

SVEUČILIŠTE U ZAGREBU  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 572

**RAČUNALNA METODA ZA ODREĐIVANJE  
PROTEINSKIH INTERAKCIJA**

Katarina Kušević

Zagreb, lipanj 2013.

*Zahvaljujem mentoru doc. dr. sc. Mili Šikiću na pomoći, vodstvu i strpljenju pri izradi diplomskog rada.*

*Od srca hvala mojim roditeljima, sestri i bratu na podršci tokom studiranja, Nineku za neprestani optimizam i gurkanje.*

*Hvala svim prijateljima i kolegama bez kojih studij ne bi bio tako zabavan i izazovan, a posebno Bruni koji mi je uvijek spremno pomagao.*

## Sadržaj

Popis slika i tablica .....	1
1. Uvod .....	3
2. Važnost računanja interakcija .....	4
3. Biokemija .....	6
3.1 Elektrostatska sila .....	7
3.2 Energija otapala .....	8
3.3 Korelacija .....	12
4. Formati podataka .....	14
4.1 PDB datoteke .....	14
4.2 Naboji atoma i otapalu dostupna područja površine .....	15
4.3 Van der Waalsovi radijusi atoma .....	17
4.4 Koeficijent privlačnosti .....	17
5. Implementacija .....	18
5.1 Opis algoritma .....	19
6. Instalacija i pokretanje .....	24
7. Grafičko sučelje i upute za korištenje .....	25
8. Rezultati .....	30
9. Zaključak .....	42
10. Sažetak .....	43
11. Literatura .....	45
Dodatak - Skice dijagrama klasa .....	47

## Popis slika i tablica

Slika 2.1. Četiri strukture proteina: primarna, sekundarna, tercijarna i kvartarna .....	4
Slika 3.1. Opća struktura aminokiselina .....	6
Slika 3.2. Otapalu dostupno područje površine molekule – ASA.....	9
Slika 3.3. Dužina slobodnog luka proteina u jednoj ravnini (označena plavom bojom).....	10
Slika 4.1. Primjer početka PDB datoteke <i>1a22.pdb</i> .....	15
Slika 4.2. Dio datoteke s nabojima atoma <i>amberCharge.txt</i> .....	16
Slika 4.3. Datoteka s ASA vrijednostima i potencijalom otapala <i>asa.txt</i> .....	16
Slika 4.4. Dio datoteke s vrijednostima Van der Waalsovih radijusa atoma <i>chothia.radii</i> .....	17
Slika 5.1. Dijagram obrazaca algoritma.....	18
Slika 5.2. Dijagram toka podataka algoritma .....	18
Slika 5.3. Dijagram klase Atom s pripadajućim atributima i metodama .....	19
Slika 5.4. Dijagram klase Chain s pripadajućim atributima i metodama .....	20
Slika 5.5. Dijagram klase Charges s pripadajućim atributima i metodama ..	20
Slika 5.6. Dijagram klase Pdb s pripadajućim atributima i metodama .....	21
Slika 5.7. Dijagram klase Molecule s pripadajućim atributima i metodama ..	22
Slika 5.8. Dijagram klase Asa s pripadajućim atributima i metodama .....	22
Slika 5.9. Dijagram klase Radius s pripadajućim atributima i metodama ...	23
Slika 5.10. Dijagram klase cAtomGrid s pripadajućim atributima i metodama .....	23
Slika 7.1. Grafičko sučelje aplikacije .....	25
Slika 7.2. Dijaloški okvir koji se pokreće pritiskom na gumb „Help“ .....	26
Slika 7.3. Dijaloški okvir koji se pokreće pritiskom na gumb „About“ .....	26
Slika 7.4. Dijaloški okvir za odabir datoteke koju se učitava.....	27
Slika 7.5. Izgled grafičkog sučelja prije pokretanja izračuna .....	27
Slika 7.6. Grafičko sučelje nakon izračuna energije solvatacije .....	28
Slika 7.7. Grafičko sučelje nakon računanja elektrostatske sile i energije solvatacije.....	29
Slika 8.1. Izgled grafičkog sučelja nakon izvršenog programa za testni primjer 1A22.pdb .....	30

Slika 8.2 Prikaz datoteke s rezultatima računanja elektrostatske sile između lanaca <i>Ele_force_between_chains.txt</i> .....	31
Slika 8.3. Prikaz datoteke s rezultatima računanja elektrostatske sile unutar lanca <i>Ele_force_inside_chain.txt</i> .....	32
Slika 8.4. Prikaz datoteke s rezultatima računanja energije solvatacije <i>Solvation_energy.txt</i> .....	33
Slika 8.5. Prikaz rezultata izračunate ukupne energije solvatacije proteina 1A22 u grafičkom sučelju .....	34
Slika 8.6. Prikaz izračunate elektrostatske sile proteina 1A22 u grafičkom sučelju .....	34
Slika 8.7. Prikaz korelacije zbroja elektrostatske sile i energije solvatacije proteina s prirodnim logaritmom koeficijenta privlačnosti točkastim dijagramom..	40
Slika 8.8. Prikaz korelacije elektrostatske sile proteina s negativnom vrijednošću prirodnog logaritma koeficijenta privlačnosti točkastim dijagramom ..	41
Tablica 3.1 Popis aminokiselina s međunarodnim i hrvatskim nazivom te troslovnim kraticama .....	7
Tablica 3.2 Vrijednosti standardnih vrijednosti otapalu dostupnih područja proteina (ASA) za aminokiseline .....	11
Tablica 8.1 Tablično ispisani rezultati izračuna elektrostatske sile i energije solvatacije, duljina trajanja izvođenja za više <i>pdb</i> datoteka.....	36
Tablica 8.2 Tablično ispisani podaci koeficijenta privlačnosti po <i>pdb</i> datotekama.....	38
Tablica 8.3 Vrijednosti izračunatih parametara analize rezultata algoritma	39
Skica A – Dijagram svih klasa paketa <i>proteinInteractions</i> – izračun elektrostatske sile.....	48
Skica B – Dijagram svih klasa paketa <i>surface</i> – izračun energije solvatacije .....	49

## **1. Uvod**

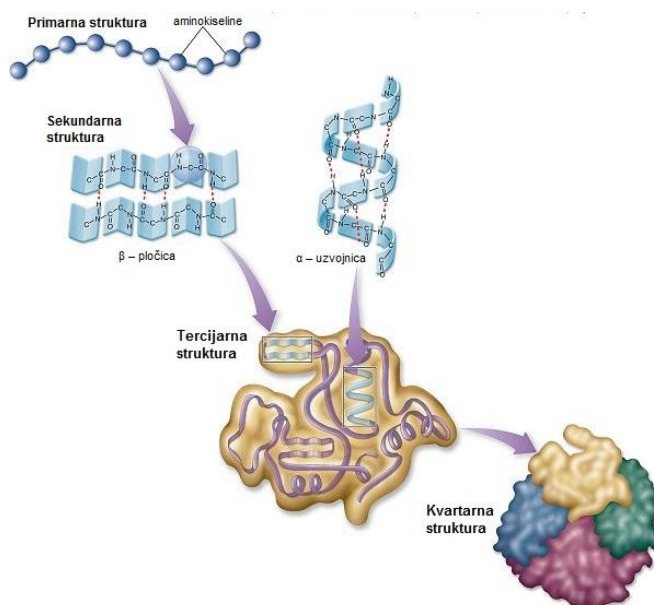
Bioinformatika je interdisciplinarno područje znanosti koje je usmjereno na primjenu računala i informacijskih tehnologija za pohranu, pristup, obradu i distribuciju informacija koji opisuju biološke molekule i procese. Bioinformatika je nastala zbog sve veće potrebe za suradnjom molekularne biologije, računarke znanosti i informacijskih tehnologija kako bi se omogućila što efikasnija obrada velike količine podataka. Bioinformatika obuhvaća tri bitna područja: razvoj novih algoritama koji se koriste za upravljanje velikim količinama podataka biološkog karaktera; analizu i interpretaciju različitih tipova podataka koji se odnose na sekvence aminokiselina, nukleotide i proteine; razvoj i implementaciju alata koji omogućuju pristup i upravljanje različitim tipovima podataka. [7]

U sklopu ovog rada potrebno je implementirati algoritam za izračun jačine elektrostatskih interakcija i energije otapanja. 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])). Nakon izračuna elektrostatske sile i energije otapanja, potrebno je rezultate korelirati s koeficijentom privlačnosti proteina.

U narednim poglavljima rada objašnjena je struktura proteina, važnost računanja interakcija, biokemijska pozadina, dan pregled podataka potrebnih za izračune, opisni su implementirani algoritmi, opisano grafičkog sučelja, dane su upute za instalaciju i korištenje te rezultati testiranja algoritma.

## 2. Važnost računanja interakcija

Proteini su makromolekule građene od različitih aminokiselina, međusobno povezanih u duge lance peptidnim vezama. Proteini su esencijalni spojevi u ljudskom organizmu koji izgrađuju sve stanice, poput mišića, kože, noktiju, unutarnjih organa. U ljudskom organizmu su jedna od najvažnijih tvari prisutnih, a sve proteine tvore kombinacije svega 20 aminokiselina. Specifične karakteristike i funkcije svakog proteina određuje broj, slijed i prostorni raspored aminokiselina u lancu proteina. Tako se tvore tisuće različitih proteina s različitim biološkim funkcijama. Za razumijevanje proteina s njihovim funkcijama potrebno je poznavati strukturu proteina koja se definira načinom savijanja (engl. *folding*) proteina. Postoje četiri strukture proteina; primarna, sekundarna, tercijska i kvartarna (Slika 2.1). Primarna struktura proteina je linearna sekvenca aminokiselina povezanih peptidnom vezom. Sekundarna struktura je lokalna prostorna organizacija aminokiselina u trodimenzionalnoj formi povezanih vodikovim vezama te ovisno o načinu na koji se aminokiseline saviju mogu biti  $\alpha$  – uzvojnica ili  $\beta$  – pločica. Tercijarna struktura je trodimenzionalna struktura polipeptidnog lanca dobivena kombinacijom sekundarnih struktura. Kvartarna struktura tj. proteinski kompleks je prostorni raspored više polipeptida koji tvore protein.



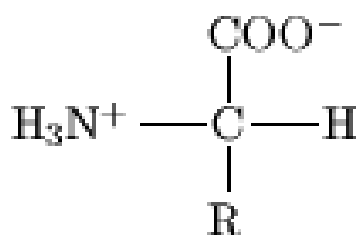
**Slika 2.1.** Četiri strukture proteina: primarna, sekundarna, tercijska i kvartarna

Proučavanje interakcija proteina važno je za shvaćanje funkcioniranja živih organizama i predstavlja informacije ključne za razumijevanje stabilnosti i funkcije proteina. Potreba za objašnjavanjem bioloških procesa i primjena novih spoznaja u drugim područjima kao što su sinteza umjetnih proteina sa željenim svojstvima dovodi do porasta važnosti računanja interakcija. Primjer primjene dobivenih spoznaja je računalno pomognuta sinteza lijekova i cjepiva, kao i projektiranje biokatalizatora i proteinskih terapija. Računanje proteinskih interakcija dovodi do boljeg razumijevanja djelovanja lijeka, a time i efikasnijeg liječenja. Razvoj pouzdanih i brzih algoritama za evaluaciju elektrostatskih energija biomolekula je važan čimbenik za ispravno razumijevanje cjelokupne funkcije proteina.[8]



### 3. Biokemija

Aminokiseline su organske molekule koje se sastoje od centralnog atoma ugljika, amino skupine  $\text{-NH}_2$ , vodikovog atoma  $\text{-H}$  i karboksilne skupine  $\text{-COOH}$  na koju se nadovezuje kiselinski ostatak, obično označavan slovom  $R$ . (Slika 3.1.) 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 3.1. 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. Aminokiseline se spajaju u velike lance polimera preko peptidnih veza. Proces povezivanja aminokiselina naziva se sinteza proteina. 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.

Pregled aminokiselina prema međunarodnoj udruzi IUPAC, koja standardizira imenovanje organskih spojeva (engl. *International Union of Pure and Applied Chemistry* [2]), dan u tablici Tablica 3.1.

**Tablica 3.1** Popis aminokiselina s međunarodnim i hrvatskim nazivom te troslovnim kraticama

IUPAC 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	Fenilalanin	PHE
Proline	Prolin	PRO
Serine	Serin	SER
Threonine	Treonin	THR
Tryptophan	Triptofan	TRP
Tyrosine	Tirozin	TYR
Valine	Valin	VAL

### 3.1 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, a zbog postojanja naboja prisutna je određena prostorna gustoća naboja.

Pri računanju elektrostatske sila između dva atoma  $F$  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  $r_{ij}$  između njih. (3.1)

$$F = c * \frac{q_i * q_j}{r_{ij}^2} \quad (3.1)$$

Elektrostatska sila aminokiseline je suma elektrostatskih sila atoma koji čine tu aminokiselinu. Ukupna elektrostatska sila  $\Delta G_d$  proteina [8] s N atoma računa se kao suma sila svih parova atoma, tj. suma svih elektrostatskih sila aminokiselina u proteinu. (3.2.)

$$\Delta G_d = c * \sum_{i=1}^N \sum_{\substack{j=1 \\ i < j}}^N \frac{q_i * q_j}{r_{ij}^2} \quad (3.2)$$

Konstanta c koja se koristi iznosi 332 kcal/mol, naboji su izračunati za svaki atom aminokiseline pojedinačno (jedinica je naboj elektrona  $e \sim 1.6 \times 10^{-19}$  C), a udaljenosti se računaju ovisno o x, y i z koordinatama pojedinih atoma. Udaljenosti koje se koriste izražene su u angstromima ( $1\text{\AA} = 10^{-10}$  m).

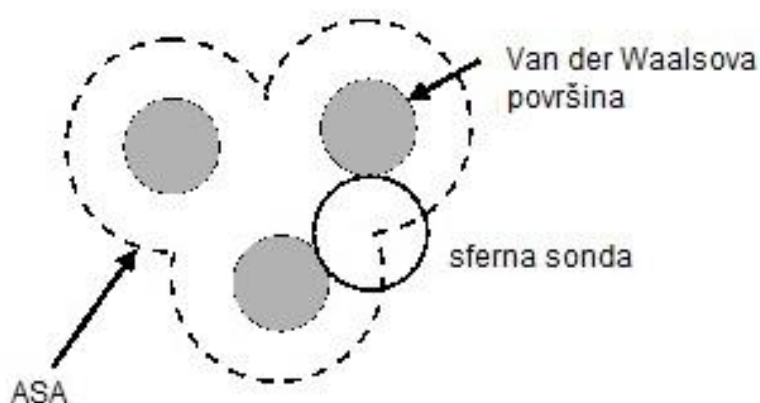
### 3.2 Energija otapala

Za proučavanje interakcija između proteina i drugih molekula, kao i za poznavanje funkcije proteina, važno je poznavati topologiju površine proteina, jer u interakcijama direktno sudjeluju neki dijelovi površine.

Otapalu dostupno područje površine proteina (engl. *Accessible Surface Area*, skraćeno ASA) je područje površine proteina na kojem je moguć kontakt proteina s otapalom u kojem se nalazi, a to je najčešće voda.

Otapalu dostupno područje površine atoma možemo definirati pomoću pozicije sferne sonde koja predstavlja molekulu otapala i kotrlja se po van der Waalsovoj površini proteina (Slika 3.2). Taj način određivanja ASA-e opisali su Lee B. i Richards F.M. 1971. godine [9]. Van der Waalsova površina proteina dobiva se tako da svaki atom zamijenimo sferom van der Waalsovog radijusa, koji

je za atome eksperimentalno određen mjerenjem prostora između para nevezanih atoma unutar kristala.

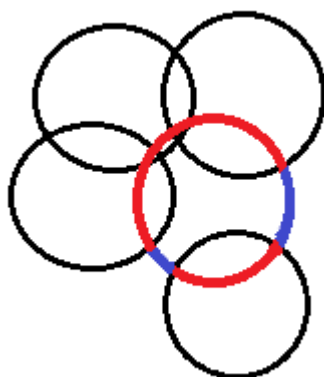


**Slika 3.2.** Otapalu dostupno područje površine molekule – ASA

Izračun otapalu dostupne površine atoma preuzet je prema diplomskom radu Josipa Mihela ([4]) i prebačen u Java programski jezik. 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  $R$ ). Kao radijus sonde uzima se iznos od 1.4 Å, koji ujedno predstavlja i radijus molekule vode. Svaki atom smatra se sferom s centrom određenim koordinatama atoma i radijusa jednakom dobivenom proširenom radijusu pojedinog atoma. Struktura proteina može se promatrati kao skup ispresijecanih sfera.

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 u svim ravninama. Na slici (Slika 3.3.) plavom bojom

predstavljena je dužina slobodnog luka atoma obojanog crvenom bojom u jednoj ravlini.



**Slika 3.3.** Dužina slobodnog luka proteina u jednoj ravlini (označena plavom bojom)

Ukupno dostupno područje površine računa se pomoću formula ((3.3.) ((3.4.) i ((3.5.), pri čemu je  $Z_i$  duljina okomice spuštene iz centra sfere na i-tu ravlinu.

$$ASA = \sum_{i=1}^n \frac{R * D * L_i}{\sqrt{R^2 - Z_i^2}} \quad (3.3.)$$

$$D = \frac{zslice}{2} + \Delta Z \quad (3.4.)$$

$$\Delta Z = \min \left\{ \frac{zslice}{2}, R - Z_i \right\} \quad (3.5.)$$

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 tablici (Tablica 3.2) prikazane su standardne vrijednosti ASA-e. Vrijednosti su dobivene korištenjem programa PSAIA (engl. *Protein Structure and Interaction Analyzer*) ([3]) i izražene su u Å<sup>2</sup>.

**Tablica 3.2** Vrijednosti standardnih vrijednosti otapalu dostupnih područja proteina (ASA) za aminokiseline

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
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

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 ( $\Delta G_S$ ) računamo za aminokiseline i baze, a ne i za metale i ostale atome koji mogu biti dio proteina. Ukupna energija otapala dobiva se množenjem potencijala otapanja pojedine aminokiseline s omjerom površine dostupne otapalu cijele aminokiseline i totalne površine dostupne otapalu. (3.6.)

$$\Delta G_S = \frac{ASA_{ALL}}{ASA_{TOT}} * p \quad (3.6.)$$

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.

### 3.3 Korelacija

Korelacija dviju varijabli govori o njihovoj povezanosti. Grafički se korelacija prikazuje točkastim dijagramom (engl. *scatter diagram*), a matematički se može računati Pearsonov  $r_p$  i Spearmanov koeficijent korelacije  $r_s$ . Korelacija može imati vrijednosti u intervalu  $-1 \leq r \leq 1$ .

Pearsonov koeficijent korelacije je parametrijski test (računa s izmjerenim vrijednostima), a Spearmanov koeficijent korelacije (rang korelacija) je neparametrijski (računa s „rank“ vrijednostima). [11]

Korelacija može biti:

- Pozitivna potpuna  $r = +1$
- Pozitivna nepotpuna  $0 < r < +1$
- Negativna nepotpuna  $-1 < r < 0$

- Negativna potpuna  $r = -1$ .

Ukoliko je korelacija nepostojeća, koeficijent iznosi  $r = 0$ .

Za biološke se varijable koeficijent može tumačiti na sljedeći način:

- $0 \leq r \leq \pm 0.25$       nema povezanosti
- $\pm 0.26 \leq r \leq \pm 0.50$     slaba povezanost
- $\pm 0.51 \leq r \leq \pm 0.75$     umjerena do dobra povezanost
- $\pm 0.76 \leq r \leq \pm 1$       vrlo dobra do izvrsna povezanost
- $r = \pm 1$               matematička povezanost.

U formuli za izračun Pearson-ovog koeficijenta korelacije (3.7) i (3.8)  $x_i$  i  $y_i$  predstavljaju vrijednosti varijabli za koje se koeficijent korelacije računa,  $\bar{x}$  je aritmetička sredina vrijednosti  $x_i$ , a  $n$  je broj koliko se varijabli uspoređuje.

$$r_p := \frac{\sum (x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 * \sum (y_i - \bar{y})^2}} \quad (3.7.)$$

$$r_p := \frac{n * \sum x_i * y_i - \sum x_i \sum y_i}{\sqrt{n * \sum x_i^2 - (\sum x_i)^2} * \sqrt{n * \sum y_i^2 - (\sum y_i)^2}} \quad (3.8.)$$



## 4. Formati podataka

### 4.1 PDB datoteke

Podatke za analizu proteinskih interakcija saznajemo iz PDB (engl. *Protein Data Bank*) 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 nuklearne magnetske rezonance (skraćeno, NMR spektroskopije) 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, kao što su ATOM, HETATM, MODEL, TITLE i 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. 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, Å)
- Faktor zauzeća (engl. *occupancy*) i temperaturni faktor
- Naziv elementa

Primjer jedne PDB datoteke dan je na slici (Slika 4.1.).

```
HEADER.....COMPLEX (HORMONE/RECEPTOR).....15-JAN-98...1A22...
TITLE.....HUMAN GROWTH HORMONE BOUND TO SINGLE RECEPTOR.....
COMPND...MOL_ID: 1;.....
COMPND...2 MOLECULE: GROWTH HORMONE;.....
COMPND...3 CHAIN: A;.....
COMPND...4 ENGINEERED: YES;.....
COMPND...5 MUTATION: YES;.....
COMPND...6 MOL_ID: 2;.....
COMPND...7 MOLECULE: GROWTH HORMONE RECEPTOR;.....
COMPND...8 CHAIN: B;.....
COMPND...9 FRAGMENT: EXTRACELLULAR DOMAIN;.....
COMPND...10 ENGINEERED: YES.....
SOURCE...MOL_ID: 1;.....
SOURCE...2 ORGANISM_SCIENTIFIC: HOMO SAPIENS;.....
SOURCE...3 ORGANISM_COMMON: HUMAN;.....
SOURCE...4 ORGANISM_TAXID: 9606;.....
SOURCE...5 EXPRESSION_SYSTEM: ESCHERICHIA COLI;.....
SOURCE...6 EXPRESSION_SYSTEM_TAXID: 562;.....
SOURCE...7 MOL_ID: 2;.....
SOURCE...8 ORGANISM_SCIENTIFIC: HOMO SAPIENS;.....
SOURCE...9 ORGANISM_COMMON: HUMAN;.....
SOURCE...10 ORGANISM_TAXID: 9606;.....
SOURCE...11 EXPRESSION_SYSTEM: ESCHERICHIA COLI;.....
SOURCE...12 EXPRESSION_SYSTEM_TAXID: 562.....
KEYWDS...COMPLEX (HORMONE/RECEPTOR), PITUITARY, HORMONE.....
EXPDTA...X-RAY DIFFRACTION.....
```

Slika 4.1. Primjer početka PDB datoteke *1a22.pdb*

Pri računanju elektrostatske sile između aminokiselina koriste se nazivi atoma, oznaka lanca proteina te x, y i z koordinate atoma.

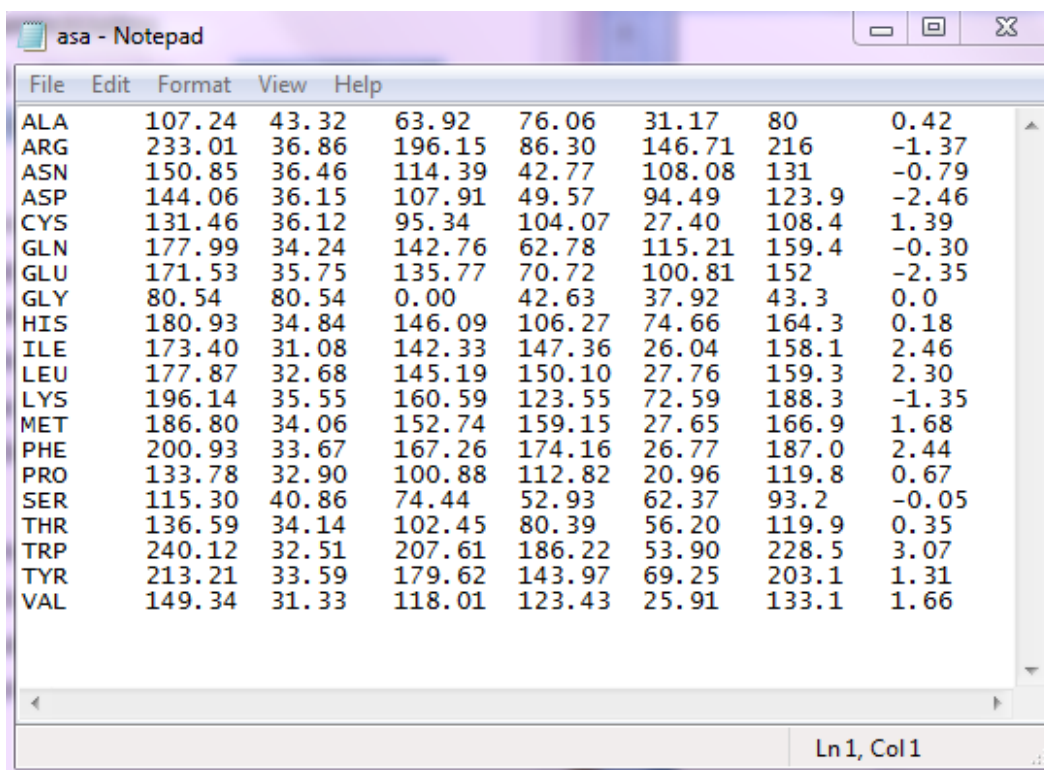
## 4.2 Naboji atoma i otapalu dostupna područja površine

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 4.2.).

```
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 4.2. Dio datoteke s nabojima atoma *amberCharge.txt*

Podaci potrebni za izračun energije otapala preuzimaju se iz datoteke *asa.txt* koja sadrži oznaku aminokiseline te zatim redom, vrijednost ukupne ASA-e svih atoma aminokiseline, ASA-e glavnog lanca aminokiseline, ASA-e bočnog lanca aminokiseline, ASA-e nepolarnog dijela, ASA-e polarnog dijela, totalne ASA-e i potencijal otapala. Datoteka je prikazana na slici (Slika 4.3.).



	ALA	ARG	ASN	ASP	CYS	GLN	GLU	GLY	HIS	ILE	LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR	VAL
1	107.24	233.01	150.85	144.06	131.46	177.99	171.53	80.54	180.93	173.40	177.87	196.14	186.80	200.93	133.78	115.30	136.59	240.12	213.21	149.34
2	43.32	36.86	36.46	36.15	36.12	34.24	35.75	80.54	34.84	31.08	32.68	35.55	34.06	33.67	32.90	40.86	34.14	32.51	33.59	31.33
3	63.92	196.15	114.39	107.91	95.34	142.76	135.77	0.00	146.09	142.33	145.19	160.59	152.74	167.26	100.88	74.44	102.45	207.61	179.62	118.01
4	76.06	86.30	42.77	49.57	104.07	62.78	70.72	42.63	106.27	147.36	150.10	123.55	159.15	174.16	112.82	52.93	80.39	186.22	143.97	123.43
5	31.17	146.71	108.08	94.49	27.40	115.21	100.81	37.92	74.66	26.04	27.76	72.59	27.65	26.77	20.96	62.37	56.20	53.90	69.25	25.91
6	80	216	131	123.9	108.4	159.4	152	43.3	164.3	158.1	159.3	188.3	166.9	187.0	119.8	93.2	119.9	228.5	203.1	133.1
7	0.42	-1.37	-0.79	-2.46	1.39	-0.30	-2.35	0.0	0.18	2.46	2.30	-1.35	1.68	2.44	0.67	-0.05	0.35	3.07	1.31	1.66

Slika 4.3. Datoteka s ASA vrijednostima i potencijalom otapala *asa.txt*

### 4.3 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 (Slika 4.4).

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
....		

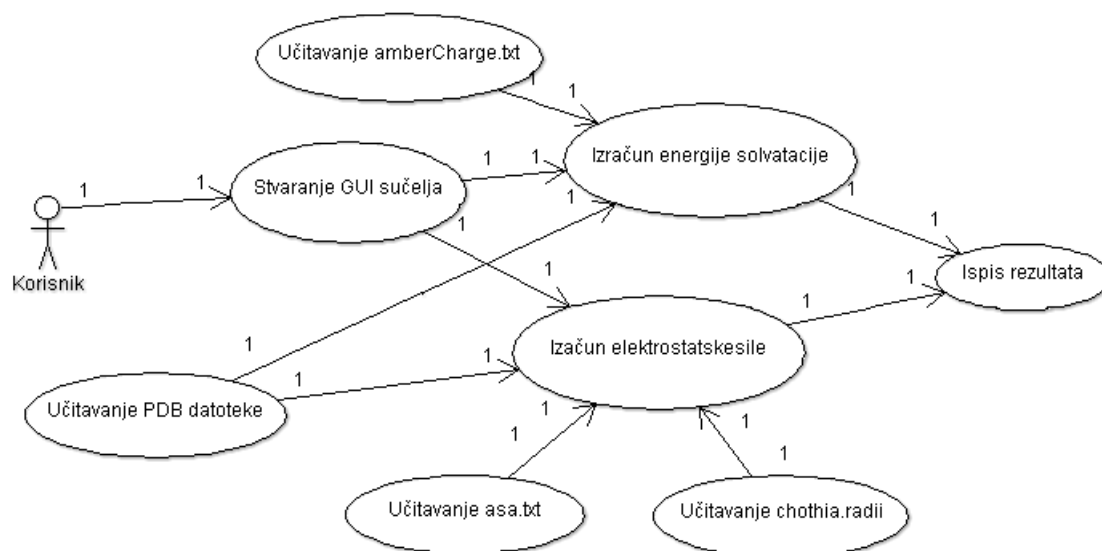
**Slika 4.4.** Dio datoteke s vrijednostima Van der Waalsovih radijusa atoma *chothia.radii*

### 4.4 Koeficijent privlačnosti

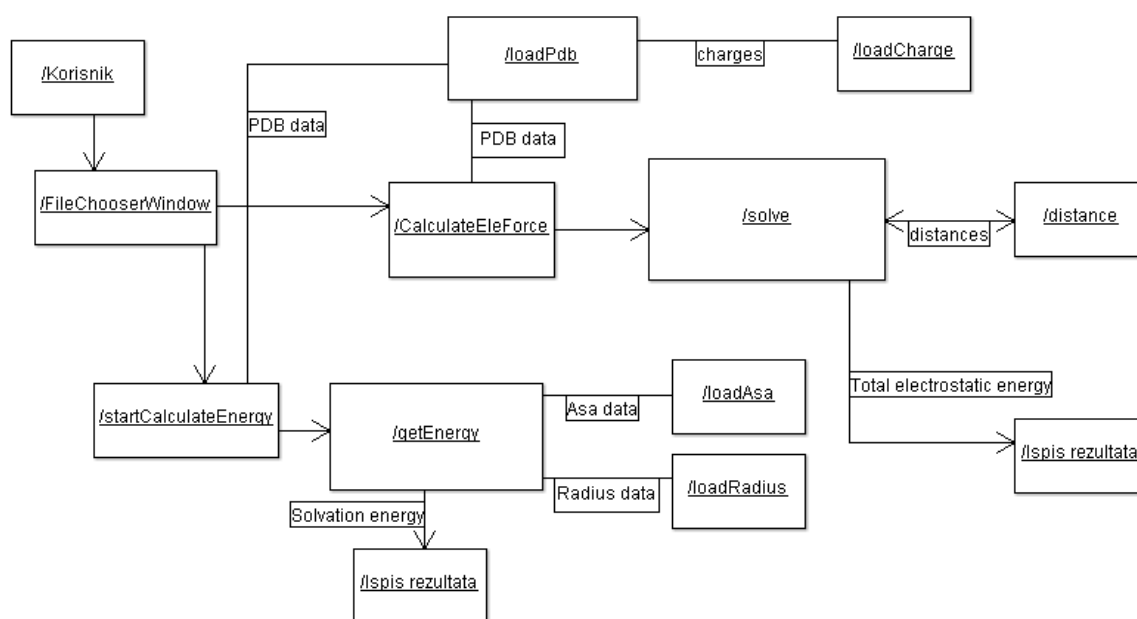
Za potrebe računanja koeficijenta korelacije između vrijednosti dobivenih napisanim algoritmom i koeficijenta privlačnosti proteina, koristi se tablica s eksperimentalno utvrđenim vrijednostima koeficijenta privlačnosti (engl. *affinity data*). Podaci su za potrebe izračuna zapisani su tablično te je za izračun korištena funkcija CORREL programa Microsoft Excel.

## 5. Implementacija

Diplomski rad je izrađen u programskom jeziku Java. Dijagram obrazaca (engl. *use case*) algoritma prikazan je na Slika 5.1., a dijagram toka podataka (engl. *data flow*) na Slika 5.2.



Slika 5.1. Dijagram obrazaca algoritma

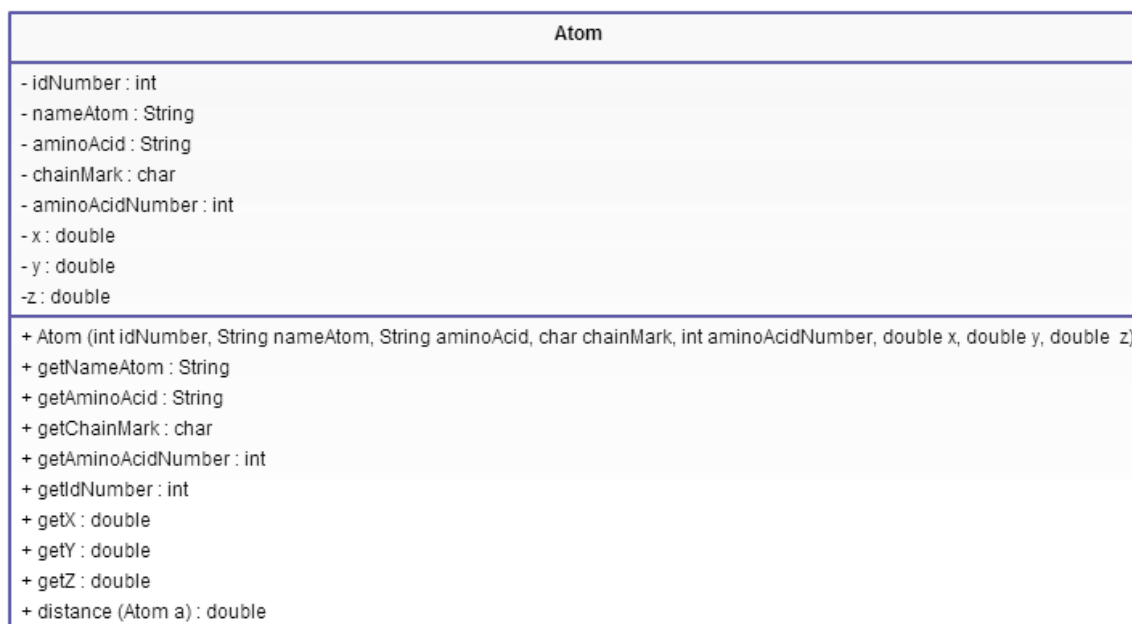


Slika 5.2. Dijagram toka podataka algoritma

## 5.1 Opis algoritma

Algoritam je podijeljen na dva dijela, jedan za računanje elektrostatske sile između aminokiselina i drugi koji računa energiju otapala. U glavnoj klasi MainProgram stvara se instanca klase FileChooserWindow koja oblikuje grafičko sučelje aplikacije. Ovisno o označenom gumbu za računanje energije solvatacije i/ili elektrostatske sile proteina, pokreću se metode klase Proteins *startCalculateEnergy* i/ili *calculateEleForce*. Prilikom pokretanja metode *calculateEleForce* stvaraju se instance klasa Pdb i Charges, a prilikom pokretanja metode *startCalculateEnergy* instance klasa Radius, Asa kao i Charges te Pdb.

Za svaki redak PDB datoteke koja se učitava, stvara se jedna instanca klase Atom (Slika 5.3.), i spremaju se podaci potrebni za računanje elektrostatske sile. Učitavanje iz PDB datoteke opisano je u metodi *loadPdb* klase Pdb. 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 jednog atoma od drugog.



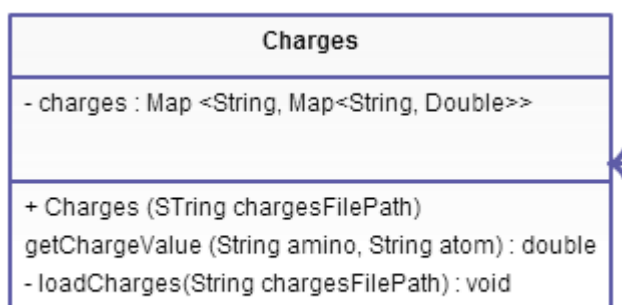
**Slika 5.3.** Dijagram klase Atom s pripadajućim atributima i metodama

Klasa koja sadrži vektor svih atoma jednog lanca naziva se *Chain*. 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 slici (Slika 5.4.).

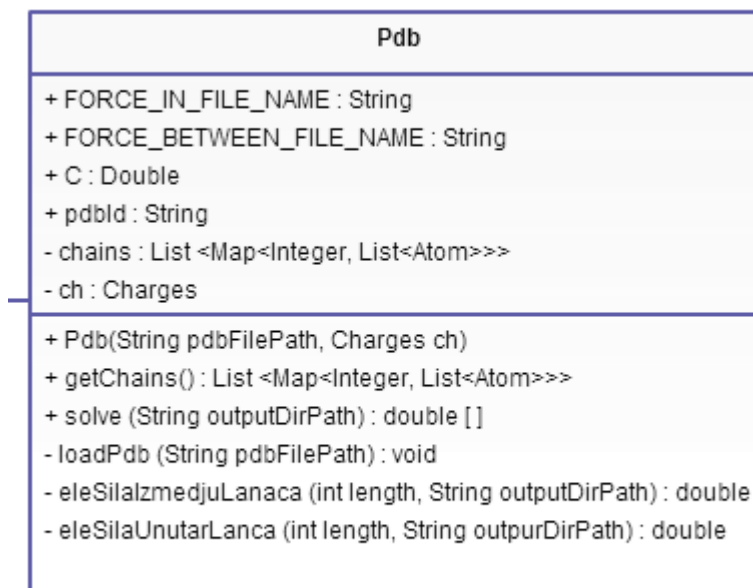


**Slika 5.4.** Dijagram klase *Chain* s pripadajućim atributima i metodama

Klasa *Pdb* sadrži svaki lanac aminokiselina koji se nalazi u učitanoj PDB datoteci. Uz atribut *chains* koji je lista instanci klase *Atom* sadrži i referencu na instancu klase *Charges* u kojoj su spremljeni učitani naboji atoma. Prilikom pokretanja izračuna elektrostatske sile se stvara instanca klase *Charges*, koja učitava naboje svih atoma iz datoteke *amberCharge.txt* te se zatim pokreće se učitavanje PDB datoteke metodom *loadPdb*. Nakon već navedenog učitavanja u instance klase *Atom*, pokreće se *solve* funkcija koja računa elektrostatske sile unutar jednog lanca proteina i između njih više (ovisno ima li ih protein) te daje ukupan zbroj tih elektrostatskih sila. Dijagrami klase *Pdb* i *Charges* prikazani su na slikama (Slika 5.5. i Slika 5.6.).



**Slika 5.5.** Dijagram klase *Charges* s pripadajućim atributima i metodama



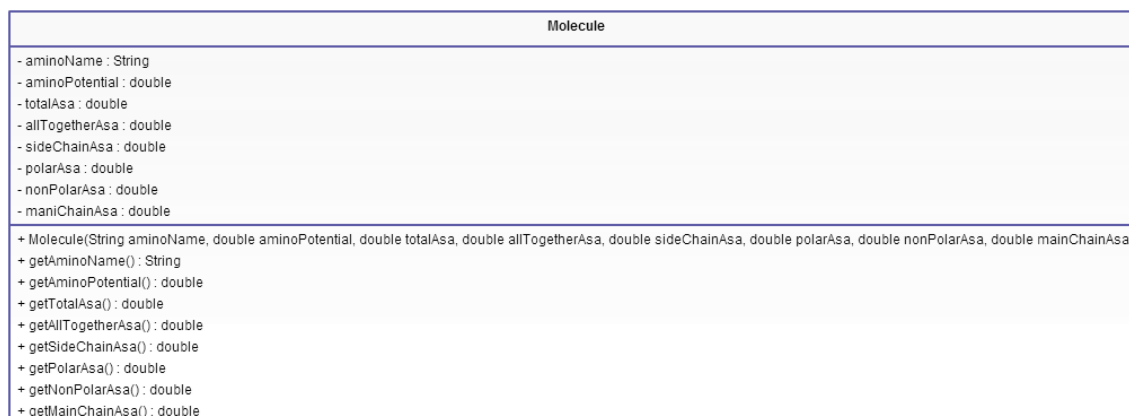
**Slika 5.6.** Dijagram klasa Pdb s pripadajućim atributima i metodama

Naboje koji su potrebni za algoritam učitavamo iz datoteke s nabojima 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 prilikom računanja u metodama *eleSilalzmedjuLanaca* i *elesilaUnutarLanca* 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  $c$  ( $c = 332 \text{ kcal/mol}$ ). Elektrostatska sila između odgovarajućih aminokiselina dobiva se zbrajanjem elektrostatskih sila između svaka dva atoma aminokiseline. Na dijagramu klase Charges (Slika 5.6.) vidi se konstruktor klase, kao i metoda *loadCharges* u kojoj se odvija učitavanje naboja.

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, podatak o površini dostupnoj otapalu bočnog lanca aminokiseline, podatak o površini polarnog dijela lanca, nepolarnog dijela lanca i površini dostupnoj otapalu glavnog lanca aminokiseline. Kasnije se u algoritmu ti podaci dohvaćaju pomoću metoda za dohvaćanje *getAminoName*, *getAminoPotential*, *getTotalAsa*,

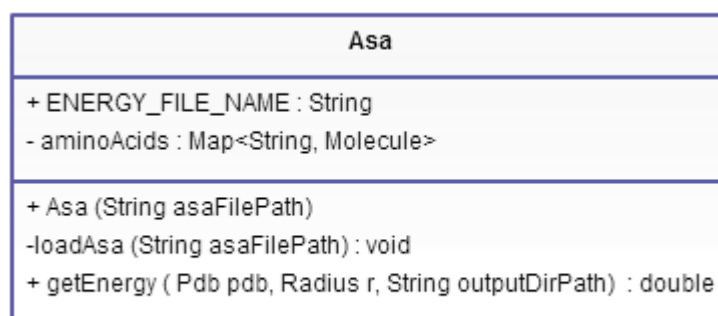


*getAllTogetherAsa*, *getSideChainAsa*, *getPolarAsa*, *getNonPolarAsa* i *getMainChainAsa*. Dijagram klase Molecule prikazan je na slici (Slika 5.7.).



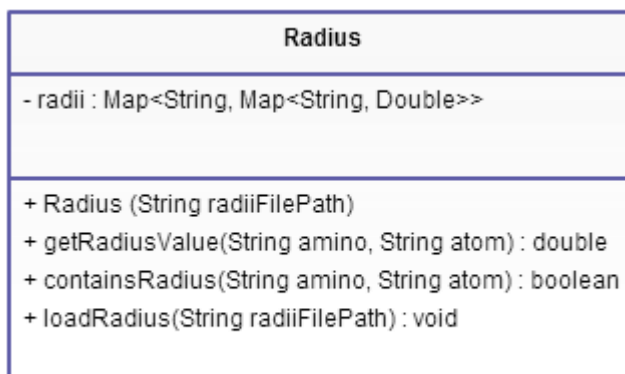
**Slika 5.7.** Dijagram klase Molecule s pripadajućim atributima i metodama

Podaci o otapalu dostupnoj površini se učitavaju iz datoteke *asa.txt* u mapu aminokiselina, prilikom čega se za svaku aminokiselinu stvara instanca klase Molecule. Klasa koja to obavlja je Asa, a metoda *loadAsa*. Metoda u kojoj se računanje energije otapanja za svaku aminokiselinu odvija je *getEnergy*. Dijagram klase Asa prikazan je na slici (Slika 5.8.)



**Slika 5.8.** Dijagram klase Asa s pripadajućim atributima i metodama

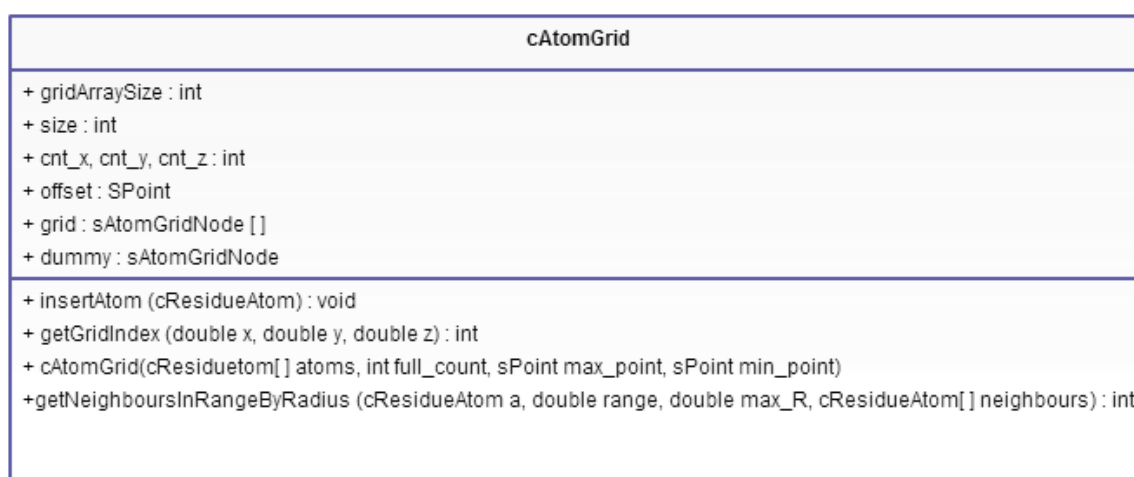
Radijusi pojedinih atoma u određenoj aminokiselini učitavaju se u metodi *loadRadius* klase Radius (Slika 5.9), te se spremaju u mapu koja za svaku aminokiselinu sprema par vrijednosti atom i radijus, a dohvaćaju se metodom *getRadiusValue*.



**Slika 5.9.** Dijagram klase Radius s pripadajućim atributima i metodama

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

Stvaranje trodimenzionalne mreže atoma implementirano je u klasi cAtomGrid. Klasa sadrži attribute poput *offset* (informacija o potrebnom pomicanju atoma unutar mreže za olakšanje izračuna, *size* (veličina stranice polja mreže) i *cnt\_x*, *cnt\_y*, *cnt\_z* (broj polja koja čine mrežu po x, y i z osima). Pretraživanje prostora oko atoma odvija se u metodi *getNeighboursInRangeByCenter*.



**Slika 5.10.** Dijagram klase cAtomGrid s pripadajućim atributima i metodama

## 6. Instalacija i pokretanje

Implementirani algoritam može se pokrenuti na bilo kojem računalu i nema posebnih zahtjeva, osim instaliranog Java okruženja (*Java Development Kit* ili *Java Runtime Environment*). Program se može pokretati na Windows i Unix platformi.

Direktorij aplikacije sadrži četiri poddirektorija: *.settings*, *bin*, *data*, i *src*. U direktoriju *data* nalaze se sve potrebne datoteke koje se zadaju kao argumenti algoritma. Direktoriji *bin* sadrži *.class* datoteke svih klasa. U direktoriju *src* nalaze se paketi *proteinInteractions* i *surface* koji sadrže *.java* datoteke svih klasa koje se koriste u algoritmu, a koje sadrže pisani kod algoritma. U mapi *src* nalazi se glavna klasa *MainProgram*, u paketu *proteinInteractions* nalaze se klase potrebne za računanje elektrostatske sile između aminokiselina, a u paketu *surface* klase potrebne za računanje energije otapala.

Algoritam se može pokretati u Java okruženju ili kao *.jar* datoteka (imenovana *proteinInteractions.jar*).

Za pokretanje algoritma iz naredbenog retka potrebno je pozicionirati se u direktoriju u kojem je *.jar* datoteka. Program se pokreće naredbom:

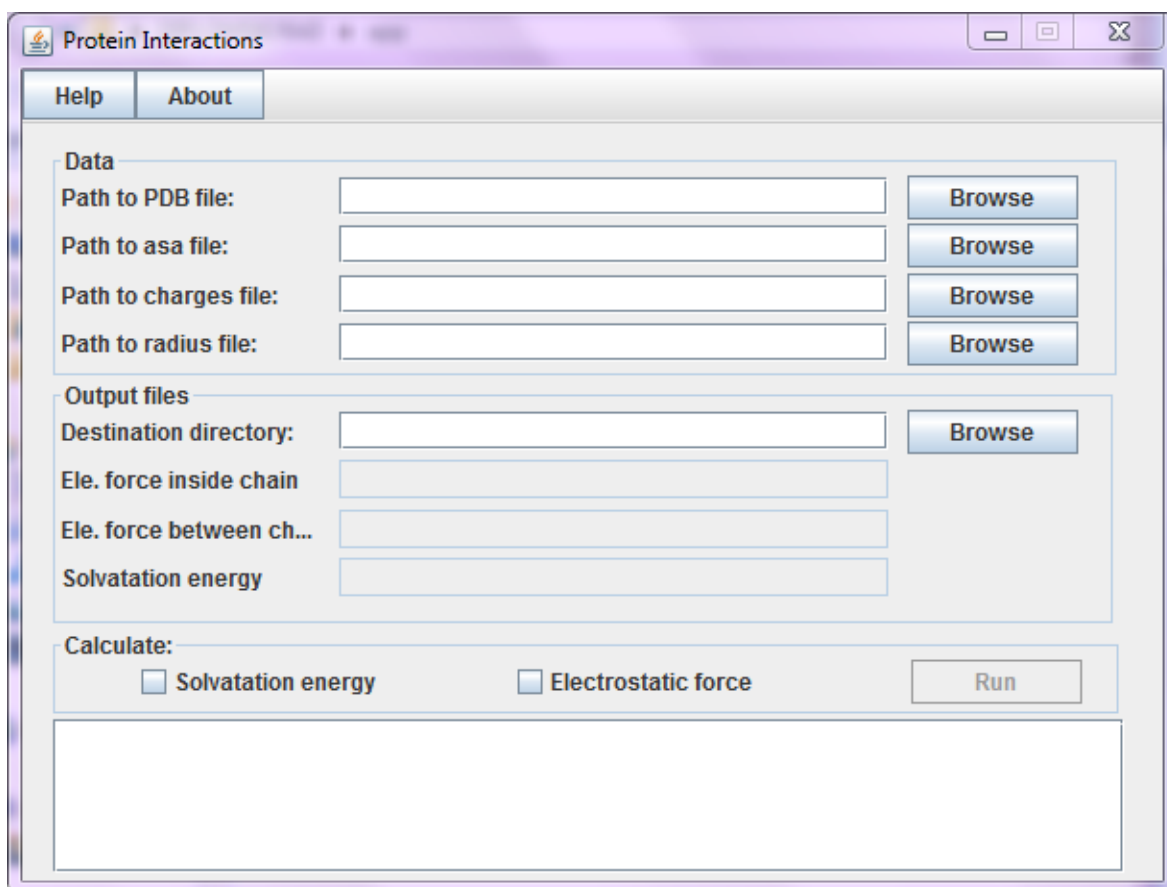
java -jar proteinInteractions.jar
-----------------------------------

Ukoliko se program pokreće iz neke od razvojnih okolina JAVA programskog jezika kao što je program Eclipse, pokreće se glavna klasa *MainProgram.java* iz paketa osnovnog paketa *src*.

Nakon što se pokrene program, otvara se prozor, tj. grafičko sučelje aplikacije nazvano *ProteinInteractions*. U sljedećem poglavlju dane su upute kako ga koristiti.

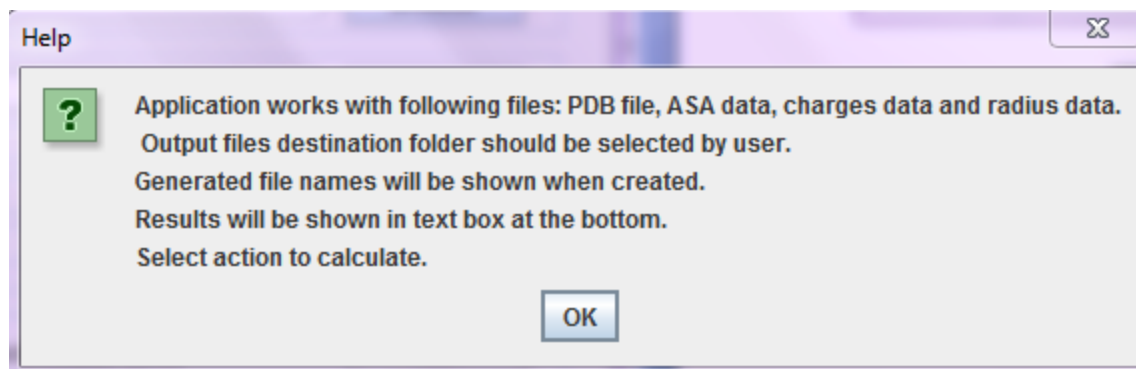
## 7. Grafičko sučelje i upute za korištenje

Nakon pokretanja programa otvara se sučelje aplikacije. Kao korisničko sučelje aplikacije koristi se grafičko, čiji je početni izgled pokazan na slici (Slika 7.1.). Računanje elektrostatske sile i energije solvatacije pokreću se učitavanjem svih datoteka s podacima potrebnima za računanje.

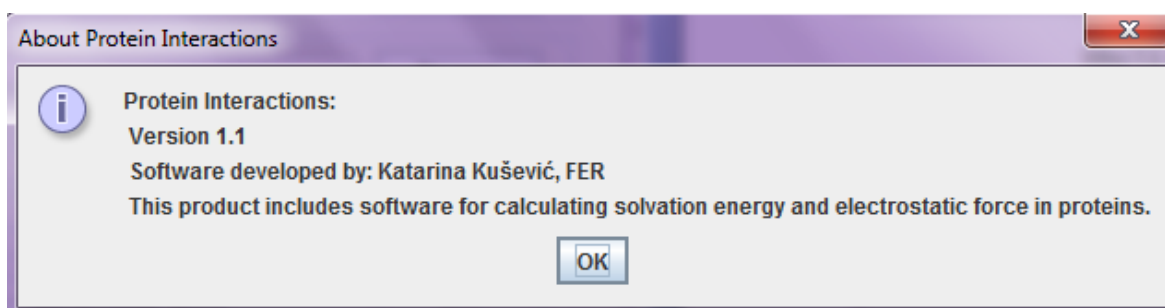


**Slika 7.1.** Grafičko sučelje aplikacije

Pritiskom na gumb „Help“ otvara se dijaloški okvir s kratkim uputama za korištenje programa (Slika 7.2), a pritiskom na gumb „About“ otvara se novi dijaloški okvir s osnovnim informacijama o verziji, autoru i podatak o kakvoj vrsti aplikacije se radi (Slika 7.3).



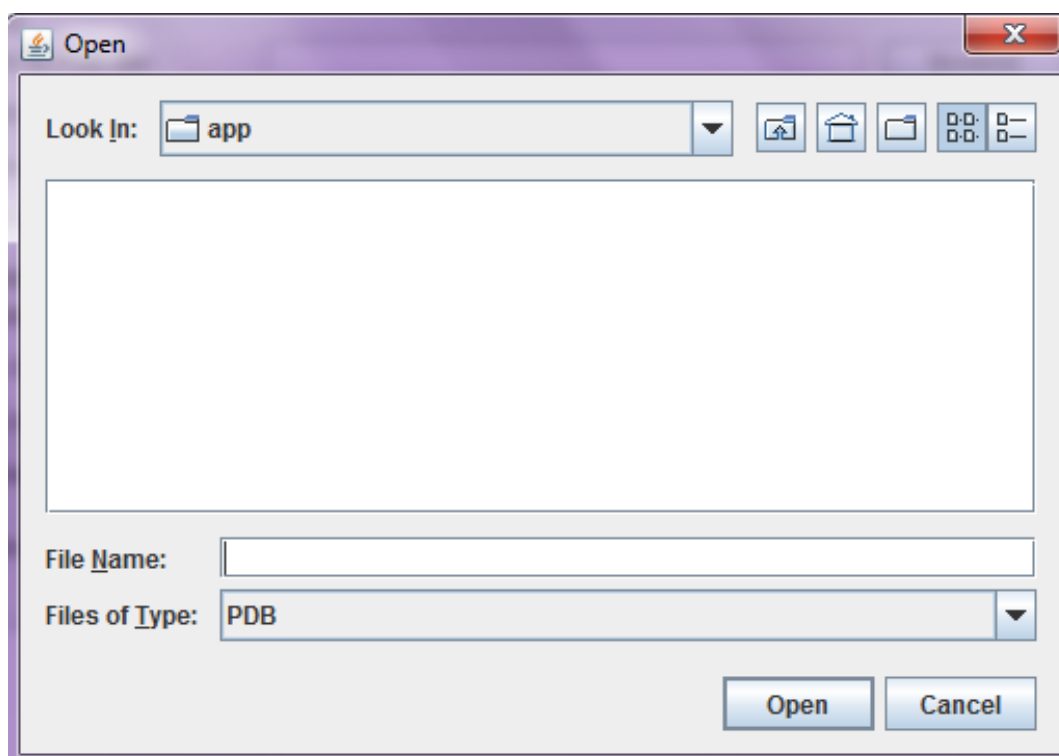
**Slika 7.2.** Dijaloški okvir koji se pokreće pritiskom na gumb „Help“



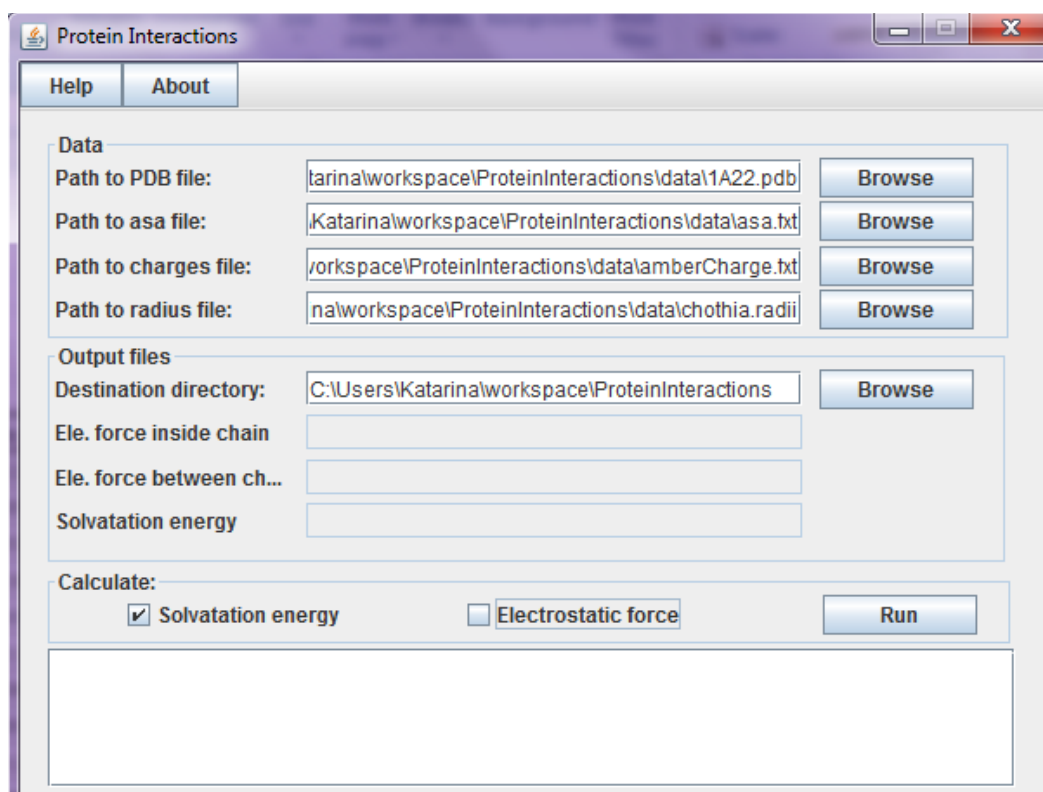
**Slika 7.3** Dijaloški okvir koji se pokreće pritiskom na gumb „About“

Grafičko sučelje podijeljeno je na tri dijela; „Data“, „Output files“ i „Calculate“. U „Data“ dijelu upisuju se putovi do datoteka potrebnih za izračun, a to su: put do PDB datoteke koja sadrži zapis o proteinu, put do datoteke s podacima o otapalu dostupnom području aminokiselina, put do datoteke s nabojima atoma u pojedinoj aminokiselini i put do datoteke s radijusima atoma. Kraj svakog polja za upis postoji gumb „Browse“. Klikom na njega, otvara se dijaloški okvir za učitavanje datoteke. Ovisno o polju za pojedinu datoteku, ograničeno je učitavanje samo ispravne ekstenzije datototeke. Primjer dijaloškog okvira dan je na slici (Slika 7.4.)

U dijelu „Output files“ korisnik mora upisati put do odredišnog direktorija za datoteke koje program stvara. Korisnik taj put može sam upisati ili otvoriti traženu mapu klikom na gumb „Browse“ pokraj tekstualnog polja. Preostala tri polja ovog dijela korisnik ne može mijenjati, na tim se mjestima pokazuje ime stvorene datoteke. Izgled grafičkog sučelja prije pokretanja programa prikazano je na slici (Slika 7.5.).

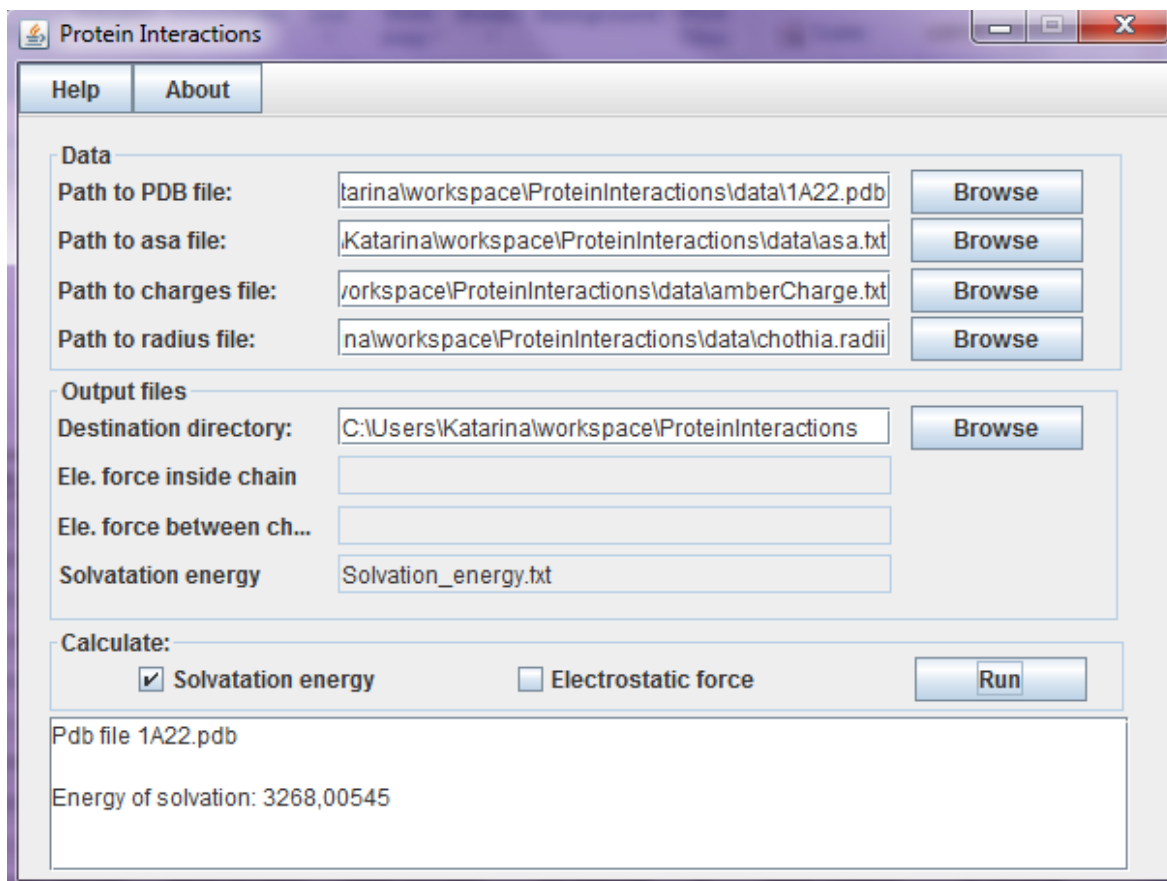


Slika 7.4. Dijaloški okvir za odabir datoteke koju se učitava



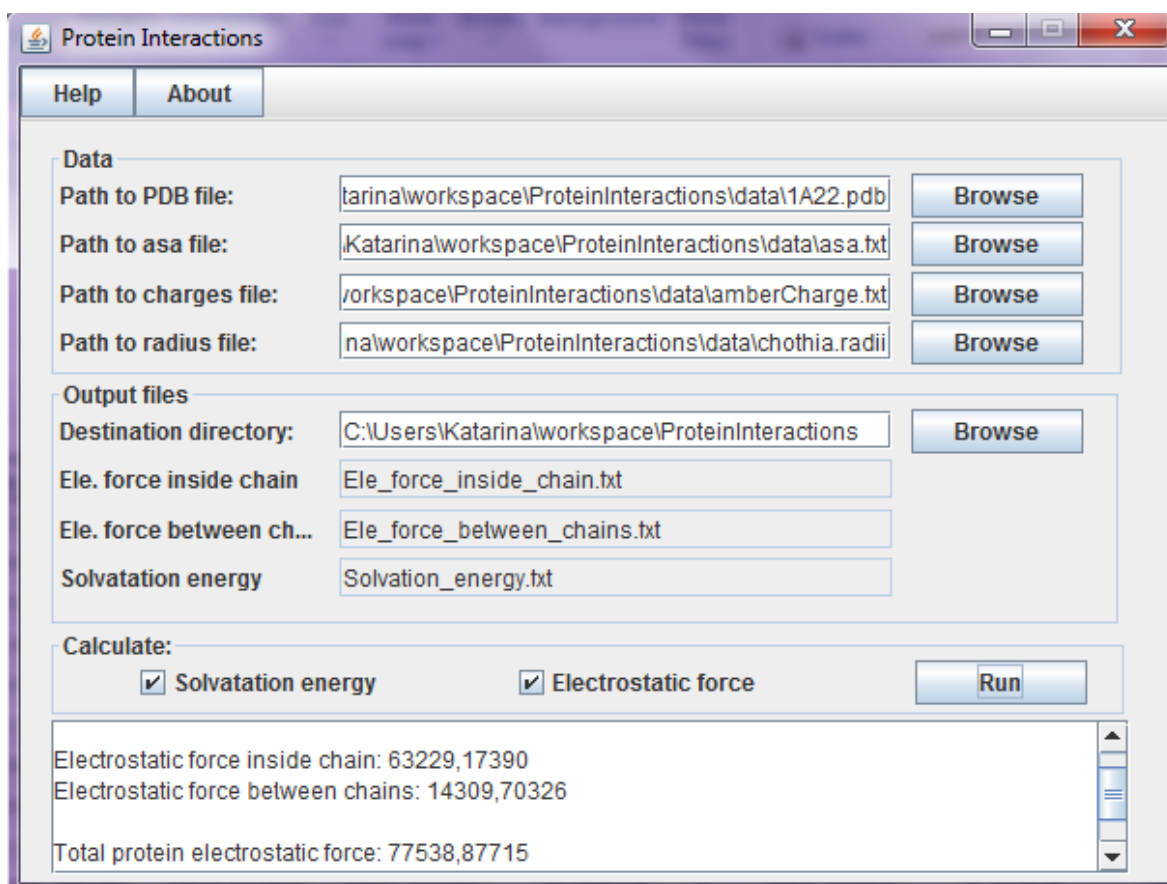
Slika 7.5. Izgled grafičkog sučelja prije pokretanja izračuna

Korisnik u dijelu „Calculate“ određuje što želi izračunati, energiju solvatacije proteina i/ili elektrostatsku silu proteina te pokreće program klikom na gumb „Run“. Ukoliko korisnik ne označi niti jedan izračun kvačicom, nije moguće kliknuti na gumb „Run“ i pokrenuti program. Primjer ispisa rezultata kada se računa samo energija solvatacije dan je na slici (Slika 7.6.)



**Slika 7.6.** Grafičko sučelje nakon izračuna energije solvatacije

Primjer ispisa rezultata kada se računa energija solvatacije i elektrostatska sila proteina dan je na slici (Slika 7.7.).



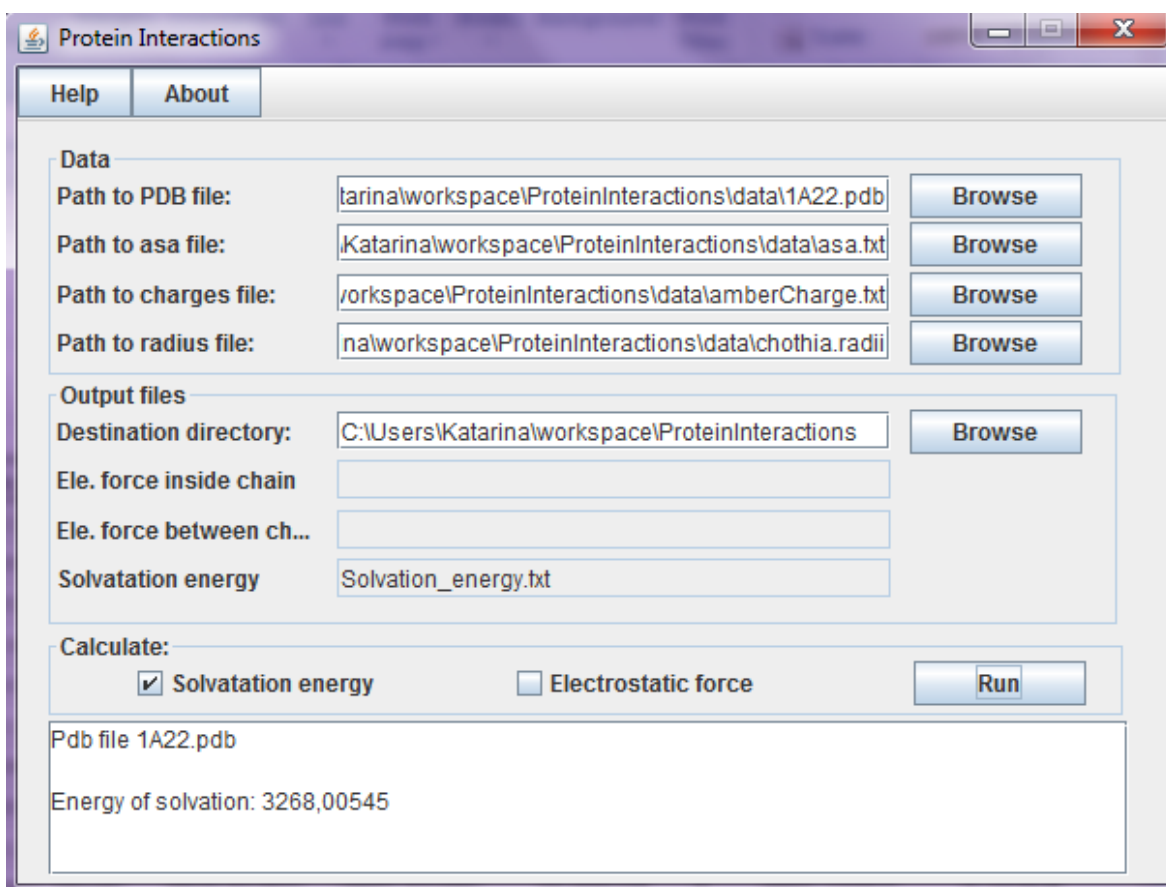
**Slika 7.7.** Grafičko sučelje nakon računanja elektrostatske sile i energije solvatacije

Program se prekida zatvaranjem prozora. Moguće je nakon izračuna vrijednosti za jednu PDB datoteku, promjenom datoteke izračunati vrijednosti za bilo koju drugu PDB datoteku bez zatvaranja i ponovnog pokretanja programa. Samo se klikom na gumb „Browse“ da programu put do nove PDB datoteke.



## 8. Rezultati

Prikazani će biti rezultati za testni primjer PDB datoteke preuzete s interneta, imenom *1A22.pdb*. Navedena PDB datoteka sadrži 2 lanca, od kojih jedan ima 1466 atoma raspodijeljenih u 191 aminokiselinu, a drugi 1578 atoma raspodijeljenih u 205 aminokiselina. Algoritam računa elektrostatske sile između svake dvije aminokiseline različitih lanaca, unutar jednog lanca i energiju otapala svake aminokiseline iz PDB datoteke (Slika 8.1).

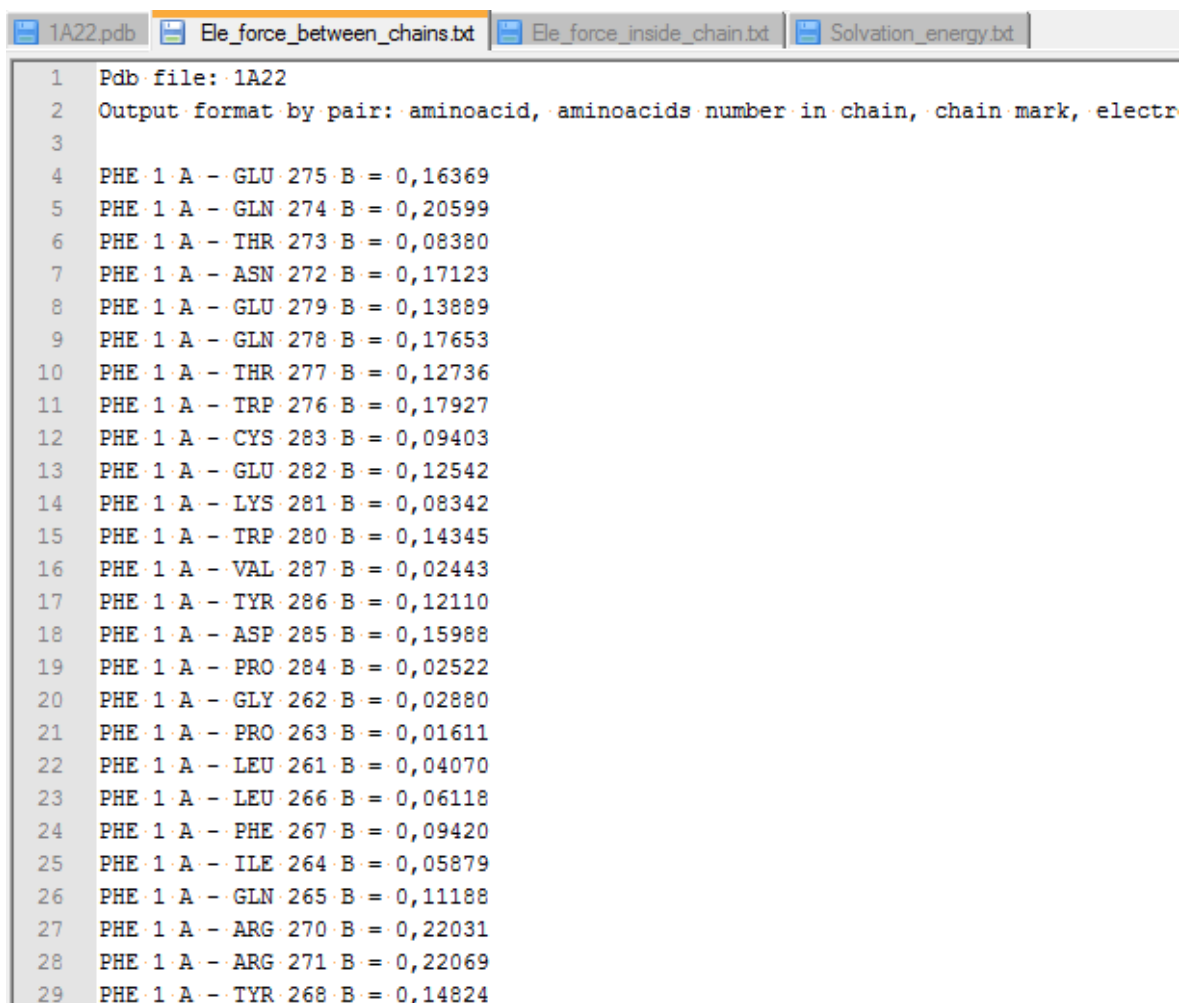


**Slika 8.1** Izgled grafičkog sučelja nakon izvršenog programa za testni primjer 1A22.pdb

Algoritam rezultate zapisuje u tri .txt datoteke, *Ele\_force\_between\_chains.txt*, *Ele\_force\_inside\_chain.txt* i *Solvation\_energy*. Računa se elektrostatska sila između svaka dva para aminokiselina u interakciji unutar lanca i između svaka dva lanca te energija otapala pojedine aminokiseline iz učitane PDB datoteke. Ispis rezultata sadrži zbroj elektrostatske sile unutar lanca za sve parove iz te datoteke, zbroj elektrostatske sile između lanaca za sve parove te datoteke, ukupni zbroj te

dvije vrijednosti (dobiva se vrijednost za protein) i zbroj energija solvatacije proteina.

Primjer ispisa rezultata algoritma za računanje elektrostatske sile između dva lanca, prikazan je na slici (Slika 8.2.). Rezultati su zapisani za svaki par aminokiselina njihove oznake, redni broj u lancu i oznaka lanca. Pri tome algoritam se izvodi za par lanaca, sve parovi aminokiselina.



```
1 Pdb file: 1A22
2 Output format by pair: aminoacid, aminoacids number in chain, chain mark, electr
3
4 PHE 1 A - GLU 275 B = 0,16369
5 PHE 1 A - GLN 274 B = 0,20599
6 PHE 1 A - THR 273 B = 0,08380
7 PHE 1 A - ASN 272 B = 0,17123
8 PHE 1 A - GLU 279 B = 0,13889
9 PHE 1 A - GLN 278 B = 0,17653
10 PHE 1 A - THR 277 B = 0,12736
11 PHE 1 A - TRP 276 B = 0,17927
12 PHE 1 A - CYS 283 B = 0,09403
13 PHE 1 A - GLU 282 B = 0,12542
14 PHE 1 A - LYS 281 B = 0,08342
15 PHE 1 A - TRP 280 B = 0,14345
16 PHE 1 A - VAL 287 B = 0,02443
17 PHE 1 A - TYR 286 B = 0,12110
18 PHE 1 A - ASP 285 B = 0,15988
19 PHE 1 A - PRO 284 B = 0,02522
20 PHE 1 A - GLY 262 B = 0,02880
21 PHE 1 A - PRO 263 B = 0,01611
22 PHE 1 A - LEU 261 B = 0,04070
23 PHE 1 A - LEU 266 B = 0,06118
24 PHE 1 A - PHE 267 B = 0,09420
25 PHE 1 A - ILE 264 B = 0,05879
26 PHE 1 A - GLN 265 B = 0,11188
27 PHE 1 A - ARG 270 B = 0,22031
28 PHE 1 A - ARG 271 B = 0,22069
29 PHE 1 A - TYR 268 B = 0,14824
```

**Slika 8.2** Prikaz datoteke s rezultatima računanja elektrostatske sile između lanaca  
*Ele\_force\_between\_chains.txt*

Primjer ispisa rezultata algoritma za računanje elektrostatske sile unutar lanca, prikazan je na slici (Slika 8.3.). Rezultati su zapisani za svaki par aminokiselina njihove oznake, redni broj u lancu i oznaka lanca. Pri tome

algoritam se izvodi za sve parove aminokiselina, za svaki lanac, svi parovi aminokiselina.

```
1 Pdb file: 1A22
2 Output format by pair: aminoacid, aminoacids number in chain, chain mark, elect:
3
4 PHE 1 A - PRO 2 A = 0,04551
5 PHE 1 A - THR 3 A = 7,03950
6 PHE 1 A - ILE 4 A = 3,08722
7 PHE 1 A - PRO 5 A = 0,46591
8 PHE 1 A - LEU 6 A = 0,97983
9 PHE 1 A - SER 7 A = 1,07821
10 PHE 1 A - ARG 8 A = 1,07031
11 PHE 1 A - LEU 9 A = 0,60812
12 PHE 1 A - PHE 10 A = 0,86907
13 PHE 1 A - ASP 11 A = 0,86757
14 PHE 1 A - ASN 12 A = 0,67010
15 PHE 1 A - ALA 13 A = 0,36304
16 PHE 1 A - MET 14 A = 0,56138
17 PHE 1 A - LEU 15 A = 0,25088
18 PHE 1 A - ALA 17 A = 0,23723
19 PHE 1 A - ARG 16 A = 0,60806
20 PHE 1 A - ARG 19 A = 0,32033
21 PHE 1 A - HIS 18 A = 0,35816
22 PHE 1 A - HIS 21 A = 0,29736
23 PHE 1 A - LEU 20 A = 0,17373
24 PHE 1 A - LEU 23 A = 0,13459
25 PHE 1 A - GLN 22 A = 0,30454
```

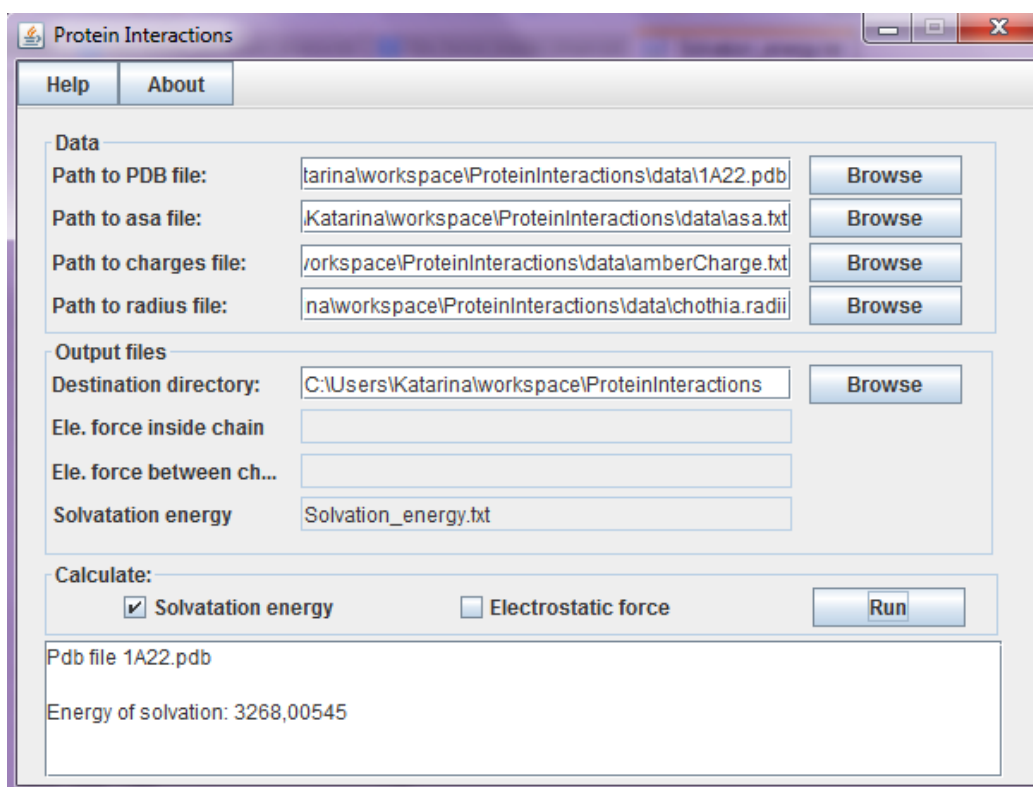
**Slika 8.3.** Prikaz datoteke s rezultatima računanja elektrostatske sile unutar lanca  
*Ele\_force\_inside\_chain.txt*

Na slici (Slika 8.4.) prikazan je primjer ispisa računanja energije otapala za aminokiseline iz PDB datoteke *1A22.pdb*. Rezultati su zapisani za svaku aminokiselinu njezina oznaka i redni broj u lancu.

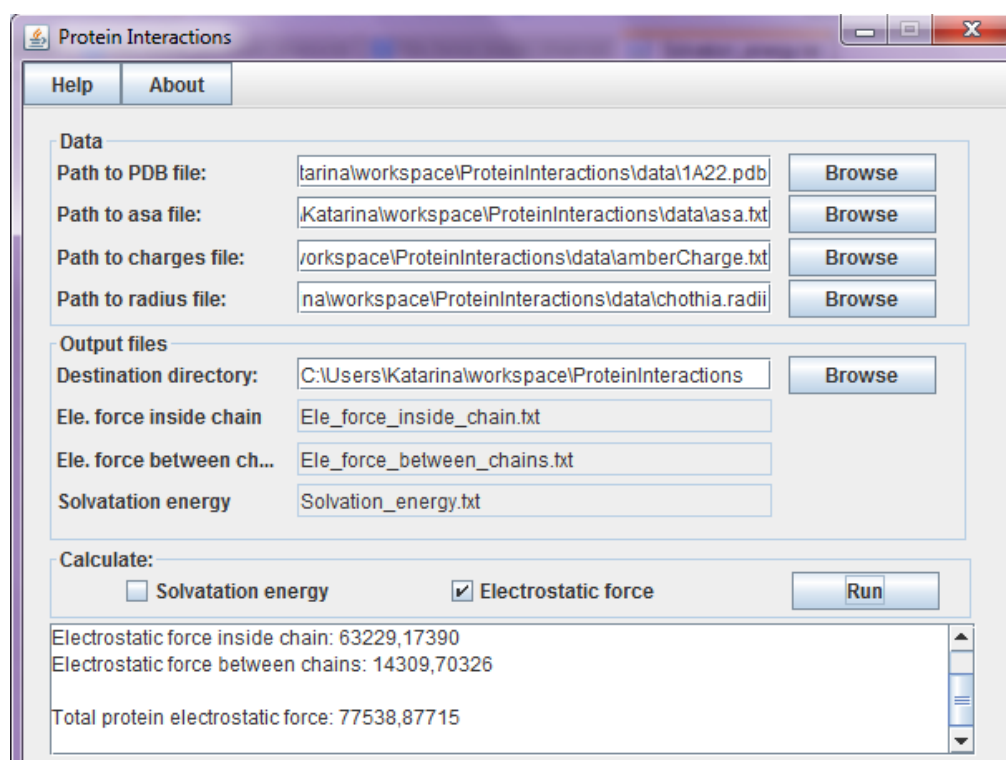
```
1 Pdb file: 1A22
2 Output format: aminoacid, aminoacids number in chain, solvation energy.
3
4 PHE 1 2,959642
5 PRO 2 1,861601
6 THR 3 1,333089
7 ILE 4 9,150411
8 PRO 5 3,867760
9 LEU 6 10,323037
10 SER 7 -0,407598
11 ARG 8 -6,059853
12 LEU 9 14,484483
13 PHE 10 13,089985
14 ASP 11 -21,582466
15 ASN 12 -6,793066
16 ALA 13 5,913837
17 MET 14 11,591788
18 LEU 15 18,332824
19 ALA 17 6,671517
20 ARG 16 -8,382549
21 ARG 19 -8,861541
22 HIS 18 1,561942
23 HIS 21 1,565728
24 LEU 20 20,635109
25 LEU 23 20,699131
26 GLN 22 -2,790652
27 PHE 25 19,762727
28 ALA 24 7,952105
```

**Slika 8.4.** Prikaz datoteke s rezultatima računanja energije solvatacije  
*Solvation\_energy.txt*

Primjer izračunate ukupne energije solvatacije dan je na slici (Slika 8.5.), a primjer izračunate ukupne elektrostatske sile proteina na slici (Slika 8.6.).



**Slika 8.5.** Prikaz rezultata izračunate ukupne energije solvatacije proteina 1A22 u grafičkom sučelju



**Slika 8.6.** Prikaz izračunate elektrostatske sile proteina 1A22 u grafičkom sučelju

Za potrebe ovog rada i izračuna korelacija izračunate su vrijednosti elektrostatske sile i energije solvatacije za 48 PDB datoteka te su dobivene vrijednosti prikazane u tablici (Tablica 8.1). Prikazane vrijednosti su:

- oznaka PDB datoteke,
- elektrostatska sila unutar lanca,
- elektrostatska sila između lanaca,
- ukupna elektrostatska sila,
- vrijeme izvođenja algoritma za izračun elektrostatskih sila,
- energija solvatacije,
- vrijeme izvođenja algoritma za izračun energije solvatacije,
- zbroj energije solvatacije i elektrostatske sile između lanaca i
- zbroj energije solvatacije i ukupne elektrostatske sile.

U nastavku tablice za svaku PDB datoteku ispisan je izmjeren koeficijent privlačnosti, te su izračunati njegov prirodni logaritam, te negativna vrijednost prirodnog logaritma koeficijenta privlačnosti (Tablica 8.2).

**Tablica 8.1** Tablično ispisani rezultati izračuna elektrostatske sile i energije solvatacije, duljina trajanja izvođenja za više *pdb* datoteka

PDB	ele.sila u.	ele. sila iz.	UK. ele. Sila	vr.izvođenja	energ.solv.	vr. izvođenja	ele. sila iz. + ene. solv.	uk.ele.sila + ene.solv.
1a22	63229,17390	14309,70326	77538,87716	2,82067	2368,00545	4,21737	16677,70871	79906,88261
1a2k	132658,33964	49096,82815	181755,16779	11,25961	8074,07340	9,40992	57170,90155	189829,24119
1a37	70010,84426	11338,25157	81349,09583	3,29882	3251,34204	4,54872	14589,59361	84600,43787
1a3b	42168,21178	3840,77123	46008,98301	1,7951621	2484,03771	2,81857	6324,80894	48493,02072
1a3e	41631,27424	3828,85749	45460,13173	1,66933	2507,96560	2,78082	6336,82309	47968,09733
1a4y	248111,37387	90220,06915	338331,44302	22,23191	19375,91195	13,31482	109595,98110	357707,35497
1abt	-13043,52004	69745,88993	56702,36989	5,86573	257,49531	40,77702	70003,38524	56959,86520
1acb	46832,87998	5762,24392	52595,12390	1,99020	4749,85128	3,14992	10512,09520	57344,97518
1ahw	204125,51631	93383,04921	297508,56552	23,25112	9413,37875	14,43679	102796,42796	306921,94427
1ak4	-9462,74528	3816,91282	-5645,83246	10,49625	3884,15087	9,18762	7701,06369	-1761,68159
1akj	96011,64973	44520,35086	140532,00059	6,46971	4091,07174	7,99434	48611,42260	144623,07233
1an1	40926,46916	4082,90460	45009,37376	1,49946	3251,15913	2,66338	7334,06373	48260,53289
1aqc	28287,41424	8435,14588	36722,56012	1,40929	2315,65786	2,68855	10750,80374	39038,21798
1atn	141752,77468	14588,47555	156341,25023	7,99015	9442,87624	7,02127	24031,35179	165784,12647
1ava	256224,27877	71934,52268	328158,80145	22,63876	12561,64663	13,10510	84496,16931	340720,44808
1avg	66887,19940	11184,80297	78072,00237	3,44958	2769,97431	5,34774	13954,77728	80841,97668
1avp	-1625,85311	-21,86938	-1647,72249	2,23976	2837,62414	3,48966	2815,75476	1189,90165
1avx	60508,17821	10520,06007	71028,23828	3,44772	3915,56882	4,77312	14435,62889	74943,80710
1avz	32593,26477	8893,46935	41486,73412	1,47849	1720,94389	3,08281	10614,41324	43207,67801
2pms	166235,88696	49850,54800	216086,43496	13,81813	5428,24131	10,00342	55278,78931	221514,67627
2pnx	8160,48875	3644,96778	11805,45653	0,40265	497,19116	1,16811	4142,15894	12302,64769
2vlp	28310,17040	5268,06150	33578,23190	1,05906	-337,11461	2,33413	4930,94689	33241,11729
3aa0	101320,83282	27898,54880	129219,38162	5,57633	3050,60048	6,58715	30949,14928	132269,98210
3aa6	101634,68440	28716,50374	130351,18814	6,60603	3090,93176	7,05063	31807,43550	133442,11990

*Računalna metoda za određivanje proteinskih interakcija*

PDB	ele.sila u.	ele. sila iz.	UK. ele. Sila	vr.izvođenja	energ.solv.	vr. izvođenja	ele. sila iz. + ene. solv.	uk.ele.sila + ene.solv.
3aaa	108851,72932	35063,75451	143915,48383	8,10549	3887,11947	7,51200	38950,87398	147802,60330
3ajb	51117,34739	2321,29696	53438,64435	2,47674	5834,93496	3,11427	8156,23192	59273,57931
3alz	105486,85357	11781,37067	117268,22424	12,15509	14683,80100	12,04394	26465,17167	131952,02524
3bzd	63963,31163	8652,86881	72616,18044	2,54804	1527,75276	3,37222	10180,62157	74143,93320
3c09	151231,91250	79843,63628	231075,54878	16,67655	10210,72342	11,21557	90054,35970	241286,27220
3csf	274439,50758	0,00000	274439,50758	17,12115	41223,62167	10,64095	41223,62167	315663,12925
3chw	103460,32178	14532,90151	117993,22329	5,21771	6925,35207	5,50293	21458,25358	124918,57536
3cx6	104696,78445	13849,20899	118545,99344	4,93250	3685,33774	5,32886	17534,54673	122231,33118
3cx7	105431,60076	14839,90507	120271,50583	5,01639	3640,45469	5,35403	18480,35976	123911,96052
3doe	51451,20346	8301,85284	59753,05630	4,86539	2124,70847	5,98737	10426,56131	61877,76477
3dpo	73800,73272	17378,77533	91179,50805	3,70567	1667,80669	4,44596	19046,58202	92847,31474
3ds0	8421,12616	1341,97803	9763,10419	1,09891	440,07175	1,60852	1782,04978	10203,17594
3eyf	127423,72582	48866,35026	176290,07608	11,15895	8479,49634	9,11213	57345,84660	184769,57242
3h8d	74327,55394	58023,52907	132351,08301	6,34179	1581,38271	7,04224	59604,91178	133932,46572
3ncb	68679,40062	13946,96689	82626,36751	3,18348	4692,12668	4,36837	18639,09357	87318,49419
3qux	128196,58919	46852,25626	175048,84545	17,50073	7976,79393	15,60910	54829,05019	183025,63938
3rt0	192782,59931	64313,23955	257095,83886	23,15256	9167,68907	18,68562	73480,92862	266263,52793
3s8v	351642,03192	58208,40052	409850,43244	28,22137	29408,34331	14,10125	87616,74383	439258,77575
3t2n	251473,17766	105497,34858	356970,52624	45,27332	25464,78728	31,52439	130962,13586	382435,31352
3uaj	241800,06561	101368,49011	343168,55572	34,00742	23616,71902	15,59023	124985,20913	366785,27474
3ssb	155763,28288	40576,03306	196339,31594	15,03448	6775,64521	14,18933	47351,67827	203114,96115
3t04	22410,38244	4226,25873	26636,64117	0,94582	1540,31792	1,99649	5766,57665	28176,95909
4er4	64807,16745	1897,88916	66705,05661	2,72420	7668,00668	3,18557	9565,89584	74373,06329
4thn	42651,46939	2267,43722	44918,90661	1,79307	2512,32676	2,75356	4779,76398	47431,23337



**Tablica 8.2** Tablično ispisani podaci koeficijenta privlačnosti po *pdb* datotekama

PDB	koef. privlačnosti	ln koef.priv.	neg. ln koef.
1a22	3,40E-10	-21,8020755	21,8020755
1a2k	1,50E-07	-15,7126305	15,7126305
1a37	1,20E-07	-15,9357741	15,9357741
1a3b	3,00E-10	-21,9272386	21,9272386
1a3e	3,50E-09	-19,4705029	19,4705029
1a4y	1,00E-15	-34,5387764	34,5387764
1abt	1,40E-06	-13,4790383	13,4790383
1acb	2,00E-10	-22,3327037	22,3327037
1ahw	3,40E-09	-19,4994904	19,4994904
1ak4	1,70E-05	-10,9822972	10,9822972
1akj	1,26E-04	-8,97922865	8,97922865
1an1	9,00E-10	-20,8286264	20,8286264
1aqc	4,56E-06	-12,2981879	12,2981879
1atn	4,50E-10	-21,5217735	21,5217735
1ava	2,20E-10	-22,2373936	22,2373936
1avg	3,00E-12	-26,5324088	26,5324088
1avp	1,20E-08	-18,2383592	18,2383592
1avx	6,00E-14	-30,4444318	30,4444318
1avz	1,58E-05	-11,0555006	11,0555006
2pms	1,03E-08	-18,3911219	18,3911219
2pnx	7,90E-06	-11,7486478	11,7486478
2vlp	5,43E-13	-28,2416671	28,2416671
3aa0	4,00E-06	-12,4292162	12,4292162
3aa6	2,60E-07	-15,1625842	15,1625842
3aaa	2,10E-08	-17,6787434	17,6787434
3ajb	4,08E-08	-17,0145838	17,0145838
3alz	5,20E-07	-14,469437	14,469437
3bzd	9,60E-05	-9,25116237	9,25116237
3c09	4,30E-08	-16,9620657	16,9620657
3csf	1,10E-06	-13,7202004	13,7202004
3chw	5,70E-08	-16,6802146	16,6802146

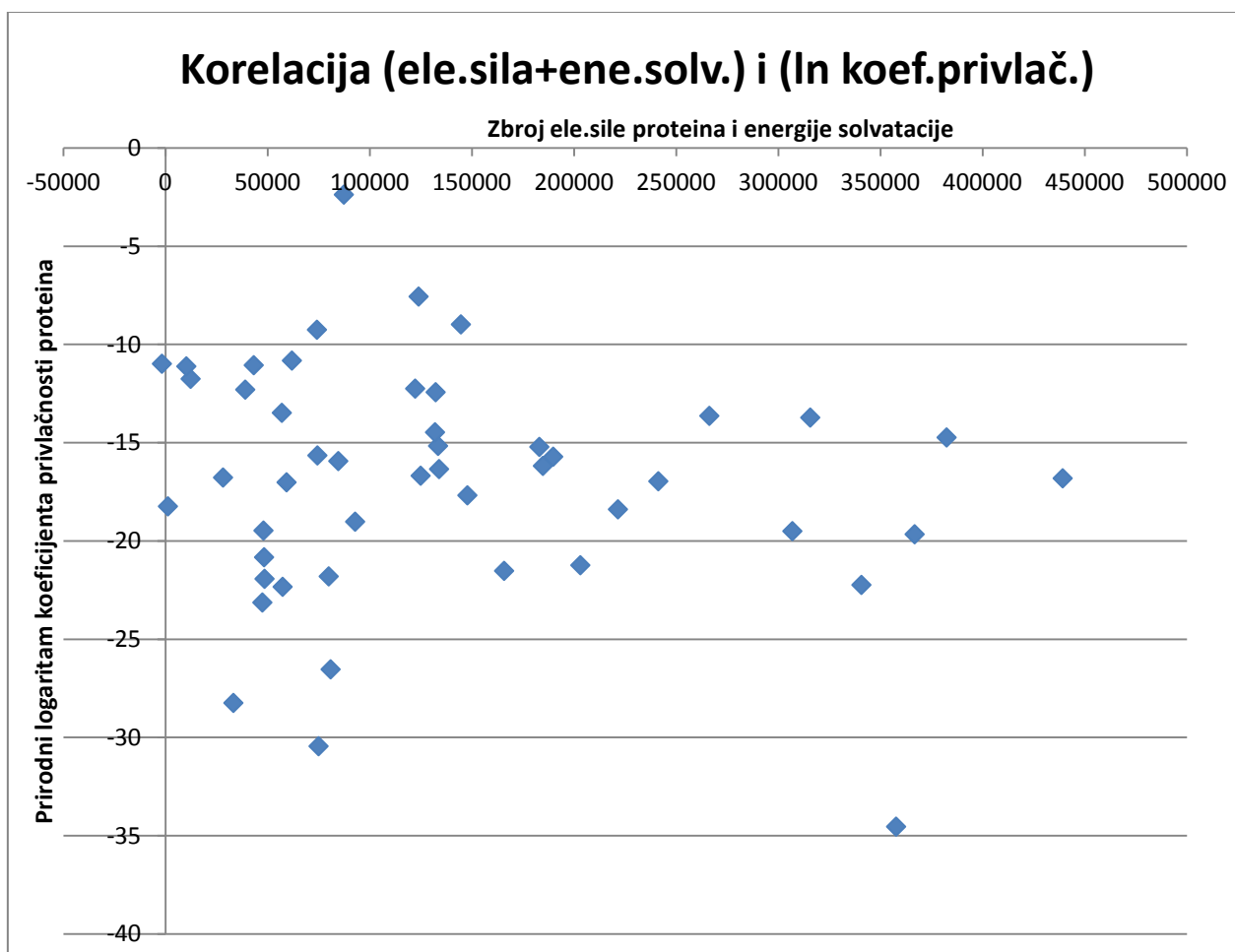
PDB	koef. privlačnosti	ln koef.priv.	neg. ln koef.
3cx6	4,80E-06	-12,2468946	12,2468946
3cx7	5,20E-04	-7,56168175	7,56168175
3doe	2,00E-05	-10,8197783	10,8197783
3dpo	5,50E-09	-19,0185177	19,0185177
3ds0	1,49E-05	-11,1141493	11,1141493
3eyf	9,40E-08	-16,1799711	16,1799711
3h8d	8,00E-08	-16,3412392	16,3412392
3ncb	9,30E-02	-2,37515579	2,37515579
3qux	2,47E-07	-15,2138775	15,2138775
3rt0	1,20E-06	-13,633189	13,633189
3s8v	5,00E-08	-16,8112428	16,8112428
3t2n	4,00E-07	-14,7323014	14,7323014
3uaj	2,90E-09	-19,6585551	19,6585551
3ssb	6,00E-10	-21,2340915	21,2340915
3t04	5,20E-08	-16,7720221	16,7720221
4er4	1,60E-07	-15,648092	15,648092
4thn	9,00E-11	-23,1312114	23,1312114

U tablici (Tablica 8.3) su prikazani parametri izračunati za datoteke navedene u tablicama.

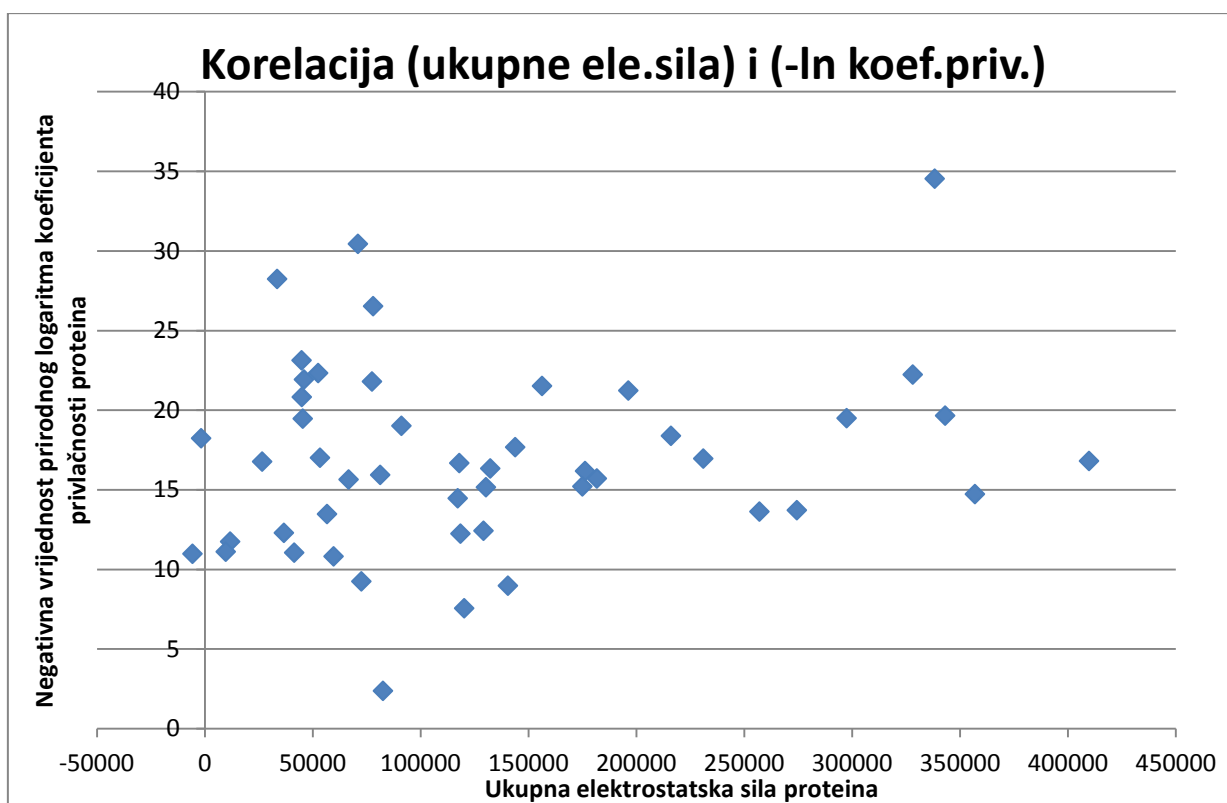
**Tablica 8.3** Vrijednosti izračunatih parametara analize rezultata algoritma

Broj obrađenih PDB datoteka	48
Prosječno vrijeme izvođenja izračuna dijela programa za izračun elektrostatske sile	9,08377 s
Prosječno vrijeme izvođenja dijela programa za izračun energije solvatacije	8,25491 s
Prosječno ukupno vrijeme izvođenja programa	8,25687 s
Koeficijent korelacije elektrostatske sile i energije solvatacije s prirodnim logaritmom koeficijenta privlačnosti	-0,166049736
Koeficijent korelacije elektrostatske sile s negativnim prirodnim logaritmom koeficijenta privlačnosti	0,169552879

Kao što se može vidjeti, korelacija između elektrostatske sile proteina i koeficijenta privlačnosti nema, tj, korelacija je neznatna. Isto vrijedi i za korelaciju zbroja ukupne elektrostatske sile i energije solvatacije proteina s koeficijentom privlačnosti. Na slikama (Slika 8.7 i Slika 8.8) prikazani su točkasti dijagrami (engl *scatter chart*) korelacije, gdje se vidi velika raspršenost izračunatih vrijednosti što također pokazuje neznatnu korelaciju.



**Slika 8.7.** Prikaz korelacije zbroja elektrostatske sile i energije solvatacije proteina s prirodnim logaritmom koeficijenta privlačnosti točkastim dijagramom



**Slika 8.8.** Prikaz korelacije elektrostatske sile proteina s negativnom vrijednošću prirodnog logaritma koeficijenta privlačnosti točkastim dijagramom

## **9. Zaključak**

Novi izazovi suvremenog doba podrazumijevaju suradnju znanstvenika s raznih područja. Spojem molekularne biologije, informacijskih tehnologija i računarske znanosti nastala je bioinformatika, koja omogućuje veću efikasnost u obradi, organizaciji i uporabi podataka o kemijskim spojevima i njihovim svojstvima. 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. Nove informacije u predviđanju proteinskih interakcija mogu olakšati stvaranje novih lijekova, objašnjavanje metaboličkih reakcija i u drugim sličnim procesima.

Zadatak ovog diplomskog rada je utvrđivanje elektrostatske sile pri interakciji aminokiselina i utvrđivanje energije otapala pomoću površine dostupne otapalu pojedine aminokiseline.

## **10. Sažetak**

### **Računalna metoda za određivanje proteinskih interakcija**

Algoritam implementiran ovim radom računa elektrostatsku silu i energiju solvatacije proteina. Elektrostatska sila računa se kao zbroj elektrostatskih sila pri međusobnim interakcijama svih atoma u svim aminokiselinama. Podaci potrebni za računanje dobivaju se učitavanjem PDB datoteka i učitavanjem naboja svakog atoma u određenoj aminokiselini. Energija solvatacije računa se kao zbroj energija solvatacije svih aminokiselina u proteinu. Za izračun energije solvatacije potrebni podaci učitavaju se iz datoteke s vrijednostima površina dostupnih otapalu i potencijalom otapala te iz datoteke s radijusom atoma svake aminokiseline. Dobivene vrijednosti se koreliraju s koeficijentom privlačnosti proteina.

Ključne riječi: proteini, elektrostatska sila, energija solvatacije, koeficijent privlačnosti, ASA, aminokiseline, JAVA, korelacija

## **Summary**

### **Computing method for determining protein interactions**

The algorithm implemented in this thesis calculates electrostatic force of a protein by calculating the force between every pair of amino acids in protein. Solvation energy is also calculated for a protein by summing the energy of each amino acid in it. Electrostatic energy is calculated as the sum of electrostatic energies in the interaction of all two pairs of atoms in 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 and from a file with atom radius. The values obtained are correlated with protein's affinity data .

Keywords: proteins, electrostatic energy, solvation energy, affinity data, ASA, amino acids, JAVA, correlation

## 11. Literatura

- [1] Internet: RCSB Protein Data Bank,  
<http://www.pdb.org/pdb/home/home.do>, ožujak 2013.
- [2] Internet: IUPAC – Nomenclature and symbolism for Amino Acids,  
<http://www.chem.qmul.ac.uk/iupac/AminoAcid/AA1n2.html#AA1>, ožujak 2013.
- [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.
- [5] K. Kušević, Implementacija algoritma za računanje proteinskih interakcija, Završni rad br. 2131, FER, 2011.
- [6] J. Petrović, Predviđanje mjesta proteinskih interakcija koristeći algoritam slučajnih šuma, Diplomski rad br. 66, FER, 2010.
- [7] Internet: M. Chitty, Overviews: „Bioinformatics, Cheminformatics and beyond“,  
<http://www.genomicglossaries.com/content/chapterinfosourcetext.asp>, svibanj 2013.
- [8] Internet: „Bioinformatics Factsheet: Bioinformatics (29.3.2004.)“,  
<http://www.ncbi.nlm.nih.gov/About/primer/bioinformatics.html>, svibanj 2013.
- [9] B. Lee, F. M. Richards, „The Interpretation of Protein Structures: Estimation of Static Accessibility“, *J. Mol. Biol.*, 55(1971), 379-400,  
[http://compbio.chemistry.uq.edu.au/mediawiki/upload/d/d5/Lee1971\\_JMolBiol55-379.pdf](http://compbio.chemistry.uq.edu.au/mediawiki/upload/d/d5/Lee1971_JMolBiol55-379.pdf), travanj 2013
- [10] Kukić P., Nielsen J.E., „Electrostatics in protein and protein – ligand complexes“, *Future Med. Chem.* (2010) 2(4), 647 – 666,  
<http://enzyme.ucd.ie/main/images/6/6a/ElectrostaticsInProteins.pdf>, lipanj 2013.
- [11] L. Bilić – Zulle, Korelacija – predavanje iz predmeta Statistika Farmaceutsko – biokemijskog fakulteta Sveučilišta u Zagrebu,  
[http://mi.medri.hr/assets/P6\\_Korelacija.pdf](http://mi.medri.hr/assets/P6_Korelacija.pdf), lipanj 2013

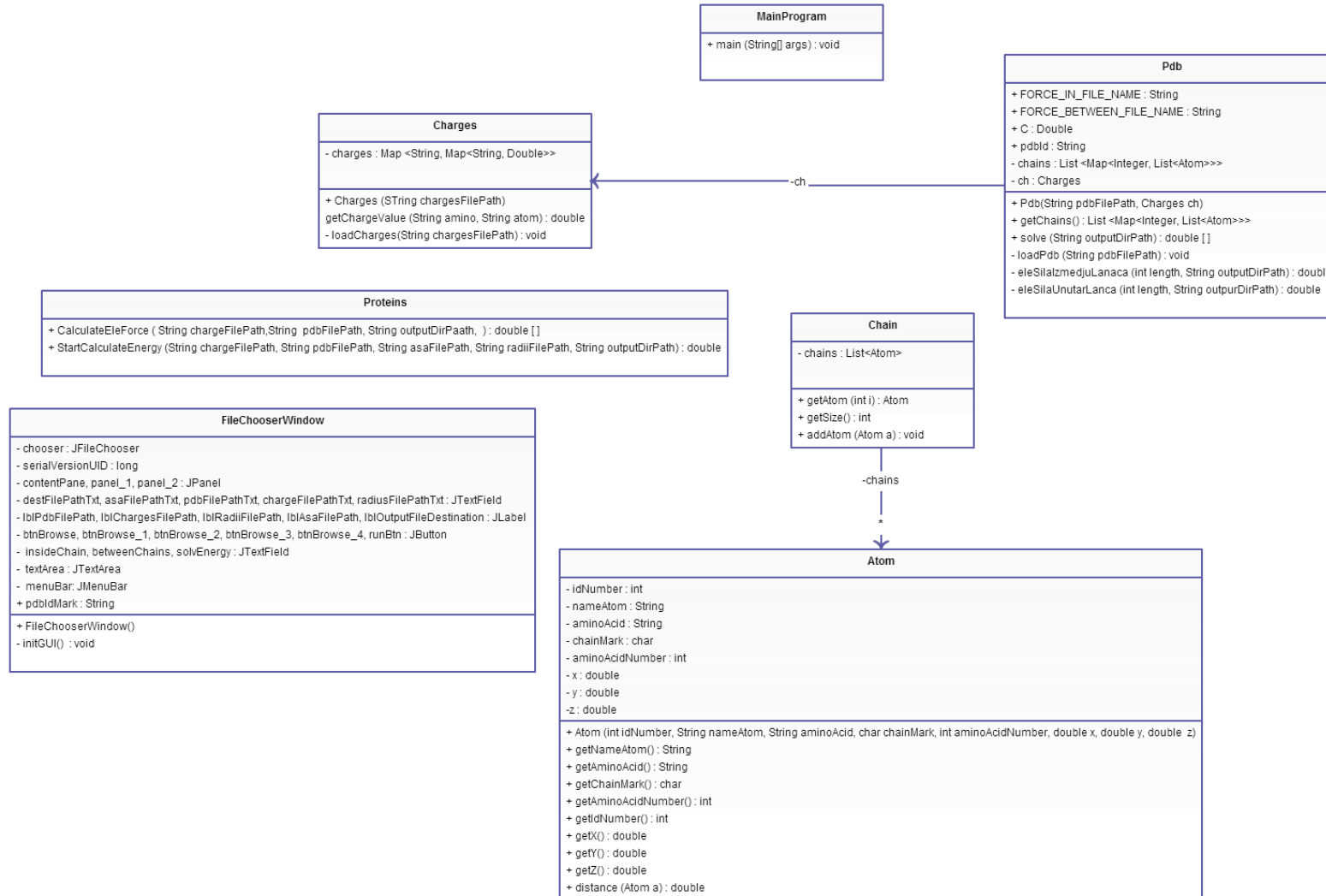


- [12] M. Udovičić, K. Baždarić, L. Bilić – Zulle, M. Petrovečki, Što treba znati kad izračunavamo koeficijent korelacije?, *Biochemia Medica* 2007; 17(1). 1 – 138, 10 – 15, <http://hrcak.srce.hr/file/19716>, lipanj 2013

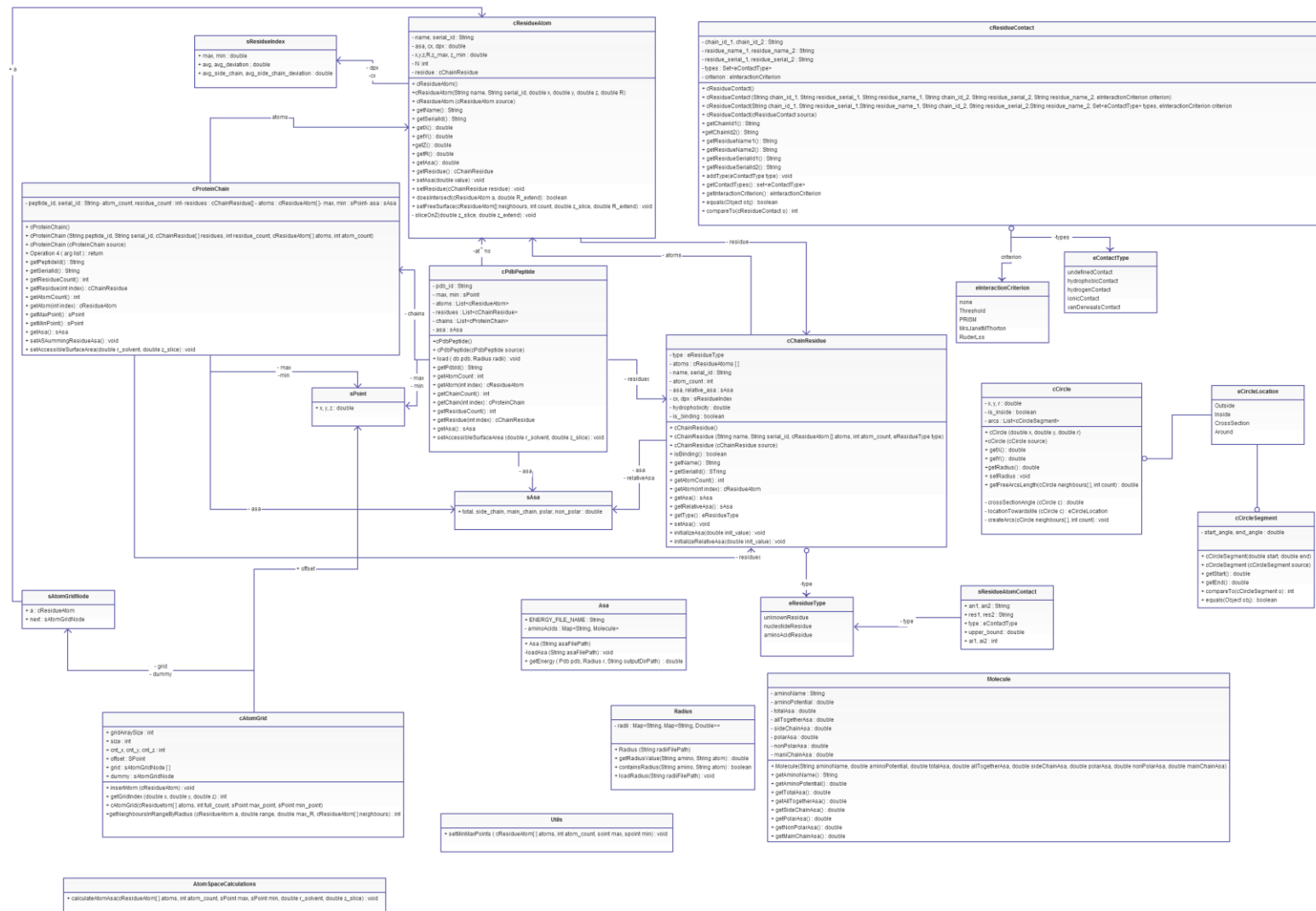
## **Dodatak - Skice dijagrama klasa**

U nastavku su dani dijagrami klasa paketa surface i proteinInteraction koje se koriste u razvijenoj aplikaciji. Na dijagramima je prikazana međusobna povezanost svih klasa aplikacije (Skica A i Skica B).

## Računalna metoda za određivanje proteinskih interakcija



**Skica A** – Dijagram svih klasa paketa *proteinInteractions* – izračun elektrostatske sile



**Skica B –** Dijagram svih klasa paketa *surface* – izračun energije solvatacije

