

SVEUČILIŠTE U ZAGREBU  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 1521

# **Alat za prijanjanje proteina: modul za pripremu prijanjanja**

Igor Čanadi

Zagreb, lipanj 2010.

*Umjesto ove stranice umetnite izvornik Vašeg rada.*

*Da bi ste uklonili ovu stranicu obrišite naredbu `\izvornik`.*

*Hvala Nini Antulov-Fantulinu, Matiji Piškorecu i Ivanu Soviću na dobroj suradnji, odlično obavljenom poslu i nezaboravnim team building-ima.*

*Hvala Mili Šikiću na ogromnoj pomoći, savjetima, pismima preporuke i ćevapima u Sofri.*

*Hvala Krešimiru Lugariću što mi je preporučio Mileta za mentora.*

*Hvala Antonu Grbinu na pomoći u implementaciji translacijskih matrica i otkrivanju čari Wolfram Mathematice.*

# SADRŽAJ

<b>1. Uvod</b>	<b>1</b>
<b>2. Prianjanje proteina - metoda <i>Spherical Polar Fourier</i></b>	<b>2</b>
2.1. Parametrizacija površine proteina . . . . .	3
2.1.1. Model s dvije ljuske . . . . .	6
2.2. Algoritam prianjanja proteina . . . . .	7
2.2.1. Diskretizacija prostora stanja . . . . .	7
2.2.2. Evaluacijska funkcija . . . . .	8
2.2.3. Traženje strukture kompleksa . . . . .	9
<b>3. Steric Scan</b>	<b>10</b>
3.1. Ideja . . . . .	10
3.2. Programska implementacija . . . . .	11
3.3. Istraživanja i rezultati . . . . .	12
3.3.1. Istraživanja na redu 5 . . . . .	14
3.3.2. Istraživanja na redu 10 . . . . .	16
3.3.3. Istraživanja na redu 15 . . . . .	18
3.4. Postignuta ubrzanja . . . . .	19
<b>4. Zaključak i diskusija</b>	<b>21</b>
<b>Literatura</b>	<b>22</b>

# 1. Uvod

Na Fakultetu elektrotehnike i računarstva razvijen je alat za prijanjanje proteina *Protein Docking Tool* (Antulov-Fantulin, 2010). On za prijanjanje koristi metodu *Spherical Polar Fourier*, koju je moguće optimizirati metodom *Steric Scan*.

U ovom radu implementirat ću novi modul na Protein Docking Tool-u koji će pokušati ubrzati proces prijanjanja metodom Steric Scan, analizirati postignuta ubrzanja i prikazati rezultate.

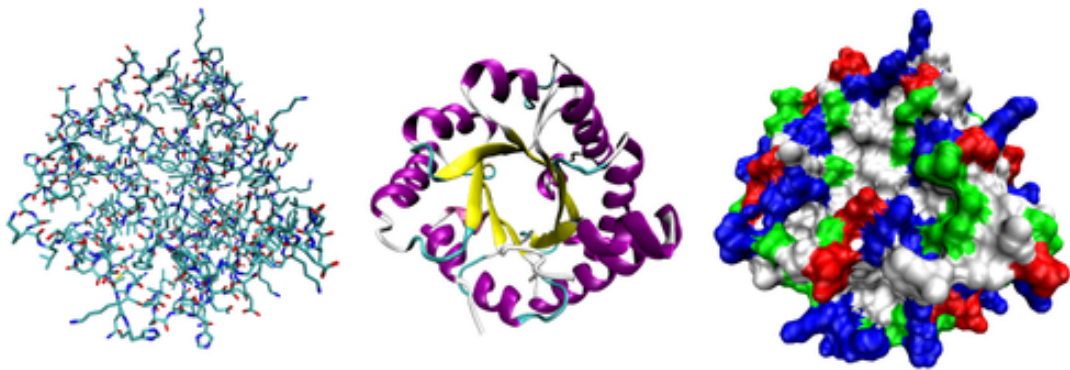
U drugom poglavlju opisat ću metodu Spherical Polar Fourier, neke karakteristike te metode i dati pseudokod glavnog dijela algoritma.

U trećem poglavlju opisat ću ideju Steric Scan-a i programsku implementaciju koja se uklopila u Protein Docking Tool. Također, istražiti ću ponašanje metode s obzirom na različite parametre te prikazati maksimalna ubrzanja koja su postignuta.

U četvrtom poglavlju dat ću zaključak i nove smjerove istraživanja s ciljem postizanja što veće efikasnosti metode Steric Scan.

## 2. Prianjanje proteina - metoda *Spherical Polar Fourier*

Proteini su esencijalni dijelovi živih organizama i sudjeluju u gotovo svakom procesu koji se odvija u stanici. Mogu se pojavljivati u različitim ulogama, u kojima im funkciju određuje njihova tercijalna struktura. Također, proteini se često međusobno povezuju te tvore proteinske komplekse, trajno ili privremeno. Kao i kod samih proteina, struktura proteinskog kompleksa određuje njegovu funkciju. Zbog toga bi bilo korisno predvidjeti strukturu proteinskog kompleksa koji bi mogao nastati vezanjem dvaju proteina. Poznavanjem te informacije, osim predviđanja efekata interakcije poznatih proteina, moguće bi bilo i stvarati umjetne proteine ili kemijske tvari te predviđati njihovo ponašanje u suradnji s drugim proteinima.



**Slika 2.1:** Neki od načina prikaza proteina. Slika preuzeta s <http://en.wikipedia.org>

Zbog toga je definiran problem prianjanja proteina, odnosno predviđanja strukture proteinskih kompleksa. Do danas razvijeno je mnogo metoda koje se međusobno natječu da budu što brže i točnije te se zbog toga svakodnevno poboljšavaju.

Problem predviđanja strukture proteinskog kompleksa uključuje ispitivanje kvalitete svih mogućih konformacija u kojima se dva proteina mogu naći. U trenutnoj implementaciji alata za prianjanje proteina ispituje se samo geometrijska komplementarnost površina, no ispitivanje je moguće provesti i korelacijom nekih drugih svojstava - elektrostatičkom ili hidrofobnošću određenih dijelova površine.

Proteini se u metodi *Spherical Polar Fourier* parametriziraju pomoću reda ortonormiranih funkcija u polarnom koordinatnom sustavu čime se postiže ubrzanje transformacija rotacije. Zbog toga se relativne pozicije dvaju proteina opisuju s pet rotacijskih i jednim translacijskim stupnjem slobode. Prva uspješna implementacija te metode opisana je u Ritchie (1998).

## 2.1. Parametrizacija površine proteina

Parametrizacija površine proteina provodi se pomoću kuglinih i radijalnih funkcija na način na koji je to opisano u Ritchie (1998) i Dokmanić (2007).

### Kugline funkcije

Kugline funkcije (engl. *spherical harmonics*)  $y_{lm}(\phi, \theta)$  zadane su u sfernim koordinatama  $(\phi, \theta)$  uz uvjet  $|m| < l$ . Dva su važna razloga zašto su one iznimno pogodne za prikaz površine proteina.

Prvo, svaka funkcija na kugli može se izraziti kao linearna kombinacija kuglinih funkcija:

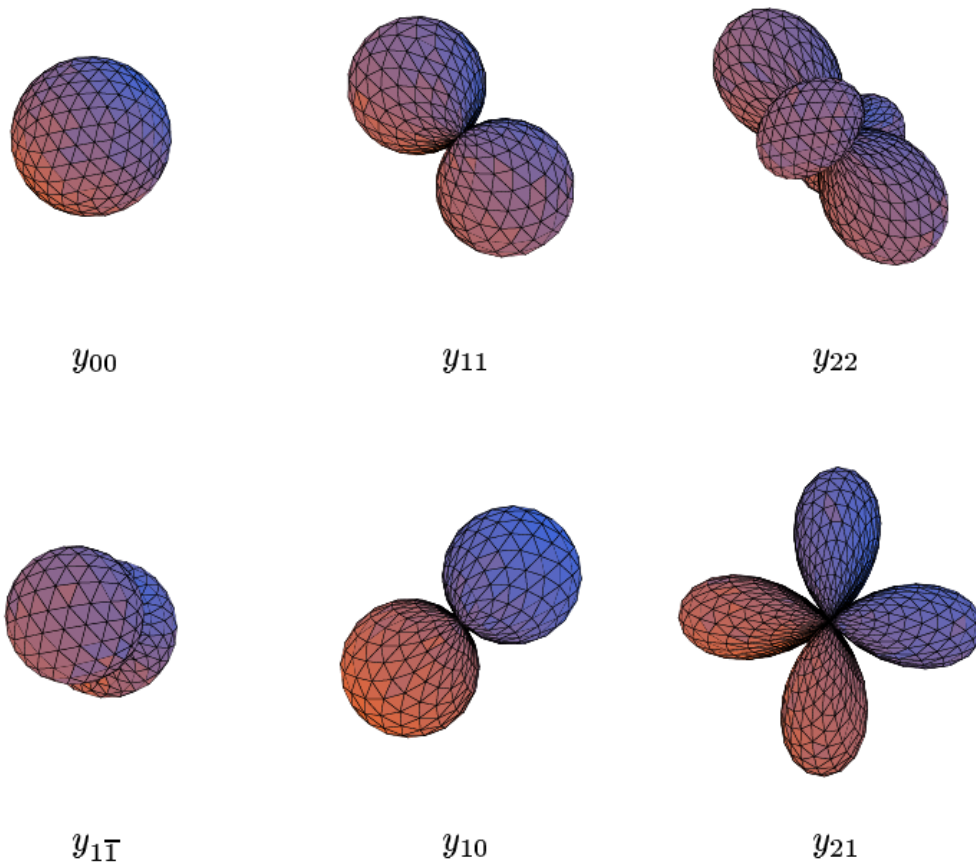
$$\mu(\phi, \theta) = \sum_{l=0}^{\infty} \sum_{m=-l}^{m=l} a_{lm} y_{lm}(\phi, \theta) \quad (2.1)$$

Sada površinu proteina možemo reprezentirati nizom koeficijenata  $(a_{0,0}, a_{1,-1}, a_{1,0}, \dots)$ . Naravno, niz koeficijenata ne može biti beskonačan, nego s određenim brojem koeficijenata postignemo bolju ili lošiju aproksimaciju površine.

Drugo važno svojstvo je što je prostor kuglinih funkcija zatvoren s obzirom na rotacije i da se kugline funkcije određenog reda  $l$  transformiraju između sebe na predvidljiv način, što nam omogućava vrlo brzu implementaciju rotiranja proteina.

### Trodimenzionalna parametrizacija

Budući da je protein trodimenzionalna struktura, nije dovoljno reprezentirati ga samo funkcijom na kugli. U tu svrhu, kugline funkcije množimo radijalnim funkcijama te proteinu dodajemo debljinu. Pogodne radijalne funkcije definirane su izrazom:



**Slika 2.2:** Prikaz nekih kuglinih funkcija. Ovdje su zbog bolje vizualizacije prikazane kao površina eksplicitno zadana u sfernim polarnim koordinatama jednadžbom  $r = y_{lm}(\phi, \theta)$ . Slika preuzeta iz Ritchie (1998).

$$R_{nl}(r) = \left[ \frac{2}{\lambda^{(3/2)} * \pi^{(1/2)}} \frac{(n-l-1)!}{(1/2)_n} \right]^{(1/2)} \exp(-\rho^2/2) \rho^l L_{n-l-1}^{(l+1/2)}(\rho^2), \quad (2.2)$$

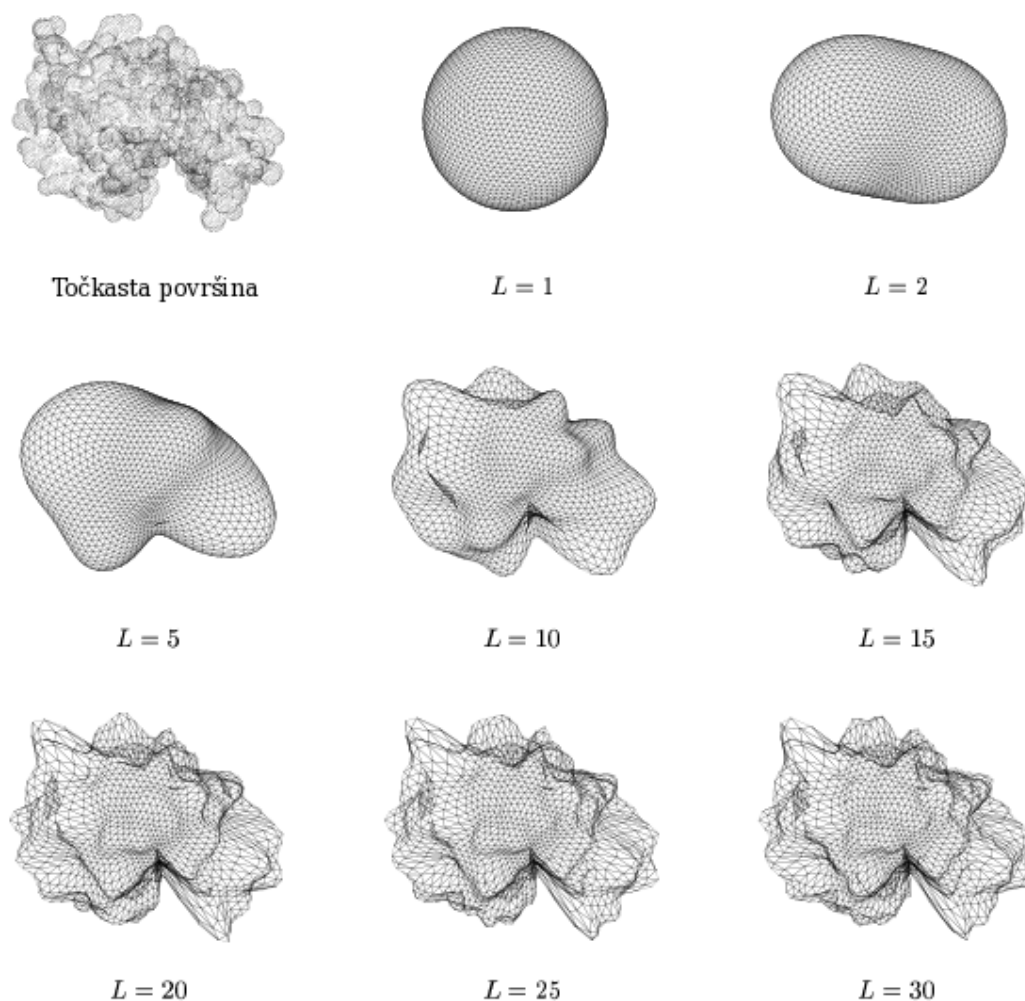
gdje je  $\rho^2 = r^2/\lambda$  s faktorom skaliranja  $\lambda$  koji se obično postavlja na oko 20.  $(x)_n$  je oznaka za umnožak  $(x)_n = x(x+1) \dots (x+n-1)$  te su  $L_{n-l-1}^{(l+1/2)}(\rho^2)$  generalizirani Laguerrovi polinomi.

Radijalne funkcije  $R_{nl}(r)$  će se sada koristiti u paru s kuglinim funkcijama  $y_{lm}(\phi, \theta)$  kako bi se opisao volumen ljuske proteina. Definiramo nove funkcije baze:

$$F_{nlm}(\underline{r}) = R_{nl}(r)y_{lm}(\phi, \theta) \quad (2.3)$$

koje su trodimenzionalne. Površinu proteina  $\sigma(\underline{r})$  sada možemo prikazati pomoću





**Slika 2.3:** Rekonstrukcija površine proteina korištenjem kuglinih funkcija s obzirom na broj koeficijenata koji se koriste za parametrizaciju. Slika preuzeta iz Ritchie (1998).

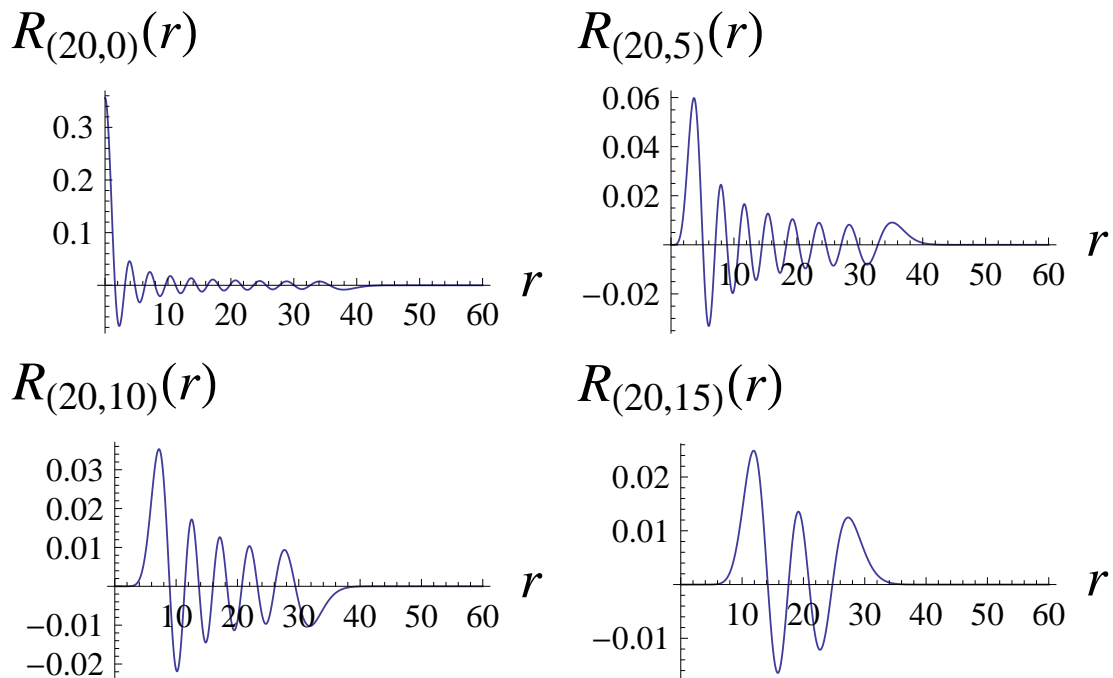
razvoja u novoj bazi

$$\sigma(\underline{r}) = \sum_{nlm}^N a_{nlm} F_{nlm}(\underline{r}) \quad (2.4)$$

Oznaka  $\sum_{nlm}^N$  je skraćeni zapis sume  $\sum_{n=1}^N \sum_{l=0}^{n-1} \sum_{m=-l}^{m=l}$ .

Parametar  $N$  zovemo redom sfernih harmonika. Za savršeni opis tijela proteina,  $N$  bi trebao težiti u beskonačnost. Međutim, za proces prijanjanja potrebna nam je samo aproksimacija ljuske proteina, tako da koristimo konačan  $N$ . Što je  $N$  veći, aproksimacija je bolja, no i proces je sporiji te zahtjeva više memorije. Obično se vrijednosti

$N$ -a kreću od 20 do 35.

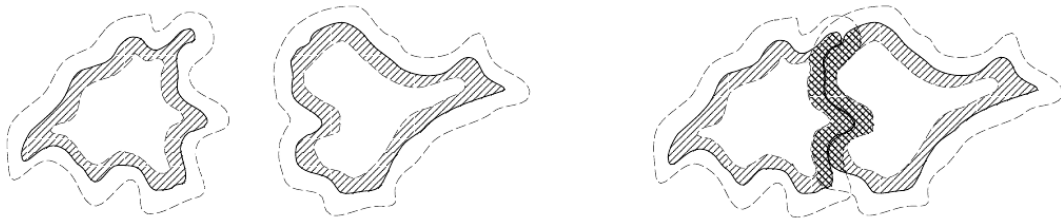


**Slika 2.4:** Prikaz nekoliko radijalnih funkcija definiranih jednažbom (2.2) s faktorom skaliranja  $\lambda = 20$ . Ove funkcije u paru s kuglinim funkcijama definiraju volumen ljuske proteina. Iz grafova je očito da ćemo dobro prikazati točke na udaljenosti do maksimalno oko 40 angstrema, jer su na tom intervalu funkcije najgušće. To je postignuto skaliranjem funkcije s parametrom  $\lambda = 20$ , zbog toga što se očekuje da će proteini prijanjati upravo na tom intervalu.

### 2.1.1. Model s dvije ljuske

Prilikom prijanjanja proteina pokušava se naći takva konformacija u kojoj će dodirna površina između dva proteina biti maksimalna. Isto tako, na neki način se moraju isključiti konformacije u kojima jedan protein ulazi u drugi. To je najlakše napraviti pomoću *modela s dvije ljuske* (engl. *double skin model*).

Svaki protein je reprezentiran pomoću dvije ljuske pa u procesu prijanjanja imamo dva niza koeficijenata za jedan protein — koeficijente  $\rho$  koji predstavljaju vanjsku ljusku i  $\tau$  koji predstavljaju unutarnju ljusku.



**Slika 2.5:** Komplementarne površine dvaju proteina najlakše je pronaći koristeći model s dvije ljuske - unutarnjom i vanjskom. Od idealne konformacije očekuje se maksimalno preklapanje unutarnje ljuske jednog i vanjske ljuske drugog proteina. Preklapanje unutarnjih ljuski je nepoželjno i algoritam ga kažnjava. Slika preuzeta iz Ritchie (1998).

## 2.2. Algoritam prijanjanja proteina

Dvije glavne komponente algoritma koje moramo definirati su:

- Pronalazak dovoljno velikog broja uniformno raspoređenih relativnih konformacija dvaju proteina (diskretizacija prostora stanja).
- Ocjena određene konformacije kako bi ih mogli međusobno uspoređivati i rangirati (evaluacijska funkcija).

### 2.2.1. Diskretizacija prostora stanja

Rotirajući jedan protein za kuteve  $(\alpha, \beta, \gamma)$ , a drugi samo za kuteve  $(\beta, \gamma)$  te translirajući ih na način da smanjujemo i povećavamo udaljenost njihovih koordinatnih sustava, možemo definirati sve moguće konformacije dviju proteina. Dobru raspodjelu kuteva  $(\beta, \gamma)$  generiramo *ikosaedralnim popločavanjem* (engl. *icosahedral tessellation*). Kut  $\alpha$  te  $R$ , udaljenost koordinatnih sustava, raspodjelujemo jednoliko po određenom intervalu.

#### Rotacije

Vrlo pogodno svojstvo proteina reprezentiranih pomoću kuglinih funkcija je da se mogu relativno brzo rotirati. To svojstvo dodatno koristimo budući da pretražujemo stanja pomoću pet rotacijskih stupnjeva slobode te samo jednim translacijskim stupnjem slobode. Translacija je u domeni kuglinih funkcija sporija operacija.

Može se pokazati (Ritchie, 1998) da su koeficijenti proteina rotiranog za Eulerove kuteve  $(\beta, \gamma)$ :

$$a'_{nlm} = \sum_{m'=-l}^l a_{nlm'} R_{mm'}^{(l)}(\beta, \gamma) \quad (2.5)$$

gdje je  $R$  element  $(m, m')$  rotacijske matrice reda  $l$  koji ovisi o kutovima  $\beta$  i  $\gamma$ . Za rotiranje oko kuta  $\alpha$  koristi se jednostavniji izraz:

$$a'_{nlm} = a_{nlm} \cos m\alpha + a_{nl\bar{m}} \sin \bar{m}\alpha \quad (2.6)$$

## Translacije

Protein transliramo samo po jednoj dimenziji, udaljenosti koordinatnih sustava jednog i drugog proteina. Izraz za udaljšavanje koordinatnih sustava za udaljšnost  $R$  jest:

$$a'_{nlm} = \sum_{n'l'm'}^N a_{n'l'm'} K_{nn'l'l'|m|}(R) \delta_{mm'} \quad (2.7)$$

gdje je  $K$  peterodimenzionalna translacijska matrica koja ovisi o radiusu  $R$ . Postupak generiranja translacijskih matrica opisan je u Ritchie (2005).

### 2.2.2. Evaluacijska funkcija

Kako će dva proteina interaktivirati kad se nađu jedan blizu drugog ovisi o raznim kemijskim i fizičkim svojstvima, njih samih ali i njihove okoline. Međutim, kod programske implementacije predviđanja njihove interakcije, ta svojstva moramo što više pojednostaviti i učiniti ih jednostavnijim za računanje. U našem programu ta funkcija ovisi o volumenu unije preklapanja vanjskih i unutarnjih ljusaka jednog i drugog proteina na sljedeći način:

$$S = \int \rho_A(\underline{r}_A) \tau_B(\underline{r}_B) dV + \int \tau_A(\underline{r}_A) \rho_B(\underline{r}_B) dV - Q \int \tau_A(\underline{r}_A) \tau_B(\underline{r}_B) dV \quad (2.8)$$

gdje je  $\rho$  vanjska ljuska, a  $\tau$  unutarnja ljuska određenog proteina. Parametar  $Q$  označava "kaznu" za prodiranje proteina jedan u drugi, što je fizički nemoguće. Obično se postavlja između 10 i 15.

### 2.2.3. Traženje strukture kompleksa

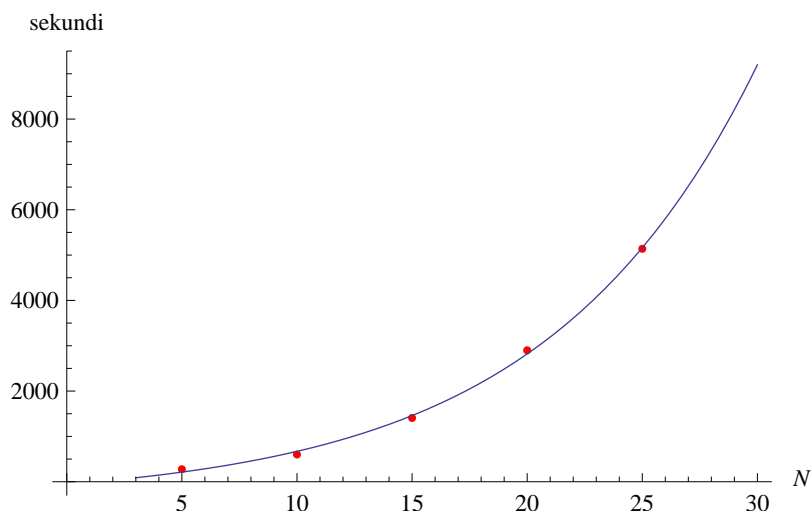
Nakon što smo sve definirali, možemo iznjeti pseudokod traženja strukture kompleksa dvaju proteina:

- 1: **za svaki**  $(\beta_1, \gamma_1)$  **radi**
- 2:     Izračunaj rotaciju proteina A za  $(\beta_1, \gamma_1)$  i zapamti.
- 3: **kraj za svaki**
- 4: **za svaki**  $(\alpha, \beta_2, \gamma_2)$  **radi**
- 5:     Izračunaj rotaciju proteina B za  $(\alpha, \beta_2, \gamma_2)$  i zapamti.
- 6: **kraj za svaki**
- 7: **za svaki**  $R$  **radi**
- 8:     **za svaki**  $(\beta_1, \gamma_1)$  **radi**
- 9:         Učitaj rotirani protein A i transliraj ga za  $R$
- 10:     **za svaki**  $(\alpha, \beta_2, \gamma_2)$  **radi**
- 11:         Učitaj rotirani protein B i evaluiraj konformaciju s transliranim proteinom A
- 12:     **kraj za svaki**
- 13: **kraj za svaki**
- 14: **kraj za svaki**

## 3. Steric Scan

### 3.1. Ideja

Jedan od glavnih parametara u proces prijanjanja je red sfernih harmonika s kojim ćemo opisati ljuske proteina. Što je taj red veći, protein će biti točnije aproksimiran, no sam proces će zahtjevati više računalnih resursa. Vrijeme izvršavanja s obzirom na različite redove se može vidjeti na slici 3.1.



**Slika 3.1:** Vrijeme izvršavanja procesa prijanjanja na određenim redovima sfernih harmonika  $N$ . Proces je na svakom redu evaluirao 149, 299, 200 konformacija. Testiranje je izvršeno na računalu s procesorom Intel Core2 Duo na 2.10GHz te s 2 GB radne memorije. Funkcija koja spaja točke na slici je  $-421.18 + e^{5.91+0.11N}$ .

S druge strane, u proces prijanjanja dolazi određeni broj konformacija proteina koje trebamo evaluirati. Evaluacija podrazumijeva rotaciju i translaciju dvaju proteina te ocjenu transformiranih proteina. Evaluacija jedne konformacije je brža na nižem redu, a sporija na višem.

Postupkom profiliranja jednog izvršavanja procesa na redu 20, ustanovljeno je da funkcija koja pretražuje sve konformacije i ocjenjuje ih troši 91.6% ukupnog vremena izvršavanja. Zaključujemo da bismo ubrzanjem te funkcije mogli postići značajnu

uštedu vremena.

Osnovna ideja metode *steric scan* je raditi proces prijanjanja u dva koraka:

1. Pokrenuti proces prijanjanja na nižem redu te izbaciti određeni broj konformacija.
2. Na manjem skupu konformacija pokrenuti proces prijanjanja na višem redu.

Cilj je izbaciti konformacije koje su fizički vrlo nemoguće (znatna penetracija proteina jednog u drugi) te one u kojima se proteini uopće ni ne dodiruju. To je moguće detektirati i na nižim redovima. Kratkom inspekcijom slike 3.1 možemo naslutiti da bismo određeno ubrzanje mogli postići izbacivanjem već i malog broja konformacija na nižim redovima.

Provjerit ćemo tu pretpostavku.

## 3.2. Programska implementacija

Implementacijom funkcije koja kao parametre prima:

1. red sfernih harmonika  $i$
2. skup konformacija koje je potrebno evaluirati.

te koja vraća ocjenjene konformacije, riješili smo problem. Tada nam se algoritam svodi na:

1. pozovi funkciju na nižem redu s punim skupom konformacija,
2. napravi novi skup konformacija koji se sastoji od određenog broja najbolje ocjenjenih konformacija,
3. pozovi funkciju na višem redu s novim, reduciranim skupom konformacija.

Konformacija se opisuje uređenom petorkom  $(R, \beta_1, \gamma_1, \beta_2, \gamma_2)$ . U svakoj toj konformaciji se ligand još rotira za kut  $\alpha$  po cijelom intervalu (loši kutevi  $\alpha$  se ne izbacuju u metodi *steric scan*).

Jedini problem prilikom implementacije te funkcije je efikasno iskorištenje mehanizma *precalculate*. Za razliku od algoritma u poglavlju 2.2.3. koji pretražuje potpuni skup konformacija, novi algoritam je:

- 1: Izračunaj parametre ljuski proteina na redu  $N$
- 2: Izračunaj sve rotacije receptora i liganda (s obzirom na ulazni skup konformacija)
- 3: Sortiraj ulazne konformacije redom po ključevima  $R, (\beta_1, \gamma_1), (\beta_2, \gamma_2)$ .
- 4: prošliR  $\leftarrow -1$
- 5: prošliReceptor  $\leftarrow \text{NULL}$
- 6: **za svaki**  $(R, \beta_1, \gamma_1, \beta_2, \gamma_2)$  iz skupa ulaznih konformacija **radi**
- 7:   **ako** prošliR  $\neq R$  **onda**
- 8:     Izračunaj translacijsku matricu na udaljenosti  $R$
- 9:     prošliReceptor  $\leftarrow \text{NULL}$
- 10:   **kraj ako**
- 11:   **ako** prošliReceptor  $\neq (\beta_1, \gamma_1)$  **onda**
- 12:     Učitaj receptor zarotiran za  $(\beta_1, \gamma_1)$  i transliraj ga na udaljenost  $R$ .
- 13:     prošliReceptor  $\leftarrow (\beta_1, \gamma_1)$
- 14:   **kraj ako**
- 15:   **za svaki**  $\alpha$  **radi**
- 16:     Učitaj ligand zarotiran za  $(\alpha, \beta_2, \gamma_2)$
- 17:     Pridruži ulaznoj konformaciji ocjenu transformiranih liganda i receptora.
- 18:   **kraj za svaki**
- 19: **kraj za svaki**

Pomoću ovog algoritma postigli smo maksimalnu efikasnost s obzirom na ulazne konformacije. Budući da ulazne konformacije sortiramo primarno po ključu  $R$ , omogućili smo da računanje translacijske matrice za svaki  $R$  radimo samo jednom. Kad se  $R$  promijeni, obrišemo staru translacijsku matricu jer znamo da nam više neće trebati i računamo novu. Isto tako, rotirane receptore transliramo samo jednom, iako ih više puta koristimo u nekim konformacijama.

### 3.3. Istraživanja i rezultati

Izlaz iz alata za prijanjanje proteina je nekoliko stotina najbolje ocjenjenih konformacija s nadom da će se u tih nekoliko stotina pronaći ona prava, koja se pojavljuje i u prirodi. Tada kažemo da smo uspješno predvidjeli strukturu proteinskog kompleksa.

Iz tog razloga, metoda steric scan nebi smjela na nižim redovima izbaciti neku od prvih konformacija zbog opasnosti da ne izbací pravu. Glavni parametar o kojem ovisi uspješnost te metode je postotak konformacija koje će preživjeti filtriranje na nižem redu. Taj parametar zvat ćemo *postotak preživjelih*. Što je postotak preživjelih veći, manja je opasnost da izbacimo pravu konformaciju, no brzina je sporija. Što više taj parametar smanjujemo, povećavamo brzinu izvođenja uz veću opasnost da na nižem redu izbacimo konformaciju koju tražimo.

Kako bismo pronašli idealnu vrijednost za postotak preživjelih koja će pronaći optimalan kompromis između brzine izvođenja i točnosti algoritma, izvršit ćemo sljedeća



istraživanja:

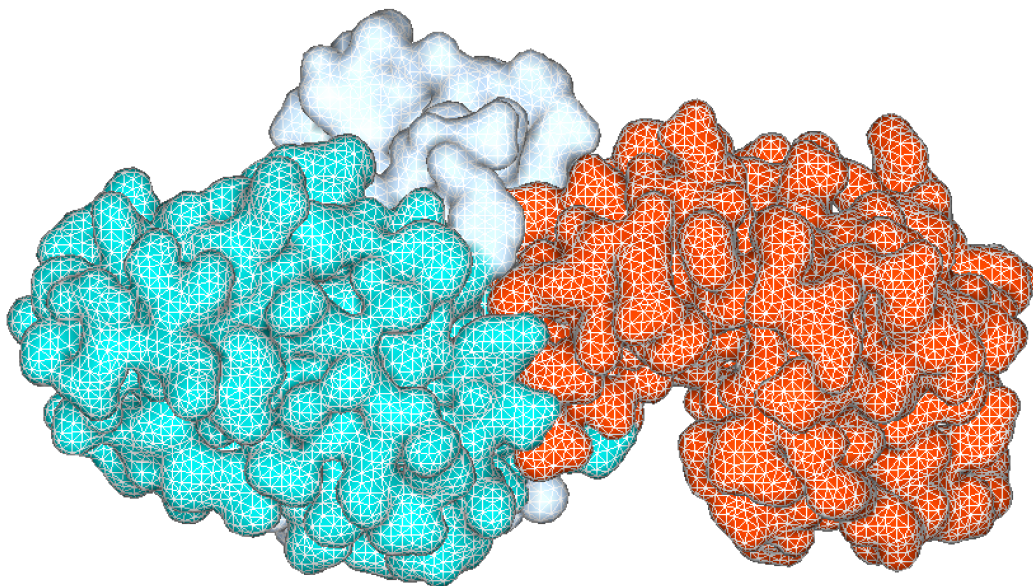
1. Koliki bi morao biti postotak preživjelih ako želimo da preživi  $X$  najboljih?
2. Ako je postotak preživjelih  $X$ , koliko konformacija od najboljih 500 će preživjeti?
3. Korelacija ocjena najboljih 3,000 konformacija na nižem i višem redu.
4. Korelacija ocjena slučajno odabranih 3,000 konformacija na nižem i višem redu.

Nakon izvršenih istraživanja, za svaki red ćemo pronaći takav postotak preživjelih da niti jedna od 500 najboljih konformacija ne bude izbačena na nižem redu (u 500 najboljih očekujemo pravu konformaciju). S tim postotkom preživjelih ćemo pokrenuti proces prijanjanja temeljen na metodi steric scan te analizirati ubrzanje.

Pod pojmom najbolje konformacije sada i u daljnjem tekstu smatramo one koje su dobile najveće ocjene na višem redu.

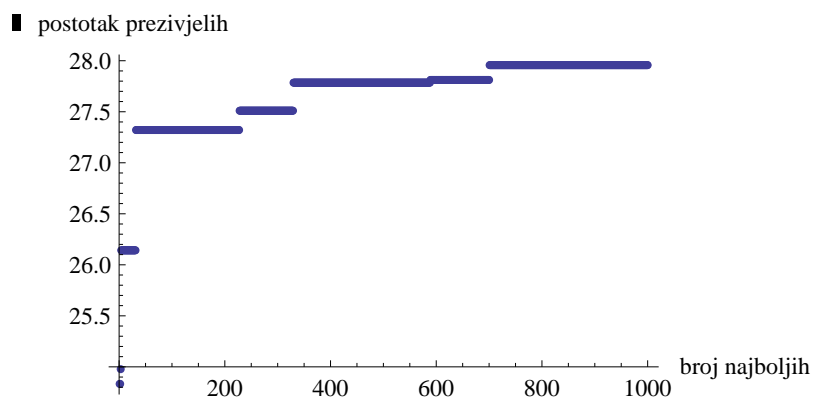
Svaki od tih istraživanja radit ćemo na nižim redovima 5, 10 i 15 te ćemo za viši red uzeti 20. U trenutnoj verziji Protein Docking Tool-a, 20 je najveći red za koji proces prijanjanja daje dobre rezultate, zbog premale preciznosti računanja translacijskih matrica.

Za ulazne proteine uzet ćemo proteine iz kompleksa 3HFL (G. H. Cohen, 1996) — antitijelo Hy/HEL-5 i lizosom iz kokošnjeg bjelanjka. Kristalografska struktura tog kompleksa, koju tražimo procesom prijanjanja, prikazana je na slici 3.2. Kompleks sadrži ukupno 4,333 atoma.

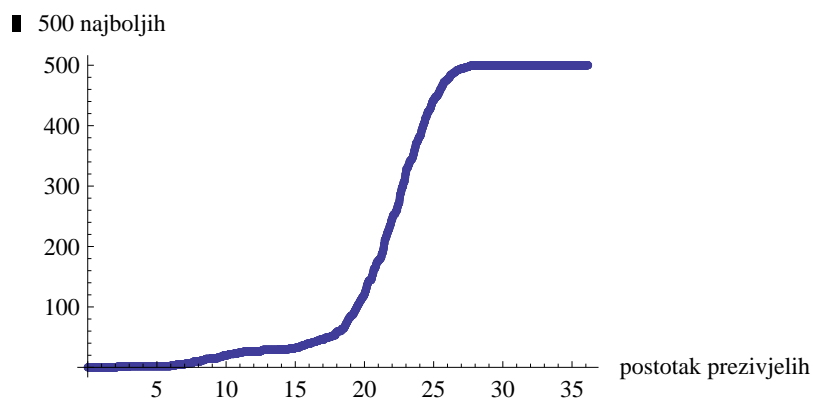


**Slika 3.2:** Kristalografska struktura 3HFL proteinskog kompleksa. Slika je generirana u alatu za prijanjanje proteina Hex.

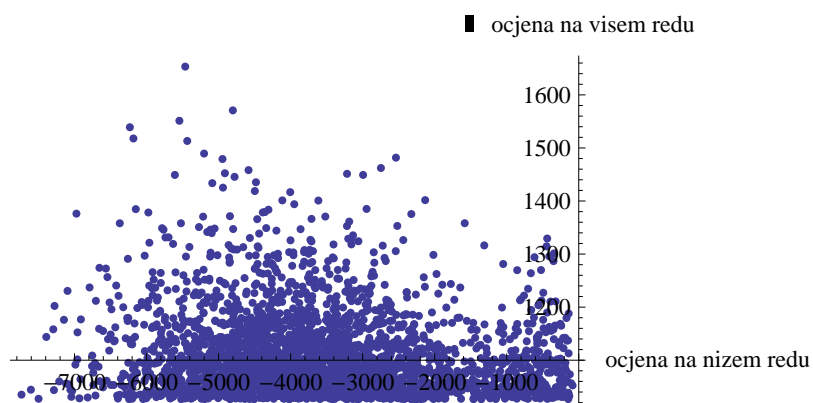
### 3.3.1. Istraživanja na redu 5



**Slika 3.3:** Redovi 5/20 — Koliki bi morao biti postotak preživjelih ako želimo da preživi  $X$  najboljih? Iz slike se vidi da bismo s postotkom preživjelih oko 28% obuhvatili najboljih 1,000 konformacija na redu 20.

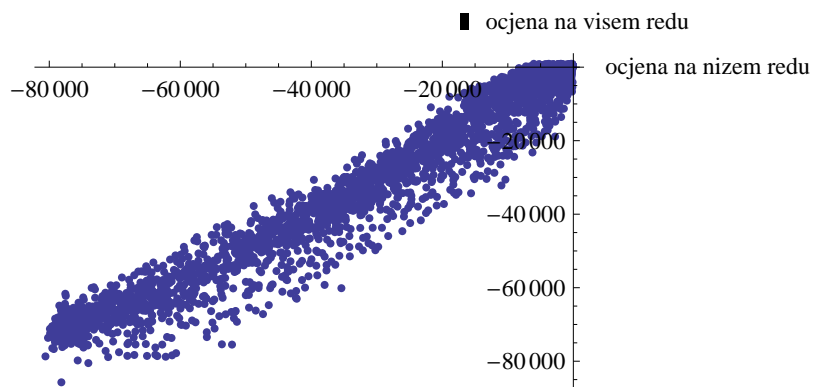


**Slika 3.4:** Redovi 5/20 — Ako je postotak preživjelih  $X$ , koliko konformacija od najboljih 500 će preživjeti? Veliki skok se događa na postotku preživjelih oko 28%.



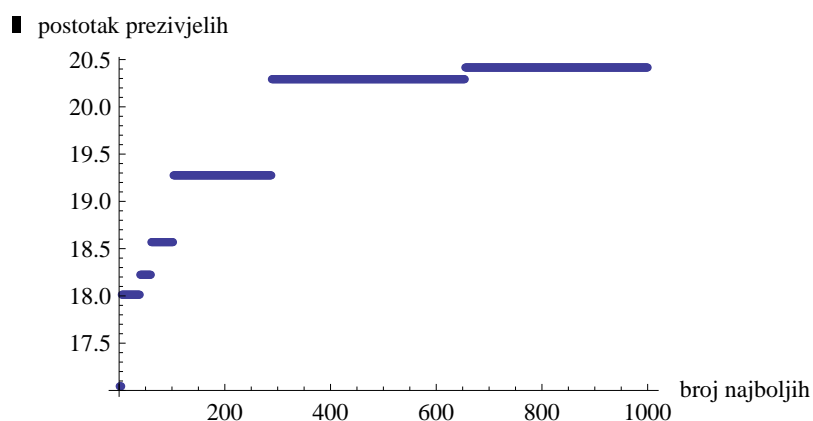
**Slika 3.5:** Redovi 5/20 — Korelacija ocjena najboljih 3,000 konformacija na nižem i višem redu. Korelacija nije velika (Pearsonov koeficijent je **-0.135192**), što je i očekivano, budući da su najbolje konformacije one u kojima se proteini dodiruju, a kod tih konformacija red sfernih harmonika, odnosno preciznost ljuski proteina igra ogromnu ulogu.

Eksperimentalno utvrđen postotak preživjelih kod kojeg će prvih 500 konformacija preživjeti filter nižeg reda 5 je **27.8%**.

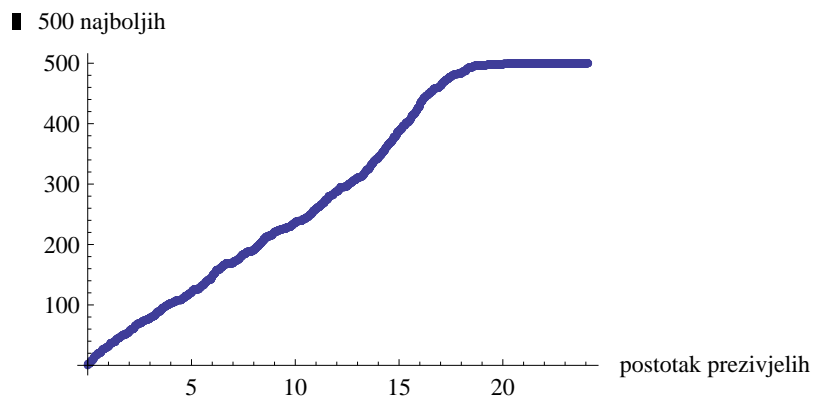


**Slika 3.6:** Redovi 5/20 — Korelacija ocjena slučajno odabranih 3, 000 konformacija na nižem i višem redu. Korelacija je velika (Pearsonov koeficijent je **0.977059**), što objašnjavamo činjenicom da se u ogromnoj većini ulaznih konformacija proteini ili uopće ne dodiruju ili penetriraju jedan u drugi. Kod takvih konformacija, red sfernih harmonika ne igra veliku ulogu u ocjeni. Na tom zaključku upravo i baziramo ideju metode steric scan.

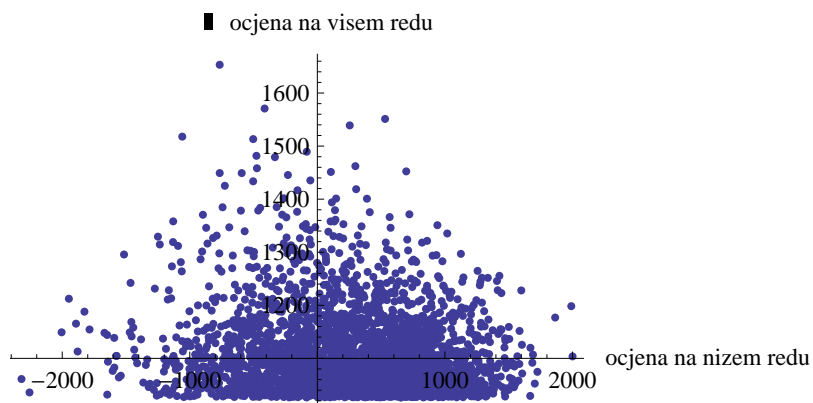
### 3.3.2. Istraživanja na redu 10



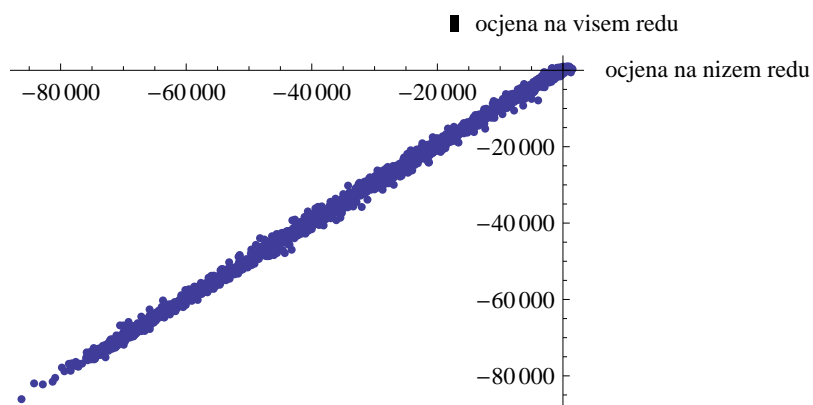
**Slika 3.7:** Redovi 10/20 — Koliki bi morao biti postotak preživjelih ako želimo da preživi  $X$  najboljih? Iz slike se vidi da bismo s postotkom preživjelih oko 20.5% obuhvatili najboljih 1,000 konformacija na redu 20.



**Slika 3.8:** Redovi 10/20 — Ako je postotak preživjelih  $X$ , koliko konformacija od najboljih 500 će preživjeti? S postotkom preživjelih oko 20% obuhvatit ćemo najboljih 500 konformacija na redu 20.



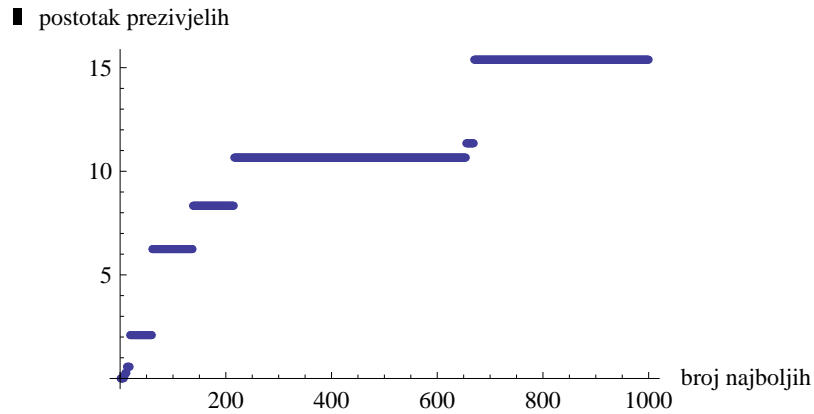
**Slika 3.9:** Redovi 10/20 — Korelacija ocjena najboljih 3,000 konformacija na nižem i višem redu. Pearsonov koeficijent korelacije je **-0.131962**, slično kao i na redu 5.



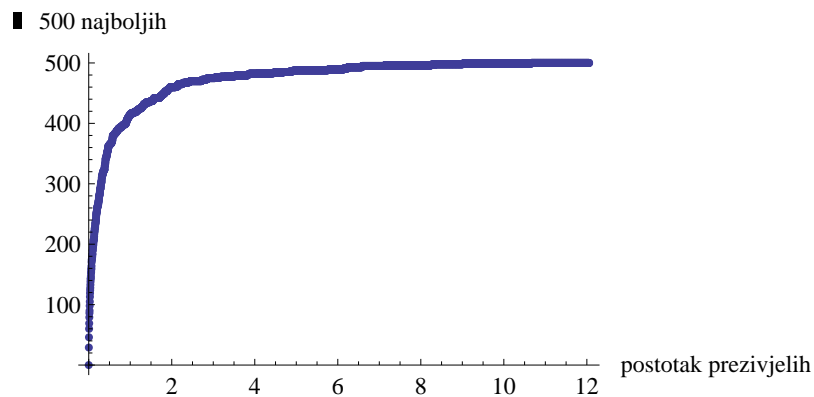
**Slika 3.10:** Redovi 10/20 — Korelacija ocjena slučajno odabranih 3,000 konformacija na nižem i višem redu. Pearsonov koeficijent je **0.999359**, što je puno bolje nego na redu 5.

Eksperimentalno utvrđen postotak preživjelih kod kojeg će prvih 500 konformacija preživjeti filter nižeg reda 10 je **20.3%**.

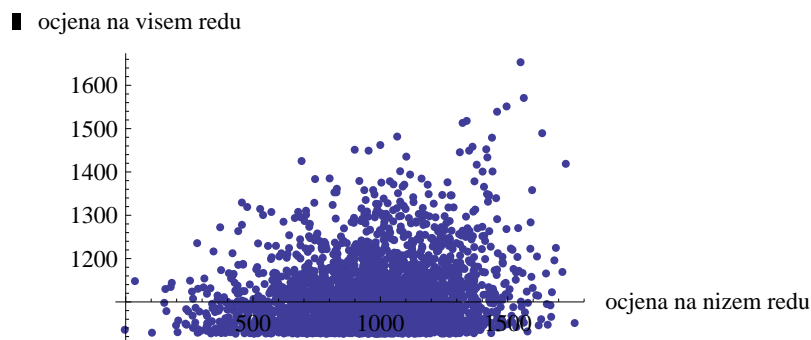
### 3.3.3. Istraživanja na redu 15



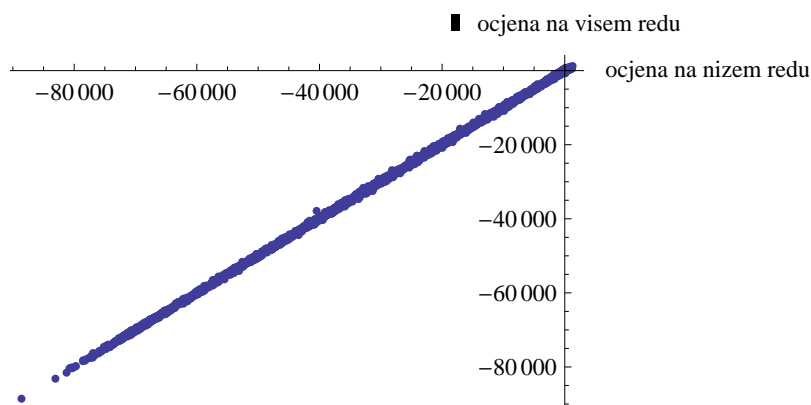
**Slika 3.11:** Redovi 15/20 — Koliki bi morao biti postotak preživjelih ako želimo da preživi  $X$  najboljih? Iz slike se vidi da bismo s postotkom preživjelih oko 12% obuhvatili najboljih 700 konformacija na redu 20.



**Slika 3.12:** Redovi 15/20 — Ako je postotak preživjelih  $X$ , koliko konformacija od najboljih 500 će preživjeti? Možemo primjetiti da bi s postotkom preživjelih već oko 2% obuhvatili veliku većinu od najboljih 500 konformacija na redu 20.



**Slika 3.13:** Redovi 15/20 — Korelacija ocjena najboljih 3, 000 konformacija na nižem i višem redu. Pearsonov koeficijent korelacije je **0.186099**. Za razliku od redova 5 i 10, koeficijent je pozitivan, no još uvijek nedostatan da proces prijanjanja na redu 15 proglasimo mjerodavnim.



**Slika 3.14:** Redovi 15/20 — Korelacija ocjena slučajno odabranih 3, 000 konformacija na nižem i višem redu. Pearsonov koeficijent korelacije je **0.999926**. Korelacija slučajno odabranih konformacija je gotovo savršena.

Eksperimentalno utvrđen postotak preživjelih kod kojeg će prvih 500 konformacija preživjeti filter nižeg reda je **10.7%**.

### 3.4. Postignuta ubrzanja

**Tablica 3.1:** Postignuta ubrzanja

Niži red	Viši red	Bez steric scan-a	Sa steric scan-om	Faktor ubrzanja
5	20	2899.43 s	1314.41 s	2.2
10	20	2899.43 s	1499.41 s	1.9
15	20	2899.43 s	2069.43 s	1.4

Nakon što smo za svaki od nižih redova 5, 10, 15 odredili postotak preživjelih kod kojeg će prvih 500 konformacija preživjeti, pokrenuli smo proces prijanjanja s implemen-

tiranom metodom steric scan. Postignuta ubrzanja prikazana su u tablici 3.1. Vremena su mjerena na računalu s procesorom Intel Core2 Duo na 2.10GHz te s 2 GB radne memorije.

Prvih 500 konformacija u oba slučaja bile su identične.



## 4. Zaključak i diskusija

Pokazale su se točne pretpostavke na kojima je razvijena ideja metode steric scan:

- Na nižem redu proces dockinga je znatno brži nego na višem redu.
- Postoji velika korelacija između ocjene na nižem i višem redu, koja je to veća što su redovi bliži jedan drugome.

Implementacijom metode dobiveni su izuzetni rezultati te je postignuto ubrzanje izvođenja od 2.2 puta, bez velikog dodatnog trošenja memorije. Od nižih redova na kojim su rađena istraživanja, 5, 10 i 15, najefikasnijim se pokazao red 5.

Istraživanje je vršeno na jednom proteinskom kompleksu (3HFL), ograničenim skupom nižih redova te uz uvjet da steric scan preživi 500 najboljih konformacija. Daljnji smjer istraživanja s ciljem šire primjenjivosti metode bio bi generalizirati zaključke o postotku preživjelih s obzirom na red i na uvjetovani broj najboljih konformacija. Također, u potrazi za najefikasnijim nižim redovima, trebalo bi provjeriti širi skup redova, koji bi mogli dati ubrzanje veće od onog postignutog na redu 5. Mogla bi se provjeriti i mogućnost trostupanjskog steric scan-a, no ne očekujem da će to dati dobre rezultate, zbog overhead-a koji svaki dodatni red unosi (računski zahtjevna parametrizacija ljusaka).

Osim istraživanja same metode steric scan, alat za prijanjanje proteina trebalo bi poboljšati na način da se implementira veća preciznost računanja translacijskih matrica korištenjem višestruke preciznosti. Tada bi bilo moguće proces prijanjanja pokrenuti na redu 25 i višima.

# LITERATURA

Piškorec Sović Šikić Antulov-Fantulin, Čanadi. Protein Docking Tool. 2010. URL <http://complex.zesoi.fer.hr/PDT.html>.

Ivan Dokmanić. Algoritam za predviđanje strukture proteinskih kompleksa korištenjem razvoja površine u redove funkcija. Diplomski rad, Fakultet elektrotehnike i računarstva, 2007.

D. R. Davies G. H. Cohen, S. Sheriff. Refined structure of the monoclonal antibody HyHEL-5 with its antigen hen egg-white lysozyme. *Acta Crystallogr D Biol Crystallogr*, 1996.

David W. Ritchie. *Parametric Protein Shape Recognition*. Doktorska disertacija, Departments of Computing Science and Molecular & Cell Biology, University of Aberdeen, 1998.

David W. Ritchie. High-order analytic translation matrix elements for real-space six-dimensional polar fourier correlations. *Journal of Applied Crystallography*, 2005.

## **Alat za prijanjanje proteina: modul za pripremu prijanjanja**

### **Sažetak**

Problem predviđanja strukture proteinskog kompleksa (prijanjanja proteina) odgovara na pitanje hoće li i kako dva proteina interagirati. Opisana je jedna od metoda prijanjanja proteina, Spherical Polar Fourier. Iznjeta je ideja Steric Scan-a s ciljem da se ubrza proces prijanjanja proteina metodom Spherical Polar Fourier. Steric Scan je implementiran te je postignuto maksimalno ubrzanje od 2.2 puta.

**Ključne riječi:** proteini, prijanjanje proteina, proteinski kompleks, Spherical Polar Fourier, Steric Scan

## **Protein Docking Tool: Steric Scan module**

### **Abstract**

The problem of predicting structure of protein complex (Protein Docking) answers the question whether and how two proteins might associate. Spherical Polar Fourier, one of the methods for Protein Docking was described. An idea of Steric Scan is developed, a method which goal is to speed up the process of Protein Docking using Spherical Polar Fourier method. Steric Scan was implemented and speed up of 2.2 times was achieved.

**Keywords:** Proteins, Protein Docking, Protein Complex, Spherical Polar Fourier, Steric Scan