

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 250

VIZUALIZACIJA MAKROMOLEKULA

Ivan Sović

Zagreb, lipanj 2008.

Veliko hvala Mili i Ani.

Sadržaj

1. Uvod.....	4
2. Standardni načini prikaza molekula	5
2.1. Prikaz kuglama i štapićima.....	5
2.2. Žičani prikaz.....	7
2.3. Kalotni prikaz	8
2.4. Prikaz površine molekule	8
2.5. Prikaz površine rekonstruirane iz zadanog grida	9
3. Ulazni podatci	11
3.1. PDB zapis strukture proteina	11
3.2. Zapis vrijednosti grida za rekonstrukciju površine.....	12
4. Opći opis alata VMol.....	13
4.1. Opis alata i njegove mogućnosti	13
4.2. Grafički prikaz dizajna alata	15
5. Struktura alata.....	16
5.1. Učitavanje PDB datoteka	16
5.2. Određivanje radijusa atoma	17
5.3. Dobivanje površine proteina.....	20
5.4. Generiranje spojne liste	21
5.5. Učitavanje datoteke koja sadrži grid i rekonstrukcija površine	22
5.6. Ostvarivanje različitih načina prikaza	25
5.7. Informacije o atomima, aminokiselinama i lancima	30
5.8. Sferno-polarnе značajke	32
5.9. Odabir bojanja.....	33
5.10. Grafičko sučelje	35
6. Način korištenja alata VMol.....	40
7. Prikazi testiranja alata.....	42
8. Zaključak.....	45
Literatura	46
Sažetak i ključne riječi	48

1. Uvod

Bioinformatika je u zadnjih desetak godina doživjela ogroman razvoj, a posebice nakon znatnog povećanja broja dostupnih struktura proteina. Prilikom proučavanja proteina, računala su postala neophodan alat. Posebno zanimljivo područje bioinformatike je prijanjanje proteina (eng. *docking*), i to zbog svoje primjene u medicini. Brzim testiranjem raznih mogućih kombinacija koje neki proteini mogu tvoriti, znatno se olakšava i ubrzava proces otkrivanja novih lijekova za razne bolesti. Upravo pri prijanjanju proteina, od ključne je važnosti i vizualizacija makromolekula. Ona nam omogućuje da pregledavamo ulazne podatke procesa prijanjanja, te nam isto tako pruža uvid u rezultate dokiranja. Bez alata ovakve namjene, provjeravanje ispravnosti rješenja bilo bi vrlo teško, ako ne i nemoguće. Baš za tu svrhu namijenjen je i program VMol, razvijen u sklopu ovog rada.

U 2. poglavlju dana je teoretska osnova za razumijevanje podataka prikazanih prilikom analize proteina. Formati zapisa ulaznih podataka koje koristi VMol opisani su u poglavlju 3. Poglavlje 4. sadrži opći opis programa razvijenog u sklopu ovog rada. U 5. poglavlju razrađena je izvedba pojedinih dijelova programa. Upute za njegovo korištenje nalaze se u poglavlju 6., a rezultati dobiveni testiranjem programa mogu se vidjeti u 7. poglavlju. Zaključak, literatura i sažetak, zajedno sa popisom ključnih riječi dani su u poglavljima 8., 9. i 10.

2. Standardni načini prikaza molekula

Postoji više metoda prikaza molekula. Neki standardni načini su: kalotni, žičani, prikaz kuglama i štapićima, te prikaz pomoću površine molekule. Također, za potrebe uvida u rezultate procesa prijanjanja proteina, potrebna je i mogućnost prikaza površine zadane molekule dobivene kuglinim funkcijama te raspoređivanjem na grid.

Kako bi se mogli razumjeti izlazni podatci ovog alata, u nastavku će biti pojašnjene upravo navedene metode prikaza.

2.1. Prikaz kuglama i štapićima

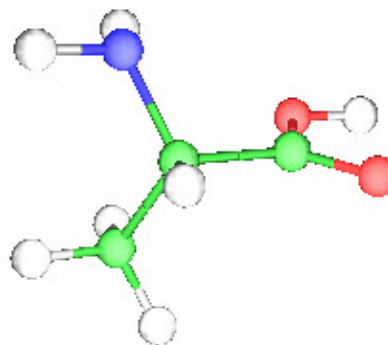
Prikaz kuglama i štapićima sastoji se od atoma sadržanih u molekuli, te veza među njima. Atomi su prikazani kuglama, te su povezani ravnim linijama koje spajaju njihova središta. Linije predstavljaju kovalentne veze među tim atomima [1]. Dvostruke i trostruke veze često se prikazuju zavojnicama koje tvore zaobljene poveznice između kugli.

Kovalentna veza je vrsta kemijskog povezivanja koje karakterizira dijeljenje para elektrona između atoma [2]. Drugim riječima, to je stabilnost između privlačenja i odbijanja atoma koja nastaje kada oni dijele elektrone.

U stvarnosti se linije koje predstavljaju veze atoma izvode pomoću žica ili krutih štapića (eng. *sticks*), dok se atomi predstavljaju kuglama (eng. *balls*), te otuda i dolazi ime ovog modela prikaza. Kod računalnog prikaza, za iscrtavanje kugli i štapića upotrebljavaju se odgovarajući prostorni modeli, čime se postiže realističnost prikaza i sličnost sa fizičkim izgledom modela molekule.



Slika 1. Izgled modela prikazanog kuglama i štapićima [3].



Slika 2. Prikaz aminokiseline Alanin pomoću kugli i štapića.

Kod ove vrste prikaza molekula kutovi među vezama, te njihove duljine, odgovaraju stvarnim odnosima unutar molekule, dok prostor kojeg atomi zauzimaju nije uopće prikazan, ili je predstavljen u nekom osnovnom obliku sa relativnim veličinama kugli. Prednost prikaza pomoću kugli i štapića jest bolji uvid u strukturu povezanosti unutar molekule, za razliku od modela molekula koji popunjavaju prostor.

Također, kao i kod drugih vrsta prikaza molekula, postoje *uobičajene* boje kojima se prikazuju pojedini atomi, te se tako omogućava njihovo lakše razaznavanje. Pod izrazom *uobičajeno* misli se na sustav boja pod nazivom CPK, što je skraćeno od imena autora: Corey, Pauling i Kultin. U tablici 1. prikazane su boje koje definira ovaj sustav [4].

Tablica 1. CPK sustav boja atoma.

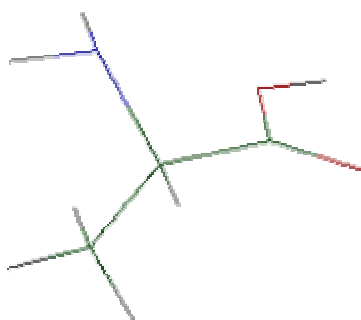
Atom	Boja
Dušik	plava
Kisik	crvena
Sumpor	žuta
Ugljik	siva, crna, zelena
Vodik	bijela
Natrij	tamnoplava
Fosfor, željezo, barij	tamnožuta

Flor, silicij, zlato	svjetložuta
Magnezij	tamnozeleno
Nikal, bakar, cink, bronca, litij	smeđa
Aluminij, kalcij, titanij, krom, mangan, srebro	tamnosiva
Boron, klor	zeleno
Jod	ljubičasto
Helij	svjetloroza
Ostalo	roza

2.2. *Žičani prikaz*

Ova vrsta prikaza je zapravo pojednostavljeni prikaz pomoću kugli i štapića. Žičani model se sastoji samo od veza između povezanih atoma. One su povučene iz središta svakog atoma molekule. Ova vrsta modela ne sadrži kugle kojima su predstavljeni atomi, te također korisniku ne daje nikakve informacije o veličini pojedinih atoma. Dijelovi veza među atomima mogu biti predstavljeni raznim bojama, čime se povećava preglednost modela, te daje korisniku informacija o atomima čija su središta povezana.

Slika 3. prikazuje žičani model molekule sa slike 2.

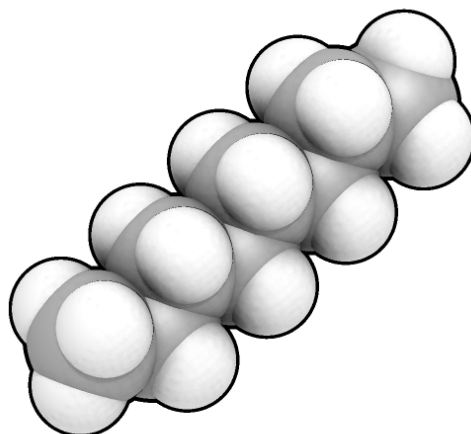


Slika 3. *Žičani prikaz molekule*

2.3. Kalotni prikaz

Kalotni modeli (eng. *Calotte models*) pripadaju skupini trodimenzionalnih modela koji popunjavaju prostor. Oni zapravo predstavljaju daljnji logični razvoj modela sa kuglama i štapićima [5]. Koriste se za detaljniji prikaz molekula. Atomi pojedinih molekula prikazani su kuglama raznih boja. Također, kugle su različitih radijusa, koji je ovisan o kemijskom elementu atoma. Veličina atoma, njihovi položaji, te veze među elementima odgovaraju njihovim odnosima u stvarnosti.

Slika 4. predstavlja kalotni prikaz molekule oktana. Ona se sastoji od 18 atoma vodika (prikazanih bijelom bojom), te 8 atoma ugljika (prikazanih sivom bojom).

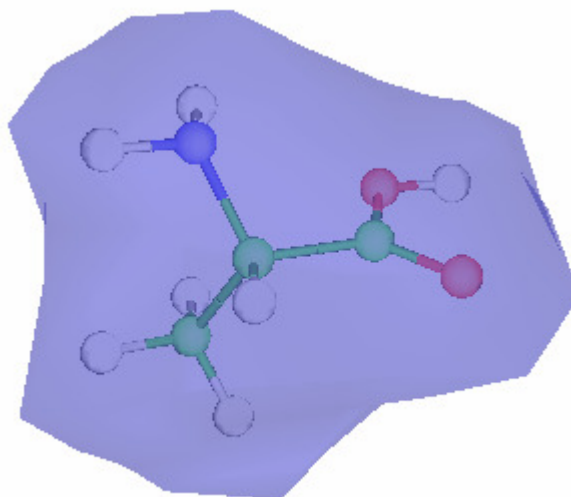


Slika 4. Kalotni prikaz jednog izomera oktana [6].

2.4. Prikaz površine molekule

Površina molekule je vanjska ovojnica te molekule, tj. prostor koji bi bio dostupan nekoj tekućini (eng. *solvent*) izvan nje. Kako bismo stvorili ovakav model prikaza, potrebne su nam informacije o položajima svih atoma unutar molekula, te njihovi radijusi.

Prikaz površine molekule je potreban, te iznimno praktičan, npr. kod alata čija je svrha prisanjanje proteina. On omogućuje korisniku jednostavno i brzo provjeravanje točnosti ponuđenog rješenja. Nedostatci ovakvog načina prikaza molekule su otežano raspoznavanje pojedinih atoma, te nedostatak informacija o vezama među njima.



Slika 5. Prikaz površine aminokiseline Alanin.

2.5. Prikaz površine rekonstruirane Marching Cubes algoritmom iz zadanog grida

Marching Cubes algoritam se primjenjuje pri rekonstrukciji površine nekog objekta iz podataka koji su dobiveni učitavanjem zadanog grida. Grid može nastati na mnogo načina, između kojih su i CT skeniranje (eng. *computed tomography*) te prijanjanje proteina.

U postupku prijanjanja proteina, zadani protein se mora rotirati i pomicati kroz prostor. U procesu računanja potrebnih podataka, protein se prvo mora *postaviti* na grid, tj. rasterizirati. Zatim slijede transformacije nad tim proteinom. Nakon što je protein rotiran i pomaknut, vrijednosti elemenata grida više nisu tako trivijalne kao što su to bile prije navedenih transformacija.

Druga spomenuta metoda - metoda CT skeniranja, primjenjuje se uglavnom u medicinske svrhe. Ona kao izlaz daje vrijednosti pojedinih elemenata grida, kao npr. u prethodnom slučaju. Iz tih vrijednosti se može rekonstruirati stvarni izgled skeniranog modela.



Slika 6. Rekonstruirani model ljudske lubanje korištenjem Marching Cubes algoritma [7].

3. Ulazni podatci

3.1. PDB zapis strukture proteina

Protein Data Bank, ili skraćeno PDB, je jedini svjetski repozitorij za procesiranje i distribuciju podataka o trodimenzionalnim strukturama velikih molekula proteina i nukleinskih kiselina [8]. Ovi podatci su dobiveni eksperimentalno, obično metodom rendgenske kristalografije ili NMR (eng. *Nuclear Magnetic Resonance*, nuklearna magnetska rezonanca) spektroskopije. Bazu podataka redovito nadopunjuju stručnjaci iz područja biologije i biokemije diljem svijeta. Ona je također javna i svima besplatno dostupna.

PDB je također i sinonim za format zapisa datoteka. On predstavlja standardnu reprezentaciju podataka makromolekularnih struktura dobivenih ranije spomenutim metodama [9]. To je pregledni tekstualni format, koji je nastao još tijekom 1970-tih godina. Podatci koje sadrži pojedini PDB dokumentastoje se, ukratko, od koordinata atoma, bibliografskih citata, primarne i sekundarne strukture, informacija, kristalografskih strukturnih faktora, te NMR eksperimentalnih podataka [10]. Tipična se PDB datoteka koja opisuje protein sastoji od nekoliko stotina, pa do tisuća linija teksta.

Primjerak PDB datoteke može sadržavati velik broj definiranih ključnih riječi. No, usprkos tome, neke od njih se češće koriste od ostalih, a neke su nam opet od bitnijeg značaja za određivanje svojstava molekula. U tablici 2. dani su opisi nekih češće korištenih, odnosno za ovaj rad bitnijih ključnih riječi.

Tablica 2. Neke češće korištene/bitnije ključne riječi u PDB formatu datoteka [11].

Ključna riječ	Opis
HEADER	Zaglavlje dokumenta
TITLE	Naziv dokumenta/molekule
AUTHOR	Ime autora danog dokumenta
REMARK	Zapisi koji započinju ovom ključnom riječi mogu sadržavati informacije u slobodnom obliku, ali također mogu i prikazati standardizirane informacije

ATOM	Linije koje započinju ovom ključnom riječi opisuju koordinate, te druge karakteristike atoma koji čine molekulu. Neke od karakteristika su: element atoma, oznaka lanca, oznaka aminokiseline kojoj pripada atom, temperaturni faktor,...
HETATM	Odnosi se na heterogene atome, tj. atome koji ne pripadaju standardnim grupama atoma. Ova ključna riječ se vrlo često koristi za opisivanje npr. molekule vode, ili atoma koji ne pripadaju molekuli proteina. Sintaksa joj je ista kao i ATOM liniji.
MODEL	Određuje serijski broj modela u slučaju kada se u istoj datoteci nalazi više struktura (što je npr. čest slučaj kod struktura određenih sa NMR metodom).
ENDMDL	Ovi zapisi su upareni sa MODEL linijama, te tako tvore individualne strukture koje se mogu pronaći u dokumentu.
TER	Označava kraj liste ATOM/HETATM linija za pojedini lanac.

3.2. Zapis vrijednosti grida za rekonstrukciju površine

Grid možemo zamisliti kao trodimenzionalnu matricu popunjenu realnim brojevima. Matrica je kubna, što znači da su njezina širina, visina i dubina jednake. Format zapisa datoteke koja sadrži grid je jednostavan, te se sastoji samo od najnužnijih podataka za uspješnu rekonstrukciju površine.

Prvi podatak u datoteci ovog formata je ukupan broj elemenata grida. Odmah za njim slijede svi elementi grida po redu, tj. njihove x, y i z koordinate, te intenzitet u toj točki. U slučaju da nam je potrebna duljina jedne stranice grida, možemo je lako dobiti iz postojećih podataka, i to tako da izračunamo treći korijen od ukupnog broja elemenata. Također, datoteka je zapisana u binarnom formatu, što olakšava njeno učitavanje, te smanjuje njezinu veličinu.

4. Opći opis alata VMol

4.1. Opis alata i njegove mogućnosti

VMol je alat namijenjen makromolekularnoj vizualizaciji. On omogućuje korisniku uvid u standardne prikaze molekula, kao što su: prikaz kuglama i štapićima, žičani model, kalotni model, te prikaz površine zadane molekule. Uz sve navedeno, VMol također omogućava i uvid u površinu dobivenu rekonstruiranjem iz zadanog trodimenzionalnog grida *Marching Cubes* algoritmom. Grid je dobiven modeliranjem površine molekule kuglinim funkcijama.

Podatci o nekoj molekuli se dobivaju učitavanjem odgovarajuće PDB (*Protein Data Bank*) datoteke. VMol danu datoteku učitava, parsira te prilagodi odabranom načinu prikaza. Unutar jedne PDB datoteke može se nalaziti i više različitih modela. U tom slučaju, VMol nudi korisniku opciju odabira koji model se trenutno želi prikazivati. Također, kako bi ovaj alat mogao prikazati i površinu rekonstruiranu s grida, potrebni podatci se učitavaju iz binarne datoteke posebnog formata. Za rekonstrukciju površine iz grida moguće je zadati i željenu isovrijednost, za slučaj da ona koju VMol ponudi nije odgovarajuća za učitani model.

Svi načini prikaza koje VMol može ostvariti međusobno se mogu superponirati. To korisniku omogućuje bolji uvid u međusobne odnose atoma i veza unutar molekule, te neke druge informacije kao što su: informacije o pojedinom atomu, lancu ili aminokiselini, te prikazivanje sferno-polarnih značajki za odabrane parametre.

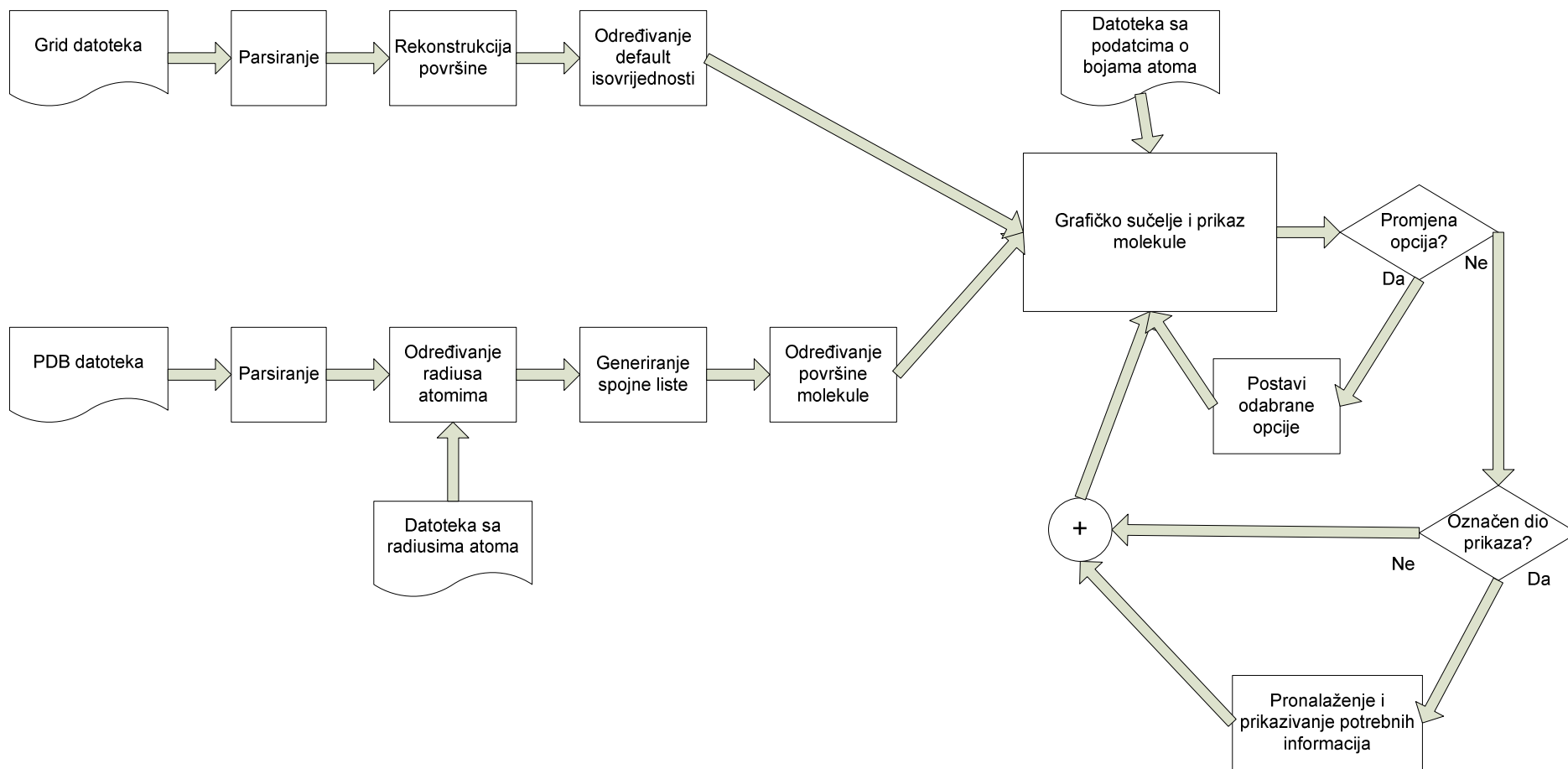
Kako bi se ovim alatom moglo interaktivno proučavati molekule i njihove značajke, VMol omogućuje sve osnovne manipulacije prikazom: pomicanje, okretanje i približavanje/udaljavanje prikaza. Za dobivanje željenih informacija, dovoljno je samo pritisnuti mišem mjesto interesa na trenutnom prikazu. Sferno-polarnu značajku se mogu dobiti samo kod prikaza površine, dok kod svih ostalih prikaza odabir željenog atoma daje informacije o njemu, aminokiselini kojoj pripada, te odgovarajućem lancu.

Dio informacija koje VMol pruža također se mogu dobiti i iz boja kojima su označeni pojedini dijelovi prikaza. Moguće odabrati jedan od nekoliko ponuđenih načina bojanja, kako bi se dobio prikaz koji najbolje odgovara korisniku. VMol

omogućuje sljedeće načine prikaza: "Default", "Atomi", "Aminokiseline", "Lanci", "Sferno polarne značajke". Njihovi opisi će biti izneseni u nastavku.

Sve navedene opcije moguće je mijenjati iz prilagodljivog grafičkog sučelja. Osim toga, za potrebe bržeg učitavanja i prikaza molekula, omogućena je i opcija određivanja putanja do željenih datoteka iz naredbenog retka.

4.2. Grafički prikaz dizajna alata



Slika 7. Dizajn alata VMol.

5. Struktura alata

5.1. Učitavanje PDB datoteka

Struktura PDB datoteke je detaljnije objašnjena u poglavlju 3.1. Spomenuto je kako je ona u tekstualnom formatu, što znači da bi za njezinu obradu prvo bilo potrebno učitavati liniju po liniju teksta. Nadalje, svaki učitani redak teksta potrebno je parsirati. Parsiranje je postupak kojim se podatci razlažu na manje i jasnije komade informacija, na kojima se mogu vršiti određene radnje [12]. Sljedeće ključne riječi su nam od važnosti pri parsiranju u svrhu trodimenzionalnog prikaza molekule: MODEL, ATOM i HETATM.. Pronalaženjem neke od ovih ključnih riječi dogode se sljedeće radnje:

- MODEL – kada parser naiđe na liniju teksta koja počinje ovom riječi, automatski se alokira (zauzme) memorija potrebna za pohranjivanje podataka o novom modelu. Dodatno, uz ovu ključnu riječ nalazi se i broj koji označava serijski broj modela. On ne mora nužno početi od nule, te se tretira više kao naziv danog modela, nego njegov redni broj.
- ATOM – linija koja počinje ovom ključnom riječi sadrži sve potrebne informacije o svakom pojedinom atomu unutar željene molekule. ATOM linije se obično nalaze unutar MODEL/ENDMDL izraza. U slučaju da se ovakva linija nađe i bez prethodno definiranog modela, pretpostavlja se da se svi atomi u danom PDB-u nalaze unutar istog modela, te se zauzme potrebna memorija za njihovo pohranjivanje. Usprkos tome, VMol ne ograničuje da se i nakon zasebne atom linije (tj. linije koja nije navedena unutar MODEL/ENDMDL izraza) ne definira novi model. Tada će se dogoditi pretpostavljeni niz radnji za taj slučaj (pogledati pojašnjenje parsiranja MODEL linije). Za svaku učitaneu ATOM liniju proširuje se lista već učitanih atoma unutar trenutnog modela. Iz te se linije zatim izdvoje sve bitne informacije, kao što su x, y i z koordinate, kemijski element, ime aminokiseline kojoj pripada, itd.
- HETATM – za liniju koja počinje ovom riječi VMol poduzima potpuno iste radnje kao i za ATOM linije. Kako je struktura podataka ovih linija također jednaka, podatci se mogu obrađivati na isti način.

Nakon upravo opisanog učitavanja podataka u memoriju, preostaju još neki koraci koje je potrebno izvesti prije nego parsiranje bude završeno. Za svaki učitani model izvrši se sljedeće:

- sortiranje atoma po lancima – u slučaju da podatci unutar PDB datoteke nisu ispravno definirani, VMol automatski sortira sve atome po oznakama lanaca. Važnost redoslijeda atoma dolazi do izražaja prilikom određivanja označenog lanca, odnosno aminokiseline.
- centriranje molekule proteina – molekula određena unutar parsirane datoteke često nije centrirana u središte Kartezijevog koordinatnog sustava. Za potrebe prikaza molekule, praktičnije je ako se ona nalazi u ishodištu, jer to omogućuje prirodnu rotaciju molekule, a osim toga osigurava da molekula neposredno nakon učitavanja ne bude izvan okvira prikaza. Centriranje molekule se provodi tako da se prvo odredi aritmetička sredina svih atoma molekule, te se zatim svi atomi pomaknu za izračunati iznos u smjeru središta koordinatnog sustava.
- generiranje spojne liste – kako bi prikazivanje molekule bilo čim brže, spojna lista se generira u postupku učitavanja PDB datoteke.
- određivanje površine molekule – kao i kod generiranja liste, površina molekule se određuje jednom, i to pri učitavanju.

5.2. Određivanje radijusa atoma

U datotekama PDB formata nije predviđena informacija o radijusu pojedinog atoma. Stoga je potrebno te podatke naknadno pridijeliti svakom atomu. Poznavanje radijusa atoma nam je potrebno kako bismo mogli vjerno prikazati učitanoj molekuli, te odrediti njenu površinu (kako za vizualizaciju, tako i za proces prijanjanja proteina). Za tu svrhu koristi se datoteka "*atmtypenumbers*", koja sadrži popis atoma i nekih jednostavnijih molekula, te njima dodijeljene radijuse. Ova datoteka se isporučuje uz MSMS - program namijenjen brzom određivanju površine molekule, kojeg i VMol koristi za tu svrhu.

Kao i kod PDB datoteka, da bismo učitali podatke iz "*atmtypenumbers*" datoteke, potrebno ju je parsirati. Datoteka se sastoji od dva dijela:

1. dio čine popisi atoma te podatci o njima. Redak iz datoteke koji sadrži ove podatke sastoji se od sljedećih zapisa, redom:

- a. ključna riječ "radius"
- b. redni broj atoma (ključ kod pretraživanja)
- c. radijus kovalentne veze atoma
- d. radijus atoma u Angstromima (1 Angstrom = 10^{-10} m)
- e. opcionalni radijus atoma za ujedinjene atome

Primjer:

```
radius      2  0.66      1.40      1.60
```

2. dio se sastoji od zapisa tipova atoma:

- a. uzorak imena aminokiseline
- b. uzorak imena atoma
- c. redni broj atoma (iz prethodnog dijela zapisa datoteke)

Primjer:

```
ARG  CD      8
```

Uzorci pod *a* i *b* u ovom dijelu zapisa zapravo su regularni izrazi. Regularni izrazi su skupine kombinacija ključeva koje omogućavaju korisniku veliku kontrolu nad onim što pretražuje [13].

Ako je učitavanje "*atmtypenumbers*" datoteke uspješno izvršeno, nakon parsiranja PDB datoteke između ostaloga slijedi i određivanje radijusa pojedinom atomu. Za svaki atom učitane molekule potrebno je usporediti imena tog atoma i aminokiseline kojoj pripada sa uzorcima dohvaćenim iz drugog dijela datoteke saradijusima. C++ nema ugrađenu podršku za regularne izraze, pa zato za uspoređivanje uzoraka postoje dvije mogućnosti: korištenje postojeće biblioteke, ili implementacija vlastitog algoritma za usporedbu. VMol koristi vlastitu implementaciju podrške za regularne izraze. Korišteni algoritam nije potpuna podrška regularnim izrazima – podržava samo slučajeve uzoraka koji se mogu naći unutar "*atmtypenumbers*" datoteke, a radi na sljedeći način:

1. ako se uzorak sastoji samo od znaka "*", ime bilo kojeg oblika je odgovarajuće, pa funkcija za usporedbu odmah vraća istinitu vrijednost (eng. *true*) i daljnje izvođenje se prekida

2. petljom se prođe kroz sve znakove uzorka; ako su pronađene uglate zagrade, a između njih postoji znak '-', lista znakova unutar zagrada se proširuje svim elementima unutar raspona znaka prije, te znaka nakon '-'.
3. zatim se petljom prođe kroz sve znakove imena koje se uspoređuje s uzorkom. Ukoliko se znak imena ispituje sa znakom uzorka '?', ispituje se da li znak imena odgovara uzorku za prethodni znak. Znak uzorka '.' označava da na tom mjestu imena može stajati bilo koji znak, dok niz znakova '*' označava niz bilo koje duljine proizvoljnih znakova. Ako su pak ovi znakovi predznačeni sa znakom '\\', oni nemaju svoje posebno značenje, nego se direktno uspoređuju s odgovarajućim znakom imena. Isto vrijedi i za sam znak '\\', tj. kada bismo htjeli usporediti znak imena sa doslovno tim znakom, moramo koristiti regularni izraz '\\\\'.
4. Ako svi znakovi imena odgovaraju znakovima uzorka, funkcija za usporedbu vraća istinitu vrijednost, a u protivnome neistinitu (eng. *false*).

Ako je ispitivani atom pronađen u listi iz drugog dijela datoteke "*atmtypenumbers*", njemu se dodjeljuje odgovarajući radijus iz prvog dijela iste datoteke. Radijus se pronalazi preko ključa navedenog unutar strukture zapisa te datoteke. Ako atom ne odgovara niti jednom uzorku, njemu se pridjeljuju posebne vrijednosti radijusa: 0.01 Angstrom za radijus atoma, te 0.001 Angstrom za kovalentni radijus.

U slučaju da je trenutni atom heterogeni, vrijede posebna pravila, i to zbog specifičnosti "*atmtypenumbers*" datoteke. Kako u aminokiselinama neki atomi imaju poseban naziv, kao npr. alfa ugljik, tako oni imaju i svoje kratice koje se koriste za njihovo imenovanje. Tako bi se alfa ugljik označavao sa CA, što bi usporedbom s uzorcima navedenim unutar "*atmtypenumbers*" datoteke davalo krivi radijus zbog korištenja istog imena za označavanje različitih atoma, u ovom slučaju kalcija (Ca). No, kalcij ne može biti dio standardne grupe atoma proteina, te se zbog toga unutar PDB datoteka nalazi u liniji pod ključnom riječi HETATM. Za njega je posebno definiran radijus, i to unutar VMol-a. Sličan slučaj se događa i sa atomom kadmija (Cd; delta ugljik = CD).

5.3. Dobivanje površine proteina

Površina proteina se u VMol-u dobiva korištenjem programa MSMS (naziv ovog programa dolazi od "*Michel Sanner's Molecular Surface*"), a on omogućuje vrlo efikasnu triangulaciju površine nedostupne otopinama [14]. Kako bi ovaj program mogao izračunati koordinate točaka površine, potrebno mu je dati listu koordinata svih atoma i njihove radijuse, i to u tekstualnom obliku. Svi retci ove datoteke imaju isti oblik, i to:

$$x \quad y \quad z \quad r$$

Gdje su x , y i z realni brojevi i predstavljaju koordinate jednog atoma molekule, a r radijus tog atoma (također je realan broj).

Kao izlaz, MSMS stvara dvije datoteke: jedna ima nastavak ".vert" i sadrži popis svih koordinata točaka na površini, normale u njima, te dodatne informacije o pripadnom atomu, dok druga datoteka ima nastavak ".face" i sastoji se od liste indeksa koordinata koje čine poligone, te dodatnih informacija o njima. Redak datoteke koja sadrži popis koordinata na površini izgleda ovako:

$$x \quad y \quad z \quad nx \quad ny \quad nz \quad d \quad num \quad f$$

U sljedećoj tablici dana su pojašnjenja navedenih podataka, odnosno onih koji su od značaja za VMol.

Tablica 3. Pojašnjenja stavaka jednog retka ".verts" datoteke.

Parametar	Objašnjenje
x, y i z	Realni brojevi, određuju koordinatu točke na površini.
nx, ny i nz	Realni brojevi, predstavljaju vektor normale na površinu u točki određenoj s prethodnim koordinatama.
num	Cijeli broj; indeks atoma iz datoteke korištene kao ulaz u MSMS program. Indeksi se kreću u rasponu od 1 do N, pri čemu je N ukupni broj atoma u ulaznoj datoteci.
d, f	Cijeli brojevi koji nisu od značaja za ovaj alat.

Nadalje, slijedi struktura jednog retka datoteke s popisom poligona:

n1 n2 n3 a b

i pojašnjenja pojedinih podataka, dana u tablici:

Tablica 4. Pojašnjenja stavaka jednog retka ".face" datoteke.

Parametar	Objašnjenje
n1	Cijeli broj; indeks prve točke poligona, koja se uzima iz datoteke koja sadrži točke površine. Indeksi se kreću u rasponu od 1 do N, pri čemu je N ukupni broj točaka u datoteci sa koordinatama.
n2	Cijeli broj; indeks druge točke poligona.
n3	Cijeli broj; indeks treće točke poligona.
a, b	Cijeli brojevi koji nisu od značaj za ovaj alat.

Neposredno nakon pozivanja programa MSMS i generiranja upravo opisanih datoteka, VMol učitava te datoteke, a pribavljene podatke drži u memoriji za daljnje korištenje. Oni će kasnije biti upotrijebljeni u postupcima prikazivanja površine, te pri određivanju odabrane točke za dobivanje sferno-polarnih značajki.

5.4. Generiranje spojne liste

Korak generiranja spojne liste atoma ključan je kako bi se mogle prikazati veze među njima. Ovim postupkom se utvrđuje koji su atomi učitane molekule proteina međusobno povezani.

Kako bi se utvrdila povezanost svih atoma, potrebno je proći kroz sve parove atoma svakog modela unutar molekule. Složenost ovakvog algoritma stoga je veća od n^2 , zbog čega je praktičnije jednom na početku odrediti sve veze te zauzeti određenu količinu memorije, nego štedjeti na memoriji i određivati veze svakom iteracijom iscrtavanja.

Prvo ćemo pojasniti kako se zapravo dva atoma nemetala vežu u jednostavnu molekulu. Veza između takva dva atoma naziva se kovalentna veza. Ona nastaje kada dva atoma dijele par elektrona. Zapravo, atomi mogu dijeliti i više od jednog para elektrona. Svaki dijeljeni par čini jednu kovalentnu vezu. Višestruke kovalentne veze nastaju ako je atomima potrebno više od jednog elektrona za nadopunjavanje njihovih valentnih ljuski [15].

Struktura podataka, koju VMol koristi za pohranjivanje informacija o pojedinom atomu, osim radijusa samog atoma sadrži i njegov kovalentni radijus. Iznos kovalentnog radijusa je definiran kao polovica udaljenosti između dvije identične atomske jezgre, koje povezuje kovalentna veza [16]. Udaljenost središta dva atoma računa se prema:

$$d = \sqrt{(a2.x - a1.x)^2 + (a2.y - a1.y)^2 + (a2.z - a1.z)^2}, \quad (1)$$

a zbroj kovalentnih radijusa ista dva atoma:

$$dc = a1.cov_r + a2.cov_r, \quad (2)$$

$$\text{ako je: } \begin{cases} d \leq dc, \text{ atomi su povezani kovalentnom vezom,} \\ d > dc, \text{ atom nisu povezani.} \end{cases}$$

Koordinate prvog atoma označene su sa $a1.x$, $a1.y$ i $a1.z$, dok $a1.cov_r$ označava njegov kovalentni radijus. Analogno tome označene su i komponente drugog atoma.

Ako su dva ispitivana atoma povezana, lista koja pohranjuje popis svih veza se proširuje, te se na njen kraj dodaje element koji sadrži pokazivače na oba atoma u toj vezi.

5.5. Učitavanje datoteke koja sadrži grid i rekonstrukcija površine

Opis formata datoteke koja sadrži grid iznesen je u poglavlju 3.2. Kako bismo učitali takvu datoteku, potrebno ju je otvoriti za binarno učitavanje, te pročitati podatke navedene u tablici 5.

Tablica 5. Detalji o učitavanju elemenata grida.

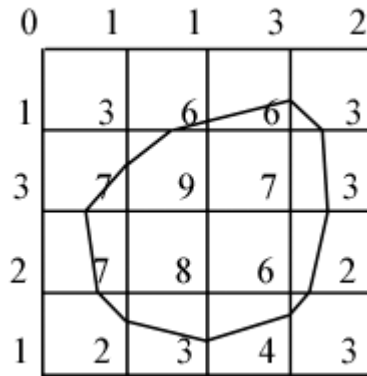
R.br.	Podatak	Veličina (bajt)	Pojašnjenje
1.	n	8 (unsigned long int)	Prvo se učitava ukupan broj elemenata grida.
2.	element[1..n]	n * 32 (veličina CGrid strukture)	Zatim se učitava n elemenata slijedno. Svaki element ima veličinu od 32 bajta zbog oblika strukture koja sadrži informacije o jednom elementu grida, a sastoji se od 4 realna broja veličine 8 bajta

Kada bismo zamislili ovaj grid na najjednostavniji način, dobili bismo jednu kocku, podijeljenu na puno manjih kockica. Svaka od ovih kockica je razapeta između 8 susjednih elemenata grida, čime za svaki njezin vrh dobivamo koordinate i određeni intenzitet. Ovakva kockica je zapravo najsitniji volumni element nekog modela, te se stoga naziva *voksel* (eng. *voxel, volume element*). U još jednom primjeru, voxel se može zamisliti kao trodimenzionalni ekvivalent piksela (eng. *pixel, picture element*) [17].

Već je spomenuto da svaka točka grida ima svoju vrijednost, tj. intenzitet. Ti intenziteti se još nazivaju i *isovrijednosti* (eng. *isovalue*). Između isovrijednosti definirane su različite *isopovršine* (eng. *isosurface*). To su površine koje se određuju za zadani referentni intenzitet, tj. referentnu isovrijednost. Ova referentna vrijednost određuje na kojem mjestu u gridu se nalazi površina:

- za isovrijednosti grida koje su manje od referentne promatrana točka se nalazi izvan volumena opisanog površinom,
- za isovrijednosti grida koje su veće od referentne promatrana točka se nalazi unutar volumena opisanog površinom.

Za bolje shvaćanje pojma isovrijednosti i isopovršine, korisno je pogledati sliku 8.:



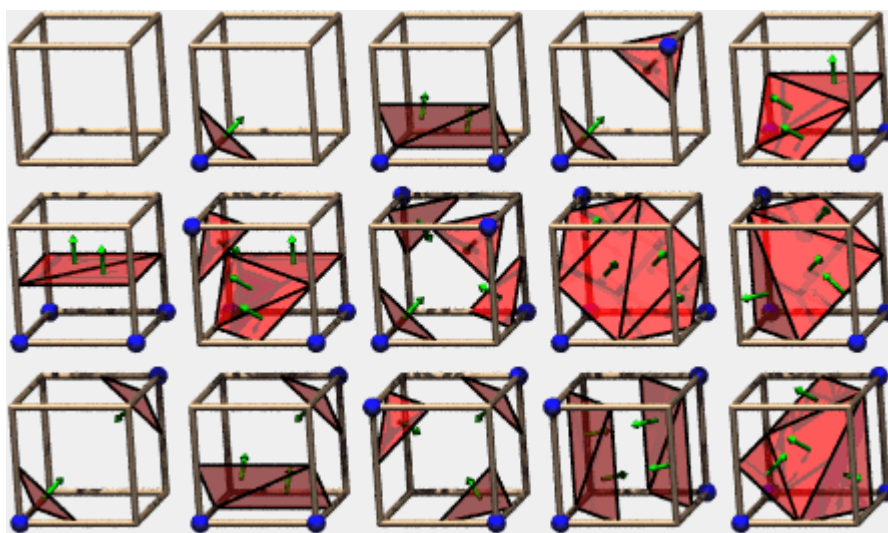
Slika 8. Dvodimenzionalni primjer određivanja isopovršine (isobare) za referentnu isovrijednost 5 [19].

Upravo određivanjem isopovršine iz dane referentne vrijednosti na trodimenzionalnom gridu bavi se *Marching Cubes* algoritam. On se zasniva na ideji interpolacije površine između vrijednosti koje padaju unutar i izvan volumena površine. Može se koristiti bilo koja vrsta interpolacije, a za potrebe alata VMol odabrana je linearna interpolacija. Marching Cubes u nekom trenutku razmatra samo jedan voksel i vrijednosti intenziteta u njegovim točkama. Zatim je potrebno iterirati kroz sve volumne elemente, te tako odrediti ukupnu isopovršinu.

Kako se jedan voksel sastoji od 8 vrhova, to nam daje ukupno 256 potencijalnih kombinacija stanja u tim vrhovima [18]. Algoritam se može pojednostaviti ako uzmemo u obzir kombinacije koje se ponavljaju pod određenim uvjetima:

- rotacija za proizvoljan broj stupnjeva oko bilo koje koordinatne osi
- zrcaljenje oblika po bilo kojoj osi
- invertiranjem stanja svih vrhova i okretanjem normala određenih poligona

Zahvaljujući ovim uvjetima, početnih 256 kombinacija može se reducirati na samo 15 različitih. Slika 9. prikazuje tih 15 kombinacija zajedno s predefiniranim skupovima poligona za te slučajeve, namijenjenim aproksimaciji površine [18].



Slika 9. Mogućih 15 kombinacija stanja u vrhovima, i njihova poligonizacija [20].

Na slici 9. plave kugle označavaju vrhove za koje je određeno da se nalaze unutar isopovršine, a zelene strelice označavaju normale odgovarajućih poligona.

Određivanjem poligona svih vokseli, te njihovim sakupljanjem u listu dobivamo trodimenzionalni model isopovršine za odabranu referentnu isovrijednost.

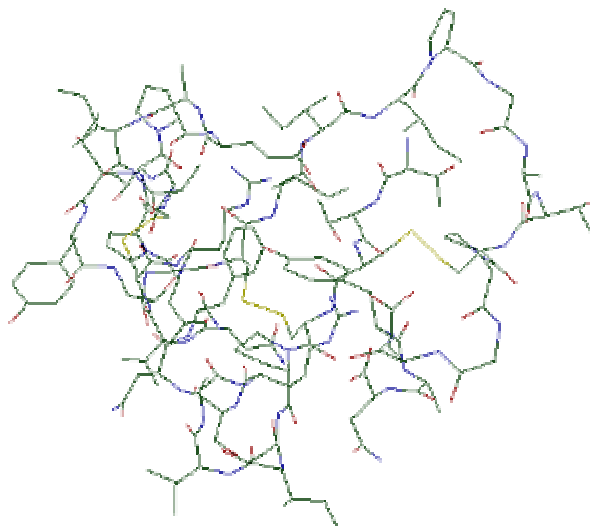
5.6. Ostvarivanje različitih načina prikaza

Nakon postupaka učitavanja željenog proteina, odnosno grida, dobivene podatke je potrebno prikazati. Neki od standardnih načina prikaza već su objašnjeni ranije, a sada će biti dano pojašnjenje njihove izvedbe u VMol-u.

5.6.1. Žičani prikaz

Ovo je ujedno i najjednostavniji način prikaza. Za odabrani model potrebno je iterirati kroz sve elemente spojne liste, te između svakog para povezanih atoma nacrtati spojnicu na trodimenzionalnom prikazu. Spojnica se sastoji od dviju linija jednakih duljina. Povlači se iz središta svakog od atoma, pa do polovice udaljenosti među njima. Razlog tome je mogućnost označavanja pojedinog dijela spojnice različitom bojom, što korisniku ovog alata daje uvid u to koji atomi čine ovu vezu. Za samo iscrtavanje na ekran, kao i kod drugih načina prikaza, koriste

se funkcije ugrađene u OpenGL (eng. *Open Graphics Library*) biblioteku. Na slici 10. može se vidjeti primjer izgleda ovog prikaza u VMol-u.



Slika 10. Žičani prikaz molekule u VMol-u.

Na slici 10. pozadinska je boja promijenjena u odnosu na ostale prikaze, kako bi se žičani model mogao jasnije vidjeti.

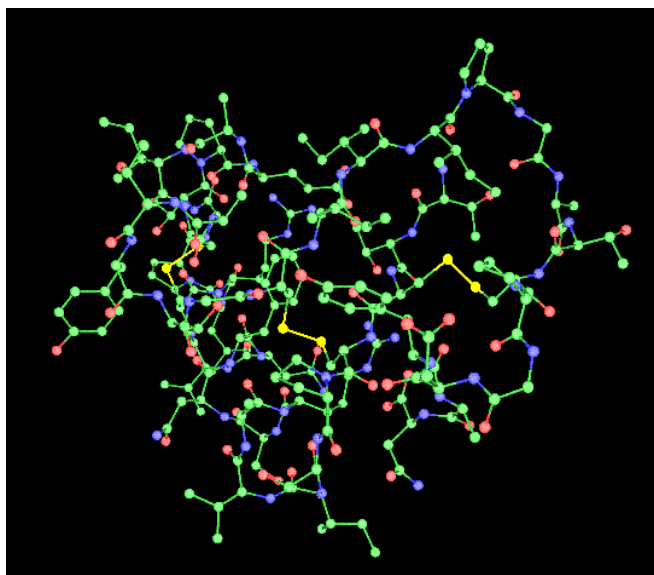
5.6.2. Prikaz kuglama i štapićima

Proširenjem žičanog prikaza trodimenzionalnim modelima dobiva se prikaz kuglama i štapićima. Spojnice više nisu izvedene pomoću jednostavnih linija, nego se na njihovim mjestima prikazuju prostorni modeli štapića. Također, u središtu svakog atoma prikazuje se model kugle. Pojedina spojnica se i u ovom načinu izvodi od dva dijela, kao i kod žičanog prikaza.

Kako bi se ovaj način prikaza mogao ostvariti, pri pokretanju VMol-a učitavaju se modeli kugle i štapića. Format zapisa ovih prostornih modela stvoren je upravo za potrebe korištenja u VMolu. Podatci u ovom formatu zapisani su u tekstualnom obliku, i sastoje se od podataka navedenih u tablici 6.

Tablica 6. Format zapisa prostornih modela kugle i štapića

Podatak	Objašnjenje
n	Cijeli broj; broj koordinata koje model sadrži.
m	Cijeli broj; ukupan broj poligona od kojih se model sastoji.
$x_i \ y_i \ z_i \ n_{xi} \ n_{yi} \ n_{zi},$ $i = 0..(n-1)$	Realni brojevi; x_i , y_i i z_i su koordinate pojedine točke modela, a n_{xi} , n_{yi} i n_{zi} su normale modela u točki.
$t_{1j} \ t_{2j} \ t_{3j},$ $j = 0..(m-1)$	Cijeli brojevi; t_{1j} , t_{2j} i t_{3j} predstavljaju indekse točaka od koji se sastoji pojedini poligon (u ovom slučaju trokut).



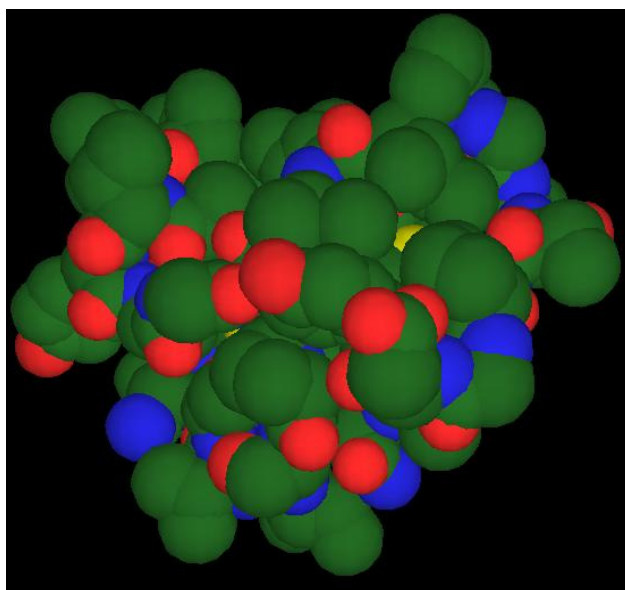
Slika 11. Prikaz kuglama i štapićima u VMol-u.

Na slici 11. prikazana je ista molekula kao i na slici 10., ali u ovom slučaju pomoću kugli i štapića.

5.6.3. Kalotni prikaz

U ovom načinu prikaza, na koordinatama svakog atoma nalaze se trodimenzionalni modeli kugli, koji predstavljaju taj atom. Veličina svake kugle

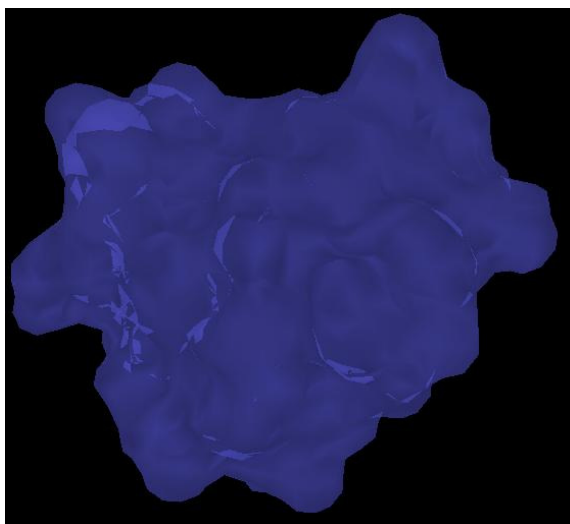
prilagođava se proporcionalno radijusu odgovarajućeg atoma kako bi se dobio realističan prikaz molekule. Kako bi mu se postavila odgovarajuća veličina, model kugle se na početku svede na jediničnu veličinu radijusa, te se nakon toga skalira po sve tri osi prema radijusu odgovarajućeg atoma. Model kugle koji se koristi kod kalotnog prikaza je isti kao i kod prikaza kuglama i štapićima, te se stoga on ne učitava ponovno. Na slici 12. dan je primjer kalotnog prikaza za istu molekulu kao i na slici 10.



Slika 12. Kalotni prikaz u VMol-u.

5.6.4. Prikaz površine molekule

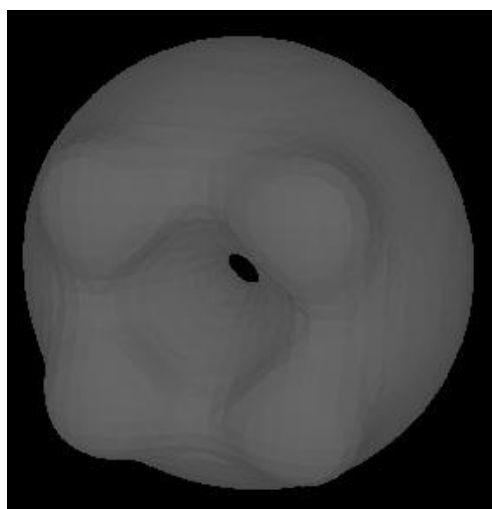
Podatci dobiveni programom MSMS sastoje se od koordinata točkica površine, te liste indeksa tih koordinata koji čine poligone. Sami ti podatci dostatni su za prikazivanje izgleda površine učitane molekule. Zbog mogućnosti superponiranja različitih načina prikaza, površina je iscrtana djelomično prozirno. Na različitim dijelovima površine mogu biti primjenjene razne boje, ovisno o odabranom načinu bojanja molekule. Kako bi VMol znao kojem atomu pripada neka točka površine, za svaku se točku iz MSMS-ovih izlaznih podataka učitava i indeks atoma kojem ona pripada. Slika 13. prikazuje površinu molekule sa slike 10.



Slika 13. Prikaz površine u VMol-u.

5.6.5. Prikaz rekonstruirane površine

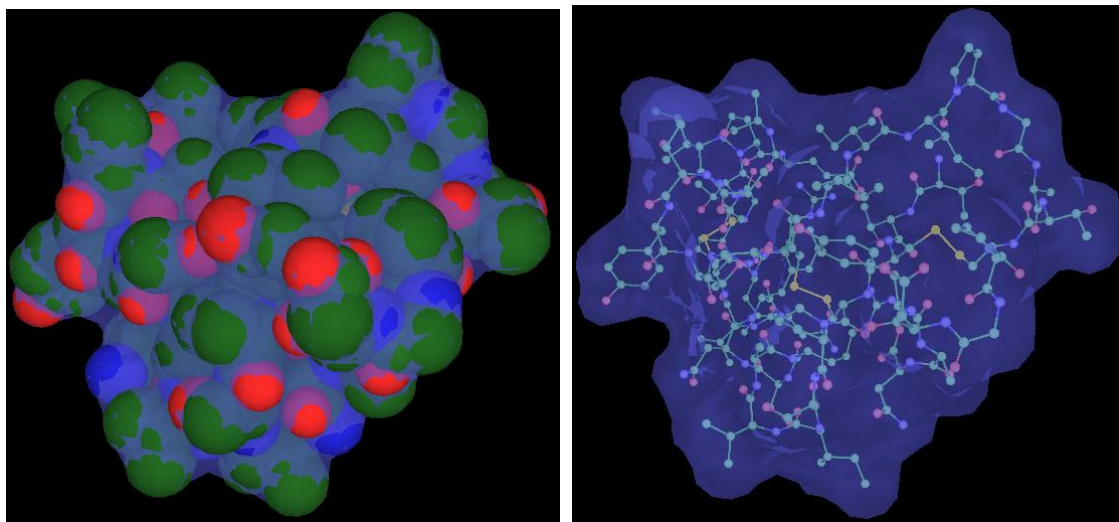
Podatci dobiveni Marching Cubes algoritmom su vrlo slični onima nastalima u postupku određivanja površine molekule. Iz tog razloga su i metode iscrtavanja ovih prikaza vrlo slične, s tom razlikom što kod prikaza rekonstruirane površine nije uključena prozirnost. Osim toga, rekonstruirana površina ne mijenja svoju boju kao što to čini površina molekule u slučajevima odabira različitih načina bojanja, zato što je objekt učitani na ovaj način neovisan o molekuli učitanoj iz PDB datoteke, te je nemoguće dobiti potrebne informacije za ostvarivanje ove radnje. Na slici 13. vidi se izgled rekonstruirane površine sa učitanoj grida.



Slika 14. Prikaz površine rekonstruirane Marching Cubes algoritmom.

5.6.6. Superponiranje prikaza

Kao što je već spomenuto, svi načini prikaza mogu se međusobno superponirati radi mogućnosti dobivanja više korisnih informacija. No, usprkos tome što su superponirani, svi prikazi nisu uvijek vidljivi. Primjer za to je ako se recimo odabere žičani i prikaz kuglama i štapićima. Prostorni modeli prikaza kuglama i štapićima prekrivaju cijelo područje oko spojnica žičanog prikaza, te tako on nije vidljiv. Slike 15. i 16. prikazuju površinu molekule superponirane kalotnom prikazu, te prikazu kuglama i štapićima.



Slika 15. Površina i kalotni prikaz superponirani. Slika 16. Površina superponirana kuglama i štapićima.

5.7. Informacije o atomima, aminokiselinama i lancima

Jedna od mogućnosti koje nudi VMol je prikazivanje informacija o atomima, aminokiselinama i molekulama. Informacije se dobivaju ovisno o odabranom atomu na trenutnom prikazu najviše razine. *Razina prikaza* označuje način određivanja željenog atoma (od najmanje važnosti do najveće):

1. žičani prikaz
2. prikaz kuglama i štapićima
3. kalotni prikaz

Iz prikaza rekonstruirane površine ne mogu se dobiti dodatne informacije, a prikaz se površine molekule koristi samo za određivanje sferno-polarnih značajki, te stoga oni nisu navedeni među razinama prikaza.

Ako se zamisli molekula prikazana pomoću kugli i štapića, te na nju superponiran kalotni model, može se uočiti da nisu vidljiva potpuno ista područja na ekranu. Kako je kalotni model prostorno popunjujući model, on zauzima veću površinu u prikazu, te onemogućuje da se iza njega vide drugi atomi (za razliku od prikaza kuglama i štapićima). To znači da se u ta dva slučaja neće nužno dobiti podatci o istom atomu ako je mišem odabrana ista točka na ekranu. Slična usporedba vrijedi i za model kuglama i štapićima te žičani prikaz, odnosno kalotni i žičani prikaz. Iz tog razloga postoje razine prikaza – one određuju koji od trenutno prikazanih modela najviše popunjuje prostor oko pojedinog atoma.

Sama izvedba biranja nekog iscrtanog objekta odvija se pomoću ugrađenih OpenGL-ovih funkcija za odabir (eng. *picking*) iscrtanog objekta. Radi boljeg uvida u trenutni odabrani atom, oko njega se iscrtavaju bridovi bijele kocke koja ga omeđuje. Također, kod prikaza informacija o aminokiselinama i lancima moguće je uključiti iscrtavanje omeđujućih kocaka za sve atome koji pripadaju toj aminokiselini, odnosno lancu. U tom slučaju, kocke će biti označene odgovarajućim bojama (isto kao i pripadajući atomi).

Informacije o pojedinom atomu koje se mogu dobiti preuzete su iz PDB datoteke učitane molekule, a uključuju:

- ime atoma,
- redni broj u PDB datoteci,
- aminokiselina kojoj pripada atom,
- lanac kojem pripada,
- identifikator segmenta atoma,
- kemijski element atoma,
- naboj,
- X, Y i Z koordinate atoma,
- zauzetost,
- temperaturni faktor,
- radijus atoma.

Podatci o određenoj aminokiselini mogu se dobiti odabirom jednog od atoma koji joj pripada, i također su dobivene učitavanjem PDB datoteke. Oni obuhvaćaju:

- ime aminokiseline,
- redni broj u PDB datoteci,
- broj atoma u toj aminokiselini,
- lista svih atoma koji pripadaju toj aminokiselini.

Detalji o odabranom lancu su slični informacijama koje se mogu dobiti za aminokiseline. Odabir lanca se također vrši odabirom nekog atoma koji mu pripada. O pojedinom lancu mogu se saznati:

- oznaka lanca,
- broj atoma u lancu,
- lista svih atoma koji pripadaju tom lancu.

5.8. Sferno-polarne značajke

Informacije o učitanoj molekuli obuhvaćaju i dobivanje njezinih sferno polarnih značajki za odabranu točku, kut elevacije i kut inklinacije. Točka se odabire na prikazu površine molekule pritiskom miša na željeno mjesto. Njena pozicija se aproksimira sa središtem poligona koji bio određen pritiskom miša. Testiranje da li je neki poligon odabran vrši se na sličan način kao i označavanje atoma na drugim prikazima, korištenjem OpenGL-ovih mogućnosti.

Sferno-polarne značajke prikazuju se kao dio površine označen crvenom bojom. Koji njeni dijelovi će biti obojani u crveno određuje se ispitivanjem svih točaka površine. One točke koje se nalaze unutar granica zadanih kutova elevacije i inklinacije u odnosu na odabranu točku označuju se crvenom bojom, a sve ostale plavom. Kako svi vrhovi pojedinog poligona ne moraju nužno ulaziti u prostor definiran zadanim kutovima, njegovi vrhovi imaju različite boje. Time se postiže efekt interpolacije boja, što aproksimirani prikaz čini točnijim.

Elevacija i inklinacija su Eulerovi kutovi koji određuju rotacije oko Y , odnosno X osi [21]. Kako bi se mogao odrediti koliki su ovi kutovi između neke dvije točke modela, potrebno ih je projicirati na XZ (za elevaciju), te YZ ravninu (za inklinaciju). Projekcija na XY ravninu se ostvaruje izjednačavanjem Y koordinate s nulom, dok je kod projekcije na YZ ravninu X koordinata jednaka nuli. Iz tih

projekcija tada je moguće odrediti kutove skalarnim produktom. U skalarni produkt ulaze vektori koji su određeni između stvarnih i projiciranih točaka, te koordinata ishodišta. Jednadžbe (3) - (8) opisuju postupak računanja kuta elevacije (α) i inklinacije (β) između dviju točaka površine:

$$\alpha_1 = \arccos\left(\frac{T_1 \cdot T_{1XZ}}{|T_1| \cdot |T_{1XZ}|}\right) \quad (3)$$

$$\alpha_2 = \arccos\left(\frac{T_2 \cdot T_{2XZ}}{|T_2| \cdot |T_{2XZ}|}\right) \quad (4)$$

$$\beta_1 = \arccos\left(\frac{T_1 \cdot T_{1YZ}}{|T_1| \cdot |T_{1YZ}|}\right) \quad (5)$$

$$\beta_2 = \arccos\left(\frac{T_2 \cdot T_{2YZ}}{|T_2| \cdot |T_{2YZ}|}\right) \quad (6)$$

$$\alpha = \alpha_1 - \alpha_2 \quad (7)$$

$$\beta = \beta_1 - \beta_2 \quad (8)$$

α_1 i α_2 predstavljaju kutove između stvarne točke (T_1 , odnosno T_2) i točke projicirane na XZ ravninu (T_{1XZ} , odnosno T_{2XZ}), a β_1 i β_2 su kutovi između stvarne točke i točke projicirane na YZ ravninu (T_{1YZ} , odnosno T_{2YZ}).

Jedna od dvije ispitne točke je u svakom slučaju točka odabrana na površini molekule, dok je druga točka ona koja se trenutno ispituje. Ako su α i β za tu točku unutar granica zadanih kutova, tu točku trebamo označiti crvenom bojom.

5.9. Odabir bojanja

U VMol-u postoji ukupno pet različitih načina bojanja prikaza. Slijede njihovi opisi i pojašnjenja:

- "Default" - ovo je standardni način. prikazuje površinu molekule jednoliko obojanu istom bojom, dok su u ostalim načinima prikaza atomi i veze među njima obojani uobičajenom predodređenom bojom za određeni kemijski element.

- "Atomi" - za razliku od "Default" načina, dijelovi površine molekule prikazani su bojama koje odgovaraju uobičajenim bojama atoma kojem taj dio površine pripada.
- "Aminokiseline" - na ovaj način atomi se bojaju prema aminokiselini kojoj pripadaju. Svakoj aminokiselini određuje se boja prema ključu, koji je zapravo ime te aminokiseline. U strukturi PDB-a, za svaki atom postoji zapis kojoj aminokiselini pripada, a on je dugačak točno tri znaka. Pomoću njih može se odrediti jedinstvena boja za svaku aminokiselinu, tako da se *ASCII* kod (eng. *American Standard Code for Information Interchange*) svakog znaka upotrijebi za određivanje pojedine komponente boje (crvene, zelene i plave). Svaka od njih se određuje tako da se odgovarajućem znaku umanji iznos njegove *ASCII* vrijednosti za iznos vrijednost znaka 'A', te podijeli sa vrijednošću znaka 'z'. Razlog tome je da se ime aminokiseline sastoji od slova, te je stoga potrebno vrijednost svakog znaka "centrirati" oduzimanjem prvog slova abecede (po *ASCII* poretku), te sve skupa skalirati za iznos zadnjeg slova abecede (također po *ASCII* poretku) kako bi vrijednosti komponenta boje bila u granicama 0-1, što je primjenjivo prilikom prikaza molekule. Površina molekule će, kao i kod "Atoma", biti obojana ovisno o tome kojem atomu pripada koji dio površine, no u ovom slučaju se neće koristiti uobičajene boje atoma, nego boje određene upravo objašnjenim postupkom.
- "Lanci" - atomi i površina se boje slično kao i kod načina "Aminokiseline", ali u ovom slučaju ovisno o lancu kojem pripadaju. Oznaka pojedinog lanca sastoji se od samo jednog znaka, što nije dovoljno kako bi se generirale komponente boje na ranije opisani način. Zato se ovaj znak koristi kao *indeks* boje, koji ima vrijednost između 0 i 255. Zbog toga postoji upravo $\sqrt{256}$ različitih kombinacija za svaki parametar boje. Zbog potrebe za korištenjem operacije modulo prilikom računanja, broj kombinacija jedne boje zaokružujemo na prvi veći cijeli broj, tj. 7. Ako sada dekadsku vrijednost *ASCII* koda oznake lanca pretvorimo u sustav s bazom 7, dobit ćemo točno tri znamenke, koje skalirane daju vrijednosti parametara boje pojedinog lanca.

- "Sferno-polarne značajke" - atomi se označavaju uobičajenim bojama, dok se površina na mjestima koje ne zadovoljavaju zadane uvjete za kutove i odabranu točku označava plavom bojom (kao kod "Default" načina), a u protivnome crvenom.

U VMol-u su boje pojedinih atoma definirane unutar datoteke "colors.ini", čiji jedan redak sadrži sljedeće podatke:

- ime kemijskog elementa
- crvenu, zelenu i plavu komponentu boje u rasponu između 0 i 1

Ova datoteka se može izmjenjivati, te se tako prikaz atoma u VMol-u može prilagoditi vlastitoj želji ili potrebi.

5.10. Grafičko sučelje

Grafičko sučelje je namijenjeno za mogućnost interaktivnog rada s aplikacijom. Ono omogućuje jednostavno mijenjanje opcija, podešavanje prikaza, te dobivanja raznih informacija. Potpuni izgled grafičkog sučelja nalazi se u datoteci "vmol.mnu", koja predstavlja skriptu namijenjenu opisu sučelja. Format te skripte sadrži podatke o svakom prozoru, te objektima na njemu. Ovako izgleda jedan redak datoteke u ovom formatu:

```
<ključna_riječ> [opcija1="vrijednost1"] [opcija2="vrijednost2"]...
```

Ključna riječ može biti jedna od sljedećih: "form" (prozor), "button" (gumb), "checkbox" (kućica s kvačicom), "option" (okrugli gumb za odabir opcija), "textbox" (kućica za unos teksta) i "picture" (slika). Neke od opcija koje se mogu koristiti su: "name" (ime objekta), "left" (njegov smještaj s obzirom na lijevi rub prozora VMol-a), "top" (smještaj s obzirom na vrh prozora), "width" (širina), "height" (visina), "text" (tekst), "visible" (vidljivost objekta), itd.

VMol učitava ovu datoteku redak po redak, a trenutno učitani redak odmah parsira. Prvo se provjerava ključna riječ. Ako je ona jednaka "form", odmah se proširuje memorija potrebna za stvaranje novog izbornika. U bilo kojem drugom

slučaju, zadani objekt se dodaje na zadnju postavljenu formu. Nakon uspješnog stvaranja prozora, učitani izbornik se može prikazati.

Funkcionalna izvedba samog sučelja ostvaruje se u VMol-u. Kada se detektira interakcija s nekim od objekata, program odmah poduzima određene radnje (kao što je npr. učitavanje PDB datoteke ako se pritisne prvi gumb "OK" na formi "Prozori" ili mijenjanje stanja objektima nakon što su pritisnuti).

Grafičko sučelje podijeljeno je na nekoliko izbornika. Sada će biti dan kratak opis svih svakog od njih.

5.10.1. Glavni izbornik

Glavni izbornik sadrži popis svih postojećih prozora u programu, te omogućuje njihovo uključivanje i isključivanje iz prikaza. Također, preko njega se može jednostavno učitati nova PDB ili grid datoteka u svakom trenutku. Za razliku od ostalih izbornika, ne može se pomicati niti isključiti.



Slika 17. Glavni izbornik.

5.10.2. Opcije prikaza

Pomoću ovog prozora mogu se uključiti/isključiti pojedini načini prikaza molekula, te odabrati koji će model iz učitanoj PDB-a biti prikazan. Osim toga, ovdje je moguće i zadati proizvoljnu isovrijednost, te omogućiti dobivanje informacija o molekuli. Također, moguće je uključiti i opciju da se prikaz vrti oko svojih osi.



Slika 18. Izbornik sa opcijama prikaza.

5.10.3. Info – Atom

Na ovom prozoru mogu se vidjeti sve dostupne informacije o odabranom atomu. Odabir se vrši pritiskom miša na željeno mjesto u prikazu molekule, uz prethodno upaljenu opciju "Info" na prozoru "Opcije prikaza".



Slika 19. Izbornik sa informacijama o odabranom atomu.

5.10.4. Info – Aminokiselina

Ovaj prozor sadrži sve informacije o aminokiselini kojoj pripada odabrani atom.

Slika 20. Izbornik sa informacijama o odabranoj aminokiselini.

5.10.5. Info – Lanac

Na slici 21. vidljiv je izbornik sa svim podacima o lancu kojem pripada odabrani atom.

Slika 21. Izbornik sa informacijama o odabranom lancu.

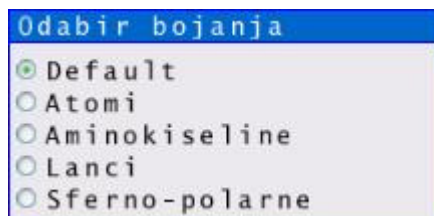
5.10.6. Sferno-polarni kutovi

Odabir minimalnog i maksimalnog kuta elevacije i inklinacije izvršava se u izborniku "Info – Sferno-polarni" (slika 22.).

Slika 22. Izbornik za postavljanje opcija sferno-polarnih značajki.

5.10.7. Odabir bojanja

U izborniku na slici 23. moguće je odabrati jedan od pet ranije opisanih načina bojanja atoma i površine molekule.



Slika 23. Izbornik sa odabirom opcija bojanja.

6. Način korištenja alata VMol

Program VMol namijenjen je korištenju pod *Windows* operativnim sustavom. Pokretanjem programa VMol otvorit će se crni prozor sa prikazanim glavnim izbornikom. Zbog preglednosti, svi drugi izbornici su početno isključeni, ali se mogu dobiti jednostavnim klikom na kućicu kraj njihovog imena u glavnom izborniku. Kada se u kućici nalazi zelena strelica, željeni izbornik je vidljiv, dok u protivnome nije.

Učitavanje PDB datoteke radi se na sljedeći način:

1. mišem se odabere kućica za unos teksta ispod oznake "PDB:", te se preko tipkovnice unese putanja do željene datoteke
2. pritiskom na tipku "OK" pokraj upravo unesene putanje, učitava se željena datoteka.

Na analogan način može se učitati i željeni grid pomoću njemu namijenjene tekst kućice i gumba "OK".

Drugi način učitavanja ulaznih podataka je pokretanjem programa iz konzole. U tom slučaju, uz poziv imena programa mogu se navesti i dodatne opcije:

```
vmol [-p putanja_pdb_datoteke] [-g putanja_grid_datoteke]  
      [-gi putanja_grid_datoteke referentna_isovrijednost]
```

Parametri u uglatim zagradama su proizvoljni, i ne moraju se nužno navoditi prilikom pokretanja programa.

Jednom kada su podatci učitani, moguće je manipulirati prikazom. Za manipulaciju se koriste postavke:

- lijevom tipkom miša (ako je isključena "Info" opcija u izborniku "Opcije prikaza") prikaz se pomiče u željenom smjeru
- desnom tipkom miša prikaz se rotira
- pritiskom na tipku *SPACE* i istovremenim korištenjem lijeve tipke miša, prikaz se može približiti, odnosno udaljiti
- ako je uključena opcija "Info", lijeva tipka miša služi za odabir željenog atoma ili točke na površini, ovisno o trenutnom prikazu

Uključivanje i isključivanje pojedinih izbornika ostvaruje se postavljanjem kvačice kraj istoimene opcije u izborniku "Opcije prikaza". U istom izborniku, pokraj opcije "Grid" nalazi se tekstualna kućica u koju je moguće upisati iznos referentne isovrijednosti grida. Pritiskom na gumb "OK" ili tipkom enter, površina se ponovno rekonstruira sa učitanoj grida, ali za novu zadanu vrijednost. Ispod oznake "Modeli" u istom prozoru, moguće je odabrati koji će se model iz trenutno učitanoj PDB-a prikazivati.

U izborniku "Info – Atom" sva su tekstualna polja namijenjena samo za čitanje, te se preko tog izbornika mogu samo primiti informacije. Za razliku od njega, "Info – Aminokiseline" i "Info – Lanac" sadrže i opciju "Označi atome", čime se u prikazu molekule mogu uključiti bridovi kocaka koje ih okružuju. Osim toga, u ova dva izbornika moguće je pregledavati sve atome navedene u postojećim listama.

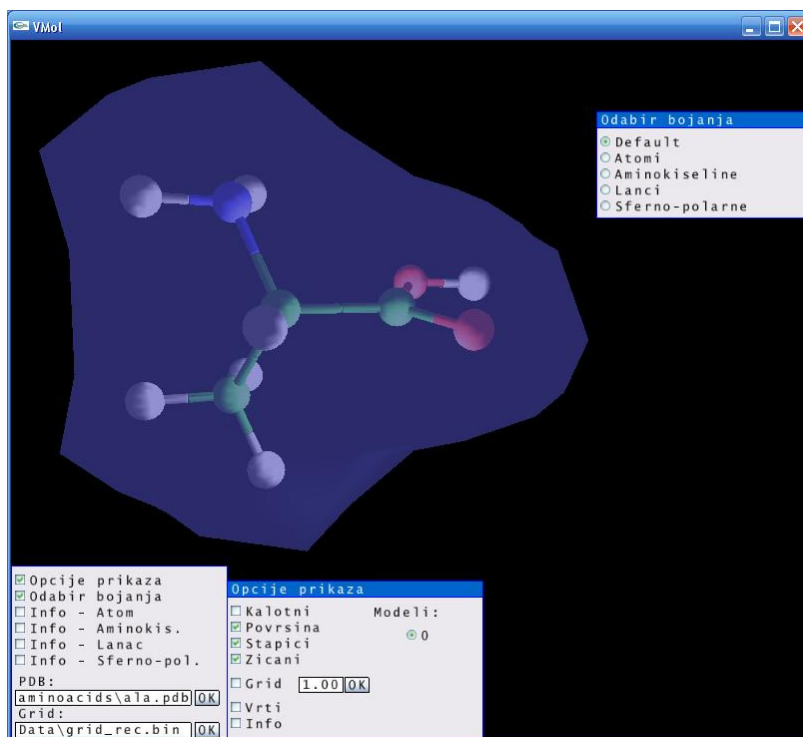
Izbornik "Info – Sferno-polarni" omogućuje podešavanje kutova kod određivanja sferno-polarnih značajki za odabranu točku. Željeni iznosi kutova se unose tako da se jednostavno upišu u ponuđena tekstualna polja.

U slučaju da je potrebno promijeniti način bojanja atoma, aminokiselina, molekula i površine, to možemo učiniti u izborniku "Odabir bojanja".

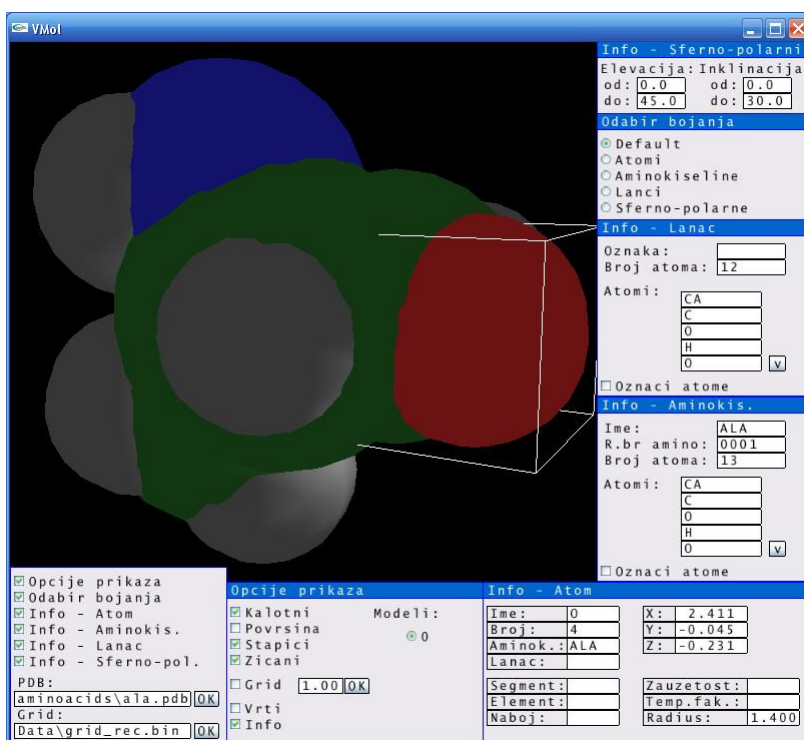
Svi prikazani izbornici, osim glavnog, mogu se pomicati po cijelom prozoru programa, i to klikom na plavu traku na vrhu nekog izbornika, te povlačenjem miša u željenom smjeru. Ako je nužno sve izbornike dovesti u početni položaj, može se koristiti tipka *TAB*. Njenim pritiskom VMol ponovno učitava skriptu sa sadržajem izbornika, te sve opcije postavlja na početne vrijednosti.

7. Prikazi testiranja alata

Slike 24. i 25. prikazuju aminokiselinu alanin:

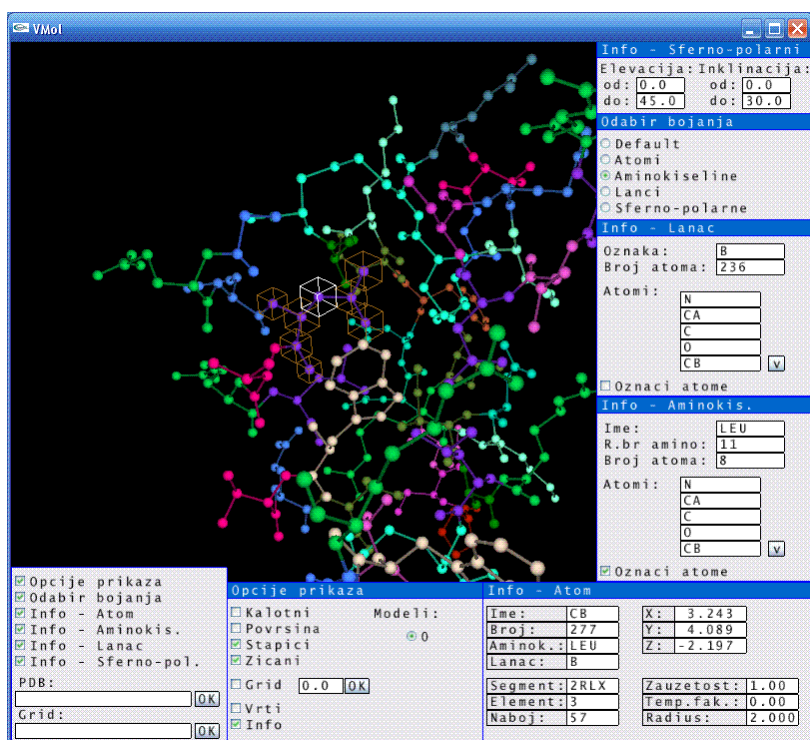


Slika 24. Prikaz površine superponirane modelu kugli i štapića.

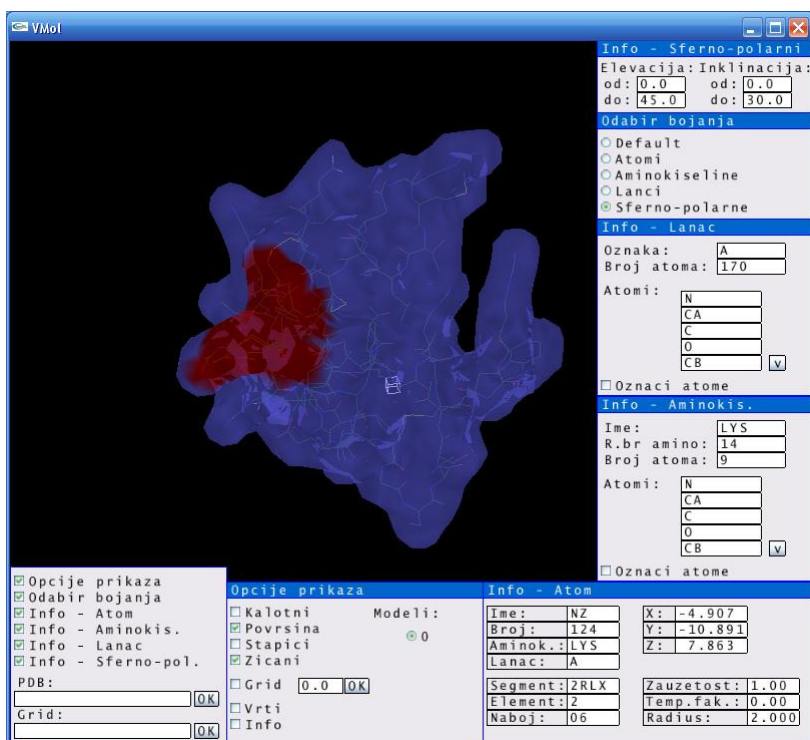


Slika 25. Kalotni model uz označen atom kisika i prikaz informacija.

Na slikama 26. i 27. dan je prikaz kompleksnije molekule, Relaxina:

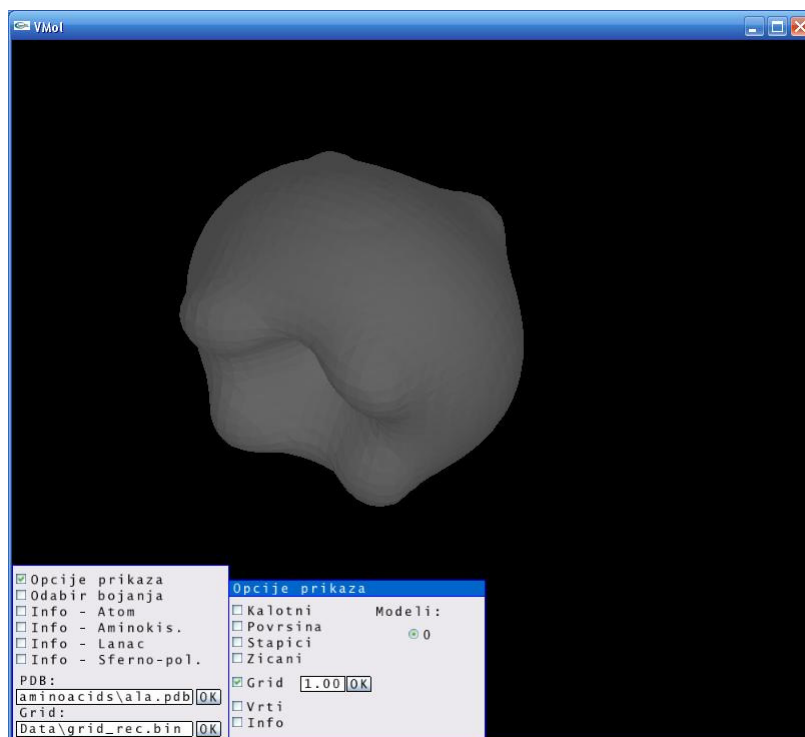


Slika 26. Prikaz molekule bojanjem aminokiselina te označavanjem atoma aminokiseline.

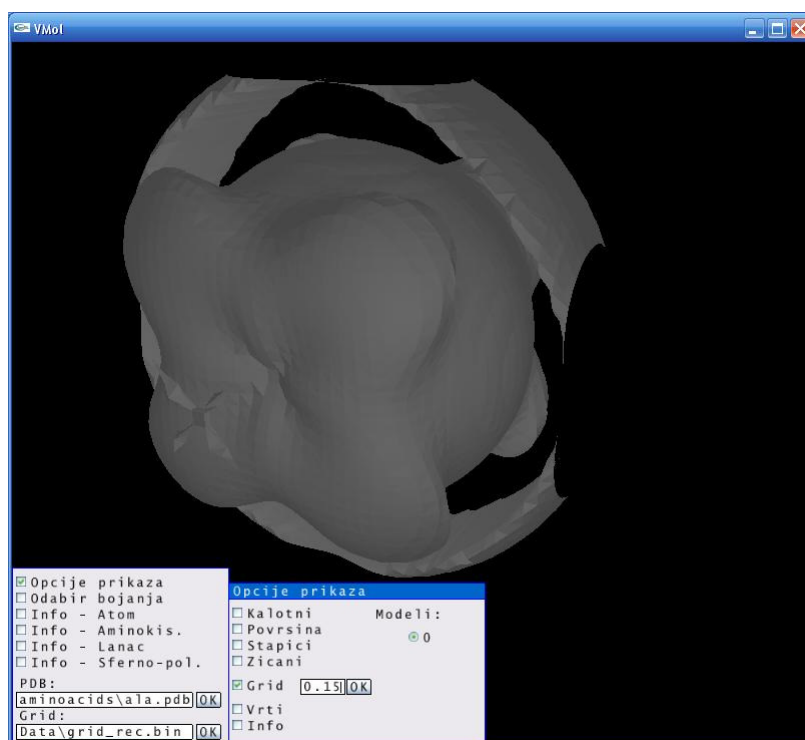


Slika 27. Sferno-polarne značajke na površini molekule.

Testiranje rekonstruirane površine iz učitanoj grida za različite referentne isovrijednosti prikazano je na slikama 28. i 29.:



Slika 28. Rekonstruirana površina za automatski određenu referentnu isovrijednost (1.0).



Slika 29. Prikaz iste rekonstruirane površine za drukčiji iznos referentne isovrijednosti (0.15).

8. Zaključak

Alat za vizualizaciju makromolekula nudi funkcionalnosti prikazivanja modela molekula, manipulaciju tim prikazom, te dobivanje korisnih informacija o atomima, aminokiselinama, lancima i sferno-polarnim značajkama površine. U postupku testiranja alata, čiji se primjeri mogu vidjeti u sklopu ovog rada, utvrđeno je da on ispravno obrađuje i prikazuje podatke iz učitane PDB datoteke, te izvodi kvalitetnu rekonstrukciju površine modelirane sfernim harmonicima. Ispravnost je utvrđena usporedbom sa rezultatima dobivenim od sličnih programskih rješenja, za iste ulazne podatke.

Ovaj program je napravljen kao samostojeća aplikacija, ali je zamišljena tako da nudi i mogućnost povezivanja s drugim programima. To je ostvareno mogućnošću pokretanja iz naredbenog retka, kada mu se mogu predati parametri s ulaznim podacima.

Primjena ovakvog alata može biti dosta široka: od jednostavnog pregledavanja molekula, pa do korištenja u svrhe bržeg i kvalitetnijeg izvođenja procesa prijanjanja proteina, te razmatranje rezultata tog postupka. U svakom slučaju, alati ovakvog tipa su neophodni u postupcima analize makromolekula.

U daljnjem radu na ovom projektu, program bi bilo dobro proširiti mogućnošću manipulacije i prikaza više proteina odjednom, kako bi se mogli vidjeti njihovi međusobni odnosi, te podobnosti pojedinih dijelova površine molekula za njihovo prijanjanje. Time bi se cijeli postupak prijanjanja proteina mogao učiniti interaktivnijim, preciznijim i bržim.

Literatura

- [1] *Ball-and-stick model*, <http://encyclopedia.thefreedictionary.com/ball-and-stick+model>, datum pristupa: 11.06.2008.
- [2] *Covalent bond*, <http://encyclopedia.thefreedictionary.com/covalent+bond>, datum pristupa 11.06.2008.
- [3] 36764.JPG, *36764*, <http://www.dkimages.com/discover/previews/741/36764.JPG>, datum pristupa: 11.06.2008.
- [4] *CPK Color Scheme*, <http://rcsb.icm.edu.pl/oca-docs/cpk.html>, datum pristupa: 11.06.2008.
- [5] *Space-filling model*, <http://encyclopedia.thefreedictionary.com/Calotte+model>, datum pristupa: 11.06.2008.
- [6] octane.qutemol.png.~.jpg, *Isomer of octane*, <http://img2.tfd.com/wiki/0/0d/octane.qutemol.png.%7e.jpg>, datum pristupa: 11.06.2008.
- [7] image020.jpg, *Image020*, http://www.fit.vutbr.cz/~spanel/segmentace/index_soubory/image020.jpg, datum pristupa: 11.06.2008.
- [8] *Introduction to the RCSB PDB Web Site*, http://www.rcsb.org/pdb/static.do?p=general_information/getting_started/index.html
- [9] *PDB File Format*, http://www.rcsb.org/pdb/static.do?p=file_formats/pdb/index.html, datum pristupa: 11.06.2008.
- [10] format3.1-20080211.pdf, *Protein Data Bank Contents Guide: Atomic Coordinate Entry Format Description*, <http://www.wwpdb.org/documentation/format3.1-20080211.pdf>, datum pristupa: 11.06.2008.
- [11] *PDB guide*, datum nastanka: 25.10.1996., <http://structure.usc.edu/pdb/>, datum pristupa: 11.06.2008.
- [12] *Reviews glossary*, http://www.nature.com/nrg/journal/v5/n4/glossary/nrg1317_glossary.html, datum pristupa: 11.06.2008.

- [13] E. Larry Lidz, *vi Editor FAQ – Glossary of Terms*,
<http://people.cis.ksu.edu/~bhoward/vi/vi90.html>, datum pristupa: 11.06.2008.
- [14] Michael Sanner, *Molecular Surfaces Computation*, datum nastanka:
05.02.1996., http://www.scripps.edu/~sanner/html/msms_home.html, datum
pristupa: 11.06.2008.
- [15] Anthony Carpi, *Atomic Bonding*, <http://web.jjay.cuny.edu/~acarpi/NSC/5-bonds.htm>, datum pristupa: 11.06.2008.
- [16] *Size of Atoms*,
<http://chemed.chem.purdue.edu/genchem/topicreview/bp/ch7/size.html>,
datum pristupa: 11.06.2008.
- [17] *Voxel terrain engine - Building the terrain*,
<http://www.codermind.com/articles/Voxel-terrain-engine-building-the-terrain.html>, datum pristupa: 11.06.2008.
- [18] James Sharman, *The Marching Cubes Algorithm*,
<http://www.exaflop.org/docs/marchcubes/>, datum pristupa: 11.06.2008.
- [19] marchingSEx.gif, *marchingSEx*,
<http://www.polytech.unice.fr/~lingrand/MarchingCubes/resources/marchingSEx.gif>, datum pristupa: 11.06.2008.
- [20] cubes1.gif, *The 15 Cube Combinations*,
<http://www.exaflop.org/docs/marchcubes/cubes1.gif>, datum pristupa:
11.06.2008.
- [21] Martin Baker, *Maths - Euler Angles*,
<http://www.euclideanspace.com/maths/geometry/rotations/euler/index.htm>,
datum pristupa: 11.06.2008.

Sažetak i ključne riječi

Naslov rada: Vizualizacija makromolekula

Brzim razvojem računala te znatnim porastom baze podataka poznatih struktura proteina, svoju veliku primjenu našli su i programi za određivanje prijanjanja proteina. U sklopu takvih programa nužno je omogućiti uvid u izgled molekula koje proučavamo.

Alat razvijen za ovaj završni rad služi vizualizaciji makromolekula. On omogućuje četiri standardna načina prikaza molekule: kalotni, žičani, model kuglama i štapićima, te prikaz površine molekule. Osim toga, VMol može i rekonstruirati površinu dobivenu sfernim harmonicima. Svi načini prikaza su detaljnije pojašnjeni u poglavljima ovog završnog rada. Ulazne podatke alat dobiva iz zadane PDB datoteke, i datoteke koja sadrži binarni grid. Osim prikaza molekule, VMol daje i informacije o odabranom atomu, aminokiselini, lancu te sferno-polarnim značajkama površine.

Na kraju ovog rada prikazano je i testiranje programa sa različitim ulaznim podacima, čime se pokazuje ispravnost korištenih postupaka, te demonstriraju mogućnosti VMol-a.

Ključne riječi: protein, prijanjanje, molekula, vizualizacija, prikaz, kalotni model, žičani model, kugle i štapići, površina molekule, grid, Marching cubes, rekonstrukcija, CPK, PDB, atom, radijus, aminokiselina, lanac, elevacija, inklinacija.

Title: Visualization of macromolecules

The study and usage of protein docking tools was boosted by the rapid computer development and fast growing number of available protein structures. Within these tools it is necessary to provide the means of observing the appearance of such structures.

Application developed as a part of this work is intended for macromolecular visualization. It provides four standard ways of displaying molecules: calotte, wireframe, balls and sticks and the surface of the molecule. In addition, VMol is able to reconstruct the surface generated with sphere harmonics. All of the available display methods are explained in more detail in the chapters of this work. VMol gets its necessary input data either from a given PDB file, or a file that holds the binary grid. Besides the display of molecules, VMol also provides information about a selected atom, amino acid, chain, and sphere-polar attributes of the surface.

Testing of this application was given near the end of this work. The results shown confirm that the methods used were correct, and also demonstrate various features of VMol.

Keywords: protein, docking, molecule, visualization, display, calotte model, wireframe model, balls and sticks, grid, Marching cubes, reconstruction, CPK, PDB, atom, radius, amino acid, chain, elevation, inclination.