

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 2131

**Implementacija algoritma za računanje
proteinskih interakcija**

Katarina Kušević

Zagreb, lipanj, 2011.

Zahvaljujem mentoru Mili Šikiću te kolegi Matiji na strpljenju i pomoći pri izradi ovog rada.

Sadržaj

Popis slika i tablica	1
1. Uvod	2
2. Biokemija	3
2.1. Proteini	3
2.2. Elektrostatska sila	5
2.3. Energija otapala	5
3. Podaci	10
3.1. PDB datoteke	10
3.2. Naboji atoma i otapalu dostupna područja površine	11
3.3. Van der Waalsovi radijusi atoma	12
4. Implementacija	13
4.1. Opis algoritma	14
5. Instalacija i pokretanje	18
6. Rezultati	19
7. Zaključak	24
8. Sažetak	25
9. Literatura	27

Popis slika i tablica

Slika 2.1. Opća struktura aminokiselina.....	3
Slika 2.2. Otapalu dostupno područje površine atoma - ASA.....	6
Slika 3.1. Primjer PDB datoteke	10
Slika 3.2. Dio datoteke s nabojima atoma	11
Slika 3.3. Datoteka s ASA vrijednostima i potencijalom otapala.....	12
Slika 3.4. Datoteka s vrijednostima Van der Waalsovih radijusa	12
Slika 4.1. Dijagram obrazaca.....	13
Slika 4.2. Dijagram toka podataka.....	13
Slika 4.3. Dijagram klase Atom.....	14
Slika 4.4. Dijagram klase Chain.....	15
Slika 4.5. Dijagram klase Pdb.....	15
Slika 4.6. Dijagram klase Charges.....	16
Slika 4.7. Dijagram klase Molecule.....	16
Slika 4.8. Dijagram klase Amino	17
Slika 4.9. Dijagram klase Asa.....	17
Slika 6.1. Primjer ispisa rezultata	19
Slika 6.2. Primjer ispisa rezultata za 1ECP.pdb	21
Slika 6.3. Primjer ispisa rezultata za 1ECP.pdb	23
Tablica 2.1. Popis aminokiselina	4
Tablica 2.2. Vrijednosti ASA-e za aminokiseline	8
Tablica 6.1. Elektrostatske sile interakcije aminokiselina	20
Tablica 6.2. Vrijednosti energije otapala.....	22

1. Uvod

Proteini su jedna od najvažnijih tvari prisutnih u ljudskom tijelu. Sastavni su dio svih stanica na Zemlji. Značajni su za rast i razvoj svih tjelesnih tkiva, izgradnju krvi, kože, mišića, kose i unutarnjih organa. Proučavanje interakcija proteina važno je za shvaćanje funkcioniranja živih organizama i predstavlja informacije ključne za razumijevanje stabilnosti i funkcije proteina te omogućuje sintezu umjetnih proteina koji će imati željena svojstva. To je područje jedno od važnijih u biokemiji i bioinformatici.

U sklopu ovog rada potrebno je implementirati algoritam za izračun jačine elektrostatskih interakcija i energije otapala ili solvatacije. Za izračun elektrostatske sile koriste se naboji dobiveni iz amber polja sila. Prostorni rasporedi i koordinate atoma unutar lanca proteina preuzeti su iz PDB zapisa koji su eksperimentalno utvrđeni i dostupni u RCSB PDB bazi (engl. *Research Collaboratory for Structural Bioinformatics – Protein Data Bank* ([1])). Pri računanju energije solvatacije koristi se područje površine dostupno otapalu preuzeto iz izračuna kojeg je u sklopu diplomskog rada 2006. godine dobio Josip Mihel, te vrijednosti potencijala otapanja i područja površine dostupne otapalu, odnosno totalne ASA-e (engl. *Accessible Solvent Area*) za pojedine aminokiseline. Udaljenosti koje se koriste izražene su u angstromima ($1\text{Å} = 10^{-10}\text{ m}$).

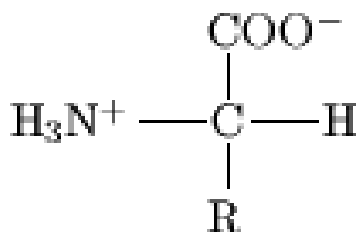
U narednim poglavljima rada objašnjena je biokemijska pozadina, struktura proteina, podaci potrebni za izračune te implementacija.

2. Biokemija

2.1. Proteini

Proteini su građeni od različitih aminokiselina, međusobno povezanih u duge lance peptidnim vezama. U ljudskom organizmu sve proteine tvore kombinacije svega 20 aminokiselina, čiji je pregled prema međunarodnoj udruzi IUPAC, koja standardizira imenovanje organskih spojeva (engl. *International Union of Pure and Applied Chemistry*), dan u Tablica 2.1. Specifične karakteristike svakog proteina određuje broj, redosljed i prostorni raspored aminokiselina u lancu proteina. Tako se tvore tisuće različitih proteina s različitim biološkim funkcijama.

Aminokiseline su organske molekule koje sadrže karboksilnu skupinu $-\text{COOH}$ i amino skupinu $-\text{NH}_2$. (Slika 2.1.) Aminokiseline se međusobno razlikuju po R skupini, odnosno bočnom ogranku koji može biti atom, jednostavna ili kompleksna molekula. Bočni ogranci aminokiselina određuju svojstva i funkciju proteina kojeg tvore.



Slika 2.1. Opća struktura aminokiselina

Povezivanjem karboksilne skupine jedne aminokiseline i amino skupine druge aminokiseline u kojem se atom ugljika veže za atom dušika uz oslobađanje molekula vode. Ta se kemijska veza naziva peptidnom vezom. Preostali dio aminokiseline naziva se ostatak aminokiseline (engl. *amino acid residue*). Ostaci aminokiselina mogu imati više različitih struktura. Važno svojstvo aminokiselina je topljivost, koja je za svaku aminokiselinu određena polarnošću bočnog ogranka. Raspodjela hidrofobnih aminokiselina, koje se slabo otapaju u vodi, i hidrofilnih, koje se dobro otapaju u vodi, ima veliku važnost u promatranju proteinskih struktura i interakcija.

IUPAC([2]) naziv aminokiseline	Hrvatski naziv aminokiseline	Troslovna kratica
Alanine	Alanin	ALA
Arginine	Arginin	ARG
Asparagine	Asparagin	ASN
Aspartic Acid	Asparaginska kiselina	ASP
Cysteine	Cistein	CYS
Glutamine	Glutamin	GLN
Glutamic Acid	Glutaminska kiselina	GLU
Glycine	Glicin	GLY
Histidine	Histidin	HIS
Isoleucine	Izoleucin	ILE
Leucine	Leucin	LEU
Lysine	Lizin	LYS
Methionine	Metonin	MET
Phenylalanine	Fenilanin	PHE
Proline	Prolin	PRO
Serine	Serin	SER
Threonine	Treonin	THR
Tryptophan	Triptofan	TRP
Tyrosine	Tirozin	TYR
Valine	Valin	VAL

Tablica 2.1. Popis aminokiselina

2.2. Elektrostatska sila

Različita elektrostatska svojstva atoma u aminokiselinama i međusobno povezanih aminokiselina utječu na prostorni raspored atoma. Djelovanjem privlačnih i odbojnih sila među atomima, oni se raspoređuju u trodimenzionalnom prostoru. Elektrostatske su interakcije važne za razumijevanje makromolekularnih interakcija. Pojedine grupe aminokiselina u proteinu sadrže često ionizirane skupine, i zbog postojanja naboja prisutna je određena prostorna gustoća naboja.

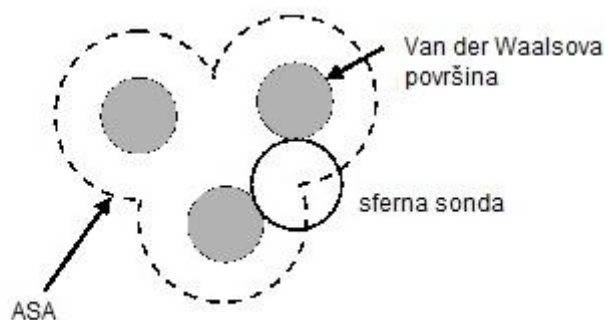
Pri računanju elektrostatske sila između dva atoma koristimo prilagođeni Coulombov zakon koji govori da je elektrostatska sila između dva točkasta naboja razmjerna umnošku iznosa oba naboja q_1 i q_2 te obrnuto razmjerna kvadratu udaljenosti r između njih. (2.1.)

$$F = \sum_a c * \frac{q_1 * q_2}{r^2} \quad 2.1.$$

Konstanta c koju koristimo iznosi -332, naboje su izračunati za svaki atom aminokiseline pojedinačno, a udaljenosti se računaju ovisno o x , y i z koordinatama pojedinih atoma. Za izračun elektrostatske sile između dvije aminokiseline zbrajamo elektrostatske sile između svih atoma tih aminokiselina.

2.3. Energija otapala

Otapalu dostupno područje površine atoma (engl. *Accessible Surface Area*, skraćeno ASA) je područje površine proteina na kojem je moguć kontakt s otapalom u kojem se protein nalazi, a to je najčešće voda. Možemo ga definirati pomoću sferne sonde koja predstavlja model otapala i kotrlja se po van der Waalsovoj površini (Slika 2.2). Taj način određivanja ASA-e opisali su Shrake A. i Rupley J.A. 1973. godine. Van der Waalsovu površinu dobijemo ako svaki atom zamijenimo sferom van der Waalsovog radijusa, koji je za atome eksperimentalno određen mjerenjem prostora između para nevezanih atoma unutar kristala. Kao radijus sonde uzima se iznos od 1.4 Å, koji ujedno predstavlja i radijus molekule vode.



Slika 2.2. Otapalu dostupno područje površine atoma - ASA

Algoritam za izračun otapalu dostupne površine atoma preuzet je iz rada Josipa Mihela ([4]). Prvo se svakom atomu unutar promatranog proteina pridruži njegov Van der Waalsov radijus, koji se učitavaju iz datoteke chothia.radii i spremaju u mapu, zatim se taj radijus proširuje za radijus sferne sonde (prošireni radijus R). Svaki atom smatra se sferom s centrom određenim koordinatama atoma i radijusa jednakom dobivenom proširenom radijusu pojedinog atoma. Područje dostupne površine svakog atoma računa se tako da se prošireni volumen atoma "izreže" po jednoj osi dvodimenzionalne ravnine (kružnice), udaljene za neku određenu vrijednost ($zslice$). Ako za promatrani atom postoji dio kružnog luka koji se ne siječe s niti jednom kružnicom susjednih atoma, atom je dostupan otapalu. Dužina slobodnog luka (L_i) određuje dostupnost atoma u i -toj ravnini, a ukupna dostupna površina atoma proporcionalna je zbroju dužina slobodnih lukova na svim ravninama. Ukupno dostupno područje površine računa se pomoću formula (2.2.) (2.3.) i (2.4.), pri čemu je Z_i duljina okomice spuštene iz centra sfere na i -tu ravninu.

$$ASA = \sum_{i=1}^n \frac{R * D * L_i}{\sqrt{R^2 - Z_i^2}} \quad 2.2.$$

$$D = \frac{zslice}{2} + \Delta Z \quad 2.3.$$

$$\Delta Z = \min \left\{ \frac{zslice}{2}, R - Z_i \right\} \quad 2.4.$$

Implementacija algoritma za računanje proteinskih interakcija

Otapalu dostupno područje površine aminokiseline dobije se tako da se zbroji ASA vrijednost svih atoma koji grade tu aminokiselinu. To područje površine naziva se ukupna ASA (engl. *total ASA*), a postoje još četiri vrijednosti ASA-e:

- ASA glavnog lanca – suma ASA svih atoma koje grade glavni lanac aminokiseline
- ASA bočnog lanca – suma ASA svih atoma koje grade bočni lanac aminokiseline
- ASA nepolarnog dijela – suma ASA svih nepolarnih atoma (svi osim dušika i kisika) koji grade aminokiselinu
- ASA polarnog dijela – suma ASA svih polarnih atoma (dušika i kisika) koji grade aminokiselinu.

U Tablica 2.2. prikazane su standardne vrijednosti ASA-e. Vrijednosti su dobivene korištenjem programa PSAIA ([3]) i izražene su u Å².

Vrsta aminokiseline	Ukupna ASA	ASA glavnog lanca	ASA bočnog lanca	ASA nepolarnog dijela	ASA polarnog dijela
ALA	107,24	43,32	63,92	76,06	31,17
ARG	233,01	36,86	196,15	86,30	146,71
ASN	150,85	36,46	114,39	42,77	108,08
ASP	144,06	36,15	107,91	49,57	94,49
CYS	131,46	36,12	95,34	104,07	27,40
GLN	177,99	34,24	142,76	62,78	115,21
GLU	171,53	35,75	135,77	70,72	100,81
GLY	80,54	80,54	0,00	42,63	37,92

Vrsta aminokiseline	Ukupna ASA	ASA glavnog lanca	ASA bočnog lanca	ASA nepolarnog dijela	ASA polarnog dijela
HIS	180,93	34,84	149,09	106,27	74,66
ILE	173,40	31,08	142,33	147,36	26,04
LEU	177,87	32,68	145,19	150,10	27,76
LYS	196,14	35,55	160,59	123,55	72,65
MET	186,80	34,06	152,74	159,15	27,65
PHE	200,93	33,67	167,26	174,16	26,77
PRO	133,78	32,90	100,88	112,82	20,96
SER	115,30	40,86	74,44	52,93	62,37
THR	136,59	34,14	102,45	80,39	56,20
TRP	240,12	32,51	207,61	186,22	53,90
TYR	213,21	33,59	179,62	143,97	69,25
VAL	149,34	31,33	118,01	123,43	25,91

Tablica 2.2. Vrijednosti ASA-e za aminokiseline

Dostupnost otapalu određuje stupanj interakcije ostatka aminokiseline (engl. *Amino acid residue*) s molekulama otapala.

Energija otapala ili solvatacije je energija koja se oslobodi kod stvaranja interakcija između čestica otapala i čestica topljive tvari, odnosno u procesu privlačenja i pridruživanja molekula otapala s molekulama ili ionima u otopini.

Energiju otapala (ΔG_s) računamo za aminokiseline i baze, a ne i za metale i ostale atome koji mogu biti dio proteina. Ukupna energija otapala dobije se množenjem potencijala otapanja pojedine aminokiseline s omjerom površine dostupne otapalu cijele aminokiseline i totalne površine dostupne otapalu. (2.5.)

$$\Delta G_S = \frac{ASA_{ALL}}{ASA_{TOT}} * p \quad 2.5.$$

U navedenoj formuli ASA_{ALL} predstavlja područje površine dostupne otapalu molekule bez atoma vodika, odnosno zbroj ASA svih atoma koje grade aminokiselinu, ASA_{TOT} predstavlja standardnu ASA vrijednost aminokiseline, a p predstavlja potencijal otapala.

3. Podaci

3.1. PDB datoteke

Podatke za analizu proteinskih interakcija saznajemo iz PDB datoteka koje je moguće dobiti iz baze Protein Data Bank, baze informacija o strukturama bioloških molekula, uključujući proteine i nukleinske kiseline. U zapisima PDB baze nalaze se podaci o prostornim koordinatama svih atoma u proteinu osim atoma vodika. Podaci o strukturi molekula dobiveni su metodama NMR spektroskopije (magnetska rezonanca) i rendgenske difrakcije te objavljeni na internetu. PDB format zapisa strukture proteina podijeljen je na dva glavna dijela: zaglavlje, s općim podacima o proteinu i zapis koordinata atoma u prostoru.

Svaka linije PDB datoteke sastoji se od maksimalno 80 znakova. Prvih šest znakova svakog retka opisuje sadržaj tog retka unaprijed određenim riječima, poravnatim lijevo i pisanim velikim tiskanim slovima. To su na primjer riječi ATOM, HETATM, MODEL; TITLE; HEADER. Oznaka ATOM određuje početak zapisa podataka o jednom atomu aminokiseline ili nukleinske kiseline, a oznaka HETATM određuje početak zapisa o vodi, metalu ili ligandu. Kraj PDB datoteke označava se riječju END. Primjer jedne PDB datoteke dan je na Slika 3.1.

HEADER	HYDROLASE				30-JUN-09				3I3B		
TITLE	E.COLI (LACZ) BETA-GALACTOSIDASE (M542A) IN COMPLEX WITH D-										
TITLE	2 GALACTOPYRANOSYL-1-ON										
ATOM	1	N	ARG	A	13	-4.140	-57.397	-5.769	1.00	34.08	N
ATOM	2	CA	ARG	A	13	-4.773	-56.197	-5.144	1.00	35.29	C
ATOM	3	C	ARG	A	13	-5.381	-55.215	-6.136	1.00	34.94	C
ATOM	4	O	ARG	A	13	-5.892	-54.165	-5.736	1.00	35.87	O
ATOM	5	CB	ARG	A	13	-3.761	-55.437	-4.294	1.00	34.68	C
ATOM	6	CG	ARG	A	13	-3.599	-55.971	-2.904	1.00	35.47	C
ATOM	7	CD	ARG	A	13	-2.868	-54.958	-2.064	1.00	34.73	C
ATOM	8	NE	ARG	A	13	-2.559	-55.478	-0.742	1.00	35.81	N
ATOM	9	CZ	ARG	A	13	-1.833	-54.828	0.158	1.00	33.88	C
ATOM	10	NH1	ARG	A	13	-1.345	-53.629	-0.129	1.00	31.85	N
ATOM	11	NH2	ARG	A	13	-1.588	-55.386	1.335	1.00	31.33	N
ATOM	12	N	ARG	A	14	-5.315	-55.541	-7.422	1.00	33.12	N
ATOM	13	CA	ARG	A	14	-5.875	-54.669	-8.453	1.00	31.81	C

.....

Slika 3.1. Primjer PDB datoteke

Nakon oznaka ATOM i HETATM dolaze stupci s raznim informacijama o svakom atomu, kao što su:

- Redni broj atoma u lancu
- Naziv atoma koji određuje njegovu ulogu i molekuli
- Naziv molekule (aminokiseline)
- Oznaku lanca proteina u kojem se atom nalazi
- x, y i z koordinate atoma u prostoru (mjerna jedinica je angstrom)
- Faktor zauzeća (engl. occupancy) i temperaturni faktor
- Naziv elementa

Pri računanju elektrostatske sile između aminokiselina koriste se nazivi atoma, oznaka lanca proteina te x, y i z koordinate atoma, i tome prilagođene PDB datoteke koje sadrže za to posebne podatke.

3.2. Naboji atoma i otapalu dostupna područja površine

Podatci o nabojima pojedinih atoma u aminokiselini preuzeti su iz datoteke amberCharge.txt koja je nastala obradom podataka o atomima svake aminokiseline i sadrži oznaku aminokiseline, oznaku atoma te naboj atoma (Slika 3.2.).

```
ACE HH31 0.076010
ACE CH3 -0.190263
ACE HH32 0.076010
ACE HH33 0.076010
ACE C 0.512403
ACE O -0.550170
ALA N -0.404773
ALA H 0.294276
ALA CA -0.027733
ALA HA 0.120802
ALA CB -0.229951
ALA HB1 0.077428
ALA HB2 0.077428
ALA HB3 0.077428
...
```

Slika 3.2. Dio datoteke s nabojima atoma

Podaci potrebni za izračun energije otapala preuzimaju se iz datoteke asa.txt koja sadrži oznaku aminokiseline te zatim redom potencijal otapala, vrijednost totalne ASA-e, ukupne ASA-e svih atoma aminokiseline i ASA-e bočnog lanca aminokiseline. Datoteka je prikazana na Slika 3.3.

ALA	0.42	80	107.24	63.92
ARG	-1.37	216	233.01	196.15
ASN	-0.79	131	150.85	114.39
ASP	-2.46	123.9	144.06	107.91
CYS	1.39	108.4	131.46	95.34
GLN	-0.30	159.4	177.99	142.76
GLU	-2.35	152	171.53	135.77
GLY	0.0	43.3	80.54	0.00
HIS	0.18	164.3	180.93	146.09
ILE	2.46	158.1	173.40	142.33
LEU	2.30	159.3	177.87	145.19
LYS	-1.35	188.3	196.14	160.59
MET	1.68	166.9	186.80	152.74
PHE	2.44	187.0	200.93	167.26
PRO	0.67	119.8	133.78	100.88
SER	-0.05	93.2	115.30	74.44
THR	0.35	119.9	136.59	102.45
TRP	3.07	228.5	240.12	207.61
TYR	1.31	203.1	213.21	179.62
VAL	1.66	133.1	149.34	118.01

Slika 3.3. Datoteka s ASA vrijednostima i potencijalom otapala

3.3. Van der Waalsovi radijusi atoma

Podaci o Van der Waalsovima radijusima atoma učitavaju se iz datoteke chothia.radii. U njoj se redom nalaze oznaka aminokiselina, oznaka atoma i Van der Waalsov radijus tog atoma te aminokiseline.

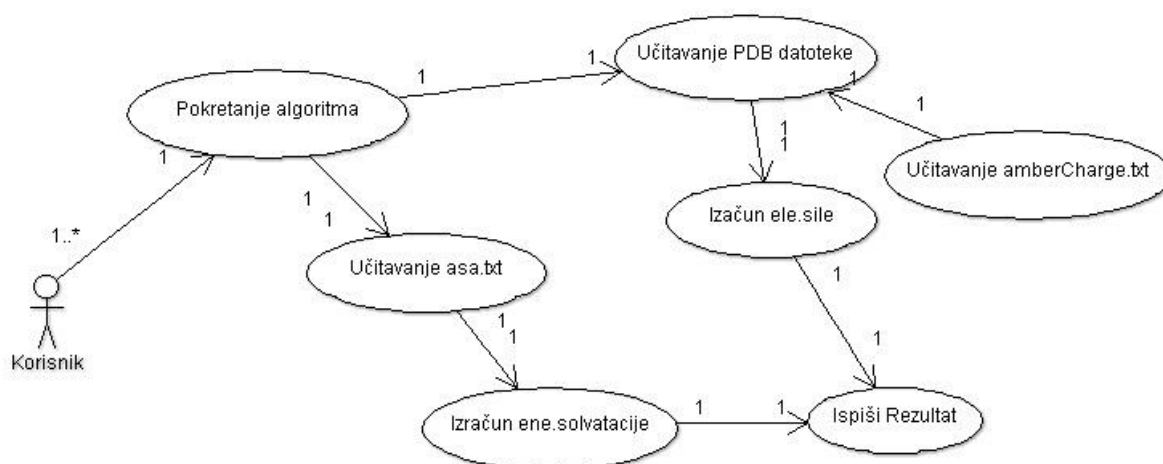
ALA	N	1.65
ALA	CA	1.87
ALA	C	1.76
ALA	O	1.40
ALA	CB	1.87
ARG	N	1.65
ARG	CA	1.87
ARG	C	1.76
ARG	O	1.40
ARG	CB	1.87

.....

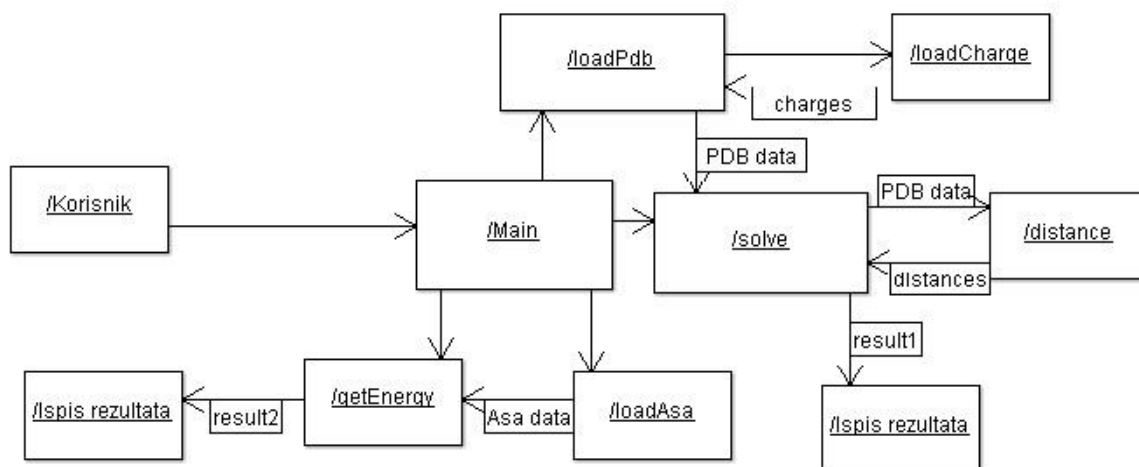
Slika 3.4. Datoteka s vrijednostima Van der Waalsovih radijusa

4. Implementacija

Algoritam je napisan u programskom jeziku C++. Pri pokretanju algoritma potrebno mu je zadati četiri argumenta, put do datoteke s nabojima, put do PDB datoteke za koju se želi izračunati elektrostatska sila, put do datoteke s ASA vrijednostima i put do datoteke s vrijednostima Van der Waalsovih radijusa. Dijagram obrazaca (engl. *use case*) algoritma prikazan je na Slika 4.1., a dijagram toka podataka (engl. *data flow*) na Slika 4.2.



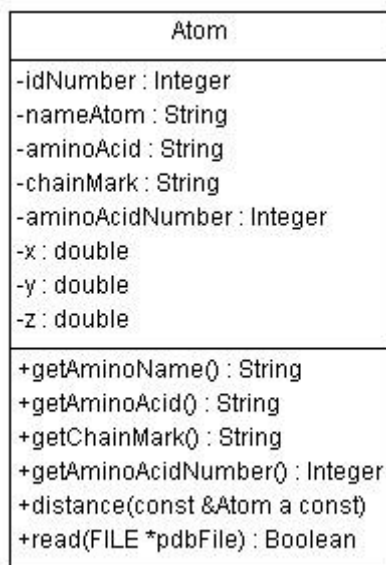
Slika 4.1. Dijagram obrazaca



Slika 4.2. Dijagram toka podataka

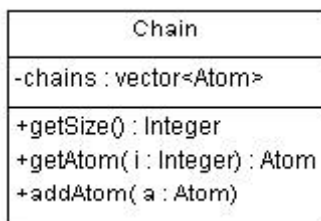
4.1. Opis algoritma

Algoritam je podijeljen na dva dijela, jedan za računanje elektrostatske sile između aminokiselina i drugi koji računa energiju otapala. U main funkciji provjerava se ispravnost argumenata i njihovog broja, te se oni prosleđuju metodama klasa koje ih koriste. Za svaki redak PDB datoteke koja se učitava, stvara se jedna instanca klase Atom (Slika 4.3.), i spremaju se podaci potrebni za računanje elektrostatske sile. Učitavanje iz PDB datoteke opisano je u metodi *read* klase Atom. Učitavaju se podaci o imenu, rednom broju atoma, oznaci aminokiseline, oznaci lanca, rednom broju aminokiseline u lancu i koordinatama atoma u prostoru. Svi su atributi klase Atom privatni i dohvaćaju se pomoću metoda *getAtomName*, *getAminoAcid*, *getChainMark* i *getAminoAcidNumber*. Kako bi se mogla izračunati elektrostatska sila, potrebno je izračunati udaljenosti svaka dva atoma aminokiseline (svaki atom sa svakim). To računanje odvija se u funkciji *distance* koja prema koordinatama računa udaljenosti.



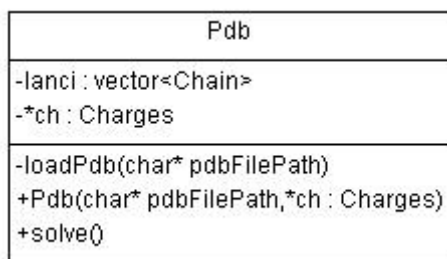
Slika 4.3. Dijagram klase Atom

Klasa koja sadrži vektor svih atoma jednog lanca naziva se Chain i pomoću nje u algoritam dohvaća atome pri računanju. Metoda za dohvaćanje pojedine instance klase Atom je *getAtom*, metoda koja dodaje Atom u vektor je *addAtom*, a metoda koja određuje veličinu lanca je *getSize*. Dijagram klase Chain prikazan je na Slika 4.4.



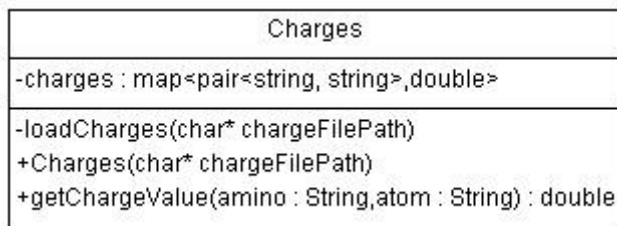
Slika 4.4. Dijagram klase Chain

Klasa Pdb (Slika 4.5.) sadrži svaki lanac aminokiselina koji se nalazi u učitanoj PDB datoteci. Sve PDB datoteke za koje će rezultati biti prikazani kasnije sadrže po dva lanca proteina. Klasa Pdb osim atributa *lanca* koji je vektor instanci klase Chain sadrži i referencu na instancu klase Charges u kojoj su spremljeni učitani naboji atoma. U *main* funkciji se konstruktorom klase Pdb stvara instanca klase Charges, koja učitava naboje svih atoma iz datoteke amberCharge.txt, i poziva se učitavanje PDB datoteke. Nakon već navedenog učitavanja u instance klase Atom, pokreće se *solve* funkcija koja računa ukupnu elektrostatsku silu interakcije dviju aminokiselina.



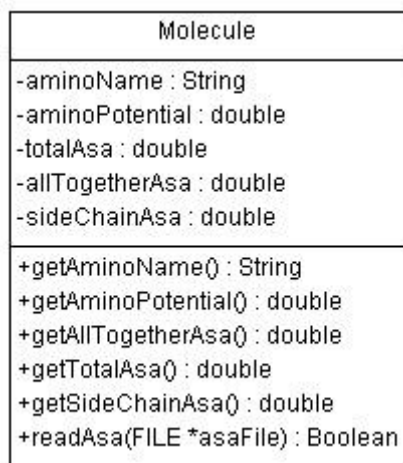
Slika 4.5. Dijagram klase Pdb

Naboje koju su potrebni za algoritam učitavamo iz datoteke s nabojsima u mapu, koja svakom paru (ime aminokiseline, ime atoma) dodjeljuje vrijednost naboja, učitavši je iz datoteke. Mapa je atribut klase Charges, a vrijednost naboja se kasnije u *solve* metodi klase Pdb dohvaća metodom *getChargeValue*. Tako spremljeni naboji množe se za svaka dva odgovarajuća atoma, dijele s kvadratom udaljenosti tih atoma i množe s konstantom c ($c = 332$). Elektrostatska sila između odgovarajućih aminokiselina dobiva se zbrajanjem elektrostatskih sila između svaka dva atoma aminokiseline. Na dijagramu klase Charges (Slika 4.6.) vidi se i konstruktor te klase, kao i metoda *loadCharges* u kojoj se učitavanje naboja odvija.



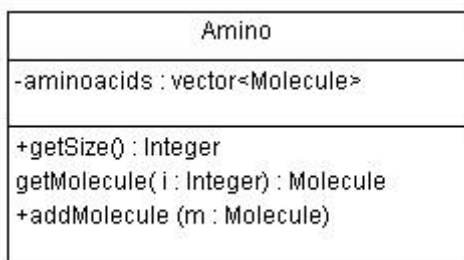
Slika 4.6. Dijagram klase Charges

Drugi dio algoritma je računanje energije otapala. Za svaki redak datoteke s ASA vrijednostima stvara se instanca klase Molecule koja sadrži podatke o imenu aminokiseline, potencijalu otapala, ukupnoj vrijednosti površine dostupne otapalu aminokiseline, podatak o ukupnoj površini svih atoma te aminokiseline i podatak o površini dostupnoj otapalu bočnog lanca aminokiseline. Ti se podaci učitavaju u privatne attribute u metodi klase *read* kojoj je ulazni argument datoteka sa svim potrebnim podacima. Kasnije se u algoritmu dohvaćaju pomoću metoda *getAminoName*, *getAminoPotential*, *getAllTogetherAsa*, *getSideChainAsa* i *getTotalAsa*. Dijagram klase Molecule prikazan je na Slika 4.7.



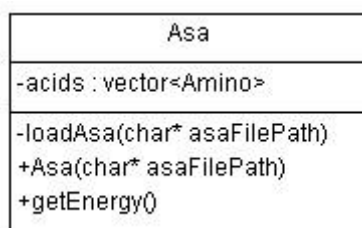
Slika 4.7. Dijagram klase Molecule

Klasa koja sadrži vektor *aminoAcids* svih učitanih aminokiselina naziva se Amino (Slika 4.8.) i pomoću nje se pronalaze podaci o pojedinoj aminokiselini pri računanju energije otapala. Veličina vektora saznaje se metodom *getSize*, podaci o pojedinoj aminokiselini se dohvaćaju metodom *getMolecule*, a u vektor se instanca klase Molecule dodaje metodom *addMolecule*.



Slika 4.8. Dijagram klase Amino

Instanca klase *Asa* sadrži vektor svih aminokiselina. Pojedina instanca te klase ima konstruktor *Asa* koji poziva metodu *loadAsa*, u kojoj se podaci o nabojima atoma učitavaju, i metodu *getEnergy*, koja računa energiju otapala svaka solvatacije za svaku aminokiselinu. U *main* funkciji se računanje energije solvatacije počinje stvaranjem instance klase *Asa* pomoću njezinog konstruktora i pozivom metode *getEnergy*. Dijagram klase *Asa* prikazan je na Slika 4.9.



Slika 4.9. Dijagram klase *Asa*

Opisani dio algoritma računa vrijednosti otapala na osnovu već definiranih ASA vrijednosti pojedinih aminokiselina, a taj izračun napravljen je i pomoću algoritma preuzetog od Josipa Mihela za računanje ASA vrijednosti svakog atoma. Vrijednosti ASA za aminokiselinu se dobiju zbrajanjem vrijednosti ASA-e svih atoma.

5. Instalacija i pokretanje

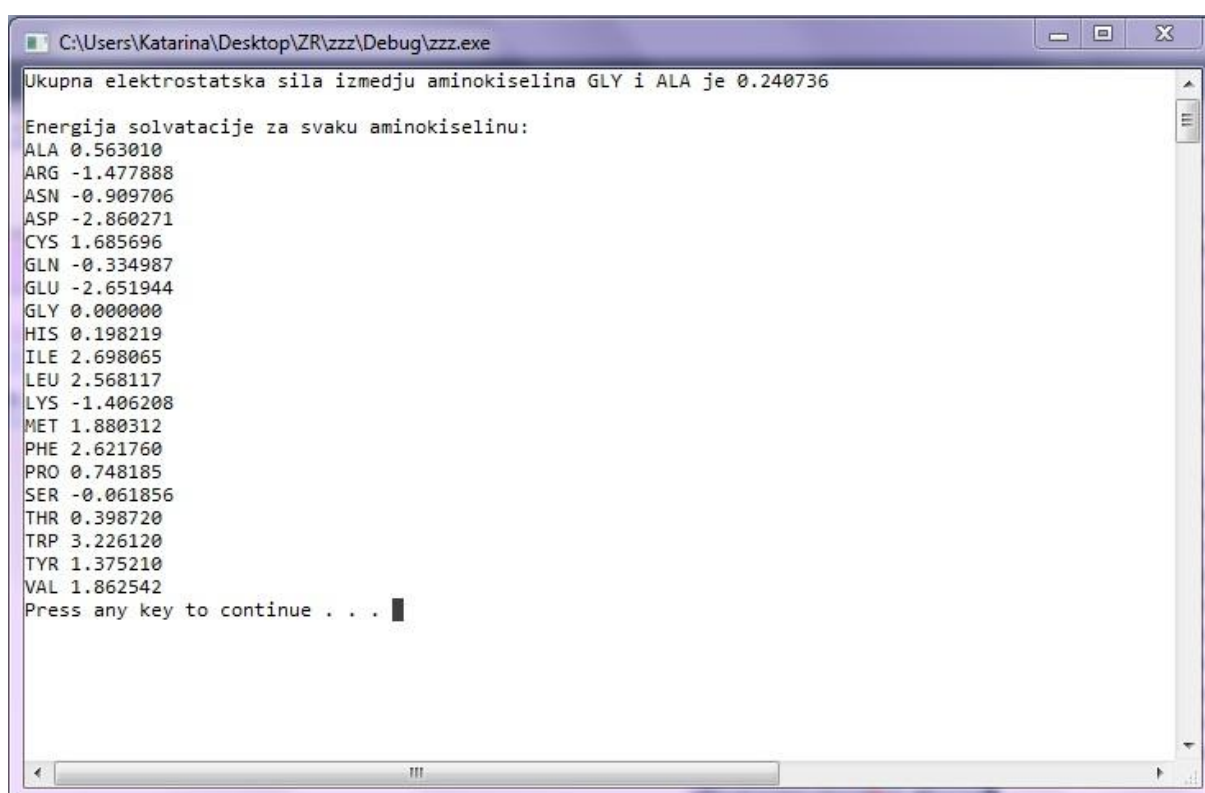
Implementirani algoritam može se pokrenuti na bilo kojem računalu i nema posebnih zahtjeva. Direktorij aplikacije sadrži tri direktorija: *dana*, *obj* i *src*. U direktoriju *data* nalaze se sve potrebne datoteke koje se zadaju kao argumenti algoritma. U direktoriju *src* nalazi se direktorij *ext* koji sadrži dio koda preuzetog iz programa Josipa Mihela za računanje ASA vrijednosti aminokiseline.

Ukoliko računalo ima neku vrstu UNIX platforme, nužno je da ima g++ kompajler i da prije se pokretanja algoritma instaliraju (naredbom *sudo apt -get install*) potrebni *library* dokumenti, a to su: *libboost -dev* i *libboost -regex -dev*. Za kompajliranje koda potrebno se nalaziti u direktoriju aplikacije, te upisati naredbu make. Program se pokreće naredbom ./mcpu sa sljedećim argumentima: putanja do *amberCharge.txt* datoteke, putanja do *PDB* datoteke, putanja do *asa.txt* datoteke i putanja do *chothia.radii* datoteke. Na UNIX x64 operacijskom sustavu potrebno je prije naredbe make upisati naredbu make clean. Kako

Ukoliko računalo ima Windows operativni sustav, algoritam se može pokrenuti iz neke od razvojnih okolina C++ programskog jezika kao što su programi Visual Studio, Visual Express te DevC++.

6. Rezultati

Algoritam kao rezultate prikazuje izračunatu elektrostatsku silu između dviju aminokiselina (koje su određene ulaznom datotekom) u interakciji te izračunate energije otapala pojedine aminokiseline. Primjer ispisa rezultata algoritma za dvije određene aminokiseline, prikazan je na Slika 6.1. Prva rečenica ispisa prikazuje za interakciju kojih se dviju aminokiselina računala elektrostatska sila, te njezinu vrijednost. Daljnji ispis prikazuje izračunate energije otapala pojedine aminokiseline zapisane za svaku aminokiselinu u novi redak.



```
C:\Users\Katarina\Desktop\ZR\zzz\Debug\zzz.exe
Ukupna elektrostatska sila izmedju aminokiselina GLY i ALA je 0.240736
Energija solvatacije za svaku aminokiselinu:
ALA 0.563010
ARG -1.477888
ASN -0.909706
ASP -2.860271
CYS 1.685696
GLN -0.334987
GLU -2.651944
GLY 0.000000
HIS 0.198219
ILE 2.698065
LEU 2.568117
LYS -1.406208
MET 1.880312
PHE 2.621760
PRO 0.748185
SER -0.061856
THR 0.398720
TRP 3.226120
TYR 1.375210
VAL 1.862542
Press any key to continue . . .
```

Slika 6.1. Primjer ispisa rezultata

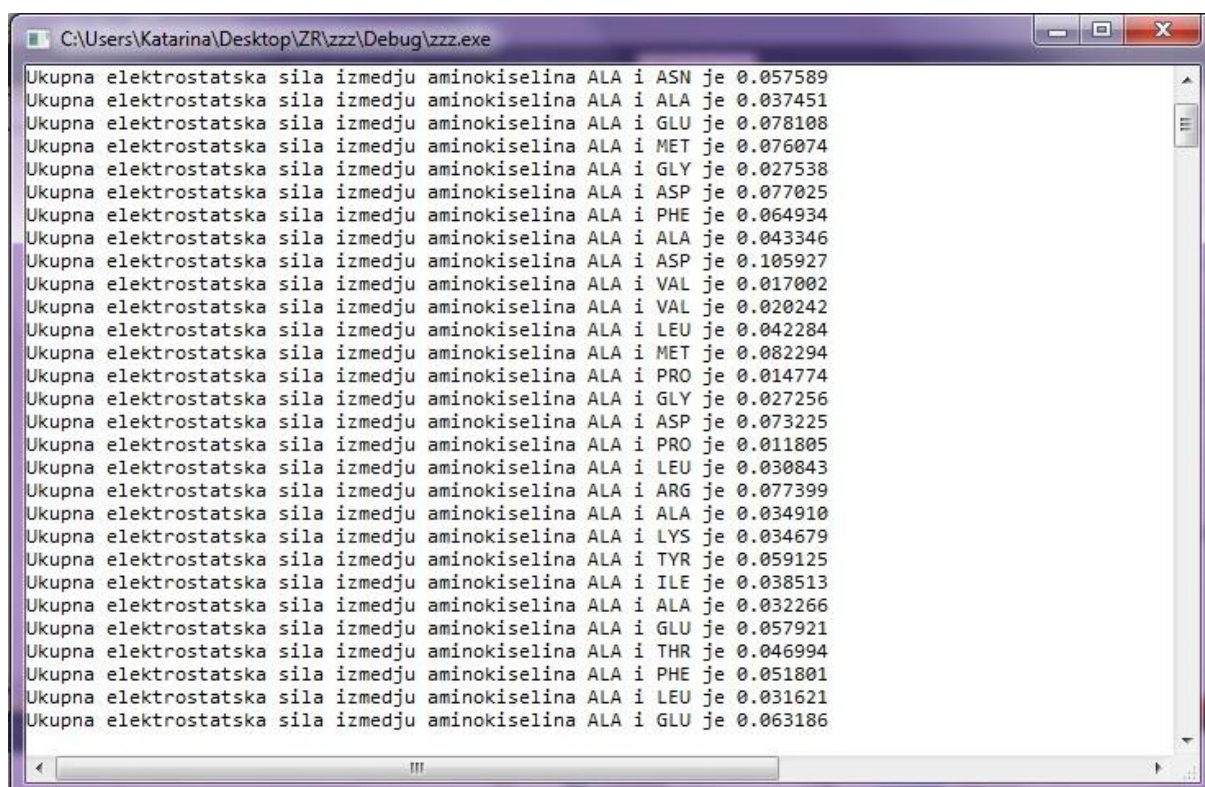
Za testne primjere interakcija aminokiselina u Tablica 6.1. su prikazane vrijednosti izračunate elektrostatske sile. Testno primjeri PDB datoteka su prilagođeni izračunu kako bi se vrijednosti elektrostatske energije mogle provjeriti. Prvi i drugi stupac tablice označavaju aminokiseline koja sudjeluju u interakciji te oznake i redni brojeve atoma tih aminokiselina koji najviše doprinose njihovoj međusobnoj interakciji. Treći stupac prikazuje koliko su ta dva atoma udaljena, i ta je udaljenost najmanja za te aminokiseline. Udaljenost je prikazana u angstromima ($1\text{\AA} = 10^{-10}\text{ m}$).

Četvrti stupac prikazuje izračunatu ukupnu elektrostatsku silu za pojedinu interakciju aminokiselina.

Aminokiselina	Aminokiselina	Udaljenost	Elektrostatska sila
GLU – 21 OE1	ASN – 53 HD22	3 Å	-3.28887
GLU – 21 OE1	ASN – 53 HD22	4 Å	-1.64684
GLU – 21 OE1	ASN – 53 HD22	5 Å	-0.98036
GLU – 21 OE1	ARH – 63 HH22	3 Å	-13.22474
GLU – 21 OE1	ARH – 63 HH22	4 Å	-8.52762
GLU – 21 OE1	ARH – 63 HH22	5 Å	-6.06626
GLU – 50 OE1	LYS - 69 HZ1	3 Å	-15.26266
GLU – 50 OE1	LYS - 69 HZ1	4 Å	-9.75089
GLU – 50 OE1	LYS - 69 HZ1	5 Å	-6.85306
GLU – 21 OE1	TYR – 56 HH	3 Å	-3.42107
GLU – 21 OE1	TYR – 56 HH	4 Å	-1.66138
GLU – 21 OE1	TYR – 56 HH	5 Å	-0.95908
ALA – 19 HB3	TYR – 57 HH	3 Å	-0.03352
ALA – 19 HB3	TYR – 57 HH	4 Å	-0.03369
ALA – 19 HB3	TYR – 57 HH	5 Å	-0.02443
GLY – 12 CA	GLY – 36 CA	3 Å	1.36264
GLY – 12 CA	GLY – 36 CA	5 Å	0.29003
GLY – 12 CA	ALA – 38 CB	3 Å	1.03633
GLY – 12 CA	ALA – 38 CB	5 Å	0.24074

Tablica 6.1. Elektrostatske sile interakcije aminokiselina

Primjer ispisa rezultata za primjer PDB datoteke preuzete s interneta, imenom 1ECP, dan je na Slika 6.2. Navedena PDB datoteka sadrži 6 lanaca, od kojih svaki ima 1793 atoma raspodijeljenih u 237 aminokiselina. Algoritam računa elektrostatske sile između svake dvije aminokiseline različitih lanaca. Dakle, za svaku aminokiselinu prvog lanca, računaju se udaljenosti sa svakom aminokiselinom drugog lanca (njih 237) i tako za svih preostalih 5 lanaca. Vrijeme izvođenja ovakvog algoritma je poprilično dugo, ali s obzirom na velik broj atoma to je predvidivo. Vrijeme izračuna za jedan par aminokiselina je otprilike jedna i pol sekunda, što za datoteku sa 1422 aminokiseline daje dugotrajno izvođenje algoritma.



```
C:\Users\Katarina\Desktop\ZR\zzz\Debug\zzz.exe
Ukupna elektrostatska sila izmedju aminokiselina ALA i ASN je 0.057589
Ukupna elektrostatska sila izmedju aminokiselina ALA i ALA je 0.037451
Ukupna elektrostatska sila izmedju aminokiselina ALA i GLU je 0.078108
Ukupna elektrostatska sila izmedju aminokiselina ALA i MET je 0.076074
Ukupna elektrostatska sila izmedju aminokiselina ALA i GLY je 0.027538
Ukupna elektrostatska sila izmedju aminokiselina ALA i ASP je 0.077025
Ukupna elektrostatska sila izmedju aminokiselina ALA i PHE je 0.064934
Ukupna elektrostatska sila izmedju aminokiselina ALA i ALA je 0.043346
Ukupna elektrostatska sila izmedju aminokiselina ALA i ASP je 0.105927
Ukupna elektrostatska sila izmedju aminokiselina ALA i VAL je 0.017002
Ukupna elektrostatska sila izmedju aminokiselina ALA i VAL je 0.020242
Ukupna elektrostatska sila izmedju aminokiselina ALA i LEU je 0.042284
Ukupna elektrostatska sila izmedju aminokiselina ALA i MET je 0.082294
Ukupna elektrostatska sila izmedju aminokiselina ALA i PRO je 0.014774
Ukupna elektrostatska sila izmedju aminokiselina ALA i GLY je 0.027256
Ukupna elektrostatska sila izmedju aminokiselina ALA i ASP je 0.073225
Ukupna elektrostatska sila izmedju aminokiselina ALA i PRO je 0.011805
Ukupna elektrostatska sila izmedju aminokiselina ALA i LEU je 0.030843
Ukupna elektrostatska sila izmedju aminokiselina ALA i ARG je 0.077399
Ukupna elektrostatska sila izmedju aminokiselina ALA i ALA je 0.034910
Ukupna elektrostatska sila izmedju aminokiselina ALA i LYS je 0.034679
Ukupna elektrostatska sila izmedju aminokiselina ALA i TYR je 0.059125
Ukupna elektrostatska sila izmedju aminokiselina ALA i ILE je 0.038513
Ukupna elektrostatska sila izmedju aminokiselina ALA i ALA je 0.032266
Ukupna elektrostatska sila izmedju aminokiselina ALA i GLU je 0.057921
Ukupna elektrostatska sila izmedju aminokiselina ALA i THR je 0.046994
Ukupna elektrostatska sila izmedju aminokiselina ALA i PHE je 0.051801
Ukupna elektrostatska sila izmedju aminokiselina ALA i LEU je 0.031621
Ukupna elektrostatska sila izmedju aminokiselina ALA i GLU je 0.063186
```

Slika 6.2. Primjer ispisa rezultata za 1ECP.pdb

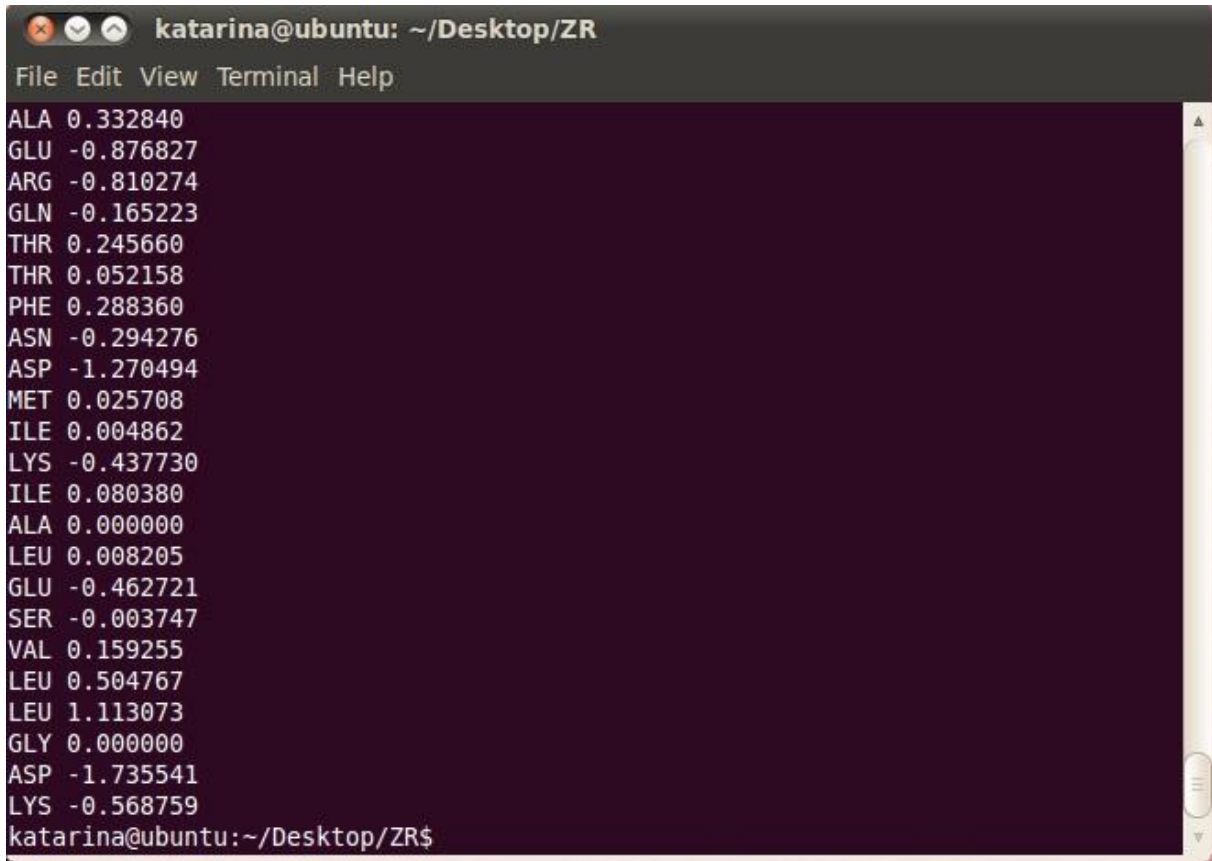
Vrijednosti energija otapala za pojedinačnu aminokiselinu prikazane su u Tablica 6.2.

Implementacija algoritma za računanje proteinskih interakcija

Oznaka aminokiseline	Energija otapala
ALA	0.563010
ARG	-1.477888
ASN	0.909706
ASP	-2.860271
CYS	1.685696
GLN	-0.334987
GLU	-2.651744
GLY	0.000000
HIS	0.198219
ILE	2.698065
LEU	2.568117
LYS	-1.406208
MET	1.880312
PHE	2.621760
PRO	0.748185
SER	-0.061856
THR	0.398720
TRP	3.226120
TYR	1.375210
VAL	1.862542

Tablica 6.2. Vrijednosti energije otapala

Na Slika 6.3. prikazan je primjer ispisa računanja energije otapala za aminokiseline iz PDB datoteke 1ECP.pdb korištenjem koda Josipa Mihela pri računanju ukupnog područja površine dostupne otapalu aminokiseline. Trajanje tog dijela algoritma je otprilike sedam do osam sekundi za PDB datoteku.



```
katarina@ubuntu: ~/Desktop/ZR
File Edit View Terminal Help
ALA 0.332840
GLU -0.876827
ARG -0.810274
GLN -0.165223
THR 0.245660
THR 0.052158
PHE 0.288360
ASN -0.294276
ASP -1.270494
MET 0.025708
ILE 0.004862
LYS -0.437730
ILE 0.080380
ALA 0.000000
LEU 0.008205
GLU -0.462721
SER -0.003747
VAL 0.159255
LEU 0.504767
LEU 1.113073
GLY 0.000000
ASP -1.735541
LYS -0.568759
katarina@ubuntu:~/Desktop/ZR$
```

Slika 6.3. Primjer ispisa rezultata za 1ECP.pdb

7. Zaključak

Razumijevanje bioloških interakcija pomaže biolozima, kemičarima i običnim ljudima razumjeti funkcioniranje svih živih organizama. Kako su proteini jedni od najvažnijih tvari u ljudskom organizmu, zanimaju nas interakcije proteina i do kakvih one reakcija dovode u tijelu. Cilj ovog algoritma je utvrđivanje elektrostatske sile pri interakciji dviju aminokiselina i utvrđivanje energije otapala pomoću površine dostupne otapalu pojedine aminokiseline. Što više novih saznanja dobijemo o proteinima, lakše ćemo moći predvidjeti neke od mogućih interakcija i to znanje upotrijebiti kako bi se razjasnili nastanci nekih bolesti, predvidjele moguće mutacije određenih uzročnika tih bolesti, kako bi se za tu bolest pronašao lijek i sl. Time se pridonosi razvoju mnogih grana medicine, farmacije, biokemije i na kraju razvoju gospodarstva, a time se produžuje životni vijek i poboljšava životni standard ljudi.

8. Sažetak

Implementacija algoritma za računanje proteinskih interakcija

Algoritam implementiran ovim radom računa elektrostatsku silu između dvije aminokiseline i energije solvatacije za pojedinu aminokiselinu. Elektrostatska sila računa se kao zbroj elektrostatskih sila pri međusobnim interakcijama svih atoma određene dvije aminokiseline. Podaci potrebni za računanje dobivaju se učitavanjem PDB datoteka i učitavanjem naboja svakog atoma u određenoj aminokiselini. Za izračun energije solvatacije svi potrebni podaci učitavaju se iz datoteke s vrijednostima površina dostupnih otapalu i potencijalom otapala svake aminokiseline.

Ključne riječi: proteini, elektrostatska sila, energija solvatacije, ASA, aminokiseline

Summary

Implementing the algorithm for computing protein interactions

The algorithm implemented in this thesis calculates electrostatic energy between two amino acids and solvents energy for each amino acid. Electrostatic energy is calculated as the sum of electrostatic energies in the interaction of all two pairs of atoms in those two amino acids. The data needed to calculate the electrostatic energy is loaded from PDB file and file with the charges of each atom in a certain amino acid. To calculate solvents energy all required data is retrieved from a file with the values of accessible solvent area and solvents potential for each amino acid.

Keywords: proteins, electrostatic energy, solvents energy, ASA, amino acids

9. Literatura

- [1] Internet: RCSB Protein Data Bank, <http://www.pdb.org/pdb/home/home.do>, svibanj 2011.
- [2] Internet: IUPAC – Nomenclature and symbolism for Amino Acids, <http://www.chem.qmul.ac.uk/iupac/AminoAcid/AA1n2.html#AA1>, svibanj 2011.
- [3] J.Mihel, M. Šikić, B. Jeren, K. Vlahoviček, "PSAIA – Protein Structure and Interaction Analyzer", Sveučilište u Zagrebu, 2008.
- [4] Mihel Josip, Alat za analizu površina proteina i mjesta proteinskih interakcija, Diplomski rad br. 1044, FER, 2006.