

**ANALIZA REZULTATOV MOLEKULSKEGA
SIDRANJA KNJIŽNICE SPOJIN NA
PROTEINSKE TARČE
VIRUSA SARS-CoV-2**

ANALIZA REZULTATOV MOLEKULSKEGA SIDRANJA KNJIŽNICE SPOJIN NA PROTEINSKE TARČE VIRUSA SARS-CoV-2

Področje: **KEMIJA**

Vrsta naloge: **Raziskovalna naloga**

Dijakinja: **Špela Žunec, 3. G**

Mentorica: **Mag. Mojca Podlipnik, GJPL**

Somentor: **Dr. Črtomir Podlipnik, FKKT, Ljubljana**

2021

Gimnazija Jožeta Plečnika Ljubljana

KAZALO

POVZETEK	4
1 UVOD	5
1.1 Namen	5
1.2 Hipoteze	5
1.3 Metode	5
2 TEORETIČNI DEL	7
2.1 Življenjski cikel virusa SARS-CoV-2	7
2.2 Molekularno sidranje	8
2.3 RxDock – program za sidranje molekul	10
2.4 SeeSAR – program za analizo kompleksov med ligandi in receptorjem	12
2.5 Proteinske tarče	14
2.6 Zbirke ligandov	15
2.7 Pravilo Lipinskega	17
3 EKSPERIMENTALNI DEL	18
3.1 Analiza ligandov, ki so rezultat sidranja z RxDock v vezavna mesta izbranih treh proteinskih tarč	18
3.2 Uporaba programa SeeSAR za določitev vezavne afinitete ligandov v vezavna mesta izbranih proteinskih tarč in za strukturno modifikacijo ligandov	19
4 REZULTATI	19
4.1 Analiza rezultatov sidranja z RxDock	19
4.1.1 Proteinska tarča 1: 3CL proteaza (3CLpro)	19
4.1.2 Proteinska tarča 2: PL proteaza (PLpro)	23
4.1.3 Proteinska tarča 3: RBD proteina Spike (krajše protein S)	25
4.1.3.1 Ligandi iz zbirke Chemotheca	25
4.1.3.2 Ligandi iz zbirke Peptidi	27
4.2 Obdelava najboljših ligandov glede na RxDock s programom SeeSAR	29
4.2.1 Proteinska tarča 1: 3CL proteaza (3CLpro)	29
4.2.2 Proteinska tarča 2: PL proteaza (PLpro)	36
4.2.3 Proteinska tarča 3: RBD proteina Spike (krajše protein S)	41
4.2.3.1 Ligandi iz zbirke Chemotheca	41
4.2.3.2 Ligandi iz zbirke Peptidi	48
5 RAZPRAVA	49
5.1 Analiza rezultatov sidranja z RxDock	49
5.2 Obdelava najboljših ligandov glede na RxDock s programom SeeSAR	50
6 ZAKLJUČEK	51

7	ZAHVALE	52
8	VIRI IN LITERATURA	52

POVZETEK

V sklopu svoje raziskovalne naloge sem analizirala rezultate molekulskega sidranja spojin iz podatkovne zbirke ChEMBL na tri proteinske tarče, ki so pomembne v življenjskem ciklusu virusa SARS-CoV-2. Prva in druga tarča sta bili proteazi 3CLpro (glavna proteaza) in PLpro (Papainu podobna proteaza), ki cepita polipeptid, ki nastane pri prepisu virusne RNA. Tretja tarča je bila receptorska domena proteina Spike, ki je pomembna za pripetje virusa na ACE2 receptor gostujoče celice. Program RxDock, ki je bil uporabljen za sidranje, iz množice spojin izloči spojine, ki se vežejo v vezavno mesto določenega receptorja (izbrane proteinske tarče), pri čemer je pomembna predvsem oblika molekule liganda. Najboljše ligande glede na program RxDock sem analizirala s programom SeeSAR in ugotovila, da ta program vezavno afiniteto med ligandom in receptorjem izračuna na drugačen način, torej so najboljši ligandi glede na SeeSAR druge spojine. Program SeeSAR namreč upošteva tudi interakcije med ligandom in receptorjem; upošteva npr. nastanek vodikovih vezi in desolvatacijo na stični površini med ligandom in receptorjem.

KLJUČNE BESEDE:

sidranje, RxDock, SeeSAR, SARS-CoV-2

1 UVOD

1.1 Namen

Namen raziskovalne naloge je iz velike množice sidranih spojin poiskati molekule, ki so najboljše ligandi in te ligande s spreminjanjem njihove strukture ob uporabi programa SeeSAR še izboljšati.

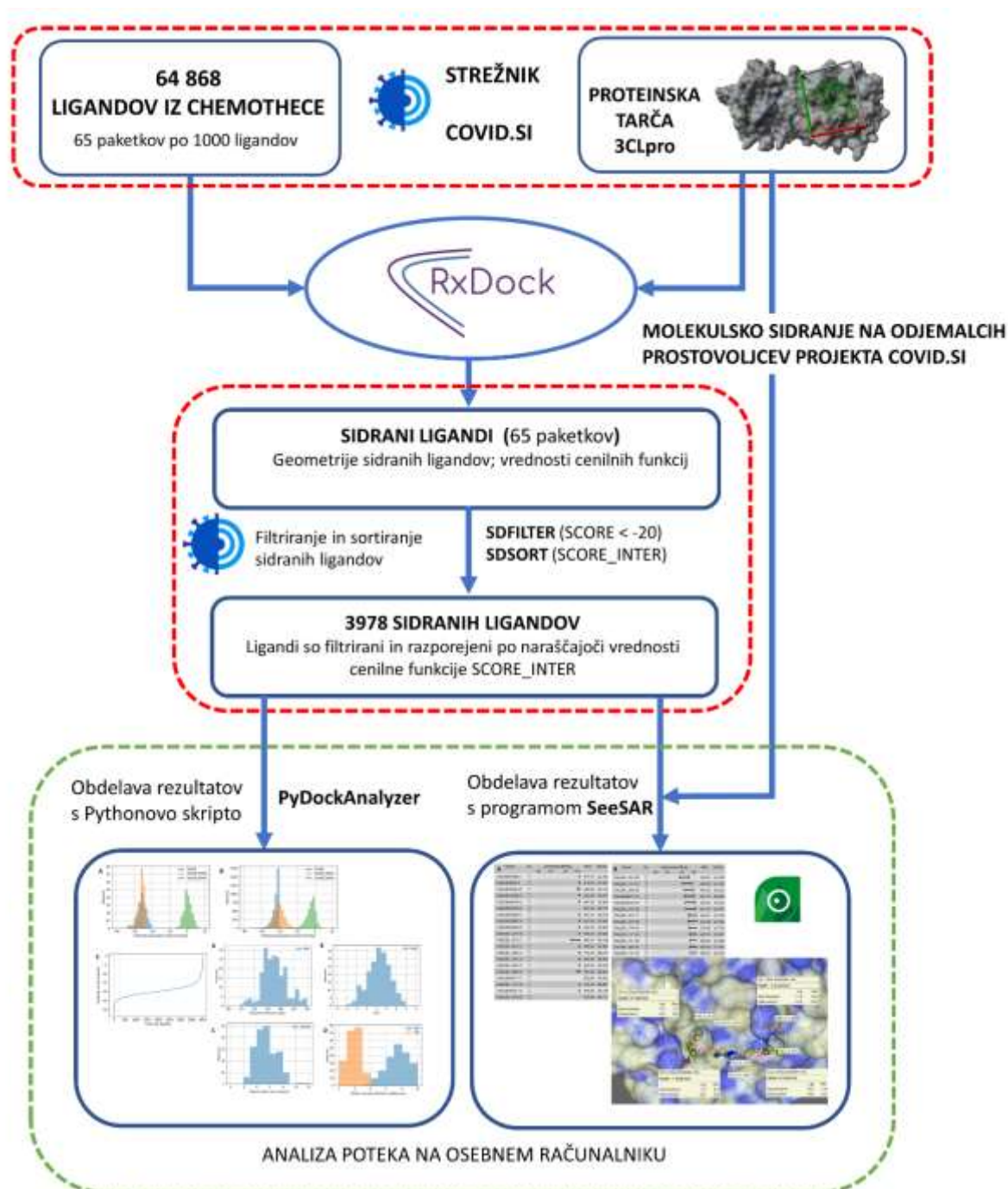
1.2 Hipoteze

V okviru svoje raziskovalne naloge sem postavila tri hipoteze:

- Molekularno sidranje RxDock lahko uporabimo za razvrščanje ligandov po vezavni afiniteti.
- Molekularno sidranje je dobro orodje za analizo interakcij med ligandom in receptorjem.
- Vezavno afiniteto ligandov z receptorji lahko z manjšimi spremembami strukture ligandov izboljšamo.

1.3 Metode

V raziskovalni nalogi sem analizirala ligande, ki so se glede na izračune opravljene s programom RxDock najuspešnejše sidrali v vezavna mesta izbranih treh proteinskih tarč. Postopek sidranja ligandov iz zbirke Chemoteca na proteinsko tarčo 3CLpro in nadaljnja analiza sidranih ligandov sta shematsko prikazana na sliki 1. Uspešno sidranih 3978 ligandov sem analizirala s pomočjo skripte PyDockAnalyzer, komplekse teh ligandov in receptorja pa s programom SeeSAR.



Slika 1: Shema postopka sidranja ligandov iz zbirke ChEMBL na proteinsko tarčo 3CLpro in njihove analize s PyDockAnalyzer-jem in SeeSAR-om.

Podobno sem s programsko skripto PyDockAnalyzer, ki jo je v Pythonu napisal zunanji mentor, analizirala tudi uspešno sidrane ligande na ostalih dveh proteinskih tarčah PLpro in RBD proteina Spike. Ligande sem razvrščala glede na vrednosti cenilnih funkcij izračunanih s programom RxDock, da bi ugotovila, kateri ligandi so najboljši ter preverila njihovo ustreznost za potencialno uporabo v obliki oralnih zdravil glede na pravilo Lipinskega.

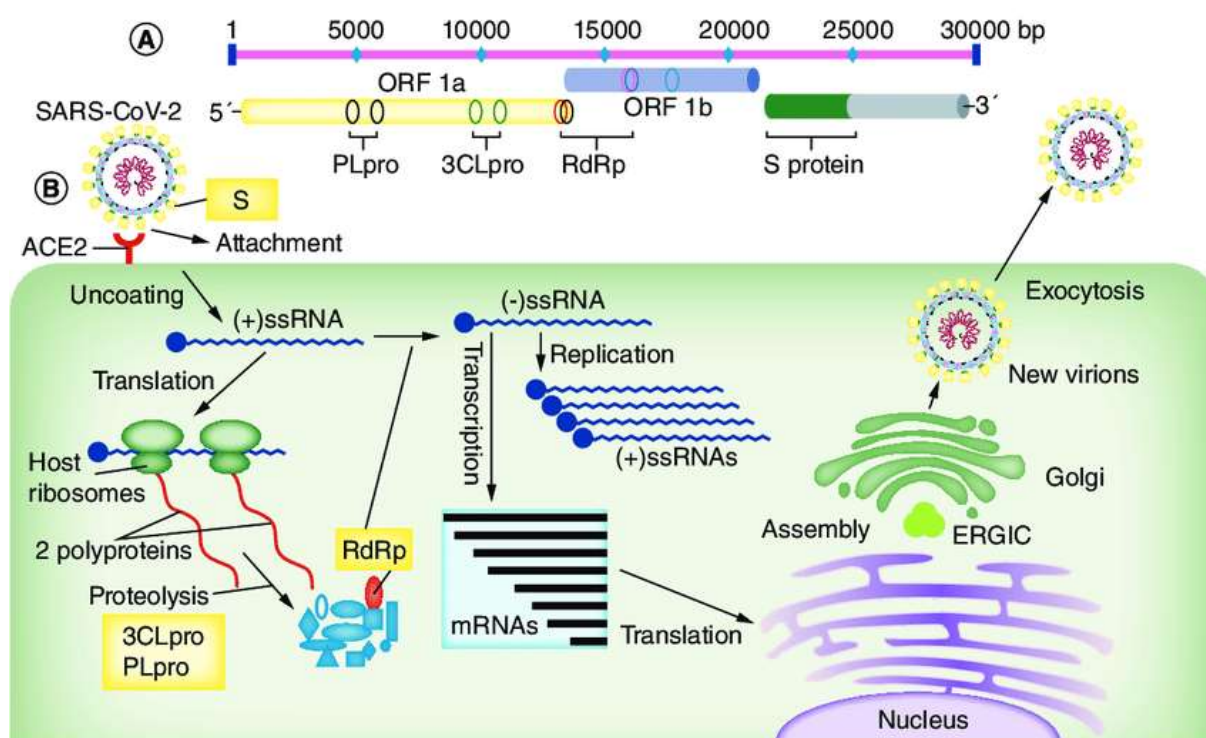
Program SeeSAR sem uporabila za izračun vezavne afinitete med ligandom in proteinsko tarčo za tistih 200 ligandov, ki so se na podlagi računov s programom RxDock, izkazali za najboljše sidrane ligande. Za vsako tarčo sem poskusila najboljša liganda iz zbirke Chemotheca s programom SeeSAR še modificirati na tak način, da bi se vezavna afiniteta izboljšala.

2 TEORETIČNI DEL

2.1 Življenjski cikel virusa SARS-CoV-2

Virus SARS-Cov-2 (iz angl. Severe acute respiratory syndrome coronavirus 2) je virus iz družine koronavirusov; to so virusi, ki povzročajo infekcije zgornjih dihal. SARS-Cov-2 povzroča bolezensko stanje covid-19. Je virus, ki se je na človeka prenesel z živali.

Na sliki 2 je prikazan genom virusa SARS-CoV-2 in nekaj stopenj življenjskega cikla tega virusa.



Slika 2: (A) Genom in (B) replikacijski cikel virusa SARS-CoV-2.

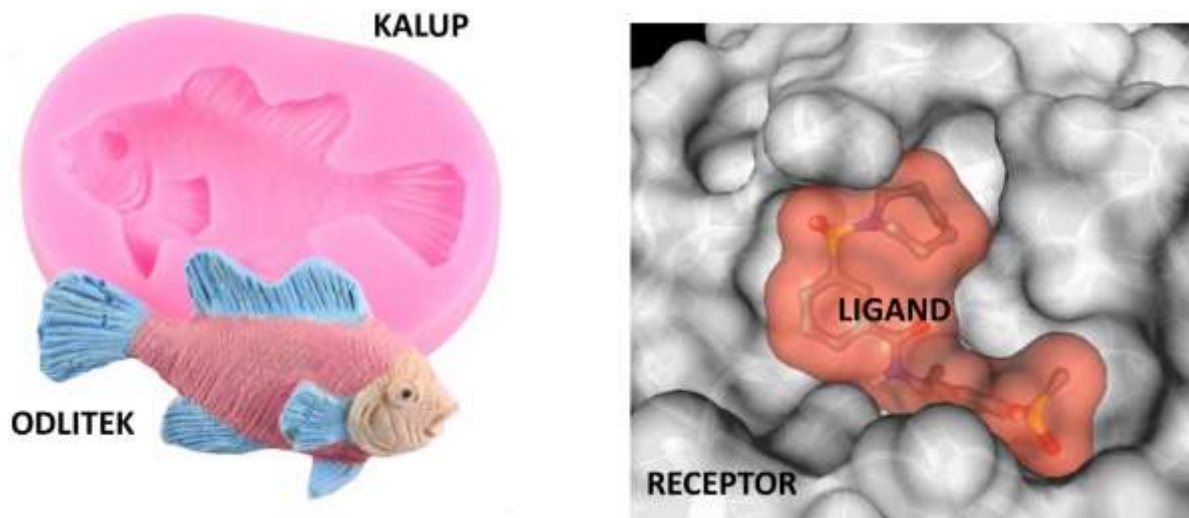
Oznake v sliki 2: 3CLpro je glavna virusna proteaza (angl. 3-Chymotrypsin-like protease); ACE2 je receptorsko mesto na celici gostitelja (ang. Angiotensin-converting enzyme 2); ERGIC je Golgijev aparat endoplazmatskega retikuluma (angl. Endoplasmic Reticulum-Golgi intermediate compartment); NC je nukleokapsid (angl. Nucleocapsid); ORF je odprt bralni okvir, tj. del genskega zapisa v molekuli RNK (angl. Open-reading frame); PLpro je virusna proteaza (angl. Papain-like protease); RdRp je virusna proteaza (angl. RNA-dependent RNA polymerase); S je protein Spike virusa SARS-CoV-2.

Iz slike 2 (A) je razvidno, da je v genomu (virusnem RNK) zapis za strukturne proteine, dva izmed njih sta 3CLpro in PLpro. Slika 2 (B) pa prikazuje življenjski cikel virusa SARS-CoV-

2. Virusni protein Spike se po principu ključ-ključavnica veže na celični receptor ACE-2 gostiteljske celice in poskrbi za spajanje virusne ovojnice s celično membrano gostitelja. Virus nato v celico vstopi z endocitozo. V celici se virusna RNK sprosti v citoplazmo gostiteljeve celice. S pomočjo gostiteljskih ribosomov nastaneta dva poliproteina, ki se kasneje cepita v več manjših proteinov, pri čemer sodelujejo celične proteaze in dve virusni proteazi 3CLpro in PLpro. Virusni genom se v gostiteljevi celici podvoja, nastanejo pa tudi virusne mRNK, ki so pomembne za nastanek virusnih proteinov. Ko nastanejo vsi sestavni deli virusa, se združijo v nove viruse, ki z eksocitozo zapustijo gostiteljsko celico.

2.2 Molekulsko sidranje

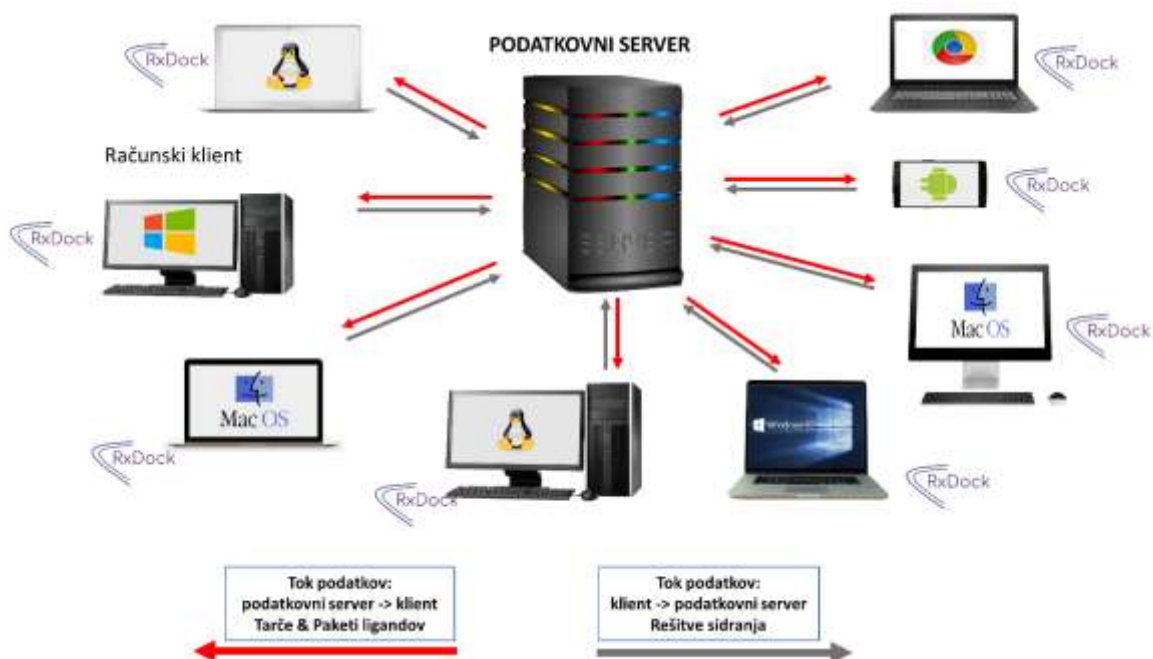
Molekulsko sidranje je metoda, ki se pogosto uporablja pri načrtovanju spojin, ki bi bile primerne za razvoj zdravil. Pri molekularnem sidranju iščemo manjše organske molekule, ki so čim bolj komplementarne biološki tarči. Običajno je biološka tarča protein, lahko pa je tudi DNA oz. RNA. Princip komplementarnosti na primeru kalupa in odlitka je prikazan na sliki 3, ligand (v tem primeru riba) naj bi se čim bolj tesno prilegal receptorju (v našem primeru kalup).



Slika 3: Prikaz principa komplementarnosti.

Poenostavljeno rečeno je molekularno sidranje optimizacijski postopek, s katerim iščemo takšno orientacijo liganda v receptorskem mestu, pri kateri je vzpostavljeno maksimalno število ugodnih in hkrati čim manjše število neugodnih kontaktov. Seštevek vseh ugodnih in neugodnih kontaktov imenujemo cenilna funkcija. Položaj liganda znotraj receptorskega mesta pa imenujemo poza. Pri postopku optimizacije spreminjamo konformacijo liganda (z vrtenjem okoli enojnih vezi) in njegovo relativno lego glede na receptor.

Rezultate molekulskega sidranja sem pridobila iz podatkovne zbirke na strežniku projekta COVID.SI (glej <https://covid.si>), kjer so zbrani rezultati sidranja velike množice ligandov (okoli 10 milijonov) na različne tarče povezane z boleznijo covid-19 oziroma z virusom SARS-CoV-2. Zbirka je del obsežnega projekta COVID.SI, pri katerem sodeluje veliko število prostovoljcev, ki so prispevali proste razpoložljive računske kapacitete na svojih računalnikih. Prostovoljci so tako s svojimi računalniki pripomogli pri iskanju potencialno zanimivih učinkovin zoper delovanje virusa SARS-CoV-2. Sodelavci projekta COVID.SI so vzpostavili svoj unikatni sistem porazdelitvenega računalništva, kjer podatkovni strežnik dodeljuje posamezne naloge računskim klientom (računalniki prostovoljcev), strežnik pa hkrati hrani vhodne podatke (zbirka ligandov, zbirka proteinskih tarč) in izhodne podatke (rešitve sidranja). Za molekulske sidranje so sodelavci projekta uporabili odprtokodni program RxDock. Na sliki 4 je prikazan sistem porazdelitvenega računalništva uporabljen za molekulske sidranje.



Slika 4: Shema izvedbe molekulskega sidranja z metodo porazdelitvenega računalništva.

Na podatkovnem serverju so pripravljene paketi z eno proteinsko tarčo in 1000 ligandi. Podatki o strukturah proteinskih tarč so v datotekah .mol2, podatki o strukturah prostih ligandov pa v datotekah .sdf. Osebni računalnik prostovoljca se poveže s serverjem in avtomatsko se začne izvajanje sidranja enega paketa z RxDock, po končanem sidranju pa se na podatkovni server vrnejo informacije o sidranih ligandih v obliki nove datoteke .sdf. Primer zapisa v eni od teh datotek prikazuje slika 5.

```
GLAVA DATOTEKE
CMLDID40686 1
OpenBabel02082117593D
librxdock.so/0.1.0
30 33 0 0 0 0 0 0 0 0999 V2000
-4.1963 3.8318 17.6890 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-5.1557 4.3471 16.6471 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-5.8125 5.3629 16.8457 O 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-5.1880 3.5403 15.5278 N 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
....
....
1 2 1 0 0 0 0
2 3 2 0 0 0 0
2 4 1 0 0 0 0
...
...
M END
KONEC BLOKOV STRUKTURNIH PODATKOV
....
> <SCORE>
-28.5572
> <SCORE.INTER>
-30.7971
BLOK Z LASTNOSTMI
> <SCORE.INTRA>
2.23986
....
$$$$
KONEC PODATKOV ZA PRVO MOLEKULO
```

Slika 5: Struktura datoteke SDF, v kateri so napisani podatki o ligandih (njihovi strukturi in izračunanih lastnostih).

V svojem raziskovalnem delu molekul ligandov nisem sidrala, sem pa obdelala nekaj rezultatov sidranja, ki so nastali v okviru projekta COVID.SI.

2.3 RxDock – program za sidranje molekul

RxDock je hiter, vsestranski in odprtokodni program za sidranje ligandov (majhne molekule) na beljakovine in nukleinske kisline. Njegova zasnova omogoča hitro virtualno reševanje (HTVS – angl. High Throughput Virtual Screening) ogromnih knjižnic spojin na biološko zanimive proteinske tarče. S postopkom HTVS lahko iz knjižnice nekaj milijonov spojin izločimo nekaj sto ligandov, ki imajo potencial za nadaljnji razvoj. RxDock lahko uporabimo tudi kot orodje za študij interakcij med ligandom in receptorjem. Celoten programski paket RxDock zahteva manj kot 50 MB prostora na trdem disku in deluje na računalnikih z operacijskimi sistemi Windows, Linux in MacOS.

Zahvaljujoč zasnovi in izvedbi samega programa ga je mogoče namestiti v računsko gručo in velik problem razdeliti na veliko število manjših nalog, ki jih procesira posamezen procesor. Odprtost in možnost enostavne paralelizacije računalniških procesov je bil poglobitni razlog, da so sodelavci projekta COVID.SI izbrali RxDock za sidranje njihove knjižnice spojin na proteinske tarče virusa SARS-CoV-2.

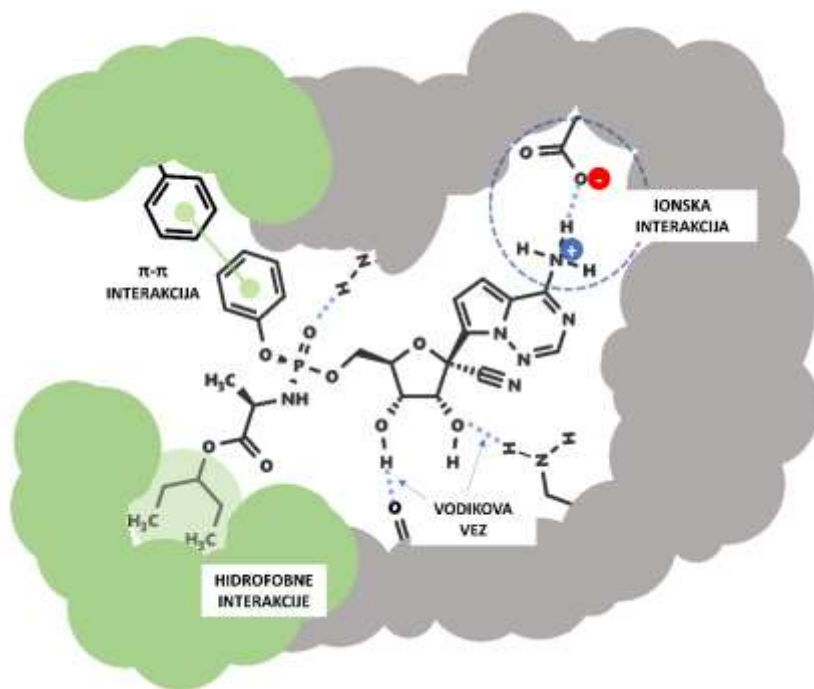
Programski paket RxDock poleg glavnega programa za sidranje vsebuje tudi nabor orodij in skript za lažjo pripravo vhodnih datotek ter naknadno obdelavo in analizo rezultatov.

Pomemben del vsakega programa za molekulske sidranje je cenilna funkcija. V primeru programa RxDock ima cenilna funkcija S^{total} takole obliko:

$$S^{total} = S^{inter} + S^{intra} + S^{site}$$

Celotna cenilna funkcija Score (v enačbi S^{total}) je sestavljena iz treh cenilnih funkcij in sicer: Score_INTER (v enačbi S^{inter}), ki predstavlja seštevek vseh interakcij med ligandom in receptorjem; Score_INTRA (v enačbi S^{intra}), ki predstavlja relativno energijo konformacije liganda glede na tisto, ki jo ima v nevezanem stanju in Score_SITE (v enačbi S^{site}), ki predstavlja relativno energijo fleksibilnih regij proteinskega receptorskega mesta glede na referenčno "nevezano" stanje. Pri sidranju ligandov na receptorska mesta proteinskih tarč virusa SARS-CoV-2 je bila upoštevana vrednost Score_SITE enaka nič, saj so predpostavili, da se struktura receptorskih mest pri sidranju ligandov ne spremeni.

Interakcije med ligandom in receptorjem so različne; lahko privlačne kot npr. so vodikove vezi, polarne interakcije (ion-ion) in disperzijske sile med nepolarnimi deli ali pa odbojne (npr. med enako nabitimi deli). Nekaj možnih interakcij med ligandom in receptorjem prikazuje slika 6.



Slika 6: Interakcije med receptorjem in ligandom.

Pri sidranju ligandov v vezavna mesta proteinskih tarč z RxDock so torej predpostavili, da se struktura proteinskih tarč ne spremeni, spreminja pa se konformacija liganda (vrtenje okoli enojnih vezi, spreminjanje dihedralnih (torzijskih) kotov) in položaj liganda v vezavnem mestu. Program RxDock išče tako pozico liganda, da bo med ligandom in receptorjem čim več ugodnih in čim manj neugodnih interakcij. Čim bolj negativni vrednosti celotne cenilne funkcije Score in cenilne funkcije Score_INTER pomenita, da se je ligand dobro vezal (sidral) v vezavno mesto receptorja.

2.4 SeeSAR – program za analizo kompleksov med ligandi in receptorjem

SeeSAR je računalniški program, ki raziskovalcem pomaga pri načrtovanju potencialno farmacevtsko zanimivih spojin. Program zajema številne funkcije, s katerimi lahko analiziramo vezavno afiniteto liganda na določen receptor (kako močno se ligand veže na receptor, tj. cenilna funkcija HYDE), lahko ga uporabimo za oceno lastnosti ADMET (angl. Absorption-Distribution-Metabolism-Excretion-Toxicity, slov. Adsorpcija-Distribucija-Metabolizem-Izločanje-Toksičnost), ligande pa lahko tudi s tem programom sidramo v receptorsko mesto.

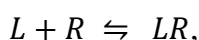
Sposobnost vezave molekule spojine na določeno proteinsko tarčo je le prvi izmed pogojev, da bi spojino potencialno lahko uporabili kot zdravilo. Idealna zdravilna učinkovina naj bi bila farmakološko aktivna, nestrupena in dovolj učinkovita, da bi jo oralno jemali enkrat na dan, po

koncu delovanja pa naj bi se varno izločila iz telesa. Poznavanje ADMET lastnosti nam torej pomaga pri razvoju in optimizaciji zdravilnih učinkovin.

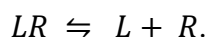
SeeSAR vsebuje tudi dve orodji, s katerimi lahko ligand simultano urejamo znotraj vezavnega mesta. Z enim urejevalnikom, tj. Editor, lahko molekuli liganda dodajamo, odvezujemo in spreminjamo atome, prekinjamo in tvorimo kovalentne vezi, itd. Z drugim urejevalnikom, tj. Inspirator, pa lahko spreminjamo sredico molekule (angl. core replacement) ali stranske verige (angl. fragment replacement) s fragmenti zbranimi v interni podatkovni zbirki.

Pomemben del programa SeeSAR je cenilna funkcija HYDE, s katero lahko kvantitativno ovrednotimo vezavno afiniteto liganda. Funkcija HYDE je sestavljena iz dveh glavnih prispevkov. Prvi prispevek program izračuna na osnovi vodikovih vezi (angl. **hydrogen bonds**), ki se tvorijo med molekulo liganda in vezavnim mestom molekule receptorja; drugi prispevek, ti. desolvacijski prispevek (angl. **desolvation**), pa je posledica sprostitve molekul vod, ki se nahajajo na stični površini liganda in proteina ob tvorbi kompleksa.

Program omogoča vizualizacijo prispevkov posameznih atomov v molekuli liganda k ocenjeni afiniteti – zeleno obarvane sfere pomenijo ugodne prispevke, rdeče obarvane sfere pa neugodne prispevke k celotni afiniteti. Velikost sfere je povezana z velikostjo prispevka, določenega atoma k skupni afiniteti. Ocena afinitete, ki jo izračuna cenilna funkcija HYDE, ima energijske enote (kJ/mol), negativna vrednost afinitete pomeni, da je interakcija med ligandom in receptorjem ugodna, nižja kot je vrednost HYDE, tj. bolj negativna, boljša je njuna vezava. Vrednosti HYDE običajno preračunajo v vrednosti disociacijske konstante, ki se običajno uporablja za vrednotenje afinitete med ligandom in receptorjem. Tvorbo kompleksa med ligandom in receptorjem lahko opišemo z enačbo ravnotežne reakcije:



disociacijo pa z izrazom:



Konstanto disociacije zapišemo:

$$K_D = \frac{[L][R]}{[LR]},$$

kjer so $[L]$, $[R]$ in $[LR]$ ravnotežne množinske koncentracije liganda, receptorja in kompleksa. Vrednost konstante podajamo z enoto mol/L. Nižja kot je vrednost konstante, močnejše se ligand

veže v receptor. Tako se ligand, katerega afiniteta je v območju pikomolarnih koncentracij (pM), predvidoma močnejše veže v receptorsko mesto kot tisti z afiniteto v območju nanomolarnih koncentracij (nM). Program SeeSAR namesto ene same vrednosti ocenjene afinitete podaja interval (Estimated affinity), znotraj katerega lahko pričakujemo, da najdemo vrednost afinitete (glej sliko 7).

Program SeeSAR in cenilna funkcija HYDE se kontinuirno izboljšujeta v okviru sodelovanja podjetja BioSOLVEIT z Univerzo v Hamburgu in podjetjem Bayer.



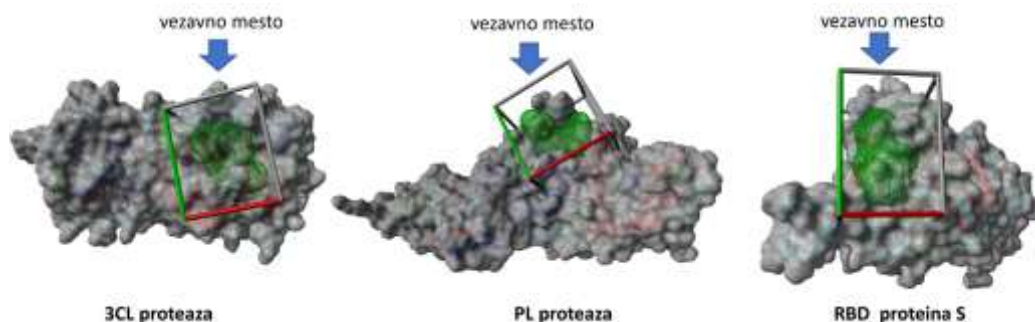
Slika 7: Prikaz na računalniškem zaslonu med uporabo programa SeeSAR.

2.5 Proteinske tarče

V okviru projekta COVID.SI so sodelavci pripravili nekaj proteinskih tarč, ki so direktno povezane z delovanjem virusa SARS-CoV-2. Večino proteinskih tarč, ki so jih uporabili za sidranje, lahko najdemo v podatkovni zbirki proteinskih struktur PDB (angl. Protein Data Bank). V zbirki PDB za določitev 3D strukture proteinov uporabljajo pretežno kristalografske metode, nekaj proteinskih struktur pa so določili z metodo NMR, v zadnjem času pa tudi s krioelektronsko mikroskopijo. Tudi proteinske tarče je bilo potrebno ustrezno pripraviti za molekularno sidranje. Ta postopek priprave tarč je precej zapleten in ga na tem mestu ne bom opisala.

Vseh proteinskih tarč, ki so jih uporabili oz. jih še uporabljajo za molekularno sidranje, je okoli šestdeset. Sama sem se lotila analize rezultatov sidranja na treh tarčah: 3CL proteaza (3CLpro),

PL proteaza (PLpro) in RBD virusnega proteina Spike (RBD – angl. receptor-binding domain). Protein Spike ali krajše protein S virusa SARS-CoV-2 se z delom RBD veže na receptor ACE2. Vse te proteinske tarče igrajo pomembno vlogo pri delovanju virusa, kar je opisano v poglavju 2.1. Na sliki 8 so predstavljene strukture proteinskih tarč z označenim vezavnim mestom sidranja.



Slika 8: Strukture proteinskih tarč (3CLpro, PLpro, RBD proteina Spike) z označenim vezavnim mestom, ki sem jih obravnavala pri svojem delu.

2.6 Zbirke ligandov

Ligandi so običajno zbrani v večjih zbirkah, zapisani so v obliki linearnega zapisa SMILES ali pa v 2D obliki v formatu sdf. Za izvedbo molekulskega sidranja potrebujemo molekule v 3D obliki, v ustrezni – smiselni konformaciji, hkrati pa morajo biti molekule v protonacijskem stanju, ki ustreza pH-ju pri fizioloških pogojih, običajno je ta pH 7,4. To običajno lahko pomeni, da so karboksilne skupine deprotonirane, aminske skupine pa protonirane. Ustrezno protonacijsko stanje ligandov izračunamo s prosto dostopnim programom Open Babel. V zbirko COVID.SI je vključenih več kot 10 milijonov spojin, v obliki dokumentov formata sdf, v katerih so informacije o 3D strukturi ligandov, ki so pripravljene za takojšnje sidranje. Večina spojin, potencialnih ligandov, v zbirki je pridobljena iz podatkovne zbirke komercialno dostopnih spojin ZINC 15.

V raziskovalni nalogi smo obravnavali dve skupini ligandov:

- molekule spojin iz zbirke Chemotheca (za vse tri proteinske tarče) in
- molekule dodekapeptidov (le za proteinsko tarčo RBD proteina Spike).

Podatkovna zbirka spojin Chemotheca je nastala kot rezultat sodelovanja več kot 25 evropskih držav. V tej zbirki je 64868 spojin, ki so označene z identifikacijskimi številkami ID (npr. spojina z oznako CMLDID35344 je bila kot 35344. dodana v to zbirko). V zbirki Chemotheca

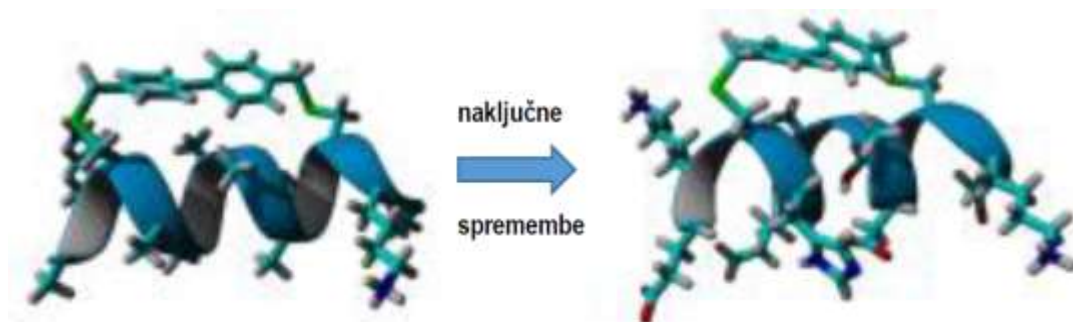
(slika 9) so zbrane spojine, ki so nastale v različnih Evropskih raziskovalnih laboratorijih in so rezultat različnih farmacevtskih raziskav.



Slika 9: Vhodna spletna stran podatkovne zbirke Virtual Chemotheca

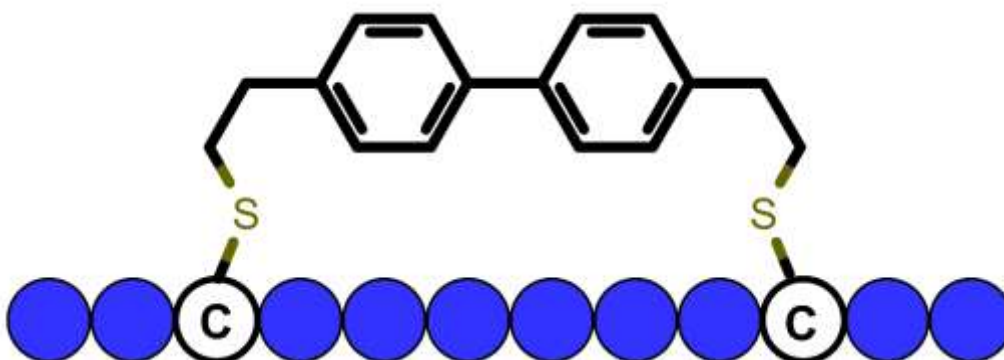
Za vsako spojino, ki se izkaže kot farmacevtsko zanimiva, kemiki sintetizirajo/izolirajo in testirajo množico spojin, za katere se lahko na koncu izkaže, da so neaktivne oz. manj aktivne. Rezultate raziskovalci običajno objavijo v znanstvenih publikacijah, potem pa se pogosto na te spojine pozabi. Ideja avtorjev zbirke Chemotheca je bila ta, da so na enem mestu zbrane informacije o neaktivnih in malo aktivnih spojin ter omogočiti drugim raziskovalcem, da spojine uporabijo pri svojih raziskavah na drugih tarčah. Zbirka Virtual Chemotheca poleg samih strukturnih informacij o spojinah nudi tudi kontaktne podatke, ki omogočijo, da se raziskovalci med seboj povežejo in izmenjujejo spojine.

Protein Spike je ključen za pripetje virusa na receptor ACE2, ki se nahaja na površini gostiteljske celice. Receptorsko mesto proteina Spike je zelo obsežno. V zbirki Chemotheca pa so predvsem manjše molekule. Raziskovalci projekta COVID.SI so zato, da bi v večji meri pokrili vezavno mesto in s tem preprečili vezavo virusnega proteina Spike na receptor ACE2 gostiteljske celice, oblikovali večje molekule; in sicer dvanajstmerne, konformacijsko zaklenjene peptide. V okviru projekta so pripravili zbirko 1000 različnih naključno generiranih dodekapeptidov, premoščenih z aromatskim sistemom (glej sliko 10).



Slika 10: Generacija konformacijsko zaklenjenih naključnih dodekapeptidov.

Peptidi so konformacijsko zaklenjeni s pomočjo bifenilnega fragmenta, ki povezuje tretji in deseti cisteinski aminokislinski ostanek. Oba omenjena cisteina zaradi svoje vloge nista podvržena mutaciji (zamenjavi), ki pa je možna na preostalih desetih mestih. Vseh možnih peptidov, ki se lahko generirajo na ta način, je 20^{10} . Na sliki 11 je prikazana idejna zasnova za načrtovanje dodekapeptidov. Mesta potencialnih mutacij (zamenjav) so označena z modro barvo.



Slika 11: Idejna zasnova za načrtovanje dodekapeptidov. Položaja dveh vezanih molekul aminokisljne cistein sta označena s C.

2.7 Pravilo Lipinskega

Pri izdelavi zdravil, ki bi jih lahko zaužili oralno, je potrebno upoštevati nekaj kriterijev. Pravilo Lipinskega pravi, da na splošno oralno aktivno zdravilo lahko krši le eno od naslednjih meril:

- največ 5 donorjev vodikovih vezi (skupno število vezi dušik-vodik in kisik-vodik)
- ne več kot 10 akceptorjev vodikovih vezi (vsi atomi dušika ali kisika)
- relativna molekulska masa manjša od 500
- porazdelitveni koeficient oktan-1-ol-voda ($\log P$), ki ne presega 5

Pri molekulah, ki se računsko izkažejo kot dobri ligandi, je torej smiselno preveriti, ali ustrezajo pravilu Lipinskega in so možni kandidati za zdravilne učinkovine.

3 EKSPERIMENTALNI DEL

3.1 Analiza ligandov, ki so rezultat sidranja z RxDock v vezavna mesta izbranih treh proteinskih tarč

S pomočjo Python-ove programske skripte PyDockAnalyzer, ki jo je napisal zunanji mentor, sem za ligande naredila več grafičnih analiz.

Kriterij uspešnega sidranja ligandov iz zbirke Chemotheca je bila vrednost celotne cenilne funkcije, izračunana s programom RxDock, manjša od -20 za proteinski tarči 3CLpro in PLpro. Pri proteinski tarči RBD proteina Spike pa smo za kriterij izbrali vrednost celotne cenilne funkcije manjšo od -10 . Kriterij smo pri tretji proteinski tarči spremenili, ker je bilo število ligandov, ki so ustrezali kriteriju celotne cenilne funkcije manjše od -20 , precej majhno (le 97 ligandov). S pomočjo programske skripte sem izrisala več histogramov. Histogrami, označeni s črko A, prikazujejo porazdelitev 200 najvišje razvrščenih ligandov glede na vrednosti Score_INTER po vrednostih posameznih cenilnih funkcij, histogrami, označeni s črko B, pa prikazujejo porazdelitev vseh uspešno sidranih ligandov po vrednostih posameznih cenilnih funkcij.

Izrisala sem tudi več grafov. Grafi, označeni so s črko C, ponazarjajo odvisnost vrednosti cenilne funkcije Score_INTER od pozicije, ki jo ima ligand v datoteki, v kateri so ligandi razvrščeni po naraščajoči vrednosti Score_INTER (od najboljšega do najslabšega liganda). Grafi, označeni so s črko D, prikazujejo vrednosti celotne cenilne funkcije Score v odvisnosti od položaja liganda v omenjeni datoteki.

Izbrane ligande sem analizirala tudi glede ustreznosti njihove uporabe kot zdravilne učinkovine. Zato sem izrisala tudi histograme, iz katerih so razvidne lastnosti relativna molekulska masa, $\log P$, število donorjev in akceptorjev vodikovih vezi ter število vrtljivih vezi.

Ker število možnih konformacij, ki jih moramo tekom sidranja preveriti, strmo narašča s številom vrtljivih vezi, preveliko število le-teh ni ugodno za samo sidranje. Preveč toge molekule, tiste brez oz. zelo majhnim številom vrtljivih vezi pa se ne morejo ustrezno prilagoditi v vezavno mesto.

Podobne analize po vrednostih cenilnih funkcij sem naredila tudi za zbirko 1000 dodekapeptidov, ki so bili možni ligandi na proteinsko tarčo RBD proteina Spike. Tudi v tem primeru sem za kriterij izbrala vrednost celotne cenilne funkcije manjšo od -10 .

Opisani histogrami in grafi so v poglavju Rezultati.

3.2 Uporaba programa SeeSAR za določitev vezavne afinitete ligandov v vezavna mesta izbranih proteinskih tarč in za strukturno modifikacijo ligandov

Za dodatno analizo rezultatov sidranja sem uporabila program SeeSAR podjetja BioSolveIT. Najboljših 200 ligandov sidranih s programom RxDock v vezavno mesto proteinskih tarč sem uporabila za izračun vezavne afinitete (Estimated affinity) s programom SeeSAR.

V program SeeSAR sem najprej uvozila strukturo proteina zapisano v datoteki PDB in nato še dvesto ligandov, ki jih je program RxDock razvrstil na najvišja mesta glede na vrednost cenilne funkcije. Nato sem morala definirati vezavno mesto proteina: vezavno mesto sem definirala glede na poze vseh v program uvoženih ligandov. Nato sem za vse ligande izračunala oceno vezavne afinitete (Estimated affinity). Nato sem primerjala razvrstitev najboljših 20 ligandov, izračunanih z RxDock in najboljših 20 ligandov, izračunanih s SeeSAR. Tem ligandom sem poskusila tudi malo spremeniti strukturo z namenom izboljšanja vezavne afinitete.

4 REZULTATI

4.1 Analiza rezultatov sidranja z RxDock

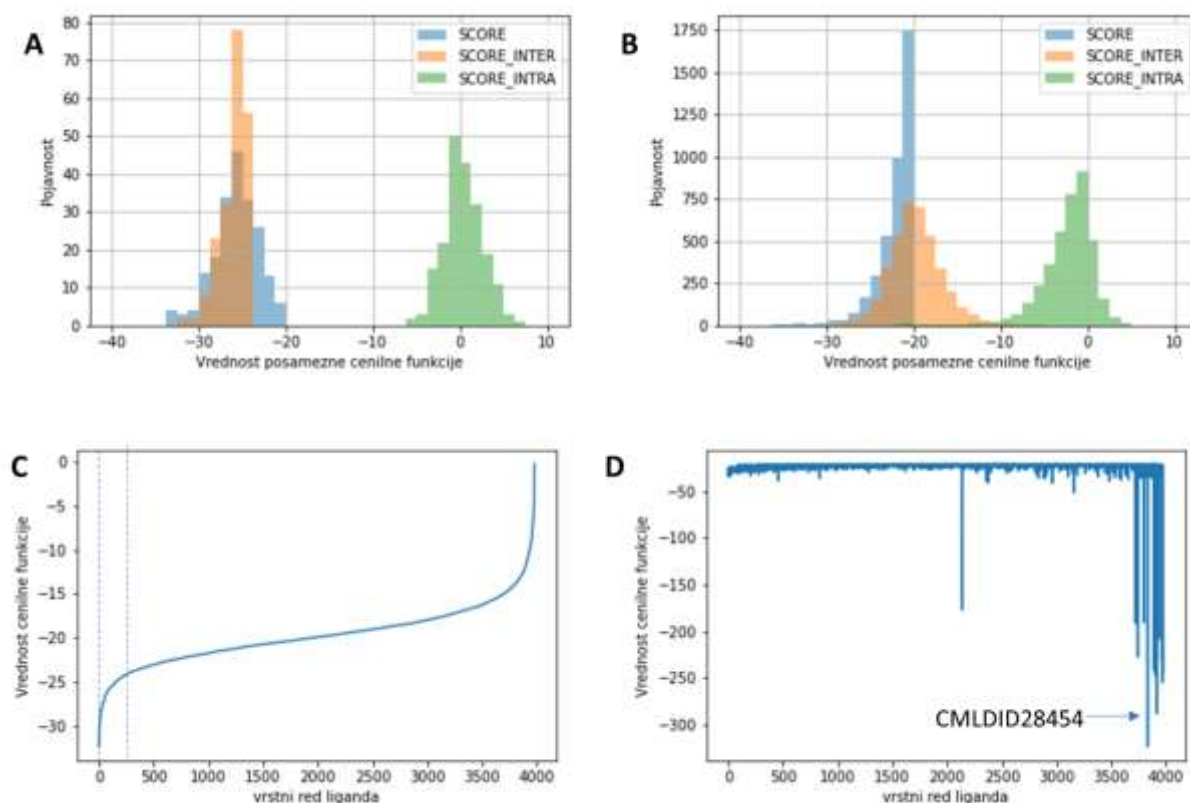
4.1.1 Proteinska tarča 1: 3CL proteaza (3CLpro)

Najprej sem se lotila analize ligandov, ki so rezultati sidranja na tarčo 3CLpro. Zunanji mentor je pripravil datoteko, v kateri so bili zbrani podatki za vse ligande iz zbirke Chemotheca, katerih vrednost celotne cenilne funkcije (Score) je bila nižja od -20 . Takšnih ligandov je bilo 3978.

Najprej sem teh 3978 ligandov razvrstila po naraščajoči vrednosti cenilne funkcije Score_INTER ter s programsko skripto PyDockAnalyzer izrisala histograme in dva grafa.

Histogrami na sliki 12 (A) predstavljajo porazdelitev najboljših 200 ligandov glede na vrednost cenilne funkcije Score_INTER po vrednostih posameznih cenilnih funkcij, histogrami na sliki 12 (B) pa prikazujejo porazdelitev vseh 3978 ligandov po vrednostih posameznih cenilnih funkcij.

Nato sem izdelala še dva grafa. Prvi graf (slika 12(C)) ponazarja odvisnost vrednosti cenilne funkcije Score_INTER od pozicije, ki jo ima ligand v datoteki, v kateri so ligandi razvrščeni po naraščajoči vrednosti Score_INTER. Drugi graf (slika 12 (D)) prikazuje odvisnost vrednosti cenilne funkcije Score od položaja liganda v omenjeni datoteki.

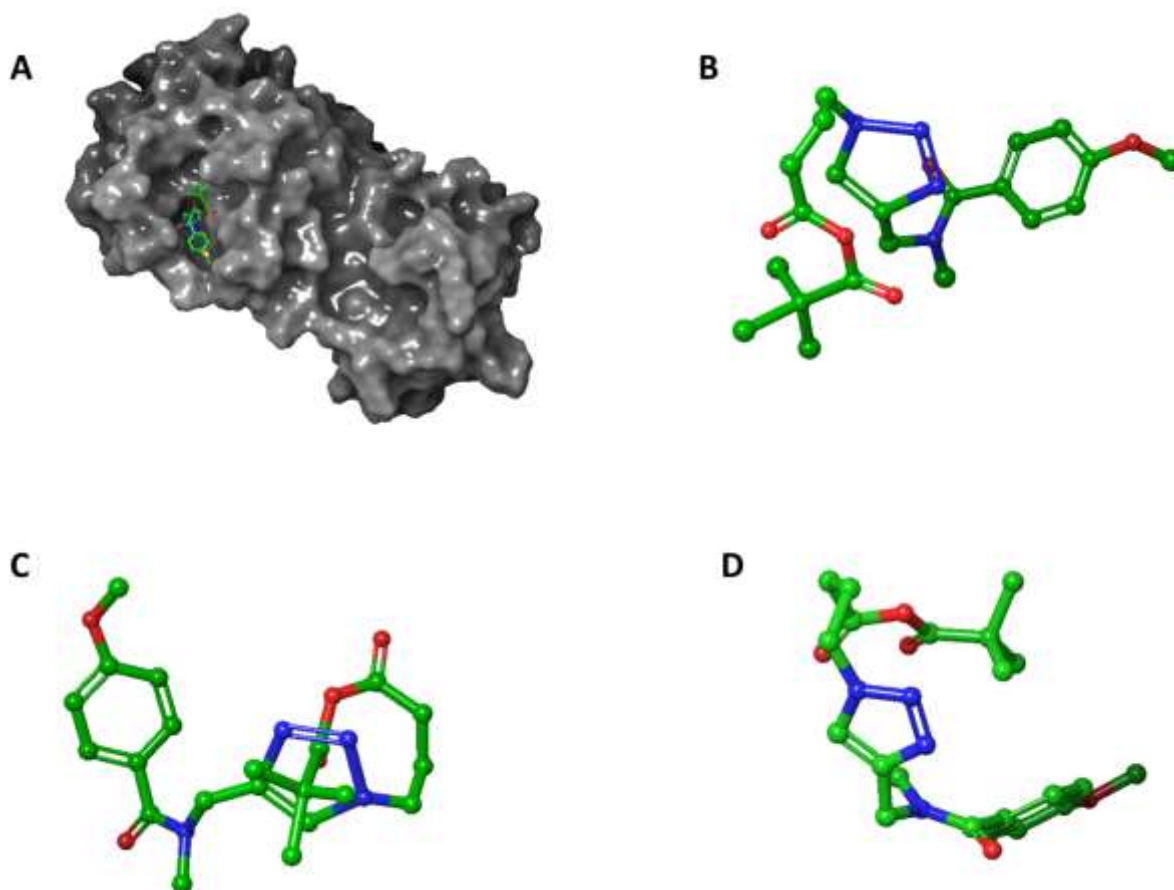


Slika 12: Analiza rezultatov sidranja ligandov iz zbirke Chemotheca v receptorско mesto proteinske tarče 3CLpro, dobljenih s programom RxDock. (A) Porazdelitev ligandov po vrednostih cenilnih funkcij (Score, Score_INTER, Score_INTRA) za najboljših 200 ligandov, razvrščenih po vrednostih cenilne funkcije Score_INTER. (B) Porazdelitev vseh 3978 ligandov po vrednostih cenilnih funkcij. (C) Vrednost cenilne funkcije Score_INTER v odvisnosti od položaja liganda v datoteki ligandov razvrščenih po vrednostih Score_INTER. (D) Vrednost cenilne funkcije Score v odvisnosti od položaja liganda v datoteki ligandov razvrščenih po vrednostih Score_INTER.

Namen sidranja je, da iz večjega nabora spojin izberemo nekaj spojin za nadaljnje analize in morebitna nadaljnja biološka testiranja. S pomočjo analize histogramov in grafov na sliki 12 lahko izoblikujemo kriterije za izbor spojin. Prvi kriterij, ki smo ga izbrali je, da smo iz množice sidranih spojin izločili tiste, ki imajo vrednost Score višji od -20 . Na histogramu za Score (slika 12(B)) je razvidno, da je pojavnost najvišja pri -20 , nato pa hitro pada. V skupini 3978 ligandov ima Score_INTER normalno porazdelitev s povprečno vrednostjo -20 , vrednosti pa se nahajajo med $-32,2$ in $-5,5$, Score_INTRA pa ima vrednosti med $-320,4$ in $7,4$ s povprečno vrednostjo $-2,5 \pm 3,6$. Iz diagrama 12 (C) je razvidno, da ima okoli 200 ligandov Score_INTER precej boljši (manjše vrednosti) od preostalih ligandov, zato sem se odločila, da bi bilo smiselno za nadaljnjo analizo izbrati le 200 ligandov (zato slika 12 (A)). Povprečna

vrednost za vrednost Score znotraj seta 200 izbranih ligandov znaša $-25,7 \pm 2,6$, medtem ko znaša za Score_INTER $-26,1 \pm 1,5$ in za Score_INTRA $0,4 \pm 2,0$.

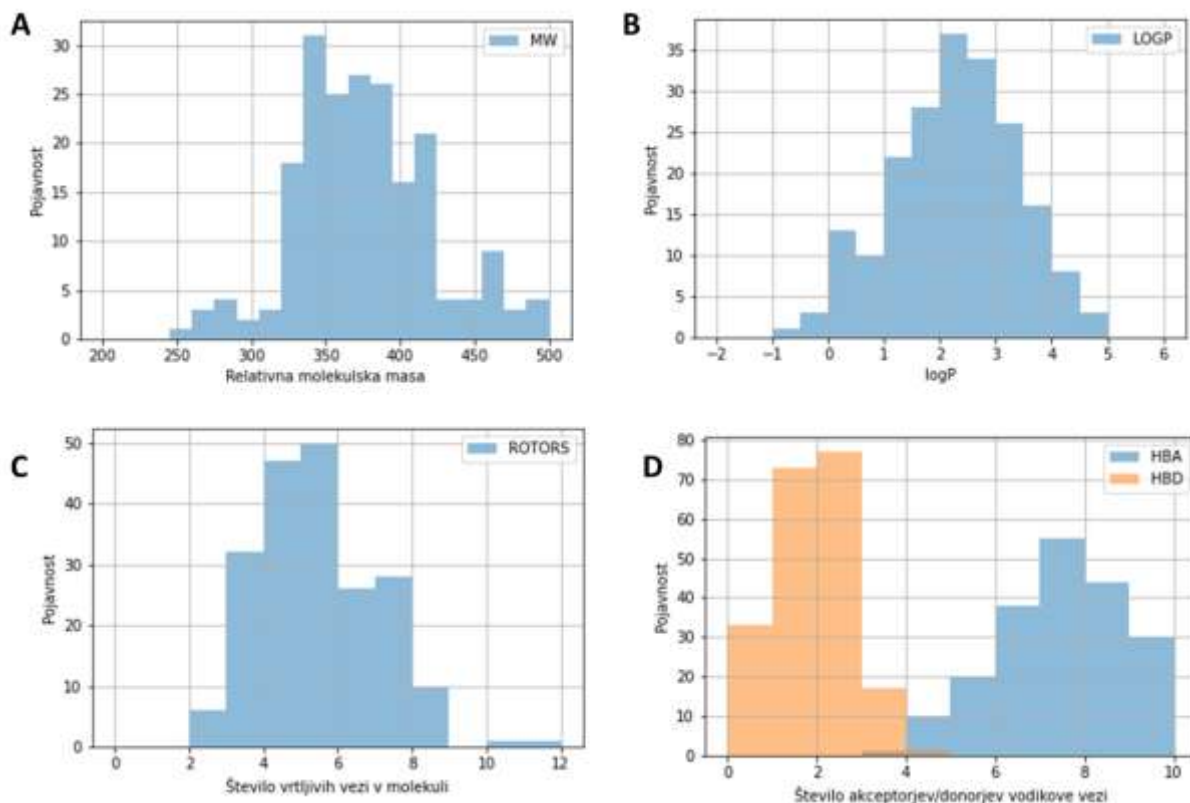
Pri analizi rezultatov sem opazila, da imajo nekatere molekule nenavadno nizko (ugodno) vrednost cenilne funkcije Score. Dodatna analiza je pokazala, da so omenjene ekstremno nizke vrednosti posledica zelo nizkih vrednosti cenilne funkcije Score_INTRA, ki je definirana kot razlika med ocenjeno energijo konformacije, kot jo ima ligand v vezanem stanju in ocenjeno energijo konformacije prostega liganda (iz vhodne datoteke). V primeru, da ima ligand na vhodu precej višjo energijo kot vezan ligand, bi to lahko pomenilo, da je nevezan ligand v neki "čudni" konformaciji. Ker sem ta efekt opazila za množico ligandov, sem se odločila, da omenjeno situacijo raziščem na primeru enega izmed ligandov, to je ligand CMLDID28454 (tj. identifikacijska številka liganda v zbirki Chemotheca). Najprej sem poiskala strukturo nesidranega liganda CMLDID28454 in ugotovila, da je geometrija liganda zelo čudna (glej sliko 13 (C)). Molekula liganda je neobičajno čudno zvit, saj ena izmed stranskih verig poteka skozi triazolni obroč. Posledica tega je, da je notranja energija molekule v prostem stanju precej višja od tiste, ki jo ima v vezanem stanju (slika 13 (B)), kar se odraža v zelo nizki vrednosti Score_INTRA, ki znaša kar -234 . Vrednosti Score_INTRA so običajno v intervalu med -10 in 10 . Ker me je zanimalo, od kje izhaja omenjena napaka, sem preverila, v kakšni konformaciji se nahaja ligand v zbirki Chemotheca. Presenečena sem ugotovila, da je ligand v napačni konformaciji zapisan že v omenjeni podatkovni zbirki. Verjetno bi se omenjeni težavi izognili, če bi na vhodu imeli ligand v optimalni konformaciji (slika 13 (D)) namesto v popačeni konformaciji (slika 13 (C)).



Slika 13: Analiza rezultatov pri vezavi liganda CMLDID28454 v vezavno mesto proteinske tarče 3CLpro. (A) Ligand v vezavnem mestu proteina. (B) Konformacija liganda, kot jo ima ligand v vezavnem mestu. (C) "Napačna" konformacija liganda, ki jo najdemo v zbirki ChEMBL. (D) Optimalna konformacija prostega (nevezanega) liganda.

Rezultati molekulskega sidranja so zelo odvisni od kvalitete vhodnih podatkov, pozorni moramo biti tako pri pripravi vhodnih struktur proteinskih tarč kot tudi ligandov. V ogromnih knjižnicah se običajno pojavijo napake, ki lahko privedejo do napačnih zaključkov.

Za najvišje razvrščenih 200 ligandov sem naredila tudi analizo razvrstitev ligandov po relativni molekularni masi, logP, številu vrtljivih vezi, številu akceptorjev in donorjev vodikovih vezi v molekulah. Porazdelitveni koeficient P opisuje porazdelitev topljenca med dvema topiloma, ki se ne mešata. Običajno je to razmerje med množinsko koncentracijo topljenca v oktanolu in množinsko koncentracijo topljenca v vodi. Porazdelitveni koeficient je merilo hidrofilitnosti in hidrofobnosti topljenca. Običajno navajamo vrednosti logP. Porazdelitveni koeficient se npr. uporablja tudi pri oceni distribucije zdravila v telesu. Rezultati omenjenega razvrščanja ligandov so prikazani v obliki histogramov (slika 14). Ti histogrami so pomembni pri razumevanju ADMET lastnosti. Preverila sem, če obravnavani ligandi ustrezajo pravilu Lipinskega.



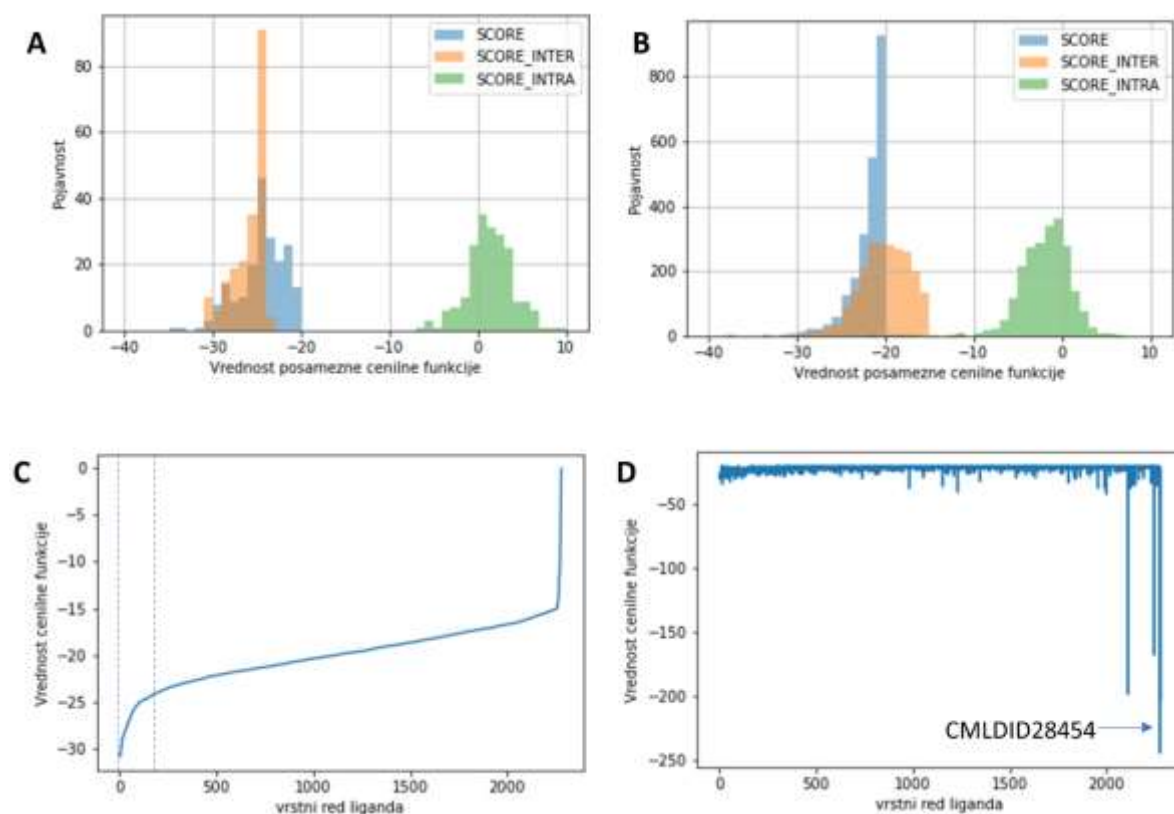
Slika 14: Razvrstitev “najboljših 200” ligandov iz zbirke Chemotheca, sidranih v receptorско mesto 3CL proteaze virusa SARS-CoV-2, glede na: (A) Relativno molekulska masa, (B) logP, (C) število vrtljivih enojnih vezi, (D) število akceptorjev/donorjev vodikove vezi.

Iz slike 14 lahko razberemo, da večina ligandov ustreza kriterijem Lipinskega, saj je relativna molekulska masa manjša od 500, vrednosti logP so manjše od 5, število donorjev vodikove vezi je manjše od 5, število akceptorjev vodikovih vezi pa manjše od 10.

4.1.2 Proteinska tarča 2: PL proteaza (PLpro)

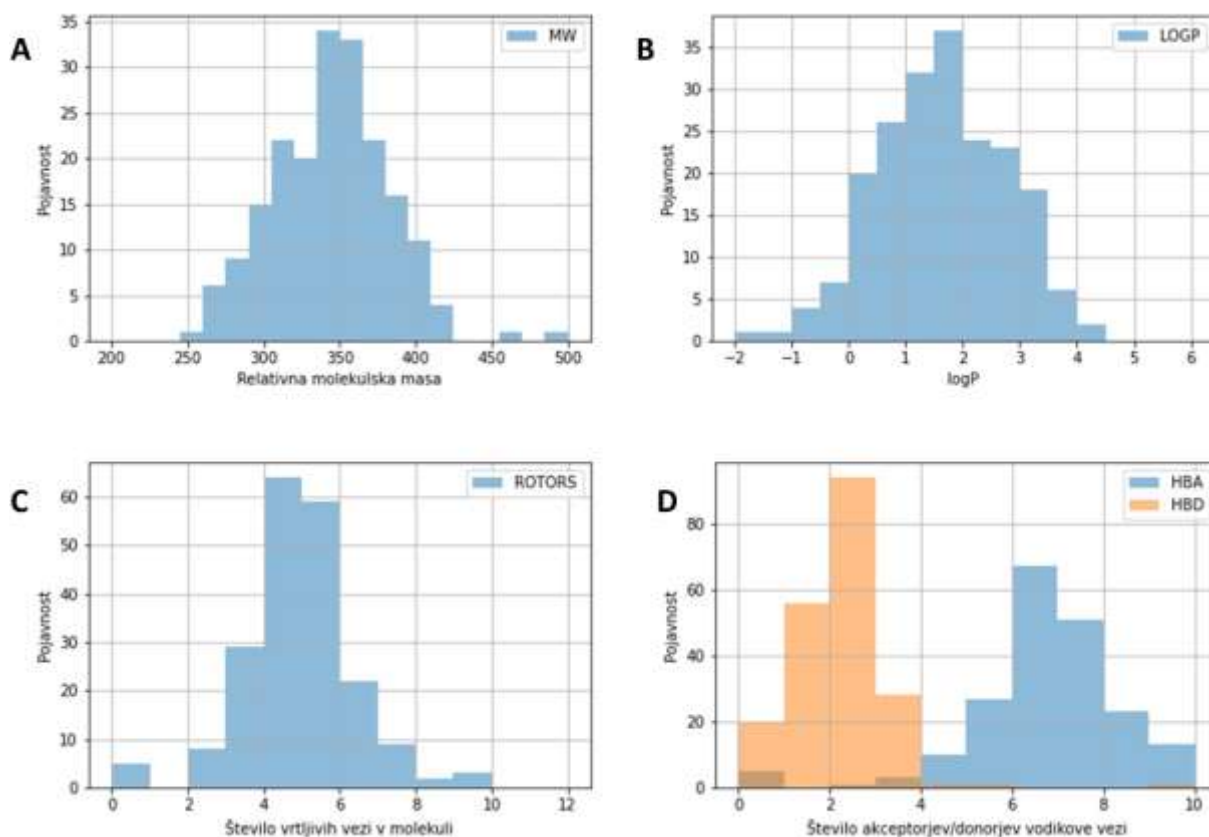
Porazdelitve po vrednostih cenilne funkcije sem izračunala za uspešno sidrane ligande iz zbirke Chemotheca v proteinsko tarčo PLpro virusa SARS-CoV-2. Porazdelitev za najvišje uvrščenih 200 ligandov je prikazana na sliki 15 (A), za vseh 2279 uspešno sidranih ligandov pa na sliki 15 (B). Slika 15 (C) ponazarja odvisnost vrednosti cenilne funkcije (Score_INTER) od pozicije, ki jo ima ligand v datoteki, kjer so ligandi razvrščeni po vrednosti Score_INTER. Na sliki 15 (D) pa je prikazana vrednost vrednosti cenilne funkcije Score v odvisnosti od položaja liganda v omenjeni datoteki.

Tudi v tem primeru sidranja ligandov iz zbirke Chemotheca sem opazila, da ima nekaj ligandov nenavadno nizke (ugodne) vrednosti cenilne funkcije Score. Izkaže se, da so tudi te vrednosti posledica “čudne” konformacije omenjenih ligandov na vhodu, že pred njihovim sidranjem.



Slika 15: Analiza rezultatov sidranja ligandov iz zbirke Chemotheca v vezavno mesto proteina PLpro, dobljenih s programom RxDock. (A) Porazdelitev ligandov po vrednosti cenilnih funkcij (Score, Score_INTER, Score_INTRA) za najvišje, po vrednosti Score_INTER, razvrščenih 200 ligandov. (B) Porazdelitev vseh 2279 ligandov po vrednostih cenilnih funkcij. (C) Vrednost cenilne funkcije Score_INTER v odvisnosti od položaja liganda v datoteki ligandov razvrščenih po vrednosti Score_INTER. (D) Vrednost cenilne funkcije Score v odvisnosti od položaja liganda v datoteki ligandov razvrščenih po vrednosti Score_INTER.

Histogrami na slikah 15 (A) in 15 (B), ki prikazujejo porazdelitev vrednosti cenilnih funkcij dobljenih pri sidranju ligandov v vezavno mesto proteina PLpro, imajo podobno obliko kot tisti za 3CLpro (glej sliko 12 (A) in 12 (B) – v poglavju 4.1.1). Na histogramu za Score (slika 15 (B)) je razvidno, da je pojavnost najvišja pri -20 , nato pa hitro pada. Score_INTER ima normalno porazdelitev s povprečno vrednostjo $-20,1$, vrednosti pa se nahajajo med $-30,8$ in $-1,6$. Score_INTRA pa zavzema vrednosti med $-242,1$ in $9,1$ s povprečno vrednostjo $-2,0 \pm 3,1$. Na podlagi diagrama (slika 15 (C)), sem se odločila, da bi bilo za dodatno analizo smiselno izbrati 200 ligandov, podobno kot pri proteinski tarči 3CLpro. Povprečna vrednost za vrednosti Score znotraj seta 200 izbranih ligandov znaša $-24,5 \pm 2,7$, medtem ko znaša za Score_INTER $-25,9 \pm 1,8$ in za Score_INTRA $-2,0 \pm 3,1$.



Slika 16: Razvrstitev “najboljših 200” ligandov iz zbirke Chemotheca, sidranih v receptorsko mesto proteinske tarče PLpro, glede na: (A) Relativno molekularno maso, (B) logP, (C) število vrtljivih enojnih vezi, (D) števila akceptorjev/donorjev vodikove vezi pri sidranju ligandov v vezavno mesto PLpro.

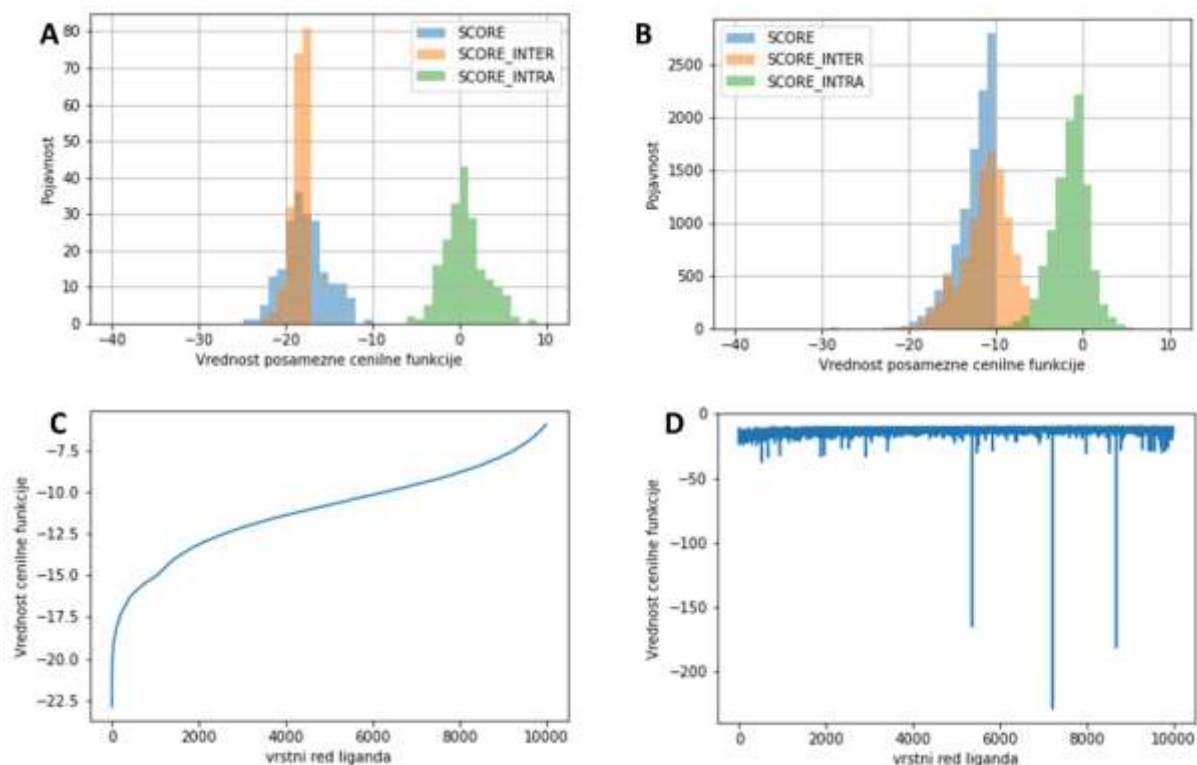
Iz slike 16 lahko opazimo, podobno kot pri 3CLpro tudi pri PLpro, da večina ligandov iz zbirke Chemotheca ustreza kriterijem Lipinskega – relativna molekularna masa je manjša od 500, logP je manjši od 5, tudi število akceptorjev vodikove vezi je manjše od 10 in donorjev manjše od 5. Le nekaj ligandov je takih, da imajo preveliko število donorjev vodikove vezi.

4.1.3 Proteinska tarča 3: RBD proteina Spike (krajše protein S)

4.1.3.1 Ligandi iz zbirke Chemotheca

Protein Spike virusa SARS-CoV-2 je tretja tarča, za katero sem analizirala rezultate sidranja ligandov iz zbirke Chemotheca. Analiza je prikazana na sliki 17. Tu sem najprej odstranila ligande z vrednostjo Score > -10 . Takih ligandov, ki so imeli vrednost cenilne funkcije manjšo od -10 , je bilo 9997, takih z vrednostjo manjšo od -15 , je bilo 1339, takih z vrednostjo manjšo od -20 , pa je bilo le 97. Porazdelitev po vrednostih cenilnih funkcij je prikazana na sliki 17 (B). Opazim lahko, da sta vrhova porazdelitev po vrednosti Score in Score_INTER pomaknjena bližje k vrhu Score_INTRA, torej drugače kot pri rezultatih sidranja ligandov na proteinskih tarčah 3CLpro in PLpro.

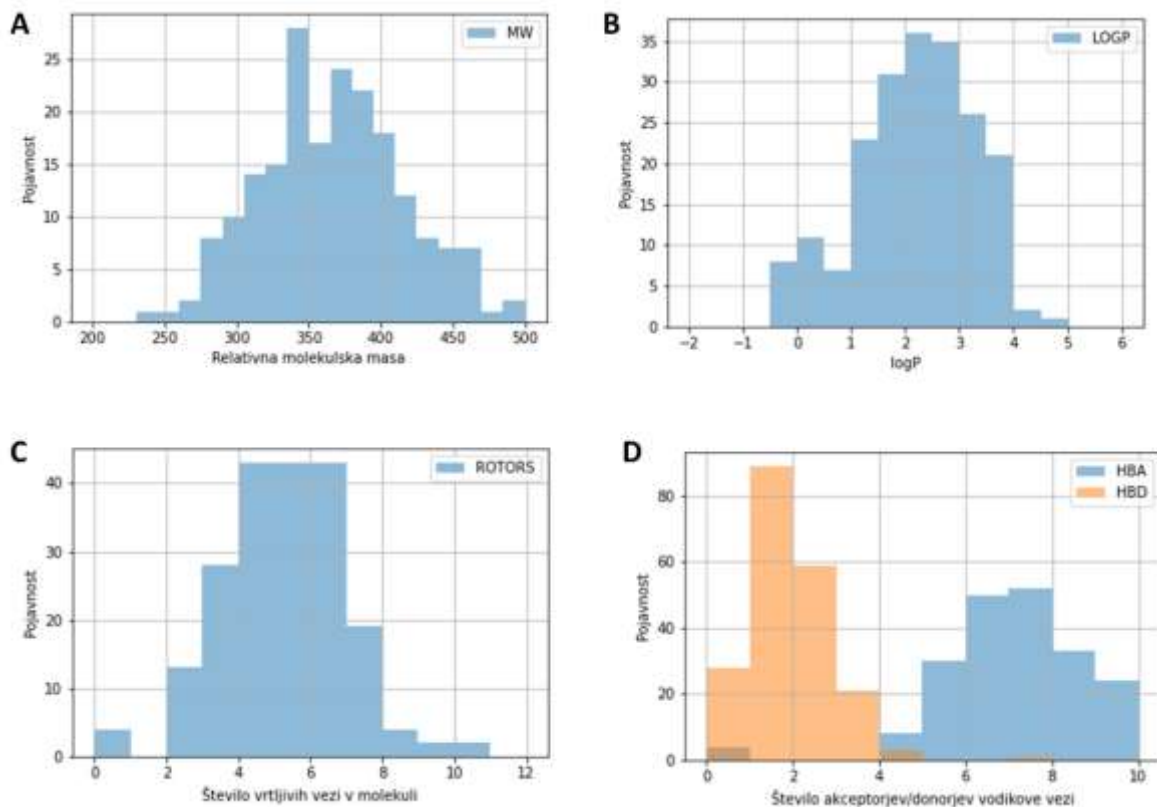
*



Slika 17: Analiza rezultatov sidranja ligandov iz zbirke ChEMBL v del proteina S, ki se veže na receptor ACE2, dobljenih s programom RxDock. (A) Porazdelitev ligandov po vrednosti cenilnih funkcij (Score, Score_INTER, Score_INTRA) za najvišje, po vrednosti cenilnih funkcij (Score, Score_INTER, Score_INTRA) za najvišje, po vrednosti Score_INTER, razvrščenih 200 ligandov. (B) Porazdelitev vseh 9997 ligandov po vrednostih cenilnih funkcij. (C) Vrednost cenilne funkcije Score_INTER v odvisnosti od položaja liganda v datoteki ligandov razvrščenih po vrednosti Score_INTER. (D) Vrednost cenilne funkcije Score v odvisnosti od položaja liganda v datoteki ligandov razvrščenih po vrednosti Score_INTER.

Pri proteinu S opazimo, da se je v vezavno mesto uspešno sidralo precej manjše število ligandov (ob kriteriju cenilna funkcija $\text{Score} < -20$). Vezavno mesto proteina S je večje in bolj odprto kot pri prejšnjih dveh proteinskih tarčah, specifičnost vezave pa je precej manjša, kar lahko razberemo na osnovi nižje povprečne vrednosti cenilne funkcije Score_INTER, kot so tiste pri drugih dveh proteinskih 3CLpro in PLpro. Na histogramu za Score (slika 17 (B)) je razvidno, da je pojavnost najvišja pri -10 , nato pa hitro pojemata. Score_INTER ima normalno porazdelitev s povprečno vrednostjo $-11,1$ vrednosti pa se nahajajo med $-22,7$ in $-6,0$. Score_INTRA pa ima vrednosti med $-219,5$ in $10,5$ s povprečno vrednostjo $-1,5 \pm 2,3$. Na podlagi diagrama (slika 17 (C)) sem se odločila, da bi bilo za dodatno analizo smiselno izbrati 200 ligandov. Povprečna vrednost za Score znotraj seta 200 izbranih ligandov znaša $-17,8 \pm 2,5$, medtem ko znaša za Score_INTER $-18,5 \pm 0,9$ in za Score_INTRA $0,7 \pm 2,4$.

Tudi pri rezultatih sidranja spojin iz zbirke Chemotheca na protein Spike opazimo tri ligande z ekstremno nizko vrednostjo cenilne funkcije Score, razlog za takšne anomalije so najbrž slabo pripravljene izhodiščne (vhodne) strukture.



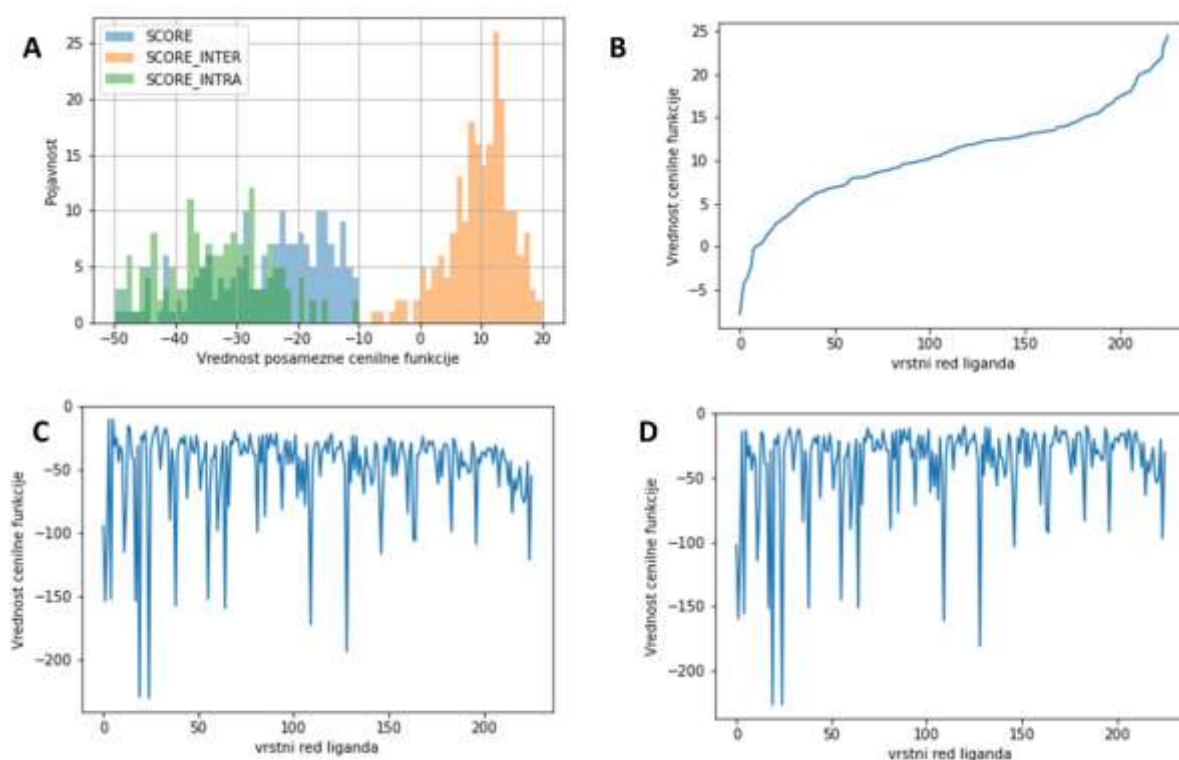
Slika 18: Razvrstitev "najboljših 200" ligandov iz zbirke Chemotheca, sidranih v receptorsko mesto proteina S virusa SARS-CoV-2, glede na: (A) Relativno molekulsko maso, (B) logP, (C) števila vrtljivih enojnih vezi, (D) števila akceptorjev/donorjev vodikove vezi.

Porazdelitev lastnosti ligandov iz zbirke Chemotheca, ki so se uspešno sidrali v vezavno mesto proteina S (slika 18) je zelo podobna tisti pri 3CLpro in PLpro, s to razliko, da se v vezavno mesto veže nekaj več ligandov z višjo molekulsko maso, to so ligandi z relativno molekulsko mase večjo od 400. Večina ligandov iz zbirke Chemotheca, ki smo jih analizirali, zadošča kriterijem, ki jih narekuje pravilo Lipinskega.

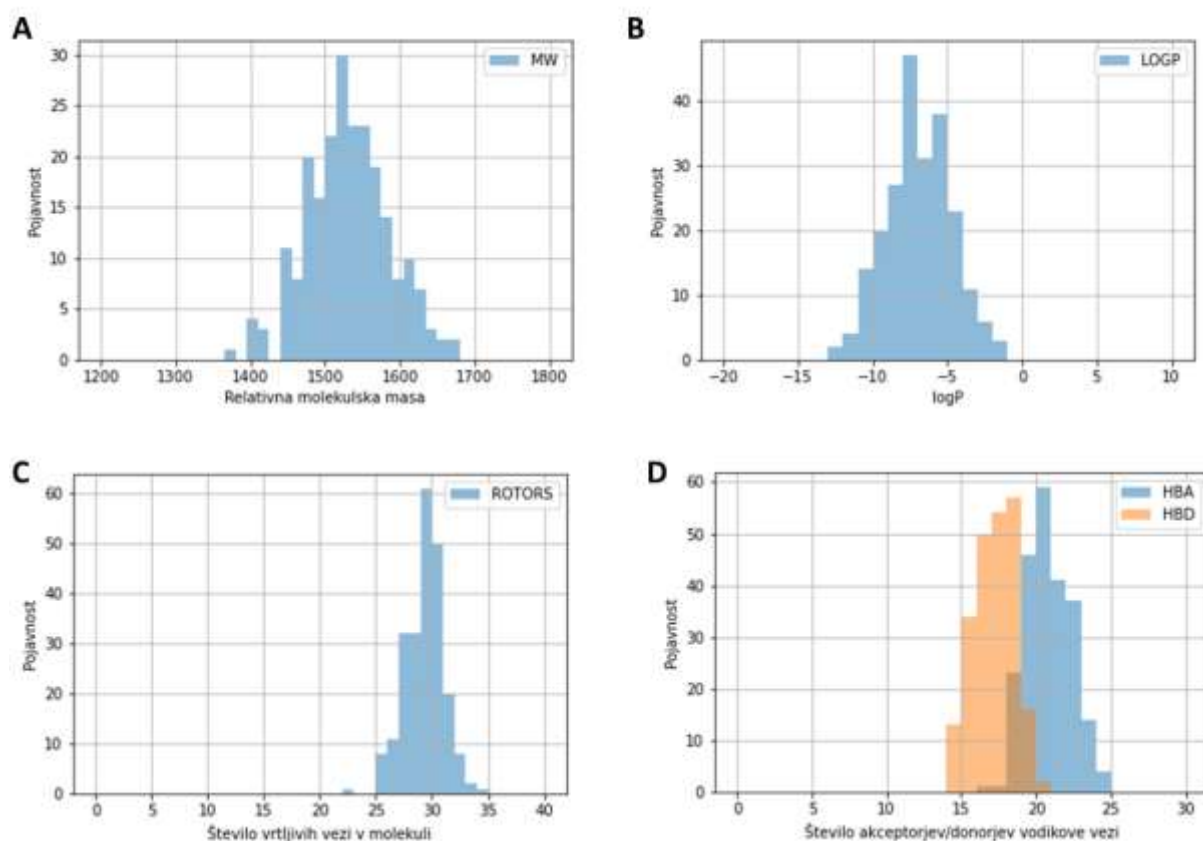
4.1.3.2 Ligandi iz zbirke Peptidi

Analiza sidranja 1000 različnih dodekapeptidov v vezavno mesto proteina Spike virusa SARS-CoV-2 je prikazana na sliki 19. Tako kot v primeru ligandov iz zbirke Chemotheca sem tudi pri dodekapeptidih odfiltrirala rezultate sidranja z vrednostjo Score > -10. Pri sidranju večjih ligandov (peptidi) imamo povsem drugačno razporeditev po vrednostih cenilne funkcije kot v primeru sidranja manjših ligandov (glej slike 12 (A), 15 (A) in 17 (A)). Za "najboljših"

200 ligandov (Slika 19 (A)) opazimo, da je Score_INTER skoraj v vseh primerih pozitiven, povprečna vrednost omenjene cenilne funkcije znaša $10,6 \pm 5,7$, minimalna vrednost je pri $-7,8$, maksimalna pa pri $24,5$. V nasprotju s porazdelitvijo vrednosti Score_INTER pri ostalih dveh porazdelitvah vrhovi niso izraziti. Povprečna vrednost za Score je $-38,3 \pm 35,8$, vrednosti pa ležijo na intervalu med $-226,8$ in $-10,0$, medtem ko je bila povprečna vrednost za Score_INTRA $-49,0 \pm 34,8$ z minimumom pri $-230,1$ in maksimumom pri $-10,3$. Takšna porazdelitev lahko pomeni, da je ugodna vrednost cenilne funkcije posledica preureditve peptida v ugodnejšo konformacijo ob vezavi. Pozitivna povprečna vrednost za Score_INTER lahko pomeni, da je vezava peptidov izrazito nespecifična. Iz slik 19 (C) in 19 (D) lahko razberemo, da ima ogromno ligandov zelo negativen Score_INTRA in posledično zelo negativno vrednost Score. Izgleda, da je konformacija liganda v vezanem stanju v teh primerih precej bolj ugodna kot konformacija v prostem stanju. Ker je šlo za peptide, ki imajo ogromno število vrtljivih vezi, je zelo težko raziskati, kaj je vzrok temu obnašanju. Morda je razlog ta, da program RxDock ob trenutnih nastavitvah ni najbolj primeren za sidranje peptidov.



Slika 19: Analiza rezultatov sidranja ligandov iz zbirke dodekapeptidov v del proteina S, ki se veže na receptor ACE2, dobljenih s programom RxDock. (A) Porazdelitev ligandov po vrednosti cenilnih funkcij (Score, Score_INTER, Score_INTRA) za 200 najvišje razvrščenih peptidov po vrednosti Score_INTER. (B) Vrednost cenilne funkcije Score_INTER v odvisnosti od položaja peptidov v datoteki, razvrščenih po vrednosti Score_INTER. (C) Vrednost cenilne funkcije Score_INTRA v odvisnosti od vrstnega reda peptidov. (D) Vrednost cenilne funkcije Score od vrstnega reda peptidov.



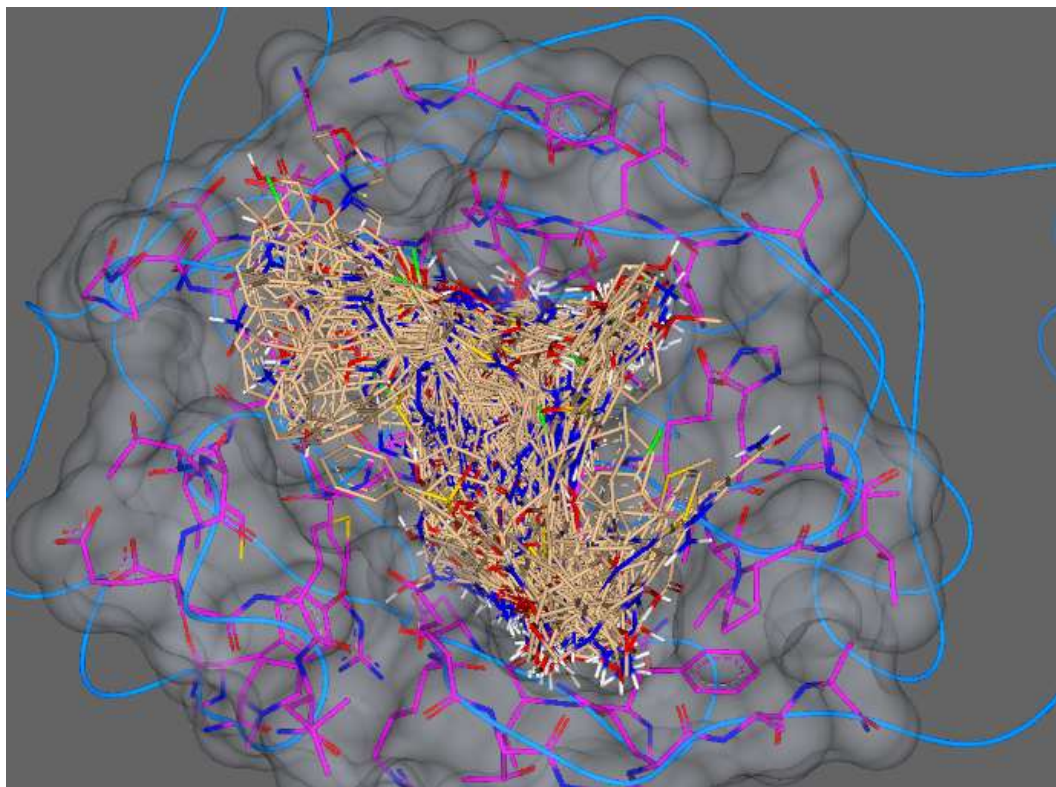
Slika 20: Razvrstitev najboljših 200 dodekapeptidov, sidranih v receptorsko mesto proteina S virusa SARS-CoV-2, glede na: (A) relativno molekularno maso, (B) logP, (C) število vrtljivih enojnih vezi, (D) število akceptorjev/donorjev vodikove vezi.

Iz slike 20 je razvidno, da peptidi kršijo v glavnem vsa pravila, ki jih je postavil Lipinski. Vendar moramo poudariti, da tu ne gre za majhne organske molekule, za katere je pravilo Lipinski napisano, marveč za peptide, ki praviloma odstopajo od omenjenega pravila. Konformacijsko zaklenjeni peptidi, ki jih opisujem v tem delu, lahko vstopajo v organizem, uvedena premostitev z bicikličnim aromatskim sistemom namreč poveča njihovo metabolično stabilnost.

4.2 Obdelava najboljših ligandov glede na RxDock s programom SeeSAR

4.2.1 Proteinska tarča 1: 3CL proteaza (3CLpro)

Na sliki 21 je prikazan rezultat sidranja dvesto ligandov iz knjižnice Chemotheca v vezavno mesto proteinske tarče 3CLpro virusa SARS-CoV-2. Prikazani ligandi so tisti, ki jih je program za sidranje RxDock razvrstil na prva mesta glede na vrednost cenilne funkcije Score_INTER.



Slika 21: Rezultat sidranja dvestotih ligandov v vezavno mesto proteinske tarče 3CLpro.

Iz slike 21 je razvidno, da se ligandi sidrajo na različne načine, na sliki so ogljikovi atomi ligandov predstavljeni s svetlo rjavo, atomi kisika z rdečo, atomi dušika pa z modro barvo. V primeru proteinske tarče 3CLpro je vezavno mesto definiralo 56 aminokislinskih ostankov. Vsak izmed 200 ligandov se je torej povezal z nekaterimi izmed teh 56 aminokislinskih delov.

Razporeditev najboljših dvajset ligandov po vrednostih cenilne funkcije Score_INTER (RxDock) in HYDE (SeeSAR) za tarčo 3CLpro je prikazana na sliki 22. V preglednicah na sliki 22 so napisana imena ligandov; identifikacijski številki liganda iz zbirke Chemotheca je dodana številka glede na razvrstitev po vrednosti Score_INTER (sidranje z RxDock-om). Vrednost Score_INTER je v sliki označena z SDF28.

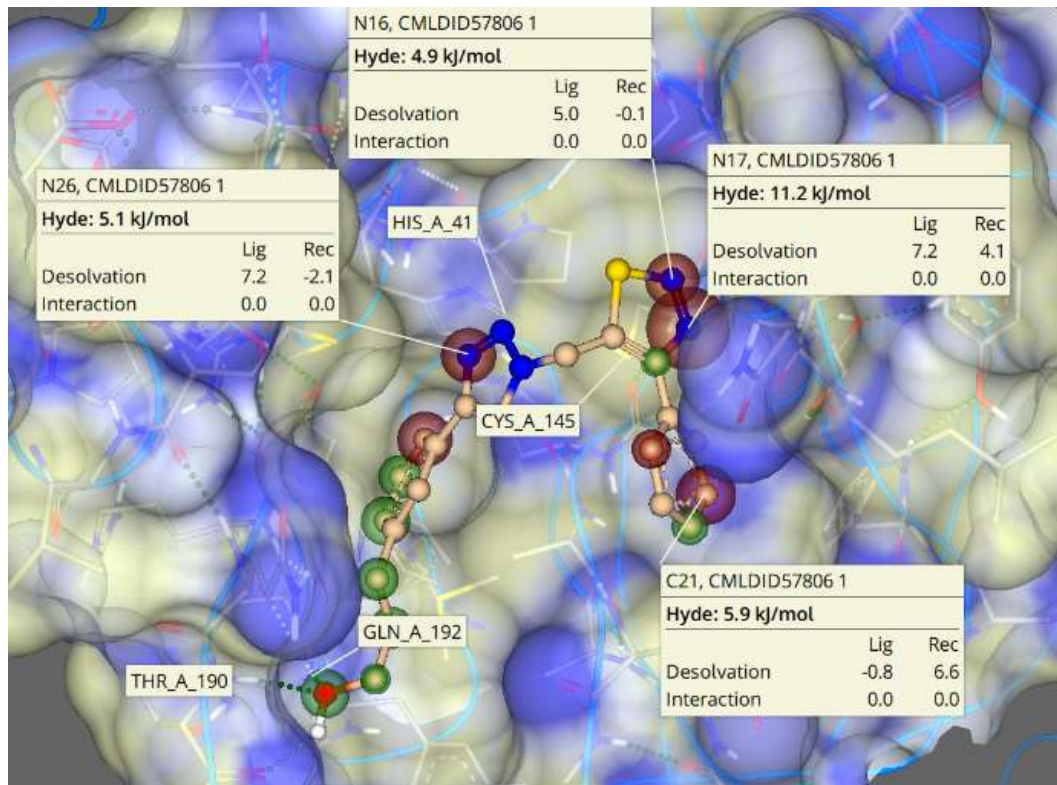
A	Name	Estimated affinity				MW	SDF28	B	Name	Estimated affinity				MW	SDF28
		pM	nM	µM	mM					pM	nM	µM	mM		
	CMLDID57806 1				▶	377.47	-32.235		CMLDID8394 165				▶	329.81	-24.798
	CMLDID9044 2				▶	412.43	-31.356		CMLDID58910 177				▶	442.92	-24.709
	CMLDID35344 3				▶	485.83	-30.742		CMLDID55434 56				▶	443.53	-26.622
	CMLDID15300 4				▶	252.30	-30.057		CMLDID8871 23				▶	421.47	-28.008
	CMLDID65703 5				▶	347.38	-29.864		CMLDID8232 99				▶	450.52	-25.620
	CMLDID37543 6					401.45	-29.772		CMLDID15437 80				▶	421.47	-25.974
	CMLDID55469 7				▶	347.37	-29.575		CMLDID15613 11				▶	493.41	-29.108
	CMLDID55867 8				▶	361.38	-29.569		CMLDID55477 39				▶	424.48	-27.352
	CMLDID54984 9				▶	362.39	-29.405		CMLDID15774 92				▶	370.45	-25.758
	CMLDID_537 10				▶	349.41	-29.340		CMLDID2212 143				▶	467.01	-25.057
	CMLDID_613 11				▶	493.41	-29.108		CMLDID4073 195				▶	389.34	-24.568
	CMLDID_472 12				▶	349.36	-28.909		CMLDID52404 35				▶	482.95	-27.502
	CMLDID_296 13				▶	412.45	-28.674		CMLDID55131 62				▶	389.37	-26.363
	CMLDID_792 14				▶	346.36	-28.668		CMLDID30928 190				▶	331.13	-24.590
	CMLDID_289 15				▶	369.39	-28.526		CMLDID16175 189				▶	462.53	-24.596
	CMLDID_098 16				▶	353.83	-28.393		CMLDID55853 104				▶	339.42	-25.588
	CMLDID8237 17					383.80	-28.352		CMLDID6011 78				▶	358.40	-26.038
	CMLDID_147 18				▶	415.47	-28.307		CMLDID35344 3				▶	485.83	-30.742
	CMLDID8765 19				▶	395.85	-28.189		CMLDID51098 16				▶	353.83	-28.393
	CMLDID_879 20					373.44	-28.171		CMLDID55819 147				▶	384.39	-24.989

Slika 22: Razporeditev najboljših 20 ligandov iz knjižnice spojin Chemotheca po (A) vrednosti cenilne funkcije Score_INTER (RxDock), v preglednici označeno z SDF28 in (B) vrednosti cenilne funkcije HYDE (SeeSAR), v preglednici označeno z Estimated affinity. Rezultati so pridobljeni s sidranjem ligandov v vezavno mesto proteina 3CLpro s programom RxDock in naknadnim izračunom afinitete s programom SeeSAR.

Ugotovila sem, da program SeeSAR povsem drugače razvrsti ligande, kot jih je razvrstil program RxDock, kar je po eni strani razumljivo, saj programa uporabljata različni cenilni funkciji. Program SeeSAR uporablja cenilno funkcijo HYDE, ki poleg samih interakcij v cenilno funkcijo vključi tudi prispevek topila, ki igra pomembno vlogo pri interakciji ligandov s proteini. Za razliko od programa SeeSAR program RxDock v verziji cenilne funkcije uporabljene v projektu COVID.SI tega prispevka ne upošteva. Najboljšega liganda glede na Score_INTER ni med najboljšimi dvajsetimi ligandi glede na ocenjeno afiniteto, izračunano s programom SeeSAR.

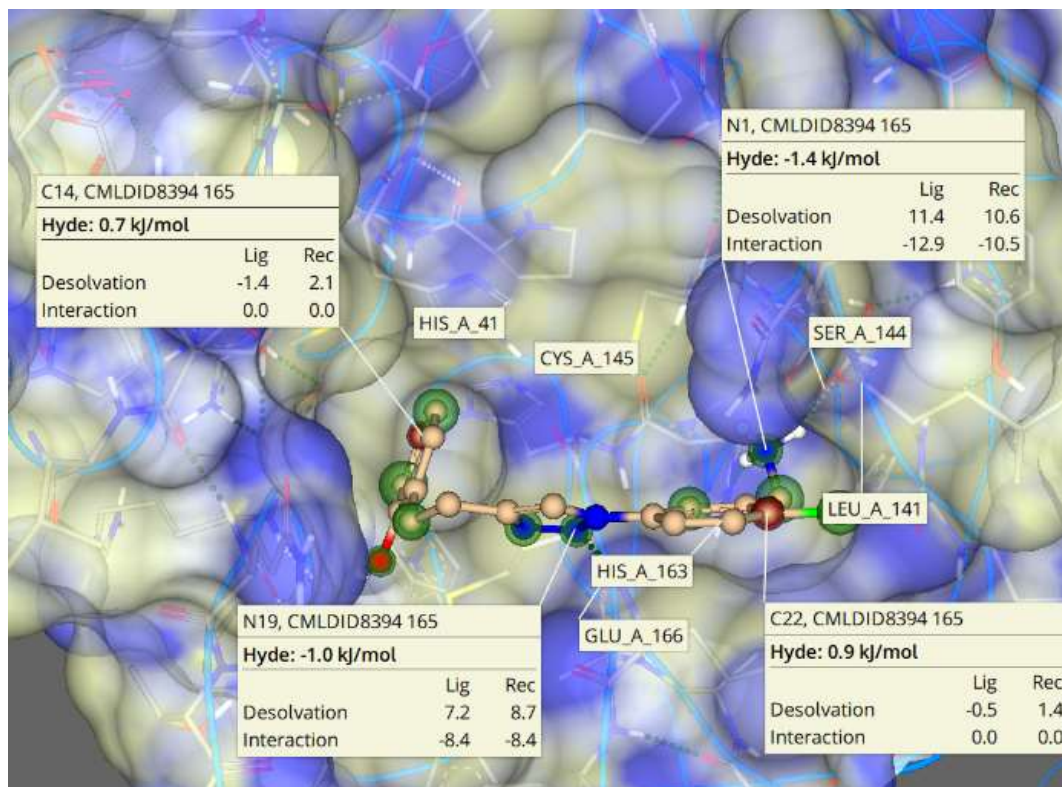
Na sliki 23 vidimo ligand z oznako CMLID57806 v vezavnem mestu tarčnega proteina, ligand je bil glede na vrednost cenilne funkcije Score_INTER (RxDock) razvrščen na prvo mesto. Iz slike je razvidno, da se ligand dobro geometrijsko prilagaja vezavnemu mestu tarčnega proteina in posledično tvori veliko število interakcij, ki se odražajo na ugodni vrednosti RxDockove cenilne funkcije. Prav tako vidimo, da ligand uspešno skriva katalitsko mesto, ki jo tvorita aminokislinska ostanka HIS 41 in CYS 145. Spojina tvori dve vodikovi vezi s samim proteinom in sicer z aminokislinskima ostankoma THR 190 in GLN 192. SeeSAR nam omogoča, da preverimo, kateri atomi imajo pozitiven (zelene sfere) in kateri negativen prispevek (rdeče sfere) k vezavni afiniteti. Atomov z negativnim prispevkom je pri tej molekuli šest. Razlog za

te neugodne atomske prispevke so neugodni desolvatacijski prispevki, na primer polarni del molekule je v kontaktu z nepolarno površino proteina in obratno, nepolarni del molekule pokriva polarno površino proteina. Analiza atomskih prispevkov k vezavni afiniteti nam je lahko v pomoč, pri načrtovanju bolj aktivnih ligandov.



Slika 23: Ligand CMLDID57806 v vezavnem mestu 3CLpro, najboljši ligand po RxDock-u

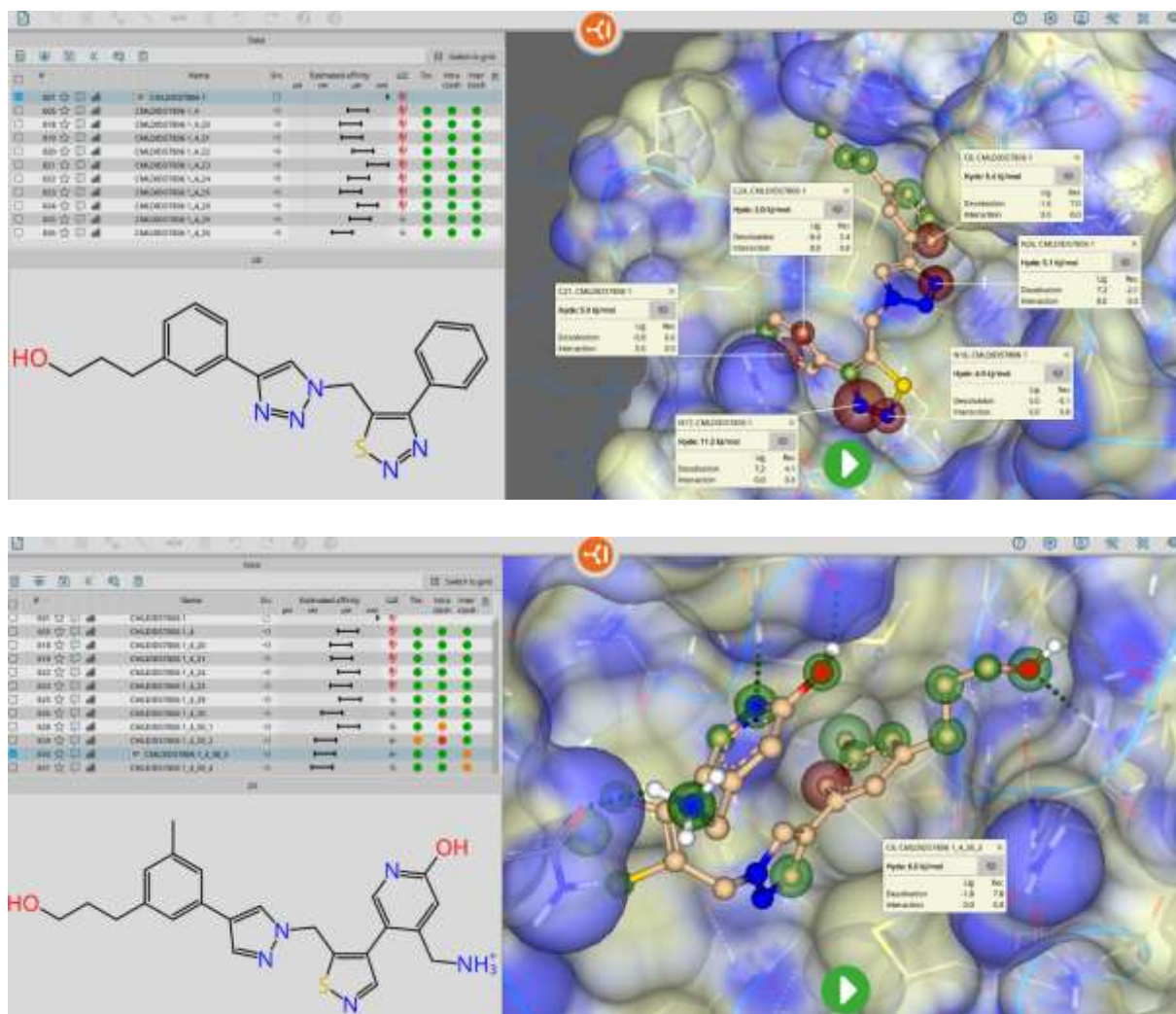
Slika 24 prikazuje vezavo liganda CMLDID8394 v vezavno mesto proteina 3CLpro, ki ga je SeeSAR prepoznal kot ligand z najboljšo ocenjeno afiniteto za vezavo na omenjen protein. Ligand tudi v tem primeru lepo zapolnjuje vezavno mesto in s tem potencialno ovira vezavo substrata na protein, katalitsko mesto (HIS 41, CYS 145) pa v tem primeru ni popolnoma zakrito. Ligand tvori štiri vodikove vezi in sicer z aminokislinskimi ostanki (GLU 166, HIS 163, LEU 151, SER 144). Rdeči sferi, ki označujeta neugodne atomske prispevke k oceni afinitete, sta dve in sicer na ogljikovih atomih C14 (0,7 kJ/mol) in na atomu C22 (0,9 kJ/mol). SeeSAR ocenjuje, da je afiniteta molekule v mikromolarnem območju. Program RxDock je razvrstil omenjeni ligand šele na 165. mesto.



Slika 24: Ligand CMLDID8394 v vezavnem mestu 3CLpro, najboljši ligand po SeeSAR-u

MODIFIKACIJA LIGANDOV

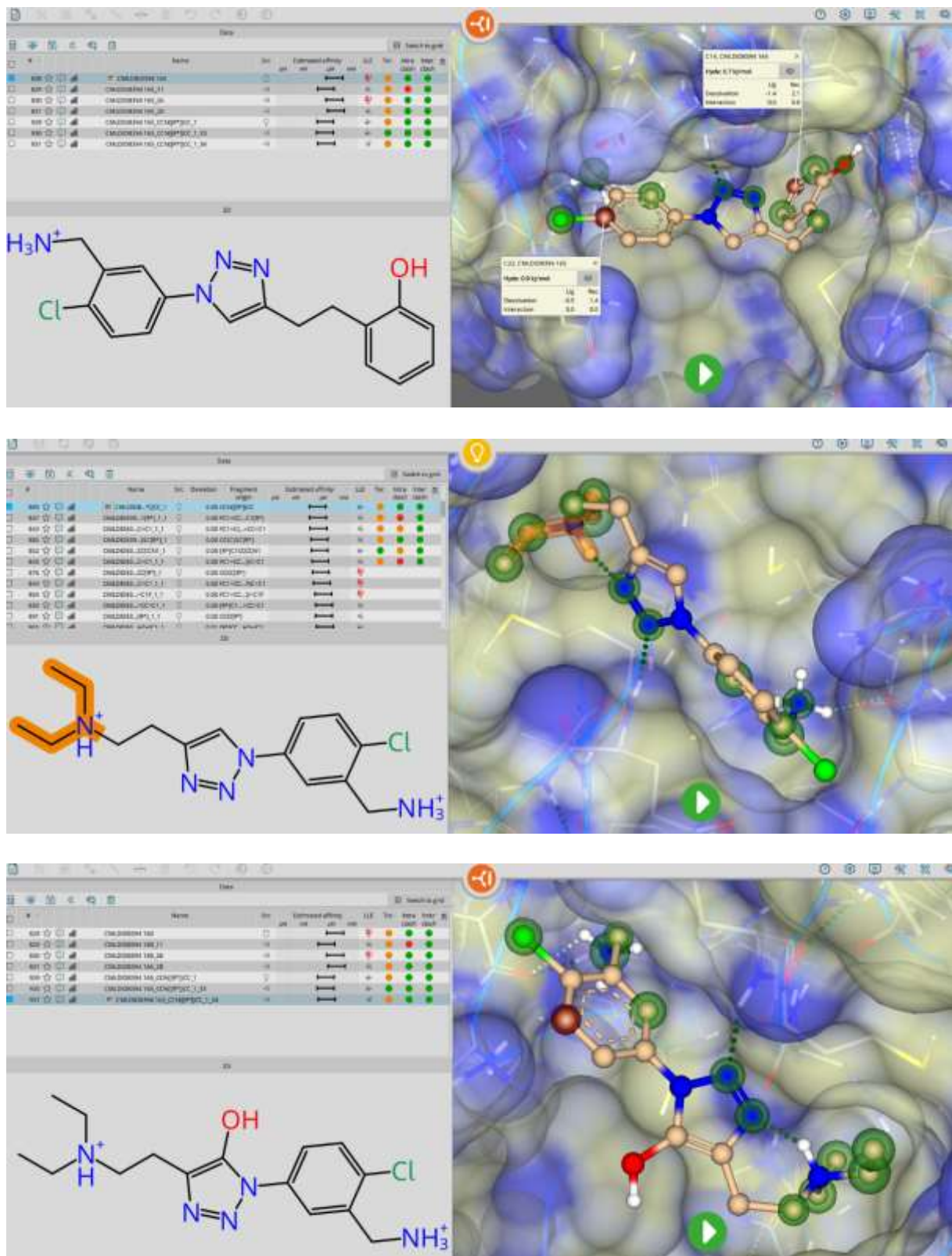
Program SeeSAR ima dve orodji, s katerima lahko molekulo liganda spreminjamo in korakoma pridemo do molekule z “boljšo” napovedano afiniteto. Ti dve orodji sta “Molecule Editor” in “Inspirator”. Na tak način sem spremenila najboljša liganda, tj. najboljšega po sidranju z RxDock-om in tistega z najboljšo ocenjeno vezavno afiniteto po SeeSAR-u. Rezultati so prikazani na slikah 25 in 26.



Slika 25: Modificiran ligand CMLDID57806 v vezavnem mestu 3CLpro, najboljši ligand po RxDock-u (zgornja slika – začetno stanje, spodnja slika – modificirano stanje)

Pri modifikaciji ligandov sem poskusila s programom SeSAR večino rdeče označenih atomov v molekuli spremeniti v zeleno obarvane, saj to običajno pomeni, da se vezavna afiniteta izboljša. Pri teh zamenjavah atomov v molekuli liganda pa je potrebno paziti, da v preoblikovani molekuli liganda ni prevelikih torzijskih napetosti (v sliki označeno Tor.) in da ni prostorskega oviranja v molekuli liganda (v sliki označeno z Intra clash) oz. med molekulama liganda in receptorja (v sliki označeno z Inter clash). Pogosto vseh rdeče obarvanih atomov ne moremo spremeniti v zeleno obarvane, še posebej so težave, če so to ogljikovi atomi, ki so del benzenovega obroča. Z njihovo zamenjavo lahko porušimo aromatsko strukturo, kar precej spremeni konformerno obliko molekule in pogosto se vezavna afiniteta precej poslabša. Poleg tega pa je smiselno spremeniti atome v molekuli liganda predvsem takrat, kadar je na teh atomih vrednost desolvatacije pozitivna in s spremembo vezanega atoma lahko dosežemo negativno (torej ugodno) vrednost desolvatacije. Pozitivne vrednosti desolvatacijskega

prispevka na receptorju ne moremo spremeniti, zato ogljikovega atoma C8 v molekuli liganda nisem zamenjala (glej sliko 25).

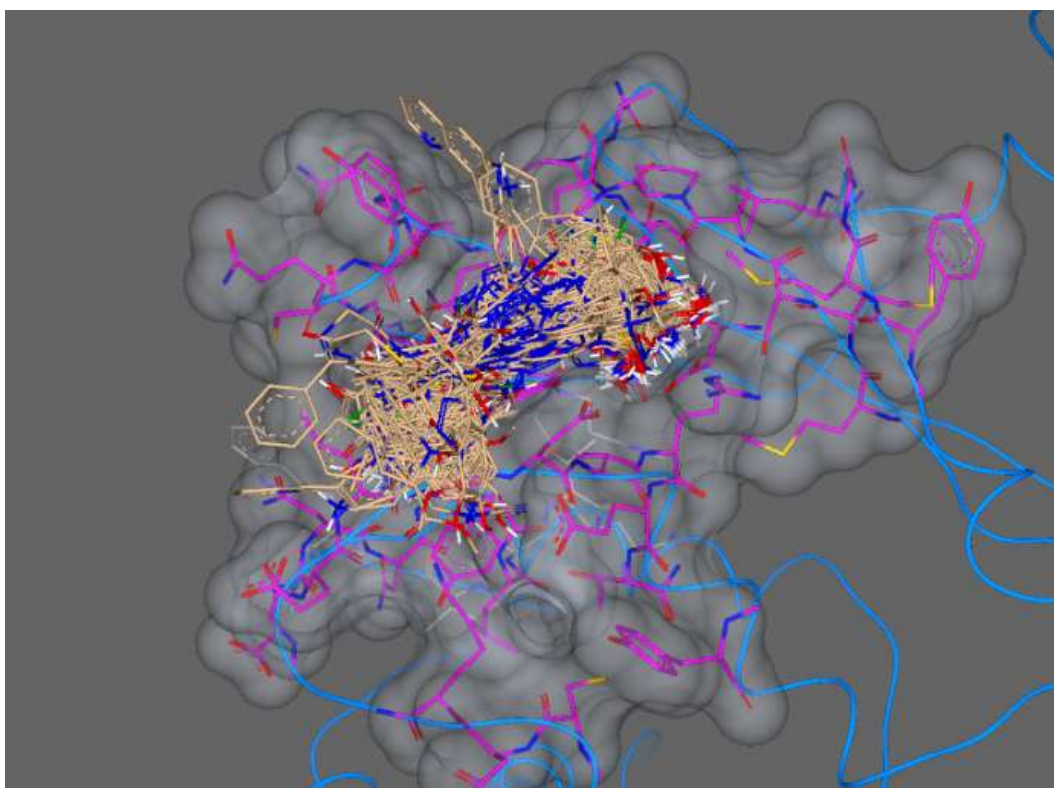


Slika 26: Modificiran ligand CMLDID8394 v vezavnem mestu 3CLpro, najboljši ligand po SeeSAR-u (zgornja slika – začetno stanje, srednja slika – menjava fragmenta (označen zamenjan fragment), spodnja slika – modificirano stanje)

Na sliki 26 je razvidno, da sem pri modifikaciji liganda CMLDID8394 uporabila orodje Inspirator, s katerim sem del molekule (benzenov obroč s hidroksilno skupino) zamenjala z novim fragmentom. Pri tej zamenjavi se je ocenjena vezavna afiniteta povečala za približno desetkrat (pomik iz μM proti nM področju), pa tudi vrednost LLE (Ligand Lipophilicity Efficiency) se je izboljšala.

4.2.2 Proteinska tarča 2: PL proteaza (PLpro)

Na sliki 27 je prikazan rezultat sidranja dvesto ligandov iz knjižnice Chemotheca v vezavno mesto proteina PLpro virusa SARS-CoV-2. Prikazani ligandi so tisti, ki jih je program za sidranje RxDock razvrstil na prva mesta glede na vrednost cenilne funkcije Score_INTER.



Slika 27: Rezultat sidranja dvestotih ligandov v vezavno mesto proteinske tarče PLpro.

Iz slike 27 je razvidno, da se ligandi sidrajo na različne načine. To smo opazili že pri proteinski tarči 3CLpro. Na sliki so ogljikovi atomi ligandov predstavljeni s svetlo rjavo, atomi kisika z rdečo, atomi dušika pa z modro barvo. V primeru proteinske tarče PLpro je vezavno mesto definiralo 46 aminokislinskih ostankov.

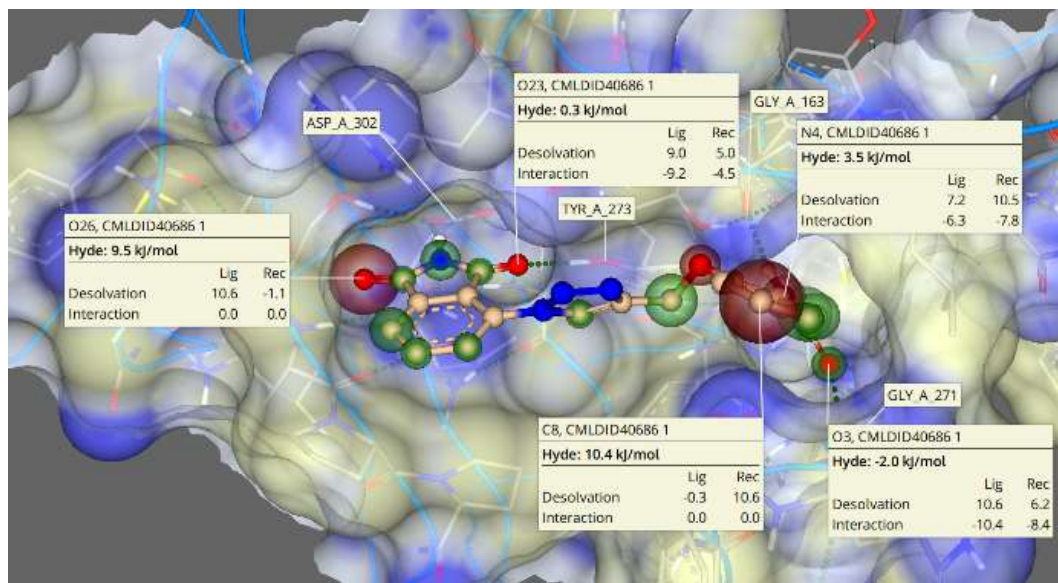
Razporeditev najboljših dvajset ligandov po vrednosti cenilnih funkcij Score_INTER (RxDock) in HYDE (SeeSAR) je prikazana na sliki 28.

A	Name	Estimated affinity				MW	SDF28
		pM	nM	μM	mM		
	CMLDID40686 1				➤	377.36	-30.797
	CMLDID44788 7				➤	354.31	-30.555
	CMLDID45558 8				➤	328.39	-30.412
	CMLDID61859 9				➤	338.33	-30.241
	CMLDID50911 10				➤	348.43	-30.060
	CMLDID64195 11				⚡	338.39	-29.955
	CMLDID51160 12				➤	286.66	-29.891
	CMLDID15241 13				⚡	299.29	-29.450
	CMLDID36466 14				⚡	297.34	-29.170
	CMLDID23100 15				➤	355.38	-29.092
	CMLDID65148 16				➤	347.33	-29.060
	CMLDID28264 17				➤	377.36	-28.921
	CMLDID46932 18				➤	292.34	-28.874
	CMLDID1846 19				➤	317.31	-28.821
	CMLDID48608 20				➤	422.47	-28.815
	CMLDID60257 21					395.40	-28.723
	CMLDID36353 22				➤	311.36	-28.664
	CMLDID36157 23				⚡	308.36	-28.558
	CMLDID50050 24				➤	302.74	-28.524
	CMLDID65570 25				⚡	341.35	-28.501

B	Name	Estimated affinity				MW	SDF28
		pM	nM	μM	mM		
	CMLDID62540 120				⚡	350.38	-24.832
	CMLDID42297 178				⚡	295.34	-24.205
	CMLDID22566 36				⚡	347.38	-27.835
	CMLDID20552 170				⚡	297.31	-24.295
	CMLDID11647 98				⚡	325.37	-25.195
	CMLDID10498 158				⚡	345.40	-24.407
	CMLDID15999 64				⚡	350.38	-26.233
	CMLDID62810 146				⚡	315.78	-24.620
	CMLDID20584 94				⚡	297.31	-25.294
	CMLDID39397 134				⚡	412.46	-24.716
	CMLDID38917 115				⚡	326.35	-24.876
	CMLDID44139 121				⚡	342.44	-24.831
	CMLDID47170 167				⚡	389.42	-24.352
	CMLDID3758 118				⚡	364.40	-24.859
	CMLDID20727 133				⚡	361.40	-24.721
	CMLDID36157 23				⚡	308.36	-28.558
	CMLDID4840 28				⚡	280.33	-28.287
	CMLDID30670 50				⚡	387.44	-27.003
	CMLDID795 195				⚡	313.38	-24.021
	CMLDID37689 63				⚡	295.36	-26.275

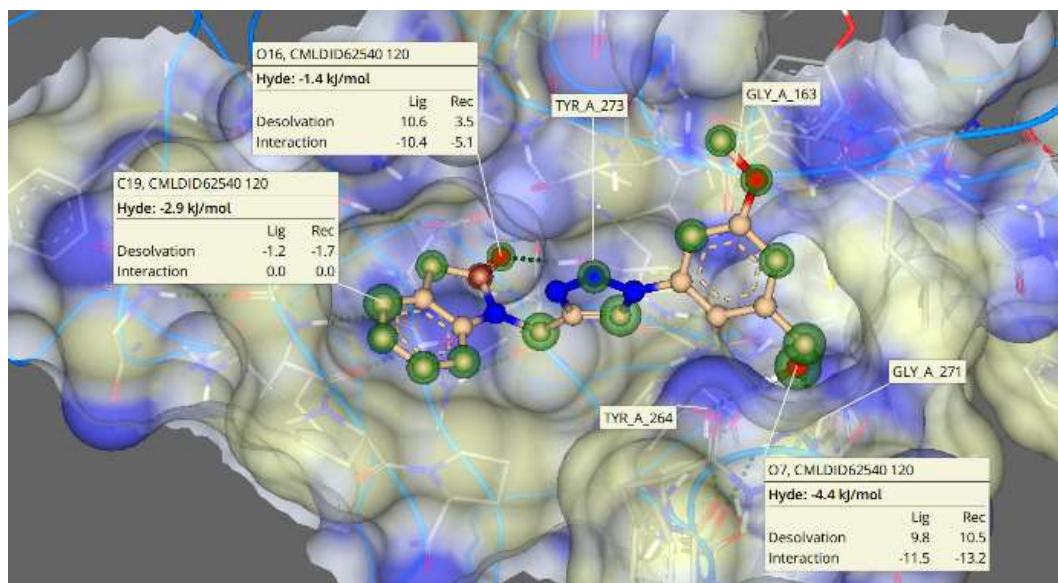
Slika 28: Razporeditev najboljših 20 ligandov iz knjižnice spojin Chemotheca po (A) vrednosti cenilne funkcije Score_INTER (RxDock), v preglednici označeno z SDF28 in (B) vrednosti cenilne funkcije HYDE (SeeSAR), v preglednici označeno z Estimated affinity. Rezultati so pridobljeni s sidranjem ligandov v vezavno mesto proteina PLpro s programom RxDock in naknadnim izračunom afinitete s programom SeeSAR.

Iz preglednic na sliki 28 lahko ugotovim, enako kot sem že ugotovila pri proteinski tarči 3CLpro, da program SeeSAR povsem drugače razvrsti ligande, kot jih je razvrstil program RxDock. Program SeeSAR uporablja cenilno funkcijo HYDE, ki v cenilno funkcijo vključi prispevek topila, program RxDock v verziji cenilne funkcije uporabljene v projektu COVID.SI pa tega prispevka ne upošteva.



Slika 29: Vezava liganda CMLDID40686 v vezavno mesto proteinske tarče PLpro.

Slika 29 prikazuje položaj liganda CMLDID40686 v vezavnem mestu PLpro. Ta ligand je razvrščen na prvo mesto glede na vrednost cenilne funkcije Score_INTER (RXDock), glede na rezultate, ki jih dobimo s programom SeeSAR, ima ta ligand afiniteto > 1 mM. Če se zanašamo na vrednost ocene afinitete dobljene s programom SeeSAR, lahko ta ligand uvrstimo med neaktivne.



Slika 30: Vezava liganda CMLDID62540 v vezavno mesto proteinske tarče PLpro.

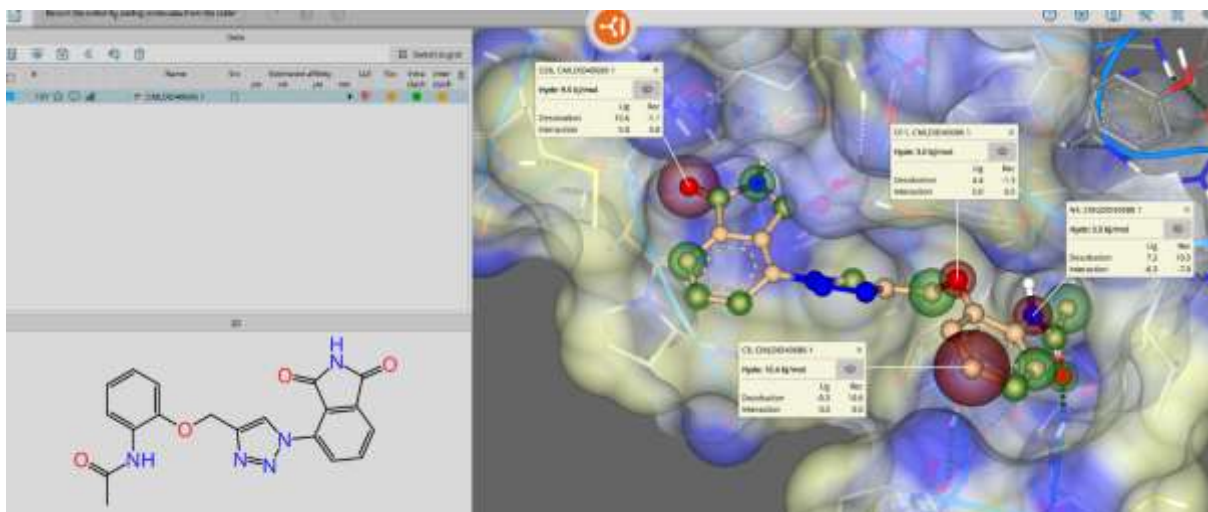
Slika 30 prikazuje vezavo liganda CMLDID62540 v vezavno mesto PLpro. Ligand je razvrščen na 120. mesto glede na vrednost cenilne funkcije Score_INTER (RxDock), glede na rezultate, ki jih dobimo s programom SeeSAR, ima ta ligand afiniteto okoli $1 \mu\text{M}$, glede na to vrednost

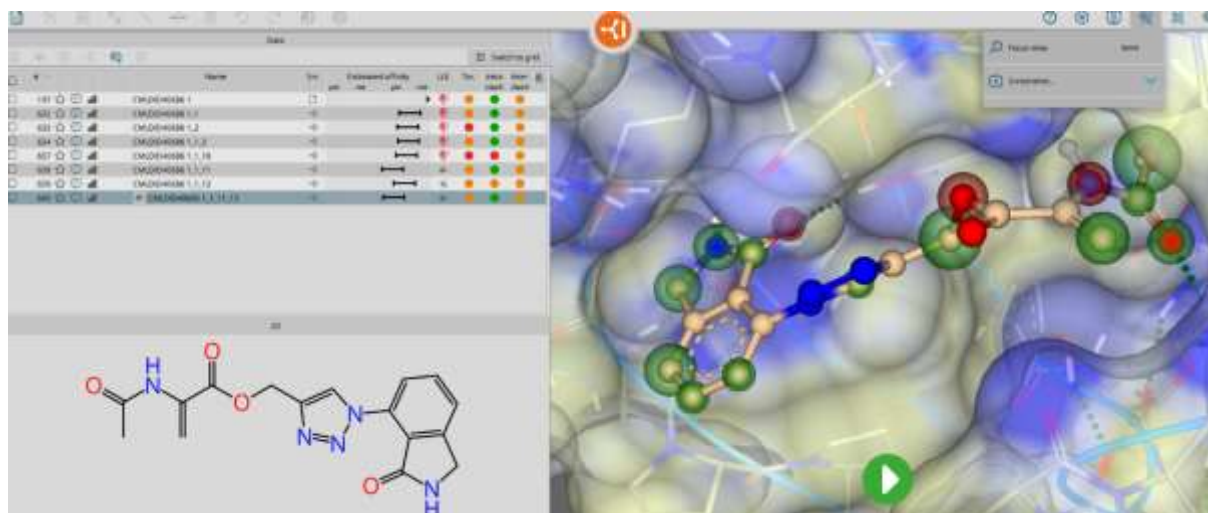
lahko ta ligand ocenimo kot potencialno uspešen zaviralec proteina PLpro virusa SARS-CoV-2.

MODIFIKACIJA LIGANDOV

Podobno kot v prejšnjem primeru (3CLpro) sem skušala z modifikacijami, ki jih omogočata orodji Editor in Inspirator izboljšati napovedane afinitete izbranih ligandov za tarčo PLpro.

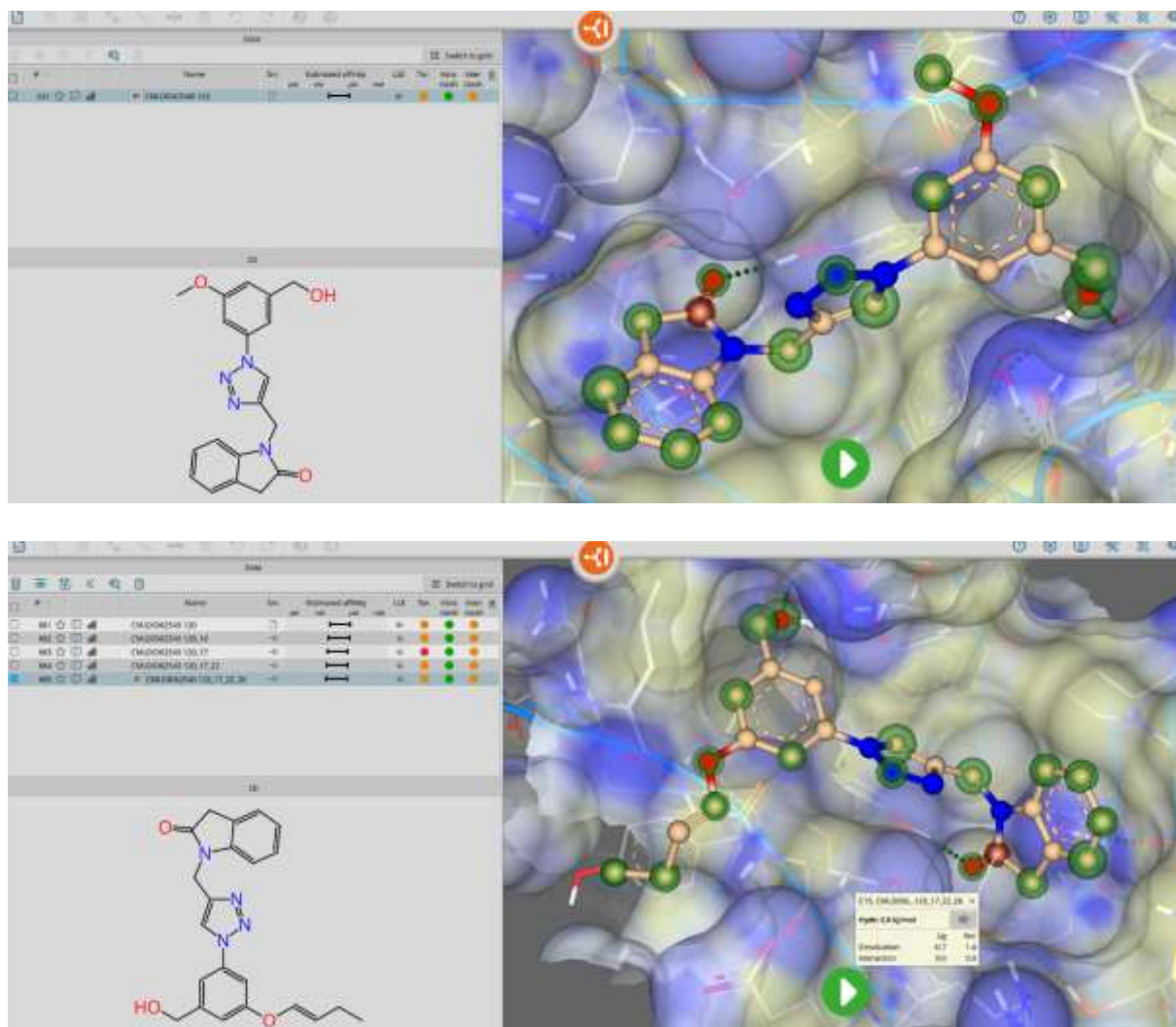
S postopnimi modifikacijami sem napovedano afiniteto za ligand, ki ga je RxDock uvrstil na prvo mesto (CMLDID40686), temeljito izboljšala. Zgornji del slike 31 prikazuje originalni ligand v vezavnem mestu PLpro, medtem, ko na spodnjem delu slike vidimo spremenjen ligand v tem istem vezavnem mestu. Iz spodnjega dela slike je razvidno, da sem se z modifikacijami znebila večine rdečih sfer in hkrati ohranila čim več zelenih sfer. Največjo izboljšavo sem dosegla z modifikacijo aromatskega sistema.





Slika 31: Modificiran ligand CMLDID40686 v vezavnem mestu proteinske tarče PLpro, najboljši ligand po RxDock-u (zgornja slika – začetno stanje, spodnja slika – stanje po modifikaciji)

Nato sem preverila kakšne so možnosti za modifikacijo liganda CMLDID40686, ki ga je najvišje uvrstil SeeSAR. Napovedano afiniteto sem uspela le rahlo izboljšati z menjavo metoksi skupine na benzenovem obroču z malo daljšim nepolarnim fragmentom, kar je razvidno iz slike 32.

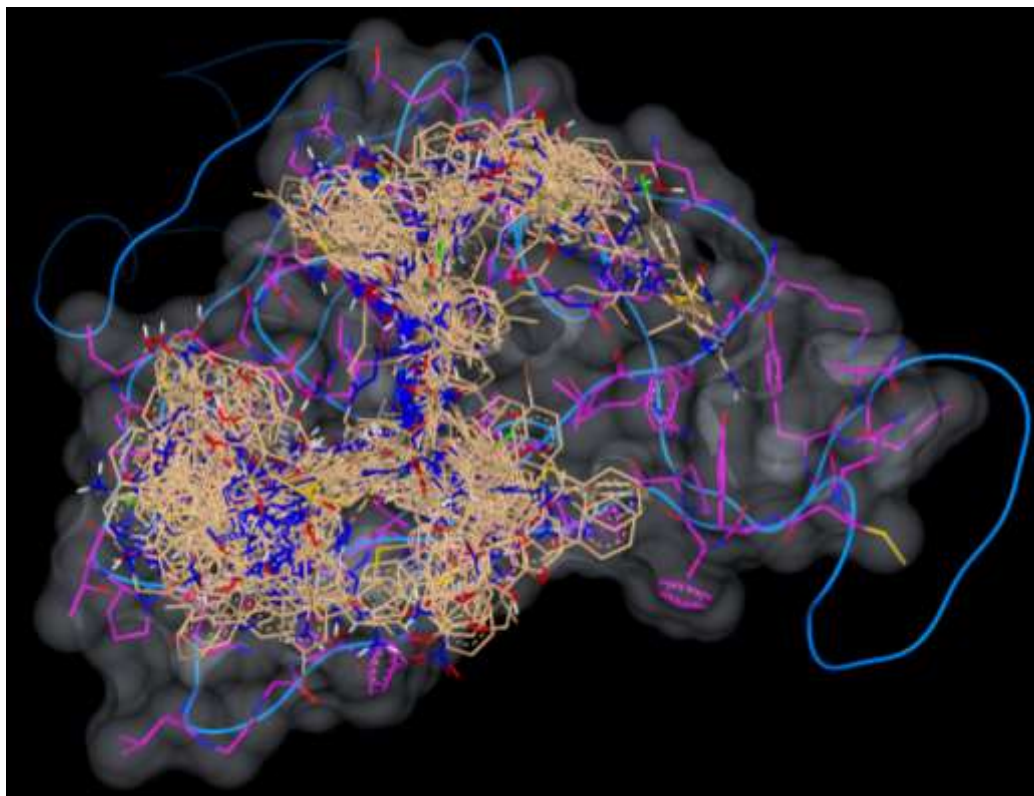


Slika 32: Modificiran ligand CMLDID62540 v vezavnem mestu proteinske tarče PLpro, najboljši ligand po SeeSAR-u (zgornja slika – začetno stanje, spodnja slika – modificirano stanje)

4.2.3 Proteinska tarča 3: RBD proteina Spike (krajše protein S)

4.2.3.1 Ligandi iz zbirke Chemotheca

Na sliki 33 je prikazan rezultat sidranja dvesto ligandov iz knjižnice Chemotheca v vezavno mesto proteina Spike virusa SARS-CoV-2. Prikazani ligandi so tisti, ki jih je program za sidranje RxDock razvrstil na prva mesta glede na vrednost cenilne funkcije Score_INTER. V primeru proteinske tarče RBD proteina Spike je vezavno mesto definiralo 50 aminokislinskih ostankov.



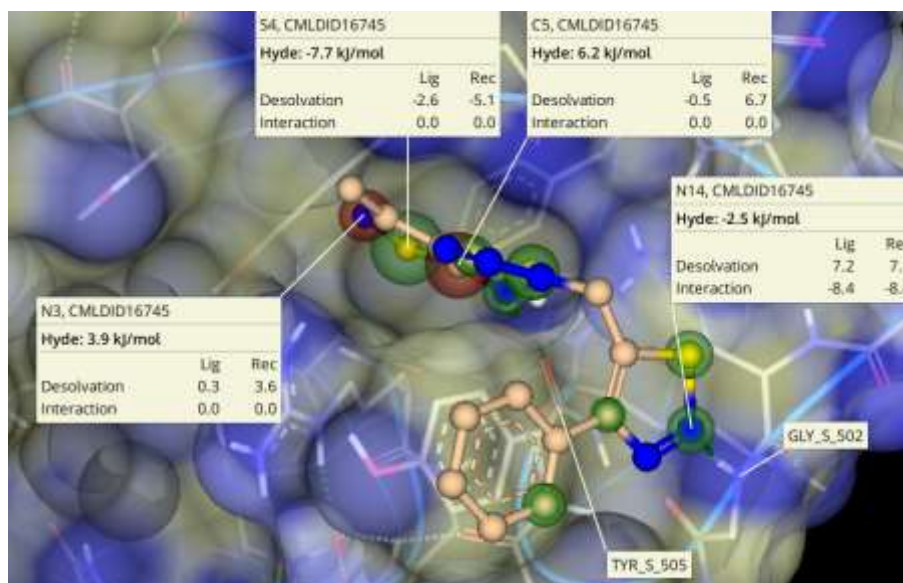
Slika 33: Rezultat sidranja dvestotih ligandov v vezavno mesto proteinske tarče RBD proteina Spike.

V preglednicah na sliki 34 sta prikazani razporeditvi ligandov glede na ocenjeno afiniteto: tj. vrednost cenilne funkcije Score_INTER (slika 34 (A)) in vrednost funkcije HYDE (slika 34 (B)) za sidranje ligandov iz zbirke ChEMBL v vezavno mesto proteina Spike virusa SARS-CoV-2. Opazimo, da programa RxDock in SeeSAR povsem drugače razvrstita ligande. Najboljši ligand po RxDock-u je precej visoko tudi po cenilni funkciji SeeSAR-a (na tretjem mestu).

Name	Estimated affinity				MW	SDF28	Name	Estimated affinity				MW	SDF28
	pM	nM	μM	mM				pM	nM	μM	mM		
CMLDID16745 1				█	355.45	-22.864	CMLDID1307 84				█	360.40	-18.403
CMLDID32079 2					402.86	-21.333	CMLDID19691 134				█	300.40	-17.902
CMLDID55131 3				▶	389.37	-21.332	CMLDID16745 1				█	355.45	-22.864
CMLDID33740 4				█	391.47	-21.015	CMLDID52572 46				█	376.42	-18.965
CMLDID48156 5					428.49	-20.951	CMLDID33720 159				█	352.42	-17.637
CMLDID49094 6				▶	345.38	-20.501	CMLDID33740 4				█	391.47	-21.015
CMLDID9012 7				▶	401.47	-20.483	CMLDID26414 185				█	343.77	-17.458
CMLDID59633 8				▶	329.33	-20.430	CMLDID31517 72				█	315.42	-18.579
CMLDID43333 9				▶	419.53	-20.384	CMLDID36305 108				█	389.46	-18.143
CMLDID57148 10					399.45	-20.358	CMLDID26316 20				█	289.36	-19.715
CMLDID12453 11					283.31	-20.291	CMLDID59693 90				█	333.35	-18.340
CMLDID31493 12				█	295.37	-20.134	CMLDID31869 69				█	315.39	-18.607
CMLDID59474 13				▶	369.45	-20.121	CMLDID34155 172				█	340.43	-17.529
CMLDID35017 14				▶	379.44	-20.091	CMLDID59617 142				█	338.37	-17.821
CMLDID16819 15				▶	403.85	-19.996	CMLDID26328 158				█	338.30	-17.649
CMLDID26560 16				▶	333.36	-19.916	CMLDID31493 12				█	295.37	-20.134
CMLDID47922 17					344.42	-19.828	CMLDID16332 110				█	399.90	-18.123
CMLDID16800 18				▶	421.89	-19.754	CMLDID35271 67				▶	338.37	-18.619
CMLDID61233 19				▶	343.36	-19.731	CMLDID60396 40				▶	381.32	-19.065
CMLDID26316 20				█	289.36	-19.715	CMLDID49330 63				▶	441.24	-18.658

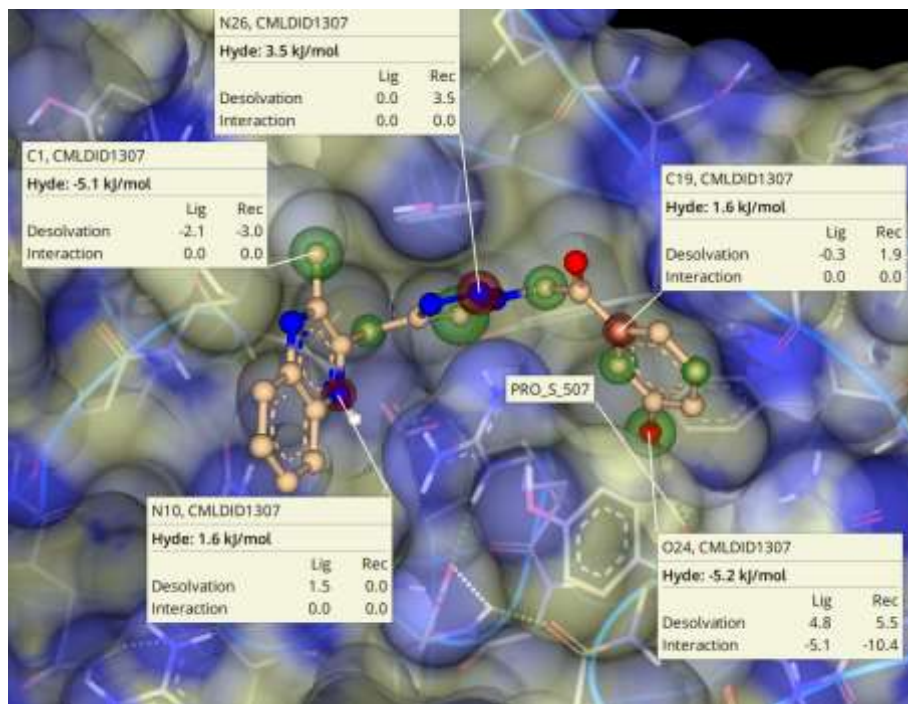
Slika 34: Razporeditev najboljših ligandov glede na vrednost cenilne funkcije Score_INTER, v preglednici označeno kot SDF28 (A) in glede na vrednost funkcije HYDE, v preglednici označeno kot Estimated affinity (B) za ligande iz zbirke Chemotheca za vezavno mesto proteina Spike virusa SARS-CoV-2.

Izmed naštetih ligandov sem s pomočjo programa SeeSAR analizirala vezavo dveh ligandov v receptorsko mesto proteina Spike. Prvi ligand je ima najnižjo vrednost cenilne funkcije Score_INTER, izračunane s programom RxDock, tj. najboljši ligand glede na RxDock (Slika 35), drugi ligand pa ima najmanjšo vrednost funkcije HYDE, torej ga je torej program SeeSAR opredelil kot najboljši ligand (Slika 36).



Slika 35: Interakcija liganda CMLDID16745 (najboljši ligand pri sidranju s programom RxDock) z receptorsko domeno virusnega proteina Spike.

Iz slike 35 je razvidno, da ligand zaseda vezavno mesto proteina, med ligandom in proteinom sta vzpostavljeni dve vodikovi vezi in sicer z aminokislinskima ostankoma GLY 502 in TYR 505. Z vizualnim pregledom atomskih prispevkov k vezavi lahko ugotovim, da imata dva atoma N3 in C5 v heterocikličnem obroču vezanega liganda neugoden prispevek k celotni vezavni afiniteti (rdeči sferi). Podrobna analiza pokaže, da je neugoden prispevek posledica desolvatacije dela receptorskega mesta v soseščini obeh atomov. Menjava teh dveh atomov ali celo menjava celotnega heterocikla bi lahko vodila k izboljšavi afinitete.

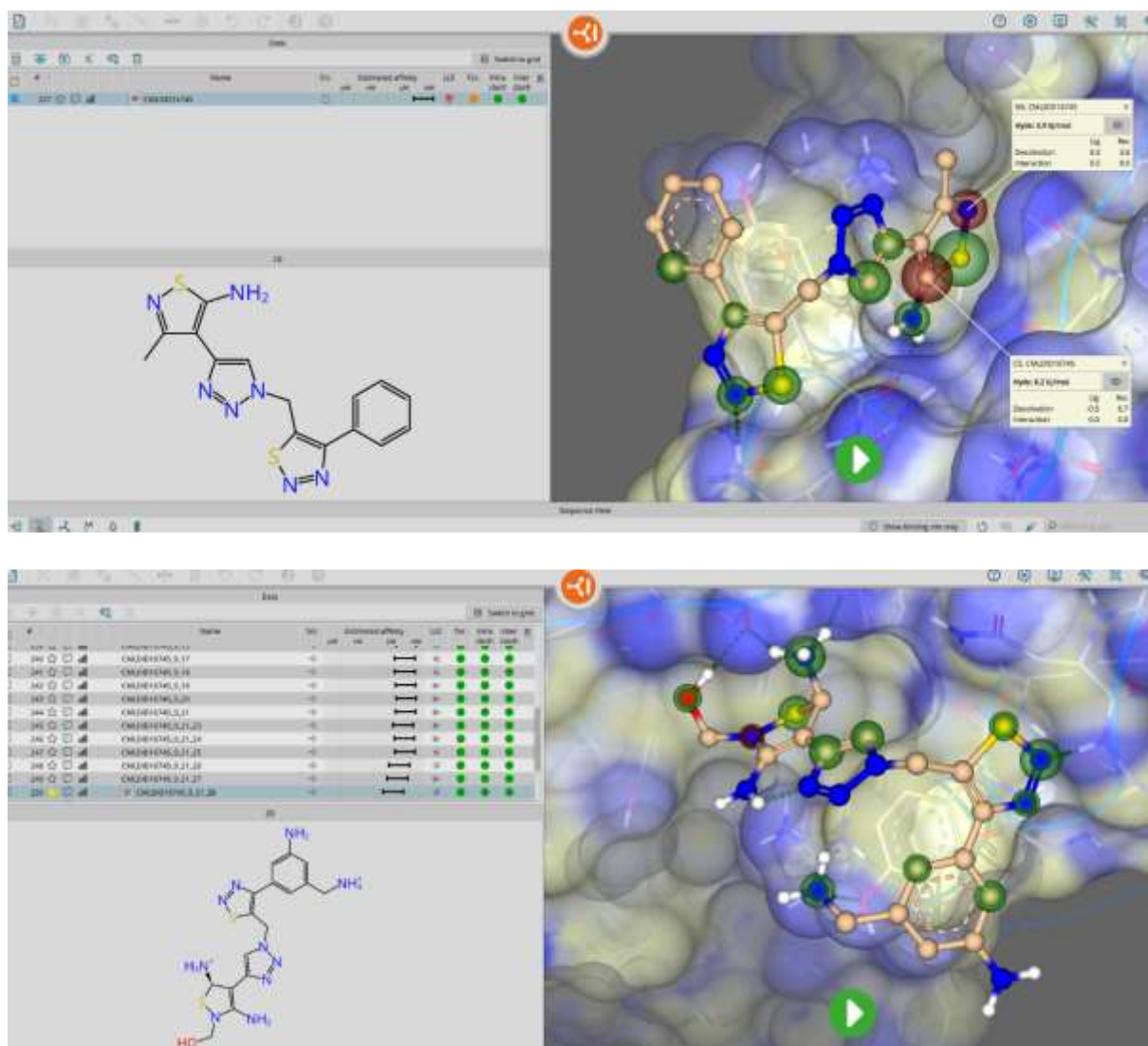


Slika 36: Interakcija liganda CMLDID1307 (najboljši ligand dobljen s programom SeeSAR) z receptorsko domeno virusnega proteina Spike.

Na sliki 36 lahko vidimo, da je ligand CMLDID1307 je lepo sidran v receptorsko domeno proteina Spike. Ligand tvori s proteinom eno vodikovo vez in sicer z aminokislinskim ostankom PRO 507. Z vizualnim pregledom sem ugotovila, da trije atomi (N10, C19 in N26) liganda neugodno vplivajo na celotno afiniteto, vendar so ti prispevki bistveno nižji kot pri ligandu CMLDID16745.

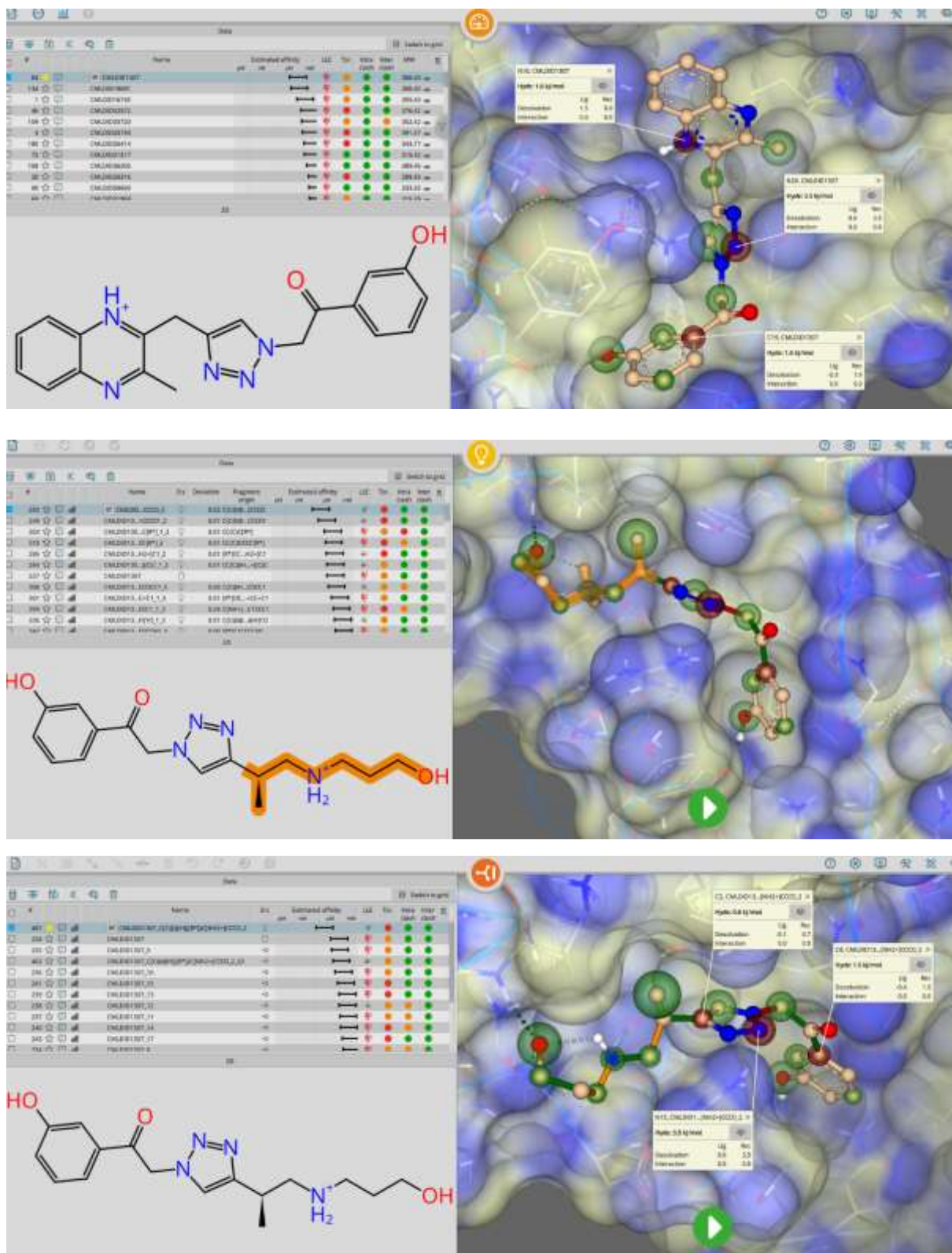
MODIFIKACIJA LIGANDOV

Spremenila sem molekuli, ki sta se izkazali za najboljša liganda vezana na RBD proteina Spike, tj. najboljšega po sidranju z RxDock-om in tistega z najboljšo ocenjeno vezavno afiniteto po SeeSAR-u. Rezultati so prikazani na slikah 37 in 38.



Slika 37: Modificiran ligand CMLDID16745 v vezavnem mestu proteina Spike, najboljši ligand po RxDock-u (zgornja slika – začetno stanje, spodnja slika – spremenjen ligand)

S postopno modifikacijo sem uspela ligand CMLDID16745, ki ga je RxDock razvrstil na prvo mesto izboljšati z dva velikostna razreda (glej sliko 37). Izboljšala pa sem tudi parameter LLE (Ligand Lipophilicity Efficiency), merilo, ki poleg same afinitete vključuje tudi oralno dostopnost spojin. Ligand sem izboljšala z vezavo dodatnih funkcionalnih skupin na sam skelet, le te tvorijo dodatne vodikove vezi z receptorjem in hkrati izboljšajo topnost liganda.



Slika 38: Modificiran ligand CMLDID1307 v vezavnem mestu proteina Spike, najboljši ligand po SeeSAR-u (zgornja slika – začetno stanje, srednja slika – zamenjan fragment, spodnja slika – modificirana molekula liganda)

Nazadnje sem skušala modificirati ligand, ki ga je SeeSAR razvrstil najvišje (glej sliko 38). Za modifikacijo liganda sem tokrat najprej uporabila orodje Inspirator, s katerim menjamo del molekule s fragmenti iz interne podatkovne zbirke (srednji del slike), nato pa poskusila z

rahlimi spremembami molekule še v orodju Editor (spodnji del slike). V tem primeru znatnega izboljšanja liganda nisem uspela doseči.

4.2.3.2 Ligandi iz zbirke Peptidi

V svoji raziskovalni nalogi sem analizirala tudi rezultate sidranja manjših peptidov, iz zbirke spojin projekta COVID.SI, v receptorsko domeno proteina S. Analiza rezultatov je prikazana na slikah 39 in 40.

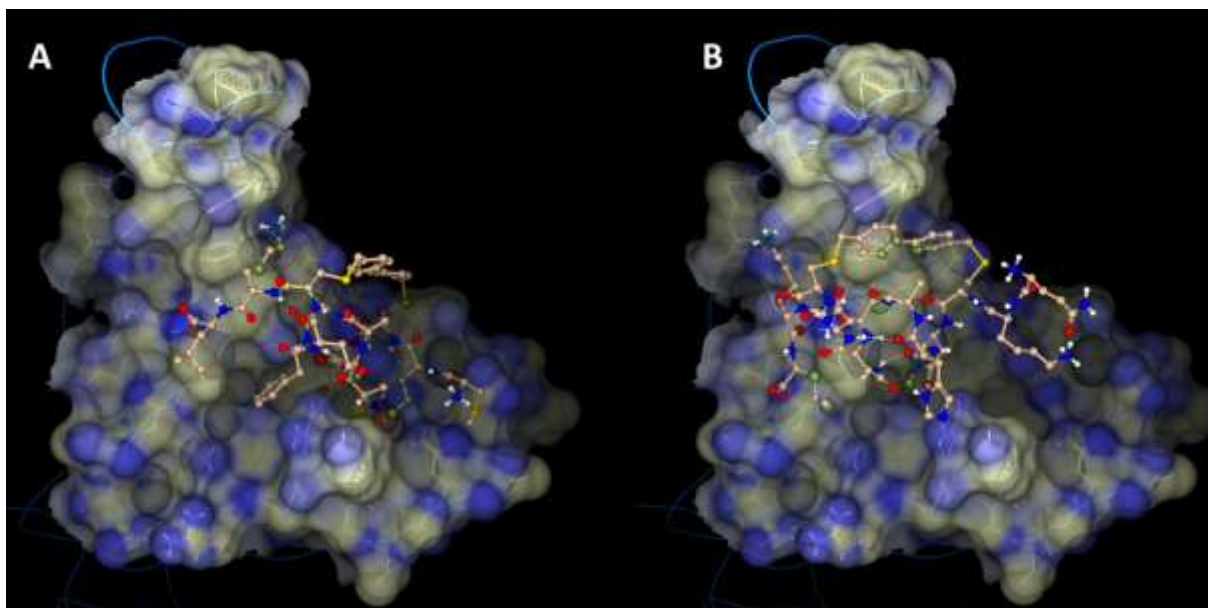
Na sliki 39 sta prikazani razporeditvi ligandov glede na vrednost cenilne funkcije Score_INTER (slika 39(A)) in glede na vrednost funkcije HYDE (slika 39 (B)) za sidranje ligandov iz knjižnice krajših peptidov v vezavno mesto proteina Spike virusa SARS-CoV-2.

Name	Estimated affinity		MW	SDF28	Name	Estimated affinity		MW	SDF28
	µM	nM				µM	nM		
peptide12_0829 1			1495.80	-7.826	peptide12_0723 122			1546.83	11.840
peptide12_0827 2			1640.94	-6.138	peptide12_0942 155			1595.88	13.180
peptide12_0828 3			1555.83	-4.346	peptide12_0175 39			1580.91	5.895
peptide12_0537 4			1569.95	-3.897	peptide12_0206 85			1532.81	9.274
peptide12_0172 5			1563.81	-3.662	peptide12_0559 129			1559.95	12.166
peptide12_0996 6			1523.72	-2.850	peptide12_0832 60			1471.80	7.909
peptide12_0884 7			1551.85	-2.359	peptide12_0215 84			1467.75	9.155
peptide12_0135 8			1485.72	-0.324	peptide12_0127 95			1546.76	9.972
peptide12_0195 9			1521.90	-0.152	peptide12_0825 161			1606.92	13.348
peptide12_0087 10			1530.87	0.175	peptide12_0695 42			1504.84	6.305
peptide12_0085 11			1472.76	0.178	peptide12_0147 94			1629.87	9.852
peptide12_0910 12			1606.90	0.246	peptide12_0869 114			1472.81	11.287
peptide12_0824 13			1556.90	0.559	peptide12_0872 72			1526.82	8.569
peptide12_0180 14			1481.79	0.750	peptide12_0174 18			1566.90	2.008
peptide12_0820 15			1564.91	1.367	peptide12_0151 135			1609.93	12.387
peptide12_0021 16			1402.71	1.382	peptide12_0213 183			1466.83	15.110
peptide12_0200 17			1555.82	1.837	peptide12_0210 192			1505.84	15.930
peptide12_0174 18			1566.90	2.008	peptide12_0692 66			1511.78	8.063
peptide12_0694 19			1533.81	2.306	peptide12_0844 117			1484.90	11.600
peptide12_0719 20			1585.90	2.686	peptide12_0910 12			1606.90	0.246

Slika 39: Razporeditev ligandov glede na vrednost cenilne funkcije Score_INTER, v preglednici označeni z SDF28 (A) in vrednost funkcije HYDE, v preglednici označeni z Estimated affinity (B) za ligande iz knjižnice peptidov vezane v vezavno mesto proteina Spike virusa SARS-CoV-2.

Iz slike 39 je razvidno, da je glede na izračune programa RxDock najboljši ligand peptide12_0829, glede na izračune programa SeeSAR pa peptide12_0723.

Slika 40 prikazuje vezavo teh dveh ligandov v receptorsko mesto proteina S. Oba peptida pokrivata lep del receptorskega mesta. Ocenjeni vrednosti afinitete teh peptidov sta v območju od 10 do 1000 µM. Na tem mestu lahko poudarim, da je bilo sidrano le 1000 izmed možnih 20¹⁰ različnih dodekapeptidov. Z večjim naborom peptidov in/ali modifikacijo obstoječih bi morda lahko našli peptid s precej boljšo afiniteto, kot je navedena za peptide iz izbrane zbirke peptidov (glej sliko 39 (B)).



Slika 40: Vezava dveh peptidov: (A) peptide12_0829 in (B) peptide12_0723 v receptorsko domeno proteina Spike virusa SARS-CoV-2.

5 RAZPRAVA

5.1 Analiza rezultatov sidranja z RxDock

Prvo hipotezo “Molekularno sidranje z RxDockom lahko uporabimo za razvrščanje ligandov po vezavni afiniteti” lahko zavrnem, saj cenilna funkcija, ki jo ima RxDock, ni dovolj natančna za pravilno razvrščanje ligandov po afiniteti. Lahko pa RxDock uporabimo za prvotno filtriranje ligandov. Najverjetnejše je za razvrščanje po vezavni afiniteti bolj primeren program SeeSAR.

Namen sidranja je, da iz ogromnih knjižnic spojin izberemo nekaj spojin za nadaljnje analize in morebitna nadaljnja biološka testiranja. S pomočjo analize histogramov in grafov prikazanih v poglavju rezultati sem poskusila postaviti kriterije za izbor spojin. Načeloma bi bil prvi kriterij vrednost celotne cenilne funkcije Score, vendar je ta vrednost močno odvisna od ustrezne konformacije vhodnih spojin. Score_INTER je manj občutljiv na kvaliteto vhodnih podatkov, zato sem 200 ligandov za nadaljnjo obravnavo izbrala s pomočjo rangiranja po vrednostih Score_INTER. Mogoča bi bila izbira ligandov z rangiranjem po vrednosti Score po izločitvi ligandov, ki imajo vrednost Score_INTRA med -5 in 5.

Ob pravih nastavitvah programa in ligandih z manjšim številom vrtljivih vezi (do 5) RxDock pravilno umesti ligand v vezavno mesto receptorja. Uporaba RxDocka takrat reproducira

eksperimentalno (kristalografsko) strukturo kompleksa med receptorjem in ligandom. Analiza dobljenih kompleksov nam pomaga razumeti način vezave liganda v receptorsko mesto, na primer predvidimo lahko, kateri atomi liganda se povezujejo z vodikovimi vezmi z aminokislinskimi ostanki receptorja.

Poleg ligandov iz Chemothece sem na tarčo proteina S virusa SARS-CoV-2 sidrala tudi 1000 naključno generiranih dodekapeptidov. Izkazalo se je, da sta vrednosti cenilnih funkcij Score in Score_INTER negativni, cenilna funkcija Score_INTRA pa skoraj v vseh primerih pozitivna. Verjetno se na te rezultate ne moremo pretirano zanesti, saj so peptidni ligandi preveč fleksibilni, da bi se pri naših nastavitvah pravilno sidrali.

5.2 Obdelava najboljših ligandov glede na RxDock s programom SeeSAR

Za dodatno analizo rezultatov molekulskega sidranja sem uporabila program SeeSAR, ki omogoča izračun afinitete z uporabno cenilne funkcije HYDE, ki za razliko od cenilne funkcije programa RxDock, upošteva tudi solvacijske prispevke.

Drugo hipotezo “Molekulsko sidranje je dobro orodje za analizo interakcij med ligandom in receptorjem” lahko potrdim. Ugotovila sem, da ima SeeSAR zelo dobra orodja za analizo in vizualizacijo omenjenih kompleksov, ki pomagajo pri razumevanju procesa vezave.

Tretjo hipotezo “Vezavno afiniteto ligandov z receptorji lahko z manjšimi spremembami strukture ligandov izboljšamo” lahko delno potrdim. Program SeeSAR ima orodji Editor in Inspirator, s katerima lahko molekulo v vezavnem mestu spreminjamo na tak način, da se napovedana afiniteta izboljša. Postopek takšne modifikacije molekule je običajno večstopenjski in zahteva ogromno inspiracije, ne vodi pa vedno do izboljšave. Z veliko potrpežljivosti mi je uspelo na ta način vsaj malo izboljšati afiniteto vsem izbranim ligandom vezanim v tarče 3CLpro, PLpro in protein S. Še najbolj uspešna sem bila pri modifikacijah liganda CMLDID57806 v vezavnem mestu 3CLpro, kjer sem neaktivni ligand z modifikacijami spremenila tako, da je bila njegova končna napovedana afiniteta v nanomolarnem območju.

6 ZAKLJUČEK

Epidemija bolezni covid-19 je v zadnjem močno zaznamovala naša življenja, zato smo se skupaj z mentorjema odločili, da bi izdelala raziskovalno nalogo na temo omenjene bolezni. Odločila sem se raziskati, kako bi računalniške metode (oz. programe) uporabila pri načrtovanju molekul, ki interagirajo s tremi proteini, ki imajo ključno vlogo v delovanju virusa SARS-CoV-2. Pri izdelavi svoje raziskovalne naloge sem obdelovala rezultate molekulskega sidranja ligandov iz knjižnice spojin Chemotheca na tri izbrane proteinske tarče (dve proteazi 3CLpro in PLpro ter receptorsko domeno proteina S), ki so nastali v okviru projekta odprte znanosti COVID.SI. Sodelavci projekta so me seznanili s tem zanimivim projektom in mi razložili, kako so prišli do rezultatov s pomočjo porazdelitvenega računalništva.

Pri sami raziskavi sem se naučila veliko novega. Spoznala sem metodo molekulskega sidranja, s katero skušamo prilagoditi ligand vezavnemu mestu. Spoznala sem program za molekulsko sidranje RxDock, sama sidrala molekul sicer nisem, sem pa se temeljito pozanimala, kako program deluje, saj je bilo to potrebno za interpretacijo rezultatov. Ugotovila sem, da interpretacija podatkov, ki jih dobimo z molekulskim sidranjem, ni enostavna in je potrebno biti zelo previden pri interpretaciji rezultatov. Podatke sidranja sem analizirala s Pythonovo skripto PyDockAnalyzer, ki jo je napisal zunanji mentor. Tekom mojega dela sem naletela na nekaj zanimivih primerov, ki sem jih argumentirano opisala v raziskovalni nalogi. Zanimiva je bila ugotovitev, da sidranje spojin z zelo čudnimi konformacijami lahko privede do zelo "dobrih" rezultatov molekulskega sidranja (zelo nizka vrednost celotne cenilne funkcije), a ti rezultati so lažni..

Nadalje sem spoznala in se naučila uporabljati program SeeSAR podjetja BiosolveIT. Omenjeni program je univerzalno orodje, ki se uporablja pri računalniškem načrtovanju farmacevtsko zanimivih spojin. Program SeeSAR sem uporabila za izračun napovedane afinitete in lastnosti ADMET. SeeSAR sem uporabila tudi za prikaz in vizualno analizo posameznih kompleksov med ligandom in proteinsko tarčo. Uporabila pa sem tudi njegovi orodji Editor in Inspirator, s katerima sem izbrane ligande v vezavnem mestu spreminjala z željo izboljšati njihovo afiniteto do izbrane tarče. S spremembami dobimo nove molekule ligandov, ki pa verjetno še niso bile sintetizirane. Če so spremembe manjše, je sinteza najbrž mogoča, če pa so velike, je najbrž potrebna sinteza od začetka in ne iz ligandov, ki so že poznani.

Molekulsko sidranje nam omogoča zožiti izbor spojin, ki bi bile primerne za nadaljnji razvoj. V idealnem scenariju bi za izbrane spojine naredili *in vitro* in/ali *in vivo* teste, s katerimi bi dobili povratno informacijo o tem, katere izmed molekul so res aktivne. Za molekule, ki se izkažejo aktivne, je potrebno narediti še toksikološke analize. Torej smo z rezultati raziskovalne naloge šele na začetku trnove in zahtevne poti, ki vodi do aktivne zdravilne učinkovine.

7 ZAHVALE

Zahvaljujem se svojim mentorjema, da sta mi bila kljub trenutni situaciji pripravljena pomagati pri izdelavi raziskovalne naloge. Mag. Mojca Podlipnik mi je pomagala pri razumevanju snovi, o kateri še nisem imela toliko znanja in pri urejanju raziskovalne naloge. Dr. Črtomir Podlipnik pa mi je pomagal pri razumevanju in delu z zame novimi programi.

8 VIRI IN LITERATURA

VIRI

https://www.researchgate.net/publication/348293095_Existing_antiviral_options_against_SARS-CoV-2_replication_in_COVID-19_patients (16. 01. 2021; 15.35)

<https://covid.si> (14. 01. 2021; 17.50)

<https://www.rxdock.org/> (25. 01. 2021; 20.30)

http://openbabel.org/wiki/Main_Page (17. 01. 2021; 10.20)

<https://zinc15.docking.org/> (26. 01. 2021; 19.15)

<http://chemotheca.unicz.it/> (20. 01. 2021; 15.40)

https://en.wikipedia.org/wiki/Lipinski%27s_rule_of_five (10. 02. 2021; 16.45)

https://en.wikipedia.org/wiki/Partition_coefficient (01. 02. 2021; 21.50)

<https://youtu.be/zvuYJTL90J8> (09. 01. 2021; 10.05)

https://www.youtube.com/watch?v=OW7UP_JlIO8 (09. 01. 2021; 10.25)

<https://www.youtube.com/user/armandohasudungan> (09. 01. 2021; 10.45)

https://www.youtube.com/watch?v=glTw0a_eyl4 (09. 01. 2021; 13.00)

<https://www.nature.com/articles/s41579-020-00468-6> (09. 01. 2021; 13.25)

https://www.youtube.com/watch?v=DD_wdvwN9rg (09. 01. 2021; 13.45)

ŽUNEC, Špela. ANALIZA REZULTATOV MOLEKULSKEGA SIDRANJA KNJIŽNICE SPOJIN NA PROTEINSKE TARČE VIRUSA SARS-COV-2. Raziskovalna naloga. Ljubljana, Gimnazija Jožeta Plečnika Ljubljana, 2021

<https://www.youtube.com/watch?v=5DGwOJXSxqg>

Where Do New Viruses Come From? - YouTube (10. 01. 2021; 10.15)