

UNIVERZITET U BEOGRADU

FARMACEUTSKI FAKULTET

Jelena N. Randelović

**Dizajniranje peptidnih i nepeptidnih  
inhibitora proteinskih interakcija  
ciklin-zavisne kinaze 9 i ciklina T1  
primenom računarskih metoda**

doktorska disertacija

Beograd, 2013

UNIVERSITY OF BELGRADE

FACULTY OF PHARMACY

Jelena N. Randelović

**Design of peptide and small molecule  
inhibitors of cyclin dependent kinase 9 /  
cyclin T1 protein-protein interaction  
using computational methods**

Doctoral Dissertation

Belgrade, 2013

**Mentori:**

dr Vladimir Savić , redovni profesor, Farmaceutski fakultet, Univerzitet u Beogradu

dr Slavica Erić, vanredni profesor, Farmaceutski fakultet, Univerzitet u Beogradu

**Komisija:**

dr Vladimir Savić , redovni profesor, Farmaceutski fakultet, Univerzitet u Beogradu

dr Slavica Erić, vanredni profesor, Farmaceutski fakultet, Univerzitet u Beogradu

dr Ivan Juranić, redovni profesor, Institut za hemiju, tehnologiju i metalurgiju, Univerzitet u Beogradu

dr Zorana Tokić Vujošević, vanredni profesor, Farmaceutski fakultet, Univerzitet u Beogradu

dr Mire Zloh, profesor, School of Pharmacy, University of Herdfordshire, UK.

Datum odbrane:

*Familiji, što je moja,  
Mentorima i kolegama, za sve divne dane  
Kejti, Majklu i Aiku, za sve naše igre,  
Bucku i Setsuni, za silikonske snove,  
i Svetu, što postoji,*

Hvala.

## Dizajniranje peptidnih i nepeptidnih inhibitora proteinskih interakcija ciklin-zavisne kinaze 9 i ciklina T1 primenom računarskih metoda

Cilj ove doktorske disertacije je razvoj direktnih inhibitora interakcija u kompleksu CDK9 / ciklin T1. Ciklin zavisna kinaza 9 u kompleksu sa ciklinom T1 čini katalitičko jezgro P-TEFb kompleksa. Inhibicija funkcije P-TEFb kompleksa je od značaja u terapiji više patoloških stanja kod ljudi: nekih tipova maligniteta, HIV-1 infekcije, hipertrofije miokarda, hroničnih inflamatornih bolesti. Postojeći inhibitori P-TEFb funkcije imaju ograničenu selektivnost dejstva, koju je moguće poboljšati primenom direktnih inhibitora CDK9 / ciklin T1 interakcije.

U ovom istraživanju urađena je *in silico* studija P-TEFb kompleksa. Ispitana je struktura, hidratisanost i dinamika vezivnih površina kompleksa. Utvrđene su energetski najznačajnije *hot spot* aminokiseline u strukturi kompleksa i upotrebljene za dalji dizajn inhibitora. Struktura slobodne CDK9, neophodna za dizajn liganada je, u nedostatku eksperimentalno utvrđene strukture, simulirana klasičnom i ubrzanom molekularnom dinamikom. Ispitana je sposobnost vezivne površine kompleksa da interaguje sa ligandima, kao i postojanje, otvaranje i postojanost privremenih hidrofobnih džepova na vezivnoj površini CDK9, pogodnih za vezivanje liganada.

Dobijene informacije o vezivnim površinama kompleksa CDK9 / ciklin T1, upotrebljene su za dizajn peptidnih i nepeptidnih inhibitora interakcije. Peptidni inhibitori su dizajnirani po ugledu na strukturu ciklina T1, ali i *de novo* dizajnom iz aminokiselina. Peptidne strukture sa najpovoljnijom energijom vezivanja za CDK9, iskorišćene su u razvoju nepeptidnih inhibitora. Predložene su strukture potencijalnih nepeptidnih inhibitora interakcije CDK9 / ciklin T1.

U okviru ovog istraživanja urađeno je i *ab initio* ispitivanje prelaznih stanja intramolekulske Heck-ove ciklizacije alilnih alkohola i predloženo je objašnjenje za eksperimentalno zapaženu regioselektivnost procesa.

Ključne reči: protein-protein interakcije, CDK9/ciklin T1, molekularna dinamika

Naučna oblast: Farmaceutske nauke

Uža naučna oblast: Farmaceutska hemija

UDK broj: 615.2/3 : 577.112.389 (043.3)    577.112.389 : 004.414.23 (043.3)

Design of peptide and small molecule inhibitors of cyclin dependent kinase 9 / cyclin T1 protein-protein interaction using computational methods

The goal of this doctoral dissertation is the development of direct inhibitors of CDK9 / cyclin T1 protein-protein interaction. Cyclin dependent kinase 9 in complex with cyclin T1 forms the catalytic core of P-TEFb protein complex. Inhibition of P-TEFb function is of significance in the therapy of several pathological states in humans: some malignancies, HIV-1 infection, cardiac hypertrophy, chronic inflammatory diseases. Existing inhibitors of P-TEFb function have a limited selectivity of action, which could be improved with the direct inhibitors of CDK9 / cyclin T1 interaction.

An *in silico* study of the P-TEFb complex was conducted in this research. The structure, hydration and dynamics of the complex binding surfaces were examined. Energetically most significant amino acids, hot spots, were determined and used in the design of inhibitors. The structure of the free CDK9, needed for the ligand design, was, in absence of an experimentally determined structure, simulated using classical and accelerated molecular dynamics. The ability of complex binding surfaces to interact with ligands was examined, as was the existence, opening and persistence of transient hydrophobic pockets on CDK9 interface surface, suitable for ligand binding.

Information obtained of the CDK9 / cyclin T1 complex binding surfaces was used in the design of peptide and small molecule inhibitors of this interaction. Peptide ligands were designed using cyclin T1 as a template structure, as well as using *de novo* design from amino acids. Peptide structures with the most favourable binding energy to CDK9 were utilised in the design of small molecule inhibitors. Structures of potential small molecule inhibitors of CDK9 / cyclin T1 interaction were proposed.

An *ab initio* study of transition states in an intramolecular Heck cyclisation of allyl alcohols was also conducted in this research. An explanation for the experimentally observed regioselectivity of the process was proposed.

Keywords: protein-protein interactions, CDK9 / cyclin T1, molecular dynamics

Academic expertise: Pharmaceutical Sciences

Field of academic expertise: Medicinal Chemistry

UDC number: 615.2/3 : 577.112.389 (043.3) 577.112.389 : 004.414.23 (043.3)

## Sadržaj

1. Uvod .....	1
1.1 Protein-protein interakcije .....	2
1.1.1 Uvod .....	2
1.1.2 Strukturne osobine protein-protein interakcija .....	5
1.1.3 Terapeutski značaj protein-protein interakcija .....	13
1.2 P-TEFb kompleks .....	21
1.2.1 Građa i funkcija P-TEFb kompleksa .....	21
1.2.2 Patološki značaj P-TEFb kompleksa .....	24
1.2.3 Inhibicija aktivnosti P-TEFb kompleksa .....	27
1.3 Protein-protein interakcije prisutne u P-TEFb kompleksu .....	34
2. Rezultati i diskusija .....	38
2.1 Inhibicija CDK9 / ciklin T1 interakcije .....	38
2.1.1 Studija P-TEFb .....	38
2.1.1.1 Priprema kompleksa .....	38
2.1.1.2 Analiza vezivnih površina .....	42
2.1.1.3 Analiza fleksibilnosti CDK9 .....	57
2.1.2 Dizajn peptidnih inhibitora .....	69
2.1.2.1 Dizajn peptida zasnovan na interakcijama u P-TEFb .....	69
2.1.2.2 Dizajn peptida iz fragmenata .....	72
2.1.2.3 Doking peptida .....	74
2.1.2.4 Procena energije vezivanja peptida MM-PBSA metodom .....	76
2.1.2.5 Procena energije vezivanja peptida SMD metodom .....	82
2.1.3 Dizajn nepeptidnih inhibitora .....	89
2.1.3.1 Pretraživanje farmakofora .....	89
2.1.3.2 Doking liganada .....	92
2.1.3.3 Procena energije vezivanja liganada .....	95
2.1.4 Dizajn inhibitora zasnovan na građi P-TEFb/Tat kompleksa .....	105
2.1.5 Zaključak .....	107
2.2 <i>In silico</i> analiza regioselektivnosti intramolekulske Heck-ove ciklizacije .....	108

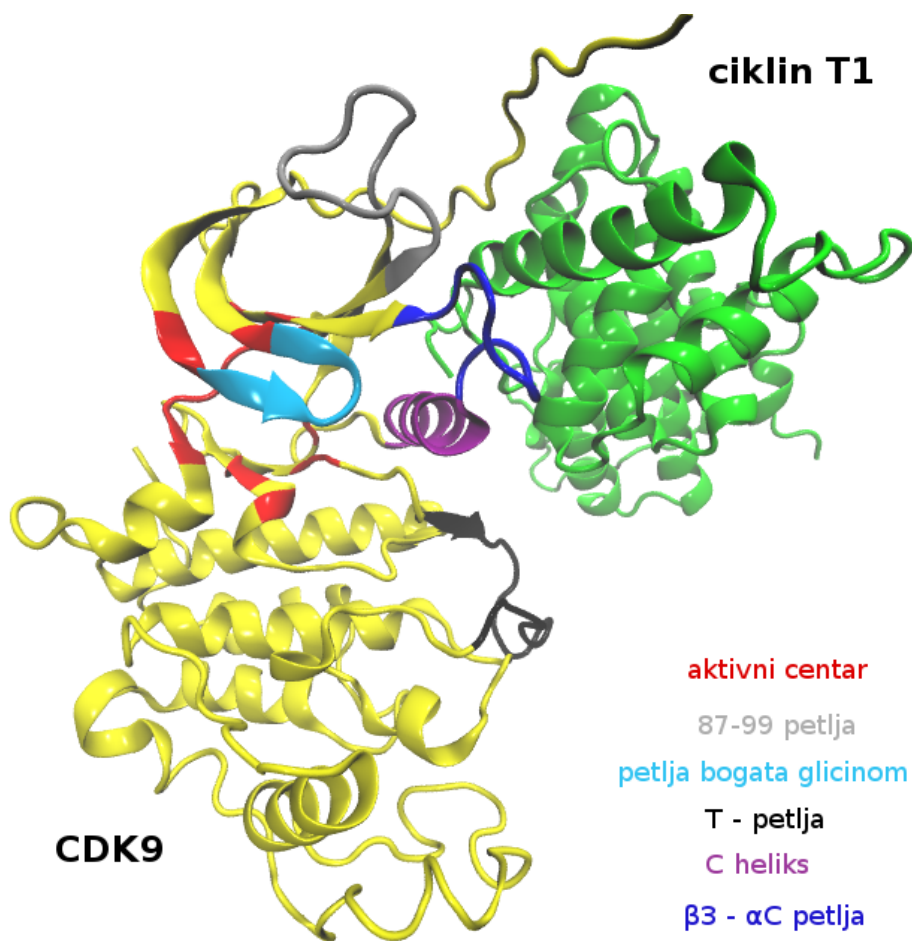
2.2.1 DFT studija prelaznih stanja procesa .....	108
2.2.2 Zaključak .....	113
3. Zaključak .....	114
4. Eksperimentalni deo .....	116
4.1 Molekularna dinamika .....	116
4.1.1 Priprema sistema .....	116
4.1.2 MD simulacije .....	120
4.2 Analiza vezivnih površina CDK9/ciklin T1 kompleksa .....	123
4.2.1 Analiza strukturnih kontakata .....	123
4.2.2 Hidratisanost vezivnih površina .....	123
4.2.3 Analiza <i>hot spot</i> aminokiselina .....	124
4.2.4 Analiza fleksibilnosti CDK9 .....	125
4.2.5 Analiza dinamike formiranja hidrofobnih džepova .....	127
4.3 Dizajn peptida iz fragmenata .....	129
4.4 Pretraživanje farmakofora .....	130
4.5 Doking liganada .....	131
4.5.1 Odabir i priprema struktura za doking .....	131
4.5.2 Doking procedura .....	132
4.5.3 Rangiranje poza liganada .....	132
4.6 MM-PBSA i MM-GBSA procedure .....	133
4.7 DFT procedure .....	135
5. Literatura .....	137
6. Prilozi .....	158
7. Biografija autora .....	165
8. Izjave .....	166



## 1. Uvod

*In silico* metode igraju značajnu ulogu u procesu otkrića i razvoja novih lekova. Primena računara u svim fazama ovog procesa omogućava brži i jednostavniji razvoj. Posebno je značajna uloga koju računarske metode igraju u otkriću novih lekova zasnovanom na strukturi njihovih ciljnih mesta (eng. *structure based design*), pošto su danas *in silico* metode jedini način za predstavljanje receptora lekova u atomskoj rezoluciji. Iako su eksperimentalne *in vitro* i *in vivo* metode nezamenjive kao krajnja potvrda rezultata razvoja novog leka, *in silico* metode se široko primenjuju i dopunjavaju saznanja dobijena eksperimentima.

U ovom istraživanju biće prikazan *in silico* dizajn inhibitora aktivnosti P-TEFb proteinskog kompleksa, inhibicijom interakcija u CDK9 / ciklin T1 kompleksu koji čini njegovo strukturno i funkcionalno jezgro (Slika 1).



**Slika 1:** CDK9 / ciklin T1 kompleks sa značajnim elementima građe

## 1.1 Protein-protein interakcije

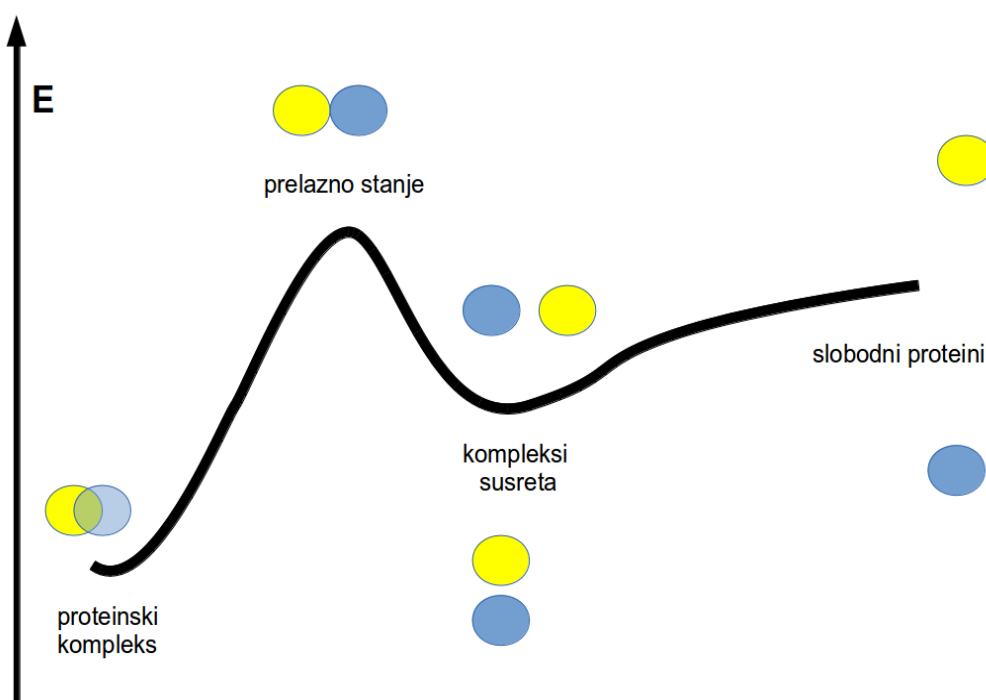
### 1.1.1 Uvod

Proteini su među najznačajnijim funkcionalnim molekulima u ćelijama. Svoju ulogu u regulaciji, kako fizioloških tako i patoloških procesa, proteini obavljaju kroz interakcije sa drugim biomolekulima, ali i sa malim organskim jedinjenjima. U živim sistemima, proteini su povezani u kompleksne mreže interakcija sa drugim proteinima, formirajući proteinske komplekse različite stabilnosti i složenosti. Procenjeno je da humani proteini stupaju u oko 650000 protein-protein interakcija (1).

Međuproteinski kompleksi mogu biti sastavljeni od više istih (homokompleksi) ili različitih (heterokompleksi) proteinskih subjedinica i javljaju se kako među slobodnim proteinima u rastvoru, tako i među membranskim receptorima. Vreme formiranja proteinskih kompleksa se meri u nanosekundama (2). U ćeliji, nivo formiranja određenog proteinskog kompleksa se reguliše ekspresijom njegovih subjedinica, njihovom kolokalizacijom, ali i uslovima sredine, kao što su prisustvo kofaktora ili pH. Konstante disocijacije ( $K_d$ ) proteinskih kompleksa su u oblasti  $10^{-4}$ – $10^{-14}$  M što odgovara slobodnim energijama vezivanja od 6–19 kcal/mol (3). U proces prepoznavanja molekula proteina prilikom građenja proteinskih kompleksa mogu biti uključeni mehanizmi konformacione selekcije, ali i indukovanog prilagođavanja (eng. *induced fit*) (4). Konformaciona selekcija podrazumeva da su u populaciji konformacija slobodnog proteina već zastupljene vezivne konformacije koje odgovaraju stanju proteina u kompleksu i da prilikom formiranja kompleksa samo dolazi do pomeranja dinamičke ravnoteže među konformacijama. Indukovano prilagođavanje se odnosi na promenu konformacija proteina nakon stupanja u kontakt sa drugim subjedinicama kompleksa.

Da bi se formirao proteinski kompleks neophodno je da dođe do sudara i prepoznavanja proteinskih subjedinica, prisutnih u niskim koncentracijama u ćeliji (4). Prvi korak u prepoznavanju predstavlja kolokalizacija u prostoru, omogućena elektrostatičkim usmeravanjem proteina. Privremeni proteinski kompleks koji se formira samo na osnovu elektrostatičke podudarnosti subjedinica naziva se kompleks susreta (eng. *encounter complex*). Kompleks susreta nije kompleks u pravom smislu te

reći, jer ga čini skup različitih orijentacija i konformacija subjedinica i u njemu još uvek nije došlo do desolvatacije vezivnih površina proteina (5). Zadržavanjem subjedinica kompleksa u relativnoj blizini, kompleks susreta smanjuje dimenzionalnost problema pronalaženja prave orijentacije molekula za vezivanje. Kompleks susreta je od konačnog oblika proteinskog kompleksa odvojen dodatnim prelaznim stanjem, u kome dolazi do desolvatacije vezivnih površina i uspostavljanja specifičnih interakcija između proteina (5) (Slika 2).



**Slika 2:** Šematski prikaz nastanka proteinskog kompleksa

Nije neophodno da po formiranju kompleksa susreta dođe do građenja proteinskog kompleksa, subjedinice kompleksa mogu da disosuju nazad u rastvor. Moguće je da kod nekih proteina uključenih u prenos elektrona kompleks susreta i predstavlja funkcionalni oblik njihove interakcije (6).

Po dužini trajanja asocijacije subjedinica, proteinski kompleksi mogu se podeliti na trajne (obligatorne) i privremene (tranzitorne) komplekse. U trajnim kompleksima, proteinske subjedinice nisu stabilne kao monomeri u rastvoru i postoje samo u kompleksu. Ovakvi kompleksi se formiraju neposredno po sintezi proteinskih lanaca u procesu formiranja konačne strukture proteina. Po formiranju, trajni kompleksi ostaju stabilni sve do svoje degradacije u ćelijama. Interakcije u njima najčešće imaju

strukturnu ulogu (formiranje kvaternerne strukture funkcionalnog proteinskog kompleksa). Privremeni kompleksi se formiraju između monomernih jedinica koje postoje samostalno u rastvoru usled dovoljne polarnosti svoje molekularne površine i često mogu obavljati svoju funkciju u ćeliji i kao monomeri. Po formiranju privremenog proteinskog kompleksa dolazi do modulacije funkcije monomera, često usled konformacionih promena monomera pri vezivanju. Privremeni kompleksi su značajno manje stabilni od trajnih. Subjedinice privremenih kompleksa disosuju i nastavljaju da postoje samostalno u rastvoru. Stanje u kome će se naći privremeni proteinski kompleksi neretko zavisi od trenutnih potreba ćelije: oni se stvaraju da bi se obavila određena vremenski ograničena funkcija i zatim disosuju. Kako su vezivne površine proteinskih subjedinica u privremenim proteinskim kompleksima u monomernom stanju u direktnom kontaktu sa vodom, one su polarnije od vezivnih površina proteina u trajnim kompleksima (7) koje se ne izlažu rastvaraču. Dodatna razlika između trajnih i privremenih proteinskih kompleksa je u fleksibilnosti njihovih vezivnih površina, tj. u stepenu promena koje se javljaju na vezivnim površinama po formiranju kompleksa. Pokazano je da vezivne površine proteina koji grade trajne komplekse podležu značajnijim promenama konformacije u odnosu na vezivne površine proteina u privremenim kompleksima (8). Razlog je energetska i entropijska cena konformacionih promena. Viši energetska i entropijska zahtevi za zauzimanje vezivnih konformacija pri vezivanju su prisutni kod trajnih kompleksa, jer se formiraju samo jednom. Privremeni kompleksi su ograničeno stabilni – zauzimanjem konformacije vezivne površine koja podseća na onu u kompleksu oni olakšavaju prolazak kroz ponovljene cikluse formiranja i disocijacije kompleksa čiji su deo (8).

Zadržavanje odlika vezivne konformacije u strukturama slobodnih proteinskih subjedinica je, u određenoj meri, prisutno kod većine proteinskih kompleksa. Informacije o rotamernom stanju bočnih nizova aminokiselina sa vezivne površine kompleksa prisutne su u konformacijama proteina u rastvoru (9). Moguće je da se objašnjenje ove pojave nalazi u manjoj konformacionoj slobodi bočnih nizova aminokiselina sa vezivnih površina proteinskih kompleksa, u poređenju sa ostatkom površine proteina (10,11). Smanjena konformaciona sloboda aminokiselina sa površina proteina uključenih u građenje kompleksa smanjuje verovatnoću nastanka interakcija koje nisu prisutne u nativnom kompleksu (10). Samim tim, povećava se specifičnost

interakcija između proteina. Pozicija čak 78% aminokiselina sa različitih vezivnih površina u proteinskim kompleksima je uspešno modelovana sa tačnošću od 1Å koristeći konformaciju slobodnog proteina kao bazu i dodatna dva najverovatnija rotamera po aminokiselini (9).

Za sve proteinske komplekse, bilo privremene, bilo trajne, zajedničke su određene strukturne specifičnosti vezivnih površina proteinskih subjedinica i interakcija među njima.

### 1.1.2 Strukturne osobine protein-protein interakcija

Proteini međusobno interaguju svojom molekularnom površinom. Nakon sinteze proteinskih lanaca i tokom formiranja viših nivoa strukture proteina, proteinski molekuli prilagođavaju svoju molekularnu površinu svom okruženju, pa su proteinska jezgra pretežno hidrofobna, njihova površina pretežno polarna, a vezivne površine kojima grade interakcije sa drugim proteinima imaju polarnost između ove dve vrednosti (12). Na vezivnim površinama se često uočavaju dve oblasti: polarni obod, koji je po sastavu sličan ostatku površine proteina i nepolarni centar interakcije, koji sadrži veći procenat hidrofobnih aminokiselina (13).

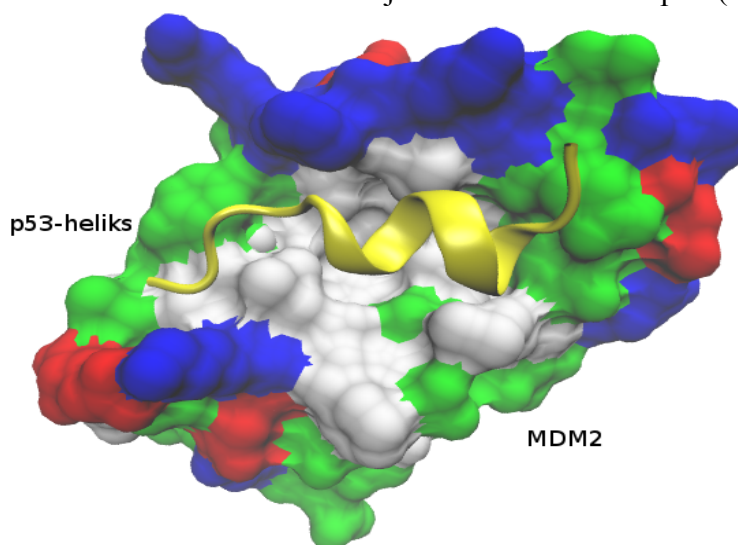
Hidrofobnost vezivnih površina je najizraženija kod homooligomernih, obligatornih kompleksa, dok su vezivne površine tranzitornih heterodimera i antitela po polarnosti najbližije površini proteina (12). Uz polarnost, molekularne površine kojima proteini međusobno interaguju, odlikuju i:

- veličina
- oblik
- tipovi interakcija
- topologija

Veličina molekularne površine uključene u protein-protein interakciju zavisi od više faktora: tipa kompleksa koji nastaje, veličine samih proteina, broja subjedinica. U proseku, veličina vezivne površine u proteinskom kompleksu je oko 2000 Å<sup>2</sup> (750–1500 Å<sup>2</sup> po proteinskom molekulu (12)), sa rasponom od 600 do 4000 Å<sup>2</sup>, koliko dostiže u oligomernim proteinskim kompleksima. Velike površine proteinskih interakcija se javljaju u kompleksima proteaza i njihovih inhibitora, kao i u

kompleksima koji uključuju G-proteine (14). Kao poređenje, tipični biološki aktivni molekuli obično zauzimaju oko 200–800 Å<sup>2</sup> na svojim proteinskim receptorima (15). Prosečnu vezivnu površinu u protein-protein interakciji izgrađuje oko 50 aminokiselina (12). Veoma retko je čitava vezivna površina proteina u interakciji izgrađena od samo jednog segmenta primarne strukture proteina, u kontinuitetu. Uobičajeno je da vezivnu površinu izgrađuju prostorno bliske aminokiseline sa većeg broja različitih većih ili manjih segmenata primarne strukture i različite sekundarne strukture.

Po obliku, vezivne površine protein-protein interakcija imaju značajno manje izraženu konkavnost i konveksnost nego što je to slučaj sa proteinskim površinama koje su namenjene vezivanju manjih supstrata (npr. hidrofobni džepovi enzima) (7). Trajni kompleksi i homodimeri pokazuju znatno veću zakrivljenost vezivnih površina od privremenih kompleksa i heterodimera (7). Komplementarnost oblika između proteinskih partnera zavisi od tipa interakcije (7). Slično zakrivljenosti, i komplementarnost vezivnih površina je bolja kod homodimera i trajnih kompleksa, u poređenju sa privremenim kompleksima i heterodimerima. Smanjena komplementarnost vezivnih površina ostavlja više potencijalnih interakcija na raspolaganju molekulima rastvarača ili inhibitora (16). Na vezivnim površinama proteina u interakciji često postoji veći broj manjih džepova (17), a ponekad se javljaju i veći hidrofobni useci, ako u vezivanju učestvuje odgovarajući rigidni element sekundarne strukture, najčešće heliks. Dobar primer je proteinski kompleks MDM2 / p53, u kome se na strukturi MDM2 proteina nalazi hidrofobni usek u koji se smešta heliks sa p53 (Slika 3).

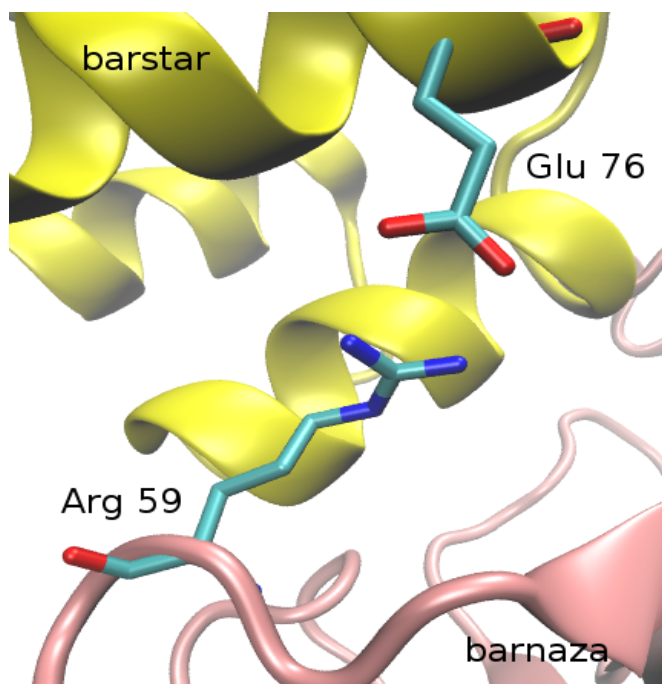


**Slika 3:** Hidrofobni usek na strukturi MDM2 proteina, u kompleksu sa heliksom iz strukture p53

Ako proteinske subjedinice kompleksa postoje samostalno u rastvoru, na njihovim vezivnim površinama se mogu stvarati privremeni džepovi (18). U otvaranju džepova na vezivnim površinama proteina učestvuje promena položaja i bočnih lanaca i polipeptidnog niza aminokiselina u blizini (19).

Svi uobičajeni tipovi molekularnih nekovalentnih interakcija zastupljeni su između proteinskih subjedinica u kompleksu. Koji tip će preovlađivati zavisi od tipa formiranog kompleksa, aminokiselinskog sastava vezivne površine i njene hidratisanosti.

Jonski parovi (soni mostovi) su relativno retka interakcija u proteinskim kompleksima. Formiraju se između terminalnih grupa kiselih i baznih aminokiselina i najveću ulogu igraju tokom elektrostatičkog usmeravanja molekula proteina u rastvoru pre samog formiranja kompleksa, kada utiču na postizanje pravilne orijentacije molekula proteina pre njihovog inicijalnog sudara. Odličan primer je proteinski kompleks barnaza / barstar (Slika 4), kome povoljne elektrostatičke interakcije vezivnih površina omogućavaju da postigne brzinu formiranja i do  $10^9 \text{ M}^{-1}\text{s}^{-1}$ , zavisno od jonske jačine okruženja (20). Usmeravajući efekat naelektrisanih grupa aminokiselina proteina je analogan onome koji je prisutan između fosfatnih grupa DNK i pozitivno naelektrisanih regiona proteina koji interaguju sa njima.



**Slika 4:** Jonski par na vezivnoj površini kompleksa barnaza/barstar

Jon-dipol i dipol-dipol interakcije se ostvaruju između molekula aminokiselina, ali i sa molekulima rastvarača.

Vodonične veze u protein-protein interakcijama mogu biti direktne ili indirektne, preko molekula vode i grade ih kako bočni nizovi, tako i poliamidni lanac proteina. U proseku se javlja podjednak broj direktnih i indirektnih vodoničnih veza na vezivnoj površini (21). Postoje značajne razlike u broju vodoničnih veza na vezivnoj površini različitih kompleksa (0 – 46), a u proseku je prisutna jedna H-veza na svakih 100 Å<sup>2</sup> (22).

Van der Valsove interakcije su značajne za integritet hidrofobnih regiona proteinskih vezivnih površina. Iako značajno slabije od polarnih interakcija, prisutne su u svim proteinskim kompleksima i njihov uticaj raste sa porastom veličine površine uključene u vezivanje. U trajnim, homodimernim i homooligomernim kompleksima sa pretežno hidrofobnim vezivnim površinama, Van der Valsove interakcije su glavna sila koja omogućava vezivanje. U proseku, oko 60% vezivne površine u proteinskim kompleksima čine hidrofobne oblasti, mada procenat varira u zavisnosti od veličine kompleksa (23). Procenat hidrofobne površine u proteinskim kompleksima je evolutivno ograničen, da bi se sprečio nastanak nespecifičnih interakcija i agregacija proteina u rastvoru (23). Uprkos tome, Van der Valsove interakcije značajno doprinose stabilnosti proteinskih kompleksa, što je dobro ilustrovano činjenicom da je prosečna vezivna površina u proteinskom kompleksu 85 – 90% zaštićena od prisustva vode (24).

Kada su molekuli vode prisutni na vezivnoj površini proteinskog kompleksa oni obično grade vodonične veze sa karbonilnim kiseonikom polipeptidnog niza, ili naelektrisanim grupama iz bočnih lanaca aminokiselina (24). U proseku se nalazi oko 10 molekula vode na svakih 1000 Å<sup>2</sup> vezivne površine (24). Ograničena komplementarnost vezivnih površina proteinskih kompleksa može biti premošćena vodoničnim vezama sa molekulima vode zarobljenim u šupljinama vezivne površine.

Proteini interaguju velikom molekularnom površinom koja bi, teoretski, mogla da pruži podjednak doprinos njihovom vezivanju u svakom svom segmentu. Međutim, vezivne površine proteina uključene u međuproteinske interakcije su, u topološkom smislu, visoko organizovane i odlikuju se modularnom građom (25). Energija vezivanja proteina u kompleksu nije podjednako raspoređena na sve aminokiseline koje grade vezivnu površinu, već je lokalizovana na mali broj značajnih aminokiselina, dok ostatak



vezivne površine obezbeđuje pogodno okruženje za interakciju ovih aminokiselina i učestvuje u inicijalnom prepoznavanju i orijentaciji proteinskih subjedinica.

Na vezivnoj površini proteina u kompleksu se izdvaja 3–9 grupa važnih aminokiselina koje su međusobno nezavisne i aditivno doprinose energiji vezivanja. U okviru jedne modularne grupe aminokiselina vezivanje je kooperativno, tj. mutacija u jednoj modularnoj grupi značajno utiče na vezivanje te grupe. Modularna građa je prisutna kod velike većine proteinskih kompleksa, osim kod onih u kojima nije postojala evolutivna mogućnost da se ona razvije mutacijama, tj. kod kompleksa uključenih u preživljavanje ćelije, kod kojih svaka mutacija dovodi do smrti ćelija. Značaj strukturne organizacije u građi vezivnih površina proteinskih kompleksa može se ilustrovati činjenicom da je ponekad dovoljna jedna ili dve mutacije na vezivnoj površini jedne od proteinskih subjedinica da se potpuno omete formiranje kompleksa. Kod kompleksa CDK9 / ciklin T1 mutacije K93L/E96K na vezivnoj površini ciklina T1 dovode do gubitka njegove sposobnosti da veže CDK9 *in vivo* (26).

Energetski značajne aminokiseline na vezivnim površinama proteinskih kompleksa koje ostvaruju ključne interakcije za vezivanje, nazivaju se *hot spot* aminokiselinama (eng. *hot spot* – ključno mesto, žarište). Otkrivene su 1995. u eksperimentalnoj studiji na hormonu rasta (27), kada se pokazalo da je mali broj aminokiselina sa centralnog dela vezivne površine kompleksa bio odgovoran za stabilnost kompleksa. Ovo zapažanje je kasnije potvrđeno na velikom broju analiziranih proteinskih kompleksa.

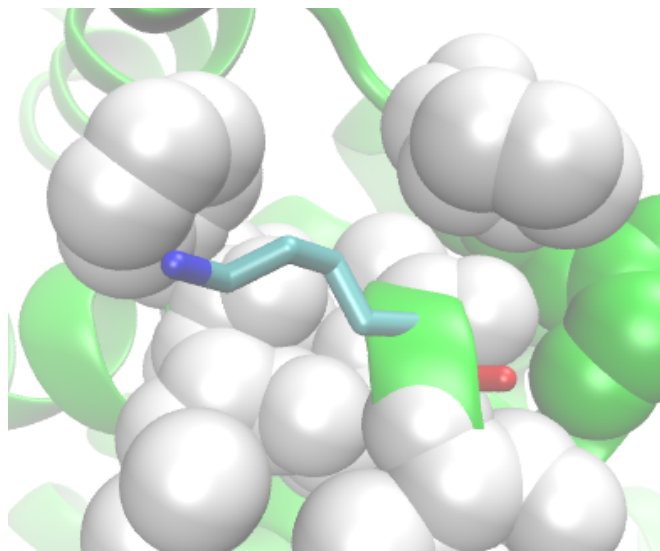
Bogan i Torn (28) su pokazali da se kao *hot spot* aminokiseline najčešće javljaju tirozin (Tyr), arginin (Arg) i triptofan (Trp), dok su leucin (Leu), valin (Val), serin (Ser), treonin (Thr) i metionin (Met) retko prisutne. Njihova studija je takođe pokazala da je izolovanost od rastvarača potreban, ali ne i dovoljan uslov da neka aminokiselina na vezivnoj površini bude *hot spot*. *Hot spot* aminokiseline se ili pružaju u energetski povoljne džepove na površini proteina, ili ulaze u strukturu istih (29). Ako protein gradi komplekse sa različitim proteinima, korišćenjem istih oblasti svoje vezivne površine, ista *hot spot* aminokiselina može biti prisutna u različitom hemijskom okruženju. Ovo zapažanje se delimično može objasniti činjenicom da su *hot spot* aminokiseline i njihovo okruženje specifično podešeni za vezivanje određenih kombinacija polarnih grupa u hidrofobnom okruženju, bez obzira na njihovu prirodu i na to da li one potiču sa

bočnih nizova aminokiselina drugih proteina ili iz struktura malih organskih molekula (30). Uz to, *hot spot* aminokiseline i aminokiseline oko njih (radijus oko 6Å) poseduju sposobnost prilagođavanja strukturama interakcionih partnera, kroz fleksibilnost svojih bočnih nizova (30).

Zastupljenost i raspored *hot spot* aminokiselina na vezivnim površinama različitih dimenzija se razlikuje. Prosečan, bazalni doprinos slobodnoj energiji vezivanja, u kompleksima sa vezivnom površinom većom od 2000 Å<sup>2</sup> je oko 3–4 cal/molÅ<sup>2</sup> (23). Proteinski kompleksi sa znatno manjom površinom pokazuju veću energetska gustinu površine (do 13 cal/molÅ<sup>2</sup>), što gotovo odgovara energetskom značaju triptofana kao *hot spot* aminokiseline (14 cal/molÅ<sup>2</sup>) (23). Ovo ukazuje da u kompleksima sa manjim vezivnim površinama *hot spot* aminokiseline zauzimaju značajan procenat površine i nose najveći deo energije vezivanja. Sa druge strane, kompleksi sa velikim vezivnim površinama mogu sebi da priušte da ne povećavaju broj *hot spot* aminokiselina u odnosu na manje komplekse, jer jedan deo energije vezivanja dobijaju sumiranjem velikog broja slabih interakcija (23).

Za utvrđivanje *hot spot* aminokiselina danas se najčešće koristi sekvencijalna mutacija aminokiselina od interesa do alanina, bilo eksperimentalno, bilo *in silico* (31). Mutacija aminokiselina do alanina utiče kako na kompleks, tako i na protein slobodan u rastvoru. Da bi se uspostavila veza između mutacije i promena u slobodnoj energiji kompleksa, neophodno je utvrditi da mutacija ne utiče značajno na slobodan protein (32). Dodatan razlog za pažljivu interpretaciju rezultata mutacija do alanina jeste činjenica da mutacija jedne aminokiseline neće uvek ispoljiti posledice samo u svom neposrednom okruženju, kroz promenu molekularnih interakcija. Efekti mutacija se mogu pojaviti i na udaljenim oblastima vezivne površine, ako je mutacija poremetila prostornu organizaciju aminokiselina na vezivnoj površini, ili dovela do promena konformacija ili fleksibilnosti drugih delova vezivne površine (33). Takođe, razlike u slobodnoj energiji vezivanja koje se dobijaju prilikom mutacija aminokiselina do alanina uzimaju u obzir samo energetska doprinos uklonjenih supstituenata bočnog niza aminokiselina, a ne i polipeptidnog niza, koji takođe može biti od značaja u vezivanju (34). Kako je procentualni udeo polipeptidnog niza u energetskom doprinosu najveći kod malih aminokiselina, one su i najređe prisutne kao *hot spot* aminokiseline (33). Kao rezultat svega navedenog, koriste se i alternativne metode za utvrđivanje aminokiselina

od značaja za interakciju, a samim tim i za inhibiciju iste. Neke od tih metoda su procena evolutivne očuvanosti ili izloženosti aminokiseline rastvaraču ili procena vezivanja molekularnih fragmenata u blizini aminokiselina od interesa, bilo eksperimentalno, ili *in silico*.



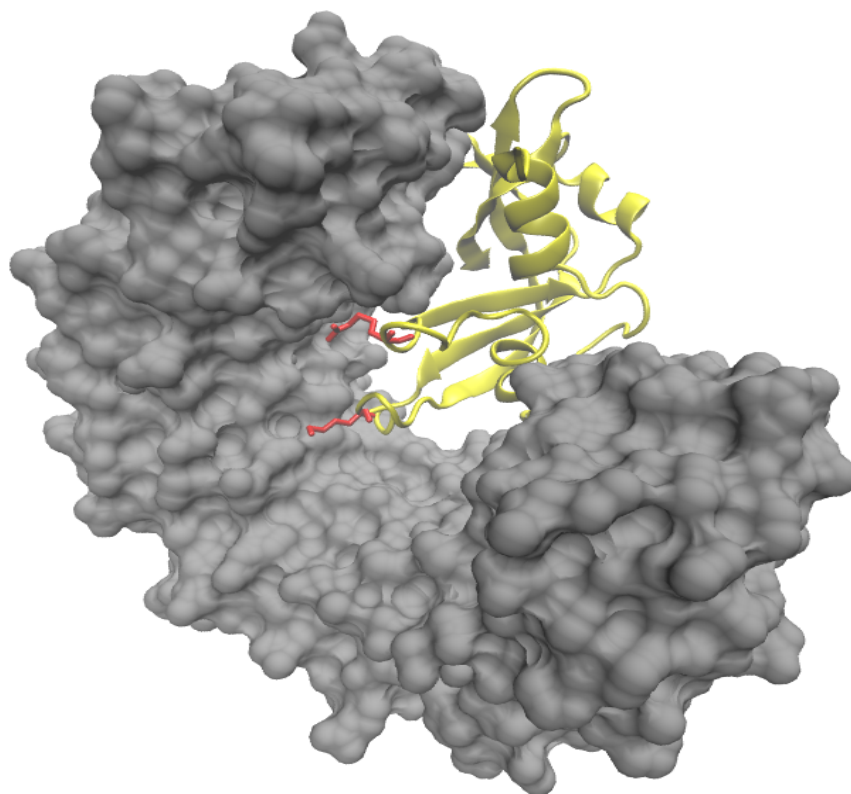
**Slika 5:** *Hot spot* aminokiselina Lys 93 (štapčićima), iz strukture ciklina T1, okružena hidrofobnim O prstenom aminokiselina (sivo)

*Hot spot* aminokiseline su veoma često okružene prstenom hidrofobnih aminokiselina (O – prsten) (28,35) (Slika 5). Uloga O-prstena je da izoluje energetske kontakte između proteina od molekula rastvarača. Aminokiseline koje grade O-prsten obično nemaju energetske značajnu ulogu u vezivanju proteina. O – prstenovi obično postoje na oba proteina u interakciji i postavljeni su naspram jedan drugome, upotpunjujući se. Teorija O-prstena je nedavno proširena i na protein-DNK komplekse (36).

*Hot spot* aminokiseline su zadužene za stabilnost jednom nastalog proteinskog kompleksa, obezbeđujući najveći deo energije vezivanja, ali u samom procesu formiranja kompleksa, značajnu ulogu imaju *anchor* aminokiseline (eng. *anchor* sidro) (37).

*Anchor* aminokiseline obezbeđuju inicijalno povezivanje proteinskih molekula u kompleks. Njihovi bočni lanci se, u strukturama slobodnih proteina koji stupaju u međusobnu interakciju, najveći deo vremena nalaze u konformaciji koju imaju po vezivanju proteina u kompleks. U prvoj fazi formiranja proteinskog kompleksa, kada je nagrađen kompleks susreta, ova već pripremljena «sidra» se smeštaju u odgovarajuće džepove na površini drugog proteina, po mehanizmu ključ-brava (37). Jednom kada

*anchor* aminokiseline obezbede inicijalno povezivanje i sterno ograniče pokretljivost površina, ostale aminokiseline na vezivnim površinama kompleksa se prilagođavaju jedna drugoj i formiraju konačan kompleks. Vrlo lep primer uloge *anchor* aminokiselina može se videti u kompleksu ribonukleaze A i njenog inhibitora (Slika 6).



**Slika 6:** *Anchor* aminokiseline, Arg 39 i Lys 94 (crveno) omogućavaju formiranje kompleksa ribonukleaze A (žuto) i njenog inhibitora (sivo).

Kao *anchor* aminokiseline javljaju se često lizin (Lys), Arg i aromatične aminokiseline. Alifatične *anchor* aminokiseline su ređe (u oko 20% ispitivanih kompleksa), a kisele aminokiseline, asparaginska (Asp) i glutaminska (Glu) kiselina veoma retke. Pošto se prilikom građenja kompleksa smeštaju u džepove izolovane od rastvarača, *anchor* aminokiseline podležu velikim promenama površine izložene rastvaraču (*solvent accessible surface area* – SASA) pri formiranju kompleksa, što se može upotrebiti za njihovo identifikovanje. One mogu, ali ne moraju biti *hot spot* aminokiseline, jer je njihova uloga da uspostave brzu i relativno specifičnu vezu između dva proteina koji interaguju dajući ostatku vezivne površine dovoljno vremena da se prilagodi. U proteinskom kompleksu može postojati jedna *anchor* aminokiselina (sa  $\Delta\text{SASA} > 100\text{\AA}^2$ ), ali ako je  $\Delta\text{SASA}$  pri formiranju kompleksa kod svih aminokiselina  $< 100\text{\AA}^2$ , neophodno je više *anchor* aminokiselina da bi se uspostavio kompleks.

### 1.1.3. Terapeutski značaj protein-protein interakcija

Kako su protein-protein interakcije često nosioci fizioloških procesa moguće je delovanjem na njih modulirati i procese koje one regulišu. Uloge protein-protein interakcija u patološkim stanjima mogu biti sumirane na sledeći način (38):

- protein-protein interakcija je direktno uključena u patološki proces, njenom inhibicijom se uspostavlja fiziološko stanje
- normalan nivo protein-protein interakcije je smanjen u patološkom procesu, neophodno je simulirati prisustvo interakcije
- fiziološka interakcija između proteina je previše izražena i uzrokuje patološko stanje, neophodna je njena inhibicija

Zbog značaja koji protein-protein interakcije imaju u živim sistemima, vezivne površine između proteina predstavljaju potencijalnu metu za razvoj novih lekova, naročito poslednjih decenija, usled porasta broja rešenih trodimenzionalnih struktura proteinskih kompleksa i razvoja metoda hemije i molekularne biologije. Mogućnost uticanja na protein-protein interakcije značajno proširuje spektar dostupnih terapeutski relevantnih proteinskih ciljnih mesta lekova.

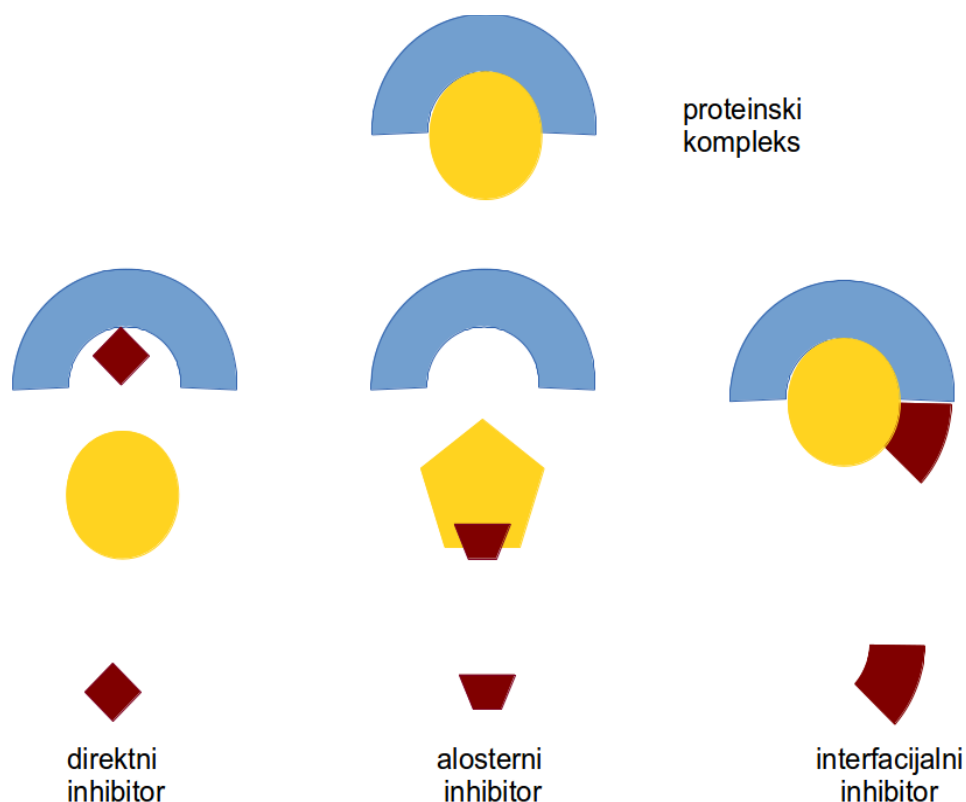
U poređenju sa klasičnim lekovima, inhibitori protein-protein interakcija pružaju više pogodnosti. Svaka protein-protein interakcija poseduje strukturne specifičnosti koje je izdvajaju od ostalih i obezbeđuju specifičnost interakcija u ćeliji, što daje i inhibitorima razvijenim za datu interakciju viši stepen selektivnosti u poređenju sa klasičnim lekovima čije ciljno mesto su aktivni centri enzima. Uz to, inhibitori protein-protein interakcija su otporniji na mutacije svojih ciljnih proteina, jer su vezivne površine proteinskih kompleksa evolutivno očuvanije od ostatka površine (39).

Sa aspekta medicine najveći značaj imaju privremeni funkcionalni proteinski kompleksi, najčešće heterodimeri, ali su oni i zahtevnija ciljna mesta za razvoj lekova, usled veće polarnosti svoje vezivne površine. Do sada je u razvoju inhibitora protein-protein interakcija najviše pažnje posvećivano interakcijama proteina uključenih u procese preživljavanja i apoptoze (p53, XIAP, proteini Bcl2 familije, menin), proteinima imunskog sistema (IL-2, TNF $\alpha$ ) ali i proteinima patogena (ZipA protein bakterijske membrane, integraza, HPV\_E2 virusni protein) (40).

Za razvoj inhibitora protein-protein interakcija moguće je izvesti nekoliko opštih preporuka:

- pogodno je odabrati hidrofobni džep ili konkavnu površinu kao ciljno mesto, ako takva postoji na vezivnoj površini (41). Nije neophodno da džep bude prisutan u vezivnoj konformaciji proteina, već se može nalaziti i u slobodnom proteinu ili nekoj od simuliranih konformacija. Prisustvo tri do pet džepova sa ukupnom zapreminom oko  $250 \text{ \AA}^3$  (17) ili pojedinačnog džepa sa zapreminom većom od  $200 \text{ \AA}^3$  (19) ukazuje na mogućnost razvoja inhibitora za taj region površine. Hidrofobnost džepa omogućuje razvoj lipofilnih inhibitora sa većom verovatnoćom prolaska ćelijske membrane.
- nije neophodno da mali molekul modulator ima veći afinitet od proteinskog partnera u interakciji – ukoliko se inhibitor manjeg afiniteta nađe u dovoljnoj koncentraciji, on može da poremeti biohemijsku ravnotežu vezivanja visokoafinitetnog proteina i ispolji svoje dejstvo (42)
- poželjno je u dizajn inhibitora uključiti strukturne karakteristike vezivne površine (16):
  - o arhitekturu i organizaciju vezivne površine, radi odabira optimalne orijentacije inhibitora i procene uticaja više istovremenih promena u njihovoj strukturi
  - o energetski značajne interakcije koje ostvaruju *hot spot* aminokiseline
  - o O-prsten (ometanjem ili oponašanjem)
  - o vezivne džepove *anchor* aminokiselina
  - o hidrofobne oblasti: desolvatacija ovih površina prilikom interakcije liganada sa njima će obezbediti povećanje afiniteta kroz povećanu entropiju oslobođenih molekula vode
  - o prisustvo i raspored molekula vode na vezivnoj površini
  - o stepen komplementarnosti vezivnih površina. Molekularne interakcije koje nisu zadovoljene u nativnom proteinskom kompleksu usled ograničene komplementarnosti vezivnih površina mogu biti iskorišćene za povećanje afiniteta liganada vezanih u toj oblasti

Navedene smernice se odnose na razvoj direktnih (ortosternih) inhibitora protein-protein interakcija, koji se vezuju za vezivnu površinu jedne od proteinskih subjedinica u kompleksu i sprečavaju formiranje kompleksa. Potencijalni nedostatak direktnih inhibitora predstavlja smanjena efikasnost inhibicije kada su u konkurenciji sa proteinskom jedinicom koja je prisutna u visokoj koncentraciji ili veoma stabilna (16). Direktni inhibitori protein-protein interakcija se najviše proučavaju i najjednostavnije ih je razviti, ali pored njih postoje i alosterni (43) i interfacijalni inhibitori (44) (Slika 7).



**Slika 7:** Tipovi inhibitora protein/protein interakcija

Alosterni inhibitori protein-protein interakcija se ne vezuju direktno za proteinske površine uključene u formiranje kompleksa, već van njih, dovodeći do konformacionih promena vezivnih površina koje ometaju njihovo vezivanje. Interfacijalni inhibitori ne ometaju formiranje proteinskog kompleksa, već se umeću u kompleks kada je on nagrađen i zadržavaju ga u nekoj od konformacija koje ne mogu da obavljaju biološku funkciju. Primer interfacijalnog inhibitora bi bio C4 peptid koji blokira funkciju CDK2 / ciklin A kompleksa (45). Iako većina inhibitora protein-protein interakcija deluje reverzibilno, razvijen je i ireverzibilni inhibitor MAGI3 kinaze /

PTEN (eng. *membrane associated guanylate kinase / phosphatase and tensin homolog*, MAGI3/PTEN) interakcije, koji po vezivanju gradi kovalentne veze sa proteinom (46).

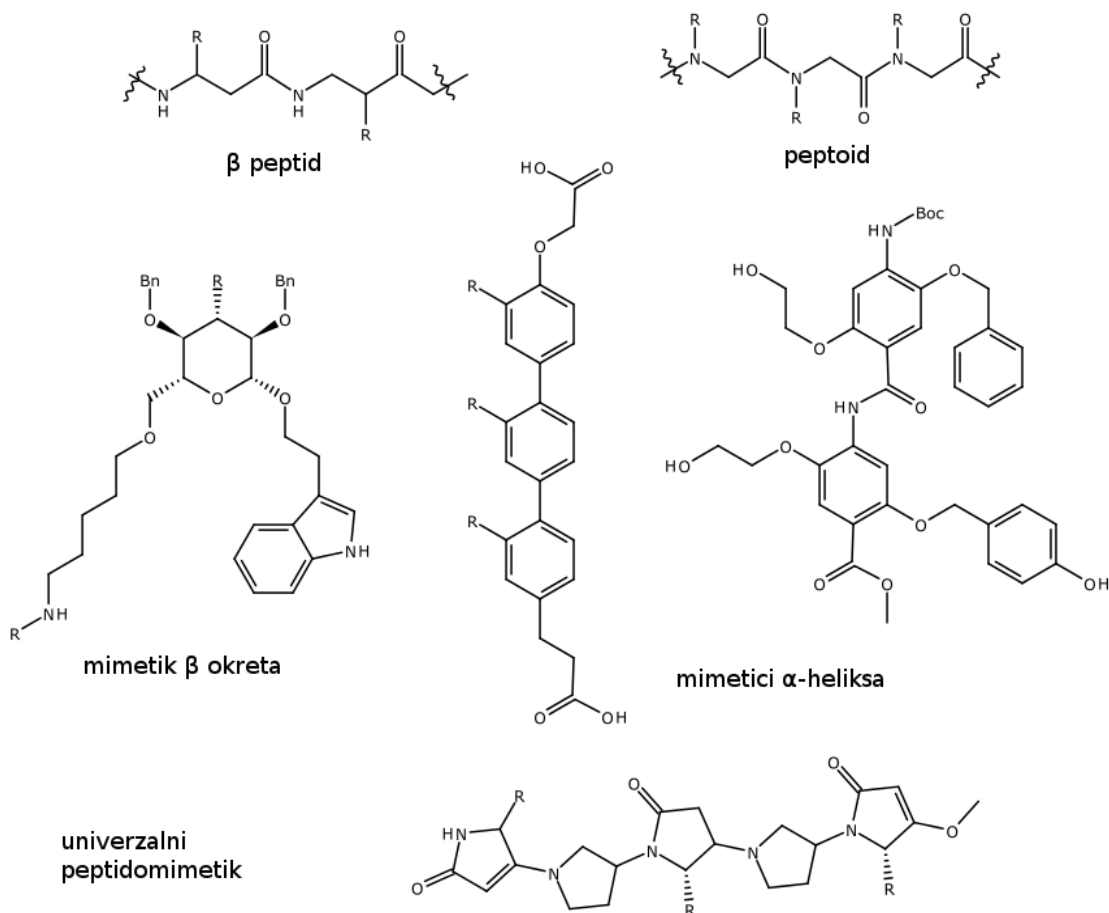
U postupku dizajna inhibitora jedne protein-protein interakcije mogu se koristiti različite eksperimentalne i *in silico* metode. Sam tok istraživanja zavisi od dostupnosti trodimenzionalne strukture proteinskog kompleksa od interesa i njegovih strukturnih karakteristika. Ako nije poznata struktura proteinskog kompleksa, jedini put za otkrivanje modulatora interakcija u njemu je *in vitro* testiranje hemijskih kolekcija prirodnih ili sintetisanih jedinjenja. Ako je struktura kompleksa poznata, bilo rešena kristalografskim metodama i NMR tehnikama ili, sa manje pouzdanosti, modelovana homologim modeliranjem, u dizajnu inhibitora se koriste njene osobine: prisustvo ili odsustvo hidrofobnih džepova, komplementarnost oblika, prisustvo i raspored molekula rastvarača, određivanje *hot spot* i *anchor* aminokiselina (16). Ako u interakciji značajno učestvuju neprekinuti peptidni fragmenti sa jednog od proteina, oni mogu biti podvrgnuti testovima vezivanja, prevedeni u peptidomimetike i dalje optimizovani. Ako to nije slučaj može se upotrebiti *de novo* racionalni dizajn inhibitora (dizajn iz fragmenata), zasnovan na prostornom rasporedu energetski značajnih kontakata na vezivnoj površini.

Primenom opisanih smernica i metoda već je uspešno razvijeno više inhibitora različitih protein-protein interakcija (39,47,48), od kojih se neki nalaze u kliničkim istraživanjima (49), ili su već odobreni za primenu u terapiji (50).

Uspešno razvijeni inhibitori protein-protein interakcija se, po svom hemijskom sastavu, mogu podeliti u dve glavne podgrupe: peptidi i njihovi derivati i nepeptidni organski molekuli. Peptidi predstavljaju prirodan izbor za oponašanje interakcija prisutnih na vezivnoj površini proteinskog kompleksa. U studiji interakcija globularnih proteina (51) pokazano je da za više od polovine ispitivanih proteinskih kompleksa na vezivnoj površini postoji jedan linearni peptidni segment dužine 5 do 10 aminokiselina koji doprinosi više od 50% energije vezivanja pri formiranju kompleksa. Peptide odlikuje jednostavniji postupak dizajna od nepeptidnih inhibitora i veća specifičnost u vezivanju, jer se obično razvijaju na osnovu jednog od proteina u interakciji. Nedostatak peptida je što imaju ograničenu sposobnost prolaska kroz ćelijske membrane i što su metabolički nestabilni. Uz to, peptidi su fleksibilni molekuli pa njihovo vezivanje neretko uključuje značajnu entropijski nepovoljnu komponentu. Derivatizacijom peptida



do peptidomimetika može se korigovati njihova stabilnost, ali i povećati lipofilnost i rigiditet, što približava ove derivate nepeptidnim strukturama, čija je lipofilnost i stabilnost obično zadovoljavajuća. U ovu svrhu peptidi se mogu ciklizovati ili se u njihovu strukturu mogu inkorporirati aminokiseline koje se ne javljaju u proteinima (D-aminokiseline,  $\beta$  i  $\gamma$  aminokiseline) ili njihovi analozi (peptoidi). Alternativni pristup dizajnu peptidomimetika jeste oponašanje rigidnih elemenata sekundarne strukture proteina bez zadržavanja elemenata polipeptidnog niza. U ovom pristupu se najčešće koriste heterociklične strukture koje prostorno pozicioniraju svoje supstituente kao elementi proteina koje oponašaju. Tako su dizajnirani mimetici  $\alpha$ -heliksa,  $\beta$ -ploče,  $\beta$  i  $\gamma$  okreta (52–55) (Slika 8). Razvijeni su i peptidomimetici koji istovremeno mogu da zauzmu konformacije koje odgovaraju elementima različitih sekundarnih struktura, tzv. univerzalni peptidomimetici (52,56).

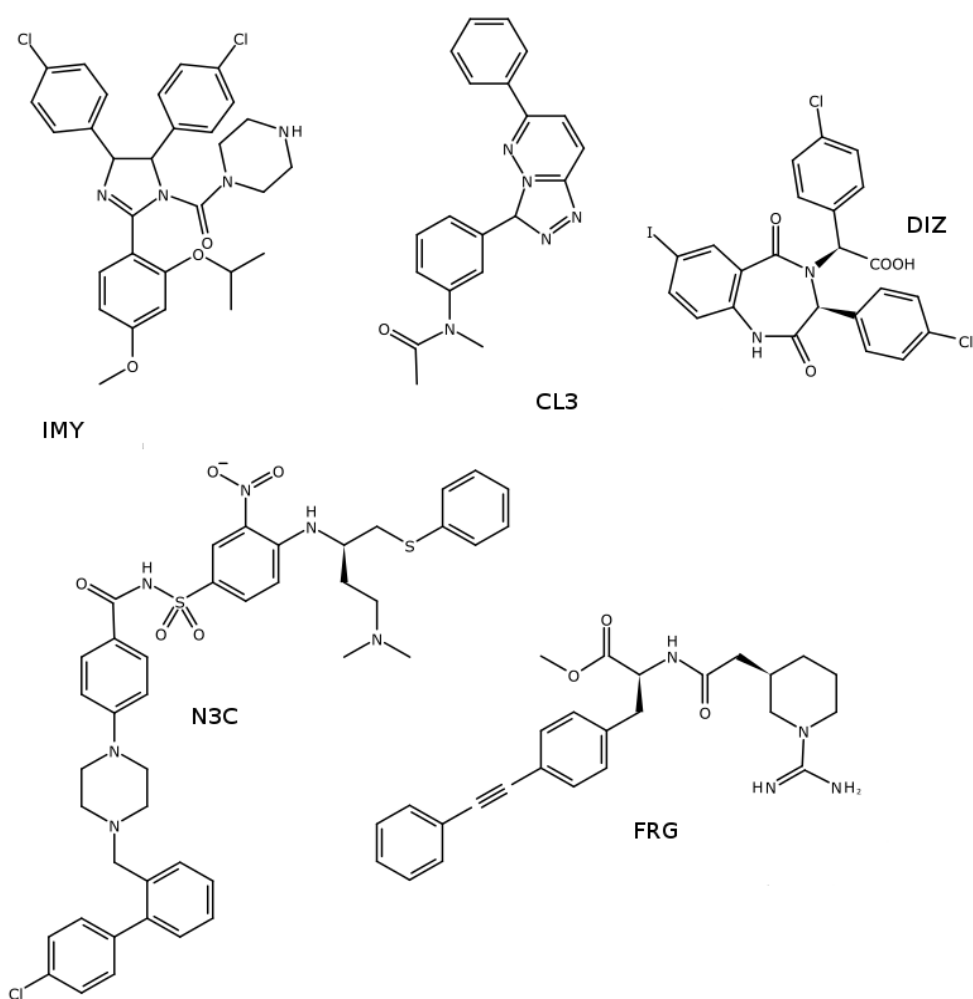


**Slika 8:** Primeri struktura peptidomimetika

Kao posebna, ograničena klasa inhibitora protein-protein interakcija izdvajaju se terapijska antitela (57). Antitela su veoma specifični inhibitori, ali mogu da moduliraju

samo ekstracelularne protein-protein interakcije, jer zbog svoje veličine i polarnosti ne dospevaju u ćelije.

Povećanje broja dostupnih nepeptidnih inhibitora protein-protein interakcija omogućilo je detaljniju analizu njihovih osobina. Formirane su i databaze poznatih inhibitora protein-protein interakcija kao što su 2p2i (58) i TIMBAL (59). Pokazano je da su, u poređenju sa klasičnim molekulima lekova, inhibitori protein-protein interakcija veći (molekulska masa > 400 g/mol), rigidniji, lipofilniji (alogP ~ 4), sa više aromatičnih prstenova ( $\geq 4$ ) i više akceptora vodoničnih veza ( $\geq 4$ ) (59,60) (Slika 9).



**Slika 9:** Primeri struktura nepeptidnih inhibitora protein-protein interakcija

Neretko, inhibitori protein-protein interakcija ne zadovoljavaju jedno ili više pravila Lipinskog o strukturnim karakteristikama lekova potrebnim za njihovu dobru resorpciju i permeabilnost. Kao rezultat, postojeće kolekcije molekula dizajnirane i

korišćene u razvoju klasičnih lekova ne pokrivaju hemijski prostor inhibitora protein-protein interakcija na odgovarajući način, što otežava otkrivanje novih inhibitora (61). Hemijske biblioteke prilagođene razvoju inhibitora protein-protein interakcija su nedavno razvijene (62). U osnovi dosadašnjih nepeptidnih inhibitora protein-protein interakcija se obično nalaze aromatični, razgranati skeleti, uz dodatne polarne ili jonske grupe (63). Analiza inhibitora je pokazala i da je njihova najznačajnija karakteristika njihov molekularni oblik, uz često prisustvo dvostrukih veza, radi smanjenja fleksibilnosti (64). Inhibitori na vezivnoj površini zauzimaju površinu koja u proseku odgovara jednom okretu  $\alpha$ -heliksa (65). U proseku inhibitori protein-protein interakcija imaju manji doprinos energiji vezivanja po atomu skeleta (1kJ po atomu) (66) od tradicionalnih lekova (1.5 kJ / atomu) (63). Strukturne osobine poznatih inhibitora protein-protein interakcija inkorporirane su u sistem za klasifikaciju i predviđanje uspešnosti novih predloženih struktura inhibitora (64).

Povećanje broja uspešnih inhibitora protein-protein interakcija dovelo je i do boljeg razumevanja građe proteinskih površina za koje se inhibitori vezuju. Statistička analiza proteinskih površina za koje se vezuju i mali organski ligandi i proteini pokazala je da se mali molekuli najčešće vezuju u blizini triptofana, histidina, tirozina i fenilalanina na vezivnoj površini, dok retko stupaju u interakcije sa nepolarnim aminokiselinama kakve su alanin, valin, leucin i izoleucin (65). Aminokiseline sa kojima interaguju inhibitori protein-protein interakcija takođe pokazuju veći stepen evolutivne očuvanosti i najčešće se nalaze u džepovima, zaklonjene od rastvarača (17). Analiza vezivnih površina je dovela i do razvoja metoda za statističko predviđanje vezivanja inhibitora za vezivnu površinu proteinskih kompleksa (67). Na sličan način, analizirani su i svi poznati proteinski kompleksi u strukturi čije vezivne površine značajnu ulogu igra prisustvo  $\alpha$ -heliksa, radi predviđanja peptidnih fragmenata strukture koji bi mogli da učestvuju u inhibiciji proučavanih struktura (68). Korak dalje predstavlja analiza koju su sprovedi Koes i Camacho (34). U svojoj studiji oni su iskoristili strukturne osobine kompleksa koji su uspešno inhibirani malim molekulima da formiraju kriterijume za pronalaženje početnih tačaka za inhibiciju u drugim proteinskim kompleksima, bez obzira na njihovu sekundarnu građu. Sve opisane studije su izvedene na rigidnim proteinskim strukturama iz PDB databaze kristalografskih struktura. Mogućnost procene uticaja konformacionih promena proteina na izgled

vezivnih površina bi dovela do još preciznijih rezultata (34). Jedna takva studija se bavila verovatnoćom otvaranja džepova na vezivnim površinama proteinskih kompleksa za koje su uspešno razvijeni inhibitori (69). Pokazano je da su se u energetski povoljnim konformacijama proteina, dobijenim polazeći od konformacije slobodnog proteina van kompleksa, džepovi otvarali samo u onim regionima vezivne površine za koje su se vezivali poznati inhibitori, a ne i svuda po površini proteina. Veličina i oblik ovih džepova bi stoga diktirali i veličinu i oblik mogućih inhibitora, dok bi se dodatno fino prilagođavanje ligandu odigravalo nakon vezivanja inhibitora (69).

Uzeti zajedno, rezultati svih navedenih istraživanja ukazuju da je moguće dizajnirati male organske molekule kao inhibitore protein-protein interakcija u proteinskim kompleksima sa različitim vezivnim površinama, koristeći svojstva njihove molekularne građe. Međutim, kako su interakcije proteina u proteinskim kompleksima evoluirale da obezbede visok stepen specifičnosti vezivanja, svaki proteinski kompleks predstavlja jedinstveno molekularno okruženje i mora biti posmatran kao takav. Kao rezultat, za uspešan dizajn inhibitora neke protein-protein interakcije, neophodno je proučiti i razumeti osobine njene molekularne površine i iskoristiti njene specifičnosti za postizanje afiniteta i selektivnosti potencijalnih inhibitora.

## 1.2 P-TEFb kompleks

### 1.2.1 Građa i funkcija P-TEFb kompleksa

Pozitivni faktor elongacije transkripcije b (P-TEFb) je proteinski kompleks koji učestvuje u regulaciji transkripcije aktivirajući RNK polimerazu II. Katalitičko jezgro P-TEFb kompleksa čini ciklin zavisna kinaza 9 (CDK9) u kompleksu sa ciklinima T1, T2a, T2b i K (70,71). U kompleksu sa različitim regulatornim proteinima, katalitičko jezgro P-TEFb kompleksa vrši fosforilaciju svojih supstrata i omogućava neometano odvijanje procesa transkripcije velikog broja gena u ćelijama. Ključna subjedinična P-TEFb kompleksa za fosforilaciju jeste CDK9 .

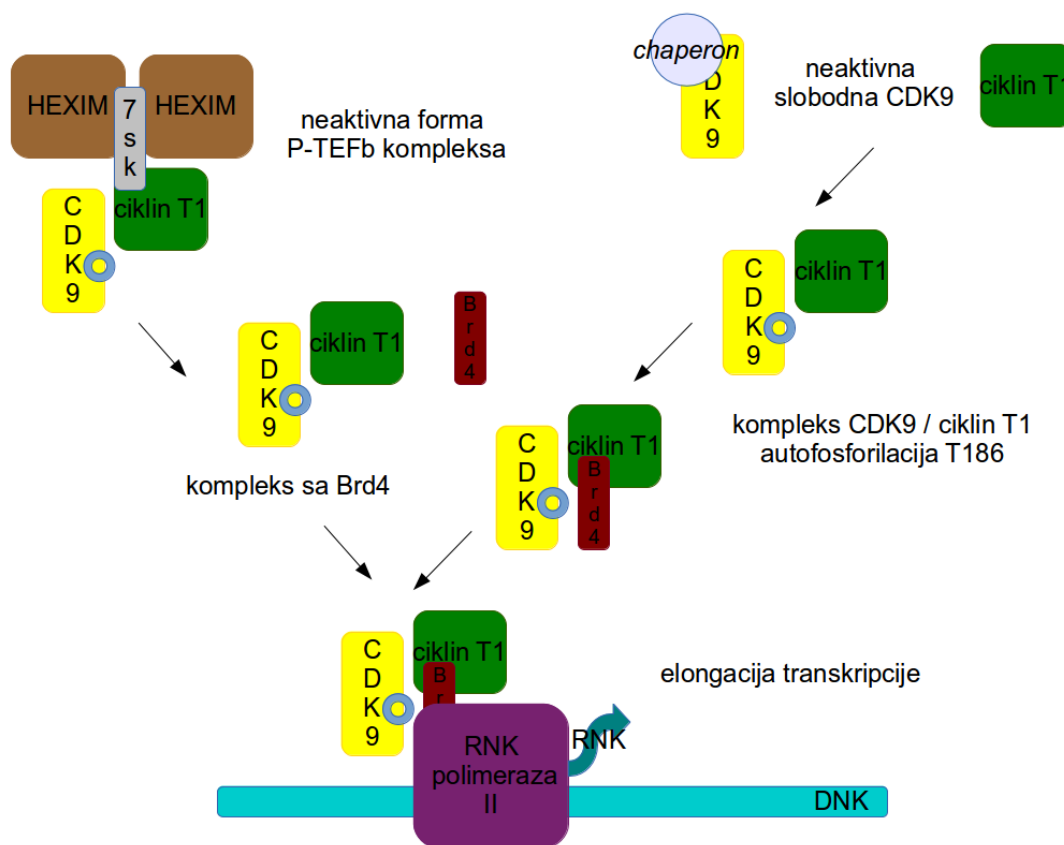
CDK9 je serin-treonin kinaza iz porodice ciklin zavisnih kinaza. Dok većina ciklin zavisnih kinaza učestvuje u odvijanju ćelijskog ciklusa, CDK9, zajedno sa CDK7 i CDK8 svoju ulogu ispoljava samo u regulaciji transkripcije. Ciklin zavisne kinaze odlikuje sličnost strukture i evolutivna očuvanost građe. CDK9 deli 33–47% primarne strukture sa drugim CDK (72,73). CDK9 je široko eksprimirana u velikom broju tkiva, najviše u terminalno diferenciranim tkivima (74). Humana CDK9 postoji u dve izoforme molekulske mase 42 i 55 kDa (CDK9\_42 i CDK9\_55). Obe izoforme grade iste komplekse sa ciklinima i razlikuju se po dodatnih 117 N-terminalnih aminokiselina u CDK9\_55 (75), čija uloga nije poznata, ali ima indikacija da su uključene u transport CDK9\_55 molekula u jedarce (76). Izoforma CDK9\_42 je dominantno prisutna u najvećem broju tkiva, mada se nivoi ekspresije dve izoforme razlikuju među različitim tipovima ćelija (76). Od značaja je pomenuti da se ekspresija CDK9 i ciklina T1 značajno povećava u procesu aktivacije T- i B-limfocita (77,78). Iako su obe izoforme CDK9 uključene u regulaciju transkripcije, njihove uloge se ne preklapaju u potpunosti. Postoje indikacije da je CDK9\_55 izoforma uključena u obnavljanje DNK, ali ne u kompleksu sa ciklinima T (79).

CDK9 je u ćeliji pretežno lokalizovana u jedru, ali se može naći i u citoplazmi (80). Novosintetisana CDK9 u ćeliji je nestabilna, stupa u prolazne interakcije sa *chaperon* (eng. *chaperon*, pratioci) proteinima HSP70, HSP90, Cdc37 (81) i stabilizuje se tek građenjem kompleksa sa ciklinima T1, T2a, T2b ili K (82), među kojima je kompleks sa ciklinom T1 najzastupljeniji (70). Kompleks CDK9 / ciklin T1 se od

kompleksa drugih ciklin zavisnih kinaza i ciklina razlikuje po manjoj vezivnoj površini, usled promenjene orijentacije ciklina T1 (83), ali je slične opšte građe (Slika 1). Građenje kompleksa sa ciklinima je neophodno za aktivaciju CDK9. U formiranom kompleksu CDK9 / ciklin dolazi do autofosforilacije Thr 186 iz T-petlje CDK9 (83). Autofosforilacija Thr 186 kod CDK9 je specifična jer se kod većine drugih CDK T-petlja fosforiliše dejstvom CDK7. CDK9 takođe autofosforiliše osam aminokiselina svog C-terminusa (tzv. C8), što omogućava transport P-TEFb kompleksa u jedro ćelije (84), omogućuje vezivanje P-TEFb / Tat kompleksa za TAR (eng. *transactivation response*) RNK (85), i utiče na redosled vezivanja supstrata na aktivnom mestu kinaze i mehanizam reakcije katalize (86). CDK9 u kompleksu sa ciklinima je stabilna. Nasuprot njoj, monomerna CDK9 u ćeliji je nestabilna i brzo se razgrađuje (81,87,88). Uz to, nivo CDK9 proteina u ćeliji zavisi od ekspresije ciklina (89,90). Kao rezultat, svoju fiziološku ulogu CDK9 obavlja samo u kompleksima sa ciklinima, pretežno kao jezgro P-TEFb kompleksa.

P-TEFb kompleks svoju ulogu u regulaciji transkripcije obavlja fosforilacijom serina 2 u karakterističnoj heptapeptidnoj sekvenci C-terminalnog domena RNK-polimeraze II (91–93), ali i fosforilacijom negativnih faktora elongacije transkripcije DSIF (eng. *DRB sensitivity inducing factor*, faktor osetljivosti na DRB) i NELF (eng. *negative elongation factor*, negativni faktor elongacije) (94). Negativni faktori elongacije su vezani za RNK-polimerazu II i sprečavaju proces elongacije transkripcije. Nakon fosforilacije od strane P-TEFb, oni disosuju i omogućavaju nesmetano odvijanje transkripcije. Dodatna uloga P-TEFb kompleksa je u regulaciji modifikacije hromatina i mRNA (95). Uloga koju će obavljati formiran molekul P-TEFb kompleksa zavisi od tipa ciklina prisutnog u njemu. Kao regulatorne jedinice P-TEFb kompleksa, ciklini za koje se veže CDK9 interaguju sa većim brojem drugih proteina koji usmeravaju katalitičku aktivnost kinaze tamo gde je potrebna. Signalna uloga ciklina u P-TEFb kompleksu odslikava se u različitosti njihove primarne strukture van vezivne površine sa CDK9 (83), koja određuje osobine molekularne površine ciklina dostupne za interakciju sa drugim molekulima. Obe izoforme CDK9 u kompleksu sa bilo kojim od svojih ciklina učestvuju u fosforilaciji C-terminalnog domena RNK-polimeraze II, ali je samo kompleks CDK9 / ciklin T2a uključen u diferencijaciju mišićnog tkiva (96), a CDK9 / ciklin K u očuvanje integriteta genoma

(97). Ciklin K sličnu ulogu ima u kompleksima sa CDK12 i CDK13 (98). Ciklini T1 i T2 imaju ograničenu sposobnost da međusobno nadoknade funkcije jer dele jedan broj gena čiju transkripciju kontrolišu (90).



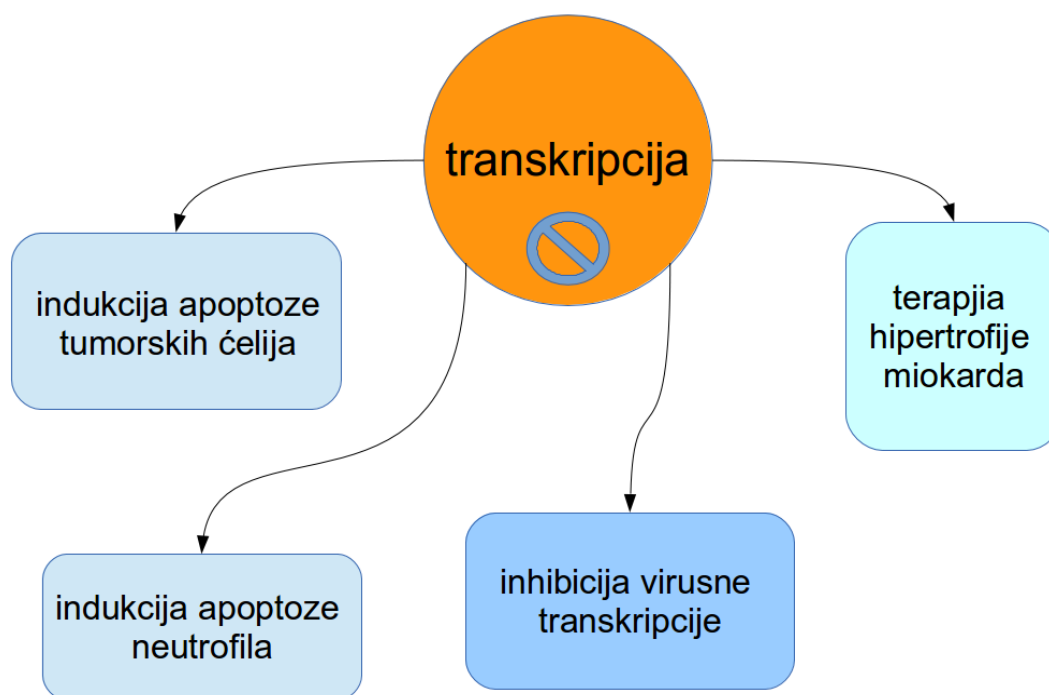
**Slika 10:** Aktivacija CDK9 / ciklin T1 kompleksa

P-TEFb kompleks u ćeliji postoji u dva oblika: katalitički aktivnom i neaktivnom (Slika 10). Katalitički aktivni oblik P-TEFb kompleksa se sastoji od kompleksa CDK9 / ciklin T sa kojim pre procesa transkripcije stupa u interakcije regulatorni jedarni BRD4 protein (eng. *bromodomain containing protein 4*, protein sa bromodomenom 4) (99,100). U neaktivnoj formi P-TEFb kompleksa se jedan molekul 7SK snRNK (eng. *small nuclear*, mala jedarna RNK, snRNK) i dva molekula HEXIM proteina (eng. *hexamethylene bis-acetamide inducible*, protein indukovano heksametilen bisacetamidom) nalaze vezani za CDK9/ciklin T kompleks. 7SK snRNK i HEXIM proteini se vezuju samo za aktivan kompleks CDK9/ciklin T1 u kome je CDK9 prethodno autofosforilisana (101,102). Na ovaj način se nastankom neaktivnog oblika P-TEFb kompleksa direktno reguliše i smanjuje aktivnost CDK9 u ćeliji. Nakon disocijacije HEXIM (HEXIM1 ili HEXIM2) proteina i 7SK snRNK CDK9 / ciklin T

kompleks je ponovo katalitički aktivan, što ukazuje da neaktivni oblik P-TEFb kompleksa služi kao ćelijski depo privremeno inaktivirane CDK9 kojim se reguliše njena aktivnost (103). Aktivnost P-TEFb kompleksa je regulisana i fosforilacijom i acetilacijom same CDK9. Acetilacija Lys 44 u sekvenci CDK9 vodi povećanju nivoa aktivnosti P-TEFb kompleksa (104).

### 1.2.2 Patološki značaj P-TEFb kompleksa

Regulacija nivoa CDK9 aktivnosti u ćeliji ima značaj u terapiji različitih patoloških stanja kod ljudi (82,105,106) (Slika 11).



**Slika 11:** Efekti inhibicije aktivnosti P-TEFb kompleksa u patološkim stanjima

Kao deo P-TEFb kompleksa, CDK9 omogućuje elongaciju transkripta velikog broja gena, a samim tim doprinosi regulaciji nivoa transkripcije u ćeliji. Deregulacija aktivnosti P-TEFb kompleksa učestvuje u više malignih stanja (82): limfomima (107), rabdomyosarkomu (108), kanceru prostate (109), PNET (eng. *primitive neuroectodermal tumor*, tumor nediferenciranog neuroektoderma) neuroblastomu (110). Smanjenje nivoa CDK9 aktivnosti u ćeliji smanjiće i nivo transkripcije, što će se najviše



odraziti na nivo kratkoživećih proteina u ćeliji kao što su regulatori ćelijskog ciklusa (ciklini D) i antiapoptotski proteini Mcl-1 (eng. *myeloid cell leukemia sequence 1*, mieloidni ćelijski faktor 1), XIAP (eng. *X-linked inhibitor of apoptosis*, inhibitor apoptoze povezan sa X hromozomom), proteini Bcl-2 familije (eng. *B-cell lymphoma 2*, limfom B-ćelija), MDM2 (eng. *mouse double minute*) i survivin (106,111). Jedan broj tumorskih ćelijskih linija preživljava zahvaljujući neprestanoj transkripciji i sintezi kratkoživećih anti-apoptotskih faktora. Primer su ćelije hronične limfocitne leukemije (112), multiplog mieloma (113), osteosarkoma (114). Kod takvih ćelija, smanjenje nivoa transkripcije, inhibicijom aktivnosti CDK9, vodi u apoptozu (112,113,115).

Aktivnost CDK9 je takođe neophodna za transkripciju metaloproteinaze MMP9, koja je uključena u metastazu tumora (116). Danas se u kliničkim ispitivanjima nalazi veći broj CDK inhibitora za koje se smatra da indukuju apoptozu tumorskih ćelija inhibicijom aktivnosti CDK9 (105,117–119). Uz to, pokazano je da inhibitori CDK9 ometaju i proces angiogeneze (120), što povećava značaj CDK9 u onkologiji. Sprečavanje procesa formiranja novih krvnih sudova od strane tumorskih ćelija (neoangiogeneze) je korisna strategija za suzbijanje rasta tumora, jer se tumorskom tkivu ograničava dotok kiseonika i hranljivih materija.

Indukcija apoptoze ometanjem transkripcije kratkoživećih proteina je primenjiva strategija i u terapiji hroničnih inflamatornih procesa kao što je reumatoidni artritis, gde CDK9 kontroliše životni vek neutrofila održavajući nivo anti-apoptotskog Mcl-1 proteina (121). Smanjenje aktivnosti CDK9 kod hroničnih inflamatornih stanja, bi, smanjenjem broja aktivnih neutrofila, smanjilo i oštećenja tkiva nastala njihovom neumerenom aktivacijom.

Ulogu P-TEFb kompleksa u regulaciji transkripcije iskorišćava jedan broj virusa za svoju replikaciju u ćelijama (82). Ovi virusi regrutuju P-TEFb kompleks za aktivaciju svoje transkripcije ćelijskom RNK-polimerazom II, pa je smanjenje aktivnosti CDK9 od koristi u terapiji jednog broja virusnih infekcija. Kod infekcije HIV-1 virusom, virusni Tat (eng. *trans-activator of transcription*, transaktivator transkripcije) protein regrutuje kompleks CDK9 / ciklin T1 ćelije domaćina da bi omogućio transkripciju virusnih gena (122). Ova interakcija je specifična za kompleks CDK9 sa ciklinom T1 (123,124). Tom prilikom Tat se vezuje za TRM (eng. *Tat:TAR recognition motif*, motiv za prepoznavanje Tat:TAR kompleksa) region ciklina T1, koji

nije prisutan u ciklinima T2 i K i koji sadrži cistein potreban za koordinaciju atoma cinka iz građe Tat proteina. Uz TRM region, Tat protein interaguje i sa T-petljom CDK9 (125). Vezivanje Tat proteina za P-TEFb povećava vezivnu površinu kompleksa CDK9 / ciklin T1 i dovodi do konformacionih promena kompleksa (125) koje povećavaju katalitičku aktivnost CDK9 i menjaju specifičnost za supstrate (126). Virusni Tat protein, inhibitorski HEXIM i aktivacioni Brd proteini se vezuju sa isti region na ciklinu (127,128). Postoje indikacije da Tat protein može da istisne HEXIM iz kompleksa sa ciklinom i tako reaktivira neaktivni P-TEFb (122,129). Aktivacija HIV-1 transkripcije pod dejstvom CDK9 zavisi i od stepena fosforilacije CDK9. Defosforilacija aminokiseline Ser 175 iz T-petlje CDK9 dejstvom proteinske fosfataze 1 dovodi do aktivacije HIV-1 transkripcije (130). Moguće objašnjenje se može naći u činjenici da Tat protein prilikom svog vezivanja pozicionira bočni niz svog Glu 9 u neposrednu blizinu ovog serina, što bi u slučaju fosforilisanog Ser 175 dovelo do nepovoljnih elektrostatskih interakcija. Više eksperimentalnih studija je pokazalo da je smanjenjem aktivnosti P-TEFb kompleksa moguće inhibirati HIV-1 transkripciju i replikaciju bez uticaja na preživljavanje i proliferaciju ćelija (89,131–133).

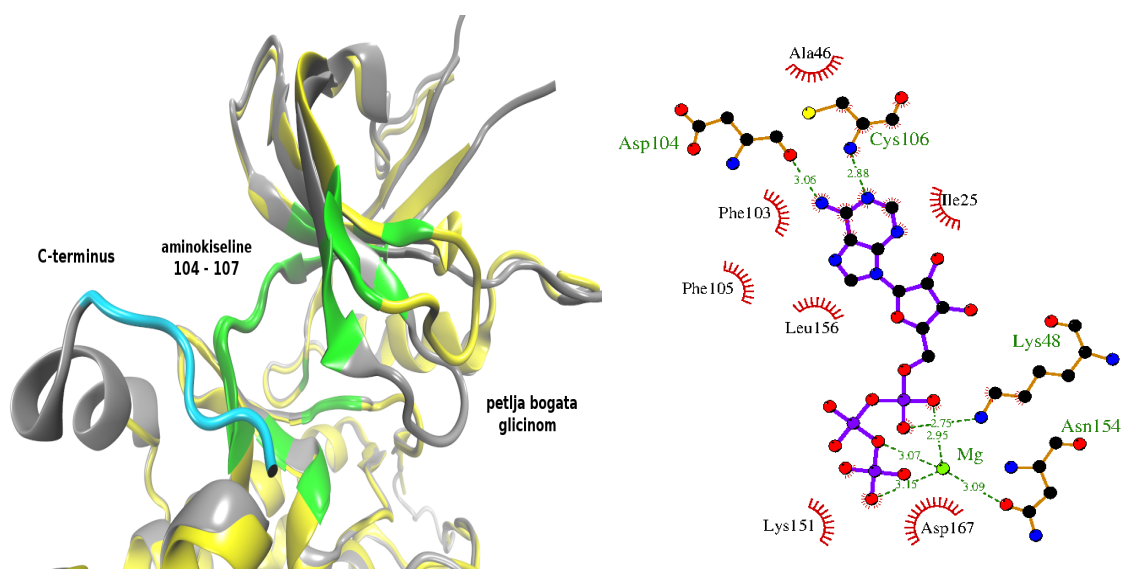
Osim u infekciji HIV-1 virusom, pokazano je da P-TEFb kompleks igra ulogu i kod infekcija Epštajn-Barovim virusom (134), humanim T-limfotropnim virusom tip 1 (135), humanim citomegalovirusom (136), herpes simplex virusima (137). Udružen antivirusni i antitumorski efekat koji se postiže inhibicijom aktivnosti P-TEFb kompleksa bio bi od koristi onkološkim pacijentima koje oslabljen imunski sistem (usled radioterapije ili primene citostatika) čini podložnim infekcijama (138).

Povećana aktivnost P-TEFb kompleksa je jedan od faktora odgovornih za razvoj hipertrofije miokarda (139), što njegovoj inhibiciji daje potencijalan terapijski značaj i u kardiologiji (140,141). Dok bi u potencijalnoj primeni inhibitora CDK9 aktivnosti u terapiji HIV infekcija cilj bio hronična terapija niskim koncentracijama inhibitora i sprečavanje replikacije virusa sa minimalnim uticajem na transkripciju ćelija domaćina, u terapiji hipertrofije miokarda bi se modifikovala transkripcija ćelija većim koncentracijama inhibitora, uz održavanje ćelija srca u životu (73).

### 1.2.3 Inhibicija aktivnosti P-TEFb kompleksa

#### *ATP – kompetitivni inhibitori*

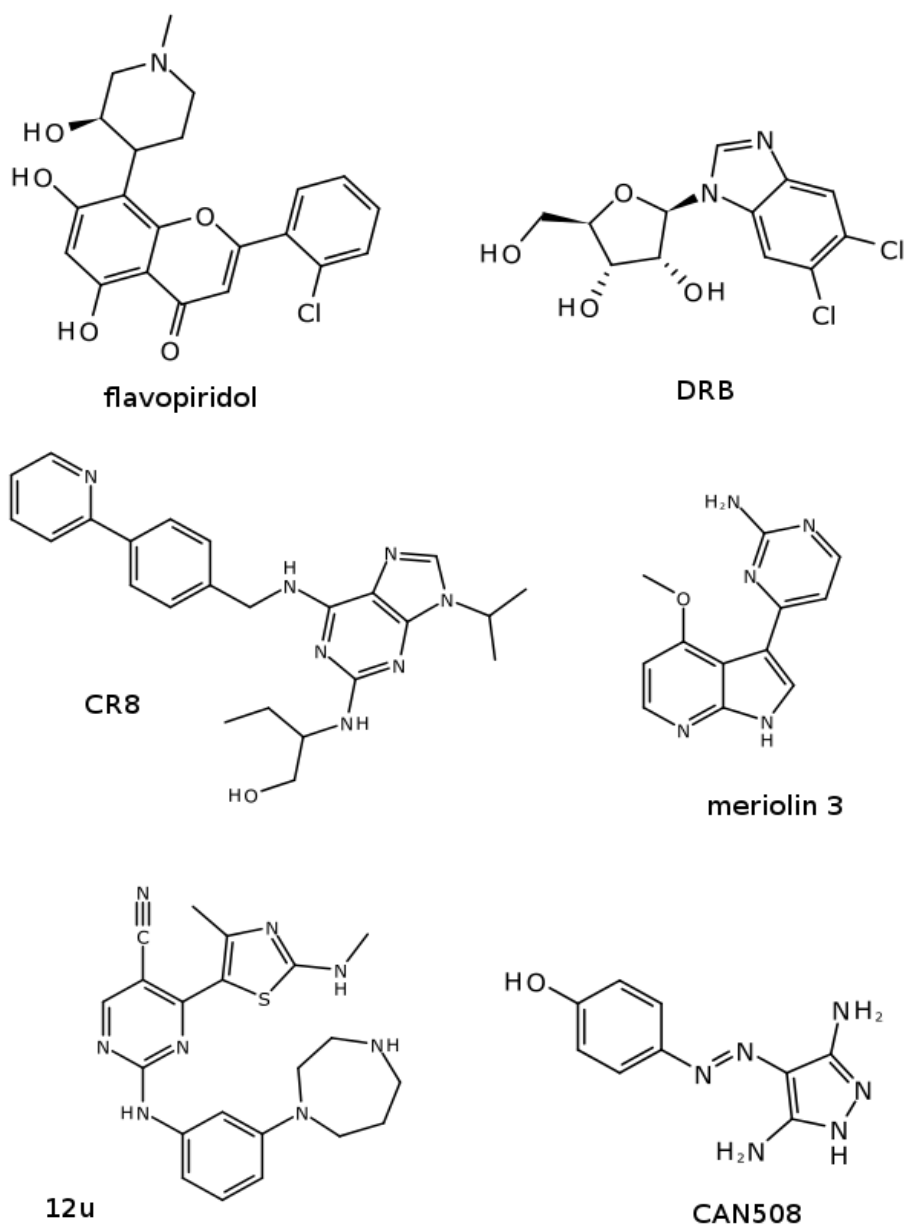
Patološki značaj P-TEFb kompleksa opravdava razvoj inhibitora njegove aktivnosti. Najviše pažnje u razvoju inhibitora P-TEFb posvećeno je katalitičkoj subjedinici kompleksa, CDK9 i njenom aktivnom centru.



**Slika 12:** Aktivni centar CDK9 (levo): Apo forma CDK9 (žuto), CDK9 iz kompleksa sa inhibitorom DRB (sivo). Aktivni centar je prikazan zelenom, a C-terminus CDK9 plavom bojom. Dijagram interakcija ATP-a sa CDK9 (desno)

Kao ciklin zavisna kinaza, CDK9 deli sa drugim proteinima ove familije evolutivno očuvan aktivni centar (142): hidrofobni džep na pregibu svoja dva domena, koji služi za vezivanje molekula ATP (Slika 12). Uz hidrofobni džep u kome se smešta purinsko jezgro ATP-a, u vezivnom mestu nalazi se i polipeptidni niz aminokiselina (aminokiseline 104–107) koji povezuje N i C terminalne domene CDK9 i formira vodonične veze sa adeninom. Ribozil-fosfatni supstituent ATP se pri vezivanju smešta u džep u čijoj neposrednoj blizini je fleksibilna petlja bogata glicinom. Fleksibilni C-terminus CDK9 se takođe nalazi u blizini aktivnog centra i njegovo prisustvo utiče na tok katalize, regulišući redosled vezivanja supstrata. Za razliku od CDK2, kod koje je vezivanje ATP i supstrata nasumično, CDK9 uvek prvo vezuje ATP, zatim vezuje, fosforiliše i otpušta svoj fosforilisani supstrat, da bi na kraju iz njenog aktivnog centra

disosovao ADP (86). Ovim mehanizmom se sprečava sukcesivno fosforilisanje više grupa u neposrednoj blizini na supstratu, što održava specifičnost interakcija sa supstratom (86). Interakcije ATP sa CDK9 su analogne njegovim interakcijama u kompleksu sa ostalim CDK. Kao rezultat, ATP-kompetitivni inhibitori CDK su relativno neselektivni, pa ATP-kompetitivni inhibitori CDK9 istovremeno inhibiraju i druge ciklin zavisne kinaze. Neki od inhibitora CDK9 ovog tipa su flavopiridol, DRB, meriolini, CR8, 12u i drugi (u (141) i reference citirane u radu) (Slika 13).



**Slika 13:** Primeri ATP kompetitivnih inhibitora CDK9

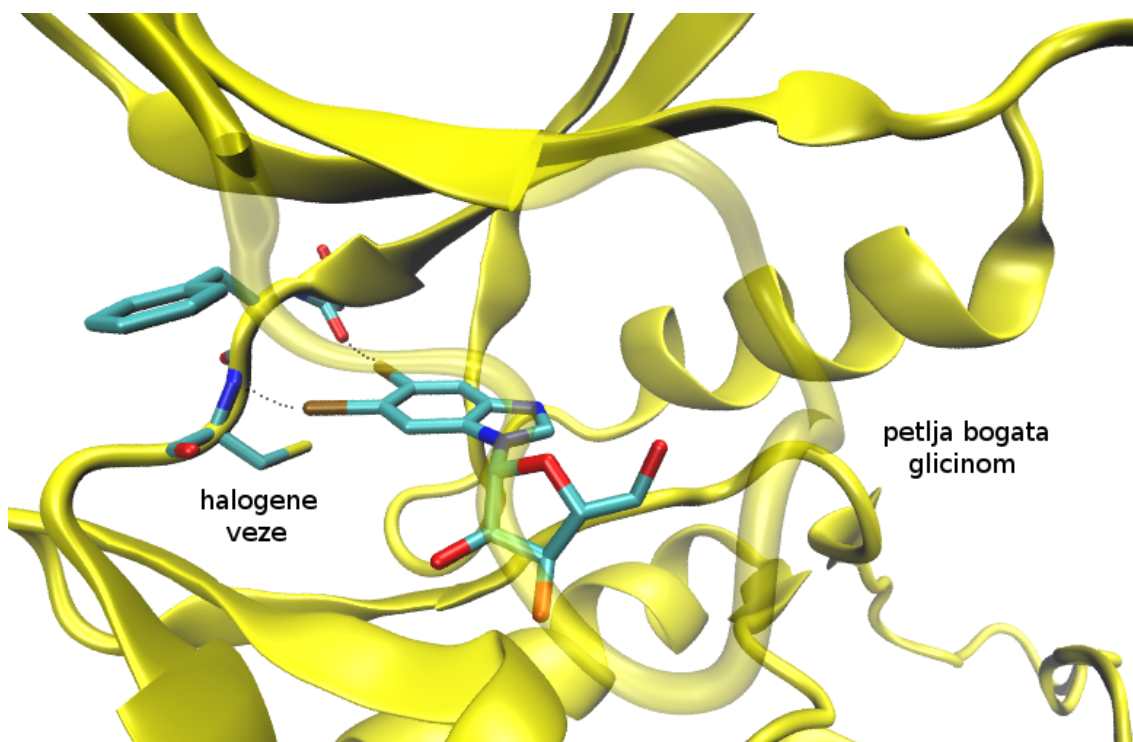
Do danas je razvijeno više od 30 inhibitora CDK9 ovog tipa (141), od kojih su neki u kliničkim istraživanjima (143). Njihov terapijski potencijal leži u terapiji različitih maligniteta. Maligno izmenjene ćelijske populacije odlikuje velika raznolikost u profilu ekspresije gena. Kod ovakvih ćelija, mogućnost istovremene inhibicije više ciklin zavisnih kinaza, ometanje ćelijskog ciklusa, supresija transkripcije i indukcija apoptoze bi smanjila stopu preživljavanja malignih ćelija efikasnije nego inhibicija samo jedne od ciklin zavisnih kinaza (144). Međutim, ograničena selektivnost inhibitora CDK povećava njihove neželjene efekte i toksičnost (105,145). Uz to, u patološkim stanjima u kojima je aktivnost CDK9 glavno ciljno mesto, kao što su virusne infekcije i hipertrofija miokarda, neophodna je primena selektivnijih inhibitora CDK9 aktivnosti (144).

Određen napredak u razvoju selektivnih ATP-kompetitivnih CDK9 inhibitora postignut je zahvaljujući razumevanju molekularnih osnova interakcija poznatih inhibitora sa aktivnim mestom CDK9, kao i strukturnih karakteristika ATP-vezujućeg džepa CDK9 koje je razlikuju od drugih ciklin zavisnih kinaza. Kao rezultat, sintetisani su selektivniji inhibitori, kao što su EXEL-8647 (struktura nije javno dostupna) (146) i CAN508 (slika 13) (147). Proučavanje molekularnih osnova interakcija selektivnijih ATP-kompetitivnih inhibitora ukazalo je da postoje specifičnosti u građi CDK9 koje se mogu iskoristiti za poboljšanje selektivnosti dejstva inhibitora.

CDK9 odlikuje veća fleksibilnost aktivnog mesta i mogućnost akomodacije većih molekula inhibitora u poređenju sa drugim CDK (148,149). Na ulazu u ATP-vezujući džep nalazi se tzv. „površina za specifičnost“ koju odlikuje prisustvo različitih aminokiselina kod različitih CDK. U CDK9 se tu nalaze male, nepolarne aminokiseline Ala 153 i Gly 112, koje dozvoljavaju prisustvo voluminoznih supstituenata u inhibitorima koji ne bi bili tolerisani u drugim CDK (141,150,151). Inhibitori iz serije 2,4 - disupstituisanih pirimidina (12u, slika 13) ispoljavaju selektivnost ka CDK9 usled veće prilagodljivosti aktivnog mesta njihovim aktivnim konformacijama i optimizovanim ostvarenim hidrofobnim interakcijama (148).

Kod CDK9 promena konformacije petlje bogate glicinom stabilizovana je pomeranjem susedne  $\beta$ 3- $\alpha$ C petlje (105,141,150). Indukcija neaktivne konformacije kinaze promenom pozicije petlje bogate glicinom je ranije opisana kod drugih kinaza. Petlja bogata glicinom preklapa aktivni centar CDK9 u prisustvu inhibitora pri čemu

aminokiselina Phe 30 uspostavlja dodatne kontakte sa inhibitorom. Ovo je primećeno u kompleksu CDK9 sa inhibitorima flavopiridolom (83), DRB (152), 2,4-disupstituisanim pirimidinima (148) CR8 (153) i CAN508 (154) (strukture inhibitora na slici 13). Ovu konformacionu promenu omogućuje prisustvo aminokiseline Leu 51 u aminokiselinskoj sekvenci CDK9. Leu 51 ispunjava hidrofobni džep nastao pomeranjem Phe 30 i stabilizuje konformaciju. U ostalim CDK se na analognoj poziciji Leu 51 nalaze bazne aminokiseline, koje su previše polarne da bi omogućile ovu stabilizaciju (150).



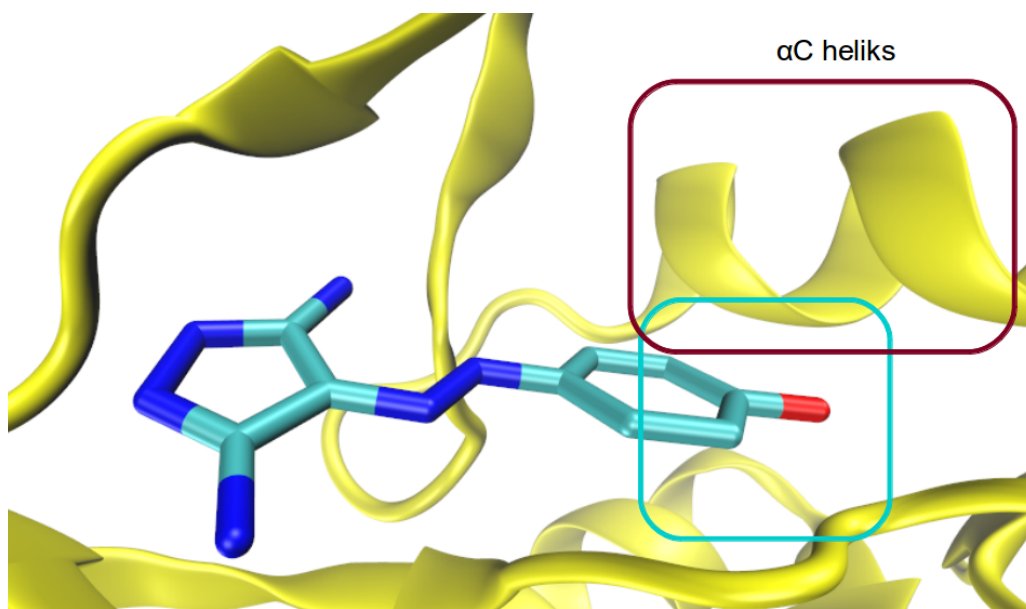
**Slika 14:** Inhibitor DRB u kompleksu sa CDK9. Atomi hlora (oker) grade halogene veze, dok petlja bogata glicinom prekriva inhibitor

U aktivnom centru CDK9 postoji mogućnost građenja halogenih veza sa molekulima inhibitora. Inhibitor DRB prilikom vezivanja za CDK9 gradi halogene veze analogne vodoničnim vezama koje nastaju prilikom vezivanja ATP-a za CDK proteine (152) (Slika 14). Međutim, veći sterni zahtevi halogenih atoma u odnosu na atome vodonika diktiraju selektivnost DRB za CDK9, čiji aktivni centar omogućava formiranje ovih veza, za razliku od aktivnog mesta CDK2 (152).

Dodatnu selektivnost u svom dejstvu inhibitori CDK9 mogu da ostvare interakcijama sa aminokiselinama C-terminusa CDK9. C-terminus CDK9 se nastavlja na njen kinazni domen i slabo je evolutivno očuvan. U aktivnoj konformaciji CDK9,

C-terminus se nalazi u neposrednoj blizini aktivnog mesta i utiče na mehanizam katalize (86). Pokazano je da neki od ATP-kompetitivnih inhibitora stupaju u interakcije sa C-terminusom, prilikom kojih dolazi do stabilizacije kompleksa CDK9 sa inhibitorima (152). Ovo je pretežno slučaj sa inhibitorima sa voluminoznijim supstituentima. Interakcija sa C-terminusom CDK9 nije od značaja kod svih grupa inhibitora, što je pokazano na seriji inhibitora na bazi 2,4-disupstituisanih pirimidina (148).

Konformaciona fleksibilnost aktivnog centra CDK9 obuhvata i položaj  $\alpha$ C heliksa. Kao posledica relativno male vezivne površine CDK9 / ciklin T1 kompleksa (oko  $900\text{\AA}^2$ )  $\alpha$ C heliks CDK9 je pokretljiviji, dajući inhibitorima priliku za dodatne interakcije u aktivnom centru i povećanu selektivnost (154). Inhibitor CAN508 prilikom svog vezivanja koristi hidrofobni džep nastao pomeranjem  $\alpha$ C heliksa (Slika 15).

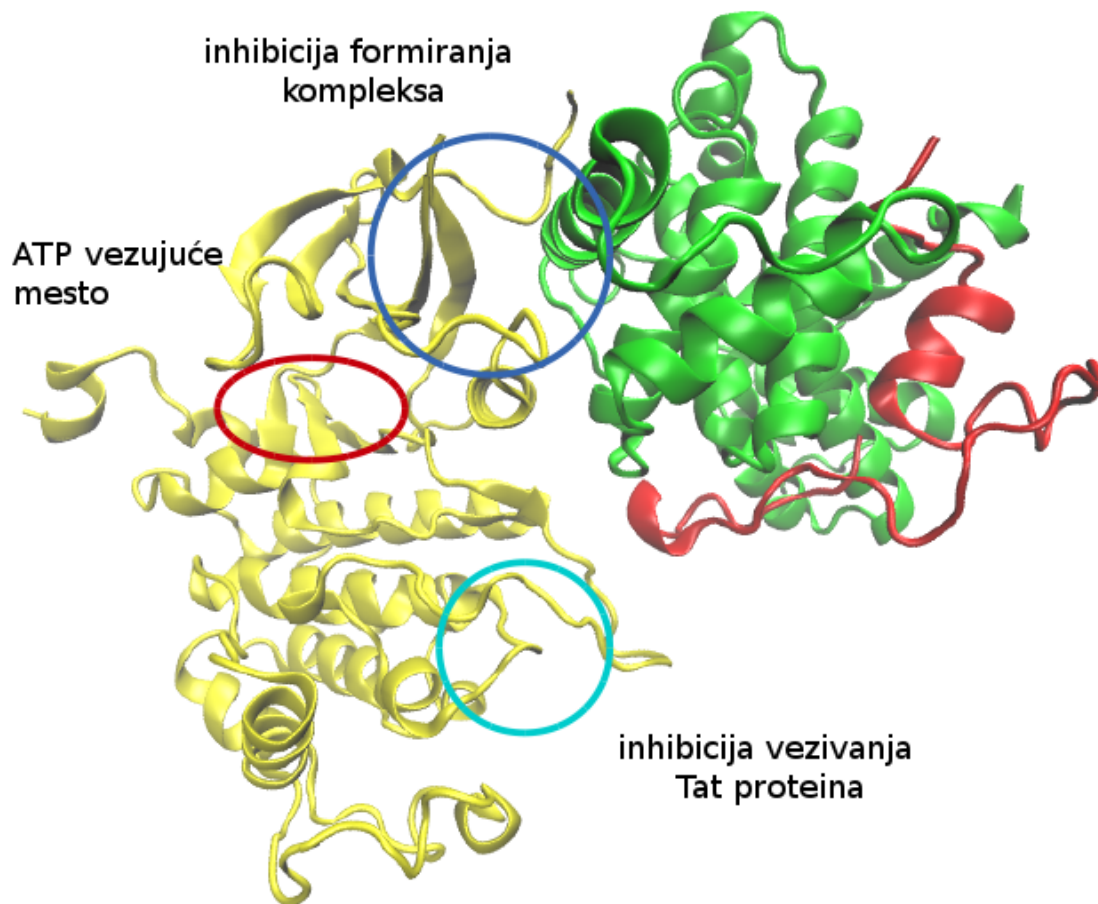


**Slika 15:** Inhibitor CAN508 u kompleksu sa CDK9 i džep nastao pomeranjem heliksa (uokviren)

Uprkos svemu navedenom, visokospecifični inhibitor CDK9 još uvek nije razvijen. Razvoj takvog inhibitora značajno otežava evolutivna očuvanost ATP-vezujućeg mesta u CDK familiji (106,141), usled koje je veliki broj istih strukturnih karakteristika vezivnog mesta prisutan kod različitih CDK. Ako se uz to uzme u obzir da ATP-kompetitivni inhibitori CDK9 moraju da ostvare inhibiciju u prisustvu visokih intracelularnih koncentracija ATP (145), alternativni, ATP-nezavisni pristupi inhibiciji aktivnosti P-TEFb kompleksa dobijaju na značaju (155).

### *ATP nekompetitivni inhibitori*

Alternativni pristup dizajnu inhibitora CDK9 predstavlja ometanje formiranja aktivnog P-TEFb kompleksa. Ova strategija je ranije već uspešno primenjena na druge enzime (156). Potencijalna ciljna mesta za razvoj ATP-nekompetitivnih inhibitora aktivnosti različitih P-TEFb kompleksa na površini CDK9 uključuju površinu za vezivanje ciklina i region T-petlje sa kojim Tat protein stupa u interakciju (Slika 16).



**Slika 16:** Ciljna mesta za razvoj ATP nekompetitivnih inhibitora aktivnosti P-TEFb kompleksa na površini CDK9 (žuto) u kompleksu sa ciklinom T1 (zeleno) i Tat proteinom (crveno). Aktivni centar CDK9 uokviren je crvenom.

U kontekstu CDK9, inhibicija formiranja kompleksa upotrebljena je na aktivnom P-TEFb-Tat kompleksu koji se formira prilikom replikacije virusa HIV-1 interakcijom virusnog Tat proteina sa P-TEFb kompleksom. Kako Tat, prilikom interakcije sa P-TEFb kompleksom ostvaruje kontakte sa T-petljom CDK9 (125),



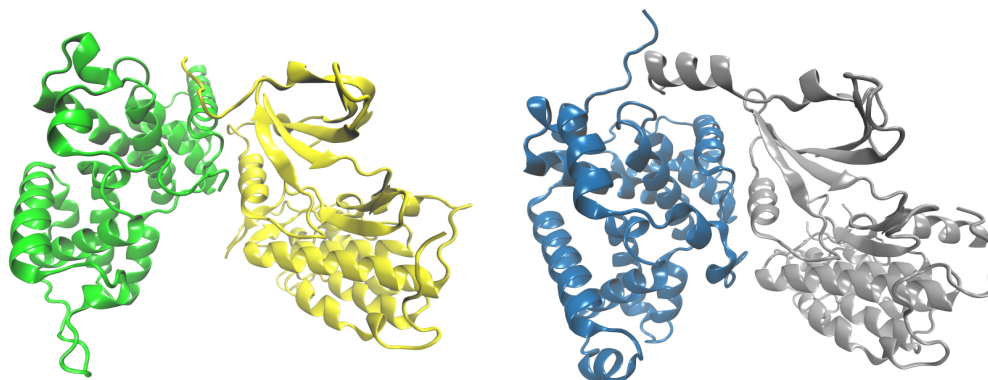
razvijena su dva inhibitora koja se vezuju za površinu CDK9. Ovi inhibitori menjaju konformaciju T-petlje i onemogućavaju formiranje aktivnog P-TEFb-Tat kompleksa, ometajući virusnu transkripciju, bez uticaja na ćelijsku aktivnost P-TEFb kompleksa (157). Da bi se postigla inhibicija ćelijske aktivnosti P-TEFb kompleksa, bilo bi potrebno ometati formiranje samog jezgra P-TEFb kompleksa, sprečavanjem vezivanja CDK9 za ciklin T1. Indirektno, to je postignuto primenom geldanomicina i ometanjem interakcija CDK9 i citoplazmatskih proteina pratioaca (81), čime je sprečeno pravilno formiranje tercijarne strukture CDK9 i formiranje P-TEFb kompleksa. Sličan efekat bi se mogao postići i direktnom inhibicijom CDK9/ ciklin T1 protein-protein interakcije. Dokaz za to predstavlja inhibicija aktivnosti CDK2 / ciklin A kompleksa, primenom peptidnog inhibitora nazvanog C4 koji ometa ovu protein-protein interakciju, vezujući se za CDK2 / ciklin A kompleks i blokirajući kompleks u neaktivnom obliku (45). Pomenuti inhibitor je razvijen na osnovu sekvence jednog od heliksa ciklina A i sastoji se od 22 aminokiseline, ali je efekat moguće postići i znatno manjim molekulima. Heksapeptid NBI1, dobijen skriningom peptida takođe inhibira nastanak CDK2 / ciklin A kompleksa (158). Oba peptida, i C4 i NBI1 su pokazala antiproliferativno dejstvo u tumorskim ćelijskim linijama (45,158). Zbog peptidne prirode C4 i NBI1, njihovo dejstvo je ograničeno stepenom unosa u ćelije, pa pokazuju značajno bolji efekat ako se koriste spregnuti sa nekim nosačem (npr. Tat proteinskom sekvencom) (45,158).

Inhibitori protein-protein interakcija CDK9 / ciklin T1 kompleksa bi postigli veću selektivnost dejstva od ATP-kompetitivnih inhibitora aktivnosti CDK9 i mogli bi da se primenjuju u svim terapijskim oblastima u kojima je neophodno regulisati aktivnost P-TEFb kompleksa, za razliku od inhibitora formiranja P-TEFb/Tat kompleksa čija primena je ograničena na HIV-1 infekciju. Već otkriveni inhibitori formiranja kompleksa drugih CDK i njihovih ciklina opravdavaju osnovnu ideju ovog pristupa inhibiciji aktivnosti P-TEFb kompleksa, dok na bitne elemente u samom procesu dizajna inhibitora upućuju skorašnja saznanja o molekularnoj građi i karakteristikama samih proteinskih kompleksa i protein-protein interakcija.

### 1.3 Protein-protein interakcije prisutne u P-TEFb kompleksu

Poznavanje trodimenzionalne strukture proteinskih kompleksa predstavlja polaznu tačku za studiju interakcija u njima na molekularnom nivou. Dodatan izvor informacija predstavljaju biohemijske i genetičke studije, kao što su studije tačkastih mutacija na proteinima, ali je pouzdan strukturni prikaz kompleksa neophodan za *in silico* simulacije samog kompleksa. U nedostatku eksperimentalnih struktura moguće je dobiti uvid u strukturu kompleksa homologim modeliranjem. Strukture dobijene ovom tehnikom su, međutim, manje pouzdane od eksperimentalno rešenih, koje se najčešće dobijaju primenom kristalografije i NMR (nuklearna magnetna rezonanca) metoda.

Trodimenzionalna struktura CDK9 / ciklin T1 kompleksa prvi put je kristalografski rešena 2008. godine (83). Dve godine kasnije rešena je struktura ovog kompleksa vezanog za Tat protein HIV-1 virusa (125). Ove dve strukture pokazale su da je kompleks CDK9 / ciklin T1 građen analogno kompleksima drugih CDK sa njihovim ciklinima, ali i da postoje određene razlike. Pregled sličnosti i razlika se može naći u referenci (83). Najveća razlika leži u manjem broju kontakata između CDK9 i ciklina T1 i orijentaciji ciklina. Ciklin T1 je u dodiru samo sa N-terminalnim domenom CDK9 i zarotiran je  $26^\circ$  u odnosu na položaj drugih ciklina u njihovim kompleksima sa CDK ćelijskog ciklusa. Još jedna transkripciona CDK, CDK8, pokazuje sličnu građu svog kompleksa sa ciklinom C, sa umerenijom rotacijom položaja ciklina ( $10^\circ$ ) (159).



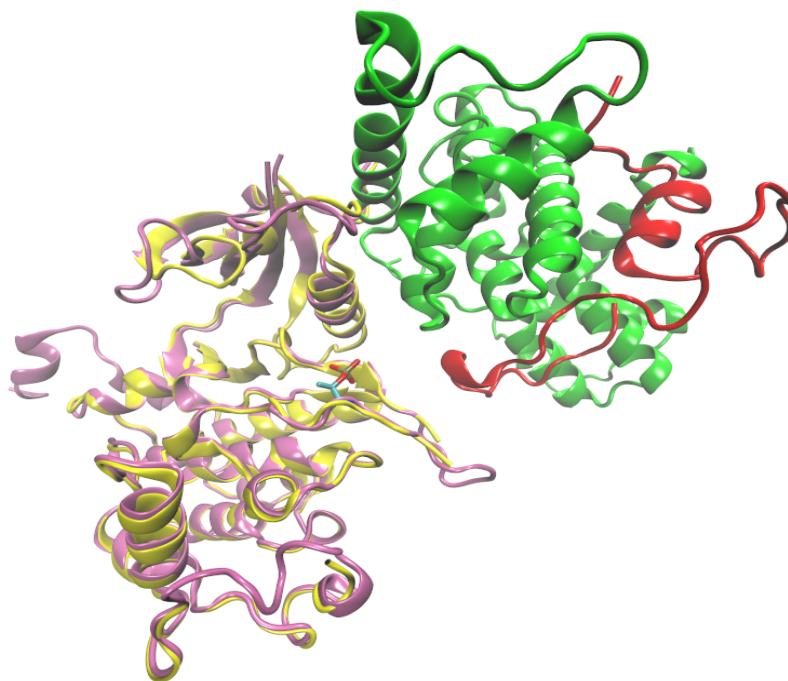
**Slika 17:** Kompleksi CDK9 / ciklin T1 (žuto/zeleno) i CDK8 / ciklin C (sivo/plavo)

I CDK9 i CDK8 imaju produžen N-terminus u odnosu na ostale članove CDK familije i njime stupaju u interakcije sa svojim ciklinima. N-terminus CDK8 je jasno

struktuiran u heliks u kompleksu sa ciklinom C i povećava vezivnu površinu kompleksa za oko  $700 \text{ \AA}^2$  (159), dok je N-terminus CDK9 nestruktuiran u kompleksu (Slika 17).

Manji broj kontakata između CDK9 i ciklina T1 daje ovom kompleksu veću konformacionu fleksibilnost (83). Uprkos tome, osnovna vezivna površina između CDK9 i ciklina T1 je strukturno analogna površinama prisutnim u interakcijama drugih ciklin zavisnih kinaza i njihovih ciklina. Čini je  $\alpha$ C heliks kinaza koji interaguje sa H3 – H5 heliksima ciklina. Zbog evolutivne očuvanosti aminokiselina koje ulaze u interakciju sa CDK9 između ciklina T1, T2 i K, kompleksi CDK9 sa ciklinima T2 i K verovatno imaju analognu strukturu rešenom kompleksu sa ciklinom T1 (83).

Smanjen broj kontakata između subjedinica CDK9/ciklin T1 kompleksa takođe omogućava olakšanu promenu konformacije kompleksa pod dejstvom regulatornih proteina (83). Tat protein HIV-1 virusa takođe deluje na ovaj način kada se veže za kompleks. Tat protein najvećim delom interaguje sa ciklinom T1, ali gradi i kontakte sa T-petljom CDK9. Vezivanje Tat proteina menja orijentaciju subjedinica kompleksa, dovodi do promena u vodoničnim vezama koje gradi fosfotreonin 186 i alosterno utiče na oblik ATP-vezujućeg mesta (125) (Slika 18).



**Slika 18:** Kompleks CDK9 / ciklin T1 / Tat (ružičasto/zeleno/crveno) i CDK9 iz kompleksa bez Tat proteina (žuto). Fosfotreonin 186 je prikazan štapićima.

Danas se u RCSB PDB (eng. *Research Collaboratory for Structural Bioinformatics Protein Databank*) bazi podataka (160) koja služi za deponovanje rešenih struktura makromolekula nalazi 17 kristalografskih struktura ljudskih CDK9 / ciklin T1 kompleksa. Ove strukture se međusobno razlikuju po prisustvu ili odsustvu inhibitora, liganada ili drugih proteina (Tabela 1).

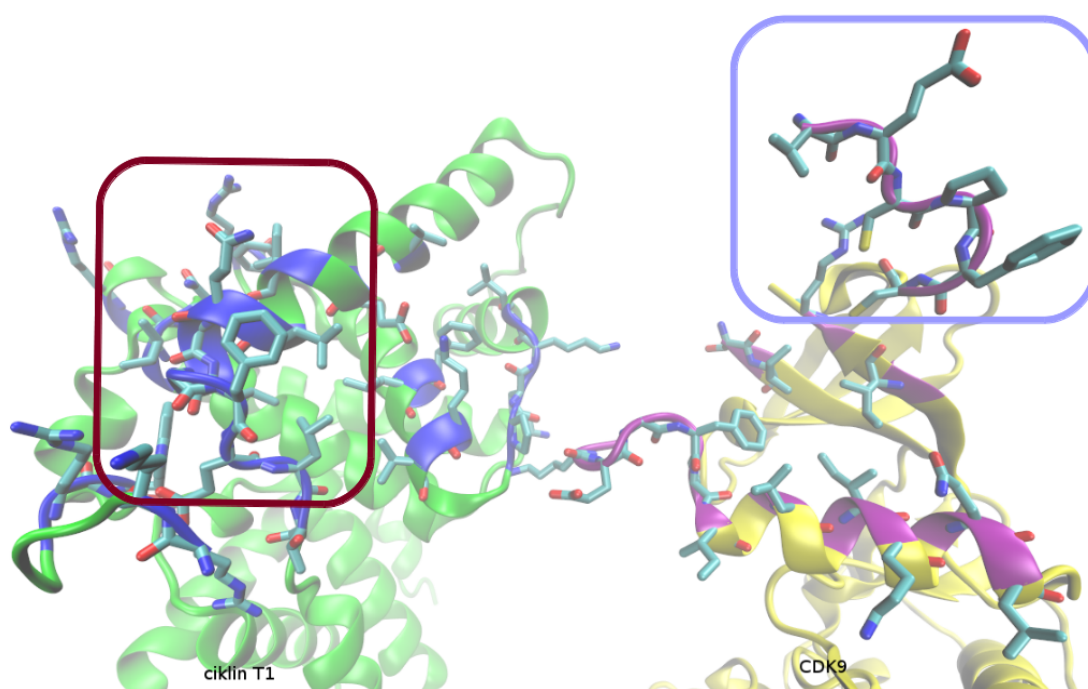
**Tabela 1:** Kristalografske strukture CDK9 / ciklin T1 kompleksa

<b>PDB ID</b>	<b>opis</b>	<b>literatura</b>
3blh	CDK9 / ciklin T1 kompleks	(83)
3blq	CDK9 / ciklin T1 kompleks, ATP	
3blr	CDK9 / ciklin T1 kompleks, flavopiridol	
3lq5	CDK9 / ciklin T1 kompleks, S-CR8	(161)
3tni	CDK9 / ciklin T1 kompleks, mutacija F241L	(86)
3tn8	CDK9 / ciklin T1 kompleks, CAN508	
3tnh	CDK9 / ciklin T1 kompleks, CAN508	
4ec8	CDK9 (puna dužina) / ciklin T1 kompleks, ATP	
4ec9	CDK9 (puna dužina) / ciklin T1 kompleks	
3mi9	CDK9 / ciklin T1 kompleks, HIV-1 Tat	(125)
3mia	CDK9 / ciklin T1 kompleks, ATP, HIV-1 Tat	
3my1	CDK9 / ciklin T1 kompleks, DRB	
4bcf	CDK9 / ciklin T1 kompleks, pirimidinski inhibitor	(148)
4bch	CDK9 / ciklin T1 kompleks, pirimidinski inhibitor	
4bci	CDK9 / ciklin T1 kompleks, pirimidinski inhibitor	
4bcj	CDK9 / ciklin T1 kompleks, pirimidinski inhibitor	
4imy	CDK9 / ciklin T1 kompleks, AFF4 protein	(162)

U svim do sada rešenim kristalografskim strukturama CDK9 se nalazi vezana za ciklin T. Vezivanje za ciklin aktivira CDK9 i stoga je u svim kristalografskim strukturama CDK9 u nekoj od konformacija koje odgovaraju aktivnoj kinazi. Vezivanje Tat-proteina za kompleks CDK9 / ciklin T1 dalje menja konformaciju CDK9 i povećava aktivnost kompleksa (125). U nekim od kristalografskih struktura inhibitori dovode do konformacionih promena u petlji bogatoj glicinom iznad aktivnog mesta CDK9, dovodeći petlju u položaj koji je opisan u neaktivnim apo kinazama. Ove male

konformacione promene, međutim, ne odslikavaju konformaciju CDK9 u njenom neaktivnom, slobodnom obliku. Uvid u izgled apo forme CDK9, koja postoji u ćelijama pre vezivanja za cikline, može se za sada dobiti samo preko njenih teorijskih modela, *in silico* simulacijama. Konformacija slobodnog ciklina T1 se ne razlikuje značajno od njegove konformacije u kompleksu sa CDK9 (83).

Presek vezivnih površina kompleksa CDK9/ciklin T1 pokazuje da u formiranju kompleksa pretežno učestvuju kratki linearni segmenti sa oba proteina. Izuzetak predstavlja N-terminus CDK9 (aminokiseline 8–13) i sekvenca aminokiselina 141–148 sa ciklina T1 (Slika 19).



**Slika 19:** Vezivne površine kompleksa CDK9/ciklin T1 (žuto/zeleno). Segmenti proteina koji učestvuju u izgradnji vezivne površine su obeleženi plavo/ljubičasto. Najvažniji linearni peptidni segmenti su uokvireni.

Pored opisanih, u P-TEFb kompleksu se još javljaju interakcije katalitičkog jezgra sa HEXIM i 7sk snRNK subjedinicama pri formiranju neaktivnog oblika kompleksa, ali trodimenzionalna struktura ovog oblika P-TEFb kompleksa još uvek nije precizno rešena.

## 2. Rezultati i diskusija

### 2.1. Inhibicija CDK9 / ciklin T1 interakcije

#### 2.1.1. Studija P-TEFb

Cilj ove doktorske disertacije je dizajn direktnih inhibitora proteinske interakcije CDK9 / ciklin T1. Inhibicija funkcije CDK9 / ciklin T1 kompleksa je od značaja u terapiji različitih patoloških stanja kod ljudi (neki tipovi tumora, HIV-1 infekcija, hipertrofija miokarda). Proučavanje strukture CDK9 / ciklin T1 proteinskog kompleksa pružiće uvid u značajne strukturne karakteristike vezivnih površina kompleksa i omogućiti identifikaciju najznačajnijih aminokiselinskih kontakata između ova dva proteina. Prvi korak u dizajnu inhibitora predstavljaju strukture kratkih peptida (5 – 8 aminokiselina) dobijene oponašanjem interakcija iz CDK9 / ciklin T1 kompleksa. Analiza energija vezivanja dobijenih peptida, zajedno sa informacijama o strukturi vezivnih površina samog CDK9 / ciklin T1 kompleksa biće upotrebene za dizajn nepeptidnih molekula, potencijalnih inhibitora CDK9 / ciklin T1 interakcije. U ovom poglavlju biće predstavljena strukturna i energetska analiza vezivnih površina CDK9 / ciklin T1 kompleksa.

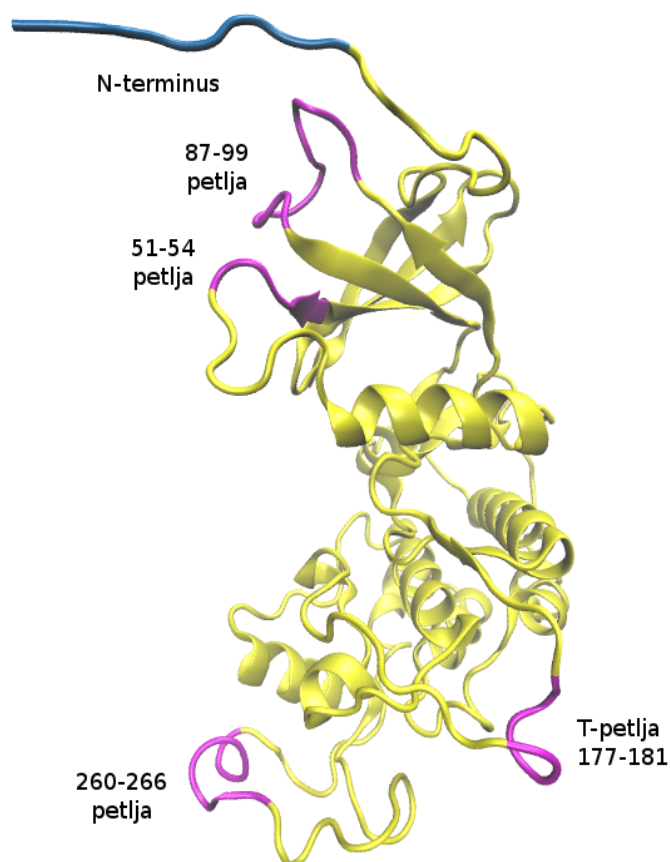
##### 2.1.1.1. Priprema kompleksa

Za *in silico* dizajn direktnih inhibitora CDK 9 / ciklin T1 interakcije korišćena je kristalografska struktura PDB ID: 3blh (83) koja predstavlja kompleks ciklina T1 sa CDK9\_42 izoformom CDK9 sa skraćenim C terminusom, bez liganada ili inhibitora. CDK9 u ovoj strukturi sadrži prvih 325 od 372 aminokiseline u građi CDK9. C-terminus CDK9 ne igra ulogu u interakciji kinaze sa ciklinom i njegovo odsustvo ne utiče na rezultate ove studije. Proteinske sekvence CDK9 i ciklina T1 u strukturi 3blh nisu potpuno identične zvaničnim sekvencama ovih proteina. Sekvenca CDK9 ne započinje metioninom, već glicil-prolinom, što je ostatak od procesa eksprimiranja ljudskog proteina u ćelijama insekata. Ova mutacija ne utiče značajno na strukturu proteina, jer se nalazi na samom kraju neuređenog i vrlo pokretljivog N-terminusa kinaze, čija elektronska gustina (a time i koordinate atoma) nije rešena ni u jednoj

kristalografskoj strukturi CDK9. N-terminus CDK9 je neuređen i u odsustvu ove mutacije. U sekvenci ciklina T1 prisutne su tri mutacije Gln77Arg / Glu96Gly / Phe241Leu za koje je pokazano da ne utiču na izgled vezivnih površina kompleksa. Mutacija Gln položaja 77 sa jedne od petlji ciklina T1 u Arg je mutacija koja se prirodno javlja u strukturi ciklina T1. Mutacija Phe položaja 241 u Leu doprinosi formiranju kristala kompleksa i nije blizu vezivnih površina kompleksa (154). Za mutaciju Glu položaja 96 do Gly je pokazano da ne utiče značajno na konformaciju kompleksa i njegovu katalitičku aktivnost (154).

Kristalografska rezolucija strukture 3blh je 2.48Å, a sama struktura zadovoljava preporučene vrednosti kvaliteta kristalografskih parametara koji pokazuju koliko dobro dobijeni model odgovara eksperimentalnim činjenicama ( $R < 0.25$ ,  $R\text{-free} < 0.4$  i  $R\text{-free} - R = 0.042 < 0.05$  (163)). Kako su do sada sve kristalografske strukture CDK9 / ciklin T1 kompleksa rešene iz kristala sa istim tipom kristalne rešetke nije bilo moguće proveriti uticaj pakovanja u kristalu na konformaciju kompleksa. Kontakti između susednih molekula proteina u kristalu mogu da izmene konformaciju molekula, ali nijedna od aminokiselina u blizini vezivnih površina proteina od interesa u ovom istraživanju ne učestvuje u formiranju takvih kontakata, što čini strukturu 3blh pogodnom za proučavanje CDK9 / ciklin T1 kompleksa.

Četiri fleksibilne proteinske petlje CDK9 u kristalografskoj strukturi 3blh nemaju rešenu konformaciju zbog nedovoljne elektronske gustine. Ove petlje uključuju aminokiseline položaja 51–54 (petlja pre  $\alpha$ C heliksa), 88–97, 177–181 (T-petlja) i 260–266. Petlje su dodate iz odgovarajućih CDK9 struktura ili modelovane homologim modeliranjem, po protokolu opisanom u odeljku 4.1.1 (Slika 20).



**Slika 20:** Položaj modelovanih petlji (ružičasto) i N-terminusa (plavo) na strukturi CDK9 (žuto)

T-petlja i petlja sa aminokiselinama 260–266 su modelovane na osnovu konformacija koje imaju u P-TEFb/Tat kompleksu. Iako vezivanje Tat proteina za P-TEFb dovodi do konformacionih promena u P-TEFb kompleksu, ove promene su lokalizovane na ciklinu T1 i regionu oko petlje 51–54 na CDK9 i ne zahvataju oblasti preuzetih petlji (125). Fosfatna grupa fosforilisanog Thr 186 iz T-petlje CDK9 je uklonjena pre simulacije. Thr 186 nije fosforilisan u neaktivnoj, slobodnoj CDK9.

Stanje protonacije svih aminokiselina u kompleksu određeno je na fiziološkom pH od 7.4, uzimajući u obzir njihovo hemijsko okruženje. Koordinate atoma prvih 7 aminokiselina N-terminusa CDK9 i prvih 5 aminokiselina N-terminusa ciklina T1 nisu rešene u kristalografskoj strukturi 3blh. Kao rezultat, pozitivno naelektrisanje koje N-terminusi proteina obično nose na fiziološkom pH je izostavljeno sa N-terminusa (Arg 5) ciklina T1 u simulaciji P-TEFb kompleksa, ali je zadržano na N-terminusu (Val

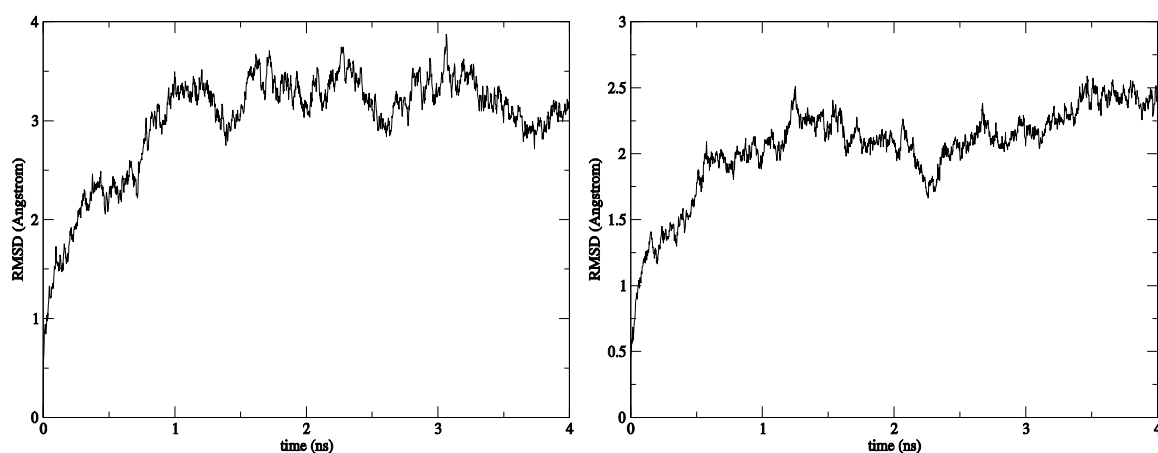


8) CDK9, zbog njegove izloženosti rastvaraču. Bočni lanac Lys 6 ciklina T1 je deprotonovan, zbog nepolarnog okruženja u kome se nalazi. Cys 10 sa površine CDK9 je zadržan u neutralnom obliku, iako je predviđeno da se na pH od 7.4 može naći i jonizovan ( $pK_a = 6.67$ ).

Za simulacije u ovom istraživanju je korišćen čitav kompleks CDK9 / ciklin T1, u studiji vezivnih površina i interakcija na njima.

S obzirom da je za razvoj inhibitora protein-protein interakcija u ovom kompleksu bilo neophodno steći uvid u strukturu CDK9 pre vezivanja ciklina, simulirana je i struktura slobodne CDK9. Struktura slobodne CDK9 je simulirana bez nedostajućih prvih 7 aminokiselina N-terminusa, ali i sa modelovanim N-terminusom u celini, jer se on nalazi u neposrednoj blizini vezivne površine kompleksa. Kao test validnosti uslova simulacije i pogodnost protokola kojim je modelovan N-terminus CDK9, izvršene su dodatne simulacije na CDK8, čiji je N-terminus jasno rešen u kristalografskoj strukturi.

Proteini su za simulacije pripremani po protokolu opisanom u poglavlju Eksperimentalni deo. Priprema je uključivala minimizaciju solvatisanog proteina radi uklanjanja nepovoljnih steričkih efekata ako postoje u strukturi, zagrevanje na temperaturu tela (310 K) i ekvilibraciju struktura molekularnom dinamikom. Simulacije su prekinute posle 4ns, kada je postignuta ekvilibracija sistema (Slika 21).



**Slika 21:** RMSD  $C\alpha$  atoma proteina u zavisnosti od dužine trajanja MD simulacije

### 2.1.1.2 Analiza vezivnih površina

Poznavanje karakteristika vezivnih površina proteinskih kompleksa i interakcija koje se na njima uspostavljaju je neophodno da bi se date interakcije među proteinima oponašale, ili omele. Kao prvi korak u dizajnu inhibitora CDK9 / ciklin T1 interakcije proučene su vezivne površine ovog kompleksa.

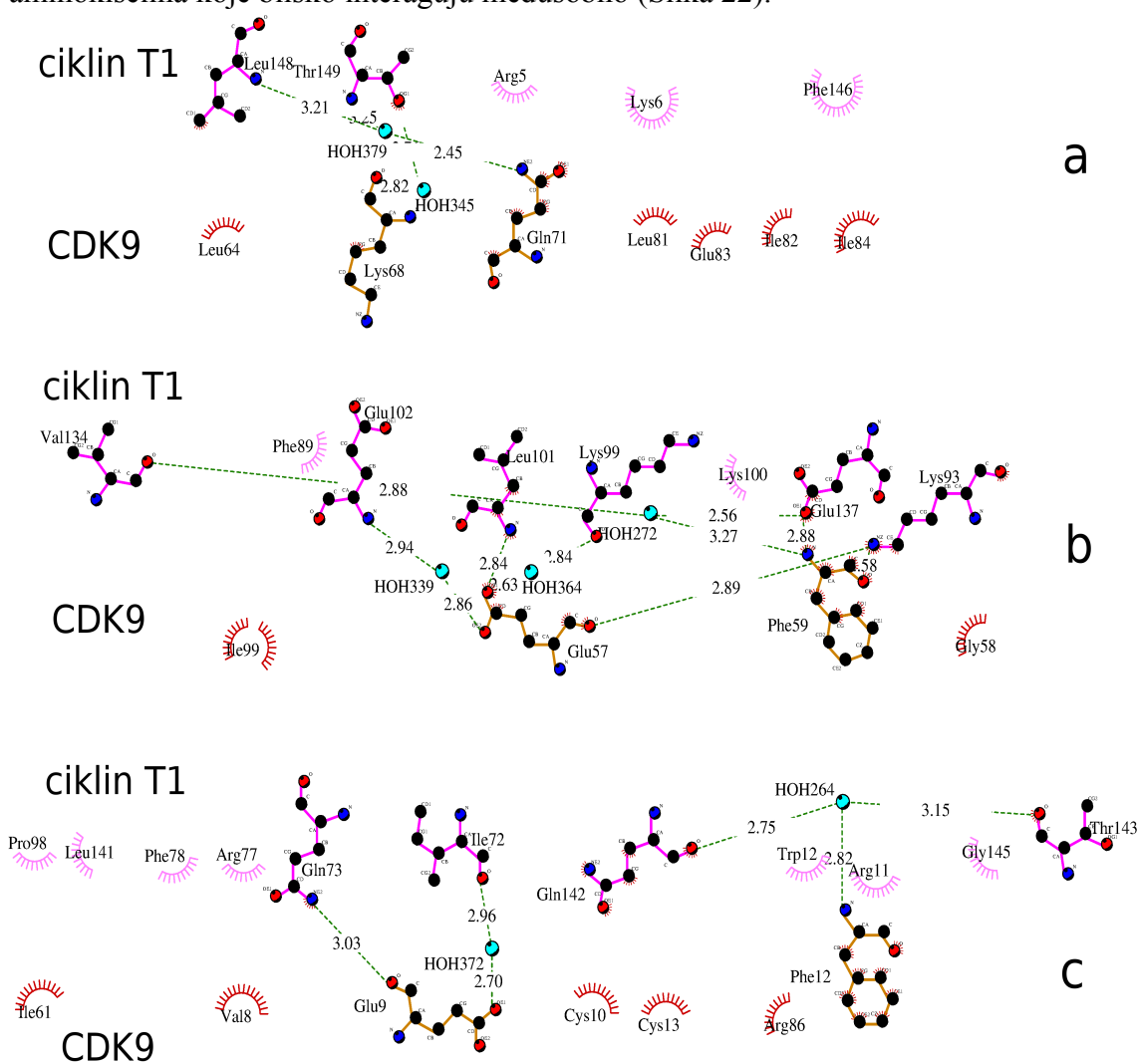
Vezivne površine kompleksa CDK9 / ciklin T1 zauzimaju površinu od  $970\text{\AA}^2$  po proteinu (142). U njihov sastav ulazi 48 – 56 aminokiselina, gde broj varira u zavisnosti od izbora geometrijskih kriterijuma za definiciju vezivne površine kompleksa, tj. od granične udaljenosti među proteinima i izbora atoma za merenje udaljenosti. U kompleksu, vezivna površina CDK9 je pretežno konkavna, a površina ciklina T1 konveksna. Komplementarnost oblika vezivnih površina proteina u kompleksu može se proceniti odnosom zapremine koju obuhvataju vezivne površine i ukupne površine (7). Za kompleks CDK9/ciklin T1 u kristalografskoj strukturi PDB ID: 3blh, ovaj odnos iznosi  $3.72\text{\AA}$ , što odgovara vrednostima za heterodimere ( $3.02 \pm 1.13\text{\AA}$ ) (7). Vezivne površine kristalografske strukture kompleksa su relativno ravne – na njima se ne izdvajaju značajni hidrofobni džepovi.

Kontaktne površine u kompleksu CDK9 / ciklin T1 se grubo mogu podeliti na dva regiona interakcija. U glavnom regionu interakcija interaguje C-heliks CDK9 sa heliksima ciklina T1 i njihovim okruženjem. Dodatni, prostorno odvojeni region interakcija uspostavlja N-terminus CDK9, koji se smešta na površinu ciklina bez kontakata sa ostatkom vezivne površine.

#### *Interakcije aminokiselina vezivnih površina CDK9 / ciklin T1 kompleksa*

Na vezivnim površinama kompleksa najzastupljenije su hidrofobne interakcije. Prisutno je svega nekoliko direktnih intermolekulskih vodoničnih veza (4–6, u zavisnosti od kriterijuma upotrebljenih za definiciju vodonične veze). Uz direktne vodonične veze, prisutno je i nekoliko vodoničnih veza koje se ostvaruju premošćavanjem rastojanja između proteina molekulom vode. Tačan broj ovakvih vodoničnih veza varira u zavisnosti od konformacije proteina i uređenosti prisutnih molekula rastvarača. Vodonične veze u građi polipeptidnih nizova proteina na vezivnoj

površini su adekvatno zaštićene od rastvarača. Prisutan je samo jedan dehidron (eng. *dehydron*), tj. samo jedna vodonična veza između amidnog azota i karbonilnog kiseonika polipeptidnog niza koja nije potpuno zaštićena od napada rastvarača pozicioniranjem nepolarnih atoma bočnih nizova aminokiselina u okruženju (164). Dehidroni su od značaja jer je zamena molekula rastvarača koji se nalaze u dodiru sa njima nepolarnim grupama energetski povoljna zbog pojačavanja polarnih interakcija (u ovom slučaju vodonične veze) u potpuno nepolarnom okruženju. Detaljniji prikaz interakcija na vezivnim površinama CDK9/ciklin T1 kompleksa ukazuje na više grupa aminokiselina koje blisko interaguju međusobno (Slika 22).



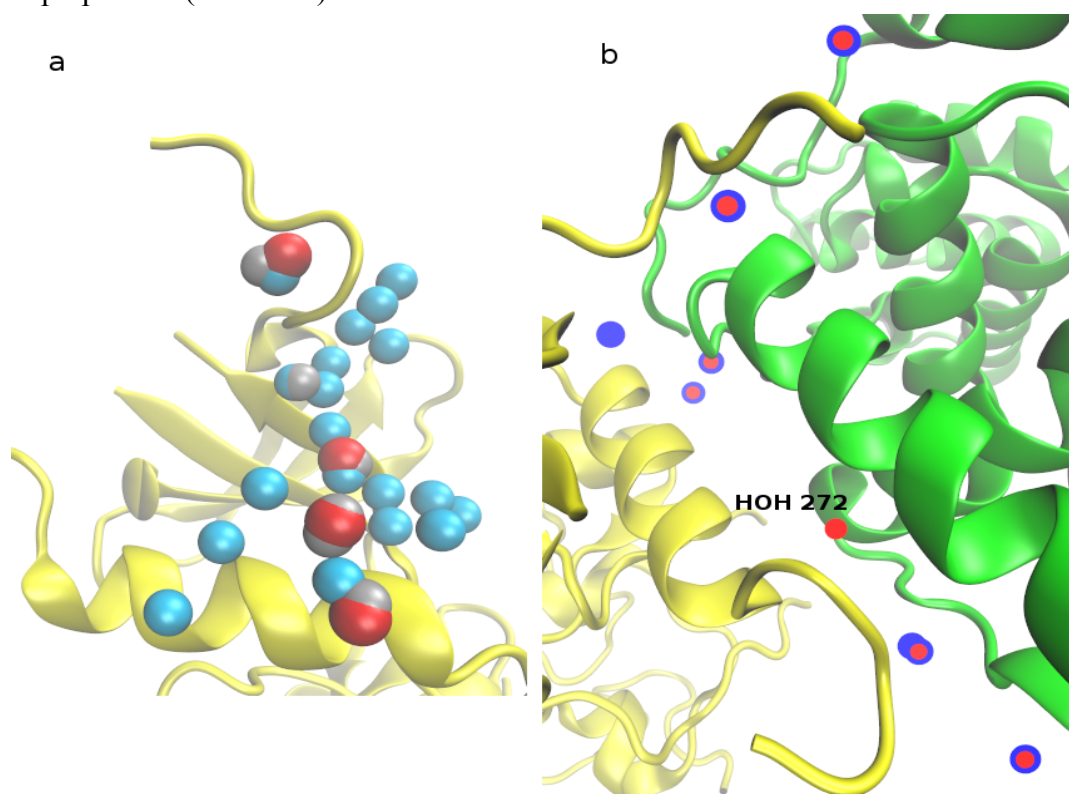
**Slika 22:** Interakcije na vezivnim površinama CDK9/ciklin T1 kompleksa. Aminokiseline iz CDK9 su označene crvenom, a aminokiseline iz ciklina ružičastom bojom. Atomi su obojeni po tipu elementa. Prikazani su molekuli vode (svetloplavo), označeni kao u PDB ID: 3blh strukturi, hidrofobne interakcije (zracima) i vodonične veze sa međuatomskim rastojanjima (zeleno).

Od grupa interakcija prisutnih na vezivnim površinama kompleksa CDK9 / ciklin T1, izdvaja se mreža direktnih vodoničnih veza koju formiraju aminokiseline Lys 93 i Glu 137 sa ciklina T1 i polipeptidni niz aminokiselina Glu 57 i Phe 59 sa CDK9 (slika 22b). Aminokiseline Glu 57 i Phe 59 se nalaze na petlji pre C-heliksa CDK9 ( $\beta 3-\alpha C$  petlja), čije reorganiziranje po formiranju kompleksa dovodi do aktivacije kinaze. Druga grupa interakcija od interesa na vezivnoj površini kompleksa jeste grupacija hidrofobnih aminokiselina u čijem centru se nalazi Phe 146 sa ciklina T1 (slika 22a). Hidrofobni region koji formiraju ove aminokiseline prisutan je u analognom obliku i u kompleksima drugih CDK i njihovih ciklina, ali se fenilalanin koji formira njegovo jezgro obično nalazi kao sastavni deo heliksa sa ciklina, što kod kompleksa CDK9/ciklin T1 nije slučaj. Phe 146 ciklina T1 je izmešten i nalazi se na okretu koji prati heliks. Od direktnih vodoničnih veza prisutna je još samo vodonična veza koja se formira između Gln 73 ciklina T1 i polipeptidnog niza Glu 9 iz N-terminusa CDK9 (slika 22c). U ostalim vodoničnim vezama sa vezivnih površina učestvuju molekuli vode.

#### *Hidratisanost vezivnih površina CDK9 / ciklin T1 kompleksa*

Prisustvo molekula vode na vezivnoj površini nekog proteinskog kompleksa mora biti pažljivo provereno jer su greške u kristalografskim strukturama moguće zbog mobilnosti molekula vode, nedovoljne rezolucije ili različitih uslova dobijanja kristala (163). Zbog toga su u ovoj studiji razmatrani samo molekuli vode koji su prisutni bar u dve kristalografske strukture kompleksa CDK9/ciklin T1. Molekuli vode smatrani su identičnim ako im se, po upoređivanju koordinata atoma, atomi kiseonika nalaze na udaljenosti manjoj od 1 Å (165). Kao ilustracija zavisnosti položaja molekula vode od kristalografske strukture koja se koristi može da posluži procena hidratisanosti vezivnih površina kompleksa CDK9 / ciklin T1 izolovanih bez prisustva drugih proteina i vezivnih površina istog kompleksa iz svih 17 rešenih kristalografskih struktura u kojima se ovaj kompleks javlja, sam ili u interakciji sa drugim proteinima, pretežno Tat proteinom. Poznato je da u interakciji CDK9 / ciklin T1 kompleksa sa Tat proteinom dolazi do konformacionih promena u kompleksu, što može uticati na raspored molekula vode. Rezultati na ograničenoj oblasti vezivne površine od interesa za razvoj liganada to

i potvrđuju. Od pet primećenih molekula vode u ovoj oblasti, jedan je prisutan samo u strukturama u kojima je P-TEFb kompleks vezan za Tat protein (Slika 23a). Zanimljivo je da je WaterDock (165), *in silico* metod za predviđanje pozicije molekula vode na površini proteina, uspešno predvideo poziciju i molekula vode prisutnog samo u Tat-vezanim konformacijama kompleksa. U razmatranje su uzeti molekuli vode čiji se atomi kiseonika nalaze u prvoj hidratacionoj sferi CDK9 (atomi na rastojanju manjem od 3.3 Å od proteina) (165,166). Kao rezultat, dobijeno je 8 lokacija sa velikom verovatnoćom za nalaženje molekula vode na vezivnoj površini kompleksa CDK9/ciklin T1. Ako se lista očuvanih kristalografskih molekula vode na vezivnoj površini kompleksa uporedi sa listom molekula vode za koje je moguće da indirektno ulaze u građu vezivne površine formiranjem premošćavajućih vodoničnih veza u kompleksu (7 molekula, slika 22) može se primetiti da se dve podgrupe podudaraju, ali ne u potpunosti (Slika 23b).



**Slika 23:** Hidratisanost vezivne površine kompleksa CDK9 / ciklin T1 (žuto / zeleno). a) Poređenje relevantnih očuvanih pozicija molekula vode iz dve grupe kristalografskih struktura (sa Tat proteinom – sivo / bez Tat proteina – crveno) sa predviđenim pozicijama vode WaterDock metodom (plavo) b) Poređenje molekula vode iz kristalografske strukture 3blh koji grade vodonične veze sa proteinima kompleksa (crveno) sa očuvanim pozicijama molekula vode iz kristalografskih struktura P-TEFb kompleksa bez prisutnog Tat proteina (plavo).

Od sedam molekula vode koji su u poziciji da grade vodonične veze sa aminokiselinama oba proteina i da tako premošćavaju vezivne površine, šest odgovara očuvanim pozicijama molekula vode iz kristalografskih struktura kompleksa. Izuzetak je molekul vode koji je u kristalografskoj strukturi PDB ID: 3blh obeležen brojem 272 i koji iako može da formira tri vodonične veze sa atomima aminokiselina Phe 59 iz CDK9 i Val 134 i Glu 137 sa ciklina T1, nije zabeležen ni u jednoj drugoj kristalografskoj strukturi P-TEFb kompleksa. Od osam očuvanih pozicija molekula vode na vezivnim površinama P-TEFb kompleksa, tri se nalaze na samoj ivici vezivne površine, u blizini  $\beta 3$ - $\alpha C$  petlje, a dva premošćavaju interakcije N-terminusa CDK9 sa ciklinom T1 u njegovoj izolovanoj oblasti vezivanja. Preostala tri očuvana molekula vode su locirana na vezivnoj površini u blizini hidrofobnog regiona sa Phe 146.

Značaj poznavanja rasporeda molekula vode na vezivnoj površini kompleksa u dizajnu inhibitora protein-protein interakcija se ogleda u mogućnosti istiskivanja molekula vode funkcionalnom grupom liganda i postizanju boljeg afiniteta liganda, kroz povoljnu entropiju procesa, usled oslobađanja molekula vode u rastvarač. Istovremeno, uspešan dizajn liganda koji uključuje remećenje čvrsto vezane i strukturane vode je značajno otežan. Primena WaterDock metoda na vezivnu površinu CDK9 u P-TEFb kompleksu ukazala je da je jedan od tri molekula vode u blizini hidrofobnog regiona pogodan za zamenu nepolarnom grupom liganda (81% verovatnoće da će taj molekul vode biti zamenjen i 80% verovatnoće da će biti zamenjen nepolarnom grupom). Za preostala dva molekula vode nije bilo moguće izvesti direktno poređenje, jer je primenom WaterDock metoda predviđen samo jedan molekul vode na njihovoj poziciji, između dva prisutna molekula i predviđena verovatnoća izmene od 0.54 (34% verovatnoće za izmenu nepolarnom grupom).

Uz WaterDock metod i vizuelno proučavanje šeme hidratacije poznatih kristalografskih struktura, pokušana je i procena hidratisanosti površine proteina praćenjem raspodele molekula vode tokom simulacije sistema molekularnom dinamikom. Problemi u ovom pristupu nastaju zbog fleksibilnosti proteina i mobilnosti molekula vode (167). Procena raspodele molekula vode u simulaciji molekularnom dinamikom se odvija na rešetki formiranoj u prostoru oko proteina. Fluktuacije položaja proteina mogu da dovedu do naizgled nedovoljne solvatacije određenih oblasti oko proteina, jer ih deo vremena u simulaciji okupiraju fleksibilni regioni proteina. Takođe,

molekuli vode osciliraju oko svojih položaja 1–2 Å (167). Uz to, Langevinov koeficijent trenja primenjen u simulacijama u ovoj studiji ( $\gamma = 2 \text{ ps}^{-1}$ ) je podešen za što potpunije uzorkovanje konformacionog prostora proteina, pa je, kao posledica, difuzija vode nešto brža nego što je u eksperimentima (168). Kao rezultat svega navedenog, raspodele gustine molekula vode iz simulacija u ovoj studiji nisu mogle biti sa sigurnošću upotrebljene za analizu hidratacije vezivnih površina P-TEFb kompleksa i CDK9. Dobijene su raspodele sa nedovoljno izraženim pikovima gustine molekula vode, a primećena je i u literaturi opisana pojava prividne nedovoljne solvatacije fleksibilnih regiona proteina.

#### *Analiza doprinosa aminokiselina energiji vezivanja P-TEFb kompleksa*

Po završenoj strukturnoj analizi vezivnih površina CDK9 / ciklin T1 kompleksa u kristalografskoj strukturi, obavljena je i analiza doprinosa aminokiselina kompleksa energiji vezivanja. Cilj analize bila je identifikacija *hot spot* aminokiselina koje značajno doprinose energiji vezivanja kompleksa. Ovakve aminokiseline su korisna polazna tačka u dizajnu potencijalnih inhibitora. *Hot spot* aminokiseline na vezivnoj površini kompleksa identifikovane su primenom tri *in silico* metode (Tabela 2).

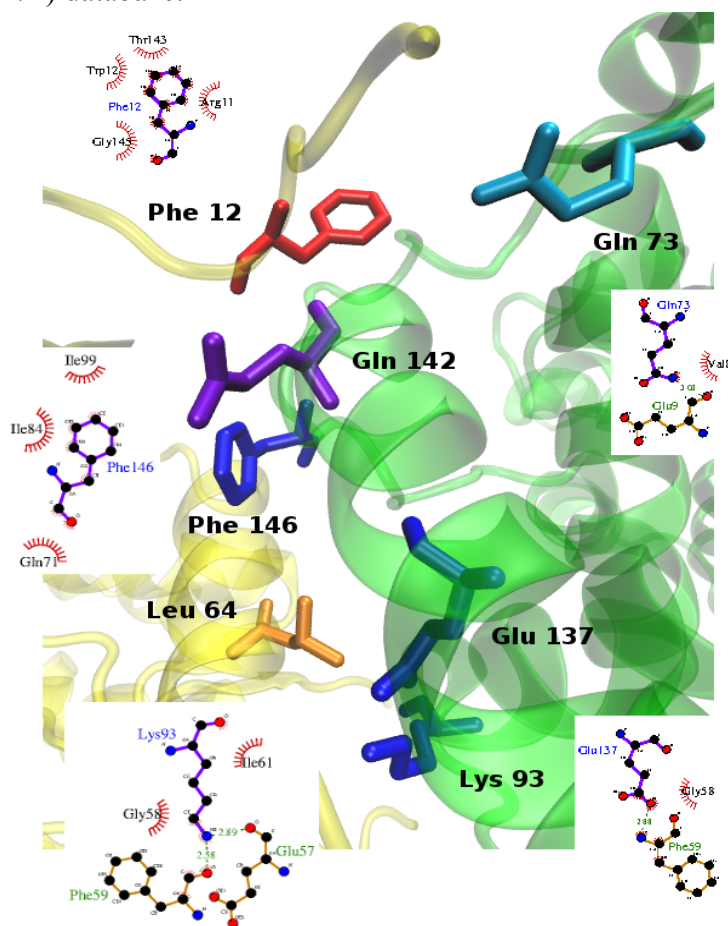
**Tabela 2:** *Hot spot* aminokiseline na vezivnoj površini CDK9 / ciklin T1 kompleksa

Metode Aminokiseline	ANCHOR		HotPoint	KFC2	Robetta	SCOWLP
	$\Delta$ SASA A2	$\Delta$ SASA %	hotspot +/-	hotspot +/-	kcal/mol	interakcije sa:
*A12 - PHE	125.8	76.7	-	-	2.21	B72, 142, 143, 145, 12, 11
A64 - LEU	86.3	61.2	-	+	1.84	B93, B94, B90, B141, B148
B73 - GLN	50	35.5	+	-	2.4	A8, A9
B93 - LYS	65.9	40.3	-	+	3.7	A61, 64, 59, 57
B137 - GLU	17.2	12.8	+	-	2.84	A59
B142 - GLN	81.5	57.8	+	-	0.75	A10, 11, 12, 13, 86, 99
B146 - PHE	91.8	55.9	+	+	2.87	A59, 67, 71, 84, 99

\* - A označava da je aminokiselina iz CDK9, a B iz ciklina T1

Primenjeno je više metoda da bi se smanjila verovatnoća greške, jer *in silico* metode utvrđivanja *hot spot* aminokiselina, iako su jeftinije i brže od eksperimenata, imaju i manji stepen tačnosti u odnosu na njih. U tabeli 2 istaknut je Robetta metod (169) koji predstavlja *in silico* verziju eksperimentalne metode u kojoj se vrši sekvencijalna mutacija aminokiselina od interesa do alanina (eng. *alanine scanning*). Aminokiselina se smatra *hot spot* aminokiselinom, po Robetta metodu, ukoliko pri

njenoj mutaciji do alanina dolazi do smanjenja energije vezivanja kompleksa za  $\geq 2$  kcal·mol<sup>-1</sup>. Robetta metod je korišćen kao glavni metod izbora *hot spot* aminokiselina u ovoj studiji. Druga dva algoritma za utvrđivanje *hot spot* aminokiselina, HotPoint (170) i KFC2 (171) primenjeni su da bi se proverilo slaganje algoritama i dobila dodatna potvrda rezultata. HotPoint i KFC2 algoritmi svoja predviđanja zasnivaju na izolovanosti aminokiseline od rastvarača po formiranju kompleksa, interakcijama poznatih *hot spot* aminokiselina i fleksibilnosti strukture oko aminokiseline. U tabeli 2 je još prikazan i stepen izolovanosti datih aminokiselina od rastvarača po formiranju kompleksa, tj. promena u molekularnoj površini aminokiseline dostupnoj rastvaraču po formiranju kompleksa ( $\Delta$ SASA) dobijen iz ANCHOR (172) databaze, kao i aminokiseline sa kojima svaka od navedenih aminokiselina interaguje po podacima SCOWLP (173,174) databaze.



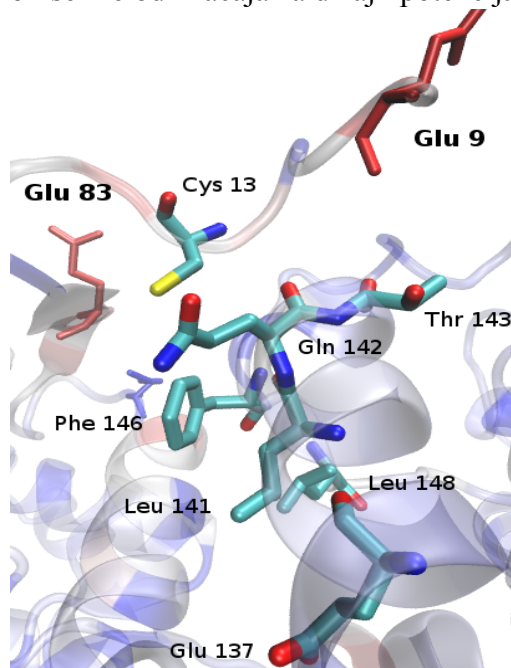
**Slika 24:** Energetski značajne (*hot spot*) aminokiseline na vezivnoj površini kompleksa CDK9/ciklin T1 (žuto/zeleno) i šematski prikaz njihovih interakcija. Crveno / plavo su označene *hot spot* aminokiseline poreklom iz CDK9 / ciklina T1, a narandžasto (Leu 64) / ljubičasto (Gln 142) dodatne aminokiseline iz teksta.



Primenom Robetta metoda identifikovano je pet aminokiselina kao *hot spot* aminokiseline na vezivnoj površini kompleksa (tabela 2, Slika 24). Jedina *hot spot* aminokiselina na vezivnoj površini kompleksa poreklom iz CDK9 jeste Phe 12, koji formira interakcije u izolovanom regionu vezivne površine na ciklinu T1. Na ciklinu T1 izolovane su četiri *hot spot* aminokiseline: Lys 93, Glu 137, Phe 146 i Gln 73. Lys 93 i Glu 137 formiraju mrežu vodoničnih veza sa polipeptidnim nizom aminokiselina Glu 57 i Phe 59 iz  $\beta 3-\alpha C$  petlje CDK9 i dva od tri upotrebljena algoritma ih smatraju važnim za formiranje kompleksa. Jedina *hot spot* aminokiselina identifikovana primenom sva tri algoritma jeste Phe 146, koji ispunjava hidrofobni džep na površini CDK9. Poslednja *hot spot* aminokiselina, Gln 73, interaguje samo sa N-terminusom CDK9, formirajući vodoničnu vezu sa Glu 9 u izolovanom sporednom regionu vezivne površine, pa nije dalje razmatrana u dizajnu inhibitora. Kao početne tačke dizajna odabrane su aminokiseline identifikovane kao *hot spot* aminokiseline primenom bar dva od tri korišćena algoritma, kao i one sa posebno zanimljivim hemijskim okruženjem na vezivnoj površini kompleksa. Za ilustraciju kriterijuma korišćenih u odabiru aminokiselina za početne tačke dizajna potencijalnih inhibitora mogu da posluže aminokiseline Leu 64 sa CDK9 i Gln 142 sa ciklina T1, od kojih je samo Gln 142 zadržan u dizajnu potencijalnih inhibitora interakcije. Obe aminokiseline su identifikovane kao energetske značajne samo jednim od tri primenjena kompjuterska algoritma. Promena slobodne energije nastanka kompleksa pri mutaciji Leu 64 u alanin je  $1.84 \text{ kcal}\cdot\text{mol}^{-1}$  što je blizu graničnoj vrednosti za *hot spot* aminokiseline ( $2 \text{ kcal}\cdot\text{mol}^{-1}$ ). Ova aminokiselina interaguje sa *hot spot* aminokiselinom Lys 93, ali samo hidrofobno i nalazi se izolovana od drugih značajnih aminokiselina CDK9. Nasuprot tome, Gln 142 se, i pored manjeg doprinosa svog bočnog lanca energiji vezivanja ( $0.75 \text{ kcal}\cdot\text{mol}^{-1}$ ) i prostorno, i u primarnoj građi ciklina T1 nalazi blizu dve *hot spot* aminokiseline, Glu 137 i Phe 146, a sa *hot spot* Phe 12 sa CDK9 interaguje vodoničnom vezom preko očuvanog molekula vode sa vezivne površine kompleksa.

Prostorno bliske aminokiseline pogodnije su za razvoj potencijalnih inhibitora od izolovanih *hot spot* aminokiselina, zbog grupisanja većeg broja povoljnih interakcija na malom prostoru dostupnom veličini inhibitora. Ovaj efekat je uzet u obzir u energetske analizi vezivnih površina kompleksa CDK9 / ciklin T1 dodatnom *in silico* analizom koja je imala za cilj da identifikuje regione blisko interagujućih *hot spot*

aminokiselina. U HotRegion (175) databazi pronađen je samo jedan takav region na vezivnim površinama kompleksa CDK9 / ciklin T1. Čine ga aminokiseline Cys 13 sa CDK9 i Glu 137, Leu 141, Gln 142, Thr 143, Phe 146 i Leu 148 sa ciklina T1 (Slika 25). Prisustvo Gln 142 u građi pronađenog energetski značajnog regiona podržava odabir Gln 142 kao aminokiseline od značaja za dizajn potencijalnih inhibitora.



**Slika 25:** Region interagujućih *hot spot* aminokiselina na vezivnoj površini CDK9 / ciklin T1 kompleksa. Obeležene su (crveno/plavo) i tri aminokiseline čije naelektrisanje igra ulogu u procesu formiranja kompleksa

Važno je istaći da se značaj neke aminokiseline sa vezivne površine proteinskog kompleksa u formiranju datog kompleksa ne može svesti samo i isključivo na njen doprinos kao *hot spot* aminokiseline. Određene aminokiseline, kako sa vezivnih površina, tako i u njihovoj okolini, imaju značaj u procesu formiranja kompleksa jer svojim naelektrisanjem doprinose formiranju povoljnog elektrostatičkog polja interakcija, koje obezbeđuje pravilnu orijentaciju proteinskih subjedinica tokom formiranja kompleksa. Uticaj naelektrisanja prisutnih aminokiselina na brzinu formiranja kompleksa proveren je *in silico* mutacijom naelektrisanja (176). Značajnim se smatraju aminokiseline na kojima promena naelektrisanja dovodi bar do četverostrukih promena u predviđenoj brzini asocijacije kompleksa. Dve aminokiseline sa vezivne površine CDK9 su identifikovane kao značajne: Glu 9 i Glu 83 (Slika 26). Dok se promenom naelektrisanja na glutamatima smanjivala brzina nastanka kompleksa, ukazujući da igraju ulogu u elektrostatskom navođenju subjedinica tokom

formiranja kompleksa, promena naelektrisanja na lizinu je povećala predviđenu brzinu asocijacije. Zanimljivo je napomenuti da je, kao celina, površina CDK9 pozitivno naelektrisana (naelektrisanje 6+), a da vezivanje sa ciklinom T1 naizgled ostvaruje fokusiranjem negativnog naelektrisanja u prostoru.

*Evolutivni značaj aminokiselina sa vezivnih površina CDK9 / ciklin T1 kompleksa*

Analiza doprinosa aminokiselina vezivnih površina CDK9 / ciklin T1 kompleksa energiji vezivanja upotpunjena je razmatranjem evolutivnog značaja aminokiselina sa vezivnih površina kompleksa. Primenjena je tehnika evolutivnog praćenja (eng. *evolutionary trace*) u kojoj se iz zastupljenosti i mutacija aminokiseline na određenom položaju u proteinu, u okviru skupa evolutivno povezanih sekvenci, izvode zaključci o njenom relativnom značaju u proteinu (177). Najznačajnijim se smatraju one aminokiseline čijom zamenom dolazi do najvećih strukturnih i funkcionalnih promena u proteinu, tj. one aminokiseline koje se nalaze u korenu evolutivnih grana različitih proteinskih struktura i koje su najviše očuvane kroz evolutivnu granu kojoj pripadaju. Tokom evolutivnog praćenja sprovedenog na kompleksu CDK9 / ciklin T1 identifikovan je, na ciklinu T1, podskup aminokiselina koje odgovaraju najznačajnijim aminokiselinama i regionima iz *hot spot* analize (16 aminokiselina), dok su na CDK9, sa vezivnih površina identifikovane pretežno aminokiseline sa funkcionalno značajnog C-heliksa kinaze (4 aminokiseline) (Tabela 3).

**Tabela 3:** Najznačajnije aminokiseline (po pet) sa vezivnih površina CDK9 / ciklin T1 kompleksa, po evolutivnom praćenju.

<b>CDK9</b>	<b>Ciklin T1</b>
Gly 58	Lys 93
Pro 60	Glu 137
Ile 67	Leu 141
Leu 81	Phe 146
	Thr 143

Po obavljenim analizama vezivnih površina kompleksa CDK9 / ciklin T1, kao konačan skup aminokiselina najpogodnijih za razmatranje u dizajnu potencijalnih inhibitora protein-protein interakcije odabrane su aminokiseline Lys 93, Glu 137, Gln

142 i Phe 146 sa ciklina T1. Kao sekundarni izbor zadržana je i *hot spot* aminokiselina Phe 12 sa CDK9, koju njen položaj na sporednoj vezivnoj površini kompleksa, u N-terminusu CDK9 čini manje pogodnom polaznom tačkom za direktnu inhibiciju formiranja kompleksa.

#### *Poređenje rezultata sa eksperimentalnim studijama iz literature*

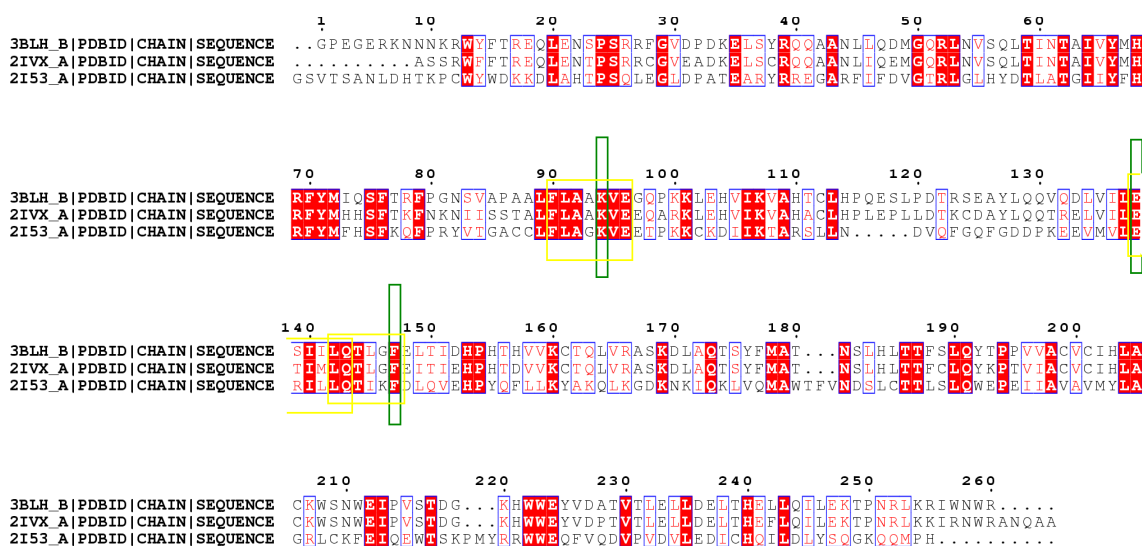
Navedeni izbor *hot spot* aminokiselina je obavljen na osnovu rezultata *in silico* metoda, jer, po dostupnim literaturnim podacima, do sada eksperimentalno nije obavljena mutaciona analiza celokupnih vezivnih površina CDK9 / ciklin T1 kompleksa. Obavljene su tri eksperimentalne mutacione studije specifičnih aminokiselina na vezivnoj površini CDK9 / ciklin T1 kompleksa. U dve studije (26,178) vezivanje CDK9 i ciklina T1 sprečeno je uvođenjem dve tačkaste mutacije u strukturu ciklina T1: mutacije Lys 93 do Leu i Glu 96 do Lys. U trećoj mutacionoj studiji (179) izvršena je izmena naelektrisanja između Lys 93 i Glu 137 sa ciklina T1, tako što je Lys 93 mutiran do Asp, a Glu 137 do Lys. I ove dve mutacije, bilo same, ili prisutne istovremeno, dovele su do ometenog formiranja CDK9 / ciklin T1 kompleksa (179). Takođe je, skraćivanjem strukture ciklina T1, eksperimentalno potvrđena uloga regiona aminokiselina 134–188 primarne strukture ciklina T1 za formiranje kompleksa. Pokazano je da molekuli ciklina T1, samo sa aminokiselinama 1–188 u svojoj strukturi, uspešno vezuju CDK9, dok je formiranje kompleksa ometeno ako ciklin T1 sadrži samo aminokiseline 1–133 u svojoj strukturi (26). U regionu 134–188 samo mali broj aminokiselina direktno učestvuje u formiranju CDK9 / ciklin T1: aminokiseline 134, 137, 139, 141–146, 148, 149.

Od aminokiselina sa vezivne površine CDK9 / ciklin T1 kompleksa koje su u ovoj studiji odabrane kao pogodne početne tačke za dizajn liganada, Lys 93 i Glu 137 su direktno podržane eksperimentima, a Gln 142 i Phe 146 su deo značajnog regiona interakcija sa ciklina (134–149) i stoga je njihovo uključivanje indirektno opravdano. Jedina eksperimentalno poznata značajna aminokiselina koja nije uključena u razmatranje u ovoj studiji jeste Glu 96. Glu 96 je u kristalografskoj strukturi korišćenoj u ovoj studiji (PDB ID: 3blh) mutiran u Gly i stoga nije bio uključen u *in silico* analize. Ova mutacija se prirodno javlja u nekim vrstama (83) i ne utiče značajno na strukturu

kompleksa, niti na njegovu sposobnost vezivanja ATP-a (154). Kako mutacija Glu 96 u Gly remeti vodonične veze u kojima Glu učestvuje, javlja se pad tačke topljenja kompleksa što ukazuje na njegovu donekle smanjenu stabilnost, ali je sama konformacija proteina u kompleksu očuvana (154).

### Specifičnost interakcija aminokiselina odabranih za inhibiciju

Dodatan važan kriterijum u analizi vezivnih površina proteinskog kompleksa jeste specifičnost interakcija koje se formiraju, naročito ako se radi o dizajnu potencijalnih inhibitora za interakciju u kojoj učestvuje protein sa većim brojem poznatih homologa ili ako se iste vezivne površine na proteinima koriste u formiranju više različitih proteinskih kompleksa. CDK9, kao član CDK familije proteina koji formira analogne komplekse sa više različitih ciklina (T1, T2a, T2b, K) ispunjava oba navedena uslova. Trodimenzionalne strukture CDK9 u kompleksu sa ciklinima T2 i K još nisu rešene, ali strukturna sličnost ciklina T1, T2 i K ukazuje da i njihovi kompleksi sa CDK9 imaju sličnu strukturu (83). Upoređivanje primarne strukture ova tri ciklina pokazuje da su sve aminokiseline izabrane u ovoj studiji prisutne u svim strukturama (Slika 26). Takođe, aminokiseline Lys 93 i Glu 137 su evolutivno očuvane između ciklina T1 i ciklina H, koji formira kompleks sa CDK7 (179).



**Slika 26:** Poređenje aminokiselinske sekvence ciklina T1, T2 i K. Očuvane *hot spot* aminokiseline Lys 93, Glu 137 i Phe 146 su uokvirene (zeleno).

Kako su iste *hot spot* aminokiseline prisutne u svim ciklinima koji grade kompleks sa CDK9, nije verovatno da će inhibitori razvijeni u ovoj studiji, koja koristi najrasprostranjeniji kompleks CDK9 sa ciklinom T1, pokazati selektivnost u odnosu na tip ciklina čije vezivanje za CDK9 ometaju. Ciklin-specifično delovanje nije presudno za terapijski efekat inhibitora aktivnosti P-TEFb kompleksa, što pokazuju ATP-kompetitivni inhibitori aktivnosti CDK9, koji podjednako inhibiraju sve podtipove CDK9 / ciklin kompleksa, a ipak pokazuju zadovoljavajući terapijski efekat.

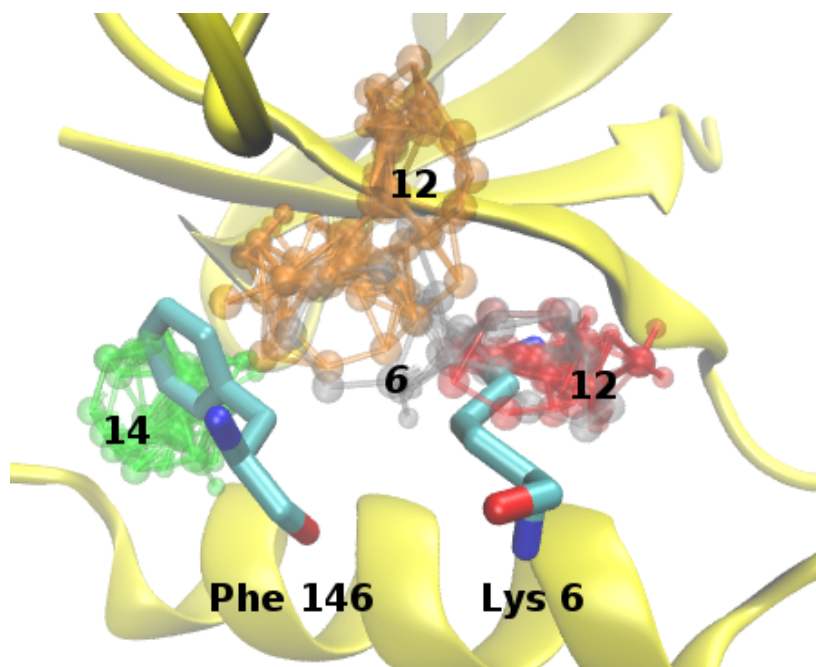
#### *Strukturni diverzitet interakcija vezivnih površina CDK9 / ciklin T1 kompleksa*

Proučavanje interakcija koje se formiraju između proteina na vezivnoj površini kompleksa CDK9 / ciklin T1 ne daje potpun uvid u vezivne sposobnosti proteinske površine na kojoj se interakcije odvijaju, jer su od molekularnih vrsta prisutni samo polipeptidni niz proteina, bočni lanci aminokiselina i rastvarač. Hemijski prostor nepeptidnih inhibitora protein-protein interakcija i organskih liganada proteina uopšte je značajno širi, pa se ne može u potpunosti mapirati samo kroz interakcije prisutne između hemijskih grupa proteina. Procena sposobnosti proteinske površine da interaguje sa različitim funkcionalnim grupama prisutnim u malim molekulima, nezavisno od njihove rasprostranjenosti u proteinima, je stoga koristan test pogodnosti izabrane oblasti proteina za razvoj inhibitora.

Literaturni rezultati jednog od metoda za predviđanje lokacija interakcija malih molekula sa vezivnom površinom proteina (67) nisu obuhvatali nijednu od 17 do sada objavljenih struktura CDK9 / ciklin T1 kompleksa. Opisani metod je zasnovan na evolutivnoj očuvanosti aminokiselina, prisustvu hidrofobnih džepova u njihovom okruženju i promeni površine aminokiseline dostupne rastvaraču po formiranju kompleksa i primenjen je na 10000 različitih proteinskih kompleksa.

Da bi bile proučene interakcije proteinske površine sa širim spektrom hemijskih grupa, vezivne površine kompleksa CDK9 / ciklin T1 u kristalografskoj strukturi kompleksa (PDB ID: 3blh) su podvrgnute *in silico* mapiranju interakcija malih organskih molekula na njima. Mapiranje se zasniva na proceni interakcija malih organskih molekula sa proteinskom površinom u različitim tačkama prostora u blizini proteinske površine, grupisanju rezultata različitih molekularnih proba i predviđanju

najverovatnijih mesta interakcije proteinske površine sa organskim ligandom sličnih osobina (FTMap procedura) (30). Na površini ciklina T1 mapiranjem nisu otkrivene pogodne tačke za interakcije malih molekula sa proteinom, ni na glavnoj vezivnoj površini kompleksa, niti na površini sa kojom interaguje N-terminus CDK9. Na vezivnoj površini CDK9 u kompleksu, mapiranjem su otkrivene četiri pogodne lokacije za interakciju sa malim organskim molekulima. Iako nijedna od četiri oblasti na vezivnoj površini CDK9 nije ispunila FTMap literaturni kriterijum za *hot spot* tačku interakcija (16 i više proba grupisanih zajedno) (30), tri od četiri grupacije proba (sa 14, 12 i 6 proba) se u prostoru preklapaju sa aminokiselinama sa ciklina T1 – *hot spot* aminokiselinom Phe 146 i Lys 6 (Slika 27).



**Slika 27:** Najpovoljnija predviđena mesta interakcija malih organskih molekula sa površinom CDK9 sa brojem grupisanih proba i prostorno preklapanje sa aminokiselinama ciklina T1

Rezultati mapiranja površine CDK9 su ostali nepromenjeni po uključivanju dodatnih rotamera aminokiselina oko identifikovanih grupacija proba, odnosno nisu identifikovani rotameri koji bi značajno promenili izgled proteinske površine. To ukazuje da identifikovane lokacije interakcija sa malim organskim molekulima ne zavise kritično od ograničenih lokalnih promena konformacije proteina. Uticaj većih konformacionih promena na izgled i osobine vezivne površine CDK9 u CDK9 / ciklin

T1 kompleksu proučavan je analizom dinamike i fleksibilnosti strukture molekularnom dinamikom.

Za fokus daljeg istraživanja izabrana je vezivna površina CDK9, a ne ciklina T1, jer je procenjeno da je ona pogodnija za vezivanje potencijalnih inhibitora. Razlog je njena konkavnost, veći broj pogodno lociranih *hot spot* aminokiselina sa ciklina T1 u poređenju sa jednom izolovanom *hot spot* aminokiselinom CDK9 i pogodni rezultati mapiranja interakcija malih molekula sa površinom CDK9.

Strukturnom i energetsom analizom statičnih kristalografskih struktura CDK9 / ciklin T1 kompleksa izdvojene su aminokiseline Lys 6, Lys 93, Glu 137, Gln 142 i Phe 146 sa ciklina T1 kao pogodne početne tačke u daljem dizajnu inhibitora interakcije. Lys 93, Glu 137 i Phe 146 su izabrane kao energetske i evolutivno značajne *hot spot* aminokiseline. Lys 6 je odabran jer se po formiranju CDK9 / ciklin T1 kompleksa smešta u hidrofobni džep u neposrednoj blizini *hot spot* aminokiseline Phe 146.

Uticaj konformacione fleksibilnosti vezivnih površina CDK9 na dobijene rezultate strukturne analize kompleksa proveren je simulacijom proteina molekularnom dinamikom. Simulacijama je omogućeno uključivanje dinamike proteina u dalji proces analize CDK9 / ciklin T1 interakcije, što je značajno poboljšalo kvalitet određivanja strukture modulatora interakcije.



### 2.1.1.3 Analiza fleksibilnosti CDK9

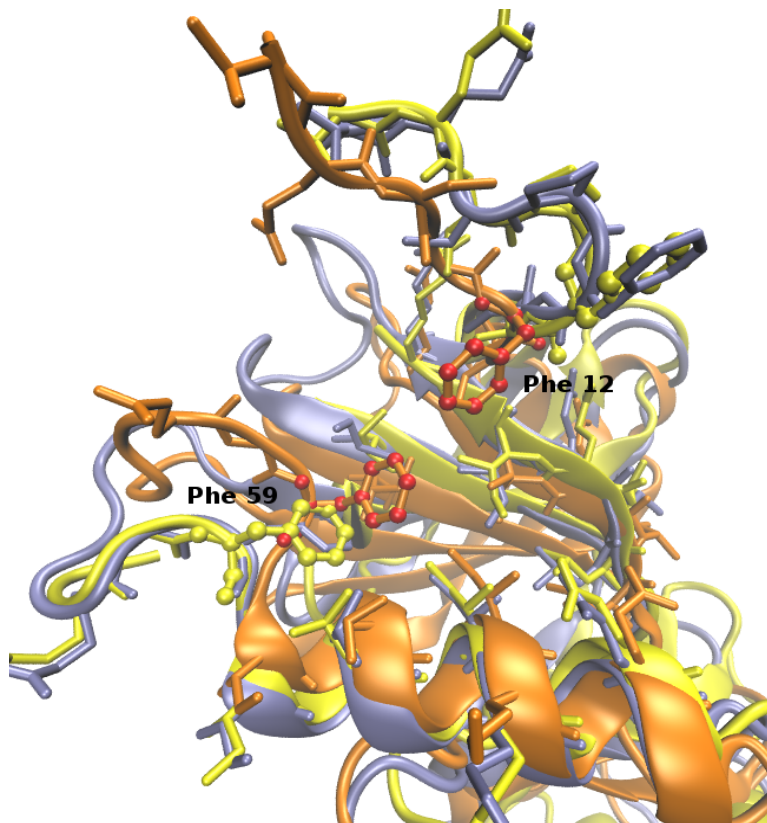
Konformacija proteina određena kristalografskom analizom predstavlja samo jednu od mogućih konformacija proteina. U kristalografskim strukturama proteini su zarobljeni u energetske minimumu koji je najviše odgovarao uslovima u toku kristalizacije i na koji često utiče pakovanje molekula proteina u kristalnoj rešetki. U rastvoru, proteini stalno menjaju svoju konformaciju, oscilirajući između velikog broja sličnih energetskih minimuma. Takođe, konformacije proteina u proteinskim kompleksima se najčešće, u većoj ili manjoj meri, razlikuju od konformacija slobodnih proteina u rastvoru. Kao rezultat, u proučavanju strukture proteina u proteinskim kompleksima i dizajnu potencijalnih inhibitora interakcija, nije dovoljno osloniti se samo na kristalografske strukture, već je neophodno steći uvid u dinamiku i fleksibilnost vezivnih površina proteina, što se može postići simulacijom struktura molekularnom dinamikom.

Molekularna dinamika je tehnika koja omogućuje simulaciju ponašanja makromolekula sa prostornom i vremenskom rezolucijom koju za sada nije dostupna ni u jednoj eksperimentalnoj tehnici. Praćenje konformacionih promena proteina u rastvoru ili membranama molekularnom dinamikom odvija se na skali od femtosekunde do stotina nanosekundi, što je čini pogodnom za relaksaciju kristalografskih struktura i razumevanje dinamike proteinske površine.

U proteinskom kompleksu, izgled vezivnih površina je relativno konstantan, uz male varijacije položaja atoma. U ovoj studiji, kompleks CDK9 / ciklin T1 simuliran je molekularnom dinamikom prevažno da bi došlo do energetske relaksacije kristalografskih struktura i njihovog prilagođavanja uslovima u rastvoru. Simulacija kompleksa CDK9 / ciklin T1 je takođe predstavljala fazu pripreme proteinskih struktura za simulaciju slobodne CDK9 u rastvoru, radi razumevanja dinamike njene vezivne površine.

U prvoj fazi istraživanja, slobodna CDK9 je simulirana klasičnom molekularnom dinamikom (4ns) i zatim je tako dobijena struktura upoređena sa kristalografskom strukturom proteina i strukturom proteina simuliranom u kompleksu. Upoređivanje je vršeno procenom kvadratne sredine odstupanja (eng. *root mean square deviation*, RMSD) atoma proteina i vezivne površine. Kao što je i očekivano, vezivna

površina simulirane strukture CDK9 u kompleksu sa ciklinom T1 je bila slična kristalografskoj (RMSD C $\alpha$  atoma aminokiselina vezivne površine 0.81 Å, a svih atoma aminokiselina 3.18 Å), dok je u simuliranoj strukturi slobodne CDK9 konformacija površine više odstupala (RMSD C $\alpha$  atoma aminokiselina vezivne površine 3.76 Å, a svih atoma 5.14 Å). Najznačajnija razlika u organizaciji vezivne površine je reorijentacija Phe 12 i Phe 59, nastala usled pomeranja N-terminusa i  $\beta$ 3- $\alpha$ C petlje CDK9 (Slika 28).



**Slika 28:** Konformacione promene vezivne površine CDK9 u simuliranim strukturama slobodne (narandžasto) i CDK9 u kompleksu sa ciklinom T1 (plavo) u odnosu na kristalografsku strukturu CDK9 iz kompleksa (žuto)

Uočene značajne konformacione promene vezivne površine CDK9 van kompleksa sa ciklinom T1 ukazale su da je neophodno pri dizajnu inhibitora interakcije sa ciklinom T1 uzeti u obzir više alternativnih konformacija CDK9. Konformacije slobodne CDK9 su korišćene u dizajnu zbog svoje postulirane dostupnosti ligandima u rastvoru, dok je konformacija CDK9 iz kompleksa CDK9 / ciklin T1 zadržana tokom dizajna jer aminokiseline vezivne površine proteina u konformaciji u kompleksu često zadržavaju rotamerna stanja kakva postoje kada je protein slobodan u rastvoru (9). Kako

kristalografska struktura slobodne CDK9 van kompleksa sa ciklinom T1 do sada nije rešena, njena struktura je dobijena simulacijom molekularnom dinamikom.

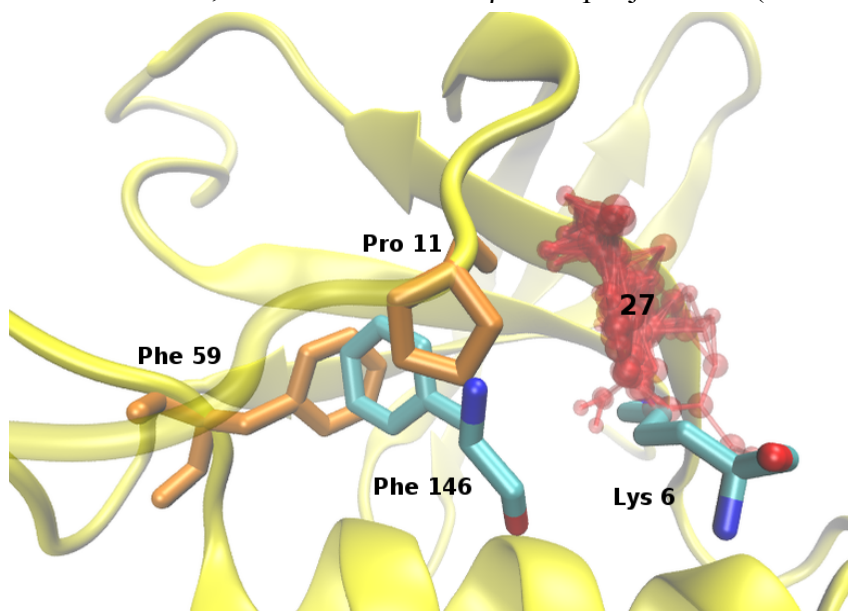
Tokom klasične molekularne dinamike simulirani molekul, u ovom slučaju CDK9 postepeno istražuje konformacioni prostor oko početnog energetskog minimuma, pri čemu stepen konformacionih promena delom zavisi od dužine trajanja simulacije, tj. od dužine posmatranja molekula. Uobičajeno trajanje simulacija molekularnom dinamikom je od par nanosekundi do par stotina nanosekundi, ako izuzmemo mali broj simulacija sprovedenih na Antonu, posebno sagrađenom kompjuteru za molekularnu dinamiku, koje dostižu i jednu milisekundu (180). Veće konformacione promene u proteinima se dešavaju u vremenskom rasponu mikro- do milisekundi, što je van dometa klasične molekularne dinamike. Kao rezultat, da bi se dobio detaljan uvid u konformacionu dinamiku nekog proteina nije dovoljno sprovesti simulaciju klasičnom molekularnom dinamikom u trajanju od nekoliko nanosekundi. Potrebno je na neki način poboljšati istraživanje konformacionog prostora proteina. Danas je razvijeno više metoda u ovu svrhu, od kojih je za ovo istraživanje izabrana aMD (eng. *accelerated molecular dynamics*, ubrzana molekularna dinamika) (181). U ubrzanoj molekularnoj dinamici se sprečava zadržavanje sistema u energetskim minimumima tako što se sistemu dodaje dovoljno energije da prevaziđe lokalne energetske barijere i istražuje ostatak konformacionog prostora. Za studiju fleksibilnosti CDK9 upotrebljena je aMD metod jer ne zahteva prethodno detaljno znanje o konformacionom ponašanju sistema a omogućuje ubrzavanje istraživanja konformacionog prostora proteina za par redova veličine (182).

Simulacije CDK9 su pripremljene i izvedene kako je opisano u poglavlju Eksperimentalni deo. Prve dve simulacije (N1 i N2) su izvedene bez prisustva prvih sedam aminokiselina N-terminusa CDK9, čije koordinate, zbog fleksibilnosti nisu rešene u kristalografskoj strukturi. Kako su dobijeni slični rezultati u obe simulacije, u nastavku će biti diskutovana samo jedna od njih (N2), osim kod značajnih razlika. Nakon simulacija, bilo je neophodno grupisati dobijene konformacije (10000) i analizirati konformacione promene do kojih je došlo.

Konformacije su prvo grubo grupisane po energiji. U rastvoru, protein najveći deo vremena provodi u konformacijama koje odgovaraju različitim minimumima potencijalne energije, pa su energetski najpovoljnije konformacije i najverovatnije.

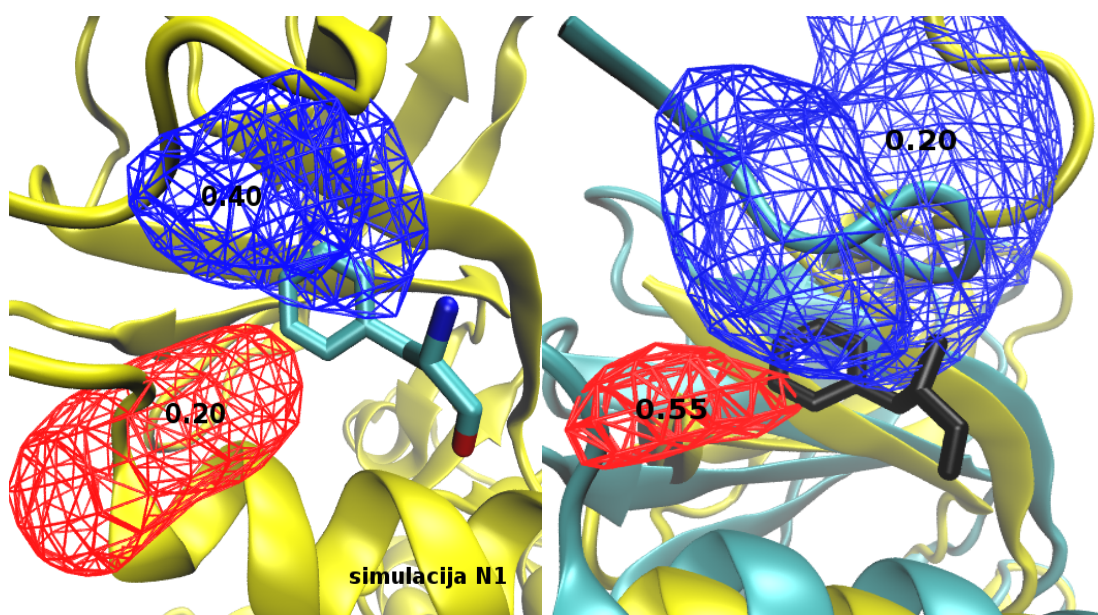
Međutim, kako aMD pristup služi prevashodno za istraživanje konformacionog prostora, dobijene konformacije u aMD simulaciji ne moraju da se podudaraju sa energetske minimumima. Rezultati simulacija su energetski korigovani po predlogu iz literature (183). Zadržano je 20% struktura sa najnižom energijom, koje su zatim grupisane po RMSD sličnosti pozicija atoma na vezivnoj površini CDK9. RMSD vrednost je definisana tako da uključuje sve atome aminokiselina sa vezivne površine osim vodonika, jer su pozicije atoma bočnih lanaca aminokiselina značajne za definisanje izgleda vezivne površine. Dodatan razlog za uključivanje svih atoma vezivne površine bio je povećanje diverziteta proteinskih struktura u kasnijem procesu predviđanja orijentacije liganada na vezivnoj površini. Deset reprezentativnih konformacija CDK9 mapirano je malim molekulima radi proučavanja promena u sposobnosti vezivne površine da interaguje sa različitim funkcionalnim grupama.

Zanimljivo, u rezultatima mapiranja deset struktura dobijenih simulacijom slobodne CDK9 bez N-terminusa, kao povoljno mesto interakcija više nije dominirao džep namenjen vezivanju *hot spot* aminokiseline Phe 146 sa ciklina T1. Kao dominantno mesto interakcija sa malim molekulima u svih 10 struktura javlja se džep u blizini Lys 6, kao *hot spot* lokacija po literaturi (18–34 grupisanih proba). Džep za Phe 146 u strukturama CDK9 sa skraćenim N-terminusom ispunjavaju aminokiseline sa N-terminusa CDK9 Pro 11, Phe 12 ili Phe 59 sa  $\beta$ - $\alpha$ C petlje CDK9 (Slika 29).



**Slika 29:** Rezultati mapiranja malim molekulima (crveno) strukture slobodne CDK9 sa skraćenim N-terminusom (žuto) i preklapanja aminokiselina CDK9 (narandžasto) sa pozicijama aminokiselina sa ciklina T1 (zeleno).

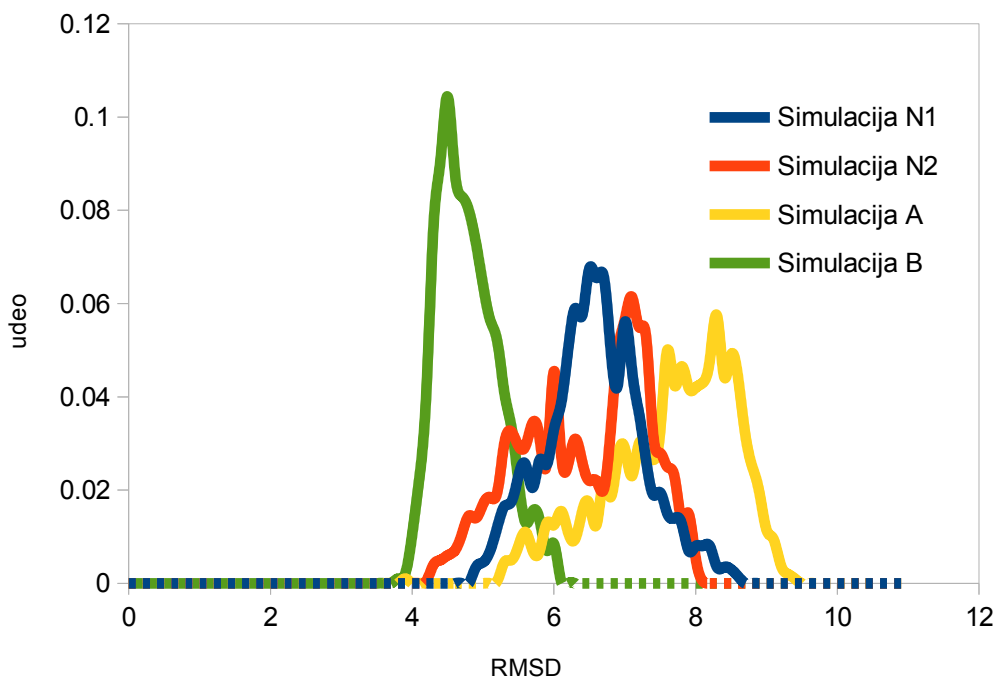
Učestalost primećene pojave ispunjavanja vezivnog mesta *hot spot* aminokiseline Phe 146 sa ciklina T1 aminokiselinama CDK9 ispitana je proverom raspodele položaja aminokiseline Phe 59 i skraćenog N-terminusa CDK9 (aminokiseline 8–13) u 2000 energetski najpovoljnijih konformacija CDK9 iz aMD simulacija N1 i N2. U simulaciji N1 aminokiseline skraćenog N-terminusa CDK9 su ispunjavale džep za vezivanje Phe 146 u 40% ispitanih konformacija, a Phe 59 20% vremena. U simulaciji N2, pokazalo se da Phe 59 zauzima džep za vezivanje Phe 146 tokom 55% vremena trajanja simulacije, dok su aminokiseline skraćenog N-terminusa zauzimale istu lokaciju oko 20% vremena trajanja simulacije (Slika 30).



**Slika 30:** Prostorno pozicioniranje aminokiselina CDK9 u džepu za vezivanje *hot spot* aminokiseline Phe 146 (štapići) sa ciklina T1 u simulacijama N1 (levo) i N2 (desno). Ošeni su prostori zadržavanja aminokiselina 8–13 (plavo, 40% / 20% vremena) i Phe 59 (crveno, 20% / 55% vremena)

Kako je tokom opisanih simulacija nedostajalo prvih sedam aminokiselina fleksibilnog N-terminusa CDK9, moguće je da uticaj aminokiselina 8–13 na izgled vezivne površine CDK9 ne odgovara stvarnosti, jer je fleksibilnost aminokiselina 8–13 povećana odsustvom prvih sedam aminokiselina. Radi provere, modeliran je N-terminus CDK9 u celosti i aMD simulacije u trajanju od 20 miliona koraka su ponovljene još dva puta, sa različitim poljima sila (simulacije A i B). N-terminus CDK9 je prisutan u celosti u ćelijama i nalazi se u položaju da interaguje sa ostatkom vezivne površine, pa je njegovo modelovanje fizički opravdano. U četiri opisane simulacije vezivna površina

CDK9 istražila je konformacioni prostor udaljen 4 do 10 Å od konformacije koju zauzima u kompleksu sa ciklinom T1 (grafik 1).



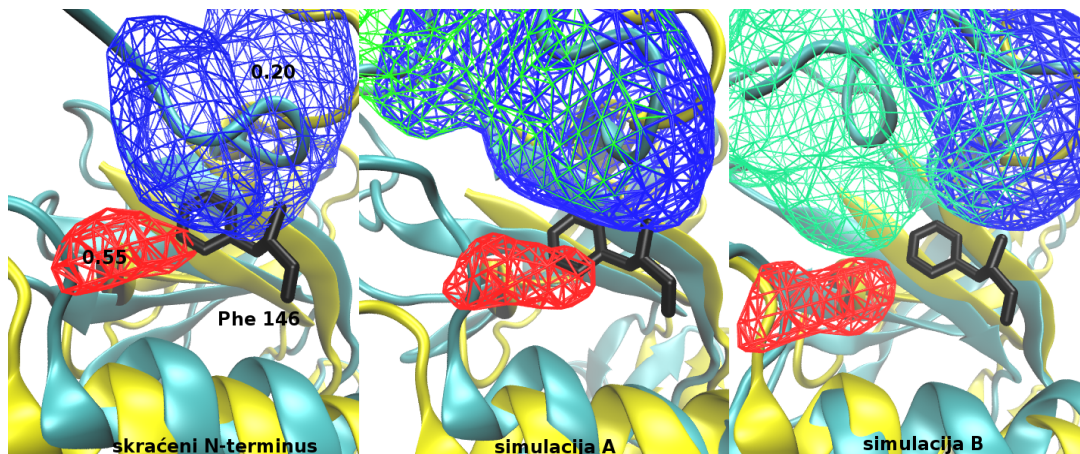
**Grafik 1:** Distribucija RMSD vrednosti atoma aminokiselina vezivne površine CDK9 u četiri simulacije u odnosu na konformaciju CDK9 u kompleksu sa ciklinom T1

Uključivanje čitavog N-terminusa u simulaciju slobodne CDK9 izmenilo je izgled vezivne površine. U simulaciji Kornel et al. poljem sila (184) (simulacija A) N-terminus je ostao nestruktuiran i interagovao je sa petljama  $\beta 3-\alpha C$  i petljom sa aminokiselinama 87–99, dok je u simulaciji sa CHARMM poljem sila (185,186) (simulacija B), N-terminus prolazno poprimao strukturu vrlo kratkog  $\alpha$ -heliksa. Formiranje heliksa u simulaciji B je bilo neočekivano, jer je N-terminus CDK9 nestruktuiran u svim rešenim kristalografskim strukturama u kojima je CDK9 prisutna i jer njegovo prisustvo nisu predvideli algoritmi za analizu sekundarne strukture proteina. Alternativna izoforma CDK9, CDK9\_55, ima produženi N-terminus koji bi mogao biti od koristi u potvrdi sekundarne strukture nedostajućeg N-terminusa CDK9. Pokušaji homologog modelovanja dodatnih 117 aminokiselina N-terminusa CDK9\_55 u ovoj studiji nisu bili uspešni zbog previše male sličnosti aminokiselinske sekvence N-terminusa sa sekvencama proteina rešene strukture. Rezultati algoritama za predviđanje sekundarne strukture primenjenih na sekvencu N-terminusa CDK9\_55 su se pokazali veoma osetljivim na tačnu dužinu sekvence na kojoj vrše analizu, predviđajući i neuređenu strukturu i heliks u regionu nedostajućih aminokiselina, pa nisu korišćeni u

analizi. U pokušaju razjašnjavanja sekundarne strukture N-terminusa CDK9, razmotrene su strukture njoj srodnih kinaza. Među CDK, N-terminalni heliks ovog tipa javlja se samo kod CDK8, koja je, kao i CDK9, transkripciona kinaza i čiji kompleks sa ciklinom C ima određenih sličnosti sa kompleksom CDK9 / ciklin T1 (159). N-terminalni heliks CDK8 je, međutim, očuvan u kristalografskoj strukturi njenog kompleksa sa ciklinom. N-terminalni heliksi su opisani kod kinaza koje ne zahtevaju fosforilaciju za svoju aktivaciju, kao što su CDK8 i NIK kinaza (eng. *NF- $\kappa$ B Inducing Kinase*, kinaza koja indukuje nuklearni faktor kB) (187). Kod ovakvih kinaza N-terminalni heliksi učestvuju u aktivaciji kinaza. CDK9 se, međutim, aktivira autofosforilacijom.

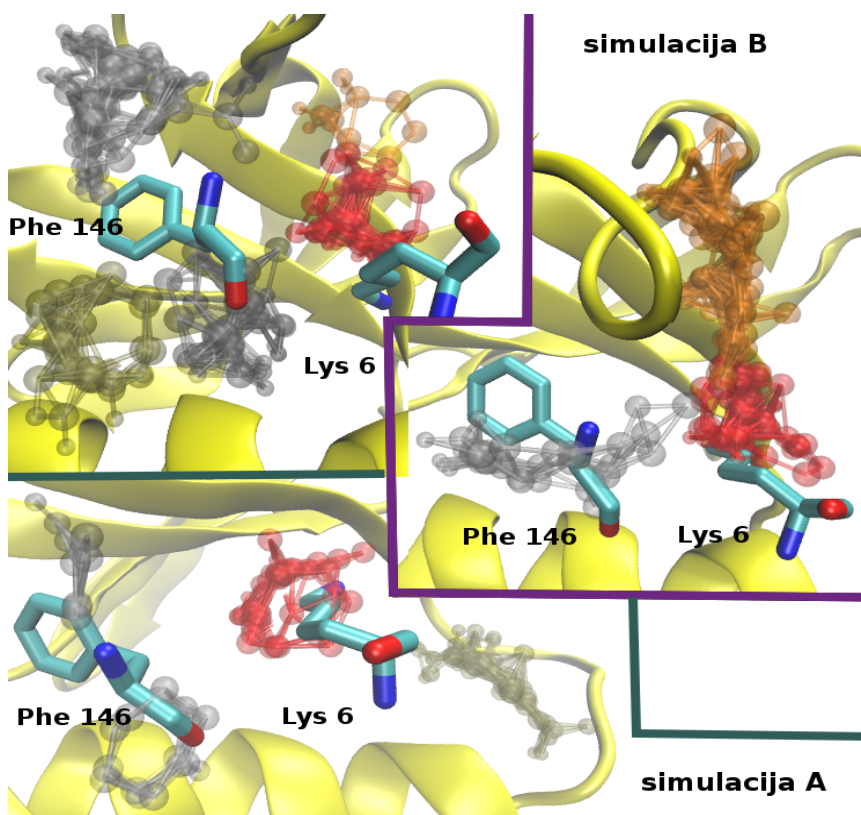
CDK8 iz kristalografskog sistema CDK8 / ciklin C, zbog eksperimentalne potvrde prisustva N-terminalnog heliksa u svojoj strukturi upotrebljena je za validaciju primenjenih uslova simulacije i njihove sposobnosti da verno modeluju N-terminus CDK struktura. N-terminus u CDK8 je modelovan istom procedurom koja je primenjena za CDK9. U aMD simulaciji Kornel et al. poljem sila došlo je do formiranja eksperimentalno potvrđenog N-terminalnog heliksa, što je ukazalo da su primenjeni uslovi simulacije odgovarajući za simulaciju N-terminusa CDK sistema. Kako pod istim uslovima (simulacija A) nije zabeleženo formiranje N-terminalnog heliksa u CDK9 sistemu, ni produžavanjem simulacije A još dodatnih 10 miliona koraka, niti ponavljanjem simulacije, N-terminalni heliks kod CDK9, ako se formira, ima najverovatnije ograničenu stabilnost.

Uključivanje svih aminokiselina N-terminusa CDK9 u simulaciju dovelo je do delimične stabilizacije oblasti vezivanja Phe 146 (Slika 31). U simulaciji sa skraćenim N-terminusom, aminokiseline N-terminusa i Phe 59 CDK9 zadržavaju se u oblasti vezivanja Phe 146 sa ciklina T1, što bi ometalo njegovo vezivanje, kao i vezivanje potencijalnih inhibitora. U simulacijama A i B, u kojima je N-terminus prisutan u celosti, pomenute aminokiseline se pretežno nalaze van oblasti vezivanja Phe 146, ostavljajući tu oblast slobodnu za pristup inhibitora.



**Slika 31:** Poređenje prostornog rasporeda aminokiseline 59 (crveno) i skraćenog i celog N-terminusa CDK9 (plavo i zeleno) tokom tri simulacije slobodne CDK9

Da je uključivanje nedostajućih aminokiselina N-terminusa CDK9 dovelo do promena izgleda vezivne površine proteina potvrđuju i rezultati mapiranja vezivne površine malim molekulima na 10 reprezentativnih struktura iz simulacija A i B. U konformacijama iz obe simulacije javljaju se grupacije proba u oblasti koja odgovara vezivanju *hot spot* aminokiseline Phe 146 iz ciklina T1, kao i u oblasti koja odgovara vezivanju Lys 6 (Slika 32).



**Slika 32:** Primeri mapiranja vezivne površine CDK9 u simulacijama A i B



U simulaciji A, u 7 od 10 reprezentativnih struktura, probe se vezuju gde i lizin 6 (crvene probe, slika 32), pri čemu se ta oblast javlja kao *hot spot* u 2 konformacije (19, 17 proba). Grupacije oko Phe 146 (sive probe, slika 32) su prisutne u 4 od 10 struktura u simulaciji A, pri čemu se nijedna ne izdvaja kao *hot spot*. U simulaciji B, grupacije proba oko Lys 6 se javljaju u 6 od 10 struktura, od čega kao *hot spot* oblast u 3 (28, 32, 17 proba), a u oblasti oko Phe 146 u 5 od 10 struktura, od čega je jedna *hot spot* oblast (23 proba). U literaturnom protokolu, pronalaženje i jedne konformacije sa *hot spot* oblašću i manjim grupacijama proba u dovoljnoj blizini je dovoljno za povoljan sud o sposobnosti vezivne površine da veže potencijalne inhibitore (30). Povećano prisustvo oblasti u kojoj se vezuje Phe 146 u rezultatima mapiranja podudara se sa većom stabilnošću džepa u kome se ova aminokiselina vezuje.

Reprezentativne strukture na kojima je izvršeno mapiranje dobijene su RMSD grupisanjem struktura iz aMD simulacija, koristeći aminokiseline prisutne na vezivnoj površini kompleksa u kristalografskoj strukturi. Prvih sedam aminokiselina N-terminusa nisu korišćene u prvom grupisanju, jer kao vrlo fleksibilne mogu da unesu značajne fluktuacije RMSD vrednosti. Uticaj izbora aminokiselina na vezivnoj površini na odabrane strukture a time i ishod mapiranja proveren je tako što su sve strukture iz simulacije B ponovo grupisane po RMSD vrednostima svih aminokiselina sa vezivne površine (i prvih sedam aminokiselina N-terminusa) i mapirane. Dobijeni su slični rezultati: okolina Lys 6 se javlja u 8 od 10 struktura, jedna od kojih je *hot spot* oblast, okolina Phe 146 prisutna u 5 od 10 struktura, nijedna *hot spot*.

Uz opisano mapiranje statičnih izabranih struktura iz simulacija, pokušana je i primena literaturnog protokola (188) za mapiranje proteinske strukture malim organskim molekulima u toku simulacije molekularnom dinamikom. Prednosti ovog protokola su što je direktno i u potpunosti uključena fleksibilnost proteinske strukture i što primena mešanog rastvarača (voda i četiri tipa molekularnih proba: 2-propanol, acetat, acetamid, protonovani izopropilamin u odnosu 7:1:1:1) indirektno podražava prisustvo nepolarnih liganada i stabilizuje hidrofobne džepove na proteinu na sličan način. Nažalost, zbog nemogućnosti reprodukcije rezultata objavljenih u literaturi, ovo mapiranje nije obavljeno.

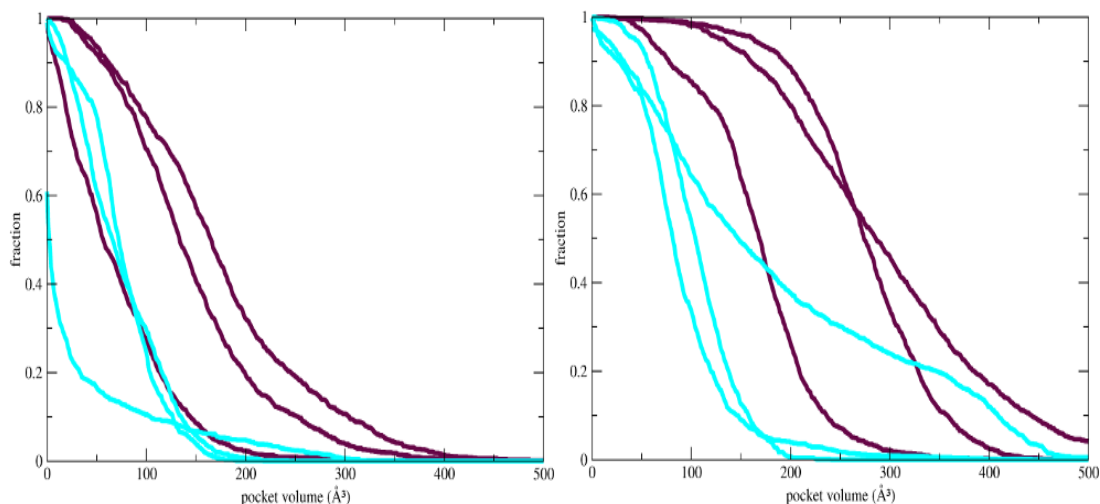
Sve obavljene analize dinamike strukture slobodne CDK9 izvedene su u cilju identifikovanja najpogodnijih oblasti na površini proteina za dizajn potencijalnih

inhibitora interakcije sa ciklinom T1. Poslednja u nizu analiza fleksibilnosti bila je analiza otvaranja i postojanosti prolaznih džepova na vezivnoj površini CDK9. Analiza je sprovedena na svih 10000 struktura iz četiri simulacije – N1 i N2 sa skraćenim N-terminusom i simulacija A i B. Glavna mesta otvaranja prolaznih džepova bila su slična u sve tri simulacije, ali su se odnosi prisutnih džepova razlikovali. U simulacijama sa skraćenim N-terminusom CDK9, džepovi su se otvarali u regionu između petlji  $\beta 3$ – $\alpha C$  i petlje sa aminokiselinama 87–99 ( 65–75% vremena trajanja simulacije), kao i oko mesta vezivanja Lys 6 sa ciklina T1 (35% vremena). Nedostajući segment N-terminusa CDK9 je moguće objašnjenje za postojanost džepova u regionu oko petlji sa kojima on interaguje. U prilog ovom objašnjenju ide činjenica da su regioni oko tih istih petlji, u simulacijama A i B u kojima je prisutan ceo N-terminus CDK9, pokazivali otvaranje džepova po 46% i 41% vremena. Ovi džepovi su obično predstavljali uzane prostore između samog N-terminusa i petlje sa aminokiselinama 87–99. I u simulaciji A i u simulaciji B džep oko vezivnog mesta za Lys 6 iz ciklina T1 bio je više otvoren nego u simulaciji sa skraćenim N-terminusom. U simulaciji A, ovaj džep je bio i najzastupljeniji, pošto je primećen tokom 66% od ukupnog vremena trajanja simulacije, dok je u simulaciji B on bio otvoren 48% vremena. Značajno otvaranje džepa u kome se vezuje Phe 146 sa ciklina T1 gotovo da nije zabeleženo u simulacijama sa skraćenim N-terminusom i dešavalo se retko u simulaciji A, u kojoj je N-terminus nestruktuiran. U simulaciji B, sa struktuiranim N-terminusom, džep u oblasti za vezivanje Phe 146 je bio otvoren 55% vremena trajanja simulacije. Ovi rezultati se podudaraju sa rezultatima mapiranja malim probama i analize prostornog rasporeda aminokiselina CDK9 u oblasti vezivanja Phe 146.

Obavljene analize ukazuju da je prisustvo svih aminokiselina N-terminusa dovelo do stabilizacije izgleda vezivne površine oko mesta vezivanja Phe 146, jednim delom zbog eliminisanja preterane fleksibilnosti aminokiselina skraćenog N-terminusa.

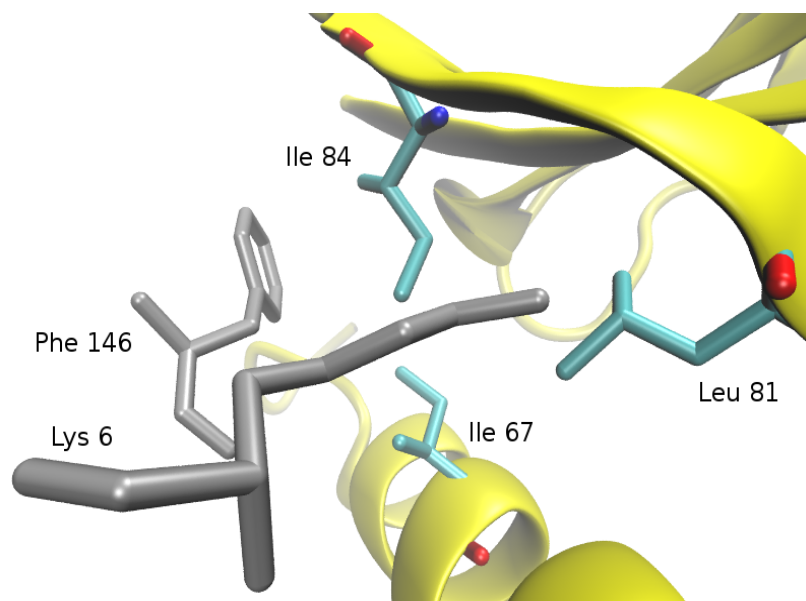
Da su oblasti vezivanja aminokiselina Phe 146 i Lys 6 pogodne za dizajn inhibitora pokazuje i dodatna analiza verovatnoće postojanja džepova na vezivnoj površini CDK9. Pokazano je da se u oblastima proteinske površine u kojima dolazi do vezivanja liganada spontano otvaraju džepovi mnogo češće nego što je to slučaj drugde na površini proteina (69). U ovom istraživanju, primenom literaturne procedure (69) ispitana je verovatnoća otvaranja džepova oko 20 aminokiselina sa vezivne površine

CDK9 i upoređena sa odgovarajućim aminokiselinama nasumično odabranim drugde na površini. Generisano je po 1000 nezavisnih konformacija CDK9, koje su zatim grupisane po veličini prisutnog džepa oko aminokiseline od interesa.



**Grafik 2:** Udeo konformacija CDK9 u kojima se otvaraju džepovi različitih zapremina kod aminokiselina 67, 81 i 84 (crvenom) i nasumično odabranih aminokiselina sa površine CDK9 u kontrolnoj simulaciji (levo) i simulaciji za analizu

Za generisanje konformacija upotrebljene su simulacije u kojima je potencirano otvaranje džepova, ako je moguće, ali i kontrolne simulacije u kojima prisustvo džepova nije uzimano u obzir. Za tri aminokiseline sa vezivne površine CDK9, Ile 67, Leu 81 i Ile 84, pokazalo se da su u njihovoj okolini prisutni džepovi značajne zapremine u velikom procentu konformacija CDK9, znatno većem nego kod aminokiselina u istim elementima sekundarne strukture drugde na površini proteina (grafik 2). Sve tri označene aminokiseline se nalaze u neposrednoj blizini mesta vezivanja aminokiselina Phe 146 i Lys 6 iz strukture ciklina T1 i stupaju u interakcije sa njima (Slika 33).



**Slika 33:** Aminokiseline sa vezivne površine CDK9 koje učestvuju u otvaranju džepova (zelenom) se nalaze u blizini vezivnog mesta aminokiselina Phe 146 i Lys 6 sa ciklina T1 (sivom)

Analiza dinamike vezivne površine slobodne CDK9 ukazala je na oblast vezivanja aminokiselina Lys 6 i Phe 146 sa ciklina T1 kao najpogodnije ciljno mesto za razvoj inhibitora interakcije. Obe primenjene metode za analizu prisustva hidrofobnih džepova ukazuju da se u toj oblasti otvara najznačajniji hidrofobni džep, koga, istovremeno, mapiranje organskim probama označava najpogodnijom oblašću na vezivnoj površini CDK9 za interakciju sa ligandima.

Nakon analize vezivnih površina CDK9 / ciklin T1 kompleksa, dizajnirani su potencijalni peptidni inhibitori interakcije, na osnovu odabranih značajnih aminokiselina sa ciklina T1: Lys 6, Lys 93, Glu 137, Gln 142 i Phe 146.

## 2.1.2 Dizajn peptidnih inhibitora

### 2.1.2.1 Dizajn peptida zasnovan na interakcijama u P-TEFb

Kao prvi korak u razvoju potencijalnih inhibitora proteinske interakcije CDK9 / ciklin T1, dizajnirani su kratki peptidi sa četiri do osam aminokiselina, čije ciljno mesto je predstavljala vezivna površina CDK9. Cilj dizajniranja peptida bio je da se kroz proučavanje njihovog vezivanja steknu dodatne informacije o interakcijama na vezivnoj površini, tj. o reproducibilnosti interakcija aminokiselina iz proteinskog kompleksa u kontekstu izolovanih peptida, bez prisustva tercijarne građe proteina koji ih orijentiše u kompleksu. Takođe, peptidi su od koristi u dobijanju početnih struktura potencijalnih inhibitora koje se mogu prevesti u stabilnije nepeptidne analoge. U ovom istraživanju, peptidi su dizajnirani koristeći dva različita pristupa.

U prvom pristupu, peptidi su dizajnirani po ugledu na građu ciklina T1, sa kojim je CDK9 u kompleksu. Izabrani su linearni fragmenti ciklina T1 koji sadrže *hot spot* aminokiseline i formirani kratki peptidi. Značajne aminokiseline uključene u građu peptida odabrane, kako je objašnjeno u prethodnom delu teksta, uključivale su Lys 93, Glu 137, Gln 142 i Phe 146 iz ciklina T1. Kriterijumi za formiranje linearnih fragmenata ciklina T1 u osnovi peptida bili su da uključuju najviše osam aminokiselina, da što više aminokiselina iz njih pripada vezivnoj površini i da se u aminokiselinskoj sekvenci peptida nalazi bar jedna, a ako je moguće više *hot spot* aminokiselina. Prilikom formiranja peptida sekundarna struktura linearnih fragmenata ciklina T1 u sklopu vezivne površine ciklina T1 je takođe uzeta u obzir, ali nije bila presudna, jer su formirani peptidi previše kratki da bi je zadržali u rastvoru, van konteksta strukture ciklina. Primenom opisanih kriterijuma izabrano je četiri peptida (Tabela 4).

**Tabela 4:** Izabrane sekvence peptida iz strukture ciklina T1

peptidi	aminokiselinske sekvence
1	LQTLGFEL
2	FLAAKV
3	ESIILQ
4	LQTLGF

u tabeli su *hot spot* aminokiseline označene tamnijim simbolima

Peptidi 1 i 4 (tabela 4) su varijacije iste sekvence. Oba peptida uključuju dve bitne aminokiseline, *hot spot* Phe 146 (F) i Gln 142 (Q) u njegovoj blizini, ali se u peptidu 1 nalaze i aminokiseline 147 i 148. Peptid 1 je uključen u analizu da bi se procenio uticaj negativnog naelektrisanja na Phe 146 (F) (C-terminus) u heksapeptidu 1 na vezivanje peptida. Ovo naelektrisanje nije prisutno u strukturi ciklina T1, kod koga se Phe 146 smešta u odgovarajući hidrofobni džep. Dodatni razlog za uključivanje oktapeptida 1 u analizu je što je Leu (L)148 iz njegove sekvence u kompleksu CDK9 / ciklin T1 u istom hidrofobnom klasteru aminokiselina kao Phe 146 i što ga je jedan od primenjenih algoritama identifikovao kao *hot spot* (HotPoint algoritam).

Peptid 2 (tabela 4) uključuje *hot spot* aminokiselinu Lys 93 i predstavlja fragment jednog od heliksa ciklina T1. Njegova sekvenca je formirana tako da se izbegne negativno naelektrisanje na Lys 93 (K) na C-terminusu peptida, ali i da se obezbedi uključivanje aminokiseline Phe 89, koja bi pružila ovom peptidu potencijalnu alternativnu orijentaciju na vezivnoj površini CDK9. Naime, Phe 89 je uključen u strukturu peptida jer je tačno tri aminokiseline udaljen od Lys 93, pa postoji mogućnost da pri vezivanju oponaša Phe 146, koga u prostoru od Lys 93 takođe deli rastojanje koje mogu da premoste tri aminokiseline u nizu. Ako bi fleksibilna struktura peptida uspela da se dovoljno prilagodi topologiji vezivne površine, postojala bi mogućnost da Phe 89 zauzme isti hidrofobni džep kao Phe 146 u kompleksu.

Peptid 3 (tabela 4) uključuje *hot spot* aminokiselinu Glu 137, kao i Gln 142 i takođe predstavlja pokušaj da se premoste dve ključne oblasti interakcija u CDK9 / ciklin T1 kompleksu: hidrofobni region oko mesta vezivanja Phe 146 (u čijoj blizini je Gln 142) i mreža vodoničnih veza koju formira Lys 93 (čiji je deo Glu 137).

Radi dodatne provere odabranih sekvenci obavljena je i *in silico* analiza čiji je cilj bio predviđanje najpogodnije peptidne sekvence u kompleksu za dalji razvoj inhibitora protein-protein interakcija. PepCrawler (189) algoritam je upotrebljen za analizu različitih peptidnih fragmenata ciklina T1 u kompleksu, njihovo vezivanje za CDK9 i procenu energije vezivanja različitih konformacija. Aminokiselinska sekvenca peptida odabranog ovim pristupom podudara se sa sekvencom oktapeptida 1 i samim tim obuhvata i sekvencu peptida 4.

Odabir aminokiselinskih sekvenci peptida 2 i 3 indirektno je podržan literaturnim podacima. U svojoj detaljnoj analizi protein-proteinskih interakcija čije

vezivne površine u svojoj građi sadrže helikse (68), Arora i Johim su izveli predviđanja mogućih inhibitornih peptida iz građe kompleksa za veliki broj struktura, među kojima je i struktura kompleksa CDK9 / ciklin T1. Oni su predvideli da heliks koji uključuje aminokiselinu Lys 93 daje 45% kontakata na vezivnoj površini, i uključuje dve aminokiseline od značaja po njihovoj analizi na razdaljini manjoj od 7Å. Kao rezultat, peptid sa sekvencom tog heliksa GNSVAPAALFLAAKVE, po analizi Arora i Johima, predstavlja dobru polaznu tačku za inhibiciju interakcije CDK9 i ciklina T1. Ovaj peptid uključuje aminokiselinsku sekvencu peptida 2 odabranog u ovoj studiji. Takođe, u okviru sekvence peptida 2 su obe značajne aminokiseline po literaturnoj analizi: Phe 89 (F) i Lys 93 (K). U istoj analizi je identifikovana i aminokiselinska sekvenca koja sadrži sekvencu peptida 3 EAYLQQVQDLVILESILQTL za koju je predviđeno da se ponaša kao slab inhibitor (68).

U drugom pristupu peptidi su dizajnirani iz izolovanih aminokiselina, nezavisno od ciklina T1, koristeći samo karakteristike vezivne površine CDK9 (objašnjeno u sledećem poglavlju).

U oba pristupa, vezivanje peptida je testirano na dve različite konformacije CDK9 - simulirane slobodne konformacije (konformacija A) i konformacije CDK9 u kompleksu sa ciklinom T1 (konformacija AB). Obe konformacije CDK9 su dobijene iz simulacija molekularnom dinamikom. Korišćenje dve različite konformacije CDK9 daje grub uvid u fleksibilnost i dinamiku vezivne površine proteina, izbegavajući veliki utrošak vremena i računarskih resursa koji bi detaljnije razmatranje fleksibilnosti CDK9 zahtevalo. Kako dizajnirani peptidi predstavljaju preliminarne inhibitore koji se dalje razvijaju, načinjeni kompromis u tretmanu dinamike proteina u ovoj fazi dizajna inhibitora je opravdan. Vezivanje dizajniranih peptida je procenjeno u prisustvu pune dinamike proteina MM-PBSA (eng. *molecular mechanics Poisson-Boltzman surface area*, molekularna mehanika sa Poisson-Bolcman tretmanom energije solvatacije) metodom.

### 2.1.2.2 Dizajn peptida iz fragmenata

U proteinskom kompleksu, proteini koriste svoju tercijarnu strukturu da pravilno pozicioniraju aminokiseline vezivnih površina jedne ka drugima i obezbede optimalnu orijentaciju bočnih lanaca za međusobne interakcije. U nedostatku proteinskog polipeptidnog skeleta i rigidnih elemenata sekundarne strukture moguće je da te iste interakcije ne bi bile i najpovoljnije koje se mogu ostvariti koristeći aminokiseline na datoj proteinskoj površini. Iz tih razloga, šire proučavanje interakcija koje su dostupne određenim aminokiselinama na datoj vezivnoj površini, van konteksta native proteinske interakcije, može da pruži dodatan uvid u povoljne početne strukture potencijalnih inhibitora, na čemu se zasniva drugi pristup dizajnu peptida u ovom istraživanju.

Kao početna tačka dizajna korišćeno je 20 prirodnih aminokiselina. Vezivna površina CDK9 je mapirana strukturama različitih aminokiselina, radi pronalaženja njihovih najpovoljnijih orijentacija na njoj, bez ograničenja uspostavljenih građom ciklina. Dizajn je sproveden na dve konformacije CDK9 – slobodnoj konformaciji A i konformaciji AB iz kompleksa sa ciklinom T1, da bi se donekle uvrstila fleksibilnost proteina u proces dizajna. Kako su u primenjenoj proceduri aminokiseline bile rigidne, u mapiranju su korišćeni njihovi rotameri, ukupno 136. Najpovoljnije orijentacije aminokiselina su zabeležene, a zatim je pokušano da se kombinacijom pozicija aminokiselina *in silico* izgrade pentapeptidi. Procedura ovog *de novo* dizajna peptida opisana je u odeljku 4.3. Ukratko, peptidi su formirani izgradnjom peptidnih veza između aminokiselina odabranih primenom genetičkog algoritma, pri čemu je cilj svake nove generacije u dizajnu bio poboljšanje energije vezivanja. Po tri peptida sa najpovoljnijom energijom vezivanja za svaku konformaciju su izabrana za dalju selekciju (Tabela 5). Za razliku od dizajna na konformaciji AB, u *de novo* dizajnu peptida na konformaciji A (slobodna CDK9) uspešno su izgrađeni samo tetrapeptidi.

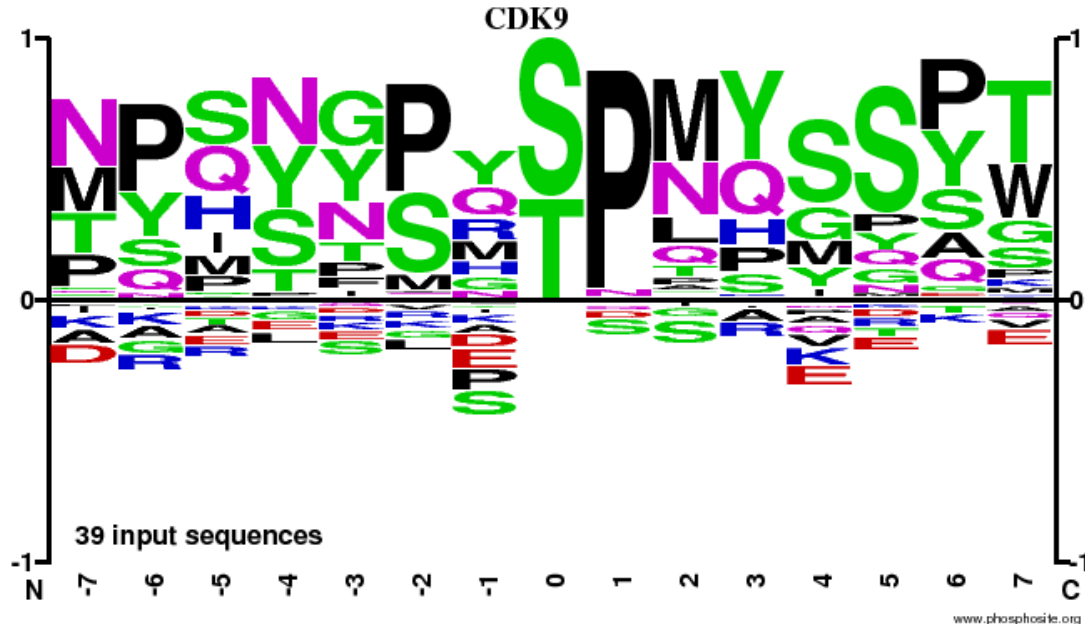
Kao negativna kontrola čitavom procesu dizajna peptida formirano je još četiri pentapeptida, nasumičnim odabirom aminokiselina. Sekvence peptida u negativnoj kontroli su: AFLIV, KIIAT, VRPHL, AAAAA i redom su označene kao peptidi 11–14. Pentapeptidi u negativnoj kontroli su korišćeni za procenu vezivanja dizajniranih peptida u odnosu na nasumično odabrane peptidne sekvence slične dužine.



**Tabela 5:** Peptidi dizajnirani iz fragmenata

peptidi	CDK9 konformacija	aminokiselinske sekvence	rang peptida u dizajnu
5	A	FWGH	1.
6	A	KPEQ	2.
7	A	KPPQ	3.
8	AB	PRWPE	1.
9	AB	PWKE	2.
10	AB	PRWKE	3.

Nakon što su formirani svi peptidi, njihove sekvence su upoređene sa preferiranim sekvencama peptidnih supstrata CDK9, da bi se smanjila verovatnoća neželjenog potencijalnog vezivanja peptida u oblasti katalitičkog džepa kinaze. CDK9, prilikom obavljanja svoje funkcije, kao i sve kinaze, pokazuje određenu selektivnost ka tipu aminokiselinskih sekvenci koje fosforiliše. Po PhosphoSitePlus (190) databazi, u 39 sekvenci peptidnih supstrata CDK9 kao glavne karakteristike ističu se očekivani serin/treonin koji se fosforiliše i prolin na poziciji +1 (Slika 34).



**Slika 34:** Frekvencija zastupljenosti aminokiselina (ordinata) u sekvencama supstrata CDK9 (apscisa), grafički prikaz sa PhosphoSitePlus® databaze, [www.phosphosite.org](http://www.phosphosite.org)  
Veličina simbola proporcionalna je verovatnoći nalaženja date aminokiseline na označenoj poziciji u sekvenci supstrata (apscisa)

Nijedan od dizajniranih peptida se nije uklapao u supstratni profil CDK9.

### 2.1.2.3 Doking peptida

Po odabiru aminokiselinskih sekvenci peptida od interesa, bilo je neophodno utvrditi najpovoljniju orijentaciju izabраниh peptida na vezivnoj površini CDK9 i proceniti njihovu energiju vezivanja. Široko primenjivana tehnika za dobijanje početnih orijentacija liganada u kompleksu sa proteinima jeste doking (eng. *docking*, pristajanje, sidrenje) u kojoj se generiše veliki broj početnih poza liganada koje se zatim grubo sortiraju po svojoj energiji interakcije sa receptorom, merenoj najčešće nekom empirijskom funkcijom čiji članovi aproksimiraju različite tipove interakcija među atomima. Doking se široko koristi kao preliminarna tehnika u dizajnu malih organskih molekula, ali je dizajn peptida znatno zahtevniji zbog njihove fleksibilnosti, tj. velikog broja stepeni slobode koje algoritam metode mora da uzme u obzir pri generisanju poza. U ovom istraživanju, doking peptida upotrebljen je kao preliminarna tehnika generisanja početnih orijentacija peptida na vezivnoj površini i za dobijanje objedinjenih preliminarnih energija vezivanja za sve dizajnirane peptide, jer je svaki od programa korišćenih u ranijim fazama dizajna koristio svoje funkcije za procenu energija vezivanja koje nisu direktno međusobno uporedive.

Za peptide 1–4, dizajnirane na osnovu aminokiselina ciklina T1, doking je bio neophodan za dobijanje početnih stanja kompleksa za dalje simulacije i procene energije vezivanja, dok su peptidi 5–10, dizajnirani iz fragmenata, već u toku formiranja dobili početne orijentacije na vezivnoj površini. Radi doslednosti i uporedivosti rezultata, i peptidi 5–10 su podvrgnuti istoj proceduri dokinga kao peptidi 1–4.

Za doking peptida upotrebljen je AutoDock Vina (191) program, čija je glavna namena doking malih organskih molekula, ali koji je u literaturi korišćen i za doking peptida na proteine, u proučavanju amiloidnih peptida (192) i kinaza (193). Prilikom dokinga, radi dobijanja što validnijih rezultata, peptidima je bilo omogućeno kako vezivanje za vezivnu površinu CDK9 u kompleksu sa ciklinom T1, tako i vezivanje u relativnom bliskom aktivnom centru CDK9. Ovo se pokazalo kao korisna odluka, jer su se neki od peptida pozicionirali dominantno, ili čak isključivo u aktivnom centru CDK9. Za te peptide se može smatrati da imaju mali afinitet ka vezivnoj površini CDK9 i kako nisu mogle biti generisane njihove početne konformacije na vezivnoj površini, oni i nisu uključeni u dalje faze istraživanja. Isključena su tri od četiri pentapeptida iz negativne

kontrole (peptidi 11, 12 i 14) i peptidi 5 (FWGH) i 10 (PRWKE) dobijeni dizajnom iz fragmenata (Tabela 6).

**Tabela 6:** Rezultati dokinga peptida na konformacije A i AB CDK9

peptidi	aminokiselinske sekvence	energije vezivanja [kcal·mol <sup>-1</sup> ]	
		konformacija A	konformacija AB
1	LQTLGFEL	7.2	6.9
2	FLAAKV	– <sup>a</sup>	6.9
3	ESIILQ	5.9	–
4	LQTLGF	–	6.4
5	FWGH	–	–
6	KPEQ	5.7	–
7	KPPQ	5.8	5.8
8	PRWPE	7	6.7
9	PWKE	6.6	6.4
10	PRWKE	–	–
11	AFLIV	–	–
12	KIIAT	–	–
13	VRPHL	7	6.7
14	AAAAA	–	–

<sup>a</sup> – označava da nisu pronađene povoljne orijentacije peptida na vezivnoj površini u najboljih 10 rezultata dokinga

Peptidi razvijeni na osnovu strukture ciklina T1 (peptidi 1–4, tabela 6) pokazali su bolje vezivanje za CDK9 konformaciju AB, što je i bilo očekivano jer je to konformacija CDK9 u kompleksu sa ciklinom T1, dok su se peptidi dobijeni *de novo* dizajnom (peptidi 6–9) vezivali za obe konformacije CDK9. Od nasumično odabranih pentapeptida iz negativne kontrole, samo je VRPHL pokazao vezivanje za CDK9. Ako se uporede energije vezivanja peptida određene aproksimativnom funkcijom u sklopu doking algoritma, može se videti da je energija vezivanja peptida 13 (VRPHL) uporediva sa energijama vezivanja dizajniranih peptida. Svi peptidi za koje je uspešno generisana orijentacija na vezivnoj površini bilo koje od dve CDK9 konformacije uključeni su u dalje istraživanje.

#### 2.1.2.4 Procena energije vezivanja peptida MM-PBSA metodom

Funkcije kojima se procenjuje energija vezivanja liganada u doking proceduri su optimizovane za brzo dobijanje aproksimativnih rezultata. Kao rezultat, neophodno je nezavisno proceniti energiju vezivanja liganada koji su pozicionirani na površini proteina u doking proceduri tačnijim metodama. U ovom istraživanju, za procenu slobodnih energija vezivanja liganada odabrane su metode MM-PBSA i MM-GBSA (eng. *molecular mechanics generalized Born surface area*, molekularna mehanika sa uopštenim Bornovim opisom solvatacije).

MM-PBSA i MM-GBSA metode u osnovi imaju sličnu formulaciju i razlikuju se samo u tretmanu energije solvatacije. Slobodna energija vezivanja, kako je definiše MM-PBSA(194) metoda, može se predstaviti kao razlika u slobodnim energijama kompleksa i slobodnog receptora i liganda, ili jednačinama:

$$\overline{\Delta G}_{\text{vez}} = \overline{\Delta E}_{\text{MM}} + \overline{\Delta G}_{\text{solv}} - T\overline{\Delta S}$$

$$\Delta E_{\text{MM}} = E_{\text{MM}}^{\text{kompleks}} - E_{\text{MM}}^{\text{protein}} - E_{\text{MM}}^{\text{ligand}}$$

$$\Delta G_{\text{solv}} = \Delta G_{\text{solv}}^{\text{kompleks}} - \Delta G_{\text{solv}}^{\text{protein}} - \Delta G_{\text{solv}}^{\text{ligand}}$$

$$\Delta S = S_{\text{kompleks}} - S_{\text{protein}} - S_{\text{ligand}}$$

$$\Delta G_{\text{solv}}^{\text{PBSA}} = \Delta G_{\text{solv}}^{\text{polarna}} + \Delta G_{\text{solv}}^{\text{nepolarna}}$$

$$E_{\text{MM}} = E_{\text{vez}} + E_{\text{ugl}} + E_{\text{torz}} + E_{\text{vdW}} + E_{\text{elec}}$$

U navedenim izrazima se  $E_{\text{MM}}$  odnosi na komponentu energije koja se procenjuje molekularnom mehanikom,  $G_{\text{solv}}$  na energiju solvatacije, a  $S$  na konformacionu entropiju, tj. na entropijsku cenu ograničavanja konformacione slobode sistema usled zaključavanja molekula u vezivne konformacije.

Za procenjivanje energije vezivanja MM-P(G)BSA metodama, potrebno je uzorkovati konformacioni prostor kompleksa i njegovih članova, što se postiže molekularnom dinamikom. Konformacije dobijene molekularnom dinamikom služe i za

procenu energije vezivanja sistema i za procenu entropijske cene vezivanja. U fazi prikupljanja konformacija u sistemu su prisutni molekuli rastvarača i direktno se uzima u obzir dinamika i prilagođavanje receptora i liganda vezanom stanju. Sama procena energije vezivanja odvija se u implicitnom rastvaraču, koji se simulira kao homogeno polje određene dielektrične konstante i jonske jačine.

Sama MM-P(G)BSA metoda se može primenjivati na dva načina – simuliranjem kompleksa i njegovih komponenti pojedinačno (metoda tri trajektorije) ili samo simuliranjem kompleksa (metoda jedne trajektorije). Metoda tri trajektorije je pogodna u slučajevima gde se očekuju drastične konformacione promene između slobodnog i vezanog stanja komponenti kompleksa. Metoda jedne trajektorije obezbeđuje tačnije potiranje unutrašnjih energija molekula u kompleksu i slobodnom stanju (jer njihova razlika mora biti nula u jednoj simulaciji) i zahteva manje kompjuterskih resursa. Ona je i primenjena u ovom istraživanju. Testirane su dve različite jonske jačine implicitnih rastvarača (koncentracija jona 100 i 150 mM). U oba slučaja dobijeni su slični rezultati.

**Tabela 7:** MM-PBSA i MM-GBSA energije vezivanja peptida za CDK9 konformacije A i AB, bez uračunavanja konformacione entropije

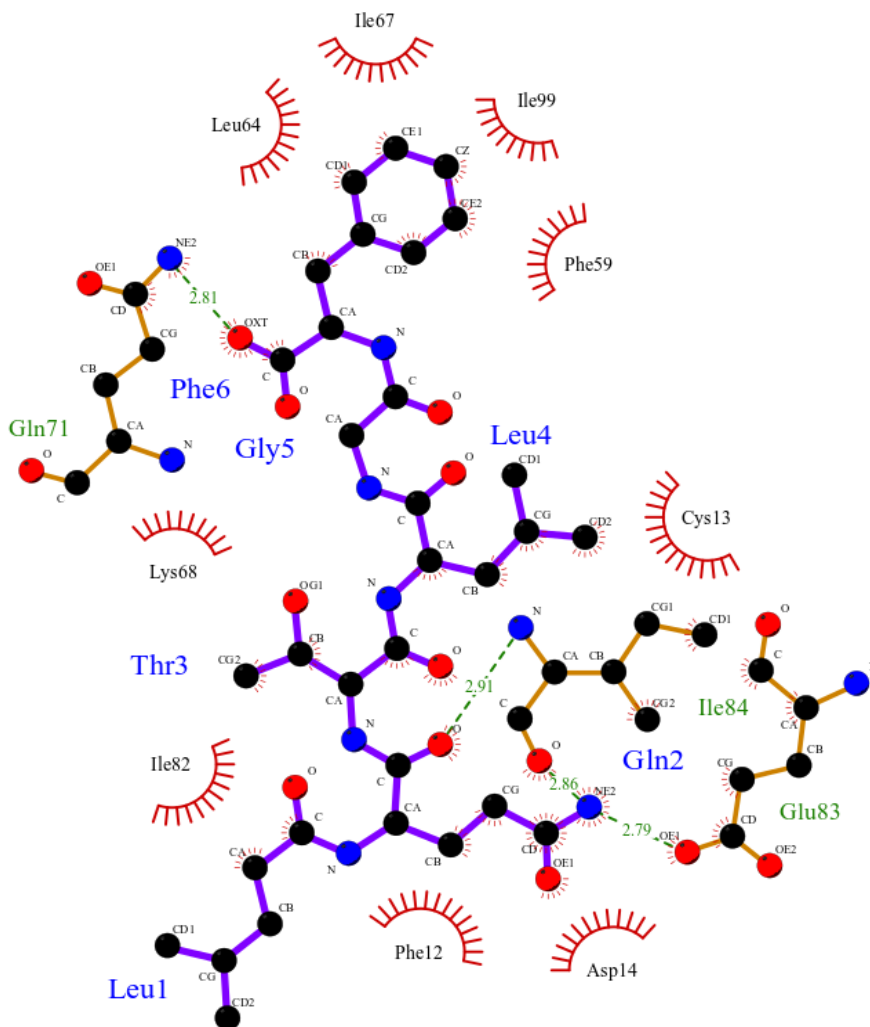
peptidi	sekvence	MM-PBSA energije [kcal·mol <sup>-1</sup> ]		MM-GBSA energije [kcal·mol <sup>-1</sup> ]	
		A	AB	A	AB
1	LQTLGFEL	-15.5±6.8	-22.2±5.7	-14.1±7.5	-23.1±5.9
2	FLAAKV	/	-21.8±5.8	/	-19.1±4.2
3	ESIILQ	-40.9±5.6	/	-37.6±4.6	/
4	LQTLGF	/	-23.4±4.4	/	-20.0±3.8
5	FWGH	/	/	/	/
6	KPEQ	-26.5±5.5	/	-24.0±4.1	/
7	KPPQ	-12.1±4.5	-9.8±3.2	-13.2±3.2	-8.3±2.4
8	PRWPE	-33.9±7.5	-14.8±3.8	-34.8±5.7	-13.8±4.0
9	PWKE	-17.7±5.1	-14.9±4.7	-17.7±5.0	-14.9±3.2
10	PRWKE	/	/	/	/
11	AFLIV	/	/	/	/
12	KIIAT	/	/	/	/
13	VRPHL	-24.1±7.0	-18.8±3.8	-27.7±4.8	-14.4±2.8
14	AAAAA	/	/	/	/

Tačnost rezultata MM-PBSA i MM-GBSA metoda se razlikuje u zavisnosti od sistema koji se istražuje, ali je MM-PBSA metoda kompjuterski zahtevnija i u proseku preciznija. U ovom istraživanju MM-PBSA metoda je korišćena kao glavna metoda za procenu energija vezivanja peptida, dok je MM-GBSA metoda korišćena za proveru slaganja rezultata (Tabela 7).

Zbog velikog utroška računarskog vremena potrebnog za njeno izračunavanje, promena konformacione entropije vezivanja je procenjena samo za 4 peptida u kompleksu sa CDK9, koristeći po pet konformacija za izračunavanje. U proseku, dobijene vrednosti promene konformacione entropije po formiranju kompleksa CDK9-peptid (TΔS), procenjene na 310 K, kretale su se od  $-20$  do  $-25 \pm 10$  kcal·mol<sup>-1</sup>. Kako je u proceni promene konformacione entropije korišćen mali broj konformacija, dobijene numeričke vrednosti nisu statistički pouzdane i imaju veliku grešku određivanja. Kao rezultat, njihovo uvrštavanje u ukupne vrednosti MM-PBSA i MM-GBSA energija nije od koristi, jer bi unelo veliku grešku u numeričke rezultate, bez značajnog poboljšanja njihovog kvaliteta. Velika greška određivanja promene konformacione entropije po formiranju kompleksa peptid-CDK9 sprečava i direktno poređenje slobodnih energija vezivanja peptida. Međutim, sve dobijene promene konformacione entropije različitih peptida su istog reda veličine i kao takve, uporedive, dozvoljavajući da peptidi budu kvalitativno rankirani po svojim energijama vezivanja utvrđenim MM-PBSA i MM-GBSA metodom, bez entropijskog doprinosa (tabela 7).

Dobijene vrednosti energija vezivanja peptida za CDK9 ukazuju da se peptidi dizajnirani na osnovu vezivne površine ciklina T1 (peptidi 1–4) bolje vezuju za CDK9 konformaciju AB, koja odgovara konformaciji CDK9 u kompleksu sa ciklinom T1, od peptida razvijenih dizajnom iz fragmenata. Najbolje vezivanje na konformaciji CDK9 iz kompleksa sa ciklinom T1 ostvaruje peptid 4 (energija vezivanja  $-23.4 \pm 4.4$  kcal·mol<sup>-1</sup>). Peptid 4 (sekvenca LQTLGF) po vezivanju uspešno oponaša vezivanje *hot spot* aminokiseline Phe 146 iz svog sastava, pozicionirajući je u isti hidrofobni džep u kome se ta aminokiselina nalazi u CDK9 / ciklin T1 kompleksu i formirajući dodatnu vodoničnu vezu preko svog C-terminalnog karboksilatnog jona. Ova vodonična veza, kao i činjenica da je heksapeptid 4 ostvario bolje vezivanje za površinu CDK9 od srodnog oktapeptida 1 pokazuje da uvođenje negativnog naelektrisanja u neposrednu blizinu *hot spot* aminokiseline Phe 146 nije negativno uticalo na njeno vezivanje, već je

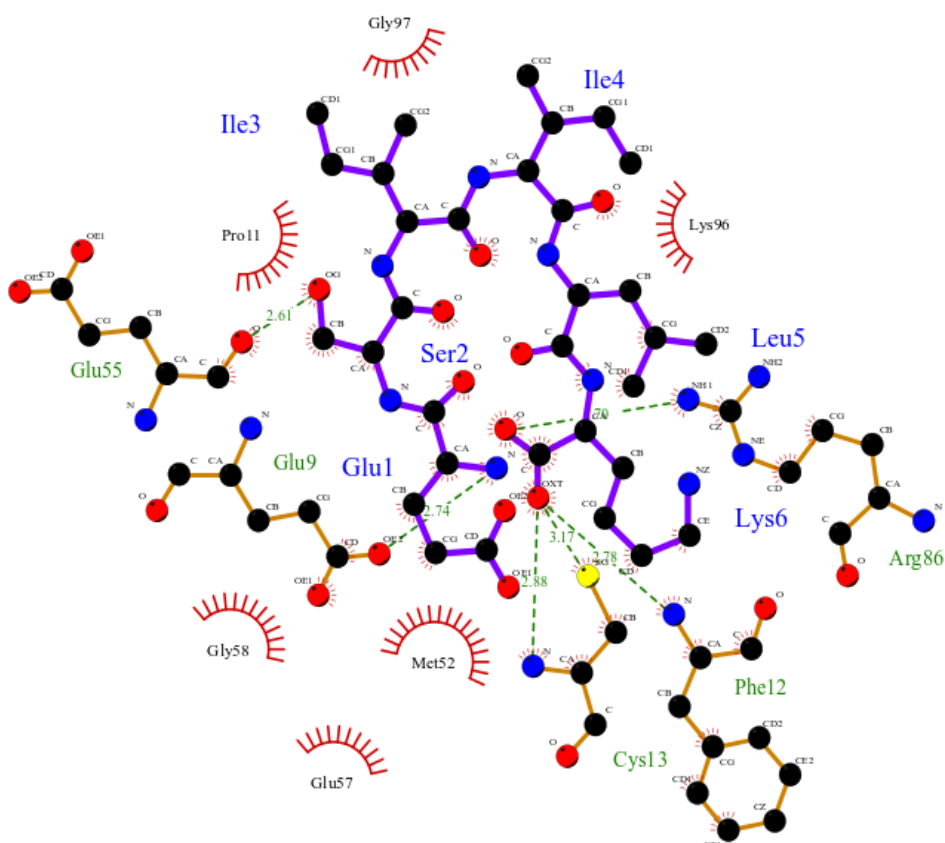
omogućilo formiranje dodatnih interakcija. Uz navedene interakcije *hot spot* aminokiseline u svom sastavu, peptid 4 pri vezivanju ostvaruje još tri vodonične veze preko glutamina u svom sastavu (Slika 35).



**Slika 35:** Šematski prikaz interakcija peptida 4 sa CDK9. Prikazane su vodonične veze (zeleno) i hidrofobne interakcije (crveno)

Peptid 4 se prilikom vezivanja za CDK9 konformaciju AB vezuje u istoj oblasti u kojoj se, pri vezivanju ciklina T1, vezuju ciklin T1 aminokiseline Phe 146 i Lys 6. Peptid 4 delom oponaša način vezivanja *hot spot* aminokiselina. Vezivanje ovog peptida je u saglasnosti sa analizom povoljnih interakcija vezivne površine CDK9 sa malim molekulima.

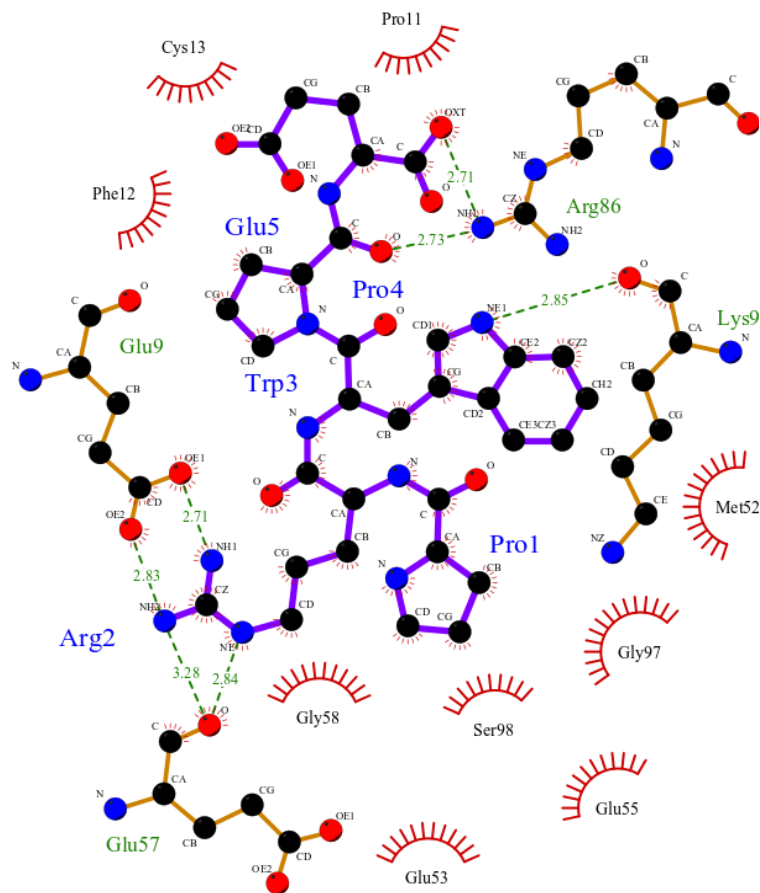
*De novo* dizajnirani peptidi (peptidi 5–10) su pokazali bolje vezivanje za površinu CDK9 konformacije A, ali je peptid sa najboljom energijom vezivanja na ovoj konformaciji, peptid 3 (sekvenca ESIIILQ), razvijen na osnovu aminokiselina ciklina T1. Energija vezivanja peptida 3 iznosila je  $-40 \pm 5.6 \text{ kcal}\cdot\text{mol}^{-1}$ . Najveći broj interakcija peptid 3 ostvaruje svojim C-terminalnim karboksilatom, koji gradi vodonične veze sa CDK9 aminokiselinama Phe 12, Cys 13 i Arg 86 (Slika 36).



**Slika 36:** Šematski prikaz interakcija peptida 3 sa CDK9. Prikazane su vodonične veze (zeleno) i hidrofobne interakcije (crveno)

Najbolje vezivanje među *de novo* dizajniranim peptidima ostvario je peptid 8 (sekvenca PRWPE) koji se vezuje za obe konformacije CDK9 korišćene u studiji, i nalazi se među najbolje rangiranim kandidatima pri vezivanju za konformaciju slobodne CDK9, sa energijom vezivanja od  $-33.9 \pm 7.5 \text{ kcal}\cdot\text{mol}^{-1}$ . Interakcijama peptida 8 sa CDK9 konformacijom A dominiraju interakcije dva arginina - arginina iz peptida 8 koji gradi vodonične veze sa Glu 9 i 57 iz CDK9 i Arg 86 iz strukture CDK9 koji interaguje sa C-terminusom peptida 8 (Slika 37).





**Slika 37:** Šematski prikaz interakcija peptida 8 sa CDK9. Prikazane su vodonične veze (zeleno) i hidrofobne interakcije (crveno)

Proučavanjem vezivanja najbolje rangiranih peptida 4, 3 i 8 može se zaključiti da su razlike u izgledu vezivne površine CDK9 između konformacija AB i A značajno uticale na raspored liganada. Peptidi 4, 3 i 8 se vezuju u dve izolovane oblasti vezivne površine. Peptid 4 interaguje sa konformacijom AB u oblasti vezivanja *hot spot* aminokiseline Phe 146 sa ciklina T1. Peptidi 3 i 8 se vezuju na konformaciji A, u sličnom regionu i ostvaruju slične interakcije svojih C-terminusa sa Arg 86. Oba ova peptida se vezuju koristeći džep na vezivnoj površini nastao između  $\beta 3$ - $C\alpha$  petlje, petlje sa aminokiselinama 87–99 i skraćenog N-terminusa. Postojanje ovog džepa je verovatno uslovljeno nefiziološkim odsustvom aminokiselina 1–7 iz kristalografske strukture CDK9 iz koje je dobijena konformacija A i sa kojom je obavljena studija vezivanja peptida. Kako je postojanje ovog džepa u prisustvu čitavog N-terminusa manje verovatno, u daljem dizajnu inhibitora CDK9 / ciklin T1 interakcije fokus je bio na oblasti vezivanja peptida 4, na čiji značaj ukazuje konsenzus primenjenih metoda.

#### 2.1.2.5 Procena energije vezivanja peptida SMD metodom

Pre prelaska na dizajn nepeptidnih inhibitora CDK9 / ciklin T1 interakcije pokušana je dodatna potvrda dobijenih rezultata MM-PBSA procene energija vezivanja peptida, korišćenjem SMD (eng. *steered molecular dynamics*, usmerena molekularna dinamika, SMD) metode.

SMD (195) je varijanta molekularne dinamike u kojoj se na jedan deo sistema koji se proučava deluje dodatnom silom, usmeravajući njegovo kretanje u određenom pravcu. Postoje dva oblika usmerene molekularne dinamike: SMD sa konstantnom brzinom kretanja odabranog dela sistema, u kome se na sistem deluje promenljivom silom podešenom tako da se uvek zadrži ista brzina kretanja, i SMD sa primenom konstantne sile, u kojoj brzina posmatranog dela sistema zavisi od trenutno prisutnih interakcija u sistemu koje se suprotstavljaju ili potpomažu željeno kretanje sistema. Praktično, primena sile na odabrani deo sistema je rešena tako što se uspostavlja zamišljena opruga između odabranog i referentnog dela sistema, koji se drži fiksiran. Zamišljenoj opruzi se dodeljuje određena elastičnost i po primeni sile dolazi do njenog rastezanja ili sabijanja.

Osim primene koju SMD ima u proučavanju usmerene dinamike različitih sistema, SMD je u literaturi korišćena i za relativno upoređivanje energija vezivanja malih molekula (196) i proteina (197). Ako se kombinuje za Jarzinskijevom jednačom, SMD se može koristiti za rigorozan tretman slobodne energije (195), ali je u ovom istraživanju primenjena samo za grubo poređenje relativnih afiniteta dizajniranih peptida.

Kada se SMD primenjuje kao metoda za upoređivanje afiniteta različitih molekula, obično se koristi varijanta metode u kojoj se posmatrani deo sistema kroz celu simulaciju kreće konstantnom brzinom, a odnos afiniteta se dobija iz odnosa visine pikova primenjene sile u zavisnosti od vremena. Naime, što je veći afinitet nekog liganda za protein za koji se vezuje, biće potrebno primeniti veću silu za odvajanje liganda od proteina, što se na grafičkom prikazu zavisnosti primenjene sile od proteklog vremena u simulaciji (profilu primenjene sile) uočava kao pik. Pod uporedivim uslovima simulacije, ligand sa većim afinitetom pokazaće i više vrednosti primenjene sile prilikom odvajanja od proteina.

SMD simulacije započete su od početnih orijentacija peptida na vezivnoj površini CDK9, dobijenih dokingom. Peptidi su sa vezivne površine CDK9 povlačeni konstantnom brzinom i beležene su vrednosti primenjene sile. U početnoj fazi istraživanja ispitano je više različitih referentnih fiksiranih segmenata sistema, brzina povlačenja peptida i konstanti elastičnosti opruga.

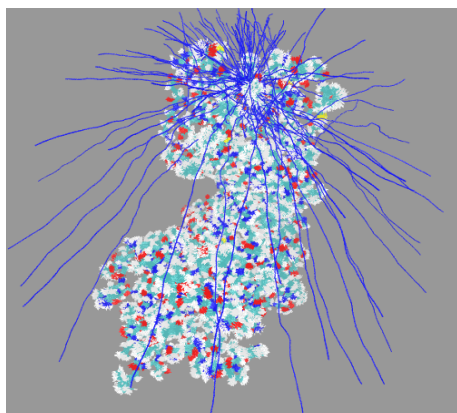
Testirana su tri segmenta fiksiranih atoma: fiksiran samo jedan atom – Ca aminokiseline 20, koja se nalazi u rigidnoj  $\beta$ -valjak strukturi udaljenoj od mesta vezivanja peptida, fiksirani svi atomi polipeptidnog niza CDK9 i fiksirana distalna strana  $\beta$ -valjka nasuprot vezivne površine (atomi polipeptidnog niza aminokiselina 20 – 25, 32 – 37, 44 – 48, 104, 106, 107 strukture CDK9). U izvođenju SMD simulacija pogodno je odabrati fiksirane referentne segmente tako da uključuju što manji broj atoma (po mogućstvu jedan) jer je svako fiksiranje koordinata sistema veštačko ograničavanje konformacione slobode sistema i može da utiče na rezultate. Sa druge strane, prilikom povlačenja liganada, naročito povlačenja većim brzinama, a samim tim i većom silom, može doći do nefizioloških deformacija proteinske strukture u blizini liganada, što se može ublažiti pažljivo odabranim fiksiranim referentnim segmentom. Kao kompromis ova dva efekta, odabrano je da se fiksira deo  $\beta$ -valjka naspram vezivne površine CDK9 (atomi polipeptidnog niza aminokiselina 20 – 25, 32 – 37, 44 – 48, 104, 106, 107 strukture CDK9).

Testirane su četiri različite brzine povlačenja peptida: 5, 10, 50 i 100 Å/ns. U odabiru brzine povlačenja peptida poželjno je odabrati što manju vrednost, jer se tako minimizuje mogućnost deformacije okolne strukture sistema usled visokih primenjenih sila, ali male brzine povlačenja značajno produžavaju vreme trajanja same simulacije. Odabrana je brzina od 10 Å/ns, kojom je postignuto odvlačenje svih peptida sa površine CDK9 u predviđenom vremenu simulacija od 4ns.

Poslednji parametar koji je potrebno razmotriti prilikom SMD simulacija jeste konstanta elastičnosti zamišljene opruge koja se povlači. Konstanta elastičnosti opruge direktno utiče na izgled dobijenog profila sila, tj. na nivo detalja o molekularnim događajima koji su ga izazvali. Što je konstanta opruge veća, opruga je kruća, što smanjuje mogućnost prilagođavanja strukture liganada promeni položaja tokom povlačenja i daje detaljan izgled profila primenjene sile. Primena krute opruge je scenario koji najmanje odgovara prirodnom procesu disocijacije liganada sa proteinske

površine, u kome se ligandi stalno prilagođavaju promenama. U ovom istraživanju testirane su tri konstante elastičnosti opruge iz literature: 1.5, 3 i 7 kcal·mol<sup>-1</sup>·Å<sup>2</sup>. Izabrana je najmanja konstanta (1.5 kcal·mol<sup>-1</sup>·Å<sup>2</sup>), koja daje elastičnu oprugu, da bi se što bolje opisalo ponašanje fleksibilnih peptida kao liganada. Gubitak molekularnih detalja na dobijenim profilima sila nije predstavljao problem, jer je cilj studije bilo samo kvalitativno rangiranje dizajniranih peptida po njihovom afinitetu za vezivnu površinu CDK9, što odgovara redosledu visina dominantnog pika na svakom profilu sila.

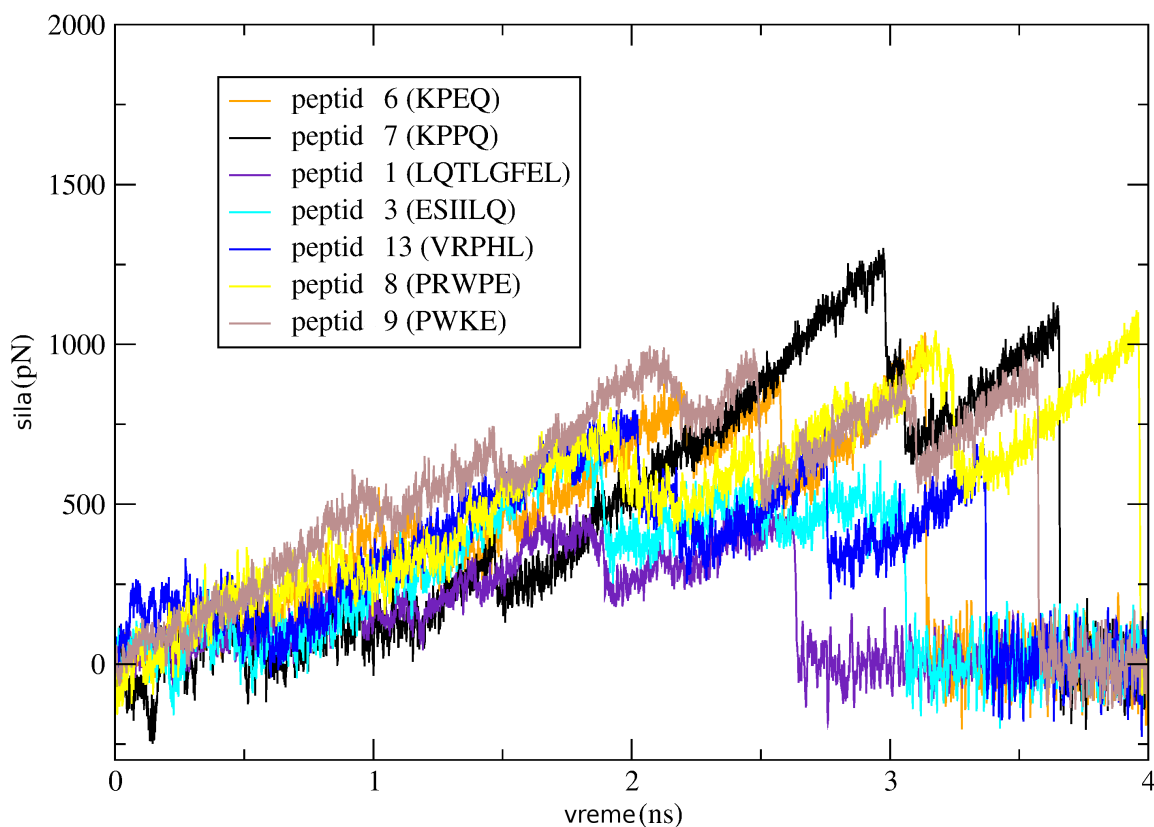
Kako je vezivna površina CDK9 sa koje su peptidi odvlačeni relativno ravna, izbor pravca u kome se vrši povlačenje, dokle god je pravac uporediv između svih peptida, nije od važnosti, što nije slučaj kada se SMD koristi na ligandima u aktivnim mestima enzima. Radi provere te tvrdnje, distribucija mogućih putanja peptida kada napuštaju vezivnu površinu CDK9 testirana je kroz 100 RAMD (eng. *random acceleration molecular dynamics*, molekularna dinamika sa nasumično orijentisanim ubrzanjem liganada) (198) simulacija. U RAMD simulacijama se ligandi ubrzavaju u nasumično odabranim pravcima i zahteva se da prilikom kretanja pređu određeno rastojanje u svakom koraku simulacije. Ako to nije slučaj, ligand je naišao na prepreku prilikom kretanja i tada se njegov pravac kretanja ponovo nasumično menja. Postupak se ponavlja sve dok ligand koji se posmatra ne napusti okruženje proteina u potpunosti, tj. dok se ne udalji na kritičnu razdaljinu. U ponovljenim RAMD simulacijama, pojava prostornog grupisanja putanja liganda prilikom napuštanja mesta vezivanja ukazuje na dominantan, najpovoljniji pravac tog procesa. Ako takav pravac postoji, njega treba uzeti kao pravac povlačenja liganada tokom SMD simulacija, jer je statistički najverovatniji.



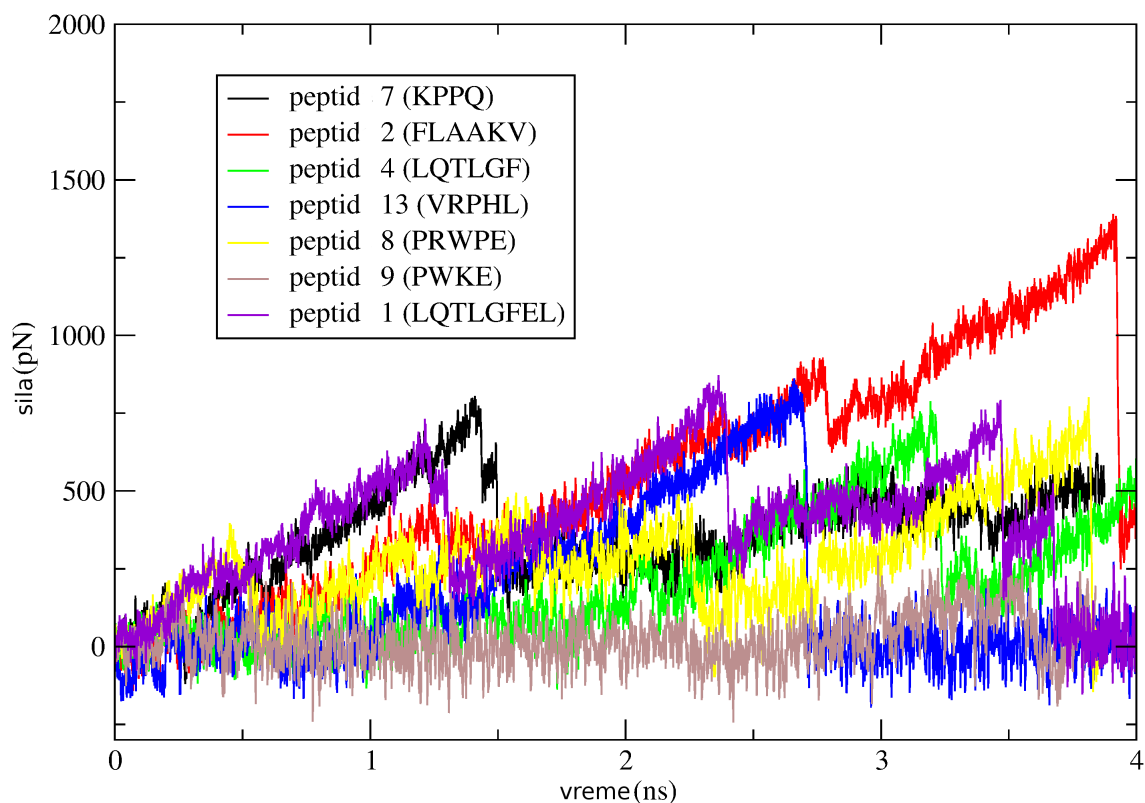
**Slika 38:** Prostorna raspodela 100 RAMD putanja peptida sa vezivne površine CDK9 (plave linije)

U 100 RAMD simulacija sprovedenih na jednom od kompleksa peptid-CDK9, nije zapaženo grupisanje RAMD putanja (Slika 38). Putanje peptida su bile nasumično raspoređene u prostoru, ukazujući da je bilo koji od odabranih pravaca kretanja peptida podjednako moguć i validan, dokle god je uporediv između različitih peptida i da izbor pravca povlačenja peptida sa površine CDK9 ne utiče kritično na rezultate dobijene SMD simulacijama.

Profili sila dobijeni povlačenjem peptida sa površina CDK9 konformacija A i AB prikazani su na graficima 3 i 4. Pikovi na graficima označavaju raskidanje pojedinačnih interakcija između peptida i CDK9. Visina pikova (primenjena sila) je proporcionalna jačini interakcije koja se raskida. Za svaki peptid, najintenzivniji pik ukazuje na trenutak u kome se peptid oslobađa sa površine proteina u rastvor. Poređenje visine najintenzivnijih pikova različitih peptida ukazuje na njihov relativni afinitet ka CDK9.



**Grafik 3:** Profili sila dobijeni SMD simulacijama povlačenja peptida sa vezivne površine CDK9 konformacije A

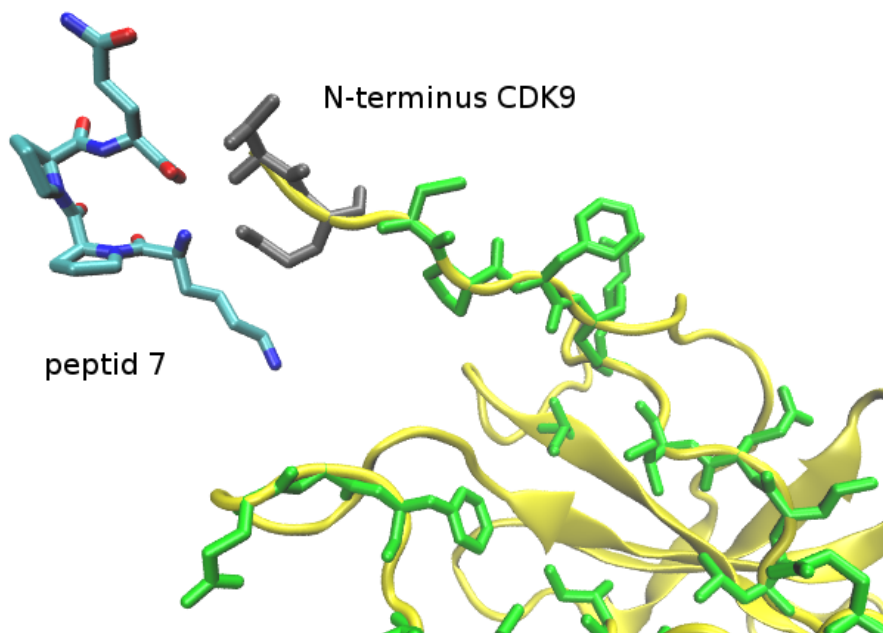


**Grafik 4:** Profili sila dobijeni SMD simulacijama povlačenja peptida sa vezivne površine CDK9 konformacije AB

Kao što grafici 3 i 4 pokazuju, rezultati SMD analize se razlikuju od redosleda peptida dobijenog MM-PBSA metodom. Zbog upotrebe različitih polja sila u analizi, direktno poređenje rezultata nije izvodljivo, ali glavni razlog neslaganja je različita evolucija sistema u SMD simulacijama u poređenju sa MD simulacijama u MM-PBSA analizi.

Naime, tokom poslednje nanosekunde klasične MD simulacije iz koje su uzorkovane konformacije sistema za MM-PBSA analizu, peptidi su ostali na vezivnoj površini CDK9, dok je u toku SMD simulacija, fleksibilni N-terminus CDK9 u blizini vezivne površine stupao u dodatne interakcije sa peptidima dok su se udaljavali sa vezivne površine. Ove dodatne interakcije ne opisuju vezivanje peptida za samu vezivnu površinu CDK9, ali se njihovim raskidanjem stvaraju dodatni, intenzivni pikovi u profilima sila, koji menjaju izgled profila. Ova pojava je najizraženija kod peptida 7 (sekvenca KPPQ) i njegove interakcije sa CDK9 konformacijom A. Energija vezivanja peptida 7 za CDK9 konformaciju A, procenjena MM-PBSA metodom je svega  $-12.1 \pm 4.5 \text{ kcal}\cdot\text{mol}^{-1}$ , ali profil primenjene sile u SMD simulaciji ovog kompleksa pokazuje

najveći pik među profilima peptida koji se vezuju za CDK9 konformaciju A, ukazujući da, među njima, peptid 7 pokazuje najveći afinitet vezivanja. Međutim, pik u pitanju ne odgovara raskidanju interakcija peptida 7 sa vezivnom površinom CDK9 konformacije A, već kasnijem molekularnom događaju u simulaciji, kada je peptid 7 već napustio vezivnu površinu proteina i stupa u dodatne interakcije sa fleksibilnim N-terminusom kinaze (Slika 39).



**Slika 39:** Interakcije peptida 7 sa N-terminusom CDK9 konformacije A (sivo). Prikazane su aminokiseline vezivne površine CDK9 (zeleno)

Sila koja je potrebna da se raskinu interakcije peptida 7 i vezivne površine CDK9 odgovara inicijalnom segmentu na profilu sile i mnogo je manja, u skladu sa niskom predviđenom MM-PBSA energijom vezivanja. Isto se može zaključiti i za vezivanje peptida 2 (FLAAKV) za CDK9 konformaciju AB. U većoj ili manjoj meri, dodatne interakcije sa fleksibilnim N-terminusom CDK9 su prisutne u SMD simulacijama svih peptid-CDK9 kompleksa. Fiksiranjem aminokiselina N-terminusa bilo bi moguće izbeći dodatne interakcije sa peptidima i njihov uticaj na profile sile u SMD simulacijama, ali bi se time veštački smanjila konformaciona sloboda regiona proteinske površine direktno uključene u interakcije sa peptidima, što bi takođe uticalo na rezultate.

Takođe, za razliku od liganada u aktivnom mestu enzima koji svi interaguju sa istim oblastima vezivne površine na sličan način, u kompleksima peptida korišćenih u

SMD studiji u ovom istraživanju nisu postojale dovoljno izražene zajedničke karakteristike vezivne površine u kontaktu sa peptidima koje bi definisale uniformnu tačku za poređenje profila sila. Zbog svega navedenog, profili sila dobijeni SMD simulacijama nisu mogli biti primenjeni za procenu razlika u vezivanju peptida za CDK9. Kao rezultat, glavni kriterijum po kome su rangirani peptidi i odabrani za dobijanje informacija o daljem dizajnu bile su predviđene MMPBSA energije vezivanja, po kojima su peptidi 4, 3 i 8 bili najbolji kandidati za dalji razvoj inhibitora.

Peptid 4 je fragment sekvence ciklina T1 sa *hot spot* aminokiselinom Phe 146. Pri vezivanju, peptid 4 oponaša način vezivanja Phe 146 i pozicionira se u hidrofobnom džepu na vezivnoj površini CDK9, u oblasti vezivanja Lys 6 i Phe 146 sa ciklina T1, u regionima u kojima je mapiranjem malim organskim probama predviđeno vezivanje liganada. Peptidi 3 i 8 se vezuju za drugi region vezivne površine CDK9, ostvarujući povoljne interakcije sa Arg 86, ali kako je vezivno mesto peptida 3 i 8 delom definisano nefiziološkim odsustvom sedam aminokiselina N-terminusa CDK9, dizajn nepeptidnih liganada je fokusiran na oblasti vezivanja peptida 4, tj. na oblast vezivanja aminokiselina Phe 146 i Lys 6 ciklina T1.

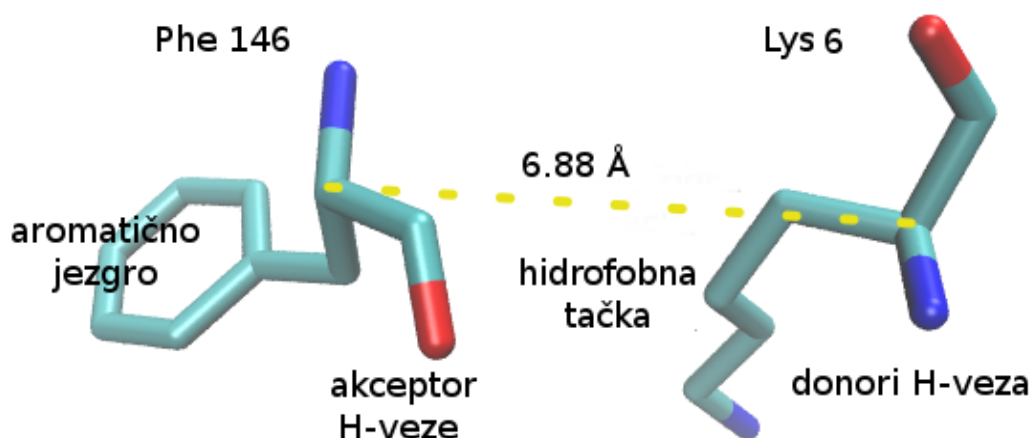


## 2.1.3 Dizajn nepeptidnih inhibitora

### 2.1.3.1 Pretraživanje farmakofora

Prvi korak u dizajnu nepeptidnih liganada bio je odabir najpovoljnijih polaznih tačaka za dizajn na vezivnoj površini CDK9. Konsenzusom analiza interakcija ostvarenih u proteinskom kompleksu CDK9 / ciklin T1, analiza dinamike i fleksibilnosti same vezivne površine CDK9 i vezivanja prethodno dizajniranih peptidnih inhibitora, naročito peptida 4, za polazne tačke definisanja strukture potencijalnih nepeptidnih inhibitora CDK9 / ciklin T1 interakcije odabrane su aminokiseline Phe 146 i Lys 6 ciklina T1 ( $C\alpha - C\alpha$  udaljenost 6.88 Å, slika 40) i interakcije koje one uspostavljaju.

Izbor ove dve aminokiseline je dodatno potvrđen u procesu definisanja farmakofore kojoj će odgovarati nepeptidni inhibitori, jer su iste dve aminokiseline predložene PocketQuery (34) metodom kao najpovoljnije polazne tačke za dizajn inhibitora sa površine ciklina T1, analizom kristalografske strukture PDB ID: 3blh. Početna farmakofora koja je definisana na osnovu ove dve aminokiseline imala je pet tačaka: aromatično jezgro koje odgovara bočnom nizu Phe 146, dva donora vodoničnih veza na azotima polipeptidnog niza i bočnog lanca Lys 6, akceptor vodonične veze na karbonilnom kiseoniku Phe 146 i hidrofobnu tačku koja odgovara ugljenicima bočnog niza Lys 6 (Slika 40).

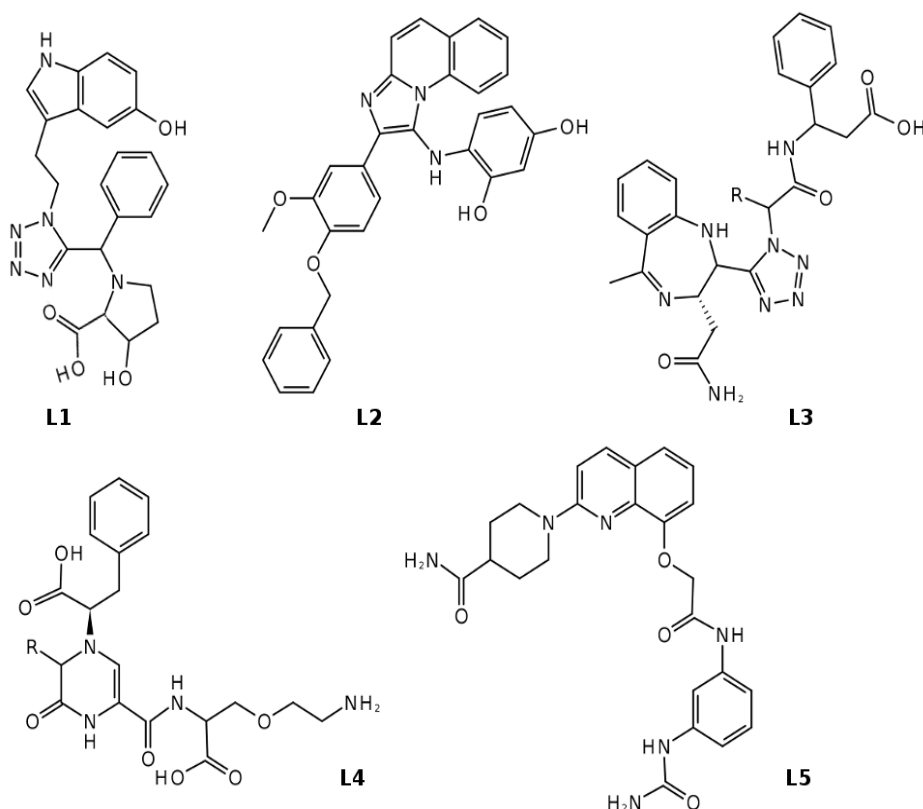


**Slika 40:** Pet tačaka farmakofore zasnovane na aminokiselinama Phe 146 i Lys 6 iz strukture ciklina T1

Primenom ove farmakofore u AnchorQuery databazi jedinjenja pronađene su tri strukture sa odgovarajućim karakteristikama. AnchorQuery databaza jedinjenja je

odabrana jer je optimizovana za razvoj inhibitora protein-protein interakcija (199). U cilju povećanja broja identifikovanih jedinjenja, ublaženi su kriterijumi farmakofore, tako što je iz njene definicije isključena hidrofobna tačka koja odgovara bočnom nizu aminokiseline Lys 6. Ova tačka je odabrana za isključenje da bi u farmakofori bilo zadržano aromatično jezgro *hot spot* aminokiseline Phe 146, koje značajno doprinosi energiji vezivanja ciklina T1 za CDK9 i sva tri elementa koja mogu da doprinesu stvaranju vodoničnih veza. Kako ugljenici bočnog niza Lys 6 učestvuju samo u Van der Valsovim interakcijama sa CDK9, smatralo se da njihovo isključivanje najmanje menja osobine farmakofore, naročito jer je u lokaciji azotovog atoma bočnog niza Lys 6 koji se ponaša kao donor vodonične veze već implicitno uključena razdaljina od polipeptidnog niza, koju isključeni ugljenikovi atomi definišu u samoj aminokiselini.

Izmenjena farmakofora je upotrebljena za pretraživanje i dobijeno je 26 mogućih struktura. Strukture su ručno podeljene po hemijskoj sličnosti i verovatnoći uspešne sinteze u pet podgrupa i za svaku podgrupu je izabrana reprezentativna struktura (strukture L1–L5, Slika 41).



**Slika 41:** Strukture odabranih liganada. R u strukturama L3 i L4 označava supstituente zamenjene metil grupom zbog sternih zahteva proteina

Dve strukture liganada, L3 i L4, su imale supstituente, 4-nitrofenil i 1-metilpropil grupu, koje su sterno ometale pravilno pozicioniranje liganada na proteinskoj površini, pa su zamenjene metil grupom. Supstituenti nisu u potpunosti eliminisani da se ne bi previše povećala fleksibilnost molekula.

U sledećem koraku dizajna procenjena je sličnost pet reprezentativnih struktura liganada sa poznatim inhibitorima protein-protein interakcija, koristeći literaturnu proceduru (64). Ukratko, za svaki od liganada je određena vrednost dva molekularna deskriptora koji opisuju molekularni oblik (RDF070m deskriptor) i rigidnost/broj nezasićenih veza ( $U_i$ , eng. *unsaturation index*, indeks nezasićenosti) i upoređena sa vrednostima koje ovi deskriptori imaju u strukturama poznatih inhibitora protein-protein interakcija ( $RDF070m \geq 13.31$ ,  $U_i \geq 4.13$ ). Molekularni deskriptori predstavljaju numeričke opise strukturnih i hemijskih karakteristika molekula, tj. svođenje molekula na matematičke izraze. Zadovoljavanje niza deskriptora se može izjednačiti sa zadovoljavanjem skupa strukturnih i hemijskih karakteristika. Primenjeno na ovo istraživanje, od izabranih struktura liganada se tražilo da pokažu određeni stepen složenosti svog oblika i rigidnost struktura. Od pet testiranih struktura, tri (L1, L2 i L3) su zadovoljile oba kriterijuma (Tabela 8).

**Tabela 8:** Klasifikacija liganada primenom molekularnih deskriptora

<b>ligandi</b>	<b>RDF070m <math>\geq</math> 13.31</b>	<b><math>U_i \geq</math> 4.13</b>	<b>odluka</b>
L1	15.373	4.248	zadovoljava
L2	13.834	5	zadovoljava
L3	18.603	4.459	zadovoljava
L4	12.178	4.087	ne
L5	12.677	4.392	1 / 2

Struktura L4 nije zadovoljila nijedan od uslova za vrednosti deskriptora i isključena je iz dalje analize. Struktura L5 je zadovoljila jedan od dva uslova i stoga je zadržana u grupi liganada za dalju analizu.

U sledećem koraku, četiri izabrane strukture liganada su dokovane za površinu CDK9 i procenjene su njihove energije vezivanja.

### 2.1.3.2 Doking liganada

Vezivanje odabranih liganada za površinu CDK9 testirano je dokingom. Primenjeni doking protokol uključuje vezivanje fleksibilnih liganada za rigidnu površinu proteina, pa je fleksibilnost proteinske strukture uključena indirektno, primenom većeg broja alternativnih konformacija CDK9, dobijenih molekularnom dinamikom. Ovaj pristup je opisan u literaturi (200). Doking je izveden na 50 različitih konformacija CDK9, dobijenih iz četiri aMD simulacije proteina (simulacije N1, N2, A i B). Tokom procedure, proveren je uticaj broja reprezentativnih konformacija iz svake simulacije, kao i uticaj načina na koji su konformacije grupisane na ishode dokinga.

Grupisanje konformacija CDK9 po sličnosti vezivne površine (po RMSD vrednostima atoma aminokiselina sa vezivne površine) obavljeno je bez (grupisanje 1) i sa uključivanjem modelovanih prvih sedam aminokiselina N-terminusa (grupisanje 2). Uticaj broja reprezentativnih konformacija CDK9 na ishod dokinga testiran je sa 10 i 20 konformacija. Uticaj je izražen kao pokrivenost konformacionog prostora proteina odabranim strukturama, tj. kao ukupan udeo broja struktura u grupacijama koje predstavljaju odabrane strukture CDK9 u 2000 struktura sa najpovoljnijom energijom iz svake od simulacija. Grupisanje konformacija je ukazivalo da se već odabirom 10 reprezentativnih struktura u svakoj simulaciji osim A (grupisanje 2) postiže zadovoljavajuća pokrivenost konformacionog prostora koji je protein posetio tokom simulacije (~ 60%) (Tabela 9). Dobijena pokrivenost konformacionog prostora ukazuje da je dovoljno upotrebiti po 10 struktura iz svake simulacije za dobro predstavljanje fleksibilnosti proteina u dokingu liganada.

**Tabela 9:** Pokrivenost posećenog konformacionog prostora CDK9 odabranim brojem reprezentativnih struktura iz aMD simulacija

simulacija	pokrivenost konformacionog prostora	
	10 struktura	20 struktura
N1 – grupisanje 1	57.70	78.50
N2 – grupisanje 1	73.30	92.80
A – grupisanje 1	79.35	93.75
A – grupisanje 2	<b>22.55</b>	<b>36.20</b>
B – grupisanje 1	88.95	97.45
B – grupisanje 2	63.05	87.70

Čak i uključivanjem 20 reprezentativnih struktura, pokrivenost konformacionog prostora simulacije A sa primenjenim grupisanjem struktura 2 nije se poboljšala značajno (svega 36.20%, tabela 9). Razlog najverovatnije leži u uticaju fleksibilnog, nestruktuiranog N-terminusa (prvih 7 aminokiselina CDK9) na ukupne RMSD vrednosti položaja aminokiselina sa vezivne površine CDK9, a time i na grupisanje konformacija. Strukture dobijene ovim grupisanjem simulacije A nisu korišćene u dokingu liganada, jer nije moglo biti isključeno da razlike u njima ne odslikavaju razlike u strukturi cele vezivne površine CDK9, već samo u položaju fleksibilnog kraja N-terminusa. Uticaj izbora načina grupisanja struktura se može videti i kod simulacije B – grupisanje 2 dovodi do većih razlika u RMSD vrednosti struktura, ali kako je u simulaciji B N-terminus delom struktuiran u okret heliksa, oscilacije RMSD vrednosti nisu tako drastične. Uticaj načina grupisanja struktura na ishod dokinga proveren je tako što su u doking uključene strukture simulacije B grupisane na oba načina. Pošto je pokrivenost konformacionog prostora proteina odabranim strukturama ukazala da je 10 struktura dovoljno za prikazivanje konformacionih promena CDK9 opaženih u simulacijama, doking je izveden na po 10 struktura iz svake simulacije.

Zbog velikog broja struktura, nije bilo izvodljivo izvesti detaljna ispitivanja energije vezivanja liganada na svakom kompleksu, već je izvršen odabir struktura. Za svaki ligand i za reprezentativnih 10 konformacija iz svake od simulacija izdvojen je kompleks sa najboljom predviđenom energijom vezivanja primenjenom aproksimativnom funkcijom za procenu energije u okviru doking procedure. Svaki od tako dobijenih kompleksa je minimizovan u implicitnom rastvaraču da bi se strukture liganda i proteina prilagodile jedna drugoj, a zatim je energija vezivanja ponovo procenjena SS-MMGBSA (eng. *single snapshot molecular mechanics generalized Born surface area*) metodom (bez računanja entropije vezivanja). Ovako dobijene, preciznije energije su iskorišćene za rangiranje konformacija liganada. Korelacija između energija vezivanja određenih funkcijom doking programa i SS-MMGBSA metodom iznosila je 0.68. Primena različitih MM-GBSA modela i koncentracija jona u okruženju molekula (0.1 i 0.15 M) nije značajno uticala na dobijene rezultate. Za svaki od liganada odabrane su dve strukture sa najpovoljnijom SS-MMGBSA energijom vezivanja i prosleđene na analizu MM-GBSA metodom (Tabela 10, obeležene strukture).

**Tabela 10:** Odnos energija vezivanja dobijenih dokingom i SS-MMGBSA metodom

<b>kompleks<sup>a</sup></b>	<b>doking energija vezivanja [kcal·mol<sup>-1</sup>]</b>	<b>SS-MMGBSA energija vezivanja [kcal·mol<sup>-1</sup>]</b>
<b>L2 – B – 2 – 1</b>	– 9.3	– 35.42
<b>L1 – B – 2 – 1</b>	–7.5	– 34.92
<b>L2 – A – 1 – 4</b>	– 8.1	– 34.40
L2 – B – 1 – 4	– 8.4	– 34.31
<b>L5 – B – 1 – 6</b>	– 8.0	– 33.50
<b>L5 – B – 2 – 1</b>	– 8.1	– 32.98
<b>L3 – B – 2 – 7</b>	– 7.1	– 29.86
L5 – A – 1 – 3	– 7.4	– 29.15
L2 – N2 – 1 – 2	– 7.2	– 27.55
L5 – N1 – 1 – 2	– 7.2	– 27.46
L5 – N2 – 1 – 3	– 7.4	– 26.63
<b>L1 – B – 1 – 1</b>	– 7.0	– 25.58
L1 – A – 1 – 4	– 7.0	– 23.59
L2 – N1 – 1 – 2	– 7.8	– 22.97
<b>L3 – N2 – 1 – 4</b>	– 6.7	– 22.42
L3 – N1 – 1 – 1	– 6.6	– 22.22
L1 – N2 – 1 – 10	– 7.0	– 21.90
L1 – N1 – 1 – 10	– 7.1	– 21.28
L3 – B – 1 – 1	– 7.0	– 20.62
L3 – A – 1 – 3	– 7.1	– 18.73

<sup>a</sup> ligand – simulacija – grupisanje – redni broj strukture

Iz tabele 10 se može videti da je bolje vezivanje liganada ostvareno na konformacijama CDK9 iz simulacija sa modelovanim N-terminusom (simulacije A i B), nego iz simulacija bez N-terminusa (simulacije N1 i N2). Od osam kompleksa liganada sa CDK9 odabranih za dalju analizu, samo je u jednom prisutna konformacija CDK9 sa skraćenim N-terminusom, i taj kompleks pokazuje najnepovoljniju energiju vezivanja liganda u SS-MMGBSA analizi.

Ako se u obzir uzme ukupno rangiranje liganada na svim konformacijama, najbolje se vezuju ligandi L5 i L2, dok se ligandi L1 i L3 vezuju slabije.

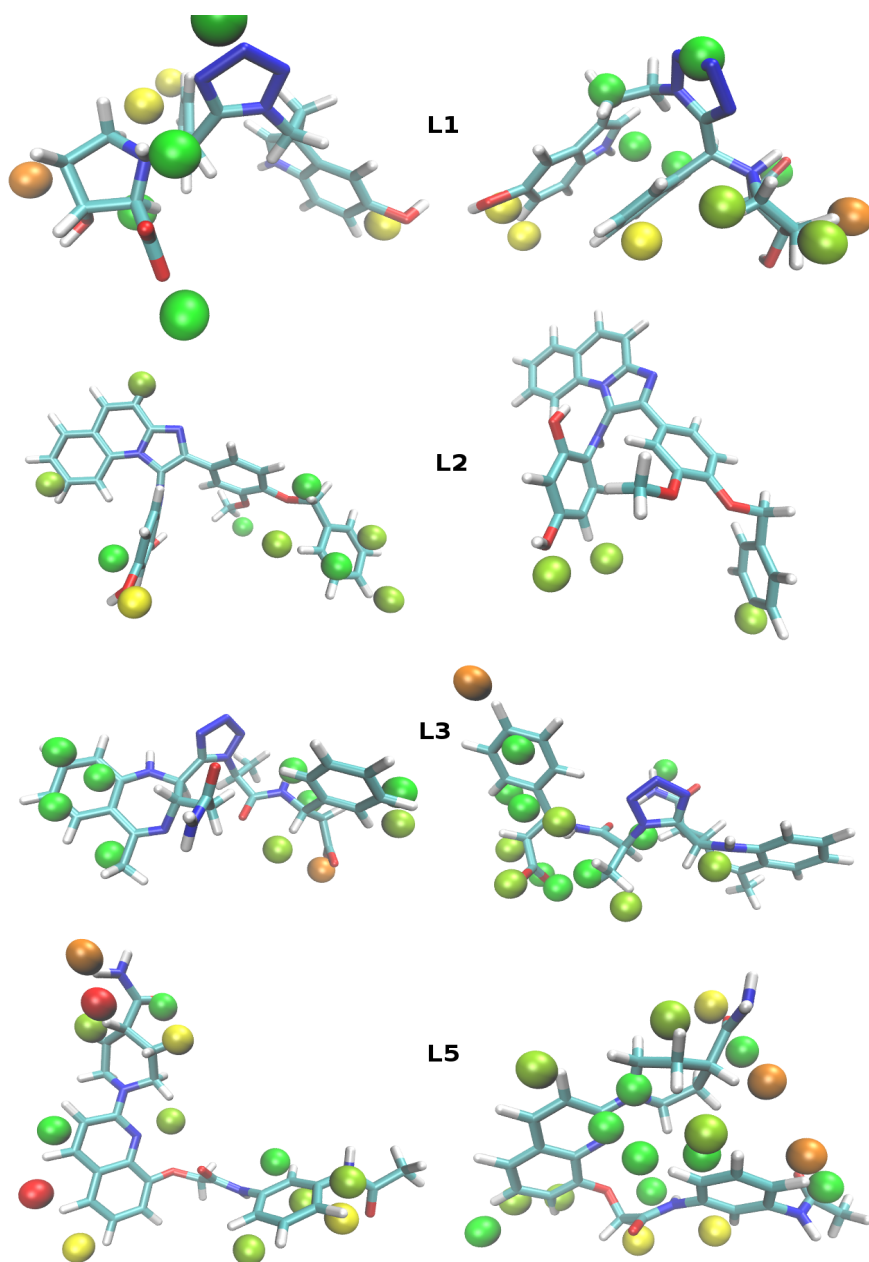
### 2.1.3.3 Procena energije vezivanja liganada

Procena energije vezivanja nepeptidnih liganada obavljena je MM-GBSA analizom, nakon simulacija liganada u kompleksu sa CDK9.

Doking liganada na površinu CDK9 je obavljen bez prisustva molekula vode, dok je u ćelijama površina proteina hidratizana. Prilikom svog vezivanja, ligandi moraju da istisnu molekule vode sa površine proteina. Istiskivanje molekula vode sa površine proteina je povoljno u hidrofobnim oblastima, ali može biti i izrazito nepovoljno, npr. ako su molekuli vode uključeni u mreže vodoničnih veza sa proteinom. Verovatnoća uspešnog vezivanja liganda za proteinsku površinu se povećava ako je izmena prisutnih molekula vode atomima liganda energetski ili entropijski povoljna. Verovatnoća izmene molekula vode atomima liganada na vezivnoj površini CDK9 procenjena je WaterDock analizom, pre same MM-GBSA analize energija vezivanja.

WaterDock analiza sprovedena na svakom od odabranih kompleksa CDK9 / ligand omogućila je da se utvrdi da li strukture liganada pri vezivanju zauzimaju prostor u kome je pogodno zadržavanje nekog od strukturno očuvanih molekula vode u datoj konformaciji proteina. Strukturno očuvani molekuli vode su čvrsto vezani za površinu proteina i postoji mala verovatnoća njihove izmene atomima liganda, tj. mala verovatnoća da će ligand moći da ostvari ispitivanu orijentaciju na površini proteina pri vezivanju za protein u rastvoru.

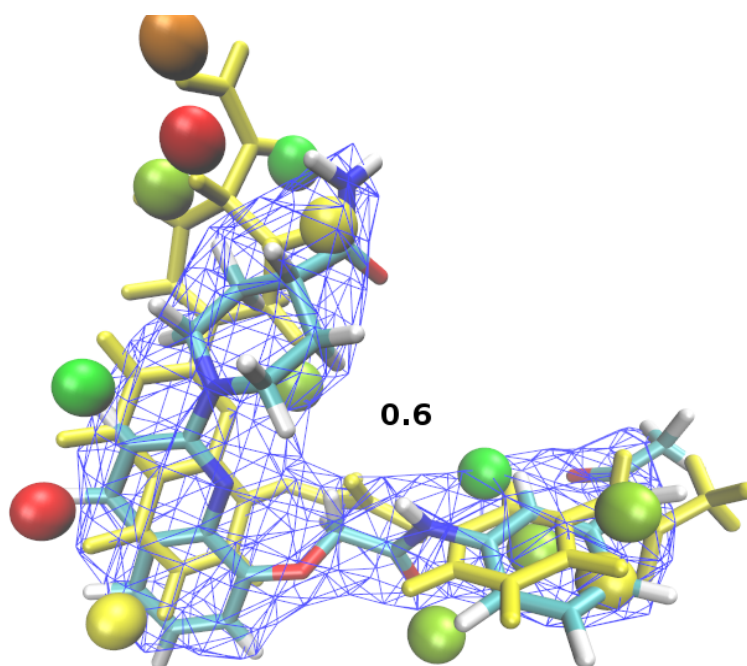
Za sve strukture liganada, osim za L5-B-1-6 strukturu, vezivanje je bilo povoljno po WaterDock analizi (Slika 42).



**Slika 42:** Položaji molekula vode (sfere, obojene po verovatnoći izmene atomom liganada, od crvene  $\leq 20\%$ , do zelene  $\geq 80\%$  verovatnoće) oko položaja liganada u ispitivanim kompleksima CDK9/ligand

L5-B-1-6 struktura liganada L5 u dokovanoj konformaciji, ostvaruje dva nepovoljna kontakta sa molekulima vode, što bi ukazalo na nepovoljan način vezivanja. Međutim, tokom simulacije molekularnom dinamikom koja je prethodila MM-GBSA analizi, ova struktura donekle menja svoju orijentaciju, oslobađajući položaje pogodne za vezivanje molekula vode, i neutrališući nepovoljne kontakte (Slika 43). Nova konformacija je i dominantna tokom simulacije, na šta ukazuje prostorni raspored atoma liganada.





**Slika 43:** Promena orijentacije liganda L5 tokom simulacije molekularnom dinamikom. Početna konformacija liganda (žuto) ostvaruje nepovoljne kontakte sa dva očuvana položaja vode (crvene sfere), dok u simulaciji ligand zauzima drugu konformaciju (u bojama atoma) u kojoj su kontakti ublaženi. Značajnost konformacione promene potvrđuje prostor koji atomi liganda zauzimaju u 60% simulacije (plava mreža).

Struktura L5-B-1-6, po promeni svoje orijentacije u simulaciji je pokazala i najpovoljniju energiju vezivanja među svih osam analiziranih kompleksa (Tabela 11). Pre simulacija, ova struktura je bila rangirana četvrta po energiji vezivanja, što potvrđuje značaj uvrštavanja konformacione fleksibilnosti liganda i receptora u proceduru procene energije vezivanja.

**Tabela 11:** Energije vezivanja liganada dobijene MM-GBSA metodom

oznake	strukture <sup>a</sup>	MM-GBSA energija vezivanja [kcal·mol <sup>-1</sup> ]
L5_1	L5 – B – 1 – 6	- 36.35 ± 3.93
L5_2	L5 – B – 2 – 1	- 35.65 ± 3.55
L1_1	L1 – B – 2 – 1	- 33.28 ± 4.55
L2_1	L2 – B – 2 – 1	- 32.60 ± 3.83
L2_2	L2 – A – 1 – 4	- 28.49 ± 2.94
L3_2	L3 – N2 – 1 – 4	- 26.42 ± 4.34
L1_2	L1 – B – 1 – 1	- 26.05 ± 3.95
L3_1	L3 – B – 2 – 7	- 21.70 ± 3.12

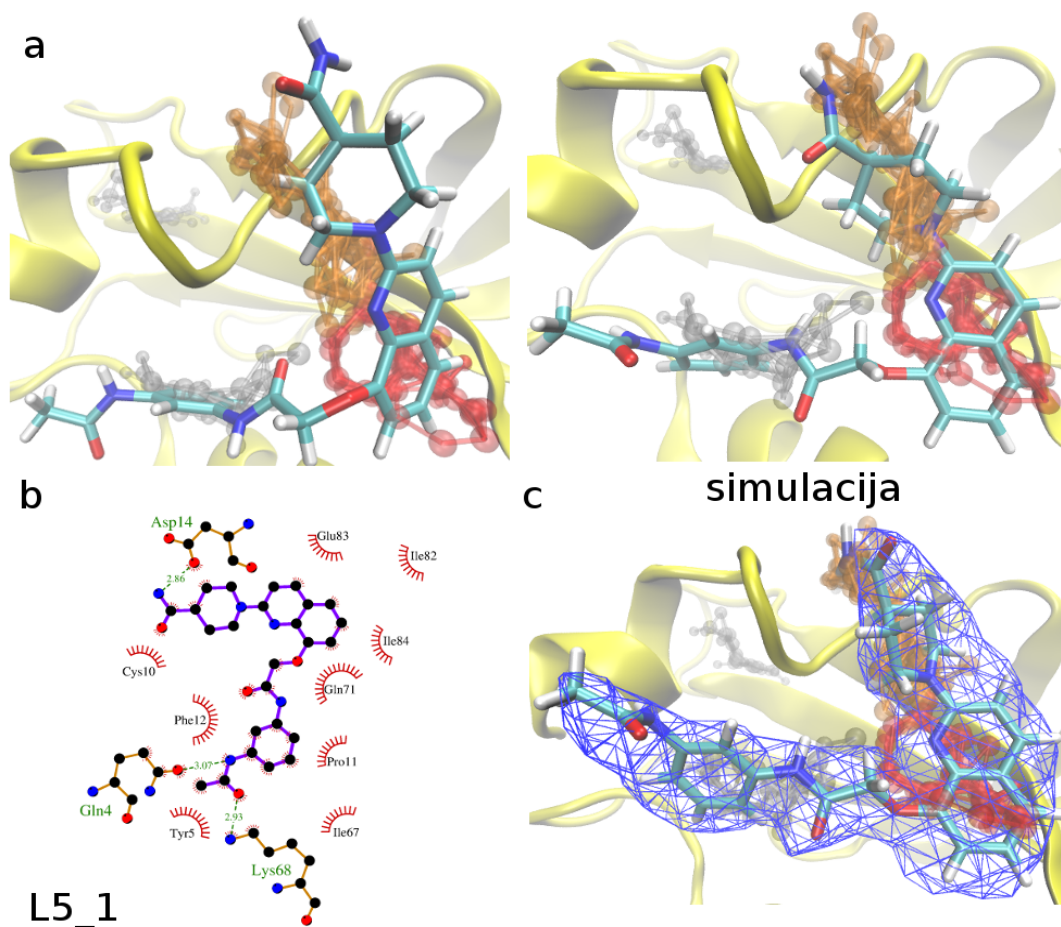
<sup>a</sup> ligand – simulacija – grupisanje – redni broj strukture

Dobijeni rezultati (tabela 11) ukazuju da najbolje vezivanje ostvaruje ligand L5, zatim ligandi L1 i L2, za oko 3 kcal·mol<sup>-1</sup> slabije, dok je vezivanje liganda L3 najnepovoljnije. Prikazani rezultati ne uključuju promenu konformacione entropije pri vezivanju liganada. Za sve ligande su dobijene uporedive vrednosti entropijskog člana energije vezivanja ( -20 kcal·mol<sup>-1</sup> ) ali sa velikim greškama određivanja ( ± 10 kcal·mol<sup>-1</sup> ) usled malog broja struktura (pet) korišćenih u izračunavanju.

MD simulacija koja je prethodila MM-GBSA određivanju energija vezivanja liganada za CDK9 omogućila je konformaciono prilagođavanje struktura liganada i proteina kompleksu, dajući bolji uvid u ostvarene interakcije. Značajne promene vezivne orijentacije dogodile su se sa strukturama liganada L1 i L3, dok su ligandi L2 i L5 pokazali manja variranja u položaju.

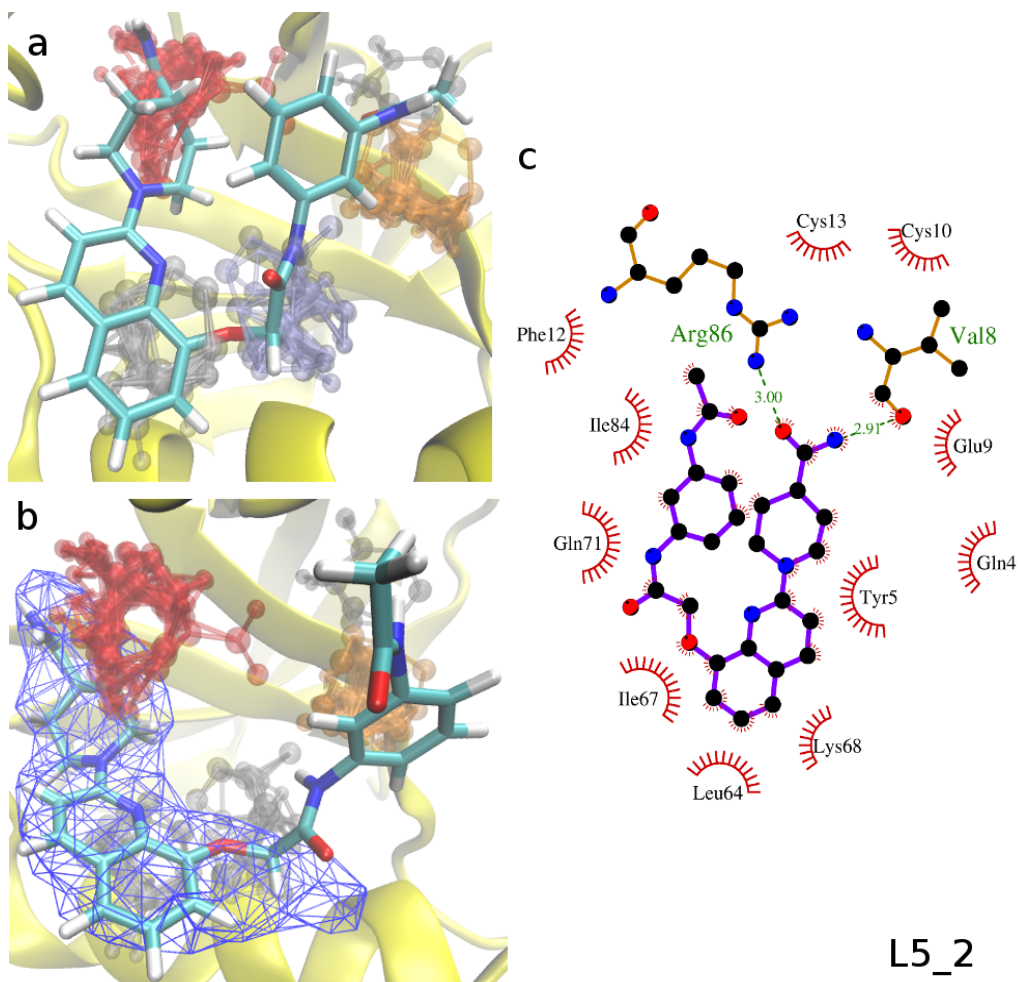
Od značaja je da i pored fluktuacija u položaju tokom simulacije, orijentacije liganada sa najpovoljnijom MM-GBSA energijom vezivanja na vezivnoj površini CDK9 prate oblasti vezivne površine CDK9 koje su predviđene kao pogodne za vezivanje malih molekula primenom FTMap metode (mapiranjem interakcija različitih organskih proba i njihovim grupisanjem), što nije slučaj sa kompleksima liganada sa nepovoljnom energijom vezivanja.

Ligand L5 ostvaruje najpovoljniju energiju vezivanja po MM-GBSA metodi od  $-36.35 \pm 3.93$  kcal·mol<sup>-1</sup>. Oba analizirana kompleksa liganda L5 pokazuju sličnu energiju vezivanja po MM-GBSA metodi. U kompleksu L5-B-1-6, ligand L5 ostvaruje tri vodonične veze, sa aminokiselinama Lys 68, Asp 14 i Gln 4 CDK9, koje nisu sve postojane u simulaciji. Ako se uporedi početna, (dokovana) pozicija liganda L5 na vezivnoj površini CDK9 u kompleksu L5-B-1-6, sa predviđenim oblastima vezivanja malih molekula FTMap metodom, može se videti da se ligand L5 pozicionira u predviđenoj oblasti vezivanja malih molekula, ali ne savršeno. Slaganje dve metode se značajno popravlja tokom MD simulacije koja prethodi MM-GBSA određivanju energije, kada ligand zauzima pozicije tri različite grupacije FTMap proba (Slika 44).



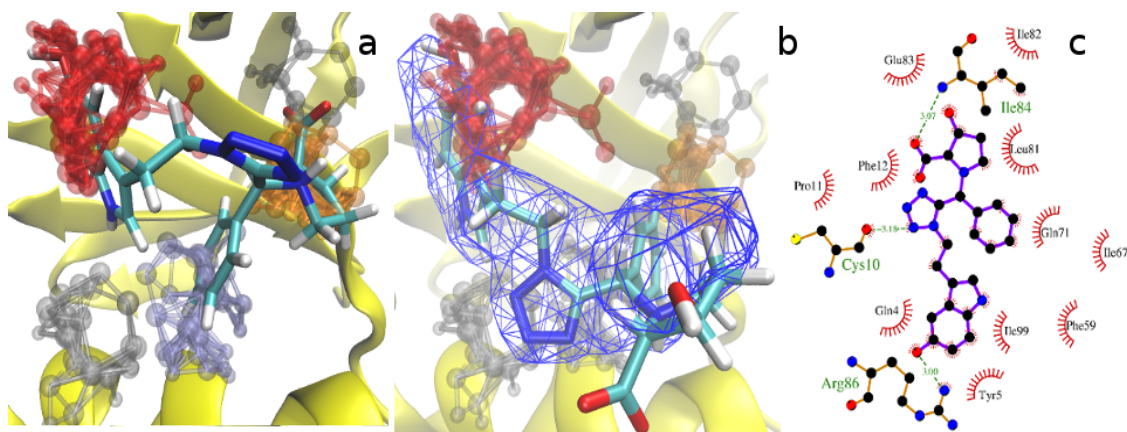
**Slika 44:** a) Slaganje orijentacije liganda L5 u kompleksu L5–B–1–6 sa FTMap probama (crveno *hot spot*, narandžasto 10 – 15 proba, sivo manje od 10), levo početna konformacija liganda, desno, konformacija u toku MD simulacije. b) šematski prikaz interakcija liganda L5 i CDK9 u kompleksu L5\_1, c) položaji atoma liganda L5 tokom MD simulacije sa prostorom koji su atomi zauzimali u bar 60% simulacije (plava mreža)

U svom drugom kompleksu, L5–B–2–1, ligand L5 ostvaruje MM-GBSA energiju vezivanja od  $-35.65 \pm 3.55 \text{ kcal}\cdot\text{mol}^{-1}$ . Najznačajnija interakcija liganda L5 u ovoj orijentaciji je vodonična veza sa Arg 86 CDK9. Ova interakcija je analogna interakcijama koje ostvaruju peptidi 3 i 8 prilikom svog vezivanja i ostaje postojana tokom simulacije. Položaj liganda L5 u kompleksu L5–B–2–1 delom odgovara oblastima vezivanja liganada predviđenim FTMap metodom, ali ne u potpunosti (Slika 45), što je u skladu sa nešto nepovoljnijom energijom vezivanja u odnosu na kompleks L5–B–1–6.



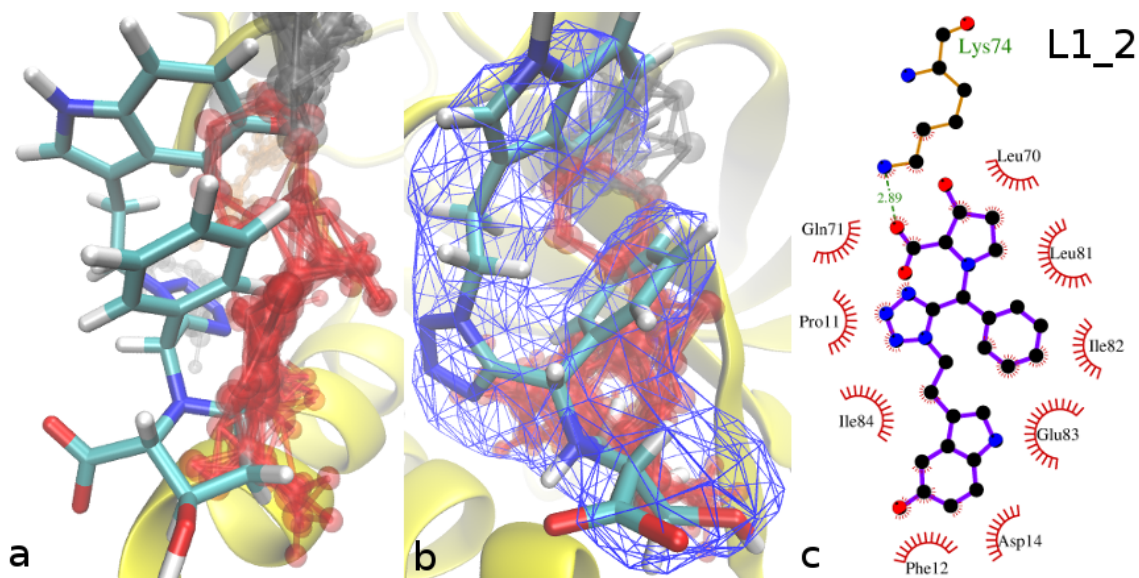
**Slika 45:** a) Slaganje početne konformacije liganda L5 u kompleksu L5–B–2–1 sa FTMap probama (crveno *hot spot*, narandžasto 10 – 15 proba, sivo manje od 10) b) položaji atoma liganda L5 tokom MD simulacije sa prostorom koji su atomi zauzimali u bar 60% simulacije (plava mreža) c) šematski prikaz interakcija liganda L5 u kompleksu L5–B–2–1

Najbolje rangirana orijentacija liganda L1 (L1–B–2–1) ima MM-GBSA energiju vezivanja od  $-33.28 \pm 4.55$  kcal·mol<sup>-1</sup>. U početnoj konformaciji u ovom kompleksu ligand L1 ostvaruje tri vodonične veze sa aminokiselinama Arg 86, Ile 84 i Cys 13 CDK9. Tokom MD simulacije postojana je samo vodonična veza sa aminokiselinom Arg 86, dok sam ligand rotira oko svoje ose i zauzima drugu konformaciju. I pored promene svog položaja tokom simulacije, ligand ostaje u kontaktu sa najznačajnijom FTMap predviđenom oblasti vezivanja tokom simulacije, što je u skladu sa povoljnom energijom vezivanja koju ligand ostvaruje (Slika 46).



**Slika 46:** a) početna i b) konformacija u toku simulacije liganda L1 u kompleksu L1–B–2–1 sa prostorom u kome se ligand zadržava tokom bar 60% simulacije (plava mreža) c) šematski prikaz interakcija liganda L1 sa CDK9 u kompleksu L1–B–2–1

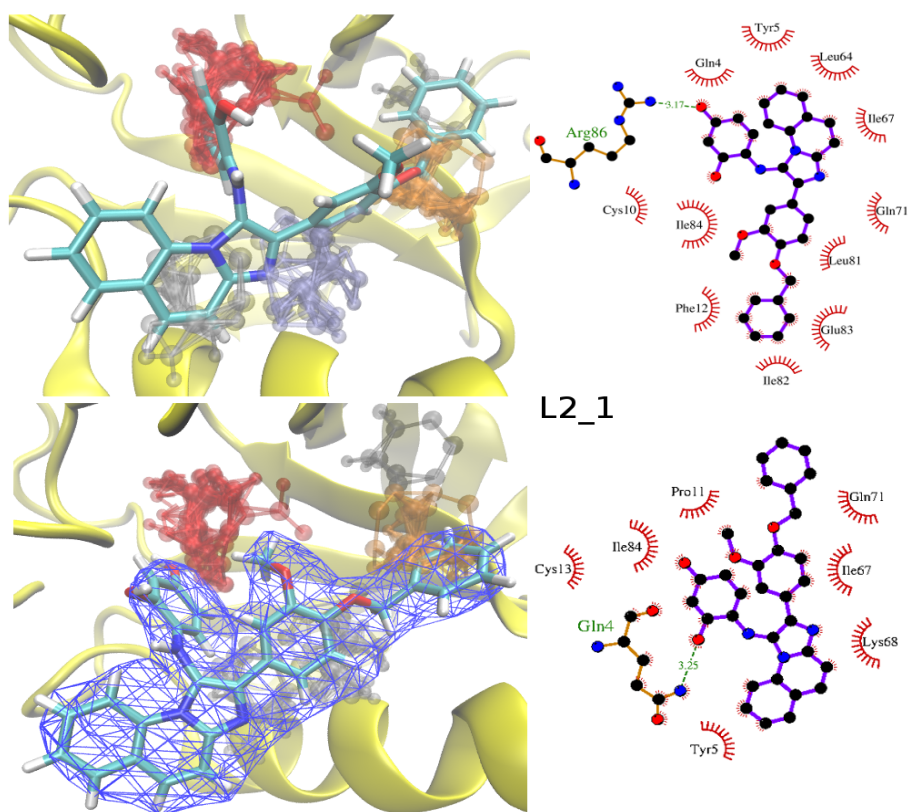
U drugom analiziranom kompleksu liganda L1, L1–B–1–1, ligand L1 ostvaruje za  $7 \text{ kcal}\cdot\text{mol}^{-1}$  slabije vezivanje ( $-26.05 \pm 3.95 \text{ kcal}\cdot\text{mol}^{-1}$ ) i formira samo jednu vodoničnu vezu sa Lys 74. Orijehtacija liganda L1 u kompleksu L1–B–1–1 ostaje stabilna tokom trajanja simulacije, ali iako se nalazi u blizini oblasti vezivne površine CDK9 značajne za vezivanje po FTMap analizi, ligand L1 ne ostvaruje značajno preklapanje sa njom (Slika 47).



**Slika 47:** Podudaranje konformacije tokom simulacije (položaj atoma liganda u bar 60% trajanja simulacije, plava mreža) liganda L1 sa značajnom vezivnom oblašću po FTMap metodi, iz dva ugla a), b). c) šematski prikaz interakcija u L1–B–1–1 kompleksu

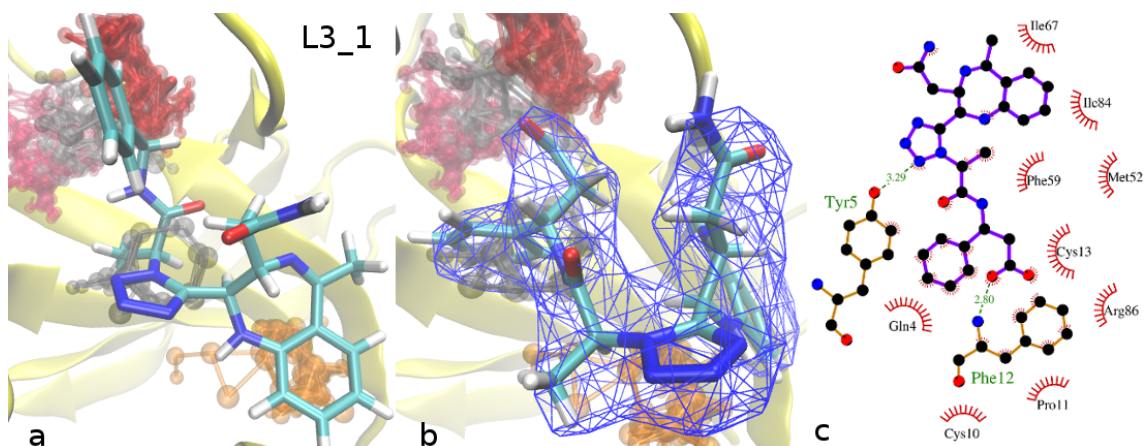
Dve vezivne konformacije liganda L2 najbolje ilustruju korist FTMap predviđanja u rangiranju liganada. Konformacija liganda L2 u kompleksu L2–B–2–1, sa MM-GBSA energijom vezivanja od  $-32.60 \pm 3.83 \text{ kcal}\cdot\text{mol}^{-1}$  se nalazi u oblastima vezivne površine CDK9 za koje FTMap metoda predviđa vezivanje liganada, što nije slučaj sa konformacijom L2–A–1–4, koja ima znatno nepovoljniju energiju vezivanja ( $-28.49 \pm 2.94 \text{ kcal}\cdot\text{mol}^{-1}$ ).

U inicijalnoj proceni energije vezivanja SS-MMGBSA metodom, kompleks L2–B–2–1 liganda L2 je pokazao najpovoljnije vezivanje, ali to nije bio slučaj u detaljnijoj MM-GBSA analizi. Razlog je u izmenjenim interakcijama liganda L2 sa CDK9 tokom simulacije. Povoljna vodonična veza sa Arg 86 koja je prisutna u početnoj konformaciji raskida se tokom simulacije i zamenjena je interakcijama sa N-terminusom CDK9 (Slika 48).



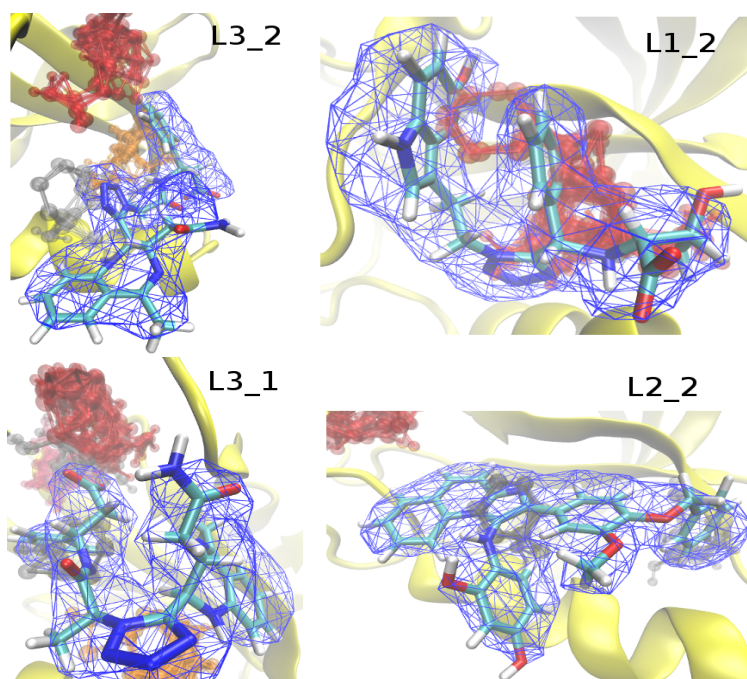
**Slika 48:** Početna (gore levo) i konformacija liganda L2 u toku simulacije (dole levo) sa šematskim prikazom interakcija u kompleksu L2–B–2–1 na početku i u toku simulacije

Od četiri predložene strukture inhibitora, ligand L3 je pokazao najslabije vezivanje za CDK9. U kompleksu L3–B–2–7 ligand L3 najveći broj interakcija ostvaruje sa fleksibilnim N-terminusom CDK9. Ove interakcije i orijentacija liganda L3 nisu stabilne tokom simulacije, već se značajno menjaju (Slika 49).



**Slika 49:** Početna a) i orijentacija tokom simulacije b) liganda L3 u kompleksu L3-B-2-7. c) šematski prikaz interakcija u kompleksu, pretežno sa N-terminusom CDK9

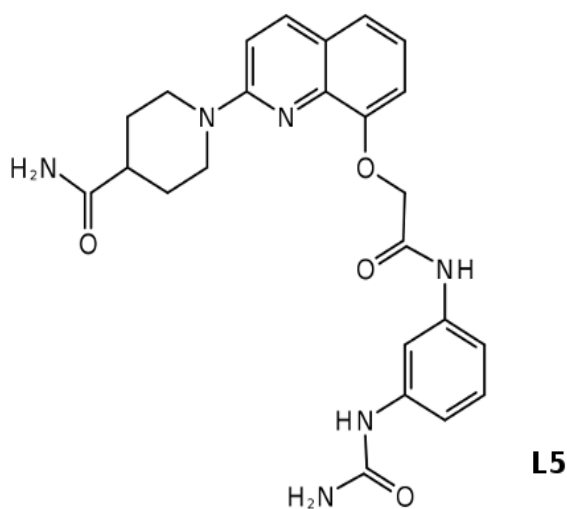
Dobro slaganje rezultata MM-GBSA procene energije vezivanja ligandata i FTMap mapiranja vezivne površine CDK9 ilustruje činjenica da se u tri od četiri kompleksa sa najnepovoljnijom MM-GBSA energijom vezivanja položaji ligandata ne podudaraju sa značajnim FTMap grupacijama proba, dok je u četvrtom slaganje marginalno (Slika 50).



**Slika 50:** Ligandi u četiri kompleksa sa najnepovoljnijom MM-GBSA energijom se prostorno ne poklapaju sa značajnim FTMap grupacijama proba (crveno) tokom MD simulacije. Prostor u kome su ligandi boravili bar 60% simulacije je označen (plavo).

Prikazani rezultati vezivanja četiri potencijalna nepeptidna inhibitora interakcije CDK9 / ciklin T1 za površinu CDK9 ukazuju na nekoliko opštih zaključaka. Od četiri najbolje rangirana kompleksa liganada sa CDK9, tri su ostvarena sa istom konformacijom CDK9. Ohrabrujuće je što je ova konformacija predstavnik najvećeg klastera sličnih konformacija od 2000 konformacija sa najpovoljnijom energijom iz simulacije B, pa postoji verovatnoća nalaženja nje ili njoj sličnih konformacija u rastvoru. Takođe, od četiri najbolje rangirana kompleksa liganada, u tri se ostvaruju vodonične veze sa aminokiselinom Arg 86 iz strukture CDK9. Ostvarene interakcije analogne su onima koje su sa tom aminokiselinom ostvarili peptidi 3 i 8, dva od tri najbolje rangirana peptidna inhibitora. Redosled dobijenih MM-GBSA energija se dobro slaže sa nezavisnom procenom vezivanja za odabrane konformacije FTMap metodom, što ukazuje da se FTMap metoda može koristiti kao dodatan preliminarni filter orijentacija liganada pre opsežnijih studija vezivanja.

Zbog najboljeg profila vezivanja za CDK9, hinolinski derivat L5 je odabran kao struktura potencijalnog inhibitora CDK9 / ciklin T1 interakcije sa najvećim mogućnostima daljeg razvoja (Slika 51).

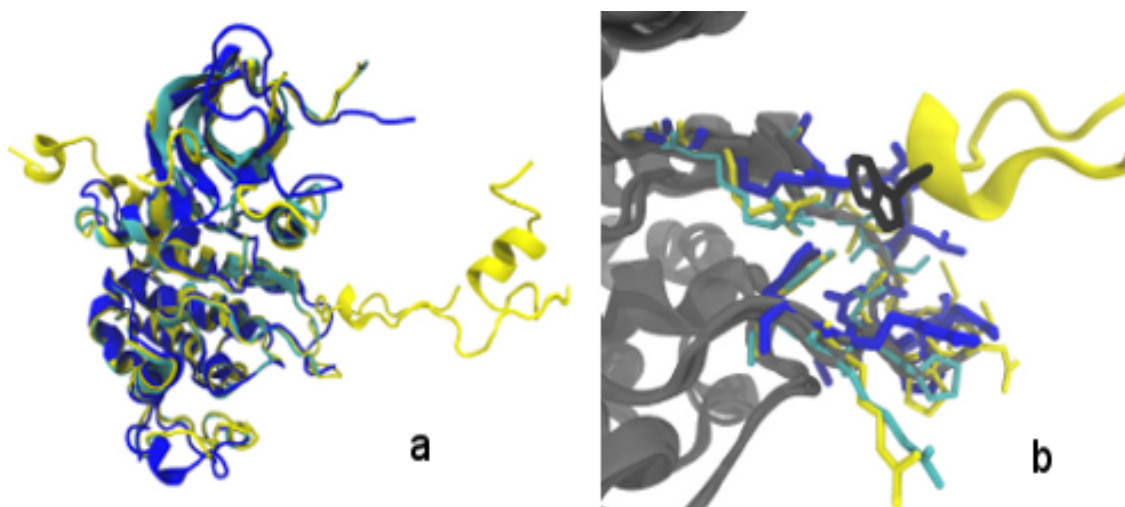


**Slika 51:** Struktura liganda L5



### 2.1.4 Dizajn inhibitora zasnovan na građi P-TEFb/Tat kompleksa

Formiranje kompleksa P-TEFb / Tat protein predstavlja važan molekularni događaj u patologiji HIV-1 infekcije. Vezivanje Tat proteina dovodi do konformacionih promena u strukturi P-TEFb kompleksa, pošto Tat protein vezivanjem za oba proteina u kompleksu povećava molekularnu površinu interakcija (125). Najvećim delom, Tat protein interaguje sa površinom ciklina T1, ali takođe ostvaruje i par kontakata svojim N-terminalnim regionom sa okolinom T-petlje na površini CDK9 (Slika 52).



**Slika 52:** a) Interakcije Tat proteina (žuta petlja), sa površinom CDK9 čije su konformacione promene označene (plavo i zeleno). b) *Hot spot* triptofan 11 (crno) na svom mestu kontakta sa T-petljom CDK9

Kako je cilj ovog istraživanja bilo formulisanje malih molekula koji bi se vezivali za površinu CDK9 i omogućili ometanje interakcija kinaze sa ciklinom T1, proučena je mogućnost razvoja molekula koji bi se vezivali za mesta interakcije Tat proteina sa CDK9 i za samu vezivnu površinu CDK9 uključenu u formiranje kompleksa sa ciklinom T1. Logika ovog pristupa zasnivala se na činjenici da je vezivno mesto Tat proteina svega 10 Å udaljeno od vezivne površine CDK9 u kompleksu sa ciklinom T1. Aminokiseline u sastavu Tat proteina na mestima kontakta sa CDK9 mogle bi da budu iskorišćene za osmišljavanje fragmenata inhibitora koji bi bili kombinovani sa delom inhibitora koji se vezuje za vezivnu površinu CDK9 u kompleksu sa ciklinom T1.

Uspeh ovog pristupa zavisi od pogodnosti mesta kontakta Tat proteina sa površinom CDK9 za razvoj inhibitora. Kada je kristalografska struktura kompleksa P-TEFb/Tat podvrgnuta *hot spot* analizi Robetta metodom, pokazalo se da je jedna od Tat aminokiselina u kontaktu sa CDK9, Trp 11, *hot spot*, sa promenom energije

vezivanja kompleksa od  $3.82 \text{ kcal}\cdot\text{mol}^{-1}$  po njenoj mutaciji u Ala.  $C\alpha - C\alpha$  udaljenost ove aminokiseline od *hot spot* aminokiseline Lys 93 iz strukture ciklina je  $10.19\text{\AA}$ , što je rastojanje koje se može premostiti sa svega tri aminokiseline u građi peptida.

Opisane strukturne karakteristike su obećavajuće i već je u literaturi pokazano (157) da je sprečavanjem interakcija Tat proteina i CDK9 u toj oblasti moguće ometati funkciju P-TEFb/Tat kompleksa, kada su dizajnirani molekuli koji su se vezivali ispod T-petlje i menjali njenu konformaciju, sprečavajući formiranje interakcija sa Tat proteinom. Međutim, dizajn inhibitora u ovom pravcu nije nastavljen u okviru ovog istraživanja, prevashodno zbog činjenice da je mesto interakcije Tat proteina previše udaljeno od glavnog regiona vezivne površine CDK9 na koji se fokusirala ova studija, što čini povezivanje fragmenata inhibitora nepraktičnim. Lys 93, koji jeste dovoljno blizu mestu interakcije sa Tat proteinom za povezivanje, iako predstavlja *hot spot* aminokiselinu, stupa u interakcije sa fleksibilnom petljom CDK9, što otežava dizajn. U analizama pogodnih oblasti vezivne površine CDK9 za interakciju sa malim molekulima, oblast vezivne površine sa kojom interaguje Lys 93 nije bila značajno zastupljena.

Iako opisani pravac dizajna nije upotrebljen u ovoj studiji, činjenica da su energetske značajne aminokiseline u strukturi Tat proteina prisutne u ograničenom regionu kojim ovaj protein interaguje sa CDK9 govori o značaju ostvarenih interakcija i potencijalnoj koristi od njihove inhibicije.

### 2.1.5 Zaključak

Direktni inhibitori CDK9 / ciklin T1 interakcije, koji su predmet dizajna opisanog u ovoj studiji bi, ometanjem formiranja P-TEFb kompleksa interferirali ne samo sa njegovom ulogom u patologiji HIV-1 infekcije, već i u drugim terapijskim stanjima, kao što su maligniteti i hipertrofija miokarda. Cilj ove studije bio je dobijanje potencijalnih struktura takvih inhibitora.

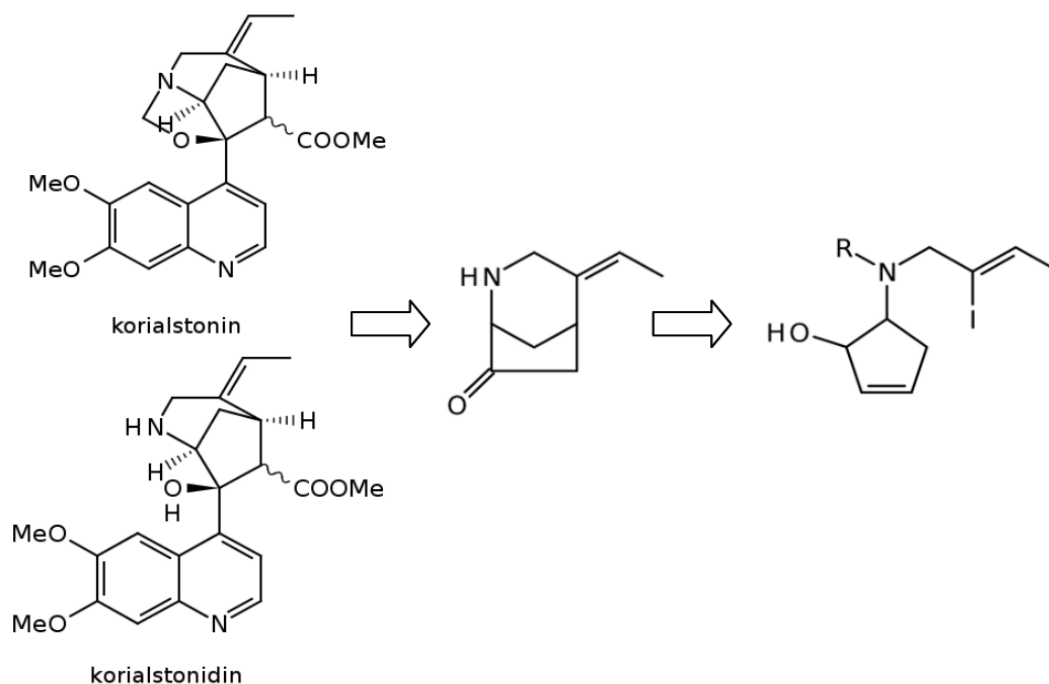
U okviru ove studije analizirana je struktura i dinamika CDK9 / ciklin T1 protein-protein interakcije. Ustanovljen je značaj modelovanja fleksibilnog N-terminusa CDK9 za uspešan dizajn inhibitora interakcije i predložene su strukture peptidnih (LQTLGF) i nepeptidnih (ligand L5, slika 51) potencijalnih inhibitora interakcije.

## 2.2 *In silico* analiza regioselektivnosti intramolekulske Heck-ove ciklizacije

### 2.2.1 DFT studija prelaznih stanja procesa

Primena *in silico* metoda za razumevanje molekularnih osnova procesa nije ograničena na biološke sisteme i molekularnu mehaniku. *Ab initio* metode se široko primenjuju u oblastima organske i neorganske hemije, kao dopuna eksperimentima u razumevanju hemijskih reakcija i struktura. Ovo poglavlje ima za cilj da ilustruje takvu primenu *in silico* metoda na primeru intramolekulske Heck-ove ciklizacije.

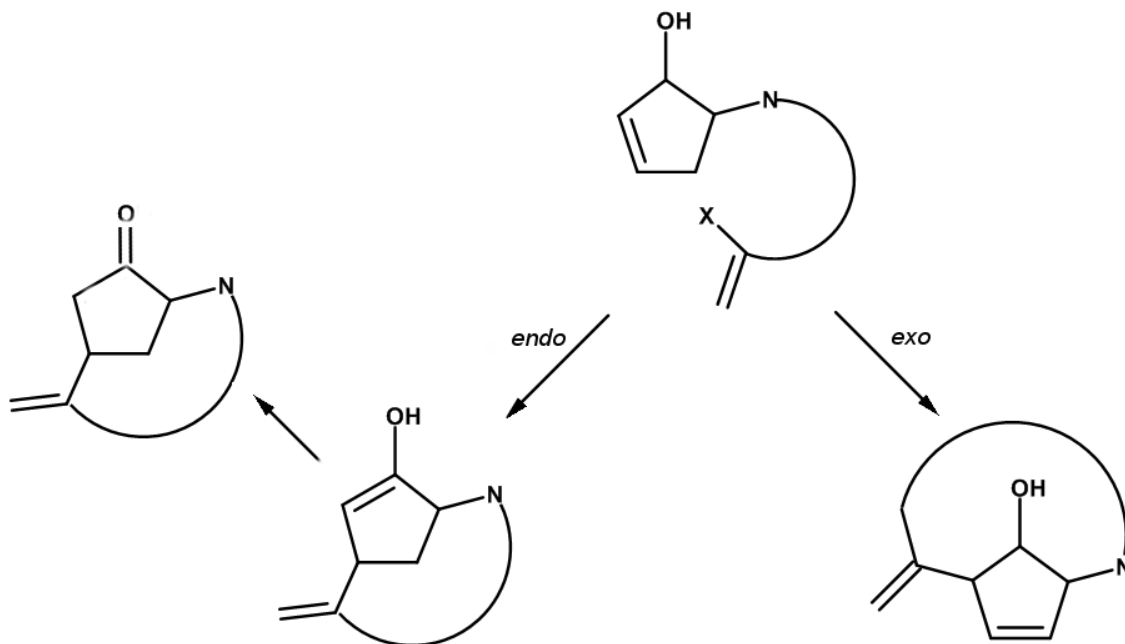
Korialstonin i korialstonidin su alkaloidi izolovani iz biljne vrste *Alstonia coriacea* sa potencijalnim antimalarijskim dejstvom (201,202). Retrosintetski pristup sintezi ova dva alkaloida, proučavan u okviru rada istraživačke grupe na Katedri za organsku hemiju Farmaceutskog fakulteta u Beogradu, ukazao je da se biciklični supstituent na hinolinskom skeletu ovih alkaloida može dobiti intramolekulskom Heck-ovom ciklizacijom alil-alkoholnih derivata ciklopentena (Slika 53) (203).



**Slika 53:** Strukture korialstonina i korialstonidina i retrosintetski pristup njihovoj sintezi

Heck-ova reakcija je paladijumom katalizovani proces u kome se iz alil, vinil, aril halogenida i analognih jedinjenja, u reakciji sa alkenima, dobijaju supstituisani alkeni definisane stereochemije. Heck-ova reakcija se široko koristi u organskoj sintezi i

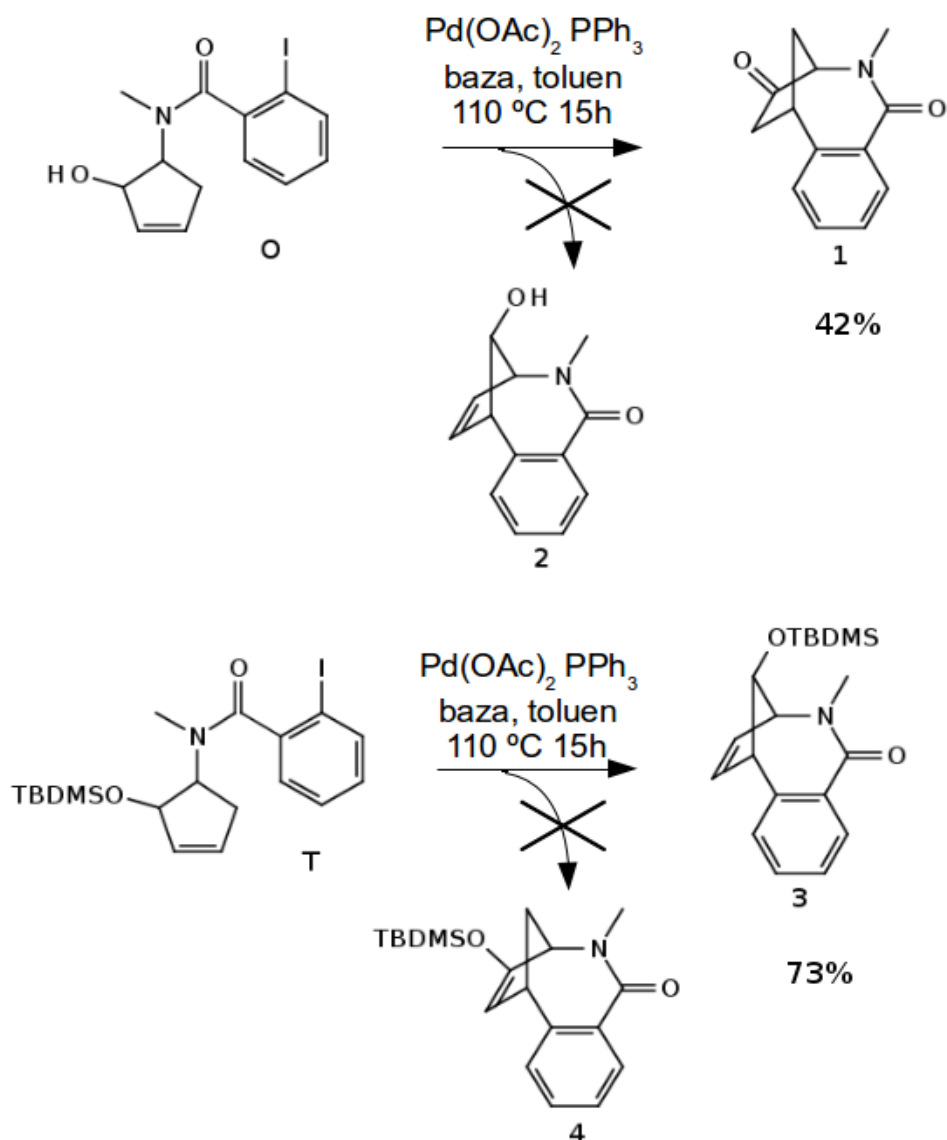
predstavlja odličan način kontrolisanog formiranja novih C–C veza u složenim molekulima sa strogim stereohemijskim zahtevima, kakvi su prirodni proizvodi. U Heck-ovoj ciklizaciji alil-alkoholnih derivata ciklopentena, proučavanoj u ovom istraživanju, moguće je dobijanje dva regioizomera, proizvoda *endo*- i *exo*- ciklizacije, u odnosu na alilnu alkoholnu grupu (Slika 54).



**Slika 54:** Proizvodi *endo*- i *exo*- ciklizacije

Kako samo *endo*- ciklizacija vodi intermedijeru u sintezi korialstonina i korialstonidina, značajno je razumevanje regioselektivnosti procesa ciklizacije.

Model sistem (jedinjenja **O**), upotrebljen je za eksperimentalno proučavanje procesa ciklizacije. Zapaženo je da regioselektivnost procesa zavisi od alkoholne grupe alil-alkoholnog fragmenta (Slika 55). Proces ciklizacije izveden je pod standardnim uslovima Heck-ove reakcije pri čemu je alkohol **O** dao proizvod **1** u prinosu od 42%. Pažljivom analizom <sup>1</sup>H NMR spektra reakcione smeše nije primećeno postojanje proizvoda **2**.

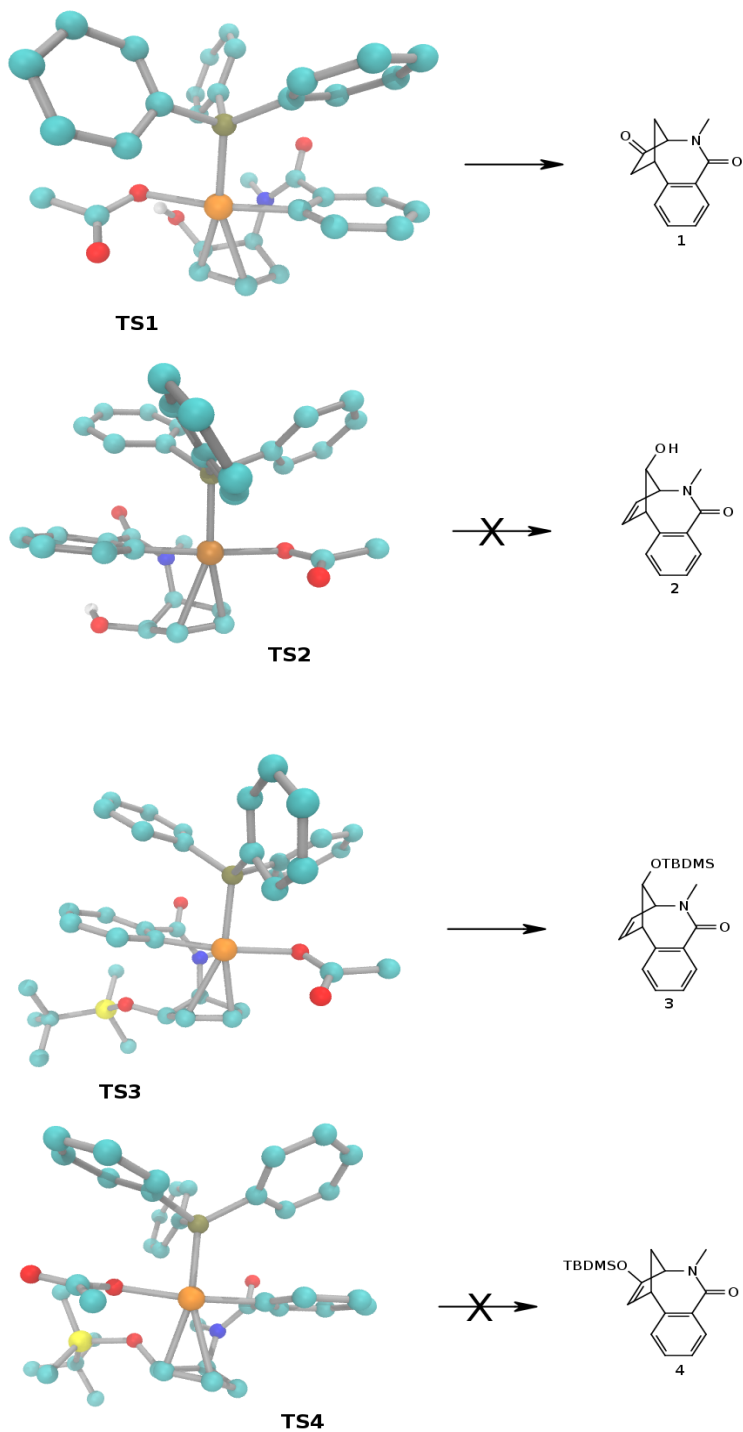


**Slika 55:** Strukture polaznih jedinjenja (O, T) i mogućih proizvoda (1 – 4) Heck-ove intramolekulske ciklizacije.

Zaštitom slobodne alkoholne grupe jedinjenja **O**, prevođenjem u *t*-butildimetilsilil etar, jedinjenje **T**, pod istim eksperimentalnim uslovima dalo je proizvod **3** u prinosu od 73% (slika 55). I u ovom slučaju nije primećeno formiranje alternativnog proizvoda **4**.

Nastanak jedinjenja **1** i **3** može se objasniti *endo* i *exo* ciklizacijom kojoj podleže polazni alilni alkohol. U zavisnosti od toga da li u migracionoj inserciji alkena u formiranju nove C–C veze učestvuje C(2) ili C(3) atom alilnog alkohola, dolazi do *endo* ili *exo* ciklizacije, pri čemu je jedinjenje **1** proizvod *endo*, a jedinjenje **3** proizvod *exo* ciklizacije.

Da bi se objasnila uočena regioselektivnost, proučavana su prelazna stanja koja vode svim teoretski mogućim proizvodima (Slika 56)



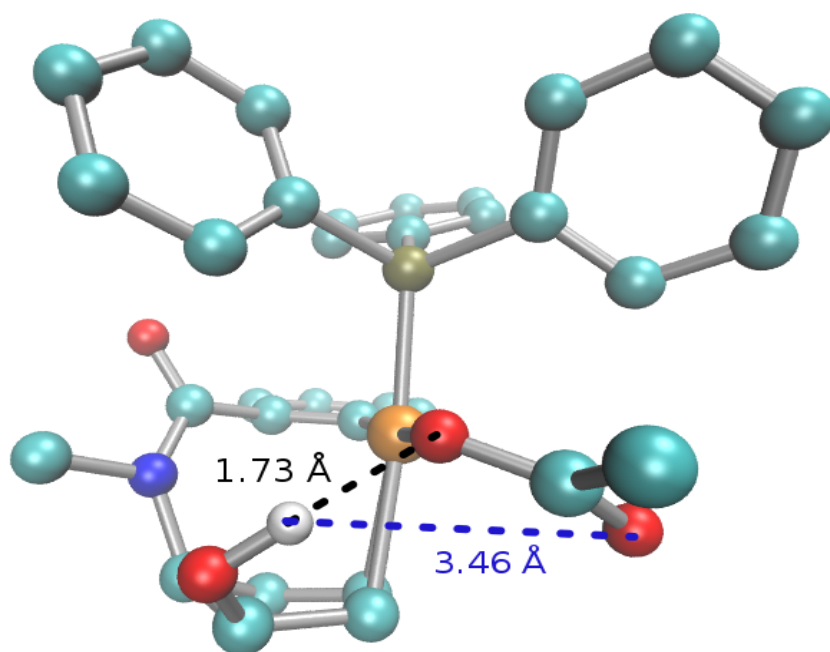
**Slika 56:** Prelazna stanja (TS1 – TS4) koja vode nastanku proizvoda 1 – 4

Uticaj stabilnosti prelaznih stanja na ishod posmatranog procesa proveren je lociranjem prelaznih stanja i *ab initio* izračunavanjem razlika u njihovim energijama.

Iako se oksidativna adicija Pd(0) dešava njegovom insercijom u C–I vezu, jodid, pod eksperimentalnim uslovima koji su korišćeni u procesima koje prati ovo istraživanje (slika 55), ne ulazi u sastav samog prelaznog stanja, već ga u njemu zamenjuje acetatni jon (204,205). Kao rezultat, prelazna stanja ovog procesa su modelovana sa acetatnim ligandom na paladijumu, a ne jodidom, tokom svih izračunavanja. Prelazna stanja su locirana DFT metodom, sa B3LYP hibridnim funkcionalom i def2-SVP i def2-TZVP setovima orbitala.

Po lociranju prelaznih stanja, dobijene strukture su verifikovane frekvencionom analizom i za svaku strukturu je dobijena samo jedna negativna, imaginarna frekvencija. Dobijene vrednosti razlika u energijama prelaznih stanja (13.6 i 3.4 kcal·mol<sup>-1</sup>) bile su u skladu sa eksperimentalnim rezultatima i ukazale su da se opažena regioselektivnost procesa može pripisati, bar delom, različitoj stabilnosti prelaznih stanja.

Kod prelaznih stanja TS1 i TS2 u kojima učestvuje alilni alkohol **O** sa slobodnom alkoholnom grupom, prelazno stanje TS1 koje vodi *endo* ciklizaciji i proizvodu **1**, pokazalo se za 13.6 kcal·mol<sup>-1</sup> stabilnije od prelaznog stanja TS2, koje vodi proizvodu **2** koji nije eksperimentalno dobijen. Povećana stabilnost prelaznog stanja TS1 se može objasniti mogućnošću formiranja vodonične veze između acetatnog liganda na paladijumu i slobodne alkoholne grupe alilnog alkohola (Slika 57), koja ne postoji u prelaznom stanju TS2.



**Slika 57:** Detalji prelaznog stanja TS1 u okolini alkoholne grupe



Kod prelaznih stanja TS3 i TS4 u kojima učestvuje jedinjenje **T** sa zaštićenom alkoholnom grupom, prelazno stanje TS3, koje vodi proizvodu *exo* ciklizacije **3**, stabilnije je za 3.4 kcal·mol<sup>-1</sup> od prelaznog stanja TS4, što je takođe u skladu sa eksperimentalnim rezultatima, jer proizvod **4**, koji bi nastao *endo* ciklizacijom kroz prelazno stanje TS4, nije eksperimentalno dobijen. Razlog uočene regioselektivnosti Heck-ovog procesa na jedinjenju **T** nije jasno vidljiv na strukturama prelaznih stanja TS3 i TS4. Ne postoje značajni sterni efekti koji bi koordinisali proces. Moguće je da je objašnjenje u povoljnim  $\pi$ - $\pi$  interakcijama fenilnih grupa u TS3 koje nisu moguće sa acetatnim ligandom u istoj poziciji, ili u različitoj hibridizaciji ugljenikovih atoma u susedstvu *t*-butildimetilsilil- grupe. Sp<sup>2</sup> hibridizovani atomi fenilne grupe u susedstvu *t*-butildimetilsilil- grupe u prelaznom stanju TS3 predstavljaju manju sternu smetnju procesu od sp<sup>3</sup> hibridizovane metil grupe acetata u TS4.

### 2.2.2 Zaključak

Uzeti zajedno, prikazani rezultati su ukazali na moguć mehanizam kojim slobodna alkoholna grupa u neposrednoj blizini koordinacionog centra paladijumovog kompleksa, utiče na eksperimentalno opaženu regioselektivnost Heck-ove intramolekularne ciklizacije na jedinjenjima **O** i **T**, kroz interakcije sa acetatnim ligandom sa paladijuma. Prevođenje ove alkoholne grupe u *t*-butildimetilsilil etar onemogućava da se ona ponaša kao donor vodonične veze i tako stabilizuje prelazno stanje koje vodi proizvodima *endo* ciklizacije, dovodeći do dobijanja *exo* regioizomera.

### 3. Zaključak

U ovom istraživanju obavljena je *in silico* studija P-TEFb kompleksa, izgrađenog od CDK9 i ciklina T1, u cilju razvoja direktnih inhibitora ove proteinske interakcije od medicinskog značaja. Proučena je struktura i hidratisanost vezivnih površina uključenih u formiranje P-TEFb kompleksa i utvrđene su energetske najznačajnije *hot spot* aminokiseline u strukturi kompleksa, među kojima su Phe 146 i Lys 93 iz strukture ciklina T1.

U nedostatku eksperimentalno utvrđene strukture slobodne CDK9 pogodne za dizajn inhibitora, konformacione promene njene vezivne površine ispitane su detaljnim simulacijama slobodnog proteina klasičnom i ubrzanom molekularnom dinamikom. Utvrđen je značaj modelovanja nedostajućih aminokiselina N-terminusa CDK9 za pravilan prikaz konformacionih promena njene vezivne površine. U okviru studija dinamike vezivne površine CDK9 ispitano je prisustvo, otvaranje i postojanost privremenih hidrofobnih džepova pogodnih za vezivanje liganada. Izolovane su dve oblasti pogodne za interakcije sa ligandima, koje odgovaraju mestima vezivanja aminokiselina Phe 146 i Lys 6 iz strukture ciklina T1.

Dobijene informacije o vezivnoj površini CDK9 iskorišćene su za dizajn peptidnih inhibitora interakcije. Peptidi su dizajnirani po ugledu na strukturu ciklina T1, ali i *de novo* dizajnom. Energija vezivanja peptida procenjena je MM-PBSA i SMD metodama. Tri peptidne strukture (aminokiselinske sekvence LQTLGF, ESILQ, PRWPE) sa najvećim predviđenim afinitetom za CDK9 poslužile su za dalji razvoj nepeptidnih inhibitora interakcije. Na osnovu informacija o načinu vezivanja peptida i osobina vezivne površine CDK9, formiran je farmakoforni model od strukturnih elemenata ciklin T1 aminokiselina Phe 146 i Lys 6 i upotrebljen za dizajn nepeptidnih inhibitora. Predložene su strukture nepeptidnih potencijalnih inhibitora interakcije i diskutovane njihove interakcije sa vezivnom površinom CDK9. Postignuto je dobro slaganje rezultata primenjenih nezavisnih *in silico* metoda. Hinolinski derivat L5 (struktura, slika 51) je odabran kao ligand sa najvećim afinitetom za vezivanje na površinu CDK9.

Ovo istraživanje ilustruje prednosti korišćenja većeg broja komplementarnih *in silico* tehnika radi poboljšanja kvaliteta dizajna inhibitora protein-protein inhibitora.

Predložene strukture potencijalnih inhibitora CDK9 / ciklin T1 interakcije mogu da posluže za dalje studije ovog proteinskog kompleksa. Sam postupak dizajna inhibitora, primenjen u ovom istraživanju, opšteg je karaktera i može se koristiti u istraživanju inhibicije drugih protein-protein interakcija od značaja u medicini.

U okviru ovog istraživanja, primenom *ab initio* metoda ispitana su prelazna stanja intramolekulske Heck-ove ciklizacije alilnih alkohola i ponuđeno je objašnjenje eksperimentalno zapažene regioselektivnosti procesa. Mogućnost formiranja vodonične veze između alilne alkoholne grupe i acetatnog liganda na paladijumu u *endo*-prelaznom stanju proučavane ciklizacije dovela je do značajne stabilizacije tog prelaznog stanja ( $13.6 \text{ kcal}\cdot\text{mol}^{-1}$ ) u odnosu na analogno *exo*-prelazno stanje (*endo* i *exo* u odnosu na alkoholnu grupu) i uticala na regioselektivnost procesa.

## 4. Eksperimentalni deo

### 4.1 Molekularna dinamika

#### 4.1.1 Priprema sistema

##### *CDK9 / ciklin T1*

Kristalografska struktura P-TEFb kompleksa, PDB ID: 3blh preuzeta je iz RSCB PDB databaze. Struktura je vizualizovana koristeći VMD softver (eng. *visual molecular dynamics*, VMD) (206) i iz nje su uklonjeni molekuli rastvarača.

Uticaj kristalnih kontakata na konformaciju proteina u kompleksu procenjen je upotrebom CryCo (207) web servisa.

U kristalografskoj strukturi 3blh, četiri petlje iz CDK9 imaju nedefinisiranu građu, jer, zbog fleksibilnosti, njihov položaj nije jasno prikazan elektronskom gustinom. Tri nedostajuće petlje su modelovane na osnovu drugih kristalografskih struktura P-TEFb kompleksa. Strukture CDK9 i ciklina T1 iz kristalografske strukture 3blh orijentisane su prema ostalim kristalografskim strukturama P-TEFb kompleksa tako da se minimizuje kvadratna sredina odstupanja (eng. *root mean square deviation*, RMSD) položaja Ca atoma proteina, a zatim su koordinate atoma nedostajućih petlji preuzete iz ostalih struktura. Konformacija petlje sa aminokiselinama 51–54 je preuzeta iz kristalografske strukture PDB ID: 3blq (83), a konformacije T-petlje (aminokiseline 177–181) i petlje sa aminokiselinama 260–266 iz kristalografske strukture P-TEFb kompleksa i Tat proteina, PDB ID: 3mi9 (125). Konformacija četvrte nedostajuće petlje, sa aminokiselinama 88–97 iz strukture CDK9 je modelirana na osnovu poznate aminokiselinske sekvence, jer koordinate njenih atoma nisu određene ni u jednoj do sada rešenoj kristalografskoj strukturi P-TEFb kompleksa. Za modeliranje petlje korišćen je MODELLER softver (208), kroz web servis ModLoop (209).

Modelovani segmenti strukture proteina su minimizovani pre simulacije, tako što je struktura proteina fiksirana, a petlje podvrgnute minimizaciji u trajanju od 20000 koraka, u vakuumu, koristeći NAMD 2.8 (210) softver. Naelektrisanja i interakcije

između atoma u simulaciji opisani su CHARMM 22/CMAP poljem sila (185,186). Strukture za simulaciju su pripremljene u VMD.

Za protonovanje proteina u kompleksu na fiziološkom pH (7.4) upotrebljena je empirijska metoda PROPKA 3.0 web servisa (211).

Priprema za simulaciju je nastavljena solvatacijom kompleksa CDK9 / ciklin T1. TIP3P model vode je upotrebljen za okruženje kompleksa. Kompleks je uronjen u četvrtastu kutiju rastvarača, pri čemu su molekuli vode ispunjavali prostor do 15 Å od bilo kog atoma kompleksa. Ukupno naelektrisanje kompleksa je neutralisano hloridnim jonima. Kompleks je minimizovan još 30000 koraka, zatim zagrejan na 310 K u toku 93 ps, sa inkrementima temperature od 0.1K i podvrgnut molekularnoj dinamici u trajanju od 4ns.

### *Slobodna CDK9*

Priprema slobodne CDK9 za simulaciju obavljena je na isti način kao i priprema kompleksa. Struktura CDK9 izdvojena je iz konformacije kompleksa nakon 4ns molekularne dinamike. Struktura je solvatisana, neutralisana, minimizovana (30000 koraka) i zagrejana na 310K, da bi zatim bila simulirana 4ns molekularnom dinamikom.

Nedostajući N-terminus CDK9 (7 aminokiselina, sekvenca MAKQYDS) je ručno modelovan u Pymol (212) softveru. Nakon modelovanja, N-terminus je minimizovan u vakuumu, pri čemu su samo prvih 15 aminokiselina N-terminusa bile minimizovane, dok je ostatak strukture bio fiksiran. Minimizacija je trajala 10000 koraka, u NAMD softveru. Nakon minimizacije N-terminusa, protein je solvatisan i neutralisan po protokolu objašnjenom ranije i molekuli rastvarača i joni su minimizovani kroz 50000 koraka uz fiksiranu strukturu proteina. Nakon minimizacije rastvarača, ceo sistem je minimizovan kroz dodatnih 50000 koraka, a zatim zagrejan na 310 K (kroz 93ps, uz inkremente temperature od 0.1 K) i podvrgnut molekularnoj dinamici u trajanju 5ns.

Pošto su protokol za izračunavanje energije vezivanja i parametrizacija nepeptidnih liganada zahtevali promenu polja sila interakcija atoma, slobodna CDK9 je pripremana za simulaciju i koristeći Kornel et al. polje sila verzija ff10 (ff99SB) (184) , sa Jang i Čitham korekcijama za jone u TIP3P modelu vode (frcmod.ionsjc\_tip3p)

(213). Priprema strukture za simulaciju je obavljena koristeći LeaP program implementiran u AmberTools 1.5 softveru (214). Solvatacija, minimizacija i zagrevanje su obavljeni na potpuno isti način kao i sa CHARMM 22/CMAP poljem sila interakcija.

### *CDK8*

Za simulaciju je upotrebljena kristalografska struktura CDK8 u kompleksu sa ciklinom C, PDB ID: 3rgf (159). Struktura CDK8 je izdvojena iz kompleksa. Četiri nedostajuće petlje (aminokiseline 28–32, 116–122, 179–194, 239–244) modelovane su MODELLER softverom kroz ModLoop web servis. Petlje su minimizovane u vakuumu kroz 5000 koraka, koristeći NAMD softver i CHARMM22/CMAP polje sila interakcija.

Protonacija aminokiselina u strukturi određena je PROPKA 3.1 web serverom. Orijentacije i protonacija histidina i asparagina proverene su Reduce (215) softverom.

Simulirane su tri varijante sistema: CDK8 sa eksperimentalnim N-terminusom, CDK8 sa modelovanim N-terminusom (ručno je modelovano prvih 10 aminokiselina) i CDK8 sa skraćenim N-terminusom (uklonjeno je prvih 10 aminokiselina iz sekvence CDK8). Sve tri varijante sistema su solvatisane i neutralisane u LeaP modulu AmberTools 1.5 programa. Pripremljene varijante CDK8 sistema su minimizovane i zagrejane analogno slobodnoj CDK9, pod identičnim uslovima simulacije.

## *Parametrizacija liganada*

Peptidni ligandi su tretirani analogno proteinima, koristeći ista polja sila interakcija atoma (ff99SB).

Nepeptidni ligandi su parametrizovani koristeći Antechamber (216) softver implementiran u Amber Tools 1.5 i acpype (217). Početne strukture nepeptidnih liganada iscertane su koristeći <http://cactus.nci.nih.gov/translate/> cactvs web servis. Strukture liganada, po tri konformacije za svaki, su prvo optimizovane na semiempirijskom nivou (PM3) koristeći MOPAC 2012 softver (218). U sledećem koraku strukture su optimizovane *ab initio* metodama, na HF/6-31G\* nivou (Hartri-Fok optimizacija, sa 6-31G\* setom orbitala), koristeći ORCA softver (219). Optimizovane strukture su pripremljene za derivatizaciju naelektrisanja R.E.D. tools softverom (220) a samo parcijalno naelektrisanje atoma je dobijeno preko R.E.D. web servera (221–223), pri čemu je parcijalno naelektrisanje atoma ujednačeno između 6 orijentacija svake od 3 konformacije molekula i izvedeno iz *ab initio* optimizacija. Ligandima su dodeljeni parametri GAFF (eng. *general Amber force field*) (224) opšteg polja sila interakcija, koje je kompatibilno sa Kornel et al. (ff99SB) poljem sila interakcija atoma za proteine i može se primenjivati zajedno sa njim. Parcijalna naelektrisanja i GAFF parametri su ligandima dodeljeni primenom acpype softvera.

#### 4.1.2 MD simulacije

##### *cMD*

Sve simulacije klasičnom molekularnom dinamikom (cMD) obavljene su u NAMD softveru. U zavisnosti od sistema primenjivana su dva polja sila, CHARMM 22/CMAP i Kornel et al. (ff10 i ff12). Ekvilibracija sistema u simulacijama procenjivana je ustaljivanjem RMSD vrednosti C $\alpha$  atoma. Po završetku simulacija kvalitet proteinskih struktura proveren je WHATIF Protein Model Check web servisom (225).

Opšti parametri simulacija su bili isti za sve sisteme. Simulacije su, ako nije drugačije naznačeno, sprovedene pod NPT uslovima (konstantan broj čestica, pritisak i temperatura) sa korakom od 1fs, i različitog trajanja 4–20ns. Opšti parametri simulacija su uključivali: PME (eng. *particle-mesh Ewald*, Evaldova rešetka čestica) uslove za izračunavanje elektrostatičkih interakcija i Langevinov termostat na 310K, sa Langevinovim koeficijentom trenja  $\gamma = 2 \text{ ps}^{-1}$ . Pritisak je održavan na 1.01325 bar koristeći Nus-Huver Langevinov piston (period 200fs, period raspada 100fs i temperatura 310K). Kada su sistemi simulirani pod Kornel et al. poljem sila, NAMD softver je prilagođen promenom parametara izvršavanja kao što je opisano u uputstvu za NAMD softver.

##### *aMD*

Sve simulacije sistema ubrzanom molekularnom dinamikom (aMD) obavljene su koristeći NAMD softver i implementaciju aMD metoda u njemu (226). U simulacijama su korišćena i CHARMM 22/CMAP i Kornel et al. polja sila atomskih interakcija. Parametri simulacija analogni su onima korišćenim u cMD simulacijama. Simulacije su trajale 20 miliona koraka kod CDK9 sistema i 7 miliona koraka kod CDK8 sistema.

Pre aMD simulacija sistemi su ekvilibrisani klasičnom molekularnom dinamikom pod NPT uslovima, u trajanju od 5ns. Konformacije iz poslednje dve nanosekunde simulacije (2000 konformacija) iskorišćene su za izračunavanje torzione i ukupne potencijalne energije sistema, primenom NAMDEnergy modula VMD softvera.



Torziona ( $V_{\text{dih}}$ ) i ukupna potencijalna energija sistema ( $V_{\text{tot}}$ ) iz simulacije klasičnom molekularnom dinamikom su potrebne za definisanje aMD parametara, kako je opisano u literaturi (227):

$$\alpha_{\text{dih}} = 4 \text{ kcal/mol} \cdot N_{\text{res}}/5$$

$$E_{\text{dih}} = V_{\text{dih}} + 4 \cdot N_{\text{res}}$$

$$\alpha_{\text{tot}} = 0.16 \text{ kcal/mol} \cdot N_{\text{all}}$$

$$E_{\text{tot}} = V_{\text{tot}} + \alpha_{\text{tot}}$$

gde je  $N_{\text{res}}$  broj aminokiselina u sistemu a  $N_{\text{all}}$  ukupan broj atoma koji se simulira.  $\alpha_{\text{dih}}$  i  $\alpha_{\text{tot}}$  predstavljaju parametre aMD metoda koji kontrolišu stepen ubrzavanja dinamike sistema. Navedeni izrazi razvijeni su za primenu sa Kornel et al. grupom polja sila interakcija. U ovom istraživanju navedeni izrazi su uz simulacije sa Kornel et al. poljem sila interakcija upotrebljeni i za dobijanje vrednosti parametara aMD simulacija sa CHARMM 22/CMAP poljem sila interakcija. Simulirani sistemi su ostali stabilni. Po završetku aMD simulacija, ponašanje sistema je praćeno kroz fluktuacije RMSD vrednosti atoma i prosečnu vrednost potencijala dovedenog sistemu u toku simulacije.

### SMD

Sve SMD (eng. *steered molecular dynamics*, vođena molekularna dinamika) simulacije su sprovedene koristeći NAMD softver i CHARMM 22/CMAP polje sila interakcija atoma. SMD simulacije su vršene na kompleksima CDK9 i 14 dizajniranih potencijalnih peptidnih inhibitora CDK9/ciklin T1 interakcija. Sistemi su za SMD simulacije pripremani polazeći od rezultata dokinga. Strukture su solvatisane, neutralisane, minimizovane (30000 koraka) i zagrejane do 310K, postupkom opisanim u odeljku 4.1.1.1.

U svrhu odabira optimalnih uslova za SMD simulacije, sprovedene su test simulacije u kojima su parametri varirani. Isprobane su tri kombinacije fiksnih atoma: fiksiran samo jedan atom –  $C\alpha$  aminokiseline 20, koja se nalazi u rigidnoj  $\beta$ -valjak strukturi udaljenoj od mesta vezivanja peptida, fiksirani svi atomi polipeptidnog niza CDK9 i fiksirana distalna strana  $\beta$ -valjka nasuprot vezivne površine (atomi polipeptidnog niza aminokiselina 20 – 25, 32 – 37, 44 – 48, 104, 106, 107 strukture CDK9). Takođe, isprobane su različite brzine povlačenja liganada (5, 10, 50 i 100 Å/ns)

i različite konstante elastičnosti (1.5, 3 i 7 kcal·mol<sup>-1</sup>·Å<sup>2</sup>). Kao optimalan, izabran je protokol u kome su fiksirani atomi polipeptidnog niza aminokiselina 20 – 25, 32 – 37, 44 – 48, 104, 106, 107 strukture CDK9, brzina povlačenja liganda iznosi 10 Å/ns, uz konstantu elastičnosti opruge od 1.5 kcal·mol<sup>-1</sup>·Å<sup>2</sup>.

Pravac povlačenja peptidnih liganada tokom SMD simulacija testiran je primenom RAMD (eng. *random acceleration molecular dynamics*, molekularna dinamika sa nasumično orijentisanim ubrzanjem molekula) simulacija, po metodi koja je implementirana u NAMD softveru. Generisano je 100 trajektorija RAMD simulacija, svaka u trajanju od 500 ps, svaka sa različitim tokom (osigurano primenom različitih nasumičnih brojeva u algoritmu na početku simulacije). Pri RAMD simulacijama korišćeni su sledeći parametri: ubrzanje liganada od 0.05 kcal·mol<sup>-1</sup>·Å·amu, minimalno rastojanje koje ligand mora preći u svakom koraku simulacije 0.006 Å i rastojanje između liganda i proteina koje se smatra dovoljnim za prekid simulacije 50 Å.

Tokom SMD simulacije, svi atomi peptidnih liganada su povlačeni, a za pravac povlačenja je odabran vektor između centra mase svih atoma liganda i fiksiranih atoma proteina. Svaka SMD simulacija je trajala 4ns (2000000 koraka od po 2fs), uz primenu PME uslova za izračunavanje elektrostatičkih interakcija.

## 4.2 Analiza vezivnih površina CDK9/ciklin T1 kompleksa

### 4.2.1 Analiza strukturnih kontakata

Kristalografska struktura CDK9 / ciklin T1 kompleksa PDB ID: 3blh je upotrebljena za analizu i definisanje kontakata na vezivnoj površini kompleksa. Vizuelna analiza kontakata na površini obavljena je u VMD i Pymol softveru. Kontakti su analizirani primenom web dostupnih databaza proteinskih struktura: SCOWLP (173,174), 2p2idb (40) i PDBePISA (228). Šema interakcija dobijena je DIMPLOT modulom LigPlot+ (229) softvera. Prisustvo dehidrona u građi kompleksa i CDK9 analizirano je web servisom WRAPPA (230).

### 4.2.2 Hidratisanost vezivnih površina

Analiza očuvanih molekula vode na vezivnim površinama kompleksa prisutnih u kristalografskim strukturama je izvedena na 17 kristalografskih struktura CDK9/ciklinT1 kompleksa, preuzetih iz RCSB PDB databaze. Korišćene su strukture sa PDB ID brojevima: 3blh, 3blq, 3blr, 3lq5, 3tni, 3tn8, 3tnh, 4ec8, 4ec9, 3mi9, 3mia, 3my1, 4bcf, 4bch, 4bci, 4bcj i 4imy. Kao kriterijum da bi se molekul vode smatrao molekulom sa vezivne površine kompleksa uzeto je da se njegov atom kiseonika nalazi do 3.3 Å udaljen od atoma CDK9 i da istovremeno bude najviše 5 Å udaljen od atoma ciklina T1 (atomi vodonika nisu razmatrani pri definisanju atoma proteina). Udaljenost od 3.3 Å od površine proteina od interesa preuzeta je iz literature (165), da bi odgovarala prvoj hidratacionoj sferi proteina (166). Primenom ovog kriterijuma nisu nađeni odgovarajući molekuli vode u strukturama: 3lq5, 3tn8, 3tnh, 3tni, 4bcf, 4bcj, 4ec8, 4ec9 i 4imy. Zbog prisustva Tat proteina eliminisane su i strukture 3mia i 3mi9, pa je konačan skup kristalografskih struktura na kojima je sprovedena analiza uključivao strukture: 3blh, 3blq, 3blr, 3my1, 4bch, 4bci. Kriterijum da bi se jedan molekul vode smatrao očuvanim na vezivnoj površini kompleksa bio je da se bar u dve kristalografske strukture na njegovoj poziciji nalaze atomi kiseonika molekula vode na međusobnoj udaljenosti manjoj od 1Å (165). Analiza je obavljena programom pisanim

u python (231) programskom jeziku (xray\_consensus\_waters.py, Prilog 1), na koordinatama atoma dobijenim iz VMD softvera.

Analiza hidratacije vezivne površine je takođe obavljena WaterDock procedurom (165). U WaterDock analizi kristalografske strukture CDK9, oblast za analizu je definisana pozicijom aminokiselina Phe146 iz ciklina T1, dok je u analizi hidratacije različitih konformacija CDK9 sa vezanim nepeptidnim ligandima, centar oblasti za analizu dodeljen centrima mase atoma liganada (bez vodonika). Procedura predviđanja verovatnoće izmene molekula vode sa vezivne površine nepolarnom grupom liganda WaterDock softverom sprovedena je sa originalnim parametrima programa.

U analizi raspodele molekula vode u molekularnoj dinamici korišćen je VolMap modul VMD softvera i python program za analizu (water\_voxel.py, Prilog 2). VolMap modul je vršio raspodelu molekula vode u prostoru bar 4 Å oko proteina. Dati rezultati su prevođeni u .pdb format prerađenom dx2pdb.py procedurom dxTuber (232) programa. Python program je upoređivao broj prisutnih molekula vode po jedinici zapremine prostora oko proteina ( $0.125 \text{ \AA}^3$ ) sa gustinom vode u simuliranom rastvaraču bez proteina ( $0.00402$  molekula vode po  $0.125 \text{ \AA}^3$ ). Nakon utvrđivanja gustine vode u posmatranoj jediničnoj ćeliji prostora, tražena je srednja vrednost gustine sa 6, odnosno 26 ćelija suseda, zavisno od protokola, a zatim su jedinične ćelije koje su pokazivale gustinu vode veću od zadate vrednosti (1.5 ili 2 puta veću od prosečne gustine čiste vode) grupisane u grupe prečnika 2.8 Å.

#### 4.2.3 Analiza *hot spot* aminokiselina

Sve analize *hot spot* aminokiselina obavljene su na kristalografskoj strukturi CDK9 / ciklin T1 kompleksa (PDB ID: 3blh). Struktura je poslata na analizu putem web servisa: Robetta, KFC2, HotPoint i databaza ANCHOR (172), SCOWLP. Rezultati su vizualizovani u VMD softveru. Regioni interagujućih *hot spot* aminokiselina dobijeni su HotRegion (175) web servisom. Značaj naelektrisanja aminokiselina u procesu formiranja kompleksa procenjen je HyPARE (176) servisom, sa simuliranom jonskom jačinom rastvora od 150 mM. Evolutivno praćenje aminokiselina sa vezivnih površina proteina obavljeno je Evolutionary Trace report\_maker web servisom (233). U proceni

specifičnosti identifikovanih interakcija primarne strukture ciklina T1, T2 i K preuzete su iz UniProt (234) databaze i upoređene primenom clustalw (235) softvera.

Mapiranje proteinske površine malim organskim molekulima i procena interakcija sprovedeni su FTMap procedurom, putem FTMap web servisa (236). Strukture su za analizu pripremljene po literaturnom protokolu (237). Protokol je ukratko uključivao izolovanje N-terminalnog domena CDK9 (primarna struktura domena je preuzeta iz CATH (238) databaze) i pokretanje FTMap procedure sa parametrima podešenim za ispitivanje površine proteina u proteinskim kompleksima, a ne aktivnih mesta kakva su u enzimima. Korišćen je literaturni kriterijum za klasifikaciju rezultata, pri čemu se značajnim smatraju grupacije od bar 16 proba malih organskih molekula, sa bar jednom manjom grupacijom na manje od 6 Å udaljenosti od njih (30). Dodatni rotameri aminokiselina oko identifikovanih grupacija proba na vezivnoj površini CDK9 uključeni su kroz modifikovanu FTMap proceduru implementiranu kroz FTFlex web servis (239).

#### 4.2.4 Analiza fleksibilnosti CDK9

Početna struktura CDK9 korišćena za analizu fleksibilnosti izolovana je nakon 4ns simulacije slobodne CDK9 molekularnom dinamikom. Sva RMSD upoređivanja i poravnavanja struktura izvedena su u VMD softveru. Nedostajuće aminokiseline N-terminusa dodate su ručno na strukturu, u Pymol softveru i N-terminus je modelovan kao  $\beta$ -sekundarna struktura. N-terminus je minimizovan 10000 koraka u vakuumu, pri čemu su aminokiseline 1–15 bile pokretne, dok je ostatak proteina fiksiran. Minimizacija je obavljena koristeći VMD i NAMD. Nakon minimizacije, proteinska struktura je pripremljena za molekularnu dinamiku i obavljena je simulacija ubrzanom molekularnom dinamikom, po proceduri iz odeljka 4.1. aMD simulacije su sprovedene paralelno sa Kornel et al i CHARMM22/CMAP poljima sila. Nakon simulacija, trajektorije su obrađene u VMD. Da bi se olakšala analiza, snimljene su samo trajektorije atoma proteina bez vodonika i snimane strukture na svaka 2 koraka. Tako je dobijeno po 10000 struktura iz svake od četiri simulacije (N1 i N2 bez N-terminusa, simulacija A i B). Tokom aMD simulacija primenjeni potencijal je čuvan svakih 50 koraka. Nakon simulacije, sve strukture su sortirane po primenjenom potencijalu i

sačuvano je po 2000 najpovoljnijih struktura iz svake simulacije. Za manipulisanje pojedinačnim strukturama iz trajektorije i obradu simulacija korišćen je wordom (240) program. Strukture iz svake simulacije su zatim grupisane po RMSD vrednosti atoma aminokiselina na vezivnoj površini CDK9 (bez vodonika). Kod struktura CDK9 sa skraćenim N-terminusom za grupisanje su upotrebljene sve aminokiseline vezivne površine (8–13, 57–59, 61, 64, 67–68, 71, 81–84, 86, 99), dok je kod struktura sa čitavim N-terminusom grupisanje rađeno još i uključujući i prvih 7 aminokiselina. Za grupisanje je korišćen QT-like algoritam iz wordom programa, pri čemu je svakom klasteru struktura dodeljen poluprečnik od 2Å. Testirane su tri vrednosti poluprečnika 1, 1.5 i 2Å. Krajnji parametri su odabrani tako da se dobije dobra pokrivenost konformacionog prostora koji je protein uzorkovao tokom simulacije, uz korišćenje malog broja reprezentativnih struktura. Ukupan konformacioni prostor energetskih minimuma koje je protein posetio aproksimiran je svođenjem na 2000 analiziranih struktura, dok je procenat pokrivenosti dobijen kao procenat ukupnog broja struktura sadržanih u izabranim klasterima. Analiza grupisanja izvedena je empirijskim pravilom lakta u Libre Office Calc programu, po kome se granica značajnosti grupacija postavlja tako da dalje dodavanje grupa podataka ne daje značajan doprinos rezultatima (tj. na grafičkom prikazu podataka se javlja pregib u toj oblasti). U ovom slučaju, pravilo lakta je eliminisalo klustere sa zanemarljivo malim brojem struktura u odnosu na dominantne klustere. Centroidi 10 najvećih klastera iz svake simulacije su poslani na FTMap analizu. Mape prostornog rasporeda određenih aminokiselina tokom simulacije izrađene su VolMap modulom VMD programa. Korišćeni su sledeći parametri: jedinične ćelije zapremine sa stranicom od 1Å, srednja vrednost rezultata prisutnosti atoma kroz čitavu simulaciju, merenje sa pravim radijusom svakog atoma.

Pokušaj dinamičnog mapiranja vezivne površine CDK9 tokom simulacije molekularnom dinamikom u mešanom rastvaraču je obavljen po literaturnoj proceduri (188), koristeći VMD i NAMD programe. Za prilagođavanje broja molekula rastvarača i molekularnih proba obliku proteina i analizu gustine proba napisani su python programi.

Za RMSD analizu vezivne površine između različitih struktura CDK9 (kristalografske strukture, strukture iz simulacija N1, N2, A i B) korišćeni su samo atomi aminokiselina 57–59, 61, 64, 67–68, 71, 81–84, 86, 99 sa vezivne površine, da bi

se izbeglo međusobno upoređivanje različitih stanja aminokiselina N-terminusa. Kod RMSD analize CDK8 struktura takođe je izbegnuto upoređivanje različitih N-terminusa, ali i aminokiselina sa C-terminalnog domena kinaze, pa su korišćeni samo atomi aminokiselina 57–61, 64, 65, 67–69, 71–73, 84, 86–88, 91–93.

#### 4.2.5 Analiza dinamike formiranja hidrofobnih džepova

Analiza dinamike formiranja hidrofobnih džepova obavljena je na po 10000 struktura iz svake od četiri aMD simulacije CDK9 (N1, N2, A i B), bez prisutnih atoma vodonika i RMSD poravnatih sa strukturom proteina sa početka simulacije. U literaturi su strukture uzorkovane na svake 2.5 ps klasične molekularne dinamike (18), dok je u ovoj studiji uzorkovanje vršeno na svaka 2 koraka ubrzane molekularne dinamike. Analiza je obavljena EPOS\_BP (241) programom. Početne grupacije struktura u pdb formatu su podeljene u 100 grupa po 100 struktura i obavljeno je po pet parcijalnih analiza za svaku simulaciju proteina, zbog tehničkih ograničenja. Analiza je obavljena koristeći napisani python program (`find_interface_pockets_EPOS.py`, Prilog 3), koji je izolovao samo džepove koji se otvaraju na vezivnoj površini u regionima od interesa. Džepovi su vizualizovani u VMD programu, koristeći kristalografske položaje aminokiselina ciklina T1 (Phe 146 i Lys6) kao orijentire za identifikaciju džepova. Procentualna otvorenost pojedinačnih džepova tokom simulacije izražena je kao srednja vrednost njihove procentualne otvorenosti tokom pet parcijalnih analiza simulacije.

Analiza verovatnoće otvaranja džepova na vezivnoj površini CDK9 u odnosu na ostatak površine molekula sprovedena je po literaturnoj proceduri (69). Sva izračunavanja su sprovedena koristeći Rosetta 3.5 program (242).

Početna struktura u analizi bila je struktura CDK9 sa modelovanim celim N-terminusom, pre pripreme za aMD simulaciju. Analiza je sprovedena na dvadeset aminokiselina koje čine vezivnu površinu CDK9 u kompleksu sa ciklinom T1 (aminokiseline 8 – 13, 57 – 59, 61, 64, 67, 68, 71, 81 – 84, 86 i 99) i dvadeset nasumično odabranih aminokiselina sa površine proteina, koje ne ulaze u građu vezivne površine (aminokiseline 119, 129, 141, 146, 154, 185, 216, 226, 232, 243, 249, 274, 281, 302, 307, 316, 21, 37 i 76). Odabir aminokiselina sa proteinske površine obavljen

je po literaturnim kriterijuma (69): aminokiseline su bar 12 Å udaljene od vezivne površine CDK9 sa ciklinom T1 i bar 12 Å udaljene jedna od druge. Za odabir aminokiselina je napisan python program (select\_surface\_residues.py, Prilog 4), čiji algoritam prvo grubo odabira među svim aminokiselinama proteina one čiji C $\alpha$  atomi su udaljeniji od 12 Å od C $\alpha$  atoma aminokiselina sa vezivne površine, zatim njih proređuje tako da ostanu samo one koje su međusobno udaljene više od 12 Å, a zatim ponavlja ceo proces selekcije uključujući sve nevodonične atome aminokiselina (sa vezivne površine i iz grupe aminokiselina sa površine koje su odabrane u prethodnom ciklusu odabira) u zahtev za zadovoljavanje udaljenosti od 12 Å. Odabrane aminokiseline su vizualno proverene.

Za svaku od odabranih aminokiselina formirano je 1000 konformacija CDK9 u kojima je formiranje džepova bilo parametar optimizacije. Formirano je i 1000 konformacija CDK9 nezavisnih od formiranja džepova. Za simulacije je korišćen „relax“ protokol (243) Rosetta programa, po literaturnoj proceduri (69). Za svaku od konformacija i svaku od aminokiselina odabranih za analizu procenjeno je postojanje i zapremina džepa koji odgovara lokaciji aminokiseline i svi rezultati su prikupljeni zajedno i prikazani histogramima. Prilikom upoređivanja rezultata vođeno je računa o tome da se upoređuju rezultati samo onih aminokiselina koje imaju istu sekundarnu strukturu (po rezultatima STRIDE web servisa (244)) i uporedive veličine džepova u blizini, po protokolu iz literature (69).



### 4.3 Dizajn peptida iz fragmenata

U *de novo* dizajnu peptida kao početni fragmenti su korišćene prirodne L-aminokiseline, čijih 134 rotamera je preuzeto iz Penultimate Rotamer Library (245) baze rotamera, dok je struktura glicina dodata ručno. U prvoj fazi dizajna vezivna površina CDK9 definisana je aminokiselinama 8–13, 57–59, 61, 64, 67–68, 71, 81–84, 86, 99 i mapirana SEED (246) (eng. *solvation energy for exhaustive docking*) softverom. SEED procenjuje interakcije molekularnih fragmenata sa rigidnom proteinskom površinom, uzimajući u obzir i energiju desolvatacije. Mapiranje je izvedeno sa originalnim parametrima u SEED programu, na konformacijama CDK9 iz simulacija slobodne CDK9 (konformacija A) i kompleksa CDK9/ciklin T1 (konformacija AB). Potrebne energije interakcija aminokiselinskih fragmenata sa proteinskom površinom da bi oni bili uključeni u formiranje peptida iznosile su –5 i –8 kcal·mol<sup>-1</sup> u grubom i finom rangiranju. Nakon odabira povoljnih pozicija aminokiselina na površini CDK9, fragmenti su prosleđeni GANDI (247) (eng. *genetic algorithm-based de novo design of inhibitors*) programu. 135 aminokiselinskih fragmenata korišćenih u SEED mapiranju korišćeni su i kao fragmenti i kao linkeri u GANDI povezivanju peptida, pri čemu je formiranje veza dozvoljeno samo duž vektora ne-vodonični atom - vodonik uključenih u izgradnju peptidnih veza. GANDI program je podešen da formira pentapeptide, evoluirajući 20 grupacija (ostrva) sa po 100 kombinacija položaja aminokiselinskih fragmenata kroz 5000 koraka. U funkciji koja procenjuje povoljnost energetskih interakcija peptida sa proteinskom površinom zanemareni su doprinosi poreklom od dvo- i trodimenzionalne sličnosti prethodno razvijenim ligandima, dok su svi ostali parametri u dizajnu ostavljeni na predloženim vrednostima samog programa. U formiranju pentapeptida korišćene su iste konformacije CDK9 i definicija vezivne površine kao u SEED mapiranju aminokiselinskim fragmentima.

U poređenju peptidnih sekvenci sa aminokiselinskim sekvencama peptidnih supstrata CDK9 sekvence supstrata su preuzete iz PhosphoSitePlus (190) databaze (PhosphoSitePlus<sup>®</sup>, [www.phosphosite.org](http://www.phosphosite.org)). Poređenje je obavljeno ručno.

#### 4.4 Pretraživanje farmakofora

Početna struktura za definisanje pogodne farmakofore na CDK9 / ciklin T1 vezivnoj površini bila je kristalografska struktura kompleksa (PDB ID: 3blh). Na ovoj strukturi su identifikovane pogodne polazne tačke za dizajn inhibitora koristeći PocketQuery web servis (34,248). Sa liste rezultata izabran je najbolji rezultat sa površine ciklina T1 (aminokiseline Phe 146 i Lys 6) i prosleđen web servisima AnchorQuery (199) i ZincPharmer (249) koji obavljaju pretraživanje farmakofora. Sama farmakofora je inicijalno definisana sa pet tačaka: aromatično jezgro Phe 146, karbonilni kiseonik Phe 146 kao akceptor vodonične veze, donori vodoničnih veza sa dva azota lizina 6 i hidrofobni region koji odgovara bočnom nizu lizina 6. Pošto su samo tri jedinjenja zadovoljavala prvu farmakoforu, iz nje je isključena tačka koja opisuje hidrofobni region bočnog niza Lys 6 i pretraživanje je obavljeno sa novom farmakoforom. Još dve dodatne farmakofore su istražene – grupacija aminokiselina Phe 146, Lys 6 i Gln 142 sa ciklina T1 i aminokiseline 8–12 sa CDK9, sa farmakofornim tačkama predloženim od strane AnchorQuery servisa. Nakon pretraživanja farmakofora, preuzete su strukture rezultata i oni su pripremljeni za dalju analizu.

Početni skup dobijenih struktura je ručno podeljen na grupe po hemijskoj sličnosti i verovatnoći efikasne sinteze, a zatim je proverena sličnost odabranih struktura sa poznatim PPI inhibitorima tako što su izračunati molekularni deskriptori  $U_i$  i  $RDF070m$  primenom e-dragon (250) programa. Vrednosti deskriptora koje zadovoljavaju molekuli slični poznatim inhibitorima protein-protein interakcija preuzete su iz literature ( $RDF070m \geq 13.31$   $U_i \geq 3.95$  ili  $4.13$ ) (64).

## 4.5 Doking liganada

### 4.5.1 Odabir i priprema struktura za doking

#### *Receptori*

Za doking peptida korišćene su samo dve konformacije CDK9 (konformacija A i konformacija AB) dobijene iz 4ns simulacija CDK9 molekularnom dinamikom. Proteinske strukture su za doking pripremljene korišćenjem AutoDockTools (251) programa koji je atomima dodelio potrebno naelektrisanje i tipove atoma.

Doking nepeptidnih liganada obavljen je na 50 konformacija CDK9, po 10 iz simulacija N1, N2, A i B i još deset iz simulacije B sa alternativnim izborom aminokiselina u grupisanju konformacija. Konformacije su grupisane na način koji je opisan u odeljku 4.2.4. Konformacije su međusobno orijentisane tako da se minimizuju RMSD razlike položaja njihovih  $C\alpha$  atoma i pripremljene za doking koristeći `prepare_receptor4.py` program u sklopu AutoDock Tools programa, kojim su na svaku od proteinskih struktura dodati vodonikovi atomi i podešena naelektrisanja i tipovi atoma.

U oba slučaja AutoDock Tools je korišćen za definisanje prostora za vezivanje liganada. Prostor je definisan tako da obuhvati čitavu vezivnu površinu i deo aktivnog mesta CDK9 kod vezivanja peptida. Kod dokinga nepeptidnih struktura korišćen je samo deo vezivne površine CDK9 oko 20 Å oko regiona od interesa. Fokusiranje na regione koji sadrže *hot spot* aminokiseline i pogodne džepove za vezivanje liganada je u skladu sa literaturom (252).

#### *Ligandi*

Strukture liganada za doking takođe su pripremljene u AutoDockTools programu. Određene su fleksibilne torzije i sve peptidne veze su tretirane kao rigidne. Na isti način su pripremljeni i peptidni i nepeptidni ligandi.

#### 4.5.2. Doking procedura

Doking svih liganada obavljen je koristeći AutoDock Vina (191) program.

##### *Peptidi*

Peptidi su dokovani u prostor veličine 24 x 30 x 34 Å sa jediničnom ćelijom stranice 1 Å, centriran na vezivnoj površini CDK9. U svakom pokušaju dokinga zadržano je po dvadeset orijentacija peptida na vezivnoj površini sa najpovoljnijom energijom interakcije. Svaki peptid je dokovan dvadeset puta, uz originalne parametre softvera, osim faktora temeljnosti (eng. *exhaustiveness*) koji opisuje broj nasumično odabranih konformacija liganada na početku doking procedure i koji je podešen na vrednost 400. Pošto su testovi na malom broju peptida pokazali da se uporedivi rezultati dobijaju i sa *exhaustiveness* parametrom sa vrednošću 64 uz znatnu uštedu računarskog vremena, dodatnih 20 pokušaja dokinga sa svakim od peptida obavljeno je sa vrednošću od 64. Rezultati dokinga vizualizovani su u VMD i Pymol programima.

##### *Ligandi*

Nepeptidne strukture su dokovane u prostor veličine 20 x 20 x 20 Å, centriran na vezivnoj površini između lokacija vezivanja aminokiselina Phe 146 i Lys 6 sa ciklina T1. U svakom pokušaju dokinga sačuvano je po 20 najpovoljnijih orijentacija liganada. Ponovljeni doking (10 puta) na istoj konformaciji pokazao je da ligand svaki put konverguje ka istoj pozi, pa doking nije ponavljan više puta, zbog velikog broja konformacija proteina na kojima je obavljan.

#### 4.5.3 Rangiranje poza liganada

Kod peptida je za dalje simulacije upotrebljena poza sa vezivne površine CDK9 sa najboljom energijom interakcije. Kod nepeptidnih liganada su za svaku simulaciju i svaki ligand izdvojene poze sa najpovoljnijom energijom interakcije i sve odabrane poze su ponovo rangirane primenom MM-GBSA metode na izolovanim konformacijama (SS-MMGBSA).

## 4.6 MM-PBSA i MM-GBSA procedure

### *MM-PBSA*

MM-PBSA metoda je korišćena za procenu energije vezivanja peptidnih liganada za CDK9. Kompleksi peptida sa CDK9 su pripremljeni za molekularnu dinamiku i simulirani pod NPT uslovima 4ns, po proceduri sličnoj proceduri opisanoj u odeljku 4.1, ali sa drugačijom minimizacijom. Za MM-PBSA proceduru kompleksa peptida nisu posebno minimizovani molekuli rastvarača i ceo sistem, već je samo ceo sistem (kompleks u neutralisanom okruženju rastvarača) minimizovan 30000 koraka. Iz poslednje nanosekunde svake od MD simulacija uzorkovano je po 100 konformacija (na svakih 10ps) koristeći traj modul AmberTools 1.5 programa. Parametri za kompleks, CDK9 i peptid u svakom kompleksu pripremljeni su koristeći LeaP modul istog programa, uz Kornel et al. polje sila ff10 (ff99SB za proteine) i modifikovane radijuse atoma prilagođene PBSA proceduri (mbondi 2). Samo izračunavanje energije vezivanja sprovedeno je koristeći automatizovanu MMPBSA.py proceduru (253) u okviru AmberTools 1.5 programa, uz različite koncentracije soli (0.1 i 0.15 M).

Promena konformacione entropije vezivanja procenjena je koristeći nmode modul AmberTools 1.5 programa, na pet konformacija iz poslednje nanosekunde svake od simulacija kompleksa, uzorkovanih svakih 200 ps. Konformaciona entropija je preračunata na 310 K, jer je standardni parametri programa računaju na 298 K.

Interakcije peptida sa CDK9 šematski su prikazane koristeći LigPlot+ program.

### *MM-GBSA*

Energije vezivanja su procenjene MM-GBSA pristupom (254), implementiranim u AmberTools programu (model igb 2). MM-GBSA metoda je primenjena za izračunavanje energije vezivanja peptida i malih organskih molekula za CDK9.

Kompleksi peptida su za MM-GBSA izračunavanja pripremani kako je opisano u prethodnom odeljku (MM-PBSA) i za obe metode korišćene su iste konformacije iz simulacija. Kompleksi malih molekula sa CDK9 su pripremani opštom procedurom opisanom u odeljku 4.1 i istim programima.

Dok su kompleksi peptida simulirani 4ns, kompleksi malih molekula sa CDK9 simulirani su 5ns klasičnom molekularnom dinamikom, a zatim je za MM-GBSA analizu upotrebljeno 200 konformacija iz poslednje dve nanosekunde simulacije. Konformacije su uzorkovane na svakih 10ps koristeći trajni modul AmberTools 13 programa. Parametri za kompleks, protein i ligande su pripremljeni koristeći LeaP modul AmberTools 13 programa, pri čemu je za proteine korišćeno Kornel et al. polje sila ff10 (ff99SB), dok su ligandi parametrizovani u okviru GAFF polja sila, kako je opisano u odeljku 4.1.1 (Parametrizacija liganada). Testirana su dva MM-GBSA modela (igb 5 i igb 2) i dve koncentracije soli (0.1 i 0.15M).

Promena konformacione entropije vezivanja je za komplekse malih molekula sa CDK9 procenjena na po pet konformacija uzorkovanih na svakih 40 od 200 konformacija upotrebljenih za MM-GBSA analizu, koristeći nmode modul AmberTools 13 programa.

#### *SS-MMGBSA*

SS-MMGBSA (eng. *single snapshot molecular mechanics generalized Born surface area*, molekularna mehanika sa opštim Bornovim pristupom energiji solvatacije na jednoj konformaciji) metoda je varijanta MM-GBSA metode koja se izvodi samo na jednoj konformaciji kompleksa od interesa. U ovom istraživanju, SS-MMGBSA metoda je upotrebljena za preciznije rangiranje dobijenih rezultata dokinga malih molekula i za odabir konformacija za detaljnu analizu.

Za SS-MMGBSA analizu, pripremljeni su parametri kompleksa, CDK9 i liganda za svaki od kompleksa koristeći LeaP modul u AmberTools 13 programu. Proteinskim strukturama su dodeljeni parametri Kornel et al. polja sila ff10 (ff99SB), a ligandima parametri GAFF polja sila. Svaki od kompleksa je zatim minimizovan u implicitnom rastvaraču (GBIS (eng. *Generalized Born Implicit Solvent*) modulom NAMD 2.9 programa) 2000 koraka. U minimizaciji su korišćeni sledeći parametri: solventDielectric 80, ionConcentration 0.2, cutoff 16.0, pairlistdist 18.0 nonbondedFreq 2, fullElectFrequency 4. Za optimizovane konformacije kompleksa izračunata je energija vezivanja MM-GBSA metodom. Promena konformacione entropije vezivanja nije uzimana u obzir u SS-MMGBSA metodi.

## 4.7 DFT procedure

Za sva izračunavanja korišćen je ORCA (219) program.

Polazne strukture prelaznih stanja Pd-kompleksa pripremljene su koristeći cactvs web servis i Facio (255) program. Prelazna stanja su locirana koristeći B3LYP hibridni funkcional (256–258) i def2-SVP set orbitala (259), sa dodatnim def2-SVP/J setom orbitala u vakuumu. RIJCOSX aproksimacija (260) je upotrebljena za ubrzavanje pronalaženja prelaznih stanja. Kako se radi o kompleksu prelaznog metala, kod kojih konvergovanje rezultata pri *ab initio* optimizacijama može da bude problem, ublaženi su kriterijumi za konvergenciju (SlowConv), ali je tražena povećana preciznost optimizacije (TightSCF). Energije sistema u lociranim prelaznim stanjima su ponovo izračunate sa širim setom orbitala (def2-TZVP) za sve atome i istim funkcionalom, u vakuumu. Frekvenciona analiza (261) je obavljena na finalnim strukturama prelaznih stanja, radi njihove validacije.

Tačni uslovi za izračunavanja bili su:

- pretraga prelaznih stanja:

```
! B3LYP/G def2-SVP def2-SVP/J RIJCOSX TightSCF SlowConv Grid4 NoFinalGrid
!OptTS
%geom Calc_Hess true
```

- izračunavanje energije prelaznog stanja:

```
! B3LYP/G def2-TZVP NormalSCF Grid4 NoFinalGrid
```

- frekvenciona analiza:

```
! B3LYP/G def2-SVP def2-SVP/J RIJCOSX VDW TightSCF SlowConv Grid4
NoFinalGrid
!NumFreq
```

## Objavljeni rezultati iz doktorske disertacije

Randjelović J, Erić S, Savić V. Computational study and peptide inhibitors design for the CDK9 – cyclin T1 complex. *J Mol Model.* **2013**, *19*, 1711–25. **M21**

Tasic G, Randjelovic J, Vusurovic N, Maslak V, Husinec S, Savic V. A highly regioselective, protecting group controlled, synthesis of bicyclic compounds via Pd-catalysed intramolecular cyclisations. *Tetrahedron Lett.* **2013**, *54(18)*, 2243–46. **M22**



## 5. Literatura

1. Stumpf MPH, Thorne T, de Silva E, Stewart R, An HJ, Lappe M, et al. Estimating the size of the human interactome. *Proc Natl Acad Sci USA*. 2008 May 13;105(19):6959–64.
2. Northrup SH, Erickson HP. Kinetics of protein-protein association explained by Brownian dynamics computer simulation. *Proc Natl Acad Sci U S A*. 1992 Apr 15;89(8):3338–42.
3. Kleanthous C. *Protein-Protein Recognition*. Oxford University Press; 2000.
4. Kastiris PL, Bonvin AMJJ. On the binding affinity of macromolecular interactions: daring to ask why proteins interact. *J R Soc Interface*. 2013;10(79):20120835.
5. Schreiber G. The Association of Protein–Protein Complexes. In: Nussinov R, Schreiber G, editors. *Computational Protein-Protein Interactions* [Internet]. CRC Press; 2009 [cited 2013 Apr 17]. p. 87–107. Available from: <http://www.crcnetbase.com/doi/abs/10.1201/9781420070071.ch5>
6. Ubbink M. The courtship of proteins: understanding the encounter complex. *FEBS Lett*. 2009 Apr 2;583(7):1060–6.
7. Jones S, Thornton JM. Principles of protein-protein interactions. *Proc Natl Acad Sci USA*. 1996 Jan 9;93(1):13–20.
8. Bhardwaj N, Abyzov A, Clarke D, Shou C, Gerstein MB. Integration of protein motions with molecular networks reveals different mechanisms for permanent and transient interactions. *Protein Sci*. 2011 Oct;20(10):1745–54.
9. Beglov D, Hall DR, Brenke R, Shapovalov MV, Dunbrack RL Jr, Kozakov D, et al. Minimal ensembles of side chain conformers for modeling protein-protein interactions. *Proteins*. 2011 Oct 18;
10. Fleishman SJ, Khare SD, Koga N, Baker D. Restricted sidechain plasticity in the structures of native proteins and complexes. *Protein Sci*. 2011 Apr;20(4):753–7.
11. Kimura SR, Brower RC, Vajda S, Camacho CJ. Dynamical view of the positions of key side chains in protein-protein recognition. *Biophys J*. 2001 Feb;80(2):635–42.
12. Lo Conte L, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. *J Mol Biol*. 1999 Feb 5;285(5):2177–98.
13. Chakrabarti P, Janin J. Dissecting protein-protein recognition sites. *Proteins*. 2002 May 15;47(3):334–43.
14. Janin J, Chothia C. The structure of protein-protein recognition sites. *J Biol Chem*. 1990 Sep 25;265(27):16027–30.

15. Smith RD, Hu L, Falkner JA, Benson ML, Nerothin JP, Carlson HA. Exploring protein–ligand recognition with Binding MOAD. *Journal of Molecular Graphics and Modelling*. 2006 May;24(6):414–25.
16. Chène P. Drugs Targeting Protein–Protein Interactions. *ChemMedChem*. 2006;1(4):400–11.
17. Fuller JC, Burgoyne NJ, Jackson RM. Predicting druggable binding sites at the protein-protein interface. *Drug Discov Today*. 2009 Feb;14(3-4):155–61.
18. Eyrisch S, Helms V. Transient pockets on protein surfaces involved in protein-protein interaction. *J Med Chem*. 2007 Jul 26;50(15):3457–64.
19. Eyrisch S, Helms V. What induces pocket openings on protein surface patches involved in protein-protein interactions? *J Comput Aided Mol Des*. 2009 Feb;23(2):73–86.
20. Schreiber G, Fersht AR. Rapid, electrostatically assisted association of proteins. *Nat Struct Mol Biol*. 1996 May;3(5):427–31.
21. Bahadur RP, Zacharias M. The interface of protein-protein complexes: analysis of contacts and prediction of interactions. *Cell Mol Life Sci*. 2008 Apr;65(7-8):1059–72.
22. Jones S, Thornton JM. Protein-protein interactions: A review of protein dimer structures. *Progress in Biophysics and Molecular Biology*. 1995;63(1):31–65.
23. Chen J, Sawyer N, Regan L. Protein-protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area. *Protein Sci*. 2013 Apr;22(4):510–5.
24. Rodier F, Bahadur RP, Chakrabarti P, Janin J. Hydration of protein-protein interfaces. *Proteins*. 2005 Jul 1;60(1):36–45.
25. Reichmann D, Rahat O, Albeck S, Meged R, Dym O, Schreiber G. The modular architecture of protein-protein binding interfaces. *Proc Natl Acad Sci USA*. 2005 Jan 4;102(1):57–62.
26. Fraldi A, Licciardo P, Majello B, Giordano A, Lania L. Distinct regions of cyclinT1 are required for binding to CDK9 and for recruitment to the HIV-1 Tat/TAR complex. *J Cell Biochem Suppl*. 2001;Suppl 36:247–53.
27. Clackson T, Wells JA. A hot spot of binding energy in a hormone-receptor interface. *Science*. 1995 Jan 20;267(5196):383–6.
28. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. *J Mol Biol*. 1998 Jul 3;280(1):1–9.
29. Li X, Keskin O, Ma B, Nussinov R, Liang J. Protein-protein interactions: hot spots and structurally conserved residues often locate in complemented pockets that

- pre-organized in the unbound states: implications for docking. *J Mol Biol.* 2004 Nov 26;344(3):781–95.
30. Kozakov D, Hall DR, Chuang G-Y, Cencic R, Brenke R, Grove LE, et al. Structural conservation of druggable hot spots in protein–protein interfaces. *PNAS* [Internet]. 2011 Aug 1 [cited 2013 Apr 17]; Available from: <http://www.pnas.org/content/early/2011/07/26/1101835108>
  31. DeLano WL. Unraveling hot spots in binding interfaces: progress and challenges. *Curr Opin Struct Biol.* 2002 Feb;12(1):14–20.
  32. González-Ruiz D, Gohlke H. Targeting protein-protein interactions with small molecules: challenges and perspectives for computational binding epitope detection and ligand finding. *Curr Med Chem.* 2006;13(22):2607–25.
  33. Zerbe BS, Hall DR, Vajda S, Whitty A, Kozakov D. Relationship between hot spot residues and ligand binding hot spots in protein-protein interfaces. *J Chem Inf Model.* 2012 Aug 27;52(8):2236–44.
  34. Koes DR, Camacho CJ. Small-molecule inhibitor starting points learned from protein-protein interaction inhibitor structure. *Bioinformatics.* 2012 Mar 15;28(6):784–91.
  35. Li J, Liu Q. “Double water exclusion”: a hypothesis refining the O-ring theory for the hot spots at protein interfaces. *Bioinformatics.* 2009 Mar 15;25(6):743–50.
  36. Ramos RM, Fernandes LF, Moreira IS. Extending the applicability of the O-ring theory to protein-DNA complexes. *Comput Biol Chem.* 2013 Jun;44:31–9.
  37. Rajamani D, Thiel S, Vajda S, Camacho CJ. Anchor residues in protein–protein interactions. *PNAS.* 2004 Aug 3;101(31):11287–92.
  38. Sharma SK, Ramsey TM, Bair KW. Protein-protein interactions: lessons learned. *Curr Med Chem Anticancer Agents.* 2002 Mar;2(2):311–30.
  39. Veselovsky AV, Archakov AI. Inhibitors of Protein-Protein Interactions as Potential Drugs. *Current Computer - Aided Drug Design.* 2007;3(1):51–8.
  40. Basse MJ, Betzi S, Bourgeas R, Bouzidi S, Chetrit B, Hamon V, et al. 2P2Idb: a structural database dedicated to orthosteric modulation of protein-protein interactions. *Nucleic Acids Res.* 2013 Jan;41(Database issue):D824–827.
  41. Fry DC. Protein-protein interactions as targets for small molecule drug discovery. *Biopolymers.* 2006;84(6):535–52.
  42. Cochran AG. Antagonists of protein–protein interactions. *Chemistry & Biology.* 2000 Apr 1;7(4):R85–R94.
  43. Arkin MR, Wells JA. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat Rev Drug Discov.* 2004 Apr;3(4):301–17.

44. Pommier Y, Cherfils J. Interfacial inhibition of macromolecular interactions: nature's paradigm for drug discovery. *Trends Pharmacol Sci.* 2005 Mar;26(3):138–45.
45. Gondeau C, Gerbal-Chaloin S, Bello P, Aldrian-Herrada G, Morris MC, Divita G. Design of a novel class of peptide inhibitors of cyclin-dependent kinase/cyclin activation. *J Biol Chem.* 2005 Apr 8;280(14):13793–800.
46. Fujii N, Haresco JJ, Novak KAP, Stokoe D, Kuntz ID, Guy RK. A Selective Irreversible Inhibitor Targeting a PDZ Protein Interaction Domain. *J Am Chem Soc.* 2003 Oct 1;125(40):12074–5.
47. Dömling A. Small molecular weight protein-protein interaction antagonists: an insurmountable challenge? *Curr Opin Chem Biol.* 2008 Jun;12(3):281–91.
48. Yin H, Hamilton AD. Strategies for targeting protein-protein interactions with synthetic agents. *Angew Chem Int Ed Engl.* 2005 Jul 4;44(27):4130–63.
49. Gandhi L, Camidge DR, Ribeiro de Oliveira M, Bonomi P, Gandara D, Khaira D, et al. Phase I study of Navitoclax (ABT-263), a novel Bcl-2 family inhibitor, in patients with small-cell lung cancer and other solid tumors. *J Clin Oncol.* 2011 Mar 1;29(7):909–16.
50. Kuritzkes D, Kar S, Kirkpatrick P. Maraviroc. *Nat Rev Drug Discov.* 2008 Jan;7(1):15–6.
51. London N, Raveh B, Movshovitz-Attias D, Schueler-Furman O. Can Self-Inhibitory Peptides be Derived from the Interfaces of Globular Protein-Protein Interactions? *Proteins.* 2010 Nov 15;78(15):3140–9.
52. Ko E, Liu J, Perez LM, Lu G, Schaefer A, Burgess K. Universal Peptidomimetics. *J Am Chem Soc.* 2011 Jan 26;133(3):462–77.
53. Saraogi I, Hamilton AD. alpha-Helix mimetics as inhibitors of protein-protein interactions. *Biochem Soc Trans.* 2008 Dec;36(Pt 6):1414–7.
54. Ahn J-M, Han S-Y. Facile synthesis of benzamides to mimic an  $\alpha$ -helix. *Tetrahedron Letters.* 2007 May 14;48(20):3543–7.
55. Hirschmann R, Nicolaou KC, Pietranico S, Salvino J, Leahy EM, Sprengeler PA, et al. Nonpeptidal peptidomimetics with  $\beta$ -D-glucose scaffolding. A partial somatostatin agonist bearing a close structural relationship to a potent, selective substance P antagonist. *J Am Chem Soc.* 1992 Nov 1;114(23):9217–8.
56. Ko E, Raghuraman A, Perez LM, Ioerger TR, Burgess K. Exploring Key Orientations at Protein-Protein Interfaces with Small Molecule Probes. *J Am Chem Soc.* 2013 Jan 9;135(1):167–73.
57. Grimme D, González-ruiz D, Gohlke\* H. Chapter 13 Computational Strategies and Challenges for Targeting Protein-Protein Interactions with Small Molecules.

- Physico-Chemical and Computational Approaches to Drug Discovery [Internet]. 2012 [cited 2013 Apr 19]. Available from: <http://pubs.rsc.org/en/content/chapter/bk9781849733533-00319/978-1-84973-353-3>
58. Bourgeas R, Basse M-J, Morelli X, Roche P. Atomic Analysis of Protein-Protein Interfaces with Known Inhibitors: The 2P2I Database. *PLoS ONE*. 2010 Mar 9;5(3):e9598.
  59. Higuero AP, Schreyer A, Bickerton GRJ, Pitt WR, Groom CR, Blundell TL. Atomic interactions and profile of small molecules disrupting protein-protein interfaces: the TIMBAL database. *Chem Biol Drug Des*. 2009 Nov;74(5):457–67.
  60. Morelli X, Bourgeas R, Roche P. Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I). *Curr Opin Chem Biol*. 2011 Aug;15(4):475–81.
  61. Pagliaro L, Felding J, Audouze K, Nielsen SJ, Terry RB, Krog-Jensen C, et al. Emerging classes of protein-protein interaction inhibitors and new tools for their development. *Curr Opin Chem Biol*. 2004 Aug;8(4):442–9.
  62. Hamon V, Brunel JM, Combes S, Basse MJ, Roche P, Morelli X. 2P2Ichem: focused chemical libraries dedicated to orthosteric modulation of protein–protein interactions. *Med Chem Commun* [Internet]. 2013 Mar 8 [cited 2013 Apr 20]; Available from: <http://pubs.rsc.org/en/content/articlelanding/2013/md/c3md00018d>
  63. Buchwald P. Small-molecule protein-protein interaction inhibitors: therapeutic potential in light of molecular size, chemical space, and ligand binding efficiency considerations. *IUBMB Life*. 2010 Oct;62(10):724–31.
  64. Reynès C, Host H, Camproux A-C, Laconde G, Leroux F, Mazars A, et al. Designing Focused Chemical Libraries Enriched in Protein-Protein Interaction Inhibitors using Machine-Learning Methods. *PLoS Comput Biol*. 2010 Mar 5;6(3):e1000695.
  65. Davis FP, Sali A. The overlap of small molecule and protein binding sites within families of protein structures. *PLoS Comput Biol*. 2010 Feb;6(2):e1000668.
  66. Wells JA, McClendon CL. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature*. 2007 Dec 13;450(7172):1001–9.
  67. Walter P, Metzger J, Thiel C, Helms V. Predicting where Small Molecules Bind at Protein-Protein Interfaces. *PLoS ONE*. 2013 Mar 7;8(3):e58583.
  68. Jochim AL, Arora PS. Systematic analysis of helical protein interfaces reveals targets for synthetic inhibitors. *ACS Chem Biol*. 2010 Oct 15;5(10):919–23.
  69. Johnson DK, Karanicolas J. Druggable Protein Interaction Sites Are More Predisposed to Surface Pocket Formation than the Rest of the Protein Surface. *PLoS Comput Biol*. 2013 Mar 7;9(3):e1002951.

70. Matthews DJ, Gerritsen ME. Targeting Protein Kinases for Cancer Therapy. 1st ed. Wiley; 2010.
71. Peng J, Zhu Y, Milton JT, Price DH. Identification of multiple cyclin subunits of human P-TEFb. *Genes Dev.* 1998 Mar 1;12(5):755–62.
72. Grana X, De Luca A, Sang N, Fu Y, Claudio PP, Rosenblatt J, et al. PITALRE, a nuclear CDC2-related protein kinase that phosphorylates the retinoblastoma protein in vitro. *Proc Natl Acad Sci U S A.* 1994 Apr 26;91(9):3834–8.
73. Marshall RM, Grana X. Mechanisms controlling CDK9 activity. *Front Biosci.* 2006;11:2598–613.
74. Bagella L, MacLachlan TK, Buono RJ, Pisano MM, Giordano A, De Luca A. Cloning of murine CDK9/PITALRE and its tissue-specific expression in development. *J Cell Physiol.* 1998 Nov;177(2):206–13.
75. Shore SM, Byers SA, Maury W, Price DH. Identification of a novel isoform of Cdk9. *Gene.* 2003 Mar 27;307:175–82.
76. Liu H, Herrmann CH. Differential localization and expression of the Cdk9 42k and 55k isoforms. *J Cell Physiol.* 2005 Apr;203(1):251–60.
77. Garriga J, Peng J, Parreño M, Price DH, Henderson EE, Graña X. Upregulation of cyclin T1/CDK9 complexes during T cell activation. *Oncogene.* 1998 Dec 17;17(24):3093–102.
78. De Falco G, Leucci E, Onnis A, Bellan C, Tigli C, Wirths S, et al. Cdk9/Cyclin T1 complex: a key player during the activation/differentiation process of normal lymphoid B cells. *J Cell Physiol.* 2008 Apr;215(1):276–82.
79. Liu H, Herrmann CH, Chiang K, Sung T-L, Moon S-H, Donehower LA, et al. 55K isoform of CDK9 associates with Ku70 and is involved in DNA repair. *Biochem Biophys Res Commun.* 2010 Jun 25;397(2):245–50.
80. Napolitano G, Licciardo P, Carbone R, Majello B, Lania L. CDK9 has the intrinsic property to shuttle between nucleus and cytoplasm, and enhanced expression of cyclin T1 promotes its nuclear localization. *J Cell Physiol.* 2002 Aug;192(2):209–15.
81. O’Keeffe B, Fong Y, Chen D, Zhou S, Zhou Q. Requirement for a kinase-specific chaperone pathway in the production of a Cdk9/cyclin T1 heterodimer responsible for P-TEFb-mediated tat stimulation of HIV-1 transcription. *J Biol Chem.* 2000 Jan 7;275(1):279–87.
82. Romano G, Giordano A. Role of the cyclin-dependent kinase 9-related pathway in mammalian gene expression and human diseases. *Cell Cycle.* 2008 Dec;7(23):3664–8.

83. Baumli S, Lolli G, Lowe ED, Troiani S, Rusconi L, Bullock AN, et al. The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation. *EMBO J.* 2008 Jul 9;27(13):1907–18.
84. Giuliana Napolitano BM. Catalytic activity of Cdk9 is required for nuclear co-localization of the Cdk9/cyclin T1 (P-TEFb) complex. *Journal of cellular physiology.* 2003;197(1):1–7.
85. Garber ME, Mayall TP, Suess EM, Meisenhelder J, Thompson NE, Jones KA. CDK9 autophosphorylation regulates high-affinity binding of the human immunodeficiency virus type 1 tat-P-TEFb complex to TAR RNA. *Mol Cell Biol.* 2000 Sep;20(18):6958–69.
86. Baumli S, Hole AJ, Wang L-Z, Noble MEM, Endicott JA. The CDK9 tail determines the reaction pathway of positive transcription elongation factor b. *Structure.* 2012 Oct 10;20(10):1788–95.
87. Garriga J, Bhattacharya S, Calbó J, Marshall RM, Truongcao M, Haines DS, et al. CDK9 is constitutively expressed throughout the cell cycle, and its steady-state expression is independent of SKP2. *Mol Cell Biol.* 2003 Aug;23(15):5165–73.
88. Kiernan RE, Emiliani S, Nakayama K, Castro A, Labbé JC, Lorca T, et al. Interaction between cyclin T1 and SCF(SKP2) targets CDK9 for ubiquitination and degradation by the proteasome. *Mol Cell Biol.* 2001 Dec;21(23):7956–70.
89. Chiu Y-L, Cao H, Jacque J-M, Stevenson M, Rana TM. Inhibition of Human Immunodeficiency Virus Type 1 Replication by RNA Interference Directed against Human Transcription Elongation Factor P-TEFb (CDK9/CyclinT1). *J Virol.* 2004 Mar;78(5):2517–29.
90. Ramakrishnan R, Yu W, Rice AP. Limited redundancy in genes regulated by Cyclin T2 and Cyclin T1. *BMC Research Notes.* 2011 Jul 26;4(1):260.
91. Marshall NF, Peng J, Xie Z, Price DH. Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J Biol Chem.* 1996 Oct 25;271(43):27176–83.
92. Price DH. P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Mol Cell Biol.* 2000 Apr;20(8):2629–34.
93. Peterlin BM, Price DH. Controlling the elongation phase of transcription with P-TEFb. *Mol Cell.* 2006 Aug 4;23(3):297–305.
94. Fujinaga K, Irwin D, Huang Y, Taube R, Kurosu T, Peterlin BM. Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Mol Cell Biol.* 2004 Jan;24(2):787–95.

95. Pirngruber J, Shchebet A, Johnsen SA. Insights into the function of the human P-TEFb component CDK9 in the regulation of chromatin modifications and co-transcriptional mRNA processing. *Cell Cycle*. 2009 Nov 15;8(22):3636–42.
96. Simone C, Stiegler P, Bagella L, Pucci B, Bellan C, De Falco G, et al. Activation of MyoD-dependent transcription by cdk9/cyclin T2. *Oncogene*. 2002 Jun 13;21(26):4137–48.
97. Yu DS, Cortez D. A role for cdk9-cyclin k in maintaining genome integrity. *Cell Cycle*. 2011 Jan 1;10(1):28–32.
98. Blazek D, Kohoutek J, Bartholomeeusen K, Johansen E, Hulinkova P, Luo Z, et al. The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev*. 2011 Oct 15;25(20):2158–72.
99. Jang MK, Mochizuki K, Zhou M, Jeong H-S, Brady JN, Ozato K. The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol Cell*. 2005 Aug 19;19(4):523–34.
100. Yang Z, Yik JHN, Chen R, He N, Jang MK, Ozato K, et al. Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol Cell*. 2005 Aug 19;19(4):535–45.
101. Li Q, Price JP, Byers SA, Cheng D, Peng J, Price DH. Analysis of the large inactive P-TEFb complex indicates that it contains one 7SK molecule, a dimer of HEXIM1 or HEXIM2, and two P-TEFb molecules containing Cdk9 phosphorylated at threonine 186. *J Biol Chem*. 2005 Aug 5;280(31):28819–26.
102. Michels AA, Bensaude O. RNA-driven cyclin-dependent kinase regulation: when CDK9/cyclin T subunits of P-TEFb meet their ribonucleoprotein partners. *Biotechnol J*. 2008 Aug;3(8):1022–32.
103. Zhou Q, Yik JHN. The Yin and Yang of P-TEFb Regulation: Implications for Human Immunodeficiency Virus Gene Expression and Global Control of Cell Growth and Differentiation. *Microbiol Mol Biol Rev*. 2006 Sep 1;70(3):646–59.
104. Fu J, Yoon H-G, Qin J, Wong J. Regulation of P-TEFb Elongation Complex Activity by CDK9 Acetylation. *Mol Cell Biol*. 2007 Jul 1;27(13):4641–51.
105. Krystof V, Baumli S, Fürst R. Perspective of cyclin-dependent kinase 9 (CDK9) as a drug target. *Curr Pharm Des*. 2012;18(20):2883–90.
106. Wang S, Fischer PM. Cyclin-dependent kinase 9: a key transcriptional regulator and potential drug target in oncology, virology and cardiology. *Trends Pharmacol Sci*. 2008 Jun;29(6):302–13.



107. Bellan C, De Falco G, Lazzi S, Micheli P, Vicidomini S, Schürfeld K, et al. CDK9/CYCLIN T1 expression during normal lymphoid differentiation and malignant transformation. *J Pathol.* 2004 Aug;203(4):946–52.
108. Simone C, Giordano A. Abrogation of signal-dependent activation of the cdk9/cyclin T2a complex in human RD rhabdomyosarcoma cells. *Cell Death Differ.* 2006 Jul 14;14(1):192–5.
109. Lee DK, Duan HO, Chang C. Androgen receptor interacts with the positive elongation factor P-TEFb and enhances the efficiency of transcriptional elongation. *J Biol Chem.* 2001 Mar 30;276(13):9978–84.
110. De Falco G, Giordano A. CDK9: from basal transcription to cancer and AIDS. *Cancer Biol Ther.* 2002 Aug;1(4):342–7.
111. Węsierska-Gądek J, Kryštof V. Selective Cyclin-Dependent Kinase Inhibitors Discriminating between Cell Cycle and Transcriptional Kinases. *Annals of the New York Academy of Sciences.* 2009;1171(1):228–41.
112. Chen R, Keating MJ, Gandhi V, Plunkett W. Transcription inhibition by flavopiridol: mechanism of chronic lymphocytic leukemia cell death. *Blood.* 2005 Oct 1;106(7):2513–9.
113. Manohar SM, Rathos MJ, Sonawane V, Rao SV, Joshi KS. Cyclin-dependent kinase inhibitor, P276-00 induces apoptosis in multiple myeloma cells by inhibition of Cdk9-T1 and RNA polymerase II-dependent transcription. *Leuk Res.* 2011 Jun;35(6):821–30.
114. Scrace SF, Kierstan P, Borgognoni J, Wang L-Z, Denny S, Wayne J, et al. Transient treatment with CDK inhibitors eliminates proliferative potential even when their abilities to evoke apoptosis and DNA damage are blocked. *Cell Cycle.* 2008 Dec 15;7(24):3898–907.
115. Polier G, Ding J, Konkimalla BV, Eick D, Ribeiro N, Kohler R, et al. Wogonin and related natural flavones are inhibitors of CDK9 that induce apoptosis in cancer cells by transcriptional suppression of Mcl-1. *Cell Death Dis.* 2011 Jul;2(7):e182.
116. Shan B, Zhuo Y, Chin D, Morris CA, Morris GF, Lasky JA. Cyclin-dependent kinase 9 is required for tumor necrosis factor- $\alpha$ -stimulated matrix metalloproteinase-9 expression in human lung adenocarcinoma cells. *J Biol Chem.* 2005 Jan 14;280(2):1103–11.
117. Gojo I, Zhang B, Fenton RG. The cyclin-dependent kinase inhibitor flavopiridol induces apoptosis in multiple myeloma cells through transcriptional repression and down-regulation of Mcl-1. *Clin Cancer Res.* 2002 Nov;8(11):3527–38.
118. MacCallum DE, Melville J, Frame S, Watt K, Anderson S, Gianella-Borradori A, et al. Seliciclib (CYC202, R-Roscovitine) induces cell death in multiple myeloma cells by inhibition of RNA polymerase II-dependent transcription and down-regulation of Mcl-1. *Cancer Res.* 2005 Jun 15;65(12):5399–407.

119. Santo L, Vallet S, Hideshima T, Cirstea D, Ikeda H, Pozzi S, et al. AT7519, A novel small molecule multi-cyclin-dependent kinase inhibitor, induces apoptosis in multiple myeloma via GSK-3 $\beta$  activation and RNA polymerase II inhibition. *Oncogene*. 2010 Apr 22;29(16):2325–36.
120. Kryštof V, Rárová L, Liebl J, Zahler S, Jorda R, Voller J, et al. The selective P-TEFb inhibitor CAN508 targets angiogenesis. *Eur J Med Chem*. 2011 Sep;46(9):4289–94.
121. Wang K, Hampson P, Hazeldine J, Krystof V, Strnad M, Pechan P, et al. Cyclin-dependent kinase 9 activity regulates neutrophil spontaneous apoptosis. *PLoS ONE*. 2012;7(1):e30128.
122. Sedore SC, Byers SA, Biglione S, Price JP, Maury WJ, Price DH. Manipulation of P-TEFb control machinery by HIV: recruitment of P-TEFb from the large form by Tat and binding of HEXIM1 to TAR. *Nucleic Acids Res*. 2007;35(13):4347–58.
123. Garber ME, Wei P, KewalRamani VN, Mayall TP, Herrmann CH, Rice AP, et al. The interaction between HIV-1 Tat and human cyclin T1 requires zinc and a critical cysteine residue that is not conserved in the murine CycT1 protein. *Genes Dev*. 1998 Nov 15;12(22):3512–27.
124. Wimmer J, Fujinaga K, Taube R, Cujec TP, Zhu Y, Peng J, et al. Interactions between Tat and TAR and human immunodeficiency virus replication are facilitated by human cyclin T1 but not cyclins T2a or T2b. *Virology*. 1999 Mar 1;255(1):182–9.
125. Tahirov TH, Babayeva ND, Varzavand K, Cooper JJ, Sedore SC, Price DH. Crystal structure of HIV-1 Tat complexed with human P-TEFb. *Nature*. 2010 Jun 10;465(7299):747–51.
126. Zhou M, Halanski MA, Radonovich MF, Kashanchi F, Peng J, Price DH, et al. Tat Modifies the Activity of CDK9 To Phosphorylate Serine 5 of the RNA Polymerase II Carboxyl-Terminal Domain during Human Immunodeficiency Virus Type 1 Transcription. *Mol Cell Biol*. 2000 Jul;20(14):5077–86.
127. Bisgrove DA, Mahmoudi T, Henklein P, Verdin E. Conserved P-TEFb-interacting domain of BRD4 inhibits HIV transcription. *PNAS*. 2007 Aug 21;104(34):13690–5.
128. Dames SA, Schönichen A, Schulte A, Barboric M, Peterlin BM, Grzesiek S, et al. Structure of the Cyclin T binding domain of Hexim1 and molecular basis for its recognition of P-TEFb. *Proc Natl Acad Sci USA*. 2007 Sep 4;104(36):14312–7.
129. Barboric M, Yik JHN, Czudnochowski N, Yang Z, Chen R, Contreras X, et al. Tat competes with HEXIM1 to increase the active pool of P-TEFb for HIV-1 transcription. *Nucleic Acids Res*. 2007 Mar;35(6):2003–12.

130. Ammosova T, Obukhov Y, Kotelkin A, Breuer D, Beullens M, Gordeuk VR, et al. Protein Phosphatase-1 Activates CDK9 by Dephosphorylating Ser175. *PLoS ONE*. 2011 Apr 21;6(4):e18985.
131. Salerno D, Hasham MG, Marshall R, Garriga J, Tsygankov AY, Grana X. Direct inhibition of CDK9 blocks HIV-1 replication without preventing T cell activation in primary human peripheral blood lymphocytes. *Gene*. 2007 Dec 15;405(1-2):65–78.
132. Flores O, Lee G, Kessler J, Miller M, Schlieff W, Tomassini J, et al. Host-cell positive transcription elongation factor b kinase activity is essential and limiting for HIV type 1 replication. *Proc Natl Acad Sci U S A*. 1999 Jun 22;96(13):7208–13.
133. Biglione S, Byers SA, Price JP, Nguyen VT, Bensaude O, Price DH, et al. Inhibition of HIV-1 replication by P-TEFb inhibitors DRB, seliciclib and flavopiridol correlates with release of free P-TEFb from the large, inactive form of the complex. *Retrovirology*. 2007 Jul 11;4(1):47.
134. Bark-Jones SJ, Webb HM, West MJ. EBV EBNA 2 stimulates CDK9-dependent transcription and RNA polymerase II phosphorylation on serine 5. *Oncogene*. 2006 Mar 16;25(12):1775–85.
135. Zhou M, Lu H, Park H, Wilson-Chiru J, Linton R, Brady JN. Tax interacts with P-TEFb in a novel manner to stimulate human T-lymphotropic virus type 1 transcription. *J Virol*. 2006 May;80(10):4781–91.
136. Feichtinger S, Stamminger T, Müller R, Graf L, Klebl B, Eickhoff J, et al. Recruitment of cyclin-dependent kinase 9 to nuclear compartments during cytomegalovirus late replication: importance of an interaction between viral pUL69 and cyclin T1. *J Gen Virol*. 2011 Jul;92(Pt 7):1519–31.
137. Durand LO, Roizman B. Role of cdk9 in the Optimization of Expression of the Genes Regulated by ICP22 of Herpes Simplex Virus 1. *J Virol*. 2008 Nov 1;82(21):10591–9.
138. Holcakova J, Tomasec P, Bugert JJ, Wang EC, Wilkinson GW, Hrstka R, et al. The inhibitor of cyclin-dependent kinases, olomoucine II, exhibits potent antiviral properties. *Antivir Chem Chemother*. 2010;20(3):133–42.
139. Sano M, Abdellatif M, Oh H, Xie M, Bagella L, Giordano A, et al. Activation and function of cyclin T-Cdk9 (positive transcription elongation factor-b) in cardiac muscle-cell hypertrophy. *Nat Med*. 2002 Nov;8(11):1310–7.
140. Sano M, Schneider MD. Cyclin-dependent kinase-9: an RNAPII kinase at the nexus of cardiac growth and death cascades. *Circ Res*. 2004 Oct 29;95(9):867–76.
141. Krystof V, Chamrád I, Jorda R, Kohoutek J. Pharmacological targeting of CDK9 in cardiac hypertrophy. *Med Res Rev*. 2010 Jul;30(4):646–66.

142. Echaliier A, Endicott JA, Noble MEM. Recent developments in cyclin-dependent kinase biochemical and structural studies. *Biochim Biophys Acta*. 2010 Mar;1804(3):511–9.
143. Malumbres M, Pevarello P, Barbacid M, Bischoff JR. CDK inhibitors in cancer therapy: what is next? *Trends Pharmacol Sci*. 2008 Jan;29(1):16–21.
144. Węsierska-Gądek J, Maurer M, Zulehner N, Komina O. Whether to target single or multiple CDKs for therapy? That is the question. *J Cell Physiol*. 2011 Feb;226(2):341–9.
145. Kirkland LO, McInnes C. Non-ATP competitive protein kinase inhibitors as anti-tumor therapeutics. *Biochemical Pharmacology*. 2009;77(10):1561–71.
146. Heuer TS. Discovery of Selective CDK9 Small Molecule Inhibitors: CDK9 Inhibition in Tumor Cells is Associated with Inhibition of Proliferation and Induction of Apoptosis. AACR-NCI- EORTC International Conference on Molecular Targets and Cancer Therapeutics. Geneva, Switzerland; 2008.
147. Krystof V, Cankar P, Frysová I, Slouka J, Kontopidis G, Dzubač P, et al. 4-arylo-3,5-diamino-1H-pyrazole CDK inhibitors: SAR study, crystal structure in complex with CDK2, selectivity, and cellular effects. *J Med Chem*. 2006 Nov 2;49(22):6500–9.
148. Hole AJ, Baumli S, Shao H, Shi S, Huang S, Pepper C, et al. Comparative structural and functional studies of 4-(thiazol-5-yl)-2-(phenylamino)pyrimidine-5-carbonitrile CDK9 inhibitors suggest the basis for isotype selectivity. *J Med Chem*. 2013 Feb 14;56(3):660–70.
149. Shao H, Shi S, Huang S, Hole AJ, Abbas AY, Baumli S, et al. Substituted 4-(Thiazol-5-yl)-2-(phenylamino)pyrimidines Are Highly Active CDK9 Inhibitors: Synthesis, X-ray Crystal Structures, Structure-Activity Relationship, and Anticancer Activities. *J Med Chem*. 2013 Feb 14;56(3):640–59.
150. Lolli G. Structural dissection of cyclin dependent kinases regulation and protein recognition properties. *Cell Cycle*. 2010 Apr 15;9(8):1551–61.
151. Wang S, Griffiths G, Midgley CA, Barnett AL, Cooper M, Grabarek J, et al. Discovery and characterization of 2-anilino-4- (thiazol-5-yl)pyrimidine transcriptional CDK inhibitors as anticancer agents. *Chem Biol*. 2010 Oct 29;17(10):1111–21.
152. Baumli S, Endicott JA, Johnson LN. Halogen bonds form the basis for selective P-TEFb inhibition by DRB. *Chem Biol*. 2010 Sep 24;17(9):931–6.
153. Bettayeb K, Oumata N, Echaliier A, Ferandin Y, Endicott JA, Galons H, et al. CR8, a potent and selective, roscovitine-derived inhibitor of cyclin-dependent kinases. *Oncogene*. 2008 Oct 2;27(44):5797–807.

154. Baumli S, Hole AJ, Noble MEM, Endicott JA. The CDK9 C-helix Exhibits Conformational Plasticity That May Explain the Selectivity of CAN508. *ACS Chem Biol*. 2012 May 18;7(5):811–6.
155. Orzáez M, Gortat A, Mondragón L, Bachs O, Pérez-Payá E. ATP-noncompetitive inhibitors of CDK-cyclin complexes. *ChemMedChem*. 2009 Jan;4(1):19–24.
156. Arkin MR, Whitty A. The road less traveled: modulating signal transduction enzymes by inhibiting their protein-protein interactions. *Curr Opin Chem Biol*. 2009 Jun;13(3):284–90.
157. Van Duyne R, Guendel I, Jaworski E, Sampey G, Klase Z, Chen H, et al. Effect of Mimetic CDK9 Inhibitors on HIV-1-Activated Transcription. *J Mol Biol*. 2013 Feb 22;425(4):812–29.
158. Canela N, Orzáez M, Fucho R, Mateo F, Gutierrez R, Pineda-Lucena A, et al. Identification of an hexapeptide that binds to a surface pocket in cyclin A and inhibits the catalytic activity of the complex cyclin-dependent kinase 2-cyclin A. *J Biol Chem*. 2006 Nov 24;281(47):35942–53.
159. Schneider EV, Böttcher J, Blaesse M, Neumann L, Huber R, Maskos K. The Structure of CDK8/CycC Implicates Specificity in the CDK/Cyclin Family and Reveals Interaction with a Deep Pocket Binder. *Journal of Molecular Biology*. 2011 Sep 16;412(2):251–66.
160. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucl Acids Res*. 2000 Jan 1;28(1):235–42.
161. Bettayeb K, Baunbæk D, Delehouze C, Loaëc N, Hole AJ, Baumli S, et al. CDK Inhibitors Roscovitine and CR8 Trigger Mcl-1 Down-Regulation and Apoptotic Cell Death in Neuroblastoma Cells. *Genes Cancer*. 2010 Apr;1(4):369–80.
162. Schulze-Gahmen U, Upton H, Birnberg A, Bao K, Chou S, Krogan NJ, et al. The AFF4 scaffold binds human P-TEFb adjacent to HIV Tat. *eLife [Internet]*. 2013 Mar 5 [cited 2013 Apr 1];2. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3589825/>
163. Davis AM, Teague SJ, Kleywegt GJ. Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angew Chem Int Ed Engl*. 2003 Jun 23;42(24):2718–36.
164. Fernández A, Scott LR. Adherence of packing defects in soluble proteins. *Phys Rev Lett*. 2003 Jul 4;91(1):018102.
165. Ross GA, Morris GM, Biggin PC. Rapid and Accurate Prediction and Scoring of Water Molecules in Protein Binding Sites. *PLoS ONE*. 2012 Mar 1;7(3):e32036.
166. Narten AH, Levy HA. Liquid Water: Molecular Correlation Functions from X-Ray Diffraction. *The Journal of Chemical Physics*. 1971 Sep 1;55(5):2263–9.

167. Henchman RH, McCammon JA. Extracting hydration sites around proteins from explicit water simulations. *J Comput Chem*. 2002 Jul 15;23(9):861–9.
168. Paterlini MG, Ferguson DM. Constant temperature simulations using the Langevin equation with velocity Verlet integration. *Chemical Physics*. 1998 Sep 15;236(1–3):243–52.
169. Kortemme T, Kim DE, Baker D. Computational alanine scanning of protein-protein interfaces. *Sci STKE*. 2004 Feb 10;2004(219):pl2.
170. Tuncbag N, Keskin O, Gursoy A. HotPoint: hot spot prediction server for protein interfaces. *Nucleic Acids Res*. 2010 Jul;38(Web Server issue):W402–406.
171. Zhu X, Mitchell JC. KFC2: a knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features. *Proteins*. 2011 Sep;79(9):2671–83.
172. Meireles LMC, Dömling AS, Camacho CJ. ANCHOR: a web server and database for analysis of protein-protein interaction binding pockets for drug discovery. *Nucleic Acids Res*. 2010 Jul;38(Web Server issue):W407–411.
173. Teyra J, Doms A, Schroeder M, Pisabarro MT. SCOWLP: a web-based database for detailed characterization and visualization of protein interfaces. *BMC Bioinformatics*. 2006 Mar 2;7(1):104.
174. Teyra J, Paszkowski-Rogacz M, Anders G, Pisabarro MT. SCOWLP classification: structural comparison and analysis of protein binding regions. *BMC Bioinformatics*. 2008;9:9.
175. Cukuroglu E, Gursoy A, Keskin O. HotRegion: a database of predicted hot spot clusters. *Nucleic Acids Res*. 2012 Jan;40(D1):D829–D833.
176. Shaul Y, Schreiber G. Exploring the charge space of protein-protein association: a proteomic study. *Proteins*. 2005 Aug 15;60(3):341–52.
177. Mihalek I, Reš I, Lichtarge O. A Family of Evolution–Entropy Hybrid Methods for Ranking Protein Residues by Importance. *Journal of Molecular Biology*. 2004 Mar 5;336(5):1265–82.
178. Bieniasz PD, Grdina TA, Bogerd HP, Cullen BR. Recruitment of cyclin T1/P-TEFb to an HIV type 1 long terminal repeat promoter proximal RNA target is both necessary and sufficient for full activation of transcription. *PNAS*. 1999 Jul 6;96(14):7791–6.
179. Fujinaga K, Irwin D, Geyer M, Peterlin BM. Optimized chimeras between kinase-inactive mutant Cdk9 and truncated cyclin T1 proteins efficiently inhibit Tat transactivation and human immunodeficiency virus gene expression. *J Virol*. 2002 Nov;76(21):10873–81.

180. Shaw DE, Dror RO, Salmon JK, Grossman JP, Mackenzie KM, Bank JA, et al. Millisecond-scale molecular dynamics simulations on Anton. Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis. 2009. p. 1–11.
181. Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J Chem Phys*. 2004 Jun 22;120(24):11919–29.
182. Pierce LCT, Salomon-Ferrer R, Augusto F. de Oliveira C, McCammon JA, Walker RC. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J Chem Theory Comput*. 2012 Sep 11;8(9):2997–3002.
183. Markwick PRL, Cervantes CF, Abel BL, Komives EA, Blackledge M, McCammon JA. Enhanced Conformational Space Sampling Improves the Prediction of Chemical Shifts in Proteins. *J Am Chem Soc*. 2010 Feb 3;132(4):1220–1.
184. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*. 2006 Nov 15;65(3):712–25.
185. MacKerell, Bashford D, Bellott, Dunbrack, Evanseck JD, Field MJ, et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins†. *J Phys Chem B*. 1998 Apr 1;102(18):3586–616.
186. Mackerell AD Jr, Feig M, Brooks CL 3rd. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem*. 2004 Aug;25(11):1400–15.
187. De Leon-Boenig G, Bowman KK, Feng JA, Crawford T, Everett C, Franke Y, et al. The Crystal Structure of the Catalytic Domain of the NF- $\kappa$ B Inducing Kinase Reveals a Narrow but Flexible Active Site. *Structure*. 2012 Oct 10;20(10):1704–14.
188. Bakan A, Nevins N, Lakdawala AS, Bahar I. Druggability Assessment of Allosteric Proteins by Dynamics Simulations in the Presence of Probe Molecules. *J Chem Theory Comput*. 2012 Jul 10;8(7):2435–47.
189. Donsky E, Wolfson HJ. PepCrawler: a fast RRT-based algorithm for high-resolution refinement and binding affinity estimation of peptide inhibitors. *Bioinformatics*. 2011 Oct 15;27(20):2836–42.
190. Hornbeck PV, Kornhauser JM, Tkachev S, Zhang B, Skrzypek E, Murray B, et al. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Research*. 2011 Dec 1;40(D1):D261–D270.

191. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 2010 Jan 30;31(2):455–61.
192. Viet MH, Ngo ST, Lam NS, Li MS. Inhibition of aggregation of amyloid peptides by beta-sheet breaker peptides and their binding affinity. *J Phys Chem B*. 2011 Jun 9;115(22):7433–46.
193. Kisseljova K, Kuznetsov A, Baudy-Floc'h M, Järv J. Effect of two simultaneous aza- $\beta$ -amino acid substitutions on recognition of peptide substrates by cAMP dependent protein kinase catalytic subunit. *Bioorg Chem*. 2011 Aug;39(4):133–7.
194. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, et al. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc Chem Res*. 2000 Dec 1;33(12):889–97.
195. Izrailev S, Stepaniants S, Isralewitz B, Kosztin D, Lu H, Molnar F, et al. Steered Molecular Dynamics. In: Deuffhard P, Hermans J, Leimkuhler B, Mark AE, Reich S, Skeel RD, editors. *Computational Molecular Dynamics: Challenges, Methods, Ideas* [Internet]. Springer Berlin Heidelberg; 1999 [cited 2013 Jun 21]. p. 39–65. Available from: [http://link.springer.com/chapter/10.1007/978-3-642-58360-5\\_2](http://link.springer.com/chapter/10.1007/978-3-642-58360-5_2)
196. Colizzi F, Perozzo R, Scapozza L, Recanatini M, Cavalli A. Single-molecule pulling simulations can discern active from inactive enzyme inhibitors. *J Am Chem Soc*. 2010 Jun 2;132(21):7361–71.
197. Cuendet MA, Michielin O. Protein-protein interaction investigated by steered molecular dynamics: the TCR-pMHC complex. *Biophys J*. 2008 Oct;95(8):3575–90.
198. Lüdemann SK, Lounnas V, Wade RC. How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms. *J Mol Biol*. 2000 Nov 10;303(5):797–811.
199. Koes D, Khoury K, Huang Y, Wang W, Bista M, Popowicz GM, et al. Enabling Large-Scale Design, Synthesis and Validation of Small Molecule Protein-Protein Antagonists. *PLoS ONE*. 2012 Mar 12;7(3):e32839.
200. Amaro RE, Baron R, McCammon JA. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J Comput Aided Mol Des*. 2008 Sep;22(9):693–705.
201. Wright CW, Allen D, Phillipson JD, Kirby GC, Warhurst DC, Massiot G, et al. *Alstonia* species: are they effective in malaria treatment? *Journal of Ethnopharmacology*. 1993 Sep;40(1):41–5.
202. Cherif A, Massiot G, Le Men-Ollivier L. Corialstonine, a novel quinoline alkaloid from *Alstonia coriacea*. *Heterocycles*. 1987;26(12):3055–8.



203. Tasic G, Randjelovic J, Vusurovic N, Maslak V, Husinec S, Savic V. A highly regioselective, protecting group controlled, synthesis of bicyclic compounds via Pd-catalysed intramolecular cyclisations. *Tetrahedron Letters*. 2013 May 1;54(18):2243–6.
204. Amatore C, Carré E, Jutand A, M'Barki MA, Meyer G. Evidence for the ligation of palladium(0) complexes by acetate ions: Consequences on the mechanism of their oxidative addition with phenyl iodide and PhPd(OAc)(PPh<sub>3</sub>)<sub>2</sub> as intermediate in the heck reaction. *Organometallics*. 1995;14(12):5605–14.
205. Jutand A. Mechanisms of the Mizoroki–Heck Reaction. In: Oestreich rtin, editor. *The Mizoroki–Heck Reaction* [Internet]. John Wiley & Sons, Ltd; 2009 [cited 2013 Jun 26]. p. 1–50. Available from: <http://onlinelibrary.wiley.com/doi/10.1002/9780470716076.ch1/summary>
206. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996 Feb;14(1):33–8, 27–8.
207. Eyal E, Gerzon S, Potapov V, Edelman M, Sobolev V. The limit of accuracy of protein modeling: influence of crystal packing on protein structure. *J Mol Biol*. 2005 Aug 12;351(2):431–42.
208. Fiser A, Do RK, Sali A. Modeling of loops in protein structures. *Protein Sci*. 2000 Sep;9(9):1753–73.
209. Fiser A, Sali A. ModLoop: automated modeling of loops in protein structures. *Bioinformatics*. 2003 Dec 12;19(18):2500–1.
210. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. *J Comput Chem*. 2005 Dec;26(16):1781–802.
211. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J Chem Theory Comput*. 2011 Feb 8;7(2):525–37.
212. The PyMOL Molecular Graphics System, Version 1.5.0.1 Schrödinger, LLC.
213. Joung IS, Cheatham, TE. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J Phys Chem B*. 2008 Jul 31;112(30):9020–41.
214. Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, et al. AMBER 11. University of California, San Francisco;
215. Word JM, Lovell SC, Richardson JS, Richardson DC. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol*. 1999 Jan 29;285(4):1735–47.

216. Wang J, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model*. 2006 Oct;25(2):247–60.
217. Silva AWS da, Vranken WF. ACPYPE - AnteChamber PYthon Parser interface. *BMC Research Notes*. 2012 Jul 23;5(1):367.
218. Stewart JJP. MOPAC2012 [Internet]. Colorado Springs, CO, USA: Stewart Computational Chemistry; 2012. Available from: <http://OpenMOPAC.net>
219. Neese F. ORCA, version 2.9, an ab initio, density functional and semiempirical program package. Max Planck-Institute for Bioinorganic Chemistry: Mülheim a.d. Ruhr, Germany; 2012.
220. Dupradeau F-Y, Pigache A, Zaffran T, Savineau C, Lelong R, Grivel N, et al. The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. *Phys Chem Chem Phys*. 2010 Jul 28;12(28):7821–39.
221. Vanquelef E, Simon S, Marquant G, Garcia E, Klimerak G, Delepine JC, et al. R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res*. 2011 Jul;39(Web Server issue):W511–517.
222. Bayly CI, Cieplak P, Cornell W, Kollman PA. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J Phys Chem*. 1993 Oct 1;97(40):10269–80.
223. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. *Gaussian 09*. Wallingford CT: Gaussian, Inc.; 2009.
224. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *J Comput Chem*. 2004 Jul 15;25(9):1157–74.
225. Rodriguez R, Chinea G, Lopez N, Pons T, Vriend G. Homology modeling, model and software evaluation: three related resources. *Bioinformatics*. 1998 Jan 1;14(6):523–8.
226. Wang Y, Harrison CB, Schulten K, McCammon JA. Implementation of accelerated molecular dynamics in NAMD. *Comput Sci Disc*. 2011 Jan 1;4(1):015002.
227. Markwick PRL, McCammon JA. Studying functional dynamics in bio-molecules using accelerated molecular dynamics. *Phys Chem Chem Phys*. 2011 Dec 7;13(45):20053–65.
228. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*. 2007 Sep 21;372(3):774–97.
229. Laskowski RA, Swindells MB. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model*. 2011 Oct 24;51(10):2778–86.

230. Fernandez A. WRAPPA [Internet]. Available from: <http://www.wrappa.org/wrapa01/>
231. Python [Internet]. Available from: <http://www.python.org/>
232. Raunest M, Kandt C. dxTuber: detecting protein cavities, tunnels and clefts based on protein and solvent dynamics. *J Mol Graph Model*. 2011 Jun;29(7):895–905.
233. Mihalek I, Reš I, Lichtarge O. Evolutionary trace report\_maker: a new type of service for comparative analysis of proteins. *Bioinformatics*. 2006 Jul 1;22(13):1656–7.
234. The UniProt Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Research*. 2011 Nov 18;40(D1):D71–D75.
235. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007 Nov 1;23(21):2947–8.
236. Brenke R, Kozakov D, Chuang G-Y, Beglov D, Hall D, Landon MR, et al. Fragment-based identification of druggable “hot spots” of proteins using Fourier domain correlation techniques. *Bioinformatics*. 2009 Mar 1;25(5):621–7.
237. Hall DR, Kozakov D, Vajda S. Analysis of protein binding sites by computational solvent mapping. *Methods Mol Biol*. 2012;819:13–27.
238. Sillitoe I, Cuff AL, Dessailly BH, Dawson NL, Furnham N, Lee D, et al. New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures. *Nucleic Acids Res*. 2013 Jan;41(Database issue):D490–498.
239. Grove LE, Hall DR, Beglov D, Vajda S, Kozakov D. FTFlex: accounting for binding site flexibility to improve fragment-based identification of druggable hot spots. *Bioinformatics*. 2013 May 1;29(9):1218–9.
240. Seeber M, Cecchini M, Rao F, Settanni G, Caflisch A. Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics*. 2007 Oct 1;23(19):2625–7.
241. Eyrisch S. EPOS\_BP - Ensemble of Pockets on Protein Surfaces with BALLPass [Internet]. Available from: <http://gepard.bioinformatik.uni-saarland.de/software/epos-bp>
242. Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Meth Enzymol*. 2011;487:545–74.

243. Qian B, Raman S, Das R, Bradley P, McCoy AJ, Read RJ, et al. High-resolution structure prediction and the crystallographic phase problem. *Nature*. 2007 Nov 8;450(7167):259–64.
244. Heinig M, Frishman D. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res*. 2004 Jul 1;32(Web Server issue):W500–502.
245. Lovell SC, Word JM, Richardson JS, Richardson DC. The penultimate rotamer library. *Proteins*. 2000 Aug 15;40(3):389–408.
246. Majeux N, Scarsi M, Apostolakis J, Ehrhardt C, Caflisch A. Exhaustive docking of molecular fragments with electrostatic solvation. *Proteins*. 1999 Oct 1;37(1):88–105.
247. Dey F, Caflisch A. Fragment-Based de Novo Ligand Design by Multiobjective Evolutionary Optimization. *J Chem Inf Model*. 2008 Mar 1;48(3):679–90.
248. Koes DR, Camacho CJ. PocketQuery: protein-protein interaction inhibitor starting points from protein-protein interaction structure. *Nucleic Acids Research*. 2012 Apr 20;40(W1):W387–W392.
249. Koes DR, Camacho CJ. ZINCPharmer: pharmacophore search of the ZINC database. *Nucleic Acids Research*. 2012 May 2;40(W1):W409–W414.
250. Tetko IV, Gasteiger J, Todeschini R, Mauri A, Livingstone D, Ertl P, et al. Virtual computational chemistry laboratory--design and description. *J Comput Aided Mol Des*. 2005 Jun;19(6):453–63.
251. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J Comput Chem*. 2009 Dec;30(16):2785–91.
252. Metz A, Pflieger C, Kopitz H, Pfeiffer-Marek S, Baringhaus K-H, Gohlke H. Hot Spots and Transient Pockets: Predicting the Determinants of Small-Molecule Binding to a Protein–Protein Interface. *J Chem Inf Model*. 2012 Jan 23;52(1):120–33.
253. Miller BR, McGee TD, Swails JM, Homeyer N, Gohlke H, Roitberg AE. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J Chem Theory Comput*. 2012 Sep 11;8(9):3314–21.
254. Onufriev A, Bashford D, Case DA. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins*. 2004 May 1;55(2):383–94.
255. Suenaga M. Facio: New Computational Chemistry Environment for PC GAMESS. *J Comput Chem Jpn*. 2005;4(1):25–32.

256. Becke AD. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics*. 1993 Apr 1;98(7):5648–52.
257. Lee C, Yang W, Parr RG. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B*. 1988 Jan 15;37(2):785–9.
258. Vosko SH, Wilk L, Nusair M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can J Phys*. 1980 Aug 1;58(8):1200–11.
259. Weigend F, Ahlrichs R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys Chem Chem Phys*. 2005 Sep 21;7(18):3297–305.
260. Neese F, Wennmohs F, Hansen A, Becker U. Efficient, approximate and parallel Hartree–Fock and hybrid DFT calculations. A “chain-of-spheres” algorithm for the Hartree–Fock exchange. *Chemical Physics*. 2009 Feb 17;356(1–3):98–109.
261. McIver JW, Komornicki A. Structure of transition states in organic reactions. General theory and an application to the cyclobutene-butadiene isomerization using a semiempirical molecular orbital method. *J Am Chem Soc*. 1972 Apr 1;94(8):2625–33.

## 6. Prilozi

### Prilog 1:

---

““““This script obtains a pdb-formatted list of consensus water sites from several xray structures pre-prepared in VMD.

The script takes as input a directory waters which contains a list of VMD pdb files with waters (oxygens) of interest. These should be obtained from aligned xray structures in VMD, by saving selections, f.e. (water and within 5 of chain A) and (water and within 5 of chain B)

Consensus waters are defined as in the WaterDock paper: Waters within 1A of each other occurring in two different X-ray structures““““

```
import math
import cmath
import random
import sys
import os
from collections import defaultdict

def dist_calc(a, b):
    dist = math.sqrt ((a[0]-b[0])**2 + (a[1] - b[1])**2 + (a[2] - b[2])**2)
    return dist

waters = defaultdict(list)
consensus_waters = defaultdict(list)

water_list = os.listdir('waters/')

for f in water_list:
    pdb = open('waters/'+f, 'r')
    for line in pdb:
        if "ATOM" in line:
            frill = ''.join(line.split())
            xCA = float(frill.split(" ")[6])
            yCA = float(frill.split(" ")[7])
            zCA = float(frill.split(" ")[8])
            t = [xCA, yCA, zCA]
            waters[f].append(t)
    pdb.close()

# print waters

for orig in waters.keys():
    for goal in waters.keys():
        if orig != goal:
```

```

    for i in range(len(waters[orig])):
        for j in range(len(waters[goal])):
            d = dist_calc(waters[orig][i], waters[goal][j])
            if d < 1.0 :
                if waters[orig][i] not in consensus_waters[orig]:
                    consensus_waters[orig].append(waters[orig][i])
waters.pop(orig)

print consensus_waters

water_sites = open("water_sites", 'w')
for tel in consensus_waters.keys():
    for i in range(len(consensus_waters[tel])):
        print >> water_sites, "ATOM %5i  0 DEN H  1  %8.3f%8.3f%8.3f      " %
(i, consensus_waters[tel][i][0], consensus_waters[tel][i][1], consensus_waters[tel][i][2])
water_sites.close()

```

---

## Prilog 2:

---

```

import os
import math

def dist_calc(voxel, seed):
    dist = math.sqrt ((voxel[1] - seed[1])**2 + (voxel[2] - seed[2])**2 + (voxel[3] -
seed[3])**2)
    return dist

R = 0.001985877534 #gas constant! kcal/K*mol
T = 300 # temperature, K

#n_zero = 0.0001775 # per 0.125 A3 (0.5**3)
n_zero = 0.004025125 # for pure water 100 A box 96603 atoms per 0.125A3
gridunit = 0.5

r_water = 1.4

n_cutoff = 2*n_zero # Twice the occupancy of bulk water

voxellist = []
neighbourhood = []
site_list = []

### make a list of all voxels

```

```

##
##voxelpdb = open("water.pdb", 'r')
##for line in voxelpdb:
##  if "ATOM" in line:
##    frill = ''.join(line.split())
##    x = float(frill.split(" ")[6])
##    y = float(frill.split(" ")[7])
##    z = float(frill.split(" ")[8])
##    occ = float(frill.split(" ")[10])
##    voxellist.append([occ, x, y, z])
##voxelpdb.close()
##
##print "Made voxellist..."
##
### find neighbours....
##avg_voxel = open('avg_vox.pdb', 'w')
##for voxel in voxellist:
##  neighbourhood.append(voxel)
##  for bypass in voxellist:
##    if voxel != bypass:
##      if (voxel[1] + gridunit >= bypass[1] >= voxel[1] - gridunit) and (voxel[2] +
gridunit >= bypass[2] >= voxel[2] - gridunit) and (voxel[3] + gridunit >= bypass[3] >=
voxel[3] - gridunit):
##        neighbourhood.append(bypass)
##  avg_occ = 0
##  # fail safe for dxtuber not printing out 0 occupied voxels:
##  while len(neighbourhood) < 27:
##    neighbourhood.append([0, "a", "b", "c"])
##  for i in range(0, len(neighbourhood)):
##    avg_occ = avg_occ + neighbourhood[i][0]
##  avg_occ = avg_occ/len(neighbourhood)
##  voxel[0] = avg_occ
##  neighbourhood = []
##  # write to file
##  print >> avg_voxel, str(voxel)
##avg_voxel.close()

print "Averaged over neighbours..."

voxelpdb = open('avg_vox.pdb', 'r')
for line in voxelpdb:
  boo = line.replace("[", "").replace("]", "").split(',')
  flo_boo = [float(i) for i in boo]
  voxellist.append(flo_boo)
voxelpdb.close()

# Calculate dG binding energy:

```



```

for voxel in voxellist:
    bind_energy = - R*T*math.log(voxel[0]/n_zero)
    voxel.append(bind_energy)
    bind_energy = 0

print "Calculated dG..."

voxellist.sort(key=lambda x: x[4])
i = 0
while 1:
    if len(voxellist) <= i:
        break
    seed = voxellist[i]
    for voxel in voxellist:
        # radii of water 1.4
        if voxel != seed and dist_calc(voxel, seed) <= r_water:
            voxellist.remove(voxel)
    i = i + 1
print len(voxellist)
for voxel in voxellist:
    if voxel[0] >= n_cutoff:
        site_list.append(voxel)

print "Clustered. Survivors:"
print len(site_list)

probe_spots = open('water_spots.pdb', 'w')
for i in range(len(site_list)):
    print >> probe_spots, "ATOM %5i  O DEN H 1  %8.3f%8.3f%8.3f 1.00%12.6f
D" %(i, site_list[i][1], site_list[i][2], site_list[i][3], site_list[i][4])
probe_spots.close()

```

---

### Prilog 3:

---

```

#script to postprocess EPOS results to find pockets involving interface residues
# requires a directory of subpockets and inputing a selection

import os

pocket_list = os.listdir('subpockets/')

selection = [8, 9, 10, 11, 12, 13, 57, 58, 59, 61, 64, 67, 68, 71, 81, 82, 83, 84, 86, 99]
hotspot_interact = [13, 59, 67, 71, 81, 82, 83, 84, 99]

interface_pockets = []
hotspot_pockets = []

```

```

for f in pocket_list:
    pdb = open('subpockets/'+f, 'r')
    for line in pdb:
        frill = ''.join(line.split())
        resnum = frill.split(" ")[5]
        if int(resnum) in selection:
            if str(f) not in interface_pockets:
                interface_pockets.append(str(f))
            if int(resnum) in hotspot_interact:
                if str(f) not in hotspot_pockets:
                    hotspot_pockets.append(str(f))
    pdb.close()

results = open("results", 'w')
print >> results, " "
print >> results, "INTERFACE POCKETS"
print >> results, " "
for i in sorted(interface_pockets):
    print >> results, i
print >> results, " "
print >> results, "HOTSPOT POCKETS"
print >> results, " "
for i in sorted(hotspot_pockets):
    print >> results, i
results.close()

```

---

#### Prilog 4:

---

```

# Script for selecting random negative control surface residues for
# drugability estimate of PPI surface on proteins, as per PLOS paper
# Druggable Protein Interaction Sites Are More Predisposed to Surface
# Pocket Formation than the Rest of the Protein Surface

```

```

import os
import math

```

```

def dist_calc(a, b):
    dist = math.sqrt ((a[1]-b[1])**2 + (a[2] - b[2])**2 + (a[3] - b[3])**2)
    return dist

```

```

interface_resnums = [8, 9, 10, 11, 12, 13, 57, 58, 59, 61, 64, 67, 68, 71, 81, 82, 83, 84,
86, 99] # INPUT NEEDED LIST!
protein = "test.pdb"

```

```

interface_data = []
residue_data = []
filt_resnums = []
interface_full = []
filt_res_full = []
selected_residues = []

pdb = open(protein, "r")
for line in pdb:
    if "ATOM" in line and "CA" in line:
        frill = ''.join(line.split())
        resnum = frill.split(" ")[5]
        xCA = float(frill.split(" ")[6])
        yCA = float(frill.split(" ")[7])
        zCA = float(frill.split(" ")[8])
        t = [resnum, xCA, yCA, zCA]
        if int(resnum) in interface_resnums:
            interface_data.append(t)
        else:
            residue_data.append(t)
pdb.close()

# eliminate residues closer than 12A to interface CA pre-filter
for int_res in interface_data:
    for resid in residue_data:
        dist = dist_calc(resid, int_res)
        if dist <= 12.0:
            residue_data.remove(resid)

# find independants
for resid in residue_data:
    for glim in residue_data:
        if resid != glim:
            dist = dist_calc(resid, glim)
            if dist <= 12.0:
                residue_data.remove(glim)

# fetch surviving resids
for resid in residue_data:
    filt_resnums.append(int(resid[0]))

# go back to protein for in-depth check:
pdb = open(protein, "r")
for line in pdb:
    if "ATOM" in line:
        frill = ''.join(line.split())

```

```

resnum = frill.split(" ")[5]
xCA = float(frill.split(" ")[6])
yCA = float(frill.split(" ")[7])
zCA = float(frill.split(" ")[8])
t = [resnum, xCA, yCA, zCA]
if int(resnum) in interface_resnums:
    interface_full.append(t)
elif int(resnum) in filt_resnums:
    filt_res_full.append(t)
pdb.close()

# check all distances again
for int_res in interface_full:
    for resid in filt_res_full:
        dist = dist_calc(resid, int_res)
        if dist <= 12.0:
            wrong_resnum = resid[0]
            for atom in filt_res_full:
                if atom[0] == wrong_resnum:
                    filt_res_full.remove(atom)

# find independants

for resid in filt_res_full:
    for glim in filt_res_full:
        if resid != glim:
            dist = dist_calc(resid, glim)
            wrong_resnum = glim[0]
            if dist <= 12.0:
                for atom in filt_res_full:
                    if atom[0] == wrong_resnum:
                        filt_res_full.remove(atom)

# print out results:
for res in filt_res_full:
    if res[0] not in selected_residues:
        selected_residues.append(res[0])

results = open("selected_residues", "w")
for i in sorted(selected_residues):
    print >> results, str(i) + " ",
results.close()

```

---

## 7. Biografija autora

Jelena Randelović rođena je 21.10.1986. u Paraćinu. Završila je srednju medicinsku školu u Kragujevcu kao đak generacije. Školske 2001/02. godine osvojila je prvo mesto na IV Saveznom takmičenju u poznavanju anatomije i fiziologije održanom u Sremskoj Mitrovici, a 2005. prvo mesto na Republičkom takmičenju iz hemije - test i eksperimentalne vežbe u Užicu, za učenike srednjih škola. Pohađala je seminare Istraživačke stanice Petnica iz hemije i fizike, astronomije i elektronike.

Farmaceutski fakultet u Beogradu, smer diplomirani farmaceut, upisala je 2005. godine, i diplomirala avgusta 2010. godine, sa prosečnom ocenom 9.68. U toku studija učestvovala je u izradi studentskog istraživačkog rada na Katedri za organsku hemiju. Od školske 2002/03. do 2009/10. godine bila je stipendista Republičke fondacije za razvoj naučnog i umetničkog podmlatka.

Oktobra 2010. upisala je doktorske akademske studije, modul farmaceutska hemija na Farmaceutskom fakultetu, Univerzitet u Beogradu, kao stipendista Ministarstva prosvete, nauke i tehnološkog razvoja Republike Srbije. Tokom doktorskih studija učestvovala je u realizaciji praktične nastave na Katedri za toksikologiju akademik *Danilo Soldatović* (2010–2013.) i Katedri za farmaceutsku hemiju (zimski semestar školske 2011/12.)

Прилог 1.

## Изјава о ауторству

Потписани-а Јелена Ранђеловић  
број индекса 38/10

Изјављујем

да је докторска дисертација под насловом

Дизајнирање пептидних и непептидних инхибитора протеинских интеракција  
циклин-зависне киназе 9 и циклина Т1 применом рачунарских метода

- резултат сопственог истраживачког рада,
- да предложена дисертација у целини ни у деловима није била предложена за добијање било које дипломе према студијским програмима других високошколских установа,
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио интелектуалну својину других лица.

У Београду, 12. 8. 2013.

Потпис докторанда

Ј. Ранђеловић

Прилог 2.

## Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора Јелена Ранђеловић

Број индекса 38/10

Студијски програм докторске академске студије, модул: фармацеутска хемија

Наслов рада Дизајнирање пептидних и непептидних инхибитора протеинских интеракција циклин-зависне киназе 9 и циклина T1 применом рачунарских метода

Ментор др Владимир Савић, др Славица Ерић

Потписани/а Јелена Ранђеловић

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла за објављивање на порталу Дигиталног репозиторијума Универзитета у Београду.

Дозвољавам да се објаве моји лични подаци везани за добијање академског звања доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

У Београду, 12. 8. 2013.

Потпис докторанда

Ј. Ранђеловић

Прилог 3.

## Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

Дизајнирање пептидних и непептидних инхибитора протеинских интеракција циклин-зависне киназе 9 и циклина Т1 применом рачунарских метода

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигитални репозиторијум Универзитета у Београду могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство
2. Ауторство - некомерцијално
3. Ауторство – некомерцијално – без прераде
4. Ауторство – некомерцијално – делити под истим условима
5. Ауторство – без прераде
6. Ауторство – делити под истим условима

У Београду, 12. 8. 2013.

Потпис докторанда

У. Панђелковић