

UNIVERZA NA PRIMORSKEM
FAKULTETA ZA MATEMATIKO, NARAVOSLOVJE IN
INFORMACIJSKE TEHNOLOGIJE

Zaključna naloga

**Razvoj spletnega uporabniškega vmesnika za program
molekularnega sidranja**

(Development of web user interface for molecular docking algorithm)

Ime in priimek: Tanja Štular

Študijski program: Bioinformatika

Mentor: doc. dr. Janez Konc

Somentor: prof. dr. Dušanka Janežič

Koper, avgust 2015

Ključna dokumentacijska informacija

Ime in PRIIMEK: Tanja ŠTULAR

Naslov zaključne naloge: Razvoj spletnega uporabniškega vmesnika za program molekularnega sidranja

Kraj: Koper

Leto: 2015

Število listov: 58

Število slik: 23

Število tabel: 2

Število prilog: 3

Število strani prilog: 5

Število referenc: 33

Mentor: doc. dr. Janez Konc

Somentor: prof. dr. Dušanka Janežič

Ključne besede: razvoj zdravil, molekularno sidranje, spletni uporabniški vmesnik, razvoj aplikacije, Candock

Izvleček:

Pri procesu odkrivanja novih zdravil imajo pomembno vlogo računalniško podprte metode in tehnologije, ki zanesljivo in hitro identificirajo aktivne spojine iz baz več milijonov spojin. Ena izmed najbolj obetavnih metod je molekularno sidranje. V zaključni nalogi je opisana implementacija spletnega uporabniškega vmesnika za program molekularnega sidranja. Uporabniški vmesnik odlikuje hiter in enostaven vnos vhodnih podatkov ter tridimenzionalni prikaz napovedanih interakcij med tarčnim proteinom in ligandi. Predstavljen je primer uporabe aplikacije za napovedovanje interakcij med proteinom Udp-n-acetylgalactopyranose mutase in testno bazo ligandov. Z uporabo razvitega spletnega vmesnika odkrijemo potencialne ligande, ki bi jih lahko nadalje eksperimentalno testirali na inhibicijo tega proteina.

Key words documentation

Name and SURNAME: Tanja ŠTULAR

Title of final project paper: Development of web user interface for molecular docking algorithm

Place: Koper

Year: 2015

Number of pages: 58

Number of figures: 23

Number of tables: 2

Number of appendices: 3

Number of appendix pages: 5

Number of references: 33

Mentor: Assist. Prof. Janez Konc, PhD

Co-Mentor: Prof. Dušanka Janežič, PhD

Keywords: drug design, molecular docking, web user interface, application development, Candock

Abstract:

Computational approaches such as molecular docking that reliably and rapidly identify active compounds from databases of millions of compounds have an important role in drug discovery process. In this work we designed and implemented a web user interface for molecular docking algorithm. This user interface is distinguished by fast and easy input of the target protein and ligands. It also features three-dimensional visualization of the results, that is the predicted interactions between the target protein and docked ligands. We present a use case of the developed user interface on the protein UDP-n-acetylgalactopyranose mutase to which we docked ligands from the prepared test database. We identify potential ligands that could be further tested for inhibition of this protein.

Zahvala

Na začetku bi se rada zahvalila vsem, ki so mi pomagali, da je delo nastalo. Velika zahvala gre mentorju, doc. dr. Janezu Koncu, za prijazne nasvete, opominjanje, hitro odzivnost in usmerjanje pri izdelavi zaključne naloge ter somentorici prof. dr. Dušanki Janežič za vso pomoč.

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

Zahvaljujem se Kemijskemu inštitutu za priložnost opravljanja študentskega dela.

Hvala moji družini, predvsem mami, ki me na moji poti tako in drugače podpira že celo življenje. Hvala prijateljem, ki so mi ves čas študija stali ob strani. Hvala vsem, ki ste verjeli vame.

Kazalo

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	Razvoj novih zdravil	2
2.1.1	Postopek načrtovanja zdravil	2
2.1.2	Virtualno reševanje	5
2.2	Molekularno sidranje	6
2.2.1	Konformacijski iskalni algoritmi	6
2.2.2	Cenilne funkcije	8
2.3	Knjižnice spojin	9
3	Uporabljene tehnologije	11
3.1	HTML	11
3.2	CSS	11
3.3	JavaScript	11
3.4	PHP	11
3.5	Ajax	12
3.6	JSmol	12
3.7	ProBiS	13
3.8	ProBiS-Database	13
3.9	Marvin JS	13
3.10	Datotečni formati	13
3.11	Open Babel	14
4	Zgradba spletne aplikacije za molekularno sidranje	15
4.1	Pregled arhitekture	15
4.2	RESTful strežnik	16
4.3	Procesi aplikacije	17

4.3.1	Vnos vhodnih podatkov	17
4.3.2	Candock algoritem za molekularno sidranje	23
4.3.3	Prikaz rezultatov molekularnega sidranja	24
5	Opis rešitve	25
5.1	Vnos vhodnih podatkov	26
5.1.1	Vnos proteina in verige	26
5.1.2	Izbira baze ligandov	26
5.1.3	Izbira vezavnih mest	28
5.1.4	Spreminanje vhodnih parametrov algoritma	28
5.1.5	Vnos elektronskega naslova	28
5.2	Zagon Candock algoritma	29
5.3	Rezultati	30
5.3.1	Prikaz rezultatov	30
5.3.2	Prenos rezultatov	32
5.3.3	Prilagoditev pogleda	32
5.3.4	Iskanje rezultatov	33
6	Primer uporabe aplikacije	34
6.1	Opis proteinske tarče	34
6.2	Vhodni podatki	35
6.3	Rezultati	36
7	Zaključek	39
	Literatura	40

Seznam tabel

2.1	Programi in konformacijski iskalni algoritmi za molekularno sidranje . .	8
2.2	Cenilne funkcije in metode	9

Seznam slik

1	Postopek načrtovanja zdravil z virtualnim rešetanjem	4
2	Shematski prikaz molekularnega sidranja	6
3	Prikaz proteina s PDB kodo 3EML v JSmolu	12
4	Arhitektura spletne aplikacije	15
5	Diagram podatkovnega toka nivo 1	17
6	Aktivnostni diagram izbire proteina in verige	18
7	Aktivnostni diagram izbire knjižnice spojin	19
8	Aktivnostni diagram izbire vezavnih mest	21
9	Aktivnostni diagram spremembe parametrov algoritma	22
10	Aktivnostni diagram prikaza rezultatov	24
11	Candock spletni uporabniški vmesnik	25
12	Risanje ligandov v Marvin JS orodju	27
13	Izbira vezavnih mest na 3EML proteinu	28
14	Spreminjanje parametrov algoritma in vnos elektronskega naslova	29
15	Primer prikaza rezultatov molekularnega sidranja	30
16	Primer prikaza rezultatov molekularnega sidranja	31
17	Primer interakcije med proteinom in ligandom. Protein je prikaz v rdeči, ligand v modri barvi.	31
18	Datoteka z rezultati	32
19	Meni v JSmolu	33
20	Vezavno mesto na proteinu 4MO2	35
21	Rezultati molekularnega sidranja za protein 4MO2 in testne baze	36
22	Prikaz interakcije med 4MO2 in spojino s kodo ChEMBL604591_59	37
23	Prikaz interakcije med 4MO2 in konformacijama ChEMBL604591 liganda	37

Seznam kratic

<i>AJAX</i>	asinhorni JavaScript in XML
<i>API</i>	vmesnik za komunikacijo med različnimi programskimi deli aplikacije
<i>ASCII</i>	ameriški standardni nabor za izmenjavo informacij
<i>CSS</i>	kaskadne stilske podloge
<i>CSV</i>	z vejico ločene vrednosti
<i>HTML</i>	označevalni jezik za oblikovanje večpredstavnostnih vsebin
<i>JavaScript</i>	objektni skriptni programski jezik
<i>JSON</i>	JavaScriptova objektna notacija
<i>PDB</i>	Protein Data Bank
<i>PHP</i>	skriptni jezik na strani strežnika
<i>REST</i>	reprezentativen prenos stanja
<i>SMILES</i>	enostavni molekularni vrstični vhodni sistem
<i>URL</i>	enolični krajevnik vira

1 Uvod

1.1 Opis problema

Napredek v organski kemiji je povzročil obstoj velikega števila potencialnih terapevtskih spojin, ki so dostopne raziskavam. Posledično se je povečala potreba po računalniških metodah in tehnologijah, ki zanesljivo in hitro identificirajo aktivne spojine iz baz več milijonov spojin. Pri procesu odkrivanja novih zdravil je potrebno med mnogimi potencialnimi spojinami identificirati tiste, ki ustrezajo določenim specifikacijam, npr. so metabolično stabilne, nimajo veliko stranskih učinkov, so zadostno topne v vodi in imajo sposobnost interakcije s tarčnim proteinom. Dandanes se nova zdravila odkrivajo s pomočjo računalniško podprtih metod, ena izmed najbolj obetavnih je molekularno sidranje. [13]

1.2 Namen in cilji naloge

Z namenom, da proces molekularnega sidranja postane prosto dostopen, lažji za uporabo, ter posledično hitrejši, smo naredili spletni uporabniški vmesnik Candock. Uporabniški vmesnik odlikuje hiter in enostaven postopek vnosa vhodnih podatkov, ter tridimenzionalni prikaz interakcije med ligandom in proteinom, kar omogoča predstavljuje rezultate in hitrejšo interpretacijo le-teh.

1.3 Pregled vsebine zaključne naloge

V začetku zaključnega dela je opisan postopek razvoja zdravil in molekularnega sidranja, kar bralcu nudi vpogled in potrebne osnove za razumevanje problema. Sledi načrtovanje aplikacije in predstavitev tehnologij, ki so bile uporabljene pri njeni izdelavi. Jedro naloge predstavlja natančen opis implementiranih lastnosti naše rešitve in njeno ovrednotenje, preko primera uporabe napovedovanja interakcij med ligandi in farmacevtsko zanimivim proteinom.

2 Pregled področja

2.1 Razvoj novih zdravil

Zdravilo je ponavadi majhna molekula, ki bodisi zavira ali aktivira funkcijo biološke makromolekule, npr. proteina. Z vplivom na delovanje proteina povzroči ugodne terapevtske učinke za bolnika. Načrtovanje novih zdravil je postopek pri katerem se išče nove molekule, katerih oblika in naboj sta komplementarna tarčni makromolekuli in posledično je med njima omogočena interakcija. Molekule, ki se vežejo na proteine imenujemo ligandi. [15]

V sodobnem pristopu načrtovanja in iskanja novih zdravilnih učinkovin se hkrati uporabljajo eksperimentalne tehnike in metode računalniške kemije, imenovane tudi *in silico* reševanje. Računalniška kemija (imenovana tudi molekularno modeliranje) uporablja matematične metode za reševanje kemijskih problemov, katere so avtomatizirane in implementirane za izvajanje na računalniku. [3] Rezultat preučevanj je molekularni model, ki predstavlja reprezentacijo kemijskega sistema, na osnovi matematičnega in fizikalnega opisa. Molekularno modeliranje ne more nadomestiti eksperimenta, saj ta predstavlja edini pravi kriterij za validacijo njegovih rezultatov, lahko pa bistveno pripomore k razumevanju procesov na atomskem nivoju in olajša načrtovanje molekul z ciljanimi lastnostmi. [19]

2.1.1 Postopek načrtovanja zdravil

Pred razvojem novih zdravil je potrebno dobro razumeti bolezen, vzroke za njen nastanek in kako bolezen spremeni izražanje genov. Sledi analiza vpliva spremenjenih genov na delovanje proteinov in študija interakcije teh proteinov z drugimi molekulami.

Identifikacija tarčne molekule

Razvoj novih zdravil je kompleksen proces (slika 1), ki se ponavadi začne z identifikacijo tarčne molekule. Tarčna molekula je lahko gen ali protein, ki je vključen v določeno bolezen. Pri postopku identifikacije tarče je pomembno, da ima izbrana tarčna molekula potencial za interakcijo z zdravilom, na primer, dobro definirano vezavno mesto

in ima pomembno vlogo pri bolezenskem procesu. V fazi identifikacije tarčne molekule se izbere molekulo, na katero želimo vplivati z zdravilom.

Iskanje zdravilne učinkovine

Cilj te faze je identificiranje obetavnih molekul (zadetkov), ki imajo potencial, da postanejo novo zdravilo. Zadetke se lahko najde na več načinov:

1. *De novo*: S prefinjenim računalniškim modeliranjem so predvidene lastnosti molekule, ki ima potencial za terapevtski vpliv. Te molekule se nato sintetizira v laboratoriju.
2. Virtualno reševanje: Virtualno ali *in silico* reševanje velja za najbolj uspešen postopek razvoja novih zdravil. Te računalniške metode omogočajo testiranje vezave milijonov spojin na tarčno molekulo in identifikacijo tistih, ki se nanjo močno vežejo. Glede na rezultate se za nadaljni razvoj izberejo najbolj obetavne spojine.
3. Eksperimentalno reševanje: Biološko aktivnost več tisočih spojin, iz dostopnih knjižnic, se lahko določi preko eksperimentalnega reševanja.
4. Biotehnologija: S pomočjo genskega inženiringa obstaja možnost preoblikovanja živega organizma, da proizvaja željeno spojino.

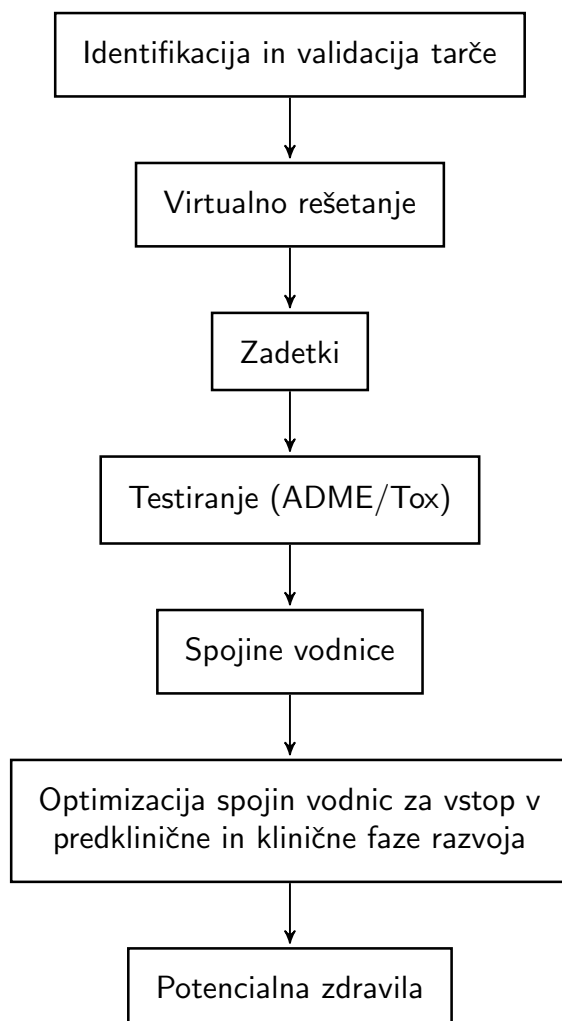
Testiranje uspešnosti spojin

Na zadetkih se izvedejo testi, ki zagotavljajo zgodnjo oceno zagotovitve kriterijev, ki so potrebni za spojine, da postanejo potencialna zdravila. Testira se absorpcija, distribucija, metabolizem, izločanje in toksikološke lastnosti (ADME/Tox). Uspešna zdravila morajo imeti sposobnost resorbiranja v krvni obtok, razpada na pravem mestu v telesu, morajo biti metabolično stabilna, se uspešno izločiti iz telesa in povzročiti čim manj stranskih učinkov. ADME/Tox testi se izvajajo v živih celicah, živalih ali z računalniškimi metodami. [26] Računalniške metode za testiranje temeljijo na predpostavki, da imajo podobne molekule zaradi strukturno podobnih fragmentov tudi podobne lastnosti, ki se jih da razstaviti na serijo med seboj aditivnih prispevkov. Na osnovi dosegljivih eksperimentalnih podatkov se s pomočjo statistične analize izpeljejo enačbe, ki se uporabljajo za napovedovanje lastnosti novih molekul. [19]

Optimizacija spojin

Zadetki, ki uspešno prestanejo testiranje, opisano v prejšnjem poglavju, se imenujejo spojine vodnice. Omenjene spojine se optimizira, da postanejo učinkovitejše in

varnejše. Optimizacija spojine se doseže s sprememinjanjem zgradbe molekule. Na primer, ustrezna sprememba v zgradbi spojine lahko povzroči, da le-ta tvori manj interakcij z netarčnimi molekulami, s čimer dosežemo manj stranskih učinkov zdravila. Pri postopku optimizacije sodelujejo biologi, kateri testirajo molekule v živem okolju in kemiki, ki sintetizirajo nove oblike spojine. Optimizirane spojine se imenujejo potencialna zdravila. [26]



Slika 1: Postopek načrtovanja zdravil z virtualnim rešetanjem

2.1.2 Virtualno rešetanje

Virtualno rešetanje je računalniški pristop, ki predstavlja *in silico* iskanje biološko aktivnih molekul v obširnih podatkovnih knjižnicah več milijonov spojin. Iz teh knjižnic se izbere manjše število potencialno aktivnih spojin, ki se jih kasneje ovrednoti še z biološkimi testi. Glavne računalniške algoritme za virtualno rešetanje delimo v dve skupini: algoritmi na osnovi ligandov in algoritmi na osnovi strukture tarčne molekule. [21]

Virtualno rešetanje na osnovi liganda

Razvoj zdravil na osnovni liganda se uporablja v primerih, ko struktura tarčne molekule ni poznana. Metoda temelji na predpostavki, da imajo molekule s podobno zgradbo, podobno biološko aktivnost. Znan mora biti vsaj en aktivni ligand, ki se primerja proti bazi molekul in identificira molekule s podobnimi farmakofornimi lastnostmi. Primer programa za virtualno rešetanje na osnovi liganda je LiSiCA, ki smo jo razvili sami in jo uporabili za napovedovanje inhibitorjev tarčnega proteina udeleženega pri nastanku Alzheimerjeve bolezni. [12]

Farmakofor je zbirka steričnih in elektronskih lastnosti, ki so potrebne za zagotovitev optimalnih interakcij s specifično biološko tarčo, ki vodi do indukcije ali inhibicije biološkega odziva. Ne predstavlja realne molekule ali zbirke funkcionalnih skupin, upošteva pa njihove lastnosti, ki vplivajo na sposobnost interakcije s preiskovano tarčo. [23]

Virtualno rešetanje na osnovi strukture tarče

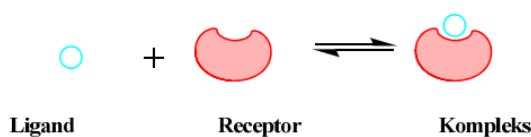
Pogoj za uporabo metode virtualnega rešetanja na osnovi strukture tarče, je poznana tridimenzionalna struktura tarčne molekule. Strukturo proteinov se določi s pomočjo literature ali s eksperimentalnimi metodami - kristalografija in nuklearna magnetna resonanca (NMR). [10]

Vsako leto je znana struktura vedno večjega števila proteinov. Na spletni strani Protein Data Base, ki je dostopna na <http://www.rcsb.org/pdb/home/home.do>, se je konec leta 2014 nahajalo 105397 znanih struktur proteinov. [30]

Najpogostejša metoda virtualnega rešetanja na osnovi strukture tarče je molekularno sidranje.

2.2 Molekularno sidranje

Molekularno sidranje se uporablja za modeliranje interakcij med majhno molekulo in proteinom, na nivoju atomov. Pristop napove preferenčno konformacijo liganda v izbranem aktivnem mestu biološke makromolekule (receptorja), ob predpostavki da le-ti tvorita stabilen kompleks. Konformacije molekule se med seboj razlikujejo le po tridimenzionalni razporeditvi atomov. Program molekularnega sidranja ponavadi sestoji iz dveh delov: konformacijskega iskalnega algoritma in cenovne funkcije. Rezultat predstavlja seznam po afiniteti razvrščenih napovedanih ligand-protein kompleksov. [14] Postopek molekularnega sidranja je prikazan na sliki 2. [19]



Slika 2: Shematski prikaz molekularnega sidranja

2.2.1 Konformacijski iskalni algoritmi

Namen iskalnih algoritmov je iskanje novih veznih konformacij liganda. Najpopolnejšo obravnavo fleksibilnosti liganda in proteina omogočajo sistematični iskalni algoritmi, ki upoštevajo vse stopnje rotacijske in konformacijske prostosti. Konformacijski prostor kompleksa sestavljajo vse možne orientacije proteina z vezanim ligandom. Zaradi prevelikega števila možnih vezavnih kombinacij pa so sistematični algoritmi računsko prezahtevni, zato večina algoritmov uporablja različne poenostavitve. V tabeli 2.1 so naštetih nekateri računalniški programi za molekularno sidranje in konformacijski iskalni algoritmi, ki jih le-ti uporabljajo. [5, 13, 16]

Algoritmi ujemanja

Algoritmi ujemanja mapirani ligand in protein predstavijo v obliki farmakoforov. Iz razdalj med posameznimi farmakofori izračunajo matriko razdalj, katera je namenjena gradnji novih konformacij ligandov. Prednost omenjenih algoritmov je nizka časovna zahtevnost.

Inkrementna izgradnja

Inkrementna izgradnja prekine vrtljive vezi v ligandu in ga razbije na več majhnih fragmentov. En del spojine (ponavadi največji) postavi v vezavno mesto, ostale fra-

gmente pa nato dodaja postopoma. Fleksibilnost liganda v aktivnem mestu doseže z generiranjem različnih usmeritev liganda.

Stohastični pristop

Stohastične metode iščejo konformacijski prostor z naključnim spreminjanjem konformacije liganda. Stohastični pristop uporablja metoda Monte Carlo in genetski algoritmi.

Metoda Monte Carlo ustvarja različne pozicije liganda preko vrtenja vezi. Če pridobljena konformacija ustreza energijskih kriterijem, se le-ta uporabi za iskanje nadaljnjih oblik molekule. Postopek se ponavlja toliko časa, dokler ni določeno napovedano število konformacij. Prednost Monte Carlo metode je, da omogoča velike spremembe v obliki spojine.

Genetski algoritem se zgleduje po zakonih evolucijske biologije. Nove strukture ligandov generira na podlagi genskih mutacij in prekrižanja kromosomov. [14]

Simulacijski pristop

Molekulska dinamika je metoda za simulacijo interakcij med atomi in molekulo v določenem časovnem obdobju. Vsak atom se premika neodvisno od ostalih, kar omogoča upoštevanje fleksibilnosti tako liganda kot proteina. Gibanje delcev opiše s pomočjo Newtonovih enačb gibanja. Atomi so predstavljeni kot delci v prostoru, ki se prosto gibljejo in med seboj zaletavajo. [4]

Tabela 2.1: Programi in konformacijski iskalni algoritmi za molekularno sidranje

Program	Algoritem za obravnavo fleksibilnosti liganda	Fleksibilnost proteina
AutoDock	Monte Carlo, genetski algoritem	ne
CHARMM	molekulska dinamika	da
DOCK	inkrementna izgradnja	ne
FlexE	inkrementna izgradnja	da
FlexX	inkrementna izgradnja	ne
FLOG	algoritem ujemanja	ne
Glide	Monte Carlo, genetski algoritem	ne
Gold	genetski algoritem	ne
LigandFit	Monte Carlo	ne
LUDI	inkrementna izgradnja	ne

2.2.2 Cenilne funkcije

Cenilne funkcije omogočajo izračun približka proste energije vezave liganda na tarčno molekulo. Uporabljajo se za oceno konformacije in orientacije liganda v vezavnem mestu. V programih za molekularno sidranje se večinoma uporabljajo cenilne funkcije, ki temeljijo na polju sil, empirične cenilne funkcije in cenilne funkcije, dobljene s pomočjo statistične mehanike. [2] V tabeli 2.2 so naštetne nekatere standardne cenilne funkcije in metode, ki jih te funkcije uporabljajo. [6, 17]

Cenilne funkcije, na osnovi polja sil

Cenilne funkcije, ki temeljijo na polju sil, ocenijo stabilnost ligand-protein kompleksa preko vsote neveznih (van der Waalsovih in elektrostatskih členov) interakcij. Nekatere funkcije pri računanju vsote upoštevajo še prisotnost vodikovih vezi, topnost in entropijo.

Empirične cenilne funkcije

Empirične cenilne funkcije temeljijo na eksperimentalno določenih interakcijah, za katere je že znana afiniteta vezave. Pri oceni vezavne energije upoštevajo koeficiente številnih členov (npr. vodikove vezi, hidrofobne interakcije, interakcije s kovinami...), katerih vrednosti se pridobi iz eksperimentalnih podatkov preko regresijske analize.

Cenilne funkcije na osnovi statistične mehanike

Cenilne funkcije, dobljene s pomočjo statistične mehanike, so osnovane na predpostavki, da so medatomske razdalje, ki se pojavljajo pogostejše, bolj ugodne od drugih razdalj. Ferkvenčna porazdelitev razdalj se pretvori v potencialne poravnave atomskih parov, ki se ocenijo s pozitivno ali negativno oceno.

Ocenjevanje na osnovi konsenza

Pri ocenjevanju na osnovi konsenza se upošteva več cenovnih funkcij skupaj. [1, 14]

Tabela 2.2: Cenilne funkcije in metode

Cenilna funkcija	Metoda	Cenilna funkcija	Metoda
AUTODOCK-score	ES, SS	DSX	SS
DOCK-score	PS, ES	PLP-score	ES
GOLD-score	PS, ES	LUDI-score	ES
FlexX-score	ES	PMF-score	PS
ChemScore	ES	CScore	KS

*PS = cenilne funkcije, ki temeljijo na polju sil

*ES = empirične cenilne funkcije

*SS = cenilne funkcije, dobljene s pomočjo statistične mehanike

*KS = ocenjevanje na osnovi konsenza

2.3 Knjižnice spojin

Pred postopkom virtualnega reševanja je zelo pomembna priprava knjižnice virtualnih molekul, kjer vsako molekulo opišemo z ustrezno tridimenzionalno konformacijo. Omejenjene konformacije se ponavadi pridobijo s pomočjo generatorjev molekulskih struktur, ki omogočajo generiranje realističnih tridimenzionalnih modelov. Programi, ki

omogočajo generiranje 3D struktur, so npr. OMEGA, Concord in Corina. [19] Pri načrtovanju in izgradnji knjižnic spojin, v katerih iščemo potencialno zanimive spojine, je potrebno paziti, da je načrtovane spojine sploh možno sintetizirati. Sposobnost sintetiziranja se oceni s pomočjo programov za načrtovanje knjižnic, kateri podajo rezultat na podlagi poznavanja ustreznih sinteznih poti. Potrebno je oceniti še smiselnost dobljenih konformacij, pravilnost aromatičnosti, stereokemije in tako dalje. [11]

3 Uporabljene tehnologije

V tem poglavju so opisane tehnologije uporabljene pri izdelavi spletne aplikacije.

3.1 HTML

HTML (*HyperText Markup Language*) je označevalni jezik namenjen izdelavi spletnih strani. Nabor opisnih simbolov, vstavljenih v datoteko, spletnemu brskalniku pove, kako naj se stran prikaže. Kodo razdelimo na elemente, ki so sestavljeni iz značk (npr. `<div>`). [27]

3.2 CSS

CSS (*Cascading Style Sheets*) je jezik, ki brskalniku pove, kako prikazati posamezne HTML elemente. Z njim nastavljamo velikost in postavitev elementov, barvo in tip pisave ter vrsto drugih funkcionalnosti. Za pravilno definicijo moramo upoštevati posamezne sintakse. [25]

3.3 JavaScript

JavaScript je interpretativni programski jezik, kar pomeni, da za zagon programa ni potrebno pretvarjanje v nižje nivojske programske jezike. Podpira tako objektno kot funkcijsko programiranje. Namenjen je razvoju interaktivnih spletnih strani. [33]

3.4 PHP

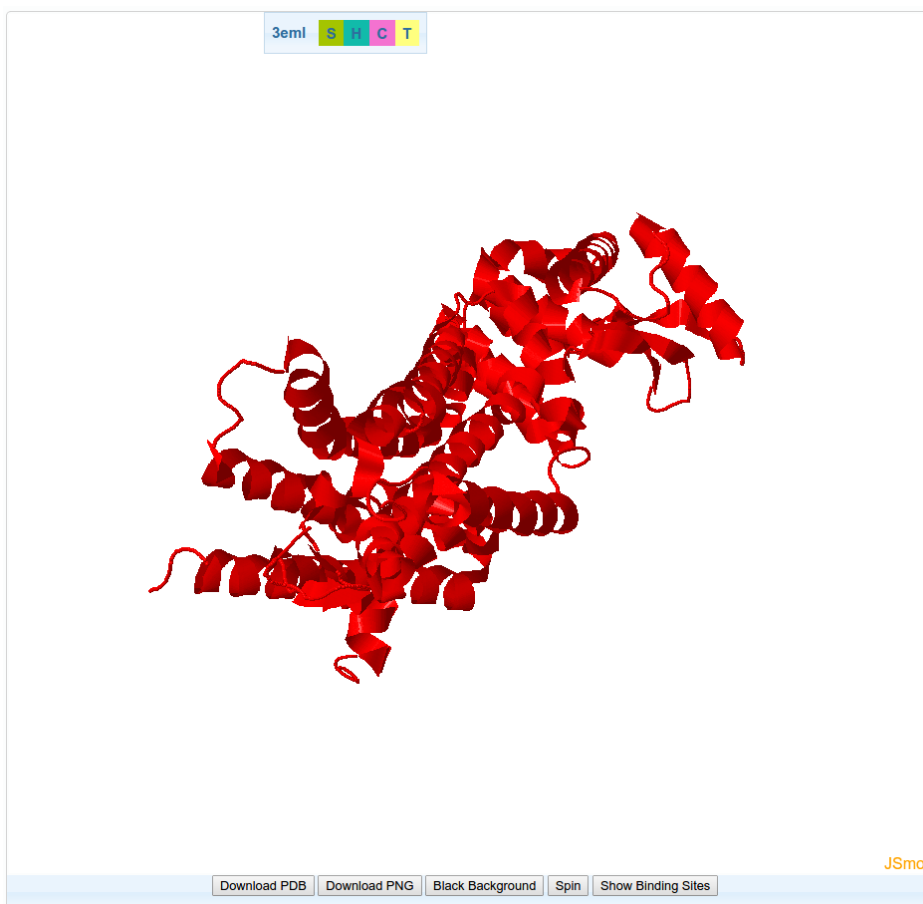
PHP (*PHP Hypertext Preprocessor*) je odprtokodni skriptni jezik, ki se uporablja za razvoj spletnih strani. PHP se izvaja izključno na strani strežnika, rezultat pa se vrne odjemalcu v jeziku HTML. [32]

3.5 Ajax

Ajax (*asynchronous JavaScript and XML*) se izvaja na strani odjemalca in omogoča izdelavo asinhronih spletnih aplikacij. Spletna aplikacija lahko v ozadju pošilja in pridobiva podatke iz strežnika. Podatki se pošiljajo v XMLHttpRequest ali JSON. Rezultati se lahko prikažejo s HTML in CSS, ki nastane znotraj brskalnika. Ajax je uporaben predvsem v primerih, ko se spreminja samo del spletne strani. [24]

3.6 JSmol

JSmol (*Jmol JavaScript Object*) je verzija Jmola (odprto kodni program za prikaz kemijskih struktur molekul v tridimenzionalni obliki), ki omogoča prikaz molekulske grafike v spletnih brskalnikih in za svoje delovanje ne potrebuje Jave temveč temelji na HTML5 tehnologiji. [7] Na sliki 3 je primer prikaza tridimenzionalne strukture proteina znotraj JSmola.



Slika 3: Prikaz proteina s PDB kodo 3EML v JSmolu

3.7 ProBiS

ProBiS algoritem napove vezavna mesta na proteinu preko lokalne strukturne poravnave. Izbrani protein primerja z bazo tridimenzionalnih struktur proteinov pri katerih s pomočjo detektiranja fizikalno-kemijskih lastnosti, poišče strukturno podobna mesta. Ostanke, ki se nahajajo na površini proteina tvorijo graf, kjer vsak od teh predstavlja svoje vozlišče. Na grafih se izvede konstrukcija podgrafov, kjer podgraf predstavlja vsa vozlišča v obsegu razdalje 15 Å. Iz podobnih podgrafov so izdelani produktni grafi na katerih algoritem poišče maksimalno kliko oziroma največji podgraf, katerega vsa vozlišča so med seboj povezana z vezmi. Vsaka maksimalna klika ustreza strukturalni poravnavi dveh proteinov, ki se oceni in projecira na izbrani protein. Lokacije najbolje ocenjenih ostankov so predvidene kot podobna vezavna mesta. [9]

3.8 ProBiS-Database

ProBiS-Database je baza, ki vsebuje preračunane strukturne poravnave proteinov, ki se nahajajo na Protein Data Bank. Funkcionalno pomembna vezavna mesta so predvidena s ProBiS algoritmom. [8]

3.9 Marvin JS

Marvin JS je ChemAxonov urejevalnik, ki združuje znanje iz področja kemije in JavaScript tehnologijo. Uporabniku omogoča prijazno orientirano orodje za hitro in enostavno risanje 2D kemijskih struktur in reakcij v spletnem okolju. [28]

3.10 Datotečni formati

V nadaljevanju so predstavljeni pdb, mol2 in smi datotečni formati.

PDB (*Protein Data Bank*) datotečni format je tekstovni format, ki opisuje tridimenzionalno kemijsko strukturo bioloških makromolekul. Datoteka vsebuje informacije o tipih atomov, atomskih koordinatah, strukturnih kristalografskih faktorjih in podatke o postopku jedrske magnetne resonance. V datoteki je lahko še zapisano ime molekule, informacije o primarni in sekundarni strukturi, referenčni podatkovni bazi, bibliografski citati... [29] Vzorčni primer datoteke v PDB formatu je v prilogi.

Mol2 format je tekstovni format, ki vsebuje vse potrebne informacije za rekonstrukcijo SYBYL (program za molekularno modeliranje) molekul. [31] Priložen je vzorčni

primer mol2 datoteke. Datoteka je razdeljen na 3 dele. Prvi del vsebuje informacije o molekuli. V drugem delu so podatki o posameznih atomih in njihovi lokaciji. Zadnji del datoteke vsebuje informacije o medatomskih vezeh.

Datotečni format smi vsebuje SMILES kode. SMILES (*simplified molecular-input line-entry system*) koda je vrstični zapis ASCII znakov, ki opisuje zgradbo molekule. [22] Primer SMILES kode:

```
Cn1cnc2n(C)c(=O)n(C)c(=O)c12
```

3.11 Open Babel

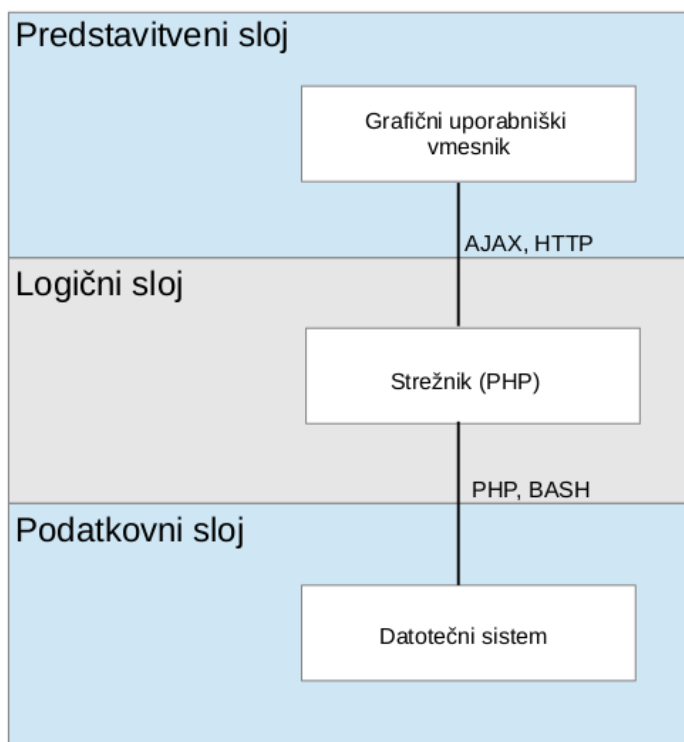
Open Babel je program, ki omogoča pretvarjanje med različnimi datotečnimi formati. [18] Primer uporabe programa za pretvorbo iz datotečnega formata mol v mol2:

```
babel -imol $vhodna_datoteka.mol -omol2 $izhodna_datoteka.mol2
```

4 Zgradba spletne aplikacije za molekularno sidranje

4.1 Pregled arhitekture

Aplikacija je sestavljena iz predstavitvenega, logičnega in podatkovnega sloja (slika 4). Na predstavitvenu sloju so uporabljene klasične tehnologije za prikaz na spletu, kot so HTML5, CSS in JavaScript. Logični sloj je napisan v jeziku PHP. Za podatkovni sloj, kamor se shranjujejo vnešeni podatki, smo v naši aplikaciji uporabili kar datotečni sistem.



Slika 4: Arhitektura spletne aplikacije

4.2 RESTful strežnik

Spletna aplikacija pri svojem delovanju pošilja povpraševanje na RESTful strežnik, ki je ločen od uporabniškega vmesnika, od katerega dobi informacije o predvidenih vezavnih mestih na proteinu. Za komunikacijo se uporablja AJAX, sporočila pa so poslana v JSON formatu.

Vhodni podatki za RESTful strežnik so koda proteina in veriga. Izvorna koda povpraševanja:

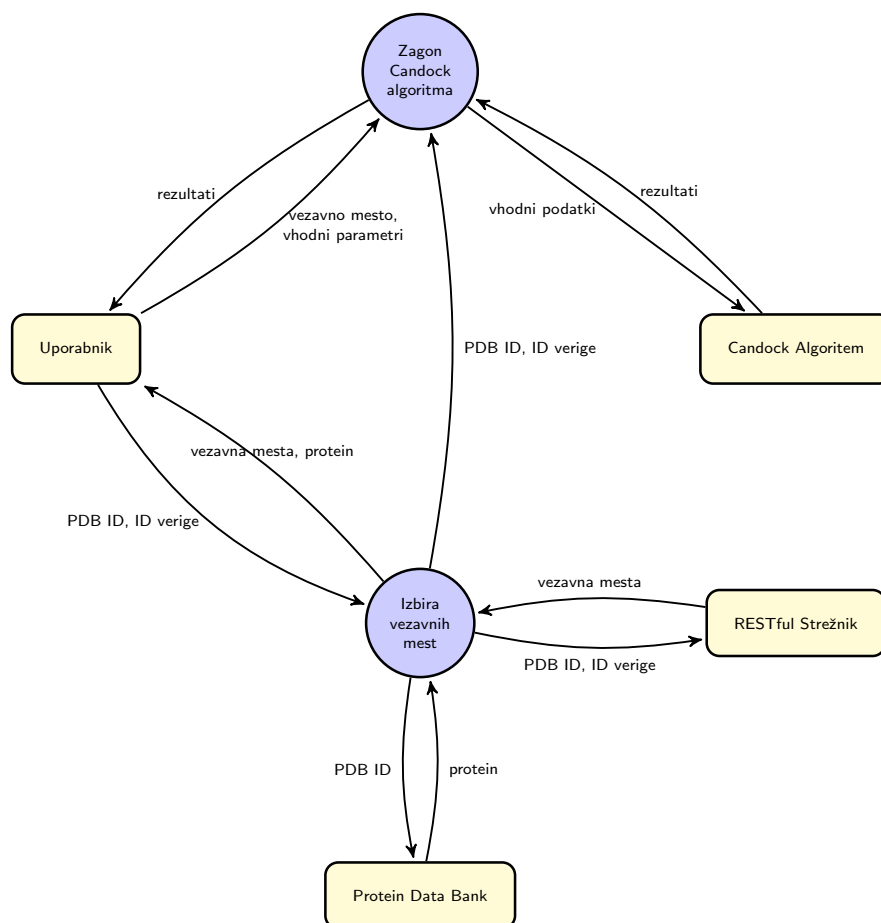
```
$.ajax({
  type: 'post',
  data: {'protein': proteinID, 'chain': chainID},
  url: 'http://tyr.cmm.ki.si/api.php',
  success: function(data) {
    var centroids = $.parseJSON(data);
    for (i = 0; i < centroids.length; i++) {
      centroids[i].push("true");
      known_centroids.push(centroids[i]);
    }
    $.fn.ligChooser.addKnownSitesProbis();
  }
});
```

RESTful strežnik kot rezultat vrne geometrijske centre vezavnih mest na proteinu in premer vezavnega mesta. Primer odgovora:

```
{["15.831", "12.169", "-36.610", "18.657"]}
```

4.3 Procesi aplikacije

Predstavljeni so procesi aplikacije in diagram podatkovnega toka nivo 1 (slika 5). Procesi aplikacije so razdeljeni na procese, ki se nanašajo na vnos vhodnih podatkov, algoritem za molekularno sidranje Candock in na rezultate molekularnega sidranja.



Slika 5: Diagram podatkovnega toka nivo 1

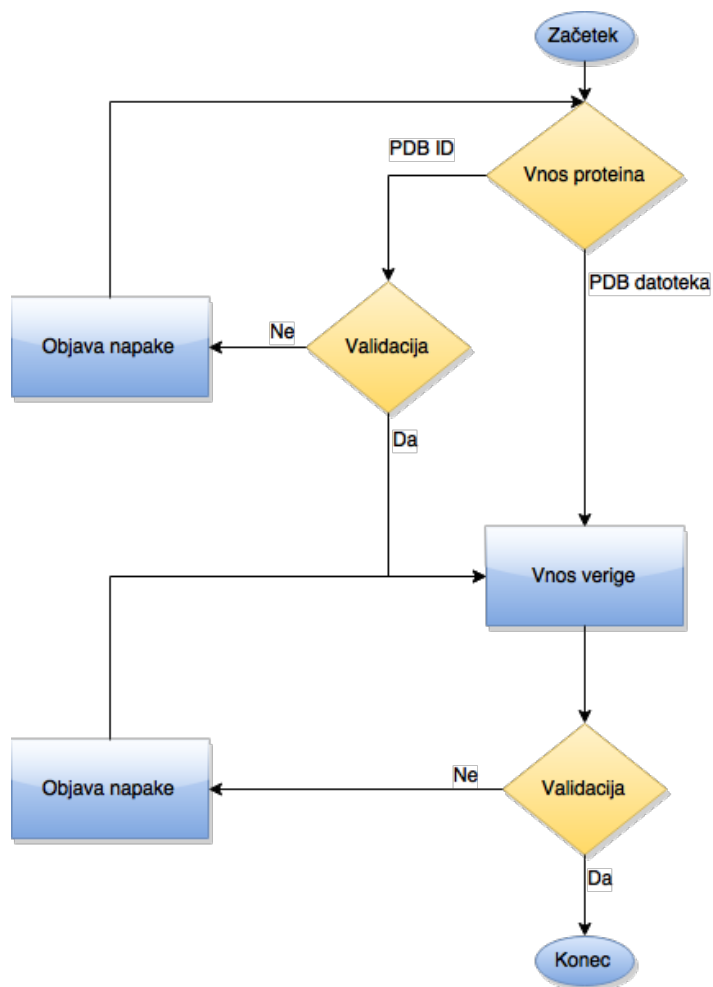
4.3.1 Vnos vhodnih podatkov

Procesi vnosa vhodnih podatkov so:

- Izbira proteina in verige
- Izbira knjižnice spojin
- Izbira vezavnih mest
- Sprememba parametrov algoritma

Izbira proteina in verige

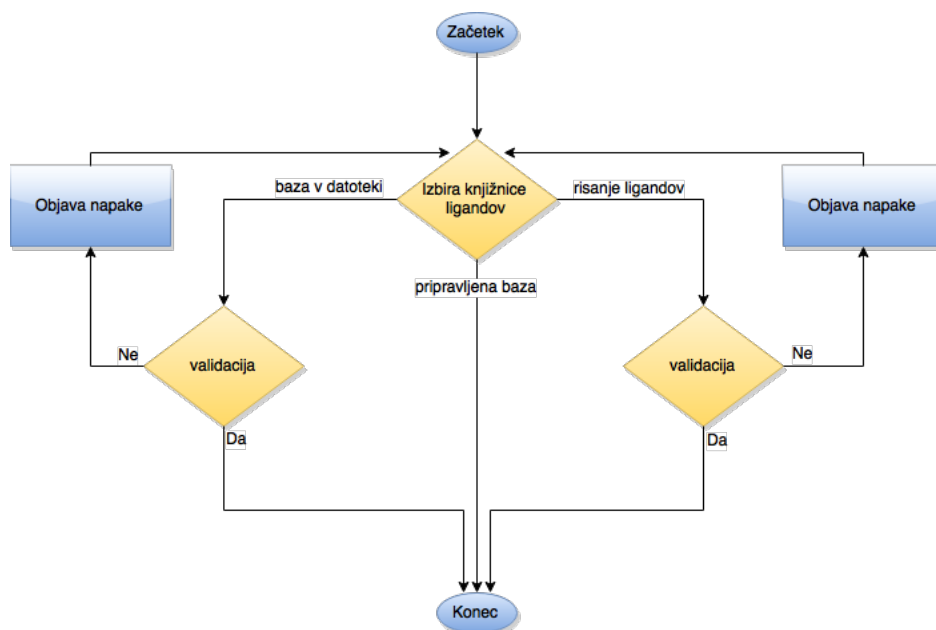
Uporabnik lahko vnese podatke o proteinu, na katerem želi izvesti molekularno sidranje s pomočjo PDB ID-ja ali preko PDB datoteke. PDB ID je koda sestavljena iz 4 znakov, ki enolično določa protein v Protein Data Bank. Za izbrani protein je potrebno določiti še tarčno verigo (chain ID) (slika 6).



Slika 6: Aktivnostni diagram izbire proteina in verige

Izbira knjižnice spojin

Knjižnico ligandov uporabnik bodisi izbere med že pripravljenimi bazami, na primer izbira lahko med Cando bazo in ZINC podatkovnimi bazami, bodisi naloži svojo knjižnico ligandov v obliki mol2 ali smi datoteki. Na razpolago ima tudi Marvin JS orodje, v katerem lahko nariše kemijske strukture ligandov, ki jih želi sidrati. Aktivnostni diagram izbire knjižnice je prikazan na sliki 7.



Slika 7: Aktivnostni diagram izbire knjižnice spojin

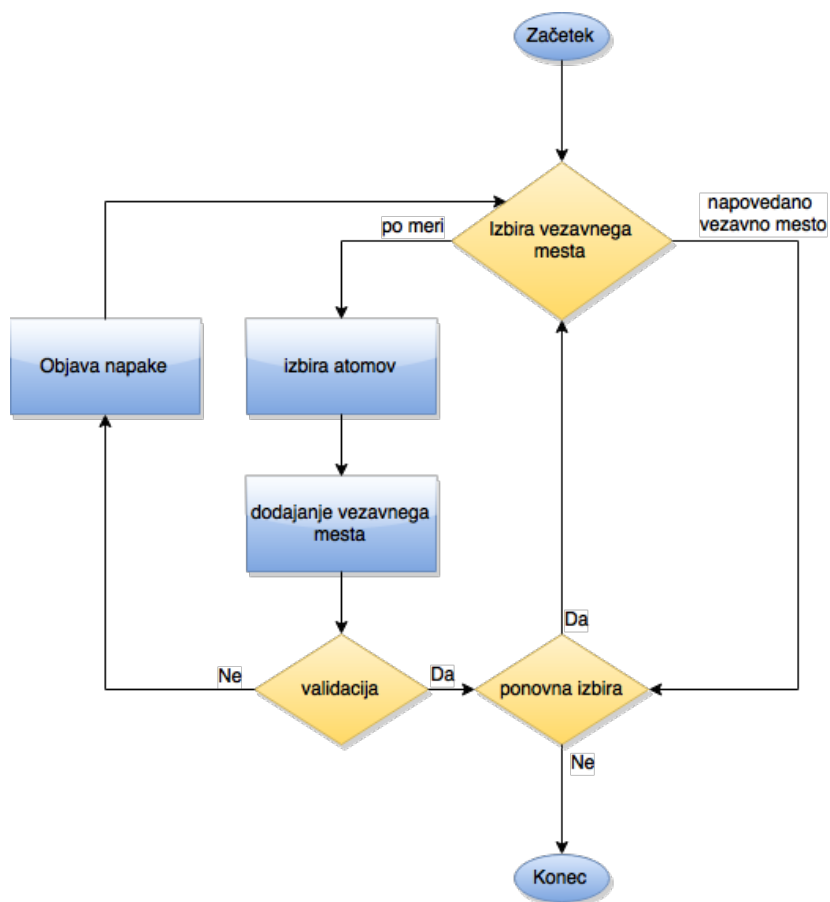
Izbira vezavnih mest

Uporabnik vezavna mesta izbira med vnaprej napovedanimi vezavnimi mesti, ki se prenesejo iz RESTful strežnika, lahko pa jih izbere sam in sicer tako, da z miško klikne na atome proteina, za katere želi, da bodo uporabljeni pri izračunu geometrijskega središča vezavnega mesta. Določi še radij vezavnega mesta. Centroid vezavnih mest se izračuna kot povprečje koordinat izbranih atomov. Izvorna koda izračuna vezavnega mesta:

```
var premer = eval( $(".output").html() * 2 );
var cen_x = 0;
var cen_y = 0;
var cen_z = 0;

// polje atomov, ki so bili izbrani v JSmolu
var atoms = Jmol.getPropertyAsArray(eval("jmolApplet" + String(opts.id)),
    "atomInfo", "selected");

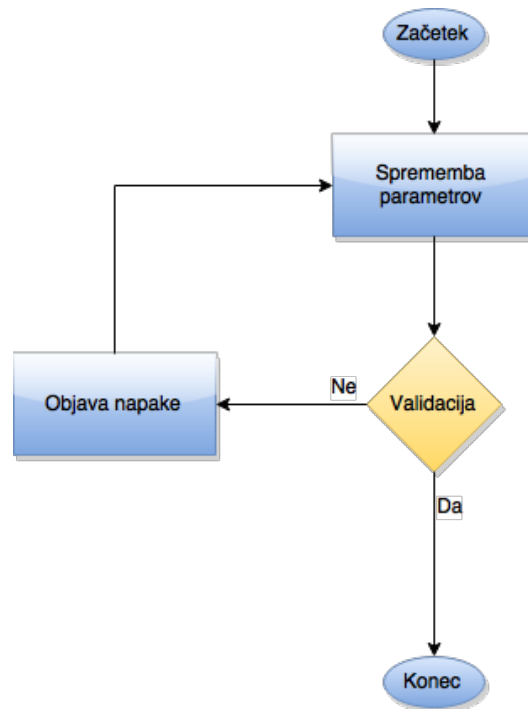
for (i = 0; i < atoms.length; i++) {
    cen_x += atoms[i].x;
    cen_y += atoms[i].y;
    cen_z += atoms[i].z;
}
cen_x /= atoms.length;
cen_y /= atoms.length;
cen_z /= atoms.length;
```



Slika 8: Aktivnostni diagram izbire vezavnih mest

Sprememba parametrov algoritma

Uporabnik lahko spremeni vhodne parametre Candock algoritma za molekularno sidranje, lahko pa pusti prednastavljene vrednosti parametrov (slika 9).



Slika 9: Aktivnostni diagram spremembe parametrov algoritma

4.3.2 Candock algoritem za molekularno sidranje

Candock algoritem se zažene v ozadju uporabniškega vmesnika. Za zagon programa so zahtevani naslednji vhodni parametri:

- **ligand**

Opis: Lokacija mol2 datoteke, ki vsebuje podatke o ligandih

Tip: String

- **receptor**

Opis: Lokacija pdb datoteke, ki vsebuje podatke o proteinu

Tip: String

- **centroid**

Opis: Lokacija txt datoteke, ki vsebuje podatke o izbranih vezavnih mestih

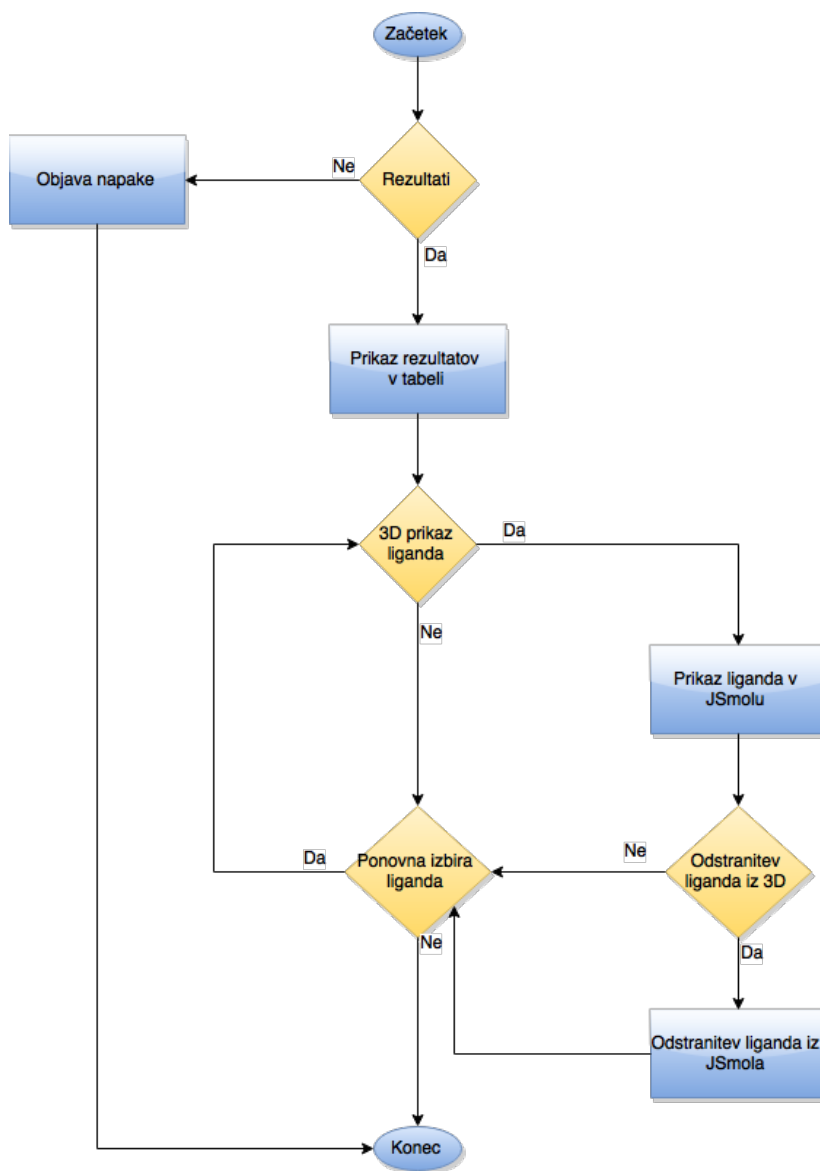
Tip: String

Ukaz za zagon programa:

```
./candock --ncpu $NCPU --ligand $ligandi --receptor $BASE/$RESULTS_DIR/$pdb_file1 --centroid
$BASE/$RESULTS_DIR/centroid.txt --num_iter $num_iter --max_seeds_to_cluster $max_seeds_to_cluster
--max_num_clus $max_num_clus --min_pts $min_pts --clus_rad $clus_rad --top_percent $top_percent
--interatomic $interatomic --excluded $excluded --grid $grid --num_bsites $num_bsites --radial
$radial --max_iter $max_iter --mini_tol $mini_tol --fftype $fftype --docked_min_pts
$docked_min_pts --docked_max_num_clus $docked_max_num_clus --docked_clus_rad $docked_clus_rad
--tol_min_coeff $tol_min_coeff --tol_max_coeff $tol_max_coeff --tol_dist $tol_dist --spin $spin
--max_num_ligands $max_num_lig --scale $scale --step $step --cutoff $cutoff --func $func --comp
$comp --ref $ref --probis_min_pts $probis_min_pts --probis_clus_rad $probis_clus_rad --neighb
$neighb
```


4.3.3 Prikaz rezultatov molekularnega sidranja

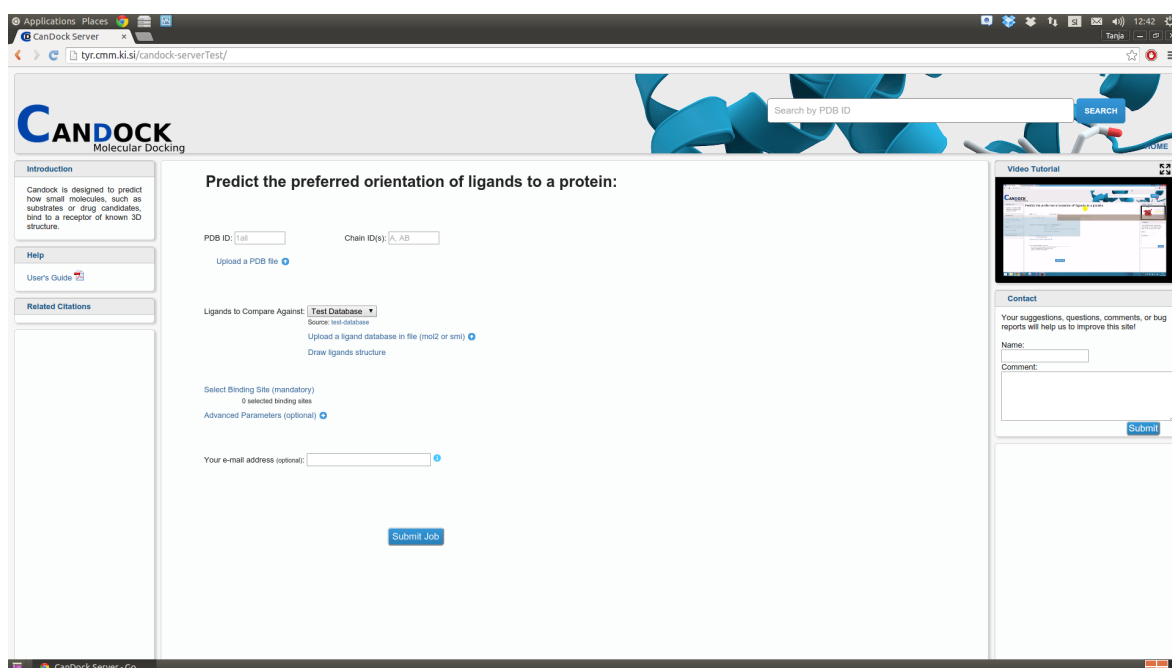
Candock algoritem vrne rezultate v dveh datotekah: energies.txt in minimized.pdb. Prva datoteka vsebuje seznam ligandov in oceno njihove proste energije vezave s tarčnim proteinom. V minimized.pdb datoteki se nahajajo koordinate atomov sidranih ligandov skupaj s pripadajočim proteinom za prikaz ligand-protein kompleksov. Na sliki 10 je prikazan proces prikaza rezultatov.



Slika 10: Aktivnostni diagram prikaza rezultatov

5 Opis rešitve

Candock spletni uporabniški vmesnik je dostopen na url naslovu: <http://tyr.cmm.ki.si/candock-serverTest/>.



Slika 11: Candock spletni uporabniški vmesnik

5.1 Vnos vhodnih podatkov

Spletni uporabniški vmesnik odlikuje hiter in enostaven vnos vhodnih podatkov. Obvezni vhodni podatki so: koda proteina in verige, baza ligandov ter vezavno mesto.

5.1.1 Vnos proteina in verige

Tarčni protein se lahko določi bodisi preko PDB ID-ja bodisi s pdb datoteko. PDB ID se zapiše v za to namenjeno okence. S klikom na *Upload a PDB file* gumb je omogočen prenos pdb datoteke. Poleg proteina mora uporabnik izbrati še njegovo verigo.

5.1.2 Izbira baze ligandov

Knjižnica ligandov je lahko določena na tri načine:

1. Izbira ZINC ali Cando knjižnice
2. Ligandi po meri
3. Risanje ligandov

ZINC knjižnica

Knjižnica ligandov je lahko izbrana iz spletne strani <http://zinc.docking.org/>. Na omejnjeni spletni strani se nahajajo prosto dostopne knjižnice spojin, izbrane spojine je mogoče tudi kupiti. Uporabnik lahko izbira med naslednjimi bazami: Leads Now, Frags Now in Drugs Now.

Cando knjižnica

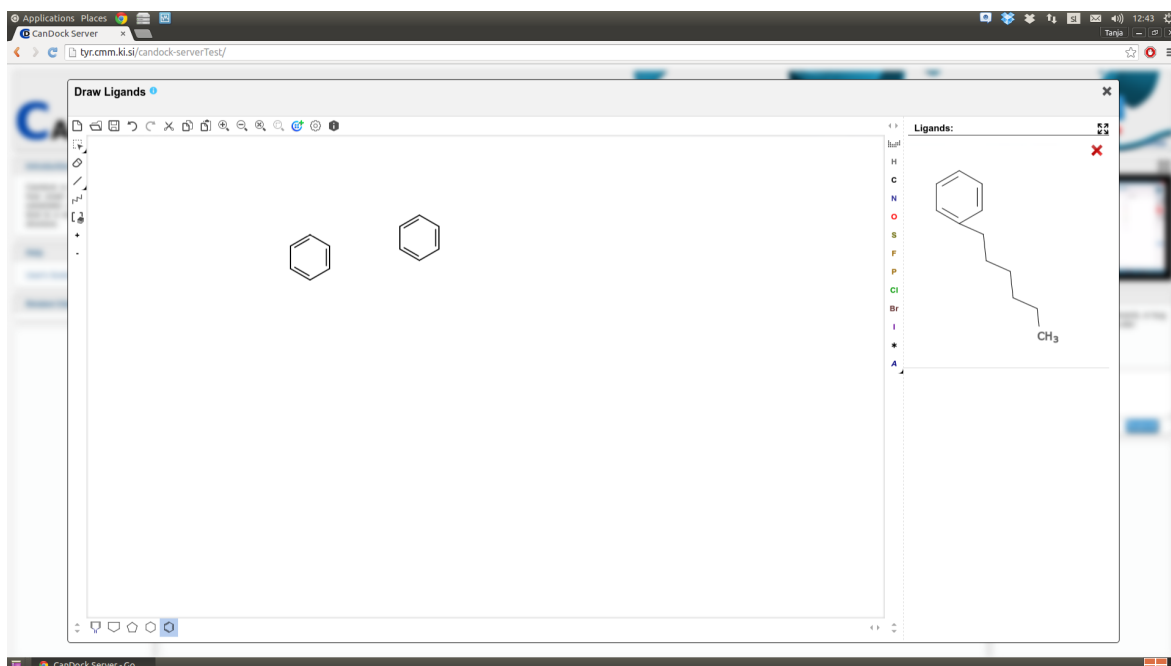
Cando knjižnica vsebuje 3421 molekul, za katere so bile predvidene ustrezne lastnosti, ki so potrebne pri kandidatnih spojinah za potencialna nova zdravila.

Ligandi po meri

S klikom na *Upload a ligand database in file (mol2 or smi)* gumb je omogočena naložitev ligandov preko datoteke. Sprejemljiva datotečna formata sta mol2 in smi.

Risanje ligandov

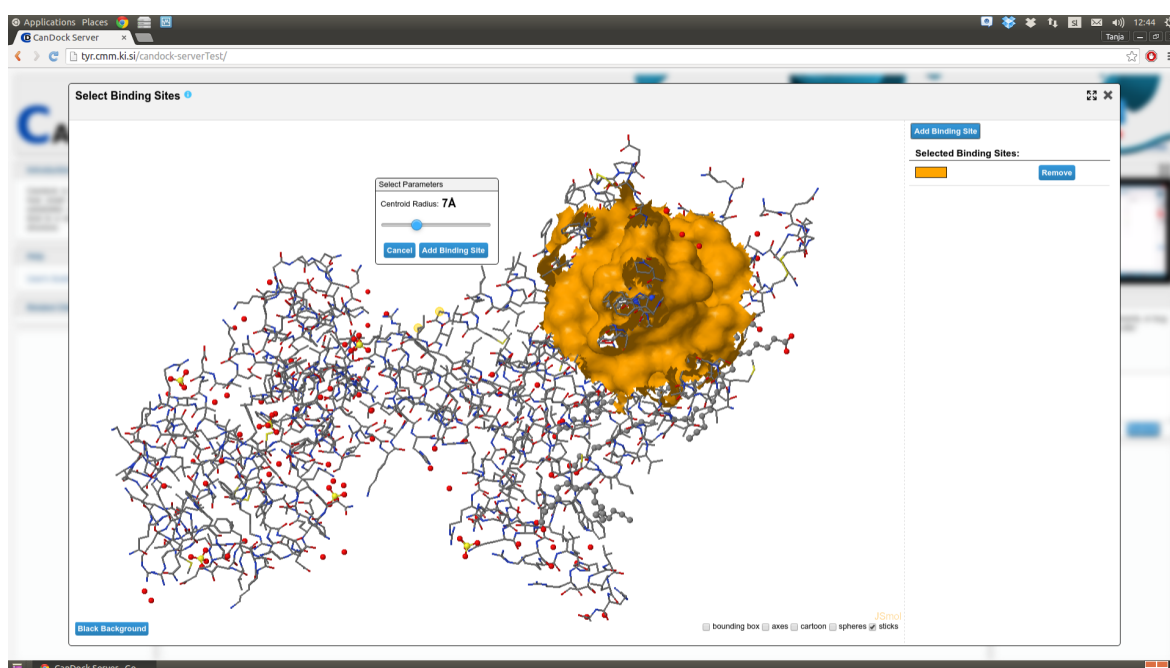
Ob kliku na *Draw ligands structure* se prikaže okno z Marvin JS urejevalnikom, ki omogoča risanje molekulskih struktur. S klikom na modri gumb z zelenim plusom, v zgornji orodni vrstici, se ligand doda v množico spojin, ki bo uporabljena v postopku molekularnega sidranja. Slike izbranih ligandov se prikažejo na desni strani. Uporabnik ligande, za katere ne želi izvesti postopka sidranja, odstrani preko klika na X gumb. Na sliki 12 je prikazana uporaba Marvin JS orodja.



Slika 12: Risanje ligandov v Marvin JS orodju

5.1.3 Izbira vezavnih mest

Za določitev vezavnega mesta je potrebno klikniti na gumb *Select Binding Sites (mandatory)*. Prikaže se okno (slika 13), ki vsebuje tridimenzionalnim prikaz proteina. Za enostavnejši postopek vnosa vhodnih podatkov so prikazana napovedana vezavna mesta. Privzeto so ta mesta izbrana za postopek molekularnega sidranja. Uporabnik izbrana aktivna mesta odstrani s klikom na gumb *Remove*. Vezavna mesta lahko doda tudi sam. Po izbiri željenih atomov se ob kliku na gumb *Add Binding Site*, prikaže okno v katerem se preko drsnika določi radij vezavnega mesta.



Slika 13: Izbira vezavnih mest na 3EML proteinu


5.1.4 Spreminanje vhodnih parametrov algoritma

Ob kliku na *Advanced Parameters (optional)* se prikaže tabela s parametri za Candock algoritem. V okencih so napisane njihove privzete vrednosti. Ob prisotnosti miške se uporabniku odpre okno, ki vsebuje opis in tip posameznega parametra. Parametri so na sliki 14.































5.1.5 Vnos elektronskega naslova


Vnos elektronskega naslova ni obvezen, je pa priporočljiv. Ker je proces molekularnega sidranja lahko časovno zahteven, se ob zaključku postopka uporabniku na vnešeni

elektronski naslov pošlje elektronska pošta z url povezavo na spletno stran, kjer se nahajajo rezultati načrtovanega procesa.

Advanced Parameters (optional) 

Algorithm Parameters

Num Iter:	<input type="text" value="1000"/>		Max Seeds To Cluster:	<input type="text" value="2000"/>	
Max Num Clus:	<input type="text" value="500"/>		Min Pts:	<input type="text" value="5"/>	
Clus Rad:	<input type="text" value="1.0"/>		Top Percent:	<input type="text" value="0.80"/>	
Interatomic	<input type="text" value="8.0"/>		Excluded:	<input type="text" value="0.8"/>	
Grid:	<input type="text" value="2.0"/>		Num Bsites:	<input type="text" value="1"/>	
Radial:	<input type="text" value="10.0"/>		Max Iter:	<input type="text" value="0"/>	
Mini Tol:	<input type="text" value="0.1"/>		Fftype:	<input type="text" value="kb"/>	
Docked Min Pts:	<input type="text" value="3"/>		Docked Max Num Clus:	<input type="text" value="10"/>	
Docked Clus Rad:	<input type="text" value="2.0"/>		Tol Min Coeff:	<input type="text" value="0.7"/>	
Tol Max Coeff:	<input type="text" value="1.0"/>		Tol Dist:	<input type="text" value="3.0"/>	
Spin:	<input type="text" value="15"/>		Max Num Ligands:	<input type="text" value="1000"/>	
Scale:	<input type="text" value="0.01"/>		Step:	<input type="text" value="0.1"/>	
Cutoff:	<input type="text" value="8"/>		Func:	<input type="text" value="radial"/>	
Comp:	<input type="text" value="reduced"/>		Ref:	<input type="text" value="mean"/>	
Probis Min Pts:	<input type="text" value="10"/>		Probis Clus Rad:	<input type="text" value="2.0"/>	
Neighb:	<input type="text"/>				

Your e-mail address (optional): 

Slika 14: Spreminjanje parametrov algoritma in vnos elektronskega naslova

5.2 Zagon Candock algoritma

Proces molekularnega sidranja med proteinom in bazo ligandov se zažene s klikom na gumb *Submit Job*.

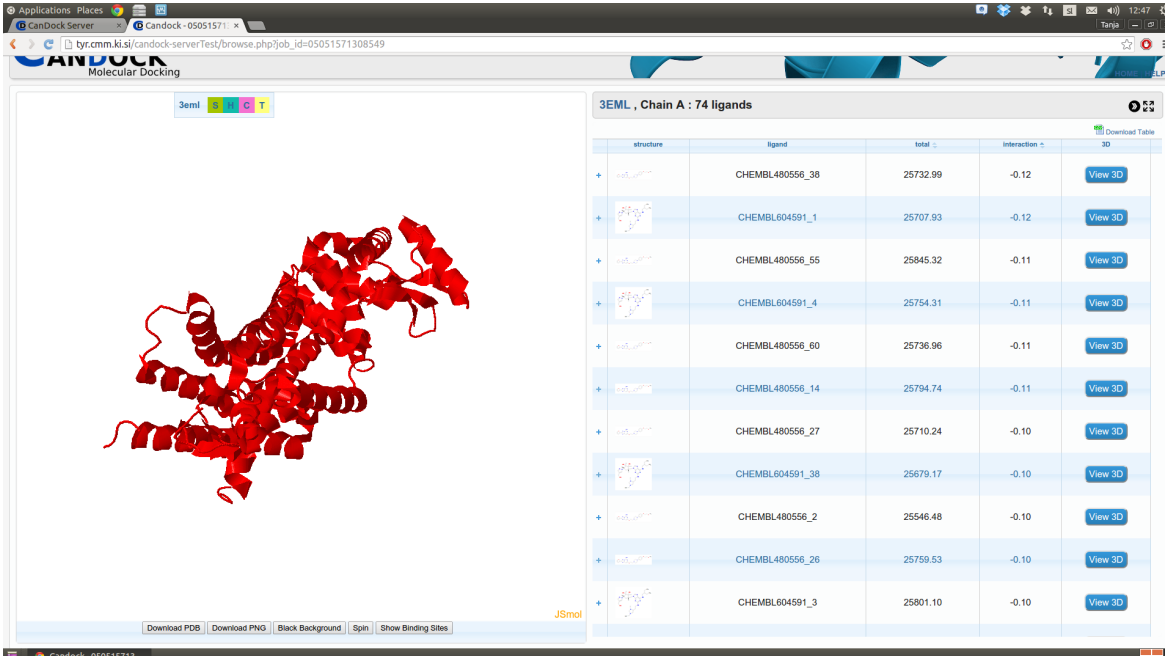
5.3 Rezultati

5.3.1 Prikaz rezultatov

Vzorčna stran (slika 15) z rezultati je dostopna na:

http://tyr.cmm.ki.si/candock-serverTest/browse.php?job_id=20071543739493

Na desni strani se nahaja seznam ligandov, urejenih po energijski oceni interakcij. Nižja energija predstavlja stabilnejši kompleks. Na levi strani je JSmol okno namenjeno tridimenzionalnemu prikazu molekul.


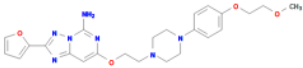


The screenshot displays the Candock Molecular Docking web interface. On the left, a 3D protein structure is shown in red. On the right, a table titled "3EML, Chain A : 74 ligands" lists the results of the docking simulation. The table has columns for "structure", "ligand", "total", "interaction", and "3D". Each row includes a plus icon for expansion and a "View 3D" button. The table data is as follows:

structure	ligand	total	interaction	3D
+	CHEMBL480556_38	25732.99	-0.12	View 3D
+	CHEMBL604591_1	25707.93	-0.12	View 3D
+	CHEMBL480556_55	25845.32	-0.11	View 3D
+	CHEMBL604591_4	25754.31	-0.11	View 3D
+	CHEMBL480556_60	25736.96	-0.11	View 3D
+	CHEMBL480556_14	25794.74	-0.11	View 3D
+	CHEMBL480556_27	25710.24	-0.10	View 3D
+	CHEMBL604591_38	25679.17	-0.10	View 3D
+	CHEMBL480556_2	25546.48	-0.10	View 3D
+	CHEMBL480556_26	25759.53	-0.10	View 3D
+	CHEMBL604591_3	25801.10	-0.10	View 3D

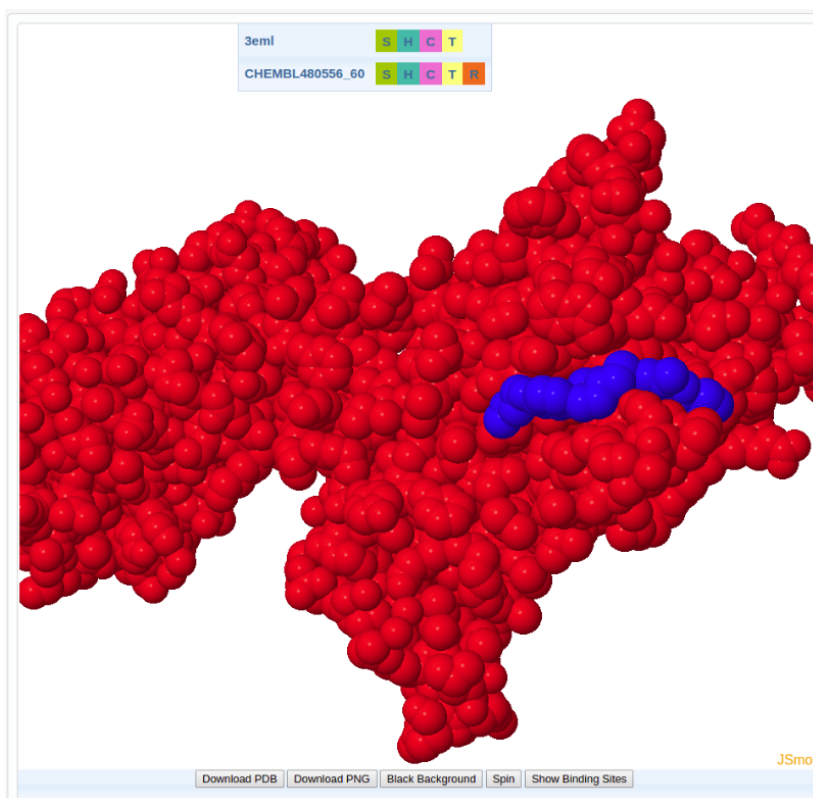
Slika 15: Primer prikaza rezultatov molekularnega sidranja

S klikom na plus ikono lahko v tabeli razširimo vrstico za posamezni ligand (slika 16). V razširjeni vrstici je podana SMILES koda in povezava na ustrezno spletno stran z opisom molekule. S klikom na sliko liganda, se le-ta poveča.

	CHEMBL604591_1	25707.93	-0.12	View 3D
SMILES: <chem>CCC(CC)CNc1c2c(nc(n1)NCCc1cn(c[nH]1)CC)n(cn2)[C@@H]1C[C@@H]([C@H]([C@H]1O)n1cc(cn1)CO</chem>		Source ▼		
CHEMBL480556_55		close		
	CHEMBL		-0.11	View 3D

Slika 16: Primer prikaza rezultatov molekularnega sidranja

S klikom na *View 3D* gumb se posamezni ligand prikaže v JSmol oknu. V tridimenzionalni grafiki vidimo interakcijo med ligandom in proteinom.



Slika 17: Primer interakcije med proteinom in ligandom. Protein je prikaz v rdeči, ligand v modri barvi.

Za prikaz vezavnih mest, ki so bila izbrana pri postopku molekularnega sidranja, je potrebno klikniti na gumb *Show Binding Sites*.

5.3.2 Prenos rezultatov

S klikom na gumb *Download Table* se prenese seznam ligandov v csv datotečnem formatu. Datoteka poleg imena liganda, SMILES kode in interakcije vsebuje še dodatne informacije, ki ocenjujejo stabilnost vezave. Na sliki 18 je prikazan primer prenesene datoteke.

Ligand	total	interaction	internal	lossion	external	torsion	smiles
CHEMBL187326_1	25618.94	-0.03	14.71	314.59	90.34	25399.4	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL187326_2	25604.7	-0.03	44.71	298.16	94.09	25367.8	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL187326_3	25614.78	-0.03	56.16	276.81	94.06	25387.8	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL187326_4	25614.39	-0.04	19.04	301.11	94.23	25400.1	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL187326_5	25603.88	-0.04	45.57	222.6	138.07	25311.3	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL121242_2	25625.18	-0.09	11.73	274.07	138.06	25201.4	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL121242_3	25769.95	-0.06	54.64	229.96	97.96	25387.5	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL121242_1	25792.07	-0.08	17.47	231.42	92.12	25391.2	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL121242_16	25687.47	-0.08	12.09	305	147.49	25223	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL121242_14	25632.02	-0.07	24.74	204.54	149.7	25261.2	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL121242_27	25945.14	-0.07	46.14	404.53	107.27	25397.3	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL121242_3	25785.96	-0.06	23.11	304.11	128.17	25240.5	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL121242_7	25628.19	-0.07	17.23	304.11	130.75	25376.2	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL121242_18	25913.6	-0.05	52.46	412.67	137.71	25310.9	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL121242_3	25792.77	-0.06	27.65	389.64	135.95	25239.6	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL121242_27	25971.14	-0.07	11.99	487.97	110.56	25366.7	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL121242_16	25689.46	-0.08	12.1	304.91	149.74	25222.8	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL121242_1	25742.35	-0.08	15.52	244.72	94.06	25388.2	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL121242_13	25785.34	-0.06	31.09	244.79	128.29	25302.3	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL121242_2	25622.15	-0.09	11.72	274.33	136.12	25200.1	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL121242_7	25623.78	-0.07	22.64	173.88	87.3	24740.1	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL121242_18	25959.52	-0.06	52.46	444.62	141.8	25311.8	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL121242_14	25647.23	-0.07	22.79	209.5	151.63	25263.4	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL121242_6	25931.5	-0.04	43	188.18	146.06	25154.3	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)O
CHEMBL489556_54	25636.18	-0.1	36.75	296.22	166.23	25147.1	COCCOC1=CC=CC=C1N1CCN(C1)CCOC1=CC=CC=C1
CHEMBL489556_100	25709.02	-0.09	23.81	280.82	126.06	25278.5	COCCOC1=CC=CC=C1N1CCN(C1)CCOC1=CC=CC=C1
CHEMBL489556_38	25732.99	-0.12	20.02	249.1	108.18	25395.9	COCCOC1=CC=CC=C1N1CCN(C1)CCOC1=CC=CC=C1
CHEMBL489556_40	25993.39	-0.09	40.38	465.4	90.19	25397.6	COCCOC1=CC=CC=C1N1CCN(C1)CCOC1=CC=CC=C1
CHEMBL489556_70	25869.14	-0.09	20.33	331.29	90.33	25427.4	COCCOC1=CC=CC=C1N1CCN(C1)CCOC1=CC=CC=C1
CHEMBL489556_64	25905.24	-0.09	19.44	196.39	130.19	25196.4	COCCOC1=CC=CC=C1N1CCN(C1)CCOC1=CC=CC=C1
CHEMBL489556_121	25654.79	-0.1	15.57	258.63	144.06	25298.7	COCCOC1=CC=CC=C1N1CCN(C1)CCOC1=CC=CC=C1
CHEMBL489556_3	25927.28	-0.09	31.54	183.18	77.72	24795	COCCOC1=CC=CC=C1N1CCN(C1)CCOC1=CC=CC=C1
CHEMBL489556_35	25845.32	-0.11	33.29	455.16	128.19	25228.8	COCCOC1=CC=CC=C1N1CCN(C1)CCOC1=CC=CC=C1
CHEMBL489556_26	25759.53	-0.1	20.62	388.35	179.96	25170.7	COCCOC1=CC=CC=C1N1CCN(C1)CCOC1=CC=CC=C1
CHEMBL489556_2	25646.48	-0.1	19.81	191.74	132.09	25203	COCCOC1=CC=CC=C1N1CCN(C1)CCOC1=CC=CC=C1
CHEMBL489556_44	25698.96	-0.09	20.02	258.03	160.23	25166.8	COCCOC1=CC=CC=C1N1CCN(C1)CCOC1=CC=CC=C1
CHEMBL489556_47	25702.88	-0.08	41.61	318.45	160.33	25182.6	COCCOC1=CC=CC=C1N1CCN(C1)CCOC1=CC=CC=C1
CHEMBL489556_60	25736.96	-0.11	28.43	232.6	106.22	25371.9	COCCOC1=CC=CC=C1N1CCN(C1)CCOC1=CC=CC=C1
CHEMBL489556_14	25794.14	-0.11	48.42	319.25	118.65	25398.7	COCCOC1=CC=CC=C1N1CCN(C1)CCOC1=CC=CC=C1
CHEMBL489556_125	25749.9	-0.07	11.61	237.6	100.2	25400.6	COCCOC1=CC=CC=C1N1CCN(C1)CCOC1=CC=CC=C1
CHEMBL489556_27	25710.24	-0.1	17.68	325.63	138.21	25228.9	COCCOC1=CC=CC=C1N1CCN(C1)CCOC1=CC=CC=C1
CHEMBL489556_17	25866.65	-0.08	18.09	276.17	82.24	25500.2	COCCOC1=CC=CC=C1N1CCN(C1)CCOC1=CC=CC=C1
CHEMBL489556_73	25992.32	-0.09	20.7	554.12	130.09	25287.6	COCCOC1=CC=CC=C1N1CCN(C1)CCOC1=CC=CC=C1
CHEMBL489556_12	25984.22	-0.06	18.05	206.78	155.87	25196.5	COCCOC1=CC=CC=C1N1CCN(C1)CCOC1=CC=CC=C1
CHEMBL17981_4	25511.64	-0.03	33.65	185.11	122.25	25170.7	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)NCCCC2
CHEMBL17981_2	25682.4	-0.04	19.37	302.23	135.84	25225	lccc(cc1[n]c(ccc2[n]c1)nc2)[C@H]1[C@@H]([C@@H]([C@H]1O)CO)O)NCCCC2

Slika 18: Datoteka z rezultati

Uporabniku je omogočen prenos trenutnih elementov znotraj JSmola v pdb datoteki in trenutni pogled elementov v JSmolu v png datoteki. Slednje naredi s klikom na *Download PDB* oziroma *Download PNG* gumb.

5.3.3 Prilagoditev pogleda

Znotraj JSmola se nahaja meni (slika 19), s seznamom molekul vidnih v tridimenzionalni grafiki, ki omogoča prilagajanje pogleda posameznih molekul.



Slika 19: Meni v JSmolu

5.3.4 Iskanje rezultatov

Iskanje po že izvedenih postopkih molekularnega sidranja se izvede z vnosom PDB ID-ja v iskalni boks in klikom na *Search* gumb. Prikaže se seznam rezultatov, ki ustrezajo iskanim zahtevam.

6 Primer uporabe aplikacije

Spletni uporabniški vmesnik Candock smo testirali na Udp-n-acetylgalactopyranose Mutase proteinu. Proces molekularnega sidranja smo izvedli na testni bazi ligandov "test-database". Omenjena baza vsebuje trinajst ligandov. Seznam molekul se nahaja v prilogi.

6.1 Opis proteinske tarče

Ime: Udp-n-acetylgalactopyranose Mutase

PDB ID: 4MO2

Razred: Izomeraze

Izvor: *Campylobacter jejuni*

Metoda generiranja 3D strukture: Rentgenska kristalografija

Veriga: A,B

Število aminokislin: 368

Piranoze-furanoze mutaze so pomembni encimi v življenjskem ciklu velikega števila mikroorganizmov. Ker so omenjeni proteini odsotni pri sesalcih predstavljajo potencialne tarče za nova zdravila. [20]

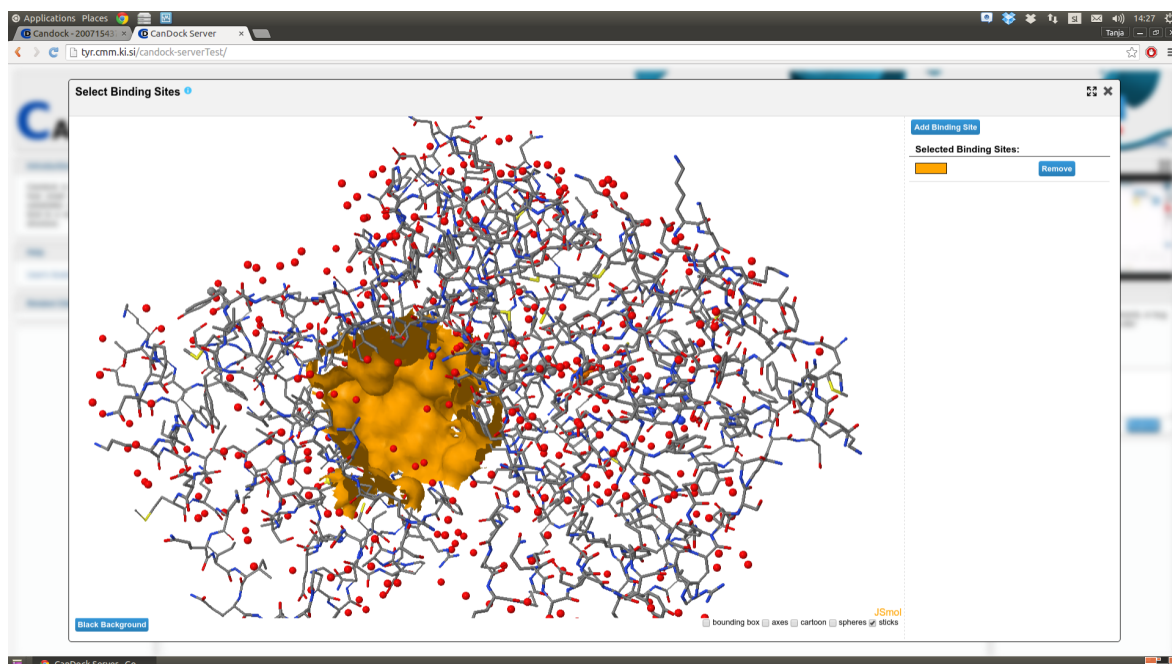
6.2 Vhodni podatki

Centroid vezavnega mesta: 15.831, 12.169, -36.610 (slika 20)

Premer vezavnega mesta: 18.657 Å

Veriga: A

Knjižnica spojin: test-database



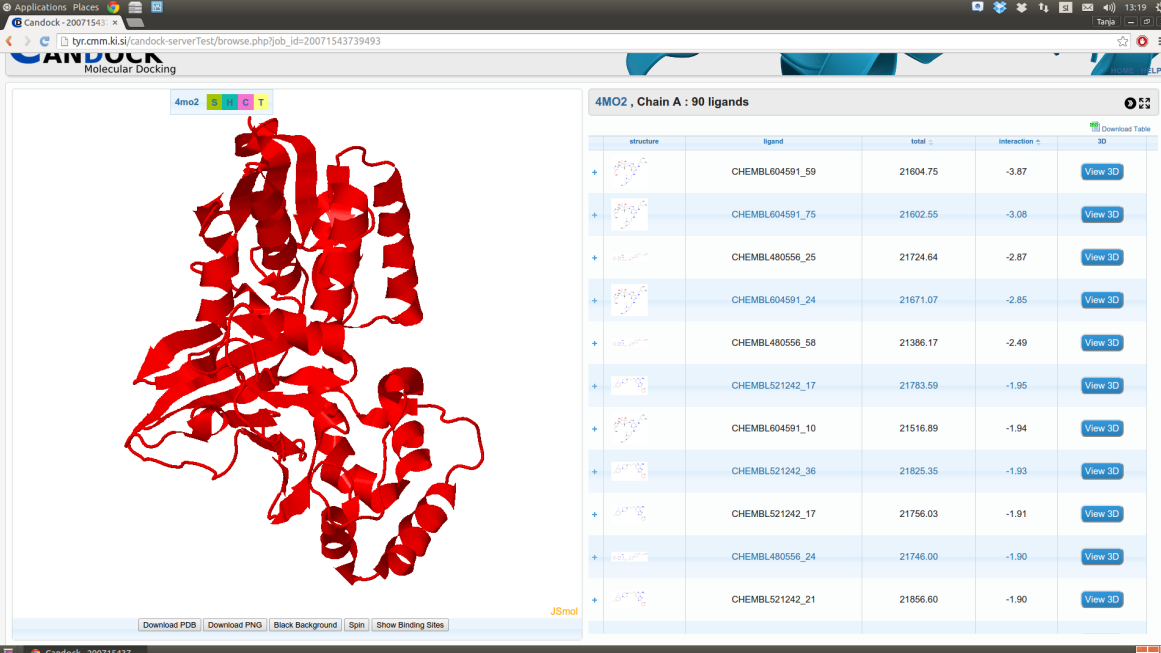
Slika 20: Vezavno mesto na proteinu 4MO2












6.3 Rezultati

Število konformacij ligandov: 90

Najbolje ocenjeni konformaciji: CHEMBL604591_59, CHEMBL604591_75

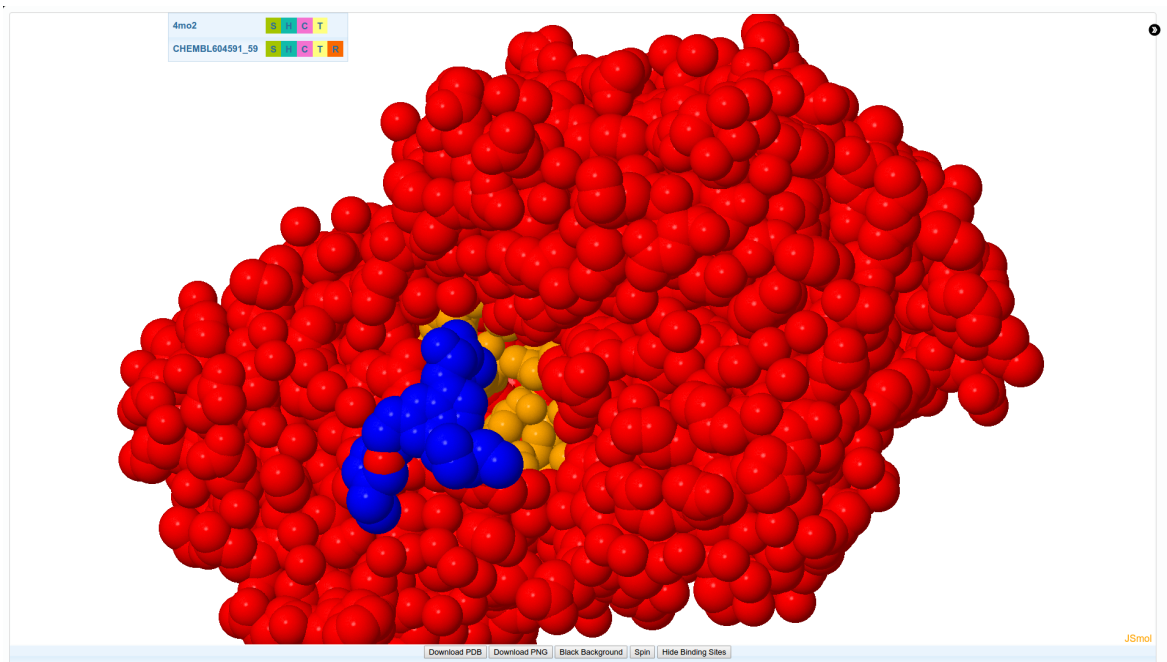
Stran z rezultati je na sliki 21.



structure	ligand	total	interaction	3D
	CHEMBL604591_59	21604.75	-3.87	View 3D
	CHEMBL604591_75	21602.55	-3.08	View 3D
	CHEMBL480556_25	21724.64	-2.87	View 3D
	CHEMBL604591_24	21671.07	-2.85	View 3D
	CHEMBL480556_58	21386.17	-2.49	View 3D
	CHEMBL521242_17	21783.59	-1.95	View 3D
	CHEMBL604591_10	21516.89	-1.94	View 3D
	CHEMBL521242_36	21825.35	-1.93	View 3D
	CHEMBL521242_17	21756.03	-1.91	View 3D
	CHEMBL480556_24	21746.00	-1.90	View 3D
	CHEMBL521242_21	21856.60	-1.90	View 3D

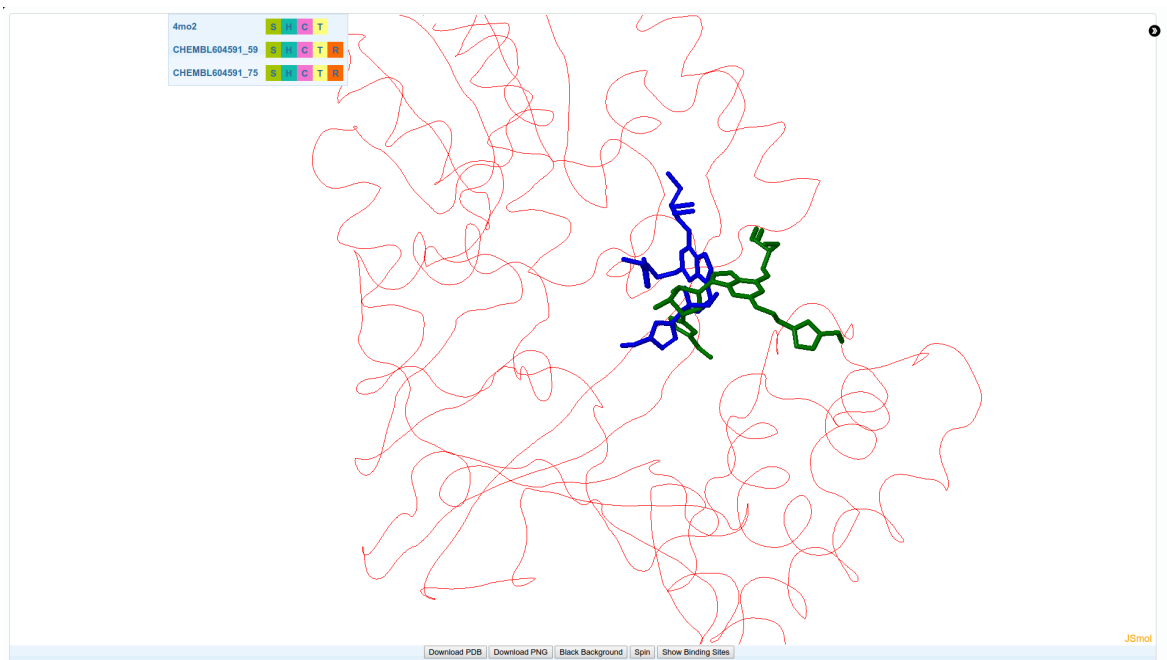
Slika 21: Rezultati molekularnega sidranja za protein 4MO2 in testne baze

Na sliki 22 je predstavljena interakcija med 4MO2 proteinom in napovedano konformacijo liganda s kodo CHEMBL604591_59. Protein je obarvan rdeče barve, ligand modro, z oranžno barvo pa je pobarvano izbrano vezavno mesto.



Slika 22: Prikaz interakcije med 4MO2 in spojino s kodo CHEMBL604591_59

Na sliki 22 sta konformaciji liganda CHEMBL604591 v vezavnem mestu proteina 4MO2.



Slika 23: Prikaz interakcije med 4MO2 in konformacijama CHEMBL604591 liganda

Rezultat molekularnega sidranja sta najbolj ocenjeni konformaciji liganda CHEMBL604591, ki ju predlagamo za eksperimentalno testiranje na inhibicijo vzročnega proteina Udp-n-acetylgalactopyranose Mutase.

7 Zaključek

V zaključni nalogi smo predstavili postopek razvoja novih zdravil z uporabo ene izmed računalniško podprtih metod, to je molekularnega sidranja. Predstavljeni sta arhitektura in razvoj spletnega uporabniškega vmesnika za program molekularnega sidranja, ki je dostopen na <http://tyr.cmm.ki.si/candock-serverTest/>. Podrobneje je opisana njegova funkcionalnost in uporaba. Z razvitim spletnim uporabniškim vmesnikom farmaceutom ponujamo enostavno, prosto dostopno in hitro orodje za izvedbo postopka molekularnega sidranja.

Za nadaljnje delo predlagam implementacijo prikaza interakcij med proteinom in ligandom v hologramu, na primer, s pomočjo Microsoftove HoloLens naprave, ki omogoča prikaz tridimenzionalnih virtualnih objektov, ki jih je mogoče upravljati s premikanjem roke.

Literatura

- [1] H. J. BÖHM, Prediction of Non-bonded Interactions in Drug Design, v: H. J. Böhm, G. Schneider, *Protein-Ligand Interactions, From Molecular Recognition to Drug Design*, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2003, 3–17. (*Citirano na strani 9.*)
- [2] H. J. BÖHM in MARTIN STAHL, The Use of scoring Functions in Drug Discovery Applications, v: H. J. Böhm, G. Schneider, *Protein-Ligand Interactions, From Molecular Recognition to Drug Design*, John Wiley & Sons, Inc. New Jersey, 2002, 41–88. (*Citirano na strani 8.*)
- [3] C. J. CRAMER, Essentials of Computational Chemistry: Theories and Models, *Theoretical Chemistry Accounts* 108(6) (2002), 367–368. (*Citirano na strani 2.*)
- [4] F. ERCOLESSI, *A molecular dynamics primer*, Spring College in Computational Physics, ICTP, Trieste, 1997. (*Citirano na strani 7.*)
- [5] R. C. GLEN in S. C. ALLEN, Ligand-protein docking: cancer research at the interface between biology and chemistry, *Curr Med Chem* 10 (2003), 763–777. (*Citirano na strani 6.*)
- [6] I. HALPERIN, B. MA, H. WOLFSON in R. NUSSINOV, Principles of Docking: An Overview of Search Algorithms and a Guide to Scoring Functions, *PROTEINS: Structure, Function, and Genetics* 47 (2002), 409–443. (*Citirano na strani 8.*)
- [7] R. M. HANSON, J. PRILUSKY, Z. RENJIAN, T. NAKANE in J. L. SUSSMAN, JSmol and the Next-Generation Web-Based Representation of 3D Molecular Structure as Applied to Proteopedia, *Israel Journal of Chemistry* 53 (2013), 207–216. (*Citirano na strani 12.*)
- [8] J. KONC, T. ČEŠNIK, J. T. KONC, M. PENCA in D. JANEŽIČ, ProBiS-Database: Precalculated Binding Site Similarities and Local Pairwise Alignments of PDB Structures, *J Chem Inf Model* 52 (2) (2012), 604–612. (*Citirano na strani 13.*)

- [9] J. KONC in D. JANEŽIČ, ProBiS algorithm for detection of structurally similar protein binding sites by local structural alignment, *Bioinformatics* 26 (9) (2010), 1160–1168. (*Citirano na strani 13.*)
- [10] P. T. LANG, T. AYNECHI, D. MOUSTAKES, B. SHOICHET, I. D. KUNTZ, N. BROOIJMANS in C. M. OSHIRO, Molecular Docking And Structure-Based Design, v: Z. Huang, *Drug Discovery Research: New Frontiers in the Post-Genomic era*, John Wiley & Sons, Inc., New Jersey, 2007, 3–23. (*Citirano na strani 5.*)
- [11] A. R. LEACH in M. M. HANN, The in silico world of virtual libraries, *Drug Discovery Today* 5(8) (2000), 326–336. (*Citirano na strani 10.*)
- [12] S. LESNIK, T. STULAR, B. BRUS, D. KNEZ, S. GOBEC, D. JANEZIC in J. KONC, LiSiCA: A Software for Ligand-based Virtual Screening and its Application for the Discovery of Butyrylcholinesterase Inhibitors, *J Chem Inf Model* (2015), Just Accepted Manuscript. (*Citirano na strani 5.*)
- [13] P. D. LYNE, Structure-based virtual screening: an overview, *Drug Discovery Today* 7(20) (2002), 1047–1055. (*Citirano na straneh 1 in 6.*)
- [14] X. Y. MENG, H. X. ZHANG, M. MEZEI in M. CUI, Molecular Docking: A powerful approach for structure-based drug discovery, *Curr Comput Aided Drug Des* 7(2) (2011), 146–157. (*Citirano na straneh 6, 7 in 9.*)
- [15] K. M. MERZ, D. RINGE in C. H. REYNOLDS, *Drug Design: Structure- and Ligand-Based Approaches*, Cambridge University Press, Cambridge, 2010. (*Citirano na strani 2.*)
- [16] M. D. MILLER, S. K. KEARSLEY, D. J. UNDERWOOD in R. P. SHERIDAN, FLOG: a system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure, *J Comput Aided Mol Des* 8(2) (1994), 153–174. (*Citirano na strani 6.*)
- [17] G. NEUDERT in G. KLEBE, DSX: a knowledge-based scoring function for the assessment of protein-ligand complexes, *J Chem Inf Model* 51 (10) (2011), 2731–2745. (*Citirano na strani 8.*)
- [18] N. M. O'BOYLE, M. BANCK, C. A. JAMES, C. MORLEY, T. VANDERMEERSCH in G-R. HUTCHISON, Open Babel: An open chemical toolbox, *Journal of Cheminformatics* (2011), 3:33. (*Citirano na strani 14.*)

- [19] A. PERDIH, *Dinamični model encima MurD iz E. coli in in silico načrtovanje novih inhibitorjev*, doktorska dizertacija, Univerza v Ljubljani, 2009. (*Citirano na straneh 2, 3, 6 in 10.*)
- [20] M. B. POULIN, Y. SHI, C. PROTSKO, S. A. DALRYMPLE, D. A. SANDERS, B. M. PINTO in T. L. LOWARY, Specificity of a UDP-GalNAc Pyranose-Furanose Mutase: A Potential Therapeutic Target for Campylobacter jejuni Infections, *Chembiochem* 15(1) (2014), 47–56. (*Citirano na strani 34.*)
- [21] G. SCHNEIDER in H. J. BÖHM, Virtual screening and fast automated docking methods, *Drug Discovery Today* 7(1) (2002), 64–70. (*Citirano na strani 5.*)
- [22] D. WEININGER, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, *J Chem Inf Comput Sci* 28(1) (1988), 31–36. (*Citirano na strani 14.*)
- [23] C. G. WERMUTH, C. R. GANELLIN, P. LINDBERG in L. A. MITSCHER, Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998), *Pure and Applied Chemistry* 70(5) (1998), 1129–1143. (*Citirano na strani 5.*)
- [24] (2015) AJAX Introduction. Dostopno 13.7.2015 na: http://www.w3schools.com/ajax/ajax_intro.asp (*Citirano na strani 12.*)
- [25] (2015) CSS Introduction. Dostopno 12.7.2015 na: http://www.w3schools.com/css/css_intro.asp (*Citirano na strani 11.*)
- [26] (2007) Drug Discovery and Development. Dostopno 19.7.2015 na: http://www.phrma.org/sites/default/files/pdf/rd_brochure_022307.pdf (*Citirano na straneh 3 in 4.*)
- [27] (2015) HTML Introduction. Dostopno 12.7.2015 na: http://www.w3schools.com/html/html_intro.asp (*Citirano na strani 11.*)
- [28] (2015) MarvinSketch in JavaScript. Dostopno 13.7.2015 na: <https://www.chemaxon.com/products/marvin/marvin-js/> (*Citirano na strani 13.*)
- [29] (2008) Protein Data Bank Contents Guide: Atomic Coordinate Entry Format Description. Dostopno 13.7.2015 na: ftp://ftp.wwpdb.org/pub/pdb/doc/format_descriptions/Format_v33_A4.pdf (*Citirano na strani 13.*)

- [30] (2015) Statistic Protein Data Bank. Dostopno 12.7.2015 na:
<http://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=total&seqid=100>
(*Citirano na strani 5.*)
- [31] (2015) Tripos Mol2 File Format. Dostopno 13.7.2015 na:
<http://www.tripos.com/data/support/mol2.pdf> (*Citirano na strani 13.*)
- [32] (2015) What can PHP do?. Dostopno 13.7.2015 na:
<https://secure.php.net/manual/en/intro-whatcando.php> (*Citirano na strani 11.*)
- [33] (2015) What is JavaScript?. Dostopno 12.7.2015 na:
https://developer.mozilla.org/en-US/docs/Web/JavaScript/Guide/Introduction#What_is_Java
(*Citirano na strani 11.*)

Priloge

Testna baza ligandov

Ime molekule	Smiles koda
CHEMBL480556	<chem>COCCOc1ccc(cc1)N1CCN(CC1)CCOc1cc2nc(nm2c(n1)N)c1ccco1</chem>
CHEMBL480556	<chem>COCCOc1ccc(cc1)N1CC[NH](CC1)CCOc1cc2nc(nm2c(n1)N)c1ccco1</chem>
CHEMBL521242	<chem>c1cc(oc1)c1nc2cc(nc(n2n1)N)SCCN1CCN(CC1)c1ccc(cc1F)F</chem>
CHEMBL521242	<chem>c1cc(oc1)c1nc2cc(nc(n2n1)N)SCC[NH]1CCN(CC1)c1ccc(cc1F)F</chem>
CHEMBL187326	<chem>c1ccc(cc1)Cn1cc(en1)c1nc(e2c(n1)n(en2)[C@@H]1[C@@H]([C@@H]([C@H](O1)CO)O)O)N</chem>
CHEMBL187326	<chem>c1ccc(cc1)Cn1cc(en1)c1nc(e2c(n1)n(en2)[C@H]1[C@@H]([C@@H]([C@H](O1)CO)O)O)N</chem>
CHEMBL22113	<chem>Cn1c(c2c(n1)nc(n1c2nc(n1)c1ccco1)NC(=O)Nc1ccc(cc1)OC)SC</chem>
CHEMBL87981	<chem>c1cc2c(cc1CCNc1nc(c3c(n1)n(en3)[C@@H]1[C@@H]([C@@H]([C@H](O1)CO)O)O)N)CCCC2</chem>
CHEMBL87981	<chem>c1cc2c(cc1CCNc1nc(c3c(n1)n(en3)[C@H]1[C@@H]([C@@H]([C@H](O1)CO)O)O)N)CCCC2</chem>
CHEMBL124788	<chem>CCCCc1nc2c(c3n1nc(n3)c1ccccc1Cl)nc[nH]2</chem>
CHEMBL55196	<chem>CCCN1c2c(c(=O)n(c1=O)CCC)[nH]c(n2)c1ccc(cc1)OCC(=O)NCCNC(=O)[C@H](Cc1cccs1)[NH3]</chem>
CHEMBL55196	<chem>CCCN1c2c(c(=O)n(c1=O)CCC)[nH]c(n2)c1ccc(cc1)OCC(=O)NCCNC(=O)[C@H](Cc1cccs1)N</chem>
CHEMBL55196	<chem>CCCN1c2c(c(=O)n(c1=O)CCC)[nH]c(n2)c1ccc(cc1)OCC(=O)NCCNC(=O)[C@@H](Cc1cccs1)[NH3]</chem>
CHEMBL55196	<chem>CCCN1c2c(c(=O)n(c1=O)CCC)[nH]c(n2)c1ccc(cc1)OCC(=O)NCCNC(=O)[C@@H](Cc1cccs1)N</chem>
CHEMBL604591	<chem>CCC(CC)CNc1c2c(nc(n1)NCCc1en(c[nH]1)CC)n(en2)[C@@H]1C[C@@H]([C@H]([C@H]1O)O)n1cc(en1)CO</chem>
CHEMBL17302	<chem>c1cc(oc1)c1nc2c3enn(c3nc(n2n1)N)Cc1ccc(cc1)F</chem>
CHEMBL68738	<chem>c1nc(e2c(n1)n(en2)[C@@H]1[C@@H]([C@H]([C@H](O1)CO)O)O)NC1CCCC1</chem>
CHEMBL222530	<chem>c1ccc2c(c1)ccc(n2)OC[C@@H]1CC[C@H]2CN(CC[N@@H]2C1)c1[nH]c(=N)n2c(n1)nc(n2)c1ccco1</chem>
CHEMBL222530	<chem>c1ccc2c(c1)ccc(n2)OC[C@@H]1CC[C@H]2CN(CC[N@@H]2C1)c1[nH]c(n2c(n1)nc(n2)c1ccco1)N</chem>
CHEMBL222530	<chem>c1ccc2c(c1)ccc(n2)OC[C@@H]1CC[C@H]2CN(CC[N@@H]2C1)c1[nH]c(=N)n2c(n1)nc(n2)c1ccco1</chem>
CHEMBL21572	<chem>Cn1c(e2c(n1)nc(n1c2nc(n1)c1ccco1)NC(=O)Cc1ccc2c(c1)OCO2)SC</chem>
CHEMBL21572	<chem>Cn1c(e2c(n1)nc(n1c2nc(n1)c1ccco1)N=C(O)Cc1ccc2c(c1)OCO2)SC</chem>
CHEMBL513030	<chem>c1cc(oc1)c1nc2c3enn(c3nc(n2n1)N)CC[NH]1CCc2c(cncn2)C1</chem>

Primer datoteke v mol2 formatu

```
@<TRIPOS>MOLECULE
ZINC39089601
  25  25  0  0  0
SMALL
USER_CHARGES

@<TRIPOS>ATOM
  1 C1      -0.0184  1.5028  0.0103 C.3  1 <0>  -0.1188
  2 C2       0.0021 -0.0041  0.0020 C.2  1 <0>  -0.0801
  3 C3      -1.1493 -0.6561  0.0041 C.2  1 <0>  -0.1548
  4 C4      -1.2057 -2.1572  0.0177 C.3  1 <0>   0.2928
  5 H1      -2.2438 -2.4816 -0.0545 H    1 <0>   0.0957
  6 N1      -0.6246 -2.6614  1.2587 N.2  1 <0>  -0.6131
  7 C5       0.6232 -2.4822  1.5514 C.2  1 <0>   0.4013
  8 C6       1.5600 -1.7901  0.6833 C.2  1 <0>  -0.1440
  9 C7       1.2862 -0.6858 -0.0139 C.2  1 <0>  -0.0453
 10 O1       1.0803 -2.9663  2.7300 O.3  1 <0>  -0.3157
 11 C8       0.1368 -3.6428  3.5632 C.3  1 <0>   0.0422
 12 O2      -0.4689 -2.6706 -1.0939 O.3  1 <0>  -0.3954
 13 C9      -0.7284 -4.0469 -1.3773 C.3  1 <0>   0.0349
 14 H2       1.0042  1.8801  0.0026 H    1 <0>   0.0684
 15 H3      -0.5455  1.8639 -0.8728 H    1 <0>   0.0738
 16 H4      -0.5286  1.8543  0.9071 H    1 <0>   0.0726
 17 H5      -2.0719 -0.0948 -0.0040 H    1 <0>   0.1299
 18 H6       2.5540 -2.2025  0.5923 H    1 <0>   0.1355
 19 H7       2.0697 -0.2679 -0.6286 H    1 <0>   0.1355
 20 H8      -0.6742 -2.9617  3.8211 H    1 <0>   0.0601
 21 H9      -0.2672 -4.5034  3.0301 H    1 <0>   0.0591
 22 H10     0.6321 -3.9793  4.4740 H    1 <0>   0.0955
 23 H11     -0.4655 -4.6524 -0.5099 H    1 <0>   0.0411
 24 H12     -1.7864 -4.1779 -1.6046 H    1 <0>   0.0432
 25 H13     -0.1313 -4.3604 -2.2337 H    1 <0>   0.0855

@<TRIPOS>BOND
  1  1  2  1
  2  1 14  1
  3  1 15  1
  4  1 16  1
  5  2  9  1
  6  2  3  2
  7  3  4  1
  8  3 17  1
  9  4  5  1
 10  4  6  1
 11  4 12  1
 12  6  7  2
```

13	7	8	1
14	7	10	1
15	8	9	2
16	8	18	1
17	9	19	1
18	10	11	1
19	11	20	1
20	11	21	1
21	11	22	1
22	12	13	1
23	13	23	1
24	13	24	1
25	13	25	1

Primer datoteke v PDB formatu

```
HEADER ISOMERASE 11-SEP-13 4M02
TITLE CRYSTAL STRUCTURE OF UDP-N-ACETYL GALACTOPYRANOSE MUTASE FROM
TITLE 2 CAMPYLOBACTER JEJUNI
COMPND MOL_ID: 1;
COMPND 2 MOLECULE: UDP-GALACTOPYRANOSE MUTASE;
COMPND 3 CHAIN: B, A;
COMPND 4 EC: 5.4.99.9;
COMPND 5 ENGINEERED: YES
SOURCE MOL_ID: 1;
SOURCE 2 ORGANISM_SCIENTIFIC: CAMPYLOBACTER JEJUNI SUBSP. JEJUNI;
SOURCE 3 ORGANISM_TAXID: 192222;
SOURCE 4 STRAIN: 11168;
SOURCE 5 GENE: CJ1439C, GLF;
SOURCE 6 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE 7 EXPRESSION_SYSTEM_TAXID: 562;
SOURCE 8 EXPRESSION_SYSTEM_STRAIN: DH5ALPHA;
SOURCE 9 EXPRESSION_SYSTEM_VECTOR_TYPE: PLASMID;
SOURCE 10 EXPRESSION_SYSTEM_PLASMID: PWM1007
KEYWDS UDP-N-ACETYL GALACTOPYRANOSE MUTASE, UNGM, CAPSULAR POLYSACCHARIDES,
KEYWDS 2 BIFUNCTIONAL, DRUG TARGET, FAD, ISOMERASE
EXPDTA X-RAY DIFFRACTION
AUTHOR S.A.DALRYMPLE,C.PROTSKO,M.B.POULIN,T.L.LOWARY,D.A.R.SANDERS
REMARK 2
REMARK 2 RESOLUTION. 2.00 ANGSTROMS.
REMARK 3
REMARK 3 REFINEMENT.
REMARK 3 PROGRAM : PHENIX
REMARK 3 AUTHORS : PAUL ADAMS,PAVEL AFONINE,VICENT CHEN,IAN
REMARK 3 : DAVIS,KRESHNA GOPAL,RALF GROSSE-
...
DBREF 4M02 B 1 368 UNP QQP8H5 QQP8H5_CAMJE 1 368
DBREF 4M02 A 1 368 UNP QQP8H5 QQP8H5_CAMJE 1 368
SEQRES 1 B 368 MET TYR ASP TYR LEU ILE VAL GLY SER GLY LEU PHE GLY
SEQRES 2 B 368 SER ILE PHE ALA TYR GLU ALA THR GLU LYS GLY TYR THR
SEQRES 3 B 368 CYS LEU VAL VAL GLU GLN ARG GLU HIS ILE GLY GLY ASN
SEQRES 4 B 368 CYS TYR THR GLU ASN ILE LYS ASN ILE ASN VAL HIS LYS
SEQRES 5 B 368 TYR GLY ALA HIS ILE PHE ARG THR SER ASP GLN ASN ILE
SEQRES 6 B 368 TRP ASP TYR MET ASN GLN PHE CYS GLU PHE ASN HIS PHE
SEQRES 7 B 368 ILE ASN SER PRO ILE ALA ILE TYR LYS ASP GLU ILE TYR
SEQRES 8 B 368 ASN LEU PRO PHE ASN MET ASN THR PHE SER LYS LEU TRP
SEQRES 9 B 368 GLY ILE LYS THR PRO ASN GLU ALA ARG LYS ILE ILE GLU
...
HET SO4 B 401 5
HET FDA B 402 53
HET GOL B 403 6
HET GOL A 401 6
HET SO4 A 402 5
HET NH4 A 403 1
HET NH4 A 404 1
HET FAD A 405 53
HETNAM SO4 SULFATE ION
HETNAM FDA DIHYDROFLAVINE-ADENINE DINUCLEOTIDE
HETNAM GOL GLYCEROL
HETNAM NH4 AMMONIUM ION
HETNAM FAD FLAVIN-ADENINE DINUCLEOTIDE
HETSYN GOL GLYCERIN; PROPANE-1,2,3-TRIOL
FORMUL 3 SO4 2(O4 S 2-)
FORMUL 4 FDA C27 H35 N9 O15 P2
FORMUL 5 GOL 2(C3 H8 O3)
FORMUL 8 NH4 2(H4 N 1+)
FORMUL 10 FAD C27 H33 N9 O15 P2
```

```

FORMUL 11 HOH *777(H2 O)
HELIX 1 1 GLY B 10 LYS B 23 1 14
HELIX 2 2 GLY B 37 CYS B 40 5 4
HELIX 3 3 ASP B 62 ASN B 70 1 9
HELIX 4 4 ASN B 96 GLY B 105 1 10
HELIX 5 5 THR B 108 GLN B 121 1 14
HELIX 6 6 ASN B 129 GLY B 139 1 11
HELIX 7 7 GLY B 139 ILE B 147 1 9
...
SHEET 1 A 5 LYS B 205 ILE B 207 O
SHEET 2 A 5 CYS B 27 VAL B 30 1 N VAL B 29 O LYS B 205
SHEET 3 A 5 TYR B 2 VAL B 7 1 N ILE B 6 O LEU B 28
SHEET 4 A 5 ALA B 223 PHE B 228 1 O VAL B 227 N LEU B 5
SHEET 5 A 5 VAL B 334 PHE B 336 1 O TYR B 335 N ILE B 226
SHEET 1 B 2 THR B 42 ILE B 45 O
...
CISPEP 1 LEU B 93 PRO B 94 0 -4.94
CISPEP 2 LEU A 93 PRO A 94 0 -2.90
SITE 1 AC1 7 LYS A 148 LYS A 160 NH4 A 404 LYS B 224
SITE 2 AC1 7 HOH B 830 HOH B 879 HOH B 953
SITE 1 AC2 34 VAL B 7 GLY B 8 GLY B 10 LEU B 11
SITE 2 AC2 34 PHE B 12 VAL B 30 GLU B 31 GLN B 32
SITE 3 AC2 34 ARG B 33 GLY B 38 ASN B 39 ALA B 55
SITE 4 AC2 34 HIS B 56 ILE B 57 ARG B 59 THR B 210
SITE 5 AC2 34 ASP B 211 PHE B 212 GLY B 230 LEU B 248
...
ATOM 1 N MET B 1 14.254 -8.949 29.430 1.00 17.53 N
ATOM 2 CA MET B 1 13.039 -9.232 28.675 1.00 22.88 C
ATOM 3 C MET B 1 12.726 -8.089 27.714 1.00 29.38 C
ATOM 4 O MET B 1 12.854 -6.918 28.069 1.00 20.99 O
ATOM 5 CB MET B 1 11.864 -9.461 29.625 1.00 23.18 C
ATOM 6 CG MET B 1 10.640 -10.075 28.969 1.00 32.26 C
ATOM 7 SD MET B 1 9.310 -10.371 30.150 1.00 36.98 S
ATOM 8 CE MET B 1 8.130 -11.253 29.132 1.00 25.99 C
ATOM 9 N TYR B 2 12.316 -8.435 26.498 1.00 18.20 N
ATOM 10 CA TYR B 2 12.025 -7.438 25.474 1.00 14.27 C
ATOM 11 C TYR B 2 10.531 -7.166 25.377 1.00 16.02 C
ATOM 12 O TYR B 2 9.716 -7.951 25.856 1.00 18.04 O
ATOM 13 CB TYR B 2 12.546 -7.902 24.109 1.00 19.85 C
ATOM 14 CG TYR B 2 14.051 -8.021 24.021 1.00 20.63 C
ATOM 15 CD1 TYR B 2 14.711 -9.133 24.532 1.00 17.99 C
ATOM 16 CD2 TYR B 2 14.812 -7.026 23.420 1.00 17.23 C
ATOM 17 CE1 TYR B 2 16.086 -9.247 24.452 1.00 21.20 C
ATOM 18 CE2 TYR B 2 16.186 -7.130 23.336 1.00 18.83 C
ATOM 19 CZ TYR B 2 16.818 -8.241 23.852 1.00 19.05 C
ATOM 20 OH TYR B 2 18.185 -8.346 23.768 1.00 24.93 O
ATOM 21 N ASP B 3 10.178 -6.045 24.757 1.00 14.59 N
...
HETATM 6112 S S04 B 401 1.203 -4.226 28.103 0.79 48.58 S
HETATM 6113 O1 S04 B 401 1.874 -4.403 26.818 0.79 48.39 O
HETATM 6114 O2 S04 B 401 -0.226 -4.479 27.944 0.79 43.15 O
HETATM 6115 O3 S04 B 401 1.759 -5.160 29.078 0.79 43.28 O
HETATM 6116 O4 S04 B 401 1.410 -2.862 28.577 0.79 40.44 O
HETATM 6117 PA FDA B 402 10.470 -13.283 3.929 1.00 11.80 P
HETATM 6118 O1A FDA B 402 10.391 -14.724 3.358 1.00 16.94 O
HETATM 6119 O2A FDA B 402 11.217 -12.266 3.231 1.00 12.40 O
HETATM 6120 O5B FDA B 402 11.172 -13.101 5.223 1.00 15.19 O
...
CONNECT 6112 6113 6114 6115 6116
CONNECT 6113 6112
CONNECT 6114 6112
CONNECT 6115 6112
CONNECT 6116 6112
CONNECT 6117 6118 6119 6120 6169
CONNECT 6118 6117
CONNECT 6119 6117
CONNECT 6120 6117 6121
CONNECT 6121 6120 6122
CONNECT 6122 6121 6123 6124
CONNECT 6123 6122 6128
CONNECT 6124 6122 6125 6126
...
MASTER 316 0 8 35 32 0 30 6 6990 2 128 58
END

```