

Sveučilište u Zagrebu
Fakultet Elektrotehnike i Računarstva

Diplomski rad br. 1163

**Napredni alat za analizu površina proteina i mjesta
proteinskih interakcija**

Tesa Orlović

Zagreb, prosinac 2008

Sadržaj

1	Uvod	3
2	Podaci	4
2.1	PDB format	4
2.2	mmCIF	5
2.3	PDBML	8
3	Metode	9
3.1	Sekundarne strukture	9
3.2	Energija otapala	11
3.3	PIADA 2	12
4	Opis sustava	18
4.1	Opći opis PSAIA sustava	18
4.2	Modul za analizu površina proteina, PSA	19
4.3	Modul za analizu proteinskih interakcija, PIA	22
5	Korištenje sustava	24
5.1	Konfiguracijska datoteka PSA modula	26
5.2	Konfiguracijska datoteka PIA modula	27
5.3	Datoteka koja sadrži popis ulaznih datoteka	29
5.4	Datoteka koja sadrži vrijednosti Van der Waalsovih radijusa atoma	29
5.5	Datoteka koja sadrži vrijednosti hidrofobnosti ostataka	29
5.6	Datoteka koja sadrži vrijednosti standardnih ASA za ostatke	30
5.7	Datoteka koja sadrži parametre za izračun energije otapala	30
5.8	Datoteka koja sadrži definicije grupa atoma i tipove interakcija	31
6	Zaključak	33
7	Dodatak A - dijagram toka PSA modula	34
8	Dodatak B - dijagram toka PIA modula	34
9	Dodatak C - primjer konfiguracijske datoteke modela PSA	34
10	Dodatak D – primjer konfiguracijske datoteke modela PIA	40
11	Popis literature	41

1 Uvod

Alat za analizu površina proteina i mjesta proteinskih interakcija (eng. *Protein Structure and Interaction Analyzer, PSAIA*) je razvijen u sklopu diplomskog rada Josipa Mihela na Fakultetu Elektrotehnike i Računarstva 2006. godine. Zbog raznih nedostataka prve verzije programa, pojavila se potreba za nadogradnjom osnovnom programa prema zahtjevima i potrebama samih korisnika.

Prva verzija programa je primala za ulaz PDB (*Protein Data Bank*) datoteke koje u sebi sadrže strukture samostalnih proteina, i proteina stvorenih interakcijom različitih proteina. Zbog razvoja sve većeg broja informatičkih programa u području analize proteina, a i velikom broju nedostataka koji taj format ima, većinom uzrokovane ljudskom pogreškom i nestandardiziranosti izgleda formata, počeli su se u sve većem broju koristiti i drugi formati prikaza proteinskih struktura od čega ćemo izdvojiti nezaobilazni XML (*Extensible Markup Language*), i malo manje poznati CIF (*Crystallographic Information File*). Velika prednost oba formata, pogotovo XML-a, je jako stroga forma, čime je uvelike olakšana elektronička obrada podataka, i donekle izbačen ljudski faktor pogreške. Oba formata će biti detaljno opisana u drugom poglavlju rada.

Iako je originalni program od početka sadržavao impresivan broj algoritama strukturne analize, trebalo je taj spektar nadopuniti algoritmom za izračunavanje energije otapala razvijen u suradnji sa Laboratorijom za kemijsku i biološku kristalografiju na Institutu Ruđer Bošković u Zagrebu, i algoritmom za izračunavanje sekundarne strukture proteina preko već postojećeg algoritma Stride, koji je napravio Europski laboratorij za molekularnu biologiju iz Heidelberga, Njemačka. Također je bilo potrebno unaprijediti algoritam za izračunavanje slobodnih površina otapalu (eng. *Accessible Solvent Area, ASA*), da uz standardne atome prihvaća i atome vode, hetatome i vodike.

Od algoritama vezanih uz traženje mjesta proteinskih interakcija, algoritam PIADA iz prve verzije je u potpunosti zamjenjen novijim, također razvijen u suradnji sa Laboratorijom za kemijsku i biološku kristalografiju na Institutu Ruđer Bošković u Zagrebu.

Formati izlaznih datoteka programa su također trebali promjenu zbog veće preglednosti i kraćeg zapisa, kao i dodavanje prikaza po atomima i hetatomima.

Performanse programa su poboljšane dodavanjem višedretvenosti.

U sklopu ovog diplomskog rada odlučili smo napraviti drugu verziju programa koja uključuje sve gore navedene stvari.

2 Podaci

Protein Data Bank (PDB) je jedinstveni svjetski repozitorij informacija o 3D strukturama bioloških molekula, uključujući proteine i nukleinske kiseline. To su molekule koje se nalaze u svim organizmima od bakterija, biljaka, pa preko životinja, do ljudi. Razumijevanje struktura molekula pomaže u razumijevanju rada svih organizama, njihovom zdravlju i bolestima, i u razvoju lijekova za izlječenje tih bolesti.

Podaci o strukturi molekula su dobiveni pomoću metoda rentgenske difrakcije i NMR spektroskopije, skupljeni i objavljeni na internetu od biologa i biokemičara diljem svijeta. PDB je predložio format zapisa tih podataka još u u 70-ima istog imena, koji je otada prihvaćen kao standard, i korišten u velikom broju programa. Razvojem tehnologije i računarnih znanosti razvio se novi format mmCIF (*macromolecular Crystallographic Information File*), koji se sve od kraja 90-ih pa do danas preporučuje kao novi standard. Kao i u drugim računarnim područjima XML se uvukao i našao svoju primjenu i u prikazu strukture proteina svojom strogom shemom i pravilima pod nazivom PDBML (*The Protein Data Bank Markup Language*), čime je uvelike olakšana elektronička obrada tih podataka. U nastavku ćemo kratko opisati svaki od tih formata.

2.1 PDB format

Svaka linija PDB datoteke sastoji se od maksimalno 80 znakova. Prvih 6 znakova opisuju sadržaj tog redka posebnim unaprijed određenim imenima, poravnata ulijevo i pisani velikih slovima (npr. ATOM, HEADER, MODEL, TITLE). Od tih oznaka bi htjeli izdvojiti ATOM, koji određuje početak zapisa podataka o jednom standardnom atomu (atoma aminokiseline ili nukleinske kiseline), i HETATM koji određuje početak zapisa podataka jednoga heterogena (vode, metali ili ligandi).

	1	2	3	4	5	6	7	8
1234567890123456789012345678901234567890123456789012345678901234567890								
ATOM	145	N VAL A	25	32.433	16.336	57.540	1.00 11.92	N
ATOM	146	CA VAL A	25	31.132	16.439	58.160	1.00 11.85	C
ATOM	147	C VAL A	25	30.447	15.105	58.363	1.00 12.34	C
ATOM	148	O VAL A	25	29.520	15.059	59.174	1.00 15.65	O
ATOM	149	CB AVAL A	25	30.385	17.437	57.230	0.28 13.88	C
ATOM	150	CB BVAL A	25	30.166	17.399	57.373	0.72 15.41	C
ATOM	151	CG1AVAL A	25	28.870	17.401	57.336	0.28 12.64	C
ATOM	152	CG1BVAL A	25	30.805	18.788	57.449	0.72 15.11	C
ATOM	153	CG2AVAL A	25	30.835	18.826	57.661	0.28 13.58	C
ATOM	154	CG2BVAL A	25	29.909	16.996	55.922	0.72 13.25	C

Slika 2-1. Primjer izgleda PDB datoteke

Kraj PDB datoteke određuje oznaka END. Format po stupcima jednog ATOM ili HETATM retka je prikazan u tablici 2-1.

Tablica 2-1. Format ATOM/HETATM redka u PDB datoteci

Redni broj stupca	Tip podataka	Opis polja
1-6	Oznaka polja	Oznaka atoma ili hetatoma (ATOM ili HETATM)
7-11	Cijeli broj	Serijski broj atoma
13-16	String	Ime atoma.
17	Znak	Oznaka alternativne lokacije.
18-20	String	Ime residuuma (molekule).
22	Znak	Oznaka lanca.
23-26	Cijeli broj	Redni broj residuuma.
31-38	Decimalni broj	X koordinata u Angstromima.
39-46	Decimalni broj	Y koordinata u Angstromima.
47-54	Decimalni broj	Z koordinata u Angstromima.
55-60	Decimalni broj	Okupiranost.
61-66	Decimalni broj	Faktor temperature.
77-78	Dva znaka	Simbol elementa, poravnan udesno
79-80	Dva znaka	Naboj atoma.

2.2 mmCIF

Od sredine 1970-ih, PDB je koristio datoteke bazirane na stupcima za pohranjivanje baze podataka proteina i nukleinskih kiselina po uzoru na druge slične formate, ograničenima tehnologijom bušenih kartica. Zbog sve većeg interesa za razvijanje baza podataka, 1990. godine je Međunarodna udruga za kaliografiju, *IUCr* (*International Union of Crystallography*) počela sa razvojem novog formata za prikaz malih molekularnih i kristalografskih struktura pod nazivom *Crystallographic Information File (CIF)*. Prva verzija je objavljena 1996. godine pod nazivom mmCIF (*macromolecular Crystallographic Information File*). RCSB (Research Collaboratory for Structure Bioinformatics) je 1998. godine preuzeo upravljanje PDB-a, i usvojio mmCIF format kao osnovni format za zapis i obradu baze podataka, koji se zadržao do danas.

Za razliku od starog PDB standarda koji je baziran na redcima sa poljima ovisnima o broju stupca, mmCIF se bazira na imenovanim poljima i petljama. Tako na primjer, zaglavlje PDB datoteke:

```
HEADER PLANT SEED PROTEIN 11-OCT-91 1CBN
```

postaje u mmCIF-u:

```
_struct.entry_id '1CBN'
_struct.title 'PLANT SEED PROTEIN'
_struct.keywords.entry_id '1CBN'
_struct.keywords.text 'plant seed protein'
_database_2.database_id 'PDB'
_database_2.database_code '1CBN'
_database_PDB_rev.rev_num 1
_database_PDB_rev.date_original '1991-10-11'
```

Petlje u mmCIF-u počinju ključnom riječi *loop_*, nakon čega slijedi redoslijed polja u tablici izraženima preko imena polja po mmCIF rječniku, i završavaju praznim redkom ili oznakom komentara. Zapisi atoma i hetatoma su smješteni u petlju pod imenom *_atom_site*, po uzoru na primjer na slici 2-2.

```

loop_
_atom_site.group_PDB
_atom_site.id
_atom_site.type_symbol
_atom_site.label_atom_id
_atom_site.label_alt_id
_atom_site.label_comp_id
_atom_site.label_asym_id
_atom_site.label_entity_id
_atom_site.label_seq_id
_atom_site.Cartn_x
_atom_site.Cartn_y
_atom_site.Cartn_z
_atom_site.occupancy
_atom_site.B_iso_or_equiv
_atom_site.auth_seq_id
_atom_site.auth_comp_id
_atom_site.auth_asym_id
_atom_site.auth_atom_id
_atom_site.pdbx_PDB_model_num

ATOM      1      N      N      .  VAL  A  1  1    6.204  16.869   4.854  7.00  49.05  1  VAL  A  N    1
ATOM      2      C      CA     .  VAL  A  1  1    6.913  17.759   4.607  6.00  43.14  1  VAL  A  CA   1
ATOM      3      C      C      .  VAL  A  1  1    8.504  17.378   4.797  6.00  24.80  1  VAL  A  C    1
ATOM      4      O      O      .  VAL  A  1  1    8.805  17.011   5.943  8.00  37.68  1  VAL  A  O    1
ATOM      5      C      CB     .  VAL  A  1  1    6.369  19.044   5.810  6.00  72.12  1  VAL  A  CB   1
ATOM      6      C      CG1    .  VAL  A  1  1    7.009  20.127   5.418  6.00  61.79  1  VAL  A  CG1  1
ATOM      7      C      CG2    .  VAL  A  1  1    5.246  18.533   5.681  6.00  80.12  1  VAL  A  CG2  1
ATOM      8      N      N      .  LEU  A  1  2    9.096  18.040   3.857  7.00  26.44  2  LEU  A  N    1
ATOM      9      C      CA     .  LEU  A  1  2   10.600  17.889   4.283  6.00  26.32  2  LEU  A  CA   1
ATOM     10      C      C      .  LEU  A  1  2   11.265  19.184   5.297  6.00  32.96  2  LEU  A  C    1
ATOM     11      O      O      .  LEU  A  1  2   10.813  20.177   4.647  8.00  31.90  2  LEU  A  O    1
ATOM     12      C      CB     .  LEU  A  1  2   11.099  18.007   2.815  6.00  29.23  2  LEU  A  CB   1
ATOM     13      C      CG     .  LEU  A  1  2   11.322  16.956   1.934  6.00  37.71  2  LEU  A  CG   1
ATOM     14      C      CD1    .  LEU  A  1  2   11.468  15.596   2.337  6.00  39.10  2  LEU  A  CD1  1
ATOM     15      C      CD2    .  LEU  A  1  2   11.423  17.268   0.300  6.00  37.47  2  LEU  A  CD2  1
....
HETATM   4057  MG  MG     .  MG  E  3  .   64.283  81.032  37.544  1.00  34.76  ?  ?  ?  ?  ?
HETATM   4058  MG  MG     .  MG  F  3  .   62.741  68.793  44.838  1.00  34.00  ?  ?  ?  ?  ?
HETATM   4059  P   PG     .  MG  G  4  .   63.837  81.063  41.074  1.00  8.68  ?  ?  ?  ?  ?
....
#

```

Slika 2-2. Primjer izgleda mmCIF datoteke

Opis imena polja petlje *_atom_site* se nalazi u tablici 2-2.

Tablica 2-2. Opis imena polja u mmCIF datoteci

Ime polja	Tip podataka	Opis polja
group_PDB	String	Oznaka atoma ili hetatoma (ATOM ili HETATM)
id	Cijeli broj	Serijski broj atoma
type_symbol	String	Simbol elementa.
label_atom_id	String	Ime atoma.
label_alt_id	String	Alternativno ime atoma.
label_comp_id	String	Oznaka residuuma.
label_asym_id	String	Ime lanca.
label_entity_id	Cijeli broj	Redni broj lanca.
label_seq_id	Cijeli broj	Redni broj residuuma.
Cartn_x	Decimalni broj	X koordinata u Angstromima.
Cartn_y	Decimalni broj	Y koordinata u Angstromima.
Cartn_z	Decimalni broj	Z koordinata u Angstromima.
occupancy	Decimalni broj	Okupiranost.
B_iso_or_equiv	Decimalni broj	Faktor temperature.
auth_seq_id	String	Alternativni redni broj residuuma.
auth_comp_id	String	Alternativna oznaka residuuma.
auth_asym_id	String	Alternativno ime lanca.
auth_atom_id	String	Alternativno ime atoma.
pdtx_PDB_model_num	Cijeli broj	Broj modela PDB-a.

2.3 PDBML

The Protein Data Bank Markup Language (PDBML) predstavlja PDB standardni format (mmCIF) koristeći XML tehnologiju. PDBML datoteke imaju istu logičku organizaciju kao mmCIF datoteke, tj. imena i značenje svih polja su jednaka (tablica 2-2). PDBML oznake počinju sa nizom znakova *PDX*:. Blok koji sadrži podatke o atomima i hetatomima proteina počinje oznakom `<PDBx:atom_siteCategory>`. Primjer izgleda PDBML datoteke se može vidjeti na slici 2-3.

```
<PDBx:atom_siteCategory>
  <PDBx:atom_site id="1">
    <PDBx:group_PDB>ATOM</PDBx:group_PDB>
    <PDBx:type_symbol>N</PDBx:type_symbol>
    <PDBx:label_atom_id>N</PDBx:label_atom_id>
    <PDBx:label_alt_id xsi:nil="true" />
    <PDBx:label_comp_id>GLY</PDBx:label_comp_id>
    <PDBx:label_asym_id>A</PDBx:label_asym_id>
    <PDBx:label_entity_id>1</PDBx:label_entity_id>
    <PDBx:label_seq_id>1</PDBx:label_seq_id>
    <PDBx:Cartn_x>82.558</PDBx:Cartn_x>
    <PDBx:Cartn_y>68.558</PDBx:Cartn_y>
    <PDBx:Cartn_z>48.878</PDBx:Cartn_z>
    <PDBx:occupancy>1.00</PDBx:occupancy>
    <PDBx:B_iso_or_equiv>37.72</PDBx:B_iso_or_equiv>
    <PDBx:auth_seq_id>14</PDBx:auth_seq_id>
    <PDBx:auth_comp_id>GLY</PDBx:auth_comp_id>
    <PDBx:auth_asym_id>A</PDBx:auth_asym_id>
    <PDBx:auth_atom_id>N</PDBx:auth_atom_id>
    <PDBx:pdbx_PDB_model_num>1</PDBx:pdbx_PDB_model_num>
  </PDBx:atom_site>
  <PDBx:atom_site id="2">
    <PDBx:group_PDB>ATOM</PDBx:group_PDB>
    <PDBx:type_symbol>C</PDBx:type_symbol>
    <PDBx:label_atom_id>CA</PDBx:label_atom_id>
    <PDBx:label_alt_id xsi:nil="true" />
    <PDBx:label_comp_id>GLY</PDBx:label_comp_id>
    <PDBx:label_asym_id>A</PDBx:label_asym_id>
    <PDBx:label_entity_id>1</PDBx:label_entity_id>
    <PDBx:label_seq_id>1</PDBx:label_seq_id>
    <PDBx:Cartn_x>81.442</PDBx:Cartn_x>
    <PDBx:Cartn_y>69.187</PDBx:Cartn_y>
    <PDBx:Cartn_z>48.108</PDBx:Cartn_z>
    <PDBx:occupancy>1.00</PDBx:occupancy>
    <PDBx:B_iso_or_equiv>35.91</PDBx:B_iso_or_equiv>
    <PDBx:auth_seq_id>14</PDBx:auth_seq_id>
    <PDBx:auth_comp_id>GLY</PDBx:auth_comp_id>
    <PDBx:auth_asym_id>A</PDBx:auth_asym_id>
    <PDBx:auth_atom_id>CA</PDBx:auth_atom_id>
    <PDBx:pdbx_PDB_model_num>1</PDBx:pdbx_PDB_model_num>
  </PDBx:atom_site>
  . . .
</PDBx:atom_siteCategory>
```

Slika 2-3. Primjer izgleda PDBML datoteke

3 Metode

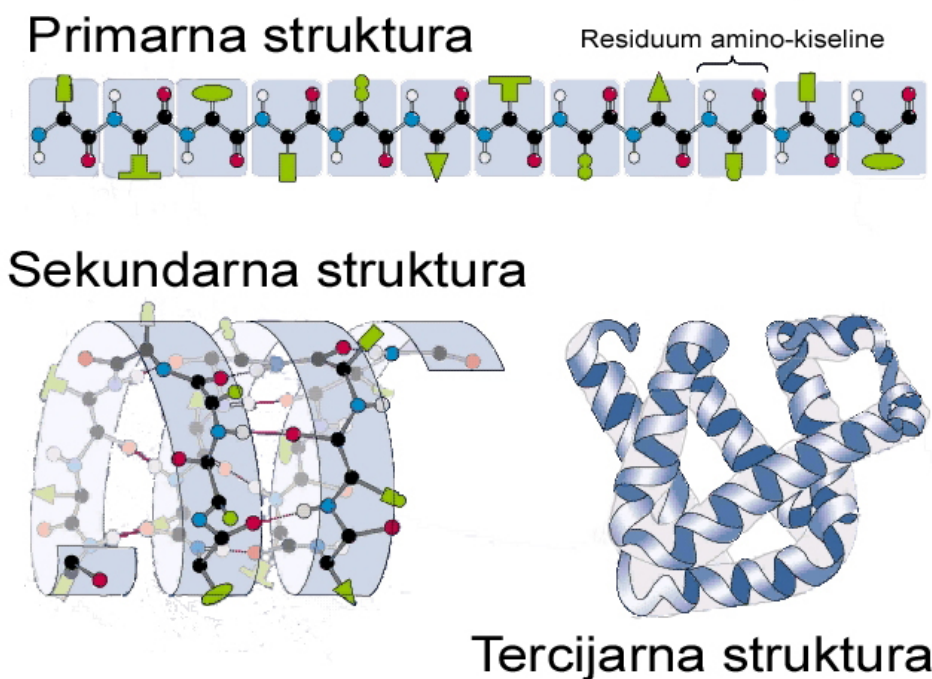
Nova verzija PSAIA donosi preinake i nadopunjenja već postojećih algoritama (dodavanje ASA za hetatome, vodike i vodu), i neke potpuno nove algoritme, koje ćemo detaljnije opisati u nastavku.

3.1 Sekundarne strukture

U biokemiji i strukturalnoj biologiji definiramo tri strukture proteina i nukleinskih kiselina: primarnu, sekundarnu i tercijarnu strukturu.

Primarna struktura sadrži atome i kemijske veze koje spajaju te atome. Po dogovoru, primarna struktura počinje amino-terminalom, N, a završava sa carboksil-terminalom, C.

Sekundarna struktura se definira kao generalna trodimenzionalna forma lokalnih segmenata biopolimerida, dok tercijalna struktura dodjeljuje atomima točne koordinate u prostoru.



Slika 3-1. Primarna, sekundarna i tercijarna struktura

Sekundarne strukture atoma se izračunavaju koristeći njihove koordinate, a bazirane su na energiji vodikovih veza i uvrtnog kuta glavnog lanca proteina. Najčešće sekundarne strukture koje se javljaju u prirodi u proteinima su α -uzvojnice (engl. α -helix) i β -ploče (engl. β -pleated sheet).

Za računanje sekundarnih struktura atoma koristimo već postojeći algoritam Stride koji je razvio Europski laboratorij za molekularnu biologiju iz Heidelberga, Njemačke.

Stride definira sedam sekundarnih struktura:

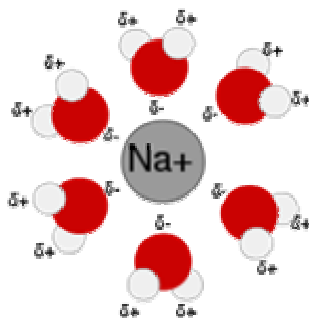
- H (α uzvojnica) – spirala sa 3 zavoja, minimalne duljine 3 ostatka
- G (3_{10} uzvojnica) – spirala sa 4 zavoja, minimalne duljine 4 ostatka
- I (π uzvojnica) – spirala sa 5 zavoja, minimalne duljine 5 ostatka
- E (Extended conformation) – produžetak β -mosta
- B (Isolated bridge) – izoliran β -most
- T (Turn) – zavoje vodikove veze (3, 4 ili 5 zavoja)
- C (Coil) – namotaj, tj. svi ostali oblici sekundarnih struktura koji ne spadaju u prvih šest

Stride prima kao ulaz jednu PDB datoteku, i set parametara opisani u tablici 3-1.

Tablica 3-1. Opis parametara Stride algoritma

Parametar	Opis parametra
-flme_datoteke	Ime izlazne datoteke. Ako ovaj parametar ne postoji, izlazni podaci se ispisuju na standardni izlaz.
-h	Uključuje podatke o vodikovim vezama u izlaznu datoteku.
-o	Ispisuje samo sažeti pregled sekundarnih struktura.
-rl1ld2..	Uzima u obzir samo lance ld1, ld2.. iz PDB datoteke
-cld1ld2..	Ispisuje sekundarne strukture samo za lance ld1, ld2 .., ali drugi lanci se i dalje čitaju i uzimaju u obzir pri izračunavanju sekundarnih struktura nabrojanih lanaca
-mlme_datoteke	Generira izlaznu datoteku tipa Molscript
-g[lme_datoteke]	Generira izlazne podatke u FASTA formatu. Ako je navedeno ime datoteke se ispisuju u nju, a inače na standardni izlaz

3.2 Energija otapala



Slika 3-2. Ion natrija otapan u molekulama vode

Energija otapala ili solvatacija je proces privlačenja i pridruživanja molekula otapala sa molekulama ili ionima otopine. Otapalo je najčešće voda (slika 3-2).

Energija otapala se računa samo za aminokiseline i baze, ne i hetatome. U tablici 3-2 prikazana su imena amino-kiselina, potencijal otapanja i totalna ASA.

Ukupna energija otapala, ΔG_s , računa se po formuli:

$$\Delta G_s = \frac{ASA_{sc}}{ASA_{TOT}} \times p$$

pri čemu je ASA_{TOT} ukupna ASA, i njene vrijednosti se uzimaju iz tablice 3-2. ASA_{sc} je ASA, tj. područje površine dostupno otapalu, bočnog lanca molekule bez atoma vodika, a p potencijal otapala čija se vrijednost također uzima iz tablice 3-2. ASA bočnog lanca se izračunava kao suma ASA svih atoma koji grade taj residuum.

Tablica 3-2. Tablica vrijednosti potencijala otapanja, i totalne ASA za pojedine amino-kiseline

Ime AA	Potencijal otapanja (p)	Totalna ASA
ALA	0,42	80
ARG	-1,37	216
ASN	-0,79	131
ASP	-2,46	123,9
CYS	1,39	108,4
GLN	-0,3	159,4
GLU	-2,35	152
GLY	0	43,3
HID	0,18	164,3
ILE	2,46	158,1
LEU	2,3	159,3
LYS	-1,35	188,3
MET	1,68	166,9
PHE	2,44	187
PRO	0,67	119,8
SER	-0,05	93,2
THR	0,35	119,9
TRP	3,07	228,5
TYR	1,31	203,1
VAL	1,66	133,1
HIS	0,18	164,3

3.3 PIADA 2

Sposobnost vezanja molekula polipeptida (nizova amino-kiselina ili nukleinskih kiselina) na druge molekule, te načina na koji oni međusobno interagiraju jako je važno za razumjevanje prirode peptida i načina na koji djeluju. Te veze omogućavaju proteinima da djeluju kao antitijela koja se vežu na viruse ili bakterije, te kao katalizatori, signalni receptori, motori, prekidači, itd. Tvar koja se veže za protein naziva se ligandom, neovisno da li je to ion ili mala molekula.

Pronalaženje mjesta proteinskih interakcija pronalazi široku primjenu na područjima biologije, kemije i biokemije, a najveću važnost ima u farmaceutskoj industriji pri razvoju lijekova.

3.3.1 Vrste veza između ostataka

Nekovalentne veze između residuuma puno (30 do 300 puta) su slabije od kovalentnih veza koje tipično tvore biološke molekule. Unatoč njihovoj slabosti veliki broj istodobno stvorenih nekovalentnih veza mogu čvrsto držati različite polipeptidne lance spojenima.

Vrste nekovalentnih veza koje se mogu formirati između ostataka su:

- Ionske veze
- Vodikove veze
- Van der Waalsove veze
- Hidrofobne veze

3.3.1.1 Van der Waalsove veze

Van der Waalsove veze nastaju uslijed fluktuacija elektronskog oblaka, i mogu nastati između bilo kojeg para residuuma, stoga je definirano da su dva residuuma povezana Van der Waalsovom vezom ako postoji barem jedan par atoma čije se jezgre nalaze na udaljenosti d manjoj od zbroja njihovih Van der Waalsovih radijusa i vrijednosti u iznosu $1,125 \text{ \AA}$.

$$d < R1 + R2 + 1,125$$

Vrijednost od $1,125 \text{ \AA}$ dobivena je aproksimacijom uzimajući u obzir moguće fluktuacije elektronskih oblaka i empirijskim putem uspoređujući dobivene vrijednosti s onima iz *i*PFAM baze. *i*PFAM je baza podataka koja sadrži domensko-domenske interakcije pronađene u zapisima PDB baze podataka.

3.3.1.2 Ionske veze

Ionske veze nastaju između suprotno nabijenih atoma zbog Coulombovih elektrostatskih privlačnih sila. Ako su polarne molekule vode nakupljene oko nabijenih iona, zbog velike dielektrične konstante vode dolazi do iznimnog smanjenja njihove međusobne privlačnosti. Snaga, odnosno sila F , ionskih veza između

polarnih atoma ovisi o udaljenosti njihovih jezgara d (snaga opada sa kvadratom udaljenosti), jačini naboja Q i dielektričnoj konstanti ϵ otapala u kojem se nalaze, te se može izraziti idućom formulom:

$$F = \frac{Q_1 \cdot Q_2}{\epsilon \cdot d^2}$$

3.3.1.3 Vodikove veze

Vodikove veze su specijalni oblik polarnih interakcija pri kojima dva elektro-negativna atoma međusobno dijele jedan elektro-pozitivan atom vodika. Taj vodik se može promatrati kao proton kojeg proton donor dijeli sa akceptorom. Za razliku od tipičnih elektrostatskih veza, vodikove veze jako su usmjerene i najjače su kada postoji ravna linija između sva tri atoma koji sudjeluju u interakciji. Prisutnost vode slabi i snagu ovih veza stvarajući vodikove veze sa molekulama koje okružuje.

3.3.1.4 Hidrofobne veze

Voda koja okružuje molekulu tjera grupe nepolarnih aminokiselina što dalje od površine molekule. Između tako grupiranih nepolarnih aminokiselina kaže se da postoje hidrofobne veze, iako je njihova privlačnost u biti stvorena zbog njihove odbojnosti prema vode.

3.3.2 Princip rada PSAIA 2 algoritma

U prvoj verziji PSAIA-e jedini dozvoljeni tip veza je veza između dva proteina. U novoj verziji programa su dozvoljeni novi tipovi interakcija:

- Veza unutar istog proteina
- Veza između proteina i nukleinske kiseline
- Veza između proteina i metala
- Veza između dvije nukleinske kiseline
- Veza između nukleinske kiseline i metala

Određeni atomi pojedinih residuuma (ostataka) proteina, nukleinskih kiselina, metala i liganada se smještaju u grupe ovisno o naboju atoma (+, -, +-, hidrofobni). Potom se te grupe koriste za određivanje postojanja i tipa interakcije ako je njihova udaljenost unutar granica određena interakcijom, te faktor udaljenosti i tip interakcije.

PIADA definira šest vrsta interakcija:

- Van der Waalsove veze
- Ionske (polarne) veze
- Hidrofobne veze
- Cistinski most
- Nepovoljne veze
- Sudar (clash)

Tablica 3-3.1 Tablica grupa ostataka i pripadnih atoma

Grupa atoma, i tip grupe	Pripadnost	Ime molekule	Tip Atoma
Grupa 11 - polar +	AA	Asn	N(H2)
	AA	Gln	N(H2)
	AA	Trp	N(H)
	AA	Bilo koja AA	N
Grupa 12 - polar +	AA	Lys	N(NH3+)
	AA	Arg	N(NH2+)
	AA	terminalna AA	N(NH3+)
	metal		K
Grupa 13 - polar +	metal		Na
	NA	ADE	N6(H2)
	NA	GUA	N2(H2)
	NA	GUA	N1(H)
	NA	THY	N3(H)
	NA	CYT	N4(H2)
Grupa 14 - polar +	NA	URA	N3(H)
	metal		Cu
	metal		Zn
	metal		Mg
	metal		Mn
	metal		Fe
	metal		Co
	metal		Ni
	metal		Cd
	metal		Ca
Grupa 21 - polar -	metal		Cr
	AA	Asn	O (C=O)
	AA	Gln	O (C=O)
	AA	Bilo koja AA	O
	NA	ADE	N1
	NA	ADE	N3
	NA	ADE	N7
	NA	GUA	N3
	NA	GUA	N7
	NA	GUA	O6
	NA	THY	O2
	NA	THY	O4
	NA	CYT	O2
	NA	CYT	N3
	NA	URA	O2
NA	URA	O4	
Grupa 22 polar -	AA	Asp	Asp O(COO-)

	AA	Glu	O(COO-)
	AA	Terminalna AA	O (COO-)
	NA	Bilo koja NA	O1P
	NA	Bilo koja NA	O2P
Grupa 23 - polar -	NA	Bilo koja NA	O4'
	NA	Bilo koja NA	O3'
	NA	Bilo koja NA	O5'
Grupa 31 - polar + -	AA	His	N(H) *
	AA	Ser	O(H) *
	AA	Thr	O(H) *
	AA	Tyr	O(H) *
	NA	Bilokoja NA	O2'(H)
	AA	Cys	S(H) *
Grupa 41 – hidrofobna	AA	Bilo koja AA	Bilo koji C atom
	NA ligand	Bilo koja NA ligand	Bilo koji C atom
Grupa 51 - cistein	AA	Cys	S

Osim slabih nekovalentnih veza, makromolekule se mogu vezati i rjeđim kovalentnim vezama npr. disulfidni mostovi između atoma sumpora dva cisteina. PIADA 2 uračunava i slučaj cistinog mosta, kako je on uvijet za postojanje par kovalentnih veza.

Nepovoljne veze su sve veze koje negativno utječu na postojanje i smanjuju snagu nekovalentnih veza između dva ostatka.

Uvijet za postojanje veze dviju molekule unutar istog proteina je određena minimalnom udaljenošću tih ostataka koju postavlja korisnik programa.

Tablica 3-2.2 Tablica interakcija

Redni broj	Tip interakcije	Partner 1	Partner 2	Udaljenost [Å]	Faktor udaljenosti	Uvijet
#Interakcija 1	Van der Waalsova	Bilo koji atom	Bilo koji atom	4	$1/r^2$	
#Interakcija 10	ionska	Grupa 11	Grupa 21	4	$1.5/r$	
#Interakcija 11	ionska	Grupa 11	Grupa 22	6	$2/r$	
#Interakcija 9	ionska	Grupa 11	Grupa 23	4	$1.5/r$	
#Interakcija 13	ionska	Grupa 12	Grupa 21	5	$2/r$	
#Interakcija 14	ionska	Grupa 12	Grupa 22	7	$3/r$	

#Interakcija 12	lonska	Grupa 12	Grupa 23	5	2/r	
#Interakcija 24	lonska	Grupa 13	Grupa 21	5	1.5/r	
#Interakcija 25	lonska	Grupa 13	Grupa 22	6	2/r	
#Interakcija 23	lonska	Grupa 13	Grupa 23	5	1.5/r	
#Interakcija 16	lonska	Grupa 14	Grupa 21	6	3/r	
#Interakcija 17	lonska	Grupa 14	Grupa 22	9	4/r	
#Interakcija 15	lonska	Grupa 14	Grupa 23	6	3/r	
#Interakcija 8	lonska	Grupa 31	Grupa 12	4	1/r	
#Interakcija 19	lonska	Grupa 31	Grupa 14	4	2/r	
#Interakcija 20	lonska	Grupa 31	Grupa 21	4	1/r	
#Interakcija 21	lonska	Grupa 31	Grupa 22	4	2/r	
#Interakcija 18	lonska	Grupa 31	Grupa 23	4	1/r	
#Interakcija 22	lonska	Grupa 31	Grupa 31	4	1/r	Ako ne postoji cistinski most
#Interakcija 26	Nepovoljna	Grupa 11	Grupa 11	4	-1/r	
#Interakcija 27	Nepovoljna	Grupa 11	Grupa 12	4	-2/r	
#Interakcija 29	Nepovoljna	Grupa 11	Grupa 13	4	-1/r	
#Interakcija 28	Nepovoljna	Grupa 11	Grupa 14	4	-2/r	
#Interakcija 30	Nepovoljna	Grupa 12	Grupa 12	5	-3/r	
#Interakcija 32	Nepovoljna	Grupa 12	Grupa 13	4	-2/r	
#Interakcija 31	Nepovoljna	Grupa 12	Grupa 14	5	-3.5/r	
#Interakcija 34	Nepovoljna	Grupa 13	Grupa 14	4	-3/r	
#Interakcija 33	Nepovoljna	Grupa 14	Grupa 14	6	-3/r	
#Interakcija 36	Nepovoljna	Grupa 21	Grupa 21	4	-1/r	
#Interakcija 37	Nepovoljna	Grupa	Grupa	4	-2/r	

		21	22			
#Interakcija 35	Nepovoljna	Grupa 21	Grupa 23	4	-1/r	
#Interakcija 39	Nepovoljna	Grupa 22	Grupa 22	5	-2/r	
#Interakcija 38	Nepovoljna	Grupa 22	Grupa 23	5	-2/r	
#Interakcija 40	Hidrofobna	Grupa 41	Grupa 41	4,7	1/r	
#Interakcija 41	Cistinski most	Grupa 51	Grupa 51	2,5	4/r	
#Interakcija 42	Sudar	Bilo koja grupa	Bilo koja grupa	2 A	-1/r ²	Ako ne postoji cistinski most

4 Opis sustava

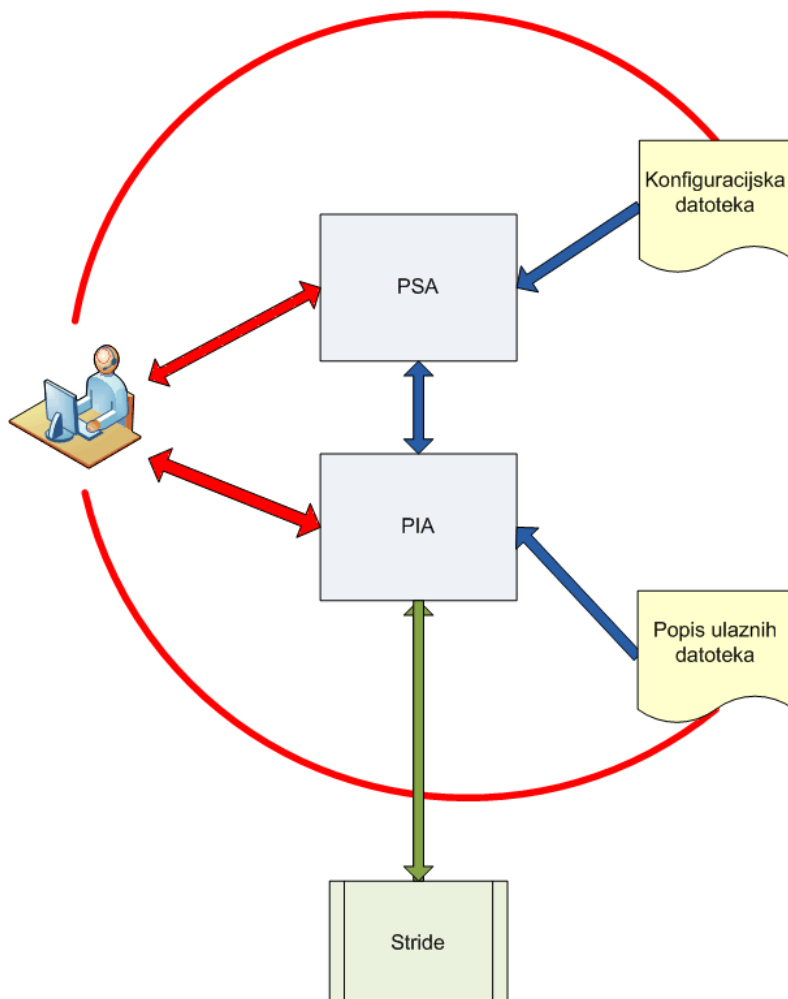
4.1 Opći opis PSAIA sustava

Protein Structure and Interaction Analyser, PSAIA, je alat razvijen za analizu geometrijskih struktura i mjesta proteinskih interakcija.

Protein Structure and Interaction Analyser, kako mu samo ime govori, se sastoji od dva dijela:

- Modul za analizu svojstava geometrijske strukture proteina (eng. *Protein Structure Analyser - PSA*)
- Modul za analizu proteinskih interakcija (eng. *Protein Interaction Analyser - PIA*)

Oba modula primaju kao ulaz jednu ili više ulaznih datoteka u PDB, PDBML ili mmCIF formatu, koje u sebi sadrži podatke o strukturama samostalnih proteina i peptida nastalih interakcijama raznih proteina. Kod izlaznih datoteka korisnik ima opciju izabrati zapis podataka u XML (*Extensible Markup Language*) formatu i/ili tabličnom zapisu.



Slika 4-1. Skica PSAIA sustava

Program je napisan u C++ programskom jeziku, što ga čini povoljnim za pokretanjem i na Microsoft Windows i Linux operativnim sustavima. Njegov multidretven dizajn omogućava uz ubrzanje rada programa, mogućnost izvođenja više analiza istovremeno i njihovo sigurno zaustavljanje u bilo kojem trenutku.

Program sadrži, uz konzolnu aplikaciju, i aplikaciju grafičkog sučelja napisanu u razvojnom alatu Qt, koji je također napisan u C++ programskom jeziku. Nakon instalacije u instalacijskom direktoriju će se nalaziti tri pokretačke datoteke: *psaia.exe*, *psa.exe* i *pia.exe*. Od toga *psaia.exe* je GUI aplikacija, a *psa.exe* i *pia.exe* konzolne aplikacije istoimenih modula.

Pri pokretanje jedne od konzolnih aplikacija potrebna su dva argumenta – put do, i ime datoteke sa postavkama, i put do, i ime datoteke sa popisom ulaznih datoteka.

4.2 Modul za analizu površina proteina, PSA

Modul za analizu geometrijske strukture, PSA, koristi se za analizu jednog ili više geometrijskih svojstava polipeptida, pri čemu se analiza može provesti na cijeloj strukturi ili zasebno za svaki lanac unutar strukture.

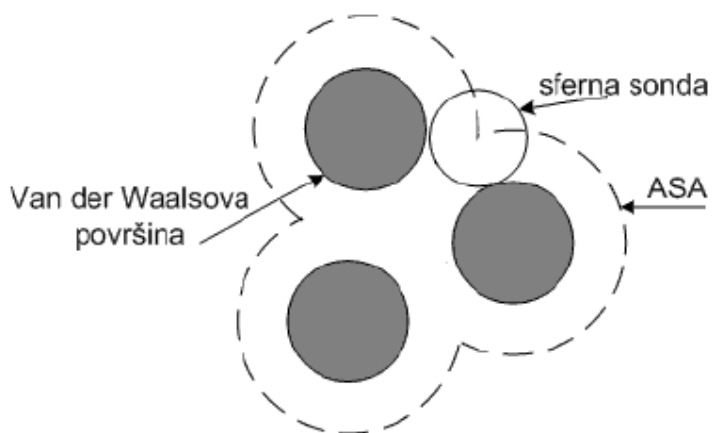
Svojstva proteina za analizu koje korisnik može odabrati su:

- Dostupna područja površine otapalu (eng. *Accessible solvent area, ASA*)
- Relativno dostupno područje površine otapalu (eng. *Relative ASA*)
- Ukopanost atoma i residuuma (eng. *Depth index*)
- Izbočenost atoma i residuuma (eng. *Protrusion index*)
- Hidrofobnost (eng. *Hyrophobicity*)
- Energija otapala (eng. *Solvatation*)
- Sekundarne strukture (eng. *Secondary structure*)

4.2.1 Otapalo dostupno područje površine, ASA

Područje površine dostupno otapalu (engl. *accessible solvent area, ASA*) opisuje područje površine proteina na kojem može doći do dodira između proteina i otapala u kojem se nalazi. Otapalo je najčešće voda.

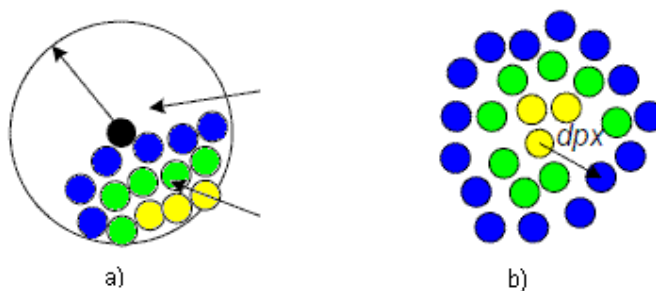
Za izračun ASA, koristi se algoritam koji su predložili Lee i Richards. Otapalu dostupna površina definirana je kao pozicija centra sferne sonde, koja predstavlja molekulu otapala, dok se ona pomiče po Van der Waalsovoj površini proteina.



Slika 4-2-1. Otapalu dostupno područje površine

4.2.2 Indeks izbočenosti atoma

Indeks izbočenosti atoma i ostataka daje nam informaciju o sveukupnim izbočenim ili jako konveksnim regija peptida. Indeks izbočenosti za svaki atom se dobiva na temelju omjera volumena što ga protein zauzima u okolici atoma i slobodnom volumenu oko istog atoma.



Slika 4-2-2. Računanje a) indeksa izbočenosti atoma; b) indeksa ukopanosti atoma.

4.2.3 Indeks ukopanosti atoma

Ukopenost atoma i aminokiselina u proteinu geometrijski je parametar koji daje uvid u unutrašnjost proteina, i definira se kao udaljenost atoma do najbližeg atoma na površini proteina. Ukopenost nam pruža mogućnost razlikovanja ostataka koji su ukopani, ali se nalaze blizu površine proteina, od onih koje su ukopani duboko u proteinu.

4.2.4 Hidrofobnost

Hidrofobnost nastaje kao rezultat razlika u međumolekularnim silama između vode i ostataka, i između ostataka i nekog drugoga medija.

4.2.5 Programska implementacija razreda modula PSA

4.2.5.1 Opis glavnih razreda

4.2.5.1.1 Razred *cPSAProgramSettings*

Pri pokretanju PSA analize, bilo to preko konzolne linije ili pritiskom gumba za pokretanje analize GUI aplikacije, prvo se u varijablu razreda *cPSAProgramSettings* učitavaju sve postavke analize ili sa ekrana za GUI, ili iz zadane datoteke sa postavkama u XML formatu. Taj razred sadrži podatke koji atributi će se izračunavati, parametre tih atributa za analizu, listu ulaznih datoteka, imena datoteka sa podacima o Van der Waalsovim radijusima, vrijednostima hidrofobnosti, standardnih ASA, parametara za izračun solvatacije; da li će se analiza provoditi za cijele lance ili odvojeno po ostacima, parametrima izlaznih datoteka: da li će se ispisivati u XML ili tabličnom formatu, ili oboje, te da li će se ispisivati vrijednosti ASA i pojedinačnih atoma i hetatoma, i nakraju da li će se i gdje bilježiti dnevnik statusa programa.

4.2.5.1.2 Razred *cPdbPeptide*

Potom se za svaku ulaznu datoteku formata PDB, XMLBL ili mmCIF, učitavaju podaci o polipeptidu u varijablu razreda *cPdbPeptide*, koja uz samu strukturu peptida podijeljenu po lancima, ostacima, atomima i hetatomima, sadrži funkcije za postavljanje vrijednosti ASA i relativnih ASA peptida, njegovih lanaca, ostatak i atoma; indeksa ukopanosti i izbočenosti, energije otapala atoma i ostataka, i sekundarnih struktura atoma. Ovisno o odabranim atributima za analizu se pokreću određene funkcije za izračunavanje vrijednosti tih atributa iz razreda *cPdbPeptide*.

4.2.5.1.3 Razred *strideInterface*

Za pronalaženje vrijednosti sekundarne strukture koristimo eksterni program Stride. Stride prima kao ulaz samo datoteke formata PDB, stoga smo morali naše peptidne strukture prilagoditi formatu koji Stride zna čitati, i nakon analize učitati Stride izlaznu datoteku, pročitati ju, i dodijeliti atomima njihove odgovarajuće vrijednosti sekundarnih struktura. Za konverziju podataka, pokretanje Stride algoritma i čitanje Strideovih izlaznih rezultata brine se razred *strideInterface*.

4.2.5.1.4 Razred *cXmlWriter*

Razred *cXmlWriter* se brine za otvaranje i prilagođavanje podataka za upis u XML formatu u izlaznu datoteku. Sve vrijednosti decimalnih brojeva su ispisivane na dvije decimale, a za slučaj nepostojanja vrijednosti koja se ispisuje, ispisuje se nepoznata vrijednost, NA (u sustavu joj je pridružena brojčana vrijednost -99.9).

Dijagram toka PSA modula se nalazi u dodatku A, dok je slika 4-2-5-2 prikaz implementacije opisanih razreda po varijablama i funkcijama.

```

--strideInterfaceIn string pdb_filename cPdbPeptide & mSecondaryStructureMap &secondary_structure_parameters)
--strideInterfaceIn cPdbPeptide & mSecondaryStructureMap &secondary_structure_parameters)
--strideInterface()
WriteTopPdb(In stream &tmp, cPdbPeptide & p, int
addSecondaryStructures(In string stride_filename, cPdbPeptide & p, mSecondaryStructureMap &secondary_structure_parameters) : int

```

```

cPdbPeptide
--pdb_id : string
--max, min : sPoint
--atoms, helatons, metal_atoms : vector<ResidueAtom>
--residues, helatm_residues : vector<ChainResidue>
--chains, helatm_chains : vector<ProteinChain>
--contacts : vector<ResidueContact>
--asa : sAsa

```

```

cPdbPeptide()
cPdbPeptide(In const cPdbPeptide &source)
+loadFromPdbStream(In Istream &pdb_stream, mVdwMap radii, set<string> &missingRadii, string flags) : unsigned int
+loadFromXmiStream(In Istream &pdb_stream, mVdwMap radii, set<string> &missingRadii, string flags) : unsigned int
+loadFromMmcStream(In Istream &pdb_stream, mVdwMap radii, set<string> &missingRadii, string flags) : unsigned int
+recommendReadyValue(In string atom_name) : double
+getPdbId() : string
+getAtomCount() : unsigned int
+getResidueAtom() : cResidueAtom
+getMetalAtomCount() : unsigned int
+getMetalAtom(In unsigned int index) : cResidueAtom
+getHeliumCount() : unsigned int
+getHelium(In unsigned int index) : cResidueAtom
+getChainCount() : unsigned int
+getChain(In unsigned int index) : cProteinChain
+getResidueCount() : unsigned int
+getResidue(In unsigned int index) : cChainResidue
+getHeliumChainCount() : unsigned int
+getHeliumChain(In unsigned int index) : cProteinChain
+getHeliumResidueCount() : unsigned int
+getHeliumResidue(In unsigned int index) : cChainResidue
+getResidueInteractionCount() : unsigned int
+getResidueInteraction(In unsigned int index) : cResidueContact
+getAsa() : sAsa
+getHydrophobicity(In mHydrophobicityMap hydro_table) : void
+getSolvation(In mSolvationMap solvation_table) : void
+getAccessibleSurfaceArea(In double r_solvant, double z_slice) : void
+getAccessibleSurfaceAreaUnbound(In double r_solvant, double z_slice) : void
+getRelativeAccessibleSurfaceArea(In mAsaMap &standart_asa) : void
+getProtrusionIndex(In double threshold, double atom_volume) : void
+getDepthIndex() : void
+getSecondaryStructure(In mSecondaryStructureMap secondary_structure_table) : void
+getClearContacts() : void
+resolvePRISMContacts(In double threshold, double intramolecular_interactions_threshold) : void
+resolveMMSAMMContacts(In double threshold, double intramolecular_interactions_threshold) : void
+resolveMMSAMMThromionContacts(In double threshold, double intramolecular_interactions_threshold) : void
+resolveRudertSSContacts(In ...) : void
+includeHelatons() : void
+separateHelatons() : void
+setHeliumAccessibleSurfaceArea(In double r_solvant, double z_slice) : void
+assignMetalToChain(In map<InteractionType &contacts_map, map<InteractionGroup &groups_map) : void

```

```

cXmiWriter
--xml_file : ostream
+XmiWriter(In const char *filename)
+cXmiWriter()
+close() : void
+isFileOpen() : bool
+writeStart(In string name, In int indent) : void
+writeElement(In string name, In value_type value, In int indent) : void
+writeEnd(In string name, In int indent, In bool endlLine) : void
+checkValue(In double value) : string
+checkValue(In string value) : string

```

```

PSA
modul

```

```

cPSAProgramSettings
+z_slice, r_solvant : double
+neighbourhood_area, v_atom : double
+asa_on, rasa_on, dpz_on, cx_on, hydro_on, solvation_on : int
+secondary_structure_on : int
+write_asa : int
+write_xml, write_table : int
+analyze_bound, analyze_unbound : int
+output_directory_name : string
+radii_filename, hydrophobicity_filename, standrt_asa_filename : string
+solvation_filename, log_filename, interactions_filename : string
+input_filenames : vector<String>
+write_log, write_standard_output, confirmation_dialog : int
+PSAProgramSettings()
+PSAProgramSettings(In const cPSAProgramSettings &source)
+cPSAProgramSettings& operator=(In const cPSAProgramSettings &source)
+newLoadSettings(In Istream &settings_stream) : int
+loadInpufFiles(In Istream &settings_stream) : int
+writePaasSettings(In ostream &stream) : void
+writeLogEntry(In ostream &log, In string entry) : void
+writeLogFilesStart(In ostream &stream) : void

```

Slika 4-2-5-2. Implementacija glavnih razreda modula PSA

4.3 Modul za analizu proteinskih interakcija, PIA

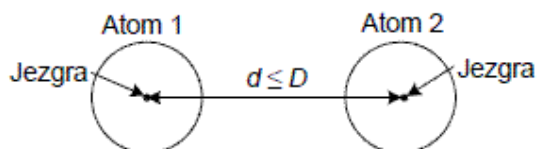
Modul za analizu proteinskih interakcija, PIA, koristi se za pronalaženje interakcijskih parova između residuuma (ostataka) peptidnih lanaca i to, dvije molekule unutar istog lanca, dvaju različitih proteinskih lanaca, proteina i nukleinske kiseline, proteina i metala, dvije nukleinske kiseline, i između nukleinske kiseline i metala.

Korisnik može odabrati jedan od četiri kriterija definiranja interakcijskih parova i mjesta interakcija:

- Interakcije na osnovu udaljenost jezgri atoma (Maximum distance)
- Interakcije na osnovu udaljenosti Van der Waalsovih radijusa (Van der Waals distance)
- LSS&Ruđer algoritam (PIADA) – princip rada algoritma je opisan u poglavlju 3.3
- Mjesta interakcije na osnovu promjene otapalu dostupnog područja površine (ASA change)

4.3.1 Interakcije na osnovu udaljenost jezgri atoma

Interakcijski parovi između proteinskih lanaca pronalaze se na osnovu udaljenosti između jezgri atoma interagirajućih residuuma. Koordinate jezgri svakoga atoma nalaze se u PDB datotekama. Ako postoji barem jedan par atoma, od kojih svaki pripada jednoj aminokiselini različitih lanaca, udaljen manje od neke određene vrijednosti D smatra se da su ta dva residuuma vezana.



Slika 4-3-1. Definiranje veze na temelju udaljenosti jezgri atoma

4.3.2 Interakcije na osnovu udaljenosti Van der Waalsovih radijusa

Interakcije na osnovu udaljenosti Van der Waalsovih radijusa između dva residuuma postoji ako postoji barem jedan par atom čiji su Van der Waalsovi radijusi udaljeni manje od neke određene vrijednosti D .



Slika 4-3-2. Definiranje veze na temelju udaljenosti Van der Waalsovih radijusa atoma

4.3.3 Mjesta proteinskih interakcija na osnovu promjene ASA

Prilikom interakcije proteinskih lanaca dolazi do kontakta dijelova njihovih površina koji tada postaju nedostupni otapalu. Računanjem promijene otapalu dostupnog područja površine residuuma prije i poslije stvaranja kompleksa, može se zaključiti koji residuumi sudjeluju u interakciji.

Smatra se da se residuum nalazi u regiji kojom proteinski lanac interagira sa drugim lancem, ako se njegovo otapalu dostupno područje površine promijeni za više od neke određene vrijednosti A .

$$\Delta ASA = ASA_{poslije} - ASA_{prije} \geq A$$

4.3.4 Programska implementacija razreda modula PIA

4.3.3.1 Opis glavnih razreda

4.3.3.1.1 Razred *cPIAProgramSettings*

Pri pokretanju PIA analize, bilo to preko konzolne linije ili pritiskom gumba za pokretanje analize GUI aplikacije, slično kao i kod PSA modula, prvo se u varijablu razreda *cPIAProgramSettings* učitavaju sve postavke analize ili sa ekrana za GUI, ili iz zadane datoteke sa postavkama u XML formatu. Taj razred sadrži podatke o kriterijima pokrenute analize, listu ulaznih datoteka, imena datoteka sa podacima o Van der Waalsovima radijusima i datoteke sa definicijama grupa atoma i interakcija, parametrima izlaznih datoteka: da li će se ispisivati u XML ili tabličnom formatu, ili oboje, te da li će se ispisivati po interakcijskim vezama, svim ostacima ili samo ostacima koji čine neku vezu, i nakraju da li će se i gdje bilježiti dnevnik statusa programa.

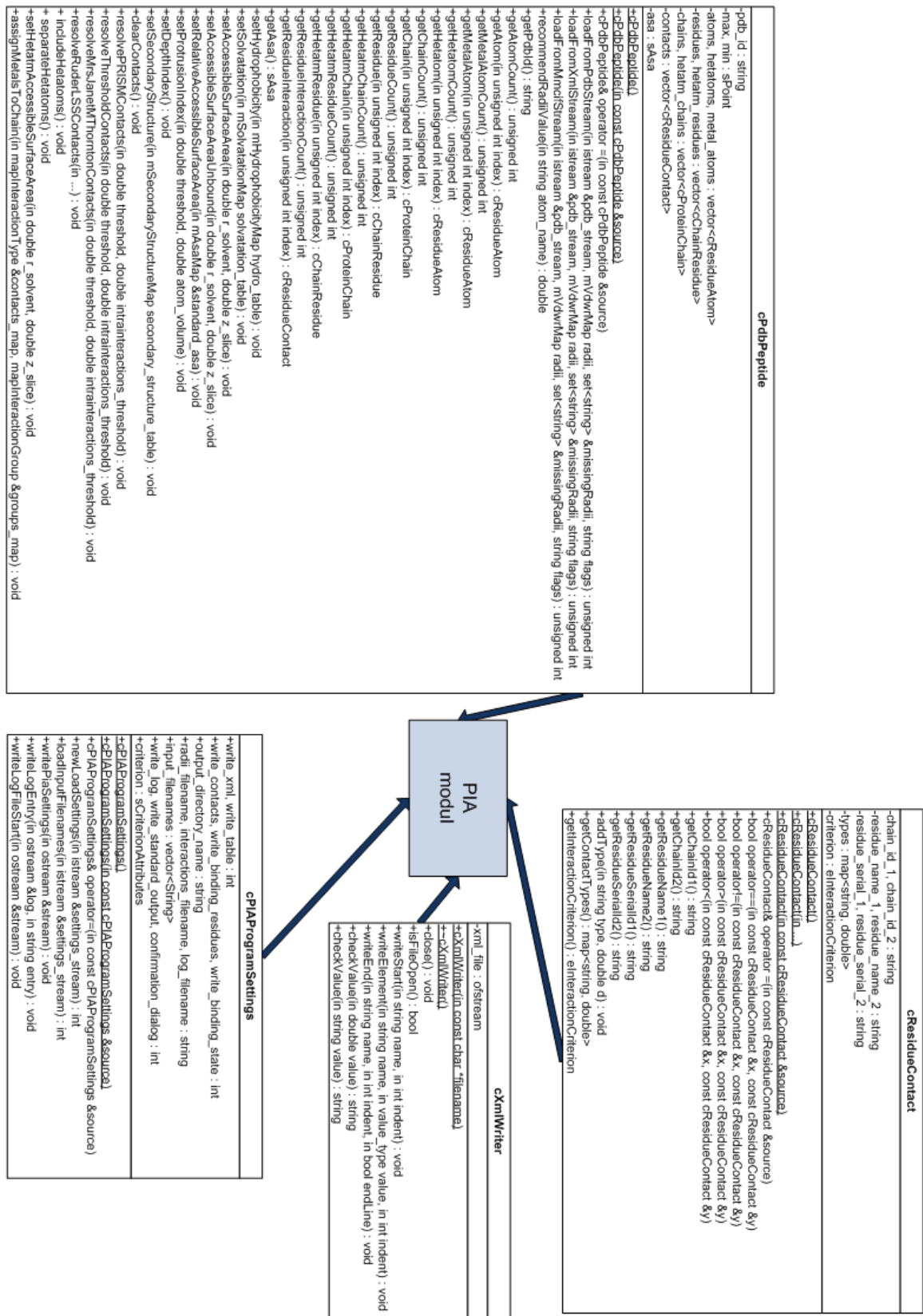
4.3.3.1.2 Razred *cPdbPeptide*

Potom se za svaku ulaznu datoteku formata PDB, XMLBL ili mmCIF, učitavaju podaci o polipeptidu u varijablu razreda *cPdbPeptide*, koja uz samu strukturu peptida podijeljenu po lancima, ostacima, atomima i hetatomima, sadrži funkcije za svaki od kriterija predviđanja peptidnih interakcija (na osnovu udaljenost jezgri atoma, Van der Waalsovih radijusa, promjene ASA i PIADA). Ovisno o odabranim postavkama analize se pokreću određene funkcije iz razreda *cPdbPeptide*.

4.3.3.1.5 Razred *cResidueContact*

Ovaj razred predstavlja parametre i funkcije interakcije između dva ostatka. Sadrži imena i serijske brojeve ostataka koji sudjeluju u toj interakciji, oznake njihovih lanaca, tipove interakcija i njihove faktore udaljenosti, i na kraju kriterij po kojem su oni pronađeni.

Dijagram toka PIA modula se nalazi u dodatku B, dok je slika 4-3-4-2 prikaz implementacije opisanih razreda po varijablama i funkcijama.



Slika 4-3-4-2. Implementacija glavnih razreda modula PIA

5 Korištenje sustava

Pri korištenju PSAIA programa, korisniku je dostupno mijenjanje i postavljanje raznih parametara vezanima uz vrijednosti parametara koje koriste razni algoritmi, parametre vezane uz sami rad algoritama i programa općenito, kao i parametre vezane uz dnevnik rada programa i izlaznih datoteka.

Uz to, vrijednosti Van der Waalsovih radijusa, standardnih ASA vrijednosti i hidrofobnosti ostataka, vrijednosti parametara za računanje energije otapala i definiciju grupa atoma i interakcija, su smješteni u odvojenim datotekama u XML formatu i također dostupni korisniku na pregled, nadopunu i, ako je potrebno, mijenjanje vrijednosti.

5.1 Konfiguracijska datoteka PSA modula

Pri pokretanju konzolne aplikacije učitavaju se podaci o analizi, parametri analize i izlaznih datoteka iz konfiguracijske datoteke. Ona započinje oznakom `<PSA:program_settings>`, i sastoji se od par dijelova: općih postavki programa, tip analize, atributi analize, postavke izlaznih datoteka i ulazne datoteke.

Primjer konfiguracijske datoteke PSA modula se nalazi u dodatku C.

Tablica 5-1. Imena i značenja oznaka za PSA konfiguracijsku datoteku

Vanjska oznaka	Imena oznaka	Vrijednost oznake
PSA:general	PSA:write_log	Ispis statusnih poruka u datoteku dnevnika programa.
	log_filename	Ime datoteke sa dnevnikom.
	PSA:write_standard_output	Ispisivanje statusnih poruka na standardni izlaz.
	confirmation_dialog	Ispisivanje potvrde podataka na ekran.
PSA:analysis_type	analyze_bound	Analiza cijelog lanca odjednom.
	analyze_unbound	Analiza svakog lanca odvojeno.
PSA:analysis_attributes	PSA:calc_asa	Računanje ASA.
	z_slice	Rezolucija podjele sfere na ravnine
	r_solvent	Radijus sferne sonde
	hetatm, water, hydrogen	Uzimanje hetatoma, voda i vodika u izračun ASA.
	PSA:calc_rasa PSA:calc_dpx	Računanje RASA. Računanje indeksa izbočenosti.

	PSA:calc_cx	Računanje indeksa ukopanosti.
	cx_threshold	Prag radijusa sfere.
	cx_volume	Volumen atoma.
	PSA:calc_hydro	Pridruživanje vrijednosti hidrofobnosti.
	PSA:calc_secondary_structure	Računanje sekundarnih struktura.
	PSA:calc_solvation	Računanje energije otapala.
PSA:output_files_options	output_dir	Direktorij u koji će se zapisivati izlazne datoteke.
	write_xml	Ispis izlaznih datoteka u XML formatu.
	write_table	Ispis izlaznih datoteka u tabličnom formatu.
	write_asa	Ispis ASA u izlazne datoteke.
	write_atoms	Ispis atoma u izlazne datoteke.
PSA:input_files_options	PSA:radii_file	Ime datoteke sa Van der Waalsovima radijusima.
	PSA:standard_asa_file	Ime datoteka sa standardnim ASA vrijednostima.
	PSA:hydro_file	Ime datoteke sa vrijednostima hidrofobnosti.
	PSA:solvation_parameters_file	Ime datoteke sa parametrima solvatacije.
	PIA:interactions_file	Ime datoteke sa grupama i tipovima interakcije.

5.2 Konfiguracijska datoteka PIA modula

Pri pokretanju konzolne aplikacije učitavaju se podaci o analizi, parametri analize i izlaznih datoteka iz konfiguracijske datoteke. Ona započinje oznakom *<PIA:program_settings>*, i sastoji se od par dijelova: općih postavki programa, kriterij pronalaženja interakcija, postavke izlaznih datoteka i ulazne datoteke.

Primjer konfiguracijske datoteke PIA modula se nalazi u dodatku D.

Tablica 5-1. Imena i značenja oznaka za PSA konfiguracijsku datoteku

Vanjska oznaka	Imena oznaka	Vrijednost oznake
PIA:general	PIA:write_log	Ispis statusnih poruka u datoteku dnevnika programa.
	log_filename	Ime datoteke sa dnevnikom.
	PIA:write_standard_output	Ispisivanje statusnih poruka na standardni izlaz.
	confirmation_dialog	Ispisivanje potvrde podataka na ekran.
PIA:contact_criterion	PIA:maximum_distance	Kriterij udaljenosti na temelju jezgri atoma
	PIA:van_der_waals_distance	Kriterij udaljenosti na temelju Van der Waalsovih radijusa
	threshold	Prag udaljenosti.
	intrainteractions_threshold	Minimalna udaljenost ostataka za računanje intrainterakcija.
	PIA:piada	Kriterij PIADA.
	PIA:asa_change	Kriterij na osnovi promjena ASA.
	delta_asa	Minimalna vrijednost promjene ASA vrijednosti.
	z_slice	Rezolucija podjele sfere na ravnine
PIA:output_files_options	r_solvent	Radijus sferne sonde
	output_dir	Direktorij u koji će se zapisivati izlazne datoteke.
	write_xml	Ispis izlaznih datoteka u XML formatu.
	write_table	Ispis izlaznih datoteka u tabličnom formatu.
	write_contacts	Ispis po interakcijama
	write_binding_residues	Ispis vezujućih ostataka.
	write_residue_binding_state	Ispis vezujućeg stanja svih ostataka.
PIA:input_files_options	PSA:radii_file	Ime datoteke sa Van der Waalsovih radijusima.
	PIA:interactions_file	Ime datoteke sa grupama i tipovima interakcije.

5.3 Datoteka koja sadrži popis ulaznih datoteka

Datoteka koja sadrži listu ulaznih PDB, XML i mmCIF datoteka ima standardnu ekstenziju .fls. Svaki njen redak sadrži točno jedan apsolutni put i ime ulazne pdb datoteke.

```
C:\Program Files\PSAIA-1.0\Examples\1lfd.pdb
C:\Program Files\PSAIA-1.0\Examples\4hhb.xml
C:\Program Files\PSAIA-1.0\Examples\1bkf.cif
```

Slika 5-3. Primjer dobro formatirane datoteke koja sadrži popis ulaznih datoteka

5.4 Datoteka koja sadrži vrijednosti Van der Waalsovih radijusa atoma

Datoteka sa vrijednostima Van der Waalsovih radijusa je u XML formatu i započinje sa oznakom `<PSAIA:radii>`. Svaki zapis u datoteci je oblika:

```
<radii name="imeOstatka_imeAtoma">vrijednostRadijusa</radii>
```

Ime zapisa se kreira pomoću imena ostatka i imena atoma ostatka odvojenim donjom crticom (_). Vrijednost Van der Waalsovih radijusa je pozitivan decimalan broj odvojen decimalnom točkom.

```
<PSAIA:radii>
  <radii name="ALA_N">1.65</radii>
  <radii name="ALA_CA">1.87</radii>
  <radii name="ALA_C">1.76</radii>
  <radii name="ALA_O">1.40</radii>
  <radii name="ALA_CB">1.87</radii>
  <radii name="ARG_N">1.65</radii>
  <radii name="ARG_CA">1.87</radii>
  <radii name="ARG_C">1.76</radii>
  ...
</PSAIA:radii>
```

Slika 5-4. Primjer dobro formatirane datoteke koja sadrži Van der Waalsove radijuse

Uz program PSAIA dostupna je datoteka Van der Waalsovih vrijednosti radijusa atoma prema Clothia.

5.5 Datoteka koja sadrži vrijednosti hidrofobnosti ostataka

Datoteka sa vrijednostima hidrofobnosti pojedinih ostataka je u također u XML formatu i započinje sa oznakom `<PSAIA:hydrophobicity>`. Svaki zapis u datoteci je oblika:

```
<hydro name="imeOstatka ">vrijednostHidrofobnosti</hydro>
```

Ime ostatka se obilježava troslovnom oznakom pisanim velikim slovima, dok vrijednosti hidrofobnosti su pozitivan ili negativan decimalan broj odvojen decimalnom točkom.

```

<PSAIA:hydrophobicity>
  <hydro name="ALA">1.800</hydro>
  <hydro name="ARG">-4.500</hydro>
  <hydro name="ASN">-3.500</hydro>
  <hydro name="ASP">-3.500</hydro>
  <hydro name="CYS">2.500</hydro>
  ..
</PSAIA:hydrophobicity>

```

Slika 5-5. Primjer dobro formatirane datoteke koja sadrži vrijednosti hidrofobnosti ostataka.

5.6 Datoteka koja sadrži vrijednosti standardnih ASA za ostatke

Standardne ASA vrijednosti ostataka su smještene u datoteku u XML formatu koja počinje oznakom `<PSAIA:standard_asa>`.

Tablica 5-6. Imena i značenja oznaka za datoteku sa vrijednostima standardnih ASA

Ime oznake	Vrijednost oznake
name	Ime ostatka
total	Totalna ASA
main_chain	ASA okosnice
side_chain	ASA bočnog lanca
non_polar	ASA nepolarnog dijela
polar	ASA polarnog dijela

Imena i značenja XML oznaka datoteke sa vrijednostima standardnih ASA se nalaze u tablici 5-6.

```

<PSAIA:standard_asa>
  <asa name="ALA">
    <total>107.24</total>
    <main_chain>43.32</main_chain>
    <side_chain>63.92</side_chain>
    <non_polar>76.06</non_polar>
    <polar>31.17</polar>
  </asa>
  <asa name="ARG">
    <total>233.01</total>
    <main_chain>36.86</main_chain>
    <side_chain>196.15</side_chain>
    <non_polar>86.30</non_polar>
    <polar>146.71</polar>
  </asa>

```

Slika 5-6. Primjer dobro formatirane datoteke koja sadrži vrijednosti standardnih ASA.

5.7 Datoteka koja sadrži parametre za izračun energije otapala

Datoteka sa vrijednostima parametara za izračun energije otapala je u XML formatu i započinje sa oznakom `<PSAIA:solvation_parameters>`. Svaki zapis u datoteci sadrži oznake opisane u tablici 5-7.

Tablica 5-7. Imena i značenja oznaka za datoteku sa parametrima energije otapala

Ime oznake	Vrijednosti oznake
name	Ime ostatka
p	Potencijal otapala
asa_total	Totalna ASA

```
<PSAIA:solvatation_parameters>
  <solvatation name="ALA">
    <p>0.42</p>
    <asa_total>80</asa_total>
  </solvatation>
  <solvatation name="ARG">
    <p>-1.37</p>
    <asa_total>216</asa_total>
  </solvatation>
  ..
</PSAIA:solvatation_parameters>
```

Slika 5-7. Primjer dobro formatirane datoteke koja sadrži parametre za izračun energije otapala

5.8 Datoteka koja sadrži definicije grupa atoma i tipove interakcija

Datoteka koja sadrži definicije grupa atoma i interakcija u XML formatu se sastoji od dva dijela: dijelu sa grupama atoma, i dijelu sa interakcijama. Datoteka započinje XML oznakom `<PIA:interaction_settings>`.

Dio sa grupama atoma započinje oznakom `<PIA:interaction_groups>`, i sadrži identifikacijski broj grupe, tip grupe i listu ostataka i pripadnih atoma (tablica 5-8-1).

```
<PIA:interaction_group id="11">
  <group_type>polar+</group_type>
  <residue>
    <name>ASN</name>
    <atom_type>ND2</atom_type>
  </residue>
  <residue>
    <name>GLN</name>
    <atom_type>NE2</atom_type>
  </residue>
  <residue>
    <name>TRP</name>
    <atom_type>NE1</atom_type>
  </residue>
  <residue>
    <name>any AA</name>
    <atom_type>N</atom_type>
  </residue>
</PIA:interaction_group>
```

Slika 5-8-1. Primjer grupe atoma; b) interakcije

Tablica 5-8-1. Imena i značenja oznaka za grupe atoma

Ime oznake	Vrijednosti oznake
id	Identifikacijski broj grupe
name	Ime ostatka
atom_type	Ime atoma ostatka

Tipovi interakcije se nalaze u drugom dijelu XML datoteke, i započinje oznakom `<PIA:interaction_type>`. Sadrži identifikacijski broj i tip interakcije, grupe atoma koji ju čine, te maksimalnu udaljenosti atoma da mogu činiti tu interakciju, i formulu za izračunavanje faktora udaljenosti.

Tablica 5-8-2. Imena i značenja oznaka za tipove interakcije

Ime oznake	Vrijednosti oznake
id	Identifikacijski broj tipa interakcije
group id="1"	Identifikacijski broj prve grupe atoma
group id="2"	Identifikacijski broj druge grupe atome
distance	Maksimalna udaljenost da bi interakcija postojala
distance_factor	Faktor udaljenosti

```

<PIA:interaction_type id="17">
  <type>polar</type>
  <group id="1">14</group>
  <group id="2">22</group>
  <distance>9</distance>
  <distance_factor>4/r</distance_factor>
</PIA:interaction_type>
<PIA:interaction_type id="15">
  <type>polar</type>
  <group id="1">14</group>
  <group id="2">23</group>
  <distance>6</distance>
  <distance_factor>3/r</distance_factor>
</PIA:interaction_type>

```

Slika 5-8-2. Primjer tipa interakcije

6 Zaključak

Razumijevanje struktura i interakcija bioloških molekula pomaže u razumijevanju rada svih organizama. To znanje nas omogućava, međuostalim, shvaćanje kako rade bolesti i pomaže u razvoju lijekova za izlječenje istih tih bolesti, i time doprinosi razvoju biokemije, medicine, i nakraju sveukupnog životnog standarda.

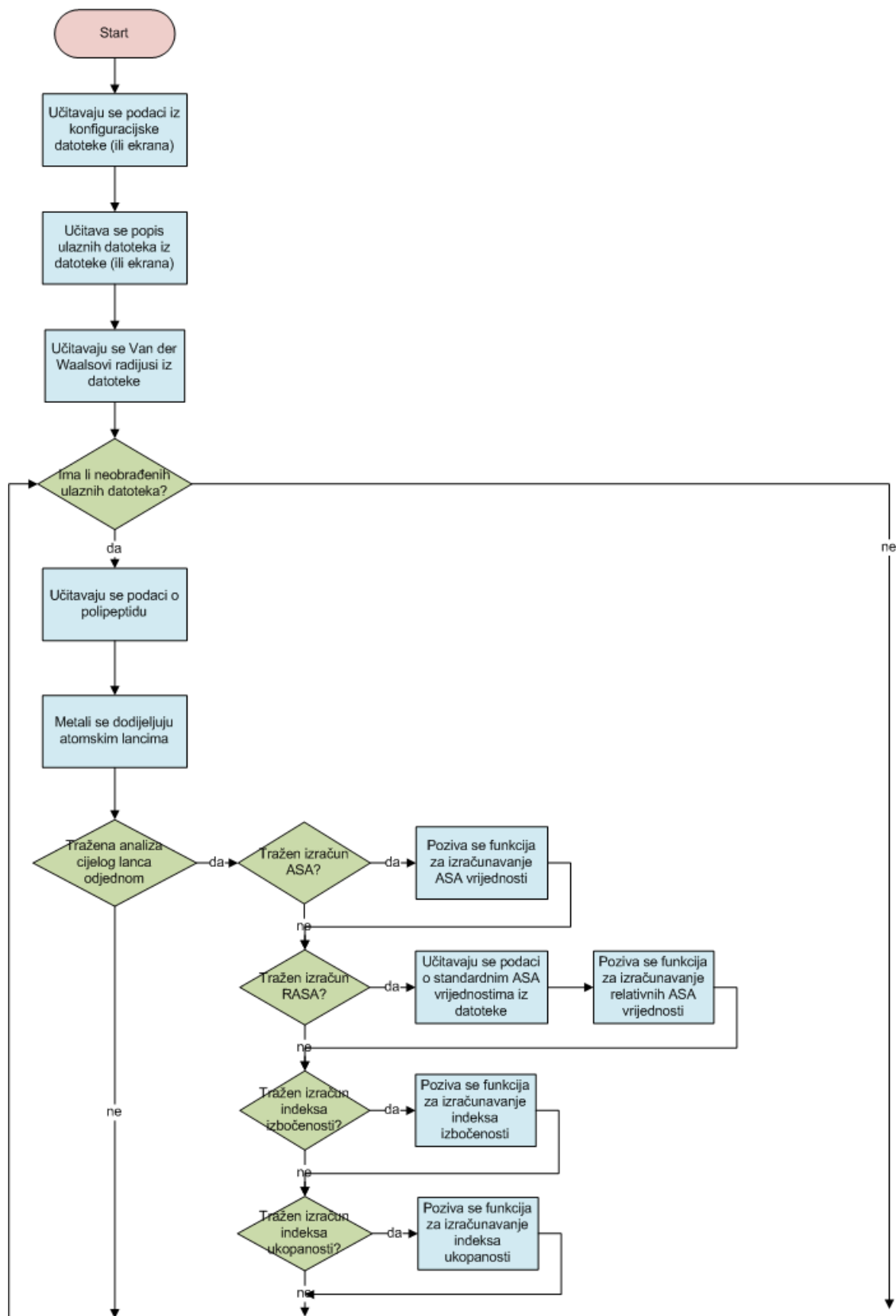
U sklopu ovog diplomskog rada i diplomskog rada Josipa Mihela razvijen je Napredni alat za analizu površina proteina i mjesta proteinskih interakcija.

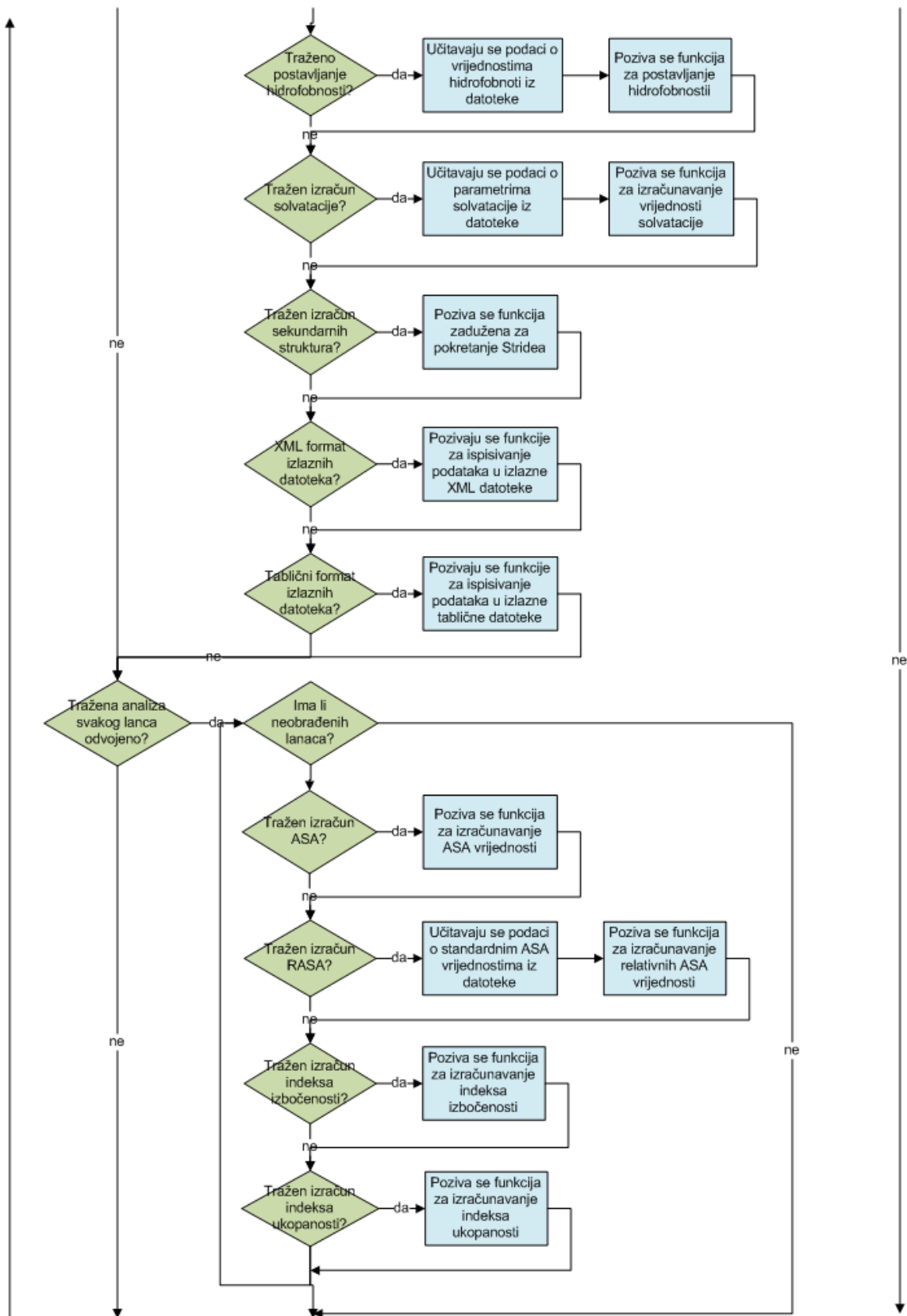
Alat prima ulazne podatke u PDB (*Protein Data Bank*), XML (*Extensible Markup Language*) ili mmCIF (*macromolecular Crystallographic Information File*) formatu.

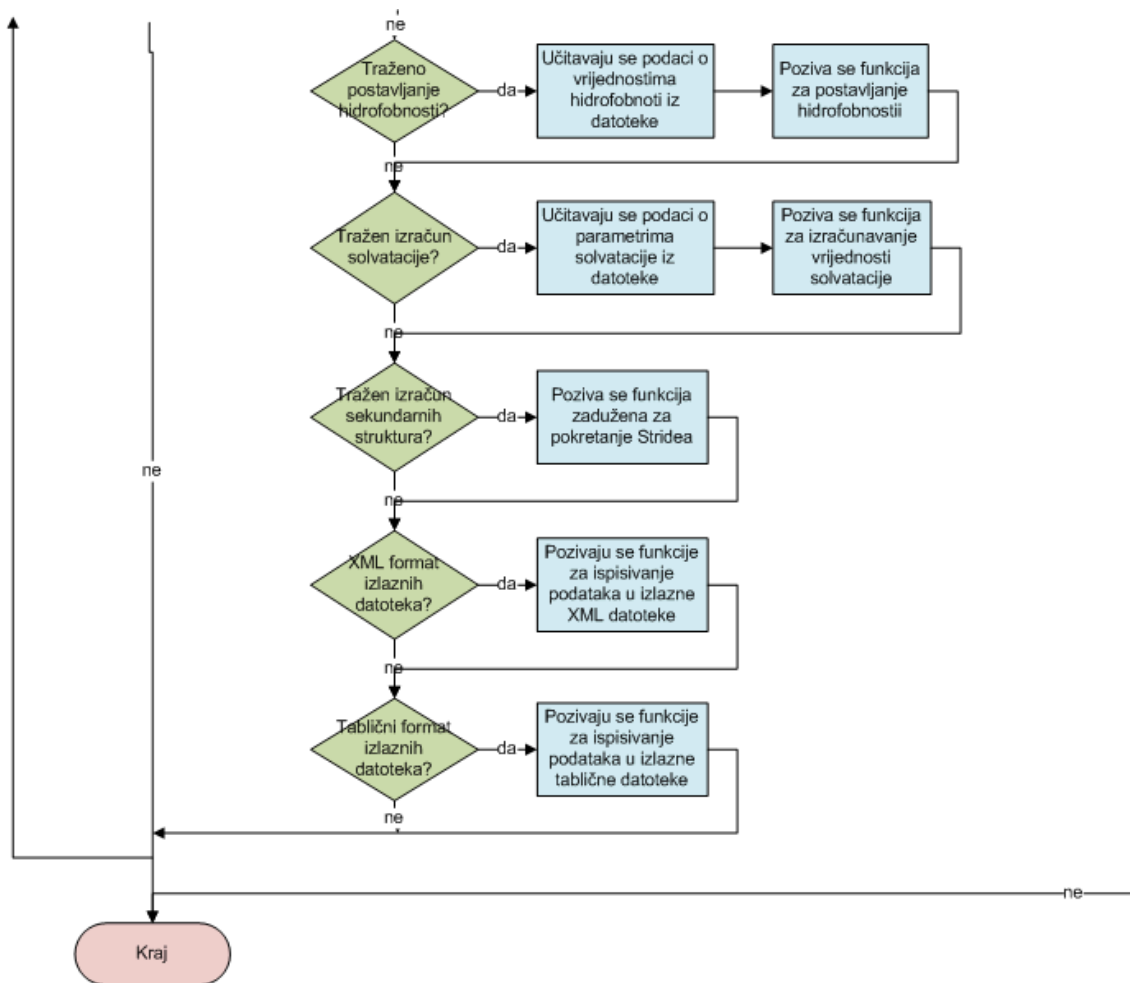
PSAIA alat uključuje mogućnost analize svih važnijih geometrijskih svojstava kao što su otapalu dostupno područje površine, ukopanost i izbočenost aminokiselinskih residuuma, energija otapala residuuma i sekundarna struktura atoma. Uključena je i mogućnost izbora više vrsta kriterija na temelju kojih se pronalaze interakcijski parovi residuuma i vezujući residuumi.

Objedinjavanje analize geometrijskih svojstava proteina i proteinskih interakcija u jedan softverski alat omogućava korisniku jednostavno pripremanje podataka za danju analizu. Korisnički prijateljsko grafičko sučelje omogućava korisniku jednostavno podešavanje rada pojedinog modula. Izlazni podaci u tabličnom i XML formatu omogućavaju korisniku ne samo lijep pregled rezultata, već i lakšu daljnju elektroničku obradu podataka.

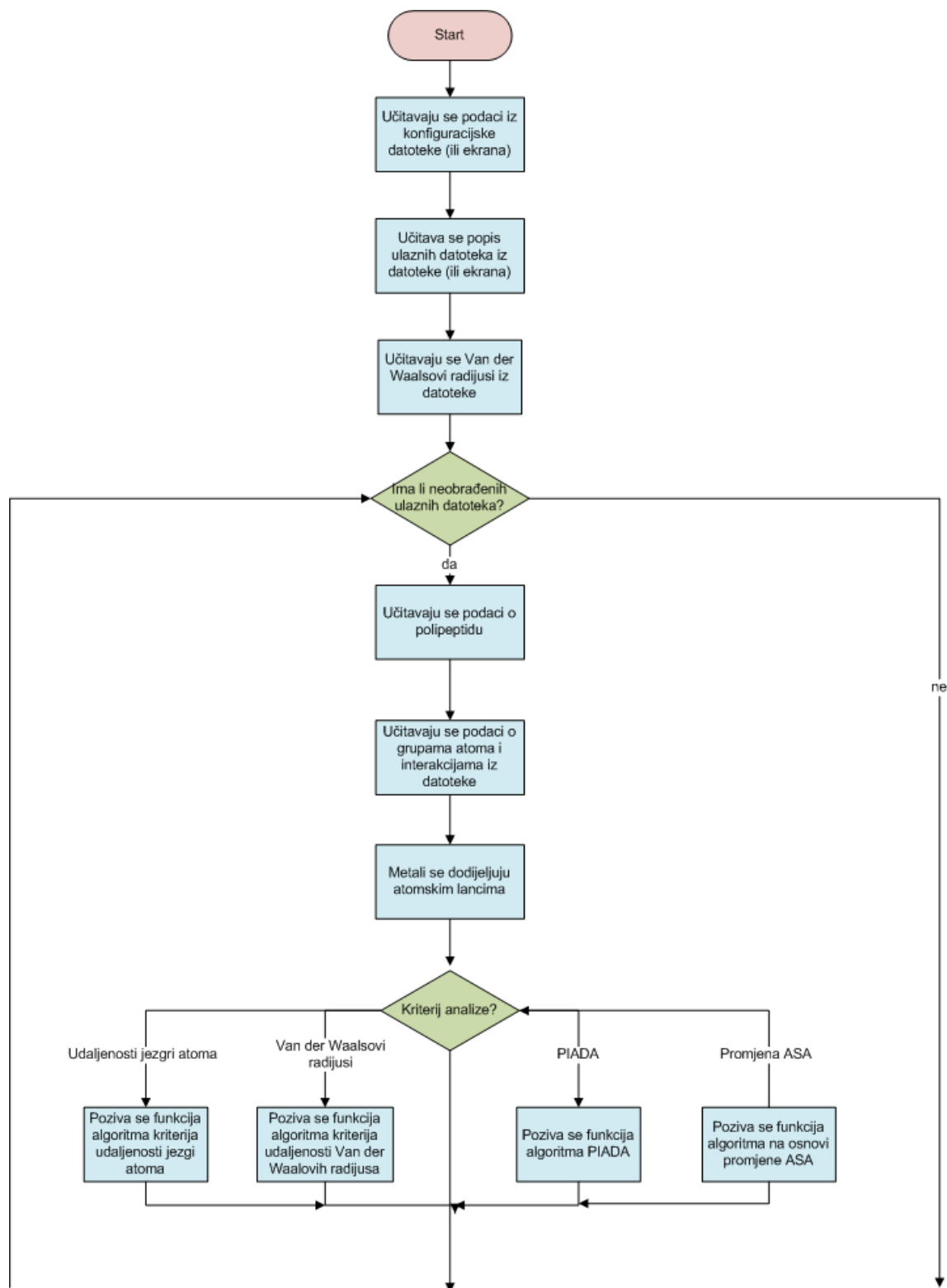
7 Dodatak A - dijagram toka PSA modula

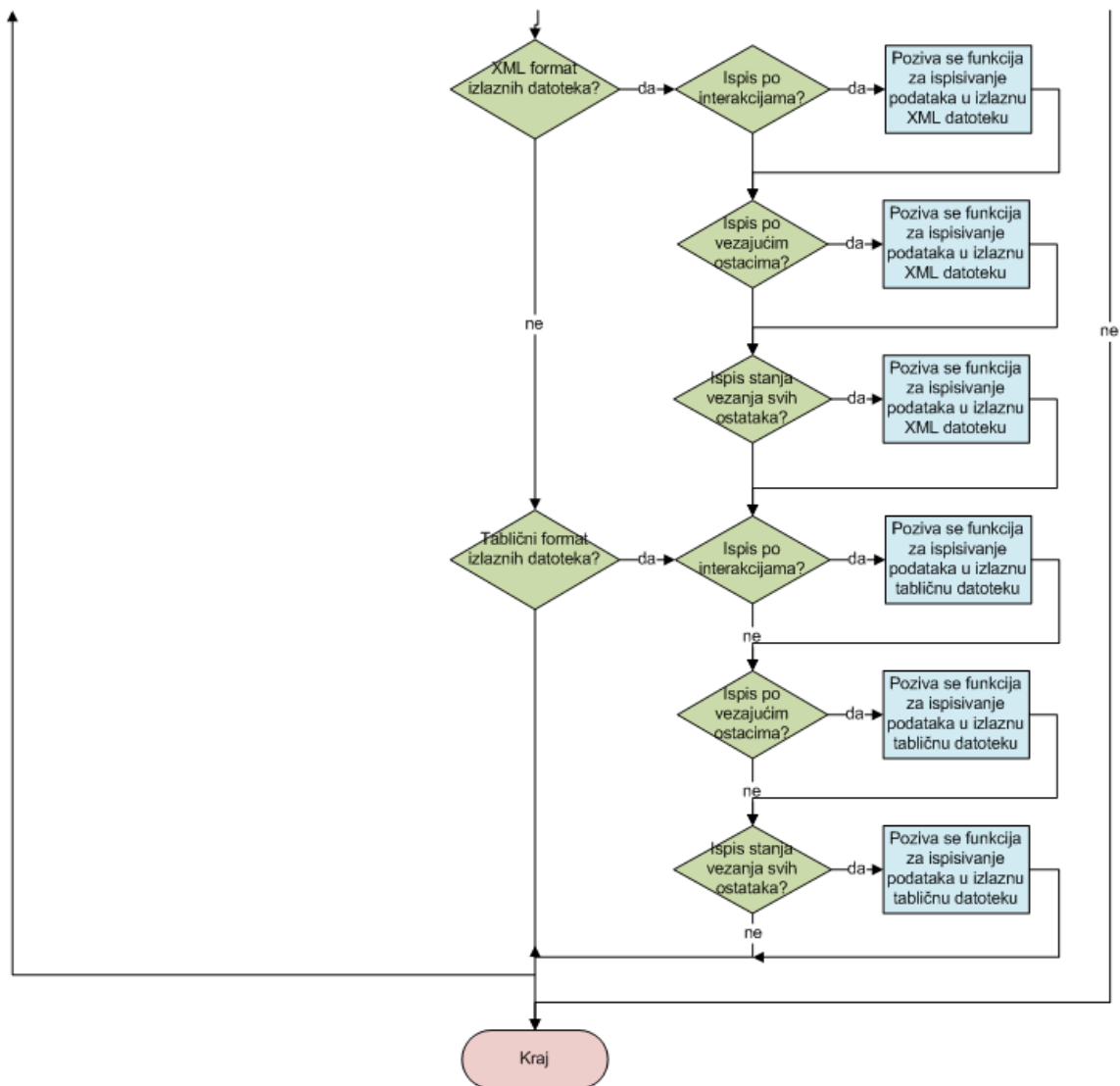






8 Dodatak B – dijagram toka PIA modula





9 Dodatak C – primjer konfiguracijske datoteke modela PSA

```
<PSAIA:configuration_file>
  <PSA:program_settings>
    <PSA:general>
      <PSA:write_log include="1">
        <log_filename>log\log.txt</log_filename>
      </PSA:write_log>
      <PSA:write_standard_output include="1">
        <confirmation_dialog>1</confirmation_dialog>
      </PSA:write_standard_output>
    </PSA:general>
    <PSA:analysis_type>
      <analyze_bound include="1"/>
      <analyze_unbound include="1"/>
    </PSA:analysis_type>
    <PSA:analysis_attributes>
      <PSA:calc_asa include="1">
        <z_slice>0.1</z_slice>
        <r_solvent>0</r_solvent>
        <hetatm include="1"/>
        <water include="1"/>
        <hydrogen include="1"/>
      </PSA:calc_asa>
      <PSA:calc_rasa include="1"/>
      <PSA:calc_dpx include="1"/>
      <PSA:calc_cx include="1">
        <cx_threshold>5</cx_threshold>
        <cx_volume>15.1</cx_volume>
      </PSA:calc_cx>
      <PSA:calc_hydro include="1"/>
      <PSA:calc_secondary_structure include="1">
      <PSA:calc_solvation include="1"/>
    </PSA:analysis_attributes>
    <PSA:output_files_options>
      <output_dir>output\</output_dir>
      <write_xml>1</write_xml>
      <write_table>1</write_table>
      <write_asa>1</write_asa>
      <write_atoms>1</write_atoms>
    </PSA:output_files_options>
    <PSA:input_files_options>
      <PSA:radii_file>example\radii.xml</PSA:radii_file>
      <PSA:standard_asa_file>example\asa.xml</PSA:standard_asa_file>
      <PSA:hydro_file>example\hydro.xml</PSA:hydro_file>
      <PSA:solvation_parameters_file>solvation.xml</PSA:solvation_parameters_file>
      <PIA:interactions_file>example\interactions.xml</PIA:interactions_file>
    </PSA:input_files_options>
  </PSA:program_settings>
</PSAIA:configuration_file>
```

10 Dodatak D – primjer konfiguracijske datoteke modela PIA

```
<PSAIA:configuration_file>
  <PIA:program_settings>
    <PIA:general>
      <PIA:write_log include="1">
        <log_filename>log\log.txt</log_filename>
      </PIA:write_log>
      <PIA:write_standard_output include="1">
        <confirmation_dialog>1</confirmation_dialog>
      </PIA:write_standard_output>
    </PIA:general>
    <PIA:contact_criterion>
      <PIA:maximum_distance include="0">
        <threshold>6.0</threshold>
        <intrainteractions_threshold>3</intrainteractions_threshold>
      </PIA:maximum_distance>
      <PIA:van_der_waals_distance include="0">
        <threshold>6.0</threshold>
        <intrainteractions_threshold>3</intrainteractions_threshold>
      </PIA:van_der_waals_distance>
      <PIA:piada include="1">
        <intrainteractions_threshold>3</intrainteractions_threshold>
      </PIA:piada>
      <PIA:asa_change include="0">
        <delta_asa>1.0</delta_asa>
        <z_slice>0.25</z_slice>
        <r_solvent>1.40</r_solvent>
      </PIA:asa_change>
    </PIA:contact_criterion>
    <PIA:output_files_options>
      <output_dir>output\</output_dir>
      <write_xml>1</write_xml>
      <write_table>1</write_table>
      <write_contacts>1</write_contacts>
      <write_binding_residues>1</write_binding_residues>
      <write_residue_binding_state>1</write_residue_binding_state>
    </PIA:output_files_options>
    <PIA:input_files_options>
      <PIA:radii_file>example\radii.xml</PIA:radii_file>
    <PIA:interactions_file>example\interactions.xml</PIA:interactions_file>
    </PIA:input_files_options>
  </PIA:program_settings>
</PSAIA:configuration_file>
```


11 Popis literature

- Mihel, Josip: **Alat za analizu površina proteina i mjesta proteinskih interakcija**
Fakultet Elektrotehnike i Računarstva, 2006
- W. C. Wimley, T. P. Creamer, and S. H. White, *Biochemistry* 35 (1996); 5109–5124.
- P. Shih, L. G. Pedersen, P. R. Gibbs, and R. Wolfenden, *J. Mol. Biol.* 280 (1998); 421–430.
- RCSB Protein Data Bank [<http://www.rcsb.org>]
- Qt library [<http://www.trolltech.com/products/qt>]
- Wikipedia [www.wikipedia.com]