

UNIVERZA NA PRIMORSKEM
FAKULTETA ZA MATEMATIKO, NARAVOSLOVJE IN
INFORMACIJSKE TEHNOLOGIJE

Zaključna naloga

Programi za napovedovanje vezavnih mest na proteinih

(Programs for predicting binding sites on proteins)

Ime in priimek: Urška Zirnstein Bekš

Študijski program: Bioinformatika

Mentor: izr. prof. dr. Janez Konc

Somentor: prof. dr. Dušanka Janežič

Koper, maj 2017

Ključna dokumentacijska informacija

Ime in PRIIMEK: Urška ZIRNSTEIN BEKŠ

Naslov zaključne naloge: Programi za napovedovanje vezavnih mest na proteinih

Kraj: Koper

Leto: 2017

Število listov: 47

Število slik: 25

Število prilog: 4

Število strani prilog: 2

Število referenc: 55

Mentor: izr. prof. dr. Janez Konc

Somentor: prof. dr. Dušanka Janežič

Ključne besede: vezavno mesto na proteinu, programi za napovedovanje vezavnih mest, načrtovanje zdravil

Izvleček:

V diplomski nalogi opišem programe, ki napovejo vezavno mesto na proteinskih strukturah. Vsak od teh programov ima svoje metode, algoritme ter programe za vizualizacijo. Programe ocenim in njihovo delovanje predstavim na primeru proteina mioglobina.

Key words documentation

Name and SURNAME: Urška ZIRNSTEIN BEKŠ

Title of final project paper: Programs for predicting binding sites on proteins

Place: Koper

Year: 2017

Number of pages: 47 Number of figures: 25 Number of appendices: 4 Number of appendix pages: 2 Number of references: 55

Mentor: Prof. Janez Konc, PhD

Co-Mentor: Prof. Dušanka Janežič, PhD

Keywords: binding sites, programs for predicting binding sites, drug discovery

Abstract:

In this work I describe programs which predict binding sites on protein structures. Each of these programs has their own methods, algorithms and programs for visualisation. I run these programs on a small protein myoglobin and evaluate the results.

Zahvala

Najprej bi se zahvalila mentorju izr. prof. dr. Janezu Koncu in somentorici prof. dr. Dušanki Janežič za pomoč, prijaznost in hitro odzivnost.

Hvala profesorjem in asistentom Fakultete za matematiko, naravoslovje in informacijske tehnologije Univerze na Primorskem, ki so mi med študijem posredovali neprecenljivo znanje.

Hvala tudi moji družini, ki me je ves čas podpirala in zaupala vame in mojemu partnerju, ki mi je stal ob strani.

Kazalo vsebine

1	Uvod	1
1.1	Opis problema	1
1.2	Namen in cilji naloge	1
1.3	Pregled vsebine zaključne naloge	1
2	Pregled področja	2
2.1	Aktivno mesto na proteinu	2
2.2	Razvoj novih zdravil s pomočjo računalnika	3
2.2.1	“De novo”	3
2.2.2	Virtualno reševanje	3
2.3	Podatkovne baze	5
2.4	Preverjanje ustreznosti zdravila	5
3	Primeri programov za virtualno reševanje	6
3.1	3DLigandSite	6
3.1.1	Phyre	6
3.1.2	Mammoth	7
3.2	RaptorX	8
3.3	COACH	9
3.3.1	I-TASSER	9
3.3.2	COACH	10
3.4	ProBiS strežnik in ProBiS vtičnik	12
3.4.1	Algoritem maksimalne klike	13
3.4.2	Molekularni sprehod	14
3.4.3	Fleksibilnost proteinov	14
4	Rezultati	16
4.1	3DLigandSite	17
4.2	RaptorX	20
4.3	COACH	23
4.4	ProBiS	24

4.4.1	ProBiS vtičnik za Pymol	26
5	Zaključek	29
	Literatura	30

Kazalo slik

1	Vezavno mesto na proteinu 4CRT, veriga A.	2
2	Potek reševanja na osnovi strukture [2].	4
3	Uporabniški vmesnik Phyre.	7
4	CRF metoda [27].	8
5	Potek delovanja programov I-TASSER in COACH [4].	9
6	TM-site [4].	10
7	S-site [4].	11
8	Metoda ločevanja s SVM [41].	12
9	Shematski potek algoritma ProBiS [5].	13
10	Deoksi hemoglobin A [43].	14
11	Oksi-mioglobin [52].	16
12	3DLigandSite: vhodni podatki.	17
13	3DLigandSite: možni ligandi.	18
14	3DLigandSite: končni rezultat.	19
15	RaptorX: vhodni podatki.	20
16	RaptorX: možni ligandi.	20
17	RaptorX: vrste struktur v proteinu.	21
18	RaptorX: končni rezultat.	21
19	Deli aminokislinskega zaporedja.	22
20	COACH: napoved ligandov.	23
21	ProBiS: najdeni ligandi.	24
22	ProBiS: rezultati.	25
23	ProBiS vtičnik za Pymol.	26
24	ProBiS: vezavna mesta in napovedani ligandi.	27
25	ProBiS pomoč.	28

Seznam kratic

<i>BioLip</i>	Base of protein-ligand interactions
<i>CRF</i>	Conditional Random Fields
<i>E – vrednost</i>	vrednost, s katero ocenimo uspešnost poravnave
<i>FG – MD</i>	Fragment-Guided Molecular Dynamics
<i>GDT</i>	Global Distance Test
<i>JSdivergenca</i>	Jensen–Shannon divergenca
<i>MTTprocedura</i>	Multiple-Template Threading procedura
<i>NCBI</i>	National Center for Biotechnology Information
<i>PDB</i>	Protein Data Bank
<i>PSFM</i>	Position-specific Frequency Matrix
<i>PSSM</i>	Position-specific Scoring Matrix
<i>RMSDmera</i>	Root-mean-square Deviation of atomic positions
<i>SCOP</i>	Structural Classification of Proteins
<i>SSFL</i>	Subsequence from the First binding residue to the Last binding residue
<i>SVM</i>	Support Vector Machines

Kazalo prilog

A Rezultati programov v sklopu programa COACH

1 Uvod

1.1 Opis problema

Napovedovanje vezavnih mest je osnova za razvoj novih zdravil. Vezavna mesta na proteinu so deli proteinov na katerih pride do interakcije z molekulami, ki ob vezavi na protein v telesu sprožijo določen odziv. Dandanes stremimo k hitremu in učinkovitemu razvoju zdravil. Eden izmed izzivov našega časa je porast bakterijske rezistence na večino antibiotikov. Naša rešitev je načrtovanje zdravil s pomočjo računalnika. V prihodnosti bomo lahko namesto navadnega računalnika za iskanje zdravil uporabljali kvantni računalnik. Kvantna tehnologija nam obljublja možnost prikaza kompleksne kvantno-mehanske simulacije molekularnih procesov. Aprila 2016 je Google prvič natančno simuliral molekulo vodika s kvantnim računalnikom. [1]

1.2 Namen in cilji naloge

V diplomski nalogi bom opisala štiri programe, ki napovejo vezavno mesto na proteinskih strukturah. Vsak od teh programov ima svoje metode, algoritme ter programe za vizualizacijo. Programe bom ocenila, izpostavila njihove prednosti in slabosti, ter njihovo delovanje predstavila na primeru proteina.

1.3 Pregled vsebine zaključne naloge

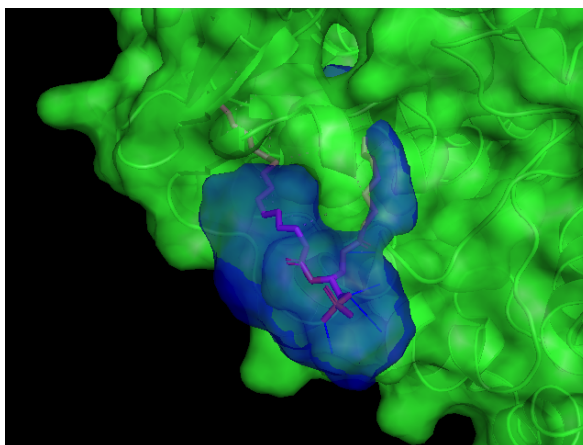
Na začetku je predstavljen potek načrtovanja zdravil z računalniškimi programi. Sledi podroben opis štirih programov 3DLigandSite [2], RaptorX [3], COACH [4] in ProBiS [5], ki z metodo virtualnega rešetanja iščejo strukture, ki so lahko potencialna zdravila. Jedro je primerjava rezultatov omenjenih programov na manjšem proteinu oksimoglobinu.

2 Pregled področja

Računalniško načrtovanje zdravil *in silico* je najbolj učinkovito v začetnih stadijih iznajdbe zdravil v nasprotju s študijami *in vitro* in *in vivo*. *In silico* študija je na primer leta 2009 uporabila programske simulacije za napoved delovanja že obstoječih zdravil v primeru zdravljenja težko ozdravljivega seva tuberkuloze. Uporabili so metodo molekularnega rešetanja, s katero so preiskali baze molekulskih spojin v iskanju potencialnega zdravila [6]. *In silico* metode vključujejo farmakofore, podatkovne baze, homologne modele, strojno učenje, podatkovno rudarjenje, računalniška orodja za analizo podatkov in drugo.

2.1 Aktivno mesto na proteinu

Aktivno mesto je del proteina, na katerega se lahko veže ligand. Navadno je to “žepek” oziroma udrtina na proteinskem površju. Če je ligand manjša molekula je lahko konkavno obrnjen v notranjost proteina, v primeru makromolekul pa je lahko le blago vdrt na površju. [7] Na sliki 1 je prikazano vezavno mesto na proteinu z oznako 4CRT.



Slika 1: Vezavno mesto na proteinu 4CRT, veriga A.

Če predvidimo vezavno mesto na proteinu oziroma proteinu pripadajoče ligande, lahko načrtno sintetiziramo zdravilo. Za to obstaja več načinov.

2.2 Razvoj novih zdravil s pomočjo računalnika

Računalniška tehnologija se zelo hitro razvija, prav tako pa tudi tehnike, s katerimi lahko preučujemo molekularni svet. Računalniške tehnike so vedno bolj uporabne v farmaciji, biokemiji, nanotehnologiji, molekularni biologiji itd.

2.2.1 “De novo”

De novo generacija gradi nove molekule z malimi fragmenti kot so karbonilne, karboksilne, amino skupine, benzenovi obroči itd., ki jih z računalniški orodji pozicionirajo, ocenijo. Končne strukture, ki so bile ustvarjene v računalniškem okolju, so kasneje sintetizirane v laboratoriju. [8]

2.2.2 Virtualno rešetanje

Virtualno rešetanje je visoko zmogljiva računalniška metoda, ki analizira obsežne podatkovne baze kemičnih spojin z namenom iskanja kandidata za zdravilo. [9]

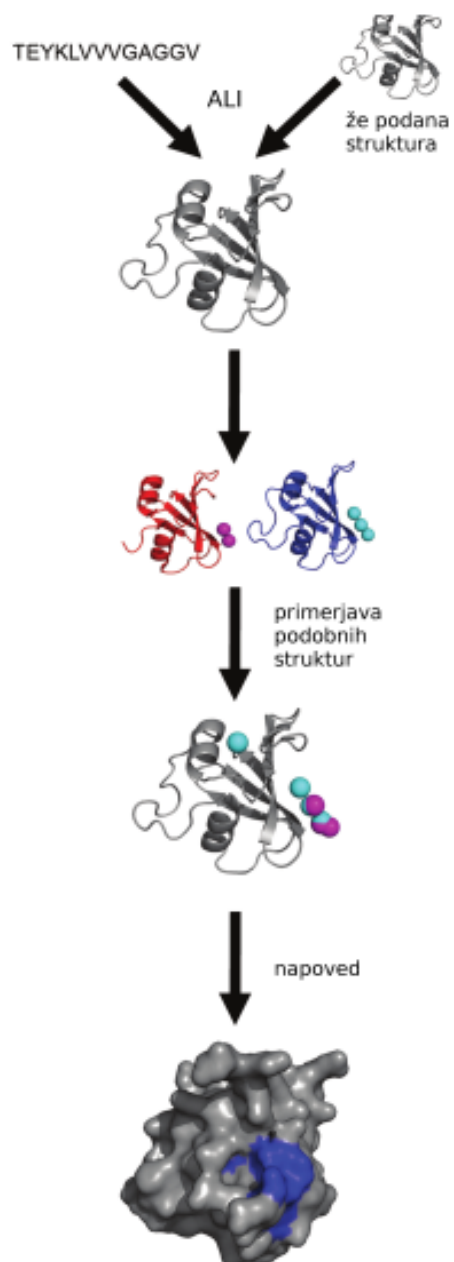
Rešetanje na osnovi liganda

Iskanje na osnovi liganda deluje na principu, da imajo ligandi, ki so podobni primerjavnemu ligandu večjo verjetnost za tvorbo kompleksa kot naključni ligandi. Navadno se ta metoda uporablja, ko nimamo na voljo 3D strukture tarčne molekule. Ta računalniška metoda sestoji iz dveh elementov: uspešne primerjalne mere in zanesljive metode uje-manja (scoring method). [10] Primer: LiSiCA PyMOL vtičnik za Windows in Linux sisteme.

Rešetanje na osnovi strukture

Pri tej metodi je bistvo analiziranje 3D strukture željene proteinske tarče. Deluje na principu, da proteinska vezavna mesta skozi evolucijo ohranjajo strukturo in aminokislinsko zaporedje. [11, 12]. Strukturna ohranjenost naj bi bila celo bolj pogosta [13]. Tudi ko se aminokislinsko zaporedje dveh struktur ne sklada, strukturna podobnost implicira na podobno funkcijo- npr. vezavno mesto. Slabost pri tej metodi je, da imajo lahko proteini kljub podobni strukturi drugačno funkcijo [5, 14]. V nalogi bom predstavila ProBiS vtičnik za PyMOL za Windows in Linux sisteme.

Na sliki 2 je primer iskanja vezavnega mesta na osnovi strukture. Podano imamo aminokislinsko zaporedje oziroma samo strukturo proteina. Dano strukturo primerjamo s podobnimi strukturami in njihovimi ligandi. Na tak način lahko predvidimo vezavno mesto.



Slika 2: Potek rešetanja na osnovi strukture [2].

Strukture makromolekul v atomski resoluciji so lahko določene z:

1. X-ray kristalografijo
2. Nuklearno magnetno resonanco (NMR)
3. Homolognim modeliranjem
4. Nevtronsko spektroskopijo
5. Simulacijo molekulske dinamike

Pogosto je v uporabi homologno modeliranje. V primeru, ko imamo protein s pomanjkljivo strukturo, jo lahko predvidimo, če obstaja homologni protein s katerim si delita vsaj 30% aminokislinskega zaporedja. [15] Tak homologni protein imenujemo predloga [16].

2.3 Podatkovne baze

Programi se v večini primerov opirajo na proteinske baze podatkov. Preko teh baz podatkov pridobimo osnovne podatke o molekuli, ki jo preučujemo na primer kemijska oznaka, opis, aminokislinsko zaporedje, 3D zgradba itd. Primeri: RCSB Protein Data Bank, BioLip, NCBI, SCOP.

2.4 Preverjanje ustreznosti zdravila

Potencialne molekule ocenimo z več testi. Tudi najbolj primerna potencialna molekula se lahko izkaže za neprimerno v nadaljnjih *in vitro* laboratorijskih testih. Kandidati so najprej ocenjeni s programi za vizualizacijo, kjer so lahko še dodatno optimizirani na primer, za povečanje afinitete. Ocenijo se tudi njihova oralna biološka razpoložljivost oziroma učinkovitost v telesu. V nadaljnjem postopku se upošteva tudi kemična in metabolična stabilnost, distribucija, metabolizem, toksičnost ter druge biokemične lastnosti. [8]

3 Primeri programov za virtualno rešetanje

Izbrala sem štiri programe, ki so neplačljivi in dosegljivi vsem. Preučila sem njihovo delovanje, ter jih preizkusila na primeru manjšega proteina.

3.1 3DLigandSite

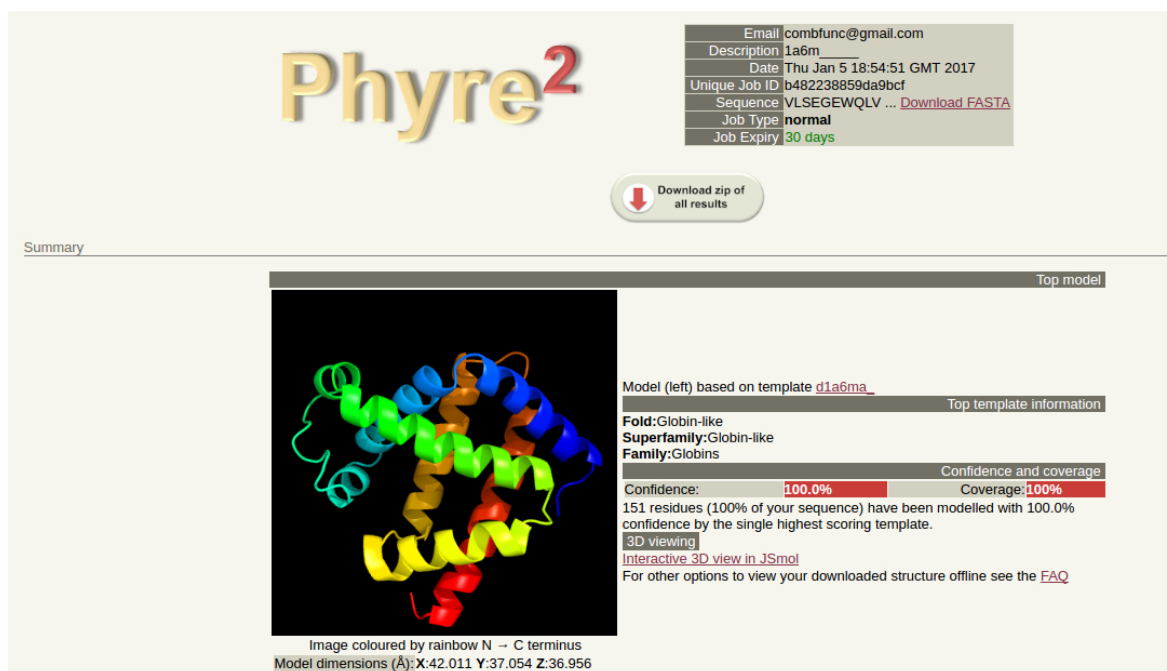
3DLigandSite išče podobne proteinske strukture in njihove pripadajoče ligande preoblikuje v strukturne modele. Na podlagi strukturnega ujemanja napove možno vezavno mesto. [2]

Strežnik za izračun uporabi več programov:

1. PHYRE (PSI-BLAST)
2. MAMMOTH

3.1.1 Phyre

S pomočjo Phyre se generira model proteina, če uporabnik poda le FASTA format. Program uporablja knjižnjico SCOP in PDB. Program Psi-Blast [17] zažene pet iteracij za pridobitev aminokislinskih homologov, končni rezultat je optimalna kombinacija vseh petih. Za napoved sekundarne strukture se zaženejo trije neodvisni programi Psi-Pred [18], SSPro [19] in JNet [20]. Vrednosti vseh treh programov so upoštevane, izračunana je končna srednja vrednost. Na koncu program Disopred [21] določi najbolj verjetne strukturno urejene in neurejene regije. Iz aminokislinskega zaporedja ustvarjen profil in sekundarna struktura sta primerjana z bazo struktur s profil-profil primerjalnim algoritmom. Ta algoritem vrne oceno primerjave. Na osnovi desetih najboljših ocen se ustvari 3D profil strukture. [2, 22] Na sliki 3 je prikazan uporabniški vmesnik Phyre.



The screenshot displays the Phyre2 web interface. At the top, the Phyre2 logo is visible. To the right, a summary box contains the following information:

Email	combfunc@gmail.com
Description	1a6m
Date	Thu Jan 5 18:54:51 GMT 2017
Unique Job ID	b482238859da9bcf
Sequence	VLSEGEWQLV ... Download FASTA
Job Type	normal
Job Expiry	30 days

Below this, a button labeled "Download zip of all results" is present. The main section is titled "Summary" and features a 3D ribbon diagram of a protein structure, colored by rainbow (N-terminus to C-terminus). To the right of the diagram, the following information is provided:

Model (left) based on template d1a6ma_

Top template information

Fold: Globin-like
Superfamily: Globin-like
Family: Globins

Confidence and coverage

Confidence: 100.0%
Coverage: 100%

151 residues (100% of your sequence) have been modelled with 100.0% confidence by the single highest scoring template.

[3D viewing](#)
[Interactive 3D view in JSmol](#)
 For other options to view your downloaded structure offline see the [FAQ](#)

At the bottom of the 3D view, it states: "Image coloured by rainbow N - C terminus" and "Model dimensions (Å): X:42.011 Y:37.054 Z:36.956".

Slika 3: Uporabniški vmesnik Phyre.

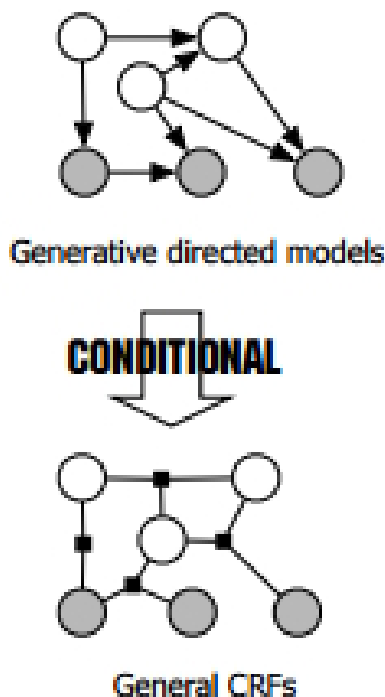
3.1.2 Mammoth

Program Mammoth identificira strukture podobne podani strukturi. Mammoth deluje na principu, da je napoved strukture uspešna, ko je modelirana struktura kar se da bolj podobna iskani strukturi, kot katerikoli drugi strukturi. Za ocenitev rezultatov uporablja P-vrednost. 25 najboljše ocenjenih možnih ligandov se nadaljne analizira. Uspešni so tisti, katerih strukturno ujemanje dosega vrednost večjo od 7. Tam, kjer je največ možnih ligandov, je tudi možno vezavno mesto. [2, 23]

3.2 RaptorX

RaptorX 3D napove strukturo proteina na osnovi prepoznavanja proteinskih zank, vezavno mesto pa s pomočjo strukturne primerjave. [3].

Najprej se za optimiziranje strategije modeliranja zažene profil-entropijska cenilna metoda. Gre za primer homolognega modeliranja. Profil-entropijska cenilna metoda je uporabna v primeru, ko za tarčni protein ne najdemo dovolj dobre predloge. Takrat se metoda bolj opira na strukturno informacijo kot na nenatančno homologno informacijo. [24] Za modeliranje interakcij med proteinskimi strukturnimi deli program uporablja verjetnostni grafični model CRF [25], ki temelji na pogojni verjetnostni porazdelitvi [26]. Na sliki 4 je enostaven primer preoblikovanja generativnega usmerjenega modela v CRF.



Slika 4: CRF metoda [27].

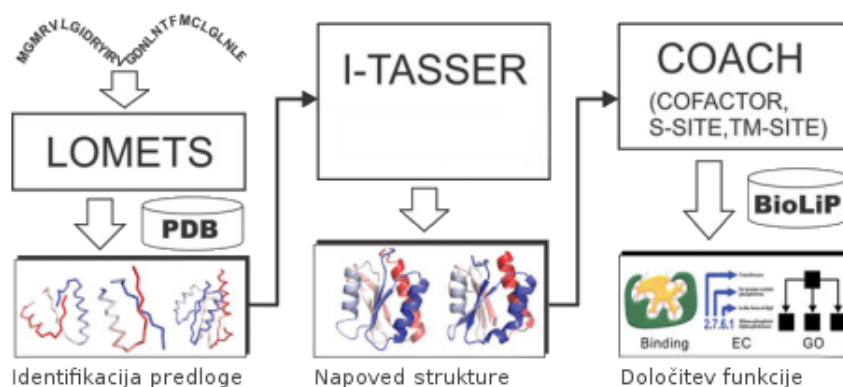
Na koncu je implementirana MTT procedura. Ker ima lahko protein več možnih homolognih predlog, MTT procedura primerja več predlog hkrati in na podlagi večih predlog napove 3D model strukture [3,28]. S predvidenim 3D modelom in strukturno primerjavo program napove možne žepke na površini proteina in pripadajoče možne ligande. [3]

3.3 COACH

Program temelji na ideji, da je rešitev za problem iskanja vezavnih mest kombinacija komplementarnih iskalnih metod, ki ena drugo dopolnjujejo, kar zagotavlja optimalen rezultat [4]. Če uporabnik ne poda 3D strukture proteina, je ta modelirana in analizirana s samostojnim programom I-TASSER [29].

3.3.1 I-TASSER

I-TASSER je hierarhičen pristop za napoved strukture in funkcije proteinov. Na podlagi aminokislinskega zaporedja in primerjave z že znanimi homolognimi strukturami v proteinski bazi (PDB, BioLiP) se določi zgradba proteinov. Sestoji iz dveh korakov; identifikacija strukturne predloge (pri tem uporablja program LOMETS [30]) in strukturna analiza (s pomočjo programov: SPICKER [31], FG-MD [32] in ModRefiner [33]). Določitev funkcije na podlagi strukture poteka preko programa COACH [29]. Njegovo delovanje je skupaj z programom I-TASSER prikazano na sliki 5.



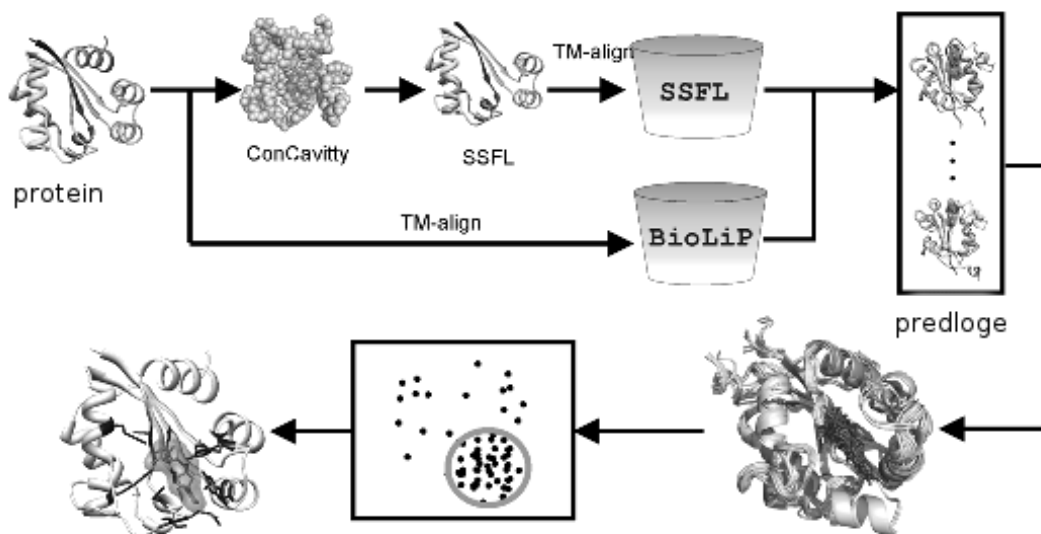
Slika 5: Potek delovanja programov I-TASSER in COACH [4].

3.3.2 COACH

Program COACH ima dve metodi; prva temelji na iskanju na osnovi strukture (TM-SITE), druga pa na primerjavi aminokislinskih zaporedij (S-SITE). TM-SITE išče vezavna mesta s primerjavo strukture z že znanimi vezavnimi mesti (knjižnjica BioLip) oziroma s strukturnimi predlogami, s katerimi program zazna podobnost na lokalnih in globalnih delih proteina. S-SITE je narejen na osnovi evolucijske prepoznavnosti vezavnih mest. Na koncu program COACH združi rezultate obeh programov s pomočjo SVM metode, ki je opisana v nadaljevanju. [4]

TM-SITE

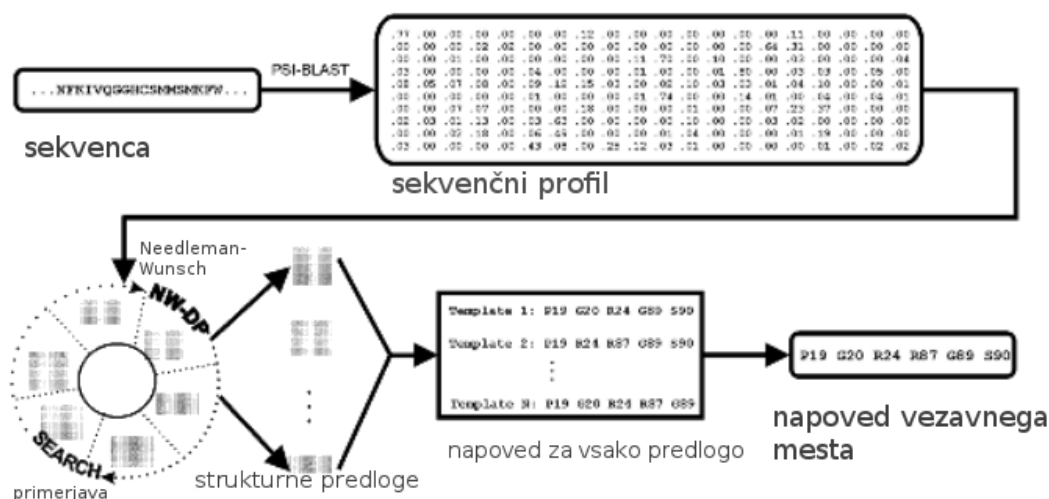
Strukturna primerjava je lahko globalna, kjer primerjamo celotni protein ali lokalna, kjer primerjamo le del proteina. TM-SITE uporablja vmesno možnost, tako imenovano SSFL, ki pregleda proteinski žep od prvega do zadnjega dela. Ta metoda naj bi zagotavljala optimalne rezultate. Predloge za SSFL se nahajajo v SSFL knjižnjici in so izračunane iz BioLip knjižnjice. Za iskanje žepkov na površju strukture se zažene program ConCavity [34]. Zatam se zažene še TM-align [35]. Rezultati TM-align-a so ocenjeni s cenilno funkcijo, ki primerja globalne, lokalne, strukturne in sekvenčne podobnosti. Za končno selekcijo vezavnih mest so izbrani ligandi iz TM-align dani v gručo na podlagi prostorske razdalje med geometričnimi centri na razdalji 4 Å. Najboljši rezultati predstavljajo možne ligande in s tem možna vezavna mesta. Potek programa TM-SITE je prikazan na sliki 6. [4]



Slika 6: TM-site [4].

S-SITE

Program S-SITE, ki temelji na primerjavi proteinskih zaporedij, najprej zažene program PSI-BLAST za obdelavo frekvence prek NCBI baze podatkov za namen konstrukcije aminokislinske sekvenčne poravnave. Ustvari se pozicijsko specifična frekvenčna matrika (PSFM). Za iskanje homolognih predlog iz BioLipa, se matrika PSFM primerja s PSI-BLAST predlogami pozicijsko specifične točkovne matrike (PSSM) v BioLip knjižnici. Pri tem se uporabi Needleman-Wunsch algoritem [36]. Najbolj ustrezni rezultati so obravnavani kot potencialna vezavna mesta kakor lahko vidimo na sliki 7. [4]



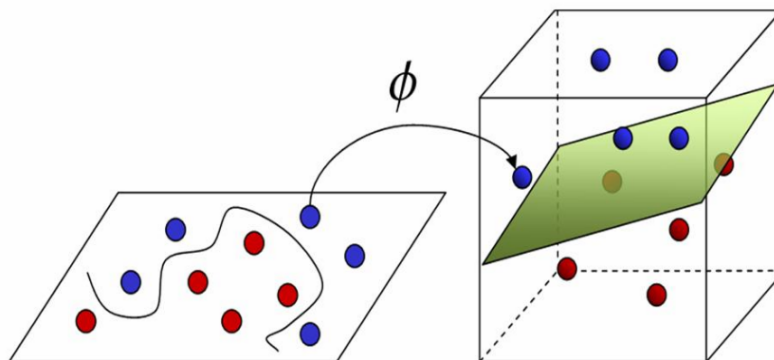
Slika 7: S-site [4].

Kontrolna programa za napoved vezavnih mest sta še FINDSITE [37], ki ravno tako temelji na primerjalnih predlogah in COFACTOR [38]- metoda, ki primerja lokalne in globalne strukturne podobnosti med preučevano strukturo in strukturnimi predlogami. [4]

COACH združuje več programov, ki napovedujejo vezavna mesta na proteinu: TM-SITE, S-SITE, COFACTOR, FINDSITE in ConCavity. Ko podamo iskano aminokislinsko zaporedje strukture, se zaženejo vsi omenjeni programi. [4] Najboljši rezultati programov se nato določijo preko SVM [39], ki vrne najbolj optimalne rešitve. SVM je metoda, ki ločuje podatke v več-dimenzionalnem prostoru. Lastnosti podatkov so preoblikovane v koordinatni prostor navadno višje dimenzije. Določena je maksimalna meja, ki ločuje podatke po dani lastnosti. [40]

SVM

Na sliki 8 je shematski prikaz delovanja SVM. V levem kvadratu so podatki v dvodimenzionalnem prostoru ločeni s krivuljo. Če jih želimo ločiti z ravno črto, jih moramo transformirati.



Slika 8: Metoda ločevanja s SVM [41].

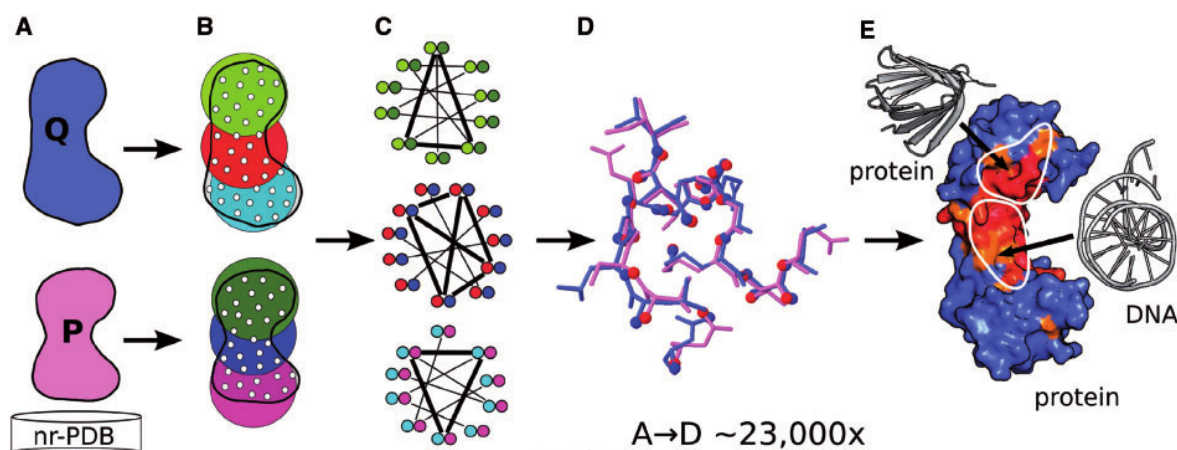
3.4 ProBiS strežnik in ProBiS vtičnik

Ko podamo proteinsko strukturo, ProBiS preišče predloge proteinov v PDB, ki imajo podobno vezavno mesto- gre za postopek molekularnega sidranja. [5] Obstaja tudi ProBiS vtičnik za program Pymol, ki takoj vrne rezultat.

Vezavna mesta so navadno ohranjena v evoluciji proteinov. ProBiS lahko vezavna mesta zazna na podlagi proteinske strukture. Ko podamo strukturo proteina, ProBiS primerja lokalno strukturo danega proteina s podatkovno bazo proteinskih struktur in išče strukturne, geometrijske in fizikalno-kemijske podobnosti. Ko najde potencialno ujemanje izračuna stopnjo ujemanja. Prednost ProBiS algoritma je, da najde podobna vezavna mesta, tudi če ima proteinska struktura drugačno zvitje. Pri tem uporablja algoritem maksimalne klike. [5]

3.4.1 Algoritem maksimalne klike

Klika je podmnožica vozlišč S v grafu, ki je sestavljen tako, da je vsak par vozlišč povezan. Maksimalna klika v omenjenem grafu je klika z največ vozlišči. Algoritem maksimalne klike [42] se precej razlikuje od že obstoječih algoritmov iskanja. V primeru ProBiSa vozlišče predstavlja funkcionalno skupino, povezava pa bližino med vozlišči. Na ta način se zajame geometrijske in fizikalno-kemijske lastnosti površine proteina. [42] Shematski potek algoritma ProBiS je prikazan na sliki 9.



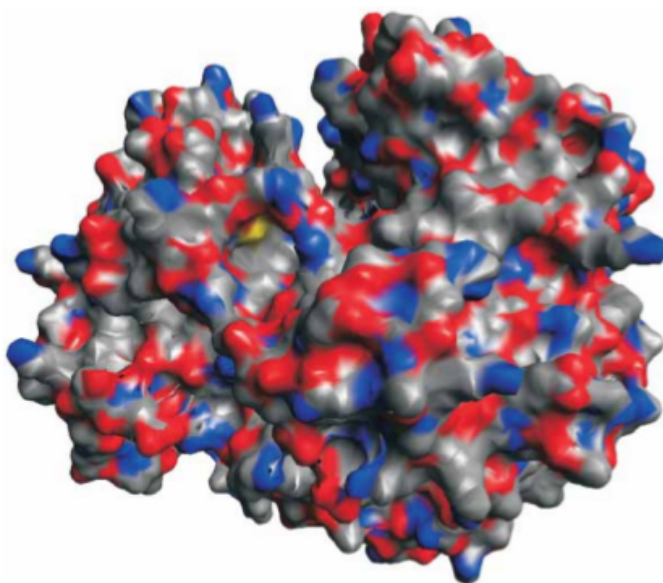
Slika 9: Shematski potek algoritma ProBiS [5].

- A Proteinska struktura (Q) se primerja z približno 23 000 strukturami (P)
- B Proteini, ki so prikazani v obliki grafa se razdelijo na podgrafe. Na podlagi matrike razdalj se izločijo podgrafi, ki si med seboj niso podobni
- C Za vsak podgraf je izračunan skupni graf. Maksimalna klika v skupnem grafu predstavlja največjo podobnost med primerjanimi podgrafi
- D Vsaka maksimalna klika določi strukturno poravnavo med dvema proteinoma. Za vsako primerjavo je izračunana E-vrednost, vektorski koti, RMSD
- E Koraki od A-D se ponovijo za vsak protein iz PDB, rezultati se shranijo v SQL podatkovno bazo [5]

3.4.2 Molekularni sprehod

Za analizo površine proteinov ProBiS uporablja algoritem, ki izračuna površino molekule dostopno topilu [43]. Površina, ki je dostopna topilu, je definirana s sferično sondo, ki se zakotali čez sferične atome [44,45]. Če obstaja površina po kateri se lahko zakotali, potem je ta ista površina dostopna tudi za molekulo topila. Ta metoda se uporablja pri analizi strukture in delovanja proteinov [46], računanje volumna in površine molekule [47], pa tudi za molekularno sidranje [48,49]. V primeru molekularnega sidranja metoda napove površinske žepke in izmeri spremembo energije ob vezavi proteina z neko molekulo. [43,47,50,51]

Na sliki 10 je prikazana molekularna površina človeškega deoksi hemoglobina A, ki vsebuje 4512 atomov. Radij sfere je v tem primeru 1.4 Å.



Slika 10: Deoksi hemoglobin A [43].

3.4.3 Fleksibilnost proteinov

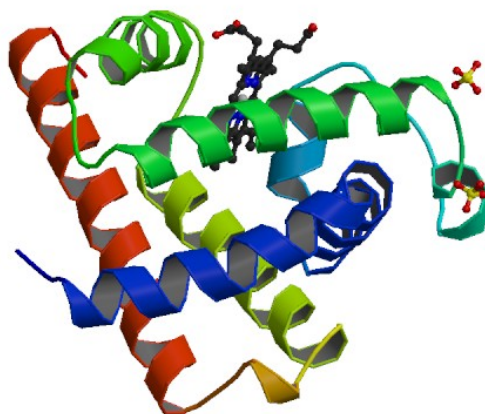
Proteini imajo lahko več konformacij. ProBiS poskuša zaznati tudi druge konformacije s primerjanjem dveh proteinov, pri čemer upošteva naslednje pogoje: razdalja med dvema ogljikovima atomoma med primerjalima proteinoma mora biti manjša od 10 Å, aminokisline v treh delih prvega proteina se morajo ujemati z deli drugega proteina in med podobnimi tremi deli prvega in drugega proteina mora imeti med ogljiki potegnjjen vektor kot manjši od 45 stopinj. [5]

Strukturna poravnava in "prstni odtisi"

Za lokalno poravnavo proBiS uporablja kot površinskega vektorja, RMSD mero, velikost primerjalnega dela proteina, E-vrednost in neposredno lokalno primerjavo dveh proteinov. Prstni odtis proteina je ponavljajoča se 3D struktura značilna za neko vezavno mesto. ProBiS prepozna prstne odtise proteina na podlagi strukturne poravnave. Na ta način lahko bistveno pripomore k hitrejši identifikaciji vezavnega mesta. [5]

4 Rezultati

Programe sem zagnala na manjšem proteinu oksi-mioglobinu s PDB kodo: 1A6M. Na sliki 11 je molekula oksi-mioglobina pridobljenega iz mišičnih vlaken kita glavača.



Slika 11: Oksi-mioglobin [52].

Mioglobin je manjši citoplazemski hemoprotein, ki se nahaja v skletenih mišičnih vlaknih. Od odkritja njegove strukture leta 1958 se trudimo v popolnosti razumeti njegovo strukturo in delovanje [53]. Njegova glavna naloga je kopičenje kisika v celicah. V primerih pomanjkanja kisika, lahko mioglobin kisik tudi sprošča in ga tako naredi dostopnega telesu [54].

4.1 3DLigandSite

Na sliki 12 so vidni podatki podani ob prijavi- elektronski naslov, unikatna številka procesa, datum, način podane strukture (v tem primeru FASTA format), strukturni model in iskanje prek programa Mammoth.

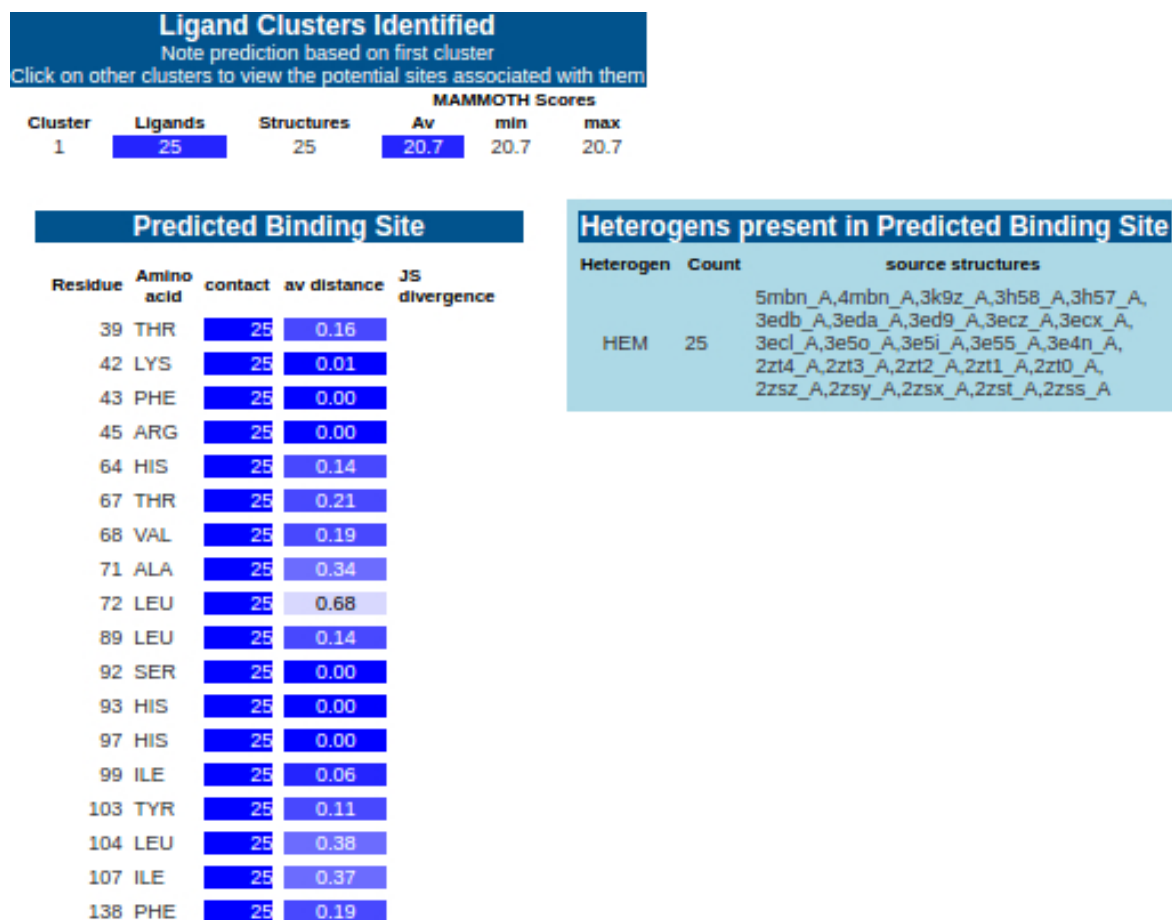
Submission Details	
Email:	urska.zirnstein@gmail.com
Unique Job identifier:	0527de04f333892a
Description:	1A6M
Date:	Mon Oct 24 20:10:07 BST 2016
Submission Type:	sequence
Query Seq:	VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAMKASED LKKHGVTVLTALGAILKKKGHAEELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP GDFGADAQGAMNKALELFRKDIAAKYKELGY

Structural Model	
Phyre2 job:	0527de04f333892a
Phyre2 template:	1a6ma_
Phyre2 confidence score:	100.0

Structural Search	
confidence data from search of structural library with Mammoth	
Average LnE:	20.69
Maximum LnE:	20.69
Min LnE:	20.69

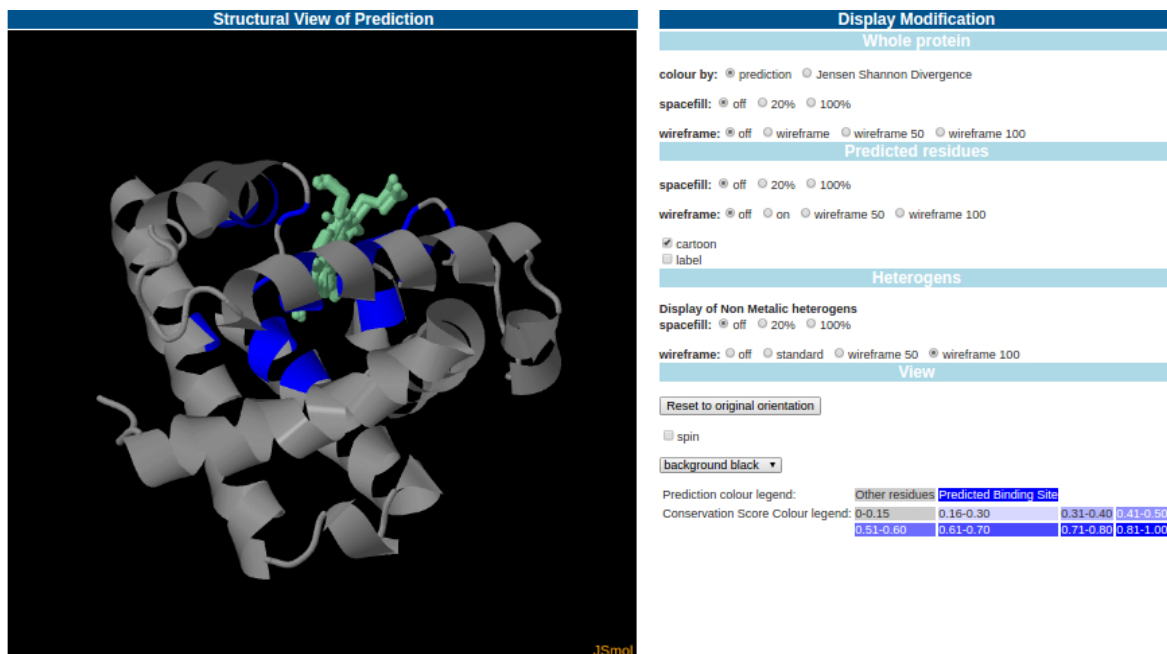
Slika 12: 3DLigandSite: vhodni podatki.

Na sliki 13 so prikazani rezultati Mammoth programa (povprečje, minimum, maksimum-ki so v tem primeru identični). Vidimo, da je našel potencialni ligand HEM. Zabeležene so vse aminokisline, ki so aktivne v vezavnem mestu. Ocenjena je povprečna razdalja, ki je ponazorjena z barvo, od temno do blede modre. Za JS divergenco ni zabeleženo nič.



Slika 13: 3DLigandSite: možni ligandi.

Na sliki 14 je prikazan končni rezultat programa 3DLigandSite, ki je vizualiziran s pomočjo Jsmol-a. Strukturo lahko premikamo in vrtime. Možnost imamo spreminjati določene lastnosti.



Slika 14: 3DLigandSite: končni rezultat.

Za izračun je program v mojem primeru potreboval približno 3 ure. Ko je bila napoved zaključena, program ni poslal elektronskega sporočila na moj elektronski naslov. Ta program ni uspešno napovedal vezavnega mesta, saj je ligand HEM že definiran v FASTA formatu proteina oksihemoglobina z oznako 1A6M.

4.2 RaptorX

Na sliki 15, lahko vidimo oznako preiskovanega proteina (1A6M), označena je tudi veriga. Število 1 pod segmentom pomeni, da je bil le ta modeliran.

1A6M | 1A6M:A|PDBID|CHAIN|SEQUENCE

Section I. Input Sequence and Domain Partition ([help](#))

1	11	21	31	41	51	61	71	81	91
VLSEGEWQLV	LHVWAKVEAD	VAGHGQDILI	RLFKSHPETL	EKFDRFKHLK	TEAEMKASED	LKKHGVTVLT	ALGAILKKKG	HHEAELKPLA	QSHATKHKIP
1111111111	1111111111	1111111111	1111111111	1111111111	1111111111	1111111111	1111111111	1111111111	1111111111
101	111	121	131	141	151				
IKYLEFISEA	IIHVLHSRHP	GDFGADAQGA	MNKALELFRK	DIAAKYKELG	Y				
1111111111	1111111111	1111111111	1111111111	1111111111	1				

Slika 15: RaptorX: vhodni podatki.

Na sliki 16 vidimo, da je program našel dva žepka. Označeni so deli proteina, ki so aktivni v vezavnem mestu. Na desni strani je osnutek rezultata, pomoč in možnost prenosa datoteke. Viden je tudi status naše poizvedbe.

Section II. Summary Prediction Results

Select domain ▼
Domain: 1 P-value: 6.73e-06 uGDT(GDT): 162(107) uSeqId(SeqId): 151(100)

Pocket	Multiplicity	Ligand	Binding residues
1	141	HEM	T39 K42 F43 R45 H64 T67 V68 A71 L72 L89 S92 H93 H97 I99 Y103 L104 I107 F138
2	70	SO4	T51 E52

Summary ([help](#))

- The input is processed as **1** domain(s)
- In total **2** predicted pocket(s)
- 1(0%)** positions predicted as disordered
- Secondary struct: **76%H, 0%E, 23%C**
- Solvent access: **29%E, 39%M, 31%B**

Download

[Download](#) detailed prediction results. To open .zip files, you may use [7-zip](#) for **Windows** or unzip for **Linux/Unix/MacOS**.

Status

Current status:	Complete
Submitted on:	2016-10-24 14:14
Scheduled on:	2016-10-24 15:02
Finished on:	2016-10-24 15:21

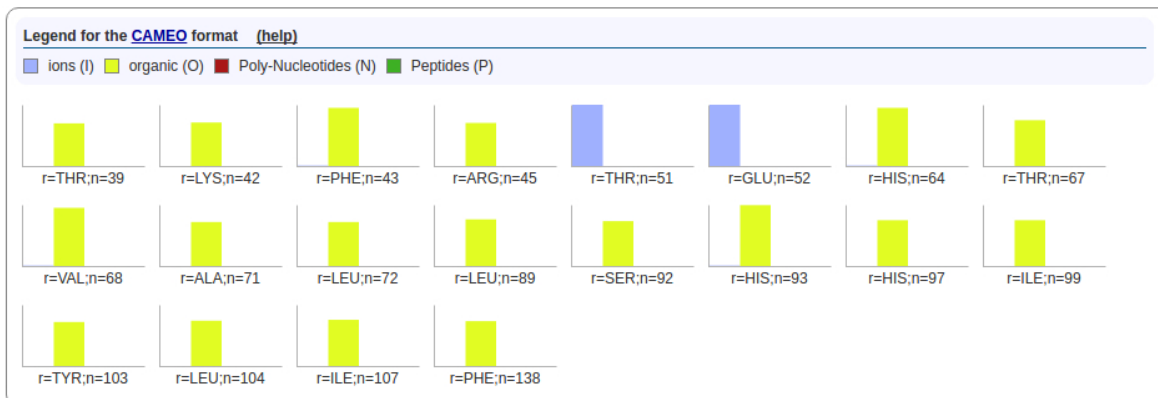
Section III. Detailed Prediction Results (you can see each result entry by clicking on it)

[+] Click to view the predicted results in the CAMEO format

[+] Click to view 3D model and binding site residues for domain 1 [1, 151] P-value:6.73e-06

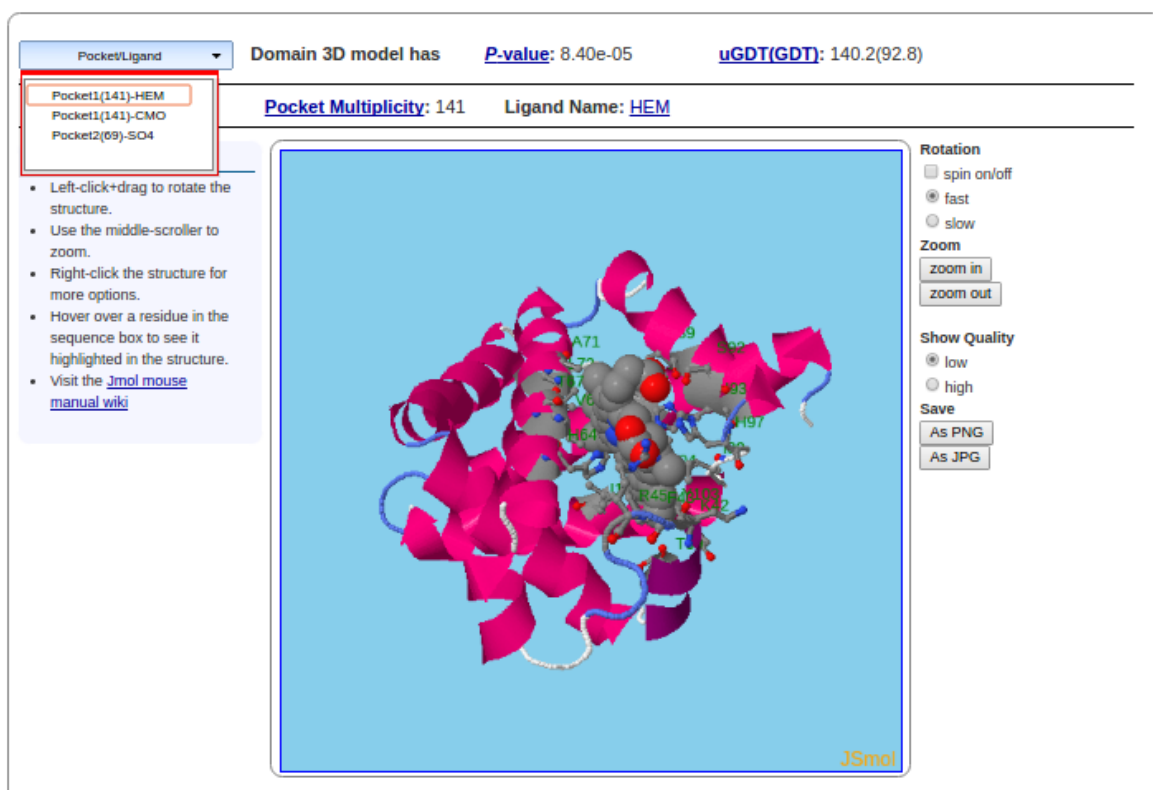
Slika 16: RaptorX: možni ligandi.

Na sliki 17 je prikazana legenda vrste struktur v preiskovanem proteinu. Vidimo, da je večina organskih struktur, dva modra stolpca sta prikazana kot iona. Polinukleotidov in peptidov program ni našel.



Slika 17: RaptorX: vrste struktur v proteinu.

Na sliki 18 je prikazan končni rezultat RaptorX programa v Jsmol-u. Zgoraj desno imamo jeziček "Pocket/Ligand" s katerim si izbiramo, katere ligande naj prikaže. Program je našel tri ligande: HEM, CMO in SO4. Prikazana je tudi P-vrednost in GDT vrednost (mera podobnosti). Uporabnik lahko spremnja hitrost vrtenja strukture in si shrani sliko v PNG oziroma JPG formatu. Za več informacij je podana povezava do navodil uporabe Jmol.



Slika 18: RaptorX: končni rezultat.

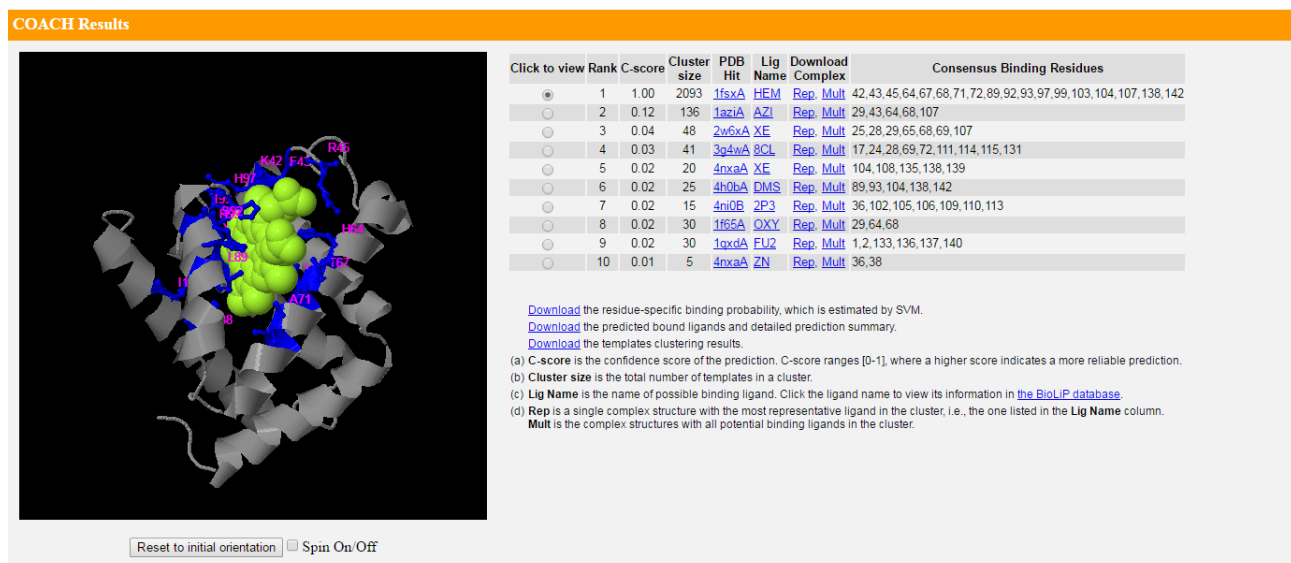
[Click here](#) to see help for this sequence box

1	11	21	31	41	51	61
VLSEGEWQLV LHVWAKVEAD VAGHGQDILI RLKFSHPETL EKFD RFRKHLK TEAEMKASED LKKHGVTVLT						
0000000000 0000000000 0000000000 0000000000 0000000000 0000000000 0000000000						
71	81	91	101	111	121	131
ALGAILKKG HHEAELKPLA QSHATKHKIP IKYLEFISEA IIVLHSRHP GDFGADAQGA MNKALELFRK						
0000000000 0000000000 0000000000 0000000000 0000000000 0000000000 0000000000						
141	151					
DIAAKYKELGY						
0000000000 0						

Program RaptorX je samostojno našel manjši ligand CMO. HEM in SO4 sta namreč že definirana v FASTA zaporedju proteina 1A6M. Za napoved je v mojem primeru potreboval 37 minut. Rezultati se na internetni strani hranijo do 14 dni.

4.3 COACH

Na sliki 20 je v levem kotu prikazan protein skupaj z ligandi. Na desni strani imamo možnost izbrati kateri ligand naj program vizualizira. V mojem primeru je našel 10 možnih ligandov. Opisani so parametri kot je C-vrednost, velikost, ime liganda. Podana je možnost prenosa podatkov. Poleg končnega rezultata programa COACH, nam strežnik vrne tudi prikaz posameznih rezultatov programov COACH: TM-SITE, S-SITE, COFACTOR in ConCavity z istimi parametri kot pri COACH-u. Rezultati so v prilogi.

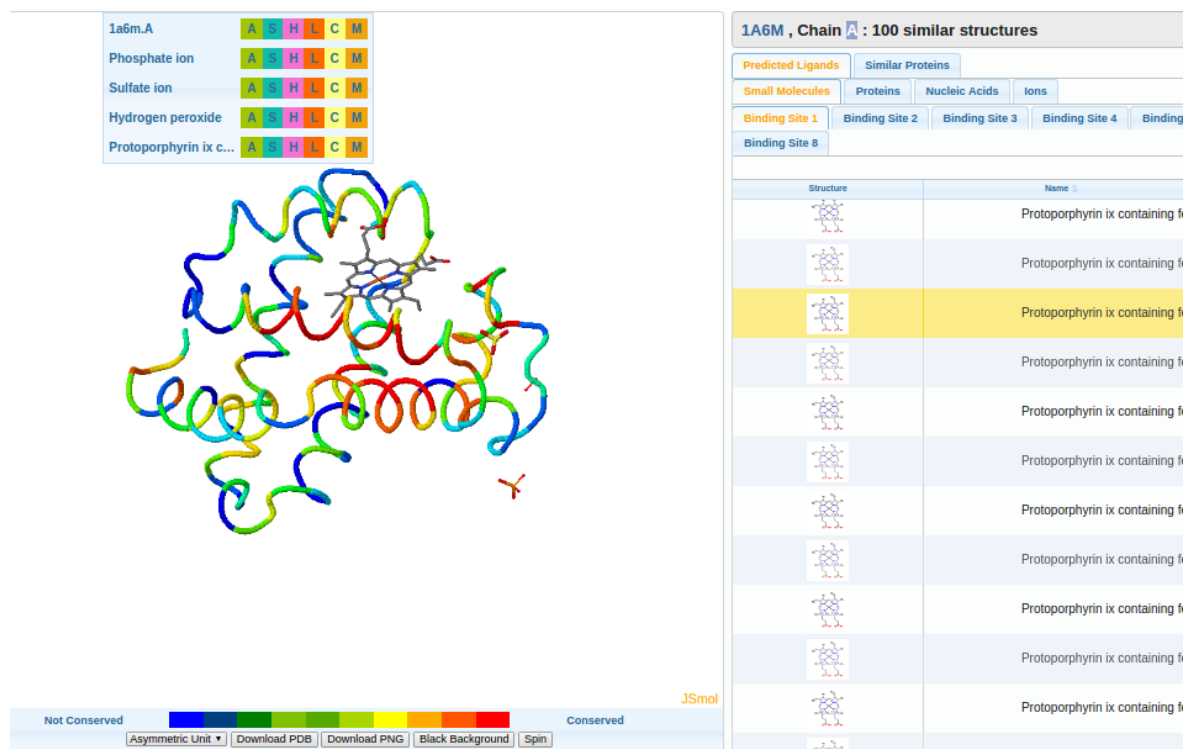


Slika 20: COACH: napoved ligandov.

Program je napovedal več ligandov. V mojem primeru je COACH za izračun potreboval 5 ur in 7minut.

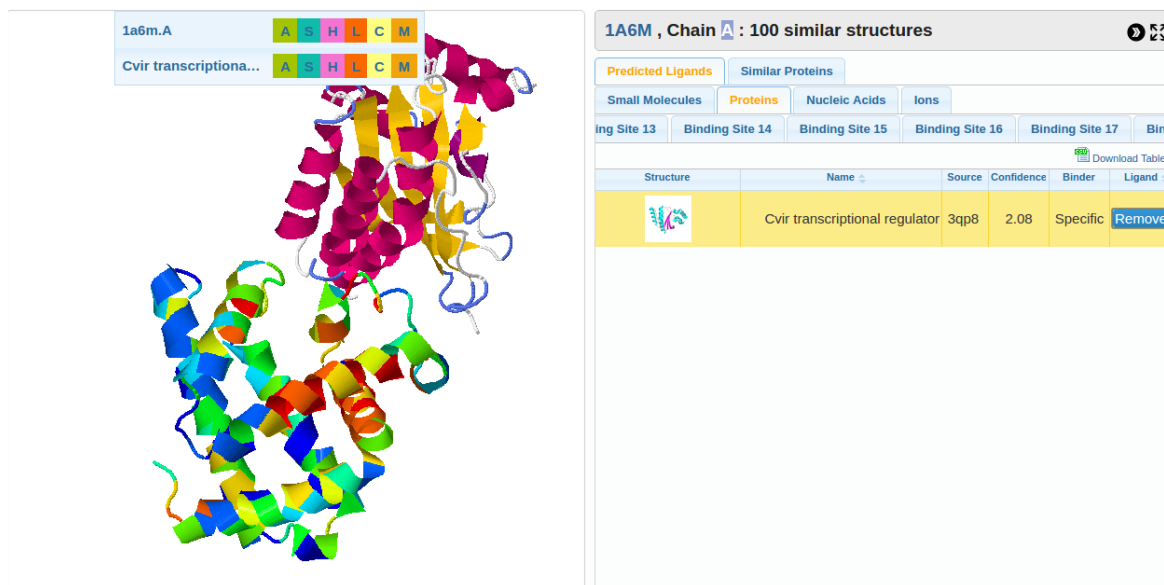
4.4 ProBiS

Na sliki 21 je v tabeli označenih 5 ligandov izmed vseh najdenih. V spodnjem delu slike je barvna razpredelnica. Skrajna modra predstavlja najbolj ne-ohranjeni del proteina, skrajna rdeča pa najbolj ohranjeni del proteina.



Slika 21: ProBiS: najdeni ligandi.

Strežnik ProBiS (slika 22) poleg majhnih molekul najde tudi proteine, posamezne nukleinske kisline in ione, ki se vežejo na preiskovani protein.

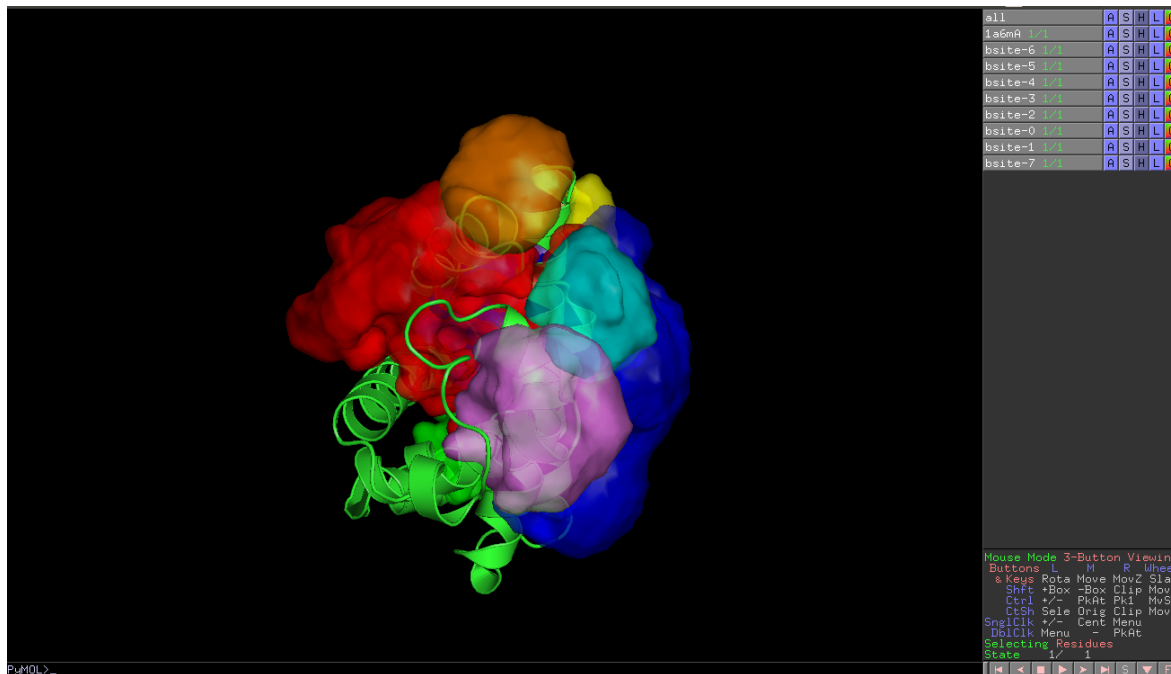


Slika 22: ProBiS: rezultati.

ProBiS je za napoved ligandov na primeru proteina 1A6M potreboval 2 uri in 19 minut. Našel je skupno 8 vezavnih mest in več malih molekul (ligandov). Za protein-protein vezavo je našel skupno 21 vezavnih mest. Nukleinskih kislin in ionov, ki bi se vezali na preiskovani protein, ni našel.

4.4.1 ProBiS vtičnik za Pymol

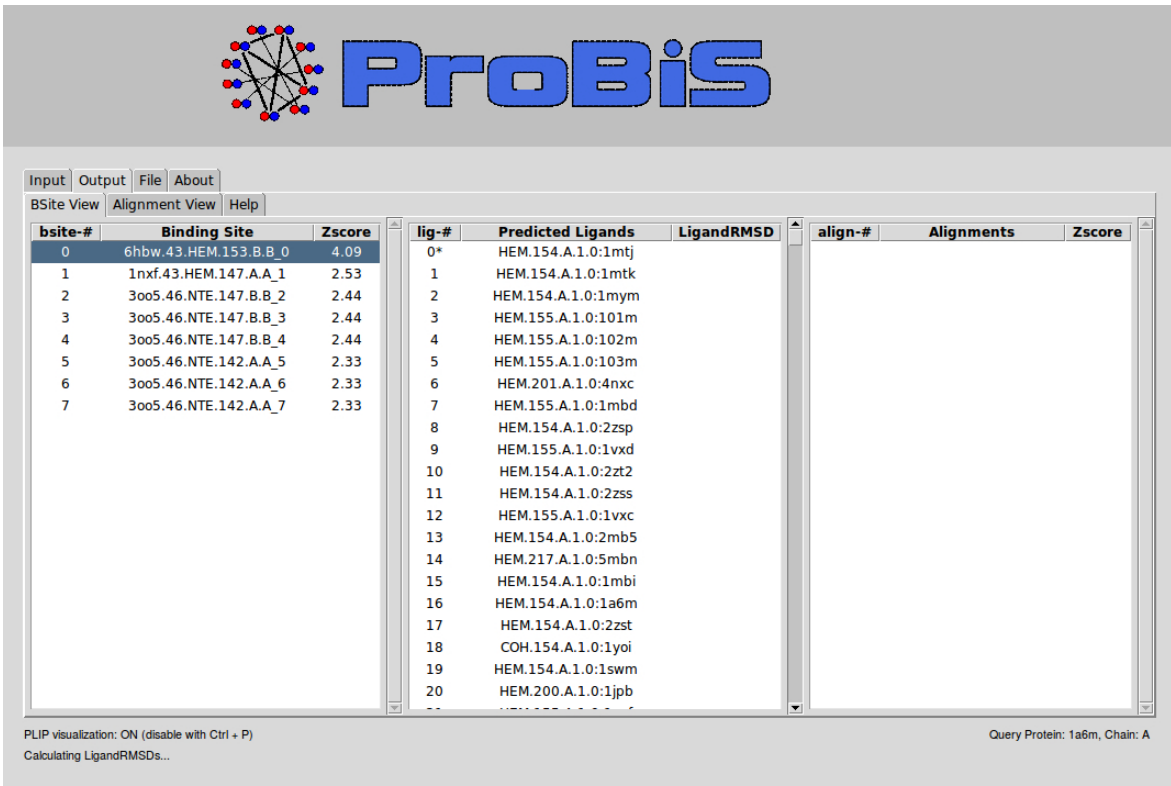
Na sliki 23 je prikazan protein (zeleno) in vezavna mesta (modra, svetlo modra, violečna in rjava). Uporabnik lahko spreminja protein preko programa Pymol [55].



Slika 23: ProBiS vtičnik za Pymol.

Pymol je zelo uporaben, brezplačen in uporabniku prijazen program za vizualizacijo in modeliranje molekul. Skupaj s ProBiS vtičnikom lahko hitro in enostavno analiziramo preučevani protein.

Na sliki 24 so v levem stolpcu vezavna mesta, na desnem stolpu pa pripadajoči ligandi. Vidimo, da je program našel 8 vezavnih mest in večje število ligandov.



The screenshot displays the ProBiS software interface. At the top, there is a logo consisting of a network of red and blue nodes connected by lines, followed by the text "ProBiS" in a large, blue, stylized font. Below the logo, there is a menu bar with "Input", "Output", "File", and "About". Underneath the menu bar, there are two tabs: "BSite View" and "Alignment View", with "BSite View" currently selected. The main window is divided into three panels. The left panel shows a table of binding sites with columns "bsite-#", "Binding Site", and "Zscore". The middle panel shows a table of predicted ligands with columns "lig-#", "Predicted Ligands", and "LigandRMSD". The right panel shows a table of alignments with columns "align-#", "Alignments", and "Zscore".

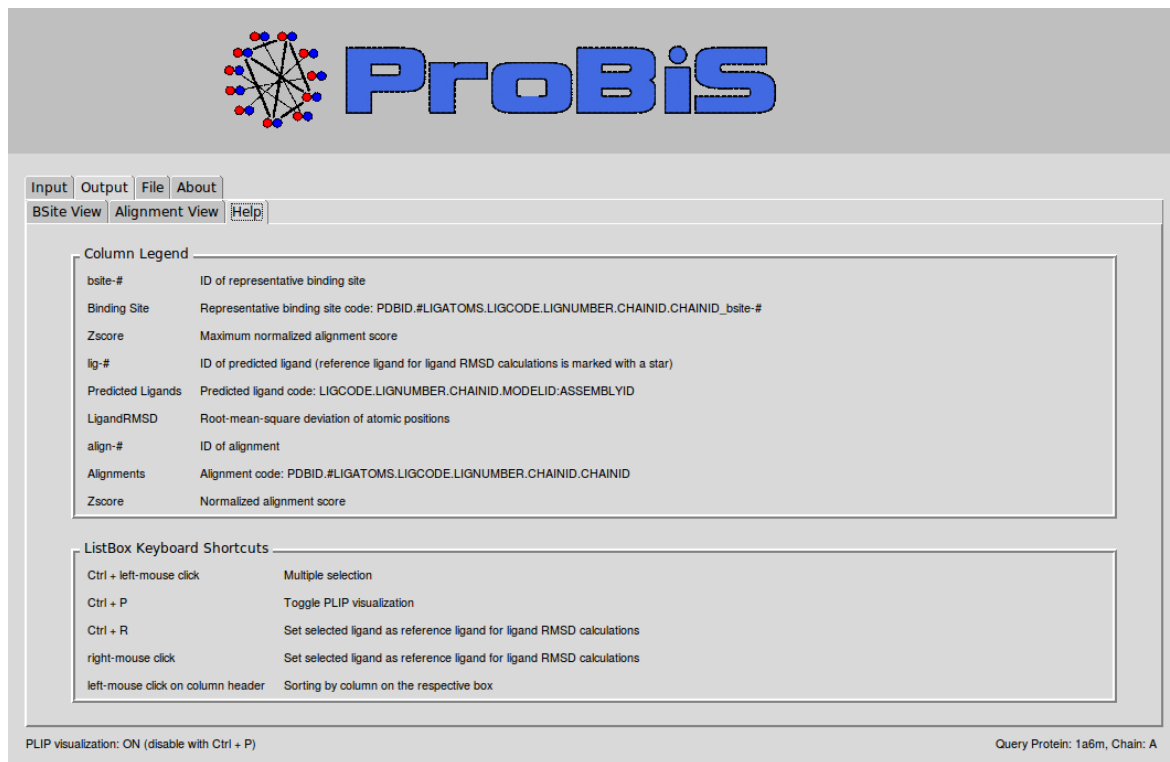
bsite-#	Binding Site	Zscore
0	6hbw.43.HEM.153.B.B_0	4.09
1	1nxf.43.HEM.147.A.A_1	2.53
2	3oo5.46.NTE.147.B.B_2	2.44
3	3oo5.46.NTE.147.B.B_3	2.44
4	3oo5.46.NTE.147.B.B_4	2.44
5	3oo5.46.NTE.142.A.A_5	2.33
6	3oo5.46.NTE.142.A.A_6	2.33
7	3oo5.46.NTE.142.A.A_7	2.33

lig-#	Predicted Ligands	LigandRMSD
0*	HEM.154.A.1.0:1mtj	
1	HEM.154.A.1.0:1mtk	
2	HEM.154.A.1.0:1mym	
3	HEM.155.A.1.0:101m	
4	HEM.155.A.1.0:102m	
5	HEM.155.A.1.0:103m	
6	HEM.201.A.1.0:4nxc	
7	HEM.155.A.1.0:1mbd	
8	HEM.154.A.1.0:2zsp	
9	HEM.155.A.1.0:1vxd	
10	HEM.154.A.1.0:2zt2	
11	HEM.154.A.1.0:2zss	
12	HEM.155.A.1.0:1vxc	
13	HEM.154.A.1.0:2mb5	
14	HEM.217.A.1.0:5mbn	
15	HEM.154.A.1.0:1mbi	
16	HEM.154.A.1.0:1a6m	
17	HEM.154.A.1.0:2zst	
18	COH.154.A.1.0:1yoi	
19	HEM.154.A.1.0:1swm	
20	HEM.200.A.1.0:1jpb	

At the bottom of the interface, there is a status bar with the text "PLIP visualization: ON (disable with Ctrl + P)" and "Calculating LigandRMSDs...". On the right side of the status bar, it says "Query Protein: 1a6m, Chain: A".

Slika 24: ProBiS:vezavna mesta in napovedani ligandi.

Na sliki 25 je prikazana pomoč pri programu ProBiS.



Slika 25: ProBiS pomoč.

Na podlagi rezultatov je ProBiS napovedal največ ligandov- malih molekul in proteinov. Pri ProBiS strežniku je potrebno poudariti, da sprejme več verig istočasno, medtem ko lahko drugi programi analizirajo le eno verigo posebej. ProBiS strežnik je za analizo vezavnega mesta potreboval približno 2 uri, ProBiS vtičnik za Pymol pa je rezultate vrnil takoj.

5 Zaključek

V nalogi sem opisala štiri brezplačne, uporabniku dosegljive programe za iskanje vezavnih mest na proteinih, ter jih preizkusila na manjšem proteinu oksi-mioglobinu. Programi uporabljajo več načinov iskanja- od iskanja podobnosti na nivoju amino-kislinskega zaporedja do analize strukture in primerjave z več knjižnicami. Nekateri programi združujejo omenjena načina za pridobitev optimalnih rezultatov. Za vizualizacijo je najbolj uporaben program Jmol. Ugotovila sem, da v mojem primeru od vseh programov izstopa ProBiS, ki je napovedal največ možnih ligandov- od manjših molekul, do proteinov. Dobre rezultate je vrnil tudi COACH, ki je vizualiziral rezultate več programov. Pri COACHu je praktično ravno to, da uporablja več komplementarnih metod. Najhitrejša sta bila RaptorX (37min) in ProBiS (2 uri). Za še bolj natančno oceno vseh programov bi bilo potrebno programe zagnati na več primerih proteinov.

Literatura

- [1] O'MALLEY, P. J. J. in ET AL., Scalable quantum simulation of molecular energies, *Physical Review X* 63 (2016), 031007. (*Citirano na strani 1.*)
- [2] WASS, MARK N., LAWRENCE A. KELLEY in MICHAEL JE STERNBERG, 3DLigandSite: predicting ligand-binding sites using similar structures, *Nucleic acids research* (2010), gkq406. (*Citirano na straneh VII, 1, 4, 6 in 7.*)
- [3] KÄLLBERG, MORTEN in ET AL., Template-based protein structure modeling using the RaptorX web server., *Nature protocols* 7(8) (2012), 1511-1522. (*Citirano na straneh 1 in 8.*)
- [4] YANG, JIANYI, AMBRISH ROY in YANG ZHANG, Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment, *Bioinformatics* 29(20) (2013), 2588-2595. (*Citirano na straneh VII, 1, 9, 10 in 11.*)
- [5] KONC, JANEZ in DUŠANKA JANEŽIČ, ProBiS algorithm for detection of structurally similar protein binding sites by local structural alignment, *Bioinformatics* 26(9) (2010), 1160-1168. (*Citirano na straneh VII, 1, 3, 12, 13, 14 in 15.*)
- [6] KINNINGS in SARAH L., ET AL., Drug discovery using chemical systems biology: repositioning the safe medicine Comtan to treat multi-drug and extensively drug resistant tuberculosis, *PLoS Comput Biol* 5(7) (2009), e1000423. (*Citirano na strani 2.*)
- [7] STANK, ANTONIA in ET AL., Protein Binding Pocket Dynamics, *Accounts of chemical research* 49(5) (2016), 809-815. (*Citirano na strani 2.*)
- [8] ANDERSON, AMY C, The process of structure-based drug design, *Chemistry & biology* 10(9) (2003), 787-797. (*Citirano na straneh 3 in 5.*)
- [9] WALTERS, W. PATRICK, MATTHEW T. STAHL in MARK A. MURCKO, Virtual screening—an overview, *Drug Discovery Today* 3(4) (1998), 160-178. (*Citirano na strani 3.*)

- [10] HAMZA, ADEL, NING-NING WEI in CHANG-GUO ZHAN, Ligand-based virtual screening approach using a new scoring function, *Journal of chemical information and modeling* 52(4) (2012), 963-974. (*Citirano na strani 3.*)
- [11] OGMEN, UTKAN, in ET AL., PRISM: protein interactions by structural matching, *Nucleic acids research* 33(suppl 2) (2005), W331-W336. (*Citirano na strani 3.*)
- [12] VALDAR, WILLIAM SJ in JANET M. THORNTON, "Protein-protein interfaces: analysis of amino acid conservation in homodimers, *Proteins: Structure, Function, and Bioinformatics* 42(1) (2001), 108-124. (*Citirano na strani 3.*)
- [13] LECOMTE, JULIETTE TJ, DAVID A. VULETICH in ARTHUR M. LESK, Structural divergence and distant relationships in proteins: evolution of the globins, *Current opinion in structural biology* 15(3) (2005), 290-301. (*Citirano na strani 3.*)
- [14] RUSSELL, ROBERT B. in ET AL., Recognition of analogous and homologous protein folds: analysis of sequence and structure conservation, *Journal of molecular biology* 269(3) (1997), 423-439. (*Citirano na strani 3.*)
- [15] HUANG, HUNG-JIN, ET AL. in , Current developments of computer-aided drug design, *Journal of the Taiwan Institute of Chemical Engineers* 41(6) (2010), 623-635. (*Citirano na strani 5.*)
- [16] MARTÍ-RENO, MARC A. in ET AL., Comparative protein structure modeling of genes and genomes, *Annual review of biophysics and biomolecular structure* 29(1) (2000), 291-325. (*Citirano na strani 5.*)
- [17] ALTSCHUL, STEPHEN F. in ET AL., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs., *Nucleic acids research* 25(17) (1997), 3389-3402. (*Citirano na strani 6.*)
- [18] MCGUFFIN, LIAM J., KEVIN BRYSON in DAVID T. JONES, The PSIPRED protein structure prediction server, *Bioinformatics* 16(4) (2000), 404-405. (*Citirano na strani 6.*)
- [19] POLLASTRI, GIANLUCA in ET AL., Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles, *Proteins: Structure, Function, and Bioinformatics* 47(2) (2002), 228-235. (*Citirano na strani 6.*)
- [20] COLE, CHRISTIAN, JONATHAN D. BARBER in GEOFFREY J. BARTON, The Jpred 3 secondary structure prediction server, *Nucleic acids research* 36(suppl 2) (2008), W197-W201. (*Citirano na strani 6.*)

- [21] WARD, JONATHAN J. in ET AL., The DISOPRED server for the prediction of protein disorder, *Bioinformatics* 20(13) (2004), 2138-2139. (*Citirano na strani 6.*)
- [22] KELLEY, LAWRENCE A. in MICHAEL JE STERNBERG, Protein structure prediction on the Web: a case study using the Phyre server, *Nature protocols* 4(3) (2009), 363-371. (*Citirano na strani 6.*)
- [23] ORTIZ, ANGEL R., CHARLIE EM STRAUSS in OSVALDO OLMEA, MAMMOTH (matching molecular models obtained from theory): an automated method for model comparison, *Protein Science* 11(11) (2002), 2606-262. (*Citirano na strani 7.*)
- [24] PENG, JIAN in JINBO XU, Low-homology protein threading., *Bioinformatics* 26(12) (2010), i294-i300. (*Citirano na strani 8.*)
- [25] PENG, JIAN in JINBO XU, Boosting protein threading accuracy., *Annual International Conference on Research in Computational Molecular Biology* Springer Berlin Heidelberg (), 2009. (*Citirano na strani 8.*)
- [26] CHANG, KUAN Y. in ET AL., Analysis and prediction of the critical regions of antimicrobial peptides based on conditional random fields., *PloS one* 10(3) (2015), e0119490. (*Citirano na strani 8.*)
- [27] SUTTON, CHARLES in ANDREW MCCALLUM, An introduction to conditional random fields., *Foundations and Trends® in Machine Learning* 4(4) (2012), 267-373. (*Citirano na straneh VII in 8.*)
- [28] PENG, JIAN in JINBO XU, A multiple?template approach to protein threading., *Proteins: Structure, Function, and Bioinformatics* 79(6) (2011), 1930-1939. (*Citirano na strani 8.*)
- [29] YANG, JIANYI in YANG ZHANG, Protein Structure and Function Prediction Using I-TASSER, *Current protocols in bioinformatics* (2015), 5-8. (*Citirano na strani 9.*)
- [30] WU, SITAO in YANG ZHANG, LOMETS: a local meta-threading-server for protein structure prediction, *Nucleic acids research* 35(10) (2007), 3375-3382. (*Citirano na strani 9.*)
- [31] ZHANG, YANG in JEFFREY SKOLNICK, SPICKER: A clustering approach to identify near?native protein folds, *Journal of computational chemistry* 25(6) (2004), 865-871. (*Citirano na strani 9.*)
- [32] ZHANG, JIAN, YU LIANG in YANG ZHANG, Atomic-level protein structure refinement using fragment-guided molecular dynamics conformation sampling, *Structure* 19(12) (2011), 1784-1795. (*Citirano na strani 9.*)

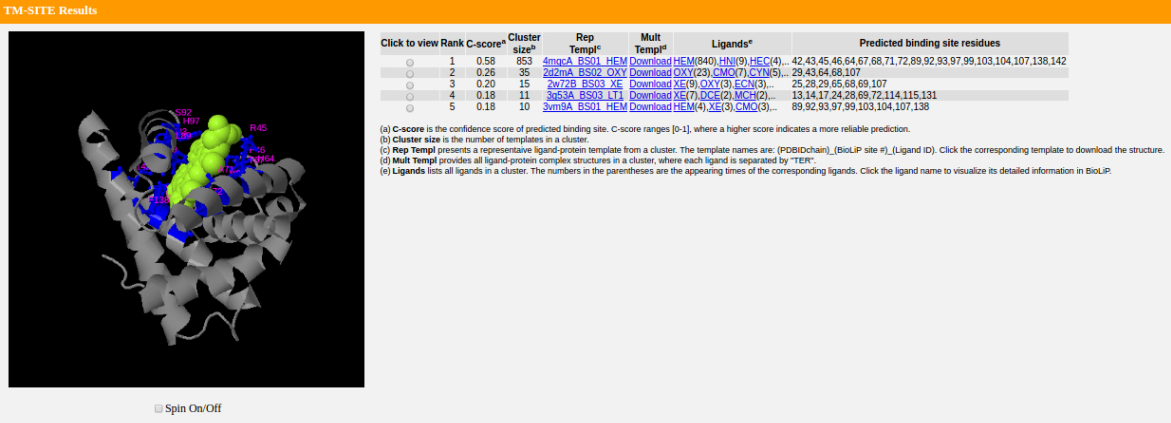
- [33] XU, DONG in YANG ZHANG, Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization, *Biophysical journal* 101(10) (2011), 2525-2534. (*Citirano na strani 9.*)
- [34] CAPRA, JOHN A. in ET AL., Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure, *PLoS Comput Biol* 5(12) (2009), e1000585. (*Citirano na strani 10.*)
- [35] ZHANG, YANG in JEFFREY SKOLNICK, TM-align: a protein structure alignment algorithm based on the TM-score, *Nucleic acids research* 33(7) (2005), 2302-2309. (*Citirano na strani 10.*)
- [36] NEEDLEMAN, SAUL B. in AND CHRISTIAN D. WUNSCH, A general method applicable to the search for similarities in the amino acid sequence of two proteins, *Journal of molecular biology* 48(3) (1970), 443-453. (*Citirano na strani 11.*)
- [37] BRYLINSKI, MICHAL in JEFFREY SKOLNICK, A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation, *Proceedings of the National Academy of sciences* 105(1) (2008), 129-134. (*Citirano na strani 11.*)
- [38] ROY, AMBRISH, JIANYI YANG in YANG ZHANG, COFACTOR: an accurate comparative algorithm for structure-based protein function annotation, *Nucleic acids research* (2012), gks372. (*Citirano na strani 11.*)
- [39] JOACHIMS in THORSTEN, Training linear SVMs in linear time, *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 217-226) (2006), ACM. (*Citirano na strani 11.*)
- [40] MANNHOLD, RAIMUND, HUGO KUBINYI in GERD FOLKERS, Virtual screening: principles, challenges, and practical guidelines, v: Ed. Christoph Sotriffer, *Machine Learning Applications for Virtual Screening*, John Wiley & Sons, 2011, Vol. 48. (*Citirano na strani 11.*)
- [41] (2012) Support vector machines. Dostopno 17.02.2017 na: <https://prateekvjoshi.com/2012/08/24/support-vector-machines/> (*Citirano na straneh VII in 12.*)
- [42] KONC, JANEZ in DUŠANKA JANEZIČ, An improved branch and bound algorithm for the maximum clique problem, *proteins* 4(5) (2007), b. (*Citirano na strani 13.*)
- [43] KONC, JANEZ, MILAN HODOŠČEK in DUŠANKA JANEZIČ, Molecular surface walk, *Croatica chemica acta* 79(2) (2006), 237-241.

- [44] CONNOLLY, MICHAEL L, Analytical molecular surface calculation, *Journal of Applied Crystallography* 16(5) (1983), 548-558. (*Citirano na straneh VII in 14.*)
(*Citirano na strani 14.*)
- [45] LEE, BYUNGKOOK in FREDERIC M. RICHARDS, The interpretation of protein structures: estimation of static accessibility, *Journal of molecular biology* 55(3) (1971), 379-400. (*Citirano na strani 14.*)
- [46] MERZEL, FRANCI in JEREMY C. SMITH, SASSIM: a method for calculating small-angle X-ray and neutron scattering and the associated molecular envelope from explicit-atom models of solvated proteins, *Acta Crystallographica Section D: Biological Crystallography* 58(2) (2002), 242-249. (*Citirano na strani 14.*)
- [47] CAVALLO, LUIGI, JENS KLEINJUNG in FRANCA FRATERNALI, POPS: a fast algorithm for solvent accessible surface areas at atomic and residue level, *Nucleic acids research* 31(13) (2003), 3364-3366. (*Citirano na strani 14.*)
- [48] RICHMOND, TIMOTHY J, Solvent accessible surface area and excluded volume in proteins: Analytical equations for overlapping spheres and implications for the hydrophobic effect, *Journal of molecular biology* 178(1) (1984), 63-89. (*Citirano na strani 14.*)
- [49] WANG, JUNMEI, ET AL. in , Solvation model based on weighted solvent accessible surface area, *The Journal of Physical Chemistry B* 105(21) (2001), 5055-5067. (*Citirano na strani 14.*)
- [50] NICHOLLS, ANTHONY in BARRY HONIG, A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation, *Journal of computational chemistry* 12(4) (1991), 435-445. (*Citirano na strani 14.*)
- [51] TOMASI, JACOPO in MAURIZIO PERSICO, Molecular interactions in solution: an overview of methods based on continuous distributions of the solvent, *Chemical Reviews* 94(7) (1994), 2027-2094. (*Citirano na strani 14.*)
- [52] (2017) 1A6M ligand. Dostopno 17.02.2017 na:
<http://www.rcsb.org/pdb/explore/explore.do?structureId=1A6M> (*Citirano na straneh VII in 16.*)
- [53] KENDREW, JOHN C. in ET AL., A three-dimensional model of the myoglobin molecule obtained by x-ray analysis, *Nature* 181(4610) (1958), 662-666. (*Citirano na strani 16.*)

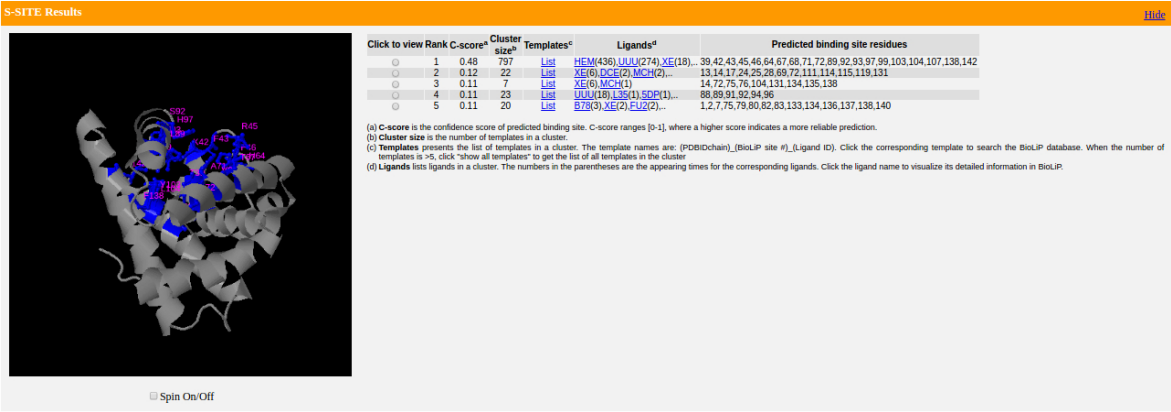
- [54] KAMGA, CHRISTELLE, SUHAS KRISHNAMURTHY in SRUTI SHIVA, Myoglobin and mitochondria: a relationship bound by oxygen and nitric oxid, *Nitric Oxide* 26(4) (2012), 251-258. (*Citirano na strani 16.*)
- [55] (2015) The PyMOL Molecular Graphics System, Version 1.8, Schrödinger, LLC.. Dostopno 08.03.2017 na: <https://www.pymol.org> (*Citirano na strani 26.*)

Priloge

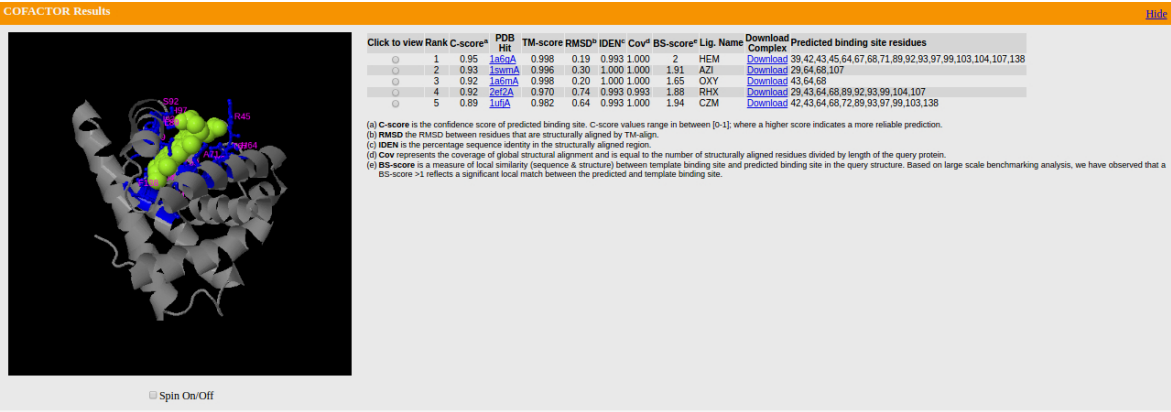
A Rezultati programov v sklopu programa COACH



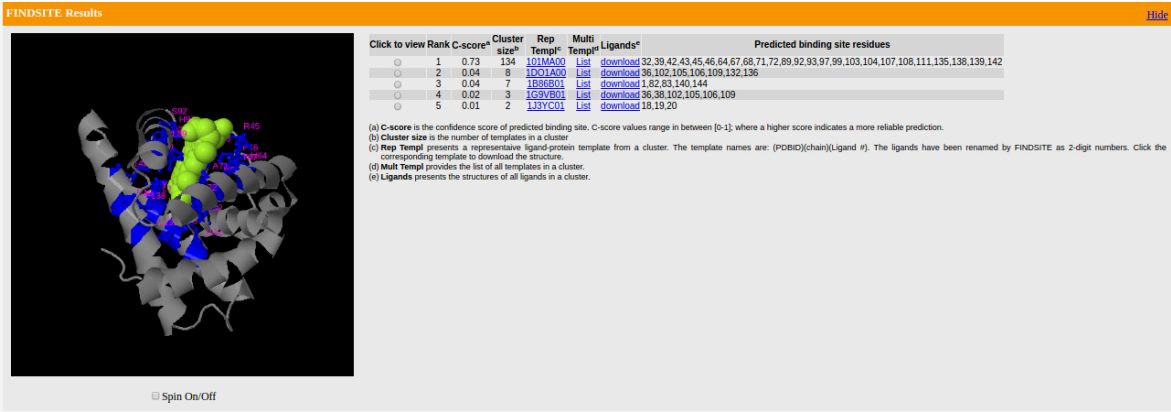
Rezultati TM-SITE.



Rezultati S-SITE.



Rezultati COFACTOR.



Rezultati ConCavity.