

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 1077

**ALGORITAM ZA PREDVIĐANJE
STRUKTURE PROTEINSKIH KOMPLEKSA
KORIŠTENJEM RAZVOJA POVRŠINE U
REDOVE FUNKCIJA**

Ivan Dokmanić

Zagreb, srpanj 2007.

Hvala mami i tati za sve, prijateljima za dobro društvo, Mili Šikiću za sjajnu kreativnu komunikaciju, i naposljetku ... hvala Bogu!

| | | |
|-------|---------------------------------------|----|
| 1 | Uvod | 1 |
| 2 | Proteini | 2 |
| 2.1 | Aminokiseline | 2 |
| 2.2 | Peptidi i proteini | 4 |
| 2.3 | Veze među makromolekulama | 4 |
| 2.3.1 | Gibbsova slobodna energija | 4 |
| 2.3.2 | Vodikova veza, topivost u vodi | 5 |
| 2.3.3 | Van der Waalsove veze..... | 6 |
| 2.3.4 | Hidrofobne interakcije..... | 6 |
| 2.3.5 | Ionske (elektrostatske) veze..... | 7 |
| 2.3.6 | Slabe interakcije | 8 |
| 2.4 | Struktura proteina | 8 |
| 2.4.1 | Primarna struktura | 8 |
| 2.4.2 | Sekundarna struktura | 8 |
| 2.4.3 | Tercijarna struktura | 9 |
| 2.4.4 | Kvartarna struktura | 9 |
| 3 | Prianjanje proteina | 10 |
| 3.1 | Formulacija | 10 |
| 3.2 | Problemi kod prianjanja | 10 |
| 3.3 | Metode prianjanja | 11 |
| 4 | Opis rada | 14 |
| 4.1 | Motivacija | 14 |
| 4.2 | Razrada zadatka | 16 |
| 5 | Matematička razrada. Kugline funkcije | 17 |
| 5.1 | Uvodna razmatranja | 17 |
| 5.2 | Kugline funkcije | 20 |
| 5.2.1 | Kompleksne kugline funkcije | 20 |
| 5.2.2 | Wignerove rotacijske matrice | 21 |
| 5.2.3 | Realne kugline funkcije..... | 22 |
| 5.2.4 | Realne rotacijske matrice | 23 |
| 6 | Točkasta površina proteina | 25 |
| 6.1 | Razne površine molekula | 25 |
| 6.1.1 | Van der Waalsova površina..... | 25 |

| | | |
|-----------|---|-----------|
| 6.1.2 | Otapalu dostupna površina (<i>solvent-accessible surface</i>) | 25 |
| 6.1.3 | Molekularna površina | 25 |
| 6.1.4 | Namjena reprezentacije površine | 26 |
| 7 | Radijalne funkcije. Ljuska proteina | 28 |
| 7.1 | Problem s plohami | 28 |
| 7.2 | Pogodne radijalne funkcije | 28 |
| 7.2.1 | Schrödingerova jednačba za vodikov atom..... | 28 |
| 7.2.2 | Orbitale Gaussovog tipa | 29 |
| 7.3 | 3-D parametrizacije | 32 |
| 7.4 | Funkcija ljuske proteina | 33 |
| 8 | Shema prianjanja | 36 |
| 8.1 | Računanje ocjene kvalitete preklapanja | 36 |
| 9 | Implementacija | 38 |
| 9.1 | Generiranje ljuski | 38 |
| 9.1.1 | SIMS – <i>smooth invariant molecular surface</i> | 38 |
| 9.1.2 | Od površine do ljuske | 38 |
| 9.1.3 | Testiranje ljuski..... | 39 |
| 9.2 | Implementacija rotacijskih matrica | 42 |
| 9.3 | Izračunavanje translacijskih matrica | 43 |
| 9.4 | Koreliranje hidrofobnosti površina | 44 |
| 10 | Testiranje i rezultati | 47 |
| 10.1 | Kompleks 1ppe | 48 |
| 10.2 | Kompleks 1eaw | 49 |
| 10.3 | Kompleks 1ay7 | 50 |
| 11 | Razmatranje rezultata | 51 |
| 11.1 | Uočeni problemi | 51 |
| 11.2 | Daljnji razvoj | 51 |
| 12 | Zaključak | 52 |
| 13 | Literatura | 53 |

1 Uvod

Još od sredine prošloga stoljeća, u središtu biologije i biokemije nalazili su se geni, odnosno struktura i funkcija nukleinskih kiselina. Razvoj tehnologije za sekvenciranje DNA pokrenuo je mapiranje ljudskog genoma, kolosalan pothvat čije posljedice, dobre ili loše, već itekako osjećamo. Kako su ti projekti privođeni kraju, istraživački fokus se počeo sve više premještati na proučavanje strukture i funkcije sljedeće karike u nizu – proteina (Halperin, 2002). Nekoliko je razloga za izuzetan volumen znanstvenog rada u području proteomike danas. Dva sasvim sigurno najvažnija su golemo povećanje javnosti dostupnih računalnih resursa, odnosno napredak u kristalografskim tehnikama. Drugi je zaslužan za postojanje nekoliko javnih baza s tisućama i tisućama riješenih 3-D struktura proteina, iz najrazličitijih živih organizama, s najrazličitijim funkcijama. Prvi je pak zaslužan za činjenicu da svatko s određenim poznavanjem biologije, tko posjeduje računalo i zna ga koristiti, te baze može pretraživati i pokušati pomoću tamo dostupnih informacija pronaći kandidate za cjepivo za npr. HIV. Dva su problema posebno interesantna, i zapravo je velika većina rada u ovom području na neki način tek *podrška* tim ključnim problemima. Radi se o prijanjanju proteina (engl. *protein docking*) i smatanju proteina (engl. *protein folding*). Kod prijanjanja se pokušava dati odgovor na pitanje hoće li dva proteina (ili protein i neka druga molekula, npr. lijek) međusobno tvoriti kompleks (hoće li reagirati), te se pokušava predvidjeti 3-D struktura eventualnog rezultirajućeg kompleksa. Kod smatanja se pokušava polazeći od slijeda aminokiselina predvidjeti 3-D struktura (tercijarna struktura, više riječi o tome kasnije) proteina. Danas je jasno kako su ti problemi srodni, i kako svaki napredak u rješavanju jednog od njih vjerojatno znači i napredak u rješavanju drugoga. Karakteristično je za biokemijske reakcije da protein prvo mora vezati ligand, bilo trajno kao kod transporta ili inhibicije, bilo kratkoročno, privremeno, kad ima katalitičku funkciju – stoga bi potpuno rješenje problema prijanjanja proteina značilo vjerojatno više nego što bi bilo mudro predvidjeti. Ovaj rad se bavi jednim mogućim pristupom problemu prijanjanja proteina.

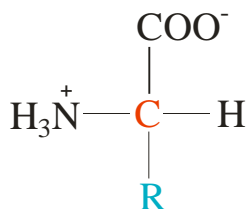
2 Proteini

Kako je razumijevanje građe i principa interakcije proteina ključno za razumijevanje problema prijanjanja proteina i rezultata opisanih u ovom radu, ovdje će biti dan kratki opis njihove strukture i funkcije.

2.1 Aminokiseline

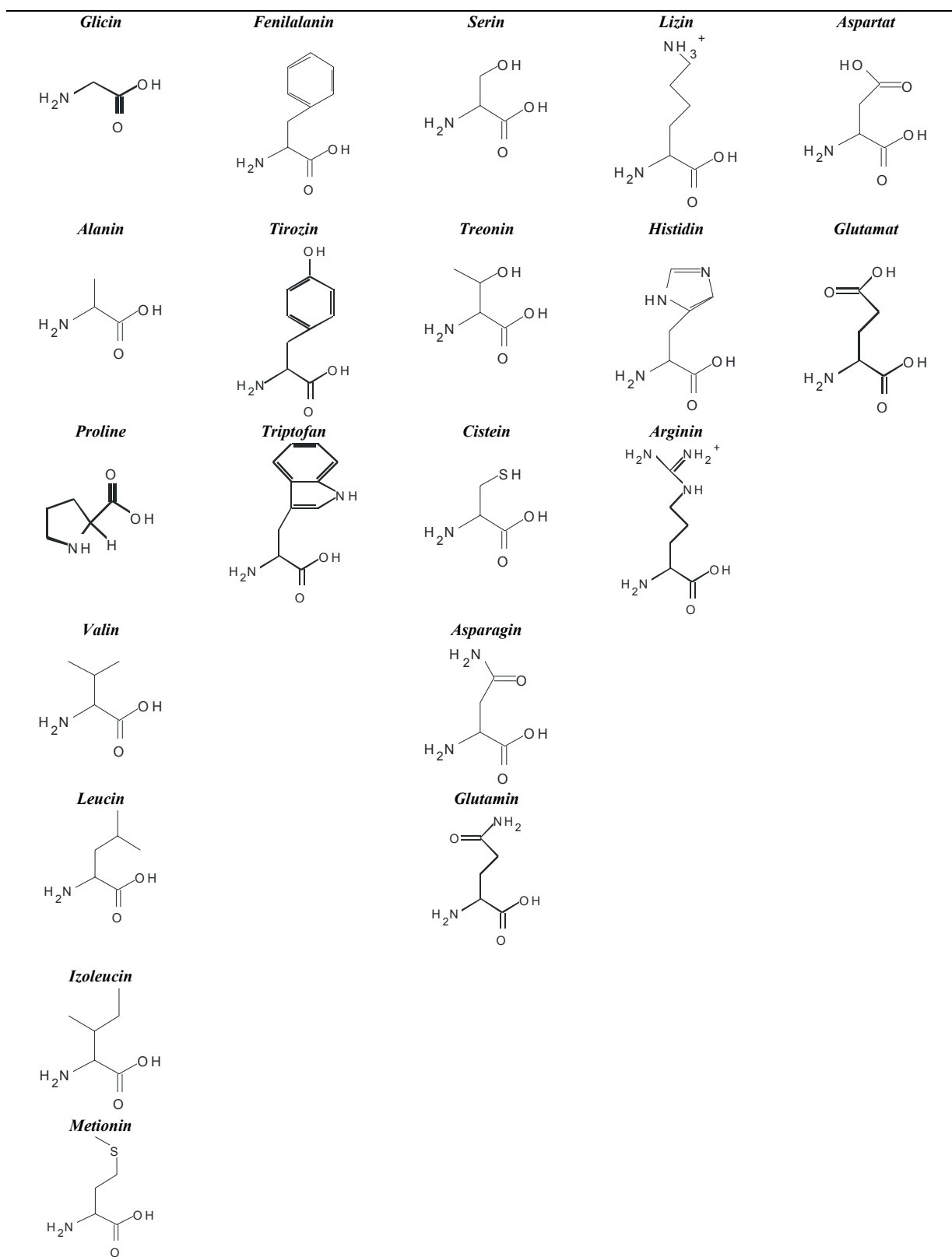
Aminokiseline su osnovne građevne jedinice proteina. Proteini svoju funkciju obavljaju zahvaljujući obliku i/ili određenim elektrokemijskim svojstvima; i jedno i drugo su izravna posljedica svojstava aminokiselina od kojih su sastavljeni.

Svi proteini svih živih bića su polimeri jednog te istog skupa aminokiselina (u daljnjem tekstu AK), odnosno njihovih reziduuma (termin koji se koristi jer se prilikom vezanja AK u proteine izlučuju elementi vode). Postoji 20 uobičajenih aminokiselina i nekoliko rjeđih.



Slika 1. Općenita struktura AK. Zajednička je svim AK osim jedne – prolin je prstenasta AK. R grupa je različita za svaku AK. Ionizacijsko stanje je prikazano kao pri pH 7.0.

Svih 20 uobičajenih AK su tzv. α -amino kiseline. Imaju karboksilnu skupinu i amino skupinu vezanu na isti C atom (α ugljik). Ono po čemu se AK razlikuju su njihove R grupe, koje su različitih struktura, veličina i električnog naboja. One određuju topivost određene AK u vodi. U Tablici 1 su navedene 20 uobičajenih AK, zajedno s njihovim strukturnim formulama. Pri biološkim pH se amino i karboksilne skupine ioniziraju, no ovdje su one prikazane bez ionizacije.



Slika 2. Strukture svih 20 uobičajenih aminokiselina, grupirane po stupcima prema svojstvima bočnog lanca. S lijeva na desno, nepolarne, aromatske, polarne (nenabijene), pozitivno nabijene, negativno nabijene.

2.2 Peptidi i proteini

Dvije aminokiseline se mogu vezati tzv. *peptidnom vezom* u *dipeptid*. Ta se veza formira eliminacijom elemenata vode iz α -karboksilne skupine jedne aminokiseline, odnosno α -amino skupine druge. Tri aminokiseline se mogu vezati pomoću dvije peptidne veze u tripeptid, četiri u tetrapeptid, itd. Kada se na ovaj način veže manji broj AK, strukturu zovemo *oligopeptid*, a kada ih se veže veći broj, nastao je *polipeptid*. Proteini pak, kao sljedeći korak, mogu imati tisuće AK reziduuma.

2.3 Veze među makromolekulama

Biološke molekule je važno promatrati u njihovom prirodnom biološkom kontekstu. Vjerojatno nema puno smisla promatrati kako se hemoglobin ponaša u benzinu, ali je zato vrlo relevantno kako se ponaša u mediju sličnom krvi. Sve veze među makromolekulama zato treba promatrati u kontekstu otapala u kojemu se one dešavaju. To otapalo je najčešće voda, i njezina uloga u vezama između makromolekula je ključna za funkcioniranje života.

2.3.1 Gibbsova slobodna energija

Fizikalni parametar koji govori koliko je neko stanje biokemijskog sustava termodinamički povoljno je Gibbsova slobodna energija sustava,

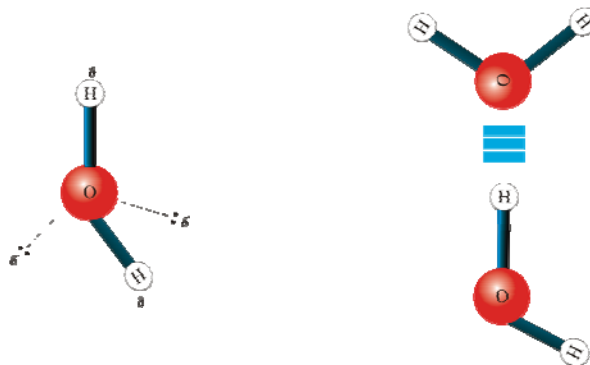
$$G = H - TS. \quad (2-1)$$

gdje je H entalpija sustava, T termodinamička temperatura, a S entropija sustava. Promjena Gibbsove energije tijekom neke reakcije pri konstantnoj temperaturi, $\Delta G = \Delta H - T\Delta S$, nam govori koliko je ta reakcija termodinamički povoljna, odnosno može li se odvijati spontano. Termodinamički sustav teži stabilnom stanju, odnosno stanju u kojem ima minimalnu Gibbsovu energiju. Članovi izraza za promjenu Gibbsove energije odražavaju kemijsku odnosno fizikalnu situaciju – ΔH , promjena entalpije, odražava vrstu i broj kemijskih veza i nekovalentnih interakcija koje se kidaju i stvaraju, dok promjena entropije ΔS odražava promjenu *neuređenosti* sustava. Tako se voda

vjerojatno nikada neće spontano zalediti budući da (iako se zaleđivanjem povećava broj vodikovih veza po jedinici volumena) je to proces koji prati veliko povećanje Gibbsove energije, zbog smanjenja entropije: struktura leda je pravilna kristalna struktura i u njoj su molekule vode uredno orijentirane za razliku od tekuće vode u kojoj su molekule relativno slobodno orijentirane (pretpostavimo da se proces odvija blizu ledišta vode i zanemarimo promjenu temperature).

2.3.2 Vodikova veza, topivost u vodi

Molekula vode se sastoji od dva vodikova atoma i jednog atoma kisika. Kisik je elektronegativniji od vodika, pa stoga jače privlači zajedničke elektronske parove. Zbog toga je molekula vode dipol, a to je i razlog zašto se molekule vode međusobno privlače. Takva veza se zove vodikova veza. Molekule vode nisu jedine koje tvore vodikove veze, one nastaju uvijek kad imamo elektronegativan atom (proton akceptor) i vodik kovalentno vezan za drugi elektronegativan atom. Vodikova veza je tipična veza između dijelova proteinskih molekula.



Slika 3. Shematski prikaz vodikove veze. Lijevo je prikazana molekula vode s pripadajućim polovima raspoređenim približno u vrhove pravilnog tetraedra. Desno je skicirano međudjelovanje dvije molekule vode.

Zbog svoje polarnosti, voda dobro otapa nabijene i polarne tvari. Za razliku od nje, nepolarna otapala poput kloroforma ili benzena te tvari loše otapaju. Tvari koje se dobro otapaju u vodi zovemo *hidrofilnima*, a one koje se u vodi loše otapaju zovemo *hidrofobnima*. Voda otapa soli poput NaCl

hidratiziranjem Na^+ i Cl^- iona tako što slabi elektrostatsku interakciju među njima zbog svoje velike relativne permitivnosti (78.5). Reakcije poput topljenja soli su i termodinamički povoljne jer, iako se entalpija malo poveća, povećanje entropije je jako veliko budući da ioni disocijacijom dobivaju puno veću slobodu gibanja, pa je ukupna promjena slobodne energije negativna.

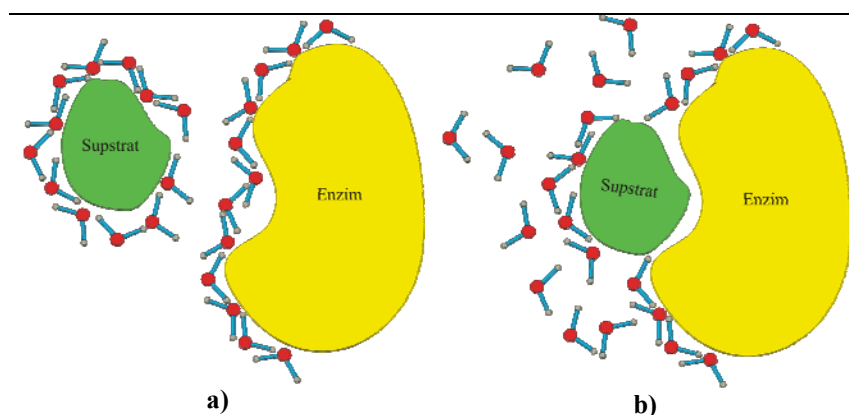
2.3.3 Van der Waalsove veze

Elektronski oblaci dva bliska atoma djeluju jedan na drugoga. Slučajne varijacije u položaju elektronskog oblaka oko jedne jezgre mogu stvoriti tranzijentni dipol koji će tada inducirati suprotni tranzijentni dipol kod drugog atoma. Dva dipola se slabašno privlače pa se i jezgre atoma približe. Takve slabe interakcije zovemo van der Waalsovima. Zbog približavanja atoma, počinju se odbijati njihovi elektronski oblaci. Tzv. van der Waalsov kontakt je ostvaren na udaljenosti na kojoj su te dvije pojave u ravnoteži. Van der Waalsov radijus nekog atoma je upravo procjena udaljenosti na koju će taj atom *dozvoliti* drugom da mu se približi.

2.3.4 Hidrofobne interakcije

Ako u vodu ulijemo nepolarnu tekućinu, formirati će se dvije odvojene faze. Voda ne može interagirati s nepolarnim molekulama, pa bi bilo energetski nepovoljno da se one odvoje – termodinamički gledano, promjena entalpije je veća, budući da uvođenje bilo kakvih čestica u vodu kida vodikove veze, a za razliku od polarnog slučaja, ovdje te veze neće biti nadomještene vezama između vode i otopljene tvari. Entropijska komponenta je također nepovoljna, budući da bi otapanje nepolarne tvari u vodi uzrokovalo smanjenje entropije sustava: molekule vode u blizini nepolarne tvari ne mogu tvoriti onoliko vodikovih veza koliko mogu bez te tvari u blizini. Sloboda orijentacije im je značajno ograničena i one tvore vrlo *urednu* ljusku oko nepolarnih čestica. Zanimljiva je situacija s česticama koje imaju i polarne i nepolarne dijelove – takve čestice u vodi formiraju strukture u kojima maksimiziraju interakciju hidrofilnih dijelova s vodom, i minimiziraju izloženost hidrofobnih dijelova vodi. Površine proteina, koji su vrlo velike molekule, sadrže mnogo polarnih ili nabijenih dijelova i mnogo nepolarnih dijelova. Ako se u blizini nađu dvije

molekule koje imaju izložene hidrofobne dijelove, energetski je povoljno da se ti dijelovi približe jedan drugome toliko da istisnu uređenu vodu koja je oko njih formirala ljusku (naravno, ukoliko oblik proteina to dozvoljava) (Lehringer, 2004).



Slika 4. Smanjenje slobodne energije oslobađanjem vezane, uređene vode oko supstrata stabilizira nekovalentno vezanje supstrata na enzim. Takva veza koja je posljedica hidrofobnosti može dodatno biti stabilizirana vodikovim mostovima i ionskim interakcijama supstrata s enzimom.

Takve interakcije zovemo hidrofobnim interakcijama, i one su jedan od osnovnih mehanizama vezanja među proteinima. Odgovorne su, između ostaloga, i za smatanje proteina.

2.3.5 Ionske (elektrostatske) veze

R grupe nekih aminokiselina sadrže skupine koje se pri biološki interesantnim pH ioniziraju. Tako se primjerice negativno nabijene ionizirane (sekundarne) karboksilne skupine aspartata i glutamata međusobno odbijaju pri pH 7.0, ali se zato privlače s pozitivnim sekundarnim amino skupinama lizina i arginina. Između tih skupina, odnosno aminokiselina, postoji coulombska elektrostatska sila koju je također potrebno uvrstiti među veze između makromolekula.

2.3.6 Slabe interakcije

Opisane veze spadaju u grupu slabih nekovalentnih interakcija. Osim njih, makromolekule se rjeđe vežu i kovalentnim vezama (disulfidni mostovi između atoma sumpora dva cistina). Primjera radi, potrebno je dovesti 410 kJ energije da bi se prekinuo jedan mol C–H kovalentnih veza, dok je za jedan mol van der Waalsovih veza dovoljno samo 4 kJ. Iako su sve navedene interakcije znatno slabije od kovalentne veze, skupni efekt velikog broja takvih interakcija može biti značajan. Formiranje svake od tih veza doprinosi smanjenju slobodne energije sustava, i svaka od tih veza dodatno stabilizira vezani kompleks. Taj je učinak pojačan činjenicom da za prekid veze između npr. dva proteina partnera nije dovoljno prekidati sve veze jednu po jednu, nego ih je potrebno prekinuti odjednom.

2.4 Struktura proteina

Nekoliko je ključnih momenata koje treba istaknuti kada se govori o strukturi proteina. Prvo, 3-D struktura proteina je određena njegovim aminokiselinskim sastavom. Drugo, funkcija proteina ovisi o njegovoj strukturi. Treće, izolirani protein obično postoji u jednoj ili malenom broju stabilnih strukturnih formi. Četvrto, najvažnije sile koje stabiliziraju strukture u kojima neki protein postoji su nekovalentne interakcije. Konačno, među golemim brojem jedinstvenih proteinskih struktura, moguće je uočiti zajedničke strukturne uzorke koji nam pomažu da bolje razumijemo prirodu proteina.

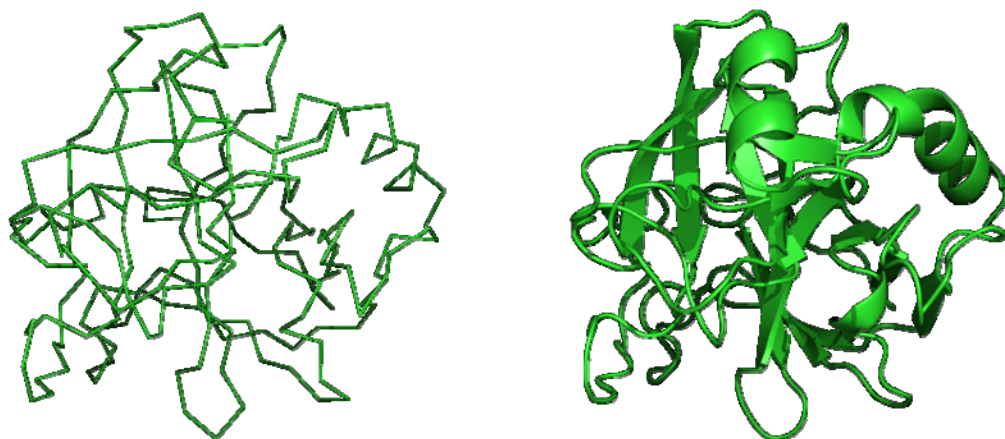
2.4.1 Primarna struktura

Kada govorimo o primarnoj strukturi proteina, mislimo na slijed aminokiselina koji čini taj protein. U uvjetima određenim otapalom, primarna struktura u potpunosti određuje sve ostale, i u budućnosti će vjerojatno biti dovoljna za potpuni opis proteina (u danom biološkom kontekstu).

2.4.2 Sekundarna struktura

Sekundarna struktura se odnosi na nekoliko uobičajenih, lokalnih *motiva* koje tvore određeni, uvijek isti sljedovi aminokiselina. Sekundarna struktura govori

o uzorcima smatanja polipeptidnog lanca. Postoji nekoliko tipova sekundarne strukture koji su posebno stabilni i koji se pojavljuju u izuzetno velikom broju proteina. To su tzv. α -zavojnice i β -ploče.



Slika 5. Lijevo, glavni lanac proteina 1avx (PDBid). Desno sekundarna struktura istog proteina. Vidljive su α -zavojnice i β -ploče (plohe sa strelicama).

2.4.3 Tercijarna struktura

Cjelokupan trodimenzionalni raspored svih atoma u proteinu naziva se tercijska struktura. Dok sekundarna struktura opisuje raspored atoma koji se u polipeptidnom lancu nalaze jedan drugome blizu, tercijska struktura uzima u obzir i mnogo udaljenije dijelove lanca. Aminokiseline koje su u lancu jako udaljene, u tercijskoj strukturi mogu biti prostorno jedna pored druge.

2.4.4 Kvarturna struktura

Neki proteini se sastoje od dva ili više odvojenih polipeptidnih lanaca (dimeri, trimeri, ...). Raspored tih podjedinica u 3-D kompleksu zove se kvarturna struktura.

3 Prianjanje proteina

3.1 Formulacija

Prianjanje je pojam koji se koristi za računalne postupke koji pokušavaju naći *najbolje* slaganje, *najbolji* kompleks dvije molekule: receptora i liganda. Problem bi mogli definirati ovako: „Uz poznavanje koordinata atoma dviju molekula, potrebno je predvidjeti njihovu *točnu* vezanu strukturu“ (Halperin, 2002). Sasvim općenito, ništa osim koordinata atoma nije poznato. Ipak, u praksi, često postoje još neke informacije poput poznavanja mjesta interakcije na jednom od partnera. Korisne su i informacije o homolognim proteinima ili poznavanje funkcije proteina ili njegove evolucije. Jedna od, za ovaj rad relevantnih, podjela problema prianjanja je ona na protein-ligand prianjanje i na protein-protein prianjanje. Kod protein-ligand prianjanja govorimo o pokušaju pravilne kompleksacije velike receptorske molekule, proteina, s malom molekulom, ligandom. Ligand je najčešće molekula lijeka odnosno molekula konstruirana kao potencijalni lijek. Protein-ligand prianjanje je daleko zrelije područje od protein-protein prianjanja, a radi se o (zbog razlike u veličini liganda i proteina) lakšem problemu. Protein-protein prianjanje je pokušaj pravilne kompleksacije dva proteina. Već sama činjenica da se radi o dvije velike molekule povećava složenost problema.

3.2 Problemi kod prianjanja

Proteini su velike molekule, koje se mogu sastojati od tisuća atoma. Jasno je da to čini tipične zadatke prianjanja proteina kao što je uspoređivanje oblika površina računalno iznimno zahtjevnima, čak i uz zanemarenje prisutnosti otapala. Razlog za to je velik broj atoma u kombinaciji s velikim brojem stupnjeva slobode – prostor pretraživanja je šesterodimenzionalan (potrebno je pretraživati 3 rotacijska i 3 translacijska stupnja slobode). Problem je tim veći što proteini nisu rigidne molekule, oni su više ili manje *fleksibilni*. Pri stvaranju kompleksa oba proteina se, ponekad značajno, strukturno mijenjaju. Kada bi u obzir uzeli svaku moguću međusobnu orijentaciju proteina partnera, zajedno sa svakom mogućom individualnom

konformacijom pojedinih proteina, problem bi postao izrazito složen. Složenost problema se može najbolje prikazati činjenicom da danas ne postoje dovoljni računalni resursi za njegovo rješavanje u razumnom vremenu. Zato većina algoritama za početak polazi od takozvanog rekonstrukcijskog (engl. *bound*) prijanjanja – iz poznatog proteinskog kompleksa se izdvoje 3-D strukture pojedinih proteina partnera, i dalje se radi s tim strukturama kao s čvrstim tijelima. Danas se istraživački napor usmjerava prema rješavanju problema fleksibilnosti, budući da je postalo jasno da će značajnija poboljšanja postojećih rezultata uslijediti tek kad se na odgovarajući način uračunaju i konformacijske promijene. Rekonstrukcijsko prijanjanje se koristi kao važan verifikacijski alat, budući da algoritmi sigurno neće dobro funkcionirati za predikcijska (engl. *unbound*) pretraživanja ako ne rade za rekonstrukcijsko.

3.3 Metode prijanjanja

Sva postojeća rješenja za prijanjanje u načelu polaze od traženja kompleksa u kojima se najbolje podudaraju oblici površina proteina partnera. Razlog za to je što se tako može formirati maksimalan broj nekovalentnih međumolekularnih veza. Nakon što se na taj način utvrdi određen ($10^3 - 10^4$) broj kompleksa kandidata, raznim metodama se ti kandidati rangiraju, te se pokušava pravilnim kombiniranjem dobivenih rezultata osigurati što viši rang *točnom* kompleksu. Sljedeća faza je obično *relaksacija* nekoliko najboljih kandidata u konkretnom polju sila, čime se u obzir uzima fleksibilnost bočnih lanaca, i donekle glavnog proteinskog lanca. Dobiveni kompleksi se mogu još jednom ocijeniti da bi se dobio konačan rang. Alati za prijanjanje se razlikuju bilo po načinu na koji definiraju/pretražuju prostor mogućih orijentacija proteina partnera (poklapanje površina), bilo po načinu na koji dobivene orijentacije ocjenjuju. Ponekad je teško definirati granicu između ove dvije faze prijanjanja, budući da se ocjenjivanje vrši implicitno, tijekom traženja orijentacija.

Postoji nekoliko postupaka koji se koriste za traženje pogodnih međusobnih orijentacija proteina. Prvi i ujedno najkorišteniji (uz varijacije na temu) je

jednostavna 3-D korelacija površina proteina. Površine proteina se mapiraju na 3-D rešetku i računanjem korelacije se dolazi do ocjene kvalitete preklapanja. Ako površinu liganda opišemo funkcijom $L(x, y, z)$ a površinu receptora s $R(x, y, z)$, izraz za ocjenu (score) preklapanja je

$$S(i, j, k) = \sum_{x, y, z} L(x + i, y + j, z + k) \cdot R(x, y, z). \quad (3-1)$$

Za računanje korelacije se koristi račun u frekvencijskoj domeni, a za računanje DFT upotrebljava se FFT algoritam, pa složenost korelacije umjesto $O(n^6)$ postaje $O(n^3 \log(n))$. Uz račun u domeni transformacije, izraz za ocjenu preklapanja postaje (Pierce, 2007)

$$S = \frac{1}{N^3} IDFT(IDFT(R) \cdot DFT(L)). \quad (3-2)$$

Prednost ove metode je činjenica da se odjednom računa ocjena za sve moguće translatorne pomake proteina. Nedostaci su to što se uvijek računa ocjena za sve točke rešetke, iako za većinu to nije potrebno (jer za velik broj točaka proteini prodiru jedan u drugoga – fizikalno nemoguće). Također, protein mora biti jednoliko pravokutno otipkan, pa nije moguće npr. površinu proteina otipkati tako da svakom uzorku pripada jednaka površina. Velik nedostatak je i činjenica da je za sve moguće rotacijske pomake potrebno računati novu FFT.

Vrlo popularna metoda je i tzv. *geometric hashing*. Kod te se metode, poznate otprije iz područja robotike i računalnog vida, protein prikazuje pomoću relativnih položaja određenih detalja na površini (udubina i izbočina), te se pomoću *hash* tablica može vrlo brzo računati kvaliteta preklapanja u određenim položajima. Nedostatak je generiranje velikog broja fizikalno nemogućih kompleksa (jedan protein prodire u drugoga), koje je kasnije potrebno eksplicitno eliminirati.

U drugom dijelu postupka se pojedini položaju ocjenjuju. Za tu svrhu je razvijen velik broj tzv. *scoring* funkcija, koje u obzir uzimaju električne karakteristike proteina, empirijski definirana polja sila, hidrofobnost površina i razne druge.

4 Opis rada

4.1 Motivacija

Svaki od postupaka korištenih za prisanjanje proteina ima svoje prednosti i mane. Prednosti su u načelu specifične za pojedini postupak, dok se među manama uvijek provlači velika složenost korištenih algoritama. Čak i kada značajno ubrzamo fazu pretraživanja kao kod *geometric hashinga*, djelomično će nam se to vratiti na naplatu u kasnijoj fazi ocjenjivanja. Bilo bi idealno sačuvati dobre strane postojećih metoda, a istovremeno razviti novi alat koji će biti brz i koji će dobru ocjenu kvalitete kompleksa moći dati već nakon pretraživanja prostora mogućih orijentacija, bez značajnog *poskupljenja* algoritma.

FFT korelacija je vrlo privlačna metoda – osim već navedenih prednosti, pruža nam mogućnost da prilikom pretraživanja koreliramo bilo koje svojstvo koje smatramo relevantnim za ocjenu kvalitete. Tako možemo točkama na površini pridružiti vrijednosti hidrofobnosti i korelirati hidrofobnost, umjesto samog oblika proteina. Možemo na postojeću rešetku mapirati i elektrostatski potencijal, pa korelirati njega. Nedostaci te metode su relativno velika složenost i potreba računanja nove korelacije za svaki rotacijski pomak. Postavlja se pitanje – možemo li površinu proteina opisati kvalitetnije nego golemim brojem koordinata točaka koje se nalaze na površini. Ako pokušamo pravokutni otvor prikazati Fourierovim redom, morati ćemo potrošiti znatno više koeficijenata i rezultat će biti znatno lošiji nego ako za bazu uzmemo npr. Rademacherovu. Wavelet transformacija daje vrlo efikasan opis polinoma nižih redova, svodeći potrebu za otipkavanjem složenih funkcija na potrebu za čuvanjem malog broja wavelet koeficijenata. Ideja je da se pronađe reprezentacija površine proteina koja će zadovoljavati sljedeće uvjete:

1. Učinkovitost reprezentacije – potrebna količina podataka za obradu mora biti manja od one bez transformiranja

2. Kvaliteta reprezentacije u računskom smislu – lako računanje kvalitete preklapanja, odnosno korelacije površina
3. Rješavanje problema rotacija – eliminacija potrebe za rotiranjem velike rešetke, mogućnost rotiranja isključivo u domeni reprezentacije (transformacije)

Još od rješavanja Schrödingerove valne jednadžbe za vodikov atom, jasno je da su kugline funkcije (sferni harmonici) pogodne u raznim klasama problema za prikazivanje različitih veličina. Koriste se u kvantnoj mehanici, klasičnoj elektrodinamici, kao i u geodeziji i raznim drugim fizikalnim disciplinama. Razvoj površine molekula u redove kuglinih funkcija je ideja koja se počela koristiti krajem 80-ih godina prošloga stoljeća, ali njihova druga pogodna svojstva su se počela iskorištavati tek kasnije. Interesantna svojstva prilikom rotacija oko koordinatnih osi potakla su nekoliko grupa na pokušaj implementacije shema za molekularnu superpoziciju i prijanjanje proteina pomoću kuglinih funkcija. Krajnji potpuni rezultat je prvi implementirao prof. David W. Ritchie, ph.d., Department of Computing science, King's College, University of Aberdeen. Mnoge grupe su, usprkos dobrom razumijevanju problema, imale problema sa stvarnim implementacijama – razlozi vjerojatno leže u lošem *prevođenju* matematičkih formula u programske jezike, ali i u problemima s numeričkom preciznosti (poglavlje 9) .

Korištenje kuglinih funkcija zadovoljava sva tri navedena uvjeta, budući da se odabirom broja članova reda lako može balansirati potrebnu preciznost prikaza s računalnim zahtjevima, izrazi za korelaciju u domeni transformacije su jednostavni i pokazano je da se kugline funkcije transformiraju među sobom prilikom rotiranja oko koordinatnih osi, drugim riječima prostor kuglinih funkcija je zatvoren s obzirom na rotacije. To znači da pravilnom transformacijom koeficijenata razvoja (čiji je broj zasigurno manji od broja točaka kojima bi prikazali površinu) možemo rotirati proteine. Pogodno je i to što kugline funkcije u sfernim polarnim koordinatama čine ortogonalnu bazu, pa je razvoj površina u njihove redove matematički vrlo jednostavan.0

4.2 Razrada zadatka

U ovom radu je opisana implementacija dokaza valjanosti koncepta prijanjanja proteina u sfernim polarnim koordinatama, uz korištenje kuglinih funkcija za razvoj površine. Za rješavanje problema potrebno je:

1. Osigurati matematički aparat za manipulaciju redovima kuglinih funkcija, odnosno specijalnih radijalnih funkcija
2. Odabrati način za definiranje *točkaste površine* proteina, odnosno za otipkavanje površine proteina i računanje vektora normala u odgovarajućim točkama
3. Izračunati koeficijente uz razne redove i stupnjeve kuglinih funkcija, odnosno radijalne funkcije
4. Proračunati rotacijske i translacijske matrice
5. Implementirati samo prijanjanje
6. Odabrati način mapiranja svojstva hidrofobnosti na površinu proteina
7. Testirati dobiveni sustav

5 Matematička razrada. Kugline funkcije

5.1 Uvodna razmatranja

Kugline funkcije (sferni harmonici) su funkcije sfernih koordinata (θ, ϕ) . Ovdje će biti korištene oznake $Y_{lm}(\theta, \phi)$ za kompleksne kugline funkcije, odnosno $y_{lm}(\theta, \phi)$ za realne. U oba slučaja, indeks m može biti negativan, uz uvjet $|m| \leq l$.

Dva su razloga zbog kojih ove funkcije dobre za prikaz površine proteina. Skup kuglinih funkcija je potpuna ortonormirana baza s obzirom na polarni integral (drugim riječima, na jediničnoj sferi). To možemo zapisati

$$\int_0^{2\pi} \int_0^\pi y_{lm}(\theta, \phi) y_{kj}(\theta, \phi) \sin \theta d\theta d\phi = \delta_{kj} \delta_{mj}. \quad (5-1)$$

Odatle slijedi da za bilo koju funkciju, μ , parametriziranu u polarnim koordinatama, (θ, ϕ) , koju prikažemo kao linearnu kombinaciju kuglinih funkcija

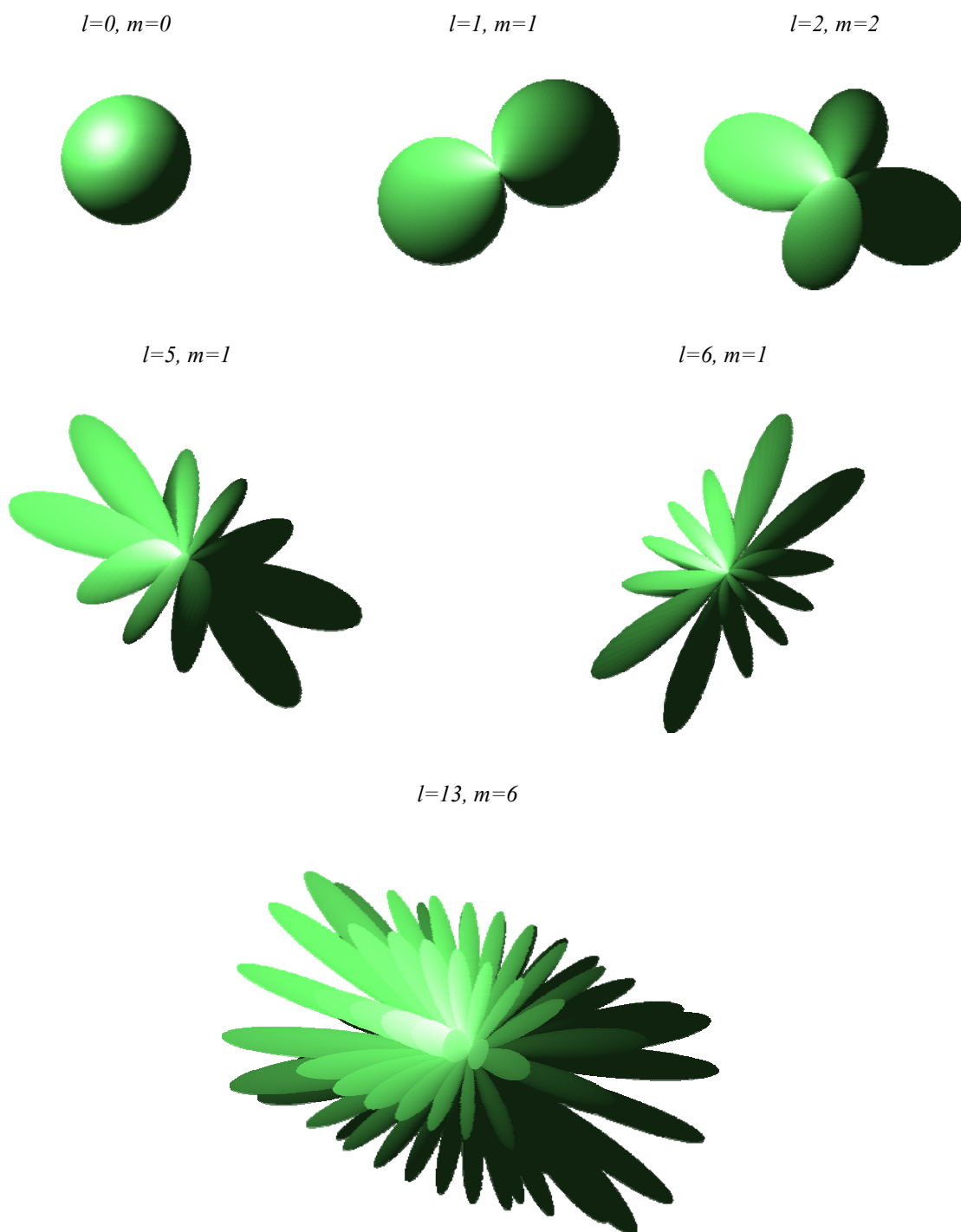
$$\mu(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l a_{lm} y_{lm}(\theta, \phi). \quad (5-2)$$

vrlo lako možemo izračunati koeficijente a_{lm} . Ako jednadžbu (5-2) pomnožimo s $y_{lm}(\theta, \phi)$ i integriramo preko cijelog prostornog kuta, dobivamo

$$\int_0^{2\pi} \int_0^\pi \mu(\theta, \phi) y_{kj}(\theta, \phi) \sin \theta d\theta d\phi = \int_0^{2\pi} \int_0^\pi \sum_{l=0}^{\infty} \sum_{m=-l}^l a_{lm} y_{lm}(\theta, \phi) y_{kj}(\theta, \phi) \sin \theta d\theta d\phi. \quad (5-3)$$

Zamjenom redoslijeda integracije i sumacije i uvažavanjem jednadžbe ortonormiranosti (5-1), dobivamo

$$a_{lm} = \int_0^{2\pi} \int_0^\pi \mu(\theta, \phi) y_{lm}(\theta, \phi) \sin \theta d\theta d\phi. \quad (5-4)$$



Slika 6. Primjeri nekih kuglinih funkcija. Zbog bolje vizualizacije, nisu prikazane na sferi, nego kako je to uobičajeno, kao površina eksplicitno zadana u sfernim polarnim koordinatama jednadžbom $r = y_{lm}(\theta, \phi)$. Iz slike je jasno zašto red kuglinih funkcija zovu sfernom polarnom analogijom Fourierovog reda, odnosno odakle naziv sferni harmonici.

Koeficijenti a_{lm} su jedinstveni, a razvoj (5-2) je strogo konvergentan u smislu da se greška (ostatak reda) monotono smanjuje kako l teži u beskonačnost. To pak znači da je većina *energije* funkcije sadržana u prvim članovima reda odnosno da prekidanjem reda nakon dovoljno velikog l možemo dovoljno vjerno prikazati funkciju $\mu(\theta, \phi)$.

Drugi važan razlog za odabir kuglinih funkcija je činjenica da je prostor kuglinih funkcija zatvoren s obzirom na rotacije, i da se kugline funkcije svakog reda l rotacijom transformiraju između sebe na predvidljiv način. To možemo zapisati

$$\hat{R}(\alpha, \beta, \gamma)y_{lm}(\theta, \phi) = \sum_{m'} y_{lm'}(\theta, \phi)R_{m'm}^{(l)}(\alpha, \beta, \gamma), \quad (5-5)$$

pri čemu $\hat{R}(\alpha, \beta, \gamma)$ predstavlja operator rotacije, a (α, β, γ) su Eulerovi kutovi. Svaka $R^{(l)}$ je realna $(2l+1) \times (2l+1)$ rotacijska matrica. Način računanja elemenata rotacijskih matrica opisan je kasnije. Iz navedenog slijedi da proizvoljnu funkciju $\mu(\theta, \phi)$ možemo rotirati da bi dobili novu funkciju, $\mu'(\theta, \phi)$ mijenjajući koeficijente razvoja. Zapišimo novu funkciju kao $\mu'(\theta, \phi) = \hat{R}(\alpha, \beta, \gamma)\mu(\theta, \phi)$ i uvrstimo odgovarajuće izraze za svaku stranu, uzimajući $l_{\max} = L$,

$$\sum_{l=0}^L \sum_{m=-l}^l a'_{lm} y_{lm}(\theta, \phi) = \sum_{l'=0}^L \sum_{m'=-l'}^{l'} a_{l'm'} \sum_q y_{l'q}(\theta, \phi) R_{qm'}^{(l')}(\alpha, \beta, \gamma). \quad (5-6)$$

Množeći obje strane s $y_{kj}(\theta, \phi)$ i integrirajući preko cijele sfere (prostornog kuta), iščezavaju svi članovi osim za $k = l = l'$ i $j = m = q$ te dobivamo izraz za koeficijente rotirane funkcije,

$$a'_{lm} = \sum_{m'} a_{lm'} R_{mm'}^{(l)}(\alpha, \beta, \gamma). \quad (5-7)$$

Ovo temeljno rotacijsko svojstvo kuglinih funkcija su prepoznali Leicester *et al.* (1994), ali nisu bili uspješni prilikom implementacije. Ritchie i Kemp (1998) su uspješno implementirali rotacije kuglinih funkcija i time potvrdili točnost matematičkih izraza.

5.2 Kugline funkcije

5.2.1 Kompleksne kugline funkcije

Laplaceova diferencijalna jednačba u sfernim polarnim koordinatama glasi

$$\nabla^2 f = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial f}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial f}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 f}{\partial \phi^2} = 0. \quad (5-8)$$

Rješenje kutnog dijela jednačbe (dobivenog jednostavnom separacijom varijabli) su kugline funkcije,

$$Y_{lm}(\theta, \phi) = \mathcal{Y}_{lm}(\mu) Z_m(\phi); \quad \mu = \cos \theta \quad (5-9)$$

gdje je

$$\mathcal{Y}_{lm}(\mu) = \left[\frac{(2l+1)(l-m)!}{2(l+m)!} \right]^{1/2} P_l^m(\mu), \quad (5-10)$$

te

$$Z_m(\phi) = \frac{1}{\sqrt{2\pi}} e^{jm\phi}. \quad (5-11)$$

$P_l^m(\mu)$ su pridruženi Legendreovi polinomi. Oni se mogu računati bilo putem rekurzivne formule,

$$P_l^l(\theta) = (-1)^l \frac{(2l)!}{l!} \left(\frac{\sin \theta}{2} \right)^l; \quad m = l, \quad (5-12)$$

$$P_l^m(\theta) = \frac{-2(m+1)\text{ctg}(\theta)P_l^{m+1}(\theta) - P_l^{m+2}(\theta)}{(l-m)(l+m+1)}; \quad m < l \quad (5-13)$$

bilo pomoću Rodriguesove formule,

$$P_l^m(\theta) = (-1)^m \frac{(1-\theta^2)^{m/2}}{2^l l!} \frac{d^{l+m}}{d\theta^{l+m}} (\theta^2 - 1)^l; \quad l \geq m \geq 0. \quad (5-14)$$

Pridruženi Legendreovi polinomi za negativne vrijednosti m se mogu izračunati pomoću relacije

$$\mathcal{G}_{\bar{m}}(\mu) = (-1)^m \mathcal{G}_{|m|}(\mu), \quad \bar{m} := -m. \quad (5-15)$$

5.2.2 Wignerove rotacijske matrice

U izrazu za rotaciju kuglinih funkcija,

$$\hat{R}(\alpha, \beta, \gamma) Y_{lm}(\theta, \phi) = \sum_{m'} Y_{lm'}(\theta, \phi) D_{m'm}^{(l)}(\alpha, \beta, \gamma), \quad (5-16)$$

unitarne rotacijske matrice $D^{(l)}(\alpha, \beta, \gamma)$ pronašao je Wigner, a elementi matrica dani su s

$$D_{m'm}^{(l)}(\alpha, \beta, \gamma) = e^{-jm'\alpha} d_{m'm}^{(l)}(\beta) e^{-jm\gamma}, \quad (5-17)$$

gdje je

$$d_{m'm}^{(l)}(\beta) = \left[\frac{(l+m')!(l-m')!}{(l+m)!(l-m)!} \right]^{1/2} \times \sum_{k=\max(0, m-m')}^{\min(l-m', l+m)} \left[\frac{(-1)^{k+m'-m} \binom{l+m}{k} \binom{l-m}{l-m'-k}}{(\cos \beta/2)^{2l+m-m'-2k} (\sin \beta/2)^{2k+m'-m}} \right]. \quad (5-18)$$

Matrice $d^{(l)}$ su višestruko simetrične:

$$d_{m'm}^{(l)}(\beta) = (-1)^{m'-m} d_{mm'}^{(l)}(\beta) = (-1)^{m'-m} d_{\bar{m}'\bar{m}}^{(l)}(\beta), \quad (5-19)$$

što znatno ubrzava izračunavanje njihovih elemenata.

5.2.3 Realne kugline funkcije

Budući da se bavimo isključivo koreliranjem realnih veličina, više smisla ima raditi s realnim kuglinim funkcijama, koje se mogu dobiti linearnim kombiniranjem kompleksnih:

$$y_{lm}(\theta, \phi) = \begin{cases} (Y_{lm}(\theta, \phi) + Y_{lm}(\theta, \phi)^*) / \sqrt{2} & m > 0 \\ Y_{l0}(\theta, \phi) & m = 0. \\ -j(Y_{l\bar{m}}(\theta, \phi) - Y_{l\bar{m}}(\theta, \phi)^*) / \sqrt{2} & m < 0 \end{cases} \quad (5-20)$$

Uvrštavanjem odgovarajućih izraza, dobiva se

$$y_{lm}(\theta, \phi) = \begin{cases} \mathcal{G}_{lm}(\theta)(\cos m\phi) / \sqrt{\pi} & m > 0 \\ \mathcal{G}_{lm}(\theta) / \sqrt{2\pi} & m = 0. \\ \mathcal{G}_{l\bar{m}}(\theta)(\sin \bar{m}\phi) / \sqrt{\pi} & m < 0 \end{cases} \quad (5-21)$$

Realne kugline funkcije se dakle mogu zapisati na način

$$y_{lm}(\theta, \phi) = \mathcal{G}_{|m|}(\mu)\varphi_m(\phi); \quad \mu = \cos\theta, \quad (5-22)$$

gdje je

$$\varphi_m(\phi) = \begin{cases} \cos m\phi / \sqrt{\pi} & m > 0 \\ 1 / \sqrt{2\pi} & m = 0. \\ \sin \bar{m}\phi / \sqrt{\pi} & m < 0 \end{cases} \quad (5-23)$$

5.2.4 Realne rotacijske matrice

Linearne kombinacije poput (5-20) čuvaju rotacijsku simetriju, pa realne kugline funkcije, kao i kompleksne, imaju povoljna rotacijska svojstva,

$$\hat{R}(\alpha, \beta, \gamma)y_{lm}(\theta, \phi) = \sum_{m'} y_{lm'}(\theta, \phi) R_{m'm}^{(l)}(\alpha, \beta, \gamma), \quad (5-24)$$

gdje su $R^{(l)}(\alpha, \beta, \gamma)$ realne rotacijske matrice. Ako linearne kombinacije (5-20) zapišemo kao sumu,

$$y_{lm}(\theta, \phi) = \sum_{m'} A_{mm'}^{(l)} Y_{lm'}(\theta, \phi), \quad (5-25)$$

može se pokazati da je za svaki l realna rotacijska matrica dana jednadžbom

$$R = AD\tilde{A}, \quad (5-26)$$

gdje je \tilde{A} Hermitska transpozicija matrice A . Uz pažljiv račun, dobiva se izraz za elemente matrice (Su, 1994), (Ritchie, 1998),

$$R_{m'm}^{(l)} = \begin{cases} d_{m'm}^{(l)}(\beta) \cos(m\gamma + m'\alpha) + (-1)^{m'} d_{\bar{m}'m}^{(l)}(\beta) \cos(m\gamma - m'\alpha) & , m' > 0, m > 0 \\ d_{0m}^{(l)}(\beta) \sqrt{2} \cos(m\gamma) & , m' = 0, m > 0 \\ (-1)^{m'+1} d_{m'm}^{(l)}(\beta) \sin(m\gamma + m'\alpha) + d_{\bar{m}'m}^{(l)}(\beta) \sin(m\gamma - m'\alpha) & , m' < 0, m > 0 \\ d_{m'0}^{(l)}(\beta) \sqrt{2} \cos(m'\alpha) & , m' > 0, m = 0 \\ d_{00}^{(l)}(\beta) & , m' = 0, m = 0 \\ (-1)^{m'+1} d_{m'0}^{(l)}(\beta) \sqrt{2} \sin(m'\alpha) & , m' < 0, m = 0 \\ (-1)^m d_{m'm}^{(l)}(\beta) \sin(m\gamma + m'\alpha) + (-1)^{m+m'} d_{\bar{m}'m}^{(l)}(\beta) \sin(m\gamma - m'\alpha) & , m' > 0, m < 0 \\ (-1)^m d_{0m}^{(l)}(\beta) \sqrt{2} \sin(m\gamma) & , m' = 0, m < 0 \\ (-1)^{m+m'} d_{m'm}^{(l)}(\beta) \sin(m\gamma + m'\alpha) + (-1)^{m+1} d_{\bar{m}'m}^{(l)}(\beta) \cos(m\gamma - m'\alpha) & , m' < 0, m < 0 \end{cases} \quad (5-27)$$

Ako odaberemo $\beta = \gamma = 0$, dobivamo samo jednu rotaciju α oko z-osi, koja ima jednostavniju formu. Budući da je

$$\hat{R}(\alpha, 0, 0)y_{lm}(\theta, \phi) = y_{lm}(\theta, \phi - \alpha), \quad (5-28)$$

uz pomoć trigonometrijskih formula

$$\cos(\alpha \pm \beta) = \cos \alpha \cos \beta \mp \sin \alpha \sin \beta \quad (5-29)$$

$$\sin(\alpha \pm \beta) = \sin \alpha \cos \beta \pm \cos \alpha \sin \beta \quad (5-30)$$

lako je pokazati da vrijedi

$$\hat{R}(\alpha, 0, 0)y_{lm}(\theta, \phi) = y_{lm}(\theta, \phi) \cos m\alpha + y_{l\bar{m}}(\theta, \phi) \sin m\alpha. \quad (5-31)$$

6 Točkasta površina proteina

6.1 Razne površine molekula

Strukture proteina partnera dobivamo iz PDB (*Protein Data Bank*) baze. PDB datoteke sadrže kartezijeve koordinate atoma od kojih je protein sastavljen. Za potrebe prisanjanja proteina, važno je odrediti funkciju površine proteina, odnosno njezine uzorke – prvo moramo definirati površinu proteina, a nakon toga pronaći način kako da je izračunamo.

6.1.1 Van der Waalsova površina

U kontekstu računanja molekularne površine, molekulom se smatra skup sfera pridruženih svakom atomu koje se međusobno presijecaju. Ako svakom atomu pridružimo njegov van der Waalsov radijus, odnosno svaki atom zamijenimo sferom van der Waalsovog radijusa, dobivamo van der Waalsovu površinu.

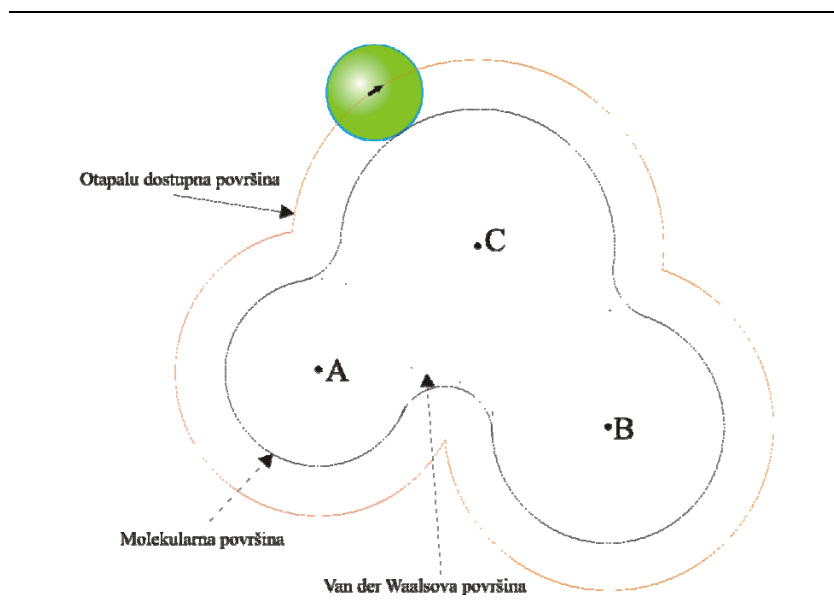
6.1.2 Otapalu dostupna površina (*solvent-accessible surface*)

Otapalu dostupnu površinu generira središte sferne probe (modela molekule otapala, najčešće vode), kada se ona kotrlja po van der Waalsovoj površini. Kao radijus probe uzima se najčešće van der Waalsov radijus vode (1.4 Å). Otapalu dostupna površina je ekvivalentna van der Waalsovoj, ako kod potonje vdW radijusima dodamo radijus probe.

6.1.3 Molekularna površina

Molekularna površina je unija kontaktnih odsječaka, kada sonda dodiruje jedan površinski atom i složenijih, konkavnih odsječaka kada sonda dodiruje više atoma. Molekularna površina obuhvaća područje proteina nedostupno otapalu, i bolja je procjena površine od van der Waalsove i otapalu dostupne za rad s hidrofobnim interakcijama. Pretpostavlja se da je tako definirana molekularna površina dobra aproksimacija dielektrične granice između polarnog otapala visoke dielektričnosti i unutrašnjosti proteina niske dielektričnosti u širokoj klasi problema (Vorobjev, 1997).

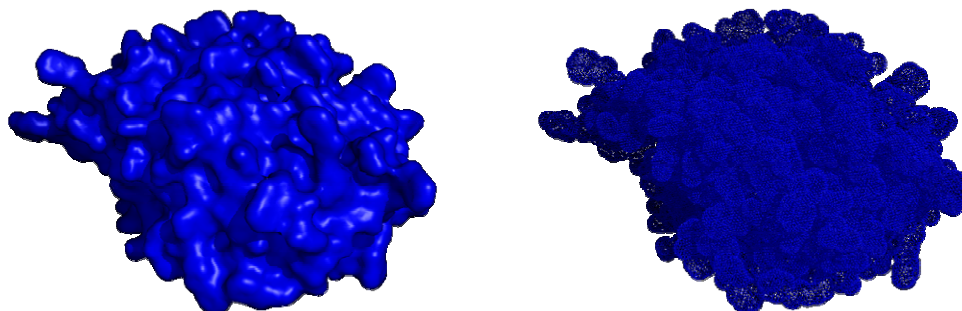
Sva tri tipa molekularne površine prikazana su na slici 7.



Slika 7. Tri tipa površina molekula za jednostavnu troatomnu molekulu. Prikazana je i sonda otapala kojom su površine generirane. Iz slike se vidi način konstrukcije otapalu dostupne površine i molekularne površine. Konkavna područja na molekularnoj površini nastaju kada sonda dodiruje više atoma istovremeno.

6.1.4 Namjena reprezentacije površine

Potreba za računanjem površina molekula pojavljuje se u alatima za molekularnu vizualizaciju, ali i u mnogim drugim primjenama gdje je potrebno imati mnogo manje uzoraka. Budući da su uzorci zapravo koordinate točaka koje se nalaze na površini molekule, često se takva reprezentacija naziva točkasta površina (engl. *dot surface*). Dobro rješenje mora dati točkastu površinu koja je maksimalno homogena, ne sadrži singularne točke, invarijantna je s obzirom na rotacije i stabilna kod malih promjena konformacije. Uvjet kontinuiranosti (nepostojanje singularnih točaka), zapravo je uvjet na normale u pojedinim točkama. Izračunate normale moraju biti kontinuirane, odnosno ne smiju naglo mijenjati smjer. Na slici 8 prikazana je razlika između površine grafičke kvalitete, i točkaste površine.



Slika 8. Površina proteina izonikotinimidilatirane alkohol dehidrogenaze (PDBid 7adh). Lijevo je prikazana površina kakva se koristi za vizualizaciju, a desno tzv. točkasta površina, odnosno određen broj uzoraka površine. U oba slučaja radi se o van der Waalsovoj površini.

7 Radijalne funkcije. Ljuska proteina

7.1 Problem s ploham

Površine, ili bolje rečeno plohe, se u sfernim polarnim koordinatama mogu parametrizirati na različite načine. Općenito, sve točke koje leže na nekoj plohi u 3-D prostoru zadane su s

$$\vec{x}(\mu, \nu) = f_x(\mu, \nu)\vec{i} + f_y(\mu, \nu)\vec{j} + f_z(\mu, \nu)\vec{k}, \quad (7-1)$$

pri čemu je problem odrediti funkcije f_x, f_y, f_z . Takve plohe su, naravno, dvodimenzionalni objekti. Pokazalo se da, iako su pogodne za probleme poput molekularne superpozicije, vrlo loše funkcioniraju u kontekstu prisanjanja proteina. Razlozi leže vjerojatno upravo u dvodimenzionalnosti – plohe nemaju debljinu. Prodiranje jednog proteina u drugi na taj način ne može biti dobro kvantificirano, za razliku od slučaja kada bi umjesto površine proteina koristili ljusku određene debljine. Izravno rješenje tog problema je uvođenje posebnih radijalnih funkcija, $R_n(r)$, kojima ćemo proširiti postojeću bazu kuglinih funkcija. Nove funkcije baze sada su

$$F_{nlm}(\underline{r}) = R_n(r)y_{lm}(\theta, \phi), \quad (7-2)$$

i one su trodimenzionalne. Neko svojstvo površine proteina, ili funkciju same površine, kojoj ćemo sada morati dodijeliti volumen, implicitno (u sfernim polarnim koordinatama) zadajemo razvojem:

$$\sigma(\underline{r}) = \sum_{nlm}^N a_{nlm} F_{nlm}(\underline{r}). \quad (7-3)$$

7.2 Pogodne radijalne funkcije

7.2.1 Schrödingerova jednadžba za vodikov atom

Pokazalo se da je konstrukciju *dobrih* radijalnih funkcija dobro početi od radijalnog dijela rješenja valne jednadžbe za vodikov atom. Schrödingerova

valna jednadžba za vodikov atom (negativni elektron koji kruži u coulombskom elektrostatskom potencijalu jezgre – protona) glasi

$$\nabla^2 \psi(\underline{x}) + V(\underline{x})\psi(\underline{x}) = E\psi(\underline{x}), \quad (7-4)$$

gdje je $V(\underline{x})$ sferno simetrična funkcija elektrostatskog potencijala jezgre, a E je energija sustava. Rješenja jednadžbe (7-4) su oblika

$$\psi(\underline{x}) = R_{nl}(r)y_{lm}(\theta, \phi); \quad n > l \geq |m|, \quad (7-5)$$

pri kvantni brojevi n, l, m određuju dozvoljena stanja atoma. Radijalne funkcije u rješenju jednadžbe su zadane izrazom

$$R_{nl}(r) = N_{nl}e^{-\rho/2}\rho^l L_{n+l}^{2l+1}(\rho), \quad (7-6)$$

gdje su $L_{n+l}^{2l+1}(\rho)$ pridruženi Laguerreovi polinomi, a ρ skalirana udaljenost od središta atoma (protona). N_{nl} je normalizacijski koeficijent. Važno je uočiti da se u biti radi o oscilatornim Laguerreovim funkcijama, koje su eksponencijalno prigušene. Te funkcije, bez određenih zahvata, nisu pogodne za modeliranje proteina, budući da se udaljenost na kojoj eksponencijalno iščezavaju kvadratno povećava kako n raste. Dobre funkcije bi bile lokalizirane upravo oko ljuske proteina i iščezavale iza nje, u prvoj aproksimaciji uvijek na istoj udaljenosti.

7.2.2 Orbitale Gaussovog tipa

Eksperimenti su pokazali da su nešto bolje funkcije, zbog relativno lakog računanja određenih integrala preklapanja (poglavlje 9), tzv. GTO funkcije (*Gaussian Type Orbitals*), koje dolaze iz rješavanja valne jednadžbe za harmonički potencijal. Normalizirane GTO funkcije zadane su s

$$R_{nl}(r) = \left[\frac{2}{\lambda^{3/2} \pi^{1/2}} \frac{(n-l-1)!}{(1/2)_n} \right]^{1/2} \exp(-\rho^2/2) \rho^l L_{n-l-1}^{(l+1/2)}(\rho^2), \quad (7-7)$$

gdje je $\rho^2 = r^2 / \lambda$ skalirani radijus. Za potrebe prijanjanja proteina korišten je $\lambda = 20$. Također, $(x)_n = \Gamma(x+n)/\Gamma(x)$. Ove funkcije su slične radijalnim valnim funkcijama za vodikov atom, uz sljedeće razlike: 1. prigušenje je Gaussovo, a ne eksponencijalno, 2. korišteni su generalizirani Laguerreovi polinomi, koji su zadani binomnim razvojem:

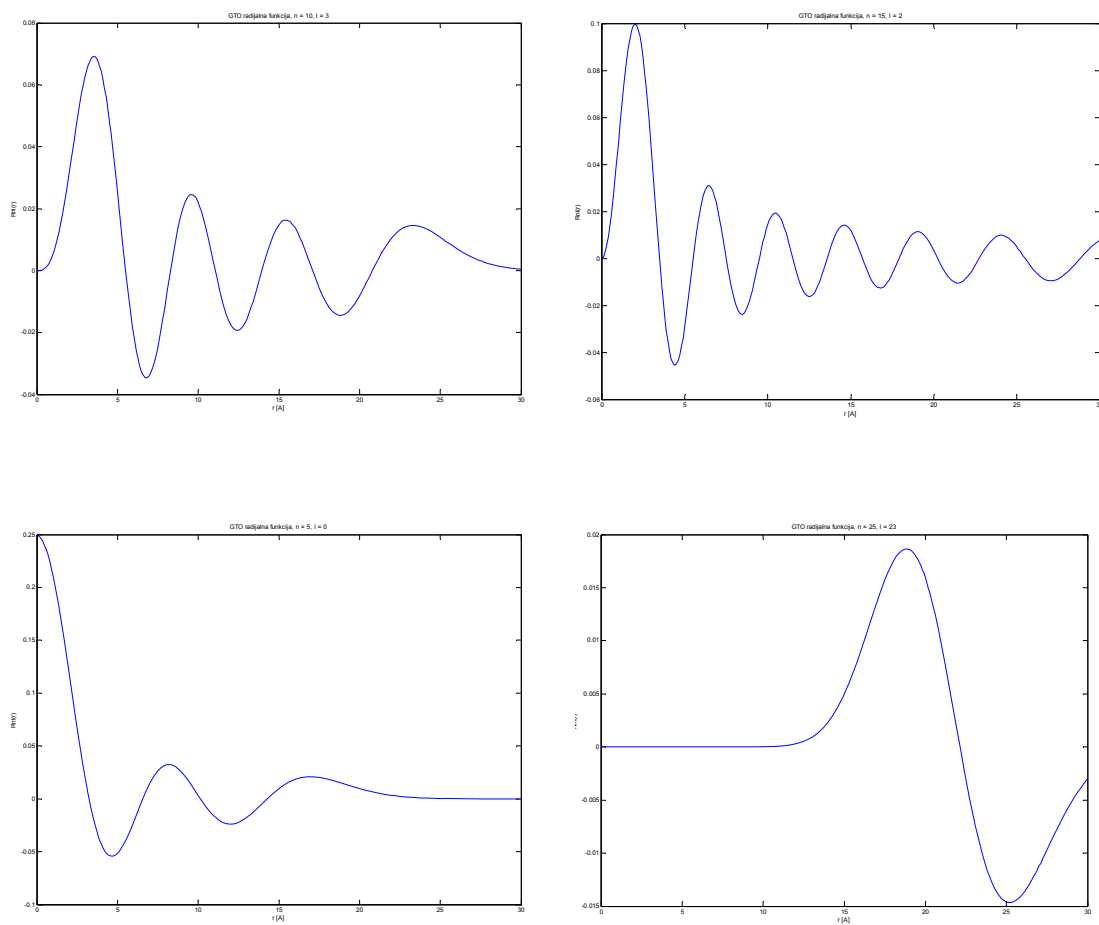
$$L_k^{(\alpha)}(x) = \sum_{j=0}^k \binom{k+\alpha}{k-j} \frac{(-x)^j}{j!}. \quad (7-8)$$

Polinomi viših redova se mogu računati pomoću rekurzivne relacije:

$$(k+1)L_{k+1}^{(\alpha)}(x) = (2k+\alpha+1-x)L_k^{(\alpha)}(x) - (k+\alpha)L_{k-1}^{(\alpha)}(x), \quad (7-9)$$

uz početne uvjete i $L_1^{(\alpha)}(x) = \alpha+1-x$.

Grafovi nekih radijalnih funkcija prikazani su na slici 9. Vidi se da broj nula radijalnih funkcija ovisi o razlici između parametara n i l .



Slika 9. Grafovi radialnih funkcija za različite vrijednosti parametara n i l .

7.3 3-D parametrizacije

U izrazu za razvoj određenog svojstva u red kuglinih funkcija i radijalnih funkcija, do određenog reda N ,

$$\sigma_N(\underline{r}) = \sum_{nlm}^N a_{nlm} R_n(r) y_{lm}(\theta, \phi), \quad (7-10)$$

zbog uvjeta $n > l \geq |m| \geq 0$, korišten je skraćeni zapis

$$\sum_{nlm}^N = \sum_{n=1}^N \sum_{l=0}^{n-1} \sum_{m=-l}^l. \quad (7-11)$$

Broj elemenata u takvoj sumaciji je $N(N+1)(2N+1)/6$. Zbog ortonormalnosti funkcija baze, koeficijenti razvoja a_{nlm} se računaju kao

$$a_{nlm} = \int \sigma(\underline{r}) F_{nlm}(\underline{r}) dV, \quad (7-12)$$

gdje F_{nlm} označava složenu funkciju baze. Ovako parametrizirane trodimenzionalne veličine se rotiraju kao i u 2-D slučaju, budući da operator rotacije ne utječe na radijalne funkcije:

$$\begin{aligned} \hat{R}(\alpha, \beta, \gamma) \sigma_N(\underline{r}) &= \sum_{nlm}^N \hat{R}(\alpha, \beta, \gamma) [a_{nlm} R_n(r) y_{lm}(\theta, \phi)] \\ &= \sum_{nlm}^N R_n(r) \hat{R}(\alpha, \beta, \gamma) [a_{nlm} y_{lm}(\theta, \phi)]. \\ &= \sum_{nlm}^N R_n(r) y_{lm}(\theta, \phi) a'_{nlm} \end{aligned} \quad (7-13)$$

Kao i prije, koeficijenti istog reda se među sobom transformiraju prema

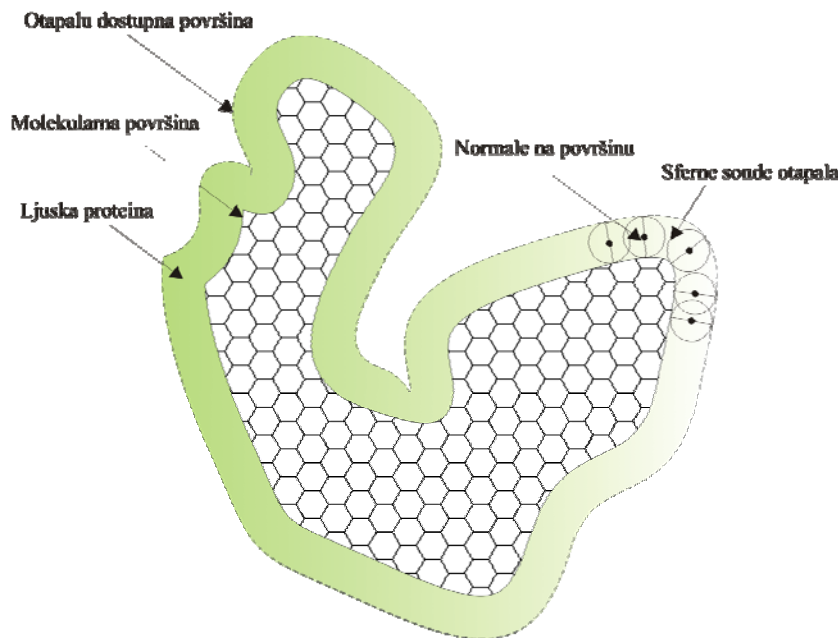
$$a'_{nlm} = \sum_{m'=-l}^l a_{nlm'} R_{mm'}^{(l)}(\alpha, \beta, \gamma). \quad (7-14)$$

7.4 Funkcija ljuske proteina

Konceptu površine proteina za potrebe 3-D parametrizacije moramo dodijeliti volumen. U tu svrhu konstruiramo *funkciju ljuske proteina*, koja opisuje volumen između molekularne površine i otapalu dostupne površine. Konstrukcija ljuske proteina skicirana je na slici 10. Funkciju ljuske zapisujemo kao

$$\rho(\underline{r}) = \begin{cases} 1; & \underline{r} \in \text{ljuske} \\ 0; & \text{ostalo} \end{cases} \quad (7-15)$$

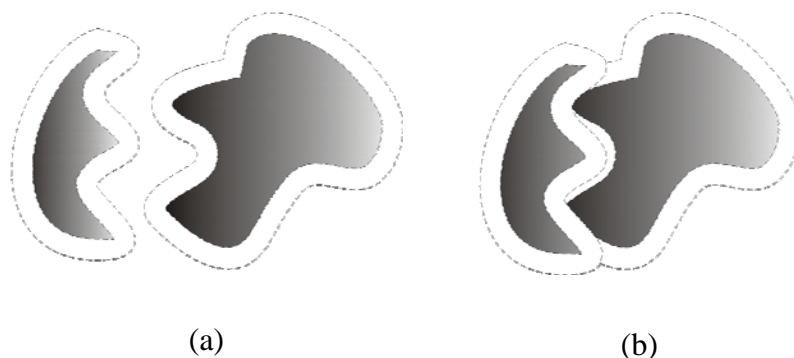
Za takvu funkciju onda računamo koeficijente razvoja prema (7-12).



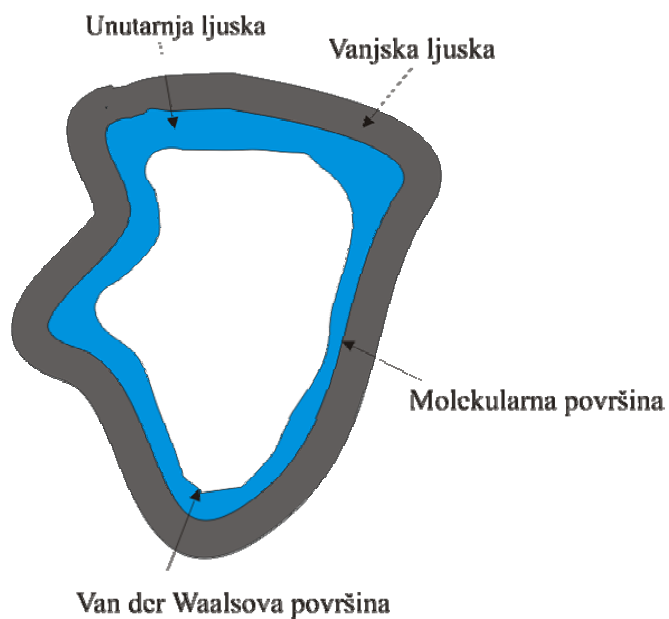
Slika 10. Ilustracija konstrukcije *ljuske* proteina. Ljuskom smatramo volumen omeđen molekularnom površinom i otapalu dostupnom površinom.

Kada bi se tražilo maksimalno preklapanje dvije ovako definirane ljuske, najpovoljniji rezultat bio bi kada bi molekularne površine bile udaljene upravo za debljinu ljuske. Takva situacija ilustrirana je na slici 11. To nije poželjno, pa se zapravo radi s modelom dvostruke ljuske. U takvom modelu osim već definirane, vanjske ljuske, definiramo i unutarnju ljusku, kao uniju van der

Waalsovih radijusa svih atoma koji se nalaze na molekularnoj površini.



Slika 11. Komplementarni proteini na optimalnom razmaku uz korištenje jedne ljuske. Ako proteine sa slike (a) približavamo do maksimalnog preklapanja ljuski, vidljivo je na slici (b) da će oni ostati razmaknuti za debljinu ljuske (njihove molekularne površine).



Slika 12. Ilustracija koncepta dvostruke ljuske proteina.

Unutarnja ljuska zadana je izrazom sličnim izrazu za vanjsku, a ilustrirana je slikom 12,

$$\tau(\underline{r}) = \begin{cases} 1; & \underline{r} \in \text{površinskog atoma} \\ 0; & \text{ostalo} \end{cases} \quad (7-16)$$

Za definiciju unutarnje ljuske mogli smo uzeti uniju van der Waalsovih volumena svih atoma u unutrašnjosti proteina, no to bi znatno usporilo kasnije račune, bez nekog vidljivog doprinosa njihovoj kvaliteti. Uz korištenje ovako zadanih ljuski proteina, ocjenu kvalitete preklapanja, odnosno ocjenu komplementarnosti oblika površina dva proteina, A i B možemo računati kao

$$S = \int \rho_A(\underline{r}_A) \tau_B(\underline{r}_B) dV + \int \tau_A(\underline{r}_A) \rho_B(\underline{r}_B) dV - Q \int \tau_A(\underline{r}_A) \tau_B(\underline{r}_B) dV, \quad (7-17)$$

pri čemu je posljednji integral zapravo kazna za preklapanje unutrašnjih ljuski. Oduzimanjem tog integrala s težinom Q zapravo kažnjavamo prodiranje jednoga proteina u unutrašnjost drugoga – fizikalno nemoguću situaciju.

8 Shema prijanjanja

U prethodnim poglavljima opisani su praktično svi alati potrebni za implementaciju prijanjanja u sfernim polarnim koordinatama. Ostaje opisati sam postupak prijanjanja te način računanja tzv. integrala preklapanja odnosno translacijskih matrica.

8.1 Računanje ocjene kvalitete preklapanja

Da bi opisani matematički aparat bio upotrebljiv za prijanjanje, potrebno je pojednostaviti izraz za ocjenu preklapanja (7-17). Članovi tog izraza mogu se računati kao

$$\int \rho_A \tau_B dV = \sum_{nlm}^N \sum_{n'l'm'}^N a_{nlm}^\rho b_{n'l'm'}^\tau \int R_{nl}(r_A) R_{n'l'}(r_B) \mathcal{G}_{l|m|}(\mu_A) \mathcal{G}_{l'|m'}(\mu_B) \varphi_m(\phi_A) \varphi_{m'}(\phi_B) dV. \quad (8-1)$$

Integrali na desnoj strani su vrlo slični određenim klasama integrala koji se često pojavljuju u kvantnomehaničkim problemima. Oni ovise isključivo o razmaku između centara koordinatnih sustava u kojima su opisane pojedine ljuske, R . Budući da se radi o integralima koji ne ovise o individualnim proteinima ili svojstvima ljusaka, oni se mogu izračunati samo jednom i spremiti za sve kasnije upotrebe. Jedan mogući način računanja tih integrala je numerička integracija u pogodnom koordinatnom sustavu. Drugi, koji je upotrijebljen u ovom radu je evaluacija analitičke formule, do koje se može doći uz nešto matematike. Te integrale interpretiramo kao matrice translacije uzduž pozitivne z-osi. Budući da su kod takve translacije ϕ i ϕ' koincidentni, azimutalne funkcije $\varphi_m(\phi)$ se mogu odvojeno integrirati, što je vrlo jednostavno zbog njihove ortonormalnosti, te se jednadžba (8-1) svodi na sumu dvodimenzionalnih integrala u (r, θ) ravnini

$$C(R) = \sum_{nlm} \sum_{n'l'm'} a_{nlm} b_{n'l'm'} T_{nl,n'l'}^{(m)}(R) \delta_{mm'}. \quad (8-2)$$

Oznake unutarnje i vanjske ljuske (ρ, τ) se podrazumijevaju. Svaki $T_{nl, n'l'}^{(m)}(R)$ interpretiramo kao matični element operatora translacije. Dvije izdvojene sume,

$$b_{nlm}^R = \sum_{n'l'} T_{nl, n'l'}^{(m)}(R) b_{n'l'm}, \quad (8-3)$$

odnosno nakon preimenovanja indeksa,

$$a_{nlm}^R = \sum_{n'l'} T_{n'l', nl}^{(m)}(R) a_{n'l'm} \quad (8-4)$$

predstavljaju translaciju tijela B u smjeru pozitivne z-osi odnosno translaciju tijela A u smjeru negativne. Korištenjem Besselove transformacije i činjenice da su funkcije orbitala Gaussovog tipa njezine svojstvene funkcije, mogu se dobiti analitički izrazi za elemente translacijskih matrica (Ritchie, 2005),

$$T_{n'l', nl}^{(m)}(R) = \sum_{k=|l-l'|}^{l+l'} A_k^{l'l|m|} \sum_{j=0}^{n-l+n'-l'-2} C_j^{(nl, n'l')} M! \exp(-R^2/4\lambda) (-R^2/4\lambda)^{k/2} L_M^{(k+1/2)}(R^2/4\lambda), \quad (8-5)$$

gdje je $M = j + (l + l' - k)/2$. Koeficijenti $A_k^{l'l|m|}$ se računaju prema izrazu

$$A_k^{l'l|m|} = (-1)^{(k+l'-l)/2+m} (2k+1) [(2l+1)(2l'+1)]^{1/2} \begin{pmatrix} l & l' & k \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & l' & k \\ m & \bar{m} & 0 \end{pmatrix}. \quad (8-6)$$

Posljednja dva simbola su Wignerovi 3-j simboli, često korišteni u kvantnoj mehanici. Koeficijenti $C_k^{(nl, n'l')}$ dani su s

$$C_k^{(nl, n'l')} = \sum_j \sum_{j'} \delta_{k, j+j'} X_{nlj} X_{n'l'j'}, \quad (8-7)$$

pri čemu je

$$X_{nlj} = \left[\frac{(n-l-1)!(1/2)_n}{2} \right]^{1/2} \frac{(-1)^{n-l-j-1}}{j!(n-l-j-1)!(1/2)_{l+j+1}}. \quad (8-8)$$

9 Implementacija

9.1 Generiranje ljuski

Da bi mogli izračunati koeficijente razvoja u red kuglinih harmonika, prvo moramo izračunati uzorke ljuski proteina. Računanje počinje od površine proteina.

9.1.1 SIMS – *smooth invariant molecular surface*

Za računanje točaka površine molekule proteina i odgovarajućih normala, korišten je SIMS algoritam (engl. *smooth invariant molecular surface*) (Vorobjev, 1997). Radi se o prvom algoritmu koji generira površinu koja zadovoljava sve uvjete postavljene u 6.1.4, što ga čini pogodnim za zahtjevnije primjene. Korištena je FORTRAN implementacija autora, koja je prilagođena za prevođenje GNU Fortran (G77) prevodiocem na Windows platformi. Svi zadaci potrebni da bi se došlo do podataka potrebnih za prijanjanje su skriptirani, tako da se i računanje molekularne površine poziva iz odgovarajuće MATLAB skripte.

9.1.2 Od površine do ljuske

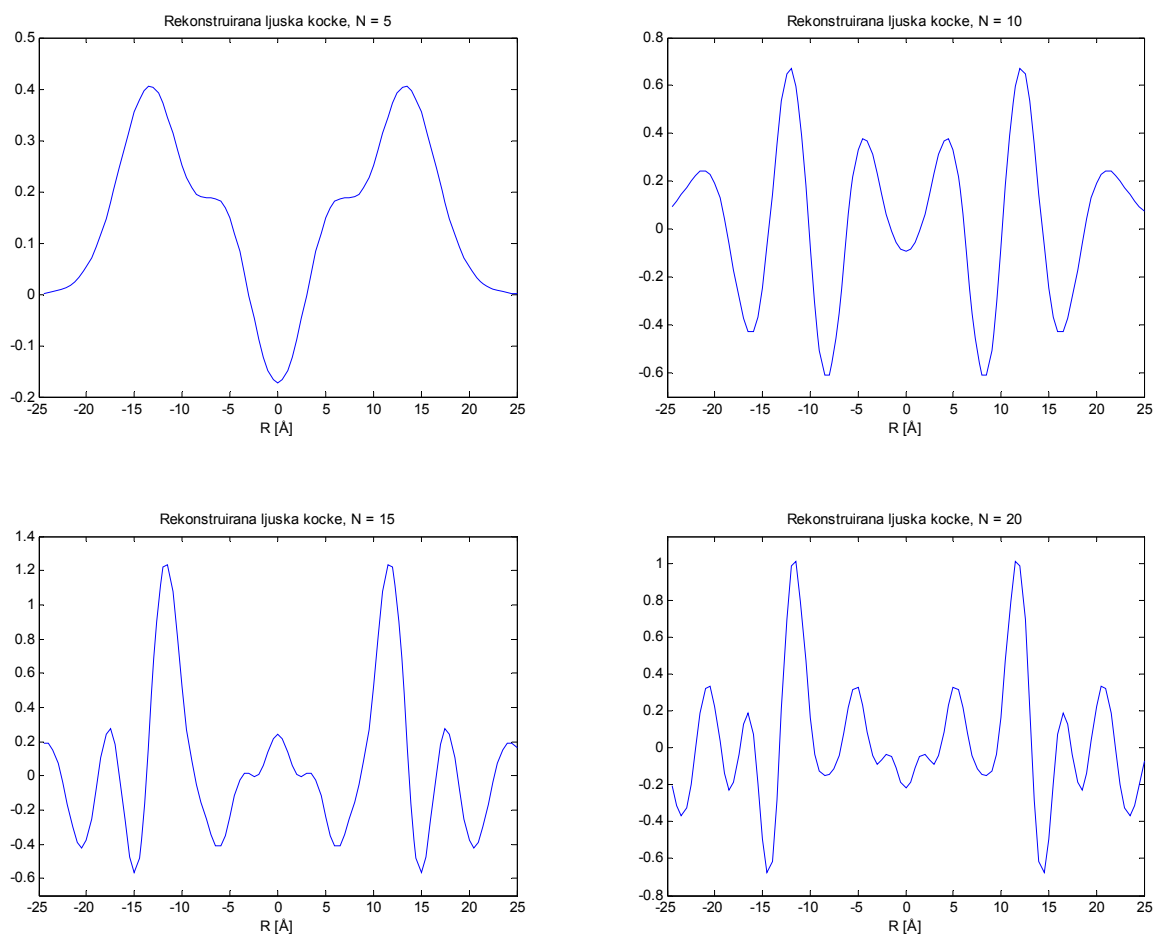
Površina proteina se prvo translacija u ishodište koordinatnog sustava, u smislu da se težište koordinata svih teških atoma (atoma koji nisu vodikovi) u proteinu poklopi s ishodištem. Za rad s PDB datotekama (iz kojih se računa *središte* proteina) korišten je Bioinformatics toolbox (MATLAB 7.4.0). Centriranje proteina je potrebno da bi se podjednako ispunila 3-D rešetka, odnosno da bi se protein nalazio u središtu alocirane mreže. Nakon toga se, da bi procijenili funkciju ljuske $\rho(r)$, na normalu u svaku točki postavi sfera čiji promjer odgovara željenoj debljini ljuske. Naravno, ako generiramo unutrašnju ljusku, uzimamo normalu suprotnog smjera. U središte svake takve kugle postavimo malu, lokalnu rešetku konstante 0.2Å. Svaka ćelija lokalne rešetke čiji centar leži unutar sfere uzima se kao uzorak ljuske koji se projicira na glavnu rešetku konstante 0.5Å. Integrali za izračunavanje koeficijenata razvoja se tada procjenjuju uz pomoć sume

$$a_{nlm}^p \cong \Delta V \sum_c R_{nl}(r_c) \mathcal{G}_{l|m}(\mu_c) \varphi_m(\phi_c); \eta_c = \cos \theta_c, \quad (9-1)$$

pri čemu se sumira preko svih ćelija koje nisu jednake 0. ΔV je volumen ćelije, a (r_c, θ_c, ϕ_c) su sferne polarne koordinate središta c -te ćelije. Budući da bi bilo presporo svaki put računati funkcije baze, korištene su unaprijed izračunate tablice funkcija $R_{nl}(r)$, $\mathcal{G}_{l|m}(\mu)$ i $\varphi_m(\phi)$ s rezolucijom od 0.1Å odnosno 0.1°. Koordinate središta ćelija se zaokružuju na odgovarajuća mjesta u tablicama i zbrajaju se umnošci odgovarajućih elemenata. Napisane su odgovarajuće MATLAB funkcije za računanje Laguerreovih i Legendreovih polinoma, odnosno za računanje radijalnih i kuglinih funkcija.

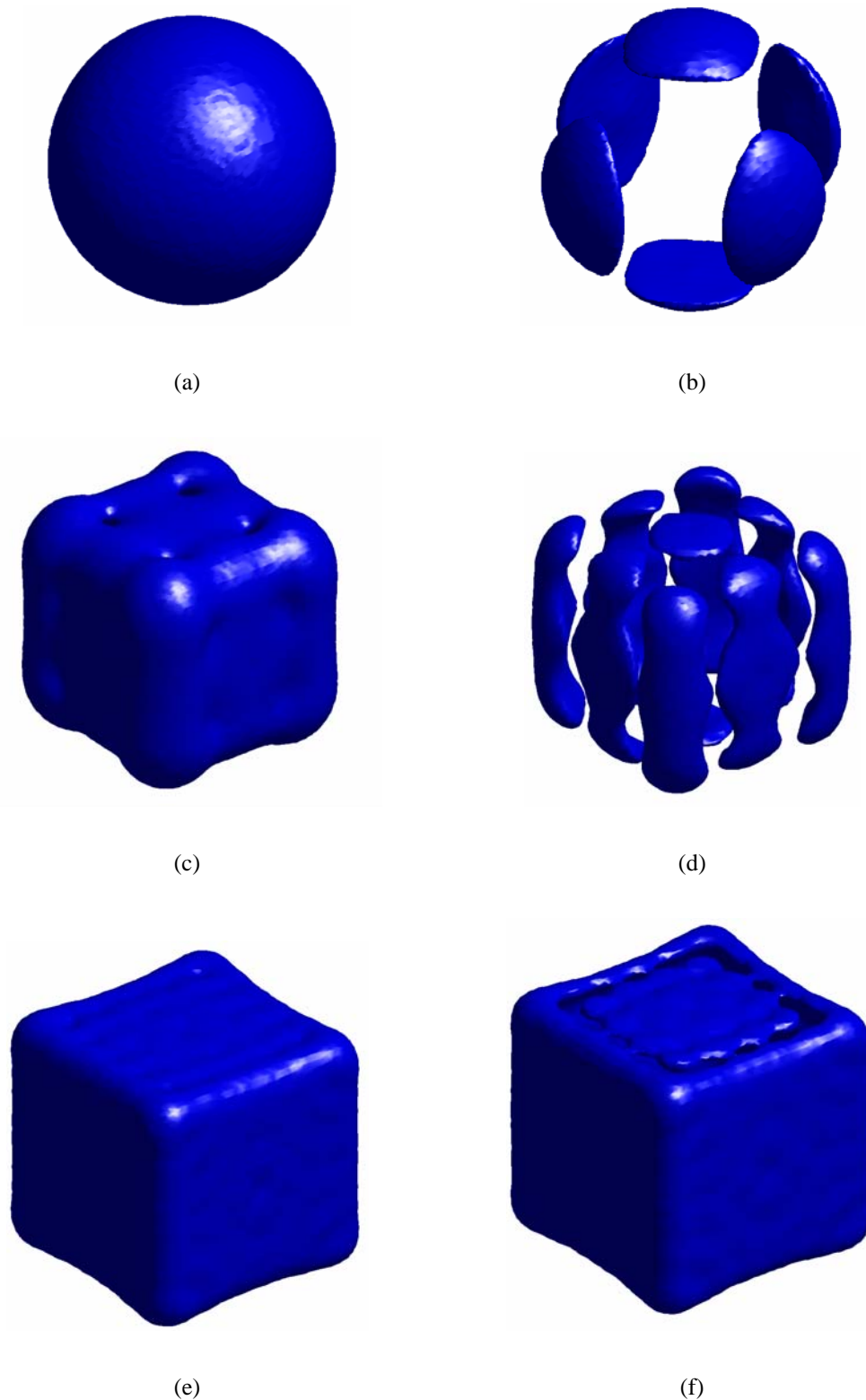
9.1.3 Testiranje ljuski

Za potrebe testiranja, implementirana je funkcija za rekonstrukciju oblika iz koeficijenata razvoja. Važno je primijetiti da je razvojem (7-10) zadana vrijednost za svaku točku u prostoru. Zbog konačnosti reda, vrijednosti praktično svih ćelija rešetke nakon rekonstrukcije biti će različite od 0, pa je vizualizacija mnogo složenija nego kad imamo jasno definiranu plohu. Također za potrebe testiranja, napisana je mala rutina koja generira ljusku kocke, debljine prema potrebi. Kocka je vrlo dobar test za cijeli sustav, budući da su kugline funkcije zapravo vrlo glatke i podsjećaju mnogo više na sferne oblike, nego na kocku. Dobra je analogija pokušaj aproksimacije pravokutnog otvora konačnim Fourierovim redom. Rekonstruirane površine vizualizirane su na dva načina. Prvi je crtanje vrijednosti rekonstruirane funkcije uzduž određenog pravca kojim smo *proboli* rešetku. Drugi je crtanje tzv. izoploha (*isosurface*). Rekonstruirane funkcije su *glatke*, zbog čega ćemo dobiti suvisle grafove ukoliko *provučemo* plohu kroz sve točke u kojima je jednaka vrijednost rekonstruirane funkcije.

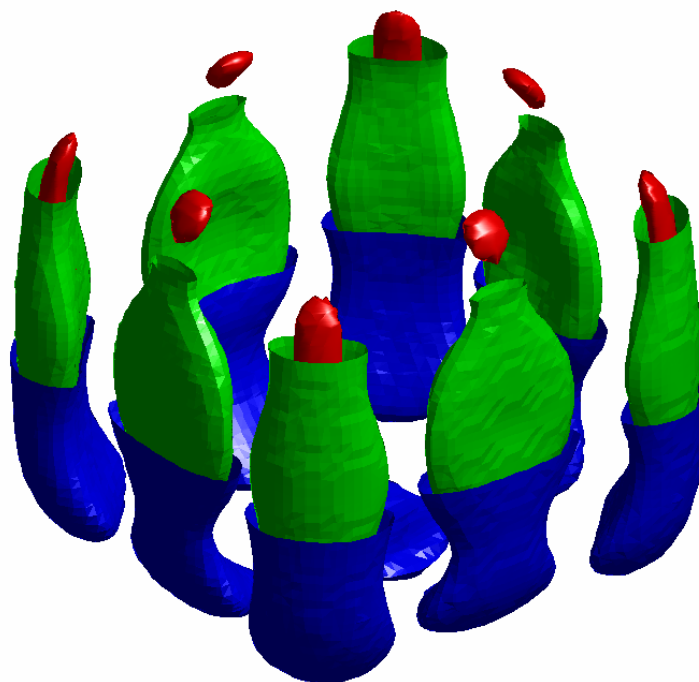


Slika 13. Iznos rekonstruirane ljuske kocke, prikazane uzduž pravca $(r, 0, 0)$, za razne vrijednosti N. Vidljivo je povećanje kvalitete lokalizacije ljuske s povećanjem N-a. Također se vidi da je razmak maksimuma 30 angstrema, što odgovara stranici kocke.

Najveća vrijednost funkcije biti će upravo u središtu rekonstruirane ljuske, a na obje strane će zatim gotovo monotono padati. Zato ćemo za svaki odabran iznos funkcije zapravo dobiti dvije plohe – po jednu sa svake strane središta ljuske.



Slika 14. Rekonstruirana kocka pri različitim N i pri različitim vrijednostima izoploha. Očigledno razvoj i rekonstrukcija funkcioniraju. Redom, (a) $N=5$, $\text{isoval}=0.2$, (b) $N=5$, $\text{isoval}=0.32$, (c) $N=10$, $\text{isoval}=0.25$, (d) $N=10$, $\text{isoval}=0.5$, (e) $N=25$, $\text{isoval}=0.45$, (f) $N=25$, $\text{isoval}=0.7$



Slika 15. Prikaz nekoliko izoploha odjednom. Crvena odgovara vrijednosti 0.7, zelena 0.6 i plava 0.5. Vidljiva je mekoća reprezentacije – najveće vrijednosti (gotovo) monotono padaju prema manjima kako se udaljavamo od stvarne lokalizacije ljuske. To svojstvo je poželjno, budući da donekle uzima u obzir (oponaša) fleksibilnost bočnih lanaca.

9.2 Implementacija rotacijskih matrica

Rotacijske matrice računaju se pomoću pažljive implementacije jednadžbi (5-27) odnosno (5-18). Budući da se radi o vrlo složenim izrazima, bilo je potrebno verificirati točnost implementacije. Vizualizacijom originalnih i rotiranih ploha pokazano je da su rotacijske matrice, kao i sama rotacija, točno implementirani. Drugi način da se ispita valjanost rotacijskih matrica je činjenica da rotacija za puni krug treba dati originalne koeficijente, pa ako komponiramo točan broj rotacija za mali kut oko neke osi, oduzimanjem početnih i završnih koeficijenata možemo provjeriti valjanost implementacije. U sljedećem odsječku prva tri parametra implementirane funkcije `wigner` odgovaraju Eulerovim kutovima, a posljednji parametar je N .

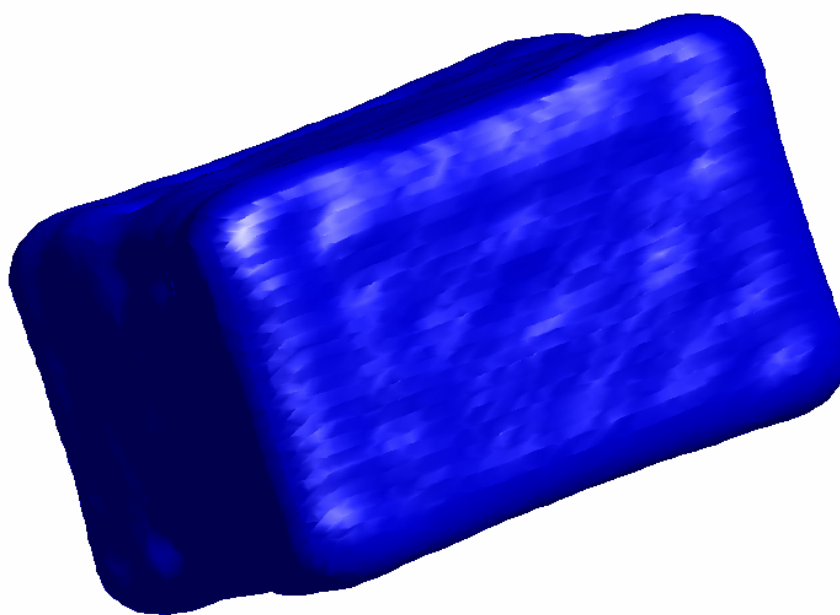
```
coeffs_cube_10_rot = coeffs_cube_10;
R = wigner(10, 0, 0, 10);
for i = 1:36
    coeffs_cube_10_rot = ...
        rotate_coeffs (R, coeffs_cube_10_rot, 10);
end

>> sum(abs(coeffs_cube_10(:) - coeffs_cube_10_rot(:)))

ans =

    3.1302e-013
```

Ovaj test je očigledno prošao uspješno – greška je zanemariva. Vizualizacija izoploha prilikom rotacije prikazana je slikom 16.



Slika 16. Rotirana izoploha kocke prikazana za $N=25$, pri vrijednosti 0.48. Rotirana je za kut $\beta=30^\circ$.

9.3 Izračunavanje translacijskih matrica

Izračun translacijskih matrica bio je računalno najzahtjevniji dio posla – ponajviše se to odnosi na zahtjeve za memorijom. Matrice su računane na pet računala zavodskog clustera. Zbog računanja izraza u kojima se zbrajaju

pribrojnici izuzetno velike dinamike eksponenata, uobičajena dvostruka preciznost (64 bita) nije bila dovoljna. Za potrebe preciznijeg računa, korišten je Multiple Precision toolbox za MATLAB, koji je u biti mex port GMP (GNU multiple precision) biblioteke. Za računanje su korišteni 256 bitni brojevi s pomičnom decimalnom točkom (osmerostruka preciznost). Valjanost izračunatih matrica može se lako provjeriti – te matrice su ortonormalne, stoga je dobra provjera evaluacija izraza za ortonormalnost,

$$\sum_{n=1}^{\infty} \sum_{l'=0}^{n'-1} T_{n'l',nl}^{(m)}(R) T_{n'l',n'l''}^{(m)}(R) = \delta_{mm} \delta_{ll''}. \quad (9-2)$$

Računanjem jednadžbe (9-2) potvrđena je ispravnost implementacije.

9.4 Koreliranje hidrofobnosti površina

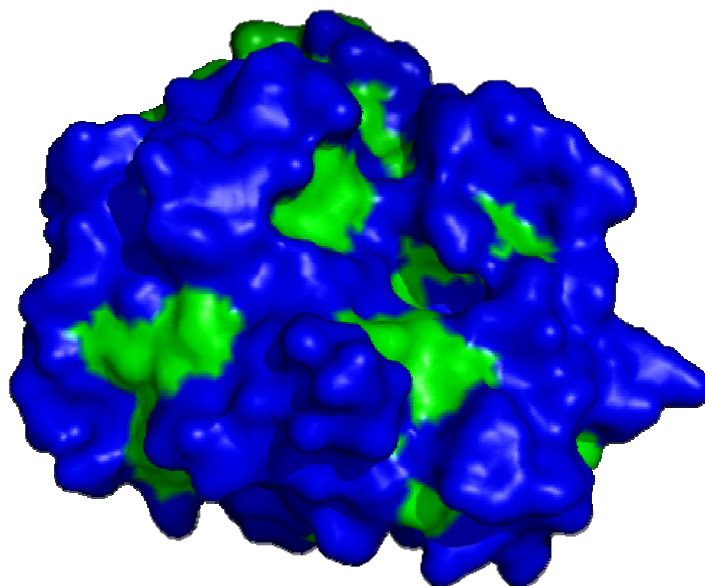
Korelacija između oblika površina je dobar način da se dođe do određenog broja orijentacija proteina partnera, od kojih su sve kandidati za točno rješenje. Iako u nekim lakšim slučajevima takva korelacija može točno izdvojiti pravi kompleks, najčešće je potrebno dobivene orijentacije ocijeniti i po nekom drugom kriteriju. Koji god kriterij odaberemo, poželjno je da ne bude računalno suviše zahtjevan, a bilo bi idealno kada bi mogli koristiti već razvijen skup alata za njegovu evaluaciju. Takvi kriteriji se uglavnom svode na maksimiziranje broja slabih interakcija između proteina partnera. Mi smo odlučili korelirati svojstvo hidrofobnosti. Kao što je obrazloženo u poglavlju 2, hidrofobne plohe će *nastojati* biti u kontaktu s hidrofobnim plohamama, budući da su takvi kontakti energetski povoljni – istiskuje se polarna voda. Nabijene ili polarne plohe radije vežu molekule vode nego što dolaze u kontakt s hidrofobnom plohom s kojom ne ostvaruju nikakvu vezu. Stoga je energetski nepovoljno da su one u kontaktu s hidrofobnom plohom.

Da bi takva korelacija bila uspješno implementirana, potrebno je odabrati način pridruživanja hidrofobnosti površini proteina. Ovdje je odlučeno da će se hidrofobnost opisati funkcijom gustoće, na jednak način na koji su opisane i ljuske. Postoji 7 izrazito hidrofobnih aminokiselina: valin, izoleucin, leucin, metionin, fenilalanin, triptofan i cistein. Za sve točke površine, koje pripadaju

atomima neke od hidrofobnih aminokiselina, *funkcija gustoće svojstva hidrofobnosti* $\chi(\underline{r})$ poprima vrijednost 1, a inače je 0:

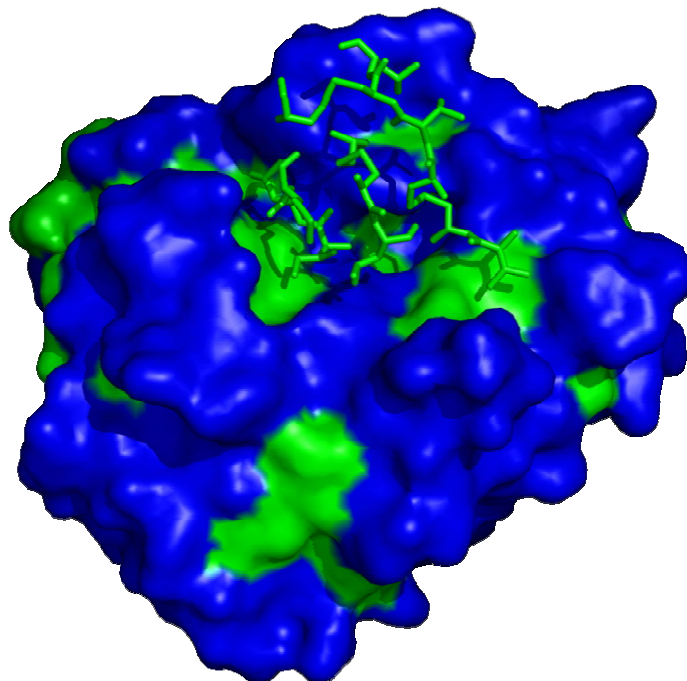
$$\chi(\underline{r}) = \begin{cases} 1, & \underline{r} \in S_{HF} \\ 0, & \text{ostalo} \end{cases},$$

pri čemu je sa S_{HF} označen skup točaka ljuske koje pripadaju nekoj hidrofobnoj aminokiselini. Ovdje je zbog jednostavnosti i konceptualne prirode eksperimenta korištena samo jedna ljuska, koja je istovremeno i unutarnja i vanjska. Budući da se radi o svojstvu pridruženom površini, funkcija $\chi(\underline{r})$ se razvija i korelira na identičan način kao i funkcije ljuske kod korelacije oblika površina. Na slici 17 prikazana je površina proteina receptora s istaknutim hidrofobnim regijama. Iako se radi o malenom proteinu – hidrofobne kontaktne plohe su često velike – može se vidjeti da postoji potencijal za hidrofobno vezanje. Na slici 18 je prikazan isti protein, ali je prikazan i ligand, i to pomoću modela sa štapovima kako bi se lakše uočila mjesta kontakta. Prikazane su samo hidrofobne aminokiseline liganda.



Slika 17. Površina proteina receptora izdvojenog iz kompleksa s PDBid 1ppe. Zeleno su prikazane hidrofobne regije – regije koje pripadaju hidrofobnim aminokiselinama.

Jasno je vidljivo kako se nasuprot hidrofobnim dijelovima receptora nalaze hidrofobni dijelovi liganda – hidrofobne plohe dolaze jedna nasuprot drugoj, sve ostalo bilo bi energetski izrazito nepovoljno.



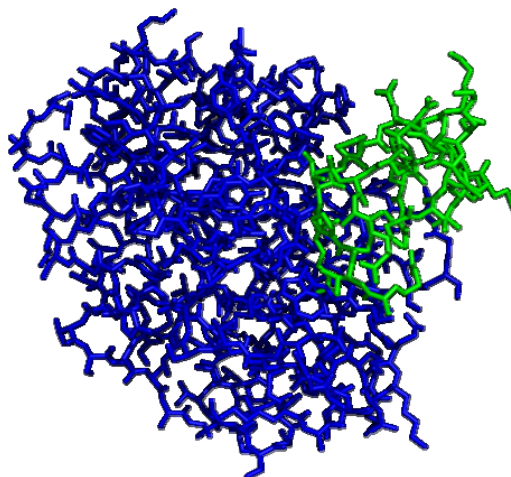
Slika 18. Površina receptora kao i na slici 17, s hidrofobnim aminokiselinama liganda prikazanim modelom štapova. Vidljiva je ispravnost pretpostavki baziranih na energetskim razmatranjima.

10 Testiranje i rezultati

Implementirani sustav je testiran na 3 proteinska kompleksa. Budući da je zadatak bio opravdati koncept prijanjanja proteina, testiranjem se je pokušalo pokazati da algoritam visoko rangira točan kompleks dva proteina u rekonstrukcijskom smislu – nije testirano nevezano prijanjanje. Također, nije korišten velik broj orijentacija kakav se koristi kod *slijepog* pretraživanja, nego je koreliran manji broj grubo raspoređenih orijentacija, pri čemu je bilo osigurano da se *točna* nalazi među njima. Razlog za takav izbor je prvenstveno sporost MATLAB implementacije – vektorizirani su svi dijelovi kôda kod kojih je to bilo moguće, no samo prijanjanje je za ozbiljnije testiranje potrebno implementirati u nekom bržem programskom jeziku. Naravno, dobar rezultat bio bi da točna orijentacija zauzme visoku poziciju. Za korištenje algoritma za prijanjanje u sfernim polarnim koordinatama, potrebno je 3 translatorna i 3 rotacijska stupnja slobode transformirati u pet rotacijskih i jedan translatorni (razmak između referentnih koordinatnih sustava). Za potrebe testiranja, pretraživana su samo 2 rotacijska i jedan translatorni, ali je bilo osigurano da će se *točan* kompleks naći među njima. Naime, proteini su prije prijanjanja rotirani i translirani tako da se spojnica njihovih centara mase poklopi s z-osi. Tako smo sigurni da ćemo uključiti i točnu orijentaciju, čak i ako variramo samo razmak između proteina (operator translacije funkcionira upravo duž z-osi).

Kod svih testiranih kompleksa, kao minimalna udaljenost proteina odabrana je nekoliko angstrema manja udaljenost od poznate stvarne udaljenosti, kako bi se izbjegle moguće greške kod testiranja fizikalno nemogućih orijentacija (ligand unutar receptora).

10.1 Kompleks 1ppe

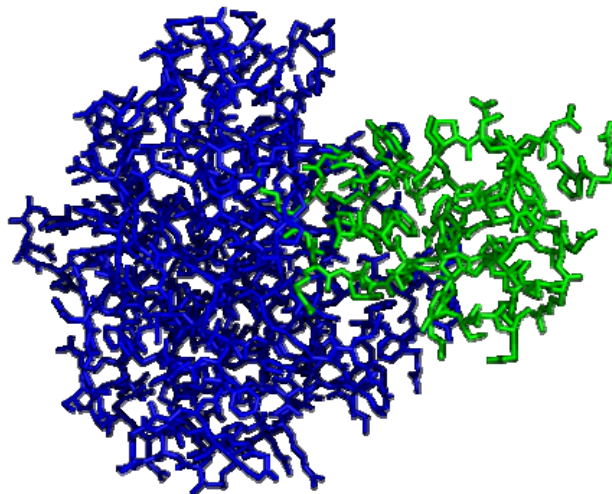


Slika 19. Kompleks PDBid 1ppe. Receptor je prikazan plavo, a ligand zeleno.

Tablica 1. Ocjene pokusnih orijentacija za kompleks 1ppe.

| rang | $\beta, ^\circ$ | r | $\alpha, ^\circ$ | $\beta, ^\circ, hf$ | R, hf | $\alpha, ^\circ, hf$ |
|------|-----------------|-----|------------------|---------------------|---------|----------------------|
| 1. | -120 | 16 | -150 | 60 | 17 | -30 |
| 2. | -120 | 16 | -180 | 60 | 17 | 0 |
| 3. | -120 | 16 | -120 | 60 | 17 | -60 |
| 4. | -120 | 16 | 150 | 60 | 17 | -90 |
| 5. | -120 | 16 | -90 | 60 | 17 | 30 |
| 6. | 0 | 18 | -90 | 30 | 17 | -30 |
| 7. | 0 | 18 | 90 | 30 | 17 | -60 |
| 8. | 0 | 18 | 60 | 30 | 17 | 0 |
| 9. | 0 | 18 | -60 | 60 | 17 | 60 |
| 10. | -30 | 18 | -60 | 60 | 18 | 90 |
| 11. | -30 | 18 | -90 | 60 | 18 | 60 |
| 12. | 0 | 18 | 120 | 60 | 18 | 120 |
| 13. | 0 | 18 | -120 | 60 | 17 | -120 |
| 14. | -30 | 18 | -30 | 60 | 18 | 150 |
| 15. | 0 | 18 | 30 | 60 | 18 | 30 |
| 16. | 0 | 18 | -30 | 30 | 17 | -90 |
| 17. | -30 | 18 | -120 | 30 | 17 | 30 |
| 18. | -30 | 18 | -180 | 60 | 18 | 0 |
| 19. | -30 | 18 | -150 | 60 | 18 | -180 |
| 20. | -120 | 16 | -60 | 60 | 18 | -90 |
| 21. | -30 | 18 | 150 | 60 | 18 | -150 |
| 22. | -120 | 16 | 120 | 60 | 18 | -30 |
| 23. | 0 | 18 | 0 | 60 | 18 | -60 |
| 24. | 0 | 18 | -150 | 60 | 18 | -120 |
| 25. | 0 | 18 | 150 | 60 | 17 | 90 |
| 26. | -30 | 18 | 0 | 30 | 17 | -120 |
| 27. | -30 | 18 | 120 | 30 | 17 | 60 |
| 28. | -30 | 18 | 30 | 60 | 17 | -150 |
| 29. | -30 | 18 | 90 | 150 | 17 | -30 |
| 30. | 0 | 18 | -180 | 150 | 17 | 0 |

10.2 Kompleks 1eaw

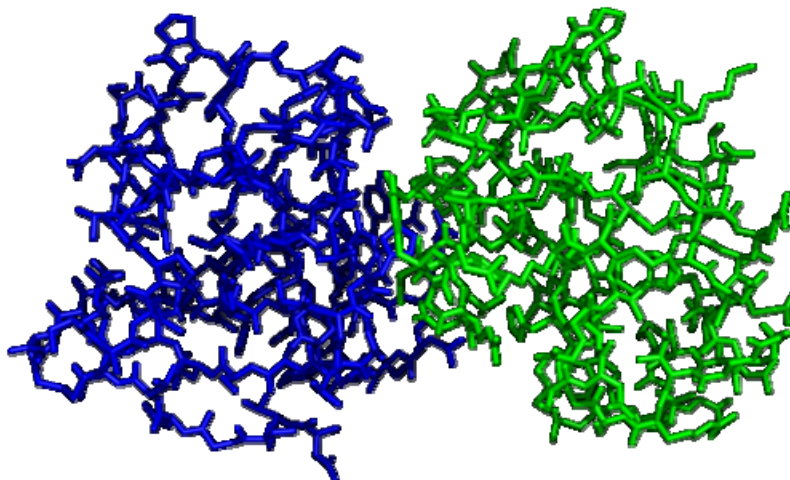


Slika 20. Kompleks PDBid 1eaw. Receptor je prikazan plavo, a ligand zeleno.

Tablica 2. Ocjene pokusnih orijentacija za kompleks 1eaw.

| rang | $\beta, ^\circ$ | r | $\alpha, ^\circ$ | $\beta, ^\circ, hf$ | R, hf | $\alpha, ^\circ, hf$ |
|------|-----------------|-----|------------------|---------------------|---------|----------------------|
| 1. | 0 | 23 | 150 | -30 | 23 | 0 |
| 2. | 0 | 23 | -150 | -30 | 23 | 30 |
| 3. | 0 | 23 | -180 | 150 | 23 | -30 |
| 4. | 0 | 23 | 120 | -30 | 23 | -30 |
| 5. | 0 | 23 | -120 | 150 | 23 | 0 |
| 6. | 0 | 23 | 90 | -30 | 23 | 90 |
| 7. | 0 | 23 | -90 | -30 | 23 | -90 |
| 8. | 0 | 23 | 60 | 150 | 23 | -60 |
| 9. | 0 | 23 | -60 | -30 | 23 | 60 |
| 10. | -30 | 23 | 120 | -30 | 23 | -120 |
| 11. | -30 | 23 | 0 | -30 | 23 | 120 |
| 12. | -30 | 23 | 90 | 150 | 23 | 30 |
| 13. | -30 | 23 | -120 | 150 | 23 | 90 |
| 14. | -30 | 23 | 30 | 150 | 23 | 60 |
| 15. | -30 | 23 | 60 | -30 | 23 | -60 |
| 16. | -30 | 23 | -150 | 150 | 23 | 120 |
| 17. | -30 | 23 | -30 | 150 | 23 | 150 |
| 18. | -30 | 23 | 150 | 150 | 23 | -90 |
| 19. | 0 | 23 | -30 | 150 | 23 | -180 |
| 20. | 0 | 23 | 30 | -30 | 23 | -150 |
| 21. | 0 | 23 | 0 | 150 | 23 | -120 |
| 22. | -30 | 23 | -90 | 150 | 23 | -150 |
| 23. | -30 | 23 | -180 | -30 | 23 | 150 |
| 24. | -30 | 23 | -60 | -30 | 23 | -180 |
| 25. | 30 | 23 | -180 | -60 | 23 | -180 |
| 26. | 30 | 23 | 150 | 120 | 23 | 0 |
| 27. | 30 | 23 | 120 | -60 | 23 | 150 |
| 28. | 30 | 23 | 0 | -60 | 23 | -60 |
| 29. | 30 | 23 | 90 | -60 | 23 | -150 |
| 30. | -30 | 24 | -120 | -60 | 23 | -90 |

10.3 Kompleks 1ay7



Slika 21. Kompleks PDBid 1ay7. Receptor je prikazan plavo, a ligand zeleno.

Tablica 3. Ocjene pokusnih orijentacija za kompleks 1ay7.

| rang | $\beta, ^\circ$ | r | $\alpha, ^\circ$ | $\beta, ^\circ, hf$ | R, hf | $\alpha, ^\circ, hf$ |
|------|-----------------|-----|------------------|---------------------|---------|----------------------|
| 1. | 0 | 23 | 150 | -30 | 23 | 0 |
| 2. | 0 | 23 | -150 | -30 | 23 | 30 |
| 3. | 0 | 23 | -180 | 150 | 23 | -30 |
| 4. | 0 | 23 | 120 | -30 | 23 | -30 |
| 5. | 0 | 23 | -120 | 150 | 23 | 0 |
| 6. | 0 | 23 | 90 | -30 | 23 | 90 |
| 7. | 0 | 23 | -90 | -30 | 23 | -90 |
| 8. | 0 | 23 | 60 | 150 | 23 | -60 |
| 9. | 0 | 23 | -60 | -30 | 23 | 60 |
| 10. | -30 | 23 | 120 | -30 | 23 | -120 |
| 11. | -30 | 23 | 0 | -30 | 23 | 120 |
| 12. | -30 | 23 | 90 | 150 | 23 | 30 |
| 13. | -30 | 23 | -120 | 150 | 23 | 90 |
| 14. | -30 | 23 | 30 | 150 | 23 | 60 |
| 15. | -30 | 23 | 60 | -30 | 23 | -60 |
| 16. | -30 | 23 | -150 | 150 | 23 | 120 |
| 17. | -30 | 23 | -30 | 150 | 23 | 150 |
| 18. | -30 | 23 | 150 | 150 | 23 | -90 |
| 19. | 0 | 23 | -30 | 150 | 23 | -180 |
| 20. | 0 | 23 | 30 | -30 | 23 | -150 |
| 21. | 0 | 23 | 0 | 150 | 23 | -120 |
| 22. | -30 | 23 | -90 | 150 | 23 | -150 |
| 23. | -30 | 23 | -180 | -30 | 23 | 150 |
| 24. | -30 | 23 | -60 | -30 | 23 | -180 |
| 25. | 30 | 23 | -180 | -60 | 23 | -180 |
| 26. | 30 | 23 | 150 | 120 | 23 | 0 |
| 27. | 30 | 23 | 120 | -60 | 23 | 150 |
| 28. | 30 | 23 | 0 | -60 | 23 | -60 |
| 29. | 30 | 23 | 90 | -60 | 23 | -150 |
| 30. | -30 | 24 | -120 | -60 | 23 | -90 |

11 Razmatranje rezultata

Za sva tri kompleksa testiran je jednak skup orijentacija. Korelirane su ljuske i hidrofobnost (stupci s oznakom *hf* u tablicama). Proteini su prije testiranja rotirani i translirani tako da se točna orijentacija nalazi na kutovima $\beta = 0$, $\alpha = 0$ i udaljenostima 18 Å za kompleks 1ppe, 24 Å za kompleks 1eaw i 23 Å za kompleks 1ay7. Kutna rezolucija pretraživanja bila je 30°, a radijalna 1 Å. Vidljivo je da su točni rezultati, ili rezultati vrlo blizu točnima vrlo visoko rangirani, što dokazuje ispravnost implementacije brojnih dijelova sustava i valjanost cjelokupnog koncepta. Važno je primijetiti da je ovdje za svaki kompleks prikazano samo prvih 30 od nekoliko tisuća ispitanih orijentacija.

11.1 Uočeni problemi

Prilikom testiranja primijećeno je da vrlo visoke ocjene dobivaju fizikalno besmislene orijentacije u kojima jedan protein prodire u drugi ili se čak u potpunosti nalazi unutar njega. Iako je takve orijentacije lako eliminirati *a priori* poznavanjem geometrije proteina partnera, takvo ponašanje je pogrešno. Dva su razloga vrlo vjerojatno odgovorni za to, te će ih biti potrebno provjeriti prije daljnjeg razvoja algoritma: 1. točkasta površina proteina sadrži točke iz unutrašnjosti proteina, i one doprinose koeficijentima razvoja, 2. translacijske matrice nisu potpuno točno izračunate, budući da je ortonormalnost provjeravana samo za manji broj indeksa – male greške se kumulativno zbrajaju i kvare rezultat.

11.2 Daljnji razvoj

Nakon rješavanja uočenih problema, potrebno je cijeli sustav ponovo napisati u nekom programskom jeziku čiji prevodilac generira brzi strojni kôd (vjerojatno C/C++) kako bi se do rezultata moglo dolaziti mnogo brže. Važno poboljšanje bilo bi uvođenje novih svojstava za korelaciju. Svojstvo koje će sljedeće biti ispitano je raspored naboja na površini proteina i mogućnost *ad hoc* korelacije gustoća pozitivnih odnosno negativnih naboja. Ako se njihove razdiobe uzmu kao dobro lokalizirane, takva korelacija bi mogla biti dobar filter za ionske interakcije.

12 Zaključak

U okviru ovog diplomskog rada izrađene su računalne implementacije kuglinih funkcija i posebnih radijalnih funkcija. Implementiran je modul za razvoj proizvoljnih 3-D svojstava u redove funkcija baze koje su formirane od kuglinih i radijalnih, kao i modul za rekonstrukciju originalnih funkcija iz koeficijenata. Napisane su funkcije za računanje Wignerovih rotacijskih matrica i matrica translacijskog operatora. Pokazano je da svaki od tih modula funkcionira prema očekivanjima. Napisane su funkcije za računanje dvostruke ljuske proteina, kao i za računanje gustoće svojstva hidrofobnosti. Razvoji ljuski proteina i gustoće hidrofobnosti korišteni su za testiranje sheme prijanjanja proteina u sfernim polarnim koordinatama. Iako su uočeni određeni problemi, dobiveni rezultati su vrlo dobri. Dokazana je valjanost koncepta, te predstoji ispravljanje uočenih pogrešaka i nastavak rada na sustavu.

Ivan Dokmanić

13 Literatura

E.M. KROVAT, T. STEINDL, T. LANGER, Recent Advances in Docking and Scoring, *Current Computer-Aided Drug Design*, 1 (2005), pp 93-102.

V. MOHAN, A.C. GIBBS, M.D. CUMMINGS, E.P. JAEGER, R.L. DESJARLAIS, Docking: Successes and Challenges, *Current Pharmaceutical Design*, 2005, 11 (2005), pp 323-333.

E. PEROLA, W.P. WALTERS, P.S. CHARIFSON, A Detailed Comparison of Current Docking and Scoring Methods on Systems of Pharmaceutical Relevance, *PROTEINS: Structure, Function and Bioinformatics* (2004), 56, pp 235-249.

G.L. WARREN et al., A Critical Assessment of Docking Programs and Scoring Functions, *Journal of Medical Chemistry*, (2006), 49, pp 5912-5931.

LEHRINGER, NELSON & COX, Principles of Biochemistry, W. H. Freeman, 4 edition, 2004.

B. PIERCE, A.T. PHILLIPS, Z. WENG, Structure Prediction of Protein Complexes, *poglavlje iz Y. XU, D. XU, J. LIANG, Computational Methods for Protein Structure Prediction and Modeling, Volume 2: Structure Prediction*, Springer, New York, 2007, pp 109-134.

R. ABAGYAN, M. TOTROV, D. KUZNETSOV, ICM – A New Method for Protein Modeling and Design: Applications to Docking and Structure Prediction from the Distorted Native Conformation, *Journal of Computational Chemistry* 5 (1994), pp 488-506.

D.W. RITCHIE, G.J.L. KEMP, Protein Docking Using Spherical Polar Fourier Correlations, *PROTEINS: Structure, Function and Genetics*, preprint (1999).

D.W. RITCHIE, Molecular Shape Recognition Using 3-D Polar Fourier Correlations, (2002), *prezentacija s predavanja*.

D.W. RITCHIE, High-order analytic translation matrix elements for real-space six-dimensional polar Fourier correlations, *Journal of Applied Crystallography*, (2005), 38, pp 808-818.

D.W. RITCHIE, Parametric Protein Shape Recognition, *Doktorski rad*, University of Aberdeen, Departments of Computing Science & Molecular and Cell Biology, Aberdeen, 1998.

B. SANDAK, H.J. WOLFSON, R. NUSSINOV, Flexible Docking Allowing Induced Fit in Proteins: Insights From an Open to Closed Conformational Isomers, *PROTEINS: Structure, Function and Genetics*, (1998), 32, pp 159-174.
