

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br.77

**ALAT ZA PRIANJANJE PROTEINA: MODULI
ZA ROTACIJU I VREDNOVANJE REZULTATA**

Nino Antulov-Fantulin

Zagreb, lipanj, 2010

Hvala mojoj obitelji Antulov-Fantulin.

Zahvaljujem se svom mentoru dr.sc. Mili Šikiću za uspješnu suradnju na ovom radu i potporu tijekom čitavog studiranja. Također hvala PDT timu (Ivanu Saviću, Matiji Piškorcicu i Igoru Čanadiju).

Sadržaj

UVOD	2
1. TEORIJSKA PODLOGA.....	3
1.1. PROSTORI FUNKCIJA	4
1.2. FOURIEROV RED	7
1.2.1. <i>Fourierov red u ravnini</i>	7
1.2.2. <i>Fourierov red na sferi</i>	8
1.2.3. <i>Sferni koordinatni sustav</i>	9
1.3. PRIDRUŽENI LEGENDROVI POLINOMI.....	11
1.4. SFERNI HARMONICI.....	13
1.4.1. <i>Ortonormirana baza</i>	13
1.4.2. <i>Realni sferni harmonici</i>	14
1.4.3. <i>Vizualizacija sfernih harmonika</i>	15
1.5. RADIJALNE FUNKCIJE	18
2. LJUSKE PROTEINA	23
2.1. RAZVOJ LJUSKE PROTEINA U BAZNE FUNKCIJE	24
2.1.1. <i>Rotacije ljuske proteina</i>	25
2.2. TRANSLACIJE LJUSKI PROTEINA.....	28
3. OCJENA KOMPLEMENTARNOSTI PROTEINA.....	31
4. IMPLEMENTACIJA.....	35
4.1. PARALELIZACIJA SUSTAVA	38
4.1.1. <i>MPI</i>	38
4.1.2. <i>Funkcionalnost paralelnog programa</i>	38
5. TESTOVI I REZULTATI.....	40
5.1. 3HFL KOMPLEKS.....	42
5.1.1. <i>Vizualizacija kompleksa u kristalografskoj strukturi</i>	42
5.1.2. <i>Test: Alfa rotacije kristalografske strukture</i>	42
5.1.3. <i>Test: Ocjena komplementarnosti po radijusu</i>	43
5.1.4. <i>Test: Detekcija kristalografske strukture</i>	44
5.1.5. <i>Test: Razvoj ljuski po z-osi</i>	45
5.2. 1AY7 KOMPLEKS	46
5.2.1. <i>Vizualizacija kompleksa u kristalografskoj strukturi</i>	46
5.2.2. <i>Test: Alfa rotacije kristalografske strukture</i>	46

5.2.3. Test: Ocjena komplementarnosti po radijusu	47
5.2.4. Test: Detekcija kristalografske strukture	48
5.2.5. Test: Razvoj ljuski po z-osi.....	49
5.3. TEST: BRZINA ALGORITMA PRIANJANJA	51
ZAKLJUČAK	53
LITERATURA.....	54
SAŽETAK	55
ABSTRACT.....	56
DODATAK.....	57
A.) POKRETANJE PARALELNOG PROGRAMA.....	57
B.) INSTALACIJA PARALELNOG PROGRAMA	58

Uvod

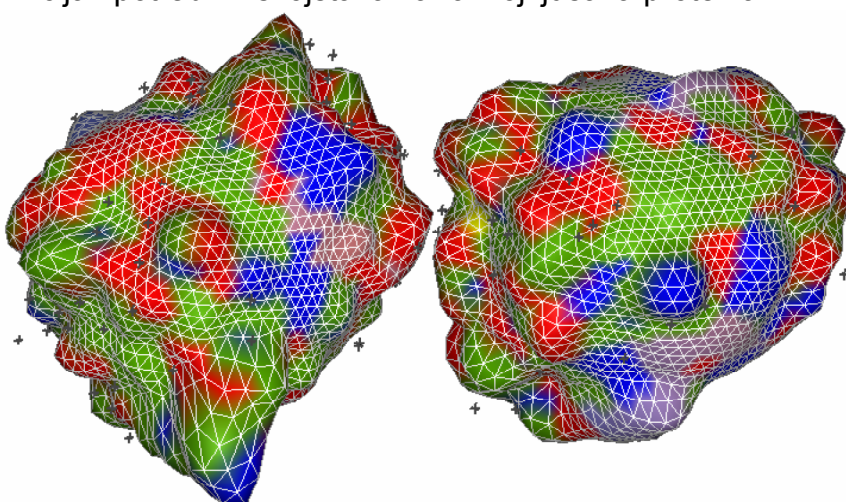
Zadnjih nekoliko desetljeća ubrzani razvoj tehnologije i znanosti u području raznih molekularnih istraživanja i genomike stvorio je enormne količine podataka povezanih sa molekularnom biologijom. Primjena metoda računarske znanosti nad tim podacima stvorio je novu interdisciplinarnu granu bioinformatike koja je rasvijetlila mnoge do tada neriješene probleme molekularne biologije.

Prianjanje proteina (engl. Protein Docking) je područje bioinformatike koje se bavi proučavanjem stvaranja mogućih kompleksa proteina. Prianjanje proteina pokušava predvidjeti mogu li dva proteina stupiti u kompleks i koja je njihova 3D struktura kompleksa na temelju svojstava geometrijske komplementarnosti površina, hidrofobnosti ili elektrostatike.

U ovom radu je korištena SPF (engl. Spherical Polar Fourier) metoda prianjanja proteina koja je opisana i prvi put implementirana u radovima Ritchiea i Kempa [1][2][3]. Ova metoda razvija ljuske proteina u redove trodimenzionalnih baznih funkcija (radijalne funkcije sa sfernim harmonicima) koje potom transformira (rotira, translata) u šest dimenzionalnom prostoru pretraživanja mogućih konformacija. Ovaj rad je nastavak rada na projektu PDT (engl. Protein Docking Tool) [4] i predstavlja nadogradnju prijašnjeg sustava [5]. U ovom radu je implementirano ubrzanje algoritma prianjanja proteina u slijednoj i paralelnoj verziji koristeći svojstva Fourierovog reda.

1. Teorijska podloga

U ovom poglavlju objašnjavaju se temeljni matematički pojmovi potrebni za razumijevanje algoritma prijanjanja proteina. U poglavlju 1.1 objašnjeni su osnovni koncepti konačno i beskonačno dimenzionalnih prostora [10]. Poglavlje 1.2. pokazuje izvod Fourierovog reda na sferi i u ravnini [13]. Poglavlje 1.3. objašnjava skup pridruženih Legendrovih polinoma koji je nužan za razumijevanje sfernih harmonika. Poglavlje 1.4. objašnjava strukturu sfernih harmonika, njihove ortonormiranosti i vizualizacije [11][12]. Poglavlje 1.5. daje pregled korištenih radijalnih funkcija i potrebnih svojstava za razvoj ljusaka proteina.



Slika 1. Razvoj površine proteina 3HFL kompleksa [6] u sumu sfernih harmonika reda 30.

1.1. Prostori funkcija

Svi se prostori mogu podijeliti na konačno dimenzionalne prostore i beskonačno dimenzionalne prostore. Je li neki prostor konačno dimenzionalan ili beskonačno dimenzionalan određujemo po broju linearno nezavisnih vektora koje možemo u njemu pronaći. Područje matematike koje izučava beskonačno dimenzionalne prostore naziva se funkcionalna analiza. Dok se Linearna algebra većinom bavi konačno dimenzionalnim prostorima.

Definicija: Prostor $L^2([a, b])$ je skup svih kvadratno integrabilnih funkcija na intervalu

$$[a, b]. L^2([a, b]) = \left\{ f : [a, b] \rightarrow \mathbb{C}; \int_a^b |f(t)|^2 dt < \infty \right\}$$

Fizikalno značenje uvjeta $\int_a^b |f(t)|^2 dt < \infty$ govori nam da je ukupna energija signala konačna.

Prostor $L^2([a, b])$ je beskonačno dimenzionalan. Npr. skup funkcija $\{1, t, t^2, t^3, \dots\} \in L^2([0, 1])$ je linearno nezavisan. Skup $\{1, t, t^2, t^3, \dots\}$ je linearno nezavisan ako i samo ako je svaki njegov konačan podskup linearno nezavisan jer za svaki k skup $\{1, t, t^2, t^3, \dots, t^k\}$ prema osnovnom teoremu algebre iščezava jedino za nul-polinom. Polinom stupnja $k : \lambda_1 + \lambda_2 t + \lambda_3 t^2 + \dots + \lambda_k t^k = 0$ jednak je nuli samo onda ako vrijedi $\lambda_1 = \lambda_2 = \lambda_3 = \dots = \lambda_k = 0$ što znači da je skup $\{1, t, t^2, t^3, \dots, t^k\}$ linearno neovisan.

Definicija: Vektorski prostor sa definiranim skalarnim produktom $\langle x, y \rangle$ zove se unitarni prostor.

U unitarnom prostoru vrijedi nejednakost Cauchy-Schwarz-Buniakowski:

$$|\langle x, y \rangle| \leq \|x\| \|y\|.$$

Definicija: Za vektore $e_1, e_2, \dots, e_n \in \mathbb{R}^n$ kažemo da čine ortogonalnu bazu ako su svi međusobno okomiti. Za svaki vektor $x \in \mathbb{R}^n$ vrijedi rastav po ortogonalnoj bazi:

$$x = \frac{\langle x, e_1 \rangle}{\|e_1\|^2} e_1 + \dots + \frac{\langle x, e_n \rangle}{\|e_n\|^2} e_n .$$

Ukoliko su norme vektora e_1, e_2, \dots, e_n jedinične onda takvu bazu zovemo ortonormiranom bazom.

Definicija: Skalarni produkt funkcija $f, g \in L^2([a, b])$ definiramo:

$$\langle f, g \rangle = \int_a^b f(x)g(x)dx .$$

Definicija: Potpun unitaran prostor nazivamo Hilbertov prostor. Ako za svaki Cauchyjev slijed u X konvergira, onda kažemo da je vektorski prostor X potpun. Slikovito rečeno znači da je vektorski prostor bez „šupljina“ u sebi.

Primjer: Neka je vektorski prostor $X \in L^2[0, 2\pi]$ sa definiranim skalarnim produktom $\langle x, y \rangle = \int_0^{2\pi} x(t)y(t)dt$, onda skup funkcija $\{\sin kt\}_{k \in \mathbb{Z}}$ čini beskonačnu ortogonalnu bazu u unitarnom prostoru X. Proizvoljni signal $x(t) \in X$ može se rastaviti u Fourierov red po sinusoidalnim harmonicima.

$$x(t) = \frac{\langle x(t), e_1(t) \rangle}{\|e_1(t)\|^2} e_1(t) + \frac{\langle x(t), e_2(t) \rangle}{\|e_2(t)\|^2} e_2(t) + \dots = \sum_{k=0}^{\infty} \frac{\langle x(t), e_k(t) \rangle}{\|e_k(t)\|^2} e_k(t), e_k(t) = \sin kt$$

Primjer: L^2 norma funkcije $f(x)$ se računa ovako $\|f(x)\| = \sqrt{\langle f(x), f(x) \rangle}$

Za funkciju $f(x) = 1$ vrijedi: $\|f(x)\|_{L^2}^2 = \int_0^{2\pi} 1^2 dx = 2\pi$,

Za funkciju $f(x) = \cos(x)$ vrijedi: $\|f(x)\|_{L^2}^2 = \int_0^{2\pi} \cos^2(x) dx = \pi$,

Za funkciju $f(x) = \sin(x)$ vrijedi: $\|f(x)\|_{L^2}^2 = \int_0^{2\pi} \sin^2(x) dx = \pi$,

$$\langle 1, \cos(nx) \rangle = \int_0^{2\pi} 1 \cos(nx) dx = 0 \Rightarrow 1 \perp \cos nx,$$

$$\langle 1, \sin(nx) \rangle = \int_0^{2\pi} 1 \sin(nx) dx = 0 \Rightarrow 1 \perp \sin nx,$$

$$\langle \sin(mx), \sin(nx) \rangle = \int_0^{2\pi} \sin(mx) \sin(nx) dx = \begin{cases} n \neq m : 0 \\ n = m : \pi \end{cases}$$

Skup funkcija $\left\{ \frac{1}{\sqrt{2\pi}}, \frac{\cos(nx)}{\sqrt{\pi}}, \frac{\sin(nx)}{\sqrt{\pi}} \right\}$ čini ortonormiran skup u Hilbertovom prostoru $L^2[a, b]$ koji je ujedno i ortonormirana baza u tom skupu. Drugim riječima svaka funkcija $f \in L^2[a, b]$ ima svoj Fourierov rastav po toj bazi:

$$f(x) = a_0 \frac{1}{\sqrt{2\pi}} + \sum_{n=1}^{\infty} \left(a_n \frac{\cos(nx)}{\sqrt{\pi}} + b_n \frac{\sin(nx)}{\sqrt{\pi}} \right).$$

$$\left\langle f(x), \frac{1}{\sqrt{2\pi}} \right\rangle = a_0 \left\langle \frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{2\pi}} \right\rangle + \sum_{n=1}^{\infty} \left(a_n \left\langle \frac{\cos(nx)}{\sqrt{\pi}}, \frac{1}{\sqrt{2\pi}} \right\rangle + b_n \left\langle \frac{\sin(nx)}{\sqrt{\pi}}, \frac{1}{\sqrt{2\pi}} \right\rangle \right) \text{ tj. zbog}$$

$$\text{ortogonalnosti vrijedi: } a_0 = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(x) dx$$

$$\left\langle f(x), \frac{\cos(mx)}{\sqrt{\pi}} \right\rangle = a_0 \left\langle \frac{1}{\sqrt{2\pi}}, \frac{\cos(mx)}{\sqrt{\pi}} \right\rangle + \sum_{n=1}^{\infty} \left(a_n \left\langle \frac{\cos(nx)}{\sqrt{\pi}}, \frac{\cos(mx)}{\sqrt{\pi}} \right\rangle + b_n \left\langle \frac{\sin(nx)}{\sqrt{\pi}}, \frac{\cos(mx)}{\sqrt{\pi}} \right\rangle \right)$$

$$\text{tj. zbog ortogonalnosti vrijedi: } a_n = \left\langle f(x), \frac{\cos(mx)}{\sqrt{\pi}} \right\rangle = \int_0^{2\pi} f(x) \frac{\cos(mx)}{\sqrt{\pi}} dx$$

$$\left\langle f(x), \frac{\sin(mx)}{\sqrt{\pi}} \right\rangle = a_0 \left\langle \frac{1}{\sqrt{2\pi}}, \frac{\sin(mx)}{\sqrt{\pi}} \right\rangle + \sum_{n=1}^{\infty} \left(a_n \left\langle \frac{\cos(nx)}{\sqrt{\pi}}, \frac{\sin(mx)}{\sqrt{\pi}} \right\rangle + b_n \left\langle \frac{\sin(nx)}{\sqrt{\pi}}, \frac{\sin(mx)}{\sqrt{\pi}} \right\rangle \right)$$

$$\text{tj. zbog ortogonalnosti vrijedi: } b_n = \left\langle f(x), \frac{\sin(mx)}{\sqrt{\pi}} \right\rangle = \int_0^{2\pi} f(x) \frac{\sin(mx)}{\sqrt{\pi}} dx.$$

1.2. Fourierov red

1.2.1. Fourierov red u ravnini

Definirajmo polinome u \mathfrak{R}^2 kao skup $\{(x, y) : x, y \in \mathfrak{R}\}$. Zatim promatrajmo harmoničke polinome za koje vrijedi $\Delta_2 p(x, y) = 0$, gdje Δ_2 predstavlja dvodimenzionalni Laplaceov operator: $\Delta_2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$. Npr. neka je $f(x, y) = x^2 - y^2$ onda je $\Delta_2 f(x, y) = 0$ i taj polinom nazivamo harmoničkim.

Za neku funkciju $f(x, y)$ kažemo da je homogeni harmonički poligon stupnja n ako vrijedi: $f(tx, ty) = t^n f(x, y)$ za sve pozitivne realne brojeve t . Npr. funkcija $f(x, y) = x^2 - y^2$ je homogeni harmonički poligon stupnja 2. Homogene funkcije se lako prikazuju u polarnim koordinatama $\Delta_2 q_n(x, y) = -n^2 q_n(x, y)$. Funkcija $f_n(x, y)$ koja je homogena sa stupnjem n može se zapisati u polarnim koordinatama ovako: $f_n(x, y) = r^n q_n(\theta)$. U polarnim koordinatama Laplaceov operator postaje:

$$\Delta_2 = \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r}. \quad \text{Homogeni harmonički polinom stupnja}$$

$n, (n \geq 0)$ zadovoljava jednadžbu: $\Delta_2 q_n(x, y) = 0$.

$$p_n(x, y) = r^n q_n(\theta)$$

$$\frac{\partial}{\partial \theta} = r^n \frac{\partial^2 q_n(\theta)}{\partial \theta^2}, \frac{\partial}{\partial r} = nr^{n-1} q_n(\theta), \frac{\partial^2}{\partial r^2} = n(n-1)r^{n-2} q_n(\theta)$$

$$\Delta_2 p_n(x, y) = \frac{1}{r^2} r^n \frac{\partial^2 q_n(\theta)}{\partial \theta^2} + n(n-1)r^{n-2} q_n(\theta) + \frac{1}{r} nr^{n-1} q_n(\theta)$$

$$\Delta_2 p_n(x, y) = r^{n-2} \left(\frac{\partial^2 q_n(\theta)}{\partial \theta^2} + n(n-1)q_n(\theta) + nq_n(\theta) \right) = r^{n-2} \left(\frac{\partial^2 q_n(\theta)}{\partial \theta^2} + n^2 q_n(\theta) \right) = 0$$

$$\Rightarrow \frac{\partial^2 q_n(\theta)}{\partial \theta^2} + n^2 q_n(\theta) = 0 \Rightarrow \frac{\partial^2 q_n(\theta)}{\partial \theta^2} = -n^2 q_n(\theta)$$

$$\Delta_{2s} = \frac{\partial^2}{\partial \theta^2}, \Delta_{2s} q_n(\theta) = -n^2 q_n(\theta)$$

Drugim riječima $q_n(\theta)$ mora biti svojstvena funkcija od kružnog Laplaceovog operatora Δ_{2s} .

Kružni Laplaceov operator Δ_{2s} je simetričan: $\langle \Delta_{2s} f(\theta), g(\theta) \rangle = \langle f(\theta), \Delta_{2s} g(\theta) \rangle$.

Simetrični operatori imaju svojstvo da svojstvene funkcije sa različitim svojstvenim vrijednostima su ortogonalne.

$$\begin{aligned} \Delta_{2s} q_n(\theta) &= -n^2 q_n(\theta) \\ \langle \Delta_{2s} q_n(\theta), q_m(\theta) \rangle &= -n^2 \langle q_n(\theta), q_m(\theta) \rangle = \langle q_n(\theta), \Delta_{2s} q_m(\theta) \rangle = -m^2 \langle q_n(\theta), q_m(\theta) \rangle \end{aligned}$$

Ako je $n \neq m \Rightarrow \langle q_n(\theta), q_m(\theta) \rangle = 0$.

Postavlja se pitanje različitih svojstvenih funkcija $q_n(\theta)$ postoji za neki n . Funkcija

$q_n(\theta)$ zadovoljava sljedeću diferencijalnu jednačinu: $\frac{\partial^2 q_n(\theta)}{\partial \theta^2} = -n^2 q_n(\theta)$.

Sva rješenja ove diferencijalne jednačine zadovoljavaju sljedeći oblik: $q_n(\theta) = A \cos(n\theta) + B \sin(n\theta)$ ili ekvivalentni oblik: $q_n(\theta) = C e^{in\theta} + D e^{-in\theta}$. Skup $\{e^{in\theta}, e^{-in\theta}\}$ predstavlja ortogonalnu bazu svojstvenog prostora.

Teorem: Svaki polinom stupnja n ograničen na kružnicu može se prikazati kao suma homogenih harmoničkih poligona stupnja do najviše n .

Polinomi ograničeni na kružnicu mogu predstaviti proizvoljnu kontinuiranu funkciju $\in L^2$ prostora. Skup eksponencijala $\{e^{in\theta}\}_{n \in \mathbb{Z}}$ razapinje $L^2(S^1)$ prostor.

1.2.2. Fourierov red na sferi

U \mathbb{R}^3 prostoru, $\{(x, y, z) : x, y, z \in \mathbb{R}^3\}$, Laplaceov operator je definiran ovako:

$$\Delta_3 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

U sfernim koordinatama:

$$\{(r, \theta, \phi), r \in \mathbb{R}_+, \phi \in [0, 2\pi), \theta \in [0, \pi)\} (x = r \sin \theta \cos \phi, y = r \sin \theta \sin \phi, z = r \cos \theta)$$

Laplaceov operator je definiran ovako:

$$\Delta_3 = \csc^2 \phi \frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial \phi^2} + \cot \phi \frac{\partial}{\partial \phi} + 2r \frac{\partial}{\partial r} + r^2 \frac{\partial^2}{\partial r^2}.$$

Neka je $p_n(x, y, z)$ homogeni harmonički polinom stupnja n onda vrijedi:

$$p_n(x, y, z) = r^n q_n(\theta, \phi) \text{ i } \Delta_3 p_n = 0.$$

$$\left(\csc^2 \phi \frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial \phi^2} + \cot \phi \frac{\partial}{\partial \phi} \right) r^n q_n(\theta, \phi) + 2nr^n q_n(\theta, \phi) + r^2 n(n-1) r^{n-2} q_n(\theta, \phi) = 0,$$

$$\Delta_{3s} = \csc^2 \phi \frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial \phi^2} + \cot \phi \frac{\partial}{\partial \phi},$$

$$\Delta_{3s} r^n q_n(\theta, \phi) + 2nr^n q_n(\theta, \phi) + n(n-1)r^n q_n(\theta, \phi) = 0,$$

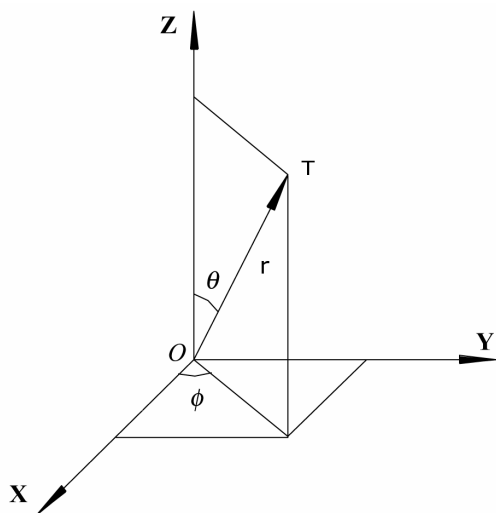
$$r^n (\Delta_{3s} q_n(\theta, \phi) + n(n+1)q_n(\theta, \phi)) \Rightarrow \Delta_{3s} q_n(\theta, \phi) = -n(n+1)q_n(\theta, \phi)$$

Iz ove relacije: $\Delta_{3s} q_n(\theta, \phi) = -n(n+1)q_n(\theta, \phi)$ zaključujemo da je $q_n(\theta, \phi)$ svojstvena funkcija cirkularnog sfernog Laplacijanovog operatora Δ_{3s} .

Sferni Laplaceov operator Δ_{3s} je simetričan što znači da su svojstvene funkcije $q_n(\theta, \phi)$ od različitih svojstvenih vrijednosti ortogonalne.

Za fiksirani n svojstveni prostor Λ_n je skup $\left\{ P_n^m(\cos \phi) e^{im\theta} / \sqrt{2\pi} \right\}_{-n \leq m \leq n}$.

1.2.3. Sferni koordinatni sustav



Neka je r udaljenost točke $T(x, y, z)$ od ishodišta koordinatnog sustava, ϕ kut između pozitivne poluosi x i projekcije radij-vektora točke T na x - y ravninu, te θ kut između pozitivne poluosi z i radij-vektora točke T . $r \in \mathbb{R}_+, \phi \in [0, 2\pi), \theta \in [0, \pi)$.

Slika 2. Prikaz sfernog koordinatnog sustava

Veza između sfernih i pravokutnih koordinata su dane sljedećim izrazima:

$$x = r \sin \theta \cos \phi,$$

$$y = r \sin \theta \sin \phi,$$

$$z = r \cos \theta,$$

$$J = \det \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial \phi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial \phi} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial \phi} \end{pmatrix} = \det \begin{pmatrix} \sin \theta \cos \phi & r \cos \theta \cos \phi & -r \sin \theta \sin \phi \\ \sin \theta \sin \phi & r \cos \theta \sin \phi & r \sin \theta \cos \phi \\ \cos \theta & -r \sin \theta & 0 \end{pmatrix}$$

$$J = \cos \theta r^2 \sin \theta \cos \theta (\cos^2 \phi + \sin^2 \phi) + r \sin \theta r \sin^2 \theta (\cos^2 \phi + \sin^2 \phi)$$

$$J = \cos \theta r^2 \sin \theta \cos \theta + r \sin \theta r \sin^2 \theta = r \sin \theta (\cos^2 \theta + \sin^2 \theta) = r \sin \theta$$

Zamjena varijabli u trostrukom integralu se provodi na sljedeći način:

$$\begin{aligned} \iiint_V f(x, y, z) dx dy dz &= \iiint_{V^*} f(r \sin \theta \cos \phi, r \sin \theta \sin \phi, r \cos \theta) |J| dr d\theta d\phi \\ &= \iiint_{V^*} f(r \sin \theta \cos \phi, r \sin \theta \sin \phi, r \cos \theta) r^2 \sin \theta dr d\theta d\phi \end{aligned}$$

Zapravo, element volumena $dV = dx dy dz$ prelazi u $dV = r^2 \sin \theta dr d\theta d\phi$

1.3. Pridruženi Legendrovi polinomi

Simbol P_l^m reprezentira pridruženi Legendrove polinome koji su definirani na domeni $[-1,1]$ a kodomena im je skup \mathbb{R} . Eksplicitna definicija:

$$P_l^m = \frac{(-1)^m}{2^l l!} \sqrt{(1-x^2)^m} \frac{d^{l+m}}{dx^{l+m}} (x^2-1)^l, l \in \mathbb{N}_0, m \in [0, l]$$

$$\int_{-1}^1 P_l^m(x) P_k^m(x) dx = \delta_{lk} \frac{2(l+m)!}{(2l+1)(l-m)!}$$

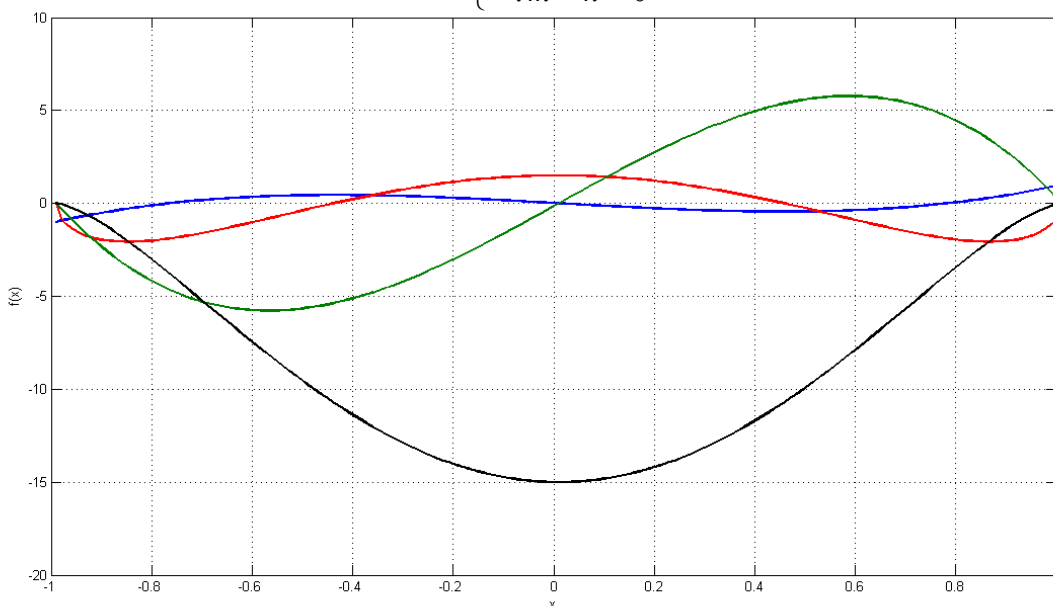
$$\int_{-1}^1 \frac{P_l^m P_n^m}{1-x^2} dx = \begin{cases} 0 : m \neq n \\ \frac{(l+m)!}{m(l-m)!} : m = n \neq 0 \\ \infty : m = n = 0 \end{cases}$$

Polinomi su ortogonalni za iste indekse m :

$$\int_{-1}^1 P_l^m(x) P_k^m(x) dx = \delta_{lk} \frac{2(l+m)!}{(2l+1)(l-m)!}$$

Također vrijedi ortogonalnost za iste l :

$$\int_{-1}^1 \frac{P_l^m P_n^m}{1-x^2} dx = \begin{cases} 0 : m \neq n \\ \frac{(l+m)!}{m(l-m)!} : m = n \neq 0 \\ \infty : m = n = 0 \end{cases}$$



Slika 3. Prikaz nekoliko osnovnih Legendrovih polinoma

Prvih nekoliko pridruženih Legendrovih polinoma:

l	m	P_l^m	l	m	P_l^m
0	0	1	2	2	$3(1-x^2)$
1	0	x	3	0	$1/2(5x^3-3x)$
1	1	$-\sqrt{1-x^2}$	3	1	$3/2(1-5x^2)\sqrt{1-x^2}$
2	0	$1/2(3x^2-1)$	3	2	$15x(1-x^2)$
2	1	$-3x\sqrt{1-x^2}$	3	3	$-15\sqrt{(1-x^2)^3}$

Za $m=0, l \in \mathbb{N}$ pridruženi Legendrovi polinomi postaju Legendrovi polinomi za koje

vrijedi sljedeće svojstvo ortogonalnosti: $\int_{-1}^1 P_l^0 P_k^0 dx = 0, l \neq k$.

Lema: Neka imamo skup ortogonalnih polinoma $\{p_0(x), p_1(x), p_2(x), \dots\}$, svaki polinom $S(x)$ stupnja n se može razviti preko polinoma linearne kombinacije polinoma: $p_0(x), p_1(x), \dots, p_n(x)$. Postoje koeficijenti a_0, a_1, \dots, a_n takvi da vrijedi :

$$S(x) = \sum_{i=0}^n a_i p_i(x).$$

Dokaz: Matematičkom indukcijom, izaberi koeficijent a_n , tako da koeficijent uz dio x^n polinoma $S(x)$ odgovara dijelu uz dio x^n polinoma $a_n p_n(x)$. Onda polinom $S(x) - a_n p_n(x)$ je stupnja $n-1$. Postupak nastaviti dalje. Koeficijente razvoja dobivamo lako:

$$S(x) = \sum_{i=0}^n a_i p_i(x) \Rightarrow \langle S(x), p_k(x) \rangle = \sum_{i=0}^n a_i \langle p_i(x), p_k(x) \rangle$$

$$\int S(x) p_k(x) dx = a_k \int p_k^2(x) dx \Rightarrow a_k = \frac{\int S(x) p_k(x) dx}{\int p_k^2(x) dx}$$

1.4. Sferni harmonici

Sferni harmonici mapiraju sferne koordinate (θ, ϕ) u skalarne vrijednosti. Sferni harmonici se obično vizualiziraju teksturama na sferi čije vrijednosti odgovaraju vrijednostima sfernih harmonika u nekoj točki na sferi. Integracija na sferi odgovora zbrajanju infinitezimalno malenih površina sfere. $\int_0^{2\pi} \int_0^\pi f(\theta, \phi) \sin(\theta) d\theta d\phi$.

Pridruženi Legendrovi polinomi mogu prikazati proizvoljnu funkciju na intervalu $[-1, 1]$ kao beskonačnu težinsku sumu polinoma.

Definicija: Sferni harmonici reda l i indeksom m definirani su ovako:

$$Y_l^m(\theta, \phi) = N_l^{|m|} P_l^{|m|}(\cos \theta) e^{im\phi}, \quad l \in \mathbb{N}_0, -l \leq m \leq l,$$

gdje $N_l^{|m|}$ predstavlja normalizacijski faktor, $P_l^{|m|}$ pridružene Legendrove polinome i $e^{im\phi}$ funkcijski dio vezan za ϕ ovisnost.

Gledajući definiciju sfernih harmonika jasno je da proizvoljno simetričnu funkciju oko z osi možemo prikazati preko pridruženih Legendrovih polinoma $P_l^{|m|}(\cos \theta)$ jer ta funkcija zbog simetrije nema ovisnosti o ϕ kutu. Zbog domene pridruženih Legendrovih polinoma koristimo preslikavanje vrijednosti $f(\theta) = \cos \theta$. Ali također trebamo mehanizam za pružanje ortogonalnosti za slučaj nesimetričnih funkcija. To postizemo kombiniranjem pridruženih Legendrovih polinoma za θ ovisnost i sinusa i kosinusa za ϕ ovisnost.

$$Y_l^m(\theta, \phi) = N_l^{|m|} P_l^{|m|}(\cos \theta) (\cos(m\phi) + i \sin(m\phi))$$

1.4.1. Ortonormirana baza

$$Y_l^m(\theta, \phi) = N_l^{|m|} P_l^{|m|}(\cos \theta) (\cos(m\phi) + i \sin(m\phi))$$

Normalizacijski faktor $N_l^{|m|}$ se može izvesti iz sljedeće jednadžbe za uvjet ortogonalnosti:

$$\int_0^{2\pi} \int_0^\pi Y_l^m(\theta, \phi) \overline{Y_l^{m'}(\theta, \phi)} \sin(\theta) d\theta d\phi = \delta_{mm'} \delta_{ll'}$$

$$\int_0^{2\pi} \int_{-1}^1 Y_l^m(\theta, \phi) \overline{Y_l^{m'}(\theta, \phi)} d(\cos \theta) d\phi,$$

$$\int_0^{2\pi} \int_{-1}^1 N_l^{|m|} N_{l'}^{|m'|} P_l^{|m|}(\cos \theta) \overline{P_{l'}^{|m'|}(\cos \theta)} e^{im\phi} \overline{e^{im'\phi}} d(\cos \theta) d\phi,$$

$$N_l^{|m|} N_{l'}^{|m'|} \int_{-1}^1 P_l^{|m|}(\cos \theta) \overline{P_{l'}^{|m'|}(\cos \theta)} d(\cos \theta) \int_0^{2\pi} e^{im\phi} \overline{e^{im'\phi}} d\phi$$

Rješavanjem integrala ovisnog o ϕ dijelu dobivamo:

$$\int_0^{2\pi} e^{im\phi} \overline{e^{im'\phi}} d\phi = \int_0^{2\pi} (\cos(m\phi) + i \sin(m\phi)) (\cos(m'\phi) - i \sin(m'\phi)) d\phi$$

$$= \int_0^{2\pi} \cos(m\phi) \cos(m'\phi) d\phi - i \int_0^{2\pi} \cos(m\phi) \sin(m'\phi) d\phi + i \int_0^{2\pi} \sin(m\phi) \cos(m'\phi) d\phi + \int_0^{2\pi} \sin(m\phi) \sin(m'\phi) d\phi$$

$$= \pi \delta_{mm'} - i \cdot 0 + i \cdot 0 + \pi \delta_{mm'} = 2\pi \delta_{mm'}$$

Rješavanjem integrala ovisnog o θ dijelu tj. korištenja uvjeta ortogonalnosti pridruženih Legendrovih polinoma dobivamo:

$$\int_{-1}^1 P_l^{|m|}(\cos \theta) \overline{P_{l'}^{|m'|}(\cos \theta)} d(\cos \theta) = \frac{2}{2l+1} \frac{(l+m)!}{(l-m)!} \delta_{ll'}$$

$$N_l^{|m|} N_{l'}^{|m'|} \frac{2}{2l+1} \frac{(l+m)!}{(l-m)!} \delta_{ll'} 2\pi \delta_{mm'} = \delta_{mm'} \delta_{ll'}$$

$$\Rightarrow N_l^{|m|} = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}}$$

1.4.2. Realni sferni harmonici

Realni sferni harmonici:

$$Y_l^m(\theta, \phi) = \mathcal{G}_{lm}(\mu) Z_m(\phi); \mu = \cos \theta$$

$$\mathcal{G}_{lm}(\mu) = \sqrt{\frac{2l+1}{2} \frac{(l-m)!}{(l+m)!}} P_l^m(\mu)$$

$$Z_m(\phi) = \frac{1}{\sqrt{2\pi}} e^{im\phi}$$

$$\int_{-1}^1 \mathcal{G}_{lm}(\mu) \mathcal{G}_{l'm'}(\mu) d\mu = \delta_{ll'}$$

$$\int_0^{2\pi} Z_m(\phi) \overline{Z_{m'}(\phi)} d\phi = \delta_{mm'}$$

Realne sferne harmonike dobivamo linearnom kombinacijom kompleksnih sfernih harmonika:

$$y^m_l(\theta, \phi) = \begin{cases} (Y^m_l(\theta, \phi) + Y^{m*}_l(\theta, \phi))/\sqrt{2} : m > 0 \\ Y^0_l(\theta, \phi) : m = 0 \\ -i(Y^{\bar{m}}_l(\theta, \phi) - Y^{\bar{m}*}_l(\theta, \phi))/\sqrt{2} : m < 0 \end{cases}$$

$$y^m_l(\theta, \phi) = \begin{cases} \mathcal{G}_{lm}(\mu)(\cos m\phi)/\sqrt{\pi} : m > 0 \\ \mathcal{G}_{lm}(\mu)(\cos m\phi)/\sqrt{2\pi} : m = 0 \\ \mathcal{G}_{l\bar{m}}(\mu)(\sin \bar{m}\phi)/\sqrt{\pi} : m < 0 \end{cases}$$

1.4.3. Vizualizacija sfernih harmonika

Sferni harmonici predstavljaju funkcije koje preslikavaju vrijednosti sa jedinične sfere na

skup realnih vrijednosti: $Y^m_l(\theta, \phi) : S^2 \rightarrow \mathfrak{R}$.

$$Y^m_l(\theta, \phi) = N^{|m|}_l P^{|m|}_l(\cos \theta)(\cos(m\phi) + i \sin(m\phi));$$

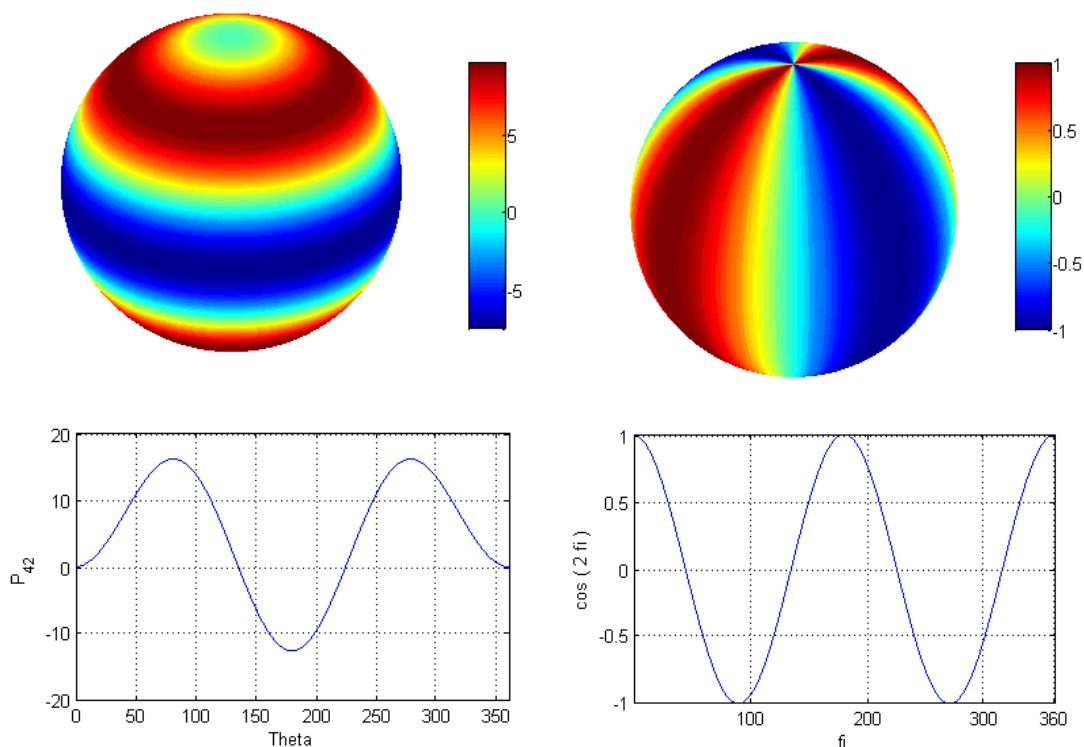
Sferni harmonici se obično vizualiziraju na jediničnoj sferi u trodimenzionalnom prostoru čiji intenzitet tj. boja na sferi predstavlja realnu vrijednost pripadne točke. Funkcionalna dekompozicija sfernih harmonika omogućava prikaz sfernih harmonika kao umnožak funkcije ovisne o kutu elevacije θ : $P^{|m|}_l(\cos \theta)$ i funkcije ovisne o azimutu ϕ : $e^{im\phi}$.

Primjer: Potrebno je vizualizirati funkciju $Y^2_4(\theta, \phi)$. Za potrebe vizualizacije možemo koristiti realni ili imaginarni dio funkcije $Y^2_4(\theta, \phi)$.

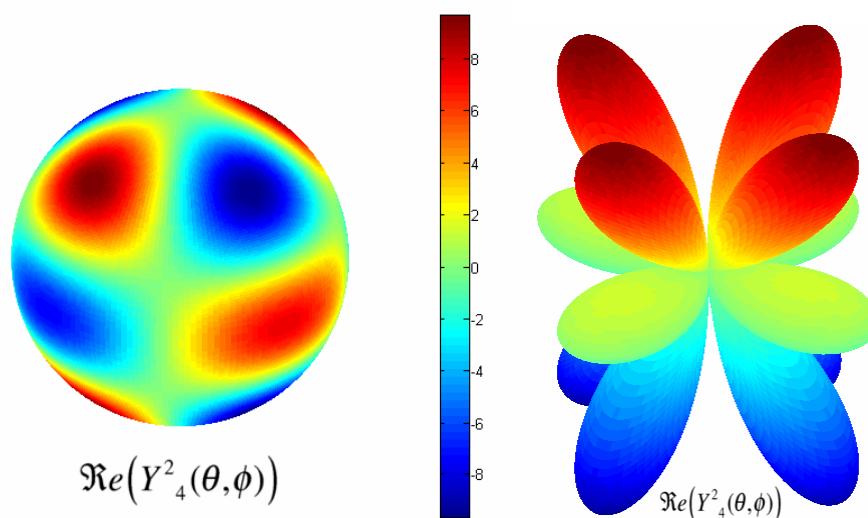
$$\Re(Y^2_4(\theta, \phi)) = N^{|2|}_4 P^2_4(\cos \theta) \cos 2\phi$$

$$y^2_4(\theta, \phi) = N^{|2|}_4 \left(\frac{25}{2} (7 \cos^2 \theta - 1) \sin^2 \theta \right) \cos 2\phi$$

$$N^{|2|}_4 = \sqrt{\frac{9 (4-2)!}{4\pi (4+2)!}}$$

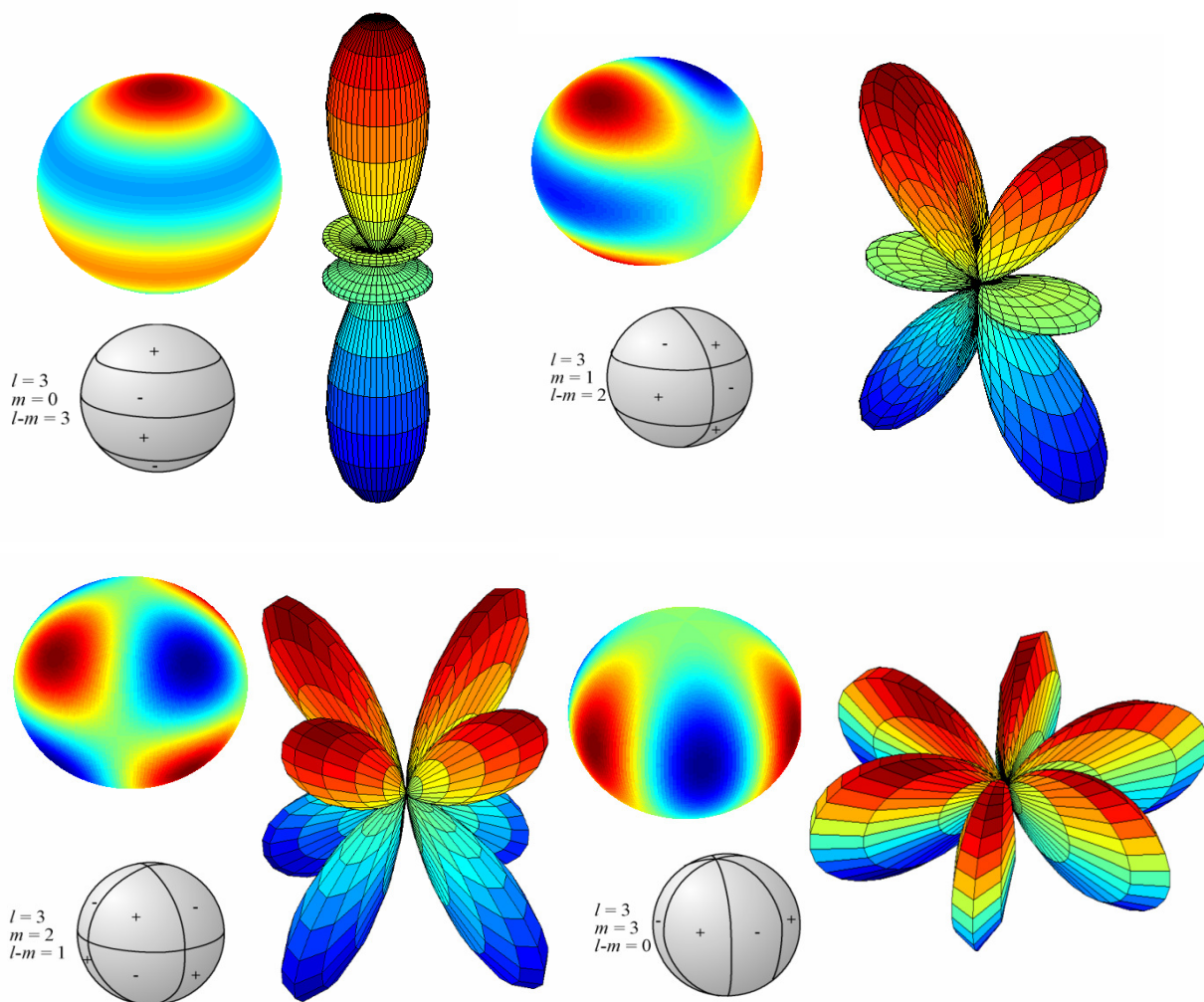


Slika 4. Funkcionalna dekompozicija sfernog harmonika na azimutni i elevacijski dio
 Kombinirajući funkciju ovisne o azimutu $\cos 2\phi$ i funkciju ovisnu o elevaciji $P_4^2(\cos \theta)$ zajedno sa normalizacijskim faktorom $N_4^{[2]}$ dobivamo realnu dio sfernog harmonika $\Re(Y_4^2(\theta, \phi))$ koji je prikazan na sljedećoj slici. Osim sfernog prikaza harmonika moguće je koristiti orbitalni prikaz harmonika gdje se nekoj vrijednosti na sferi (θ, ϕ) umjesto intenziteta na sferi dodaje radijalni dio r .



Slika 5. Prikaz sfernog harmonika na sferi i u orbitalnom prikazu

Sferni harmonici (Laplaceovi) $y_m^l(\theta, \phi)$ se mogu uspješno vizualizirati preko nodalnih linija. Nodalne linije predstavljaju skup točaka na sferi gdje sferni harmonik iščezava. $y_m^l(\theta, \phi) = 0$. Nodalne linije sfernih harmonika se sastoje od skupova kružnica koje su složene horizontalno ili vertikalno. Broj nodalnih linija u horizontalnom smjeru odgovara broju nultočaka pridruženih Legendrovih polinoma tj. broju $l - |m|$, dok broj nodalnih linija u vertikalnom smjeru odgovara broju nultočaka trigonometrijskih funkcija sinus i kosinus tj. broju $2|m|$. Kada je parametar m sfernih harmonika $y_m^l(\theta, \phi)$ jednak nuli onda govorimo o zonalnih sfernim harmonicima koji su cirkularno simetrični jer nemaju ovisnost o azimut kutu ϕ . Kada su parametri l i m jednaki $l = |m|$ onda govorimo o sektoralnih dok se ostali sferni harmonici zovu teseralni sferni harmonici.



Slika 6. Prikaz nekoliko tipova sfernih harmonika na sferi, orbitalnom prikazu i pomoću nodalnih linija

1.5. Radijalne funkcije

Kako bi mogli opisati proizvoljne 3D objekte u prostoru proširili smo skup 2D baznih sfernih harmonika sa 1D baznih radijalnim funkcijama.

$$F_{nlm} = R_n(r)y_l^m(\theta, \phi)$$

Stoga površina proteina dodajemo volumen i dobivamo trodimenzionalne ljuske proteina koje se mogu razviti u red 3D baznim funkcijama F_{nlm} . Odabir pripadnih radijalnih funkcija $R_n(r)$ utječe na svojstva uspješnosti rastavljanja funkcija ljuske u sumu baznih funkcija. Kako bi radijalne funkcije uspješno prikazale ljuske proteina potrebno je postaviti neke zahtjeve na te funkcije:

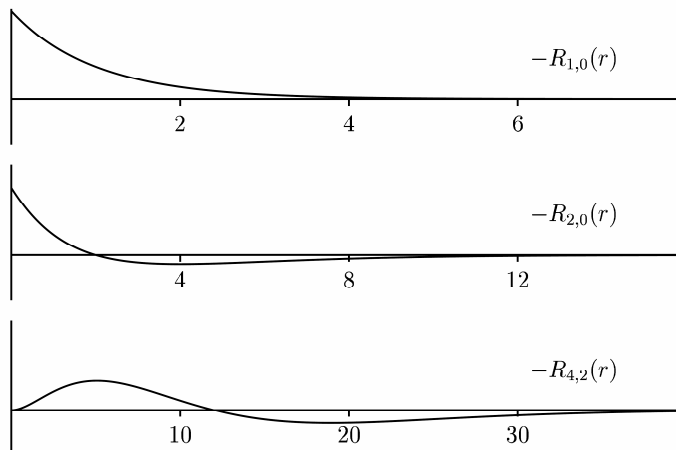
- Ortonormiranost: radijalne funkcije moraju tvoriti skup ortonormiranih funkcija
- Eksponecijalno opadanje: radijalne funkcije bi trebale imati svojstvo eksponecijalnog opadanja vrijednosti funkcije s obzirom na udaljenost od stvarne površine proteina
- Skaliranost: srednja vrijednost radijusa svih radijalnih funkcija bi se trebala skalirati kako bi odgovarala srednjem radijusu proteina (otprilike 15 Å ili 20 Å)

Pokazano je da se Laguerrovi polinomi $L_{nl}(r)$ pojavljuju u rješenju Schrödingerove jednadžbe $\nabla^2\psi(x) + V(x)\psi(x) = E\psi(x)$, gdje $V(x)$ predstavlja elektrostatsku potencijalnu energiju i E energiju sustava. Rješenje ove jednadžbe je: $\psi(x) = R_{nl}(r)y_l^m(\theta, \phi), n > l \geq |m|$, gdje cijeli kvantni brojevi n, l, m predstavljaju dopuštenja stanja elektrona u atomu. Radijalna funkcija je predstavljena ovako:

$R_{nl}(r) = N_{nl}e^{-\rho/2}\rho^l L_{n-l}^{2l+1}(\rho)$, gdje $L_{n-l}^{2l+1}(\rho)$ predstavljaju Laguerrove polinome i parametar ρ skaliranu udaljenost. Valna funkcija $\psi(x)$ za vodikov atom se sastoji od eksponecijalne funkcije pada pomnožene sa Laguerrovim polinomom i

sfernim harmonikom. N_{nl} je normalizacijski faktor zbog kojeg su Laguerrove

funkcije ortogonalne s obzirom na polarni integral: $\int_0^{\infty} R_{nl}(r)R_{n'l'}(r)r^2 dr = \delta_{nn'}$.



Slika 7. Prikaz nekoliko funkcija rješenja radijalnog dijela Schrödingerove jednadžbe. Slika iz [1].

Neki primjeri grafova radijalnih funkcija $R_{nl}(r) = N_{nl}e^{-\rho/2} \rho^l L_{n-l}^{2l+1}(\rho)$ su prikazani na slici. Zadovoljen je uvjet ortonormiranosti ali eksponencijalni pad je prespor i prostire se duž velikog prostora, pogotovo za veće vrijednosti parametra n . Također se može pokazati da srednja vrijednost kod ovih radijalnih funkcija se povećava kao n^2 što nije dobro za reprezentaciju površine proteina. Ova se funkcija treba skalirati tako da joj srednja vrijednost radijusa bude puno bliža srednjem radijusu proteina i da ne ovisi o parametrima n, l .

Kako bi smo dobro skalirali prethodne radijalne funkcije prvo ćemo pogledati uvjet ortonormiranosti za Laguerrove polinome:

$$\int_0^{\infty} e^{-r} r^q L_p^q(r) L_{p'}^q(r) dr = \delta_{pp'}; p, p' \geq q \geq 0$$

Postavimo li uvjet $q=2l$ dobivamo novi polarni integral:

$$p' = p \Rightarrow \int_0^{\infty} e^{-r} r^{2l} L_p^{2l+2}(r) L_p^{2l+2}(r) r^2 dr = \int_0^{\infty} e^{-r} r^{2l} \left(L_p^{2l+2}(r) \right)^2 r^2 dr = \delta_{pp'}$$

također ako:

$$n = n', l = l' \Rightarrow$$

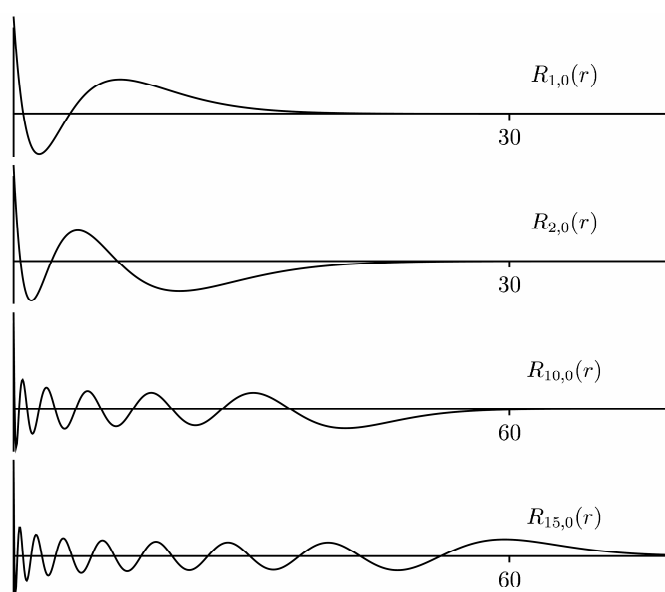
$$\int_0^{\infty} R_{nl}(r) R_{n'l'}(r) r^2 dr = \int_0^{\infty} (R_{nl}(r))^2 r^2 dr = \delta_{nn'}$$

onda vrijedi:

$$\int_0^{\infty} e^{-r} r^{2l} (L_p^{2l+2}(r))^2 r^2 dr = \int_0^{\infty} (R_{nl}(r))^2 r^2 dr \Rightarrow R_{nl}(r) = e^{-r/2} r^l L_p^{2l+2}(r); l < n$$

zbog uvjeta: $q \leq p \Rightarrow p = n + l + 1$;

$$R_{nl}(r) = e^{-r/2} r^l L_{n+l+1}^{2l}(r)$$



Slika 8. Prikaz probnih radijalnih funkcija $L_{n+l+1}^{2l+2}(r)$. Slika iz [1].

Prikaz ponašanja novih radijalnih funkcija je vidljiv na prethodnoj slici. Vidimo da funkcije višeg reda imaju više nultočaka ali konvergencija se događa tek za velike vrijednosti radijusa. Ove funkcije imaju preveliko rasipanje i potrebno ih je prilagoditi tako da se većina nultočaka nalazi blizu površine proteina i eksponencijalni pad koji počinje nekoliko angstrema od površine. Jedan način je

promjena varijable ρ na ovaj način : $\rho = \frac{1}{3} \left(\frac{r}{r_0} \right)^3$ gdje parametar r_0 omogućuje

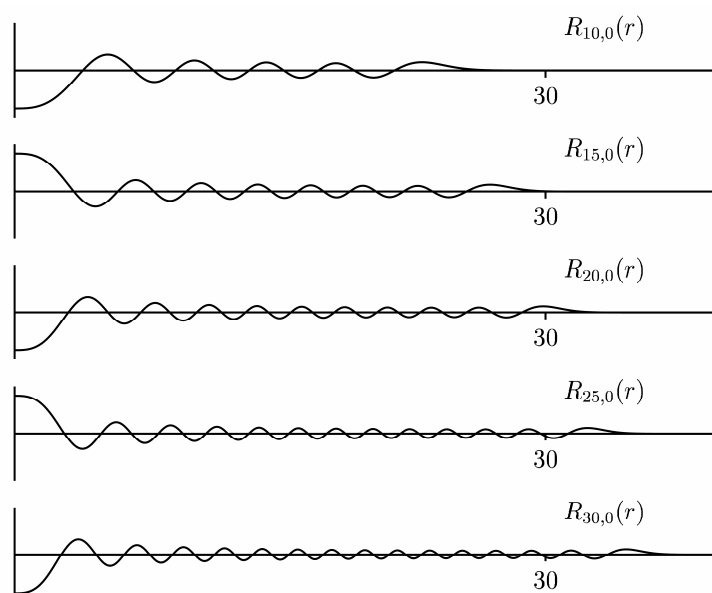
skaliranje. Normalizirane skaliranje radijalne funkcije sada izgledaju ovako:

$$R_{nl}(r; r_0) = r_0^{-3/2} e^{-\rho/2} \rho^l L_{n+l-1}^{2l}(\rho); l < n .$$

Potrebno je procijeniti parametar r_0 tako da srednja vrijednost od r za pripadnu radijalnu funkciju \bar{r}_{nl} bude jednaka srednjem radijusu proteina. Iz kvante fizike je poznato da vjerojatnost da se elektron pronađe na radijusu r , na kvantnim stanjima n, l : $P_{nl}(r) = R_{nl}(r)^2 r^2$. Stoga očekivana vrijednost slučajne varijable r se računa ovako:

$$\bar{r}_{nl} = \int_0^{\infty} R_{nl}(r; r_0) r R_{nl}(r; r_0) r^2 dr$$

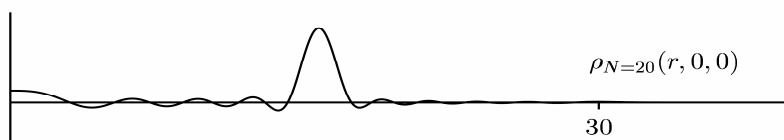
Ovaj integral se može riješiti numeričkom integracijom za razne parametre n i r_0 . Pokazano je da je optimalno izabrati parametar $r_0 = 5$ jer tada srednja vrijednost radijalnih funkcija iznosi: $r_{10,0} = 17.3, r_{15,0} = 19.9, r_{20,0} = 22.0, r_{25,0} = 23.7, r_{30,0} = 25.2$ što odgovara srednjem radijusu proteina koji iznosi otprilike od 15 Å do 20 Å.



Slika 9. Prikaz skaliranih radijalnih funkcija $L_{n+l-1}^{2l}(\rho)$ sa parametrom $r_0 = 5$. Slika iz [1].

Grafovi skaliranih radijalnih funkcija pokazuju da su nultočke ravnomjerno raspoređene po radijusima. Većina energije proteina će se nalaziti blizu njegove površine što je željeno ponašanje radijalnih funkcija. Zašto želimo očekivanu vrijednost od r da bude na otprilike 15 Å za sve vrijednosti n, l ? Zato što opisujemo površinu proteina čija se površina određene debljine nalazi na prosječno 15 Å sa sumom baznih funkcija čije radijalne funkcije također imaju očekivanu vrijednost na 15 Å.

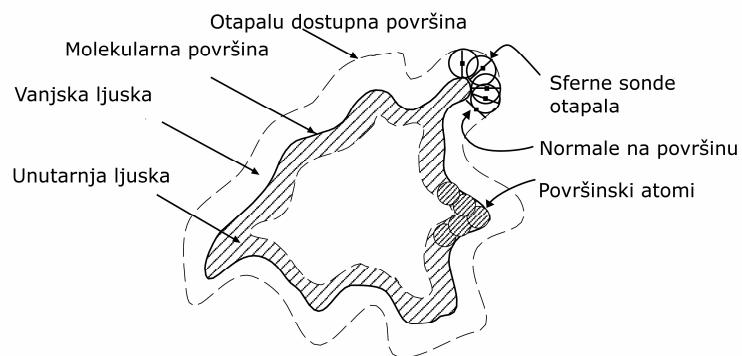
U suprotnom slučaju kada bismo imali radijalne funkcije čija očekivana vrijednost se nalazi na puno većim vrijednosti onda bih trebali uzeti puno više baznih funkcija kako bi dobro opisali površinu. Kako bismo napravili test kako radijalnih funkcija možemo reprezentirati protein uzet ćemo kuglu radijusa 15 Å i debljinom ljuske 1.4 Å.



Slika 10. Rekonstrukcija ljuske kugle debljine 1.4 Å i radijusa 15 Å. Slika iz [1].

2. Ljuske proteina

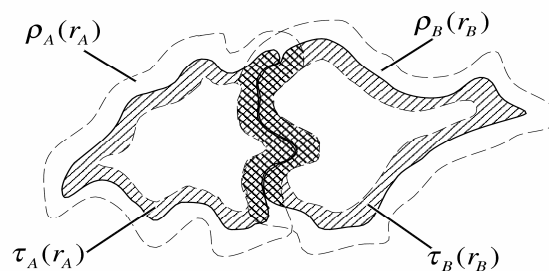
Vanjska ljuska proteina definira se kao volumen omeđen plohamo molekularne površine proteina i otapalu dostupne površine. Otapalu dostupna površina dobiva se prolaskom sferne sonde otapala po molekularnoj površini. Unutarnja ljuska proteina definira se kao unija van der Waalsovih atoma koji se nalaze unutar molekularne površine.



Slika 11. Shematski prikaz vanjske i unutarnje ljuske proteina. Slika iz [1].

Komplementarnost proteina definira se kao integral preklapanja unutarnjih i vanjskih ljusaka proteina i penaliziranog integral preklapanja unutarnjih ljusaka proteina.

$$S = \int \rho_A(r_A) \tau_B(r_B) dV + \int \rho_B(r_B) \tau_A(r_A) dV - Q \int \tau_A(r_A) \tau_B(r_B) dV$$



Slika 12. Shematski prikaz preklapanja ljuski proteina. Slika iz [1].

2.1. Razvoj ljuske proteina u bazne funkcije

Proizvoljno tijelo u trodimenzionalnom prostoru $\sigma(r)$ može se prikazati kao beskonačna suma 3D baznih funkcija F_{nlm} .

$$\sigma(r) = \sum_{n=1}^{\infty} \sum_{l=0}^{n-1} \sum_{m=-l}^l a_{nlm} F_{nlm}$$

$$F_{nlm} = R_n(r) y_l^m(\theta, \phi)$$

Koristeći konačan broj baznih funkcija red prelazi u sumu:

$$\sigma(r) = \sum_{n,l,m}^N a_{nlm} F_{nlm},$$

$$\sum_{nlm}^N \equiv \sum_{n=1}^N \sum_{l=0}^{n-1} \sum_{m=-l}^l$$

Koeficijente ljuske računamo lako preko svojstava ortogonalnosti baznih funkcija :

$$\int F_{nlm} F_{n'l'm'} dV = \delta_{nn'} \delta_{ll'} \delta_{mm'}$$

$$\int \sigma(r) F_{n'l'm'} dV = \int \sum_{nlm}^N a_{nlm} F_{nlm} F_{n'l'm'} dV$$

$$\Rightarrow a_{nlm} = \int \sigma(r) F_{nlm} dV$$

$$dV = r^2 \sin \theta dr d\theta d\phi$$

Broj koeficijenata jedne ljuske proteina reda N jednak je broju $N(N+1)(2N+1)/6$.

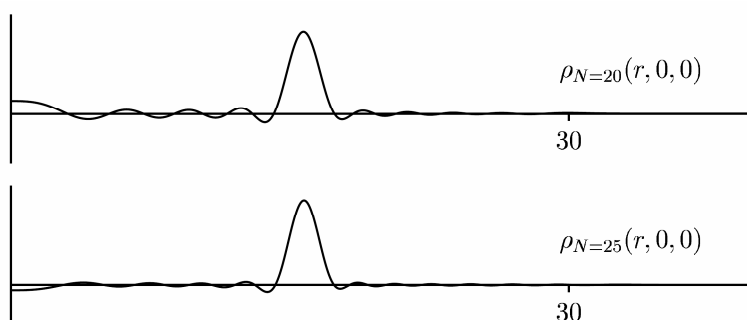
$$\sum_{n=1}^N \sum_{l=0}^{n-1} (2l+1) = \sum_{n=1}^N \left(2 \frac{(n-1)n}{2} + n \right) = \sum_{n=1}^N n^2 = N(N+1)(2N+1)/6$$

Integrale koeficijenta baznih funkcija aproksimiramo sumama:

$$a_{nlm} \approx \Delta V \sum_c R_{nl}(r_c) y_l^m(\theta_c, \phi_c)$$

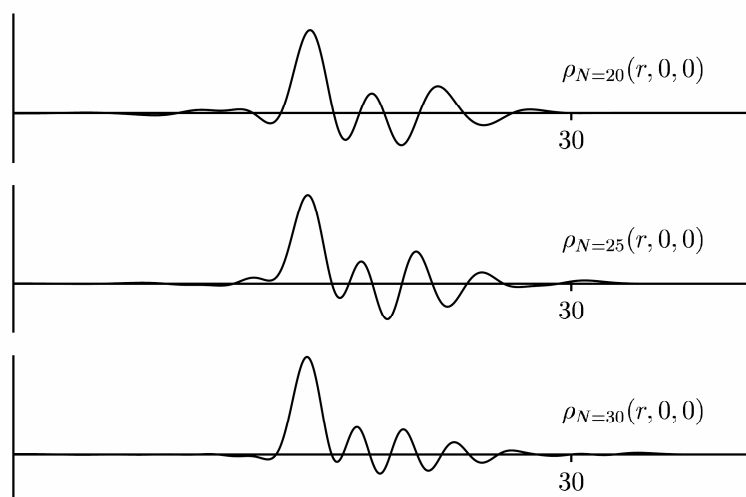
Sumacija se događa preko svih točaka ljuske grida koje su različite od nule. Svaka točka grida ima volumen ΔV i sferne koordinate (r_c, θ_c, ϕ_c) .

Razvijanjem i rekonstrukcijom sfere radijusa 15 Å i debljine ljuske 1.4 Å u sumu baznih funkcija vidimo jako lijepu reprezentaciju ljuske sfere na slici. Vidimo maksimum gustoće ljuske na 15 Å i eksponencijalni pad gustoće izvan područja.



Slika 13. Rekonstrukcija ljuske kugle. Slika iz [1].

Prikaz rekonstrukcije ljuske kocke smještene u ishodištu debljine 1.4 Å i duljine brida 30 Å. Vidimo da je teže reprezentirati kocku nego sferni oblik pomoću baznih funkcija.



Slika 14. Rekonstrukcija ljuske kocke. Slika iz [1].

2.1.1. Rotacije ljuske proteina

Moguće je pokazati da se funkcije sfernih harmonika prilikom rotacije transformiraju ovako:

$$\hat{R}(\alpha, \beta, \gamma) y_l^m(\theta, \phi) = \sum_{m'} y_l^{m'}(\theta, \phi) D_{m'm}^{(l)}(\alpha, \beta, \gamma)$$

gdje $\hat{R}(\alpha, \beta, \gamma)$ predstavlja rotacijski operator u Eulerovim kutovima (α, β, γ) .

Unitarna realna rotacijska matrica $D_{m'm}^{(l)}(\alpha, \beta, \gamma)$ dimenzija $(2l+1) \times (2l+1)$ se računa ovako :

$$D_{m'm}^{(l)}(\alpha, \beta, \gamma) = e^{-im'\alpha} d_{m'm}^{(l)}(\beta) e^{-im'\gamma}$$

gdje se $d_{m'm}^{(l)}$ [15][14] elementi računaju ovako:

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

Ovakva rotacijska svojstva znače da se proizvoljna funkcija $\sigma(\theta, \phi)$ može rotirati u novu funkciju $\sigma'(\theta, \phi)$ promjenom samo koeficijenata ekspanzije a_{nlm} . Dokaz slijedi direktno iz svojstva ortogonalnosti:

$$\sigma'(\theta, \phi) = \hat{R}(\alpha, \beta, \gamma) \sigma(\theta, \phi)$$

ekspanzijom u koeficijente baznih funkcija dobivamo:

$$\sum_{l=0}^L \sum_{m=-l}^l a'_{lm} y_{lm}^m(\theta, \phi) = \sum_{l'=0}^L \sum_{m'=-l'}^{l'} a_{l'm'} \sum_q R_{qm'}^{(l)} y_{l'}^q(\theta, \phi)$$

množenjem obadvije strane sa $y_k^j(\theta, \phi)$ i integriranjem po sferi svi članovi zbog ortogonalnosti nestanu osim članova $k=l=l'$ i $j=m=q$.

$$\int \sum_{l=0}^L \sum_{m=-l}^l a'_{lm} y_{lm}^m(\theta, \phi) y_k^j(\theta, \phi) dV = \int \sum_{l'=0}^L \sum_{m'=-l'}^{l'} a_{l'm'} \sum_q R_{qm'}^{(l)} y_{l'}^q(\theta, \phi) y_k^j(\theta, \phi) dV$$

$$a'_{kj} = \sum_{m'} R_{jm'}^{(k)} a_{km'} \text{ ili } a'_{lm} = \sum_{m'} a_{lm'} R_{mm'}^{(l)}$$

Trodimenzionalne bazne funkcije se rotiraju na sličan način jer rotacijski operatori ne utječu na radijalne funkcije:

$$\begin{aligned}\hat{R}(\alpha, \beta, \gamma)\sigma_N(\underline{r}) &= \sum_{nlm}^N \hat{R}(\alpha, \beta, \gamma) [a_{nlm} R_{nl}(r) y_l^m(\theta, \phi)] \\ \hat{R}(\alpha, \beta, \gamma)\sigma_N(\underline{r}) &= \sum_{nlm}^N R_{nl}(r) \hat{R}(\alpha, \beta, \gamma) [a_{nlm} y_l^m(\theta, \phi)] \\ \hat{R}(\alpha, \beta, \gamma)\sigma_N(\underline{r}) &= \sum_{nlm}^N a'_{nlm} R_{nl}(r) y_l^m(\theta, \phi)\end{aligned}$$

Koeficijenti rotiranog proteina se računaju ovako:

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

2.1.1.1. Rotacija oko z-osi

Realni sferni harmonici su definirani ovako :

$$y_l^m(\theta, \phi) = \begin{cases} \mathcal{G}_{lm}(\mu)(\cos m\phi)/\sqrt{\pi} & : m > 0 \\ \mathcal{G}_{lm}(\mu)(\cos m\phi)/\sqrt{2\pi} & : m = 0 \\ \mathcal{G}_{l\bar{m}}(\mu)(\sin \bar{m}\phi)/\sqrt{\pi} & : m < 0 \end{cases}$$

Rotacija sfernih harmonika za Eulerove kuteve $\beta = \gamma = 0$ i proizvoljni α kut ima jednostavnu formu:

$$\begin{aligned}\hat{R}(\alpha, 0, 0)y_l^m(\theta, \phi) &= y_l^m(\theta, \phi - \alpha) \\ y_l^m(\theta, \phi - \alpha) &= \mathcal{G}_{lm}(\mu)(\cos(m\phi - m\alpha))/\sqrt{\pi} \\ y_l^m(\theta, \phi - \alpha) &= \frac{\mathcal{G}_{lm}(\mu)}{\sqrt{\pi}} (\cos m\phi \cos m\alpha + \sin m\phi \sin m\alpha) \\ y_l^m(\theta, \phi - \alpha) &= \frac{\mathcal{G}_{lm}(\mu)}{\sqrt{\pi}} \cos m\phi \cos m\alpha + \frac{\mathcal{G}_{lm}(\mu)}{\sqrt{\pi}} \sin m\phi \sin m\alpha \\ y_l^m(\theta, \phi - \alpha) &= y_l^m(\theta, \phi) \cos m\alpha + y_l^{\bar{m}}(\theta, \phi) \sin m\alpha\end{aligned}$$

2.2. Translacije ljuski proteina

$$I_{nlmn'l'm'} = \int R_{nl}(r_A) y_l^m(\theta_A, \phi_A) R_{n'l'}(r_B) y_{l'}^{m'}(\theta_B, \phi_B) dV$$

Prebacivanjem ovog integrala u izduženi sferoidni (engl. prolate spheroidal) koordinatni sustav račun se znatno pojednostavljuje. Ishodište globalnog koordinatnog sustava izabrano je na polovici udaljenosti težišta proteina na z-osi. Težište prvog proteina je postavljeno na poziciju $-R/2$ po z-osi a težište drugog na $+R/2$ udaljenosti od ishodišta. Veza između sfernog koordinatnog sustava i izduženo sferoidnog (engl. prolate spheroidal) sustava dana je sljedećim relacijama:

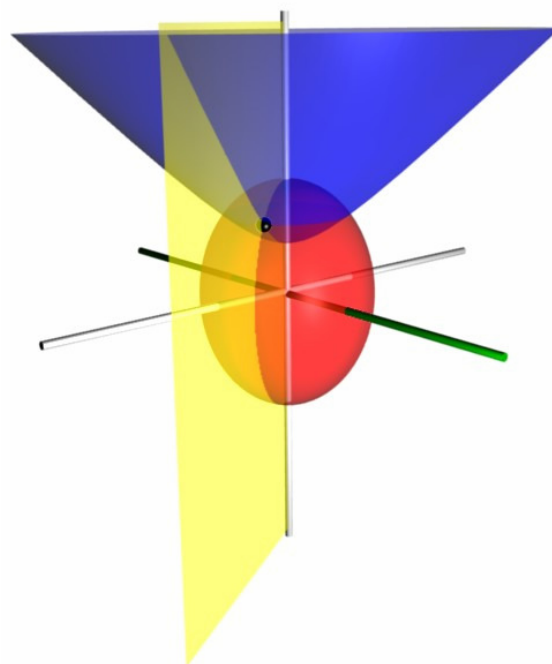
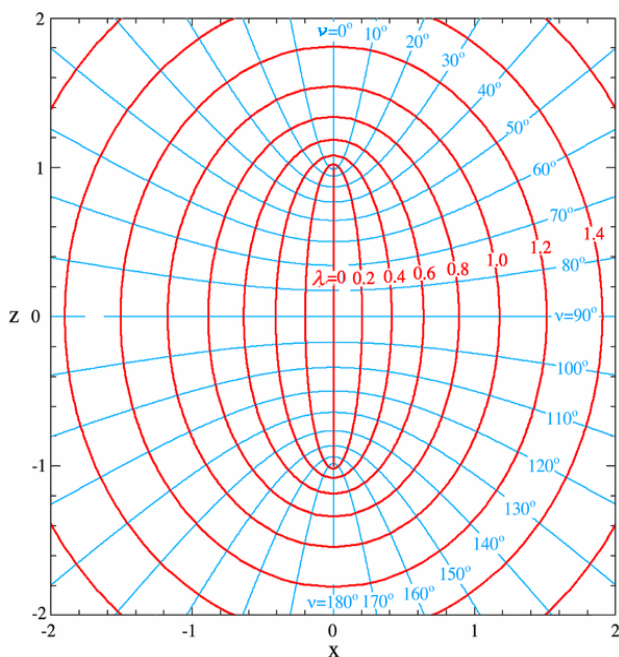
$$\lambda = \frac{r_A + r_B}{R}, \nu = \frac{r_A - r_B}{R}, \phi = \phi_A = \phi_B$$

Izduženo sferoidni (engl. prolate spheroidal) koordinatni sustav je trodimenzionalni ortogonalni sustav koji nastaje rotacijom sferoida oko glavnih osi gdje se nalaze njihova žarišta. Svaka točka je opisana uređenom trojkom (λ, ν, ϕ) . Veza sa kartezijevim koordinatnog sustavom (x, y, z) dana je sljedećim relacijama:

$$x = a \sinh \lambda \sin \nu \cos \phi$$

$$y = a \sinh \lambda \sin \nu \sin \phi$$

$$z = a \cosh \lambda \cos \nu$$



Slika 15. Prikaz izduženog sferoidnog sustava (engl. prolate spheroidal).

Volumni integral u novom koordinatnom sustavu računa se ovako:

$$\int dV = \int_1^\infty d\lambda \int_{-1}^1 dv \int_0^{2\pi} (R/2)^3 (\lambda^2 - v^2) d\phi$$

$$I_{nlmnl'm'} = \int R_{nl}(r_A) y_l^m(\theta_A, \phi_A) R_{n'l'}(r_B) y_{l'}^{m'}(\theta_B, \phi_B) dV$$

$$I_{nlmnl'm'} = (R/2)^3 \int_1^\infty d\lambda \int_{-1}^1 dv R_{nl}(r_A) R_{n'l'}(r_B) \int_0^{2\pi} y_l^m(\theta_A, \phi_A) y_{l'}^{m'}(\theta_B, \phi_B) (\lambda^2 - v^2) d\phi$$

$$A = \int_0^{2\pi} y_l^m(\theta_A, \phi_A) y_{l'}^{m'}(\theta_B, \phi_B) (\lambda^2 - v^2) d\phi = (\lambda^2 - v^2) \int_0^{2\pi} \mathcal{G}_{l|m|}(\cos \theta_A) \varphi_m(\phi_A) \mathcal{G}_{l'|m'|}(\cos \theta_B) \varphi_{m'}(\phi_B) d\phi$$

$$\varphi_m(\phi) = \begin{cases} \cos m\phi / \sqrt{\pi} > 0 \\ 1/\sqrt{2\pi} : m = 0 \\ \sin m\phi / \sqrt{\pi} : m < 0 \end{cases}$$

$$m, m' > 0 : A = \mathcal{G}_{l|m|}(\mu_A) \mathcal{G}_{l'|m'|}(\mu_B) (\lambda^2 - v^2) \int_0^{2\pi} \cos m\phi / \sqrt{\pi} \cos m'\phi / \sqrt{\pi} d\phi = \mathcal{G}_{l|m|}(\mu_A) \mathcal{G}_{l'|m'|}(\mu_B) \delta_{mm'}$$

analogno se pokaže za ostale kombinacije da vrijedi:

$$\int_0^{2\pi} \varphi_m(\phi) \varphi_{m'}(\phi) = \delta_{mm'}$$

$$A = \mathcal{G}_{l|m|}(\mu_A) \mathcal{G}_{l'|m'|}(\mu_B) (\lambda^2 - v^2) \delta_{mm'}$$

$$I_{nlmnl'm'} = (R/2)^3 \int_1^\infty d\lambda \int_{-1}^1 dv R_{nl}(r_A) R_{n'l'}(r_B) A$$

$$I_{nlmnl'm'} = (R/2)^3 \int_1^\infty d\lambda \int_{-1}^1 dv R_{nl}(r_A) R_{n'l'}(r_B) \mathcal{G}_{l|m|}(\mu_A) \mathcal{G}_{l'|m'|}(\mu_B) (\lambda^2 - v^2) \delta_{mm'}$$

$$I_{nlmnl'm'} = K_{nlmnl'm'} \delta_{mm'}$$

Elementi translacijskih matrica se računaju ovako ovako:

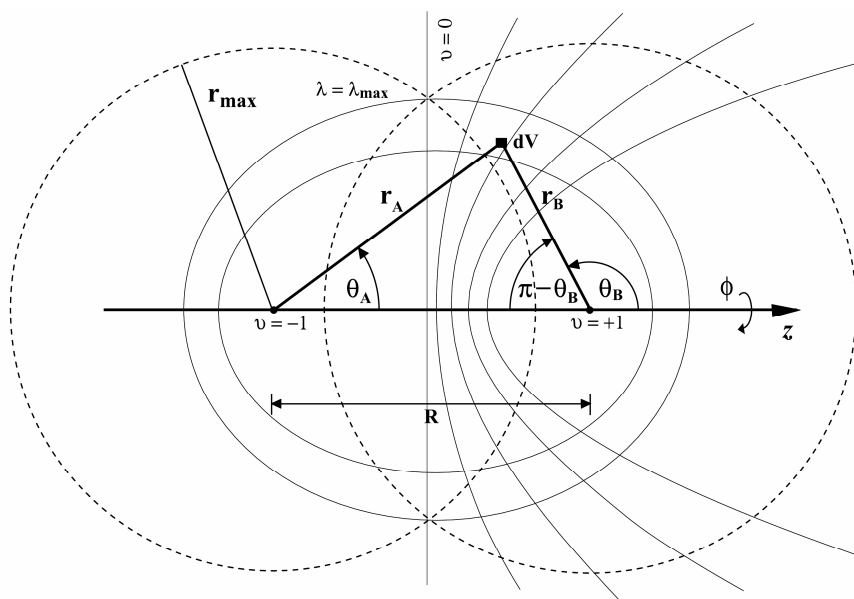
$$K_{nlmnl'm'}(R) = (R/2)^3 \int_1^{\lambda_{\max}} d\lambda \int_{-1}^1 dv R_{nl}(r_A) R_{n'l'}(r_B) \mathcal{G}_{l|m|}(\mu_A) \mathcal{G}_{l'|m'|}(\mu_B) (\lambda^2 - v^2) dv$$

tj. integrale računamo metodom numeričke integracije:

$$K_{nlmnl'm'}(R) \approx (R/2)^3 \sum_i \sum_j R_{nl}(r_A) R_{n'l'}(r_B) \mathcal{G}_{l|m|}(\mu_A) \mathcal{G}_{l'|m'|}(\mu_B) (\lambda_i^2 - v_j^2) \Delta_\lambda \Delta_v$$

Numerička integracija se može računati preko 200 odsječaka po Δ_λ i Δ_v .

Translacijske matrice se izračunaju jednom i spremne na disk.



Slika 16. Prikaz odnosa koordinata težišta proteina u izduženo sferoidnom sustavu (engl. prolate spheroidal)

Iz relacija za vezu sa sfernim koordinatnim sustavom lako se dobiju izrazi za r_A i

r_B :

$$\lambda = \frac{r_A + r_B}{R} \Rightarrow r_A = \lambda R - r_B$$

$$v = \frac{r_A - r_B}{R} \Rightarrow r_B = r_A - vR$$

$$r_A = \lambda R - r_A + vR \Rightarrow r_A = \frac{(\lambda + v)R}{2}$$

$$r_B = r_A - vR = \frac{(\lambda + v)R}{2} - vR = \frac{R(\lambda - v)}{2}$$

Iz slike se lako dobiju relacije za μ_A i μ_B :

$$r_B^2 = r_A^2 + R^2 - 2r_A R \cos \theta_A = \frac{(\lambda + v)^2 R^2}{4} + R^2 - 2 \frac{(\lambda + v)R^2}{2} \cos \theta_A$$

$$\frac{R^2(\lambda - v)^2}{4} = \frac{(\lambda + v)^2 R^2}{4} + R^2 - 2 \frac{(\lambda + v)R^2}{2} \cos \theta_A$$

$$R^2(\lambda - v)^2 = (\lambda + v)^2 R^2 + 4R^2 - 4(\lambda + v)R^2 \cos \theta_A$$

$$-4\lambda v = 4 - 4(\lambda + v) \cos \theta_A \Rightarrow \cos \theta_A = \frac{1 + \lambda v}{(\lambda + v)} = \mu_A$$

analogno se izvodi:

$$\cos(\pi - \theta_B) = \frac{1 - \lambda v}{\lambda - v} = \mu_B$$

3. Ocjena komplementarnosti proteina

Ocjenjujemo komplementarnost dva proteina koristeći model dvostruke ljuske ocjenjujemo pomoću sljedećeg izraza:

$$S = \int \rho_A(r_A) \tau_B(r_B) dV + \int \rho_B(r_B) \tau_A(r_A) dV - Q \int \tau_A(r_A) \tau_B(r_B) dV$$

gdje $\rho_A(r_A)$ predstavlja element vanjske ljuske proteina A koji se nalazi na r_A radij vektoru od težišta proteina A, $\tau_A(r_A)$ element unutarnje ljuske proteina A koji se nalazi na radij vektoru r_A od težišta proteina A, $\rho_B(r_B)$ element vanjske ljuske proteina B koji se nalazi na radij vektoru r_B od težišta proteina B i $\tau_B(r_B)$ element unutarnje ljuske proteina B koji se nalazi na radij vektoru r_B od težišta proteina B. Parametar Q predstavlja negativni penetracijsku konstantu koja kažnjava prodiranje unutarnjih ljusaka proteina. U simulacijama je korišten faktor $Q = 11$.

Ekspanzijom funkcija gustoće ljusaka proteina u red baznih funkcija dobivamo novu ocjenu komplementarnosti:

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

Integrali na desnoj strani jednadžbe ne ovise o individualnim proteinima nego samo o udaljenosti težišta R pa se oni mogu jednom izračunati i poslije ih koristiti. Ti integrali se mogu riješiti numeričkom metodom u izduženo sferoidnom (engl. prolate spheroidal) sustavu ili analitičkim putem [7] i spremiti u translacijske matrice $K_{nl'n'|m|} \delta_{mm'}$.

$$\int R_{nl}(r_A)R_{n'l'}(r_B)\mathcal{Y}_{|m|}(\mu_A)\mathcal{Y}_{|m'|}(\mu_B)\varphi_m(\phi_A)\varphi_{m'}(\phi_B)dV = K_{nl'n'l'|m|}(R)\delta_{mm'}$$

stoga se integral preklapanja vanjske i unutarnje ljuske može zapisati ovako:

$$\int \rho_A \tau_B dV = \sum_{nlm}^N \sum_{n'l'm'}^N a_{nlm}^{\rho} b_{n'l'm'}^{\tau} K_{nl'n'l'|m|}(R)\delta_{mm'}$$

Koristeći rotacijska svojstva sfernih harmonika, koeficijenti funkcije gustoće ljuske se računaju ovako:

$$a_{nlm}^{\rho} = \sum_{m'=-l}^l a_{nlm}^{\rho} R^{(l)}_{mm'}(0, \beta_1, \gamma_1),$$

$$b_{n'l'm'}^{\tau} = \sum_{m'=-l}^l b_{n'l'm'}^{\tau} R^{(l)}_{mm'}(0, \beta_2, \gamma_2).$$

Rotacija oko z-osi ima posebno lagan izračun:

$$B_{nlm}^{\rho} = b_{nlm}^{\rho} \cos m\alpha_2 + b_{nlm}^{\rho} \sin m\alpha_2$$

Analogno tako se izvode formule za računanje ostalih ljusaka: $a_{nlm}^{\tau}, B_{nlm}^{\tau}$.

Korištenjem izraza br. moguće je računati ocjenu komplementarnosti dvije konformacije sa asimptotskom složenošću od $O(N^5)$. Zbog velike vremenske složenosti potrebno je algoritam ocjene konformacije presložiti kako bi mu se složenost smanjila.

Ocjenu konformacije moguće je zapisati na drugačiji način:

$$\int \rho_A \tau_B dV = \sum_{n'l'm'}^N \left(\sum_{nlm}^N a_{nlm}^{\rho} K_{nl'n'l'|m|}(R)\delta_{mm'} \right) B_{n'l'm'}^{\tau}$$

pa se unutarnja suma može smatrati kao promijenjeni koeficijenti za molekulu A koja je translatairana za udaljenost R:

$$A_{n'l'm'}^{\rho}(R) = \sum_{nlm}^N a_{nlm}^{\rho} K_{nl'n'l'|m|}(R)\delta_{mm'}$$

Koristeći ovakav način ocjena konformacija se može napraviti sa asimptotskom složenošću $O(N^3)$ ovako:

$$\int \rho_A \tau_B dV = \sum_{n'l'm'}^N A_{n'l'm'}{}^{\rho}(R) B_{n'l'm'}{}^{\tau}$$

Algoritam prijanjanja proteina pretražuje šest dimenzionalni prostor mogućih parova konformacija dva proteina. Šest dimenzionalni prostor mogućih parova konformacija opisuje se pomoću uređene šestorke $(\beta_1, \gamma_1, R, \alpha_2, \beta_2, \gamma_2)$. Uređena trojka (β_1, γ_1, R) opisuje poziciju proteina receptora koji je rotiran za Eulerove kutove (β_1, γ_1) i translatican po z-osi za R udaljenost. Uređena trojka $(\alpha_2, \beta_2, \gamma_2)$ opisuje poziciju proteina liganda koji je rotiran za pripadne Eulerove kutove. Šestodimenzionalni prostor konformacije se diskretizira određenim metodama i dobije se skup uređenih šestorki $\{(\beta_1, \gamma_1, R, \alpha_2, \beta_2, \gamma_2)_i\}$ koji se potom evaluiraju ocjenom komplementarnosti.

Kako bi se skup izabranih konformacija brzo pretražio koristimo sljedeće izraze:

$$b_{nlm}{}^Q = b_{nlm}{}^{\rho} - Q b_{nlm}{}^{\tau}$$

$$S(R, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2) = \sum_{nlm}^N \left(A_{nlm}{}^{\rho}(R) \left(b_{nlm}{}^{\tau} \cos m\alpha_2 + b_{nlm}{}^{\tau} \sin \bar{m}\alpha_2 \right) + A_{nlm}{}^{\tau}(R) \left(b_{nlm}{}^Q \cos m\alpha_2 + b_{nlm}{}^Q \sin \bar{m}\alpha_2 \right) \right)$$

očito je da je prethodna formula za ocjenu komplementarnosti ekvivalentna formuli br. što pokazujemo raspisivanjem formule:

$$\begin{aligned} A_{nlm}{}^{\tau}(R) \left(b_{nlm}{}^Q \cos m\alpha_2 + b_{nlm}{}^Q \sin \bar{m}\alpha_2 \right) &= A_{nlm}{}^{\tau}(R) \left(\left(b_{nlm}{}^{\rho} - Q b_{nlm}{}^{\tau} \right) \cos m\alpha_2 + \left(b_{nlm}{}^{\rho} - Q b_{nlm}{}^{\tau} \right) \sin \bar{m}\alpha_2 \right) \\ &= A_{nlm}{}^{\tau}(R) \left(b_{nlm}{}^{\rho} \cos m\alpha_2 + b_{nlm}{}^{\rho} \sin \bar{m}\alpha_2 \right) - Q A_{nlm}{}^{\tau}(R) \left(b_{nlm}{}^{\tau} \cos m\alpha_2 + b_{nlm}{}^{\tau} \sin \bar{m}\alpha_2 \right) \\ &= A_{nlm}{}^{\tau}(R) \left(B_{nlm}{}^{\rho} \right) - Q A_{nlm}{}^{\tau}(R) \left(B_{nlm}{}^{\tau} \right) \end{aligned}$$

pa konačno vidimo da izraz za ocjenu komplementarnosti je ekvivalentan početnom izrazu:

$$S(R, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2) = \sum_{nlm}^N \left(A_{nlm}{}^{\rho}(R) \left(B_{nlm}{}^{\tau} \right) + A_{nlm}{}^{\tau}(R) \left(B_{nlm}{}^{\rho} \right) - Q A_{nlm}{}^{\tau}(R) \left(B_{nlm}{}^{\tau} \right) \right)$$

Izlučivanjem dijelova izraza uz $\cos m\alpha_2$ i $\sin \bar{m}\alpha_2$ dobivamo sljedeći izraz:

$$S(R, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2) = \sum_{n=1}^N \sum_{l=0}^{n-1} \sum_{m=-l}^l \left(A_{nlm}^{\prime\rho}(R) \left(b_{nlm}^{\prime\tau} \cos m\alpha_2 + b_{nlm}^{\prime\tau} \sin \bar{m}\alpha_2 \right) + A_{nlm}^{\prime\tau}(R) \left(b_{nlm}^{\prime\varrho} \cos m\alpha_2 + b_{nlm}^{\prime\varrho} \sin \bar{m}\alpha_2 \right) \right) \\ = \sum_{n=1}^N \sum_{l=0}^{n-1} \sum_{m=-l}^l \left(\cos m\alpha_2 \left(A_{nlm}^{\prime\rho}(R) b_{nlm}^{\prime\tau} + A_{nlm}^{\prime\tau}(R) b_{nlm}^{\prime\varrho} \right) + \sin \bar{m}\alpha_2 \left(A_{nlm}^{\prime\rho}(R) b_{nlm}^{\prime\tau} + A_{nlm}^{\prime\tau}(R) b_{nlm}^{\prime\varrho} \right) \right)$$

zamjenom redoslijeda sumacija dobivamo sljedeći izraz:

$$S(R, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2) = \sum_{m=-L}^L \sum_{(n,l) \in R^{(m)}} \cos m\alpha_2 \left(A_{nlm}^{\prime\rho}(R) b_{nlm}^{\prime\tau} + A_{nlm}^{\prime\tau}(R) b_{nlm}^{\prime\varrho} \right) + \\ \sin \bar{m}\alpha_2 \left(A_{nlm}^{\prime\rho}(R) b_{nlm}^{\prime\tau} + A_{nlm}^{\prime\tau}(R) b_{nlm}^{\prime\varrho} \right); L = N - 1$$

ovaj korak je matematički moguće napraviti samo ako očuvamo redoslijed sumacije po indeksima n, l koji su elementi skupa $R^{(m)} = \{(n, l)_i\}$.

Skupove $R^{(m)} = \{(n, l)_i\}$ generiramo tako da u svakoj iteraciji $\sum_{n=1}^N \sum_{l=0}^{n-1} \sum_{m=-l}^l$ ove sume u

skup $R^{(m)}$ dodajemo uređeni član (n, l) .

Stoga ocjena komplementarnosti pojednostavljeno izgleda ovako:

$$S(R, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2) = \sum_{m=-L}^L Q_m^+ \cos m\alpha_2 + Q_m^- \sin \bar{m}\alpha_2$$

članovi Q_m^+, Q_m^- predstavljaju Fourierov spektar funkcije ocjene komplementarnosti proteina $S(\alpha_2)$.

Spektar funkcije $S(\alpha_2)$ se računa ovako:

$$Q_m^+ = \sum_{(n,l) \in R^{(m)}} \left(A_{nlm}^{\prime\rho}(R) b_{nlm}^{\prime\tau} + A_{nlm}^{\prime\tau}(R) b_{nlm}^{\prime\varrho} \right) \\ Q_m^- = \sum_{(n,l) \in R^{(m)}} \left(A_{nlm}^{\prime\rho}(R) b_{nlm}^{\prime\tau} + A_{nlm}^{\prime\tau}(R) b_{nlm}^{\prime\varrho} \right)$$

Konačno zbog svojstva $\cos m\alpha_2 = \cos \bar{m}\alpha_2$ i $\sin m\alpha_2 = -\sin \bar{m}\alpha_2$ formula ocjene komplementarnosti se svodi na realni Fourierov red:

$$S(\alpha_2) = Q_0^{++} + \sum_{m=1}^L Q_m^{++} \cos m\alpha_2 + Q_m^{--} \sin m\alpha_2$$

$$Q_m^{++} = Q_m^+ + Q_m^+ (1 - \delta_{m0})$$

$$Q_m^{--} = Q_m^- - Q_m^-$$

4. Implementacija

U ovom radu implementirana je brza metoda pretraživanja prostora konformacija u slijednoj i paralelnoj verziji. Kod je pisan u C++ i korišten je MPI standard prilikom paralelizacije.

Pseudokod pretraživanja šest dimenzionalnog prostora konformacija i postupak računanja prijanjanja proteina prikazan je ovdje:

```
Docking::DockingQsearch()
(1)   calculateAndSaveAllRotations();
      Foreach R
(2)       Učitaj  $K_{nln'l|m|}(R)$  translacijske matrice
          Foreach  $(\beta_1, \gamma_1)$ 
(3)           rotiraniReceptor = UčitajRotiraniReceptor  $(\beta_1, \gamma_1)$ ;
(4)           transliraj( $K_{nln'l|m|}(R)$ , rotiraniReceptor);
          Foreach  $(\beta_2, \gamma_2)$ 
(5)           rotiraniLigand = UčitajRotiraniLigand  $(\beta_2, \gamma_2)$ ;
(6)           Izračunaj spektar  $\{Q_m^{++}, Q_m^{--}\}, \forall m$ 
          Foreach  $\alpha_2$ 
(7)           
$$S(\alpha_2) = \sum_{m=1}^{N-1} (Q_m^{++} \cos m\alpha_2 + Q_m^{--} \sin m\alpha)$$

          endfor
          endfor
      endfor
endfor
end
```

Tablica 1. Pregled vremenske i prostorne složenosti metode pretrage prostora konformacija

Odsječak	Vremenska složenost	Memorijska složenost
(1)	$O(M) \times O(N^3)$	$2 \times M \times N^3 \times 8$ bajta
(2)	$O(N^5)$	$N^2(N+1)^2(2N+1)/12 \times 8$ bajta
(3)	$O(1)$	1 pokazivač = 4 bajta
(4)	$O(N^5)$	0
(5)	$O(1)$	1 pokazivač = 4 bajta
(6)	$O(N^3)$	$6 \times N \times 8$ bajta
(7)	$O(N)$	$2 \times \alpha_2 \times N \times 8$ bajta

M predstavlja broj različitih kuteva (β, γ) koji se evaluiraju prilikom pretrage u šest dimenzionalnom prostoru konformacija. N predstavlja maksimalni stupanj baznih funkcija F_{nlm} razvoja ljuski. Broj različitih kuteva rotacije po z-osi predstavlja oznaka $|\alpha_2|$, broj različitih radijusa predstavlja oznaka $|R|$ a oznaka c predstavlja konstantu.

Ukupan broj koraka algoritma pretraživanja prostora konformacija i prijanjanja proteina može se opisati sljedećom formulom:

$$\approx \underbrace{c \times M \times N^3}_{\text{Sve Rotacije}} + |R| \times \left(\underbrace{c \times N^5}_{\text{Učitavanje Translacijskih Matrica}} + M \times \left(\underbrace{c \times N^5}_{\text{Translacija}} + M \times \left(\underbrace{c \times N^3}_{\text{Računanje Spektra}} + |\alpha_2| \times \underbrace{N \times c}_{\text{Računanje Ocjene}} \right) \right) \right)$$

U praksi broj različitih kuteva M jednak je broju 20×2^{2T} gdje T predstavlja teselacijsku konstantu. Mi obično koristimo $T = 2$ i onda imamo 320 različitih kuteva (β, γ) . Zbog toga vrijedi da je $M \gg N$ također red N je najčešće 20 pa vrijedi da je $N^2 \approx M$ i broj koraka algoritma približno iznosi:

$$\approx \underbrace{|R| \times M}_{\text{BrojTranslacija}} \times c \times N^5 + |R| \times M \times M \times c \times N^3$$

Vidimo da složenost algoritma ne ovisi o parametru $|\alpha_2|$ za parametre

$M \approx 320, N = 20$.

4.1. Paralelizacija sustava

4.1.1. MPI

MPI (engl. Message Passing Interface) definira standard komunikacije porukama, odnosno razmjenu podataka među procesima. Komunikacija porukama engl. (message passing) je najkorišteniji model paralelnog programiranja. Broj procesa u MPI programu je po definiciji konstantan (u osnovnoj inačici standarda nisu predviđeni mehanizmi stvaranja odnosno gašenja procesa). Svi procesi izvode isti program - SPMD model (engl. Single Program Multiple Data), međutim mogu se pokrenuti i različiti programi na različitim procesima -MPMD model (engl. Multiple Program Multiple Data). MPMD model se uvijek može simulirati uvrštavanjem više funkcija u jedan SPMD program.

MPI standard podržava četiri načina komunikacije:

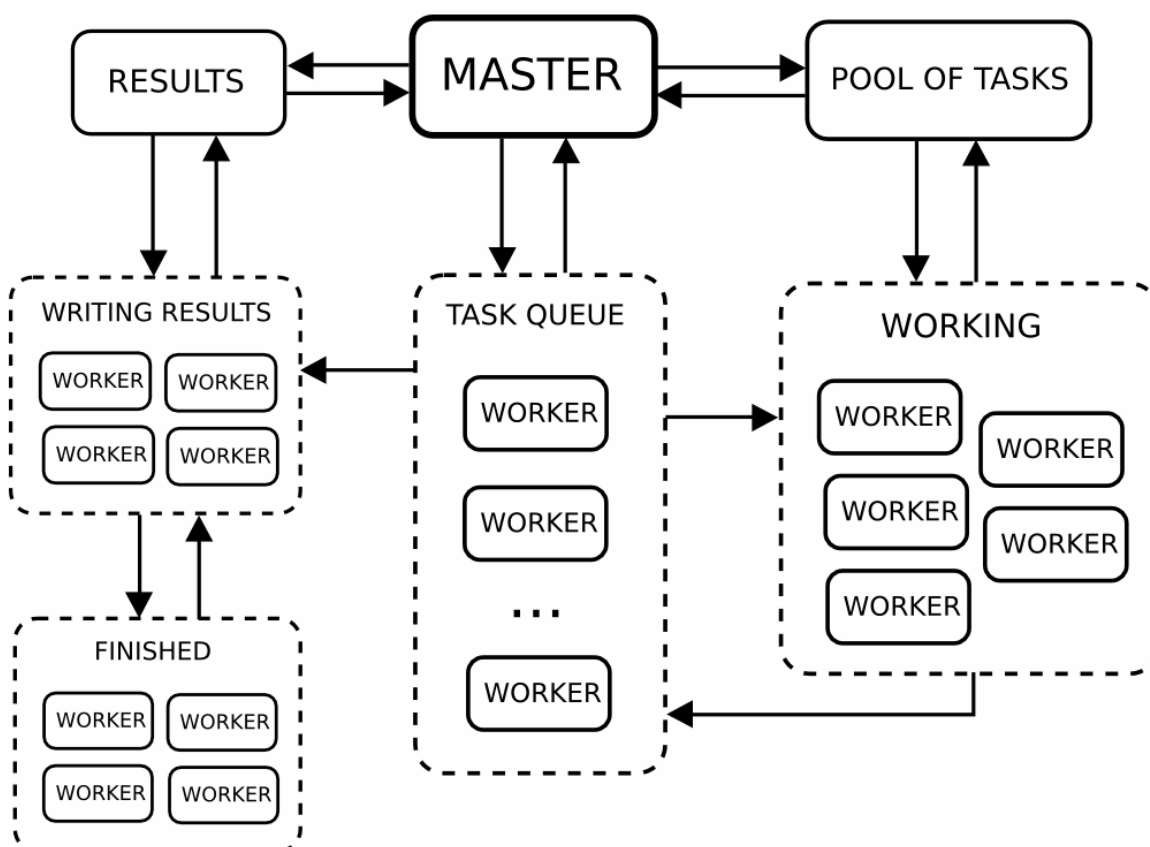
- point-to-point komunikacija između dva određenog procesa
- collective komunikacija između grupa procesa
- probe funkcije za asinkronu komunikaciju
- communicator mehanizam za razvoj modularnih paralelnih programa

4.1.2. Funkcionalnost paralelnog programa

Zbog modularne arhitekture PDT sustava omogućena je paralelizacija sustava ugradnjom samo jednog dodatnog modula. Program *docking_MPI_main.cpp* predstavlja taj dodatni paralelni modul ostvaren pomoću MPI standarda koji poziva članske metode klase *docking*. Svaki proces može jednu od dvije uloge *master* ili *worker*.

Postoji samo jedan proces *master* koji je odgovoran da dodjelu zadataka procesima *worker*. Zadatak se definira kao proces evaluacije prijanjanja proteina liganda i receptora za određeni radijus. Proces *worker* mora poslati zahtjev procesu *master* kako bi dobio zadatak. Proces *master* čeka zahtjeve za dodjelom zadataka i nakon primitka određenog zahtjeva dodjeljuje prvi neodrađeni zadatak

tom procesu. Ovaj način dodjele zadataka omogućuje visoku skalabilnost (engl. scalability), modularnost (engl. modularity) i ubrzanje. Paralelni program također omogućava promjenu zrnatosti zadatka (engl. granularity) i ujednačavanja opterećenja pojedinih procesa (engl. load balancing). Nakon što su svi procesi završili svoje zadatke i skup neodrađenih zadataka kod procesa *master* je prazan pokreće se zapisivanje rezultata na disk. Svaki proces *worker* zapiše skup najboljih konformacija u datoteku sa nazivom koji ima sufiks broja procesa koji je napravio datoteku. Potom proces *master* učitava datoteke svih procesa *worker* i napravi glavnu datoteku najboljih konformacija sa nazivom koji je zapisan u datoteci config.txt pod parametrom scores_output_file. Svaki proces *worker* također pošalje procesu *master* najbolji score za određeni radius. Proces *master* potom napravi datoteku RadiusScores.txt u kojoj se nalaze za svaki radius najbolji score.



Slika 17. Shematski prikaz paralelnog dijela sustava.

5. Testovi i rezultati

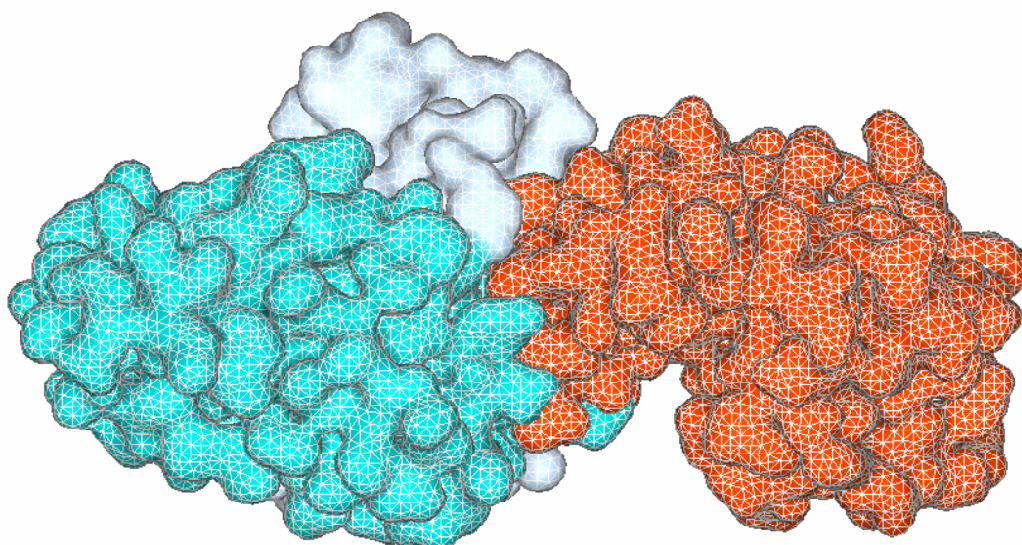
U ovom poglavlju prikazani su rezultati testova prijanjanja proteina nad kompleksima 3HFL i 1AY7 i testovi brzine izvođenja algoritma. Svaki kompleks je prvo vizualiziran u kristalografskoj strukturi u alatu Hex [6] i zatim podvrgnut sljedećim testovima:

- Test: Alfa rotacije kristalografske strukture – U ovom testu se receptor translacija po z-osi na udaljenost približno jednakoj kristalografskoj udaljenosti R_k i zatim se vrši 360 rotacija liganda po α_2 alfa kutu. Rezultati pokazuju da su maksimumi ocjene komplementarnosti kada je $\alpha_2 \approx 0$. Taj rezultat pokazuje da je kristalografska struktura ($\beta_1 = \gamma_1 = \beta_2 = \gamma_2 = \alpha_2 = 0$ i udaljenost proteina $R \approx R_k$) najbolje ocijenjena.
- Test: Ocjena komplementarnosti po radijusu- U ovom testu pretražuje se čitav skup uređenih šestorki koji reprezentiraju moguće pozicije konformacija proteina $\{(R, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2)\}$. Za određeni R prikazuje se $\max\{S(R, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2)\}, \forall(\beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2)$, gdje S predstavlja funkciju ocjene komplementarnosti.

- Test: Detekcije kristalografske strukture – Ovaj test je najvažniji izlaz algoritma prijanjanja proteina. U njemu se dobije datoteka rangiranih konformacija po ocjeni geometrijske komplementarnosti. Prijanjanje proteina je to uspješnije što se kristalografska struktura ($\beta_1 = \gamma_1 = \beta_2 = \gamma_2 = \alpha_2 = 0$ i udaljenost proteina $R \approx R_k$) nalazi više rangirana [8].
- Test: Razvoj ljuske po lokalnoj z-osi- Ovaj test vrši inverznu transformaciju kristalografske strukture iz sume baznih funkcija u prostornu domenu po z-osi. Na temelju ovog testa moguće je vidjeti gustoće vanjskih i unutarnjih ljuski proteina.

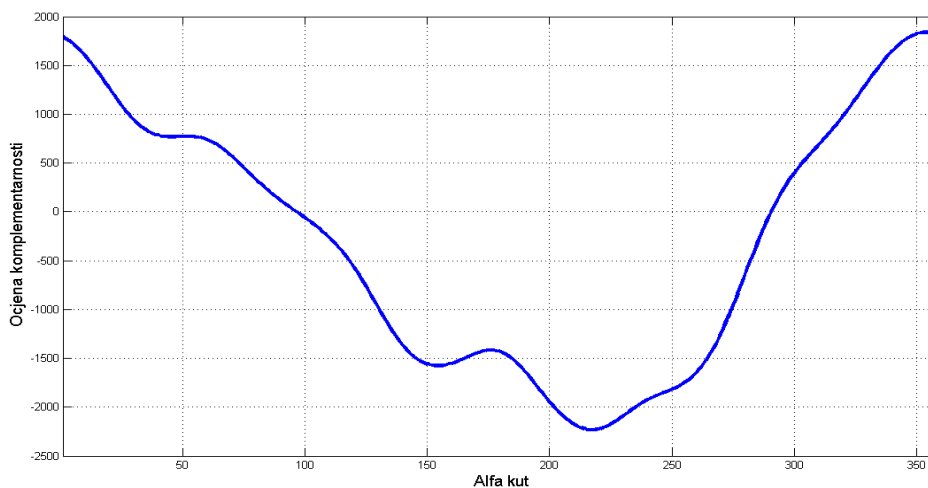
5.1. 3HFL kompleks

5.1.1. Vizualizacija kompleksa u kristalografskoj strukturi



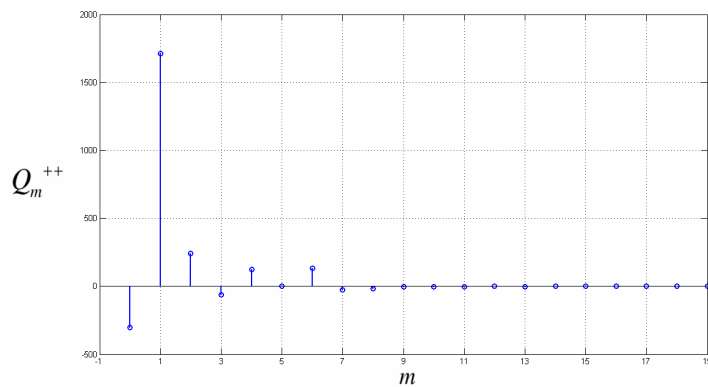
Slika 18. Prikaz 3HFL kompleksa u kristalografskoj strukturi $R_k = 31.5 \text{ \AA}$

5.1.2. Test: Alfa rotacije kristalografske strukture

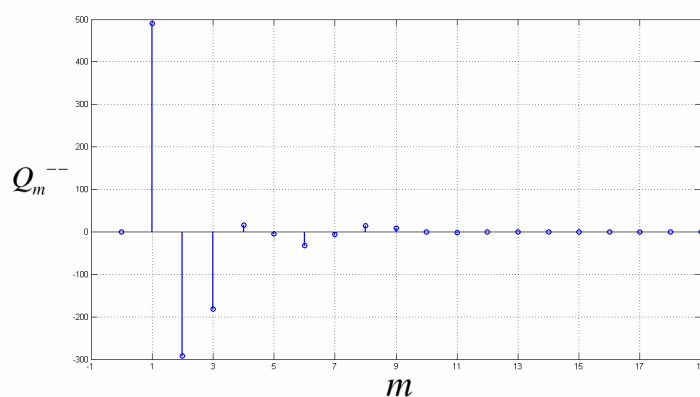


Slika 19. Prikaz funkcije komplementarnosti ovisne o alfa kutu. 3HFL kompleks, radijus

$$R = 33.5 \text{ \AA}, \beta_1 = \gamma_1 = \beta_2 = \gamma_2 = 0.$$

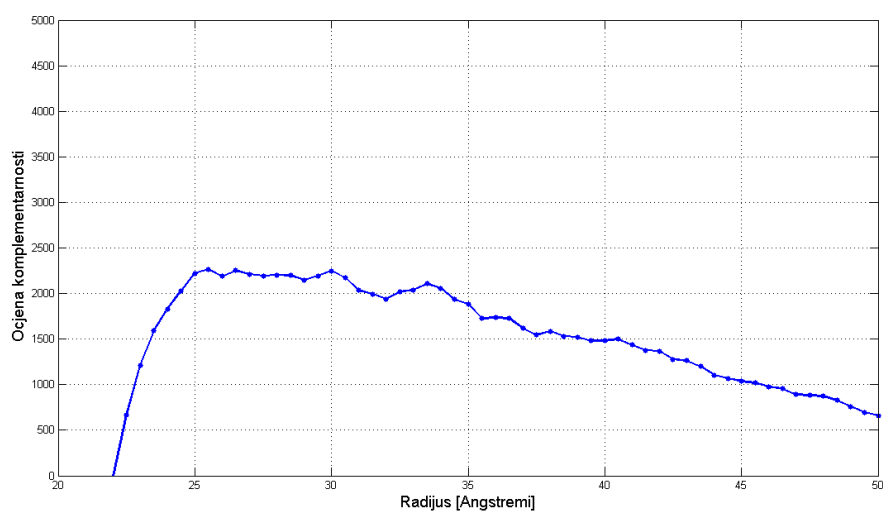


Slika 20. Prikaz spektra Q_m^{++} za 3HFL kompleks, $R = 33.5 \text{ \AA}$, $\beta_1 = \gamma_1 = \beta_2 = \gamma_2 = 0$.



Slika 21. Prikaz spektra Q_m^{--} za 3HFL kompleks, $R = 33.5 \text{ \AA}$, $\beta_1 = \gamma_1 = \beta_2 = \gamma_2 = 0$.

5.1.3. Test: Ocjena komplementarnosti po radijusu



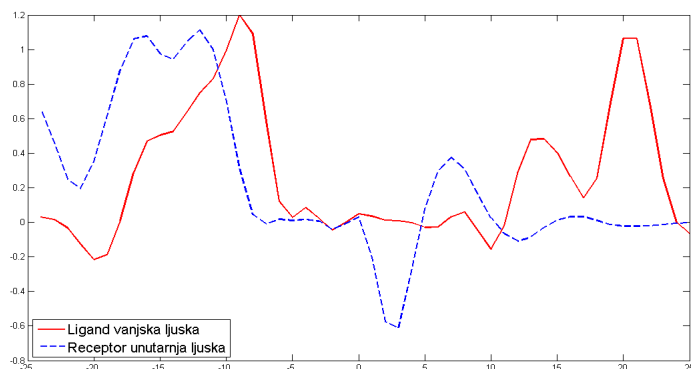
Slika 22. Prikaz maksimalne ocjene komplementarnosti za neki radijus 3HFL kompleksa.

5.1.4. Test: Detekcija kristalografske strukture

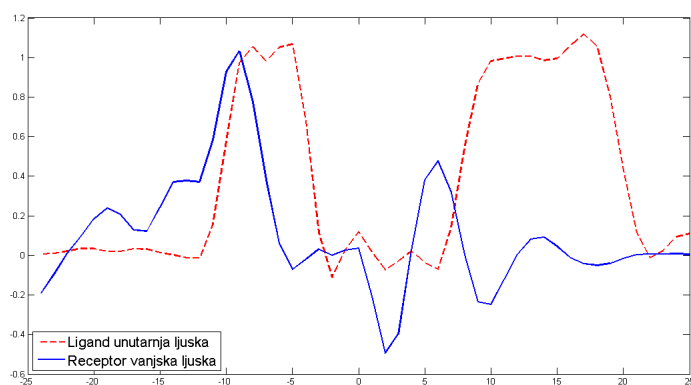
```
# MPI worker...
# ligand inner skin name: data/coef/aligned_3hfl/3hfl_fv_inner.coef
# ligand outer skin name: data/coef/aligned_3hfl/3hfl_fv_outer.coef
# receptor inner skin name: data/coef/aligned_3hfl/3hfl_ly_inner.coef
# receptor outer skin name: data/coef/aligned_3hfl/3hfl_ly_outer.coef
# orderOfCoefficients: 20
# translationPath: data/trans
# tessellationOrder: 2
# alphaSize: 36
# penalty: 5.000000
# scoresSize: 200
# radiusStart: 0.000000
# radiusEnd: 50.100000
# rank score R Rec_beta1 Rec_gamma1 Lig_alpha2 Lig_beta2 Lig_gamma2
1 2.265719E+03 25.000000 1.188061 2.180512 5.061455 2.304166 2.092092
2 2.253399E+03 26.000000 0.701644 1.509053 1.919862 2.364075 2.239600
3 2.247107E+03 29.500000 2.219089 2.379460 0.698132 2.536327 4.031964
4 2.220737E+03 24.500000 1.188061 2.180512 5.061455 2.304166 2.092092
5 2.212125E+03 26.500000 0.701644 1.509053 1.919862 2.364075 2.239600
6 2.204171E+03 27.500000 1.713638 4.758610 5.759587 2.552373 3.859802
7 2.198174E+03 28.000000 1.713638 4.758610 5.759587 2.536327 4.031964
8 2.194878E+03 29.000000 2.219089 2.379460 0.698132 2.536327 4.031964
9 2.193544E+03 27.000000 2.516365 4.365308 4.188790 0.165954 1.260680
10 2.192953E+03 26.500000 1.399017 1.949618 6.108652 2.383800 1.802562
...
459 1.924862E+03 27.000000 0.639462 1.675352 5.235988 1.482143 1.449533
460 1.924835E+03 29.500000 1.659449 4.591126 5.759587 2.362947 4.302874
461 1.924708E+03 32.500000 0.000000 0.000000 6.108652 0.000000 0.000000
462 1.924635E+03 27.000000 0.757793 1.802562 2.094395 2.586919 2.276545
```

Slika 23. Prikaz izlaznih rezultata algoritma prijanjanja proteina. Crveno označena konformacija je bliska kristalografskoj strukturi $\alpha_2 = 6.108 \approx 2\pi$.

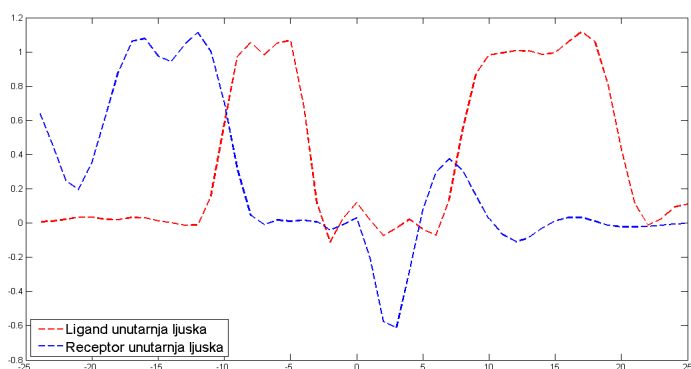
5.1.5. Test: Razvoj ljuski po z-osi



Slika 24. Prikaz razvoja vanjske ljuske liganda i unutarnje ljuske receptora po z-osi 3HFL kompleksa , $R = 33.5 \text{ \AA}$, $\beta_1 = \gamma_1 = \beta_2 = \gamma_2 = \alpha_2 = 0$.



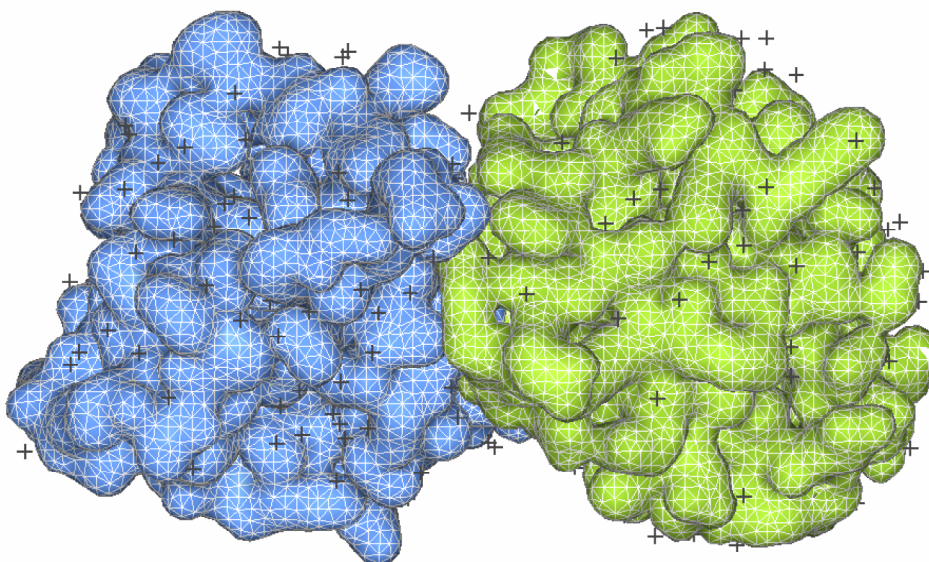
Slika 25. Prikaz razvoja unutarnje ljuske liganda i vanjske ljuske receptora po z-osi 3HFL kompleksa , $R = 33.5 \text{ \AA}$, $\beta_1 = \gamma_1 = \beta_2 = \gamma_2 = \alpha_2 = 0$.



Slika 26. Prikaz razvoja unutarnje ljuske liganda i unutarnje ljuske receptora po z-osi 3HFL kompleksa , $R = 33.5 \text{ \AA}$, $\beta_1 = \gamma_1 = \beta_2 = \gamma_2 = \alpha_2 = 0$.

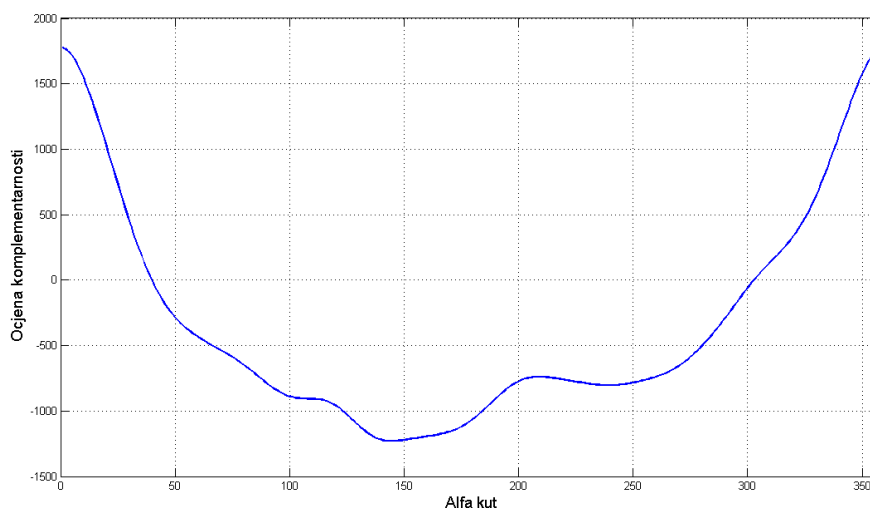
5.2. 1AY7 kompleks

5.2.1. Vizualizacija kompleksa u kristalografskoj strukturi



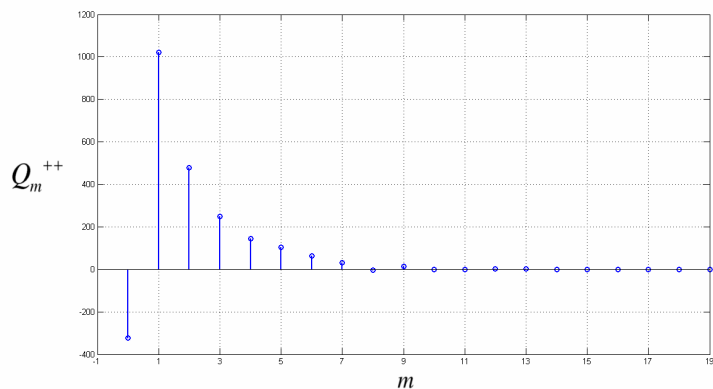
Slika 27. Prikaz 1AY7 kompleksa u kristalografskoj strukturi $R_k = 24.6 \text{ \AA}$

5.2.2. Test: Alfa rotacije kristalografske strukture

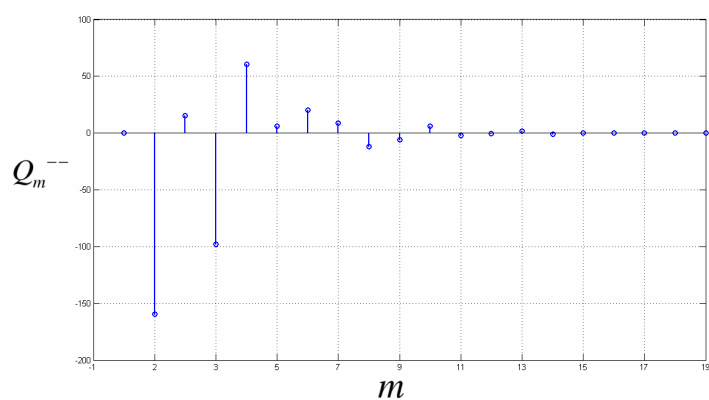


Slika 28. Prikaz funkcije komplementarnosti ovisne o alfa kutu. 1ay7 kompleks,

$$R = 26.0 \text{ \AA}, \beta_1 = \gamma_1 = \beta_2 = \gamma_2 = 0.$$

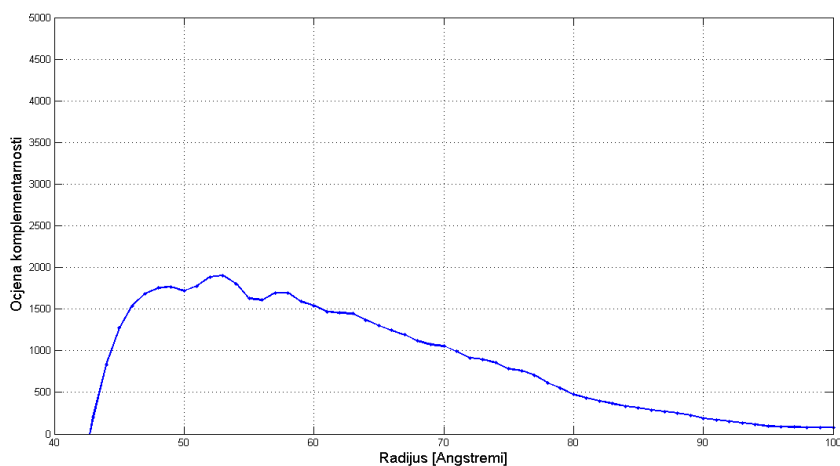


Slika 29. Prikaz spektra Q_m^{++} za 1ay7 kompleks, $R = 26.0 \text{ \AA}$, $\beta_1 = \gamma_1 = \beta_2 = \gamma_2 = 0$.



Slika 30. Prikaz spektra Q_m^{--} za 1ay7 kompleks, $R = 26.0 \text{ \AA}$, $\beta_1 = \gamma_1 = \beta_2 = \gamma_2 = 0$.

5.2.3. Test: Ocjena komplementarnosti po radijusu



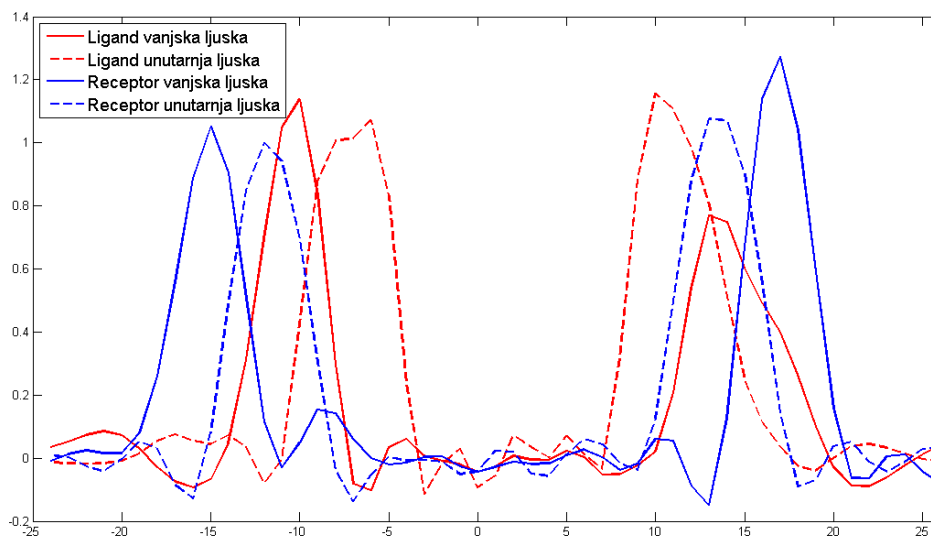
Slika 31 Prikaz maksimalne ocjene komplementarnosti za neki radijus 1ay7 kompleksa

5.2.4. Test: Detekcija kristalografske strukture

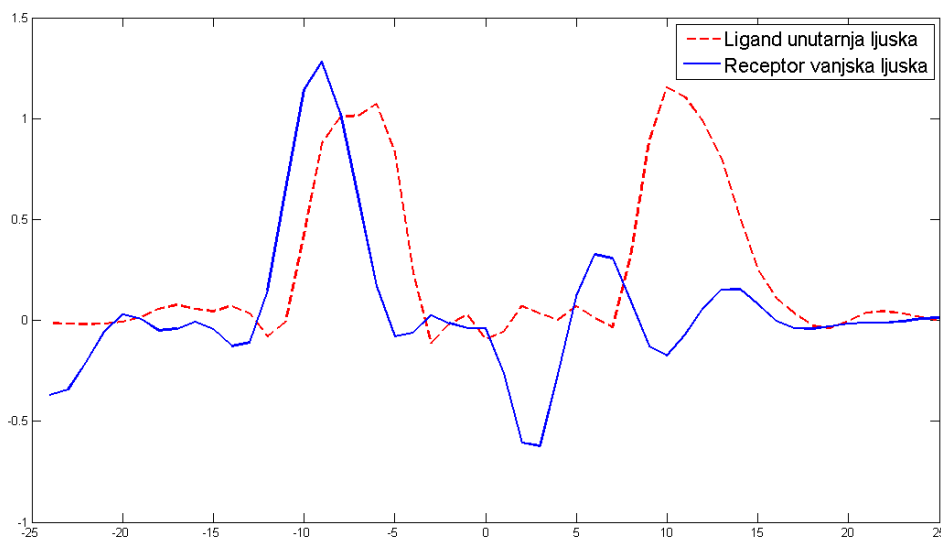
```
# MPI worker...
# ligand inner skin name: data/coef/aligned_1ay7/1ay7_receptor_inner.coef
# ligand outer skin name: data/coef/aligned_1ay7/1ay7_receptor_outer.coef
# receptor inner skin name: data/coef/aligned_1ay7/1ay7_ligand_inner.coef
# receptor outer skin name: data/coef/aligned_1ay7/1ay7_ligand_outer.coef
# orderOfCoefficients: 20
# translationPath: data/trans
# tessellationOrder: 2
# alphaSize: 36
# penalty: 5.000000
# scoresSize: 200
# radiusStart: 0.000000
# radiusEnd: 50.100000
# rank score R Rec_beta1 Rec_gamma1 Lig_alpha2 Lig_beta2 Lig_gamma2
1 1.901485E+03 26.000000 0.992221 0.925602 0.174533 0.512037 5.924324
2 1.890433E+03 26.000000 0.000000 0.000000 1.396263 0.079630 1.417110
3 1.886236E+03 25.500000 0.992221 0.925602 0.174533 0.512037 5.924324
4 1.883168E+03 25.500000 0.992221 0.925602 0.174533 0.530232 6.096806
5 1.877598E+03 26.000000 0.992221 0.925602 0.174533 0.530232 6.096806
6 1.874980E+03 25.500000 1.043159 0.757253 0.349066 0.512037 5.924324
7 1.842738E+03 26.000000 0.000000 0.000000 1.221730 0.088399 1.098233
8 1.840552E+03 26.000000 0.000000 0.000000 1.745329 0.094952 1.723015
9 1.840327E+03 25.500000 1.043159 0.757253 0.349066 0.530232 6.096806
10 1.838707E+03 26.000000 1.043159 0.757253 0.349066 0.512037 5.924324
...
20 1.779496E+03 25.500000 0.000000 0.000000 1.221730 0.088399 1.098233
21 1.778391E+03 26.000000 0.000000 0.000000 0.000000 0.000000 0.000000
22 1.777278E+03 25.000000 1.043159 0.757253 0.349066 0.512037 5.924324
```

Slika 32. Prikaz izlaznih rezultata algoritma prijanjanja proteina. Crveno označena konformacija je bliska kristalografskoj strukturi.

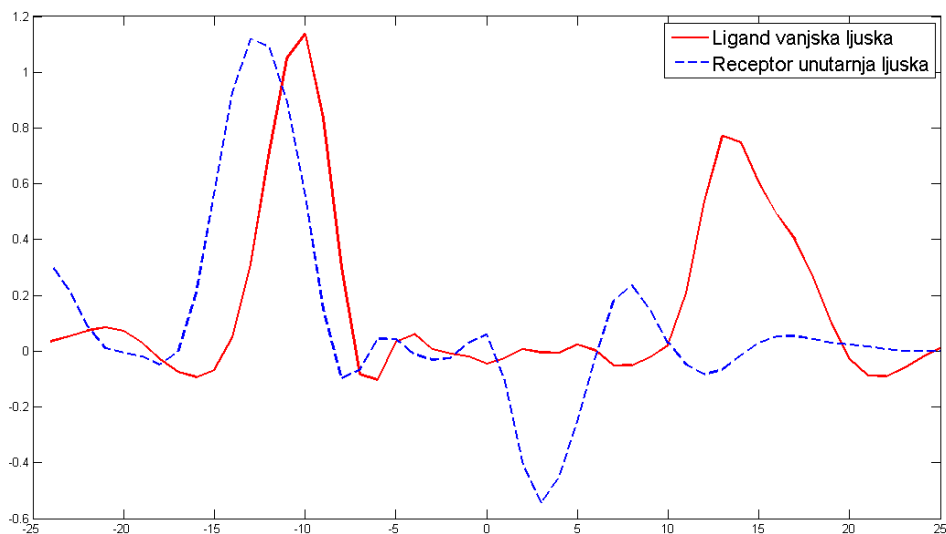
5.2.5. Test: Razvoj ljuski po z-osi



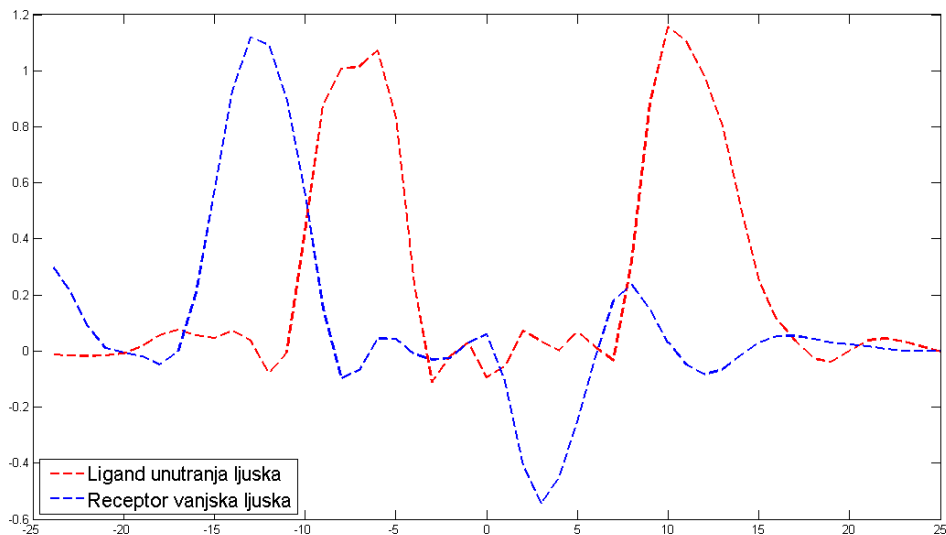
Slika 33. Prikaz ljuski 1AY7 kompleksa koji su centrirani u ishodište.



Slika 34. Prikaz razvoja unutarnje ljuske liganda i vanjske ljuske receptora po z-osi 1AY7 kompleksa , $R = 26.0 \text{ \AA}$, $\beta_1 = \gamma_1 = \beta_2 = \gamma_2 = \alpha_2 = 0$.



Slika 35. Prikaz razvoja vanjske ljuske liganda i unutarnje ljuske receptora po z-osi 1AY7 kompleksa , $R = 26.0 \text{ \AA}$, $\beta_1 = \gamma_1 = \beta_2 = \gamma_2 = \alpha_2 = 0$.



Slika 36. Prikaz razvoja unutarnje ljuske liganda i unutarnje ljuske receptora po z-osi 1AY7 kompleksa , $R = 26.0 \text{ \AA}$, $\beta_1 = \gamma_1 = \beta_2 = \gamma_2 = \alpha_2 = 0$.

5.3. Test: Brzina algoritma prijanjanja

Ukupno vrijeme čekanja u paralelnoj verziji se može opisati ovako:

$$T = \max \{ T_{komunikacija}^i + T_{računanje}^i + T_{čekanje}^i \}$$

gdje $T_{komunikacija}^i$ predstavlja vrijeme potrebno za prijenos podataka za i -ti MPI proces, $T_{računanje}^i$ predstavlja vrijeme potrebno i -tom MPI procesu za računanje prijanjanja za jedan radijus (zadatak) i $T_{čekanje}^i$ predstavlja vrijeme potrebno i -tom MPI procesu za čekanje u redu.

Zbog $T_{komunikacija}^i \ll T_{računanje}^i$ i $T_{komunikacija}^i \ll T_{čekanje}^i$ vrijedi:

$$T = \max \{ T_{računanje}^i + T_{čekanje}^i \}.$$

Ako $T_{računanje}^i + T_{čekanje}^i \approx const = T_1, \forall i$ onda se formula ukupnog vremena izvođenja svodi na sljedeći jednostavni oblik:

$$T \approx \left\lceil \frac{N}{p} \right\rceil T_1$$

gdje p predstavlja broj raspoloživih procesora, N broj ukupno različitih radijusa u prostoru pretraživanja i T_1 vrijeme potrebno za računanje prijanjanja jednog radijusa na jednom MPI procesu [5].

Tablica 2. Prikaz brzine izvođenja algoritma prijanjanja na realnim uvjetima

Ukupno vrijeme izvođenja	Brzi algoritam prijanjanja (vidi poglavlje 4.)	Normalni algoritam prijanjanja (bez računanja spektra)
Test slijedne verzije	18.9 sekundi	139.6 sekundi
Test paralelne verzije	44.5 sekundi	642 sekundi =10 m 42 sekunde

Prilikom računanja brzine izvođenja algoritma prijanjanja korištena su dva testa:

- Test slijedne verzije – ovaj test podrazumijeva računanje prijanjanja proteina za jedan radijus na 3HFL proteinu. Korištena je konstanta teselacije $T = 2$ pa je ukupan broj mogućih (β, γ) konformacije bio 320 na redu $N = 20$. Broj ispitanih alfa kuteva bio je 36. Stoga je ukupan broj testiranih konformacija bio : $|(\beta_1, \gamma_1)| \times |(\beta_2, \gamma_2)| \times |\alpha_2| = 320 \times 320 \times 36 = 3686400$.
- Test paralelne verzije – ovaj test uključuje pretraživanje prostora konformacija kao u testu slijedne verzija ali za 100 različitih radijusa na 51 slobodnih MPI procesa. Ukupan broj ispitanih konformacija bio je:

$$|(\beta_1, \gamma_1)| \times |(\beta_2, \gamma_2)| \times |\alpha_2| \times |R| = 320 \times 320 \times 36 \times 100 = 368640000$$

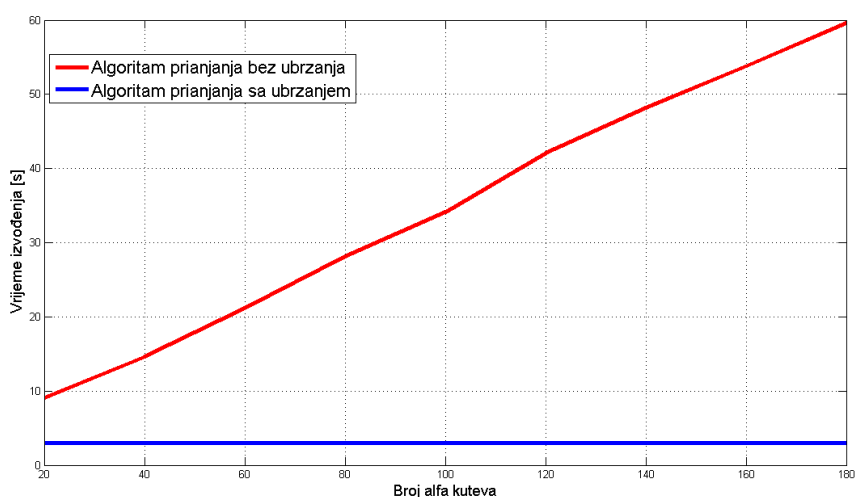
Prilikom računanja brzine izvođenja korištena su dva algoritma:

- Brzi algoritam prijanjanja koji koristi računanje spektra $\{Q_m^{++}, Q_m^{--}\}$.

$$\text{Broj koraka ovog algoritma} \approx \underbrace{|R| \times M}_{\text{BrojTranslacija}} \times c \times N^5 + |R| \times M \times M \times c \times N^3$$

- Normalni algoritam prijanjanja bez računanja spektra

$$\text{Broj koraka ovog algoritma} \approx \underbrace{|R| \times M}_{\text{BrojTranslacija}} \times c \times N^5 + |R| \times M \times M \times c \times N^3 \times |\alpha_2|$$



Slika 37. Prikaz potrebnog vremena računanja normalnog i ubranog algoritma za parametar teselacije $T=1$

Zaključak

Kao član tima projekta PDT (engl. Protein Docking Tool) [4] razvili smo osnovnu inačicu sustava za prijanjanje proteina i proširenu verziju sustava PPDT (engl. Parallel Protein Docking Tool) [5]. PDT sustav kao ulaz dobiva dvije PDB datoteke (engl. Protein Data Bank) koje potpuno reprezentiraju trodimenzionalnu strukturu receptora i liganda. Sustav automatski generira ljuske proteina koje potom rastavlja u sumu trodimenzionalnih baznih funkcija. Uzastopnim translacijama i rotacijama ljuski proteina razvijenih u bazne funkcije pretražujemo šest-dimenzionalni prostor mogućih konformacija liganda i receptora. Sve konformacije se ocjenjuju prema geometrijskoj komplementarnosti površina proteina i potom rangiraju.

U sklopu ovog diplomskog implementiran je slijedni i paralelni brzi algoritam pretraživanja prostora mogućih konformacija prilikom procesa prijanjanja proteina. Brzina algoritma je od iznimne važnosti za proces prijanjanja algoritma jer se ispituje velik broj mogućih konformacija. Ubrzanje sustava implementacijom brzog algoritma u odnosu na normalnu verziju algoritma prijanjanja u praktičnim primjerima slijednog algoritma je nešto više od sedam puta. Dok se u paralelnom algoritmu ubrzanje osjeti više ako je broj računala ograničen. Paralelna verzija sustava sa brzim algoritmom prijanjanja ispita 368 milijuna mogućih konformacija red 20 na pedeset računala u nešto više od 44 sekunde. Što predstavlja izvrstan rezultat.

Točnost sustava provjerena je na proteinskim kompleksima 3HFL i 1AY7. U oba proteinska kompleksa uspješno je detektirana kristalografska struktura što je pravi test profesionalnih alata za prijanjanje proteina. Sustav se može proširiti sa računanjem elektrostatike koja potencijal u prostoru prikazuje kao red trodimenzionalnih baznih funkcija koje utječu na prostornu gustoću naboja na proteinu koja je isto prikazana redom baznih funkcija. Informacije o hidrofobnim regijama proteina se mogu ugraditi imaginarni dio kompleksnih sfernih harmonika. Konačna ocjena proteinskog kompleksa bi se ocijenila kao težinska kombinacija geometrijske komplementarnosti, hidrofobnosti i elektrostatike [9].

Literatura

- [1] Ritchie, D. W., Parametric Protein Shape Recognition, PhD Thesis, Departments of Computing Science and Molecular and Cell Biology, University of Aberdeen (1998)
- [2] Ritchie, D.W., Kemp, G.J.L., Fast Computation, Rotation, and Comparison of Low Resolution Spherical Harmonic Molecular Surfaces, (1999) *J. Comp. Chem.* 20(4), 383-395.
- [3] Ritchie, D.W., Kemp, G.J.L., Protein Docking Using Spherical Polar Fourier Correlations, (2000) *PROTEINS: Struct. Funct. Genet.* 39, 178-194
- [4] Antulov-Fantulin, N., Čanadi, I., Piškorec, M., Sović, I., Šikić, M., Protein Docking Tool, <http://complex.zesoi.fer.hr/PDT.html>, (2010)
- [5] Sović, I. Antulov-Fantulin, N., Čanadi, I., Piškorec, Šikić, M., Parallel Protein Docking Tool, 33. Međunarodni skup za informacijsku i komunikacijsku tehnologiju, elektroniku i mikroelektroniku – MIPRO 2010, (2010)
- [6] Ritchie, D. W., HEX 5.1., <http://www.loria.fr/~rithied/hex/>
- [7] Ritchie, D. W., High Order Analytic Translation Matrix Elements For Real Space Six-Dimensional Polar Fourier Correlations, *Journal of Applied Crystallography*. 38: pages 808-818
- [8] Ritchie, D. W., Evaluation of Protein Docking Predictions Using Hex 3.1 in CAPRI Rounds 1 and 2, D.W. Ritchie (2003) *PROTEINS: Struct. Funct. Genet.* 52(1), 98-106
- [9] Ritchie, D. W., Recent progress and future directions in protein-protein docking, 2008., *Curr. Prot. Pep. Sci.* 9(1), 1-15.
- [10] Prof.dr.sc. Neven Elezović, Doc. dr. sc. Mario Krnić 1. poglavlje Unitarni prostori, kolegij: „Fourierova analiza“, Fakultet elektrotehnike i računarstva, ak.god. 2009./2010.
- [11] William Elwood Byerly, An Elementary Treatise on Fourier's Series and Spherical, Cylindrical, and Ellipsoidal Harmonics, Harvard University, 1893.
- [12] Volker Schönefeld, Spherical Harmonics, http://heim.c-otto.de/~volker/prosem_paper.pdf, 2005.
- [13] Martin J. Mohlenkamp, A User's Guide to Spherical Harmonics, May 10, 2010, <http://www.math.ohiou.edu/~mjm/research/uguide.pdf>
- [14] Miguel A. Blanco, M. Flórez, M. Bermejo, Evaluation of the rotation matrices in the basis of real spherical harmonics, 2010, <http://www1.elsevier.com/homepage/saa/eccc3/paper48/eccc3.html>
- [15] Sakurai J.J., Modern Quantum Mechanics, Revised Edition, 1994

Sažetak

Prijanjanje proteina (engl. Protein Docking) proučava mogućnosti stvaranja proteinskih kompleksa. U ovom radu korištena je SPF metoda (engl. Spherical Polar Fourier) koja ljuske proteina rastavlja u redove trodimenzionalnih baznih funkcija.

U prvom poglavlju objašnjeni su matematički fundamenti potrebni za razumijevanje algoritma prijanjanja proteina.

U drugom poglavlju opisuju se ljuske proteina, njihov razvoj u red baznih funkcija, rotacija i numerička metoda translacije ljusaka tj. generiranja translacijskih matrica.

Poglavlje tri opisuje način ocjene komplementarnosti površina proteina. U ovom poglavlju se daje potpuni izvod brzog algoritma prijanjanja proteina koji je korišten u ovom radu.

Poglavlje četiri opisuje implementaciju slijedne i paralelne verzije brzog algoritma za prijanjanje proteina.

Poglavlje pet opisuje provedene testove i dobivene rezultate. Provedeni su testovi nad 3HFL i 1AY7 proteinskim kompleksima. Detektirane su kristalografske strukture što upućuje na točnost implementirane metode. Također su pokazana dobivena ubrzanja sa novim brzim algoritmom prijanjanja proteina. Dobivena su ubrzanja preko sedam puta u slijednom algoritmu u praktičnim uvjetima. U paralelnom algoritmu su dobivena ubrzanja do 14 puta za praktične uvjete i ispitano je 136 milijuna konformacija po kompleksu.

Abstract

Spherical harmonics are the angular portion of a set of solutions to Laplace's equation. Combined with orthogonal Radial functions like associated Legendre polynomials together form three dimensional basis functions in Hilbert space. Therefore arbitrary three dimensional function like protein skins can be expanded to basis function series. Docking algorithm evaluates geometric complementarity of ligand and receptor by rotating and translating them in space. In this thesis fast algorithm of docking search was implemented. By calculating Fourier series on score function of twist angle sequential docking algorithm speeds up to more than seven times. Test were made on two sets of proteins complexes 3HFL and 1AY7 with sequential and parallel version. In both cases crystalografic structure was detected which implies that system is running well.

Ključne riječi na hrvatskom: prijanjanje proteina, sferni harmonici, radijalne funkcije

Ključne riječi na engleskom: Protein Docking, Spherical Harmonics, Radial functions.

DODATAK

a.) Pokretanje paralelnog programa

U datoteci `machines` nalazi se popis računala na kojima se mogu pokretati MPI procesi. Stoga ta datoteka mora biti osvježena, koristite naredbu „`qstat -f`“ kako bi provjerili status računala na klasteru. Radi automatizacije procesa testiranja i pokretanja paralelnog programa napravljena je skripta `StartMPIDockingScript.sh`. Pokretanjem skripte korisnik unosi naziv testa koji se provodi, opis testa koji se provodi i broj procesa na kojem će se pokrenuti paralelni program. Rezultati svakog testa spremaju se u direktorij `data/tests/` u datoteku koja ima naziv testa koji je korisnik upisao i sufiksom jedinstvenog rednog broja testa. Za svaki test skripta automatski dodjeli jedinstveni broj testa koji odgovara broju napravljenih testova u direktoriju `data/tests/` uvećanog za jedan. Skripta `StartMPIDockingScript.sh` automatski ispisuje vrijeme trajanja testa u konzolu. Nakon završetka testa sve datoteke pojedinih procesa `woker` su objedinjene u datoteku najboljih konformacija sa nazivom koji je zapisan u datoteci `config.txt` pod parametrom `scores_output_file`. Nakon toga datoteke pojedinih procesa `woker` više nisu potrebne pa se automatski obrišu. Automatski se u datoteku testa zapisuje opis testa `about_test.txt` , datoteka `config.txt` , datoteka `scores_output_file` i datoteka najboljih konformacija.

b.) Instalacija paralelnog programa

Potrebno je prethodno instalirati neku MPI biblioteku npr. Open MPI library. Također je potrebno instalirati GSL biblioteku (*GNU Scientific Library*). Za kompajliranje paralelnog programa potrebno je pokrenuti makefile koji se nalazi u direktoriju MPIDocking. Za kompajliranje paralelnog programa `docking_MPI_main.cpp` potrebno je koristiti *mpiCC* wrapper kompajler npr. "`mpiCC -o docking $(OBJS) $(USER_OBJS) $(LIBS)`". Poželjno je isključiti „-static“ opciju statičkog kompajliranja. Potrebno je također linkati GSL biblioteku tako da u varijablu LIBS u datoteci `object.mk` unutar direktorija MPIDocking/ postavimo putanje do biblioteka, include datoteka i imena biblioteka npr. LIBS := -L/home/pdt/local/lib -l/home/pdt/local/include/ -lgsl -lgslcblas". Također je potrebno koristiti *mpiCC* wrapper kompajler u datoteci `subdir.mk` direktorija MPIDocking/src/ i zastavicu putanje do include datoteka biblioteke GSL. Poželjno je također koristiti zastavicu `-O3` za optimizaciju prilikom kompajliranja koja ubrza program do 3 puta. Iako smo uspjeli kompajlirati program, prilikom njegovog pokretanja program traži dinamičke biblioteke *shared library* u sistemskim direktorijima. Ukoliko GSL *shared library* (instaliran lokalno) nije u sistemskim direktorijima, program se ne može pokrenuti. Potrebno je postaviti putanju do biblioteke u varijablu `LD_LIBRARY_PATH` i to zapisati u skriptu `bashrc`.