

UNIVERZA V LJUBLJANI
FAKULTETA ZA FARMACIJO

MARKO PUCKO
MAGISTRSKA NALOGA
ENOVITI MAGISTRSKI ŠTUDIJSKI PROGRAM FARMACIJA

Ljubljana, 2018

UNIVERZA V LJUBLJANI
FAKULTETA ZA FARMACIJO

Univerza *v Ljubljani*
Fakulteta *za farmacijo*



Marko Pucko

**OPTIMIZACIJA NAČRTOVANJA KONFORMACIJSKIH
KNJIŽNIC ZA UPORABO V VIRTUALNEM REŠETANJU**

**CONFORMATIONAL LIBRARY DESIGN OPTIMIZATION
IN VIRTUAL SCREENING**

**MAGISTRSKA NALOGA
ENOVITI MAGISTRSKI ŠTUDIJSKI PROGRAM FARMACIJA**

Ljubljana, 2018

Magistrsko nalogo sem opravljal na Fakulteti za farmacijo, na katedri za farmacevtsko kemijo pod mentorstvom prof. dr. Stanislava Gobca, mag. farm. in somentorstvom doc. dr. Marka Jukiča, mag. farm.

Zahvala

Zahvalil bi se mami, atiju in sestri za vso podporo v teh letih študija. Upam, da se boste zdaj lahko malo bolj sprostiti in poskrbeli za svoje zdravje ter dobro počutje. Pri tem vam bom pomagal kolikor se bo le dalo.

Hvala Ana za vse lepe trenutke in tudi za pomoč ter motivacijo v težkih časih. Ni bilo vedno vse enostavno, je pa vsaj bilo zanimivo ☺. Hvala, da prinašaš sonce v moje dni.

Hvala vsem prijateljem, da ste me prenašali toliko let in vmes niste pozabili name. Ne vem točno kako smo pristali skupaj, ampak sem zelo vesel da smo ☺. Zakon ste!

Hvala vsem članom komisije, še posebej pa tebi Marko! Hvala za priložnost, za pomoč, za spodbudo in za zaupanje.

Brez vseh vas mi to ne bi nikoli uspelo. Hvala, da ste v mojem življenju. Rad vas mam!

Zdaj pa gremo naprej! Akcija!

Izjava

Izjavljam, da sem magistrsko nalogo izdelal samostojno pod mentorstvom prof. dr. Stanislava Gobca, mag. farm. in somentorstvom doc. dr. Marka Jukiča, mag. farm.

Marko Pucko

Ljubljana, 2018

Komisija za zagovor:

Predsednica komisije: prof. dr. Marija Bogataj, mag. farm.

Mentor: prof. dr. Stanislav Gobec, mag. farm.

Somentor: doc. dr. Marko Jukič, mag. farm.

Članica komisije: asist. dr. Nika Lovšin, univ. dipl. kem.

KAZALO VSEBINE

KAZALO VSEBINE.....	II
POVZETEK	IV
ABSTRACT	V
SEZNAM OKRAJŠAV	VI
1 UVOD.....	1
1.1 REŠETANJE VISOKIH ZMOGLJIVOSTI (HTS).....	1
1.2 VIRTUALNO REŠETANJE (VS).....	1
1.2.1 VRSTE VIRTUALNEGA REŠETANJA	2
1.2.2 CENILNE FUNKCIJE	5
1.2.3 KONFORMACIJE	6
1.2.4 GENERATOR KONFORMACIJ – OMEGA2.....	8
1.2.5 FRED.....	9
1.3 OCENA SIDRANJA	10
1.3.1 DUD-E.....	11
1.4 FORMAT VHODNIH PODATKOV	11
1.5 ROC KRIVULJA (RECEIVER OPERATING CHARACTERISTIC CURVE ANALYSIS).....	12
1.6 KNIME	13
1.7 UPORABLJENE TARČE	13
2 NAMEN DELA	15
3 MATERIALI IN METODE	16
4 EKSPERIMENTALNO DELO	18
4.1 PRIDOBITEV TARČ.....	18
4.2 PRIPRAVA TARČ IN LIGANDOV	20
4.3 TVORBA KONFORMACIJ Z OMEGO	23
4.4 SIDRANJE V FRED-u.....	25
4.5 OBDELAVA PODATKOV V KNIME	26
4.5.1 PRIMER IZDELAVE ROC KRIVULJE V KNIME	26
4.5.2 DESKRIPTORJI V KNIME.....	27
5 REZULTATI in RAZPRAVA	29
5.1 ACES.....	29
5.2 ANDR.....	31
5.3 CAH2.....	33
5.4 FA10.....	35

5.5	FNTA.....	37
5.6	HIVPR	39
5.7	MK14.....	41
5.8	PGH2	43
5.9	PDE5A.....	45
5.10	TRY1	47
5.11	VGRF2	49
5.12	PREGLED SPREMEMB AUC.....	51
6	SKLEP.....	56
7	LITERATURA.....	57

POVZETEK

Rešetanje visokih zmogljivosti je v današnjem času še vedno najpogostejša metoda rešetanja kemijskih knjižnic. Gre za zelo kompleksen in dolgotrajen postopek zato se uporablja komplementarne metode kot je virtualno rešetanje. Virtualno rešetanje se lahko izvaja tudi kot samostojen proces, kjer gre za kombinacijo različnih računalniških pristopov in orodij. Zelo je pomembno, da poznamo delovanje teh orodij do potankosti in parametre, ki vplivajo na rezultate naših eksperimentov. V okviru magistrskega dela smo preverjali vpliv parametra števila izračunanih konformacij na rezultat sidranja.

Sidranje spada med vrste rešetanj, ki temeljijo na osnovi tarče. Iz baze podatkov DUD-E smo pridobili strukturne informacije o terapevtskih tarčah s kristaliziranimi ali kokristaliziranimi ligandi ter pripadajočimi knjižnicami aktivnih in neaktivnih spojin. S programskim paketom OEDocking smo ustrezno pripravili tarče in definirali vezavna mesta za izvajanje eksperimentov sidranja. S programom Omega, smo izvedli pripravo 3D virtualnih knjižnic spojin, to je generiranje različnega števila konformacij: 1, 25, 250 in 2500. Nato smo s programom Fred iz računalniškega paketa OEDocking izvedli postopek sidranja pripravljenih 3D virtualnih knjižnic v izbrane terapevtske tarče. Eksperiment sidranja smo za vsako tarčo ponovili trikrat in rezultate zbrali v programu KNIME s pomočjo ROC krivulj. Tarče z najmanjšimi vezavnimi mesti glede na prostornino in najbolj rigidnimi kristaliziranimi ligandi so se odrezale kot konformacijsko neodvisne. Imele so zelo majhne vrednosti površine pod krivuljo (angl. Area under the curve - AUC) in pa zelo majhne prirastke ob prehodu na večje število konformacij. Prav tako se je izkazala pomembna primerjava podobnosti po deskriptorjih med aktivnimi in neaktivnimi spojinam ter med uporabljeno knjižnico spojin in kristaliziranim ligandom. Boljše rezultate in konformacijsko odvisnost pa smo pokazali v primeru večjih vezavnih mest. Ugotovili smo, da je pri večini primerov prišlo do prirasta AUC pri uporabi 25 konformacij in ta velikostni rang smo določili za optimalen ob upoštevanju rezultatov in porabe časa, ter računske moči.

ABSTRACT

High-throughput screening is still the most used method for screening chemical libraries. It's a very complex and long procedure and for that reason we use complementary methods like virtual screening. Virtual screening can be used as a standalone method which consists of several computational approaches and software tools. It's very important that we understand these tools and all the parameters that affect the outcome of our experiments. As part of this master's thesis we observed the influence of the number of conformations on the outcome of virtual screening.

Docking is a structure based drug design type of screening. We used the DUD-E database to obtain structural data on therapeutic targets and their ligands along with a database with corresponding active and inactive compounds. With the help of OEDocking software we prepared the targets and defined binding regions for our experimental docking. We prepared 3D virtual libraries that contain a different amount of conformations per compound: 1, 25, 250, 2500. Then we executed the docking with Fred (OEDocking). We docked the compounds into corresponding targets 3 times and each time we used randomly selected compounds. Targets with the smallest volume of binding regions and the most rigid crystalized ligands were the worst performers. They had the lowest area under the curve (AUC) and also the AUC gain when the conformation number was increased was minimal. The comparison of descriptors between active and inactive compounds and also between the library and crystalized ligands was very important. With bigger binding regions we got better results and a bigger conformational independence. We observed an AUC increase with most targets at 25 conformations so we selected this size as optimal considering computational time.

SEZNAM OKRAJŠAV

2D	dvodimenzionalen
3D	tridimenzionalen
ADMET	absorpcija, distribucija, metabolizem, izločanje in potencialna toksičnost
AUC	površina pod ROC krivuljo (angl. Area under the curve)
ChEMBL	baza podatkov o bioaktivni molekulah, ki imajo zdravilom podobne lastnosti
DUD-E	»Database of useful decoys: enhanced«; baza podatkov o receptorjih, aktivni spojinah in spojinah vabah
g/mol	gram na mol
HTS	angl. high-throughput screening; rešetje visokih zmogljivosti
LogP	logaritemska vrednost porazdelitvenega koeficienta
MM	molekulska masa
PDB	angl. Protein Data Bank; podatkovna baza, ki vsebuje informacije o strukturah proteinov, nukleinskih kislin in kompleksov ligand-tarča
RMSD	angl. Root-mean-square deviation; kvadratni koren povprečja kvadrata standardnih odklonov
ROC krivulja	angl. receiver operating characteristic; je metoda ocenjevanja ali primerjave sposobnosti testov, kjer imamo na voljo dve vrednosti oz. populaciji
SMILES	angl. simplified molecular-input line-entry system; gre za enovrstični način zapisa molekul, ki vsebuje informacije o njihovi strukturi
VS	virtualno rešetje (angl. virtual screening)
ZINC	baza podatkov o komercialno dostopnih spojinah

1 UVOD

1.1 REŠETANJE VISOKIH ZMOGLJIVOSTI (HTS)

Dandanes se za odkrivanje novih zdravilnih učinkovin večinoma še vedno uporablja metodo rešetanja visokih zmogljivosti (HTS), ki je tudi najpogostejši vir zadetkov. Gre za avtomatizirani postopek, kjer z ustreznimi metodami preverjajo potencialno aktivnost velike količine spojin na dan. Ob pridobitvi spojin zadetkov se le-te vodi skozi proces optimizacije do spojin vodnic. Tem vodnicam nato spet z modifikacijo strukture izboljšajo lastnosti, ki vplivajo na ADMET: absorpcija, distribucija, metabolizem, izločanje in potencialna toksičnost. To so ključni koraki, ki surove potencialne spojine pretvorijo v vodnice in na koncu v nove zdravilne učinkovine. Ta proces je dolgotrajen in drag saj trenutno najboljše HTS metode dovoljujejo rešetanje 10^6 spojin na dan ampak le majhen odstoten zadetkov napreduje v nadaljnje raziskave, kar predstavlja veliko ekonomsko breme. Zato tudi ni presenečenje da je bila potreba po alternativni ali vsaj komplementarni metodi velika. Pomembna je učinkovita identifikacija zadetkov v zgodnjih stopnjah razvoja za optimalno izrabo virov in časa. To nam lahko omogoči virtualno rešetanje.⁽¹⁾

1.2 VIRTUALNO REŠETANJE (VS)

Z razvojem računalništva je pridobilo na veljavi tudi tehnologija virtualnega rešetanja (VS), ki igra pomembno vlogo pri iskanju novih potencialnih spojin na različnih tarčah. Lahko se izvaja kot samostojen proces ali pa kot komplementarni proces HTS, ki služi kot vir podatkov o vezavah spojin. Gre za kombinacijo različnih pristopov, kjer imamo na voljo veliko orodij in programov, ki jih lahko vključimo v proces virtualnega rešetanja. Znanje o uporabi teh orodij je ključnega pomena. Še vedno pa moramo ohraniti misel, da gre za tehnologijo, ki je komplementarna že prisotnim eksperimentalnim metodam.

Gre za fleksibilen proces, ki poleg zadetkov nudi razširitev kemijskega prostora, saj lahko identificiramo molekule s popolnoma novimi strukturnimi ogrodji. Pred začetkom VS moramo določiti parametre in tako postaviti meje. Ti parametri lahko ključno vplivajo na sam potek VS, zato moramo biti pri izbiri posebej previdni. Nekateri od teh so ključni za izbrana orodja in metodologije, zato moramo orodja poznati do potankosti.⁽¹⁾

Pri VS je najbolj pomembna oplemenitev knjižnice in navsezadnje tudi identifikacija zadetkov. VS obsega od iskanja glede na farmakofor pa do avtomatiziranega sidranja in

iskanja glede na podobnost struktur (angl. shape based). Vse te metode »rudarijo« po bazah podatkov, ki so predhodno filtrirana z namenom da se znebimo neželenih spojin ali pa izberejo obogateno skupino molekul, kjer je poudarek na določenih lastnostih spojine. Ta baza podatkov oz. molekul mora biti ustrezno pripravljena, kjer se upošteva stereokemija, naboji, tautomerija, konformacije molekul in ostalo. Vsi ti faktorji vplivajo na izkupiček rešetanja ampak še vedno nismo v popolnoma prepričani popolnem doprinosu oziroma vplivu vsakega atributa. Konformacija liganda je tak parameter, ki ključno vpliva na rezultate VS in mu bomo posvetili pozornost v tem magistrskem delu.

1.2.1 VRSTE VIRTUALNEGA REŠETANJA

Virtualno rešetanje delimo na dva dela: rešetanje, ki temelji na strukturi liganda in pa rešetanje, ki temelji na osnovi tarče.

1.2.1.1 REŠETANJE NA OSNOVI STRUKTURE LIGANDA

Ena izmed glavnih prednosti rešetanja na osnovi liganda je, da za njen potek ni potreben fizični obstoj tarče zato imamo na razpolago več kemijskega prostora. Potreben pa je obstoj aktivnega liganda s potrjeno biološko aktivnostjo. Ločimo več načinov rešetanja:

- **Rešetanje s pomočjo farmakoforov**

Temelji na konceptu farmakofora, ki je zbirka steričnih in elektronskih lastnosti molekule, ki so potrebne za zagotovitev optimalnih molekularnih interakcij s specifično biološko tarčo, kar vodi do biološkega odziva. V primerjavi s prstnimi odtisi lahko tu pridobimo strukturno različne spojine a s podobno razporeditvijo funkcionalnih spojin.

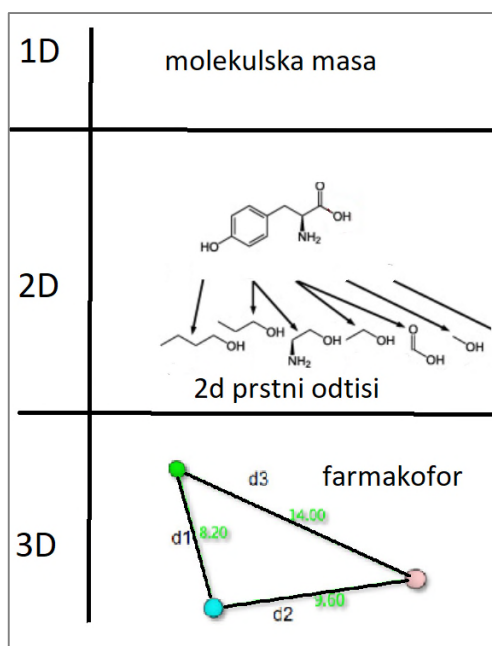
- **Rešetanje na osnovi podobnosti**

Ta princip temelji na osnovi teorije, da strukturno podobne molekule izkazujejo tudi podobne lastnosti. S takim načinom se lahko zelo omejimo, saj raziščemo majhen del kemijskega prostora. Podobnost molekul lahko definiramo na več načinov: 1D, 2D in 3D podobnost, ki jih izrazimo s pomočjo deskriptorjev (Slika 1). Deskriptorji so lastnosti molekule, ki jih lahko določimo in izmerimo. 1D so najbolj osnovni in najbolj splošni ter se uporabljajo samo za grobo filtriranje (npr. molekulska masa). 2D deskriptorji se nanašajo na strukturne značilnosti molekule (npr. osnovna struktura, 2D prstni odtisi, število obročev itd.). Pri 3D deskriptorjih pa imamo še

informacijo o relativnem mestu strukturnih elementov molekule (npr. 3D prstni odtis, oblika in volumen molekule).

- **Reševanje z računalniško podprtim učenjem**

Pri tej obliki reševanja imamo na voljo modelni set aktivnih in neaktivnih spojin s pomočjo katerih program določi pogoje za iskanje novih spojin. Obstaja več načinov računalniško podprtega učenja, med katere spada tudi uporaba umetne inteligence ⁽²⁾.



Slika 1 Primeri različnih tipov deskriptorjev molekul

1.2.1.2 REŠEVANJE NA OSNOVI TARČE - SIDRANJE

Poznamo več metod reševanja na osnovi tarče in sidranje je med njimi najpogosteje uporabljena. Osnova tega reševanja je znanje o tridimenzionalni strukturi tarče, ki jo pridobimo z različnimi metodami (NMR, rentgenska kristalografija). Temu sledi sidranje izbrane knjižnice spojin in ocenjevanje sidranih ligandov. Pri sidranju je zelo pomembna relativna pozicija liganda glede na tarčo in vezavna konformacija liganda. S sidranjem poskušamo najti vse možne poze liganda v aktiven mestu tarče. Dobljene poze in njihovo prileganje z vezavnim mestom nato program oceni z različnimi cenilnimi funkcijami in jih nato razvrsti po ugodnosti vezave med izračunano konformacijo liganda in prilegajočo tarčo ^(2,3).

Ligandi in njihove tarče niso rigidne strukture, kar moramo upoštevati pri našem iskanju. Sidranje si poenostavimo tako, da tarčo velikokrat obravnavamo kot rigidno, pri čemer pa spreminjamo translacijo, rotacijo in konformacijo liganda v kontekstu prostora pripadajoče

terapevtske tarče. V magistrskem delu se bomo posvetili ligandu in njegovemu konformacijskem vplivu na potek virtualnega rešetanja. Če se posvetimo ligandu, potem imamo več možnosti za začetka našega iskanja: generiranje knjižnic konformacij in postopna izgradnja liganda.

1.2.1.2.1 Generiranje knjižnic konformacij

Prva sidranja so bila opravljena z rigidnimi ligandi. Fleksibilnost liganda so lahko testirali tako, da so naredili knjižnico z vsemi rigidnimi pozami ⁽⁴⁾. V realnosti seveda ligandi niso rigidni in njihova poza lahko močno variira, kar seveda vpliva na njihovo interakcijo s tarčo. Tudi konformacija istega liganda, ki jo dobimo iz kristalografskih podatkov iz PDB v dveh različicah istega proteina lahko variira.⁽⁵⁾ Razlika v pozi liganda je prikazana s kartezijski koordinatami v 3D prostoru. Program postavi ligand v različne poze in tako ustvari rigidne verzije iz katerih tvori knjižnico konformacij. S to bazo molekul potem izvajamo sidranje. Programi preberejo konformacije, ki so v svoji specifični pozi in jih izpostavijo obliki in specifikacijam vezavnega mesta tarče. Obstajajo številni programi, ki znajo relativno dobro generirati vezavno konformacijo ligandov (npr. Omega od OpenEye) ⁽⁶⁾. Dobra stran takih knjižnic konformacij je ta, da so enostavno prenosljive med različnimi program za sidranje.

1.2.1.2.2 Postopna izgradnja liganda

Pri postopni izgradnji program razstavi molekulo na manjše fragmente. Najprej poteka sidranje osnovnega fragmenta, nato pa sledi postopna izgradnja liganda. Poza liganda med tem postopkom je izbrana s pomočjo cenilnih funkcij, kjer boljši rezultat določa obliko molekule. Izgradnja liganda poteka v prisotnosti tarče. Slabost tega postopka je ta, da ima veliko molekul enake osnovne fragmente in je zaradi tega lahko postopek časovno potraten. Proces je možno pospešiti tako, da se vse poze fragmentov izračunajo predhodno in se nato shranijo za kasnejšo uporabo. Na ta način poteka generiranje konformacij pri programu Omega, ki ga bomo uporabili pri našem delu.

1.2.1.2.3 Stohastična metoda/Naključna metoda

Obstajajo tudi alternativne metode generiranja konformacij, kjer poteka generiranje konformacij med samim sidranjem. Takšno sidranje traja dlje in je računsko bolj zahtevno in je v tem delu ne bomo uporabljali. Pri tej metodi je vsako sidranje edinstveno in rezultanti niso 100% ponovljivi. Ločimo dva glavna algoritma v uporabi: Monte Carlo in genetski

algoritem. Pri obeh metodah pride do izvajanja naključnih sprememb na ligandu, kjer je cilj dobiti najboljši možen rezultat.

Z Monte Carlo načinom spreminjamo parametre (rotacija vezi, rotacija in translacija molekule) posamično in hkrati poteka ocena vezave s cenilno funkcijo. Če je rezultat nove poze boljši od prejšnje, potem se ta poza ohrani in se išče naslednja naprej. Hkrati se uporablja postopek »simulated annealing« kjer program simulira postopni padec temperature z namenom, da ujamemo kompleks ligand-tarča v energetske minimumu. Začetno zvišanje temperature omogoči prehod čez energetske bariere z namenom zajema širšega konformacijskega prostora. Pri genetskem algoritmu najdemo sledove Darwinove teorije evolucije. Poza liganda predstavlja kromosom in njene lastnosti so geni. Z »evolucijo« preko mutacij se naključno ustvarjajo nove generacije oziroma nove konformacije, ki se ocenijo s cenilno funkcijo in ohranijo ali zavržejo. ⁽⁷⁾

1.2.2 CENILNE FUNKCIJE

Po umestitvi ligandov v vezavno mesto tarče sledi ocena vezave konformacij. Obstaja mnogo različnih cenilnih funkcij in tehnik ocenjevanja vezave kompleksa ligand-tarča, ki se razlikujejo po hitrosti in natančnosti. Izbira vrste cenilne funkcije je odvisna od količine podatkov oz. kompleksov, ki jih hočemo obdelati. Pri ogromni količini pride v poštev manj rigorozna cenilna funkcija, ki nam omogoči, da uspemo obdelati podatke v smiselnem času. Na razpolago imamo več tipov cenilnih funkcij:

- empirične (angl. empirical),
- cenilne funkcije na osnovi polja sil (angl. force field)
- cenilne funkcije, ki delujejo na osnovi statistične mehanike (angl. knowledge based).

Vse te funkcije se uporabljajo z namenom optimizacije lege liganda v vezavnem mestu in na koncu seveda tudi razvrstitev po rezultatih vezave.

Empirične cenilne funkcije:

Ta tip cenilnih funkcij uporablja določen modelni set kompleksa ligand-protein za katerega so znani podatki o vezavi in eksperimentalno določeno 3D strukturo. To je tudi slabost te cenilne funkcije zaradi vprašljivosti napovedne moči izven domene modela, ki je bil uporabljen za postavitve te funkcije.

Cenilne funkcije, ki delujejo na osnovi statistične mehanike (angl. knowledge based):

Osnovane so na znanih informacijah, ki se nanašajo na kompleks ligand-tarča. Te ocenjujejo podobnost s pomočjo strukturnih informacij, ki jih najpogosteje dobijo preko postopka z rentgensko kristalografijo. Čim večkrat najdemo ligand in tarčne atome na določeni razdalji en od drugega, bolj je interakcija ugodna. Vsem tem interakcijam se pripiše energija in končni rezultat je potem vsota vseh teh interakcij znotraj kompleksa.

Cenilne funkcije na osnovi polja sil (angl. Force field):

Osnovane so na ne-vezavnih interakcijah kot so van der Waals-ove interakcije, ki so definirane preko Lennard-Jones-ovega potenciala in elektrostatske interakcije, ki jih definiramo preko Coulombove funkcije. Njihova slabost je ta, da le s težavo opišejo entropijsko stanje sistemov, ki jih ocenjujejo. ⁽⁸⁾

1.2.3 KONFORMACIJE

Večina molekul, ki so farmakološko zanimive ima rotacijske vezi. To so enojne vezi, ki omogočajo prosto rotacijo. Pomembna je tudi konformacija v cikličnih sistemih pri čemer so izključeni rigidni in planarni aromatski ali heteroaromatski ciklični in večciklični sistemi. Izključene so določene vezi kot je C-N vez zaradi visoke energijsko-rotacijske pregrade ali pa na primer amidna vez, ki se obravnava za rigidno zaradi delokalizacije elektronov. Z vrtenjem okrog ene rotacijske vezi lahko dobimo več kot eno konformacijo z zelo podobno energijo. Zaradi tega je ključno da pri raziskavah vključimo več konformacij iste molekule. Paziti moramo na zmernost uporabe več konformacij, saj z uporabo ne le podaljšamo računski čas ampak tudi povečamo število lažno pozitivnih rezultatov in s tem vplivamo na pomembno razmerje med resnično pozitivnimi in lažno pozitivnimi rezultati in posledično na napovedno moč *in silico* eksperimenta.

1.2.3.1 Računalniško generiranje konformacij

Orodja potrebujejo vsaj eno 3D konformacijo na spojino, da je omogočen proces reševanja. Ena izmed možnosti je, da orodja sama generirajo konformacije. Slabost tega je, da je za ta korak potrebna relativno velika računsko moč, kar podaljša čas celotnega postopka. Druga možnost pa je, da orodju že na začetku priskrbimo bazo podatkov z vsemi konformacijami. Z ločitvijo teh dveh procesov imamo možnost večjega nadzora nad bazo podatkov in s tem omogočeno bolj natančno prilaganje nastavitvev pri tvorbi konformacij

Znano je, da bioaktivna konformacija ni enaka tisti z najnižjo energijo kar postavlja procesu generiranja konformacij še posebno pomembnost in težo.

Poznamo več tipov generatorjev konformacij:

- generatorji, ki so osnovani na pravilih in podatkih (rule-data based)
- sistematska iskanja
- numerične metode
- naključne metode
- genetski algoritmi

Kljub tej razdelitvi pa je večina aktualnih tehnologij danes hibrid med več vrstami in jih v splošnem lahko delimo na sistematske in naključne.

V splošnem algoritmi v programih za računalniško generiranje 3D knjižnic najprej identificirajo vezi, ki omogočajo rotacijo. Sledi generiranje konformacij ustreznega števila konformacij (Enačba 1), kjer se ponavadi dolžine in koti vezi ohranijo, spreminja pa se le torzijski kot rotacijskih vezi. Torzijski kot je kot med dvema ploskvama, kjer vsako tvorijo po trije sosednji atomi.

$$N = \left(\frac{360}{k}\right)^n$$

Enačba 1 Enačba za izračun števila konformacij; N = število možni konformacij; k = število torzijskih kotov; n = število rotacijskih vezi

$$RMSD = \sqrt{\frac{\sum_i^N (\Delta d)^2}{N}}$$

Enačba 2 Enačba za izračun RMSD vrednosti; N = število atomov, ki niso vodik; d = razdalja med istima atomoma med dvema konformacijama;

Zaradi eksponentne rasti konformacij so pomembne metode, ki hitro odkrijejo neželene konformacije (sterično nemogoče lege molekul, nerealni koti, napeti obročni sistemi) v veliki prednosti. Program mora tudi identificirati skoraj identične konformacije in se jih znebiti oziroma jih preprečiti. To omogočimo z dodelitvijo meje z RMSD vrednostjo (Enačba 2), s katero izrazimo podobnost med dvema konformacijama. Ob superpoziciji dveh molekul določimo razdalje med istimi atomi obeh molekul. Te razdalje (ali koordinate)

uporabimo za izračun RMSD. Večja je RMSD vrednost, bolj sta ti dve konformaciji različni. (9,10)

1.2.4 GENERATOR KONFORMACIJ – OMEGA2

V magistrski nalogi smo se osredotočili na parameter generiranja konformacij pri virtualnem rešetanju pri čemer smo se še posebej omejili na program OMEGA2, ki je tipični predstavnik postopka generiranja konformacij. Je generator, ki tvori konformacije iz knjižnic fragmentov.

Omega2 spada med komercialne generatorje. Razlike med komercialnimi generatorji so mnogo manjše kot med prosto dostopnimi generatorji.⁽¹¹⁾ Sposobnost generatorja Omega2 so v preteklosti primerjali z drugimi komercialnimi in tudi nekaterimi prosto dostopnimi generatorji (npr. Rdkit). Testirali so jih pri izvedbi generacije 50 in 250 konformacij na ligand s privzetimi nastavitvami. Njihovo natančnost so ocenili z RMSD vrednostjo, ki so jo dobili s primerjavo eksperimentalnih in generiranih vrednosti. Med komercialnimi programi pri nastavitvi tako za 50 kot 250 konformacij je Omega bila najbolj natančna in tudi najbolj hitra. Še posebej velika razlika je bila pri generiranju 50 konformacij. Uporabljeni ligandi so imeli različno število rotacijskih vezi in z večanjem števila teh vezi se je zmanjševala tudi natančnost teh generatorjev. Pri spojinah do 16 rotacijskih vezi je imela Omega najmanjši padec natančnosti. Izkazala se je pa tudi pri generiranju konformacij z geometrijskimi napakami kot so napačne dolžine in koti vezi ali napake pri aromatskih sistemih. Kljub temu, da obstaja veliko orodij za generiranje konformacij pa so primerjalni testi pokazali, da so se komercialni generatorji izkazali za bolj robustne v primerjavi s prosto dostopnimi. (5,10–12)

Omega uporablja hitro, sistematično in »knowledge-based« metodo iskanja konformacij. Proces je sicer sestavljen iz petih korakov, kjer pa se prva dva izvajata le enkrat. Proces generiranja konformacij z Omega2 je sestavljen iz petih glavnih korakov:

- **Priprava baze podatkov o fragmentih:** Ta baza podatkov o fragmentih izhaja iz zbirke komercialni spojin. Te spojine se fragmentirajo na manjše dele in informacije o konformacijah teh fragmentov so pridobljene iz že predhodno generiranih knjižnic.
- **Vzorčenje torzijskih kotov (angl. torsional sampling):** Vsaki rotacijski vezi se pripiše set torzijskih kotov, pri katerih naj bi se vez vzorčila. Ta priporočila izhajajo iz eksperimentalni kristalografskih struktur iz PDB ali iz preračunanih 3D struktur.

- **Ustvarjanje 3D strukture:** Tukaj Omega izhaja iz knjižnice fragmentov, ki je bila ustvarjena pri prvem koraku. Fragmenti z najnižjo energijo se nato sestavijo v celoto dokler ni seznam fragmentov prazen, energija sestavljenega liganda presega zastavljeno mejo ali pa smo dosegli izbrano število konformacij.
- **Primerjava torzijskih kotov (angl. Torsion driving):** Po izgradnji konformacij je vsaka rotacijska vez izpostavljena postopku, kjer se torzijski koti primerjajo s tistimi v knjižnici torzijskih kotov. Tu se ustvari še večji konformacijski ansambel za vsako molekulo.
- **Vzorčenje:** Vsi konformerji, ki jih dobimo po zgornjih štirih korakih so razvrščeni po energijah (po oceni s cenilno funkcijo MMFF94). Konformacije znotraj katerih prihaja do večjih trčenj oz. prekrivanj se izločijo. Konformacije pri katerih je RMSD vrednost nad določeno mejo so sprejete.
- **Stereokemija:** Vhodne informacije o spojinah so lahko v več formatih. Če uporabimo format, ki ne vsebuje 3D strukturne informacije o spojinah (npr. format SMILES) lahko uporabimo funkcijo »flipper«, ki pregleda stereokemijo spojin in vsak stereo izomer obravnava kot ločeno molekulo. Program Omega2 ne preračunava tautomerov in ionizacije spojin. ^(10,13,14)

1.2.5 FRED

Za izvedbo eksperimentalnega in silico dela smo se v magistrski nalogi osredotočili na orodje Fred. Gre za orodje identičnega proizvajalca kot pri Omega2 (OpenEye Scientific Software, Inc.), kar nam omogoča učinkovit in hiter protokol dela z minimalnimi programskimi navzkrižji in napakami.

Program FRED se poslužuje sistematičnih metod pri sidranju ligandov. Uporablja cenilni funkciji Chemgauss3 in Chemgauss4, kjer z metodo funkcije po Gaussu oceni komplementarnost in interakcije liganda z vezavnim mestom. Interakcije med dvema paroma atomoma opiše kot stopničasto funkcijo, ki jo nato prenese na mrežo in le-to ovije z Gaussovo funkcijo. S tem nam omogoči sidranje v manjši ločljivosti in zmanjša napako, ki je povezana z približkom rigidizacije proteina. S Chemgauss3 lahko ovrednotimo: vodikove vezi med ligandom in proteinom, interakcije med kovinskimi ioni in kelatorji ter komplementarnost med ligandom in aktivnim mestom. Chemgauss4 je modifikacija Chemgauss3, ki ima izboljšano ponazoritev vodikovih vezi in kelatorjev.

Gre za hiter program za sidranje, ki sidra vnaprej pripravljene knjižnice konformacij ligandov. Konformacije prebere iz datoteke in jih sidra v vezavno mesto tarče ter jih nato oceni.

Potek delovanja programa:

- Označitev možnih rotacij in translacij vsakega sidranega liganda v vezavnem mestu
- Zavrnitev konformacij liganda, ki so preveč oddaljene ali v stiku z vezavnim mestom
- Zavrnitev konformacij zaradi omejitvenih faktorjev (*ang.* constraint), ki jih postavi uporabnik na začetku
- Ocenitev ostalih konformacij z uporabo cenilne funkcije Chemgauss3 ali 4
- Razvrstitev konformacij po ugodnosti vezave

Sledijo pa še tudi koraki optimizacije:

- Program označi oziroma oštevilči bližnje položaje vsake konformacije tako da naredi en pozitiven in en negativen korak za vsako translacijsko in rotacijsko prostostno stopnjo. Pri tem uporabi polovično resolucijo od tiste v korakih pred optimizacijo, ki jo določimo na začetku sidranja
- Vsako konformacijo oceni s cenilno funkcijo Chemgauss4
- Najbolje ocenjene konformacije program obdrži
- Program preračuna povprečni rezultat vseh ligandov in to označi kot najboljši rezultat liganda ter jo primerja z drugimi molekulami v knjižnici⁽¹⁵⁾

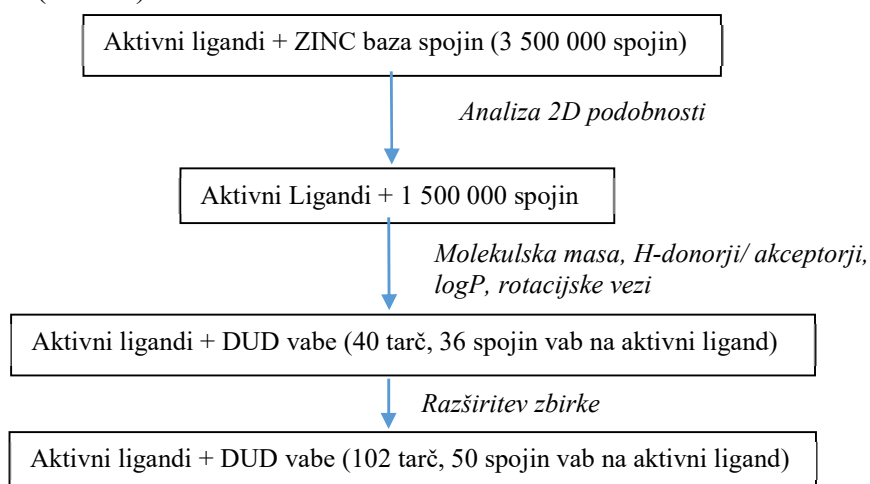
1.3 OCENA SIDRANJA

Sidranje se lahko ocenjuje preko obogatitve knjižnic. Čim višja je koncentracija resnično aktivnih ligandov med najboljšimi zadetki po sidranju tem boljša je obogatitev. Za izvedbo takega testa potrebujemo knjižnico oz. zbirko podatkov, ki vsebujejo tarče s pripadajočimi aktivnimi in neaktivnimi spojinami. Takšna baza podatkov na portalu DUD-E. V magistrskem delu smo za ocenjevanje uporabili ROC krivulje (receiver operating characteristic curve analysis).

1.3.1 DUD-E

Za vir naše podatkovne knjižnice smo uporabili portal DUD-E. Ustvarjen je bil za ocenjevanje in primerjavo procesov in programov za sidranje. Na portalu najdemo tarče in pripadajoče aktivne ter neaktivne spojine. Baza vsebuje 22886 aktivni spojin za 102 tarč, kar je povprečno 224 na tarčo in 50 neaktivnih spojin za vsako aktivno spojino. Spojine vabe oziroma neaktivne spojine imajo podobne fizikalno-kemijske lastnosti kot aktivne spojine, razlikujejo se pa v 2D topologiji oz. geometriji. Ločimo pa dve vrsti neaktivni spojin. En tip je podprt z eksperimentalnimi podatki, drugi tip pa je računalniško generiran. Osnova za DUD-E spojine je podatkovna baza ZINC, ki vsebuje komercialno dostopne spojine.

Iz baze ZINC so na temelju 2D podobnosti spojin ustvarili filter (fingerprinti). Za tem je sledilo računanje lastnosti filtrirane baze, kjer so prioritizirali nekaj pomembnejših funkcionalnih spojin in drug-like deskriptorjev: molekulska masa, število donorjev in akceptorjev vodikovih vezi, število rotacijske vezi, logP. Tako je nastala zbirka DUD in nadalje DUD-E (Slika 2).



Slika 2 Pregled nastanka DUD-E baze.

Gre za daleč najbolj obsežno in javno dostopno zbirko spojin za testiranje zmožnosti sidranja programa , ki ima hkrati tudi raznoliko zbirko tarč. ^(16,17)

1.4 FORMAT VHODNIH PODATKOV

Vsi uvoženi podatki o strukturah bodo v formatu SMILES. Gre za zelo kompakten zapis za spojine, ki je sposoben dvodimenzionalno strukturo molekule ohranjati v obliki enovrstičnega zapis znakov. Kljub temu je sposoben ohraniti informacijo o kiralnosti

spojine. Atomi se označujejo s črkami, vezi in ostale strukturni elementi pa s simboli. Pri tem kiralnost označujemo s poševnico (Preglednica I) ⁽¹⁸⁾ ^(19,20).

Preglednica I Primer SMILES zapisa in strukturne formule iste spojine

SMILES zapis	<chem>CC(=O)OCCC(/C)=C\C[C@H](C(C)=C)CCC=C</chem>
Strukturna formula	

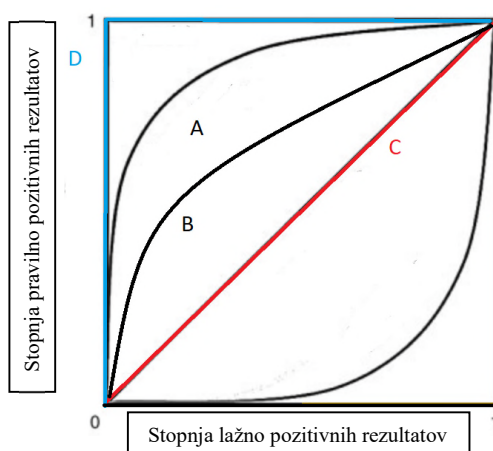
1.5 ROC KRIVULJA (RECEIVER OPERATING CHARACTERISTIC CURVE ANALYSIS)

Uspešnost molekulskega sidranja in oceno obogatitve knjižnice smo ocenili s pomočjo ROC krivulj, ki je zelo hitra, enostavna in objektivna metoda ocenjevanja sposobnosti določenega testa, kjer imamo na voljo dve vrednosti oz. populaciji.

Po procesu VS spojine razvrstimo po rezultatih vezave in z ROC krivuljami ocenimo, kolikšen je delež identificiranih resnično pozitivnih spojin v primerjavi z deležem neaktivnih.

V idealnem primeru se vse aktivne spojine odrežejo bolj od neaktivnih spojinami in so tako na seznamu postavljene pred njimi. V tem primeru bo krivulja iz začetne točke šla navpično do zgornjega levega kota dokler niso prebrane vse aktivne spojine. Za tem pridejo na vrsto neaktivne, torej bo krivulja potovala do zgornjega desnega kota. Vrednost AUC bo v tem primeru 1 (Slika 3). Pri naključnem modelu potuje krivulja po diagonali iz izhodišča do zgornjega desnega kota in bo imela vrednost AUC 0,5 (Slika 3). Pri najslabšem modelu VS, kjer pride do večje obogatitve neaktivnih spojin in bodo le-te postavljene pred vsemi aktivnimi, bo vrednost AUC enaka 0. V tem primeru potuje krivulja najprej do desnega spodnjega oglišča nato do zgornjega desnega oglišča (Slika 3).

Pri veliki količini podatkov in modelov pa uspešnost testiranja lažje primerjamo s pomočjo površine pod krivuljo (AUC). Vrednost AUC se načeloma giblje med 0.5 in 1. Vrednost 1 predstavlja idealen primer, vrednost 0,5 pa dobimo ob naključni izbiri spojin. ⁽²¹⁾



Slika 3 ROC krivulje; D je idealen model; C je naključni model; A je model, ki je boljši od naključnega; B je model, ki je slabši od A ampak boljši od naključja.

1.6 KNIME

KNIME je brezplačen in odprtokodni program, ki omogoča obdelavo ogromne količine podatkov. Pri našem delu ga bomo uporabljali pri tvorbi shem z vozlišči, s pomočjo katerih bomo obdelovali rezultate sidranja in analizirali vhodne spojine. ⁽²²⁾

1.7 UPORABLJENE TARČE

Tarče, ki smo jih izbrali preko portala DUD-E na podlagi velikosti števila aktivnih in neaktivnih spojin:

- ACES (PDB 1e66): je encim acetilholinesteraza, ki katalizira razgradnjo živčnega prenašalca acetilholin. ⁽²³⁾
- ANDR (PDB 2am9): predstavlja receptorsko domeno za vezavo androgenih hormonov (hARLBD). ⁽²⁴⁾
- CAH2 (PDB 1bcd): je encim karbonska anhidraza II, ki katalizira pretvorbo ogljikovega dioksida in vode v bikarbonatni ion. ⁽²⁵⁾
- FA10 (PDB 3kl6): predstavlja koagulacijski faktor Xa, ki sodeluje pri pretvorbi protrombina v trombin. ⁽²⁶⁾

- FNTA (PDB 3e37): je encim farnesiltransferaza, ki katalizira esencialno posttranslacijsko lipidno modifikacijo več kot 60 proteinov.⁽²⁷⁾
- PDE5A (PDB 1udt): je encim fosfodiesteraza 5, ki katalizira razgradnjo sekundarnega prenašalca cGMP.⁽²⁸⁾
- PGH2 (PDB 3ln1): je encim ciklooksigenaza 2, ki katalizira pretvorbo arahidonske kisline v endoperokside.⁽²⁹⁾
- TRY1 (PDB 2ayw): je encim trebušne slinavke tripsin, ki katalizira presnovo beljakovin.⁽³⁰⁾
- VGFR2 (PDB 2p2i): je kinazna domena vaskularno-endotelijskega receptorja 2 rastnega faktorja (VEGFR), ki sodeluje pri angiogenezi.⁽³¹⁾
- MK14 (PDB 2qd9): je z mitogenom aktivirana protein kinaza, ki je udeležena v celični odzivih na različne dražljaje.⁽³²⁾
- HIVPR (PDB 1xl2): je HIV-1 proteaza, ki je udeležena pri hidrolizi peptidne vezi v retrovirusih.⁽³³⁾

2 NAMEN DELA

V okviru magistrske naloge bomo pripravili različne knjižnice spojin, s katerimi bomo preverili vpliv ključnega parametra števila konformacij na rezultate virtualnega rešetanja. S pomočjo portala DUD-E in PDB bomo najprej izbrali tarče in njim pripadajoče aktivne ter neaktivne spojine. Izbira tarč ne sme biti prevelika, ker smo omejeni s procesorskim časom. Pri izbiri tarč v bazi podatkov DUD-E bomo pozorni, da izberem take, ki imajo večje število aktivnih in neaktivni spojin, kar nam bo omogočilo statistično relevantno izvedbo eksperimentov. Po pridobitvi podatkov bo sledila priprava tarč in pripadajočih ligandov (programi Pymol, OEDocking). Iz pridobljenih podatkov bomo izdelali knjižnice kjer bomo za vsako tarčo posebej naključno izbrali dvesto aktivni spojin in pet tisoč neaktivni spojin vab. Sledila bo generacija konformacij, ki je variabilna stopnja pri naših eksperimentih. Dobljene knjižnice konformacij bomo nato sidrali v tarčo s programom Fred. Rezultate sidranja bomo ustrezno zbrali in analizirali v programu KNIME in Excelu. Za vsako tarčo bomo celoten postopek opravili trikrat ter tako dodatno zagotovili statistično signifikantnost rezultatov. Na koncu bomo s pomočjo ROC krivulj in obdelavo podatkov poskušali ugotoviti vpliv števila generiranih konformacij na zmožnost identifikacije aktivnih molekul z *in silico* poskusom sidranja. Predpostavljamo, da z večanjem števila konformacij povečamo možnost odkritja pravilno pozitivnih rezultatov oziroma resnično aktivnih spojin. Naš cilj je ugotoviti primerno število generiranih konformacij glede na tip spojin oziroma tarč ter pripraviti optimiziran protokol sidranja z uporabo 3D konformacijskih knjižnic za uporabo z računalniškega paketa OEDocking (OpenEye Scientific Software Inc).

3 MATERIALI IN METODE

Pri računskih operacijah smo uporabljali tri računalniške sisteme. Dva od teh računalnikov sta serverja v lasti Fakultete za farmacijo in uporabljena sta bila za pripravo in izvedbo dela. Uporabljen je bil tudi osebni računalnik za končno analizo in zapis podatkov.

Računalnik številka 1 (»laseer«):

- Procesor: 4 x 8 jeder, ki računajo s frekvenco 2,0 GHz; podjetje AMD; model Opteron 6128
- Pomnilnik oz. RAM: 32 GB
- Trdi disk: 2 x 1000 GB HD disk
- Operacijski sistem: Ubuntu Linux

Računalnik številka 2 (»plazma«):

- Procesor: 4 x 16 jeder, ki računajo s frekvenco 2,1 GHz; podjetje AMD; model Opteron 6272
- Pomnilnik oz. RAM: 32 GB
- Trdi disk: 2 x 1000 GB HD disk
- Operacijski sistem: Ubuntu Linux

Računalnik številka 3 (»yoga«):

- Procesor: 2 x 2 jeder, ki računajo s frekvenco 1,2 GHz; podjetje Intel; model Core M-5Y71
- Pomnilnik oz. RAM: 8 GB
- Trdi disk: 500 GB SSD
- Operacijski sistem: Windows 10

Uporabljena programska oprema:

- Za obdelavo podatkov: KNIME 3.4.2, KNIME GmbH
- Za pretvarjanje med vrstami podatkov: OpenBabel 2.3.1

- Ustvarjanje konformacij spojin: Omega 2.5.1.4 (OpenEye Scientific Software Inc)
- Za pregled spojin: VIDA 4.3.0.4 (OpenEye Scientific Software Inc.) in Pymol 2.2 (Schrödinger LLC)
- Sidranje: Fred (OpenEye Scientific Software Inc.), ki je del paketa OEDocking v3.2.0
- Za ustrezno pripravo receptorjev: MakeReceptor, ki je del paketa OEDocking v3.2.0 (OpenEye Scientific Software Inc.)

4 EKSPERIMENTALNO DELO

4.1 PRIDOBITEV TARČ

S spletne strani DUD-E (<http://dude.docking.org/targets>) smo izbrali 11 različnih tarč (Preglednica II) in njim pripadajoče ligande ter seveda ustrezne knjižnice aktivnih ter neaktivnih spojin. Pri izbiri smo bili pozorni na zadostno velikost knjižnice, da se že na začetku nismo omejili s majhno bazo izhodnih podatkov. Računalniške kalkulacije smo namreč večkrat ponovili z naključno izbiro spojin iz baze podatkov za določeno tarčo in tako potrdili dobljene rezultate. Tako ima v povprečju vsaka izbrana tarča 468 aktivnih in 29529 neaktivnih spojin vab.

Preglednica II Seznam izbranih tarč s portala DUD-E in pripadajoče število aktivnih ter neaktivnih spojin^(34,35)

Ime tarče	PDB šifra	Tarča	Aktivne spojine	Spojine vabe/ neaktivne spojine
ACES	1e66	Acetilholinesteraza	453	26250
ANDR	2am9	Androgeni receptor	269	14350
CAH2	1bcd	Karbonska anhidraza 2	492	31172
FA10	3kl6	Koagulacijski faktor Xa	537	28325
FNTA	3e37	Farnesiltransferaza	592	51500
PDE5A	1udt	Fosfodiesteraza 5a	398	27550
PGH2	3ln1	Ciklooksigenaza 2	435	23150
TRY1	2ayw	Tripsin 1	449	25980
VGFR2	2p2i	Receptor rastnega faktorja vaskularnega endotelija 2	409	24950
MK14	2qd9	MAP kinaza p38 alfa	578	35850
HIVPR	1xl2	HIV proteaza tipa 1	536	35750

Ob pridobitvi knjižnic spojin smo za vsako posamezno tarčo ločeno obravnavali tako aktivne spojine kot vabe. Vse knjižnice so v formatu *.ism in vsebujejo naslednje podatke o spojini: SMILES koda, ZINC koda, ChEMBL koda (samo za aktivne spojine) (Slika 4). Tako ima vsaka spojina informacije v svoji vrstici.

```
1 Clc5ccc4c(NCCCCNC(=O)CCCc1c[nH]c2cccc12)c3CCCCc3nc4c5 374466 CHEMBL225021
2 O=C(CNc2c1CCCCc1nc3cccc23)NCCC5CCN(Cc4cccc4)CC5 508668 CHEMBL486901
3 C(CCCCNc2c1CCCCc1nc3cccc23)CCCNc4ccnc5cccc45 630282 CHEMBL1082738
```

Slika 4 Primer prvih treh vrstic/ spojin s SMILES, ZINC in ChEMBL šifro spojine

Za vsako tarčo smo nato ustvarili dve manjši knjižnici, torej eno za aktivne spojine in eno za vabe.

Za vsako knjižnico smo naključno izbrali dvesto aktivnih in pet tisoč neaktivni spojin ter tako dobili razmerje 1:25. To smo ponovili 3-krat in tako pridobili 3 različne knjižnice za vsako tarčo. Spodaj so uporabljeni ukazi, ki smo jih vnesli v ukazno vrstico za naključni izbor:

```
shuf aktivne.ism
```

```
head -X > aktivne_Y.ism
```

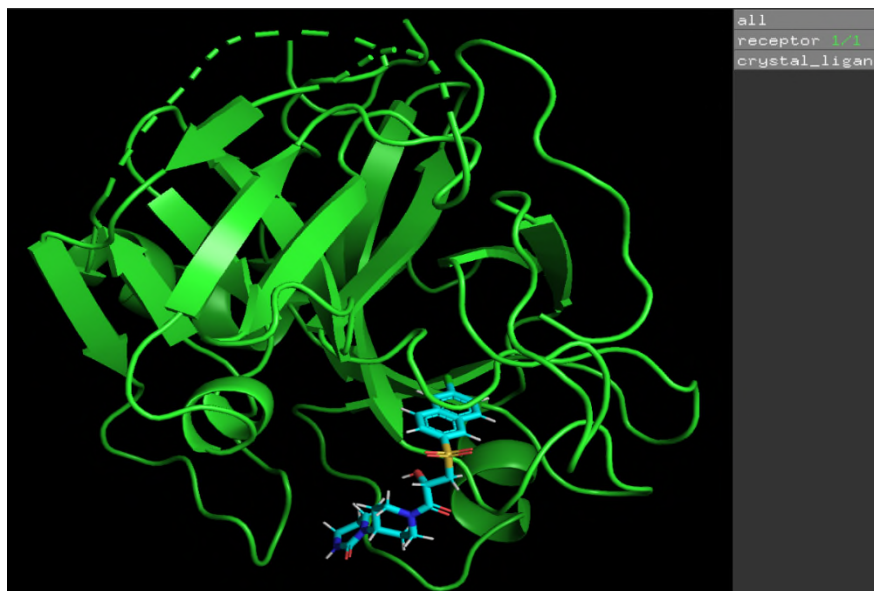
```
shuf neaktivne.ism
```

```
head -X > neaktivne_Y.ism
```

Z ukazom »shuf« v ukazni vrstici smo premešali vse vrstice v datoteki, ki jo vnesemo za ukazom. Z ukazom »head -X« pa izberemo X število vrstic od začetka seznam in jih shranimo v ločeni datoteki, ki jo poljubno poimenujemo (npr. aktivne_Y.ism). Na koncu dobimo šest datotek s spojinami na vsako tarčo, ki smo jih nato uporabili pri generiranju konformacij.

4.2 PRIPRAVA TARČ IN LIGANDOV

Skupaj s spojinami smo z DUD-E strani dobili tudi datoteke s tarčami in pripadajočimi ligandi, kot jih najdemo v bazi PDB. V program Pymol (Slika 5) smo uvozili tako tarčo kot tudi ligand. Temu je sledil izvoz združenih struktur.



Slika 5 Primer tarče (fa10) in pripadajočega liganda, združena v programu Pymol pred skupnim izvozom v datoteko

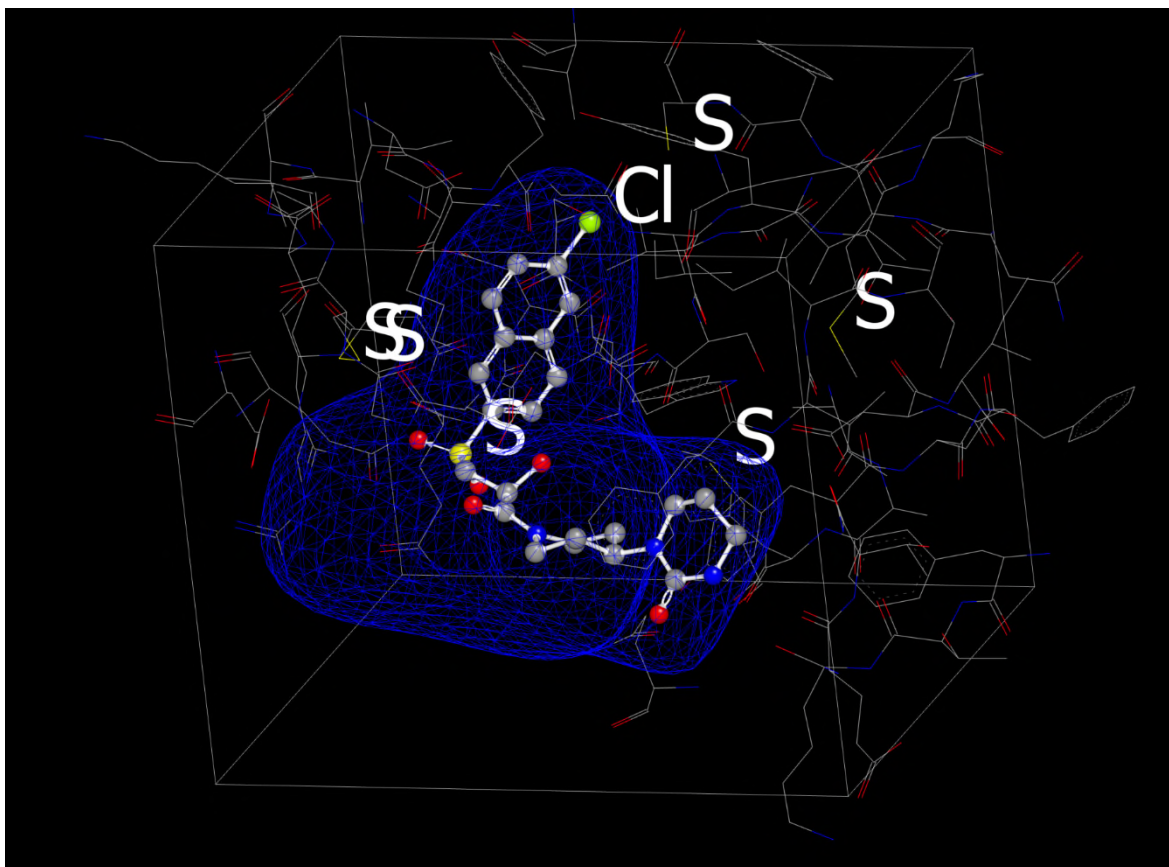
Naslednji korak je bila priprava samega vezavnega žepa v tarči. Za to smo uporabili MakeReceptor program, ki je del paketa OEDocking (OpenEye). Gre za grafični vmesnik, kjer med pripravo najprej specificiramo glavni kristaliziran ligand in dele tarče okrog liganda s katerimi interagira in so sestavni del vezavnega žepa. Sledi izbira velikosti in oblike »box volume« okrog liganda, kjer je naš namen omejiti območje, ki ga bo program preiskal in določil obliko žepa. V naslednjem koraku program določi obliko vezavnega žepa (izbrana opcija »Balanced«). Pri definiciji območja sidranja smo bili pozorni na to, da je bilo vključeno celotno aktivno mesto, kjer nastopa ligand v interakciji z izbranim mestom. Če je bilo potrebno smo ponoviti prejšnje korake in spremeniti velikost izbranega območja. Z njo določimo volumen območja, v katerem se mora nahajati vsaj en atom liganda. Na tem mestu izberemo še možnost »outer contour« s katero označimo območje, v katerem se v času sidranja morajo nahajati vsi atomi molekule. Možnost »Outer contour« pa moramo uporabiti, da bo sidranje končano v razumnem času in se jo ne izbere le, če aktivno mesto nima definirane oblike. V našem primeru so vse tarče imele definirano aktivno mesto (Slika

6). Izračunane vrednosti »outer contour« (Preglednica III) so se gibale med 500 Å³ in 2000 Å³.

S tem smo ustvarili obliko in velikost mesta sidranja, ki predstavlja območje z veliko možnostjo interakcije med ligandom in tarčno ter hkrati majhno možnostjo neugodnih trkov struktur.

Preglednica III Tarče in volumni v programu Make receptor

Ime tarče	ACES	ANDR	CAH2	FA10	FNTA	PDE5A	PGH2	TRY1	VGFR2	MK14	HIVPR
Vezavno mesto (Å ³)	4505	4222	3352	7170	10322	9975	5622	13065	6705	14314	9582
"Outer contour" (Å ³)	554	349	296	1060	2326	1094	415	1880	881	2172	1359



Slika 6 Kristaliziran ligand v aktivnem mestu z izbranim volumnom škatle in območjem »Outer contour« (program Make Receptor)

V zadnjem koraku lahko določimo še uporabo omejitvenih faktorjev (*ang.* constraint), ki jih program upošteva pri omejevanju prostora interakcij med malo molekulo in tarčo. Le-ti ne vplivajo na rezultat vezave ampak na algoritem izbire poze med sidranjem. Poze, ki ne zadostijo zahtevam so zavržene. Pri našem delu se nismo posluževali omejitvenih faktorjev (*ang.* constraint).

4.3 TVORBA KONFORMACIJ Z OMEGA

Po pripravi knjižnice spojin za vse receptorje je sledila tvorba konformacij s programom Omega. Za to smo uporabili naslednjo serijo ukazov, ki smo jih izvedli v ukazni vrstici:

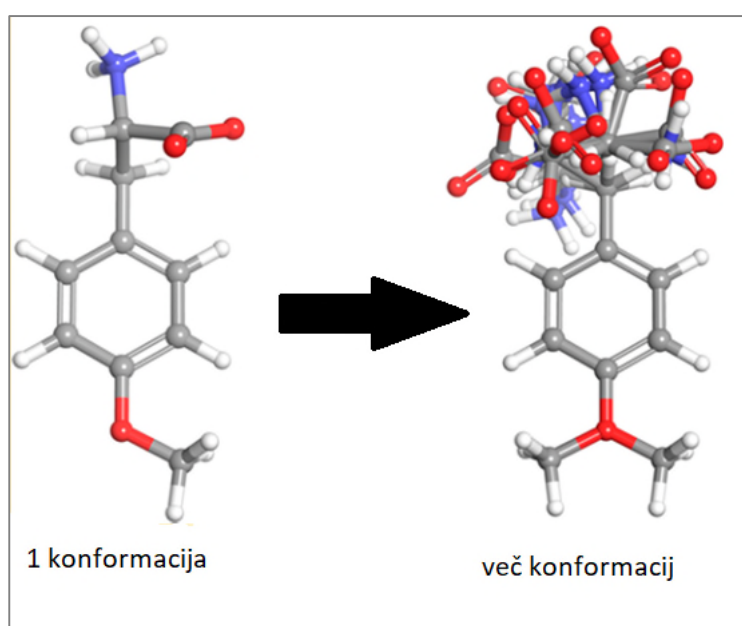
```
omega2 -in knjižnica.ism -out konformacije.oeb -mpi_np X -flipper true -rms 0.5 -maxconfs Y -progress percent -prefix Z
```

- Omega2* S tem ukazom smo zagnali program Omega
- in knjižnica.ism:* Tu navedemo ime ali lokacijo knjižnice na trdem disku, v kateri se nahajajo spojine, pri katerih bo program računal konformacije. Ker je bil ukaz izveden v mapi datoteke, ni bilo potrebno vpisati celotne poti.
- out konformacije.oeb* Tu navedemo ime in format izhodne datoteke v kateri bodo potem shranjene preračunane konformacije spojin.
- mpi_np X* Namesto X navedemo število jeder procesorja, ki jih pri preračunavanju želimo uporabljati. Izbrali smo ji ali 30/ 32 ali 62/ 64 glede na uporabljeni računalnik. Nikoli nismo popolnoma zasedli vseh jeder procesorja z namenom da ga ne preobremenimo ali povzročimo kako napako.
- flipper true* Pri tem ukazu imamo na izbiro dve opciji in to sta *true* (vklop) ali pa *false* (izklop) opcije »flipper«. Z vklopom flipperja program pri vsaki molekuli preveri, kje se nahajajo kiralni centri. Pri najdenih *n* centrih program ustvari 2^n stereo izomere. Tako počne pri vseh molekulah, ki imajo do maksimalno šest kiralnih centrov. Pri molekulah z več kiralnimi centri pa se ustvari naključen set 64 izomerov. V našem primeru so bile molekule že prej kiralno definirane.
- rms 0.5* Kvadratni koren povprečja na kvadrat. Program po superpoziciji ustvarjenih konformacij preveri kakšne so razlike v razdaljah med atomi. Če so le te manjše od zastavljene meje, ustvarjeno konformacijo zavrne in gre na naslednjo. Po privzetih nastavitvah je to 0,5.
- maxconfs Y* Namesto »Y« vnesemo maksimalno število konformacij, ki jih hočemo, da jih Omega ustvari na eno spojino. Gre za ključni parameter pri izvedbi. Uporabili smo možnosti 1, 25, 250 in 2500.
- progress percent* S tem vklopimo prikazovanje napredka računalniške operacije v »%«.

-prefix Z

S tem ukazom določimo predpono datotek, ki bodo ustvarjene in shranjene v postopku.

Postopek smo ponovili pri vsaki tarči šestkrat (trikrat za aktivne spojine in trikrat za neaktivne spojine) za vsako iskanje pri določenem številu konformacij. Proces smo zagnali pri številu konformacij 1, 25, 250 in 2500 torej je bilo vse skupaj 24 vrstic ukazov na tarčo oziroma 264 vrstic ukazov. Konformacije so bile na koncu zapisane v *.oeb tipu dokumenta, ki je najbolj splošen in najpogosteje uporabljen zapis za tovrstne podatke. Podatke smo si pozneje ogledali v programu Vida in Pymol (Slika 7).



Slika 7 Primer liganda z eno konformacijo in primer istega liganda z več konformacijami, ki so naložene ena na drugo.

4.4 SIDRANJE V FRED-U

Sidranje smo zagnali s sledečim ukazom:

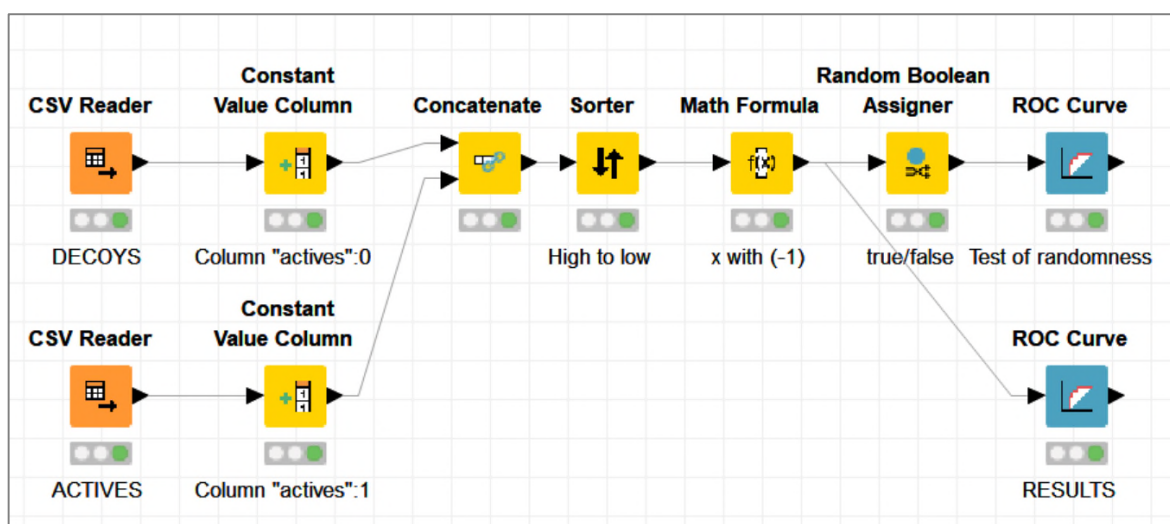
```
fred -receptor X.oeb -dbase Y.sdf -dock_resolution High -mpi_np 30/62 -prefix Z -hitlist_size 0
```

<i>-fred</i>	S tem ukazom zaženemo program Fred.
<i>-receptor X.oeb</i>	Namesto »X« vnesemo pot do datoteke s kompleksom receptor-ligand. V našem primeru smo uporabili samo ime datoteke, ker smo ukaz zagnali z mesta, kjer se je le-ta nahajala.
<i>-dbase Y.sdf</i>	Namesto »Y« vnesemo željeno knjižnico s konformacijami, ki smo jih pred tem ustvarili z generatorjem konformacij Omega.
<i>-dock_resolution</i>	S tem parametrom vplivamo na ločljivost sidranja. Pri možnosti »High« je to pri 1Å (rotacijske in translacijske nastavitve).
<i>-mpi_np 30/62</i>	Tu določimo število jeder računalnika, ki jih bomo uporabili za preračunavanje. Izbrali smo 30 oziroma 62.
<i>-prefix Z</i>	Namesto »Z« vnesemo ime po katerem bomo prepoznali datoteke z rezultati ob koncu sidranja.
<i>-hitlist_size 0</i>	Po privzetih nastavitvah Fred razvrsti rezultate sidranja ligandov od najboljšega do najslabšega in jih pri tem ohrani 500, vse ostale pa zavrže. Z izbiro »0« smo to onemogočili in tako bo Fred izvozil vse rezultate. S tem dobimo celovito sliko sidranja vseh ligandov in konformacij.

4.5 OBDELAVA PODATKOV V KNIME

Po sidranju smo uporabili KNIME za vizualno programiranje in obdelavo vhodnih ter izhodnih podatkov o spojinah. S programom smo oblikovali shemo, ki je sestavljena iz več manjših vozlišč, ki so med seboj povezana zaporedno ali vzporedno. Vsako vozlišče izvaja specifično operacijo pri čemer imamo cel čas izvajanja celotnega algoritma vpogled v vhodne, izhodne podatke in lahko natančno nadzorujemo operacijo ki se izvaja nad podatki v vsaki točki algoritma. Omogočajo nam različne operacije kot so funkcije, kalkulacije, vizualizacije ali katerikoli drugi načini obdelave podatka. Pri našem delu ga bomo v glavnem uporabljali za pridobitve ROC krivulj rezultatov sidranja in pa tudi za izračun ter zapis različnih deskriptorjev, ki jih bomo analizirali pozneje še tudi v programu Excel.

4.5.1 PRIMER IZDELAVE ROC KRIVULJE V KNIME

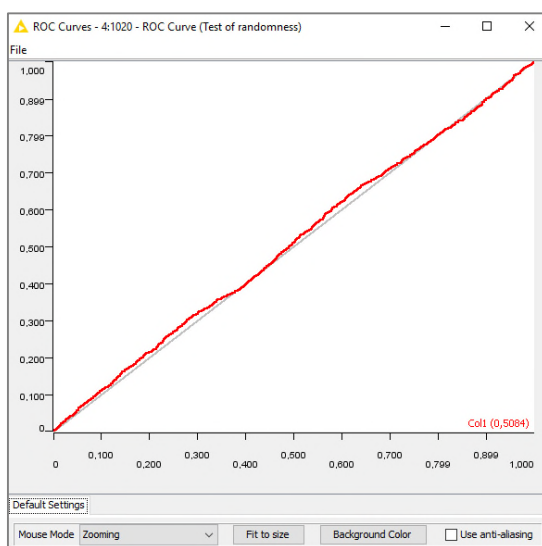


Slika 8 Shema vozlišč v programu KNIME za pridobitev ROC krivulj.

Opis sheme (Slika 8) narejene s programom KNIME:

- »**CSV Reader**«: To vozlišče je bralnik datotek *.csv in drugih formatov. Jaz sem uporabil *.txt datoteke, v katerih so bili shranjeni rezultati sidranja. Uporabili smo dve identični vozlišči, kjer smo eno uporabili za aktivne spojine, drugo pa za spojine vabe. Med dodatnimi nastavitvami smo morali določiti, da je program ustrezno prebral SMILES zapis molekul.
- »**Constant value column**«: To vozlišče omogoča, da nadomestimo ali dodamo stolpec, ki vsebuje informacijo v vsaki vrstici. Tukaj smo dodali stolpec tako med aktivne in neaktivne spojine z vneseno informacijo, ki nakazuje njihovo aktivnost. Neaktivnim spojinam smo dodelili vrednost »0«, aktivnim spojinam pa »1«.

- »Concatenate«: S tem vozliščem smo samo združili oba seznama spojin v enega.
- »Sorter«: Tukaj smo razvrstili vse rezultate skupaj s spojinami od najboljšega proti najslabšemu rezultatu.
- »Math formula«: Tukaj lahko izvedemo matematično operacijo na izbranih podatkih. V naše primeru smo vse rezultate pomnožili z (-1), tako da je bil zdaj najboljši rezultat največje število (na vrhu seznama), najslabši rezultat pa najmanjše število (na dnu seznama).
- »Random boolean assigner«: Tukaj spet dodamo stolpec s podatkom, ki vsebuje eno izmed dve vrednosti: true, false. Te vrednosti se ne nanašajo na lastnosti spojin ampak služi le kot test naključnosti za ROC krivuljo (Slika 9). Obe sta dodeljeni s 50% verjetnostjo, tako da bi morala vrednost AUC pri tej ROC krivulji biti 0,5.
- »Roc Curve«: Pri tem vozlišču se ustvari ROC krivulja. Čim večja je AUC krivulje, večja bo frekvenca aktivnih ligandov med najboljšimi zadetki. Krivulja se vedno začne v točki (0,0) in konča v (1, 1). Spojine začne vozlišče brati od zgoraj navzdol. Za vsako aktivno spojino se pomakne krivulja gor, za vsako vabo pa v desno smer.



Slika 9 ROC krivulja naključnega modela.

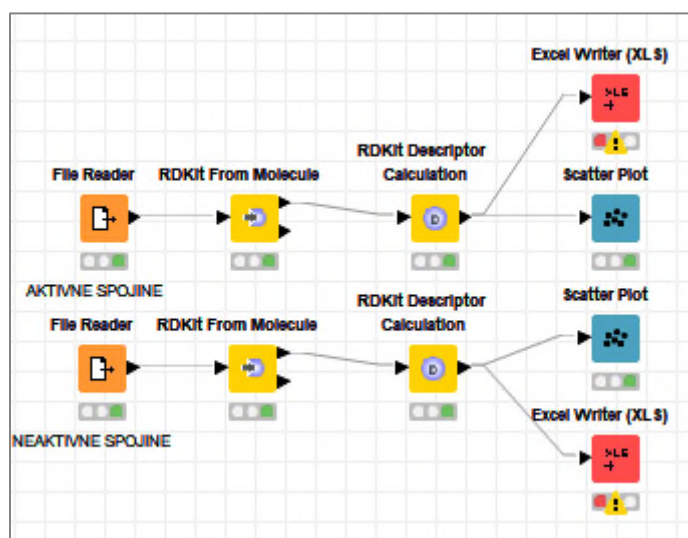
4.5.2 DESKRIPTORJI V KNIME

Uporabili smo dve shemi, da lahko obravnavamo aktivne in neaktivne spojine ločeno. Opis sheme (Slika 10):

- »File reader«: S tem vozliščem smo prebrali SMILES zapise aktivnih in neaktivnih spojin

»Rdkit from molecule«: Tukaj poteka pretvorba iz zapisa SMILES v zapis, ki je berljiv in s tem sposoben obdelave za naslednjo vozlišče.

- »Rdkit descriptor calculator«: Tu poteka izbira in izračunava deskriptorjev iz seznama, ki jih je sposobno to vozlišče izračunati.
- »Excel Writer«: Podatki o deskriptorji se zapišejo v format, ki je berljiv s programom Excel.
- »Scatter plot«: To vozlišče tvori raztresen grafikon, ki nam omogoča analizo izbranih deskriptorjev ter pregled in odkritje trendov.

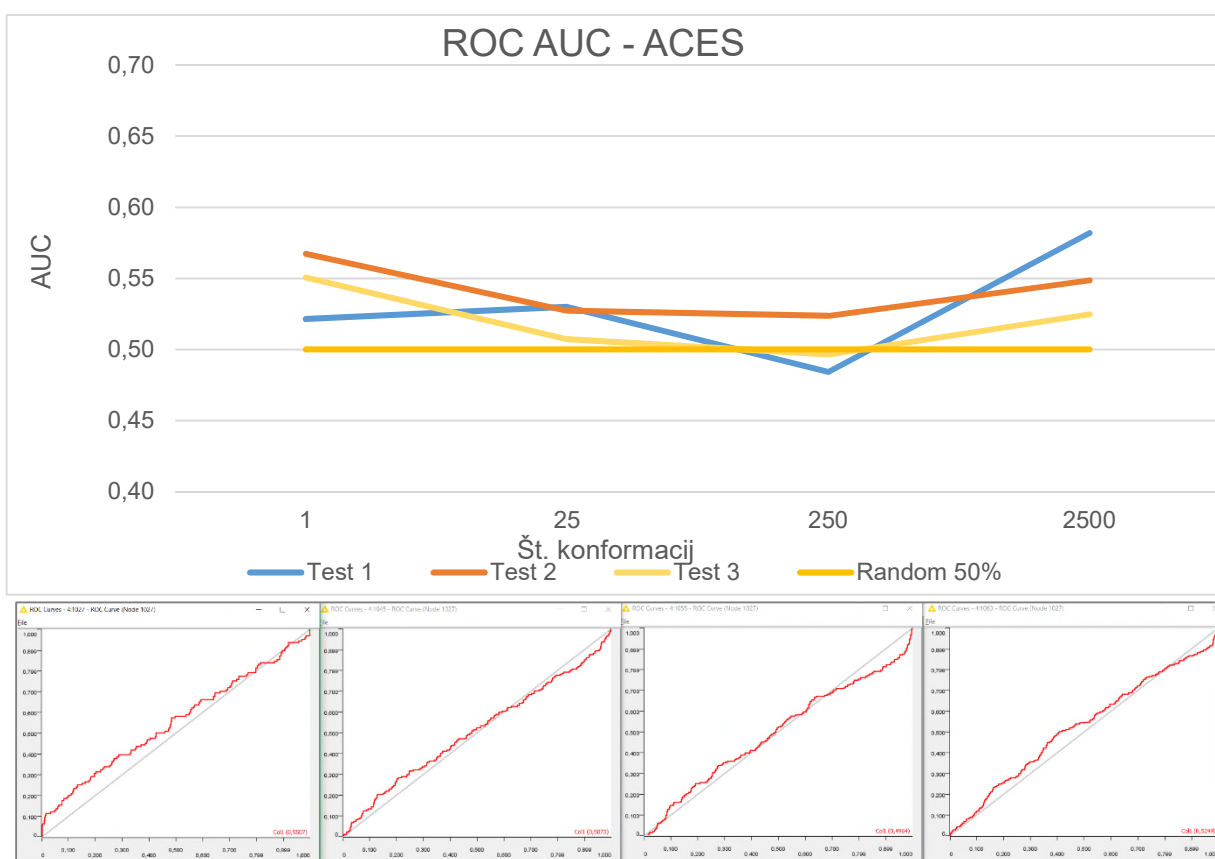


Slika 10 Shema vozlišč v programu KNIME za pridobitev deskriptorjev in ogleda raztresenih grafikonov.

5 REZULTATI IN RAZPRAVA

Rezultate sidranja smo vnesli v ustvarjeno shemo v programu KNIME, s pomočjo katerega smo izračunali ROC krivulje. Z ločeno shemo smo tudi preračunali deskriptorje za vhodne knjižnice spojin. Ob tem smo podatke obdelali tudi v programu Excel, kjer smo ustvarili grafe porazdelitve MM in rotacijskih vezi za aktivne in neaktivne spojine. Podatke o vezanih ligandi smo pridobili s spletne strani PDB (<https://www.rcsb.org/>).

5.1 ACES (PDB 1E66)

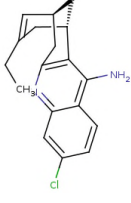


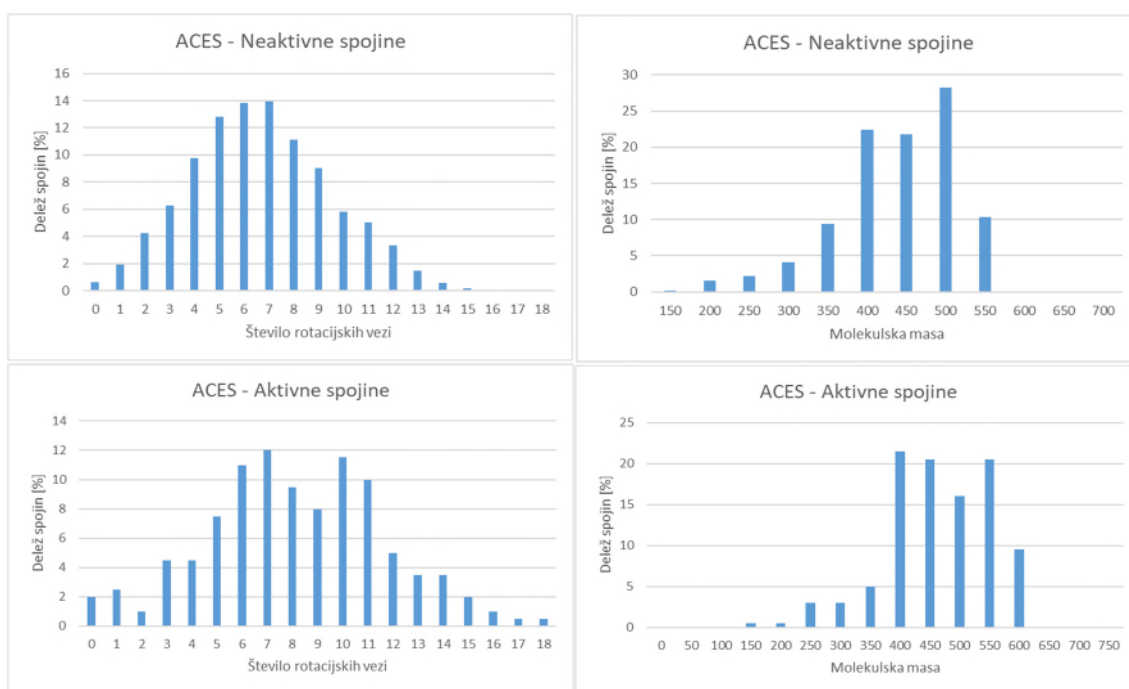
Slika 11 Zgornja slika predstavlja nihanja vrednosti AUC ob spremembi števila konformacij za vse tri teste. Označena je tudi vrednosti 0,5, ki predstavlja naključni model. Na spodnjih slikah so prikazane ROC krivulje za 1, 25, 250 in 2500 konformacij (od leve proti desni).

V primeru ACES lahko vidimo, da v večini primerov vrednost AUC pade pri povečanju števila konformacij in je le zmerno nad vrednostjo 0,5 (Slika 11). Potek ROC krivulje je pri vseh treh testih zelo blizu krivulje naključnega modela, kar pomeni, da pri tem eksperimentu ne more identificirati spojin. To se pokaže tudi pri velikosti AUC, ki ni v nobenem primeru večja od 0,6 in je tako minimalno večja od AUC naključnega modela. V enem primeru se AUC spusti celo pod 0,5. Pri ogledu strukture kristaliziranega liganda (Preglednica IV) smo

ugotovili, da gre za rigidno strukturo molekule, ki ima relativno majhno molekulsko maso v primerjavi z drugimi ligandi (Slika 12), ki smo jih uporabili pri testu.

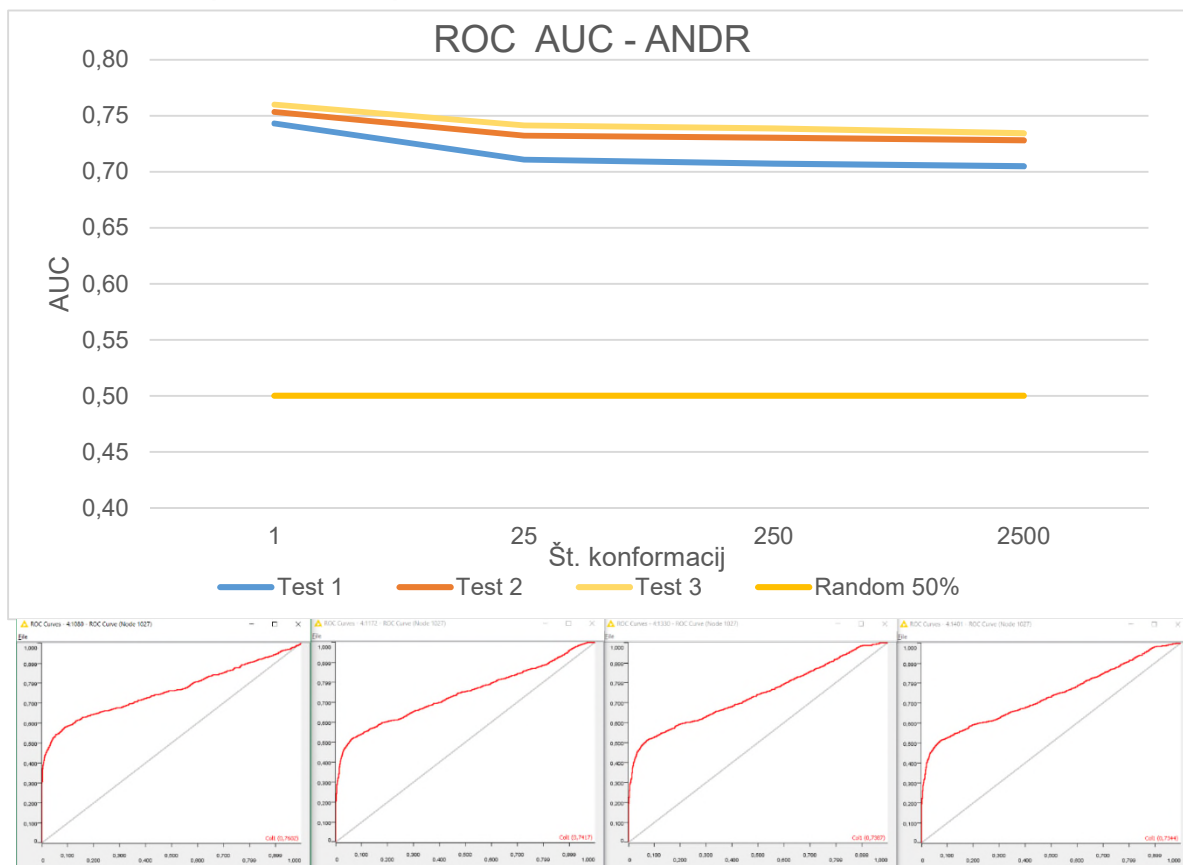
Preglednica IV V spodnji tabeli najdemo strukturo in deskriptorje kristaliziranih ligandov.⁽³⁵⁾

Pripadajoča tarča	ACES	
Molekulska formula	C ₁₈ H ₁₉ ClN ₂	
Molekulska masa (g/mol)	298,81	
Število atomov	40	
Število vezi	43	
Število aromatskih vezi	11	
Število kiralnih centrov	2	
Kiralni centri	C11, C6	
Rotacijske vezi	1	



Slika 12 Razporeditev aktivnih in neaktivnih spojin po številu rotacijskih vezi na levi strani in po MM na desni strani.

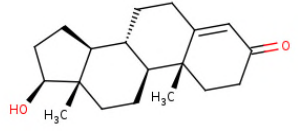
5.2 ANDR (PDB 2AM9)

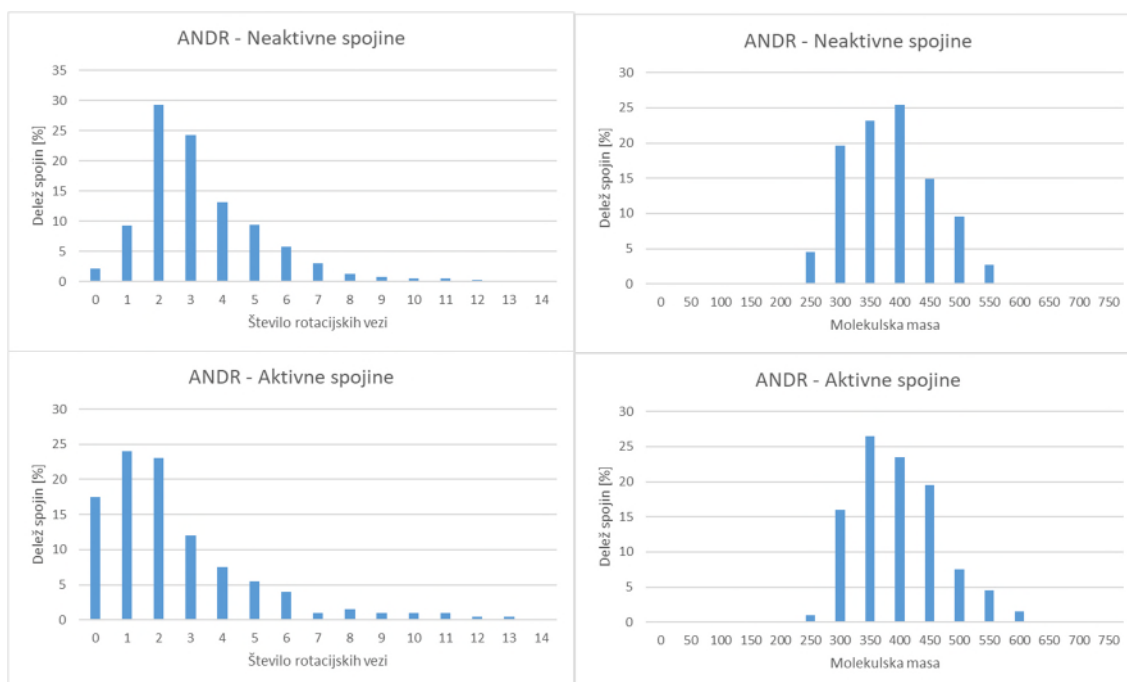


Slika 13 Zgornja slika predstavlja nihanja vrednosti AUC ob spremembi števila konformacij za vse tri teste. Označena je tudi vrednosti 0,5, ki predstavlja naključni model. Na spodnjih slikah so prikazane ROC krivulje za 1, 25, 250 in 2500 konformacij (od leve proti desni).

Vsi trije testi so tukaj proizvedli zelo podobne rezultate. Po ROC krivuljah (Slika 13) lahko vidimo zelo hiter vzpon same krivulje pri vseh uporabljenih konformacijah. AUC je relativno visok in se giblje okrog vrednosti 0,75. Pri dvigu števila konformacij ni praktično nobenega prirastka in tudi ne padca. Podobno kot pri ACES je tudi tukaj kristaliziran ligand v večini ciklične strukture z majhno fleksibilnostjo (Preglednica V). Spojine se tu po deskriptorjih že bolj približajo vezanemu ligandu (Slika 14). Opazimo lahko večji delež molekul z 0 in 1 rotacijsko vezjo med aktivnimi (17,5% in 24%) kot neaktivnimi spojinami (2% in 9%).

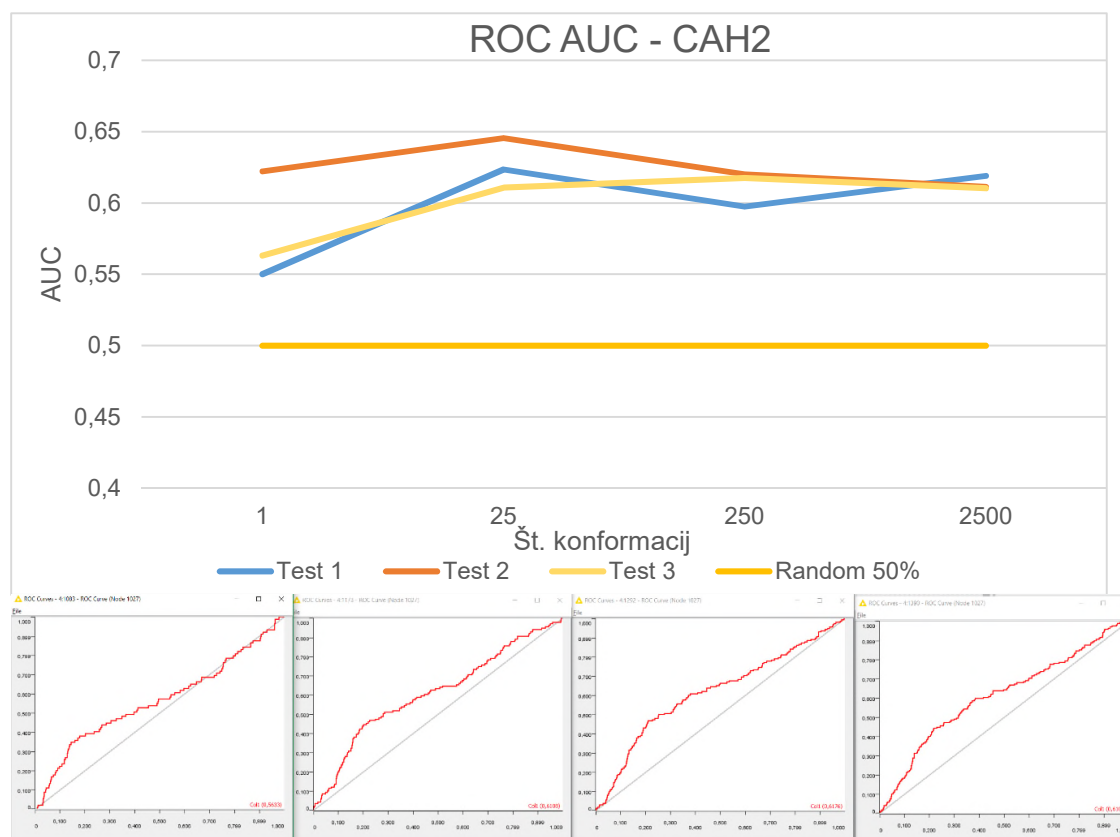
Preglednica V V spodnji tabeli najdemo strukturo in deskriptorje kristaliziranih ligandov. (35)

Pripadajoča tarča	ANDR	
Molekulska formula	C ₁₉ H ₂₈ O ₂	
Molekulska masa (g/mol)	288,42	
Število atomov	49	
Število vezi	52	
Število aromatskih vezi	0	
Število kiralnih centrov	6	
Kiralni centri	C10, C13, C14, C17, C8, C9	
Število rotacijskih vezi	0	



Slika 14 Razporeditev aktivnih in neaktivnih spojin po številu rotacijskih vezi na levi strani in po MM na desni strani.

5.3 CAH2 (PDB 1BCD)

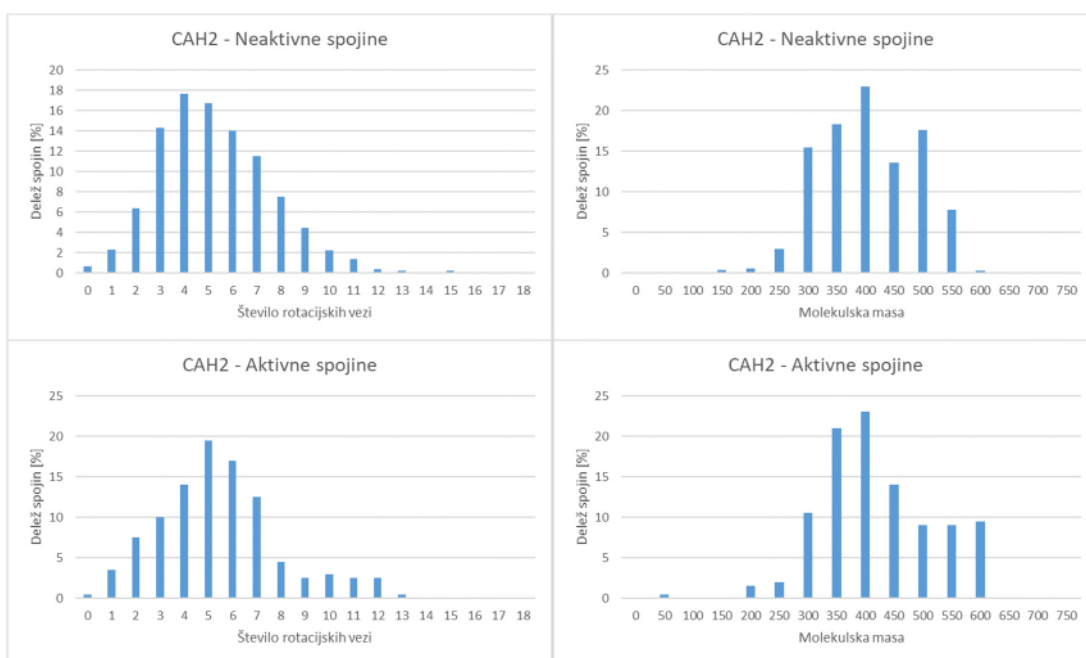


Slika 15 Zgornja slika predstavlja nihanja vrednosti AUC ob spremembi števila konformacij za vse tri teste. Označena je tudi vrednosti 0,5, ki predstavlja naključni model. Na spodnjih slikah so prikazane ROC krivulje za 1, 25, 250 in 2500 konformacij (od leve proti desni).

ROC krivulja se giblje med vrednostjo 0,55 in 0,65. Vidimo zelo majhne prirastke in padce pri spremembi konformacij (Slika 15). Glede na velikost in lastnosti kristaliziranega liganda (Preglednica VI) lahko pričakujemo opažene rezultate. Gre za majhno molekulo z molekulska maso 149 g/mol, ki je tudi najmanjša na našem testu. Zelo velika razlika se pojavi v primerjavi z testiranimi pripadajočimi neaktivnimi in aktivnimi spojina, kjer je povprečna molska masa le-teh okrog 380 g/mol (Slika 16). Kljub temu vseeno lahko opazimo večji prirastek AUC v primerjavi z ANDR in ACES primeroma. Tudi volumen vezavnega mesta je najmanjši na našem testu.

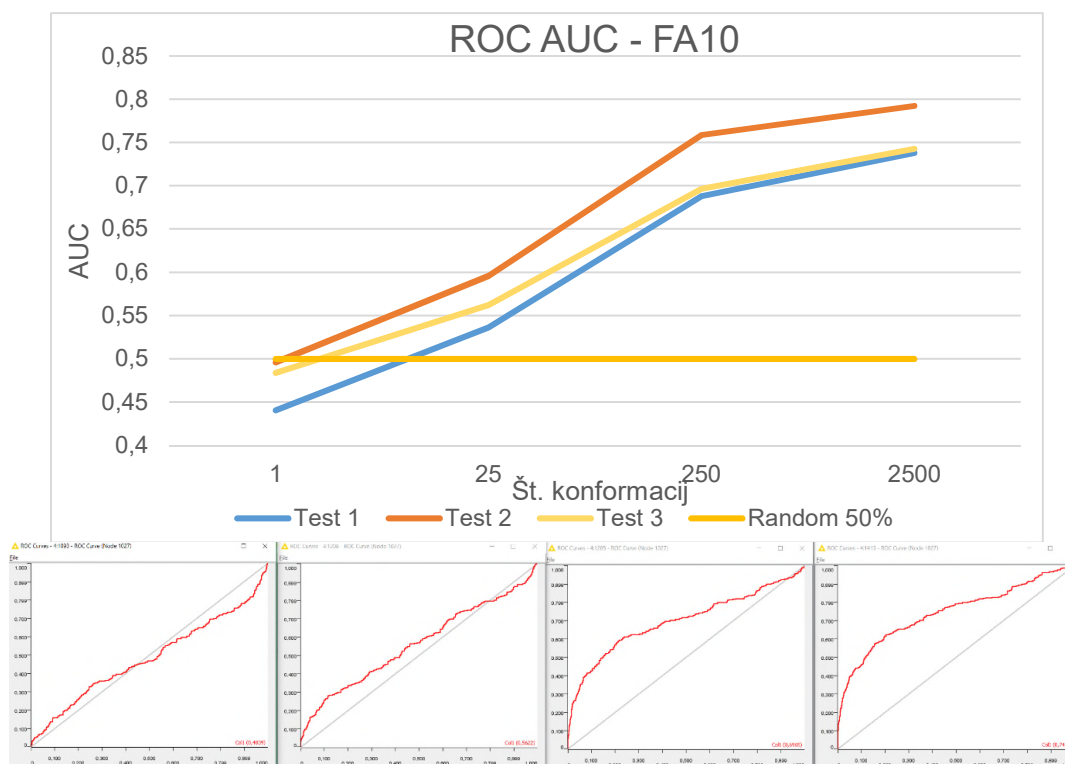
Preglednica VI V spodnji tabeli najdemo strukturo in deskriptorje kristaliziranih ligandov⁽³⁵⁾

Pripadajoča tarča	CAH2	
Molekulska formula	C H ₂ F ₃ N O ₂ S	
Molekulska masa (g/mol)	149,09	
Število atomov	10	
Število vezi	9	
Število aromatskih vezi	0	
Število kiralnih centrov	0	
Kiralni centri	/	
Število rotacijskih vezi	0	



Slika 16 Razporeditev aktivnih in neaktivnih spojin po številu rotacijskih vezi na levi strani in po MM na desni strani.

5.4 FA10 (3KL6)



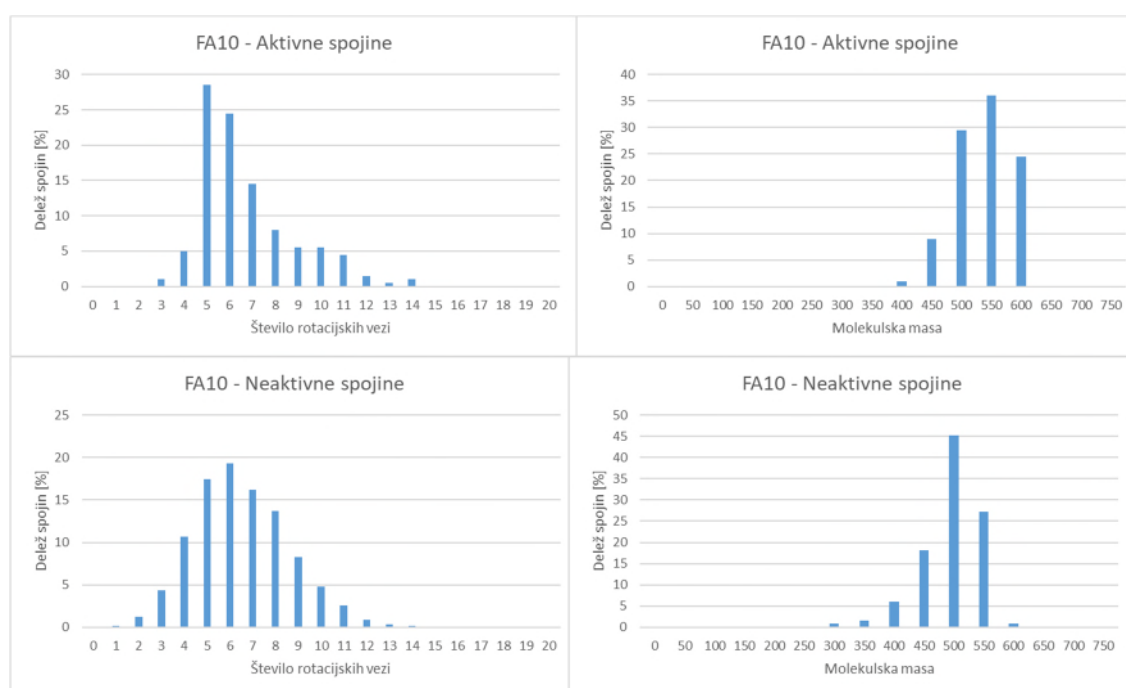
Slika 17 Zgornja slika predstavlja nihanja vrednosti AUC ob spremembi števila konformacij za vse tri teste. Označena je tudi vrednosti 0,5, ki predstavlja naključni model. Na spodnjih slikah so prikazane ROC krivulje za 1, 25, 250 in 2500 konformacij (od leve proti desni).

V primeru tarče FA10 lahko opazimo, da s povečanjem števila konformacij opazimo visok dvig AUC. Začetna krivulja se ne razlikuje veliko od naključnega modela (vrednost AUC okrog 0,5). Pri prvem prehodu dobimo prirastek za 20% na vrednost okrog 0,55, pri drugem pa za 25% predhodnega primera na vrednost okrog 0,7. Presenetilo nas je to, da je prirastek pri prehodu na 250 konformacij večji od prirastka na prehodu do 25. Tako smo pri drugem testu dobili rezultat AUC kar 0,79 z izjemno hitrim narastkom krivulje na samem začetku (Slika 17).

Pri pregledu podatkov lahko vidimo, da je povprečje MM neaktivnih spojin 471 g/mol, aktivne spojine imajo pri tem v povprečju MM 514 (Preglednica XVII). Pri pregledu razlik (Slika 18) smo opazili, da ima velik delež aktivnih MM med 500 in 600 kar je okrog 60% (25% med 550 in 600), med tem ko je delež neaktivnih tukaj okrog 28% (<1% med 550 in 600). Razlike se pojavijo tudi pri številu rotacijskih vezi. Pri neaktivnih vidimo normalno porazdelitev le-teh med 2 in 13, pri aktivnih pa jih ima več kot 50% 5 oz. 6 rotacijskih vezi.

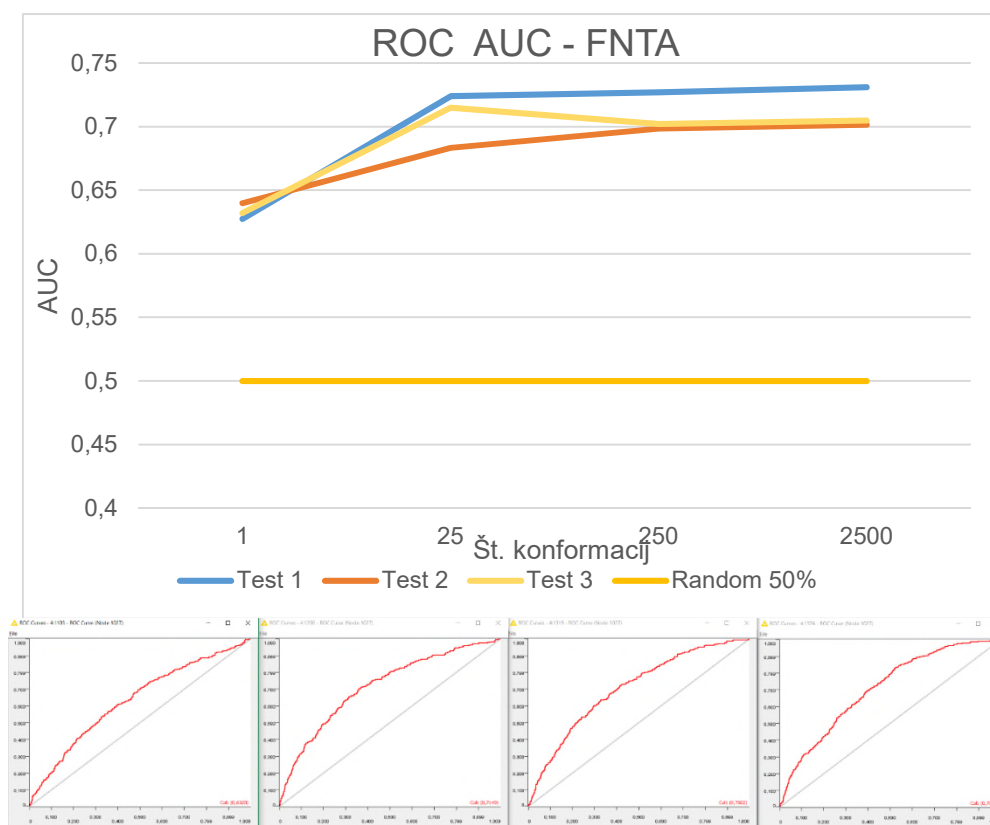
Preglednica VII V spodnji tabeli najdemo strukturo in deskriptorje kristaliziranih ligandov⁽³⁵⁾

Pripadajoča tarča	FA10	
Molekulska formula	C ₂₂ H ₂₆ Cl N ₃ O ₅ S	
Molekulska masa (g/mol)	479,98	
Število atomov	58	
Število vezi	61	
Število aromatskih vezi	11	
Število kiralnih centrov	1	
Kiralni centri	C2	
Število rotacijskih vezi	5	



Slika 18 Razporeditev aktivnih in neaktivnih spojin po številu rotacijskih vezi na levi strani in po MM na desni strani.

5.5 FNTA (3E37)



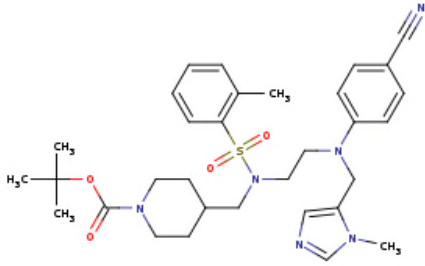
Slika 19 Zgornja slika predstavlja nihanja vrednosti AUC ob spremembi števila konformacij za vse tri teste. Označena je tudi vrednosti 0,5, ki predstavlja naključni model. Na spodnjih slikah so prikazane ROC krivulje za 1, 25, 250 in 2500 konformacij (od leve proti desni).

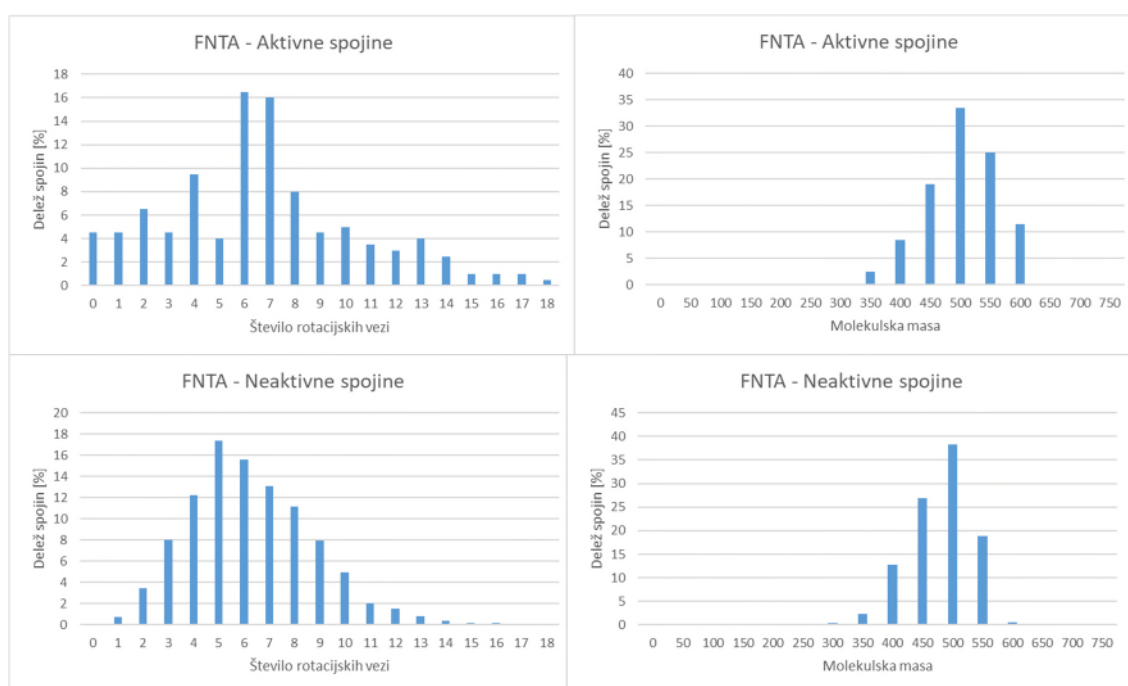
Pri FNTA lahko vidimo dvig AUC za približno 0,1 na vrednost okrog 0,7 pri prvem prehodu konformacij (Slika 19). Za tem krivulja večinoma stagnira okrog vrednosti 0,7. Krivulja ohranja svojo obliko pri vseh primerih konformacij, razen pri manjšem dvigu naklona pri 25 konformacijah.

Velikih razlik med povprečnimi vrednostmi MM in rotacijskih vezi med aktivnimi in neaktivnimi ni (Preglednica XVII), lahko pa vidimo razliko pri distribuciji spojin glede na število rotacijskih vezi (Slika 20). Pri neaktivnih lahko vidimo normalno distribucijo, med tem ko so pri aktivnih približno v enakih odstotkih (okrog 5%) razporejeni med 0 in 13 rotacijskih vezi. Večji delež aktivih (okrog 15%) najdemo le pri aktivnih spojinah z 5 in 6 rotacijskih vezi. Okrog 6% aktivnih spojin pa ima celo več kot 13 rotacijskih vezi (tudi do 18).

Preglednica VIII V spodnji tabeli najdemo strukturo in deskriptorje kristaliziranih ligandov⁽³⁵⁾

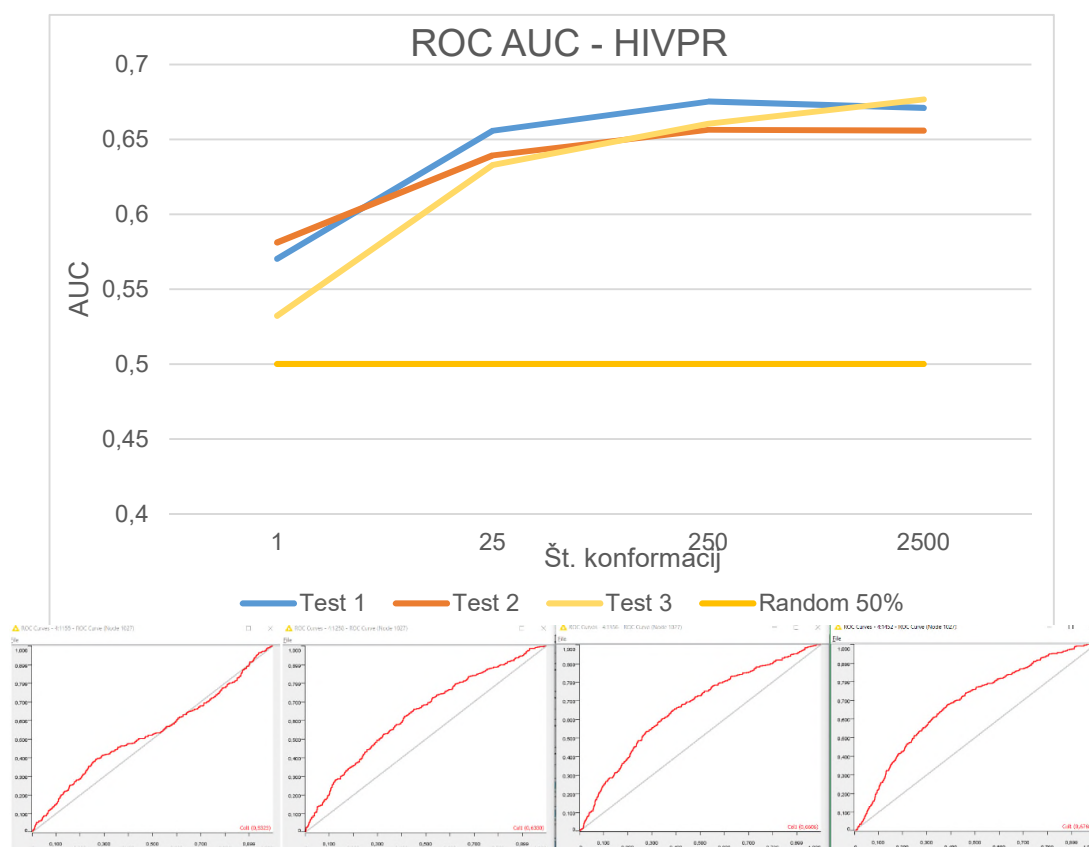
Pripadajoča tarča	FNTA
Molekulska formula	C ₃₂ H ₄₂ N ₆ O ₄ S
Molekulska masa (g/mol)	606,78
Število atomov	85
Število vezi	88
Število aromatskih vezi	17
Število kiralnih centrov	0
Kiralni centri	/
Število rotacijskih vezi	10





Slika 20 Razporeditev aktivnih in neaktivnih spojin po številu rotacijskih vezi na levi strani in po MM na desni strani.

5.6 HIVPR (PDB 1XL2)

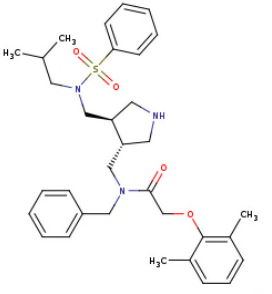


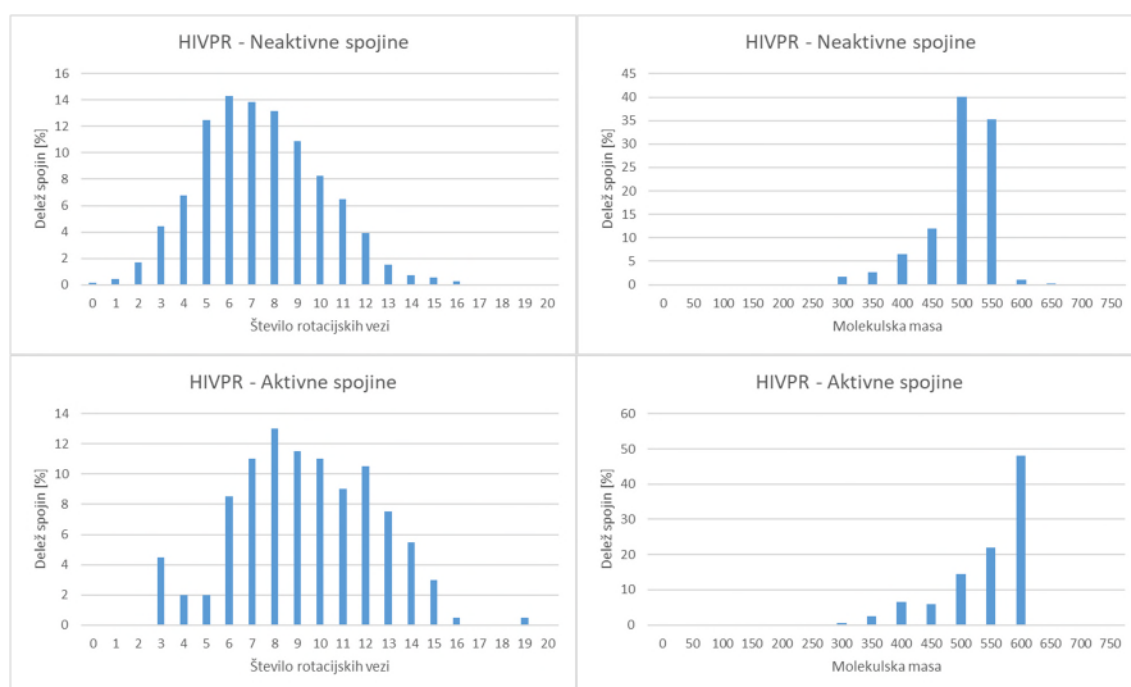
Slika 21 Zgornja slika predstavlja nihanja vrednosti AUC ob spremembi števila konformacij za vse tri teste. Označena je tudi vrednosti 0,5, ki predstavlja naključni model. Na spodnjih slikah so prikazane ROC krivulje za 1, 25, 250 in 2500 konformacij (od leve proti desni).

Največji preskok opazimo pri 25 konformacijah, kjer se AUC poveča za približno 0,1 (Slika 21). Za tem stagnira okrog vrednosti 0,65. V primeru HIVPR ne vidimo veliko posebnosti, razen relativno nizek AUC, ki doseže plato pri 25 konformacijah. ROC krivulja narašča kot pričakovano s povečevanjem konformacij, kjer so tudi prirastki vedno manjši. Oblika krivulje se »izboljša« po prvem povečanju konformacij kar opazimo s povečanjem naklona na samem začetku in odmaknitev od naključne krivulje v osrednjem delu.

Pri primerjavi lahko vidimo višjo povprečno MM aktivnih (522 g/mol) od neaktivnih (473 g/mol) (Preglednica XVII) ter pomik distribucije rotacijskih vezi aktivnih proti višjim vrednostim (Slika 22), kar se pozna tudi pri povprečnih vrednosti (9,4 za aktivne in 7,4 za neaktivne). S temi vrednostmi se aktivne spojine zelo približajo lastnostim kristaliziranega liganda, ki pa ima kar 13 rotacijskih vezi (Preglednica IX).

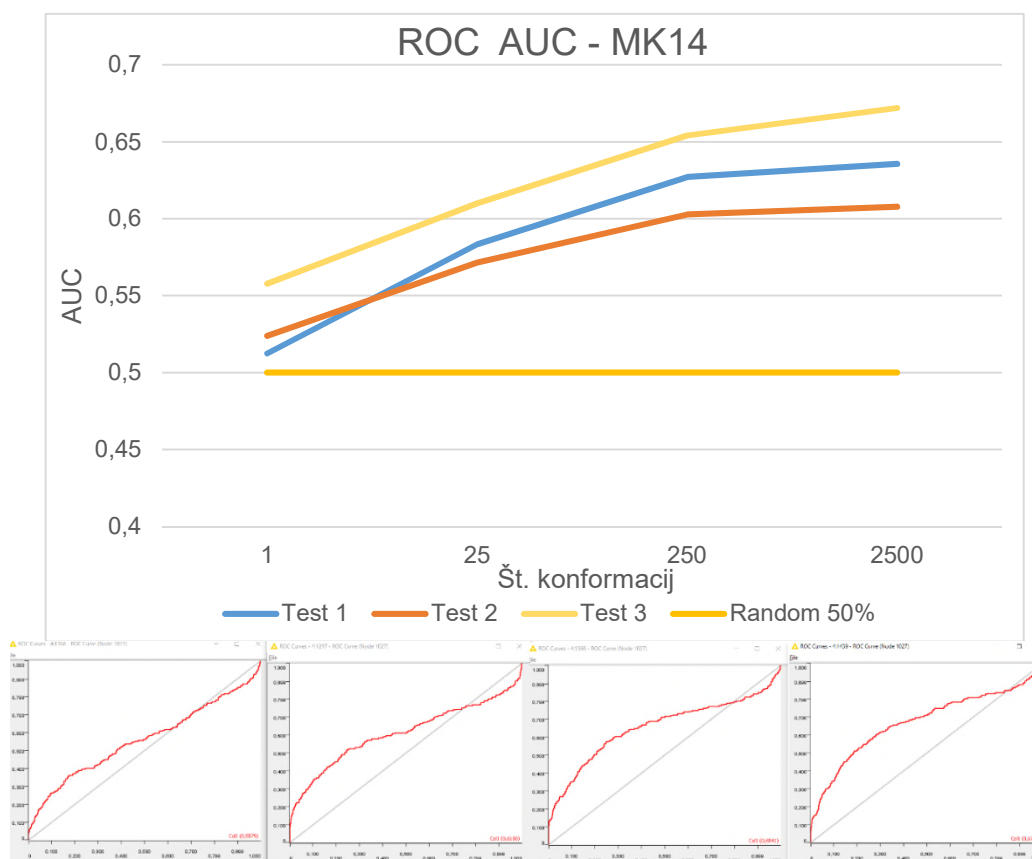
Preglednica IX V spodnji tabeli najdemo strukturo in deskriptorje kristaliziranih ligandov⁽³⁵⁾

Pripadajoča tarča	HIVPR	
Molekulska formula	C ₃₃ H ₄₃ N ₃ O ₄ S	
Molekulska masa (g/mol)	577,78	
Število atomov	84	
Število vezi	87	
Število aromatskih vezi	18	
Število kiralnih centrov	2	
Kiralni centri	C21, C25	
Število rotacijskih vezi	13	



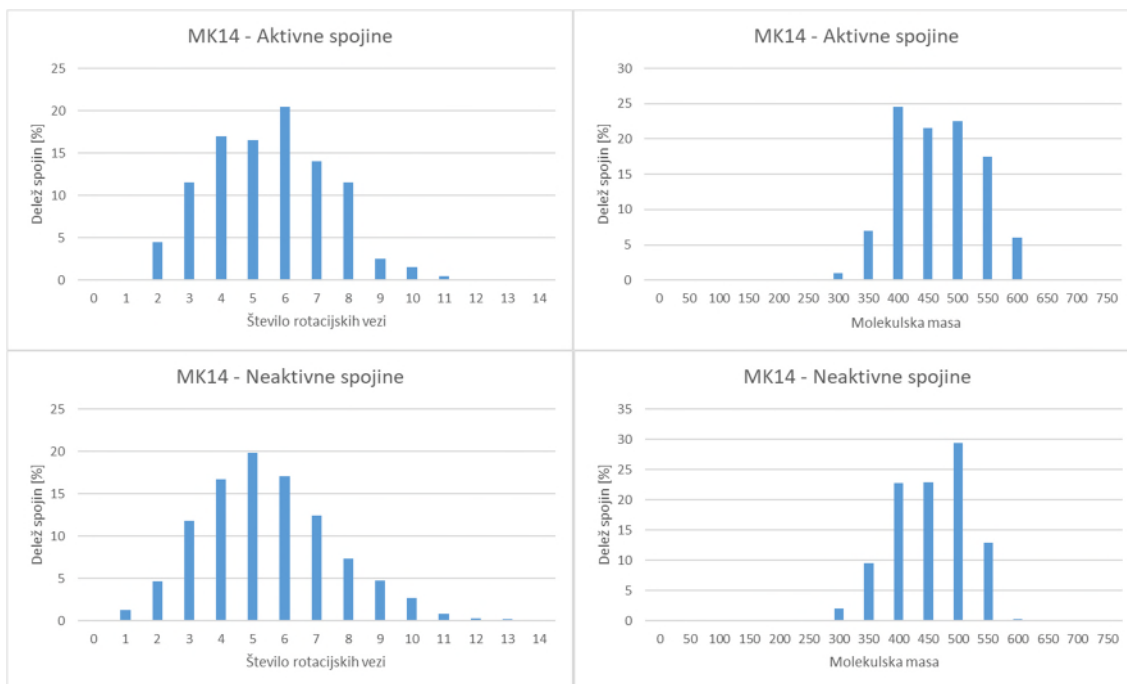
Slika 22 Razporeditev aktivnih in neaktivnih spojin po številu rotacijskih vezi na levi strani in po MM na desni strani.

5.7 MK14 (PDB 2QD9)



Slika 23 Zgornja slika predstavlja nihanja vrednosti AUC ob spremembi števila konformacij za vse tri teste. Označena je tudi vrednosti 0,5, ki predstavlja naključni model. Na spodnjih slikah so prikazane ROC krivulje za 1, 25, 250 in 2500 konformacij (od leve proti desni).

Ob pogledu na ROC krivulje vidimo prirastke AUC, ki se postopoma zmanjšujejo (Slika 23). Na začetku je ROC krivulje blizu naključnega modela a se pri dvigu konformacij poveča. Od idealnega primera nas loči le velikost prirastka, saj so le-ti relativno majhni in tako imamo na koncu največji AUC samo med 0,60 in 0,65. Naklon krivulje se poveča le na začetku krivulje s prvim povečanjem konformacij, med tem ko proti koncu ostane na krivulji naključnega modela. Razlike in distribucije deskriptorjev med aktivnimi in neaktivnimi so minimalne (Preglednica XVII, Slika 24). Večje razlike v MM so pa med knjižnico in ligandom. MK14 ima največje vezavno mesto na našem testu (Preglednica XVII).

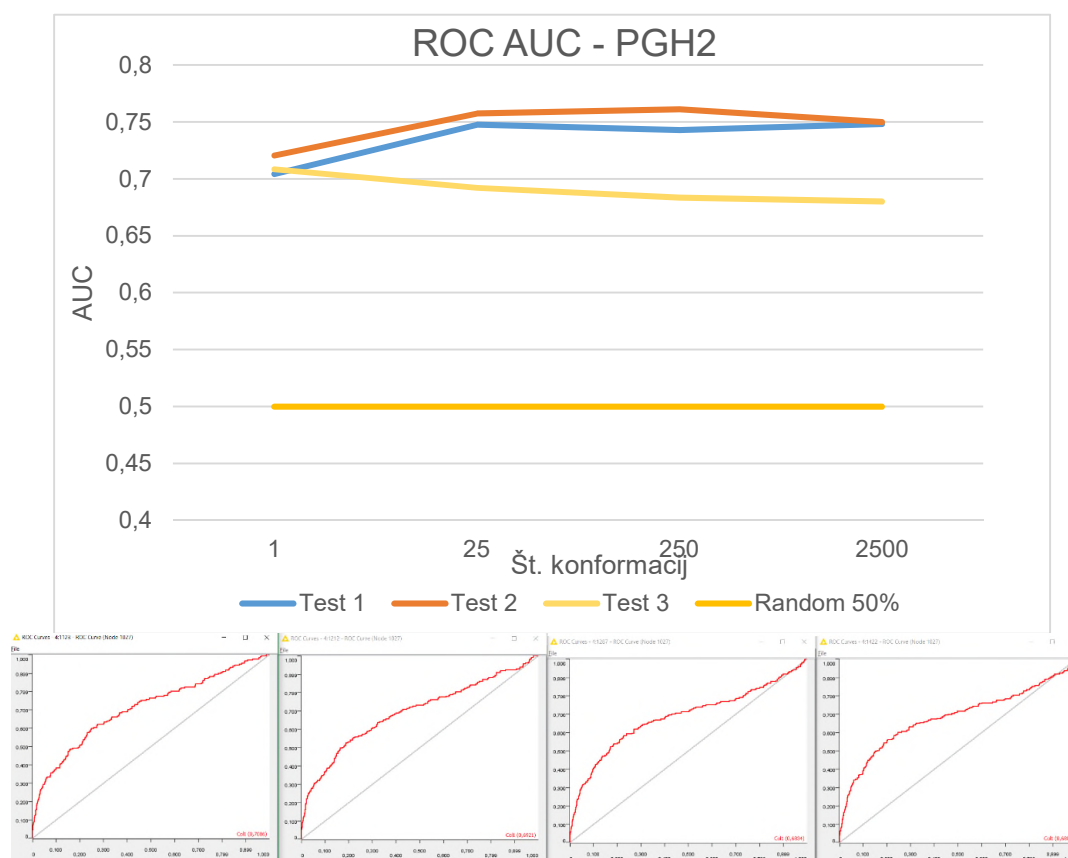


Slika 24 Razporeditev aktivnih in neaktivnih spojin po številu rotacijskih vezi na levi strani in po MM na desni strani.

Preglednica X V spodnji tabeli najdemo strukturo in deskriptorje kristaliziranih ligandov (35)

Pripadajoča tarča	MK14	
Molekulska formula	$C_{27}H_{24}F_2N_6O_5$	
Molekulska masa (g/mol)	550,51	
Število atomov	64	
Število vezi	69	
Število aromatskih vezi	12	
Število kiralnih centrov	2	
Kiralni centri	C23, C37	
Število rotacijskih vezi	4	

5.8 PGH2 (3LN1)

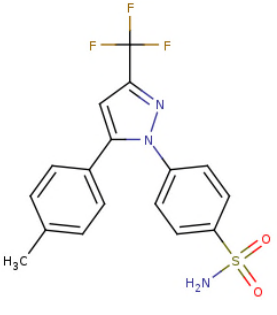


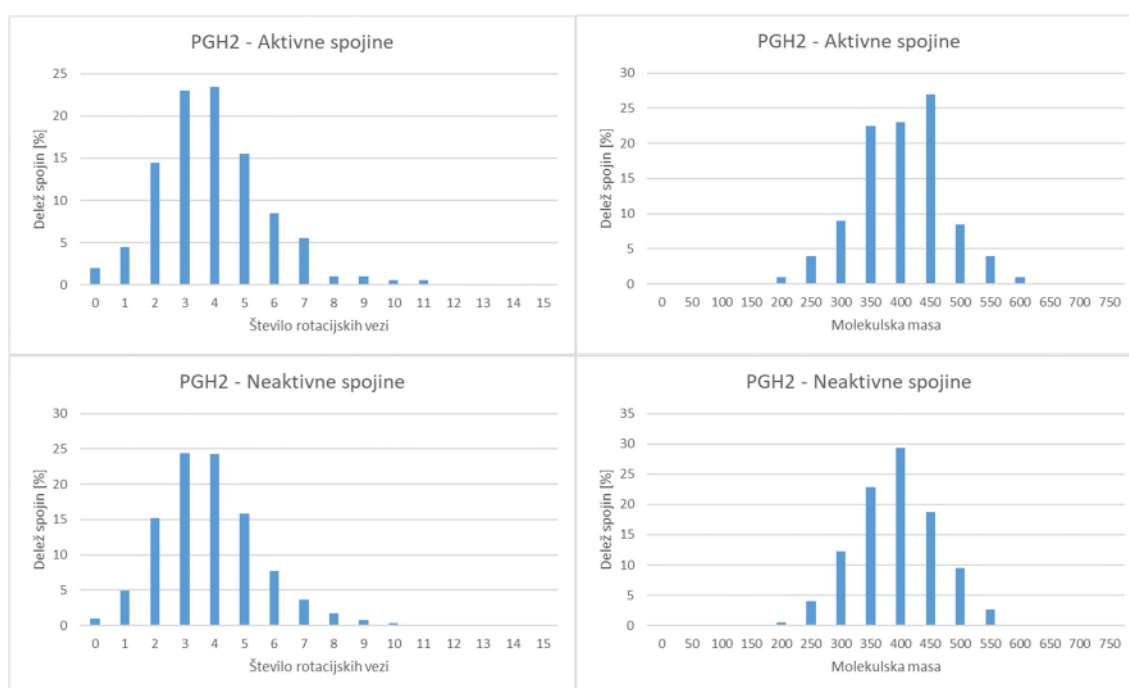
Slika 25 Zgornja slika predstavlja nihanja vrednosti AUC ob spremembi števila konformacij za vse tri teste. Označena je tudi vrednosti 0,5, ki predstavlja naključni model. Na spodnjih slikah so prikazane ROC krivulje za 1, 25, 250 in 2500 konformacij (od leve proti desni).

Krivulja ima že relativno velik AUC pri konformaciji 1 (okrog 0,7), ki se potem še minimalno poveča na vrednost okrog 0,75 pri prehodu na 25 konformacij (Slika 25). Pri tej vrednosti AUC krivulja tudi stagnira. Tokrat se je pri testu številka 3 pojavil odstop, kjer se AUC ob vsakem koraku zmanjšuje kar je različno od drugih dveh testov.

Volumen vezavnega mesta je med najmanjšimi, kar velja tudi za kristaliziran ligand. Le-ta ni popolnoma rigidna saj ima 3 rotacijske vezi (Preglednica XVII, Preglednica XIV) za razliko od prvih treh primerov na našem testu. Aktivne in neaktivne spojine se po deskriptorjih ne razlikujejo veliko (Slika 26). MM in število rotacijskih vezi uporabljene knjižnice so podobne kristaliziranemu ligandu, kar bi podpiralo razlago o višjem AUC (Preglednica XVII).

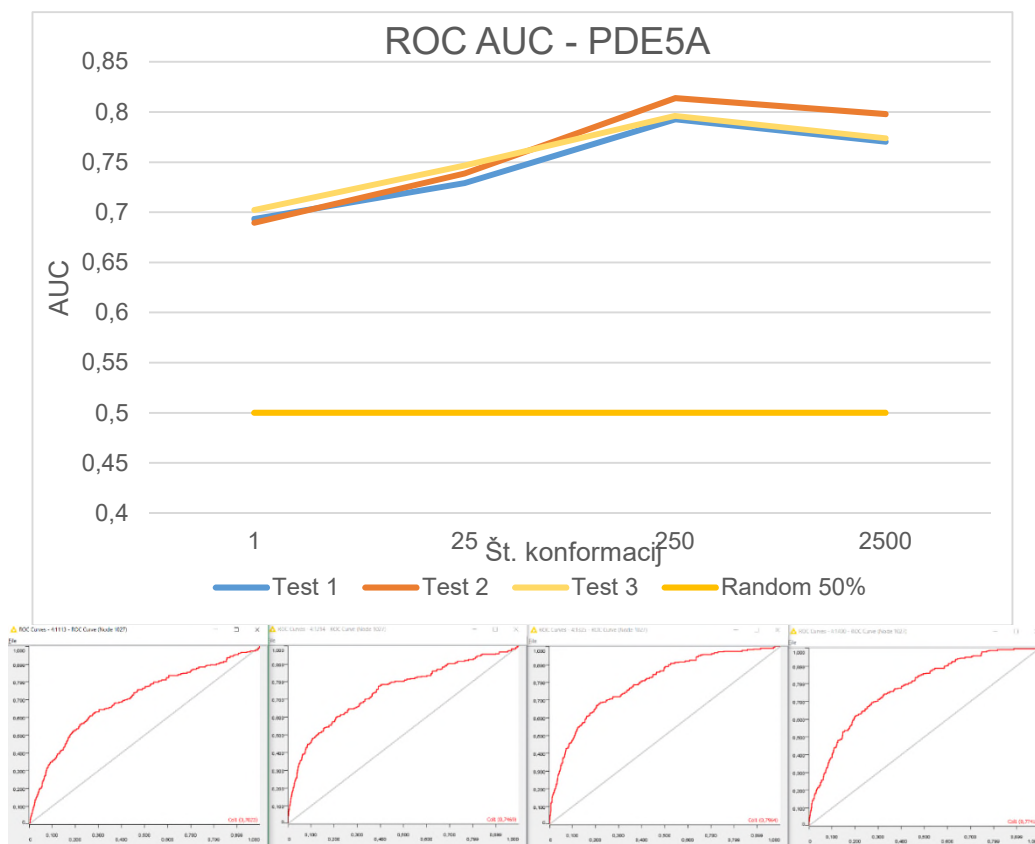
Preglednica XI V spodnji tabeli najdemo strukturo in deskriptorje kristaliziranih ligandov⁽³⁵⁾

Pripadajoča tarča	PGH2	
Molekulska formula	C ₁₇ H ₁₄ F ₃ N ₃ O ₂ S	
Molekulska masa (g/mol)	381,37	
Število atomov	40	
Število vezi	42	
Število aromatskih vezi	19	
Število kiralnih centrov	0	
Kiralni centri	/	
Število rotacijskih vezi	3	



Slika 26 Razporeditev aktivnih in neaktivnih spojin po številu rotacijskih vezi na levi strani in po MM na desni strani.

5.9 PDE5A (PDB 1UDT)

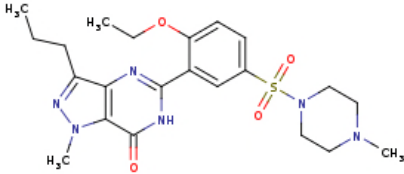


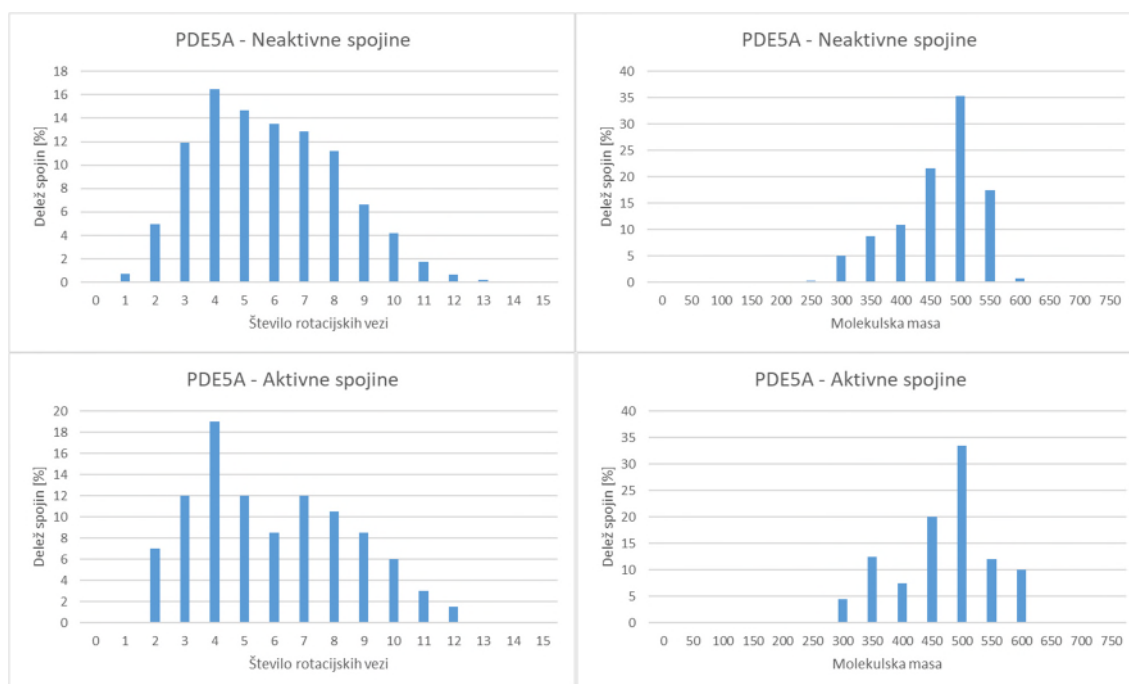
Slika 27 Zgornja slika predstavlja nihanja vrednosti AUC ob spremembi števila konformacij za vse tri teste. Označena je tudi vrednosti 0,5, ki predstavlja naključni model. Na spodnjih slikah so prikazane ROC krivulje za 1, 25, 250 in 2500 konformacij (od leve proti desni).

Začetni AUC je relativno visok (0,7) in ob prehodih narašča za vrednost okrog 0,05 vse do vrednosti okrog 0,8, ki jo doseže pri 250 konformacijah. Pri prehodu na 2500 pa se AUC rahlo zniža. Oblike ROC krivulj so podobne tistim, ki jih najdemo pri FNTA in se jim poveča naklon po preskoku števila konformacij. Razlika je v tem, da imajo večji AUC že od začetka (Slika 27).

Če primerjamo MM in rotacijske vezi spojin v knjižnici s kristaliziranim ligandom lahko vidimo, da so si podobne (Preglednica XVII). Prav tako ni veliko razlik v distribuciji med aktivnimi in neaktivnimi spojinami (Slika 28). Ta tarča ima srednje veliko vezavno mesto.

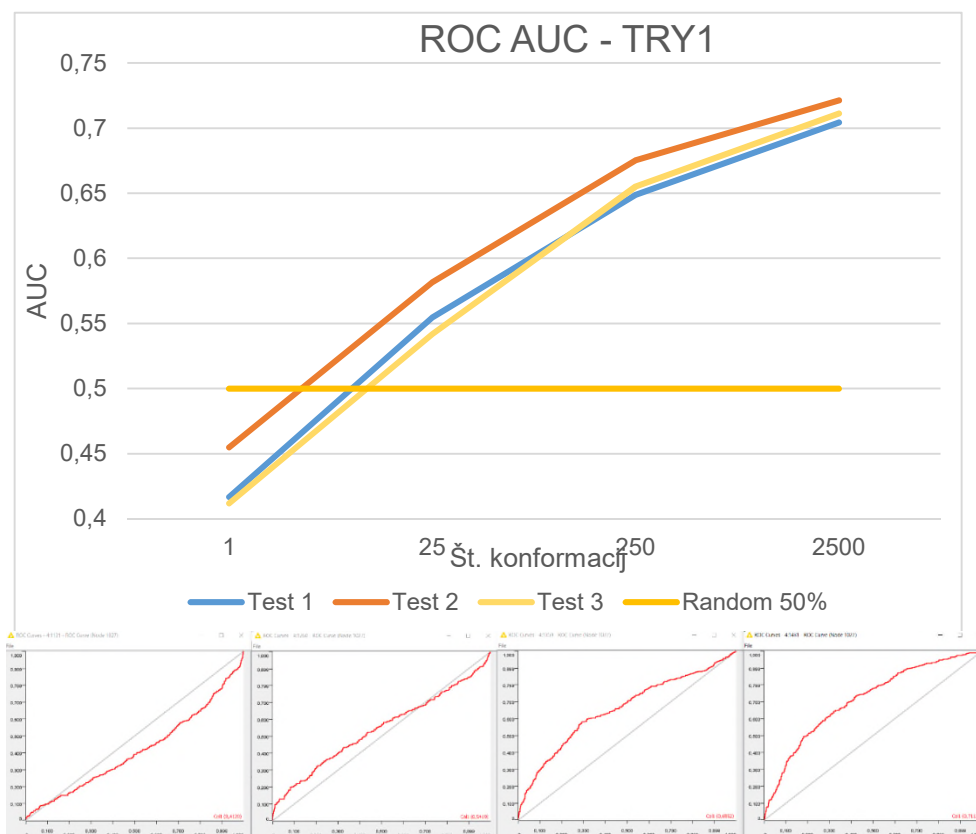
Preglednica XII V spodnji tabeli najdemo strukturo in deskriptorje kristaliziranih ligandov⁽³⁵⁾

Pripadajoča tarča	PDE5A	
Molekulska formula	C ₂₂ H ₃₀ N ₆ O ₄ S	
Molekulska masa (g/mol)	474,58	
Število atomov	63	
Število vezi	66	
Število aromatskih vezi	17	
Število kiralnih centrov	0	
Kiralni centri	/	
Število rotacijskih vezi	7	



Slika 28 Razporeditev aktivnih in neaktivnih spojin po številu rotacijskih vezi na levi strani in po MM na desni strani.

5.10 TRY1 (PDB 2AYW)

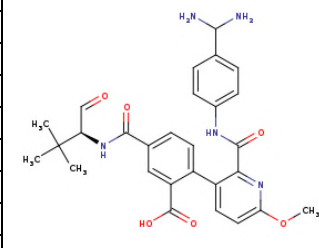


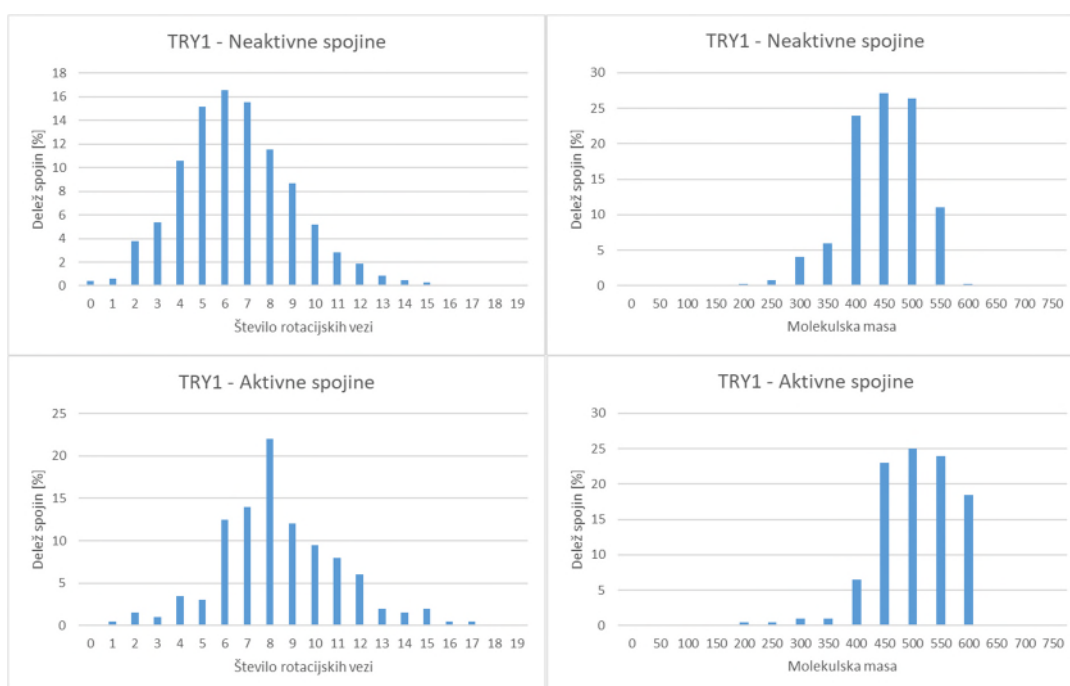
Slika 29 Zgornja slika predstavlja nihanja vrednosti AUC ob spremembi števila konformacij za vse tri teste. Označena je tudi vrednosti 0,5, ki predstavlja naključni model. Na spodnjih slikah so prikazane ROC krivulje za 1, 25, 250 in 2500 konformacij (od leve proti desni).

Pri TRY lahko vidimo, kako bi idealen primer lahko zgledal. Pri konformaciji 1 je krivulja pod mejo naključnega modelu. Z vsakim sledečim dvigom števila konformacij prihaja do velikega prirastka AUC (več kot 0,1), ki pri 2500 konformacijah doseže vrednost okrog 0,7. Prirastek se manjša z večanjem števila konformacij, kar ustreza idealnemu primeru (Slika 29).

Pri primerjavi aktivnih in neaktivnih spojin lahko opazimo razlike tako pri MM kot pri rotacijskih vezeh. V povprečju imajo aktivni MM 482 g/mol, neaktivni pa 425 g/mol in to se opazi tudi pri distribuciji. Aktivne spojine imajo tudi v povprečju za 2 rotacijski vezi več od neaktivnih in so tako v obeh deskriptorjih bolj podobni kristaliziranemu ligandu. Vezavno mesto je med največjimi na testu (Preglednica XVII).

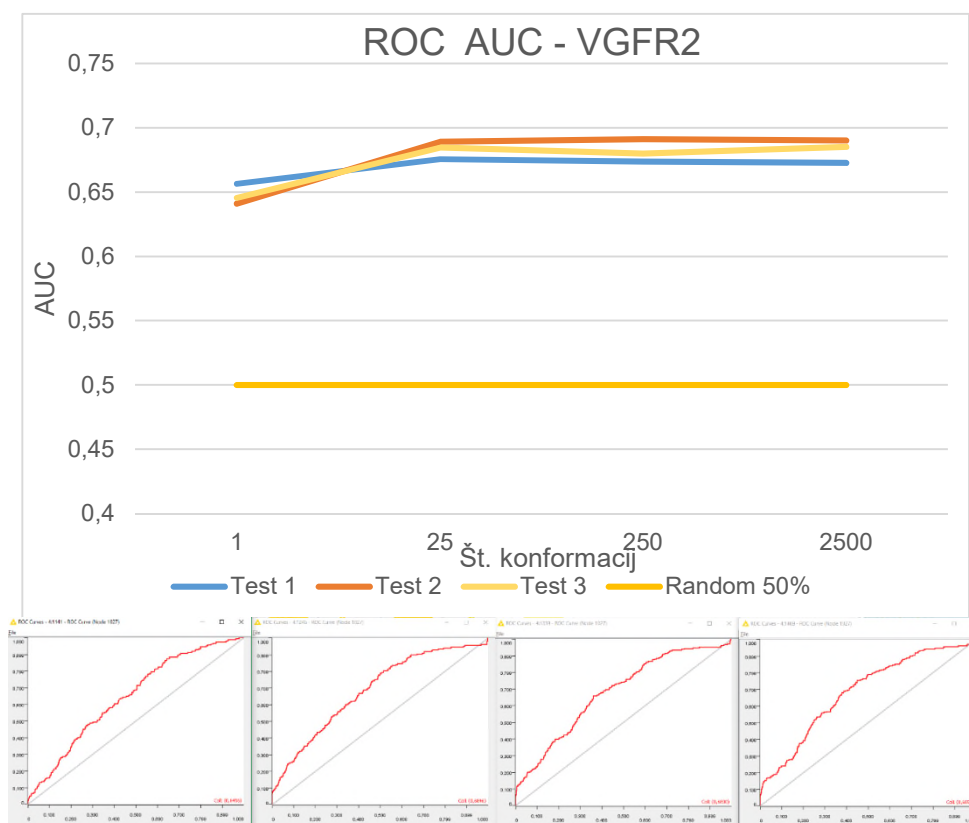
Preglednica XIII V spodnji tabeli najdemo strukturo in deskriptorje kristaliziranih ligandov⁽³⁵⁾

Pripadajoča tarča	TRY1	
Molekulska formula	C ₂₈ H ₃₁ N ₅ O ₆	
Molekulska masa (g/mol)	533,58	
Število atomov	70	
Število vezi	72	
Število aromatskih vezi	19	
Število kiralnih centrov	1	
Kiralni centri	C34	
Število rotacijskih vezi	9	



Slika 30 Razporeditev aktivnih in neaktivnih spojin po številu rotacijskih vezi na levi strani in po MM na desni strani.

5.11 VGRF2 (PDB 2P2I)

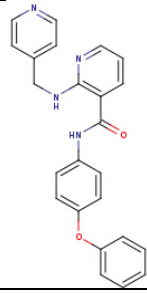


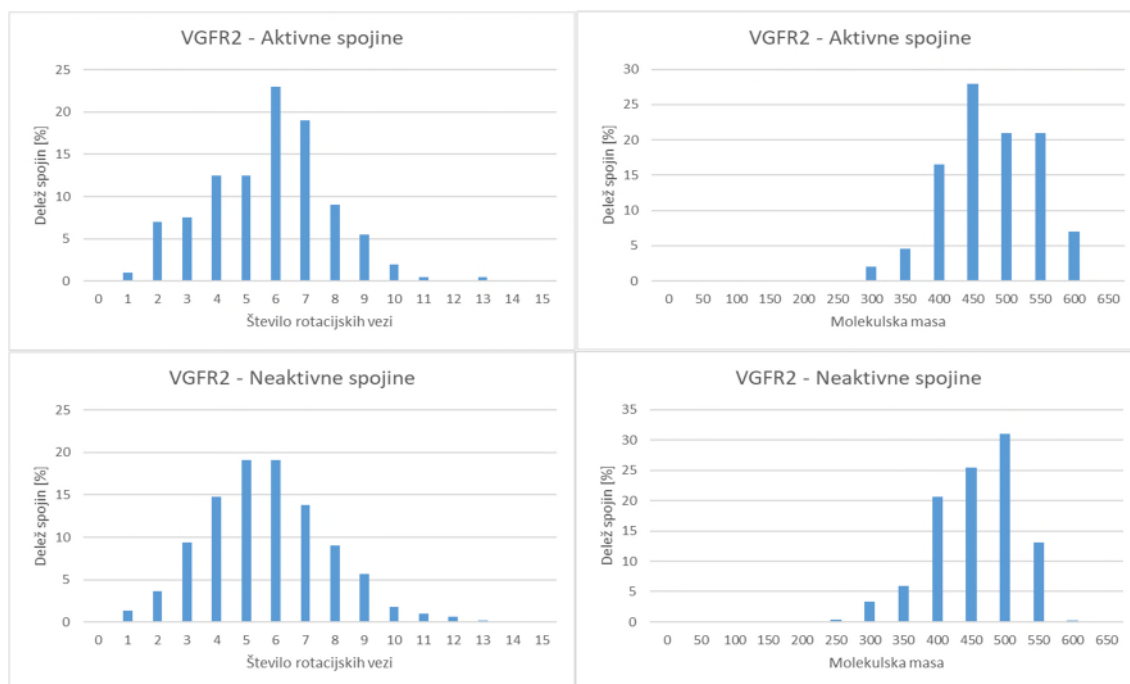
Slika 31 Zgornja slika predstavlja nihanja vrednosti AUC ob spremembi števila konformacij za vse tri teste. Označena je tudi vrednosti 0,5, ki predstavlja naključni model. Na spodnjih slikah so prikazane ROC krivulje za 1, 25, 250 in 2500 konformacij (od leve proti desni).

Do manjšega dviga AUC (manj kot 0,1) pride le ob prvem prehodu. VGRF je še en primer, kjer ROC krivulja narašča minimalno in zelo zgodaj doseže svoj plato. To se zgodi že pri prvem preskoku števila konformacij. Tudi vrednosti AUC niso pretirano visoki in se gibljejo med 0,65 ter 0,68 (Slika 31).

V primerjavi deskriptorjev aktivnih in neaktivnih spojin ne najdemo prevelikih razlik, razen pri MM, ki je pri aktivnih rahlo višja. Če primerjamo MM cele pripadajoče knjižnice spojin (433 g/mol) s kristaliziranimi ligandom lahko vidimo, da ima ligand rahlo manjšo MM (396 g/mol). Tudi pri primerjavi rotacijskih vezi ima knjižnica skoraj za 2 rotacijski vezi manj kot pa pripadajoči ligand. Vezavno mesto je po velikosti med srednje velikimi (Preglednica XIV, Preglednica XVII).

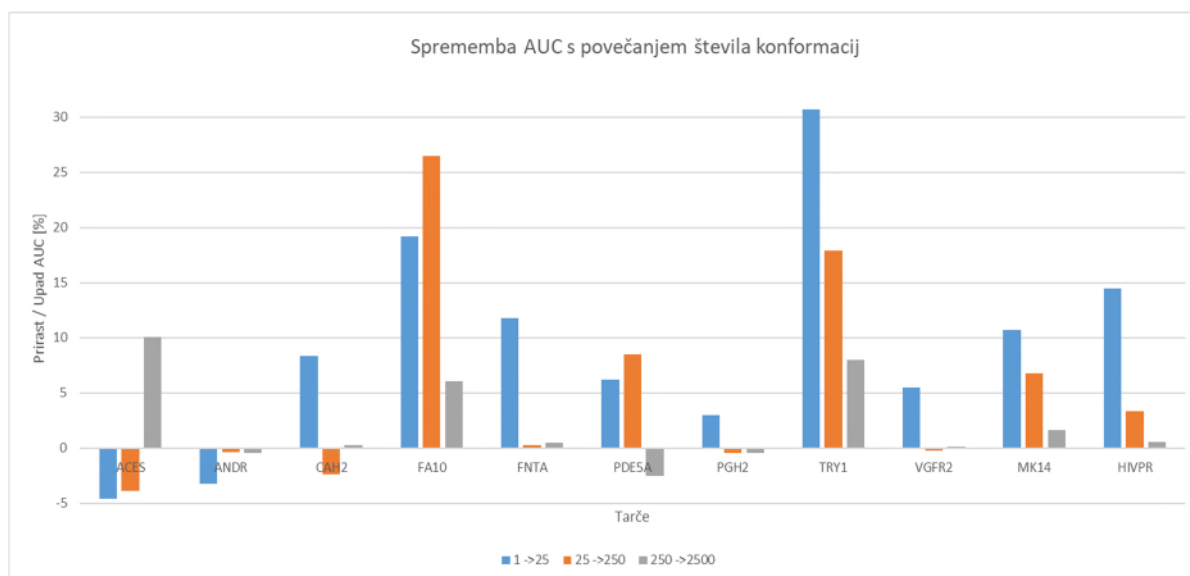
Preglednica XIV V spodnji tabeli najdemo strukturo in deskriptorje kristaliziranih ligandov⁽³⁵⁾

Pripadajoča tarča	VGFR2	
Molekulska formula	C ₂₄ H ₂₀ N ₄ O ₂	
Molekulska masa (g/mol)	396,44	
Število atomov	50	
Število vezi	53	
Število aromatskih vezi	24	
Število kiralnih centrov	0	
Kiralni centri	/	
Število rotacijskih vezi	7	



Slika 32 Razporeditev aktivnih in neaktivnih spojin po številu rotacijskih vezi na levi strani in po MM na desni strani.

5.12 PREGLED SPREMEMB AUC



Slika 33 Graf spremembe vrednosti AUC v odstotkih, pri prehodu iz manjše na večje število konformacij za vsako tarčo.

Pri tvorbi konformacij z Omega smo izbrali naslednje število konformacij: 1, 25, 250, 2500. S temi izbirami zastopamo vse pomembnejše velikostne razrede, ki se uporabljajo v praksi. V zgornjem grafu (Slika 33) lahko vidimo, da v večini primerov (9 od 11) prihaja do povečanja AUC pri 25 konformacijah. Le pri petih primerih naraste AUC tudi pri 250 konformacijah in približno pri 4 primerih naraste AUC pri največjem številu konformacij. A tukaj gre za relativno majhne prirastke v primerjavi z manjšim številom konformacij. Opazimo lahko, da prve tri tarče (ACES, ANDR, CAH2) zelo izstopajo zaradi veliki padcev AUC. Od teh treh izstopa ANDR, ki ima kljub majhnim prirastkom velik AUC, med tem ko ostali dve tarči imata AUC skoraj 0,5.

Ob podrobnejšem pregledu podatkov vidimo, da gre za 3 primere, kjer je vezani ligand majhna in/ali rigidna molekula. Poleg tega imajo pripadajoče aktivne in neaktivne spojine veliko večjo MM in tudi večje število rotacijskimi vezmi. Opazimo tudi, da gre v teh primerih za tarče, ki imajo zelo majhno vezavno mesto (Preglednica XVII). V takih primerih sam protokol sidranja verjetno ni ustrezen.

Druga skupina podatkov vključuje primera FA10 in TRY1. Pri teh dveh tarčah imamo največje prirastke AUC glede na ostale tarče na našem testu. Pri TRY1 vidimo celo zniževanje velikosti prirasta AUC, kar bi bilo pričakovati pri idealnih primerih. Pri TRY1 lahko vidimo večje razlike med deskriptorji znotraj knjižnice kot pri FA10. V obeh primerih

pa so aktivne spojine bolj podobne kristaliziranemu ligandu v primerjavi z neaktivni spojinami. Oba primera spadata med srednje velika oziroma velika vezavna mesta. Vsi ti pogoji bi lahko bili razlog, da sta med redkimi tarčami na testu, ki imata prirastek konformacij tudi pri zadnjem prehodu od 250 na 2500.

V tretjo skupino bi lahko uvrstili preostale tarče, ki imajo srednje velike prirastke pri 25 konformacijah in manjše ali ničte prirastke pri 250 in 2500 konformacijah. V tej skupini so se bolje odrezali primeri s srednjimi ali velikimi vezavnimi mesti. Pri tem malce izstopa MK14, ki ima največje vezavno mesto na testu in ustrezne prehode a vseeno relativno majhne vrednosti AUC.

V primerih, kjer je večja razlika med deskriptorji aktivnih in neaktivnih spojin, so bili rezultati boljši. Sklepamo da je to razlog za večjo diskriminacijo knjižnice in boljši rezultat vezave. Najboljši rezultati so se pojavili, če je bila prisotna večja podobnost z deskriptorji kokristaliziranega liganda.

Ugotovili smo, da so rezultati sidranja boljši, če uporabimo več kot 1 konformacijo za vsako spojino. Samo v enem primeru (ANDR), je bila velikost AUC že pri 1 konformaciji zelo visoka, ki se potem ni spreminjala z dvigom konformacij, kar pripisujemo velikosti vezavnega mesta in lastnostim pripadajoče knjižnice v primerjavi s kristaliziranim ligandom. Pri večini tarč so bili prisotni prirastki tako pri 25 kot tudi 250 konformacijah (Slika 33) a so le-ti pri številu 250 večinoma manjši. Če upoštevamo porabo računskega časa in uporabo knjižnic večjih velikosti, je vsekakor boljša uporaba števila konformacij v velikostnem rangju 25 konformacij. Pri 2500 konformacijah je prirastkov malo in le-ti so tudi zelo majhni ali celo ničti. Pri vsem tem pa moramo upoštevati tudi same specifične lastnosti tarče kot je na primer volumen vezavnega mesta. Pri skrajnih vrednostih se namreč pojavljajo odstopi. Vsekakor je priporočljiva individualna obravnava tarč, ki nam pomaga pri določitvi parametrov sidranja. Če imamo poleg tarče na voljo tudi znani ligand, lahko še z znanjem o lastnostih liganda oblikujemo oz. ustrezno filtriramo knjižnico in tako izboljšamo izid sidranja.

Preglednica XV Preglednica tarč s pripadajočimi vrednostmi AUC vseh treh testov, povprečje vrednosti AUC testov za posamezno število konformacij, sprememba AUC v odstotkih pri prehodu na večje število konformacij, in povprečni RSD testov.

Tarča	Število konformacij	Test 1	Test 2	Test 3	Povprečje	Sprememba AUC (%)	Povprečni RSD (%)
ACES	1	0,5214	0,5674	0,5507	0,5465	/	3,56
	25	0,5300	0,5273	0,5073	0,5215	-4,5685	
	250	0,4843	0,5238	0,4964	0,5015	-3,8412	
	2500	0,5820	0,5486	0,5249	0,5518	10,0366	
ANDR	1	0,7434	0,7537	0,7602	0,7524	/	1,84
	25	0,7110	0,7324	0,7417	0,7284	-3,1985	
	250	0,7073	0,7306	0,7387	0,7255	-0,3890	
	2500	0,7050	0,7282	0,7344	0,7225	-0,4135	
CAH2	1	0,5502	0,6223	0,5633	0,5786	/	3,82
	25	0,6237	0,6456	0,6108	0,6267	8,3132	
	250	0,5976	0,6203	0,6176	0,6118	-2,3722	
	2500	0,6190	0,6113	0,6104	0,6136	0,2833	
FA10	1	0,4407	0,4963	0,4839	0,4736	/	5,61
	25	0,5363	0,5958	0,5622	0,5648	19,2413	
	250	0,6881	0,7587	0,6965	0,7144	26,5006	
	2500	0,7380	0,7925	0,7428	0,7578	6,0654	
FNTA	1	0,6276	0,6398	0,6320	0,6331	/	2,06
	25	0,7240	0,6834	0,7149	0,7074	11,7353	
	250	0,7270	0,6986	0,7022	0,7093	0,2592	
	2500	0,7311	0,7016	0,7049	0,7125	0,4606	

Preglednica XVI Nadaljevanje preglednice tarč s pripadajočimi vrednostmi AUC vseh treh testov, povprečje vrednosti AUC testov za posamezno število konformacij, sprememba AUC v odstotkih pri prehodu na večje število konformacij, in povprečni RSD testov.

Tarča	Število konformacij	Test 1	Test 2	Test 3	Povprečje	Sprememba AUC (%)	Povprečni RSD (%)
MK14	1	0,5125	0,5240	0,5579	0,5315	/	3,95
	25	0,5835	0,5717	0,610	0,5884	10,7125	
	250	0,6271	0,6029	0,6541	0,6280	6,7358	
	2500	0,6357	0,6077	0,6720	0,6385	1,6613	
HIVPR	1	0,5704	0,5813	0,5323	0,5613	/	2,64
	25	0,6558	0,6394	0,6330	0,6427	14,5012	
	250	0,6754	0,6564	0,6606	0,6641	3,3295	
	2500	0,6710	0,6558	0,6768	0,6679	0,5621	
VGFR2	1	0,6565	0,6411	0,6456	0,6477	/	1,17
	25	0,6758	0,6891	0,6846	0,6832	5,4704	
	250	0,6738	0,6911	0,680	0,6816	-0,2244	
	2500	0,6727	0,6901	0,6852	0,6827	0,1516	
TRY1	1	0,4169	0,4549	0,4120	0,4279	/	3,74
	25	0,5548	0,5817	0,5419	0,5595	30,7369	
	250	0,6490	0,6754	0,6552	0,6599	17,9457	
	2500	0,7045	0,7214	0,7112	0,7124	7,9562	
PGH2	1	0,7042	0,7205	0,7086	0,7111	/	3,85
	25	0,7476	0,7573	0,6921	0,7323	2,9860	
	250	0,7427	0,7610	0,6834	0,7290	-0,4506	
	2500	0,7480	0,7499	0,6802	0,7260	-0,4115	
PDE5A	1	0,6935	0,6898	0,7023	0,6952	/	1,18
	25	0,7291	0,7389	0,7469	0,7383	6,1997	
	250	0,7928	0,8138	0,7964	0,8010	8,4925	
	2500	0,7704	0,7980	0,7741	0,7808	-2,5177	

Preglednica XVII Preglednica z volumni in rangi vezavnega mesta, MM liganda in knjižnice spojin ter rotacijske vezi knjižnice spojin.

Ime tarče	Volumen vezavnega mesta (Å)	Zunanja površina – Outer contour (Å)	MM kristaliziranega liganda (g/mol)	Rang po velikosti vezavnega mesta	Rang po velikosti zunanje površine	Povprečna MM aktivnih spojin (g/mol)	Povprečna MM neaktivnih spojin (g/mol)	Število rotacijskih vezi – aktivne spojine	Število rotacijskih vezi – neaktivne spojine
ACES	4505	554	298,81	9	8	439,42	415,35	8,15	6,67
ANDR	4222	349	288,42	10	10	372,16	358,23	2,49	3,31
CAH2	3352	296	149,09	11	11	395,80	381,30	5,44	5,27
FA10	7170	1060	479,98	6	6	514,17	471,42	6,71	6,53
FNTA	10322	2326	606,78	3	1	479,87	454,88	6,73	6,28
PDE5 A	9975	1094	474,58	4	5	444,64	439,87	5,93	5,80
PGH2	5622	415	381,37	8	9	377,12	368,30	3,91	3,82
TRY1	13065	1880	533,58	2	3	482,15	424,94	8,41	6,54
VGFR 2	6705	881	396,44	7	7	451,15	432,35	5,75	5,68
MK14	14314	2172	550,51	1	2	442,78	429,79	5,50	5,45
HIVP R	9582	1359	577,78	5	4	522,40	472,85	9,39	7,38

6 SKLEP

Ker je sidranje eden od najpogostejših pristopov pri virtualnem rešetanju smo v okviru magistrske naloge poskušali ovrednotiti vpliv parametrov priprave konformacijskih knjižnic pri izvedbi eksperimentov sidranja rigidnih ligandov.

Uspešnost sidranja smo ocenili z uporabo ROC krivulj in pri tem smo pričakovali več idealnih primerov prirastka površine pod ROC krivuljami ob večanju števila konformacij v vhodnih knjižnicah spojin. Spoznali smo, da imata tako tarča kot knjižnica spojin pomembno vlogo pri uspešnosti sidranja, pri sami izvedbi sidranja pa moramo tudi prilagoditi sam protokol posamezni tarči. Paziti moramo na ključne parametre, kot je velikost aktivnega mesta ali strukturne posebnosti kristaliziranih ligandov npr. velikost ali sterična rigidnost molekul. Pri tarčah z majhnim vezavnim mestom, so prirastki ROC AUC nižji, kot pri tarčah z večjim vezavnim mestom. Ob kombinaciji majhnega vezavnega mesta in rigidnimi ligandi, pa so ROC AUC prirastki najnižji. Najboljše rezultate lahko opazimo ob uporabi večjih vezavnih mest ter knjižnic spojin v večjo molsko maso ter prisotnostjo rotacijskih vezi kar smo opazili iz izračunov molekulskih deskriptorjev za uporabljene knjižnice spojin. Če se deskriptorji aktivnih spojin v testnih knjižnicah še približajo vrednostim deskriptorjev vezanega liganda pa pričakovano dosežemo najboljše rezultate. V takem primeru program za rešetanje najlažje prepozna aktivne spojine in jih loči od neaktivnih. Kar se tiče števila konformacij, smo lahko ugotovili, da je število 25 izračunanih konformacij najugodnejše oziroma najugodnejši velikostni razred konformacij, saj z izračunom večjega števila konformacij potrebujemo veliko procesorskega časa in pri večini primerov ne dosežemo signifikantnega prirastka k velikosti ROC AUC-ja. Na našem testu je pri tej količini prišlo to povečanja AUC-ja kar pri 9 od 11 tarč.

S to diplomsko nalogo smo ugotovili pomembnost vpliva parametra števila konformacij spojin na izid sidranja ter optimizirali sam protokol priprave konformacijskih knjižnic za izvedbo eksperimentov sidranja.

7 LITERATURA

1. Schneider G, Baringhaus K-H, Kubinyi H. *Molecular Design: Concepts and Applications*. 1 edition. Weinheim: Wiley-VCH, 2008.
2. Andricopulo A, Guido R, Oliva G. Virtual Screening and Its Integration with Modern Drug Design Technologies. *Current Medicinal Chemistry* 2008; 15(1): 37–46.
3. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Reviews Drug Discovery* 2004; 3(11): 935–949.
4. Kearsley SK, Underwood DJ, Sheridan RP, Miller MD. Flexibases: A way to enhance the use of molecular docking methods. *Journal of Computer-Aided Molecular Design* 1994; 8(5): 565–582.
5. Hawkins PCD. Conformation Generation: The State of the Art. *Journal of Chemical Information and Modeling* 2017; 57(8): 1747–1756.
6. Kirchmair J, Wolber G, Laggner C, Langer T. Comparative performance assessment of the conformational model generators omega and catalyst: A large-scale survey on the retrieval of protein-bound ligand conformations. *Journal of Chemical Information and Modeling* 2006; 46(4): 1848–1861.
7. Riniker S, Landrum GA. Better Informed Distance Geometry: Using What We Know to Improve Conformation Generation. *Journal of Chemical Information and Modeling* 2015; 55(12): 2562–2574.
8. Ferrara P, Gohlke H, Price DJ, Klebe G, Brooks CL. Assessing Scoring Functions for Protein–Ligand Interactions. *Journal of Medicinal Chemistry* 2004; 47(12): 3032–3047.
9. Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. *Journal of Medicinal Chemistry* 2002; 45(12): 2615–2623.
10. Schwab CH. Conformations and 3D pharmacophore searching. *Drug Discovery Today: Technologies* 2010; 7(4): e245–e253.
11. Friedrich N, Meyder A, Kops CDB, Sommer K, Flachsenberg F, Rarey M, Kirchmair J. High-Quality Dataset of Protein-Bound Ligand Conformations and Its Application to Benchmarking Conformer Ensemble Generators. 2017.
12. Friedrich N, Kops CDB, Flachsenberg F, Sommer K, Rarey M, Kirchmair J. Benchmarking Commercial Conformer Ensemble Generators. 2017: 2719–2728.
13. Hawkins PCD, Skillman AG, Warren GL, Ellingson BA, Stahl MT. Conformer generation with OMEGA: Algorithm and validation using high quality structures from the protein databank and cambridge structural database. *Journal of Chemical Information and Modeling* 2010; 50(4): 572–584.
14. <https://docs.eyesopen.com/omega/omegatheory.html#stereochemistry-enumeration>. OMEGA Theory — OMEGA, v3.0.1.2., dostopano avgusta 2018
15. FRED and HYBRID — OEDocking, v3.2.0.2. https://docs.eyesopen.com/oedocking/fred_and_hybrid_theory.html., dostopano avgusta 2018
17. Huang N, Shoichet BK, Irwin JJ. Benchmarking sets for molecular docking. *Journal of Medicinal Chemistry* 2006; 49(23): 6789–6801.
18. Martin YC. Let's not forget tautomers. *Journal of Computer-Aided Molecular Design* 2009; 23(10): 693.
19. https://docs.eyesopen.com/oedocking/fred.html#cmdoption-fred-dock_resolution., dostopano avgusta 2018

20. <http://www.daylight.com/dayhtml/doc/theory/theory.smiles.html>., dostopano avgusta 2018
21. Triballeau N, Acher F, Brabet I, Pin J-P, Bertrand H-O. Virtual Screening Workflow Development Guided by the “Receiver Operating Characteristic” Curve Approach. Application to High-Throughput Docking on Metabotropic Glutamate Receptor Subtype 4. *Journal of Medicinal Chemistry* 2005; 48(7): 2534–2547.
22. KNIME Essentials. <https://www.knime.com/>, dostopano avgusta 2018
23. Dvir H, Wong DM, Harel M, Barril X, Orozco M, Luque FJ, Muñoz-Torrero D, Camps P, Rosenberry TL, Silman I, Sussman JL. 3D structure of Torpedo californica acetylcholinesterase complexed with huprine X at 2.1 Å resolution: Kinetic and molecular dynamic correlates. *Biochemistry* 2002; 41(9): 2970–2981.
24. Pereira de Jésus-Tran K, Côté P-L, Cantin L, Blanchet J, Labrie F, Breton R. Comparison of crystal structures of human androgen receptor ligand-binding domain complexed with various agonists reveals molecular determinants responsible for binding affinity. *Protein Science: A Publication of the Protein Society* 2006; 15(5): 987–999.
25. Håkansson K, Liljas A. The structure of a complex between carbonic anhydrase II and a new inhibitor, trifluoromethane sulphonamide. *FEBS letters* 1994; 350(2–3): 319–322.
26. Fujimoto T, Imaeda Y, Konishi N, Hiroe K, Kawamura M, Textor GP, Aertgeerts K, Kubo K. Discovery of a tetrahydropyrimidin-2(1H)-one derivative (TAK-442) as a potent, selective, and orally active factor Xa inhibitor. *Journal of Medicinal Chemistry* 2010; 53(9): 3517–3531.
27. Hast MA, Fletcher S, Cummings CG, Pusateri EE, Blaskovich MA, Rivas K, Gelb MH, Van Voorhis WC, Sebt SM, Hamilton AD, Beese LS. Structural basis for binding and selectivity of antimalarial and anticancer ethylenediamine inhibitors to protein farnesyltransferase. *Chemistry & Biology* 2009; 16(2): 181–192.
28. Sung B-J, Hwang KY, Jeon YH, Lee JI, Heo Y-S, Kim JH, Moon J, Yoon JM, Hyun Y-L, Kim E, Eum SJ, Park S-Y, Lee J-O, Lee TG, Ro S, Cho JM. Structure of the catalytic domain of human phosphodiesterase 5 with bound drug molecules. *Nature* 2003; 425(6953): 98–102.
29. Wang JL, Limburg D, Graneto MJ, Springer J, Hamper JRB, Liao S, Pawlitz JL, Kurumbail RG, Maziasz T, Talley JJ, Kiefer JR, Carter J. The novel benzopyran class of selective cyclooxygenase-2 inhibitors. Part 2: the second clinical candidate having a shorter and favorable human half-life. *Bioorganic & Medicinal Chemistry Letters* 2010; 20(23): 7159–7163.
30. Sherawat M, Kaur P, Perbandt M, Betzel C, Slusarchyk WA, Bisacchi GS, Chang C, Jacobson BL, Einspahr HM, Singh TP. Structure of the complex of trypsin with a highly potent synthetic inhibitor at 0.97 Å resolution. *Acta Crystallographica. Section D, Biological Crystallography* 2007; 63(Pt 4): 500–507.
31. Hodous BL, Geuns-Meyer SD, Hughes PE, Albrecht BK, Bellon S, Bready J, Caenepeel S, Cee VJ, Chaffee SC, Coxon A, Emery M, Fretland J, Gallant P, Gu Y, Hoffman D, Johnson RE, Kendall R, Kim JL, Long AM, Morrison M, Olivieri PR, Patel VF, et al. Evolution of a highly selective and potent 2-(pyridin-2-yl)-1,3,5-triazine Tie-2 kinase inhibitor. *Journal of Medicinal Chemistry* 2007; 50(4): 611–626.
32. Murali Dhar TG, Wroblewski ST, Lin S, Furch JA, Nirschl DS, Fan Y, Todderud G, Pitt S, Doweiko AM, Sack JS, Mathur A, McKinnon M, Barrish JC, Dodd JH, Schieven GL, Leftheris K. Synthesis and SAR of p38α MAP kinase inhibitors based on heterobicyclic scaffolds. *Bioorganic & Medicinal Chemistry Letters* 2007; 17(18): 5019–5024.
33. Specker E, Böttcher J, Lilie H, Heine A, Schoop A, Müller G, Griebenow N, Klebe G. An old target revisited: two new privileged skeletons and an unexpected binding mode for

- HIV-protease inhibitors. *Angewandte Chemie (International Ed. in English)* 2005; 44(20): 3140–3144.
34. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *Journal of Medicinal Chemistry* 2012; 55(14): 6582–6594.
 35. RCSB PDB: <https://www.rcsb.org/>, dostopano avgusta 2018
 36. DUD-E: <http://dude.docking.org/targets>, avgust 2018