

Sveučilište u Zagrebu  
Prirodoslovno - matematički fakultet  
Fizički odsjek

Molekularno dinamičke (MD) simulacije i  
njihova primjena na kompleks mutanta  
enzima dipeptidil peptidaze III i Keap1  
proteina

Lana Kralj

Zagreb, 2017.

Ovaj rad izrađen je u Laboratoriju za biokemiju proteina i molekulske modeliranje Zavoda za organsku kemiju i biokemiju Instituta „Ruđer Bošković” pod vodstvom prof. dr.sc. Sanje Tomić i predan je na natječaj za dodjelu Rektorove nagrade u akademskoj godini 2016./2017.

## Popis kratica

Ala	Alanin, oznaka još i A
Arg	Arginin, oznaka još i R
Cys	Cistein, oznaka još i C
DPP III	dipeptidil peptidaza III
Glu	Glutamat, oznaka još i E
Gly	Glicin, oznaka još i G
Gln	Glutamin, oznaka još i Q
hDPP III	ljudska dipeptidil peptidaza III
His	Histidin, oznaka još i H
Leu	Leucin, oznaka još i L
R703H	mutant DPP III
Ser	Serin, oznaka još i S
Thr	Treonin, oznaka još i T
wt DPP III	divlji tip proteina DPP III
yDPP III	kvaščeva dipeptilpeptidaza III

# Sadržaj

<b>1</b>	<b>Uvod</b>	<b>1</b>
1.1	Cilj i opseg rada . . . . .	1
<b>2</b>	<b>Molekularno dinamičke (MD) simulacije</b>	<b>3</b>
2.1	Izbor molekularnog sistema . . . . .	3
2.2	Metoda najstrmijeg spusta (Steepest descent method) . . . . .	4
2.3	Polje sila . . . . .	5
2.4	MD simulacije . . . . .	8
2.5	Ograničenja konstantnog tlaka i temperature u sistemu- <i>NVT</i> ansambl i <i>NPT</i> ansambl . . . . .	10
2.5.1	Ograničenja temperature . . . . .	10
2.5.2	Ograničenja tlaka . . . . .	12
2.5.3	Ostala ograničenja . . . . .	13
2.6	Algoritmi za integraciju jednažbi gibanja . . . . .	14
2.7	Rubni uvjeti . . . . .	14
2.8	Točnost i korisnost MD simulacija . . . . .	15
<b>3</b>	<b>Biološki sustav</b>	<b>16</b>
3.1	Dipeptidil peptidaza III . . . . .	16
3.1.1	Peptidaze . . . . .	16
3.1.2	DPP III klasifikacija . . . . .	17
3.1.3	Struktura Dipeptidil peptidaze III . . . . .	19
3.1.4	Mutacije DPP III . . . . .	21
3.2	Kompleksi Nrf2-Keap1 i DPP III-Keap1 . . . . .	23
<b>4</b>	<b>Metodologija</b>	<b>26</b>
4.1	Priprema sustava za provođenje MD simulacija . . . . .	26
4.2	Priprema strukture i solvatacija . . . . .	26
4.3	Minimizacija . . . . .	27
4.4	Ekvilibracija . . . . .	27
4.5	Produktivne MD simulacije . . . . .	27
4.6	Analiza rezultata . . . . .	28
4.6.1	RMSD . . . . .	28
4.6.2	Rgyr . . . . .	28

4.7	Vizualizacija i obrada rezultata . . . . .	29
<b>5</b>	<b>Rezultati i rasprava</b>	<b>30</b>
5.1	Koordinacija cinkovog atoma . . . . .	30
5.2	RMSD . . . . .	33
5.3	Rgyr . . . . .	34
5.4	Udaljenosti pojedinih aminokiselinskih parova $d_1$ i $d_2$ . . . . .	35
5.5	Interakcija DPP-III i Keap proteina . . . . .	38
<b>6</b>	<b>Zaključak</b>	<b>42</b>
<b>7</b>	<b>Zahvale</b>	<b>43</b>
<b>8</b>	<b>Literatura</b>	<b>44</b>
<b>9</b>	<b>Sažetak</b>	<b>47</b>
<b>10</b>	<b>Summary</b>	<b>49</b>

# 1 Uvod

Biofizika je interdisciplinarna znanost koja se bavi direktnom primjenom fizikalnih metoda na biološke sustave. U biofizici se koriste simulacije dinamike molekularnih sistema na računalu, molekularno dinamičke (MD) simulacije pomoću kojih se konstruira model realnog biokemijskog sustava nakon čega se izračunavaju različita korisna svojstva sistema. Teorijski princip na kojem su zasnovane MD simulacije je u potpunosti baziran na fizici. U najkraćim crtama, MD simulacije rade tako da se rješavanjem jednadžbi gibanja za molekularni sustav dobivaju trajektorije svih atoma u sistemu iz čega se dobivaju informacije o dinamici tog sistema koje se mogu usporediti s eksperimentalno izmjerenim svojstvima ili se koristiti za predviđanje ponašanja sistema pri ekstremnim, eksperimentalno neostvarivim uvjetima .

Dipeptidil peptidaza III (DPP III) je protein koji spada u grupu peptidaza, podklasu egzo-peptidaza. DPP III se sastoji od dvije domene, gornje i donje domene međusobno odvojene velikim procjepom koji ukazuje na moguću podložnost ovog proteina na strukturne promjene. DPP III je zabilježen je u različitim zdravim i kancerogenim humanim tkivima. Također, pokazano je da djeluje kao aktivator Nrf2-Keap1 signalizacijskog puta koji djeluje kao mehanizam obrane stanice od kancerogena. Poznato je da se transkripcijski faktor Nrf2 veže za Keap1 protein putem ETGE strukturnog motiva. Međutim, u nekim slučajevima se umjesto Nrf2 faktora za Keap1 protein vežu i drugi proteini koji sadrže ETGE strukturni motiv među kojima je i DPP III. Na taj način sprečava se formacija Nrf2-Keap1 signalizacijskog puta koji postaje nefunkcionalan pa je zbog toga kompleks DPP III-Keap1 zanimljiv za proučavanja. Osim nemutirane DPP III koja se još naziva divlji tip proteina, (eng. wild type), zanimljivo je proučavati i različite mutacije tog proteina i komplekse koje oni tvore.

## 1.1 Cilj i opseg rada

Cilj ovog rada jest opisati fizikalni princip rada molekularno dinamičkih (MD) simulacija. Opisuje se način na koji se u MD simulacijama modelira biološki sustav, preko odabira stupnjeva slobode koji će najrealnije opisati dani biološki sistem do modeliranja i odabira adekvatnog potencijala koje je sastavni dio MD simulacija a čija je uloga najrealniji mogući opis interakcija kojima komponente biološkog sistema međudjeluju. Nadalje, opisuju se algoritmi koji vrše integra-

ciju jednadžbi gibanja za biološke sisteme i način na koji se u sustav uvode različita ograničenja kojima su realni biološki sistemi često izloženi (npr. zadani tlak, zadana temperatura). Također je dan opis kako se u sistem uključuje otapalo, konkretno molekule vode, budući da su realni biološki sustavi često u prirodnom okruženju okruženi vodom.

Nakon izlaganja principa rada MD simulacija, cilj je bio napraviti njihovu primjenu na realnom biološkom sustavu. U ovom radu je po prvi puta provedena MD simulacija na kompleksu R703H-Keap1 a cilj mu je provesti MD simulacije na realnom biološkom sistemu te izložiti kvantitativnu i kvalitativnu analizu rezultata dobivenih direktno iz MD simulacija kako bi se utvrdili molekularni detalji interakcije. Nadalje, u ranijim ispitivanjima grupe pod vodstvom dr. Tomić je MD simulacijama podvrgnut kompleks divljeg tipa DPP III i Keap1 proteina , pa su ti podaci uspoređeni s rezultatima za kompleks mutanta DPP III proteina R703H i Keap1 proteina.

## 2 Molekularno dinamičke (MD) simulacije

### 2.1 Izbor molekularnog sistema

Polazište kod MD simulacija jest izbor adekvatnog molekularnog modela koji će se koristiti za opis realnog sistema i adekvatnog polja sila čija je zadaća opis interakcija između atoma u sistemu. Kod tekućina, makromolekula, otopina itd. suočavamo se sa sistemima s mnogo čestica u kojima nije moguća redukcija problema na samo nekoliko stupnjeva slobode, nego je potrebno uzeti u obzir i analizirati više stupnjeva slobode da bi se adekvatno opisala svojstva tog sistema. Prilikom korištenja određenog modela za predviđanje svojstava molekularnog sistema, potrebno je uzeti u obzir različita pojednostavljenja, aproksimacije i ograničenja klasičnih tehnika kompjuterskih simulacija. Prvo, molekularni sistem je modeliran kao sistem točkastih masa odgovarajućih naboja čije je međudjelovanje opisano odgovarajućom jednadžbom unutar zadanog polja sila. Gibanje tih točkastih masa je vrlo dobro opisano zakonima klasične mehanike što postavlja sljedeća ograničenja kod klasičnog modeliranja: nije moguće adekvatno opisati gibanja molekula na vrlo niskim temperaturama (0-10 K), detaljno gibanje vodikovih atoma (kod kojih su izraženi kvantni efekti) niti kemijske reakcije. Drugo, broj atoma koji mogu biti simulirani na računalu je ograničen. Zbog toga je kod izbora modela iz kojeg će se izračunavati određena svojstva bitno eksplicitno uključiti samo one stupnjeve slobode o kojima svojstva od interesa ovise, a ostale stupnjeve slobode je potrebno izbaciti iz sistema. Na primjer, kod proučavanja smatanja proteina, konfiguracijski prostor koji zauzimaju svi atomi proteina je prevelik. Zato se napravi redukcija veličine konfiguracijskog prostora na način da se reducira broj promatranih stupnjeva slobode tako da se na primjer cijele grupe atoma (aminokiselinski ostaci) reprezentiraju kao jedna ili dvije kuglice (3 ili 6 stupnjeva slobode)<sup>1</sup>. Postoje različiti molekularni modeli kao što je prikazano u Tablici 1. Koji će model biti korišten, odnosno koji stupnjevi slobode će biti uzeti u obzir a koji zanemareni, ovisi o fizikalnim veličinama koje se žele modelirati. Prije izbora modela, potrebno je odrediti svojstva molekularnog sistema koja će biti od interesa, procijeniti veličinu konfiguracijskog prostora, odrediti traženu točnost željenih svojstava i procijeniti dostupno vrijeme izračunavanja. Utvrđi li se nakon toga da je dostupna snaga računala dovoljna za dovoljno točno izvršavanje željene simulacije, moguće je odabrati prikladan molekularni model.



Tablica 1: Zadržani/ izbačeni stupnjevi slobode ovisno o modelu

Model koji se simulira	Čestice čije se interakcije zadržavaju	Čestice čije se interakcije zanemaruju
kvantnomehanički	jezgre, elektroni	nukleoni
svi atomi	dipoli	subatomske čestice
otopljena tvar	atomi otopljene tvari	atomi otapala

## 2.2 Metoda najstrmijeg spusta (Steepest descent method)

U slučaju makromolekule optimizacije se zbog velikog broja stupnjeva slobode provode metodama "prve derivacije", tj. metodama najstrmijeg spusta (steepest descent method) i konjugiranih gradijenata (conjugated gradient). Prije bilo kakvih opsežnijih istraživanja potrebno je geometriju pripremljenog modela optimizirati što se postiže minimizacijom energije danom izrazom (2.4). Metoda najstrmijeg spusta je iterativni algoritam prvog reda a koristi se za pronalaženje lokalnog minimuma funkcije i potencijalne energije. Ilustracija principa rada tog algoritma prikazana je na Slici 1. Princip rada se sastoji od toga da se minimum funkcije pronalazi putem koraka u kojima se pronalazi vrijednost negativnog gradijenta funkcije u danoj točki. U teoriji, jednostavnije bi bilo mijenjati pojedine stupnjeve slobode molekula dok se ne pronade energijski najpovoljnija konfiguracija, međutim postupak je neefikasan te se zbog toga koristi Metoda najstrmijeg pada. Sljedeći primjer opisuje princip rada ovog algoritma. Uzmimo funkciju više varijabli  $f(X)$ . Pretpostavi se da postoji neka početna vrijednost varijabli  $X_0$ . Početne vrijednosti se može mijenjati u mnogo smjerova, ovisno o dimenziji. Na primjer, ako je prostor jednodimenzionalan, postoji samo jedna varijabla,  $x$ , i ta se može povisiti ili sniziti u jednom smjeru). Primjenom opera-

tora gradijenta na funkciju  $f(X)$ ,  $\nabla f(X)$ , dobije se smjer u kojem se  $f$  najviše promijeni, budući da je gradijent derivacija duž svakog smjera  $X$ . Slijedi da će gradijent dati smjer najvećeg porasta funkcije  $f(X)$  a da će se uzimanjem negativne vrijednosti tog gradijenta dobiti minimum funkcije. Dakle, algoritam je baziran na izvrednjavanju relacije (2.1)

$$X_{k+1} = X_k - \lambda \nabla f(X_k) \quad (2.1)$$

gdje je  $\lambda$  mali pozitivan broj koji osigurava stabilnost algoritma jer on određuje veličinu koraka (engl. step size) čime se postiže da algoritam pravi male korake. Pri optimalnim uvjetima (dobro izabran  $\lambda$ ), sigurno vrijedi da je

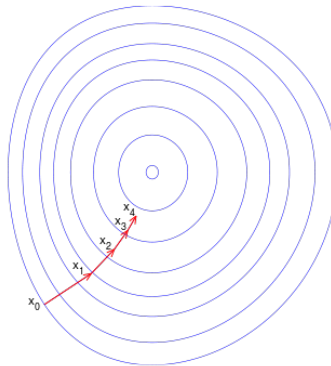
$$f(x_n) > f(x_{n+1}), \quad (2.2)$$

odnosno

$$f(x_1) > f(x_2) > \dots \quad (2.3)$$

što zapravo vodi na konvergenciju niza  $x_n$  ka ciljanom lokalnom minimumu i algoritam se ponavlja sve dok ne dođe do konvergencije niza  $x_n$ .

Točna vrijednost  $\lambda$  ovisi o funkciji koja se izvrednjuje međutim bitno je izabrati dobru vrijednost jer u slučaju premale vrijednosti, funkciji će trebati previše vremena da konvergira a u slučaju prevelike vrijednosti  $\lambda$ , doći će do divergencije.



Slika 1: Ilustracija principa rada Gradient descent metode

### 2.3 Polje sila

Drugi dio je pitanje izbora odgovarajućeg polja sila koje će se za dani model koristiti. Polje sila je zajedničko ime za skup parametara koji se pridružuju

različitim tipovima atoma i funkcije koja korištenjem tih parametara omogućava računanje potencijalne energije promatranog sustava na temelju njegove geometrije. Parametri su dobiveni eksperimentalno i kvantno mehaničkim računima. Ključna pretpostavka za korištenje polja sila je prenosivnost parametara, odnosno pretpostavka da skup parametara izveden na određenom skupu molekula može poslužiti za opisivanje njima sličnih molekula. Tipična funkcija kojom se računa potencijalna energija ima formu (2.4):

$$V = V_1 + V_2 + V_3 + V_4 + V_5 + V_6 + \sum V_7 \quad (2.4)$$

gdje su redom:

$V_1$ - potencijalna energija koja se javlja u molekuli kao posljedica odstupanja duljine veza od ravnotežne vrijednosti definirana kao:

$$V_1 = \frac{k_s}{2} [l - l_0]^2 \quad (2.5)$$

gdje je  $k_s$  empirijski parametar koji predstavlja konstantnu sile istezanja veze,  $l$  je duljina veze, a  $l_0$  je referentna vrijednost za duljinu veze koja je u danom polju sila tabelirana kao optimalna.  $V_1$  ovdje ima formu harmoničkog potencijala koji je dobar za aproksimaciju malih pomaka od ravnotežnog položaja  $l_0$ , dok je za veće odmake od ravnoteže potrebno koristiti Morseov potencijal koji daje bolju aproksimaciju gibanja molekule a oblik mu je tipa  $V_{Morse} \approx [1 - e^{-a(l-l_0)}]^2$  gdje je  $a$  parametar koji opisuje širinu potencijala.

$V_2$ - doprinos potencijalnoj energiji od promjene valentnih kutova definiran kao:

$$V_2 = \frac{k_b}{2} [\theta - \theta_0]^2 \quad (2.6)$$

gdje je  $k_b$  empirijski parametar koji predstavlja konstantu sile promjene kuta,  $\theta$  valentni kut, a  $\theta_0$  referentna vrijednost valentnog kuta.

$V_3$ - doprinos potencijalnoj energiji od promjene torzijskih kuteva definiran kao:

$$V_3 = \sum_n \frac{V_n}{2} (1 + \cos(n\phi - \gamma)) \quad (2.7)$$

gdje je  $V_n$  rotacijska barijera,  $n$  multiplicitet rotacijske barijere,  $\phi$  vrijednost torzijskog kuta, a  $\gamma$  pomak u fazi.

$V_4$ -potencijalna energija koja se javlja kod planarnih sustava uslijed odstupanja atoma iz ravine definirana kao:

$$V_4 = \frac{k_\xi}{2} \xi^2 \quad (2.8)$$

gdje je  $\xi$  kut odstupanja iz ravnine, a  $k_\xi$  pripadni empirijski parametar.

$V_5$ - potencijalna energija elektrostatske interakcije, koja se računa pomoću Kulonovog izraza:

$$V_5 = \frac{q_1 q_2}{4\pi\epsilon r_{12}} \quad (2.9)$$

gdje su  $q_1$  i  $q_2$  naboji atoma, a  $\epsilon$  dielektrična konstanta u mediju. Kako je dielektrična konstanta makroskopsko svojstvo medija, postavlja se pitanje koju vrijednost dielektrične konstante uzimati kada naboji nisu u vakuumu i nalaze se međusobno na dovoljno malenoj udaljenosti da broj molekula otapala među njima nije dovoljno velik da bi koristili makroskopsko svojstvo medija. Zbog toga se uvodi prostorno ovisna dielektrična konstanta dana relacijom:

$$\epsilon = \frac{\epsilon_0 - \epsilon_c}{r_{eff} - r_{con}} + \epsilon_c \quad (2.10)$$

gdje je  $r_{con}$  najmanja udaljenost na kojoj se naboji mogu naći,  $r$  udaljenost naboja,  $\epsilon_c$  vrijednost dielektrične konstante pri najmanjoj udaljenosti na kojoj se naboji mogu naći  $r_c$ ,  $\epsilon_0$  vrijednost dielektrične konstante koja se uzima kada su naboji na dovoljnoj udaljenosti da se može koristiti makroskopsko svojstvo medija.

$V_6$ - potencijal koji potječe od Van der Waalsovih interakcija a računa se pomoću izraza (2.8) :

$$V_6 = \frac{A}{r^{12}} - \frac{C}{r^6} \quad (2.11)$$

gdje su  $A$  i  $C$  empirijski dobiveni parametri ovisni o vrsti atoma između kojih se računa interakcija, a  $r$  je udaljenost među tim atomima.

$\sum V_7$ -miješani članovi koji se javljaju zbog nemogućnosti neovisnog promatranja različitih doprinosa ukupnom potencijalu.

U literaturi postoji mnogo različitih polja sila. Ona se mogu klasificirati prema

sljedećem: a) tipu molekula na kojima će se provesti simulacije, na primjer ugljikohidrati, lipidi, polipeptidi, polinukleotidi,... b) tipu okoliša u kojem se spoj od interesa nalazi, na primjer: plin, vodena otopina, itd., c) rasponu temperatura u kojem se događa efektivna interakcija, d) tipu članova interakcije u polju sila, npr. članovi koji opisuju torziju, rastezanje veza, itd. i e) tipu prilagodbe parametara odnosno na koje se veličine parametri prilagođavaju. Izbor određenog polja sila će ovisiti o svojstvima sistema koja se simulacijama žele ispitati. Također, potrebno je naglasiti da vrijednost potencijalne energije molekule izračunate pomoću polja sila nema direktno fizikalno značenje jer je upotrebom različitih polja sila moguće dobiti različite vrijednosti potencijalne energije iste reprezentacije sistema. Zbog toga računi ostvareni različitim poljima sila nisu međusobno usporedivi. Međutim, vrijednosti energija dobivene za različite reprezentacije sistema korištenjem istog polja sila jesu međusobno usporedive.

## 2.4 MD simulacije

Nakon što je izabran molekularni model i polje sila, sljedeći korak je odabir metode pomoću koje će se što bolje pretražiti ansambl svih mogućih reprezentacija sistema. To se može napraviti različitim metodama, od kojih svaka ima prednosti i mane koje ovise o formi i tipu funkcije energije interakcije  $V(r)$ , broju stupnjeva slobode (veličini sistema), i tipovima stupnjeva slobode (Kartezijeve koordinate, duljine veza, kutevi vezanja itd.). Postoje metode pretraga koje skeniraju ukupan konfiguracijski prostor molekularnog sistema i metode čiji je cilj generiranje reprezentativnog ansambla konfiguracija. One se dijele na dva tipa: Non-step metode koje generiraju nepovezanu seriju slučajnih konfiguracija i Step metode koje generiraju novu konfiguraciju iz prethodne. MD simulacije koje su temelj ovog rada spadaju u Step metode. Tijekom MD simulacija generira se ansambl konfiguracija molekularnog sistema, tzv. trajektorija. Pri tome se gibanje čestice (atoma) u molekularnom sustavu opisuju nizom jednadžbi gibanja (Newtonovih ili Langevinovih). Newtonova jednadžbe gibanja za  $i$ -ti atom u sistemu dana je izrazom (2.12) i (2.13):

$$\frac{d^2 r_i(t)}{dt^2} = m_i^{-1} F_i \quad (2.12)$$

$$F_i = -\frac{\partial V(r_i, \dots, r_N)}{\partial r_i} \quad (2.13)$$

gdje je  $F_i$  sila na  $i$ -ti atom, a  $t$  je vrijeme. Budući da se računa gradijent potencijalne energije, zahtjev je da funkcija  $V(r)$  bude diferencijabilna funkcija koordinata atoma  $r_i$ . Integracija jednadžbe (2.12) se provodi u malim vremenskim intervalima (time steps)  $\Delta t$ , tipično 1-10 fs za molekularne sisteme. Newtonove jednadžbe gibanja (2.12), diferencijalne jednadžbe drugog reda, mogu se zapisati u obliku kao dvije diferencijalne jednadžbe prvog reda (2.14) i (2.15):

$$\frac{dv_i(t)}{dt} = m_i^{-1} F_i(r_i(t)) \quad (2.14)$$

$$\frac{dr_i(t)}{dt} = v_i(t) \quad (2.15)$$

gdje su sa  $r_i(t)$  označeni položaji  $i$ -te čestice u ovisnosti o vremenu  $t$ , a sa  $v_i(t)$  brzine  $i$ -te čestice u ovisnosti o vremenu  $t$ . Sila koja djeluje na  $i$ -tu česticu  $F_i$  dobivena je kao negativan gradijent funkcije potencijalne energije, u skladu s relacijom (2.13). Da bi se dobila rješenja jednadžbi (2.14) i (2.15), potrebno ih je numerički integrirati. U nastavku je opisan takozvani leap frog algoritam za integraciju jednadžbi (2.14) i (2.15) koji se provodi u malim vremenskim intervalima  $\Delta t$ . Leap frog algoritam je jedan od najtočnijih i najstabilnijih algoritama za proučavanje molekularne dinamike sistema od našeg interesa. Polazišne jednadžbe leap-frog sheme su zapravo kombinacija Taylorovih razvoji funkcija  $v_i(t)$  i  $r_i(t)$  do drugog reda u vremenskom trenutku  $t = t_n$  danih jednadžbama (2.16) i (2.17)

$$v_i(t_n + \Delta t/2) = v_i(t_n) + [dv_i/dt]_{t_n} \Delta t/2 + [d^2 v_i(t)/dt^2]_{t_n} (\Delta t/2)^2/2! + O(\Delta t^3) \quad (2.16)$$

$$v_i(t_n - \Delta t/2) = v_i(t_n) - [dv_i/dt]_{t_n} \Delta t/2 + [d^2 v_i(t)/dt^2]_{t_n} (\Delta t/2)^2/2! + O(\Delta t^3) \quad (2.17)$$

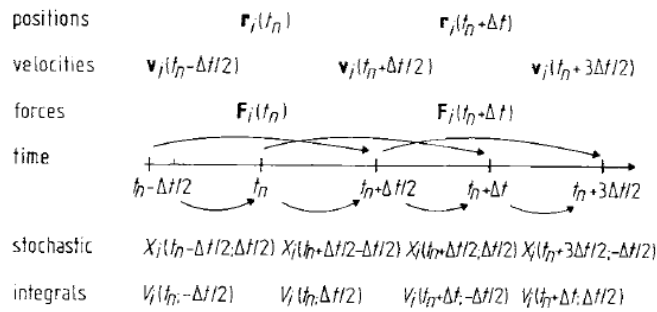
i jednadžbe (2.15) što daje jednadžbe (2.18) i (2.19) koje su baza leap frog sheme:

$$v_i(t_n + \Delta t/2) = v_i(t_n - \Delta t/2) + m_i^{-1} F_i(r_i(t_n)) \Delta t + O(\Delta t^3) \quad (2.18)$$

$$r_i(t_n + \Delta t) = r_i(t_n) + v_i(t_n + \Delta t/2) \Delta t + O(\Delta t^3). \quad (2.19)$$

Način na koji radi leap frog shema prikazan je na Slici 2. Ukratko, radi se o tome da se kontinuirane varijable (kao što su brzina i položaj) aproksimiraju njihovim vrijednostima u konačnom, diskretnom nizu vremenskih točaka  $[t_n -$

$\Delta t/2, t_n, t_n + \Delta t/2, t_n + \Delta t, \dots]$ . Dakle, algoritmi Step metoda rade tako da aproksimiraju kontinuiranu vremensku evoluciju varijabli od interesa diskretnim promjenama tih varijabli od jedne vremenske točke do druge. Specijalno, kod leap frog algoritma, položaji se određuju u točkama  $[t_n, t_n + \Delta t, t_n + 2\Delta t, \dots]$ , a brzine u točkama  $[t - \frac{\Delta t}{2}, t + \frac{\Delta t}{2}, t + \frac{3\Delta t}{2}, \dots]$ , odnosno o vremenu ovisni vektori položaja i brzina „preskaču” jedan drugog, od čega i potječe ime metode.



Slika 2: Leap Frog SHEMA. Slika je preuzeta iz Literature<sup>1</sup>.

## 2.5 Ograničenja konstantnog tlaka i temperature u sistemu- *NVT* ansambl i *NPT* ansambl

Integracijom Newtonovih jednadžbi gibanja za klasičan sistem s brojem čestica  $N$  i volumenom  $V$ , ukupna energija simuliranog sistema  $E$  i ukupan linearni moment  $p$  ostaju očuvani, pri čemu će simulacije generirati takozvani mikrokanonski ansambl u kojem su broj čestica  $N$ , volumen  $V$  i ukupna energija  $E$  konstantni što je neprikladno budući da su u većini realnih eksperimenata konstantne veličine temperatura  $T$  i tlak  $p$ . Zbog toga se javlja potreba za postavljanjem temperature i tlaka kao nezavisnih varijabli, što se može napraviti uvođenjem termostata i barostata čime se više neće generirati nepoželjni mikrokanonski ansambl već ili ansambl s konstantnom temperaturom (*NVT* ansambl u kojem je konstantan broj čestica  $N$ , volumen sistema  $V$  i temperatura sistema  $T$ ) ili ansambl s konstantnim tlakom i temperaturom (*NPT* ansambl).

### 2.5.1 Ograničenja temperature

Kontroliranje temperature u molekularnoj dinamici tako da ona bude konstantna moguće je izvršiti različitim metodama. Jedna od njih je metoda u kojoj se os-

novnim atomskim stupnjevima slobode sistema koji se simulira dodaje dodatan stupanj slobode  $s$  koji predstavlja termostat na način da se ukupnom Hamiltonijanu sistema umjetno dodaju članovi kinetičke i potencijalne energije tog dodatnog stupnja slobode, te se na taj način u sustav uključuje termostat. Član koji opisuje kinetičku energiju koja potječe od tog dodatnog stupnja slobode  $s$ , osigurava toplinski tok iz termostata u sistem i obrnuto.

Drugi način osiguravanja konstantne temperature u MD simulacijama jest metoda slabog sprezanja gdje se željena temperatura sustava održava "uranjanjem" sustava u vanjski termostat neke konstantne temperature  $T_0$ . Temperatura je direktno povezana s ukupnom kinetičkom energijom sustava preko jednadžbe (2.20) iz koje se vidi da je kinetička energija proporcionalna brzini:

$$T(t) = \frac{1}{(3N - n)k_B} \sum_i \frac{|p_i(t)|^2}{m_i} \quad (2.20)$$

gdje je  $N$  ukupan broj atoma,  $n$  broj stupnjeva slobode koji se zanemaruju,  $p_i = m_i v_i$  linearni impuls  $i$ -tog atoma,  $k_B$  Boltzmannova konstanta iznosa  $k_B = 1.38065 \cdot 10^{-23} \text{ J/K}$ . Zbog toga bi najjednostavniji način kontrole temperature bio skaliranje brzina pojedinih čestica pri svakom koraku integracije. Međutim, skaliranje se ne može provesti naglo zbog prirodnih fluktuacija temperature, već ga je potrebno provesti postupno na način da se u ukupnu jednadžbu gibanja doda dodatni član koji se zove član konstante sprege  $\tau_t$  a određuje koliko su sustav koji se simulira i termostat međusobno povezani. Veći  $\tau_t$  znači da su sustav i termostat slabije povezani, odnosno potrebno je duže vrijeme za postizanje željene temperature  $T_0$ . Faktor sprezanja je potrebno odabrati tako da on bude dovoljno malen da se izbjegn timermetnje sistema, a dovoljno velik da se postignu zadovoljavajući rezultati.

$$m_i \frac{dv_i}{dt} = F_i + \frac{p_i(t)}{\tau_t} \left[ \frac{T_0}{T(t)} - 1 \right] \quad (2.21)$$

Brzine se tada skaliraju nakon svakog koraka integracije tako da je brzina promjene temperatura proporcionalna razlici temperatura sukladno relaciji (2.22):

$$\frac{dT(t)}{dt} = \tau_t^{-1} [T_0 - T(t)] \quad (2.22)$$

iz koje direktno slijedi da će se nakon svakog koraka integracije u trajanju  $\Delta t$  temperatura promijeniti za točno



$$\Delta T = \frac{\Delta t}{\tau_t} [T_0 - T(t)] \quad (2.23)$$

Skalirajući faktor za brzine  $\lambda$  koji opisuje tok topline u sustav ili iz sustava dan je relacijom (2.24):

$$\lambda = \left[ 1 + \frac{\Delta t}{\tau_t} \left( \frac{T_0}{T\left(t - \frac{\tau_t}{2}\right)} - 1 \right) \right]^{\frac{1}{2}} \quad (2.24)$$

U molekularnoj dinamici se koriste tri vrste termostata: Berendsenov, Langevinov i Andersonov. Berendsenov i Langevinov termostat se uključuju u svakom vremenskom koraku, dok se Andersonov termostat uključuje u simulacije svakih nekoliko stotina koraka<sup>2</sup>. U simulacijama provedenim na projektu je korišten Langevinov termostat, te je stoga potrebno opisati princip na kojem on djeluje. Langevinova jednadžba gibanja (2.25) opisuje gibanje sistema na način da uključuje dodatne članove koji opisuju silu trenja proporcionalnu koeficijentu trenja  $i$ -te čestice  $\gamma_i$  i nasumičnu, stohastičku silu  $R_i$  koja opisuje slučajne sudare između čestica sistema:

$$\frac{d^2 r_i(t)}{dt^2} = m_i^{-1} F_i + m_i^{-1} R_i - \gamma_i dr_i(t)/dt \quad (2.25)$$

Kako je koeficijent trenja  $\gamma_i$  pozitivan broj, sila trenja na sistem utječe tako da on gubi kinetičku energiju, odnosno utječe na smanjenje temperature sustava. S druge strane, stohastička sila je direktna posljedica sudara čestica i ona doprinosi povećanju kinetičke energije u sustavu. Dakle, nasumični sudari u kombinaciji sa silom trenja održavaju temperaturu sustava na nekoj zadanoj vrijednosti referentne temperature  $T_0$ . Slijedi dakle iz Langevinove jednadžbe gibanja da Langevinov termostat mijenja jednadžbu gibanja u vremenskim intervalima  $\Delta t$  tako da se impulsi  $p_i$  čestica nakon svakog koraka integracije mijenjaju za faktor (2.26)<sup>3</sup>:

$$\Delta p_i = \left( \frac{\partial U(r_i)}{\partial r_i} - \gamma p_i + R_i \right) \Delta t \quad (2.26)$$

## 2.5.2 Ograničenja tlaka

S druge strane, često je poželjno provesti simulacije uz uvjet konstantnog tlaka. Za izotropni sistem volumena  $V$ , tlak je skalar definiran jednadžbom (2.27)

$$P(t) = \frac{2}{3V} [E_{kin} - e(t)] \quad (2.27)$$

gdje je  $e$  virijal definiran kao (2.28):

$$e(t) = -\frac{1}{2} \sum_{\alpha < \beta} (r_{\alpha}(t) - r_{\beta}(t)) F_{\alpha\beta}(t) \quad (2.28)$$

gdje je  $r_{\alpha}(t) - r_{\beta}(t)$  relativni položaj vektora centra mase molekula  $\alpha$  i  $\beta$  a  $F_{\alpha\beta}$  sila kojom atom  $\alpha$  djeluje na atom  $\beta$ . Metode za izvršavanje MD simulacija pri konstantnom tlaku slične su metodama za izvršavanje MD simulacija pri konstantnoj temperaturi. Konkretno, radi se o uvođenju barostata s referentnim tlakom  $P_0$  u sustav čime se osigurava konstantan tlak u sustavu. Kao i kod regulacije temperature, pri regulaciji tlaka koristi se metoda slabog sprežanja, s razlikom da se u slučaju regulacije tlaka skaliraju položaji atoma umjesto brzina atoma budući da je tlak proporcionalan razlici položaja dvaju atoma što se vidi iz relacija (2.27) i (2.28). Povezivanje sistema koji prolazi simulacije s barostatom je ponovno opisano konstantom sprežanja  $\tau_p$ . Ponovno vrijedi što je veći  $\tau_p$ , sprežanje je slabije. Brzina promjene tlaka nakon svakog koraka integracije je proporcionalna razlici tlakova što opisuje relacija (2.29):

$$\frac{dP(t)}{dt} = \tau_p^{-1} [P_0 - P(t)] \quad (2.29)$$

odnosno tlak se nakon svakog integracijskog koraka u trajanju  $\Delta t$  promijeni za faktor  $\Delta P$  dan s (2.30)

$$\Delta P = \frac{\Delta t}{\tau_p} [P_0 - P(t)]. \quad (2.30)$$

### 2.5.3 Ostala ograničenja

Osim postavljanja ograničenja fiksnog tlaka ili temperature u sustav, poželjno je ograničiti i druge stupnjeve slobode, na primjer, duljine molekularnih veza ili kuteva između veza s ciljem uštede vremena koje je računalu potrebno za provođenje simulacija. Primjenom ograničenja moguće je smanjiti vrijeme izračuna. Kod MD simulacija poželjno je postići čim dulji vremenski korak  $\Delta t$  a on je ograničen maksimalnom frekvencijom gibanja  $v_{max}$  koja je prisutna u sistemu. Dakle, uklanjanjem stupnja slobode s najvećom frekvencijom, moguće je produljiti vremenski korak  $\Delta t$ . Primjena ograničenja u sistemu je korisna u slučajevima kad: "zamrznuti" stupnjevi slobode ne utječu značajno na svojstva sustava koji se promatra, kad su frekvencije "zamrznutih" stupnjeva slobode

puno veće od ostalih frekvencija u sistemu tako da se postigne značajno produljenje koraka  $\Delta t$  i kad algoritmi preko kojih se ograničenja uvode u sustav nisu prekomplikirani za danu snagu računala.

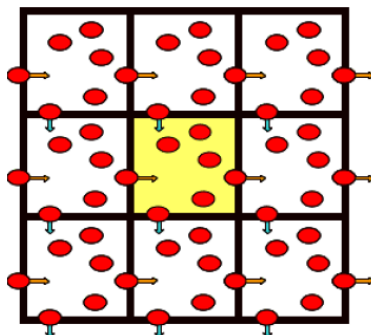
## 2.6 Algoritmi za integraciju jednadžbi gibanja

Postoji više pristupa za integraciju jednadžbi gibanja molekularnog sistema s uključenim ograničenjima a mogu se svrstati u pristup koji uključuje formulaciju preko generaliziranih koordinata ili formulaciju preko Kartezijevih koordinata od kojih je svaka pogodna za određene molekule.

U slučaju fleksibilnijih sistema, pristup preko generaliziranih koordinata ne daje dobre rezultate te se za to koriste Kartezijeve koordinate. Integracija jednadžbi gibanja provodi se korištenjem Kartezijevih koordinata. Najčešće metode za integraciju takvih jednadžbi gibanja su matrice metode od kojih je vrlo rasprostranjena SHAKE metoda<sup>1</sup> pogodna za tretiranje makromolekula.

## 2.7 Rubni uvjeti

Kod simulacija realnih sistema konačne veličine, potrebno je u razmatranja uključiti molekule otapala kojima je sistem prirodno okružen jer se tako dobiva vjerodostojniji model. Jedan od načina na koji se može riješiti problem uključivanja otapala u sustav je tzv. tehnika periodičkih rubnih uvjeta koja je prikazana na Slici 3. Cjelokupni sistem koji će biti simuliran stavi se u na primjer kubičnu kutiju koja predstavlja jediničnu ćeliju a tretira se kao da je okružena sa svake strane sa svojim identično translatiranim kopijama. Korištenjem periodičkih rubnih uvjeta, u osnovnoj simulacijskoj kutiji očuvana je masa, broj čestica, ukupna energija i linearni moment. Korist ovakvog pristupa je u tome što se minimiziraju rubni efekti koji u nekim slučajevima mogu bitno utjecati na sustav budući da svaka molekula koja tijekom simulacije izađe iz središnje jedinične ćelije, ulazi u neku od svojih kopija, odnosno zamjenjuje ju njezina periodička slika čime se zapravo rješava problem rubnih uvjeta jer se postiže zatvorenost sustava jer sustav "ne vidi" vakuum unutar udaljenosti do koje se računaju interakcije od interesa. Zbog toga je bitno da periodička kutija bude dovoljno velika, odnosno da strukturu koja se simulira okružuje dovoljno molekula vode, tako da molekula ne može interagirati sama sa svojim periodičkim slikama.



Slika 3: Periodički rubni uvjeti

## 2.8 Točnost i korisnost MD simulacija

Logički se nameće pitanje točnosti i korisnosti MD simulacija. Točnost predikcija dobivenih putem simulacija će ovisiti uglavnom o dva faktora. Prvi je točnost izbora molekularnog modela i polja sila, a drugi je koliko je konfiguracijski prostor dobro definiran. Najbolji pokazatelj točnosti predikcija dobivenih simulacijama je zapravo direktna usporedba predikcija s eksperimentalno izmjerenim rezultatima i zbog toga bi se rezultati dobiveni simulacijama trebali usporediti s eksperimentalno određenim rezultatima kad god je to moguće. Štoviše, direktna eksperimentalna mjerenja se i dalje preferiraju u odnosu na rezultate simulacija. Dakle, cilj računalnih simulacija zapravo nije u tome da one u potpunosti zamijene eksperimentalna mjerenja već u njihovoj mogućnosti da nadopune eksperiment i da daju predikcije mjerenja u uvjetima tlaka i temperature u kojima nije moguće direktno obaviti eksperimentalna mjerenja.

### 3 Biološki sustav

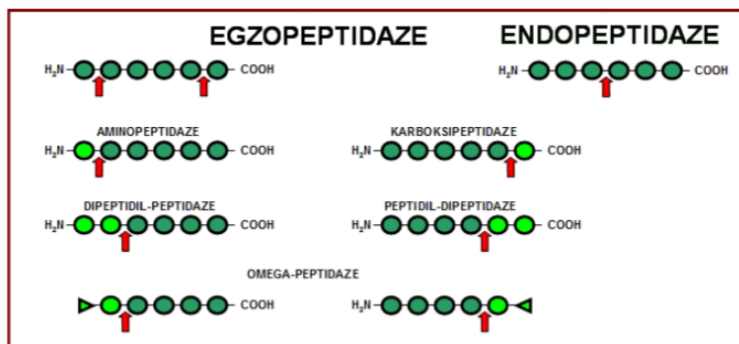
Signalizacijski put Keap1-Nrf2 u stanicama djeluje kao regulator staničnog odgovora na egzogene i endogene stresne poremećaje uzrokovane reaktivnim oksidativnim vrstama (ROS) i elektrofilima <sup>4</sup>. Keap1-Nrf2 kompleks kontrolira ekspresiju skoro 500 različitih gena. Transkripcijski faktor Nrf2 veže se kao monomer za homodimerni Keap1 protein putem dva motiva, DLG i ETGE. Pronađeno je da se proteini koji sadrže ETGE motiv natječu sa Nrf-om u vezanju za Keap1 <sup>5</sup>. Takav protein je i DPP III. Eksperimenti su pokazali da se DPP III veže za Kelch domenu Keap1 proteina putem ETGE motiva <sup>5</sup>. U prethodnim ekperimentima ispitan je afinitet vezanja divljeg tipa DPP-a III s Keap1 proteinom kao i molekularni detalji interakcije putem MD simulacija te je ustanovljena promjena konformacije proteina uzrokovana njegovim vezanjem <sup>6</sup>. Cilj ovog rada je provođenjem molekularno dinamičkih simulacija na kompleksu koji se sastoji od mutiranog tipa DPP-a III i Keap1 proteina usporediti interakciju između nemutiranog DPP-a III sa Keap1 proteinom i interakciju mutiranog tipa DPP-a III i Keap proteina.

#### 3.1 Dipeptidil peptidaza III

##### 3.1.1 Peptidaze

Enzimi su proteini koji djeluju kao katalizatori kemijskih reakcija. Enzimi se dijele prema vrsti kemijske reakcije koju kataliziraju na klase i podklase. Peptidaze su podklasa enzima koje pripadaju klasi hidrolaza koje kataliziraju reakciju hidrolize (reakcije razgradnje molekula kovalentnih složenih kemijskih spojeva s vodom). Imaju bitnu ulogu u katabolizmu proteina, a sudjeluju i u drugim važnim biološkim procesima poput morfogeneze tkiva, angiogeneze i apoptoze. Peptidaze djeluju tako da razgrađuju proteine i peptide na način da kataliziraju hidrolizu peptidnih veza koje povezuju aminokiseline u polipeptidnom lancu. Peptidaze mogu cijepati peptidne veze koje se nalaze na kraju lanca proteina (egzopeptidaze) ili unutrašnje peptidne veze (endopeptidaze) što je prikazano na Slici 4. Egzopeptidaze se klasificiraju prema broju aminokiselina koje odcjepljuju i prema kraju polipeptidnog lanca sa kojeg odcjepljuju a to može biti N-kraj ili C-kraj (svaka aminokiselina ima amino i karboksilnu skupinu, a lanac aminokiselina se povezuje peptidnim vezama koje se stvaraju

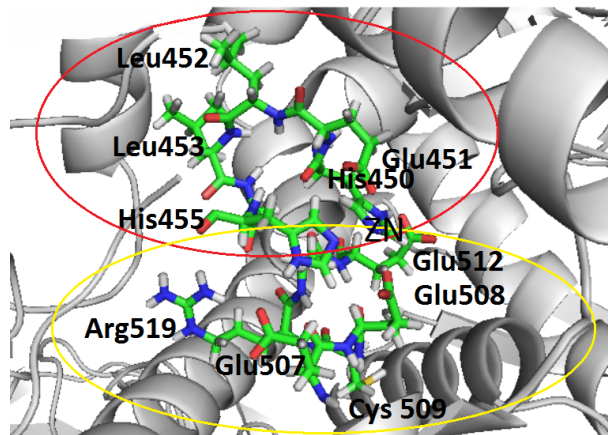
u reakciji dehidracije na način da se karboksilna skupina jedne aminokiseline spoji sa aminogrupom sljedeće aminokiseline, što znači da polipeptidni lanac ima dva kraja- aminoskupinu, N-kraj, i karboksilnu skupinu, C-kraj.). Dipeptidil peptidaze odcjepljuju dvije aminokiseline s N- kraja polipeptidnog lanca.



Slika 4: Podjela peptidaza prema mjestu cijepanja. Mjesto hidrolaze označeno je crvenom strelicom. Slika je preuzeta iz Literature<sup>7</sup>.

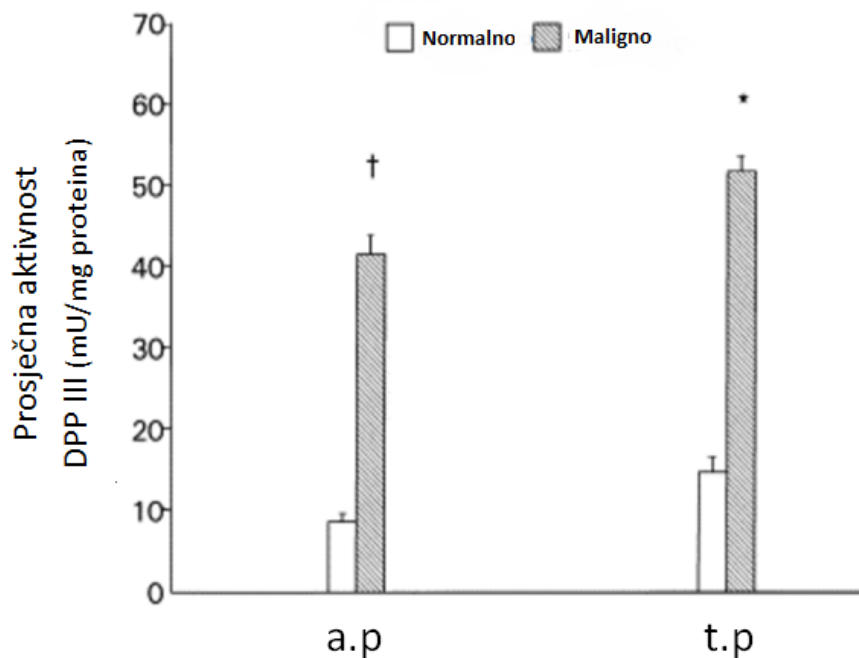
### 3.1.2 DPP III klasifikacija

Dipeptidil peptidaze tipa III su jedna od četiri tipa iz grupe dipeptidil peptidaza a čine zasebnu porodicu proteina koja je klasificirana na temelju statističke sličnosti aminokiselinskih sljedova svojih članova odgovornih za vezanje metalnog kationa i katalizu i jedinstvenog strukturnog motiva (strukturni motivi su kratki segmenti ukupne proteinske strukture koji kod proteina opisuju vezu između sekundarnih proteinskih struktura, obično se sastoji od 3-10 aminokiselina<sup>8</sup>), heksapeptida HEXXGH koji je odgovoran za vezanje cinka i katalitičku aktivnost enzima. Specifično, kod DPP-a III strukturni motiv HEXXGH ima oblik HELLGH (što znači da su aminokiseline koje ga tvore redom H-Histidin, E-Glutamat, L-Leucin, L-Leucin, G-Glicin, H-Histidin). Osim HELLGH strukturnog motiva, kod DPP-a III očuvan je i EECRAE motiv (tvore ga aminokiseline E-Glutamat, E-Glutamat, C-Cistein, R-Arginin, A-Alanin, E-Glutamat). Oba katalitička motiva dio su katalitičke domene. Pojedine aminokiseline iz navedenih strukturnih motiva sudjeluju u koordinaciji cinkovog iona  $Zn^{2+}$ . Enzimi tipa Dipeptil peptidaze III formiraju grupu aminopeptidaza molekularnih masa 80-85 *kDa*, s izoelektričnom točkom blizu pH 4,5.



Slika 5: Uvećano; strukturni motivi proteina DPP III. EECRAE strukturni motiv zaokružen je žuto, a HEXXGH strukturni motiv karakterističan za obitelj M49 metalopeptidaza kojima pripada DPP III ljubičasto. Mjesto gdje se nalazi cinkov ion označeno je oznakom ZN. Slika je generirana pomoću programa Pymol.

DPP III je široko prisutna kod organizama. Kod sisavaca je uočena u hipofizi, mozgu, eritrocitima, koži i jetri<sup>9</sup>, dakle u stanicama raznih tkiva pa se zbog toga pretpostavlja da sudjeluje u procesima unutarstaničnoga metabolizma proteina i peptida. Također je zabilježena kod nekih nižih eukariota. Povišena aktivnost i razina DPP-a III u humanom tkivu uočena je kod raka endometrije (vrsta raka koja se pojavljuje u maternici) i kod raka ovarija što zapravo sugerira da bi se mjerenje aktivnosti enzima DPP III u humanim tkivima mogla koristiti i kao potencijalni indikator endometrijskih i ovarijskih karcinoma, odnosno kao tumorski biomarker<sup>9</sup>. Neke farmakološke studije povezuju DPP III sa mehanizmima regulatora boli zato što su niske razine aktivnosti DPP-a III uočene u cerebrospinalnoj tekućini pacijenata koji su patili od akutnih bolova<sup>10</sup>. DPP III katalizira razgradnju prirodnih neuropeptida kao što su L-encefalin i endomorfina. Pokazano je da DPP III djeluje kao aktivator Keap1-Nrf2 signalizacijskog puta koji djeluje kao mehanizam obrane stanice od kancerogena<sup>5</sup>. DPP III je široko prisutna kod organizama. Kod sisavaca je uočena u hipofizi, mozgu, eritrocitima, koži i jetri<sup>9</sup>, dakle u stanicama raznih tkiva pa se zbog toga pretpostavlja da sudjeluje u procesima unutarstaničnoga metabolizma proteina i peptida. Također je zabilježena kod nekih nižih eukariota.



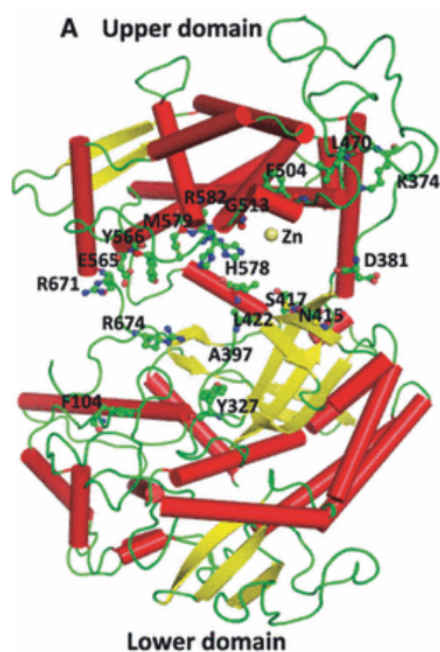
Slika 6: Usporedba aktivnosti hDPP-a III u raku endomerija i normalnog tkiva; opažene su znatno povišene aktivnosti kod kancerogenog tkiva. Slika je preuzeta iz Literature<sup>9</sup>.

### 3.1.3 Struktura Dipeptidil peptidaze III

Sljed ljudske DPP III je duljine 737 aminokiselina. 3D truktura DPP III se sastoji od dvije domene, gornje i donje koje su međusobno odvojene velikim procjepom. Gornja domena se uglavnom sastoji od  $\alpha$  zavojnica i to je C-terminalna domena, a donja od  $\alpha$  zavojnica i  $\beta$  ploča (oblici sekundarnih proteinskih struktura) i ona je N-terminalna domena. Aktivno mjesto kod proteinskih enzima predstavlja mali broj aminokiselina smještenih u unutrašnjosti proteinske molekule a njihov prostorni raspored je takav, da omogućava specifično vezanje molekule supstrata. Naime, svaki supstrat koji se veže za enzim ima točno određeni oblik. Aktivno mjesto enzima je modelirano tako da se supstrat savršeno prostorno uklopi u njega onemogućujući tako vezivanje molekula neodgovarajućeg oblika. Aktivno mjesto je, uz koenzime koji se vežu na aktivno mjesto, odgovorno za katalitičku sposobnost enzima. Aktivno mjesto s ionom cinka kao koenzimom je kod DPP-a III smješteno na donjem dijelu kata-

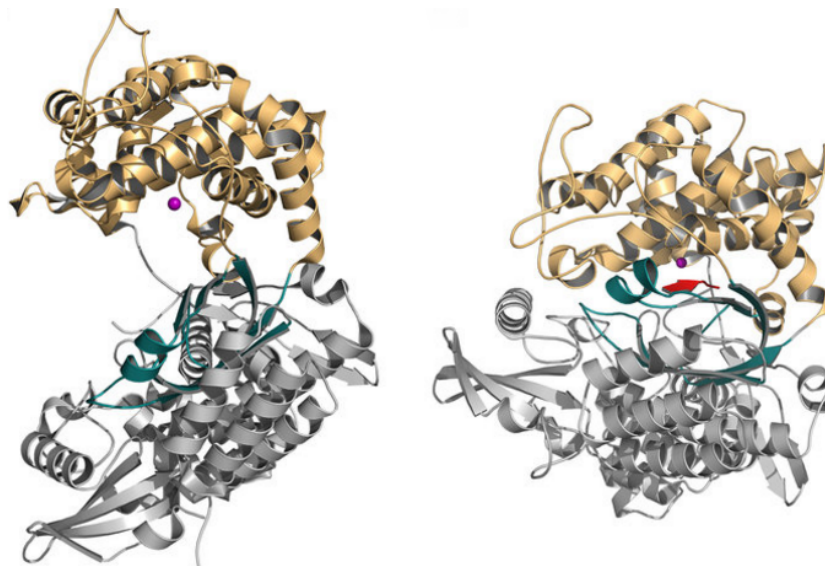


litičke gornje domene i usmjereno ka pukotini. Katalitički motiv HEXXGH dio je gornje domene. Aminokiseline koje su dijelovi karakterističnih strukturnih motiva zajedno s molekulom vode i koordiniraju cinkov ion. Pomoću rendgenske strukturne analize 2008. godine određena je 3D kristalna struktura mutanta DPP-a III iz kvasca *Saccharomyces cerevisiae*, yDPP III koja predstavlja strukturni prototip za cijelu M49 obitelj metalopeptidaza, a u bazi podataka proteina Protein Data Bank (PDB) nalazi se pod kodom 3CSK. Struktura je prikazana na Slici 7.



Slika 7: Kristalna struktura kvašćeve DPP III (yDPP III, PDB 3CSK). Cinkov ion označen je kao žuta kuglica.  $\alpha$  zavojnice obojene su crveno, a  $\beta$  ploče žuto. Slika je preuzeta iz Literature<sup>11</sup>.

Nešto kasnije određena je i kristalna struktura ljudske DPP III (hDPP III). Objavljena je 2009. godine u bazi podataka proteina Protein Data Bank pod kodom 3FVY<sup>12</sup>. Humana (hDPP III) i kvašćeva (yDPP III) prostorna građa je vrlo slična. hDPP III postoji u dva prostorna oblika, konformacije, otvorenoj i zatvorenoj a one su prikazane na Slici 8. MD simulacije u ovom radu provest će se samo na zatvorenoj konformaciji.

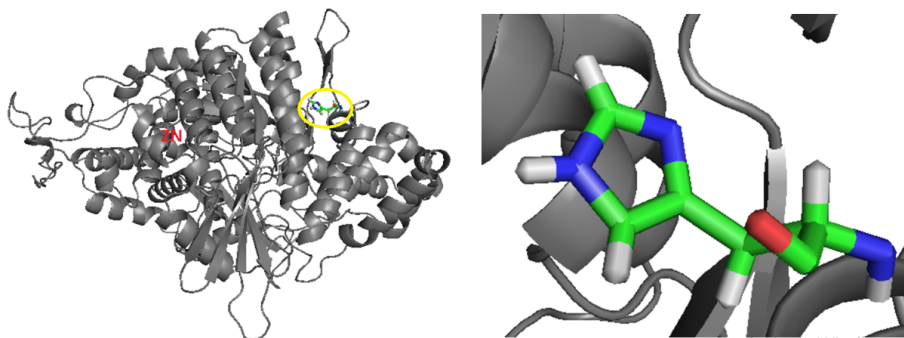


Slika 8: Lijevo: otvorena konformacija hDPP III Desno: Zatvorena konformacija hDPP III. Gornje domene su u oba slučaja prikazane narančasto, a donje sivo. Cinkov ion je prikazan kao ljubičasta kuglica. Slika je preuzeta iz Literature <sup>13</sup>, a dobivena je pomoću programa Pymol.

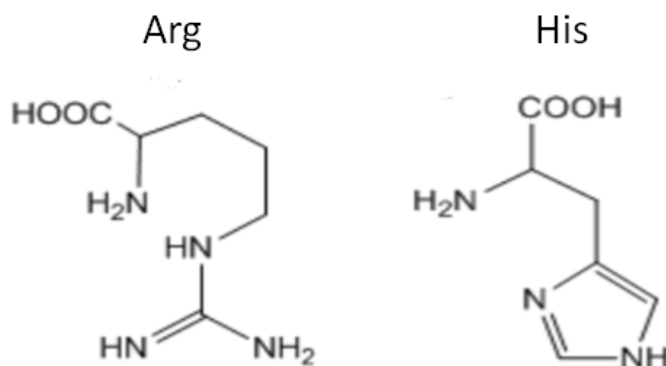
### 3.1.4 Mutacije DPP III

Osim nemutiranog ili divljeg tipa DPP III proteina, postoje i različiti mutanti. Da bi se "napravio" mutant dovoljno je zamijeniti samo jednu aminokiselinu u aminokiselinskom lancu. Mutacije se mogu pokazati korisnim, štetnim ili neutralnim. Na primjer, hDPP III ima u svojoj strukturi ukupno šest Cisteina, od čega se tri (Cys509, Cys519 i Cys654) nalaze u gornjoj, a tri (Cys19, Cys147 i Cys176) u donjoj domeni. U projektu kojeg su proveli M.Abramić i suradnici<sup>12</sup>, napravljeno je šest mutiranih formi enzima DPP III na način da se svaki pojedini cistidin (Cys) nezavisno supstituirao alaninom (Ala) da bi se ispitala njihova svojstva. U ovom radu je MD simulacijama podvrgnut kompleks koji se sastoji od mutiranog tipa hDPP III proteina u kojem je na mjestu 703 originalni Arginin (Arg) zamijenjen Histidinom (His) i Keap1 proteina. Mutacija je prikazana na Slici 9., a strukture Histidina i Arginina na Slici 10. Rezultati interakcije su uspoređeni sa rezultatima interakcije divljeg tipa DPP-a III sa Keap1 proteinom koja je napravljena u radu gdje je kombinacijom molekularnog dokiranja

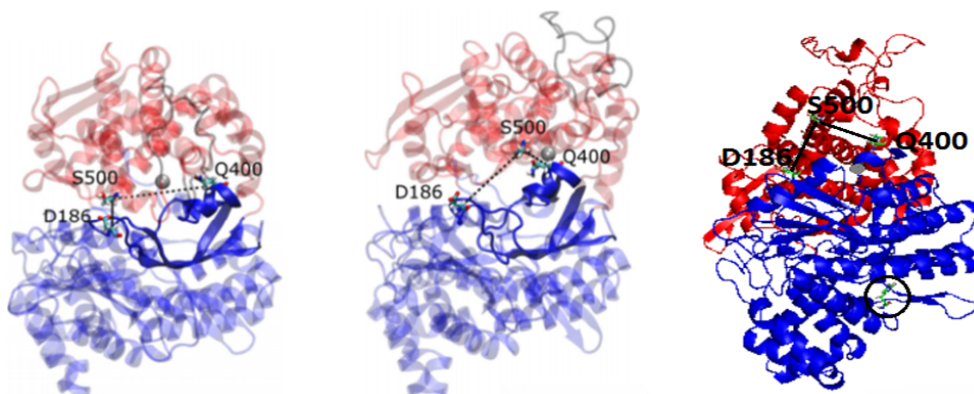
(eng. molekular docking) i MD simulacija ispitano kolika je zapravo promjena konformacije samog proteina uzrokovana njegovom mutacijom i vezanjem za Keap1<sup>6</sup>. Usporedba samih struktura eksperimentalno određene forme, kao i zatvorenih formi divljeg tipa DPP III i mutanta R703H, dana je na Slici 11.



Slika 9: Mutacija DPP III proteina. Mjesto na kojem se nalazi cinkov ion istaknuto je oznakom ZN, a mjesto mutacije istaknuto je žutim krugom (lijevo). Uvećano mjesto mutacije gdje je Arginin iz divljeg tipa zamijenjen Histidinom (desno). Slika je generirana pomoću programa Pymol.



Slika 10: Strukturne formule Arginina i Histidina. Slika je preuzeta iz Literature<sup>14</sup>.

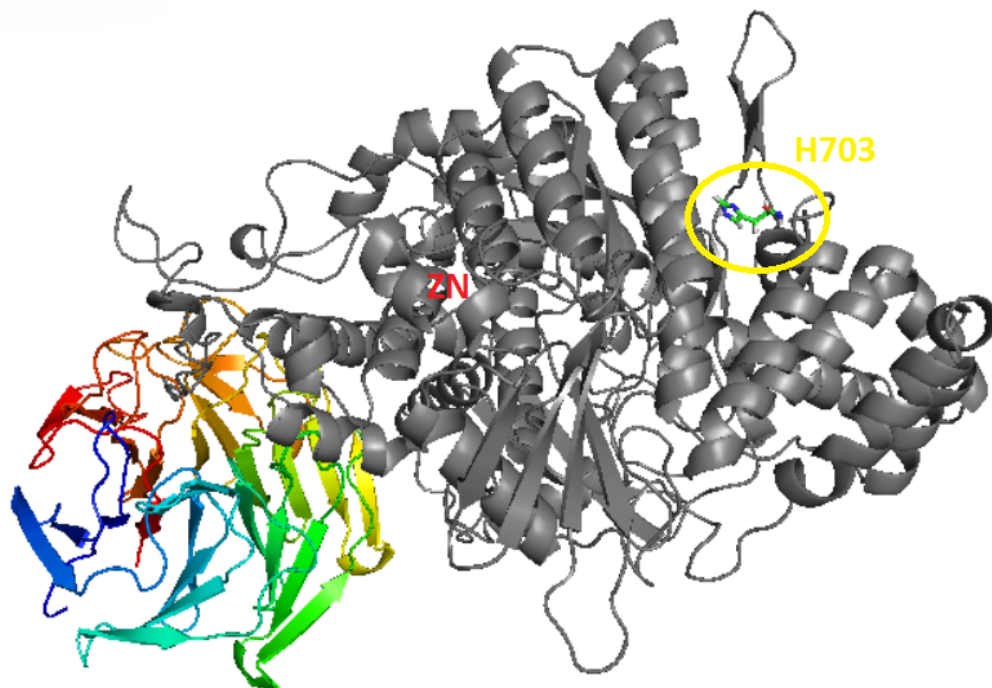


Slika 11: Usporedba struktura zatvorene forme DPP III proteina i mutanta R703H. Gornja domena (aminokiseline 337-374 i 422-668) obojana je crveno, a donja (aminokiseline 1-336, 375-421 i 669-726) plavo. Cinkov ion prikazan je kao siva sfera. Istaknute su aminokiseline S500, Q400 i D186 koje definiraju udaljenosti  $d_1$  i  $d_2$  koje će se pratiti tijekom simulacija a prikazane su u štapičastoj reprezentaciji. Lijevo je prikazana zatvorena forma divljeg tipa DPP III proteina određena eksperimentalno, rentgenskom difrakcijom. U sredini je prikazan isti protein dobiven MD simulacijama. Desno je prikazana zatvorena forma mutanta DPP III proteina, R703H. Mjesto mutacije zaokruženo je crnom bojom. Slike lijevo i sredina preuzete su iz Literature <sup>15</sup>, dok je slika mutanta generirana pomoću programa Pymol.

### 3.2 Kompleksi Nrf2-Keap1 i DPP III-Keap1

Keap1-Kelch-like ECH-associated protein je humani protein kodiran Keap1 genom. Najpoznatiji je po interakciji sa proteinom Nrf2 koji ima ulogu transkripcijskog faktora. Nrf2 je protein koji se u humanim stanicama nalazi u latentnom stanju, a oslobađa se putem Nrf2 aktivatora koji pokreće Nrf2 signalizacijski put. Oksidativni stres i šteta u stanicama koja nastaje od slobodnih radikala su uzročnici mnogih bolesti tipa raka, dijabetesa i arteroskleroze. Aktivacijom Nrf2 proteina u jezgri, započinje proizvodnja antioksidativnih enzima (npr. Katalaze, Glutaciona) koji su dovoljno jaki da neutraliziraju i do milijon slobodnih radikala po sekundi, svake sekunde. Keap1-Nrf2 signalizacijski put ima centralnu ulogu u zaštiti stanica od oksidativnog i ksenobiotičkog stresa<sup>16</sup>. Kad se stanica ne nalazi pod stresnim okolnostima, Nrf2 je u stanici konstantno inaktiviran preko E3 ligaze (jedna od ligaza koje mogu pokrenuti proces degradacije proteina)

čija se podjedinica, Cullin3 (Cul3) povezuje s Keap1 proteinom tvoreći kompleks Cul3-Keap1 koji na Nrf2 djeluje tako da ga brzo razgrađuje u proteasomama u kojima se inače proteolizom razgrađuju nepotrebni ili oštećeni proteini. S druge strane, kad je stanica izložena elektrofilicnom ili oksidativnom stresu, dolazi do modifikacije reaktivnih cisteinskih ostataka što vodi padu aktivnosti E3 ligaze, stabilizaciji Nrf2 proteina, povezivanju Nrf2 proteina s Keap1 proteinom i pokretanju citoprotektivnih gena. Kod različitih tipova humanih tumora, primjećene su mutacije na genima Nrf2 i/ili Keap1 proteina. Te mutacije ometaju aktivnost Keap1-Nrf2 kompleksa. Također, povišena ekspresija Nrf2 gena u kancerogenim stanicama kreira okruženje pogodno za ubrzan stanični rast <sup>17</sup>. Biokemijskim i strukturnim analizama utvrđeno je da se Nrf2 veže za Keap1 putem ETGE strukturnog motiva. Također, utvrđeno je da se i drugi proteini koji sadrže ETGE strukturni motiv, između kojih je i DPP III natječu sa Nrf2 proteinom za vezanje za Keap1. Pokazano je da se DPP III veže za Kelch domenu Keap1 proteina putem ETGE motiva i na taj način sprečava formiranje Keap1-Nrf2 kompleksa. U radu u kojem su se proučavale interakcije divljeg tipa DPP III proteina sa Keap1 proteinom <sup>6</sup> ispitani su molekularni detalji interakcije ta dva proteina pomoću MD simulacija. U ovom radu je napravljena analiza koja ispituje vezanje mutanta R703H s Keap1 proteinom te su rezultati uspoređeni s rezultatima dobivenim za divlji tip DPP III proteina. Kompleks koji je podvrgnut MD simulacijama prikazan je na Slici 12.



Slika 12: Kompleks mutanta DPP III-Keap1 proteina. Mutant DPP III R703H obojan je sivo, dok je Keap1 obojen prema spektru. Mjesto na kojem se nalazi cinkov ion istaknuto je oznakom ZN, a mjesto mutacije zaokruženo je žuto. Slika je generirana pomoću programa Pymol.

## 4 Metodologija

### 4.1 Priprema sustava za provođenje MD simulacija

Prije provođenja samih MD simulacija, potrebno je pripremiti 3D strukturu proteina, odnosno potrebno je provesti parametrizaciju i generirati neophodne datoteke koji izabrani program za simulaciju zahtijeva te ga minimizirati i uravnotežiti. Cilj tih simulacija je napraviti optimizaciju sustava, odnosno dobiti stabilan sustav bez većih fluktuacija vrijednosti izabranih parametara sustava (na primjer, temperature ako se simulira sustav pri konstantnoj temperaturi).

### 4.2 Priprema strukture i solvatacija

Struktura koja će se simulirati se definira u nekoliko koraka: unos same strukture, definicija polja sila koje će se koristiti, dodavanje iona, dodavanje molekula vode što uključuje specifikaciju tipa vode i gustoća kocki vode koja će se koristiti, solvatacija kojom se postiže stanje proteina najbližije njegovom prirodnom stanju.

Strukture koje su korištene su originalno preuzete iz PDB-a- Protein Data Base-a, a konačne strukture korištene u ovom eksperimentu preuzete su iz ranijih radova <sup>6</sup> i takve direktno učitane u program. Postoje različite klase polja sila dostupna u različitim programima. U ovom radu korišteno je jedno od specijaliziranih polja sila dostupnih u programskom paketu *Amber* razvijena za biomakromolekule. Polje sila koje je korišteno u svim simulacijama jest ff03 <sup>18</sup> polje sila. Nakon unosa same strukture i definicije polja sile, dodane su molekule vode. Kako bi se smanjilo vrijeme potrebno za simulaciju sustavu u kojem najveći dio pripada atomima vode kojih je najviše, za okruživanje proteina korištene su molekule vode tipa TIP3P. Tip molekula vode TIP3P specifičan je po tome što su uklonjeni određeni stupnjevi slobode koji ne utječu značajno na ponašanje sustava, naime naboji su raspoređeni u tri točke koje predstavljaju središta atoma O i H, a stupnjevi slobode koji su uklonjeni su duljine veza i kutevi koji se tijekom simulacije ne mijenjaju. Također, pretpostavlja se da je potencijal za te molekule vode se sastoji od članova koji uključuju Lennard Jonesov potencijal i Coulombov član <sup>19</sup>. Sustav je još neutraliziran dodavanjem natrijevih ( $Na^+$ ) iona. Neutralizacija i solvatacija sustava provedena je u programu *Leap* koji se nalazi unutar programskog paketa *Amber*. Nakon

toga generirane su datoteke za topologiju i koordinate sustava koje se koriste u daljnjim koracima minimizacije, ekvibracije i simulacija. Datoteke koje opisuju topologiju i koordinate sustava sadrže podatke tipa ukupnog broja atoma u sustavu, ukupnog broja različitih atoma u sustavu, ukupnog broja atoma koje sadrže vodikove veze u sustavu i slično.

### 4.3 Minimizacija

Provođenjem simulacije minimizacije pomoću programa *Sander* koji se nalazi unutar programskog paketa *Amber* optimizirana je geometrija proteina. Minimizacija sistema provedena je u tri koraka zato da se čim više smanje greške, odnosno da se čim bolje uklone intramolekulska naprezanja. Nakon provođenja minimizacija, postignut je najbolji i najtočniji mogući energetski raspored molekula vode oko proteina (optimizacijom udaljenosti kovalentnih veza, valentnih kuteva i sl. postignuta je minimizacija energije sistema).

### 4.4 Ekvibracija

Nakon simulacija minimizacije, provode se simulacije ekvibracije (uravnoteženja) sustava odnosno njegovih parametara- tlaka, temperature, energije i slično. Simulacija ekvibracije omogućuje dodavanje kinetičke/ potencijalne energije u sistem, zagrijavanje sustava, postavljanje tlaka i slično. Ekvibracija je provedena u dva koraka. U prvom koraku ekvibracije koristi se *NVT* ansambl. U tom koraku je položaj atoma proteina fiksiran, dok je kretanje molekula otapala nesputano. Ograničenja su u sistem uvedena na način da se ograniče vibracije svih veza, uključujući vodik putem SHAKE algoritma. Tijekom prve ekvibracije ustav je zagrijan s početne temperature 0K na konačnu temperaturu 300 K, pri čemu je korišten *NVT* ansambl. Vremenski korak u simulacijama ekvibracije iznosio je  $1 fs$ . U drugj ekvibraciji se koriste uvjeti konstantnog tlaka i temperature *NPT* ansambl, pri čemu se provodi podešavanje gustoće vode.

### 4.5 Produktivne MD simulacije

Uravnotežena struktura kompleksa R703H-Keap1 podvrgnuta je četirima nezavisnim MD simulacijama (korištenjem različitih nasumičnih brojeva za generiranje početnih brzina sustava, definirano s varijablom  $ig=-1$ ). U tim simulacijama



primjenjeni su uvjeti konstantnog tlaka i temperature (1 atm i 300 K, korištenjem *NPT* ansambla). Temperatura je održavana konstantnom korištenjem Langevinovog termostata, s frekvencijom sudara  $1 \text{ ps}^{-1}$  dok se tlak regulira Berendsenovim barostatom. Duljina vremenskog koraka iznosila je  $2 \text{ fs}$ . Primjenjeno je ff03 polje sila. Sveukupno, simulacije R703H DPP III rezultirale su trajektorijom dugom 50000 ps.

## 4.6 Analiza rezultata

Analiza simulacije molekulske dinamike uključuje kvalitativnu i kvantitativnu analizu. Promjene u strukturi i dinamici promatranog sistema koje su opažene vizualizacijom potrebno je ispitati i opisati mjerenjem različitih strukturnih varijabli i njihovih promjena (udaljenosti, radijusi giracije i slično) ili dinamičkih svojstava kao fluktuacija.

### 4.6.1 RMSD

Srednji korijen kvadrata udaljenosti, *RMSD* vrijednost (eng. root mean square deviation) je mjera prosječne udaljenosti ekvivalentnih atoma dviju struktura što znači da je to zapravo mjera sličnosti dviju struktura. Ta vrijednost se obično upotrebljava kao pokazatelj sličnosti dviju konformacija iste molekule, ali može se upotrijebiti i za usporedbu sličnosti različitih molekula. Dakle, manja *RMSD* vrijednost odgovara većoj sličnosti dviju struktura<sup>20</sup>. Iste konformacije imat će *RMSD* vrijednost nula. *RMSD* je dana matematički relacijom (4.1):

$$RMSD = \sqrt{\frac{1}{N} \sum_i \delta_i^2} \quad (4.1)$$

gdje je  $N$  ukupan broj atoma koji se uspoređuju a  $\delta_i$  udaljenost između dva ekvivalentna atoma.

### 4.6.2 Rgyr

Radijus giracije, *Rgyr* se općenito matematički definira kao srednji kvadrat udaljenosti dijelova nekog objekta od centra mase tog objekta ili osi rotacije tog objekta, dok se u biofizici definira kao srednji kvadrat sume udaljenosti pojedinih monomera (aminokiselina) od centra mase makromolekule. Promjena vrijednosti radijusa giracije u ovisnosti o vremenu se prati tijekom simulacija

s ciljem opažanja promjene strukture molekule jer ta promjena zapravo opisuje koliko se struktura molekule tijekom simulacija otvara ili zatvara. Radijus giracije  $R_g$  se izračunava prema formuli (4.2):

$$R_{gyr} = \sqrt{\frac{\sum_i^{N_{uk}} r_i^2 m_i}{\sum_i^{N_{uk}} m_i}} \quad (4.2)$$

gdje je  $r_i$  udaljenost  $i$ -tog promatranog monomera od centra mase strukture,  $m_i$  masa tog monomera, a  $N_{uk}$  ukupni broj monomera u ispitivanoj strukturi. Statistička analiza Statistička analiza geometrijskih parametara provedena je u programu Qtiplot specijaliziranom za statističke proračune, a na temelju numeričkih podataka dobivenih direktno iz MD simulacija. Općenito, za srednje vrijednosti  $\bar{x}$  i pripadne pogreške  $M_x$  (srednje kvadratne pogreške aritmetičke sredine ili nepouzdanosti) općenitog parametra  $x$  uzete su vrijednosti izračunate u programu prema relaciji (4.3).

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad M_n = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}}, \quad M_n = \frac{m}{\sqrt{n}} \quad (4.3)$$

gdje je  $n$  ukupan broj mjerenja,  $x_i$   $i$ -to mjerenje, a  $m$  standardna devijacija uzorka.

## 4.7 Vizualizacija i obrada rezultata

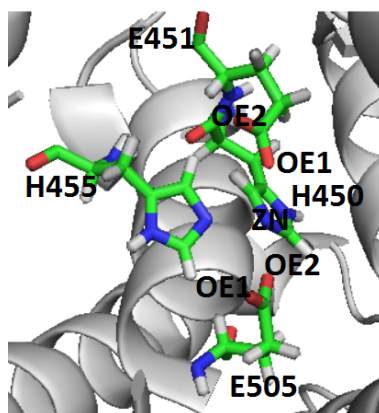
Slike koje prikazuju grafičke prikaze rezultata generirane su pomoću programa Grace, 2D programa razvijenog za crtanje grafova i namijenjenog primarno za upotrebu na operacijskom sustavu Linux. Slike za vizualizaciju sistema generirane su pomoću programa Pymol. Statistička analiza podataka napravljena je u programima Excel i Qtiplot.

## 5 Rezultati i rasprava

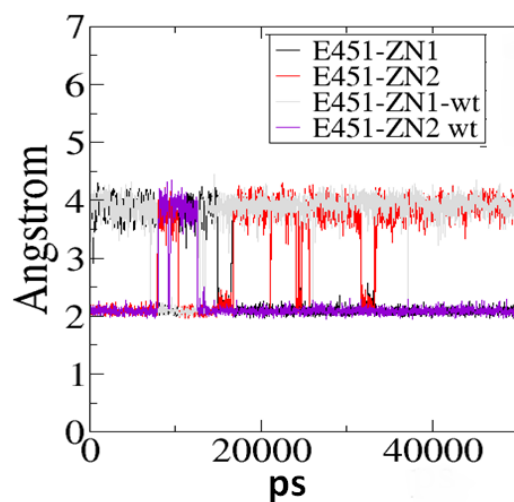
U ovoj sekciji dan je pregled analize koja je provedena nad rezultatima dobivenim nakon završenih MD simulacija. Cilj je utvrditi utjecaj mutacije R703H na strukturu i dinamiku DPP III, te ispitati utječe li ta mutacija na interakcije DPP III s Kelch domenom proteina Keap1. U tu svrhu rezultati analize MD simulacija mutanta R703H uspoređeni su s rezultatima dobivenim za divlji tip DPP III proteina u ranijem istraživanju<sup>6</sup>.

### 5.1 Koordinacija cinkovog atoma

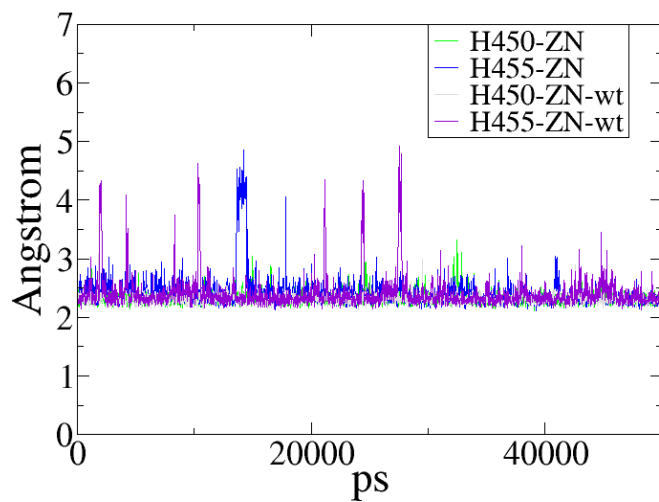
Koordinacija cinkovog atoma vrši se pomoću aminokiselina koje su dio *HELLGH* i *EECRAE* strukturnih motiva DPP III proteina. Provedena je analiza za aminokiseline H450, E451, E505 i H455 koje su dio *EECRAE* strukturnog motiva i za aminokiselinu E505 koja je dio *HELLGH* strukturnog motiva. Glutamati E451 i E505 u svojoj strukturi sadrže karboksilnu skupinu koja sadrži dva kisikova atoma pa se koordinacija cinka vrši putem jednog ili drugog kisikovog atoma što je za vizualizaciju prikazano na Slici 13. Zbog toga su za te aminokiseline rezultati za svaki kisikov atom, OE1 i OE2, prikazani odvojeno. Rezultati su prikazani grafički na Slici 14, Slici 15. i Slici 16. Rezultati dobiveni za koordinaciju cinkovog iona su u skladu s očekivanjima, odnosno, koordinacija cinkovog atoma je ostala sačuvana tijekom simulacija. Rezultati za Histidin H450 su očekivani jer se na grafičkom prikazu vidi da oba Histidina koordiniraju ion. Nadalje, vidi se da Glutamati E451 divljeg tipa DPP III tijekom trajanja simulacija od 5000 ps konstantno monodentno koordiniraju cinkov ion, dok je u slučaju mutanta koordinacija bidentatna. Kod prikaza rezultata za Glutamat E505 vidi se da se kisikovi atomi OE1 i OE2 izmjenjuju u koordinaciji cinkovog iona.



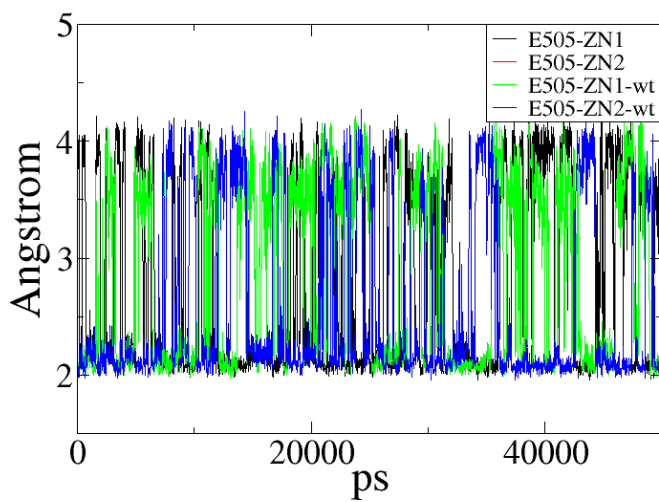
Slika 13: Prikaz aminokiselina koje vrše koordinaciju cinkovog atoma. Aminokiseline su prikazane u štapičastoj reprezentaciji i označene standardnim kraticama (E- Glutamat, H- Histidin), dok su sa OE1 I OE2 označeni kisikovi atomi koji su dio karboksilne skupine glutamata, a koji sudjeluju u koordinaciji. Mjesto na kojem se nalazi cinkov atom istaknuto je oznakom ZN. Slika je generirana pomoću programa Pymol.



Slika 14: Koordinacija aminokiselinom E451.



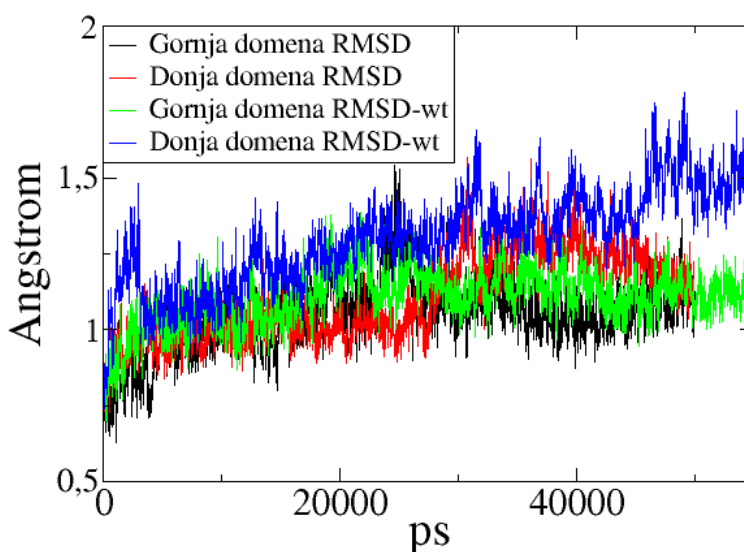
Slika 15: Koordinacija aminokiselina H450 i H455.



Slika 16: Koordinacija aminokiselinom E505.

## 5.2 RMSD

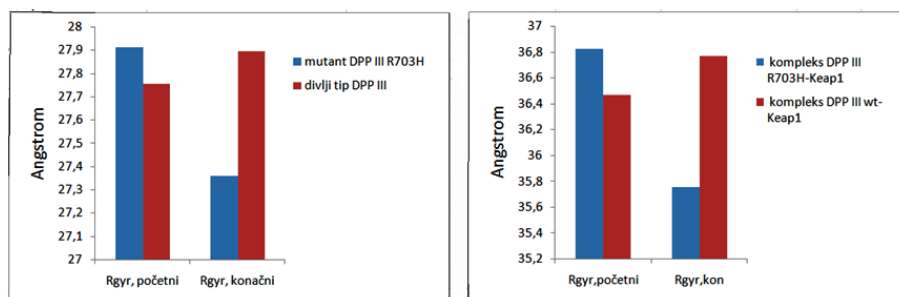
U ovom radu se praćenjem promjene *RMSD* vrijednosti DPP-a III tijekom simulacija dolazi do informacija o promjeni proteinske strukture tijekom simulacija. Rezultati su prikazani na Slici 17. Analiza *RMSD* profila upućuje na stabilnost cjelokupne strukture R703 H i divljeg tipa DPP tijekom 50000 ps simulacija, dakle i gornja i donja domena ostaju stabilne tijekom simulacija. *RMSD* vrijednosti svake domene se sveukupno promijene za manje od 1.5 Å tijekom cijelog trajanja simulacija, a uzme li se u obzir da iste strukture imaju *RMSD* vrijednost nula, slijedi da struktura ostaje stabilna.



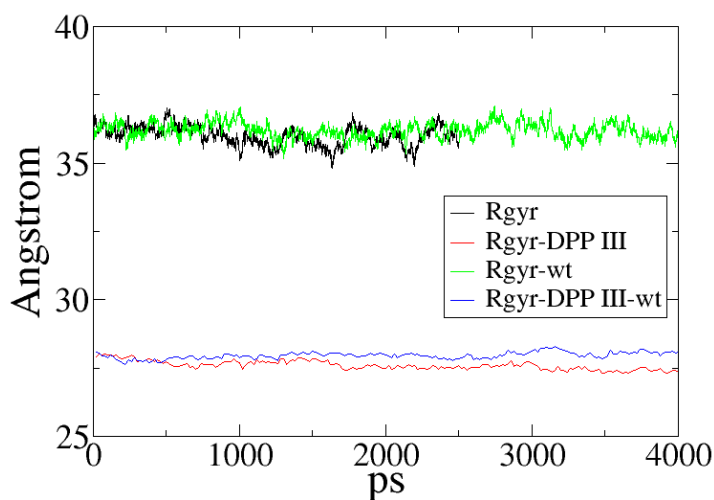
Slika 17: Prikaz promjena *RMSD* vrijednosti za odabrane dijelove kompleksa DPP III-Keap1 u vremenu trajanja simulacija od 50000 ps. Prikazane su *RMSD* vrijednosti za mutant R703H (crna i crvena linija) i za divlji tip (zelena i plava linija) istog proteina. Izračuni su provedeni odvojeno za različite dijelove proteina; gornja domena (crno- nemutirani tip, zeleno-divlji tip) koja se sastoji od aminokiselina 337-374 i 422-668 i donja domena (crveno- nemutirani tip, plavo- divlji tip) koja se sastoji od aminokiselina 1-336, 375-421, 669-726. Slika je generirana pomoću programa Grace.

### 5.3 Rgyr

Radijus giracije zapravo opisuje stupanj zatvorenosti proteina. Smanji li se radijus giracije tijekom simulacija, može se reći da dolazi do povećanja globularnosti proteina. Rezultati promjena radijusa giracije dani su na Slikama 18., 19. i Tablici 2. Analiza rezultata za promjene radijusa giracije ukazuje na blago zatvoreniju formu kompleksa kojeg tvori R703H s Keap1 proteinom od kompleksa divljeg tipa i Keap1 proteina. U slučaju mutanta R703H, došlo je do smanjenja radijusa giracije s početne vrijednosti na početku produktivne MD faze s 27.9117 na 27.7531, dok se u slučaju divljeg tipa DPP III radijus giracije promijenio s početne vrijednosti 27.7531 na 27.8929. Na grafikonima se dobro vide odnosi promjena radijusa giracije za različite forme. Vežanje mutanta R703H s Keap1 proteinom ima nešto veći utjecaj na promjenu same strukture molekule od vežanja divljeg tipa s Keap1 proteinom. Potrebno je napomenuti da su na Slici 18. početni i konačni odnosi radijusa giracija dani za ilustraciju. Na tom prikazu izgleda kao da je promjena jako velika, no treba uzeti u obzir da je raspon cijele ordinate  $1 \text{ \AA}$ , tako da su te promjene blage.



Slika 18: Grafikoni prikazuju odnose vrijednosti radijusa giracije na početku i na kraju MD simulacija, lijevi za same proteine, a desni za komplekse. Vrijednosti su očitane iz numeričkih podataka dobivenih simulacijama.



Slika 19: Grafički prikaz promjene radijusa giracije  $R_{gyr}$  u vremenu trajanja MD simulacije od 4000 ps. Crna krivulja odgovara radijusu giracije cjelokupnog kompleksa R703H- Keap1. Crvena krivulja odgovara radijusu giracije mutanta R703H. Zelena (samo DPP III) i plava linija (cijeli kompleks) analogno odgovaraju istim veličinama za divlji tip proteina i Keap1 proteina. Slika je izrađena pomoću programa Grace.

Tablica 2: prosječne vrijednosti radijusa giracije i njihove nepouzdanosti

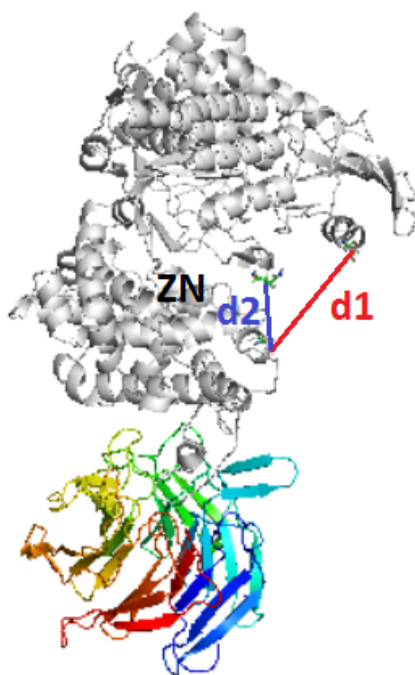
Parametar	Mutant R703H-Keap1	DPP III,R703H	DPP III (wt) - Keap1	DPP III, divlji tip (wt)
$R_{gyr}, \text{\AA}$	$(35.988 \pm 0.007)$	$(27.463 \pm 0.005)$	$(36.81 \pm 0.01)$	$(27.749 \pm 0.004)$

#### 5.4 Udaljenosti pojedinih aminokiselinskih parova $d_1$ i $d_2$

S ciljem promatranja dinamike kompleksa, analizirane su udaljenosti između odabranih aminokiselinskih parova. Prva udaljenost,  $d_1$  odgovara udaljenosti između  $C\alpha$  atoma parova aminokiselina Asp186-Ser500. Druga udaljenost odgovara udaljenosti između  $C\alpha$  aminokiselinskog para Gln400-Ser500. Udaljenosti  $d_1$  i  $d_2$  definiraju međusobnu orijentaciju dviju proteinskih domena. Vizualizacija danih udaljenosti dana je na Slici 20., dok je na Slici 21. dan grafički prikaz mijenjanja udaljenosti  $d_1$  i  $d_2$  u vremenu. Odnosi početnih i konačnih



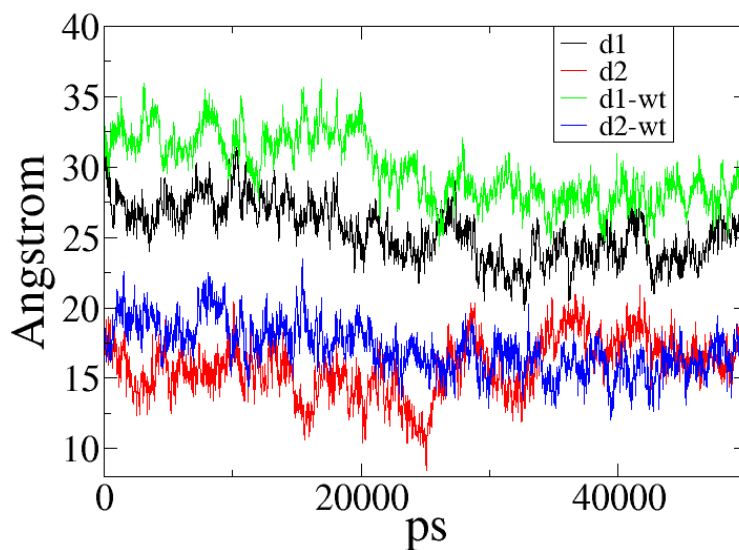
udaljenosti prikazani su na grafikonima na Slici 22. Te su udaljenosti izabrane tako, da jedna aminokiselina pripada gornjoj, a druga donjoj domeni. Aminokiselina Ser500 (S500) dio je gornje, dok su aminokiseline Asp186 (D186) i Gln400 (Q400) dijelovi donje domene DPP III. Iz rezultata se vidi da je mutant R703H DPP III proteina nešto zatvorenije strukture, budući da je prosječna  $d_1$  udaljenost zamjetno manja kod R703H nego kod divljeg tipa cijelo vrijeme trajanja simulacija što ukazuje na manji razmak između domena. Konačne vrijednosti  $d_1$  R703H se smanje za oko 25% početne vrijednosti, dok se kod divljeg tipa vrijednosti razlikuju tek oko 5%. Analizom numeričkih vrijednosti parametara  $d_1$  i  $d_2$  dobije se da  $d_1$  fluktuiraju oko srednjih vrijednosti ( $25.43 \pm 0.04$ ) za mutant, ( $29.03 \pm 0.04$ ) za divlji tip, a  $d_2$  oko ( $15.78 \pm 0.04$ ) za mutant te ( $15.75 \pm 0.07$ ) za divlji tip tijekom 50000 ps trajanja MD simulacija.



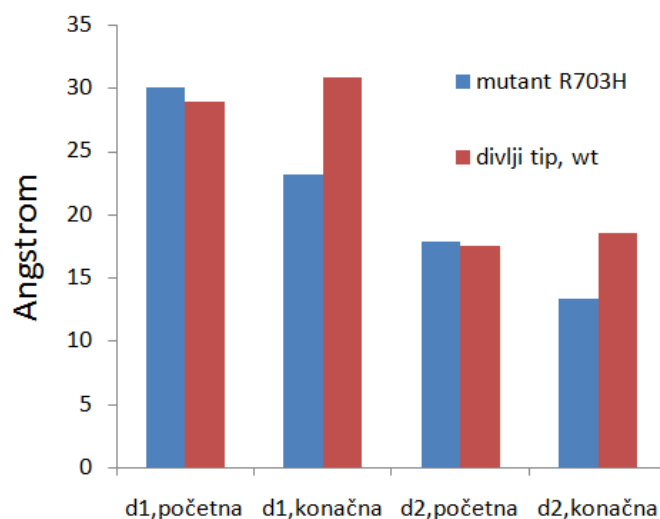
Slika 20: Vizualizacija promatranih udaljenosti  $d_1$  i  $d_2$ . Udaljenost  $d_1$  označena je crvenom linijom; udaljenost  $d_2$  označena je plavom linijom. Udaljenosti su prikazane na razini cijelog kompleksa R703H-Keap1. Mutant R703H obojan je sivo, dok je Keap1 obojan prema spektru. Mjesto gdje se nalazi cinkov ion označeno je oznakom ZN. Slika je generirana pomoću programa Pymol.

Tablica 3: Statistički izračunate vrijednosti za parametre  $d_1$  i  $d_2$

Parametar	Mutant R703H	Divlji tip (wt)
$d_1(D186 - S500)/\text{\AA}$	$(25.43 \pm 0.04)$	$(29.03 \pm 0.04)$
$d_2(Q400 - S500)/\text{\AA}$	$(15.78 \pm 0.04)$	$(15.75 \pm 0.07)$



Slika 21: Grafički prikaz promjene udaljenosti  $d_1$  i  $d_2$  u vremenu. Dani su rezultati nakon provođenja 5000 ns MD simulacija. Crvena i plava krivulja odgovaraju udaljenostima za mutant R703H, dok zelena i crna krivulja prikazuju rezultate za divlji tip. Graf je izrađen pomoću programa Grace.

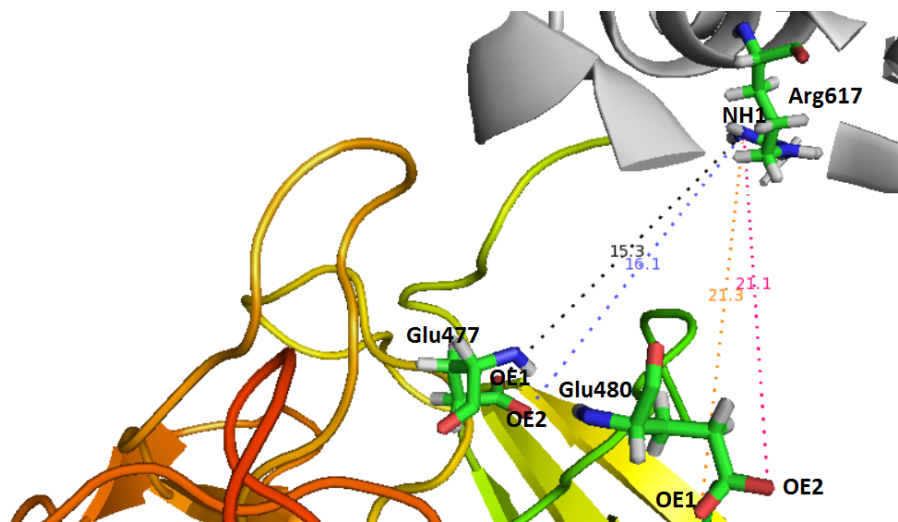


Slika 22: Grafikon s početnim i konačnim vrijednostima parametara  $d_1$  i  $d_2$ , posebno za mutant R703 H, a posebno za divlji tip.

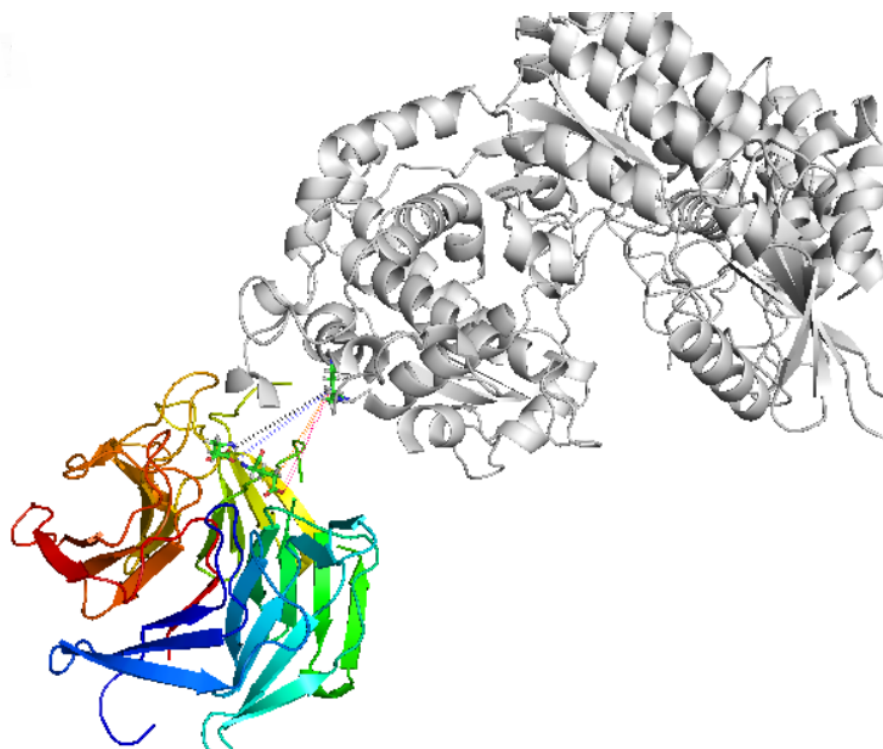
## 5.5 Interakcija DPP-III i Keap proteina

Potrebno je analizirati i same interakcije putem kojih se R703H i divlji tip DPP III vežu za Keap1 protein. Praćene su interakcije između različitih aminokiselina od kojih jedna pripada mutantu DPP III proteina, a druga Keap1 proteinu. Budući da nije svejedno između kojih se atoma pojedinih aminokiselina promatraju udaljenosti, ponekad je potrebno gledati više atoma iste aminokiseline, kao što je slučaj s Glutamatima, za koje se promatra interakcija s atomima kisika OE1 i OE2. Pogodno je promotriti interakcije između aminokiselina  $Arg617^{R703H} - Glu477 - OE1^{Keap}$ ,  $Arg617^{R703H} - Glu477 - OE2^{Keap}$ ,  $Arg617^{R703H} - Glu480 - OE2^{Keap}$ ,  $Arg617^{R703H} - Glu480 - OE2^{Keap}$ , zbog toga što su Glutamati Glu477 i Glu480 dio ETGE strukturnog motiva (ranije je navedeno (poglavlje 3.3) da se za Keap1 protein, osim Nrf2 faktora, mogu vezati i ostali proteini koji sadrže ETGE motiv). Vizualizacija odabranih aminokiselinskih parova dana je na Slici 23. i na Slici 24. Rezultati su prikazani na Slici 25. i Slici 26. Prvo je bitno uočiti da simulacije predviđaju da se ostvaruju interakcije aminokiselina iz *ETGE* motiva R703H i Kelch domene Keap1 proteina, odnosno da R703H interagira s Keap1 proteinom putem istih aminokiselina kao i divlji tip. Nadalje, na grafičkim prikazima na Slikama 25. i 26. vidi

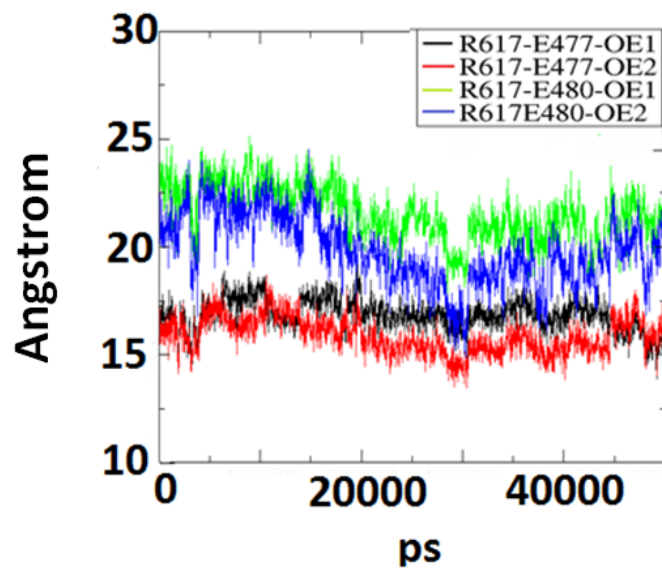
se da mutant R703H ostvaruje jače interakcije s Keap1 proteinom od divljeg tipa, što pokazuju manje oscilacije u udaljenosti tijekom vremena u slučaju mutanta nego u slučaju divljeg tipa. Kod divljeg tipa proteina su interakcije stabilne tijekom prvih 25000 ps simulacija, nakon čega se primjećuju zamjetne fluktuacije. Vidi se da na 25000 ps trajanja simulacija dokazi do skoka od čak 10 Å. Kod mutanta takvih fluktuacija nema, grafovi su kontinuitani i stabilni i fluktuiraju za manje od 3 Å tijekom cijelog vremena trajanja simulacija od 50000 ps. Zbog toga se zaključuje da je mutant uravnoteženiji, dok je kod divljeg tipa gibanje tranzijentno.



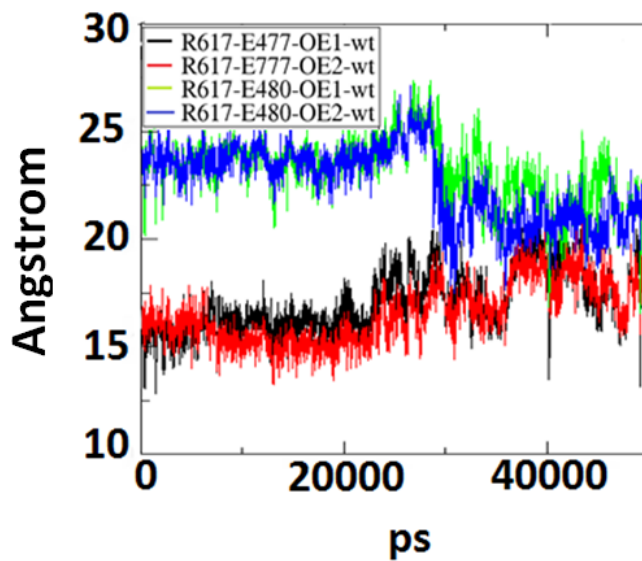
Slika 23: Prikazane su udaljenosti odabranih aminokiselinskih parova između kojih se promatra interakcija tijekom simulacija. Označeni su točni atomi pojedinih aminokiselina koji se promatraju, kao i udaljenosti u jedinici Å izmjerene pomoću alata u programu Pymol. R703H je prikazan sivo, a Keap1 je obojan prema spektru. Slika je generirana pomoću programa Pymol.



Slika 24: Promatrane udaljenosti prikazane na razini cijelog kompleksa. R703H je prikazan sivo, a Keap1 je obojan prema spektru.



Slika 25: Prikazane su promjene udaljenosti odabranih aminokiselinskih parova u ovisnosti o vremenu MD simulacija za R703H. Slika je generirana pomoću programa Grace



Slika 26: Prikazane su promjene udaljenosti odabranih aminokiselinskih parova u ovisnosti o vremenu MD simulacija za divlji tip. Slika je generirana pomoću programa Grace.

## 6 Zaključak

Na položaju 703. aminokiselinskog lanca ljudske DPP III iz arginina u histidin, napravljen je mutant R703H kako bi se provedbom MD simulacija ispitala njegova interakcija s Keap1 proteinom. Analizirana su svojstva mutanta u odnosu na divlji tip proteina i promjene u strukturi enzima uzrokovane mutacijom koje utječu na vezanje s Keap1 proteinom. Obradom rezultata MD simulacija utvrđena je interakcija R703H s Kelch domenom Keap1 proteina putem ETGE strukturnog motiva, što znači da R703H interagira s Keap1 proteinom putem istih aminokiselina kao i divlji tip. Razlika u interakciji divljeg tipa DPP III- Keap1 i R703H- Keap1 je u jačini veza koje se tijekom interakcije ostvaruju budući da su simulacije pokazale da R703H ostvaruje nešto jače interakcije s Keap1 od divljeg tipa. Nadalje, uočena je konformacijska promjena R703H u odnosu na divlji tip proteina. Statističkom analizom geometrijskih parametara utvrđeno je da R703H ima nešto zatvoreniju strukturu u odnosu na divlji tip proteina. Praćenjem koordinacije cinkovog atoma tijekom simulacija i analizom RMSD vrijednosti zaključeno je da vezanje R703H s Keap1 proteinom ne utječe na koordinaciju cinkovog iona i aktivnog mjesta enzima kao ni na stabilnost domena koje se tijekom simulacija neznatno mijenjaju kao što je to slučaj i kod divljeg tipa. Zaključak je stoga da mutant R703H putem *ETGE* strukturnog motiva također tvori stabilan kompleks sa Kelch domenom Keap1 proteina što daje osnovu za daljnja računalna i biokemijska istraživanja u kojima bi se točnije odredio afinitet vezanja R703H mutanta za Keap1 u usporedbi s afinitetom divljeg tipa. Na taj bi se način moglo utvrditi može li se pojavnost mutacije povezati sa povećanjem transkripcijske aktivnosti Nrf2 gena u tumorskim stanicama, koja vodi većoj otpornosti stanica na elektrofilne i oksidativne stresove što znači i smanjenu efikasnosti uništavanja tumorskih stanica kemoterapijom.

## **7 Zahvale**

Zahvaljujem se voditeljici projekta prof.dr.sc. Sanji Tomić na ukazanoj prilici za izradu ovog rada i uključivanje u projekt na Institutu „Ruđer Bošković”, kao i na svim korisnim savjetima i uloženom vremenu.

Zahvaljujem i mag.biol.mol Sari Matić s Instituta „Ruđer Bošković” na velikoj pomoći pri provedbi simulacija te uloženom trudu i strpljenju.

Na kraju, izrada ovog rada nebi bila moguća bez preporuke, savjeta i suradnje doc.dr.sc. Selme Supek kojoj ovim putem posebno zahvaljujem.



## 8 Literatura

- [1] W.F. van Gunsteren, H.J.C. Berendsen, *Computer Simulation of Molecular Dynamics: Methodology, Application, and Perspective in Chemistry*, 994.-995.
- [2] A.W. Gotz, M.J. Williamson, D. Xu, D. Poole, S. Le Grand, R.C. Walker, Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born, *Journal of Chemical Theory and Computation* (2012), 154.
- [3] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. Di Nola, J.R. Haak, *Molecular dynamics with coupling to an external bath*, *The Journal of Chemical Physics* (1984), 3685.-3686.
- [4] E. Kansanen, H.-K. Jyrkkanen, A.-L. Levonen, *Activation of stress signaling pathways by electrophilic oxidized and nitrated lipids*, *Free Radical Biology and Medicine* 52 (2012), 973–982.
- [5] B. E. Hast, D. Goldfarb, K. M. Mulvaney, M. A. Hast, P. F. Siesser, F. Yan, D.N. Hayes, M. B. Major, *Proteomic analysis of ubiquitin ligase KEAP1 reveals associated proteins that inhibit NRF2 ubiquitination.*, *Cancer Res.* 73 (2013), 2199.-2210.
- [6] M. Gundić, A. Tomić, R.C. Wade, M. Matovina, Z. Karačić, S. Kazazić, S. Tomić, *Human DPP III – Keap1 Interactions: a Combined Experimental and Computational Study*, *Croat. Chem. Acta* (2016), 217.- 227.
- [7] D. Agić, *New inhibitors of metallopeptidases of the family M49*, doktorska disertacija (2013), 5.
- [8] R. Singh, M. Saha, *Identifying structural motifs in proteins*, *Pacific Symposium of Biocomputing* (2003), 1.
- [9] Š. Šimaga, D. Babić, M. Osmak, J. Ilić-Forko, Lj. Vitale, D. Miličić, M. Abramić, *Dipeptidyl peptidase III in malignant and non-malignant gynaecological tissue* (1998), *European Journal of Cancer*, 402.-403.

- [10] P. Kumar Baral, N. Jajčanin-Jozić, S. Deller, P. Macheroux, M. Abramić, K. Gruber, The First Structure of Dipeptidyl-peptidase III Provides Insight into the Catalytic Mechanism and Mode of Substrate Binding, *The Journal of Biological Chemistry* (2008), 1.
- [11] S.C. Prajapati, S.S. Chauhan, Dipeptidyl peptidase III: a multifaceted oligopeptide N-end cutter, *FESB Journal* (2011), 3262.
- [12] A. Dong, E. Dobrovetsky, A. Seitova, B. Duncan, L. Crombet, M. Sundstorm, C.H. Arrowsmith, A.M. Edwards, C. Bountra, A. Bocharov, D. Cossar, Crystal structure of human dipeptidyl peptidase III (2009)
- [13] P. Kumar, V. Reithofer, M. Reisinger, S. Wallner, T. Pavkov-Keller, P. Macheroux, K. Gruber, Substrate complexes of human dipeptidyl peptidase III reveal the mechanism of enzyme inhibition, *Nature* (2016), 2.
- [14] P. Bhandare, P. Madhavan, B. M. Rao, N. Someswar Rao, Determination of arginine, lysine and histidine in drug substance and drug product without derivatisation by using HILIC column LC technique, *Journal of Chemical and Pharmaceutical Research* (2010), 584.- 585.
- [15] A. Tomić, Primjena računalnih pristupa različitih stupnjeva složenosti u svrhu razumijevanja strukture, dinamike i aktivnosti ljudske dipeptidil-peptidaze III, *Doktorska disertacija* (2015), 142.-143.
- [16] E. Kansanen, S.M. Kuosmanen, H. Leinonen, A.L. Levonen, The Keap1-Nrf2 pathway: Mechanisms of activation and dysregulation in cancer, *Redox Biology* (2012), 45.-46.
- [17] N.F. Villeneuve, A. Lau, D.D. Zhang, Regulation of the Nrf2-Keap1 Antioxidant response by the Ubiquitin Proteasome System: An Insight Into Cullin Ring Ubiquitin Ligases (2010), 1700.-1701.
- [18] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang,

P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman, A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, *J. Comput. Chem.*(2003), 1999-2012.

[19] P. Mark, L. Nilsson, Structure and Dynamics of the TIP3P, SPC and SPC/E Water Models at 298 K, *J. Phys. Chem*, (2001), 9954.-9956.

[20] V. N. Maiorov, G. M. Crippen, Significance of Root- Mean- Square- Deviation in Comparing Three Dimensional Structures of Globular Proteins, *Journal of Molecular Biology*, (1994), 625.

## 9 Sažetak

### **Molekularno dinamičke (MD) simulacije i njihova primjena na biološki kompleks mutanta enzima dipeptidil peptidaze III i Keap1 proteina**

Lana Krajc

Molekularno dinamičke (MD) simulacije su računalne simulacije čiji se princip rada bazira na fizikalnim metodama. U najkraćim crtama, njihov princip rada sastoji se od rješavanja Newtonovih jednadžbi gibanja za dan molekularni sistem čime se dobivaju podaci za izračunavanje različitih korisnih svojstava biološkog sustava.

Dipeptidil peptidaza III (DPP III) klasificira se kao egzo-peptidaza te spada u grupu metalopeptidaza. Zanimljiva je za ispitivanja jer je zabilježena u različitim tkivima kod ljudi. Posebno, uočena je povećana aktivnost tog proteina i nekih njegovih mutacija u nekim tipovima raka kod ljudi. Pokazano je da se DPP III natječe sa Nrf2 transkripcijskim faktorom u vezanju za Kelch domenu Keap1 proteina putem ETGE strukturnog motiva i da na taj način utječe na aktivaciju Nrf2 faktora odnosno na njegovu aktivnost u kancerogenim tkivima, što je zanimljivo za ispitivanja budući da Nrf2-Keap1 signalizacijski put standardno djeluje kao mehanizam obrane stanice od kancerogena.

U ovom radu je po prvi puta napravljena računalna simulacija vezanja mutanta DPP III proteina pod nazivom R703H s Keap1 proteinom. Glavni cilj bio je primjenom MD simulacija na kompleksu R703H i Keap1 proteina pokazati primjenu fizikalnih metoda na biološke sisteme, utvrditi detalje interakcije kompleksa R703H- Keap1 i napraviti usporedbu rezultata sa divljim tipom DPP III proteina. Najbitniji rezultat dobiven simulacijama jest da se R703H, kao i divlji tip proteina, povezuje i interagira s Keap1 proteinom, i to sa njegovom Kelch domenom a putem ETGE strukturnog motiva kao i divlji tip. Nadalje, uočene su neke promjene u strukturi R703H u odnosu na divlji tip proteina- statističke analize ukazuju da je R703H nešto zatvorenije strukture. Također, utvrđeno je da vezanje Keap1 proteina s R703H ne utječe na stabilnost pojedinih domena samog proteina niti na koordinaciju cinkovog iona. Može se dakle zaključiti da mutant također tvori stabilan kompleks sa Kelch domenom Keap1 proteina što daje bazu za daljnja istraživanja u kojim se točno slučajevima R703H veže s Keap1 proteinom te na taj način onemogućuje formaciju Nrf2-Keap1 kompleksa, te na njegovu prisutnost i ulogu u različitim tipovima raka kod ljudi.

**Ključne riječi:** molekularno dinamičke (MD) simulacije / 'protein-protein interakcije' / dipeptidil peptidaza III/ mutant R703H / Keap1 protein

## 10 Summary

### **Molecular dynamics (MD) simulations and their application to mutant dipeptidyl peptidase and Keap1 protein complex**

Lana Kralj

Molecular dynamics (MD) simulations are computer simulation that work on physical principles and methods. The method of molecular dynamics consists of solving Newton's equations of motion for given biological system which results in trajectories for all atoms in the system. These results are then used for obtaining different useful properties of simulated system.

Dipeptidyl peptidase III (DPP III) is classified as exopeptidase and belongs to the zinc metallopeptidase group. This protein is interesting for researching because its presence, as well as presence of its mutations, was observed in different healthy and cancerogenous human tissues. It was determined in earlier researches that DPP III competes with Nrf2 transcription factor for binding to Keap1 protein via an ETGE motif which influences Nrf2 activation and thereby its activity in cancer. This fact makes DPP III- Keap1 complex interesting because the role of Nrf2-Keap1 signaling pathway in organism is to protect cells from cancerogens.

In this experiment, MD simulations of binding DPP III mutant, called R703H, with Keap1 protein was made for the first time. Main goals of this experiment were to apply MD simulations to the R703H- Keap1 complex to show the use of physical methods for prediction of the dynamics of real biological system, to elucidate the molecular details of interaction R703H with Keap1 protein and to compare the results with the results obtained for the wild type of DPP III protein from previous experiment. One of the most important results is that simulations predicted that R703H binds and interacts with Kelch domain of Keap1 protein via an ETGE structure motif - same as the wild type. Also, some differences in structure between mutant R703H and the wild type of DPP III protein were observed. Statistical analysis suggested that R703H has more closed structure than the wild type of DPP III. Also, it was determined that formation of R703H-Keap1 complex does not influence the structure of DPP III domains and the zinc ion coordination. It can be concluded that although there are some differences in structures between R703H and wild type of DPP III protein, R703H also forms stable complexes with the Keap1 protein. This fact provides

the basis for further experiments to investigate in which cases will R703H bind for Keap1 instead of Nrf2 factor, as well as to examine its role in different types of human cancers.

**Key words:** molecular dynamics (MD) simulations / 'protein- protein interactions'/ dypeptidil peptidase III/ mutant R703H/ Keap1 protein