

Kemijski odsjek, Prirodoslovno-matematički fakultet,
Sveučilište u Zagrebu

Molekulsko modeliranje biomakromolekula

Branimir Bertoša,
Andrea Hloušek-Kasun, Natalia Mrnjavac

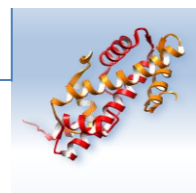


Zagreb, 2020.

Sadržaj

Uvod	1
1. Vizualizacija 3D strukturâ bioloških makromolekula	3
Vježba 1: Pretraživanje baze PDB (<i>Protein Data Bank</i>).....	6
Vježba 2: Vizualizacija struktura nukleotida	7
Vježba 3: Vizualizacija strukture hemoglobina	10
2. Polje sila	13
Vježba 4: Upoznavanje strukture polja sila	18
Vježba 5: Parametrizacija molekula u programskom paketu <i>Amber</i>	19
3. Kompletiranje i protoniranje bioloških makromolekula	22
Vježba 6: Protoniranje esteraze EstA iz bakterije <i>Pseudomonas aeruginosa</i> korištenjem programa <i>What If</i>	24
Vježba 7: Kompletiranje strukture i protoniranje ljudske N-acetil-glukozamin kinaze korištenjem programa <i>SWISS-MODEL</i>	25
4. Solvatacija, periodični rubni uvjeti i elektroneutralnost	29
Vježba 8: Priprema 3D strukture za simulaciju molekulske dinamike (MD)	311
Vježba 8.1 Priprema 3D strukture za MD simulaciju, solvatacija i periodični rubni uvjeti u programu <i>Gromacs</i>	32
Vježba 8.2 Priprema 3D strukture za MD simulaciju, solvatacija i periodični rubni uvjeti u programu <i>Amber</i>	35
5. Minimizacija energije (optimizacija geometrije)	39
Vježba 9.1 Minimizacija energije (optimizacija geometrije) programom <i>Gromacs</i> ..	40
Vježba 9.2 Minimizacija energije (optimizacija geometrije) programom <i>Amber</i>	42
6. Priprema i pokretanje simulacije molekulske dinamike (MD)	45
Vježba 10: Simulacije uravnoteženja (ekvilibracije)	47
Vježba 10.1 Simulacije uravnoteženja programom <i>Gromacs</i>	47
Vježba 10.2 Simulacije uravnoteženja programom <i>Amber</i>	49
Vježba 11: Simulacija molekulske dinamike (MD).....	51
Vježba 11.1 Simulacije molekulske dinamike programom <i>Gromacs</i>	51
Vježba 11.2 Simulacije molekulske dinamike programom <i>Amber</i>	52
7. Analiza simulacije molekulske dinamike (MD)	55
Vježba 12: Vizualizacija MD trajektorije	58
Vježba 12.1 Vizualizacija MD trajektorije dobivene programom <i>Gromacs</i>	58
Vježba 12.2 Vizualizacija MD trajektorije dobivene programom <i>Amber</i>	60
Vježba 13: Definiranje poskupova sustava	62
13.1 Definiranje podskupova sustava u programu <i>Gromacs</i>	62
Vježba 13.2 Definiranje podsustava u programu <i>Amber</i>	63
Vježba 14: Izračun <i>RMSD</i> vrijednosti	65
Vježba 14.1 Izračun <i>RMSD</i> vrijednosti u programu <i>Gromacs</i>	65

Vježba 14.2 Izračun <i>RMSD</i> vrijednosti u programu <i>Amber</i>	66
Vježba 15: Izračun <i>RMSF</i> vrijednosti	69
Vježba 15.1 Izračun <i>RMSF</i> vrijednosti u programu <i>Gromacs</i>	69
Vježba 15.2 Izračun <i>RMSF</i> vrijednosti u programu <i>Amber</i>	70
Vježba 16: Izračun radijusa giracije.....	72
Vježba 16.1 Izračun radijusa giracije u programu <i>Gromacs</i>	72
Vježba 16.2 Izračun radijusa giracije u programu <i>Amber</i>	73
Vježba 17: Mjerenje udaljenosti središta masa	74
Vježba 17.1 Mjerenje udaljenosti središta masa u programu <i>Gromacs</i>	74
Vježba 17.2 Mjerenje udaljenosti središta masa u programu <i>Amber</i>	75
Vježba 18: Analitička metoda po izboru	76
Vježba 18.1 Analitička metoda po izboru u programu <i>Gromacs</i>	76
Vježba 18.2 Analitička metoda po izboru u programu <i>Amber</i>	77
8. Molekulsko uklapanje (<i>docking</i>).....	78
Vježba 19: Molekulsko uklapanje	80
9. Dodatak.....	88
9.1 Izrada skripte za parametrizaciju sustava potprogramom <i>tleap</i>	88
9.2 Pokretanje paralelnih simulacija na više procesora.	89
9.3 Pokretanje simulacija na grafičkoj kartici	90
9.4 Izrada <i>cpptraj.in</i> skripte	91
10. Popis literature	92



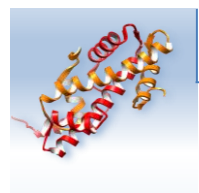
Uvod

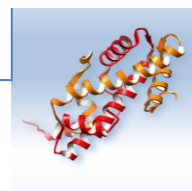
Cilj ovog nastavnog priručnika jest upoznati čitatelja s osnovama molekuskog modeliranja bioloških makromolekula (biomakromolekula). Priručnik sadrži teorijsku osnovu te detaljne upute za pripremu, pokretanje i analizu računalnih simulacija prilagođenih izučavanju molekulskih sustava, s naglaskom na izučavanje biomakromolekula.

Priručnik je organiziran u 8 poglavlja te sadrži dodatak i popis literature. Svako poglavlje započinje teorijskom osnovom koja je neophodna za pravilnu uporabu računalnih tehnika predstavljenih u danom poglavlju, kao i za adekvatnu interpretaciju dobivenih rezultata. Nakon teorijske osnove slijede vježbe s detaljnim uputama za njihovo izvođenje koje korisnika obučavaju za praktičnu primjenu predstavljene računalne tehnike. Na kraju svake vježbe su zadaci koji služe za provjeru usvojenih znanja i razumijevanja dobivenih rezultata.

U svom prvom izdanju priručnik je izvorno nastao kao nastavni priručnik u sklopu kolegija Strukturna računalna biofizika. U svom drugom izdanju i dalje je zadržao svoju osnovnu namjenu, ali je proširenim sadržajem namijenjen svim studentima diplomskog studija Kemije i diplomskog studija Molekularne biologije koji su zainteresirani za molekulska modeliranje biomakromolekula. Shodno tome, priručnik je prvenstveno fokusiran na biofizičke, odnosno biokemijske sustave, ali upute za računalne metode obuhvaćene priručnikom mogu se iskoristiti i za njihovu primjenu na izučavanje drugih sustava. Biomakromolekule koje se u priručniku koriste su: esteraza EstA iz bakterije *Pseudomonas aeruginosa*, ljudska N-acetil-glukozamin kinaza, heksokinaza iz organizma *Sulfolobus tokodaii*, hemoglobin te nukleotidi koji grade DNA. Računalni programi koji se koriste u sklopu priručnika većinom su besplatni i dostupni studentima. Neke od programa potrebno je instalirati na lokalno računalo, poput programa: *VMD*, *Grace*, *Gromacs*, *Amber*, *AutoDock* i *AutoDock Vina*. Drugi su mrežno dostupni, poput programa: *What If*, *H++*, *MolProbity* i *SWISS-MODELLER*.

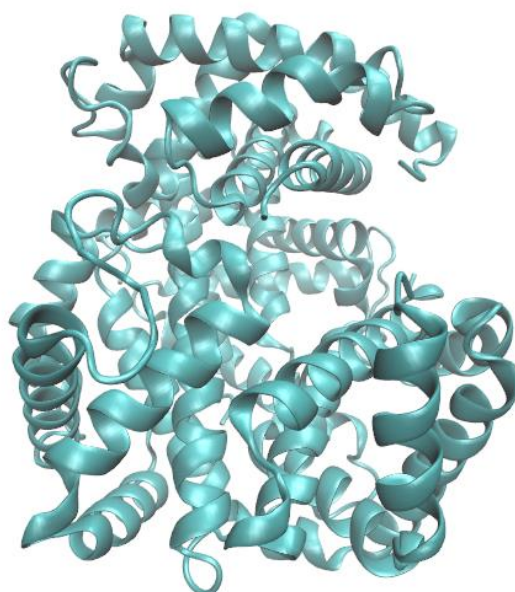
Na kraju priručnika nalazi se popis literature te dodatak sa skriptama koje korisniku mogu olakšati korištenje predstavljenih programa, kao i upute za izvođenje računa na grafičkoj kartici i paralelno povezanim procesorima.



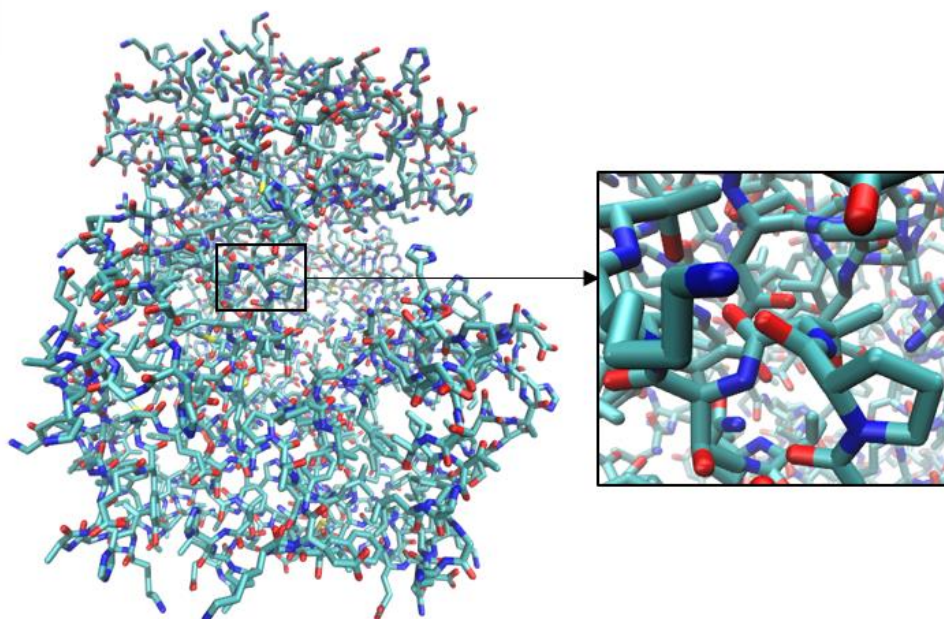
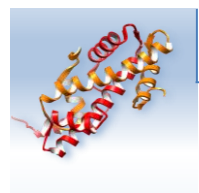


1. Vizualizacija 3D strukturâ bioloških makromolekula

Vizualizacija sustava koji je predmet istraživanja vrlo je važan korak u računalnom istraživanju kemijskih, biokemijskih i bioloških sustava. Odabir prikaza modela molekula koje izučavamo može značajno utjecati na našu percepciju tog sustava, te stoga može utjecati i na: odabir računalne metode, pripremu sustava za simulaciju, analizu simulacije pa čak i na interpretaciju rezultata simulacije. Utjecaj vizualizacije osobito je izražen kod kompleksnih sustava, poput bioloških makromolekula. Stoga je važno izabrati program za vizualizaciju koji je jednostavan za korištenje, ali pruža dovoljno različitih opcija koje je neophodno kombinirati kako bi se ostvarila odgovarajuća vizualizacija sustava. Primjerice, prikaz samo sekundarnih strukturâ proteina modelom vrpce (engl. *Ribbons*) (slika 1) omogućit će adekvatnu percepciju cjelokupne strukture, ali neće dopustiti uvid u detalje interakcija između aminokiselinskih ostataka. S druge strane, prikaz koji uključuje sve atome i veze među njima, omogućava jasan uvid u interakcije između aminokiselinskih ostataka, ali otežava percepciju cjelokupne strukture makromolekule (slika 2).



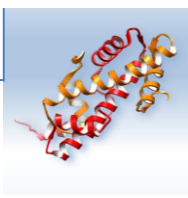
Slika 1. Katalitička domena esteraze EstA iz bakterije *Pseudomonas aeruginosa*.



Slika 2. Katalitička domena esteraze EstA iz bakterije *Pseudomonas aeruginosa*, u izdvojenom okviru uvećan je dio strukture kako bi se bolje vidjele moguće interakcije između aminokiselina.

Vizualizacija modela molekula zahtijeva poznavanje njihovih trodimenzijskih (3D) koordinata. Najvjerodostojniji način dolaska do početnih 3D koordinata molekule od interesa jest korištenje struktura dobivenih eksperimentalnim metodama, poput kristalografije ili NMR spektroskopije. U slučaju malih molekula, odnosno molekula s nekoliko desetaka ili stotinjak atoma, početne 3D modele moguće je i izgraditi *in silico* bez upotrebe eksperimentalnih podataka. Naime, postoji mnoštvo programa koji korisniku omogućavaju da na vrlo jednostavan i intuitivan način, uz malo kemijskog znanja, *ab ovo* izgradi modele molekula koji su kemijski korektni i mogu poslužiti kao osnova za pripremu računalnih simulacija. Međutim, u slučaju bioloških makromolekula takav bi pothvat bio odveć složen i gotovo nemoguć. U slučaju bioloških makromolekula neophodno je osloniti se na eksperimentalne podatke kako bi se generirao početni 3D model strukture.

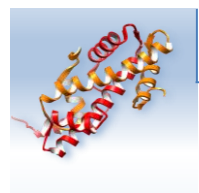
Baza PDB (*Protein Data Bank*) sadrži koordinate 3D struktura bioloških makromolekula dobivenih različitim eksperimentalnim metodama. Baza je besplatna i može joj se pristupiti preko poveznice: <https://www.rcsb.org/>. Svaka struktura pohranjena u bazi PDB dobiva svoju jedinstvenu oznaku koja se naziva PDB kôd i sastoji se od četiri znaka koji su kombinacija slova i brojeva. Baza PDB



ima vrlo intuitivno sučelje koje omogućava višestruke mogućnosti pretraživanja pohranjenih struktura. Osim po PDB kôdu, strukturu biološke makromolekule moguće je pronaći po imenu makromolekule, po ligandu koji je prisutan u strukturi, po autorima koji su strukturu riješili, po aminokiselinskoj sekvenci, itd. Datoteka u kojoj je pohranjena struktura u bazi PDB ima ekstenziju .pdb (dalje u tekstu .pdb datoteka). Osim prostornih koordinata strukture molekule (slika 3), .pdb datoteka sadrži i mnoštvo informacija vezanih uz rješavanje strukture, poput metode kojom je određena struktura, rezolucije na kojoj je riješena, publikacije u kojoj je objavljena i sl. Detaljnije informacije o strukturi .pdb datoteke mogu se pronaći na poveznici: <http://www.wwpdb.org/documentation/file-format>.

	BROJ ATOMA	TIP ATOMA	AMINOKISELINA	LANAC	BROJ AMINOKISELINE U LANCU	KOORDINATE			OKUPIRANOST	B-FAKTOR	KEMIJSKI ELEMENT
ATOM	1	N	MET	A	1	10.285	19.678	85.047	1.00	47.22	N
ATOM	2	CA	MET	A	1	11.717	19.659	84.640	1.00	41.14	C
ATOM	3	C	MET	A	1	11.864	19.618	83.121	1.00	38.13	C
ATOM	4	O	MET	A	1	11.173	20.336	82.399	1.00	38.82	O
ATOM	5	CB	MET	A	1	12.423	18.445	85.244	1.00	40.91	C
ATOM	6	CG	MET	A	1	11.806	17.111	84.846	1.00	42.21	C
ATOM	7	SD	MET	A	1	13.051	15.849	84.539	1.00	26.57	S
ATOM	8	CE	MET	A	1	13.674	15.607	86.199	1.00	47.33	C
ATOM	9	N	MET	A	2	12.769	18.763	82.652	1.00	31.13	N
ATOM	10	CA	MET	A	2	13.039	18.603	81.228	1.00	29.47	C

Slika 3. Struktura .pdb datoteke.



Vježba 1: Pretraživanje baze PDB (*Protein Data Bank*)

Pretragom baze PDB potrebno je pronaći strukture: (i) esteraze EstA iz bakterije *Pseudomonas aeruginosa* i (ii) ljudske N-acetil-glukozamin kinaze, a zatim pronađene .pdb datoteke spremiti na lokalno računalo. Datoteke je moguće otvoriti tekstualnim editorom te različitim programima za vizualizaciju. Za početak je potrebno pronađene strukture otvoriti samo tekstualnim editorom.

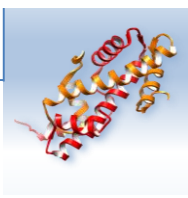
Zadatak 1.1 Pronaći .pdb datoteke esteraze EstA iz bakterije *Pseudomonas aeruginosa* i ljudske N-acetil-glukozamin kinaze (NAG kinaze) u kompleksu s N-acetil-glukozaminom (NAG) u bazi PDB i odgovoriti na zadana pitanja.

Upute:

Preko mrežne poveznice <https://www.rcsb.org/> može se pristupiti bazi PDB te upisivanjem imena tražene makromolekule pronaći dostupne .pdb datoteke. Pronađenu .pdb datoteku je potrebno spremiti na lokalno računalo i otvoriti u tekstualnom editoru (primjerice, u *Linux* operativnom sustavu mogu se koristiti *gedit* ili *vim*, a u *Windows* operativnom sustavu *notepad* ili *wordpad*) kako bi se pronašlo odgovore na zadana pitanja.

Pitanja:

1. Koji je PDB kôd datoteke koja sadrži strukturu esteraze EstA iz bakterije *Pseudomonas aeruginosa*, a koji datoteke koja sadrži strukturu ljudske NAG kinaze u kompleksu s NAG?
2. Kojom metodom i pri kojoj rezoluciji je riješena struktura: a) esteraze EstA iz bakterije *Pseudomonas aeruginosa*, b) ljudske NAG kinaze u kompleksu s NAG?
3. Koje su reference publikacija u kojima su navedene strukture objavljene?
4. Koje aminokiseline nedostaju u svakoj od navedenih struktura, odnosno njihove koordinate nisu određene rendgenskom strukturnom analizom?
5. Ima li koji od navedenih proteina disulfidni most? Ako ima, između kojih aminokiselina je ostvaren?



Vježba 2: Vizualizacija struktura nukleotida

Cilj vježbe je vizualizirati i spariti odgovarajuće parove nukleotida te prikazati vodikove veze koje se među njima ostvaruju. Za vizualizaciju molekula u sklopu ovih uputa koristit će se program *VMD (Visual Molecular Dynamics)* [1] koji je besplatan i vrlo rasprostranjen u znanstvenoj zajednici.

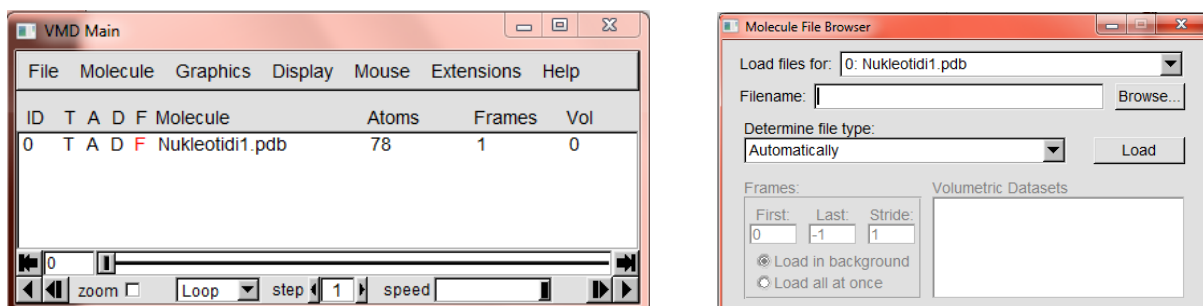
Zadatak 1.2 Prikazati strukture zadanih nukleotida modelom kuglica i štapića (*CPK model*) te ih spariti na način da ostvare maksimalan broj vodikovih veza.

Upute:

Najprije je potrebno pokrenuti program *VMD* i učitati strukture nukleotida (datoteka: *Nukleotidi1.pdb*). Program *VMD* se pokreće upisivanjem u terminal naredbe:

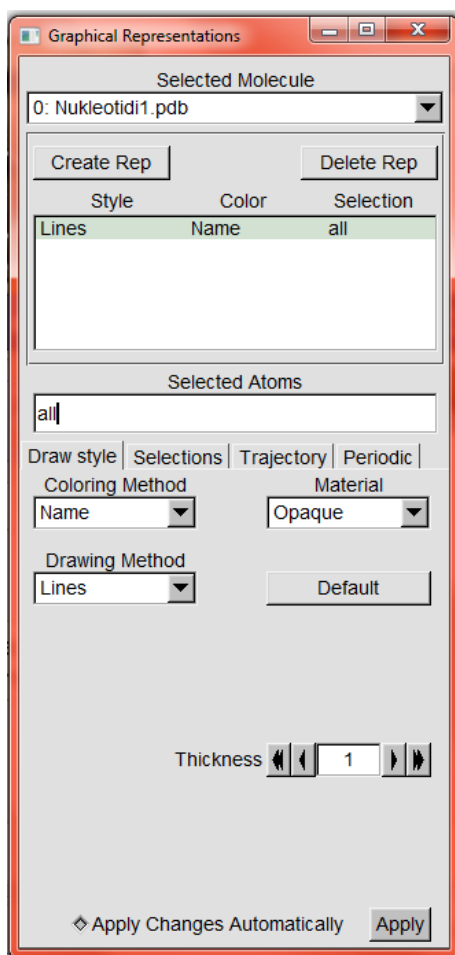
vmd

U prozoru pod nazivom ***VMD Main*** potrebno je otvoriti prozor za novu molekulu (***File*** → ***New Molecule***), kliknuti na ***Browse*** te pronaći datoteku *Nukleotidi1.pdb* i odabrati ***Load*** (slika 4).

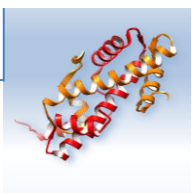


Slika 4. Osnove grafičkog sučelja programa *VMD*. Lijevo prozor „***VMD Main***“ iz kojeg se mogu otvoriti svi ostali prozori, desno prozor „***Molecule File Browser***“.

Strukture nukleotida mogu se prikazati različitim molekulskim prikazima. Za promjenu načina prikaza molekule potrebno je u glavnom prozoru **VMD Main** kliknuti na **Graphics** i izabrati **Representations**. Otvorit će se prozor **Graphical Representations** (slika 5) u kojem se nalazi padajući izbornik **Drawing Method** koji omogućava odabir prikaza modela molekule (npr. *Lines*, *Bonds*, *CPK*, *VDW* ...). U padajućem izborniku nazvanom **Coloring Method** može se odabrati način bojenja atoma. Novi se prikaz, uz zadržavanje postojećeg, postiže opcijom **Create Rep**, a briše se opcijom **Delete Rep**. U prostor **Selected Atoms** upisuje se dio strukture koji želimo prikazati ili za koji želimo izabrati željeni prikaz (npr. *all*, *protein*, *resid 1*, *resname GLU*, itd.).



Slika 5. Grafičko sučelje **Graphical representations** programa *VMD* u kojem se odabiru prikazi modela molekula.



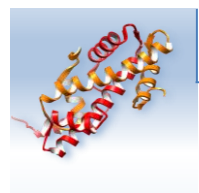
Za sparivanje nukleotida putem vodikovih veza potrebno je uključiti prikaz vodikovih veza te rotirati strukturu jednog nukleotida u odnosu na drugi.

Vodikove veze su vidljive samo ako postoje u sustavu te ukoliko je izabran prikaz **Hbonds** u izborniku **Drawing Method**. Debljina prikaza vodikovih veza može se mijenjati opcijom **Line Thickness**. Ukoliko je prikaz vodikovih veza jedini uključeni prikaz, atomi molekule nisu prikazani te je stoga uz vodikove veze neophodno izabrati i još jedan način prikaza korištenjem **Create Rep** opcije.

U programu *VMD* početno je uključena **rotacija** (pokazivač miša ima oblik strelice) čitavog sustava. Pritiskom slova „C” na tipkovnici pokazivač miša poprima oblik križa te omogućava odabir atoma koji će se nalaziti u **središtu rotacije**. Pritiskom na slovo „R” ponovo se uključuje rotacija. Pritiskom na slovo „T” uključuje se **translacija** sustava (pokazivač miša ima oblik ukrštenih strelica). Uključivanje **translacije jedne molekule u odnosu na drugu** postiže se pritiskom na slovo „R”, a odmah zatim na broj „7” na tipkovnici (pokazivač miša dobit će oblik znaka plus). Zatim je potrebno kliknuti na atom molekule koju se želi translirati. Sve dok je lijeva tipka miša stisnuta, odabranu molekulu je moguće translirati u odnosu na ostale molekule. Rotacija jedne molekule u odnosu na drugu postiže se tako što se drži pritisnuta tipka „SHIFT” u načinu rada koji se koristi za translaciju jedne molekule u odnosu na drugu.

Pitanja:

1. Koliko je vodikovih veza postignuto sparivanjem zadanih nukleotida?
2. Napraviti slike sparenih nukleotida s označenim vodikovim vezama, pri čemu je potrebno isprobati barem tri različita prikaza modela molekula.
3. Ponoviti isti postupak za sparivanje nukleotida s bazama adeninom i timinom, potrebne koordinate nalaze se u datoteci Nukleotidi2.pdb, te odgovoriti na prva dva pitanja u tom slučaju.



Vježba 3: Vizualizacija strukture hemoglobina

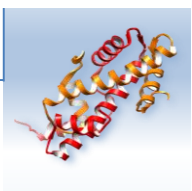
Vizualizacija proteinskih struktura puno je zahtjevnija od vizualizacije malih molekula, poput modela molekula nukleotida koji su korišteni u prošloj vježbi. Zbog kompleksnosti proteinske strukture te mnoštva atoma koji ju grade, potrebno je kombinirati različite prikaze modela molekula kako bi se struktura vizualizirala na odgovarajućoj razini. U ovoj vježbi vizualizirat će se struktura hemoglobina, na način da se promotre elementi sekundarne strukture, ali i vezanje molekule kisika na prostetičku skupinu.

Zadatak 1.3 Vizualizirati strukturu hemoglobina na način da se prikažu elementi sekundarne strukture, te izmjeriti strukturne parametre koji opisuju vezanje molekule kisika.

Upute:

Pokrenuti program *VMD* i učitati 3D strukturu hemoglobina (datoteka: Hemoglobin.pdb) i promijeniti molekularni prikaz hemoglobina. Za mijenjanje molekularnog prikaza potrebno je u glavnom prozoru **VMD Main**, u izborniku **Graphics**, odabrati **Representations** (**VMD Main** → **Graphics** → **Representations**). Zatim u padajućem izborniku **Drawing Method** izabrati odgovarajući prikaz proteinskog dijela molekule. Svaki prikaz može se doraditi korištenjem dodatnih opcija (npr. za prikaz *NewCartoon* može se izabrati transparentnost materijala, debljina prikaza i sl.).

Za vizualizaciju elemenata sekundarne strukture hemoglobina potrebno je izabrati odgovarajući prikaz (npr. *NewCartoon*, *Cartoon*, *Ribbons*, itd.) i primijeniti ga na izabrani dio strukture (u prostor **Selected Atoms** upisuje se „protein“). Za vizualizaciju prostetičke skupine potrebno je izabrati novi prikaz korištenjem opcije **Create Rep** te upisati: „**resname HEM**“, u prostor za odabir (**Selected Atoms**), a zatim izabrati odgovarajući prikaz (npr. *Bonds*, *Licorice*, *CPK*, itd.). Vizualizacija molekule kisika ostvaruje se upisivanjem: „**resname OXY**“ u prostor za odabir atoma novog prikaza te odabirom odgovarajućeg molekularnog prikaza. Važno je napomenuti da u liniji za upis selekcija funkcioniraju osnovni Booleovi operatori, a ukoliko se želi nabrojati više argumenata za istu ključnu riječ (*resname*, *resid*, *name*, itd.), dovoljno je između njih ostaviti razmak. Primjerice,

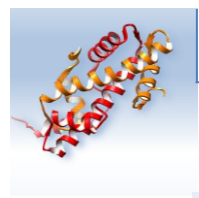


prikaz svih histidina u proteinskoj strukturi hemoglobina može se postići upisivanjem „**resname HIS**“, u prostor za odabir atoma. Prikaz svih atoma koji su na udaljenosti manjoj ili jednakoj 5 Å od prostetičke skupine hema postiže se upisivanjem „**within 5 of resname HEM**“, u prostor za odabir atoma. Upisivanjem „**residue as within 5 of resname HEM**“, prikazuju se sve aminokiseline koje imaju barem jedan atom na udaljenosti manjoj ili jednakoj 5 Å od prostetičke skupine hemoglobina. Prikaz samo histidina koji se nalaze na udaljenosti manjoj ili jednakoj 5 Å od prostetičke skupine hemoglobina postiže se upisivanjem izraza: „**resname HIS and within 5 of resname HEM**“.

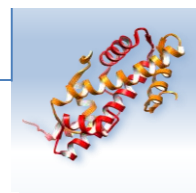
Označavanje atoma u vizualizacijskom prozoru programa *VMD* vrši se pritiskom na slovo **L**, zatim na broj **1**, nakon čega je potrebno kliknuti mišem na željeni atom. Udaljenosti između atoma u vizualizacijskom prostoru se mjere pritiskom na slovo **L**, a zatim na broj **2**, nakon čega je potrebno klikom miša izabrati dva atoma čiju udaljenost se želi izmjeriti. Pritiskom na slovo **L**, a odmah zatim na broj **3** uključuje se mjerenje kuta koji zatvaraju tri atoma. Pritiskom na slovo **L**, a odmah zatim na broj **4** uključuje se mjerenje torzijskog kuta koji zatvaraju četiri atoma. Izmjerene vrijednosti se pojavljuju u vizualizacijskom prostoru, a može ih se pronaći i tako što se u prozoru **VMD Main**, u izborniku **Graphics**, odabere **Labels (VMD Main → Graphics → Labels)** i zatim u padajućem izborniku izabere odgovarajuća opcija. U ovom izborniku mogu se odabrane vrijednosti obrisati, sakriti ili prikazati na grafu u slučaju analize trajektorije.

Pitanja:

1. Napraviti slike prikaza modela molekule hemoglobina na kojima se jasno razaznaju sekundarne strukture proteina, dok su prostetičke skupine i molekule kisike te proksimalni i distalni histidini prikazani modelom kuglica i štapića.
2. Koliko α -zavojnica ima jedna podjedinica hemoglobina?
3. Po čemu su specifične aminokiseline 141 i 289? Napraviti sliku na kojoj se te aminokiseline jasno razaznaju u odnosu na ostale.



4. Koja je prva aminokiselina N-terminalnog kraja β -podjedinice hemoglobina?
5. Kolika je udaljenost između iona željeza u prostetičkoj skupini hema i njemu bližeg atoma kisika u molekuli kisika?
6. Koliki je kut koji zatvaraju atomi molekule kisika i ion željeza?



2. Polje sila

Polje sila (engl. *force field*) zajedničko je ime za skup parametara i funkciju koja korištenjem tih parametara omogućava računanje potencijale (konformacijske) energije molekule odnosno sustava. Parametri su dobiveni eksperimentalno (polja sila prve generacije) i kvantno-mehaničkim računima (polja sila druge generacije). Ključna pretpostavka za korištenje polja sila je prenosivost parametara, odnosno pretpostavka da skup parametara izveden na određenom skupu molekula može poslužiti za opisivanje njima sličnih molekula unutar istog polja sila. Pojam **tip atoma** definira atomski broj, formalni naboj te vrste kemijskih veza dotične vrste atoma.

Funkcija kojom se računa potencijalna energija molekule sastavljena je od različitih doprinosa (1):

$$E = E_{\text{stretch}} + E_{\text{bend}} + E_{\text{tors}} + E_{\text{oop}} + E_{\text{el}} + E_{\text{vdw}} + \sum E_{\text{cross}} \quad (1)$$

Svaki doprinos u izrazu (1), osim posljednjeg, ovisi o samo jednoj vrsti interne koordinate i ima vlastitu funkciju po kojoj se računa.

E_{stretch} – potencijalna energija koja se javlja u molekuli kao posljedica odstupanja duljine veza od ravnotežnih vrijednosti:

$$E_{\text{stretch}} = \frac{k_s}{2} (l - l_0)^2 \quad (2)$$

k_s je empirijski parametar koji predstavlja konstantu sile istezanja veze, l je duljina veze, dok je l_0 referentna vrijednost za duljinu veza koja je u danom polju sila tabelirana kao optimalna. Neka polja sila, poput MM3 ili MM4 polja sila, koriste i kubični ili čak i više članove reda u izrazu (2). Osim izraza za harmonijski oscilator, za opis energije uslijed promijene duljine veza neka polja sila koriste izraz za anharmonijski potencijal:

$$E_{\text{stretch}} = D_e \left(e^{[-a(l-l_0)]} - 1 \right)^2 - D_e \quad (3)$$

u kojem je D_e energija disocijacije molekule, dok parametar a određuje strminu krivulje potencijalne energije. Izraz (3) ne koristi se često jer razlika između harmonijskog i anharmonijskog potencijala dolazi do izražaja tek pri jačem istezanju veza. Kako se metodama temeljenim na polju sila ne mogu opisivati procesi kod kojih veze pucaju ili nastaju, već su veze između atoma uglavnom blizu ravnotežne vrijednosti, izraz za harmonijski potencijal sasvim je dostatan za

opisivanje istezanja veza, a računalno je znatno manje zahtjevan od anharmonijskog.

E_{bend} – potencijalna energija koja se javlja u molekuli kao posljedica promjene valentnih kutova:

$$E_{\text{bend}} = \frac{k_b}{2} (\theta - \theta_0)^2 \quad (4)$$

u kojem je k_b empirijski parametar koji predstavlja konstantu sile promjene kuta, θ valentni kut, a θ_0 referentna vrijednost valentnog kuta.

E_{tors} – potencijalna energija koja se javlja u molekuli kao posljedica promjene torzijskih kutova:

$$E_{\text{tors}} = \sum_{n=0}^N \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \quad (5)$$

u kojem je V_n rotacijska barijera, n njen multiplicitet, ϕ vrijednost torzijskog kuta, a γ pomak u fazi. S obzirom da je promjena torzijskog kuta najčešći uzrok konformacijskih promjena, ovaj doprinos ukupnom potencijalu molekule iznimno je važan.

E_{oop} – potencijalna energija koja se javlja kod planarnih sustava uslijed odstupanja atoma iz ravnine:

$$E_{\text{oop}} = \frac{1}{2} k_\xi \xi^2 \quad (6)$$

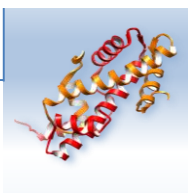
u kojem je ξ kut odstupanja iz ravnine, a k_ξ odgovarajući empirijski parametar. S obzirom da metode temeljene na polju sila ne uzimaju elektrone eksplicitno u obzir, potrebno je uvesti ovaj član kako bi se osigurala planarnost sustava kod kojih je ona uzrokovana elektronskom strukturom (npr. fenilni prsten kod fenilalanina i tirozina, indolni prsten kod triptofana, itd.).

Nevezne interakcije (elektrostatske i van der Waalsove) između atoma u molekuli računaju se samo za atome koji su odvojeni s tri ili više kovalentnih veza.

E_{el} – potencijalna energija koja se javlja u molekuli kao posljedica elektrostatskih interakcija najčešće se računa korištenjem izraza za električni potencijal točkastih naboja:

$$E_{\text{el}} = \frac{q_1 q_2}{4\pi\epsilon r_{12}} \quad (7)$$

gdje su q_1 i q_2 (parcijalni) atomski naboji, a ϵ dielektrična konstanta medija. Najslabiju kariku u ovom izrazu predstavlja upravo dielektrična konstanta koja je makroskopsko svojstvo medija. Stoga se postavlja pitanje koju vrijednost



dielektrične konstante upotrijebiti kada naboji nisu u vakuumu i nalaze se međusobno na dovoljno maloj udaljenosti da broj molekula otapala među njima nije dostatno velik da bi koristili makroskopsko svojstvo medija. Taj problem donekle je riješen upotrebom dielektrične konstante koja je prostorno ovisna (engl. *distance dependence*) te se vrši njena korekcija s obzirom na udaljenost na kojoj se naboji nalaze:

$$\varepsilon = \frac{\varepsilon_0 - \varepsilon_c}{r_{\text{eff}} - r_{\text{con}}} (r - r_{\text{con}}) + \varepsilon_c \quad (8)$$

ε_c je vrijednost dielektrične konstante (ε) pri najmanjoj udaljenosti na kojoj se naboji mogu naći (r_{con}), a ε_0 vrijednost koja se uzima kada su naboji na dovoljnoj udaljenosti da možemo koristiti makroskopsko svojstvo medija ($r \geq r_{\text{eff}}$).

E_{vdw} – potencijal koji se javlja u molekuli kao posljedica van der Waalsovih interakcija, najčešće se računa korištenjem Lennard–Jonesovog izraza:

$$E_{\text{vdw}} = \frac{A}{r^{12}} - \frac{C}{r^6} \quad (9)$$

u kojem su A i C empirijski dobiveni parametri ovisni o vrsti atoma između kojih se računa interakcija, a r udaljenost među njima. Osim Lennard–Jonesovog izraza, koriste se još i Buckinghamov izraz te Hillov izraz.

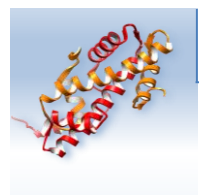
ΣE_{cross} – miješani članovi koji se javljaju zbog nemogućnosti neovisnog promatranja različitih doprinosa ukupnom potencijalu.

Rastavljanje ukupnog potencijala na zbroj različitih doprinosa prilikom čega svaki od njih ovisi samo o jednoj vrsti interne koordinate nije u potpunosti moguće jer različite interne koordinate nisu međusobno potpuno neovisne, te stoga uvijek ostaju i „miješani“ članovi koji ovise o više različitih internih koordinata. Primjerice, miješani član koji se javlja kao posljedica nemogućnosti neovisnog promatranja savijanja valentnog kuta (θ) i istezanja veza (l) koje ga tvore:

$$E_{\text{cross.S.-B.}} = \frac{k_{\text{S.B.}}}{2} [(l_1 - l_{10}) + (l_2 - l_{20})] (\theta - \theta_0) \quad (10)$$

gdje je $k_{\text{S.B.}}$ empirijski određena konstanta, l_1 i l_2 su duljine veza koje zatvaraju promatrani valentni kut, l_{10} i l_{20} su njihove optimalne vrijednosti, θ i θ_0 su valentni kut i njegova optimalna vrijednost. Takvih članova potencijala ima mnogo te se u praksi koristi aproksimacija prema kojoj se „miješani“ članovi, u potpunosti ili djelomično, zanemaruju.

Pridruživanje određenih parametara atomima u molekuli omogućava računanje potencijalne energije molekule ovisno o njezinoj konformaciji. Upravo

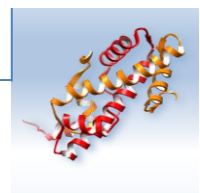


konformacijska analiza i jest glavna zadaća metoda koje se temelje na polju sila. Naime, vrijednost potencijalne energije molekule izračunata korištenjem polja sila nema izravno fizikalno značenje. Upotrebom različitih polja sila moguće je dobiti različite vrijednosti energije za istu konformaciju iste molekule, stoga vrijednosti energija izračunate različitim poljima sila nisu međusobno usporedive. Međutim vrijednosti energija izračunate za različite konformacije molekule korištenjem istog polja sila jesu međusobno usporedive te omogućuju kvantitativnu konformacijsku analizu.

Polja sila međusobno se razlikuju s obzirom na broj i vrstu atoma za koje posjeduju parametre. Postoje općenita polja sila, koja posjeduju parametre za veliki broj tipova atoma, i specijalizirana polja sila, koja su prilagođena za određenu klasu molekula za koje posjeduju vrlo precizne parametre. Specijalizirana polja sila koja se koriste za modeliranje organskih molekula su MM2, MM3, CVFF, CFF i Tripos, za biološke makromolekule (prvenstveno proteine i nukleinske kiseline) najčešće se upotrebljavaju AMBER, CHARMM, GROMOS i OPLS, dok su najpoznatija općenita polja sila MMFF, COMPASS, ESFF i UFF. Pored nabrojanih postoje i mnoga druga manje poznata polja sila od kojih neka imaju vrlo usku primjenu poput polja sila za zeolite ili polja sila za anorganske kompleksne spojeve (YETI).

Osim po broju i vrsti parametara, polja sila razlikuju se i po izrazu koji se koristi za računanje (steričke) energije molekule. Tako polja sila s kompleksnijom funkcijom trebaju manji broj parametara koji su pri tome lakše prenosivi. Složenost funkcije pridonosi preciznosti računanja energije sustava, ali i povećava vrijeme računanja. Neka polja sila veću preciznost postižu tako što nastoje uzeti u obzir čim više "miješanih" članova, što povećava vrijeme računanja. Takva su polja sila uglavnom namijenjena manjim molekulama. Polja sila koja se koriste za biopolimere uglavnom u potpunosti zanemaruju „miješane“ članove, što im omogućava relativno veliku brzinu računa s obzirom na veličinu sustava.

Svi parametri potrebni za pokretanje računalnih simulacija temeljenih na polju sila nalaze se u datoteci (ili datotekama) koja nosi ime izabranog polja sila. Pri tome postoji nekoliko podvrsta istog polja sila koja se malo razlikuju u određenim parametrima koji su s vremenom izmijenjeni ili nadodani. Tako se u programskom paketu *Gromacs* mogu naći G45a3, G43a1, G43a2, G43b1, itd., a u programskom paketu *Amber* ff03, ff99, ff99SB i druga polja sila. Razlike između nabrojanih polja sila su relativno male, ali mogu doći do izražaja pri simuliranju



određenih klasa bioloških makromolekula. Stoga je odabir adekvatnog polja sila vrlo važan korak za koji je potrebno imati određenog iskustva i konzultirati literaturu.

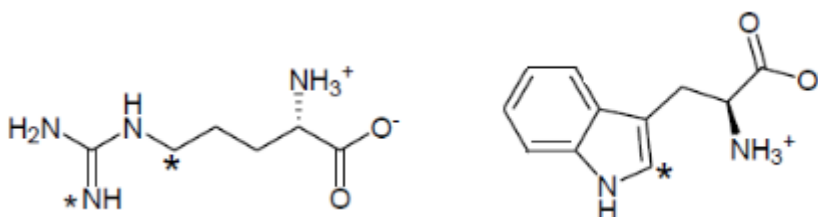
Vježba 4: Upoznavanje strukture polja sila

Cilj vježbe je upoznati se s osnovnom strukturom polja sila koja se koriste u programskim paketima *Gromacs* i *Amber*.

Zadatak 2.1 Upoznati se sa strukturom polja sila na primjeru Gromos G45a3 polja sila.

Upute:

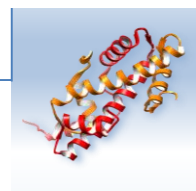
Na lokalnom računalu, u direktoriju **ffG45a3** nalaze se datoteke povezane s G45a3 poljem sila, potrebno ih je otvoriti u tekstualnom editoru te proučiti njihov sadržaj. Najprije je potrebno pronaći različite tipove atoma (napomena: svi tipovi atoma istog kemijskog elementa imaju pridruženu istu molekulska masu i grupirani su zajedno). Zatim je potrebno pronaći parametre koji opisuju istezanje veza, savijanje kutova, promjenu torzijskih kutova i nevezne interakcije.



Slika 6. Strukture arginina i triptofana, prikazani su samo polarni atomi vodika.

Pitanja:

1. Koji su tipovi atoma definirani za kisik u Gromos G45a3 polju sila?
2. Koji su tipovi atoma pridruženi zvjezdicom označenim atomima na slici 6?
3. Koje su optimalne duljine veza koje zvjezdicom označeni atomi na slici 6 ostvaruju sa susjednim atomima?
4. Kolika je konstanta sile za savijanje valentnog kuta šesteročlanog prstena triptofana?
5. Koji su naboji pridruženi atomima označenim zvjezdicom na slici 6?



Vježba 5: Parametrizacija molekula u programskom paketu Amber

Prilikom modeliranja bioloških makromolekula često je potrebno uz strukturu biološke makromolekule uzeti u obzir i ligand za koji parametri nisu dostupni u standardnom polju sila. U takvoj situaciji neophodno je generirati parametre koji će omogućiti računalno tretiranje liganda. *Antechamber* je potprogram unutar programa *Amber* koji sadrži niz alata za parametrizaciju organskih molekula i nekih metalnih centara u proteinima. Pri tome se koristi polje sila GAFF koje sadrži općenite tipove atoma i koristi semiempirijske metode za generiranje naboja. Ukoliko ligand spada u skupinu poznatijih organskih spojeva, *antechamber* omogućuje njegovu potpunu parametrizaciju. Ponekad nakon primjena potprograma *antechamber* nedostaju još neki parametri koji se naknadno mogu generirati potprogramom *parmchk2*. Za parametrizaciju metalnih centara najčešće se koristi potprogram *Metal Center Parameter Builder (MCPB)*.

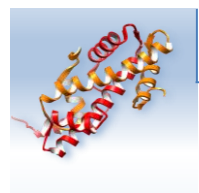
Zadatak 2.2 Parametrizirati molekulu N-acetil-glikozamina (NAG) iz .pdb datoteke oznake 2ch5 analizirane u vježbi 1. Za početak je potrebno izdvojiti molekulu N-acetil-glikozamina (NAG) koji se nalazi u kompleksu s ljudskom N-acetil-glukozamin kinazom u zasebnu .pdb datoteku korištenjem tekstualnog editora. Izdvojenu molekulu prvo protonirati, a zatim parametrizirati u potprogramu *antechamber*.

Upute:

Otvoriti 2ch5.pdb datoteku u tekstualnom editoru. Pronaći atome koji čine ligand NAG i pridruženi su lancu A, odnosno imaju oznaku da pripadaju lancu A. Željene atome označiti i kopirati u zasebnu .pdb datoteku nazvanu **NAG_original.pdb**. Kako bi se ligand uspješno parametrizirao, potrebno ga je prvo ispravno protonirati. Protoniranje će se provesti potprogramom *reduce* naredbom:

reduce NAG_original.pdb > NAG_H.pdb

Datoteku dobivenu protoniranjem potrebno je dodatno optimizirati za rad u *Amberu* naredbom:



pdb4amber -i NAG_H.pdb -o NAG_H_clean.pdb

Datoteku NAG_H_clean.pdb otvoriti u vizualizacijskom programu *VMD* te provjeriti je li molekula liganda ispravno protonirana.

Parametrizaciju ispravno protonirane molekule NAG pokrenuti naredbom:

antechamber -fi pdb -fo mol2 -i NAG_H_clean.pdb -o NAG_H_clean.mol2 -c bcc -nc 0 -j 4

Osnovne opcije potprograma *antechamber* korištene u ovoj vježbi objašnjene su u tablici 1. Dodatne opcije moguće je pronaći u uputama programa *Amber* na poveznici: <https://ambermd.org/doc12/Amber16.pdf>.

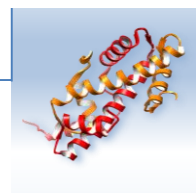
Dobivenu .mol2 datoteku potrebno je provjeriti u *parmchk* potprogramu koji će ispisati sve dodatne modifikacije polja sila (.frcmod) nužne za potpunu parametrizaciju liganda:

parmchk2 -f mol2 -i NAG_H_clean.mol2 -o NAG_H_clean.frcmod

U konačnici, dobivene datoteke NAG_H_clean.mol2 i NAG_H_clean.frcmod sadrže sve potrebne parametre koji omogućuju računalno tretiranje NAG liganda u *Amberu*.

Tablica 1. Objašnjenja *antechamber* opcija korištenih za parametrizaciju molekule NAG.

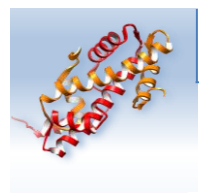
<i>antechamber</i> opcija	Značenje
-i	Ime ulazne datoteke
-fi	Format ulazne datoteke
-o	Ime izlazne datoteke
-fo	Format izlazne datoteke
-c	Metoda dodavanja naboja (u sklopu ove vježbe koristit će se AM1-BCC metoda čije je oznaka bcc)
-nc	Ukupan naboj molekule
-j	Metoda prediviđanja tipa atoma i tipa veze



Liste svih podržanih formata datoteka i metoda dodavanja naboja prikazuju se unosom naredbe **antechamber -L**.

Pitanja:

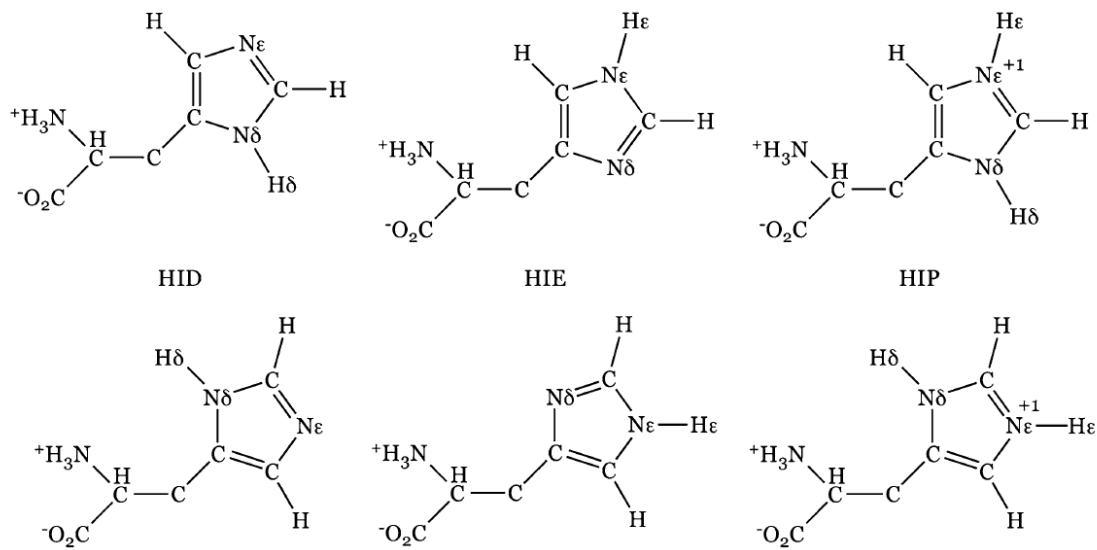
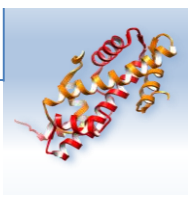
1. Vizualizirati protoniranu molekulu NAG u programu *VMD* i napraviti sliku s metodom prikaza *Licorice* u kojem će radijus prikazane veze iznositi
2. Kako je u vizualizacijskom programu prikazana dvostruka veza između ugljika i kisika?
2. Koje informacije o strukturi molekule sadrži *.mol2* datoteka, a, u pravilu, ih nema u *.pdb* datoteci?
3. Koju informaciju sadrži nastala *.fromod* datoteka?
4. Kojim se još metodama može dodati naboj u potprogramu *antechamber*?
5. U direktoriju `$AMBERHOME/dat/leap/parm` tekstualnim editorom otvoriti *gaff.dat* polje sila. Potrebno je pronaći opise i naboje slijedećih tipova atoma: *c3*, *n3*, *ss*.



3. Kompletiranje i protoniranje bioloških makromolekula

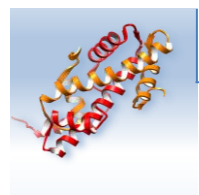
Prvi korak pripreme sustava za simulaciju molekulske dinamike sastoji se od kompletiranja strukture. Naime, strukture bioloških makromolekula dobivene kristalografijom najčešće nemaju koordinate vodikovih atoma, a ponekad im nedostaju i koordinate pojedinih aminokiselinskih ostataka ili čak čitavih strukturnih elemenata poput omče ili zavojnice. Stoga je prvi korak pripreme sustava za simulaciju kompletiranje strukture, što uključuje protoniranje molekule i izgradnju strukturnih dijelova koji nedostaju.

Protoniranje nije nužno jednoznačno u slučaju svih aminokiselinskih ostataka. Različita protonacijska stanja pojedinih aminokiselina mogu značajno utjecati na mreže vodikovih veza u strukturi biološke makromolekule. Mreže vodikovih veza u velikoj mjeri utječu na strukturna i dinamička svojstva bioloških makromolekula, a često predstavljaju i važan aspekt mehanizama kojima biološke makromolekule obnašaju svoje funkcije. Stoga je protoniranje početne strukture od velike važnosti za vjerodostojnost i uspješnost simulacija molekulske dinamike bioloških makromolekula. Aminokiselina čije protoniranje predstavlja najveći izazov jest histidin čiji imidazolni prsten ima 3 moguća protonacijska stanja (slika 7). Nadalje, u većini kristalografski riješenih struktura bioloških makromolekula, rezolucija nije dovoljno dobra da bi se sa sigurnošću moglo odrediti u kojem od dva rotamera se nalazi imidazolni prsten histidina. Rotameri se međusobno razlikuju za 180° , a položaji donora i akceptora vodikove veze značajno se razlikuju u rotamerima (slika 7). Iz svega navedenog proizlazi da histidin može imati čak 6 različitih stanja koja mogu imati različitu ulogu u potencijalnoj mreži vodikovih veza. Kao i u slučaju histidina, i kod glutaminske te asparaginske kiseline rezolucija većine kristalografski riješenih struktura bioloških makromolekula onemogućava jednoznačno određivanje rotamera u kojem se nalazi aminokiselinski ostatak. Rotacijom aminokiselinskih ostataka asparagina i glutamina za 180° akceptor i donor vodikove veze međusobno zamijene mjesta što može uzrokovati prekid ili neuspostavljanje mreže vodikovih veza u kojoj ta aminokiselina sudjeluje. Slijedom svega navedenog, jasno je da su protonacijsko stanje i orijentacija aminokiselinskih ostataka početne strukture od velike važnosti za uspješnost simulacija.



Slika 7. Protonacijska stanja i rotameri imidazolnog prstena histidina. Označeni su nazivi dušikovih atoma imidazolnog prstena i dani su uobičajeni nazivi protonacijskih stanja. Slika je preuzeta iz literaturnog izvora [2].

Postoji više programa i različitih algoritama koji pokušavaju odrediti optimalno protonacijsko stanje i orijentacija aminokiselinskih ostataka početne strukture. Većina algoritama temelji se na iteracijskom postupku kojim se mijenja orijentacija i protonacijsko stanje aminokiselinskih ostataka s ciljem pronalaska strukture u kojoj je postignuta energijski optimalna mreža vodikovih veza. Primjeri takvih programa su *What If* [3], *H++* [4-6] i *Mol Probit* [7] čije je korištenje opisano u ovom priručniku.



Vježba 6: Protoniranje esteraze EstA iz bakterije *Pseudomonas aeruginosa* korištenjem programa *What If*

Strukturu esteraze EstA iz bakterije *Pseudomonas aeruginosa* koja je pronađena pretragom baze *PDB* potrebno je protonirati korištenjem programa *What If* [3]. Program *What If* moguće je lokalno instalirati, ali za potrebe ove vježbe preporuča se korištenje mrežno dostupne verzije programa kojem se može pristupiti preko poveznice: <https://swift.cmbi.umcn.nl/whatif/>.

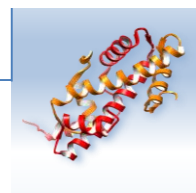
Zadatak 3.1 Dodati polarne atome vodika u strukturu esteraze EstA iz bakterije *Pseudomonas aeruginosa* i odgovoriti na zadana pitanja.

Upute:

Korištenjem predloženog sučelja i odabirom *What If* servera dodati polarne atome vodika u strukturu esteraze EstA iz bakterije *Pseudomonas aeruginosa*. Pri tome paziti da se spremne i informacije o protonacijskim stanjima i promjenama orijentacija pojedinih aminokiselina.

Pitanja:

1. Nabrojite histidine esteraze EstA iz bakterije *Pseudomonas aeruginosa* koje je program *What If* protonirao na: a) N_{δ} , b) N_{ϵ} , c) oba dušikova atoma imidazolnog prstena.
2. Nabrojite histidine esteraze EstA iz bakterije *Pseudomonas aeruginosa* koje je program *What If* reorijentirao za 180° (engl. *flip*).
3. Nabrojite glutaminske i asparaginske kiseline esteraze EstA iz bakterije *Pseudomonas aeruginosa* koje je program *What If* reorijentirao za 180° .



Vježba 7: Kompletiranje strukture i protoniranje ljudske N-acetil-glukozamin kinaze korištenjem programa SWISS-MODEL

Dijelovi strukture koji nedostaju u izvornoj .pdb datoteci izgradit će se korištenjem mrežno dostupnog programa za homologno modeliranje strukture proteina SWISS-MODEL [8]. U homolognom modeliranju, 3D model proteina od interesa dobiva se ekstrapolacijom eksperimentalnih podataka evolucijski srodne proteinske strukture koja služi kao predložak. Izgradnja 3D modela sastoji se od četiri koraka: (i) pronalaženja strukturnih predložaka, (ii) sravnjivanja aminokiselinskih sljedova proteina od interesa i strukturnih predložaka, (iii) izgradnje modela i (iv) procjenjivanja kvalitete dobivenih modela. Za traženje strukturnih predložaka koriste se SWISS-MODEL knjižnice (STML) koje sadrže podatke o svim, do sad, riješenim strukturama proteina[8].

Nakon završetka modeliranja mrežno dostupnim programom SWISS-MODEL potrebno je dohvatiti strukturu koja je ocijenjena kao najkvalitetnija te ju pohraniti na lokalno računalo. Kompletiranu strukturu ljudske N-acetil-glukozamin kinaze (NAG kinaze) potrebno je protonirati korištenjem mrežno dostupnog programa H++ [4][5][6] na poveznici: <http://biophysics.cs.vt.edu/>. Program H++ računa pK_a vrijednosti ionizabilnih skupina u makromolekulama te ih protonira ovisno o okruženju u kojem se skupina nalazi i u ovisnosti o zadanoj pH vrijednosti na kojoj se sustav želi protonirati.

Zadatak 3.2 Korištenjem mrežno dostupnog programa SWISS-MODEL izgraditi dijelove koji nedostaju u strukturi ljudske N-acetil-glukozamin kinaze čija je .pdb datoteka pohranjena u vježbi 1 te odgovoriti na postavljena pitanja.

Upute:

Pristupiti bazi PDB preko poveznice: <https://www.rcsb.org/>, pronaći strukturu ljudske N-acetil-glukozamin kinaze u kompleksu s NAG iz vježbe 1 te pohraniti njen aminokiselinski slijed u obliku formata FASTA. Format FASTA je tekstualni format koji se koristi u bioinformatički za prikaz nukleinskih ili aminokiselinskih sljedova, pri čemu su nukleotidi ili aminokiseline zapisani jednoslovnim kodovima. Jednostavnost formata FASTA omogućava jednostavnu manipulaciju i analizu u

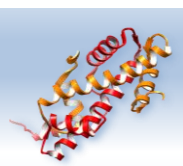
različitim programima za obradu teksta i programskim jezicima poput *R*, *Python*, *Ruby* i *Perl*.

Preko poveznice: <https://swissmodel.expasy.org/interactive> pristupiti mrežno dostupnom programu *SWISS-MODEL*. Otvorit će se sučelje kao na slici 8.

The screenshot shows the 'Start a New Modelling Project' page on the SWISS-MODEL website. The page includes a navigation bar with 'Modelling', 'Repository', 'Tools', 'Documentation', 'Log in', and 'Create Account'. The main content area has a 'Target Sequence(s)' input field with a placeholder 'paste your target sequence(s) or UniProtKB AC here'. Below this field are buttons for '+ Upload Target Sequence File...' and 'Validate'. There are also input fields for 'Project Title' (defaulting to 'Untitled Project') and 'Email' (defaulting to 'Optional'). At the bottom of the form are two large blue buttons: 'Search For Templates' and 'Build Model'. To the right, a 'Supported Inputs' dropdown menu is visible, listing 'Sequence(s)', 'Target-Template Alignment', 'User Template', and 'DeepView Project'. A footer note states: 'By using the SWISS-MODEL server, you agree to comply with the following terms of use and to cite the corresponding articles.'

Slika 8. Interaktivno sučelje mrežno dostupnog programa *SWISS-MODEL* za homologno modeliranje strukture proteina.

Iz prethodno pohranjene datoteke formata FASTA kopirati aminokiselinski slijed koji pripada proteinskom lancu A ljudske NAG kinaze u prozor *Target Sequence(s)*. Imenovati projekt „NAG kinaza“ te odabrati opciju *Build Model*. Nakon nekoliko minuta otvorit će se nova stranica s nekoliko ponuđenih modela. Odabirom opcije **Structure Assessment** u prozoru pojedinog modela otvara se nova stranica s detaljnijim informacijama o dobivenoj strukturi, dok se klikom na *PDB* kod ispod podnaslova **Template** prikazuju informacije o predlošku na temelju kojeg je model izgrađen. Potrebno je pohraniti .pdb datoteku najboljeg modela na lokalno računalo.

**Pitanja:**

1. Koje aminokiseline nedostaju u lancu A ljudske NAG kinaze? Napraviti sliku u programu *VMD* iz koje je vidljivo na kojem mjestu nedostaje dio strukture lanca A.
2. Koji je protein poslužio kao predložak za dobivanje najboljeg modela? Pogledati sravnjene aminokiselinske sljedove predložka i proteina od interesa klikom na PDB kod predložka. Komentirati kako je izgrađen najbolji model.
3. Koliko je lanaca u pohranjenoj .pdb datoteci najboljeg modela?
4. Zašto je potrebno imati cjelovitu strukturu proteina za simulacije molekulske dinamike?

Zadatak 3.2 Dodati polarne atome vodika u strukturu kompletirane ljudske NAG kinaze programom *H++* i odgovoriti na zadana pitanja.

Upute:

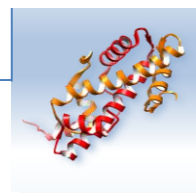
Otvoriti pohranjenu .pdb datoteku kompletirane strukture ljudske NAG kinaze iz prethodnog zadatka u tekstualnom editoru te izdvojiti aminokiseline koje pripadaju lancu A u zasebnu datoteku nazvanu NAG-kinase-A.pdb. Programu *H++* pristupa se putem poveznice: <http://biophysics.cs.vt.edu/>, za izvođenje zahtjevnijih računa potrebno se registrirati. Nakon registracije odabrati opciju **Process A Structure**. Priložiti NAG-kinase-A.pdb datoteku korištenjem opcije **Browse**, a zatim pokrenuti obradu odabirom opcije **Process File**. Novootvoreni prozor nudi mogućnost odabira dielektrične konstante, ionske jakosti, pH vrijednosti, itd. Potrebno je izmijeniti pH vrijednost na 5,6 te odabrati opciju **Process**. Nakon nekoliko minuta program će generirati datoteke (.pdb, .top i .crd) koje je potrebno pohraniti na lokalno računalo. Datoteke .top i .crd preimenovati u: NAG-kinase-H.crd i NAG-kinase-H.top. Kako bi se u programu *Amber* mogli pravilno učitati svi dodani vodikovi atomi, potrebno je na temelju datoteka NAG-kinase-H.top i NAG-kinase-H.crd generirati novu .pdb datoteku naredbom:

```
ambpdb -p NAG-kinase-H.top -c NAG-kinase-H.crd > NAG-kinase_protonated.pdb
```

Dobivena **NAG-kinase-protonated.pdb** datoteka spremna je za pripremu sustava za simulaciju molekulske dinamike programom *Amber*.

Pitanja:

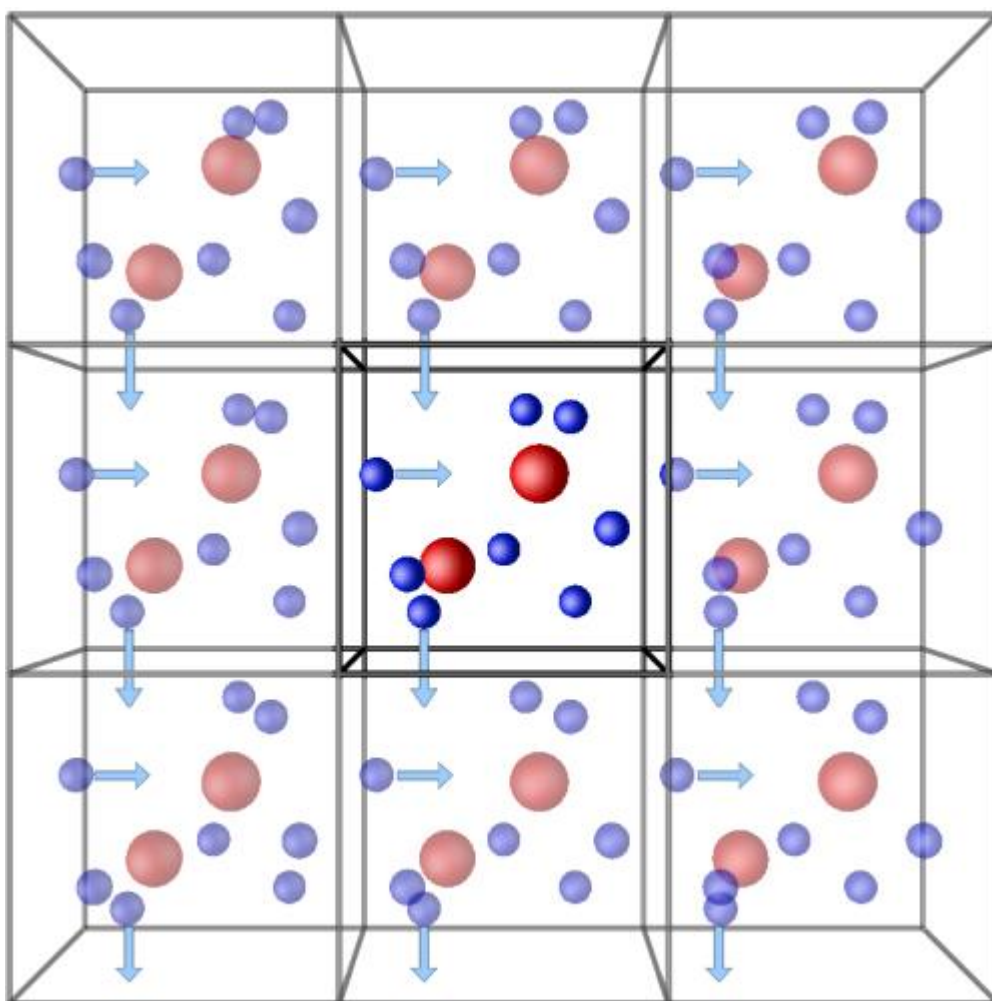
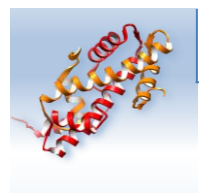
1. U rezultatima programa *H++* iščitati ukupan naboj proteina pri $\text{pH} = 5,6$.
2. Vizualizirati dobivenu datoteku NAG kinase-protonated.pdb u programu *VMD*. Napraviti sliku na kojoj će histidini 37, 148 i 245 biti istaknuti prikazom *Licorice*. Komentirati njihovo protonacijsko stanje. Kako su ti isti histidini imenovani u NAG-kinase-protonated.pdb datoteci?
3. Napraviti sliku na kojoj će prikazom *Licorice* biti istaknut bočni ogranak 135 i sve aminokiseline s kojima ostvaruje vodikovu vezu. Namjestiti debljinu isprekidane linije kojom se prikazuje vodikova veza na 6.



4. Solvatacija, periodični rubni uvjeti i elektroneutralnost

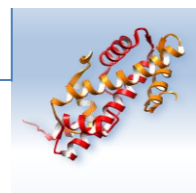
Utjecaj otapala na strukturu i dinamiku bioloških makromolekula vrlo je izražen i ne smije se zanemariti tijekom računalnih simulacija. Postoje različiti načini tretiranja otapala tijekom simulacija molekulske dinamike (MD). **Implicitni modeli** tretiraju otapalo kao kontinuum određene dielektrične konstante i računalno su manje zahtjevni, ali i manje vjerodostojni u odnosu na eksplicitne modele. Kod **eksplicitnih modela**, molekule otapala su eksplicitno prisutne tijekom simulacije. Kako bi se racionalizirala potrošnja računalnog vremena, molekule otapala često su predstavljene modelima kod kojih nisu dozvoljeni svi stupnjevi slobode gibanja koje molekula otapala inače ima. Primjerice, TIP3P model molekulu vode prikazuje kao kruto tijelo koje posjeduje dipolni, a ne kvadrupolni moment, te koje nema mogućnost vibracije veza i savijanja valentnog kuta, već samo mogućnost translacije i rotacije. Uzevši u obzir broj molekula vode potrebnih za adekvatnu solvataciju biološke makromolekule, ovakvim modelom postiže se značajna ušteda računalnog vremena.

U slučaju eksplicitnih modela otapala potrebno je riješiti problem **rubnih uvjeta**, odnosno spriječiti gubitak molekula vode tijekom simulacije. Najčešće korištena tehnika za rješavanje problema otapala su **periodični rubni uvjeti** (engl. *periodic boundary conditions – PBC*). Kod periodičnih rubnih uvjeta središnji poliedar (najčešće kvadar, ali moguće je koristiti i krnji oktaedar, ikozaedar...) replicira se u prostoru na način da bude sa svih strana okružena svojim replikama. Svaka molekula koja tijekom simulacije izađe s jedne strane središnje ćelije u susjednu repliku, s druge strane će u nju ući iz odgovarajuće replike, čime se osigurava zatvorenost sustava i rješava problem rubnih uvjeta (slika 9).



Slika 9. Periodični rubni uvjeti. Istaknut je središnji poliedar (ćelija), ostale ćelije su njegove replike (preuzeto s: <http://isaacs.sourceforge.net/phys/pbc.html>).

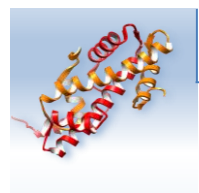
Elektroneutralnost sustava postiže se dodavanjem odgovarajućeg broja aniona ili kationa. U nekim slučajevima, uz postizanje elektroneutralnosti sustava, kationi i anioni se dodaju i radi postizanja ionske jakosti pri kojoj se želi provesti simulacija. Najčešće se dodaju natrijevi i kloridni ioni za koje postoje vjerodostojni parametri u većini polja sila.



Vježba 8: Priprema 3D strukture za simulaciju molekulske dinamike (MD)

Neophodno je pripremiti 3D strukturu željene makromolekule za MD simulaciju, odnosno provesti parametrizaciju makromolekule i generirati datoteke koje programi *Gromacs* ili *Amber* zahtijevaju za MD simulaciju. Početna 3D struktura makromolekule najčešće je zapisana u obliku .pdb datoteke u kojoj se nalaze: koordinate atoma, redni brojevi i oznake atoma, oznake i redni brojevi aminokiselina, odnosno nukleotida ili liganada, te oznaka lanca (slika 3). Ostale informacije također se mogu naći u .pdb datoteci, ali nisu nužne za pripremu strukture za MD simulaciju programima *Gromacs* ili *Amber*, a neke od tih informacija mogu ponekad čak i stvarati probleme prilikom parametrizacije. U svrhu pripreme .pdb datoteke za MD simulaciju programima *Gromacs* ili *Amber*, potrebno je otvoriti .pdb datoteku u tekstualnom editoru i urediti ju prema sljedećim uputama:

- Provjeriti je li struktura željene makromolekule kompletna. Ukoliko nedostaju dijelovi strukture, izgraditi ih korištenjem mrežno dostupnog programa *SWISS-MODEL* (opisano u vježbi 7).
- Ukloniti sve strukturne elemente za koje nema parametara u željenom polju sila poput: posttranslacijskih modifikacija, liganda, kofaktora, iona, itd. Strukturni elementi koji nisu standardno parametrizirani u željenom polju sila zahtijevaju dodatnu parametrizaciju.
- Strukturne elemente za koje standardno polje sila nema parametre, primjerice ligand, potrebno je izdvojiti i spremiti u zasebne .pdb datoteke kako bi se odvojeno provela njihova parametrizacija (opisano u vježbi 5).
- Ukoliko .pdb datoteka sadrži zapise o povezanosti atoma, oni bi mogli stvarati probleme pri pripremi sustava korištenjem programa *Gromacs* ili *Amber* te ih je u tom slučaju potrebno ukloniti.
- Na kraju svakog aminokiselinskog lanca potrebno je staviti oznaku **TER** ukoliko već nije stavljena. Ova oznaka označava kraj kovalentno vezanih atoma.



Utjecaj otapala tijekom MD simulacija modelirat će se korištenjem eksplicitnog modela otapala i periodičnih rubnih uvjeta. Elektroneutralnost će se postići dodavanjem natrijevih ili kloridnih iona.

Vježba 8.1 Priprema 3D strukture za MD simulaciju, solvatacija i periodični rubni uvjeti u programu *Gromacs*

Program *Gromacs* zahtijeva dvije nove datoteke, `.top` i `.gro`, koje se jednostavno mogu dobiti iz `.pdb` datoteke. Datoteka **`.gro`** je datoteka s koordinatama atoma makromolekule, donekle slična `.pdb` datoteci. Datoteka **`.top`** opisuje topologiju sustava, odnosno sadrži parametre koji su pridruženi sustavu na temelju odabranog polja sila.

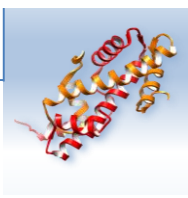
Zadatak 4.1 Pripremiti `.top` i `.gro` datoteke potrebne za pripremu MD simulacije katalitičke domene esteraze EstA iz bakterije *Pseudomonas aeruginosa* programom *Gromacs*.

Upute:

Koristeći `.pdb` datoteku esteraze EstA iz bakterije *Pseudomonas aeruginosa* koju ste protonirali programom *What If* u vježbi 6, potrebno je izdvojiti katalitičku domenu (aminokiseline 1-306) od transmembranske domene (aminokiseline 307-622). U tu svrhu, u tekstualnom editoru je potrebno izdvojiti samo retke koji se odnose na koordinate atoma koji pripadaju aminokiselinama 1-306 u posebnu datoteku. Tako pripremljena `.pdb` datoteka (**`EstA.pdb`**) već se nalazi na lokalnom računaru radi usporedbe uspješnosti editiranja originalne `.pdb` datoteke. Na temelju `.pdb` datoteke koja sadrži samo katalitičku domenu esteraze EstA iz *Pseudomonas aeruginosa* potrebno je generirati `.top` i `.gro` datoteke naredbom:

`gmx pdb2gmx -f EstA.pdb -p EstA.top -o EstA.gro`

opcija `-f` određuje ulaznu datoteku, opcije `-p` i `-o` određuju imena izlaznih `.top` i `.gro` datoteka. Pri izvršavanju naredbe potrebno je odabrati polje sila i model molekula otapala. Za izvođenje ove vježbe preporuča se odabir Amber99sb polja sila i TIP3P modela molekula vode. Ukoliko se u pripremu sustava kreće sa strukturom koja je protonirana programima *H++* ili *What If* često je potrebno na



kraju naredbe dodati opciju **-ignh** kako bi se oznake vodikovih atoma prevele u one koje *Gromacs* može učitati. U slučaju pripreme sustava koji sadrži disulfidne mostove potrebno je dodati još i opciju **-ss**. Sve ostale raspoložive opcije potprograma *pdb2gmx* i njihova objašnjenja mogu se pronaći naredbom:

gmx pdb2gmx -h

Opcija **-h** dostupna je kod većine naredbi programa *Gromacs*.

Dobivene datoteke potrebno je otvoriti i provjeriti u tekstualnom editoru. Kako bi se pronašli odgovori na postavljena pitanja, otvoriti *.gro* datoteku programom *VMD*.

Pitanja:

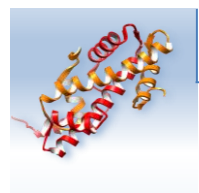
1. S kojom datotekom iz prethodnih vježbi pokazuje sličnosti *.top* datoteka? Zašto?
2. Koje su glavne razlike *.gro* i *.pdb* datoteke? Što označava prva brojka u *.gro* datoteci? Što predstavljaju tri brojke u zadnjem retku *.gro* datoteke?
3. Koja opcija naredbe *pdb2gmx* nam omogućava odabir protonacijskog stanja za određenu aminokiselinu u slučaju da nismo prethodno protonirali sustav?

Zadatak 4.2 Provesti solvataciju katalitičke domene esteraze EstA iz bakterije *Pseudomonas aeruginosa* programom *Gromacs* koristeći eksplicitni model otapala i periodične rubne uvjete.

Upute:

Solvatacija proteina eksplicitnim modelom otapala uz korištenje periodičnih rubnih uvjeta u programu *Gromacs* provodi se u nekoliko koraka. U prvom koraku se naredbom **gmx editconf** izabire veličina i oblik prizme koja će se replicirati u prostoru:

gmx editconf -f EstA.gro -bt cubic -d 1 -o EstA_box.gro



opcijom `-bt` specificira se oblik simulacijskog poliedra, a opcijom `-d` udaljenost u nm između atoma proteina i granica poliedra. Umjesto opcije `-d`, moguće je zadati dimenzije poliedra opcijom `-box`, ali ta se opcija preporuča jedino u slučajevima kada se iz opravdanog razloga želi koristiti upravo određena veličina. U središtu poliedra nalazi se makromolekula, a ostatak je potrebno ispuniti molekulama otapala, što se postiže naredbom:

```
gmx solvate -cp EstA_box.gro -cs spc216.gro -p EstA.top -o EstA_sol.gro
```

opcijom `-cs` specificira se model molekula otapala. Kako bi sustav bio električki neutralan, potrebno je adekvatan broj molekula otapala zamijeniti kationima (najčešće natrija) ili anionima (najčešće klora). To se može učiniti na dva načina. Prvi način je zamjena molekula vode ionima u `.gro` datoteci korištenjem tekstualnog editora. Primjerice, jednu molekulu vode koja je zadana s tri atoma:

15976SOL	OW51436	7.876	7.872	7.968
15976SOL	HW151437	7.906	7.800	7.906
15976SOL	HW251438	7.837	7.832	8.051

potrebno je zamijeniti jednim ionom natrija:

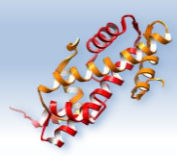
15976NA	NA51436	7.876	7.872	7.968
---------	---------	-------	-------	-------

ili jednim ionom klora:

15976CL	CL51436	7.876	7.872	7.968
---------	---------	-------	-------	-------

Pri tome je neophodno izmijeniti ukupan broj atoma u sustavu u `.gro` datoteci, kao i broj molekula otapala i broj iona koji su navedeni u `.top` datoteci. Broj iona koje je potrebno dodati ovisi o ukupnom naboju sustava.

Drugi način neutralizacije naboja sustava podrazumijeva korištenje naredbe **gmx genion** i opisan je u sljedećoj vježbi budući da je za njega neophodno korištenje `.tpr` datoteke.

**Pitanja:**

1. Koji su oblici simulacijskog polierda dostupni u programu *Gromacs*? Koji od ponuđenih oblika je najprikladniji za globuralni protein i zašto?
2. Zašto udaljenost ruba poliedra od atoma makromolekule (opcija `-d` tijekom naredbe `editconf`) ne smije biti premala, ali nije poželjno ni da je prevelika?
3. Gdje je u `.top` datoteci specificirano koji će se model molekula vode koristiti?
4. Vizualizirati sustav u programu *VMD* i napraviti: a) sliku sustava s adekvatnim prikazima molekula otapala i makromolekule; b) sliku na kojoj se vide barem tri replike simulacijske ćelije.

Vježba 8.2 Priprema 3D strukture za MD simulaciju, solvatacija i periodični rubni uvjeti u programu *Amber*

U *Amberu* se parametrizacija sustava provodi potprogramima *tLeap* ili *xLeap*. Za razliku od potprograma *tLeap*, *xLeap* sadrži grafičko sučelje. Zbog veće jednostavnosti korištenja u ovoj će se vježbi parametrizacija provesti u potprogramu *tLeap*.

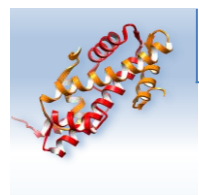
Parametrizacijom se generiraju dvije nove datoteke **.prmtop** i **.inpcrd**. Datoteka `.prmtop` opisuje topologiju sustava, odnosno sadrži parametre koji su pridruženi sustavu na temelju odabranog polja sila, dok `.inpcrd` sadrži informacije o početnim koordinatama atoma u sustavu.

Zadatak 4.3 Potprogramom *tLeap* pripremiti `.prmtop` i `.inpcrd` datoteke potrebne za simulaciju ljudske NAG kinaze u kompleksu s NAG u programu *Amber*.

Upute:

U već pripremljenoj datoteci nazvanoj `tLeap_14SB` sadržaja:

```
source leaprc.protein.ff14SB
loadamberparams gaff.dat
source leaprc.water.tip3p
```



navedena su imena polja sila i parametara koji će se koristiti za pripremu sustava. Leap.prtotein.ff14SB je ime glavnog polja sila koje će se koristiti za parametrizaciju proteina, gaff.dat je polje sila za parametrizaciju malih molekula, a leaprc.water.tip3p sadrži parametre za model molekule vode. Naredbom:

tleap -f tleap_14SB

otvara se potprogram *tleap* te se učitavaju sva polja sila i parametri navedeni u *tleap_14SB* datoteci.

Učitavanje kompletirane i protonirane strukture ljudske NAG kinaze u potprogram *tleap* vrši se naredbom:

enz = loadpdb NAG-kinase-protonated.pdb

Ovom se naredbom učitanoj .pdb datoteci pridružuje oznaka *enz*, koja će se koristiti dalje tijekom pripreme sustava. Ako se učitava više .pdb datoteka, važno je paziti da im se ne dodijeli ista oznaka. Nakon učitavanja, preporučuje se napraviti provjeru učitane .pdb datoteke naredbom:

check enz

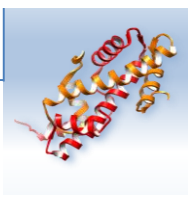
Ukoliko nešto nije uredu s ulaznom datotekom, program će odmah ispisati pogreške „ERRORS“ ili upozorenja „WARNINGS“. Pogreške se ne mogu zanemariti i parametrizacija ne može biti nastavljena dok se one ne isprave. Nakon učitavanja enzima potrebno je učitati ligand NAG čiji su parametri dobiveni u vježbi 5. Budući da se njegovi parametri nalaze u dvije zasebne datoteke .mol2 i .frcmod učitavaju se dvjema naredbama:

NAG= loadmol2 NAG_H_clean.mol2

loadamberparms NAG_H_clean.frcmod

Učitani ligand provjeriti naredbom:

check NAG



Sjedinjavanje koordinata i parametara liganda i enzima u kompleks postiže se naredbom:

cplx= combine {enz NAG}

Ponovno provjeriti dobiveni kompleks: **check cplx**

Neutralizacija naboja sustava postiže se dodavanjem kloridnih ili natrijevih iona. Ukupni naboj sustava provjerava se naredbom:

charge cplx

Ovisno o ukupnom naboju promatranog sustava dodaju se Na⁺ ili Cl⁻ ioni naredbom:

addions cplx Na+ 0

ili

addions cplx Cl- 0

Ovom naredbom *Amber* još jednom provjerava naboj sustava te ovisno o njemu dodaje ione natrija ili klora dok se ne postigne elektroneutralnost sustava. Solvatacija dobivenog kompleksa eksplicitnim modelom otapala (vode) uz korištenje periodičnih rubnih uvjeta u programu *Amber* provodi se naredbom:

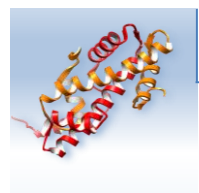
solvatebox cplx TIP3PBOX 10

Oznaka TIP3PBOX sadrži informaciju o obliku simulacijskog poliedra i modelu molekula vode (TIP3P), dok broj 10 određuje dimenzije simulacijske ćelije u angstromima.

Topologija i koordinate pripremljenog sustava ispisuju se naredbama:

saveamberparm cplx cplx-NAG-kinase.prmtop cplx-NAG-kinase.inpcrd

savepdb cplx cplx-NAG-kinase.pdb

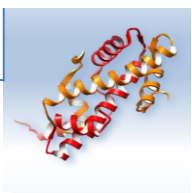


Parametrizacija se može provesti i korištenjem skripte koja sadrži sve potrebne naredbe za program *tLeap*, te imena datoteka. Izrada takve skripte opisana je u dodatku.

Nakon parametrizacije je nužno provjeriti dobiveni kompleks vizualizacijom *.pdb* datoteke u programu *VMD*.

Pitanja:

1. Koji su oblici simulacijskog poliedra dostupni u programu Amber? Koji od ponuđenih oblika je najprikladniji za globuralni protein i zašto?
2. Koja su sve polja sila i parametri zapisani u datoteci *tLeap_14SB* koju učitavamo u program *tLeap*?
3. Zašto udaljenost ruba poliedra od atoma makromolekule ne smije biti premala, ali nije poželjno ni da je prevelika?
4. Koliki je bio naboj same NAG kinaze, a koliki cijelog kompleksa prije neutralizacije sustava?
5. Vizualizirati sustav u programu *VMD* i napraviti: a) sliku sustava s adekvatnim prikazima molekula otapala, makromolekule, liganda i dodanog iona; b) sliku na kojoj se vide barem tri replike simulacijske ćelije.



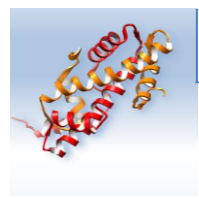
5. Minimizacija energije (optimizacija geometrije)

Simulaciji molekulske dinamike uobičajeno prethodi jedan ili više ciklusa računa minimizacije energije, odnosno optimizacije geometrije, tijekom kojih se konformacija (makro)molekule ne mijenja značajno, već se sustav relaksira u jedan od lokalnih minimuma bliskih početnoj strukturi na plohi potencijalne energije. Prilikom optimizacije geometrije nastoje se ukloniti izrazito nepovoljne interakcije koje nastaju uslijed premalih ili prevelikih udaljenosti između atoma, optimiziraju se udaljenosti kovalentnih veza, valentnih kutova i sl.

Optimizacija funkcije više varijabli, u ovom slučaju funkcije polja sila, računski je zahtjevan zadatak za čije se rješavanje primijenjuju numeričke metode. Algoritmi koji se pri tome koriste temelje se na iteracijskom postupku kojim se postupno približavamo minimumu funkcije, a dijele se s obzirom na red derivacije funkcije potencijalne energije koje pri njezinoj optimizaciji koriste na optimizacijske metode nultog, prvog i drugog reda.

Metode nultog reda, poput metode simpleks, tijekom minimizacije energije ne provode deriviranje njezine funkcije. Međutim, takve metode su vrlo robusne i rijetko se koriste.

Metode prvog reda, poput metode najstrmijeg spusta (engl. *steepest descent*) i metode konjugiranih gradijenata (engl. *conjugate gradient*), koriste samo prvu derivaciju funkcije polja sila. Takve metode baziraju se na činjenici da se na temelju smjera gradijenta funkcije polja sila u nekoj točki može procijeniti smjer u kojem se nalazi minimum na plohi potencijalne energije, dok iznos gradijenta upućuje na udaljenost početne točke od minimuma. Metoda najstrmijeg spusta spada u robusnije metode optimizacije prvog reda. Postoji više izvedbi metode najstrmijeg spusta koje se razlikuju s obzirom na način određivanja promjene koordinata između dva uzastopna iteracijska koraka (npr. linijsko pretraživanje, metoda proizvoljnog koraka). Metoda konjugiranih gradijenata koristi informacije o gradijentima funkcije potencijalne energije polja sila dva uzastopna iteracijska koraka i manje je robusna od metode najstrmijeg spusta. S obzirom na to da se prilikom optimizacije geometrije molekule uvijek kreće od robusnije metode prema manje robusnim metodama, metoda konjugiranih gradijenata u praksi slijedi nakon primjene metode najstrmijeg spusta. U tom slučaju, primjena metode najstrmijeg spusta služi da bi se početna geometrija



molekule samo ugrubo približila minimumu, a metodom konjugiranih gradijenata se zatim provodi finije ugađanje geometrije s ciljem dodatnog približavanja minimumu na plohi potencijalne energije.

Metode optimizacije drugog reda, koriste i drugu derivaciju funkcije koja sadrži informaciju o zakrivljenosti plohe potencijalne energije u nekoj točki. Primjer takve metode je Newton-Raphsonova metoda i njene brojne modifikacije. Metode drugog reda koriste se kada je molekula već blizu minimuma energije i služe za finu optimizaciju geometrije molekule, odnosno za precizno nalaženje minimuma na plohi potencijalne energije. Takve metode računski su neisplative u slučajevima kada je jedini cilj optimizacije geometrije molekule njeno pripremanje za simulacije molekulske dinamike.

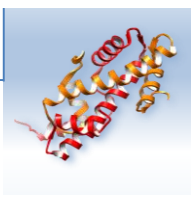
Vježba 9.1 Minimizacija energije (optimizacija geometrije) programom *Gromacs*

Kod programa *Gromacs*, sve specifikacije uvjeta vezanih uz račun koji se provodi nalaze se u **.mdp** datoteci koja je ulazna datoteka za potprogram **mdrun**. U **.mdp** datoteci specificirane su: vrste izlaznih datoteka i učestalost njihova zapisa, izabrani algoritmi, granične vrijednosti za nevezne interakcije; u slučaju MD simulacija i temperatura, tlak, vremenski korak, broj koraka, itd. Detalje o sadržaju **.mdp** datoteke može se pronaći na mrežnoj stranici: <http://manual.gromacs.org/>

Zadatak 5.1 Provesti minimizaciju energije (optimizaciju geometrije) katalitičke domene esteraze EstA iz bakterije *Pseudomonas aeruginosa* koristeći 3000 koraka algoritma najstrmijeg spusta (engl. *steepest descent (SD)*).

Upute:

Na lokalnom računalu nalazi se **.mdp** datoteka **MIN.mdp** koja je pripremljena za račun minimizacije energije. Potrebno je provjeriti jesu li sve opcije u danoj **.mdp** datoteci specificirane u skladu sa zadatkom te, po potrebi, određene opcije modificirati. Značenje pojedinih opcija može se provjeriti na gore navedenoj poveznici.



Na temelju modificirane .mdp datoteke te .gro i .top datoteka napravljenih u jednoj od prethodnih vježbi, potrebno je generirati .tpr datoteku naredbom:

```
gmx grompp -f MIN.mdp -c EstA_sol.gro -p EstA.top -o EstA_EM.tpr
```

Datoteka .tpr je binarna datoteka i to je jedina datoteka koju potprogram **mdrun** zahtijeva za pokretanje računa minimizacije ili simulacije.

Ukoliko sustav nije elektroneutralan, pri izvršavanju naredbe **gmx grompp** dobiva se opaska vezana uz ukupni naboj. U tom slučaju, potrebno je naboj sustava neutralizirati naredbom **gmx genion**:

```
gmx genion -s EstA_EM.tpr -p EstA.top -o EstA_sol_Na.gro -neutral -pname NA
```

opcija **-neutral** osigurava dodavanje neophodnog broja iona za elektroneutralizaciju sustava, opcije **-pname** ili **-nname** određuju vrstu kationa ili aniona koji se dodaju u sustav. Opcijama **-np** i **-nn** moguće je odabrati željeni broj kationa i aniona koji će se dodati u sustav u slučaju da želimo koristiti određenu ionsku jakost sustava. Nakon ovog koraka **neophodno je ponoviti** naredbu **gmx grompp** kako bi napravili novu .tpr datoteku iz .gro i .top datoteka koje sada sadrže i dodane ione.

Koristeći tako napravljenu .tpr datoteku, potrebno je pokrenuti račun minimizacije energije (optimizacije geometrije) naredbom:

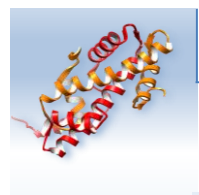
```
mdrun -s EstA_EM1.tpr -deffnm EstA_EM &
```

oznaka **&** omogućava da se račun odvija u pozadini, opcijom **-deffnm** biramo ime koje će izlazne datoteke imati (s različitim ekstenzijama). Ukoliko bude zadovoljen iteracijski kriterij koji je zadan u .mdp datoteci, račun će završiti i prije zadanih 3000 koraka. Naredbom:

```
g_energy -f EstA_EM.edr -o EstA_EM_energyX.svg
```

moguće je analizirati promjene energije pojedinih interakcija tijekom minimizacije energije. Dobiveni grafovi lako se mogu vizualizirati u programu *Grace*:

```
xmgrace EstA_EM_energyX.svg
```



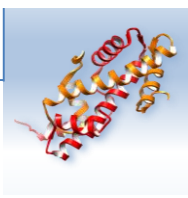
Pitanja:

1. Kolika je energija sustava nakon optimizacije geometrije, a kolika na početku?
2. U programu *VMD* napraviti sliku na kojoj su preklopljene i različito obojane strukture proteina prije i nakon optimizacije.
3. Ponoviti sliku koja se traži u prethodnom pitanju koristeći različite prikaze i bojenje atoma u nastojanju da se istakne dio strukture kod kojeg je zamijećena najveća razlika uslijed optimizacije geometrije.
4. U programu *Grace* grafički prikazati promjene energije: a) istežanja veza, b) savijanja kutova, c) elektrostatskih interakcija tijekom računa optimizacije geometrije sustava.
5. Grafički prikazati promjene potencijalne energije tijekom računa optimizacije geometrije sustava.

Vježba 9.2 Minimizacija energije (optimizacija geometrije) programom *Amber*

Sve specifikacije i uvjeti izvođenja računa u programu *Amber* zapisani su u ulaznoj .in datoteci. Ulazna datoteka za minimizaciju energije u programu *Amber* sadrži informacije o učestalosti zapisa izlaznih datoteka, izabranim algoritmima, broju koraka minimizacije, graničnim vrijednostima za nevezne interakcije...

U ovoj vježbi će se minimizacija energije ljudske NAG kinaze u kompleksu s NAG provesti u pet ciklusa optimizacije geometrije čije su specifikacije i uvjeti zapisani u pet zasebnih .in datoteka. U prvom ciklusu optimiziraju se samo koordinate molekula otapala, dok su promjene koordinata atoma makromolekule i liganda otežane primjenom harmonijskog potencijala sa konstantom sile od $100 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. U drugom ciklusu minimizacije energije isti je harmonijski potencijal primijenjen na položaje svih atome makromolekule i liganda, osim na položaje atoma vodika. Cilj drugog ciklusa je relaksacija položaja vodikovih atoma. U trećem i četvrtom ciklusu harmonijski potencijal primjenjuje se samo na položaje atoma proteinske okosnice. Iznos harmonijskog potencijala u trećem ciklusu iznosi $100 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, a u četvrtom $50 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. U posljednjem ciklusu minimizacije energije nema ograničenja na položaje niti jednog atoma u sustavu.



Zadatak 5.2 Provesti minimizaciju energije (optimizaciju geometrije) ljudske NAG kinaze u kompleksu s NAG u pet ciklusa. Svaki se ciklus sastoji od 200 koraka algoritma najstrmijeg spusta nakon čega slijedi 800 koraka algoritma konjugiranih gradijenata.

Upute:

Na lokalnom računalu nalazi se pet .in datoteka: **min1.in**, **min2.in**, **min3.in**, **min4.in** i **min5.in** pripremljenih za izvođenje pet ciklusa minimizacije energije. Provjeriti u tekstualnom editoru jesu li sve opcije u danoj .in datoteci specificirane u skladu sa zadatkom te ih po potrebi modificirati. Obratiti pozornost slaže li se broj aminokiselina simuliranog sustava sa upisanim brojem aminokiselina koje se fiksiraju u pojedinim ciklusima minimizacije. Značenje pojedinih opcija može se provjeriti u uputama programa *Amber* na poveznici:

<https://ambermd.org/doc12/Amber16.pdf> .

Na temelju modificirane ulazne .in datoteke te .prmtop i .inpcrd datoteka dobivenih u prethodnoj vježbi pokrenuti prvi ciklus minimizacije naredbom:

```
sander -p cplx-NAG-kinase.prmtop -c cplx-NAG-kinase.crd -ref cplx-NAG-kinase.inpcrd -i min1.in -o cplx-NAG-kinase_min1.out -r cplx-NAG-kinase_min1.rst -inf cplx-NAG-kinase_min1.info
```

Računom minimizacije nastaju tri nove datoteke: **.out**, **.rst** i **.inf**. Datoteka .out sadrži informacije o stanju sustava (temperaturi, tlaku, potencijalnoj energiji itd.) u svakom koraku koji se ispisuje. Datoteka .inf sadrži informacije o stanju sustava u posljednjem koraku za koji je minimizacija načinjena, a datoteka .rst sadrži koordinate sustava u tom koraku. Svaki idući ciklus minimizacije pokreće se po završetku prethodnog pri čemu se koristi odgovarajuća .in datoteka, a izlazna .rst datoteka prethodnog koraka postaje ulazna u sljedećem koraku. Primjerice:

```
sander -p cplx-NAG-kinase.prmtop -c cplx-NAG-kinase_min1.rst -ref cplx-NAG-kinase_min1.rst -i min2.in -o cplx-NAG-kinase_min2.out -r cplx-NAG-kinase_min2.rst -inf cplx-NAG-kinase_min2.info
```

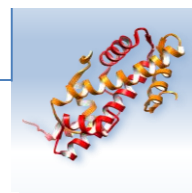
Upute za pripremu skripti te izvođenje računa minimizacije u paralelnoj izvedbi na više procesora potprogramom *sander.mpi* nalaze se u dodatku.

Po završetku svakog ciklusa minimizacije potrebno je otvoriti izlaznu *.out* datoteku tekstualnim editorom i provjeriti kako je tekla minimizacija energije te je li provedeno svih 1000 koraka ili je eventualno ranije zadovoljen iteracijski kriterij. Nakon svakog završenog koraka može se generirati odgovarajuća *.pdb* datoteka iz *.rst* datoteke naredbom:

```
ambpdb -p cplx-NAG-kinase.prmtop -c cplx-NAG-kinase_min5.rst > cplx-NAG-kinase_min5_rst.pdb
```

Pitanja:

1. Kolika je energija sustava nakon optimizacije geometrije u petom ciklusu, a kolika na početku?
2. U programu *VMD* napraviti sliku na kojoj su preklapljene i različito obojane strukture kompleksa prije optimizacije, nakon trećeg i nakon petog ciklusa minimizacije.
3. Ponoviti sliku koja se traži u prethodnom pitanju koristeći različite prikaze i bojenje atoma u nastojanju da se istakne dio strukture kod kojeg je zamijećena najveća razlika uslijed optimizacije geometrije.
4. Zašto je potrebno optimizirati geometriju sustava prije pokretanja računa molekulske dinamike?



6. Priprema i pokretanje simulacije molekulske dinamike (MD)

Molekulska dinamika (MD) je računalna metoda koja simulira gibanje atoma u sustavu na temelju čega se nastoji opisati ponašanje sustava na molekularnoj razini kroz zadani vremenski period. Osnova molekulske dinamike je da se iz trenutačnih položaja atoma, njihovih brzina te ubrzanja koje dobivaju zbog silâ koje na njih djeluju, može predvidjeti njihov raspored u prostoru tijekom vremena. Molekulska dinamika koristi Newtonove jednadžbe gibanja kako bi simulirala ponašanje sustava na danoj temperaturi te generira niz reprezentacija sustava tijekom zadanog vremenskog intervala. Ukoliko je taj vremenski interval dovoljno dugačak, a barijere plohe potencijalne energije relativno niske, ansambl generiranih konformacija prema ergotskoj pretpostavci odgovara ansamblu, odnosno raspodjeli konformacija, sustava s velikim brojem istovrsnih molekula.

U molekularnoj dinamici temperatura je prisutna kao konstantan ili promjenjiv parametar, uslijed kojeg atomi imaju određenu kinetičku energiju te se gibaju (11):

$$E_k = \frac{3}{2} N k_B T \quad (11)$$

gdje je N – broj atoma u sustavu, k_B – Boltzmannova konstanta ($1,38 \cdot 10^{-23} \text{ J K}^{-1}$), T – temperatura (K). Početne brzine atomima u sustavu najčešće se pridružuju prema Maxwell–Boltzmannovoj raspodjeli za danu temperaturu. Uslijed gibanja atomi napuštaju svoje početne položaje što dovodi do promjene konformacije molekule, odnosno do promjene položaja sustava na plohi potencijalne energije. Deriviranjem izraza za potencijalnu energiju (tj. izraza za energiju ukupne interakcije svih atoma u sustavu) i korištenjem drugog Newtonovog zakona, dobiva se skup Newtonovih jednadžbi gibanja čijim rješavanjem se dolazi do vremenske zavisnosti položaja atoma u sustavu, odnosno do trajektorije. Kako je sila koja djeluje na jedan atom u molekuli ovisna o položaju tog atoma, ali i položaju svih ostalih atoma s kojima je u interakciji, skup Newtonovih jednadžbi gibanja za atome u molekuli nema analitičko rješenje te je potrebno primijeniti numeričke metode rješavanja. Postoje različiti algoritmi koji omogućavaju numeričko rješavanje Newtonovih jednadžbi gibanja i dobivanje trajektorije, najčešći su *Verletov* algoritam i tzv. *leap-frog* algoritam, te njihove inačice poput

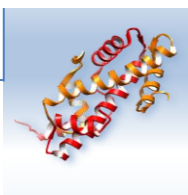
Beemanovog algoritma ili *velocity Verlet* algoritma. Sve numeričke metode za rješavanje jednadžbi gibanja temelje se na aproksimaciji da je sila koja djeluje na atom konstantna tijekom odabranog vremenskog koraka. Čim je manji odabrani vremenski korak, aproksimacija je prihvatljivija. Uobičajeno se koristi vremenski korak od 1 fs (10^{-15} s) što odgovara desetini periode istezanja C–H veze. Korištenjem različitih tehnika, poput *SHAKE*, *Rattle* i *Lincs* algoritama, kojima se fiksiraju određeni stupnjevi slobode tijekom simulacije, vremenski korak se može povećati.

Rezultat MD simulacija su **trajektorije** koje omogućavaju promatranje sustava u vremenskoj ovisnosti na molekularnoj razini, odnosno atomskoj razini, a time i promatranje **mikroskopskih** svojstava sustava. Stoga se rezultati simulacija često koriste za interpretaciju i razumijevanje eksperimentalno dobivenih **makroskopskih** svojstava molekula.

Glavnom dijelu MD simulacije, tijekom kojeg se skupljaju informacije o sustavu i prate njegove promjene, prethodi simulacija uravnoteženja (ekvilibracija) koja je potrebna kako bi se sustav stabilizirao i uravnotežio. Tijekom simulacije uravnoteženja s posebnom se pažnjom prate parametri poput temperature, kinetičke, potencijalne i ukupne energije, tlaka, brzina, gustoće otapala, itd. Cilj simulacije uravnoteženja je dobiti stabilan sustav, bez većih fluktuacija vrijednosti izabranih parametara sustava (gustoće otapala, temperature ako je simulacija pri konstantnoj temperaturi, tlaka ako je pri konstantnom tlaku i sl.). Tijekom simulacije uravnoteženja, koja prethodi MD simulaciji s konstantnom temperaturom, vrijednosti brzina se skaliraju puno češće nego što je to uobičajeno tijekom MD simulacija. Kako su temperatura i brzina povezane preko izraza za kinetičku energiju (12), skaliranjem brzina održava se željena temperatura sustava:

$$E_k = \frac{3}{2} N k_B T = \sum_{i=1}^N \frac{m_i v_i^2}{2} \quad (12)$$

gdje je m – masa čestice, v – brzina čestice. Tijekom glavne simulacije skaliranje brzina, a time i kontrola temperature čestica, vrši se puno rjeđe nego tijekom simulacije uravnoteženja.



Vježba 10: Simulacije uravnoteženja (ekvilibracije)

Prije pokretanja produkcijske faze MD simulacije, potrebno je provesti simulacije uravnoteženja (ekvilibracije) čiji je cilj uravnoteženje sustava, odnosno njegove energije, temperature, tlaka, gustoće otapala, itd.

Prvo će se provesti *NVT* ($N, V, T = \text{konst.}$; tzv. kanonski ansambl) ekvilibracija tijekom koje će položaji atoma proteina biti fiksirani primjenom harmonijskog potencijala sa konstantom sile, dok će kretanje molekula otapala biti nesputano. Volumen simulacijske kutije će se održavati konstantnim, a temperatura će se povisivati od 0 K do konačne temperature na kojoj se sustav želi simulirati. Zatim slijedi *NPT* ($N, p, T = \text{konst.}$; tzv. izotermno-izobarni ansambl) ekvilibracija koja će se provoditi pri konstantnom tlaku (101 300 Pa) i temperaturi, pri čemu se ne primjenjuje nikakvo ograničenje na koordinate atoma u sustavu.

Vježba 10.1 Simulacije uravnoteženja programom *Gromacs*

Zadatak 6.1 Pripremiti i provesti *NVT* i *NPT* simulacije uravnoteženja prethodno minimiziranog sustava katalitičke domene esteraze EstA iz bakterije *Pseudomonas aeruginosa* u trajanju od po 200 ps (*NVT*) i 100 ps (*NPT*) programom *Gromacs*.

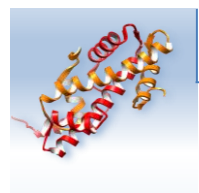
Upute:

Odgovarajuće .mdp datoteke za simulacije uravnoteženja (NVT.mdp i NPT.mdp) su pripremljene i nalaze se na lokalnom računaru. Potrebno ih je otvoriti u tekstualnom editoru, provjeriti specificirane opcije i, po potrebi, neke od njih izmijeniti. Na temelju NVT.mdp i .top datoteka, te .gro datoteke dobivene optimizacijom geometrije sustava (prethodna vježba), potrebno je pripremiti .tpr datoteku naredbom:

gmx grompp -f NVT.mdp -c EstA_EM.gro -p EstA.top -o EstA_NVT.tpr

Korištenjem dobivene .tpr datoteke pokreće se *NVT* simulacija uravnoteženja naredbom:

mdrun -s EstA_NVT.tpr -deffnm EstA_NVT -x EstA_NVT &



Opcijom `-x` biramo zapis tzv. prijenosnog oblika trajektorije s ekstenzijom `.xtc`. Takav oblik je smanjene preciznosti, ali zauzima manje računalne memorije i brže se učitava u programe za vizualizaciju, pa je stoga vrlo praktičan.

Nakon završetka *NVT* ekvibracije, potrebno je pokrenuti *NPT* ekvibraciju koja će se na nju nastavljati. U tu svrhu koristimo datoteku `.cpt` (*checkpoint*) koja sadrži zapis posljednje simulacijom dobivene reprezentacije sustava sa svim koordinatama, brzinama i energijama u visokoj preciznosti te omogućava ponovno pokretanje simulacije od te točke ili produljivanje simulacije. Datoteka `NPT.tpr` priredit će se naredbom:

```
gmx grompp -f NPT.mdp -c EstA_NVT.gro -p EstA.top -t EstA_NVT.cpt -o EstA_NPT.tpr
```

NPT simulacija uravnoteženja pokreće se naredbom:

```
mdrun -s EstA_NPT.tpr -deffnm EstA_NPT -x EstA_NPT &
```

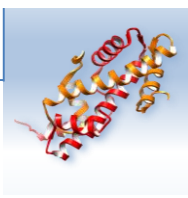
Naredbom:

```
g_energy -f EstA_NPT.edr -o EstA_NPT_energyX.xvg
```

moguće je dobiti grafove vremenske ovisnosti različitih parametara sustava (potencijalne i kinetičke energije, temperature, tlaka, itd.) koje je zatim jednostavno vizualizirati u programu *Grace*:

```
xmgrace EstA_NPT_energyX.xvg
```

Naravno, iste naredbe mogu se koristiti i za analizu *NVT* simulacije uravnoteženja.

**Pitanja:**

1. Koje su razlike između NVT.mdp i NPT.mdp datoteka?
2. Grafički prikazati vremenske ovisnosti: a) temperature, b) volumena, c) tlaka, d) gustoće otapala, e) kinetičke energije, f) potencijalne energije sustava tijekom simulacija uravnoteženja.
3. Koliko često se zapisuje stanje sustava u trajektoriju tijekom simulacija uravnoteženja?
4. Koliki je vremenski korak korišten tijekom simulacija uravnoteženja?
5. Koji termostat je korišten tijekom simulacija uravnoteženja?
6. Kolike su dimenzije simulacijske ćelije (poliedra) prije i nakon simulacija uravnoteženja?

Vježba 10.2 Simulacije uravnoteženja programom Amber

Zadatak 6.2 Pripremiti i provesti *NVT* i *NPT* simulacije uravnoteženja prethodno minimizirane strukture ljudske NAG kinaze u kompleksu s NAG u trajanju od 200 ps (*NVT*) i 100 ps (*NPT*) programom *Amber*.

Upute:

Odgovarajuće .in datoteke za simulacije uravnoteženja (*NVT.in* i *NPT.in*) su pripremljene i nalaze se na lokalnom računaru. Potrebno ih je otvoriti u tekstualnom editoru te provjeriti specificirane opcije i, po potrebi, neke od njih izmijeniti. Obratiti pozornost na to slaže li se broj aminokiselina simuliranog sustava s upisanim brojem aminokiselina čije je kretanje ograničeno u *NVT* fazi ekvibracije. Na temelju *NVT.in* i .prmtop datoteka, te .rst datoteke dobivene minimizacijom sustava u prethodnoj vježbi pokrenuti *NVT* ekvibraciju naredbom:

```
sander -p cplx-NAG-kinase.prmtop -c cplx-NAG-kinase_min5.rst -ref  
cplx-NAG-kinase_min5.rst -i NVT.in -o cplx-NAG-kinase_NVT.out -r cplx-  
NAG-kinase_NVT.rst -x cplx-NAG-kinase_NVT.mdcrd -inf cplx-NAG-  
kinase_NVT.info
```

Nakon *NVT* ekvilibracije sustava nastaju četiri izlazne datoteke: *.out*, *.inf.*, *.rst* i *mdcrd*. Za razliku od *.rst* datoteke dobivene nakon minimizacije, nova *.rst* datoteka, uz informacije o finalnim koordinatama sustava i dimenziji simulacijske kutije, sadrži i konačne brzine atoma sustava. Datoteka *.mdcrd* sadrži koordinate sustava tijekom trajektorije.

NPT ekvilibracija pokreće se korištenjem *.prmtop* i *.rst* datoteka dobivenih *NVT* ekvilibracijom te *NPT.in* datoteke naredbom:

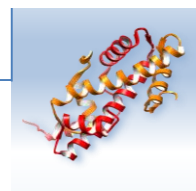
```
sander -p cplx-NAG-kinase.prmtop -c cplx-NAG-kinase_NVT.rst -ref cplx-NAG-kinase_NVT.rst -i NPT.in -o cplx-NAG-kinase_NPT.out -r cplx-NAG-kinase_NPT.rst -x cplx-NAG-kinase_NPT.mdcrd -inf cplx-NAG-kinase_NPT.info
```

Upute za izvođenje navedenih računa na više procesora paralelno korištenjem programa *sander.mpi* nalaze se u dodatku.

Po završetku *NVT* i *NPT* ekvilibracije provjeriti dobivene *.out* datoteke u tekstualnom editoru te vizualizirati strukturu uravnoteženog sustava u vizualizacijskom programu *VMD*.

Pitanja:

1. U programu *VMD* napraviti sliku na kojoj su preklopljene i različito obojane strukture kompleksa prije i nakon *NVT* i *NPT* ekvilibracije. Istaknuti dijelove strukture u kojima je zamijećena najznačajnija promjena.
2. Do koje se konačne temperature povisuje temperatura sustava tijekom *NVT* ekvilibracije?
3. Koje su osnovne razlike u ulaznim *NVT.in* i *NPT.in* datotekama?



Vježba 11: Simulacija molekulske dinamike (MD)

Nakon uravnoteženja sustava simulacijama uravnoteženja (ekvilibracije), slijedi MD simulacija koja se često naziva i produkcijskom fazom. Poželjno je da simulacija traje čim duže kako bi bila zadovoljena ergotska hipoteza. Učestalost spremanja stanja sustava tijekom simulacije, odnosno vremenski razmak dvaju susjednih stanja sustava u trajektoriji, ovisi o raspoloživim računalnim resursima te o detaljnosti kojom se želi proučiti sustav. Svrha ove simulacije jest istraživanje strukturnih i dinamičkih svojstava sustava na mikroskopskoj razini.

Vježba 11.1 Simulacije molekulske dinamike programom Gromacs

Zadatak 6.3 Pripremiti i provesti MD simulaciju prethodno uravnoteženog sustava katalitičke domene esteraze EstA iz bakterije *Pseudomonas aeruginosa* u trajanju od 10 ns.

Upute:

Potrebno je modificirati .mdp datoteku s ciljem pripreme MD simulacije u trajanju od 10 ns uz vremenski korak od 1 fs te spremanje stanja sustava u datoteku trajektorije svaku ps simulacije. Odgovarajuća .mdp datoteka za MD simulaciju se nalazi na lokalnom računaru. Potrebno ju je otvoriti u tekstualnom editoru te provjeriti specificirane opcije i, po potrebi, neke od njih izmijeniti. Na temelju MD.mdp i .top datoteka, te .gro i .cpt datoteka dobivenih simulacijama uravnoteženja sustava (prethodna vježba), potrebno je pripremiti .tpr datoteku za MD simulaciju naredbom:

```
gmx grompp -f MD.mdp -c EstA_NPT.gro -p EstA.top -t EstA_NPT.cpt -o EstA_MD.tpr
```

MD Korištenjem dobivene .tpr datoteke pokreće se MD simulacija naredbom:

```
mdrun -s EstA_MD.tpr -deffnm EstA_MD -x EstA_MD &
```

Kao i u prethodnoj vježbi, naredbom **g_energy** i programom *Grace* moguće je dobiti grafove vremenske ovisnosti različitih parametara sustava (potencijalne i kinetičke energije, temperature, tlaka, ...).

Pitanja:

1. Koje su razlike između MD.mdp datoteke i .mdp datoteka korištenih za simulacije uravnoteženja?
2. Grafički prikazati vremenske ovisnosti: a) temperature, b) volumena, c) tlaka, d) gustoće otapala, e) kinetičke energije, f) potencijalne energije sustava tijekom MD simulacije.
4. Koliko često se zapisuje stanje sustava u datoteku trajektorije tijekom MD simulacije?
5. Koji termostat je korišten tijekom MD simulacije

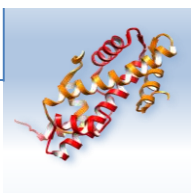
Vježba 11.2 Simulacije molekulske dinamike programom
Amber

Zadatak 6.4 Pripremiti i provesti MD simulaciju prethodno uravnoteženog sustava ljudske NAG kinaze u kompleksu s NAG u trajanju od 10 ns.

Upute:

Odgovarajuća MD.in datoteka za MD simulaciju nalazi se na lokalnom računalu. Potrebno ju je otvoriti u tekstualnom editoru te provjeriti specificirane opcije i, po potrebi, neke od njih izmijeniti. Korištenjem MD.in i .prmtop datoteka, te .rst datoteke dobivene NPT ekvibracijom sustava pokrenuti MD simulaciju naredbom:

```
sander -p cplx-NAG-kinase.prmtop -c cplx-NAG-kinase_NPT.rst -ref cplx-NAG-kinase_NPT.rst -i MD.in -o cplx-NAG-kinase_MD.out -r cplx-NAG-kinase_MD.rst -x cplx-NAG-kinase_MD.mdcrd -inf cplx-NAG-kinase_MD.info
```



Budući da je računalno zahtjevna, MD simulaciju je poželjno izvoditi u dva uzastopna intervala od po 5 ns te na više procesora paralelno ili na grafičkoj kartici. Upute za takvo izvođenje MD simulacija nalaze se u dodatku.

Zadatak 6.5 Analizirati rezultate molekulske dinamike grafičkim prikazom promjene temperature, tlaka, ukupne energije i potencijalne energije u vremenu.

Upute:

Kao što je već spomenuto u prethodnim poglavljima, informacije o promjenama u sustavu tijekom simulacija zapisane su u izlaznoj .out datoteci koja se može otvoriti tekstualnim editorom. Kako bi se dobile zasebne informacije o promjeni pojedinih parametara poput temperature, tlaka, energije sustava u vremenu koristi se već pripremljena skripta ***process_mdout.perl*** koja dolazi s programskim paketom *Amber*.

Za početak, potrebno je kreirati novi direktorij ANALIZA koji sadrži kopiju izlaznih NVT.out, NPT.out i MD.out datoteka. U navedenom direktoriju pokrenuti skriptu *process_mdout.perl* naredbom:

`$AMBERHOME/bin/process_mdout.perl NVT.out`

Ponoviti istu naredbu za NPT.out i MD.out datoteku. Nastale datoteke NVT.EPTOT, NPT.EPTOT i MD.EPTOT koje sadrže informacije o promjeni potencijalne energije sustava u vremenu mogu se jednostavno sjediniti naredbom **cat**:

`cat NVT.EPTOT NPT.EPTOT MD.EPTOT > NVT_NPT_MD.EPTOT`

Rezultate grafički prikazati u programu *Grace* naredbom:

`xmgrace NVT_NPT_MD.EPTOT`

Grafovi promjene temperature, tlaka te potencijalne, ukupne i kinetičke energije tijekom trajektorije daju informacije o mogućim neželjenim promjenama u sustavu tijekom simulacija. Graf promjene temperature u vremenu treba sadržavati polagani uspon do željene temperature u *NVT* fazi ekvibracije, nakon

čega se vrijednost temperature ne mijenja značajno do kraja produktivne faze. Ustaljeni grafovi s blagim promjenama očekuju se i u slučaju prikaza promjene potencijalne i kinetičke energije u vremenu. Bilo kakvi oštri skokovi i jako velike promjene vrijednosti ukazuju na problem u simulacijskom protokolu, lošu početnu strukturu, preveliki simulacijski korak, lošu parametrizaciju itd.

Pitanja:

1. Grafički prikazati vremenske ovisnosti: a) temperature, b) volumena, c) tlaka, d) kinetičke, potencijalne i ukupne energije sustava tijekom simulacija uravnoteženja i produkcijske faze. Vremenske ovisnosti kinetičke, potencijalne i ukupne energije prikazati na istom grafu.
2. Komentirati vremensku ovisnost temperature sustava tijekom simulacija uravnoteženja i produkcijske faze.
3. Iz grafa očitajte kada je postignuta najniža vrijednost potencijalne energije, a zatim precizno odredite trenutak u kojem je postignuta najniža vrijednost potencijalne energije korištenjem naredbe **awk** za koju su upute dane na dnu stranice.
4. Koje su razlike u simulacijskim protokolima (.in datotekama) korištenim za simulacije uravnoteženja i produktivnu fazu molekulske dinamike?
5. Koji je termostat korišten tijekom produktivne faze MD simulacije?
6. Koliko se često zapisuje stanje sustava u .out datoteku trajektorije tijekom produkcijske faze?

Upute za korištenje awk naredbe:

Najniža vrijednost energije se može jednostavno pronaći naredbom awk*:

```
awk '{if($2<min) {min=$2;print $1" "min}}' NVT_NPT_MD.EPTOT
```

Posljednja ispisana vrijednost odgovara najnižoj energiji (npr: 4567.000 -567.98).

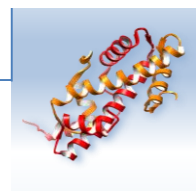
Precizno određivanje koraka tijekom trajektorije u kojem je postignuta najniža energija može se pronaći naredbom grep[§] u odgovarajućoj OUT datoteci, primjerice:

```
grep -567.98 MD.out
```

***awk** je jezik za skriptiranje koji dolazi kao dio *Linux* operativnog sustava. Koristi se za jednostavnu manipulaciju podacima iz terminala.

§**grep** naredba je također dio *Linux* operativnog sustava, a koristi se za jednostavno i brzo pretraživanje tekstualnih datoteka u terminalu. Pretraživanje se temelji na unosu riječi ili fraze koja se potom pretražuje u linijama teksta definirane datoteke, primjerice:

```
grep trazena_rijec ime_datoteke.txt
```

7. Analiza simulacije molekulske dinamike (MD)

Analiza simulacije molekulske dinamike uključuje **kvalitativnu** i **kvantitativnu** analizu. Sve promjene u strukturi i dinamici makromolekule koje su opažene vizualizacijom potrebno je kvantificirati mjerenjem promjene strukturnih varijabli (udaljenosti, kutovi, radijus giracije, itd.) ili dinamičkih svojstava (npr. fluktuacije) tijekom simulacije.

Vizualizacija trajektorije vrlo je važan korak u analizi MD simulacije. Kombiniranje različitih molekulskih prikaza, ubrzavanje i usporavanje vizualizacije trajektorije, promatranje cjelokupne strukture ili fokusiranje na manje dijelove strukture, samo su neke od opcija koje je potrebno koristiti za kvalitetnu analizu trajektorije putem vizualizacije. Bilo kakve promjene u strukturi, opažene tijekom vizualizacije, potrebno je prikazati slikama te kvantificirati mjerenjem odgovarajućeg strukturnog svojstva analizom trajektorije generirane tijekom simulacije ili simulacija. Također, promjene u dinamičkim svojstvima (npr. veća rigidnost ili fleksibilnost dijela strukture) potrebno je kvantificirati (npr. mjerenjem fluktuacija aminokiselina). Postoje različiti programi koji omogućavaju vizualizaciju trajektorija dobivenih MD simulacijama. Jedan od najčešće korištenih programa u akademskoj zajednici je program *Visual Molecular Dynamics (VMD)* [1] koji je predstavljen u prvom poglavlju ovog priručnika. Osim velikog broja različitih prikaza modela bioloških makromolekula, što uvelike pridonosi kvalitetnoj analizi trajektorije, program *VMD* pruža i puno opcija za kvantifikaciju određenih strukturnih varijabli, poput mjerenja udaljenosti, kutova i sl. Nadalje, programom *VMD* jednostavno je mijenjati format datoteke u kojoj je trajektorija zapisana, primjerice prevesti trajektoriju iz formata u kojem ju *Gromacs* zapisuje (.trr ili .xtc) u format u kojem ju *Amber* zapisuje (.crd) ili obrnuto. Prevođenje trajektorije iz jednog u drugi format osobito je korisno prilikom provođenja kvantitativne analize, ovisno o tome koje se opcije za analizu trajektorije pojedinog programskog paketa žele koristiti.

Kvantitativna analiza MD trajektorija temelji se na kvantitativnom određivanju promjena strukturnih i dinamičkih svojstava sustava tijekom MD simulacije. Osnovne analitičke tehnike uključuju određivanje promjene korijena srednjeg kvadrata udaljenosti (*RMSD*), analizu fluktuacija (*RMSF*) i računanje radijusa giracije (*Rg*) tijekom MD simulacije.

Korijen srednjeg kvadrata udaljenosti, odnosno *RMSD* (engl. *Root Mean Square Deviation*), mjera je prosječne udaljenosti ekvivalentnih atoma dvije strukture, a time i mjera sličnosti dviju struktura. Najčešće se koristi kao statistički pokazatelj sličnosti dviju konformacija iste molekule, ali može se upotrijebiti i za uspoređivanje različitih molekula. Manja *RMSD* vrijednost predstavlja veću sličnost, shodno tome, *RMSD* vrijednost između istih konformacija je nula. Računa se prema formuli:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2} \quad (13)$$

gdje je N ukupan broj atoma koje uspoređujemo, a δ_i udaljenost između dva ekvivalentna atoma. Često je jedna od prvih kvantitativnih analiza trajektorije upravo računanje *RMSD* vrijednosti proteinske okosnice tijekom MD simulacije. Pri tome se sukcesivne konformacije dobivene tijekom simulacije sravnjuju i uspoređuju s početnom konformacijom. Na takav način se dolazi do informacije o mijenjanju proteinske strukture tijekom simulacije.

Analiza fluktuacija temelji se na računanju fluktuacija pojedinih atoma, odnosno *RMSF* (engl. *Root-Mean-Square Fluctuation*) vrijednosti, tijekom MD simulacije. *RMSF* vrijednost je mjera uprosječenog fluktuiranja pojedinog atoma, aminokiseline ili dijela proteina tijekom MD simulacije. *RMSF* vrijednosti pokazuju koliko neki atom, aminokiselina ili proteinska regija fluktuiraju tijekom simulacije što omogućava istraživanje njihovih dinamičkih svojstava, odnosno omogućava identifikaciju fleksibilnih i rigidnih dijelova strukture makromolekule. Računa se prema izrazu:

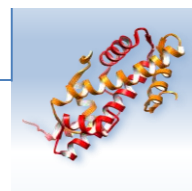
$$RMSF = \sqrt{\frac{1}{t} \sum_{i=1}^t (r_i - r_{ref})^2} \quad (14)$$

gdje je r_i položaj atoma u trenutku i tijekom simulacije, r_{ref} položaj atoma u referentnoj strukturi, t ukupno vrijeme simulacije tijekom kojeg se računa *RMSF* vrijednost. Iz *RMSF* vrijednosti jednostavno se mogu izračunati B-faktori koji se koriste u kristalografiji korištenjem formule:

$$B_j = \frac{8\pi}{3} RMSF_j \quad (15)$$

gdje je $RMSF_j$ vrijednost fluktuacije j -tog atoma tijekom MD simulacije.

Radius giracije (Rg) općenito se može definirati kao korijen srednje kvadratne udaljenosti dijelova nekog objekta od središta mase tog objekta (ili,

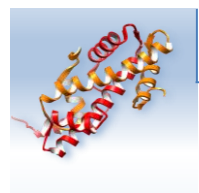


ponekad, osi rotacije). U biofizičkim sustavima, radijus giracije se može definirati kao korijen srednje sume kvadrata udaljenosti pojedinih monomera (atoma ili čitavih aminokiselina ili nukleotida) od središta mase biopolimera, odnosno biološke makromolekule. Računa se prema formuli:

$$Rg = \sqrt{\frac{\sum_{i=1}^{N_{ak}} r_i^2 m_i}{\sum_{i=1}^{N_{ak}} m_i}} \quad (16)$$

gdje je r_i udaljenost i -tog monomera od središta mase biopolimera, m_i njegova masa, a N_{ak} broj monomera u biopolimeru. Radijus giracije predstavlja vrlo korisnu procjenu volumena biološke makromolekule ili dijela biološke makromolekule.

U sklopu ovog priručnika, osnovne kvantitativne analitičke tehnike, koje se uobičajeno rade pri analizi MD simulacije, provest će se koristeći oba programska paketa, *Gromacs* i *Amber*. Pri tome će se za analize programom *Gromacs* koristiti trajektorija simulacije katalitičke domene esteraze EstA iz bakterije *Pseudomonas aeruginosa*, a za analize programom *Amber* trajektorija simulacije ljudske NAG kinaze u kompleksu s NAG. Za vizualizaciju grafova dobivenih analizom MD trajektorije koristit će se besplatno dostupni program *Grace*, ali može se koristiti i bilo koji drugi odgovarajući program, poput programa *Excel* koji je sastavni dio *MS Office* paketa. Osim prethodno objašnjenih osnovnih kvantitativnih analitičkih tehnika, postoji i niz drugih tehnika i opcija koje pružaju oba programska paketa. Mogućnosti i alati za analizu MD simulacija su vrlo brojni, a izbor određene analitičke tehnike ovisi o strukturnim i dinamičkim svojstvima sustava koje se želi istražiti. Primjerice, može se pratiti stvaranje i razmatanje sekundarnih struktura tijekom simulacije, promjena otapalu dostupne površine, nastanak i cijepanje vodikovih veza, međusobno pomicanje dijelova strukture makromolekule, interakcije s otapalom, itd.



Vježba 12: Vizualizacija MD trajektorije

Vizualizacija trajektorije vrlo je važan korak u analizi MD simulacije. U ovoj vježbi vizualizirat će se MD trajektorije dobivene programima *Gromacs* i *Amber*. Kombiniranjem različitih prikaza modela molekula i atoma, ubrzavanjem i usporavanjem vizualizacije trajektorije, promatranjem cjelokupne strukture ili fokusiranjem na manje dijelove strukture, potrebno je pronaći strukturne i dinamičke promjene simuliranih sustava. Pri tom je poželjno isprobavati čim više opcija koje nudi program *VMD* koji će se koristiti za vizualizaciju trajektorija.

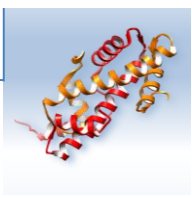
Vježba 12.1 Vizualizacija MD trajektorije dobivene programom *Gromacs*

Zadatak 7.1 Vizualizirati trajektoriju MD simulacije katalitičke domene esteraze EstA iz bakterije *Pseudomonas aeruginosa* u programu *VMD* i odgovoriti na zadana pitanja.

Upute:

Potrebno je učitati trajektoriju MD simulacije u program *VMD*. Prvo se mora učitati datoteka koja sadrži topologiju sustava: u glavnom prozoru programa *VMD* (***VMD main***) odabrati opciju ***New molecule*** → ***Browse*** → *EstA.top* → ***OK*** → ***Load***. Trajektorija se učitava desnim klikom na *EstA.top* u glavnom prozoru, zatim → ***Load Data Into Molecule*** → ***Browse*** → odabrati *.xtc* ili *.trr* datoteku → ***Load***. Ukoliko je trajektorija rascjepkana u više *.xtc* ili *.trr* datoteka, učitati ih jednu po jednu prema redoslijedu nastajanja. Nakon učitavanja trajektorije preporučljivo je centrirati protein tijekom simulacije kako bi se lakše uočile značajne promjene tijekom trajektorije. Centriranje sustava se postiže odabirom opcije ***Extensions*** → ***Analysis*** → ***RMSD Visualizer Tool*** → ***Backbone*** → ***ALIGN***.

U prozoru ***Graphical Representations*** odabrati željeni molekularni prikaz (za početak je preporučljivo odabrati prikaz *NewCartoon*) te pogledati trajektoriju i nastojati uočiti promjene u strukturi i dinamici katalitičke domene enzima EstA. Nakon što je uočena promjena dijela proteinske okosnice, preporuča se promjena prikaza s ciljem detaljnijeg proučavanja uočene promjene. Esteraza EstA spada u



skupinu serinskih esteraza te sadrži katalitičku trijadu u aktivnom mjestu. Svakako je potrebno obratiti pažnju na postojanost vodikovih veza katalitičke trijade tijekom MD simulacije. Mjerenje udaljenosti u programu *VMD* opisano je u prvoj cjelini ovih vježbi (vježba 3). Odabrana udaljenost u izborniku ***VMD Main***, zatim ***Graphics***, onda ***Labels*** (***VMD Main*** → ***Graphics*** → ***Labels***), može se prikazati na grafu i spremiti u obliku .xvg formata kojeg čita program *Grace*. Na isti način grafički se može prikazati i promjena bilo koje druge strukturne varijable (npr. kut, torzijski kut) tijekom trajektorije.

Ponekad se dogodi da tijekom simulacije protein „izađe“ iz simulacijske ćelije (poliedra) što znatno otežava vizualizaciju sustava u programu *VMD*. U tom je slučaju najbolje prvo modificirati dobivenu trajektoriju opcijom *gmx trjconv* koja centrira protein u simulacijskoj ćeliji:

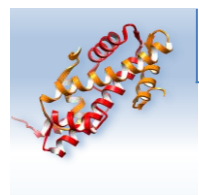
```
gmx trjconv -s EstA_MD.tpr -f EstA_MD1.trr -pbc mol -center 0 -o EstA_MD1_center.trr
```

Opcija *gmx trjconv* nudi još cijeli niz modifikacija trajektorije poput: izrezivanja dijela trajektorije, spremanje točno određenog *frame*-a u obliku .pdb datoteke, prebacivanje iz jednog formata u drugi, promjena vremenskog koraka trajektorije itd. Sve mogućnosti navedene naredbe mogu se izlistati upisivanjem:

```
gmx trjconv -h
```

Pitanja:

1. Koje promjene u konformaciji katalitičke domene esteraze EstA se uočavaju tijekom MD simulacije? Potrebno ih je prikazati slikama.
2. Središnja aminokiselina katalitičke trijade esteraze EstA je His289. Koje su druge dvije aminokiseline katalitičke trijade esteraze EstA?
3. Napraviti sliku katalitičke domene esteraze EstA s istaknutim aminokiselinama katalitičke trijade i označenim vodikovim vezama.
4. Jesu li vodikove veze katalitičke trijade postojane tijekom MD simulacije katalitičke domene esteraze EstA?



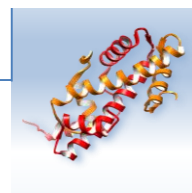
Vježba 12.2 Vizualizacija MD trajektorije dobivene programom *Amber*

Zadatak 7.2 Vizualizirati trajektoriju MD simulacije ljudske NAG kinaze u kompleksu s NAG u programu *VMD* i odgovoriti na zadana pitanja.

Upute:

Potrebno je učitati trajektoriju MD simulacije ljudske NAG kinaze u kompleksu s NAG u programu *VMD*. Prvo se učitava datoteka koja sadrži toplogiju sustava: u glavnom prozoru programa *VMD* odabrati opciju **New molecule** → **Browse** → *cplx-NAG-kinase.prmtop* → **OK** → **Load**. Trajektorija se učitava desnim klikom na *cplx-NAG-kinase.prmtop* u glavnom prozoru → **Load Data Into Molecule** → **Browse** → odabrati *cplx-NAG-kinase_MD.mdcrd* datoteku → u traci **Determine File Type** odabrati format NetCDF (AMBER, MMTK) → **Load**. Ukoliko je trajektorija rascjepkana u više .mdcrd datoteka, učitati ih jednu po jednu prema redosljedu nastajanja. Nakon učitavanja trajektorije preporučljivo je centrirati protein tijekom simulacije kako bi se lakše uočile značajne promjene tijekom trajektorije. Centriranje sustava se postiže odabirom opcije **Extensions** → **Analysis** → **RMSD Visualizer Tool** → **Backbone** → **ALIGN**.

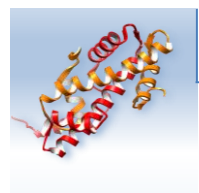
Potrebno je obratiti pažnju na postojanost vodikovih veza između NAG i aktivnog mjesta enzima tijekom MD simulacije. Kako bi se lakše uočile interakcije između aktivnog mjesta i NAG molekule, najbolje je ligand i aminokiseline u neposrednoj blizini liganda prikazati opcijom *Lines* ili *Licorice*, dok je ostatak proteina prikazan modelom vrpci. To se postiže otvaranjem prozora *Graphical Representations* i upisivanjem: „**within 4 of resname NAG**“, u traku **Selected atoms** čime se odabiru svi atomi koji su na udaljenosti manjoj od 4 Å od NAG. Ukoliko se žele uključiti čitave aminokiseline koje imaju barem jedan atom na udaljenosti manjoj od 4 Å od NAG, potrebno je koristiti odabir: „**same residue as within 4 of resname NAG**“. Kako bi se ovaj prikaz primijenio na cijelu trajektoriju potrebno je u istom prozoru odabrati karticu **Trajectory** i označiti opcije **Update Selection Every Frame** i **Update Color Every Frame**. Za praćenje vodikovih veza koje se ostvaruju između NAG i susjednih aminokiselina, koristi se isti odabir (**within 4 of resname NAG**), ali je potrebno pod opcijom **Drawing methods** odabrati opciju *H-bonds*. Mjerenje udaljenosti u programu



VMD opisano je u prvoj cjelini ovih vježbi (vježba 3). Odabrana udaljenost može se prikazati na grafu odabirom **VMD Main** → **Graphics** → **Labels** → **Bonds** → označiti željenu vezu → **Graph** → **Graph**. Dobiveni graf može se pohraniti u .xvg formatu koji se može otvoriti programom *Grace* koji je korišten u prethodnim vježbama.

Pitanja:

1. Koje aminokiseline aktivnog mjesta ljudske NAG kinaze ostvaruju interakciju s NAG tijekom MD simulacije? Napraviti sliku na kojoj su te aminokiseline istaknute prikazom *Licorice* te prikazati izmjerene udaljenosti između aminokiselina i NAG.
2. U programu *VMD* napraviti graf udaljenosti između NAG i Asp 152.
3. Dolazi li do značajnog pomicanja NAG u aktivnom mjestu ljudske NAG kinaze tijekom MD simulacije? Zašto?



Vježba 13: Definiranje poskupova sustava

Često je prilikom kvantitativne analize trajektorije MD simulacije potrebno izabrati samo manje dijelove sustava na čije se promjene tijekom simulacije želimo usredotočiti. Odabir tih dijelova (podsustava) u programu *Gromacs* omogućava **.ndx** datoteka, dok se u programu *Amber* koristi opcija **Amber atom mask**.

13.1 Definiranje podskupova sustava u programu *Gromacs*

Većina potprograma u programskom paketu *Gromacs* prihvaća **.ndx** datoteku kao ulaznu datoteku uključivanjem opcije **-n**.

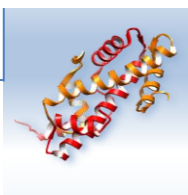
Zadatak 7.3 Napraviti **.ndx** datoteku za katalitičku domenu esteraze EstA iz bakterije *Pseudomonas aeruginosa* te definirati dijelove sustava koji su potrebni kako bi se našli odgovori na postavljena pitanja.

Upute:

Za stvaranje **.ndx** datoteke koristi se naredba:

```
gmx make_ndx -f EstA_MD.gro -o index.ndx
```

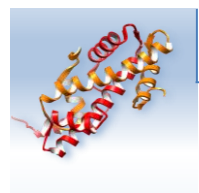
u slučaju da već postoji **.ndx** datoteka koju se želi nadograditi, ona se učitava u gornjoj naredbi opcijom **-n**. Nakon pokretanja gornje naredbe potrebno je specificirati nove grupe koje predstavljaju dijelove podsustava. Sintaksa prepoznaje Booleove operatore (AND=&, OR=|, NOT=!), aminokiseline koje se žele uključiti u novu grupu definiraju se s „r“, atomi s „a“, proteinski lanci s „chain“, itd. Detaljne informacije o sintaksi ove naredbe i definiranju pojedinih grupa dobivaju se upisivanjem „help“ nakon pokretanja gornje naredbe. Nakon definiranja željenih grupa, odnosno dijelova podsustava, program se zatvara upisivanjem „q“ i pritiskom tipke „Enter“.

**Pitanja:**

1. Napraviti index.ndx datoteku za katalitičku domenu esteraze EstA iz bakterije *Pseudomonas aeruginosa* i definirati novu grupu koja sadrži:
a) aminokiselinu 15, b) aminokiseline 26 i 37, c) regiju koja sadrži aminokiseline 211-221, d) aminokiseline katalitičke trijade, e) sve serine, f) sve Ca atome.
2. Iz index.ndx datoteke očitati ukupan broj atoma vode i iona.
3. U datoteci index.ndx spojiti grupe 18 i 21. Koje aminokiseline esteraze EstA sadrži nova grupa? Koliki je broj atoma nove grupe?

Vježba 13.2 Definiranje podsustava u programu Amber

Najčešći način definiranja atoma i bočnih ogranaka u programskom paketu *Amber* je korištenjem oznaka **mask**. Prilikom korištenja oznaka *mask*, aminokiseline se označuju oznakom „:”, atomi oznakom „@” ili pak AMBER tipovima atoma definiranim u polju sila korištenjem oznake „@%”. Oznaka „:*” označuje sve bočne ogranke, a „@*” sve atome. Osnovne oznake se mogu kombinirati s binarnim operatorima „&” (i), „|” (ili), „!” (ne) i oznakama udaljenosti „<” (unutar) i „>” (izvan). Razmaci oko oznaka operatora nisu bitni. Primjeri dodjeljivanja *Amber* oznaka *mask* prikazani su u tablici 2.



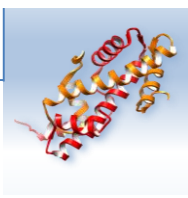
Tablica 2. Primjeri dodjeljivanja *Amber* oznaka *mask* atomima i bočnim ograncima.

Dodjeljivanje amber oznaka <i>mask</i> bočnim ograncima	
:1-10	= bočni ogranci od 1 do 10
:1,3,5	= bočni ogranci 1, 3 i 5
:1-3,7-20,29	= bočni ogranci od 1 do 3, od 7 do 20 i 29
:LYS	= svi lizini
:ALA,ARG	= svi arginini i alanini
Dodjeljivanje amber oznaka <i>mask</i> atomima	
@12,17	= atomi 12 i 17
@54-85	= svi atomi od 54 do 85
@12, 44-54	= atom 12 i atomi od 44 do 54
@CA	= svi atomi oznake CA (C-alfa atomi)
@CA,C,N	= svi atomi oznaka CA, C, N (atomi proteinske okosnice)
Primjeri kompleksnijih odabira	
:12@CA	= CA atom u bočnom ogranku 12
:@C= & !@CA,C	= svi ugljikovi atomi osim C-alfa i karbonilnog atoma okosnice
:2-57&!(@H=)	= svi atomi osim vodikovih atoma u bočnim ograncima od 2 do 57
:1 <@2.4	= svi atomi unutar 2.4 Å od bočnog ogranka 1

Zadatak 7.4. Proučiti način dodjeljivanja *Amber* oznaka *mask* prikazan u tablici 2 te odgovoriti na zadana pitanja.

Pitanja:

1. Korištenjem *Amber* oznaka *mask* definirati samo atome proteinske okosnice od bočnog ogranka 5 do 150.
2. Korištenjem *Amber* oznaka *mask* definirati sve atome osim vodikovih u aminokiselinama 5, 16, 17 i 18.



Vježba 14: Izračun RMSD vrijednosti

Promjena *RMSD* vrijednosti proteinske okosnice tijekom MD simulacije uobičajeno je jedna od prvih kvantitativnih analiza trajektorije. Naime, iz dobivenog grafa može se procijeniti je li se sustav uravnotežio ili se još mijenja pa je potrebno nastaviti simulaciju. U ovoj vježbi, potrebno je izračunati i grafički prikazati promjenu *RMSD* vrijednosti proteinske okosnice tijekom MD simulacija programima *Gromacs* i *Amber*.

Vježba 14.1 Izračun RMSD vrijednosti u programu *Gromacs*

Zadatak 7.5 Grafički prikazati *RMSD* vrijednosti proteinske okosnice tijekom MD simulacije katalitičke domene esteraze EstA iz bakterije *Pseudomonas aeruginosa*.

Upute:

Graf *RMSD* vrijednosti tijekom trajektorije u programu *Gromacs* generira se naredbom:

```
gmx rms -s EstA_MD.tpr -f EstA_MD.xtc -o EstA_RMSD.xvg
```

opcijom `-s` odabire se referentna konformacija s kojom će se uspoređivati sve ostale konformacije. Nakon upisivanja naredbe potrebno je odabrati skup atoma prema kojima će se vršiti usporedba, preporuča se odabir atoma proteinske okosnice. Dobivenu `.xvg` datoteku potrebno je učitati u program *Grace*:

```
xmgrace EstA_RMSD.xvg
```

gdje se može spremati kao slika odabirom opcije „Print to File“ te odabirom vrste slikovne datoteke. Navedene upute za vizualizaciju `.xvg` datoteka i njihovo spremanje u obliku slika koristit će se i u narednim vježbama.

Pitanja:

1. Grafički prikazati *RMSD* vrijednosti proteinske okosnice tijekom MD simulacije katalitičke domene esteraze EstA iz bakterije *Pseudomonas aeruginosa*.
2. Grafički prikazati *RMSD* vrijednosti proteinske okosnice tijekom *NVT* i *NPT* simulacija uravnoteženja katalitičke domene esteraze EstA te sve grafove spojiti u jedan.
3. Napraviti *RMSD* graf samo za regiju koju tvore aminokiseline 231-241. Koji element sekundarne strukture tvore aminokiseline te regije?
4. Kako se na temelju promjene *RMSD* vrijednosti proteinske okosnice tijekom simulacije može zaključivati o uravnoteženosti sustava?
5. Na temelju grafa *RMSD* vrijednosti tijekom simulacije pronaći strukture koje su na početku i na kraju većeg skoka *RMSD* vrijednosti te ih preklopiti u programu *VMD*.

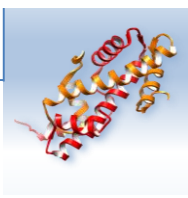
Vježba 14.2 Izračun *RMSD* vrijednosti u programu *Amber*

Zadatak 7.6 Ako je MD simulacija provedena tako da su generirane dvije uzastopne trajektorije, sjediniti ih u jednu cjelovitu trajektoriju formata *.binpos* koji zauzima manje računalne memorije od *.mdcrd* formata. Korištenjem cjelovite trajektorije izračunati i grafički prikazati *RMSD* vrijednosti proteinske okosnice tijekom MD simulacije ljudske NAG kinaze u kompleksu s NAG te odgovoriti na zadana pitanja.

Upute:

U programskom paketu *Amber* sve analize i transformacije trajektorija provode se u potprogramu *cpptraj* koji se može koristiti na interaktivni način ili putem skripte (priprema *.in* skripte za *cpptraj* analizu opisana je u dodatku). U terminalu datoteke koja sadrži trajektorije potrebno je započeti interaktivni način rada unosom naredbe:

cpptraj



Za početak je potrebno učitati topologiju simuliranog sustava:

parm cplx-NAG-kinase.prmtop

zatim i trajektorije:

trajin cplx-NAG-kinase_MD_1.mdcrd

trajin cplx-NAG-kinase_MD_1.mdcrd

Dvije uzastopne trajektorije sjediniti u jednu formata .binpos naredbom:

trajout cplx-NAG-kinase_MD_10ns.binpos

Kako bi cpptraj započeo s obradom ili analizom trajektorija potrebno je nakon unosa svih naredbi upisati:

run

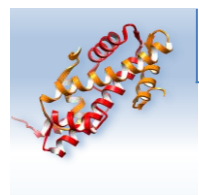
Po završetku obrade trajektorije ili analize iz interaktivnog načina rada izlazi se naredbom:

quit

Budući da će se *RMSD* vrijednosti izračunati koorištenjem cjelovite trajektorije .binpos formata, potrebno je izaći iz interaktivnog načina rada i započeti novi u kojem će se naredbom *trajin* učitati cjelovita trajektorija cplx-NAG-kinase_MD_10ns.binpos. Vrijednost *RMSD* je potrebno izračunati samo za atome proteinske okosnice (CA, C i N) stoga je u naredbi potrebno definirati navedeni podskup atoma korištenjem *Amber* oznake *mask*:

rmsd tofirst :1-344@CA,C,N mass out RMSD_NAG-kinase_backbone.dat

Opcija *tofirst* nalaže da se kao referentna struktura koristi prva struktura u trajektoriji. Ukoliko se kao referentna struktura želi koristiti neka druga struktura spremljena u obliku .pdb datoteke, potrebno je odabrati opciju **reference** i zatim

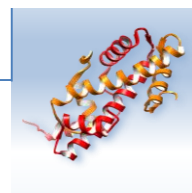


navesti ime .pdb datoteke . Nakon izlaska iz cpptraj interaktivnog načina rada grafički prikazati izračun *RMSD* vrijednosti naredbom:

xmgrace RMSD_NAG-kinase_backbone.dat

Pitanja:

1. Grafički prikazati *RMSD* vrijednosti proteinske okosnice tijekom MD simulacije ljudske NAG kinaze u kompleksu s NAG.
2. Grafički prikazati *RMSD* vrijednosti proteinske okosnice tijekom *NVT* i *NPT* simulacija uravnoteženja ljudske NAG kinaze u kompleksu s NAG te sve grafove spojiti u jedan.
3. Na temelju dobivenog *RMSD* grafa komentirajte stabilnost i eventualne konformacijske promjene proteina.
4. Na temelju grafa *RMSD* vrijednosti tijekom simulacije pronaći strukture koje su na početku i na kraju većeg skoka *RMSD* vrijednosti. Preklopiti navedene strukture u programu *VMD* te istaknuti njihove razlike.
5. Napraviti *RMSD* graf u kojem će referentna struktura biti struktura u onom trenutku trajektorije u kojem je zabilježena najniža vrijednost potencijalne energije.



Vježba 15: Izračun *RMSF* vrijednosti

Računanje i grafički prikaz *RMSF* vrijednosti tijekom MD trajektorije uobičajeni je način identifikacije fleksibilnih i rigidnih dijelova simuliranog sustava. Dobiveni graf pokazuje nam koliko neki atom, aminokiselina ili proteinska regija fluktuiraju tijekom simulacije. Često se informacije dobivene na temelju analize *RMSF* vrijednosti koriste za detaljniju analizu trajektorije putem vizualizacije. Naime, *RMSF* graf nam ukazuje na fleksibilnije dijelove strukture što nas upućuje na koje dijelove strukture treba posebno obratiti pažnju tijekom vizualizacije kako bi se uočile strukturne promjene.

Vježba 15.1 Izračun *RMSF* vrijednosti u programu *Gromacs*

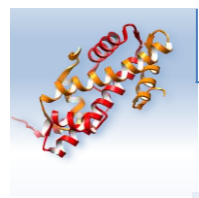
Zadatak 7.7 Napraviti graf *RMSF* vrijednosti aminokiselina katalitičke domene esteraze EstA iz bakterije *Pseudomonas aeruginosa* tijekom MD simulacije.

Upute:

U sklopu programa *Gromacs*, graf *RMSF* vrijednosti tijekom trajektorije može se dobiti naredbom:

```
gmx rmsf -s EstA_MD.tpr -f EstA_MD.xtc -o EstA_RMSF.xvg -res
```

opcija `-res` određuje računanje prosječne *RMSF* vrijednosti za čitavu aminokiselinu, bez uključivanja te opcije dobivaju se *RMSF* vrijednosti za pojedinačne atome. Nakon pokretanja naredbe potrebno je izabrati dio sustava za koji će se *RMSF* vrijednosti računati, preporuča se odabir „Protein“. Korištenjem gore navedene naredbe i opcije `-oq` dobiva se .pdb datoteka u kojoj se nalaze izračunate vrijednosti B-faktora.



Pitanja:

1. Grafički prikazati *RMSF* vrijednosti za MD simulaciju katalitičke domene esteraze EstA iz bakterije *Pseudomonas aeruginosa*.
2. Koristeći `index.ndx` datoteku napravljenu pod 1. napraviti *RMSF* graf koji sadrži vrijednosti samo za regiju između Thr211 i Ser221.
3. Grafički prikazati *RMSF* vrijednosti za MD simulaciju katalitičke domene esteraze EstA iz bakterije *Pseudomonas aeruginosa* u periodu 500-1000 ps.
4. Vrijednosti dobivene pod 1. i pod 2. prikazati na istom grafu korištenjem programa *Grace*.
5. Koje regije katalitičke domene esteraze EstA najviše fluktuiraju tijekom MD simulacije? Na temelju vizualizacije trajektorije pokušati pronaći objašnjenje za povećanu fleksibilnost pronađenih regija.
6. Napraviti `.pdb` datoteku katalitičke domene enzima EstA koja sadrži vrijednosti B-faktora izračunate na temelju *RMSF* vrijednosti. Dobivene vrijednosti usporediti s onima u datoteci koja je pronađena pretraživanjem baze PDB u vježbi 1.

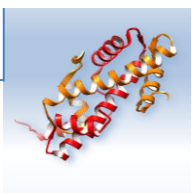
Vježba 15.2 Izračun *RMSF* vrijednosti u programu *Amber*

Zadatak 7.8 Napraviti graf *RMSF* vrijednosti aminokiselina tijekom MD simulacije ljudske NAG kinaze u kompleksu s NAG te odgovoriti na zadana pitanja. Prilikom izračuna koristiti cjelovitu **`cplx-NAG-kinase_MD_10ns.binpos`** trajektoriju dobivenu u prethodnoj vježbi.

Upute:

Izračun *RMSF* vrijednosti će se ostvariti korištenjem potprograma *cpptraj*. Potrebno je otvoriti interaktivni način rada, učitati topologiju sustava `cplx-NAG-kinase.prmtop` i cjelovitu trajektoriju `cplx-NAG-kinase_MD_10ns.binpos`. Naredba za izračun *RMSF* vrijednosti aminokiselina glasi:

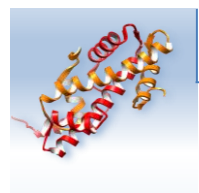
`rmsf tofirst :1-344 byres out RMSF_NAG-kinase_byres.dat`



Primijetiti da je, kao i u prethodnoj vježbi, ligand izostavljen iz izračuna. Opcija *byres* označava računanje prosječne *RMSF* vrijednosti za čitavu aminokiselinu, bez uključivanja te opcije dobivaju se *RMSF* vrijednosti za pojedinačne atome.

Pitanja:

1. Grafički prikazati *RMSF* vrijednosti za MD simulaciju ljudske NAG kinaze u kompleksu s NAG.
2. Koje aminokiseline najviše fluktuiraju? U programu *VMD* napraviti sliku strukture kompleksa nakon MD simulacije s istaknutim aminokiselinama kod kojih je primijećena najveća fluktuacija.
3. Grafički prikazati *RMSF* vrijednosti atoma liganda za cijelu MD simulaciju. Koji atomi najviše fluktuiraju? Zašto?
4. Napravite *RMSF* graf u kojem će referentna struktura biti struktura u onom trenutku trajektorije u kojem je zabilježena najniža vrijednost potencijalne energije. Preklopiti dobiveni graf s grafom iz prvog zadatka.



Vježba 16: Izračun radijusa giracije

Radijus giracije predstavlja procjenu volumena biološke makromolekule ili dijela biološke makromolekule. Također se može koristiti i kao kvantitativni pokazatelj promjena određenog dijela strukture.

Vježba 16.1 Izračun radijusa giracije u programu *Gromacs*

Zadatak 7.9 Savladati računanje radijusa giracije tijekom analize trajektorije MD simulacije katalitičke domene esteraze EstA iz bakterije *Pseudomonas aeruginosa* korištenjem programskog paketa *Gromacs*.

Upute:

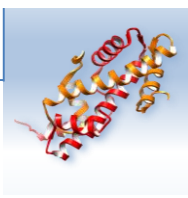
Računanje radijusa giracije u programu *Gromacs* omogućava naredba **gmx gyrate**. Računanje promjene radijusa giracije tijekom trajektorije provodi se naredbom:

gmx gyrate -s EstA_MD.tpr -f EstA_MD.xtc -o EstA_RG.svg

Ukoliko se želi računati radijus giracije čitavog proteina, nakon pokretanja naredbe potrebno je izabrati grupu koja definira protein. Ukoliko se želi računati radijus giracije samo manjeg segmenta strukture (npr. jedne domene), potrebno je taj segment prethodno definirati kao zasebnu grupu u .ndx datoteci, u gornju naredbu uključiti opciju -n i učitati .ndx datoteku te izabrati tu grupu nakon pokretanja naredbe.

Pitanja:

1. Napraviti graf promjene radijusa giracije katalitičke domene esteraze EstA iz bakterije *Pseudomonas aeruginosa* tijekom MD simulacije.
2. Definirati aminokiseline koje tvore aktivno mjesto esteraze EstA iz bakterije *Pseudomonas aeruginosa* te napraviti graf promjene radijusa giracije izabranih aminokiselina tijekom MD simulacije.



Vježba 16.2 Izračun radijusa giracije u programu Amber

Zadatak 7.10 Savladati računanje radijusa giracije tijekom analize trajektorije MD simulacije ljudske NAG kinaze u kompleksu s NAG te odgovoriti na zadana pitanja. Prilikom izračuna koristiti cjelovitu **cplx-NAG-kinase_MD_10ns.binpos** trajektoriju.

Upute:

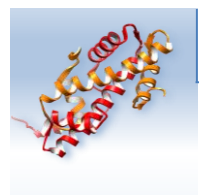
Potrebno je otvoriti interaktivni način rada potprograma *cpptraj*, učitati topologiju sustava *cplx-NAG-kinase.prmtop* te cjelovitu trajektoriju *cplx-NAG-kinase_MD_10ns.binpos*. Naredba za izračun radijusa giracije za cijeli enzim NAG kinazu bez liganda glasi:

radgyr (:1-344)&!@H out RoG_NAG-kinase.dat mass nomax

Primijetiti da su iz izračuna isključeni atomi vodika.

Pitanja:

1. Napraviti graf promjene radijusa giracije enzima ljudske NAG kinaze simuliranog u kompleksu s NAG i u izračun uključiti atome vodika. Na istom grafu prikazati promjene radijusa giracije proteina u čiji su izračun uključeni atomi vodika i izračun iz kojeg su isključeni atomi vodika. Da li se dobiveni grafovi jako razlikuju? Zašto?
2. Definirati barem tri aminokiseline koje interagiraju s NAG te napraviti graf promjene radijusa giracije tijekom trajektorije za te aminokiseline.



Vježba 17: Mjerenje udaljenosti središta masa

Često se prilikom analize trajektorije uočava konformacijska promjena koja se može opisati kao pomak jednog dijela proteinske strukture (proteinske domene, α -zavojnice, omče, itd.) u odnosu na drugu. Kvantificiranje takvog rezultata zahtijeva mjerenje udaljenosti između središta masa definiranih strukturnih segmenata tijekom simulacije.

Vježba 17.1 Mjerenje udaljenosti središta masa u programu Gromacs

Zadatak 7.11 Savladati mjerenje udaljenosti između središta masa strukturnih segmenata tijekom analize trajektorije MD simulacije u programu *Gromacs* te odgovoriti na postavljena pitanja.

Upute:

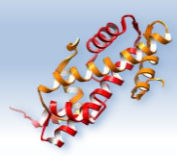
Mjerenje udaljenosti središta masa određenih regija sustava može se postići naredbom **gmx dist** u sklopu programa *Gromacs*. Pri tome je neophodno prethodno definirati željene regije u **index.ndx** datoteci. Upisivanjem naredbe:

```
gmx dist -s EstA_MD.tpr -f EstA_MD.xtc -n index.ndx -oall EstA_dist.xvg
```

potrebno je izabrati definirane grupe koje sadrže regije od interesa. Promjena udaljenosti izabranih središta masa tijekom simulacije prikazana je u .xvg datoteci koja se može grafički prikazati korištenjem programa *Grace*.

Pitanja:

1. Izmjeriti udaljenost središta masa regija esteraze EstA iz bakterije *Pseudomonas aeruginosa* definiranih aminokiselinama 211–221 i 231–241 duž trajektorije te ju grafički prikazati.
2. Izmjeriti udaljenost središta masa aminokiselina katalitičke trijade esteraze EstA iz bakterije *Pseudomonas aeruginosa* te ih usporediti s prethodno mjerenim udaljenostima vodikovih veza.



3. Izmjeriti udaljenost regije definirane aminokiselinama 231–241 od središta mase enzima EstA iz bakterije *Pseudomonas aeruginosa* analizom generirane trajektorije.
4. Izmjeriti udaljenost pod 3. samo za period od 500. ps do 700. ps simulacije.

Vježba 17.2 Mjerenje udaljenosti središta masa u programu Amber

Zadatak 7.12 Savladati mjerenje udaljenosti između središta masa dijelova 3D strukture tijekom analize trajektorije MD simulacije ljudske NAG kinaze u kompleksu s NAG u programu *Amber* i odgovoriti na postavljena pitanja.

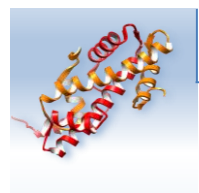
Upute:

Otvoriti interaktivni način rada potprograma *cpptraj*, učitati topologiju sustava *cplx-NAG-kinase.prmtop*, te također učitati i cjelovitu trajektoriju *cplx-NAG-kinase_MD_10ns.binpos*. Naredba za izračun udaljenosti između bočnih ogranaka 33 i 212 glasi:

distance :33 :212 out distance_Leu33-Arg121.agr noimage

Pitanja:

1. Prema uputama izračunati udaljenost između bočnih ogranaka Leu33 i Arg121. Dobiveni graf prikazati programom *Grace*.
2. Mijenja li se značajno udaljenost između bočnih ogranaka Phe208 i Leu39 koji s jedne strane zatvaraju aktivno mjesto? Grafički prikazati izračunatu udaljenost. S obzirom na svojstva aminokiselina, objasniti dobiveni rezultat.
3. Pokušati pronaći elemente strukture koji se značajno udaljavaju/približavaju tijekom simulacije. Izračunati njihovu udaljenost i grafički ih prikazati. Napraviti sliku na kojoj će navedeni elementi strukture biti istaknuti.



Vježba 18: Analitička metoda po izboru

U sklopu ovih uputa predstavljene su osnovne i najčešće korištene metode analize trajektorija MD simulacija. Postoji još niz drugih tehnika za analizu trajektorije MD simulacije koje se u određenim slučajevima mogu pokazati kao izuzetno korisne. Neke od tih tehnika u sklopu programa *Gromacs* navedene su na mrežnoj stranici: <http://manual.gromacs.org/>. U slučaju programskog paketa *Amber* niz drugih tehnika za analizu trajektorije se može pronaći u uputama dostupnima na poveznici: <https://ambermd.org/doc12/Amber16.pdf>.

Vježba 18.1 Analitička metoda po izboru u programu *Gromacs*

Zadatak 7.13 Samostalno izabrati metodu kvantitativne analize trajektorije MD simulacije, koja nije bila obrađena u sklopu ovih uputa, te ju provesti na trajektoriji generiranoj tijekom MD simulacije katalitičke domene esteraze EstA iz bakterije *Pseudomonas aeruginosa*.

Upute:

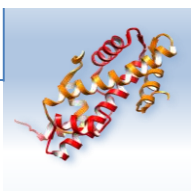
Posjetiti gore navedenu mrežnu stranicu te pročitati kratak opis navedenih potprograma programskog paketa *Gromacs*. Izabrati potprogram koji omogućava analizu trajektorije koja bi mogla biti primjerena za analizu MD simulacije katalitičke domene esteraze EstA iz bakterije *Pseudomonas aeruginosa*.
Upisivanjem:

naredba –h

doći do uputa za korištenje dane naredbe, odnosno izabranog potprograma. Zatim provesti izabranu analizu.

Pitanja:

1. Prikazati rezultate dobivene primjenom izabrane tehnike na trajektoriji MD simulacije katalitičke domene esteraze EstA iz bakterije *Pseudomonas aeruginosa*.



2. Interpretirati dobivene rezultate s obzirom na strukturna i dinamička svojstva katalitičke domene esteraze EstA iz bakterije *Pseudomonas aeruginosa*.

Vježba 18.2 Analitička metoda po izboru u programu Amber

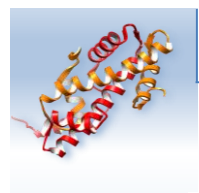
Zadatak 7.14 Samostalno izabrati metodu kvantitativne analize trajektorije koja nije bila obrađena u sklopu ove skripte te ju primijeniti na MD simulaciji ljudske NAG kinaze u kompleksu s NAG.

Upute:

Putem poveznice <https://ambermd.org/doc12/Amber16.pdf> otvoriti upute za korištenje programskog paketa Amber te u potpoglavlju V. *Analysis of simulations* (sekcija V.29. *cpptraj*) odabrati metodu kvantitativne analize po izboru. Odgovoriti na zadana pitanja.

Pitanja:

1. Prikazati rezultate dobivene primjenom izabrane tehnike na trajektoriji MD simulacije ljudske NAG kinaze u kompleksu s NAG.
2. Interpretirati dobivene rezultate s obzirom na strukturna i dinamička svojstva ljudske NAG kinaze u kompleksu s NAG.



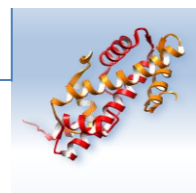
8. Molekulsko uklapanje (*docking*)

Metoda molekulskog uklapanja (engl. *docking*) omogućava računalno istraživanje nekovalentnog vezanja liganda za makromolekulu. Osobito je korisna u slučajevima kada poznajemo strukturu makromolekule, a želimo istražiti koje male molekule i u koja vezna mjesta u strukturi makromolekule bi se mogle vezati. Takve situacije osobito su česte u farmaceutskim istraživanjima gdje su računi molekulskog uklapanja izuzetno popularni. Metoda molekulskog uklapanja često se koristi i kao metoda koja prethodi simulacijama molekulske dinamike. Naime, najpovoljniji načini vezanja liganda na makromolekulu koji su dobiveni metodom molekulskog uklapanja često služe kao početne strukture za simulacije molekulske dinamike.

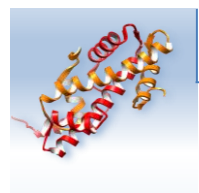
U svrhu upoznavanja metode molekulskog uklapanja za pripremu sustava i analizu rezultata koristit će se programski paket *AutoDock*, točnije grafičko sučelje potprograma *AutoDock Tools* (ADT) [9]. Za provođenje simulacije uklapanja koristit će se program *AutoDock Vina* [10]. Programi *AutoDock* i *AutoDock Vina* svoje izračune temelje na semi-empirijskom polju sila parametriziranom na temelju velikog broja poznatih protein-inhibitor kompleksa čija je konstanta inhibicije (K_i) poznata. Korištenjem jednažbe pridruženog polje sila procjenjuje se afinitet vezanja liganda na makromolekulu u dva koraka. U prvom se koraku određuje se energija prijelaza liganda i makromolekule iz slobodnog u vezano stanje (kompleks), dok se u drugom koraku računaju potencijalne energije različitih konformacija nastalog kompleksa. Doprinosi energiji vezanja u sklopu *AutoDock* polja sila su: disperzne sile, vodikove veze, elektrostatske interakcije i energija desolvatacije.

Priprema sustava za račun molekulskog uklapanja započinje učitavanjem .pdb datoteka liganda i makromolekule te njihovim prevođenjem u datoteke formata **.pdbqt** u programu *AutoDock*. Navedeni .pdbqt format sadrži: polarne vodikove atome, parcijalne atomske naboje, tipove atoma i informacije o rotabilnim dijelovima molekule. Za tipičan račun molekulskog uklapanja nužno je pripremiti dvije zasebne .pdbqt datoteke, pri čemu jedna sadrži informacije o ligandu, a druga o makromolekuli.

Postoje različite izvedbe računa molekulskog uklapanja. U tipičnoj simulaciji uklapanja sve se aminokiseline makromolekule fiksiraju, međutim *AutoDock* i



AutoDock Vina pružaju mogućnost definiranja fleksibilnih dijelova receptora. Definiranje fleksibilnih dijelova izrazito je korisno u slučajevima kad bočni ogranci pojedinih aminokiselina sprječavaju pravilno uklapanje liganda. Tada, prilagođavanje veznog mjesta na prisutnost liganda omogućava optimalno vezanje. No, valja imati na umu da su takve simulacije zahtjevnije i mogu dovesti do nerealnih konformacija receptora.



Vježba 19: Molekulska uklapanje

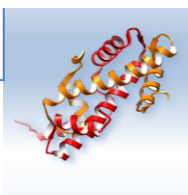
Cilj ove vježbe je smjestiti N-acetil-glukoziamin (NAG) iz vježbe 5 u aktivno mjesto heksokinaze iz organizma *Sulfolobus tokodaii* (PDB kod: 2E2O). Ljudska NAG kinaza korištena u Amber vježbama i heksokinaza iz *Sulfolobus tokodaii* sadrže visoki stupanj homologije te je pokazano da ova heksokinaza uz glukozu može vezati i NAG. Nakon molekuskog uklapanja potrebno je usporediti komplekse koje NAG ostvaruje s NAG kinazom i heksokinazom njihovim preklapanjem u programu *VMD*. Komentirati nekovalentne interakcije koje se ostvaruju između NAG i aktivnog mjesta tih dvaju enzima.

Zadatak 8.1 Metodom molekuskog uklapanja smjestiti N-acetil-glukoziamin iz vježbe 5 u aktivno mjesto heksokinaze iz organizma *Sulfolobus tokodaii*.

Upute:

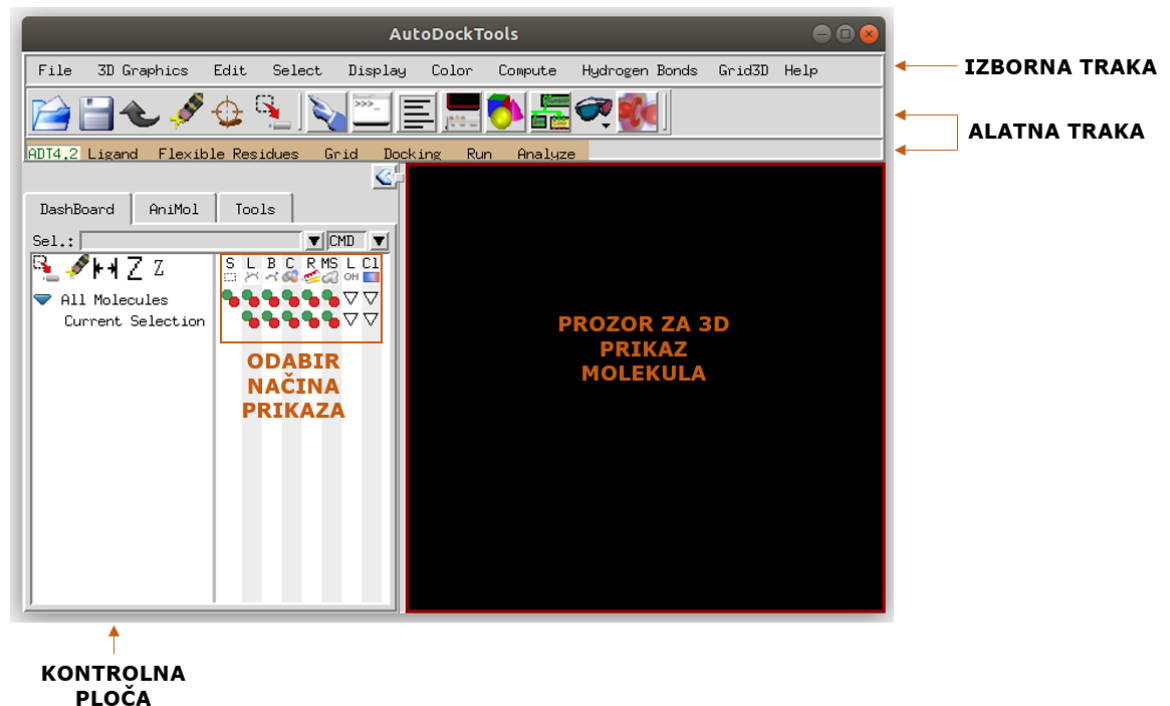
Za početak je potrebno stvoriti novi direktorij nazvan Docking na radnoj površini lokalnog računala te u njega kopirati .pdb datoteku koja sadrži koordinate N-acetil-glukoziamina iz vježbe 5. Iz baze PDB podataka pohraniti datoteku 2E2O.pdb, otvoriti je u tekstualnom editoru i ukloniti sve linije osim onih koje sadrže koordinate atoma koji pripadaju aminokiselinskom lancu A heksokinaze iz *Sulfolobus tokodaii*. Važno je napomenuti da program *AutoDock* može adekvatno obrađivati samo cjelovite strukture stoga je prije učitavanja makromolekule nužno provjeriti nedostaju li pojedini dijelovi strukture te ju kompletirati ako je potrebno. Za protoniranje* će se koristiti mrežno dostupni program *MolProbity* [7], koji je dostupan preko poveznice: <http://molprobity.biochem.duke.edu/>. *MolProbity* je program za protoniranje makromolekule kompatibilan s programom *AutoDock*. Putem navedene poveznice pristupiti programu *MolProbity* te opcijom **Choose File** učitati prethodno pripremljenu .pdb datoteku. Nakon učitavanja datoteke otvara se novi prozor s ponuđenim alatima u kojem je potrebno odabrati opciju **Add Hydrogens**.

* Makromolekula se može protonirati i nakon učitavanja u potprogramu *AutoDock*. Za potrebe ove vježbe odabran je način protoniranja putem mrežno dostupnog programa pošto se smatra preciznijim i nudi mogućnost promjene orijentacije aminokiselina (a odabran je i iz edukativnih razloga kako bi korisnik priručnika usvojio korištenje još jednog programa koji omogućava protoniranje).



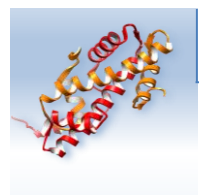
Program *MolProbity* nudi opciju protoniranja uz promjenu orijentacije nekih aminokiselina (**Asn/Glu/His flip**) koje je potrebno označiti i odabirom **Start adding H** generirati protoniranu strukturu. Dobivenu datoteku imenovati 2e2oH.pdb i premjestiti u direktorij Docking.

Upisivanjem naredbe **adt** u terminalu direktorija Docking otvara se grafičko sučelje programa *AutoDock Tools* prikazano na slici 10.



Slika 10. Izgled gafičkog sučelja potprograma *AutoDock Tools*.

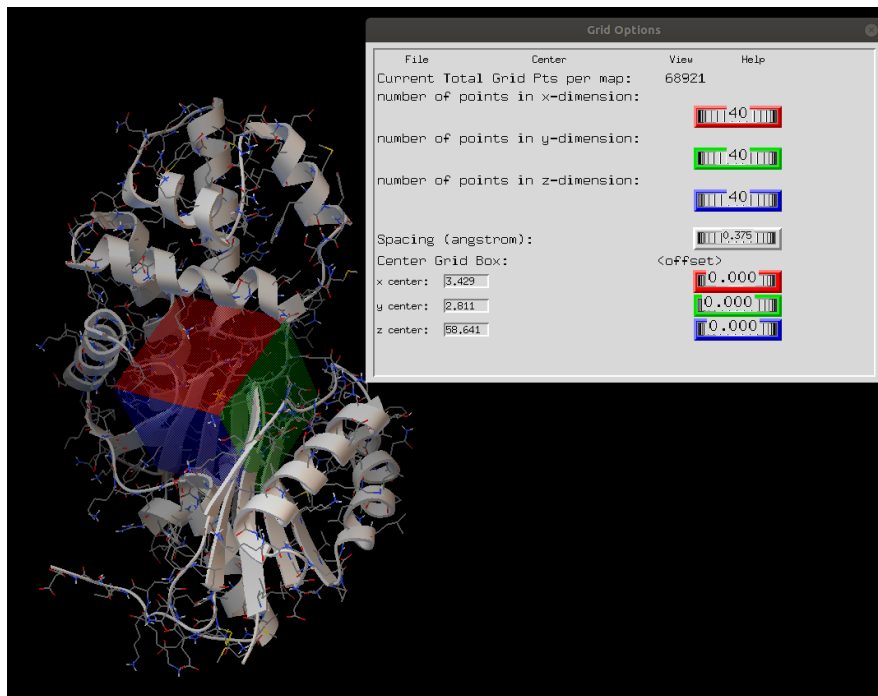
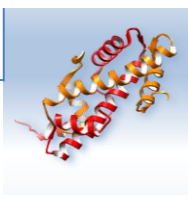
Protonirana struktura heksokinaze učitana se desnim klikom na **All molecules** na kontrolnoj ploči, zatim **Read molecule**. U prozoru za odabir datoteka potrebno je kliknuti na **PBQT** i kod odabira formata prikazanih datoteka odabrati **all files** kako bi se pojavio cijeli sadržaj radnog direktorija. Odabrati datoteku 2E2O.pdb i kliknuti **Open**. Učitana struktura je u prozoru za 3D prikaz prikazana linijama, a označavanjem kružića ispod slova R na kontrolnoj ploči pojavljuje se *Cartoon* način prikaza sekundarnih struktura. Prikazana molekula može se **rotirati lijevim pritiskom miša**, **translatirati desnim pritiskom miša** te **povećati pomicanjem srednjeg kotačića** na mišu. Odabir cijele makromolekule postiže se označavanjem kvadratića ispod slova S na kontrolnoj ploči. Za pohranu strukture u .pdbqt formatu potrebno je kliknuti na opciju **Grid**



u alatnoj traci, zatim **Macromolecule** → **Choose** → 2e2oH → **Select Molecule** → upisati novo ime 2e2oH.pdbqt → **Save**.

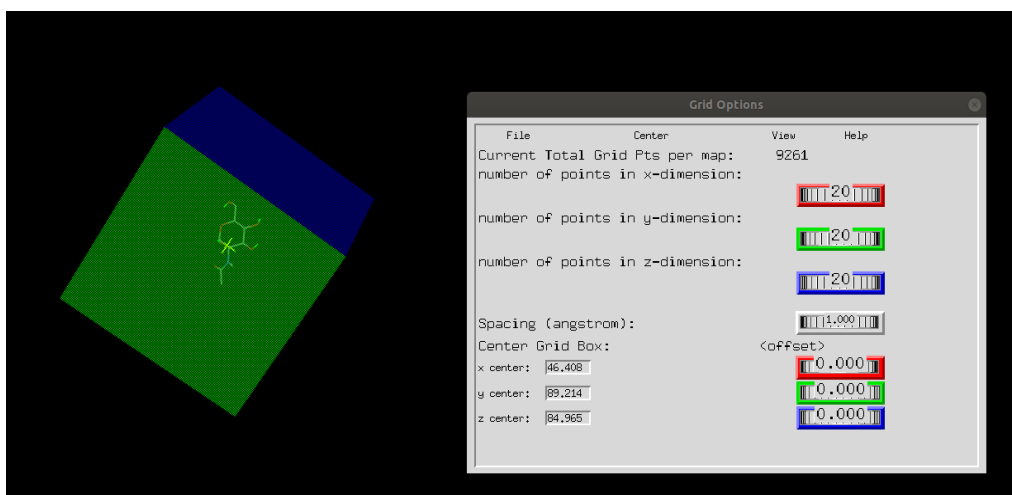
Struktura liganda učitava se odabirom opcije **Ligand** u alatnoj traci, zatim **Input** → **Open** → ponovno kod odabira formata prikazanih datoteka kliknuti **all files** → NAG.pdb → **Open**. Struktura liganda koja se učitava ne smije sadržavati vodikove atome, a ako ih i sadrži, oni se neće moći učitati. Prikaz veza liganda oko kojih je tijekom simulacije molekulskog uklapanja dozvoljena rotacija postiže se odabirom: **Ligand** → **Torsion Tree** → **Choose Torsions**. U novom prozoru prikazuje se broj rotabilnih veza. Veze oko kojih je dozvoljena rotacija obojane su zeleno, a veze s nedozvoljenom rotacijom crveno. Ljubičasto su prikazane veze koje su trenutno naznačene kao nerotabilne, ali se mogu promijeniti u rotabilne. Za potrebne ove vježbe neće se mijenjati broj rotabilnih veza, stoga je prozor potrebno zatvoriti klikom na **Done**. Koordinate liganda se u obliku .pdbqt datoteke pohranjuju klikom na: **Ligand** → **Output** → **Save as PDBQT** → upisati novo ime NAG.pdbqt → **Save**. Nakon pohranjivanja koordinata potrebno je još jednom definirati ligand: **Grid** → **Set Map Types** → **Choose Ligand** → NAG → **Select Ligand**.

Idući korak pripreme sustava podrazumijeva definiranje dijela makromolekule na kojem se želi istražiti mjesta za vezanje liganda. To se postiže definiranjem 3D mreže (**Grid Box**). Važno je napomenuti da uspješnost smještanja liganda uvelike ovisi o pravilnom definiranju 3D mreže. Po veličini i obliku mora omogućavati rotaciju liganda u svim smjerovima, no ne treba biti prevelika kako bi se račun fokusirao na specifično vezanje liganda u određeno vezno mjesto ili određeni dio makromolekule. Stoga poznavanje mjesta vezanja liganda uvelike ubrzava proces molekulskog uklapanja. Prikaz 3D mreže otvara se odabirom **Grid** → **Grid Box**.

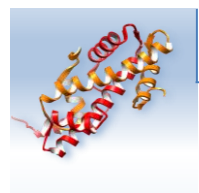


Slika 11. Izgled prozora za 3D prikaz molekula nakon otvaranja opcije **Grid box**. U gornjem desnom kutu prikazan je prozor **Grid options** u kojem se mijenjaju i definiraju sve opcije 3D mreže.

Ukoliko je ligand u početnoj strukturi smješten u vezno mjesto, preporuča se 3D mrežu centrirati na ligand. U primjeru koji se izvodi u sklopu ove vježbe, 3D mreža se centrira na NAG odabirom liganda u opciji **Center** prozora **Grid Options**, čime se dobiva prikaz kao na slici 12.



Slika 12. Izgled prozora za 3D prikaz molekula nakon definiranja liganda kao središta 3D mreže. U gornjem desnom kutu prikazan je prozor **Grid options** s promijenjenim početnim vrijednostima kako bi se postigla veličina 3D mreže koja omogućuje rotaciju liganda u svim smjerovima.

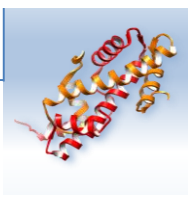


Prve tri trake za rotiranje u prozoru **Grid Options** omogućavaju promjenu dimenzija 3D mreže u x, y i z smjeru. Opcija **Spacing (Angstroms)** omogućava određivanje razmaka između točaka 3D mreže. Prije namještanja dimenzija 3D mreže potrebno je kod opcije **Spacing (Angstroms)** namjestiti vrijednost 1 kao što je prikazano na slici 12. Dimenzije u x, y i z smjeru se mogu mijenjati na dva načina: pomicanjem traka za rotiranje lijevo ili desno, te desnim klikom na brojčanu vrijednost i unošenjem željene vrijednosti. Nakon namještanja dimenzija 3D mreže potrebno ju je na pravilan način smjestiti u aktivno mjesto makromolekule. To je najlakše postići tako da se u opciji **Center** prozora **Grid options** odabere makromolekula, a zatim se rotiranjem zadnje tri trake za rotaciju 3D mreža smjesti u aktivno mjesto. Informacije o aktivnom mjestu enzima potražiti u radu u kojem je objavljena struktura heksokinaze iz organizma *Sulfolobus tokodaii*. Nakon pozicioniranja 3D mreže potrebno je zapisati njene dimenzije te koordinate središta zapisane ispod *Center Grid Box*. Zapisani parametri potrebni su za izradu ulazne datoteke kojom će se pokrenuti račun molekuskog uklapanja programom *AutoDock Vina*.

Program *AutoDock Vina* nema grafičko sučelje, stoga se svi potrebni parametri za provođenje simulacije uklapanja generiraju programom *AutoDock Tools*, nakon čega se definiraju u ulaznoj skripti **vina_input.txt**. Skripta mora sadržavati imena ulaznih .pdbqt datoteka, dimenzije 3D mreže i koordinate njenog središta, broj načina vezanja koji će se spremiti, itd. Već pripremljena datoteka vina_input.txt nalazi se na lokalnom računaru, potrebno ju je otvoriti u tekstualnom editoru te prilagoditi pojedine parametre i imena datoteka. Nakon modifikacije ulazne datoteke, račun molekuskog uklapanja se pokreće iz terminala u radnom direktoriju naredbom:

vina -config vina_input.txt

Po završetku računa u prozoru će se ispisati energija vezanja pojedinih načina vezanja. Dobiveni načini vezanja analiziraju se u *AutoDock Tools* grafičkom sučelju odabirom: **Analyze** → **Dockings** → **Open Autodock vina result ...** → odabrati izlaznu .pdbqt datoteku s rezultatima → **Open** → **Single molecules with multiple conformations** → **OK**. U prozoru za 3D prikaz pojavljuje se NAG ligand



smješten u aktivno mjesto makromolekule. Pritiskom tipke → ili ← lista se kroz dobivene načine vezanja liganda.

Energijski najpovoljniji način vezanja odabrati (označavanjem kvadratića pored imena datoteke s rezultatima u kontrolnoj ploči) te ga pohraniti u radni direktorij: **File** → **Save** → **Write .pdb**. Ukoliko se želi ponoviti simulacija uklapanja s ciljem poboljšavanja rezultata potrebno je navedeni način vezanja pohraniti kao .pdbqt datoteku te ju koristiti kao ulaznu datoteku u sljedećem računu molekuskog uklapanja.

Datoteke .pdb najpovoljnijeg načina vezanja liganda i makromolekule sjediniti u jedinstvenu .pdb datoteku upisivanjem naredbe:

cat 2e2o_H.pdb mode_1.pdb > hexokinase_NAG_complex.pdb

Pitanja:

1. Koliko je iznosila energija najpovoljnijeg načina vezanja NAG u aktivno mjesto heksokinaze iz organizma *Sulfolobus tokodaii*?
2. Vizualizirati dobiveni kompleks u programu *VMD* i napraviti sliku na kojoj je heksokinaza iz *Sulfolobus tokodaii* prikazana načinom prikaza *New Cartoon*, a NAG načinom prikaza *Licorice*.
3. Sadrži li pohranjena .pdb datoteka najboljeg načina vezanja NAG liganda atome vodika? Ako ne, predložiti način kojim bi se dodali.
4. Protonirati NAG u pohranjenoj .pdb datoteci metodom po izboru. Datoteku .pdb protoniranog liganda sjediniti s 2e2o_H.pdb datotekom. Otvoriti kompleks u programu *VMD* te prikazati vodikove veze koje se ostvaruju između liganda i aktivnog mjesta enzima. Komentirati točnost ovakvog načina dodavanja atoma vodika.

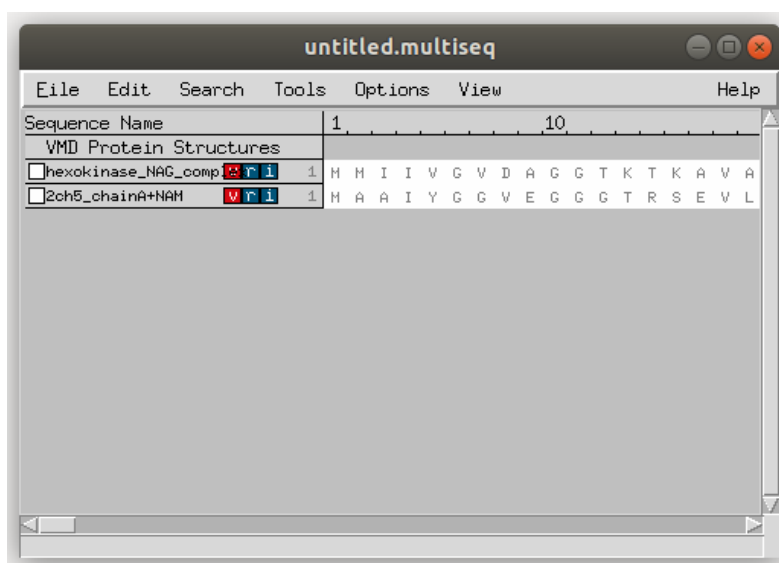
Zadatak 8.2 Strukturu kompleksa heksokinaze iz organizma *Sulfolobus tokodaii* i NAG usporediti sa strukturom kompleksa ljudske NAG kinaze i NAG, preklapanjem u programu *VMD*. Usporediti nekovalentne interakcije koje se ostvaruju između NAG i aktivnog mjesta dvaju enzima.

Upute:

Za početak je potrebno pripremiti .pdb datoteku koja sadrži koordinate ljudske NAG kinaze u kompleksu s NAG. U tu svrhu iskoristiti .pdb datoteku koja sadrži kompletiranu strukturu ljudske NAG kinaze dobivenu u vježbi 7. Navedenu datoteku otvoriti u tekstualnom editoru te izbrisati sve osim atoma koji pripadaju lancu A i njemu pripadnom NAG ligandu.

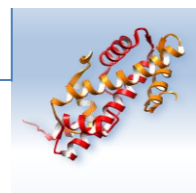
U programu *VMD* prvo učitati strukturu heksokinaze iz organizma *Sulfolobus tokodaii* u kompleksu s NAG. Potom odabirom **File** → **New molecule...** → **Browse** → odabrati prethodno pripremljenu datoteku s koordinatama kompleksa NAG i ljudske NAG kinaze → **Load**.

Preklapanje dviju struktura postiže se odabirom opcije **Extensions** → **Analysis** → **Multiseq** u glavnom prozoru. Prije otvaranja **Multiseq** prozora prikazanog na slici 13, potrebno je odabrati direktorij u koji će se pohraniti privremene datoteke. Odabirom opcije **Tools** → **Stemp Structural Alignment** u **Multiseq** prozoru generirat će se preklapljene strukture dvaju kompleksa.



Slika 13. Izgled **Multiseq** prozora za uspoređivanje sekvenci i 3D struktura homolognih proteina u programu *VMD*.

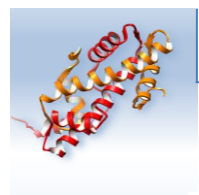
Uz preklapanje struktura, **Multiseq** opcija nudi i mogućnost analiziranja sravnjenih aminokiselinskih sljedova dvaju (ili više) homolognih proteina. Sravnjene aminokiselinske sekvence ljudske NAG kinaze i heksokinaze iz *Sulfolobus tokodaii* mogu se prikazati odabirom: **Sequence Alignment** → **ClustalW Multiple Alignment**. Isticanje očuvanih aminokiselina postiže se



odabirom: **View** → **Sequence Similarity** → **Blosum 30**. Klikom na pojedinu aminokiselinu u **Multiseq** prozoru prikazuje se njezin redni broj u strukturi proteina.

Pitanja:

1. U programu *VMD* napraviti sliku preklopljenih struktura ljudske NAG kinaze i heksokinaze iz *Sulfolobus tokodaii* u kompleksima s NAG. Različitim načinima prikaza i bojanja istaknuti sličnosti u načinu vezanja NAG liganda.
2. Srvniti aminokiselinske sekvence ljudske NAG kinaze i heksokinaze iz *Sulfolobus tokodaii* te komentirati očuvanost aminokiselina koje čine aktivno mjesto enzima.
3. Na preklopljenim strukturama ljudske NAG kinaze i heksokinaze iz *Sulfolobus tokodaii* istaknuti sve očuvane aminokiseline prikazom *Licorice*. Komentirati njihovu moguću ulogu i potencijalni razlog očuvanosti.



9. Dodatak

9.1 Izrada skripte za parametrizaciju sustava potprogramom

tleap

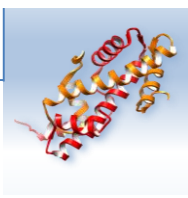
Parametrizacija sustava opisana u vježbi 8.2 se može znatno ubrzati ukoliko se prije pokretanja potprograma *tleap* pripremi skripta koja će sadržavati: imena svih potrebnih polja sila, naredbe i imena odgovarajućih datoteka potrebnih za parametrizaciju. Skripta se jednostavno može izraditi u jednom od tekstualnih editora, primjerice *gedit*. Važno je imati na umu da se skripta mora nalaziti u direktoriju koji sadrži sve navedene datoteke ili je potrebno definirati put (staze) do tih datoteka unutar skripte. U slučaju vježbe 8.2, skripta nazvana **tleap_parametrizacija** bi sadržavala slijedeće naredbe i imena datoteka:

```
source leaprc.protein.ff14SB
loadamberparams gaff.dat
source leaprc.water.tip3p
enz = loadpdb NAG-kinase-protonated.pdb
check enz
NAG= loadmol2 NAG_H_clean.mol2
loadamberparms NAG_H_clean.frcmod
check NAG
cplx= combine {enz NAG}
check cplx
addions cplx Na+ 0
addions cplx Cl- 0
solvatebox cplx TIP3PBOX 10
saveamberparm cplx cplx-NAG-kinase.prmtop cplx-NAG-kinase.inpcrd
savepdb cplx cplx-NAG-kinase.pdb
```

Skripta se pokreće naredbom:

tleap -f tleap_parametrizacija

Izrazito je važno da se u slučaju ovakvog načina parametrizacije provjere sve obavijesti o mogućim greškama ili nepravilnostima ispisanima u terminalu po



završetku. Sve obavijesti se također zapisuju i u **leap.log** datoteci koja nastaje i nadopunjava se prilikom svakog ponovnog pokretanja potprograma *tleap*.

9.2 Pokretanje paralelnih simulacija na više procesora.

U svrhu pokretanja simulacije na više procesora koristi se Amber potprogram *mpirun*. Za pokretanje takvih računa potrebno je dolje navedenu skriptu iskopirati u novu .txt datoteku te ju prilagoditi vlastitim potrebama. Budući da će se skripta pokrenuti koristeći bash terminal, u prvim linijama označenim narančastom bojom navedeni su putevi (staze) do svih potrebnih potprograma koji će omogućiti paralelno izvođenje simulacija. Putevi do simulacijskih programa i potrebnih datoteka nisu isti na svim računalima stoga je prvo potrebno na svom računalu u tekstualnom editoru otvoriti ~/.bashrc datoteku i iskopirati puteve do navedenih potprograma u skriptu:

```
#!/bin/bash
export AMBERHOME=/home/andrea/amber16
export CUDA_HOME=/usr/local/cuda-8.0
export PATH="/home/andrea/amber16/bin:$PATH"
```

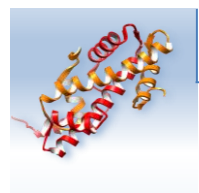
```
#!/usr/bin/orted
/usr/bin/mpirun -np 8 taskset -c 0,1,2,3,4,5,6,7 $AMBERHOME/bin/sander.MPI -p
struktura.prmtop -c struktura.rst -ref struktura.rst -i parametri_za_simulaciju.in -o
izlazna_datoteka.out -r izlazna_datoteka.rst -x izlazna_datoteka.mdcrd
```

* **-np** oznaka se stavlja ispred broja procesora na kojima se žele izvoditi računi. U ovom je slučaju simulacija puštena na 8 procesora. Poželjno je da broj procesora koji se koristi za račune bude potencija broja 2, dakle 2, 4, 8, 16 ...

-c oznaka se stavlja ispred rednih brojeva procesora kojima se želi pridružiti simulacija. Brojanje kreće od 0 te će se stoga u ovom slučaju simulacija izvoditi na *procesorima* 1, 2, 3, 4, 5, 6, 7, 8.

Navedena skripta pokreće se iz terminala naredbom:

source ime_skripte.txt



9.3 Pokretanje simulacija na grafičkoj kartici

Simulacije na grafičkoj kartici izvode se potprogramom *pmemd*. Skripta za pokretanje simulacije nalazi se niže u tekstu, potrebno ju je iskopirati u novu .txt datoteku i modificirati sukladno vlastitim potrebama. Opcijom **export CUDA_VISIBLE_DEVICES=0** simulacija se pokreće na grafičkoj kartici označenoj brojem 0. U slučaju računala s jednom grafičkom karticom ona je uvijek označena brojem 0.

Skripta:

```
export CUDA_VISIBLE_DEVICES=0
```

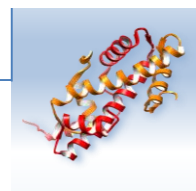
```
pmemd.cuda -p struktura.top -c struktura.rst -ref struktura.rst -i  
parametri_za_simulaciju.in -o izlazna_datoteka.out -r izlazna_datoteka.rst -x  
izlazna_datoteka.crd
```

Navedena skripta pokreće se iz terminala naredbom:

source ime_skripte.txt

U slučaju izvođenja računa na grafičkoj kartici dobro je pratiti njezin status, posebice temperaturu. Učestalo zagrijavanje grafičke kartice na temperature više od 85°C moglo bi dovesti do kvara. Status grafičkih kartica može se jednostavno pratiti u terminalu programom *nvidia-smi* koji se pokreće upisivanjem naredbe:

nvidia-smi



9.4 Izrada cpptraj.in skripte

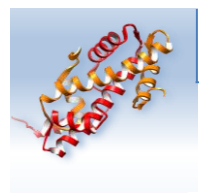
Kao što je opisano u vježbi 14.2., u potprogramu *cpptraj* postoji interaktivni način rada, te zadavanje naredbi putem skripte. Svaka *cpptraj* skripta mora sadržavati topologiju sustava i trajektoriju koja će se analizirati.

Primjer jedne *cpptraj* skripte za modificiranje trajektorije i izračun rmsd vrijednosti:

```
parm sustav.prmtop
trajin sustav_NVT.crd
trajin sustav_NPT.crd
trajin sustav_MD.crd
center :1
autoimage origin
trajout NVT_NPT_MD.binpos
rmsd tofirst :1 mass out RMSD_ADPr.dat
```

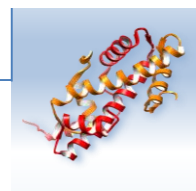
Navedena skripta se u terminalu pokreće naredbom:

cpptraj -i cpptraj_skripta.in



10. Popis literature

- [1] W. Humphrey, A. Dalke, K. Schulten, "VMD: Visual molecular dynamics" *J. Mol. Graph.* **14(1)** (1996) 33–38.
- [2] M. O. Kim, S. E. Nichols, Y. Wang, J. A. McCammon, "Effects of histidine protonation and rotameric states on virtual screening of M. tuberculosis RmlC" *J. Comput. Aided. Mol. Des.* **27(3)** (2013) 235–246.
- [3] G. Vriend, "WHAT IF: A molecular modeling and drug design program" *J. Mol. Graph.* **8(1)** (1990) 52–56.
- [4] R. Anandakrishnan, B. Aguilar, A. V. Onufriev, "H++ 3.0: Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations" *Nucleic Acids Res.* **40** (2012) 537–541.
- [5] J. Myers, G. Grothaus, S. Narayanan, A. Onufriev, "A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules" *Proteins Struct. Funct. Genet.* **63(4)**, (2006) 928–938.
- [6] J. C. Gordon, J. B. Myers, T. Folta, V. Shoja, L. S. Heath, A. Onufriev, "H++: A server for estimating pKas and adding missing hydrogens to macromolecules" *Nucleic Acids Res.* **33**, (2005) 368–371.
- [7] C. J. Williams, B. J. Hintze, J. J. Headd, N. W. Moriarty, V. B. Chen, S. Jain, M. G. Prisant, S. M. Lewis, L. L. Videau, D. A. Keedy, L. N. Deis, W. B. III Arendall, V. Verma, J. S. Snoeyink, P. D. Adams, S. C. Lovell, J. S. Richardson, "MolProbity: More and better reference data for improved all-atom structure validation" *Protein Sci.* **27(1)** (2018) 293–315.
- [8] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, "SWISS-MODEL: Homology modelling of protein structures and complexes" *Nucleic Acids Res.* **46** (2018) W296–W303.
- [9] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, "AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility" *J. Comput. Chem.* **30(16)** (2009) 2785–2791.
- [10] O. Trott, A. J. Olson, "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading" *J. Comput. Chem.* **31(2)** (2010) 455–461.



Temeljna literatura:

- R. Leach, *Molecular Modelling: Principles and Applications*, 2nd edition, Prentice Hall, 2001.
- C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models*, 2nd edition, J. Wiley, Chichester, 2004.
- R. Huey, G. M. Morris, A. J. Olson, D. S. Goodsell, „A semiempirical free energy force field with charge-based desolvation“ *J. Comput. Chem.* **28** (2007) 1145–1152.

Mrežne stranice (pristupljeno 12. 5. 2020.):

- Protein Data Bank (PDB): <https://www.rcsb.org/>
- SWISS-MODEL: <https://swissmodel.expasy.org/interactive>
- What If: <https://swift.cmbi.umcn.nl/servers/html/index.html>
- H++: <http://biophysics.cs.vt.edu/>
- MolProbity: <http://molprobity.biochem.duke.edu/>
- Model Box Periodic Boundary Conditions:
<http://isaacs.sourceforge.net/phys/psc.html>