalne aminokiseline, kako je i uobičajeno).

$p$  – (U kontekstu reprezentacije simulacija mapama trajektorija) Pomak, definiran kao Euklidska udaljenost centra mase aminokiseline u vremenu  $t$  od pozicije u vremenu  $t_r$ .

DNA – Deoksiribonukleinska kiselina, molekula bitna za život koja je česta meta proučavanja računalnim metodama.

RNA – Ribonukleinska kiselina, molekula bitna za život koja je česta meta proučavanja računalnim metodama.

PDB – engl. *Protein Data Bank*, mrežna baza podataka kristalnih struktura biomakromolekula.

HRP – engl. *Horseradish Peroxidase*, enzim peroksidaza hrena, dobro istraživani i biotehnoški koristan enzim.

mHRP – Rekombinantna varijanta enzima HRP (posjeduje određene mutacije).

AH – Proizvoljna oznaka za primjer supstrata enzima peroksidaze hrena koji posjeduje proton (označen s H).

TAL – engl. *Transcription activator-like effector*, protein s ulogom efektor/aktivatora.

MntR – Transkripcijski faktor moduliran s  $Mn^{2+}$  ionima, protein bakterije *Bacillus subtilis*.

*holo* – forma proteina s vezanom pomoćnom skupinom (ion, molekula, kofaktor...); *holo* forma

*apo* – forma proteina bez vezane pomoćne skupine (ion, molekula, kofaktor...); *apo* forma

$C_\alpha$  – Alfa ugljikov atom okosnice aminokiseline.

.xtc – Nastavak imena datoteke koja sadrži trajektorije simulacije MD.

.gro – Nastavak imena datoteke koja sadrži topologiju sustava, iz programa GROMACS.

.pdb – Nastavak imena datoteke koja sadrži koordinate strukture sustava (npr. protein, solvatirana molekula...), oblikom sukladan standardu PDB baze podataka.

.nc – Nastavak imena datoteke koja sadrži trajektorije, iz programa AMBER.

.prmtop – Nastavak imena datoteke koja sadrži topologiju sustava, iz programa AMBER.

.csv – engl. *Comma Separated Values*, često korišteni format spremanja podataka, korišten i u programu Microsoft Excel.

dr. – Drugo.

sl. – Slično.

tzv. – Takozvano.

npr. – Na primjer.

*et al.* – latin. *et alia*, i svi ostali. Način navođenja postojanja i referenciranja još autora publikacije, pored glavnog spomenutog.

cca. – latin. *circa*, otprilike.

latin. – Latinski jezik.

engl. – Engleski jezik.

## 9. Izjava nekorištenja alata umjetne inteligencije

Autor izjavljuje da u sklopu pisanja glavnog djela teksta nije niti na ikoji način izravno korišten ikakav alat umjetne inteligencije poput ali ne isključivo *ChatGPT*, *Google Bard* i dr.

## 10. Životopis

### Osobni podatci

**Ime i prezime:** Matej Kožić

**Datum rođenja:** 24.12.1998

**Mjesto rođenja:** Grad Zagreb

### Obrazovanje

- 2005–2013** Osnovna škola Nikole Tesle, Zagreb
- 2013–2017** Prirodoslovna škola Vladimira Preloga, Ekološki tehničar, Zagreb
- 2017–2021** Sveučilišni prijediplomski studij kemija, Prirodoslovno-matematički fakultet Zagreb
- 2021–2024** Sveučilišni diplomski studij kemija, smjer istraživački, studij fizikalne kemije i biokemije, Prirodoslovno-matematički fakultet Zagreb
- 2023** Međunarodna studentska razmjena, CEEPUS ekurzija, Fakultet za kemiju i kemijsku tehnologiju, Ljubljana, Slovenija

### Nagrade i priznanja

- 2017** Nagrada Faust Vrančić za osvojeno prvo mjesto na državnom natjecanju ekoloških tehničara
- 2022** Rektorova nagrada za individualni znanstveni rad „*In silico* istraživanje utjecaja glikana i mutacija na rekombinantni enzim peroksidaza hrena“
- 2023** Nagrada za najbolji *e*-poster, *Computational Chemistry Day 2023*
- 2023** Digitalni inovacijski inkubator, *hackathon*, – osvojeno prvo mjesto u kategoriji
- 2023** *6th CASSINI hackathon* – lokalni pobjednik na području Hrvatske

### Sudjelovanja u popularizaciji znanosti

- 2017** Dan i noć na PMF-u, Otvoreni dani Kemije – volonter
- 2023** Dan i noć na PMF-u, Otvoreni dani Kemije – predavanje Ljepota organske mašinerije: estetika života; poveznica na video snimku probe: <https://www.youtube.com/watch?v=bVwz1QDIbRw> (pristupljeno 2024.01.20)
- 2023** *SFeraKon*, sci-fi konvencija – predavanje *The Art of Organomachinery*

### Sudjelovanja na znanstvenim skupovima

1. Postersko izlaganje, 7. Simpozij studenata kemije, Zagreb, Prirodoslovno-matematički fakultet, Kemijski odsjek, 2022.
2. Postersko izlaganje, *Computational Chemistry Day*, Zagreb, Prirodoslovno-matematički fakultet, Kemijski odsjek, 2023.

### Publikacije

1. M. Kožić, B. Bertoša, *NAR genom. bioinform.* **6** (2024) 1–8, DOI: 10.1093/nargab/lqad114.
2. S. Škulj, M. Kožić, A. Barišić, A. Vega, X. Biarnes, I. Piantanida, I. Barišić, B. Bertoša, *CSBJ* **23** (2024) 742–751, DOI: 10.1016/j.csbj.2024.01.001.