Razvoj 3D-QSAR modela za tip I inhibitore Src kinaze

Stošić, Lidija

Master's thesis / Diplomski rad

2021

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Rijeka / Sveučilište u Rijeci**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:193:795263

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-29**

Image not found or type unknown

Repository / Repozitorij:

Repository of the University of Rijeka, Faculty of Biotechnology and Drug Development - BIOTECHRI Repository

zir.nsk.hr

UNIRI DIGITALNA KNJIŽNICA

Image not found or type unknown

SVEUČILIŠTE U RIJECI

ODJEL ZA BIOTEHNOLOGIJU

Diplomski sveučilišni studij

Istraživanje i razvoj lijekova

Lidija Stošić

Razvoj 3D-QSAR modela za tip I inhibitore Src kinaze

Diplomski rad

Rijeka, 2021.

SVEUČILIŠTE U RIJECI

ODJEL ZA BIOTEHNOLOGIJU

Diplomski sveučilišni studij

Istraživanje i razvoj lijekova

Lidija Stošić

Razvoj 3D-QSAR modela za tip I inhibitore Src kinaze

Diplomski rad

Rijeka, 2021.

Mentor rada: izv. prof. dr. sc. Sanja Koštrun

UNIVERSITY OF RIJEKA

DEPARTMENT OF BIOTECHNOLOGY

Graduate program

Drug research and development

Lidija Stošić

Development of 3D-QSAR models for Type I Src kinase inhibitors

Master's thesis

Rijeka, 2021.

Mentor: Sanja Koštrun, PhD, assoc. prof.

Diplomski rad obranjen je dana 25.10.2021.

pred povjerenstvom:

1. doc. dr. sc. Željko Svedružić (predsjednik)

2. doc. dr. sc. Jelena Ban (članica)

3. izv. prof. dr. sc. Sanja Koštrun (mentorica)

Rad ima 69 stranica, 31 sliku, 6 tablica i 31 literaturnih navoda.

Diplomski rad je izrađen u znanstveno-istraživačkom centru Fidelta d.o.o., Zagreb, pod mentorstvom izv. prof. dr. sc. Sanje Koštrun.

ZAHVALA

Ovim putem želim izraziti veliku zahvalu svojoj mentorici izv. prof. dr. sc. Sanji Koštrun na strpljenju, znanju, razumijevanju, trudu, pomoći i vremenu prilikom izrade ovog diplomskog rada. Zahvaljujem Vama i Fidelti d.o.o. na ukazanoj prilici tijekom ovog osjetljivog razdoblja pandemije.

Želim se posebno zahvaliti mojoj obitelji i bližnjima zato što su uvijek bili uz mene i pružili mi podršku tijekom studija.

SAŽETAK

ZNAČAJ: Src kinaza je tirozin kinaza koja sudjeluje u prijenosu signala povezanih s brojnim staničnim procesima, kao što su proliferacija, diferencijacija, ili adhezija. Src kinaza poželjna je terapijska meta jer ima prekomjernu ekspresiju, visoku aktivnost i jak utjecaj na razvoj metastaza u raznovrsnim tipovima karcinomima.

METODE: Razvili smo kvantitativne i kvalitativne 3D-QSAR modele koji se temelje na molekulskim poljima koristeći literaturno dostupne Src inhibitore tipa I.

REZULTATI: Razvijena su 4 odvojena modela za 4 klase strukturno sličnih analoga; *PP2 analozi, Dasatinib analozi, Saracatinib analozi, Zajednički skup,* od kojih se model za skup *Saracatinib analoga* pokazao kao najbolji model visoke točnosti i stabilnosti s prihvatljivom prediktivnom moći (PLS: q^2=0.63; SVM: q^2=0.60) i dobrim koeficijentom korelacije (PLS: r^2=0.92, r_{tes}^2=0.85; SVM: r^2=0.95, r_{test}^2=0.28). Dodatno, cilj je bio usporediti pristup temeljen na ligandima (engl. LBDD, *ligand-based drug design*) s pristupom temeljenim na strukturi (engl. SBDD, *structure-based drug design*). Spojevi korišteni za izgradnju 3D-QSAR modela uklopljeni su u DFG-"in" konformaciju Src kinaze koristeći literaturno dostupne kristalne strukture, te su uklopljene komplekse minimizirani MM-GBSA metodom. Niska korelacija (r=0.48) eksperimentalnih aktivnosti i Gibbsove energije interakcije dobivene MM-GBSA metodom, ukazuje kako pristup temeljen na strukturi ne poboljšava predviđanje aktivnosti Src inhibitora u odnosu na pristup temeljen na ligandu. U posljednjem koraku dobiveni rezultati su uspoređeni s prethodno objavljenim studijama te je pokazano da je predstavljeni model za saracatinib skup podataka bolji od do sada objavljenih studija te da MM-GBSA metoda nije korisna za efikasno predviđanje aktivnosti.

ZAKLJUČAK: Opisali smo odnose između strukture i aktivnosti za tip I inhibitore Src kinaza koristeći robusnije i pouzdanije modele s boljom

prediktivnom moći za dizajn, filtriranje i prioritiziranje virtualnih biblioteka potencijalnih SRC inhibitora. Predstavljeni rezultati mogu se korsititi za planiranje sinteze novih i aktivnijih inhibitora.

KLJUČNE RIJEČI: Kinazni inhibitori, SRC kinaza, CADD, 3-D QSAR, Molekulsko uklapanje

ABSTRACT

BACKGROUND: Src kinase is a tyrosine kinase that participates in the transmission of signals associated with numerous cellular processes such as proliferation, differentiation, or adhesion. Src is a desirable therapeutic target because it is overexpressed and highly active in various cancers, and it influences the development of metastasis as well.

METHODS: Field-based quantitative and qualitative 3D-QSAR models were developed for literature-available Type I Src inhibitors.

RESULTS: Four separate models have been developed for 4 classes of structurally similar analogues; *PP2 analogues, Dasatinib analogues, Saracatinib, analogues, Common set*, of which the model for the *Saracatinib analogue* set proved to be the best model of high accuracy and stability with acceptable predictive power (PLS: q^2=0.63; SVM: q^2=0.60), and good correlation coefficient (PLS: r^2=0.92, r_{tes}^2=0.85; SVM: r^2=0.95, r_{test}^2=0.28). Additionally, the aim was to compare the ligand-based approach (LBDD) with the structure-based drug design (SBDD). Compounds used to build the 3D-QSAR model were incorporated into the DFG-"in" conformation of Src kinase literature-available crystal structures, and the incorporated complexes were minimized by the MM-GBSA method. The low correlation (r=0.48) of the experimental activities and the Gibbs interaction energy obtained by the MM-GBSA method indicates that the structure-based approach does not improve the prediction of Src inhibitor activity compared to the ligand-based approach. Finally, the obtained results were compared with previously published studies and we concluded that the saracatinib data set is better than the previously published studies, and that the MM-GBSA method is not useful for an efficient prediction of activity.

CONCLUSION: We described the structure-activity relationships for type I Src kinase inhibitors using more robust and reliable models with better predictive capability; design; filtering and prioritization of virtual libraries

of potential Src inhibitors. The presented results can be used to plan the synthesis of new and more active inhibitors.

KEY WORDS: Kinase inhibitors, Src kinase, CADD, 3D QSAR, Molecular docking

SADRŽAJ

1. UVOD 1

1.1. TIROZIN KINAZE 1
1.1.2. RTK I NRTK – GRAĐA I MEHANIZAM DJELOVANJA 2
1.2. NRTK SRC KINAZE 4
1.2.1. FUNKCIJA NRTK SRC KINAZA 4
1.2.2. STRUKTURNE KARAKTERISTIKE NRTK SRC KINAZA 5
1.2.3. C-SRC KINAZE 6
1.2.3.1. Struktura c-Src kinaze 7
1.2.3.2. Regulacija aktivnosti c-Src kinaze 8
1.2.3.3. Uloga c-Src kinaze u regulaciji staničnih procesa i nastajanju tumora 10
1.3. TIROZIN KINAZE KAO BIOLOŠKE METE U LIJEČENJU KARCINOMA 11
1.3.1. ZNAČAJKE ATP VEZNOG MJESTA 11
1.4. INHIBITORI TIROZIN KINAZA 13
1.4.1. NEKOVALENTNI INHIBITORI 13
1.4.2. KOVALENTNI INHIBITORI 16
1.4.3. REZISTENCIJA I SELEKTIVNOST INHIBITORA SRC KINAZA 17
1.5. *IN SILICO* DIZAJN NOVIH LIJEKOVA 17
1.5.1. DIZAJN LIJEKOVA PREMA STRUKTURI (SBDD) 21
1.5.1.1. Molekulsko uklapanje 22
1.5.2. DIZAJN LIJEKOVA PREMA LIGANDU (LBDD) 26
1.5.2.1. QSAR 27
1.5.2.2. Analiza temeljena na molekulskim poljima 29
1.5.2.2.1. XED polje sile 30
1.5.2.2.2. Izračunavanje polja za procjenu molekularnih interakcija 31

2. CILJ RADA 35

3. MATERIJALI I METODE 36

4. REZULTATI I RASPRAVA 37

4.1. KVANTITATIVNI 3D-QSAR MODELI 38
4.2. FARMAKOFORNI MODELI 50
4.3. KLASIFIKACIJSKI 3D-QSAR MODELI 54

5. ZAKLJUČAK 64

6. LITERATURA 66

7. ŽIVOTOPIS 70

1. UVOD

Karcinom predstavlja veliki zdravstveni problem širom svijeta od kojeg svake godine obolijeva nekoliko milijuna ljudi. Među vodećim je uzrocima smrti, a potreba za brzi razvoj djelotvornog lijeka sve je veća. Kao privlačne mete lijekova pokazale su se Src kinaze, iz razloga što je njihova prekomjerna ekspresija i visoka aktivnost utvrđena u širokom spektru karcinom kod ljudi. Za brži razvoj i otkrivanje novih inhibitora Src kinaza koriste se, između ostalog, i *in silico* metode.

1.1. Tirozin kinaze

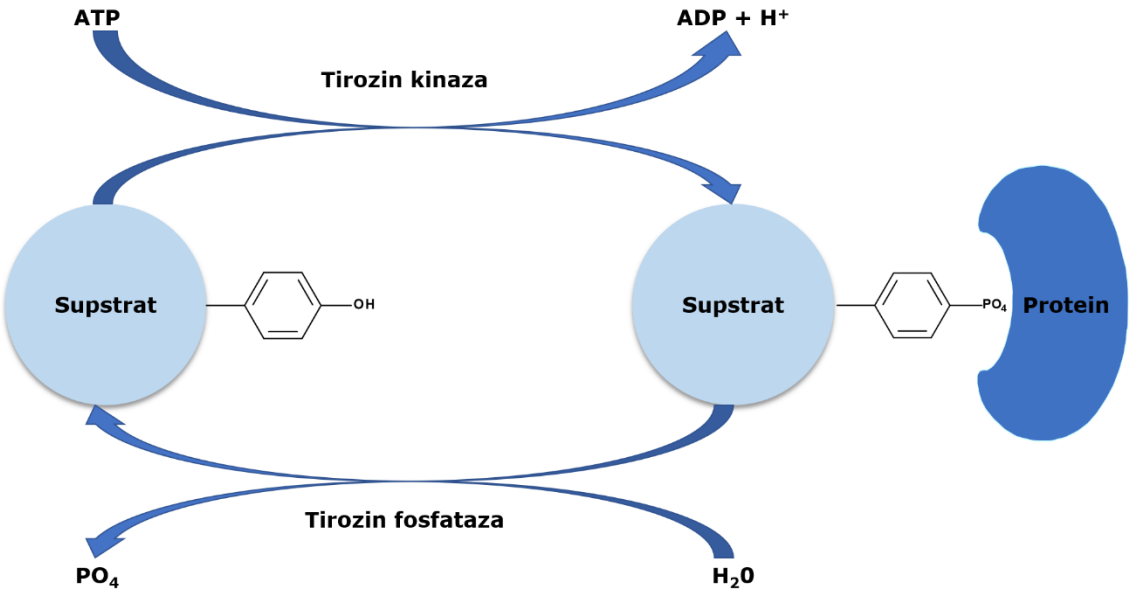
Proteini tirozin kinaze velika su i strukturno raznolika grupa od 90 enzima koji se nalaze u humanom genomu, a dijele se na 58 receptorskih tirozin kinaza (RTK) (npr. Receptor za epidermalni faktor rasta (EGFR), trombocitni faktor rasta (PDGFR), receptor vaskularnog endotelnog faktora rasta (VEGF), receptor inzulinskog receptora (IR)…), te na 32 ne-receptorske tirozin kinaze (NRTK) (npr. SRC, ABL, FAK, JAK…) (1). RTK su transmembranski proteini koji se obično nalaze kao monomeri i imaju važnu ulogu u regulaciji staničnog rasta, diferencijaciji, migraciji i preživljavanju. NRTK su citoplazmatski proteini s funkcijom prijenosa unutarstaničnih signala koji potječu od izvanstaničnog receptora te reguliraju veliki broj staničnih funkcija, kao što su stanično preživljavanje, dioba ili razmnožavanje i adhezija, ekspresija gena, imunološki odgovor, itd. (2). Svi enzimi posjeduju katalitičku domenu, samostalnu podjedinicu od 250-300 aminokiselina, koja sadrži šesnaest visoko sačuvanih elemenata linearne sekvence od kojih se nekoliko pokazalo važnim za katalitičku aktivnost te domene (3).

Tirozin kinaze imaju važnu ulogu u signalnim kaskadama, u različitim biološkim procesima te u odgovoru na unutarnje i vanjske podražaje. Unatoč strogoj regulaciji u stanicama, moguće su pojave mutacija,

prekomjerne ekspresije i autokrine-parakrine stimulacije, što dovodi do malignosti. Selektivnim inhibitornima tirozin kinaze moguće je blokirati onkogenu aktivnost u stanicama raka, što predstavlja obećavajući pristup za terapiju. Također, tirozin kinaze su mete za angiogenezu, koja je glavni faktor u rastu i proliferaciji stanica raka (1).

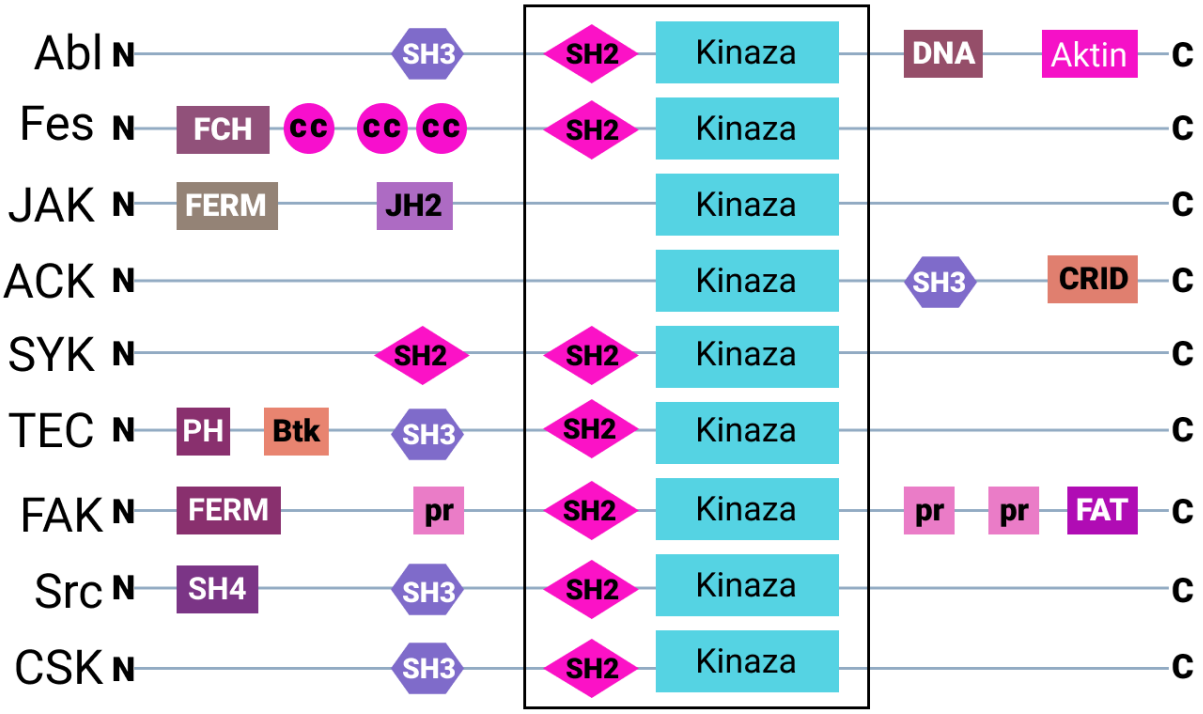
1.1.2. RTK i NRTK – građa i mehanizam djelovanja

Tirozin kinaze su enzimi koji selektivno fosforiliraju, tj. prenose γ-fosfatazu ATP-a na hidroksilnu skupinu tirozina na supstratu što dovodi do aktivacije ili inaktivacije proteina (Slika 1.) (4). RTK strukturno su podijeljene na izvanstaničnu domenu koja se sastoji od jedne ili više domena za vezanje liganda, transmembransku zavojnicu koja prolazi plazma membranu i unutarstaničnu podjedinicu koja posjeduje tirozin kinaznu aktivnost. RTK prenose izvanstanične signale u citoplazmu, a do njihove aktivacije dolazi kada se na izvanstaničnu domenu veže ligand. Ligandi su izvanstanične signalne molekule koje potiču dimerizaciju receptora (izuzev inzulinskog receptora) i njihovu stabilizaciju postižu različito. Npr. hormon rasta i receptor hormona rasta tvore 1:2 kompleks gdje se jedan ligand veže za dva receptora, dok VEGF i VEGFR tvore 2:2 kompleks gdje se dva liganda vežu istodobno za dva receptora. Nakon fosforilacije aktivirana petlja u katalitičkoj domeni TK mijenja konformaciju iz zatvorene u otvorenu i time je omogućen pristup ATP-u i supstratu. Tada dolazi do prijenosa γ-fosfataze iz Mg-ATP kompleksa na hidroksilnu skupinu tirozina na supstratu i na staničnim proteinima koji sudjeluju u transdukciji signala (1).



Slika 1. Signalizacijski put tirozin kinaze i fosfataze. Fosforilacija supstrata tirozin kinazom (gore) i defosforilacija tirozin fosfatazom (dolje). Po uzoru na (4).

NRTK se, zbog sličnosti u sekvenci, klasificiraju u devet podskupina, gdje im je zajednička kinazna domena poznata još kao SH1 (Slika 2.). Također, posjeduju nekoliko dodatnih signalnih domena ili domene koje imaju protein-protein interakcije (npr. SH2, SH3 i PH domena). NRTK sastoje se od približno 300 podjedinica, gdje N-terminalni kraj ima pet β-ploča i jednu α-zavojnicu, a velika citoplazmatska C-terminalna domena ima uglavnom α-zavojnice. ATP se veže između dvije domene, a tirozinska sekvenca proteinskog supstrata povezuje se sa ostacima C-terminalne domene. Mehanizam aktivacije NRTK složeniji je zbog heterologne protein-protein interakcije koja je potrebna kako bi došlo do transfosforilacije (2, 5, 8).



Slika 2. Podjela i građa NRTK. N-amino terminalni kraj, SH3: SRC 3 homologna domena, SH2: SRC 2 homologna domena, kinaza: katalitička domena (SH1), DNA: domena koja veže DNA, aktin: domena koja veže aktin, FCH: homologna domena proteina u interakciji s Fes/Fer/Cdc-42, CC: motiv snop zavojnica, FERM: domena četiri točke – ezrin, radixin, mosein, JH2: Janus homologna domena (domena pseudokinaze), CRIB: Cdc42/Rac-interaktivna domena, PH: Pleckstrin homologna domena, Btk: motiv cinkovog prstena tipa Btk, pr: prolinom bogata regija, FAT: domena koja cilja fokalnu adheziju, SH4: SRC homologna 4 domena, C: C-terminalni kraj. Po uzoru na (2).

1.2. NRTK SRC kinaze

c-Src kinaza, poznata kao proto-onkogen c-Src kinaza, član je obitelji Src tirozin kinazne (engl. *Src family kinases*, SFK). SFK uključuje nekoliko srodnih kinaza: Blk, Fyn, Hck, Fgr, Lyn, Lck, c-Yes, Yrk, Frk (poznata kao Rak), Srm i c-Src.

1.2.1. Funkcija NRTK SRC kinaza

Zbog svoje prisutnosti u različitim tkivima, SRC kinaze uključene su u mnoštvo staničnih mehanizama poput diferencijacije, proliferacije,

apoptoze, reorganizacije citoskeleta, regulacije transkripcije te migracije, a sudjeluju i u pravilnom funkcioniranju imunosnog i živčanog sustava (2, 5). Igraju važnu ulogu u staničnoj adheziji, morfologiji i pokretljivosti stanica te resorpciji kosti. Kontrolirane su receptorskim protein-tirozin kinazama, integrin receptorima, G-protein združenim receptorima (GPCR), receptorima citokina, antigenima i Fc-združenim receptorima te receptorima steroidnih hormona (2, 7).

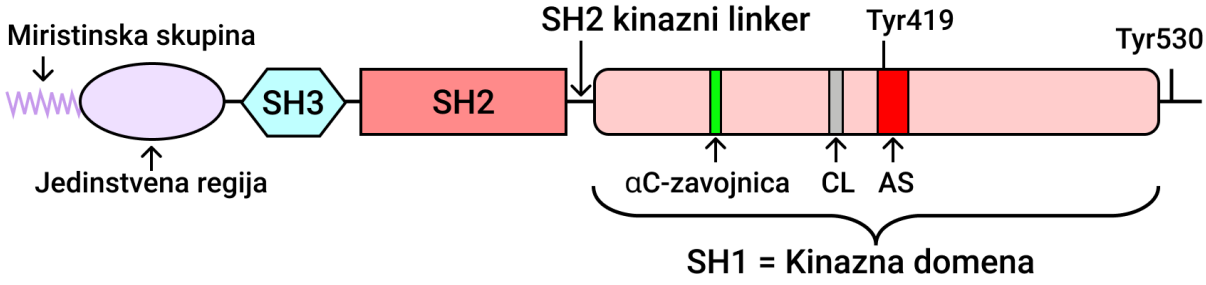
1.2.2. Strukturne karakteristike NRTK SRC kinaza

Proteini SFK skupine imaju sličnu strukturu i prema redoslijedu od N-amino do C-karboksi kraja sadrže šest funkcionalnih domena:

1) **kratko membransko sidro tj. SH4 očuvana miristinska N-terminalana domena**. Aminokiselinski ostataci sadrže signale za modifikaciju masnih kiselina. Glicin na amino kraju kovalentno je vezan za miristilin, a pomoću njega se SRC kinaze vežu na membranu. Ne-miristilizirane SRC kinaze ne mogu se vezati za membrane.
2) loše očuvana **"jedinstvena" domena**. Funkcija joj nije poznata u potpunosti, a razlikuje se između članova obitelji.
3) **SH3 domena** sadrži približno 60 aminokiselina i nalazi se u mnogim signalnim molekulama. Ova domena je ne-katalitička i sudjeluje u protein-protein interakcijama na specifične sekvence proteina bogate prolinom. Neophodna je za interakcije s proteinskim supstratima, a također osigurava intramolekularne veze koje kontroliraju katalitičku aktivnost, lokalizaciju proteina u stanici i povezanost s proteinskim metama. SH3 domena SRC ima kuglastu strukturu kojoj je jedna strana blago hidrofobna i sadrži skupine kiselih ostataka.
4) **SH2 domena** sadrži približno 100 aminokiselina i veže se za fosforilirane ostatke $Tyr^{527/530}$ (527 u pilećoj, 530 u humanoj SRC

kinazi) na vlastitoj C-terminalnoj regulatornoj domeni ili na drugim proteinima. SH2 domena svakog člana SFK ima različiti afinitet prema peptidima zbog različitog slijeda aminokiselina, a tri do pet aminokiselinskih ostataka nakon tirozina određuje specifičnosti pojedinih domena.

5) **kinazna domena (SH1)** je najočuvanija domena u svim TK. Ova domena odgovorna je za katalitičku aktivnost molekule. Sadrži mnoštvo poddomena visoke homologije u serin-treonin kinazama te mjesto za vezanje specifičnog supstrata i hidrofobni džep za vezanje ATP-a.
6) kratka regulatorna **C-terminalna domena** ima značajnu ulogu u regulaciji aktivnosti SRC kinaza. Sve SFK imaju C-terminalno područje od 15-19 aminokiselinskih ostataka sa očuvanim ostatkom $Tyr^{527/530}$ čija fosforilacija dovodi do vezanja za SH2 domenu. Time se postiže "zatvorena" konformacija c-Src kinaze i rezultira inhibicijom aktivnosti (Slika 3.) (2, 6, 7, 8).



Slika 3. Organizacija c-Src kinaze. Po uzoru na (9).

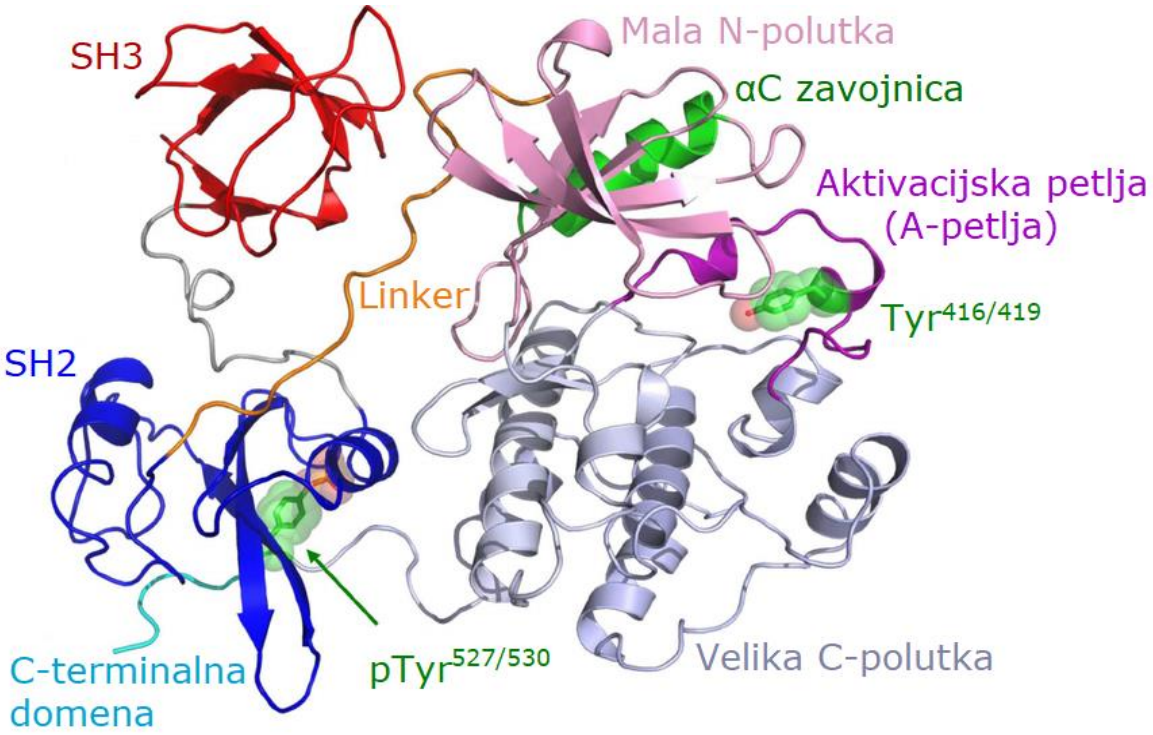
1.2.3. c-Src kinaze

Prvo otkriveni proto-onkogen i najviše istraživana NRTK je c-Src kinaza (6). Pileća c-Src kinaza građena je od 533 aminokiselina, dok humana sadrži tri dodatne aminokiseline (ukupno 536) koje su umetnute na dva mjesta unutar jedinstvene domene. Pozitivno regulatorno mjesto autofosforilacije u c-Src kinazi je na $Tyr^{416/419}$ u katalitičkoj podjedinici

gdje dolazi do aktivacije, a negativno regulatorno fosforilacijsko mjesto je na $Tyr^{527/530}$ u C-terminalnoj domeni gdje dolazi do inhibicije aktivnosti (10).

1.2.3.1. Struktura c-Src kinaze

c-Src kinaza posjeduje ukupno sedam funkcionalnih domena, a dodatna je SH2 linker domena od 14 aminokiselina koja spaja SH2 i SH3 podjedinice s kinaznom domenom (Slika 3.) (6). Prema obliku SH3 domena je β-bačva s pet antiparalelnih β-nabranih ploča i dvije istaknute petlje (RT i n-Src). Površina joj je bogata aromatskim i hidrofobnim ostacima koji prepoznaju specifične sekvence bogate prolinom i tvore zavojnice poliprolina tipa II (6, 9). SH2 domena sastoji se od tročlane β-ploče s jednom α-zavojnicom sa svake strane. Ovi elementi sekundarne strukture tvore dva džepa - jedan za prepoznavanje i koordiniranje fosfotirozina te drugi za vezanje jednog ili više hidrofobnih ostataka C-terminalno od fosfotirozina. Pojedini proteini, koji stupaju u interakciju sa SH2 sadrže optimalno vezujuću strukturu aminokiselinskog slijeda pYEEI. Premda, ne moraju svi imati fosforiliranu sekvencu. Kinazna domena podijeljena je u dvije polutke - malu N-terminalnu i veliku C-terminalnu. Polutke su povezane zglobnom (engl. *hinge*) regijom. Strukturne podjedinice N-polutke sastoje se od peteročlane antiparalelne β-ploče, petlje koja veže nukleotide (P-petlja ili G-bogata petlja) gdje se odvija transfer fosfata putem očuvane sekvence bogate glicinom koja omogućava interakciju s γ-fosfatom molekule ATP-a i jedne α-zavojnice (αC zavojnica). C-polutka je pretežito α-zavojita s aktivacijskom petljom (A-petlja) te uključuje katalitičku domenu koja se smjestila između velike i male polutke. A-petlja bitan je dio u kinaznoj domeni jer može ometati vezanje ATP-a i/ili supstrata. Također, sadržava regulatorni tirozinski ostatak ($Tyr^{416/419}$) bitan za stabilizaciju konformacije. Uzvodno od A-petlje nalazi se visoko očuvani DFG motiv (aspartat-fenilalanin-glicin) (Slika 4.) (6, 7, 9, 10).



Slika 4. Sekundarna struktura c-Src kinaze. Po uzoru na (11).

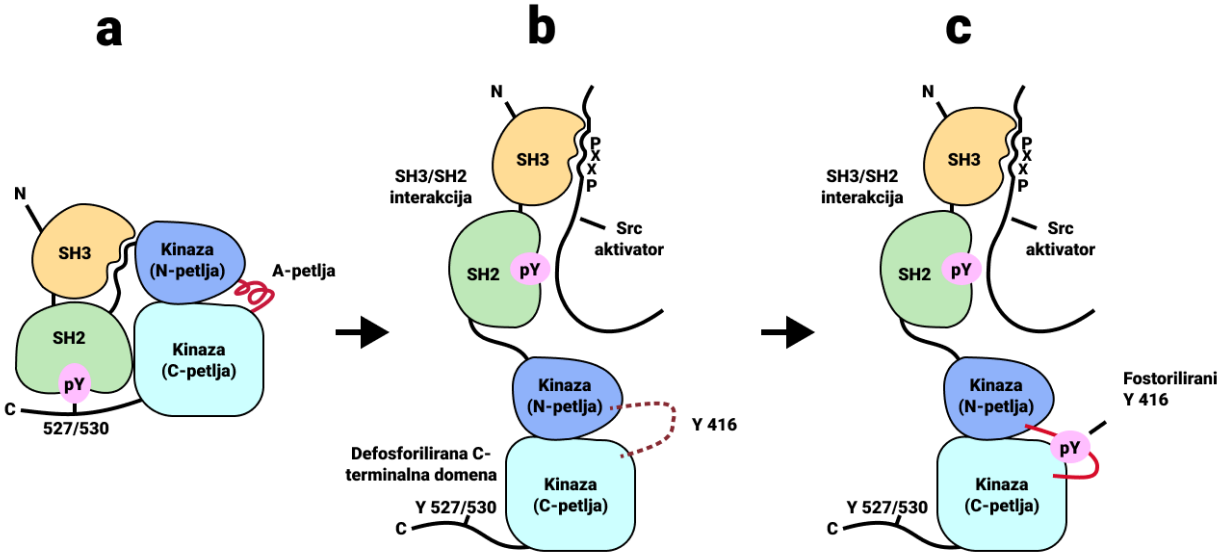
Jedna od funkcija SH2 i SH3 domena je da intramolekularnim interakcijama utječu na djelovanje i aktivnost c-Src kinaze. Također, pomicanjem tih domena aktivira se enzim. Njihova prisutnost u proteinskoj građi omogućava interakciju s SH2 i SH3 domenama Src-kinaze i time potiče migraciju enzima unutar stanice (7, 9).

1.2.3.2. Regulacija aktivnosti c-Src kinaze

Kinaza c-Src, kao što je ranije navedeno, može zauzeti aktivnu i neaktivnu konformaciju. Kod aktivne tj. "otvorene" enzimske konformacije najprije dolazi do defosforilacije $Tyr^{527/530}$ na C-terminalnom kraju koja narušava intramolekularnu interakciju sa SH2 domenom. SH2 se oslobađa i takva otvorena konformacija omogućava autofosforilaciju $Tyr^{416/419}$ na aktivacijskoj petlji. Navedeno rezultira pomicanjem $Tyr^{416/419}$, olakšanim vezanjem supstrata i aktivacijom c-Src. Zatim, αC zavojnica zauzima orijentaciju prema aktivnom mjestu, a s aminokiselinskim ostatkom

glutamata ($Glu^{310/313}$) tvori solni most s lizinom ($Lys^{295/298}$). Solni most služi za pozicioniranje bočnog ogranka $Lys^{295/298}$ radi koordinacije α i β-fosfata ATP molekule. DFG motiv, koji se nalazi na početku aktivacijske petlje, zauzima DFG-"in" konformaciju, u kojem je pokrajnji lanac aspartata usmjeren prema aktivnom mjestu kako bi vezao magnezijev kation koji izravno stupa u interakciju s atomom kisika β-fosfataze ATP-a. U isto vrijeme, pokrajnji lanac fenilalanina usmjeren je van aktivnog mjesta, a pozicija P-petlje zauzima otvorenu konformaciju.

U neaktivnoj, tj. "zatvorenoj", enzimskoj konformaciji, c-Src ima defosforilirani $Tyr^{416/419}$, a $Tyr^{527/530}$ je fosforiliran specifičnom c-Src kinazom (CSK). CSK i CSK homologna kinaza (CHK) su tirozin kinaze koje fosforiliraju $Tyr^{527/530}$ te time inhibiraju aktivnost c-Src uključivanjem SH2 u intramolekularnu interakciju s fosfotirozinom. Istodobno, dolazi do interakcije između SH3 domene i SH2 linkera, koja potiče promjene u orijentaciji αC zavojnice. Pomak αC zavojnice pomiče $Glu^{310/313}$, prekida se solni most između $Glu^{310/313}$ i $Lys^{295/298}$ te inhibira vezanje ATP-a. Također, inhibiran je i transfer fosfata na supstrat zatvorenom konformacijom P-petlje i DFG-"out" motivom (Slika 5).



Slika 5. Mehanizam aktivacije c-Src kinaze – a) neaktivna struktura; b) defosforilacija Tyr527/530; c) aktivna struktura. Po uzoru na (7).

Na staničnoj razini, vezanje SH2 domene c-Src kinaze na tirozin fosforilirane receptore faktora rasta pomiče $Tyr^{527/530}$, dopuštajući fosforilaciju $Tyr^{416/419}$ i aktivaciju c-Src. Nasuprot tome, disocijacija c-Src kinaze od receptora faktora rasta omogućuje promjenu intramolekulskih interakcija i zauzimanje neaktivne konformacije (7, 8, 9, 10, 12).

1.2.3.3. Uloga c-Src kinaze u regulaciji staničnih procesa i nastajanju tumora

Zbog sudjelovanja u brojnim signalnim putevima, c-Src kinaza uključena je u regulaciju različitih staničnih procesa kao što je preživljavanje, pokretljivost, diferencijacija i proliferacija stanica. Pri tome, stupa u interakciju s više RTK (VEGF, EGFR, PDGF, HGF, IGF-1…), drugim staničnim receptorima (CSF-1, integrinima, staničnim adhezijskim molekulama), receptorima sterodinih hormona, ligandima izvanstaničnog matriksa (ECM), komponentama čiji putevi su regulirani heterotrimernim

G proteinima, FAK, Shc i dr. Također, sudjeluje u kaskadnim aktivnostima koje aktiviraju STAT puteve i uključuje regulaciju RAS/RAF/MEK/ERK/MAPK puteva. Povećana aktivacija c-Src uzrokuje progresivni razvoj karcinoma s abnormalnom staničnom proliferacijom, angiogenezom, apoptozom, staničnom adhezijom, migracijom i invazijom. Osim toga, otkriveno je da ima važnu ulogu u drugim patološkim stanjima poput hematopoetskih poremećaja, epilepsije, dijabetesa, Parkinsonove bolesti, ishemijskih ozljeda, učenja i pamćenja. Na temelju navedenih predodžbi razvijaju se inhibitori TK, ponajviše za c-Src kinazu jer se smatra od iznimne kliničke važnosti i ključnom biološkom metom za liječenje raka (2, 13, 14).

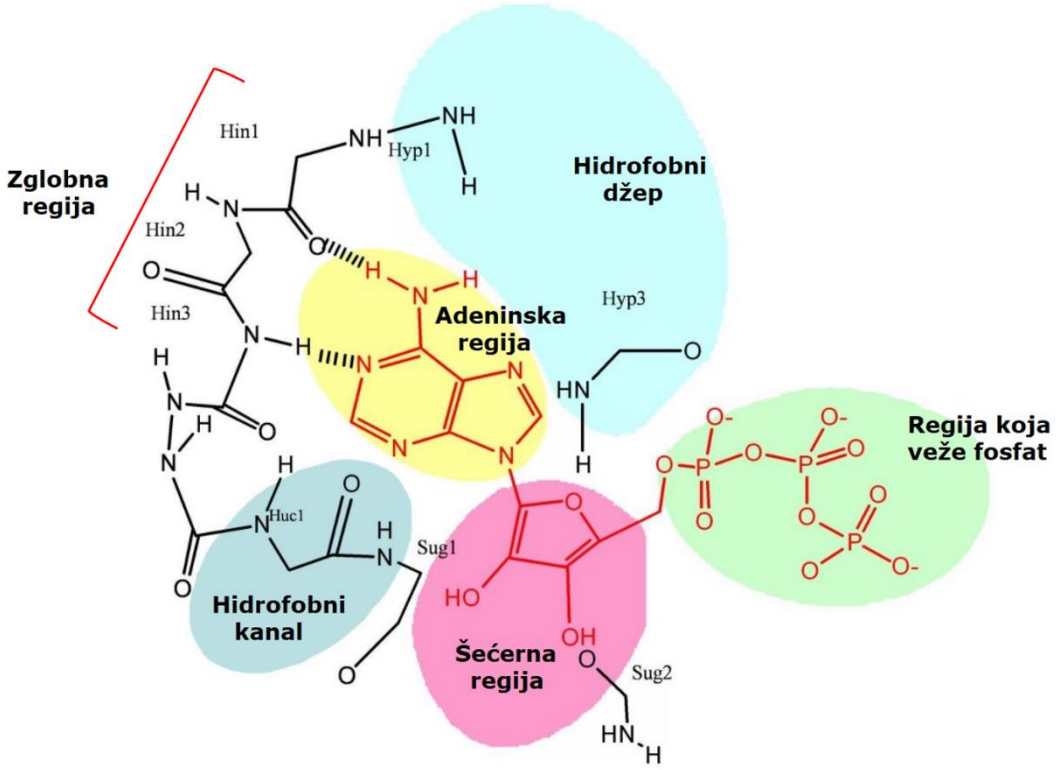
1.3. Tirozin kinaze kao biološke mete u lječenju karcinoma

U skladu s napretkom molekularnog razumijevanja procesa nastajanja karcinoma, kinaze su prepoznate kao potencijalne biološke mete za razvoj novih lijekova. Inhibitori kinaza mogu biti kompetitivni, nekompetitivni ili alosterički. Razvoj kompetitivnih inhibitora, koji se vežu na ATP vezno mjesto, smatrao se izazovom zbog visoke koncentracije ATP-a u stanici, nedovoljnog razumijevanja regulacije kinazne aktivnosti i visoke homologije ATP veznog mjesta, tj. niske selektivnosti prema pojedinačnim kinazama te posljedično toksičnosti (1, 13, 15).

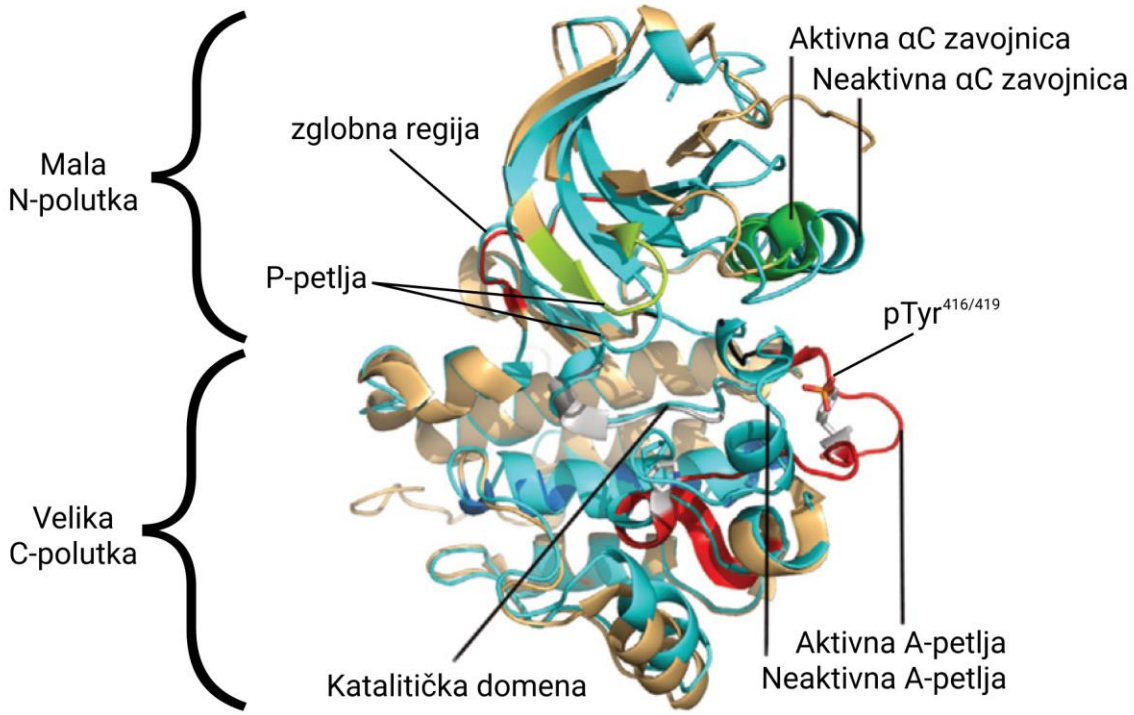
1.3.1. Značajke ATP veznog mjesta

ATP se veže unutar dubokog rascijepa između dva režnja tirozin kinaze, pri čemu se fosfatna skupina prenosi na supstrat koji je vezan za aktivacijsku petlju C-terminalnog režnja. Iako je ATP vezno mjesto očuvano, građa u okolnim regijama od mjesta vezanja ATP-a daje ključnu raznolikost za razvoj novih lijekova (1, 13). Mjesto vezanja građeno je od:

- **Adeninske regije** koja sadrži dvije ključne vodikove veze koje nastaju interakcijom N-1 i N-6 amino skupine adeninskog prstena s karboksilnom i amino skupinom zglobne (engl. *hinge*) regije. Mnogi snažni inhibitori stvaraju jednu ili obje od tih vodikovih veza.
- **Šećerne regije** (džep riboze) koja predstavlja hidrofilno područje.
- **Hidrofobnog džepa** koji se nalazi između C-terminalnog režnja i šećerne regije. ATP s njim ne ostvaruje interakcije, ali igra ključnu ulogu u selektivnosti inhibitora.
- Dubokog **hidrofobnog kanala** koji se nalazi uz zglobnu i adeninsku regiju te je okružen aminokiselinama N-terminalnog režnja (npr.Thr$^{339/342}$, Lys$^{295/298}$). ATP s njim ne ostvaruje interakcije, ali se koristi za optimiranje specifičnosti inhibitora.
- **Regije koja veže fosfat** i koristi za poboljšanje selektivnosti inhibitora (Slika 6. i 7.) (1, 8, 13).



Slika 6. ATP vezno mjesto protein kinaza. Po uzoru na (1).



Slika 7. Prikaz kinazne domene c-Src kinaze u aktivnom (različite boje) i neaktivnom (svijetlo plavo) obliku. Po uzoru na (9).

1.4. Inhibitori tirozin kinaza

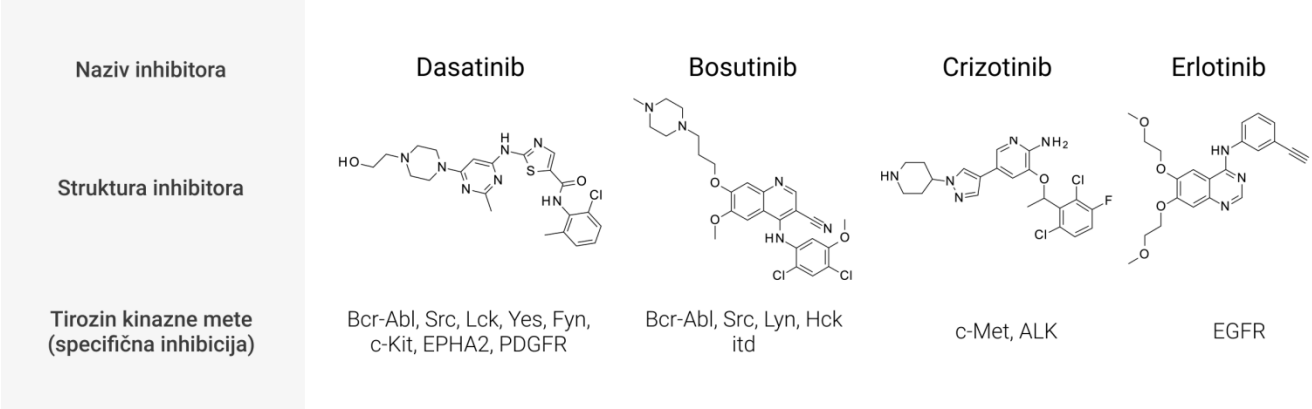
Većina inhibitora TK koji se koriste u kliničkoj praksi vežu se u ATP vezno mjesto te prema načinu vezanja razlikujemo nekovalentne i nekovalentne inhibitore.

1.4.1. Nekovalentni inhibitori

Nekovalentni kinazni inhibitori klasificiraju se na one koji se vežu ili ne na zglobnu regiju. Ovisno o tome i konformaciji DFG petlje, dijele se na reverzibilni tip 1, 1.5, 2 te inhibitorni tip 3 i 4.

Inhibitori tipa 1 su kompetitivni inhibitori koji se vežu za aktivnu (DFG-"in") konformaciju kinaze. Tada se A-petlja nalazi u otvorenoj

konformaciji, aspartat DFG motiva koordinira fosfate ATP-a, a fenilalanin se nalazi duboko u hidrofobnom džepu veznog mjesta. Selektivnost se postiže optimiranjem interakcije inhibitora s aminokiselinskim ostatkom "čuvara ulaza" (engl. *gatekeeper*) koji se kod različitih kinaza razlikuje prema veličini, obliku i polarnosti aminokiselinskog ostatka. Uspjeh tip 1 inhibitora u klinici pokazuje da je moguće optimirati selektivnost unatoč očuvanom mjestu vezanja ATP-a. To potvrđuje činjenica da predstavljaju veliku većinu dosada odobrenih lijekova. Primjeri inhibitora tipa 1 su dasatinib, bosutinib, crizotinib, erlotinib itd. (Slika 8.).

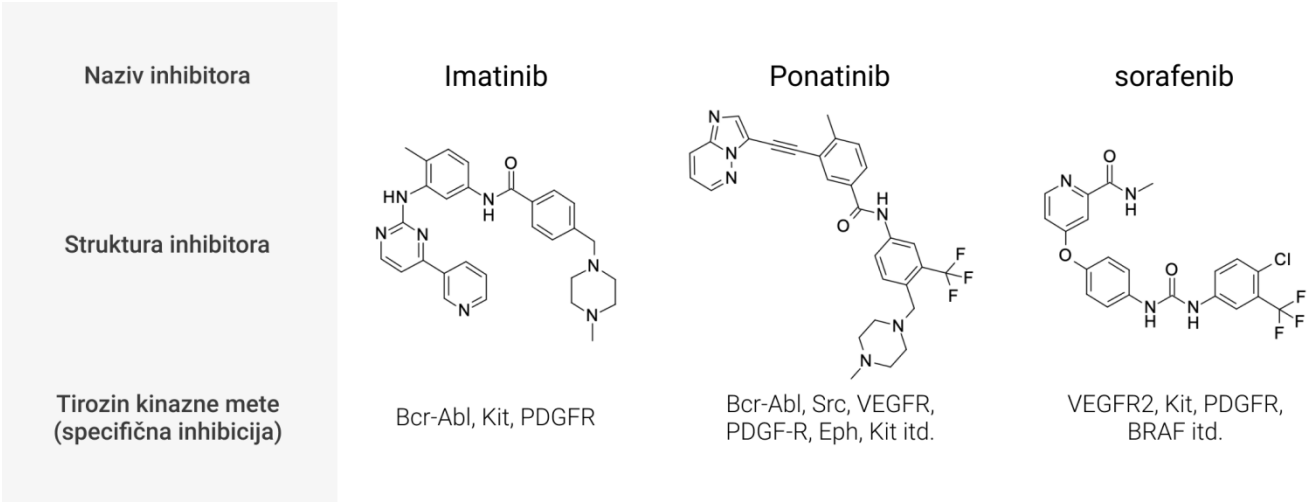


Slika 8. Tip 1 inhibitori Src kinaza i njihove mete.

Inhibitori 1.5 su podvrsta tipa 1 i vežu se na aktivnu DFG-"in" konformaciju kinaze. Razlikuju se od tipa 1 po tome što je αC-zavojnica orijentirana od veznog mjesta ATP-a.

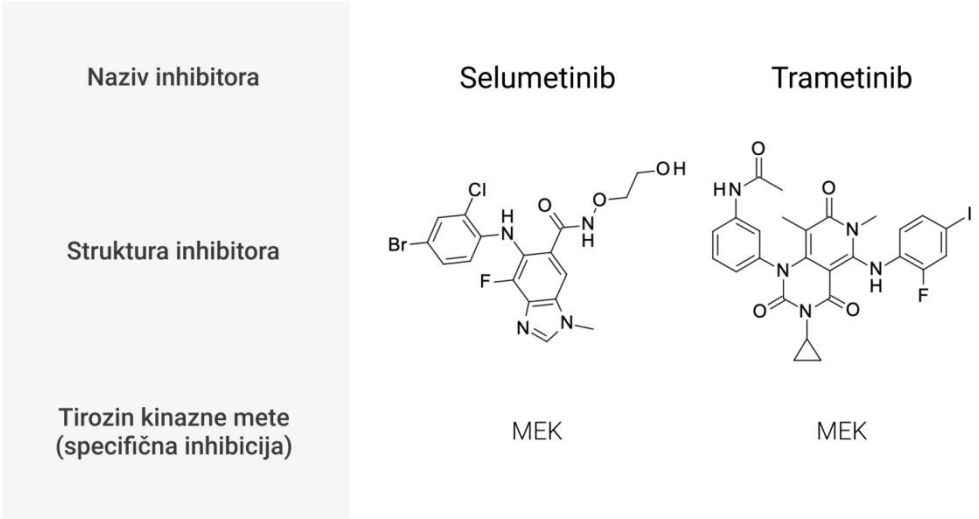
Inhibitori tipa 2 su kompetitivni inhibitori koji se vežu za neaktivno stanje konformacije gdje su bočni lanci DFG motiva orijentirani van veznog mjesta za ATP (DFG-"out") te su još uvijek u kontaktu sa zglobnom regijom. Vanjsko pomicanje DFG motiva stvara veliki hidrofobni džep na koji ciljaju, uz ATP vezno mjesto, dijelovi inhibitora kako bi kinaza ostala u naktivnom stanju. αC-zavojnica, G-petlja i A-petlja mogu zauzeti različite konformacijske oblike, zbog čega je ovaj tip inhibitora više selektivan.

Primjeri inhibitora tipa 2 su imatinib, ponatinib, sorafenib i sunitinib (Slika 9.).



Slika 9. Tip 2 inhibitori Src kinaza i njihove mete.

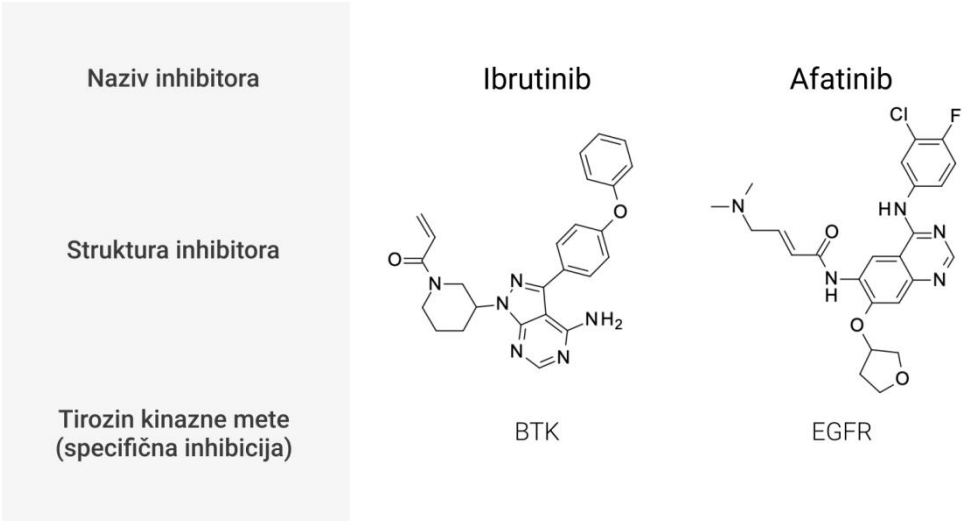
Inhibitori tipa 3 su alosterički inhibitori koji se ne vežu za ATP vezno mjesto već uz njega. Vežu se na aktivnu DFG-"in" konformaciju te time dopuštaju istovremeno vezanje inhibitora i ATP-a. Neki inhibitori tipa 3 stvaraju polarne interakcije s fosfatnom okosnicom ATP-a, što rezultira ATP-nekompetitivnim načinom djelovanja. Drugi inhibitori neizravno se natječu s ATP-om tako što se vežu za međusobno isključujuće konformacije. αC-zavojnica, G-petlja i A-petlja mogu generirati različite konformacije kinazne domene. Jedinstvene kombinacije svih strukturnih domena stvaraju strukturnu raznolikost koja se može koristiti za razvoj specifičnih i selektivnih inhibitora koji u odnosu na tip 1 i 2 inhibitore pokazuju najveću selektivnost pri vezanju u aktivno mjesto i regulatorni mehanizam koji je jedinstven za određenu kinazu. Primjeri inihibitora tipa 3 su selumetinib, refametinib, trametinib itd. (Slika 10.) (15, 16, 17).



Slika 10. Tip 3 inhibitori Src kinaza i njihove mete.

1.4.2. Kovalentni inhibitori

Kovalentni inhibitori mogu se vezati za ATP vezno mjesto ili u njegovoj blizini. Ovisno o reaktivnosti kovalentnih inhibitora vezanje može biti reverzibilno ili ireverzibilno. Inhibitori najčešće tvore kovalentne veze sa ne-katalitičkim ostacima Lys ili Cys u ili oko ATP veznog mjesta. Unatoč tome što su pokazali impresivne kliničke rezultate, mutacije Lys ili Cys mogu spriječiti vezanje kovalentnih inhibitora i njihovu učinkovitost. Primjeri kovalentnih inhibitora, koji se primarno vežu na BTK i VEGFR, su ibrutinib i afatinib (Slika 11.) (15,16).



Slika 11. Kovalentni inhibitori Src kinaza i njihove mete.

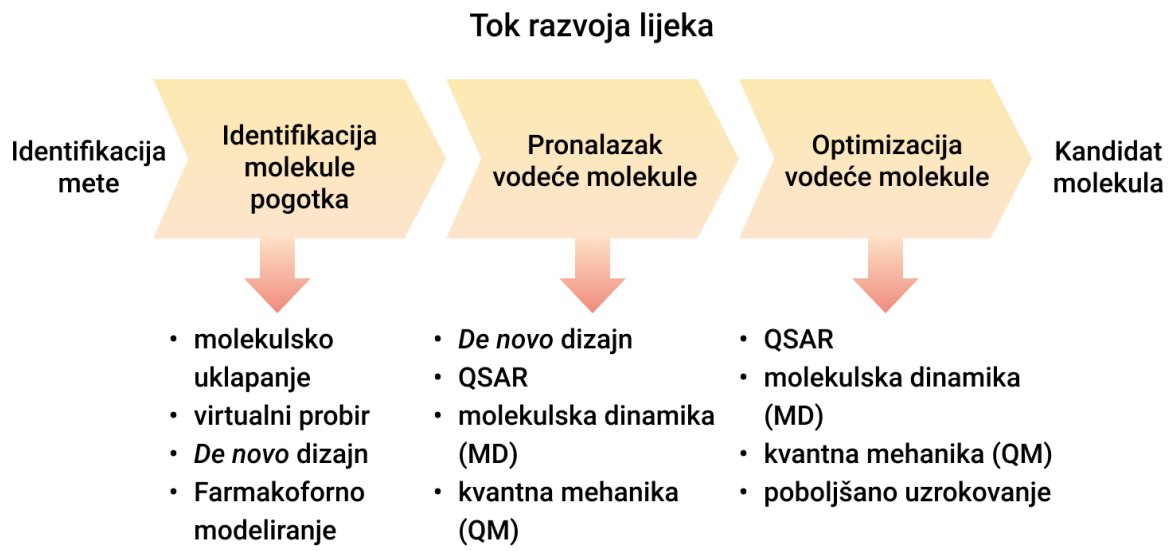
1.4.3. Rezistencija i selektivnost inhibitora Src kinaza

Dva izazova koja se javljaju kod razvoja inhibitora c-Src kinaze su rezistencija i selektivnost. Pokazalo se da inhibitori c-Src i srodnih kinaza izazivaju mehanizme višestruke rezistencije na kemoterapiju i zračenje. Osim smanjene akumulacije lijeka, dolazi i do promjena u ciljnim proteinima, mutacija koje smanjuju vezanje inhibitora, pojačanog metabolizma lijekova CYP enzimima (citokrom P450), smanjene apoptičke signalizacije i povećanog DNA oštećenja. Također, rezistencija se može pojaviti još i kompenzacijskim mehanizmima ili smanjenjem afiniteta kinaze prema njezinim inhibitorima.

Važno svojstvo kinaznih inhibitora je selektivnost, no veliki broj sličnih proteina u organizmu i različite strukturne forme istih otežava njihov razvoj. c-Src inhibitori imaju ograničenu selektivnost u odnosu na druge SFK i ostale protein kinaze zbog visoke homologije unutar kinazne domene (16).

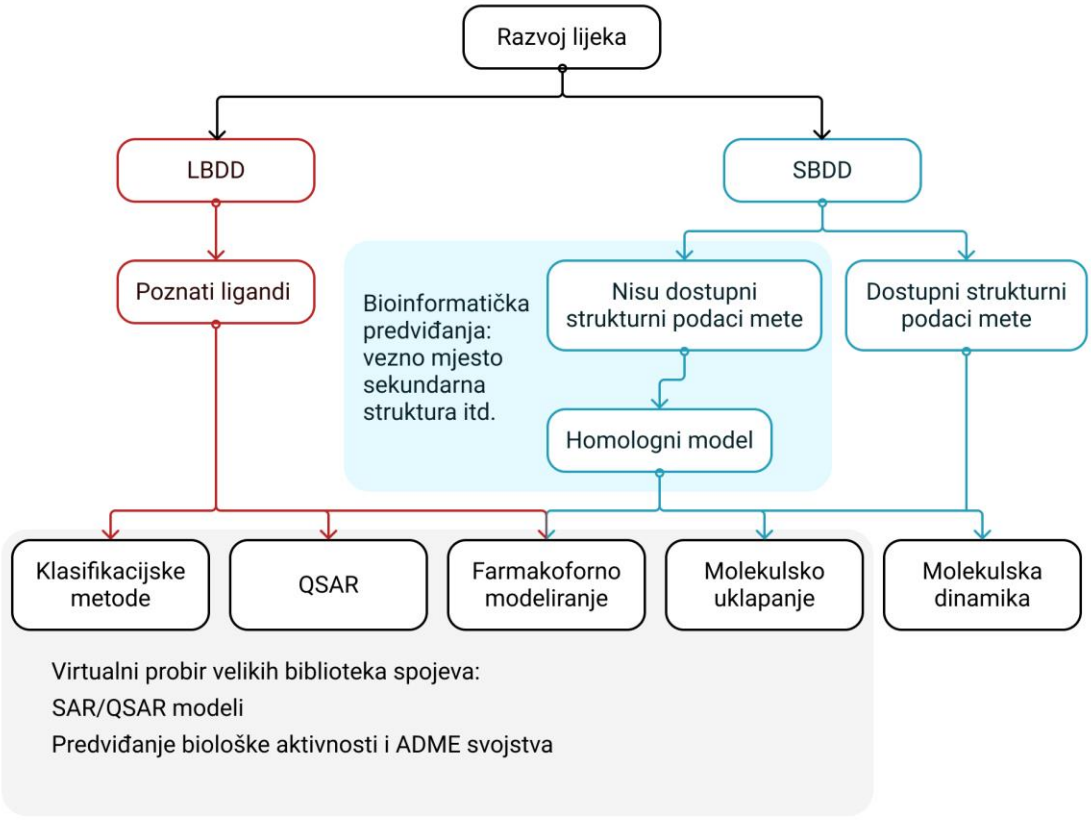
1.5. *In silico* dizajn novih lijekova

Razvoj novih lijekova je dugotrajan, složen i vrlo skup proces te je uspješnost razvoja početnog spoja u odobreni i komercijalno dostupni proizvod vrlo niska. *In silico* (računalne) metode pridonose ubrzanju procesa i financijskom smanjivanju troškova, a različite metode primjenjuju se gotovo u svim fazama otkrivanja i razvoja lijeka (Slika 12.).



Slika 12. Tok razvoja lijeka. Po uzoru na (19).

Računalno potpomognuti dizajn lijekova (engl. *Computer-aided drug design*; CADD) proces je razvoja novih lijekova u kojem se sve faze dizajna i analize novih molekula provode pomoću računalnih programa. Dva su osnovna pristupa u racionalnom dizajniranju lijekova – dizajn lijekova temeljen na ligandu (engl. *ligand-based drug design*; LBDD) i dizajn lijekova temeljen na strukturi (engl. *structure based drug design*; SBDD). Izbor najprikladnijeg pristupa ovisi o dostupnosti informacija na početku razvoja lijeka, npr. dostupnost 3D strukure biološke mete, dostupnost poznatih liganda itd. (Slika 13.).

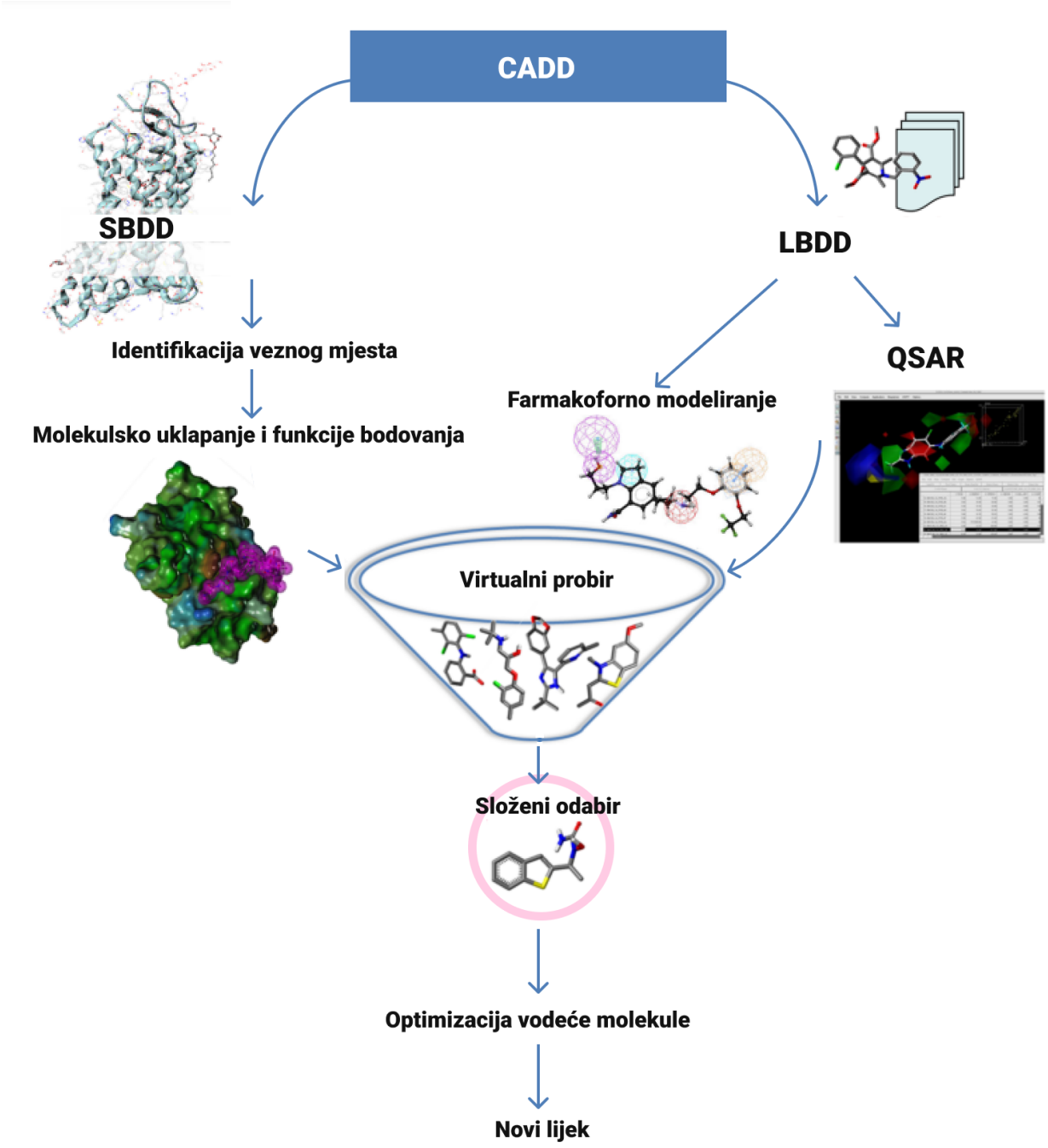


Slika 13. Različiti pristupi korišteni u razvoju lijeka. Po uzoru na (21).

LBDD pristup koristi se ukoliko 3D strukturne informacije o proteinu nisu dostupne ili u potpunosti pouzdane. Ovaj pristup temelji se na informacijama o aktivnosti poznatih liganada za koje je poznato da stupaju u interakciju s biološkom metom. Nasuprot, SBDD pristup koristi se kada su dostupni strukturni podaci ciljnog proteina. Određivanjem kristalnih struktura kompleksa biološke mete i inhibitora dobiva se uvid u bioaktivnu konformaciju liganada te specifične interakcije koje stvara s proteinom. Osim za predviđanje strukture proteina, razvoja modela za predviđanje i/ili klasifikaciju biološke aktivnosti novih liganada, CADD se koristi i za izgradnju modela koji predviđaju i/ili klasificiraju farmakokinetički (PK) i toksikološki (ADMET) profil spojeva, kako bi se izbjegli spojevi s nepoželjnim farmakokinetičkim ili toksičnim svojstvima (18, 19, 21).

Virtualni probir (engl. *virtual screening*; VS) ili *in silico* probir je računalna metoda koja služi za probir velikog broja molekula. Uspješno se koristi za razvoj novih lijekova uz probir visoke propusnosti (engl. *High-throughput screening*; HTS). Primjenjuje se za provjeru virtualnih biblioteka spojeva koristeći različite računalne pristupe (LBDD ili SBDD) za identifikaciju molekula za koje je vjerojatno da će se vezati za protein te ih je vrijedno testirati ili sintetizirati. U virtualnom probiru mogu se koristiti biblioteke komercijalno dostupnih spojeva u svrhu odabira manjeg skupa spojeva za biološka testiranja ili virtualne biblioteke potencijalno zanimljivih *target* molekula koje se ovim pristupom mogu filtrirati i prioritizirati za sintezu. Time se postiže glavni cilj – omogućava se brza i isplativa evaluacija velikih virtualno složenih baza podataka kako bi se pronašle potencijalno zanimljive molekule za sintezu i daljnje proučavanje te povećala vjerojatnost za njihovu biološku aktivnost (Slika 14.) (18, 19, 20).

Razvojem rendgenske strukturne analize visoke učinkovitosti (engl. *high-throughput x-ray analysis*) razvijen je i pristup istraživanja lijekova zasnovan na molekulskim fragmentima (engl. *fragment based drug discovery*; FBDD), koji se primijenjuje za razvoj malih molekula koje se prema građi sastoje od više organskih fragmenata. Svaki fragment ima zasebno mjesto vezanja koje je dovoljno blizu kako bi se kemijski fragmenti mogli povezati u relativno povoljan položaj. Njihova selekcija vrši se iz virtualne biblioteke spojeva, a uobičajeno imaju nisku molekularnu masu (<300Da), slabi afinitet vezanja za protein i jednostavnu kemijsku strukturu. Osim *in silico* metoda, FBDD pristup koristi i eksperimentalne metode koje uključuju kristalografiju, nuklearnu magnetsku rezonancu (engl. *nuclear magnetic resonance*; NMR), SPR (engl. *surface plasmon resonance*), ITC (engl. *isothermal titration calorimetry*) itd. (22, 23).



Slika 14. Pregled CADD procesa. Po uzoru na (20).

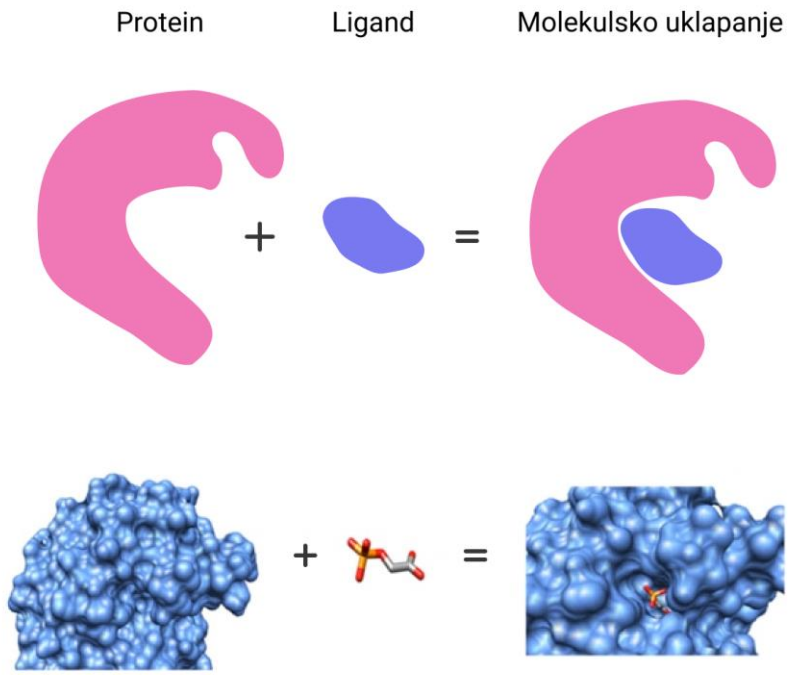
1.5.1. Dizajn lijekova prema strukturi (SBDD)

Napretkom eksperimentalnih tehnika, poput kristalografije, nuklearne magnetne rezonance, elektronske mikroskopije itd., proširila su se znanja o strukturama proteina i time je SBDD metoda doživjela ubrzani razvoj.

Strukturne informacija proteina i funkcije bodovanja (engl. *scoring functions)* pomažu SBDD metodi u dizajnu aktivnijih i selektivnijih liganada. Glavni cilj metode je razumjeti kako se ligand veže na aktivno mjesto proteina na temelju specifičnih protein-ligand interakcija (19, 24). SBDD primjenjuje širok raspon računskih metoda, kao što su metode temeljene na polju sila, npr. molekulsko uklapanje (engl. *molecular docking*) i molekulska dinamika (engl. *molecular dynamics; MD*). Također, primjenjuje kvantnu mehaniku (engl. *quantum mechanics; QM, QM/MM*) i *de novo* dizajn lijekova (19). Proces se sastoji od identifikacije proteina, a zatim se kristalografijom ili NMR-om određuje njegova struktura. Struktura mete određuje se homolognim modeliranjem, ako postoji određena struktura homolognog proteina i ukoliko se struktura ne može odrediti eksperimentalno. Nakon toga, određuje se mjesto vezanja i elektrostatska svojstva. Proučava se struktura kompleksa protein-ligand, a za bolje razumijevanje kompleksa pridonose informacije o međumolekulskim interakcijama, konformacijama vezanja i karakterizaciji nepoznatih veznih mjesta. Konačni cilj procesa sinteza je potencijalnih bioaktivnih liganada te provođenje bioloških testova (21, 24, 25).

1.5.1.1. Molekulsko uklapanje

Molekulsko uklapanje najčešće je korištena SBDD tehnika, a služi za brzo i ekonomično predviđanje načina vezanja (orijentacije) i afiniteta vezanja malih molekula (liganada) prema ciljnim proteinima. Ova tehnika simulira proces molekulskog prepoznavanja predviđanjem slobodne energije vezanja i geometrije interakcije veznog kompleksa protein-ligand te tako predviđa najbolju orijentaciju liganda u 3D prostoru (Slika 15.) (18, 26).



Slika 15. Prikaz molekulskog uklapanja liganda i protein.

Razvoj računalnih metoda te pristup bazama eksperimentalno određenih 3D struktura proteina i malih molekula povećali su učinkovitost ove tehnike, te olakšali primjenu u studijama molekularnih interakcija uključenih u protein-ligand vezanje. Male molekule mogu se uklopiti na tri različita načina:

- Rigidnim uklapanjem – strukture proteina i liganda se ne optimiraju.
- Fleksibilnim uklapanjem – strukture proteina i liganda su fleksibilne.
- Uklapanjem fleksibilnog liganda – struktura liganda je fleksibilna dok se struktura proteina ne optimira.

Obično, glavni cilj molekulskog uklapanja je identifikacija liganda koji se najpovoljnije veže unutar veznog mjesta proteina te određivanje energetski najpovoljnije orijentacije njegovog vezanja (poze). Poza vezanja odnosi se ili na konformaciju molekule liganda unutar mjesta

vezanja ciljnog proteina koja je eksperimentalno potvrđena ili na računalno modelirano hipotetičku konformaciju.

Razvijen je niz programa za protein-ligand molekulsko uklapanje, a neki od njih su AutoDock, Glide, DOCK, GOLD, FlexX itd. Dva su osnovna koraka u procesu molekulskog uklapanja: algoritam pretraživanja konformacijskog prostora i funkcije bodovanja. Algoritam pretraživanja generira niz različitih konformacija i orijentacija vezanja liganda. Funkcije bodovanja koriste se zatim za procjenu afiniteta vezanja generiranih poza, a potom slijedi njihovo rangiranje i identifikacija najboljeg načina vezanja. Idealan algoritam pretraživanja treba biti brz i učinkovit, a funkcije bodovanja s dovoljnom točnošću predvidjeti fizikalno-kemijska svojstva molekula i termodinamiku interakcija. Dakle, dobro uklopljeni položaj karakterizira povoljno bodovanje koje bi trebalo odražavati i odgovarati eksperimentalnim podacima, ako su dostupni. Iako se ova dva koraka smatraju individualnim problemima, oni su povezani. To je vidljivo kada se u obzir uzme energija vezanja kao primarni kriteriji za konformacijski odabir veznog liganda, koja se koristi za odabir pozicije među svim mogućim varijantama vezanja istih. Zbog stvaranja kompleksa protein-ligand dolazi do promjene energije koja je dana Gibbs-ovom slobodnom energijom (ΔG) i konstantom vezanja. Energija vezanja kompleksa predviđa se procjenom interakcija uključenih u protein-ligand vezanje.

Funkcije bodovanja podijeljene su u tri kategorije: funkcije temeljene na polju sila, empirijski temeljene funkcije i funkcije temeljene na znanju. Budući da svaka od njih ima svoje prednosti i nedostatke, a niti jedna nije savršena u smislu točnosti i opće primjenjivosti, uvedena je strategija bodovanja pomoću konsenzusa (engl. *consensus score/c-score*) kako bi se poboljšala vjerojatnost pronalaska točnijih rješenja kombiniranjem bodovanja iz više funkcija bodovanja. Stvara se konsenzusno bodovanje kao prosječna ili linearna kombinacija funkcija bodovanja. Navedeno se pokazalo korisnijim za bodovanje i rangiranje spojeva jer konsenzusno

bodovanje pokazuje bolje slaganje sa eksperimentalnim rezultatima (18, 19, 20, 26).

Funkcije temeljene na polju sila koriste eksperimentalne i kvantno mehaničke (QM) proračune funkcionalnih oblika i polja sila za procjenu afiniteta vezanja. Za pojednostavljivanje postupaka procjenjuje se samo jačina nekovalentnih interakcija između molekula (entalpijski doprinos) u kompleksu, no to nije dovoljno dobro rješenje jer za točniju procjenu afiniteta vezanja treba uključivati promjene u intramolekulskim interakcijama nakon vezanja protein-ligand kompleksa i u interakcijama uključenima u otapalu te entropijske učinke. Učinak otapala može se opisati eksplicitnim molekulama vode ili korištenjem implicitnih (ili kontinuiranih) modela otapala npr. Poisson-Boltzmann modelom površine (engl. *Poisson-Boltzmann surface area;* PBSA) i modelom generalizirane Born površine (engl. *generalized Born surface area*; GBSA). (26).

Popularni pristupi za izračunavanje slobodne energije vezanja liganda za protein su MM-PBSA (engl. molecular mechanics, *Poisson-Boltzmann surface area*) i MM-GBSA (engl. molecular mechanics, generalized Born model and solvent accessibility). Ove računalne metode uključuju simulacije molekulske dinamike (MD) i implicitne modele otapala. Također, mogu biti korisne za poboljšanje rezultata molekulskog uklapanja i virtualnog probira ili za razumijevanje uočenih afiniteta (27). Kod ovih metoda predviđanje afiniteta vezanja postiže se proračunima slobodne energije koji uzimaju u obzir sve termodinamičke pojave poput dinamike/fleksibilnosti proteina, eksplicitnog uključivanja otapala i razlike između interakcija u protein-ligand kompleksu i njihovim interakcijama s vodom i protuionima u nevezanim oblicima (26).

Detaljnija slika vezanja liganda može se dobiti simulacijskim pristupima. Molekulska dinamika je metoda računalne simulacije gibanja atoma i molekula koja se temelji na Newtonovim zakonima gibanja. Kretanje čestica se izračunava u određenom vremenskom razdoblju i istražuje se

evolucija sustava. Također, ova metoda omogućava praćenje dinamike sustava u različitim otapalima. Primjena simulacije molekulske dinamike vrlo je važna za povezivanje fleksibilnosti biomolekularnog sustava s prepoznavanjem liganda. Nadalje, moćan je alat za identificiranje dodatnih mjesta vezanja koja se ne mogu otkriti drugim strukturnim metodama. Zbog svega navedenog, smatra se poveznicom između teorijskog i eksperimentalnog pristupa (24).

1.5.2. Dizajn lijekova prema ligandu (LBDD)

Kao što je ranije spomenuto, LBDD metoda primjenjuje se kada je struktura proteina nepoznata ili nije u potpunosti pouzdana. U tom slučaju za izradu farmakofornog modela koriste se eksperimentalni podaci aktivnih spojeva. Metoda se bazira na pretpostavci da strukturno slični spojevi pokazuju sličan biološki odgovor i slične interakcije s biološkom metom. Skup spojeva treba obuhvaćati širok raspon aktivnosti da bi se stvorio pouzdani model probira temeljen na ligandu. Uobičajene metode LBDD dizajna su kvantitativni odnosi strukture i aktivnosti (engl. *quantitative structure–activity relationship;* QSAR), farmakoforno modeliranje, analiza temeljena na molekulskim poljima, 2D ili 3D procjena sličnosti (20, 21).

Prikaz molekula temelj je LBDD pristupa. Za predstavljanje strukturnih i fizikalno-kemijskih svojstava molekula koriste se molekularni deskriptori (strukturne, steričke i elektronske značajke). To su numeričke vrijednosti poput molekulske mase, elektronegativnosti, dostupne površine, geometrije, indeksa aromatičnosti, polarizabilnosti, svojstva solvacije itd. Definiranje i izračunavanje molekulskih deskriptora široko je znanstveno područje koje uključuje niz različitih teorija i pristupa (19, 21).

LBDD metoda oslanja se na prethodno znanje aktivnog liganda na temelju čega se iz velikih knjižnica kemijskih spojeva pronalaze slični spojevi. Nakon identifikacije, spojevi se rangiraju prema koeficijentu sličnosti, te

se za odabrane spojeve izračunavaju molekulski deskriptori. Češće korištene kategorije molekularnih deskriptora su 2D deskriptori ("kemijski otisci", topološki deskriptori, molekularna svojstva) i 3D deskriptori (farmakoforna svojstva, svojstva ovisna o konformaciji molekule, molekularni oblik/volumen), koji služe za pretraživanje sličnosti jednog ili više aktivnih spojeva. Ovaj pristup vrlo je popularan zbog malih troškova, ali i jednostavnije uporabe od SBDD metode (19, 20, 21). Dodatna je vrijednost u izgradnji modela kombinacija LBDD i SBDD pristupa, pri čemu se struktura biološke mete može koristiti za određivanje oblika aktivnog mjesta unutar kojeg se preklapaju molekule ili za određivanje bioaktivne konformacije koja se koristi kao referentna konformacija za preklapanje novih molekula.

1.5.2.1. QSAR

QSAR modeli opisuju korelaciju između strukturnih i kemijskih svojstava s biološkom aktivnošću te se razvija statistički model za predviđanje bioloških svojstava novih spojeva (20, 24). Cilj je pronaći statistički značajnu vezu između kemijske strukture i kontinuiranog (pIC50, pEC50 itd.) ili kategorijskog/binarnog (aktivne, neaktivne, toksične, netoksične itd.) biološkog/toksikološkog svojstva primjenom tehnika regresije i klasifikacije. Za izgradnju oba modela potreban je što veći skup podataka aktivnih spojeva. Izgradnja klasifikacijskog modela jednostavnija je od regresijskog, iz razloga što za njega nisu potrebne eksplicitne vrijednosti bioloških aktivnosti (izražene kao IC_{50}, EC_{50} itd.) kao kod regresijskog, no i razina predviđanja također je ograničena, tj. molekule se mogu predvidjeti samo kao aktivne ili neaktivne (21).

Za QSAR pristup potrebno je poznavanje aktivnosti svakog spoja te korištenje niza molekularnih deskriptora. Pomoću matrice podataka razvija se model koji opisuje odnos između svojstava molekule opisanih deskriptorima i njihove biološke aktivnost, pri čemu se mogu koristiti

linearne i nelinearne metode. Deskriptori se izračunavaju iz fizikalno-kemijskih i strukturnih svojstava molekula kao što su broj atoma, vrste i udaljenost atoma, molekularna masa, elektronegativnost, aromatičnost i dr. Osim što je potreban skup podataka dovoljne veličine da se razviju robusni i pouzdani modeli, potreban je pravilan odabir spojeva za test set i set za učenje. Koraci u razvoju QSAR modela prikazani su na slici 16. (18, 20, 24).



Slika 16. Opći koraci za razvoju QSAR modela. Po uzoru na (18).

2D-QSAR analiza predstavlja kvantitativni odnos između biološke aktivnosti skupa spojeva i 2D molekularnih/strukturnih (topoloških) deskriptora spojeva. 2D deskriptori opisuju kako su atomi povezani u molekuli, vrstu veze i interakciju pojedinih atoma. Za izradu modela koriste se jednostavne statističke tehnike temeljene na regresiji, npr. parcijalna linearna regresija (engl. *Partial least squares method*, PLS) i multipla linearna regresija uz metodu potpornih vektora (engl. *support vector machines,* SVM) (18).

3D-QSAR analizom istražuje se kvantitativni odnos između biološke aktivnosti i 3D strukture proučavanih molekula opisanih 3D deskriptorima (molekularne površine, volumena, elektrostatskih ili steričkih potencijala itd.) pomoću statističkih metoda korelacije (npr. PLS, SVM). Kada su potrebni složeniji izračuni poput deskriptora molekulskih polja i deskriptora temeljenih na farmakofornim svojstvima, primjenjuju se složenije 3D-QSAR metode kao što su CoMFA (usporedna analiza molekulskih polja, engl. *comparative molecular field analysis*) i CoMSIA (usporedna analiza indeksa molekularne sličnosti, engl. *comparative molecular similarity indices analysis*) 3D-QSAR (18, 24). Osim ovih metoda razvijeni su i napredniji višedimenzionalni modeli (4D/5D/6D…) koji uzimaju u obzir energetske doprinose vezanja liganda te uključuju energiju solvacije i konformacijsku entropiju. Unatoč tome, CoMFA jedna je od, do sada, najčešće korištenih 3D-QSAR metoda (20).

1.5.2.2. Analiza temeljena na molekulskim poljima

Analiza i preklapanje molekula na temelju molekulskih polja primijenjena je u Cresset računalnom paketu. Cresset tehnologija koristi XED polja sila u dizajnu bioaktivnih molekula. XED (engl. *extended electron distribution*) pristup poboljšava tradicionalno molekulsko polje sila korištenjem složenog opisa atoma koji naboj stavlja van atomskih centara. Time je omogućen detaljniji opis elektrostatskih polja i bolje predviđanje intermolekularnih interakcija. XED polje sila dobro modelira utjecaj supstituenata na aromatske sustave, promjenu gustoće naboja kod složenih aromata i intermolekularne interakcije malih molekula, proteina i molekula vode.

Osim elektrostatskih interakcija za molekularno prepoznavanje bitni su hidrofobnost i oblik molekule. Elektrostatsko okruženje oko liganda ili proteina opisuje se kao potencijal molekularnih interakcija (engl. *MIP, molecular interaction potentional)* ili polja. Opisivanje molekula prema

molekulskim poljima, umjesto prema strukturi, omogućuje uspoređivanje strukturno različitih molekula te bioizosternih skupina.

1.5.2.2.1. XED polje sile

Glavni fokus Cresset programskog paketa je točnim izračunom elektrostatskih potencijala prikazati molekulska polja. Vrlo točne elektrostatske potencijale mogu dati kvantno-mehanički izračuni, koji su u većini slučajeva prespori za praktično korištenje na većem broju molekula. Kao rezultat toga, potrebna je točnija metoda izračunavanja elektrostatskih potencijala u kontekstu molekularne mehanike.

Aproksimaciju naboja u centru atoma (engl*. ACC, atom-centred charge*) koristi većina standardnih polja. Skupom točkastih djelomičnih naboja smještenih na jezgrama aproksimira se elektrostatski potencijal. Temeljni model za izračunavanje djelomičnih naboja jedan je točkasti naboj po atomu. Ova metoda može dobro izračunati elektrostatski potencijal za veće udaljenosti (npr. dipolni moment), ali daje loše rezultate pri opisivanju elektrostatskog potencijala u blizini molekularne površine. Razlog tome je što atomi nemaju nabijene sfere već imaju nepodijeljene elektronske parove, π orbitale, σ rupe itd. Osim toga, atomi i molekule polarizirani su i mijenjaju svoje elektrostatsko ponašanje kao odgovor na vanjska električna polja. ACC model ne pokriva niti jedan od ovih efekata, ali zato novija polja sila, kao što je AMOEBA, rješavaju problem postavljanjem eksplicitnih multipola i polarizacijskih funkcija na atome, što daje realniji elektrostatski potencijal. Ova polja sila se dobro ponašaju na proteinima, ali zbog problema s prenosivošću parametara su neprikladni za modeliranje liganda. XED polje sile bio je prvi pokušaj rješavanja ovih elektrostatskih problema, a to je učinjeno postavljanjem dodatnih monopola na atome.

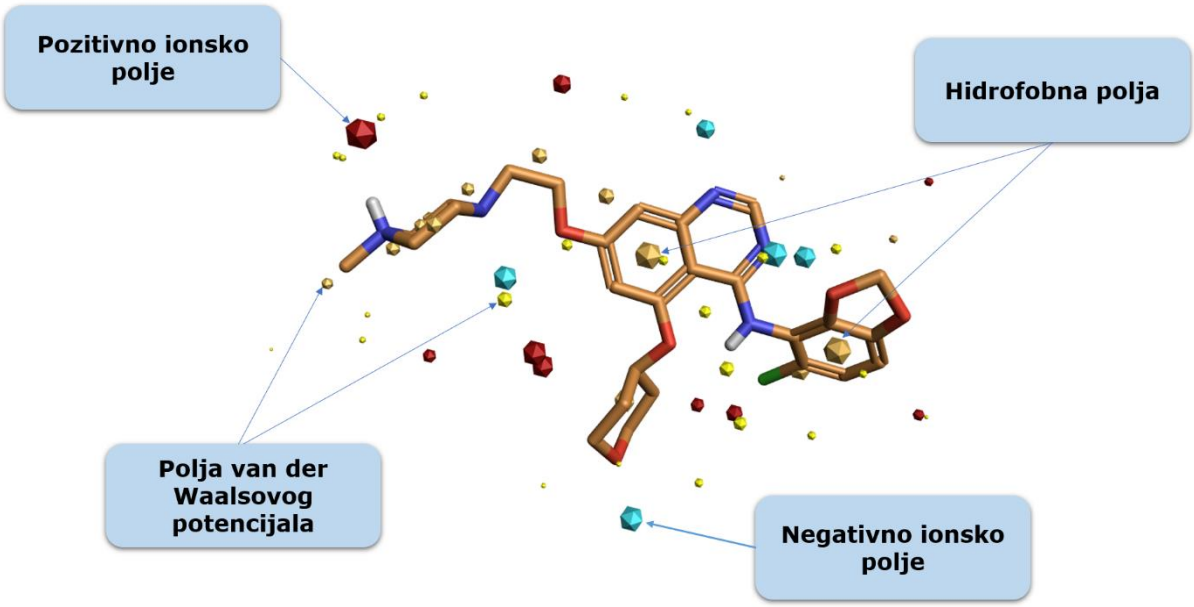
Dodatne monopolske točke ili XED-ovi tretiraju se unutar polja sila kao atomi s nultim van der Waalsovim radijusom. One dolaze s potencijalima istezanja veza i savijanja kuta te se mogu kretati pod utjecajem vanjskih

(i intramolekularnih) elektrostatskih potencijala, dopuštajući izravno modeliranje polarizabilnosti. Dakle, dodatne monopolske točke nisu postavljene u rigidnu geometriju u odnosu na njihov matični atom. Složeniji unutarnji elektrostatski model omogućuje da se intramolekularne elektrostatske/orbitalne interakcije, poput anomernog učinka, modeliraju bez uvođenja posebnih torzijskih parametara.

1.5.2.2.2. Izračunavanje polja za procjenu molekularnih interakcija

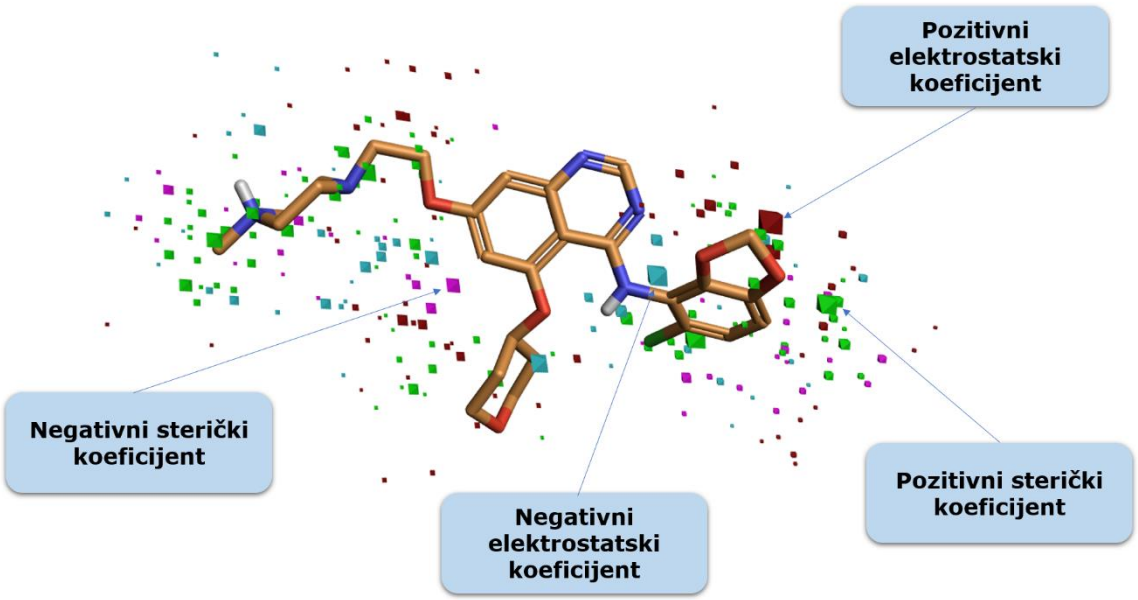
Cresset tehnologija sličnosti 3D liganda uspoređuje molekule prema njihovim molekularnim elektrostatskim potencijalima (engl*. Molecular electrostatic interaction potentials, MIPs*) ili poljima (engl*. fields*). MIP molekule je skalarno polje gdje je vrijednost u svakoj točki u prostoru energija interakcije nabijenog atoma probe (s van der Waalsovim parametrima kisika) s molekulom. Izračunavaju se pomoću XED polja sila. Nadalje, vrijednost polja postavljena je na nulu u točkama gdje je van der Waalsova energija interakcije pozitivna i veća od apsolutne vrijednosti elektrostatske energije.

Zbog jednostavnosti prikaza, izračunava se maksimum molekulskog elektrostatskog polja te se svaki takav ekstrem naziva točka polja (engl*. field point*). Skup točaka polja definiran je jedinstveno za bilo koju danu molekularnu konformaciju. Točke polja prikazuju se kao sfere u boji gdje je veličina točke polja određena veličinom polja – jača polja dobivaju veće sfere. Time se omogućuje vizualni prikaz gdje molekula može ostvariti lokalno maksimalnu elektrostatsku interakciju s drugom molekulom. Plava polja odgovaraju područjima gdje je negativno elektrostatsko polje, crvena pozitivnim elektrostatskim poljima, žuta predstavlja polje van der Waalsovog potencijala, a narančasta odgovara području hidrofobnosti (Slika 17). Prisutnost točaka polja sugerira da je to područje liganda koje može stvoriti povoljnu interakciju s veznim mjestom proteina.



Slika 17. Različite vrste molekulskih polja prikazane na Saracatinib skupu podataka na molekuli saracatiniba.

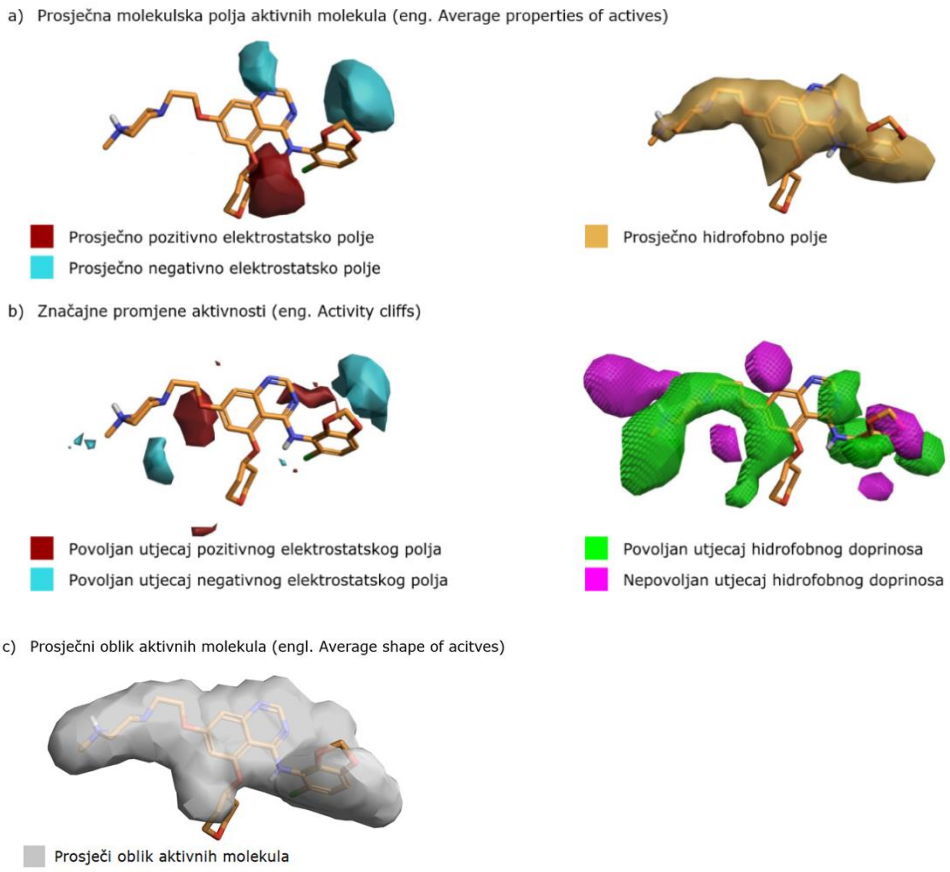
Ovaj program koristi se za dizajn novih biološki aktivnih molekula na temelju struktura poznatih liganada (LBDD) te odnosa strukture i aktivnosti (engl. *structure-activity relationship*, SAR). Pri razvoju 3D-QSAR modela koristimo statističke metode za pronalaženje odnosa između molekulske strukture opisane strukturnim deskriptorima (vektor brojeva koji predstavlja svaku molekulu) i biološkim aktivnostima molekula. Cresset koristi deskriptore temeljene na elektrostatskim molekularnim poljima i steričkim svojstvima za karakterizaciju svake molekule i izgradnju 3D-QSAR modela. Kao i većina 3D-QSAR metoda za generiranje deskriptora, program zahtijeva zadovoljavajuće preklapanje molekula s referentnom strukturom, pri čemu se detaljno pretražuje konformacijski prostor molekule i pronalaze energetski povoljne konformacije koje se najbolje preklapaju s referentnom strukturom. Primjer 3D-QSAR modela prikazan je na slici 18.



Slika 18. Pozitivni i negativni elektrostatski i sterički koeficijeni određeni 3D-QSAR modelom na Saracatinib skupu podataka na molekuli saracatiniba.

Osim kvantitativni, mogu se razvijati i kvalitativni statistički modeli poput "*Activity atlas*" metode koja se temelji na analizi vjerojatnosti odnosa strukture i aktivnosti (SAR) skupa preklopljenih spojeva kao funkcije njihovih elektrostatskih svojstava, hidrofobnosti i oblika. Rezultat kvalitativnog modela je poopćena verzija farmakofornog modela, tj. prosjek elektrostatskih i hidrofobnih molekulskih polja za set aktivnih molekula kao što je prikazano na slici 19. Prosječna molekulska polja aktivnih molekula prikazuju prosjek pozitivnog (crveno) i negativnog (plavo) elektrostaskog polja te hidrofobnog polja (žuto) (Slika 19. a). "*Activity atlas*", također, omogućava određivanje onih molekulskih polja koja pri maloj promjeni značajno utječu na promjenu aktivnosti (engl. *Activity cliffs*). "*Activity cliff*" analiza pokazuje regije, gdje će pozitivnije polje (u crvenim regijama) i/ili negativnije polje (u plavoj regiji) povećati aktivnost. "*Activity cliff*" analiza hidrofobnih regija pokazuje gdje je povećanje hidrofobnih interakcija povoljno (zelena područja), odnosno nepovoljno (magenta regije) za biološku aktivnost. Analiza oblika pokazuje regije u kojima je sterički efekt povoljan, zelena, odnosno

nepovoljan, magenta. Primjer "*Activity cliff*" analize prikazan je na slici 19. b). Pozitivnije polje može se tumačiti i kao "manje negativan" te stavljanjem jakih donora H-veza u tu regiju može se poboljšati aktivnost. Suprotno tome, stavljanje snažnih akceptora H-veza vjerojatno će smanjiti aktivnost (28, 29).



Slika 19. Generalizirana farmakoforna svojstva za saracatinib skup podataka na referentnoj molekuli saracatiniba određena „Activity atlas" modulom. a) Prosječna molekulska polja aktivnih molekula (engl. Average properties of actives); b) Sažetak značajne promjene aktivnosti (engl. Activity cliffs); c) Prosječni oblik aktivnih molekula (engl. Average shape of actives).

2. CILJ RADA

U ovom diplomskom radu glavni je cilj razviti 3D-QSAR modele temeljene na molekulskim poljima za Tip I inhibitore Src kinaze. Prikupit će se literaturno dostupni podaci o aktivnostima Src inhibitora koji će se zatim detaljno pročistiti s obzirom na biološke metode određivanja aktivnosti i strukturnu raznolikost. Zatim, cilj je razviti kvantitativne i kvalitativne modele za predviđanje biološke aktivnosti Src inhibitora koji će ujedno pružiti i mehanistički uvid u strukturno važne parametre za stvaranje protein-ligand kompleksa.

Nadalje, cilj je dobivene 3D-QSAR modele usporediti s pristupom temeljenim na strukturi, gdje će Src inhibitori biti molekulski uklopljeni u literaturno dostupne kristalne strukture DFG-"in" konformacije Src kinaze. Predviđanje biološke aktivnosti funkcijom vrednovanja iz molekulskog uklapanja te ΔG vrijednostima protein-ligand kompleksa minimiziranog MM-GBSA metodom, usporedit će se s rezultatima dobivenim 3D-QSAR modelima temeljenim na molekulskim poljima.

Dobiveni rezultati također će biti uspoređeni s prethodno objavljenim 3D-QSAR modelima, koji su razvijeni koristeći CoMSIA metodu za saracatinibu slične analoge, te MM-GBSA i WaterMap pristupom za dasatinibu slične analoge purina (AP23464).

3. MATERIJALI I METODE

Poznati inhibitori Src kinaze s njihovim biološkim aktivnostima preuzeti su iz ChEMBL (https://www.ebi.ac.uk/chembl/) baze podataka. Preuzeto je 5020 spojeva koji su pročišćeni u programu DataWarrior. Zatim, preostalih 701 spojeva pretvoreno je u .sdf format, učitano u Schrödinger programski paket i pripremljeno pomoću LigPrep protokola.

Dostupne kristalne strukture kompleksa Src kinaze s inhibitorima te strukturne informacije o načinu vezanja unutar aktivnog mjesta preuzete su iz PDB baze podataka (engl. *Protein Data Bank*). Pročišćene i pripremljene su u Schrödinger programskom paketu pomoću protokola za pripremu proteina.

Glide je protokol molekulskog uklapanja (engl. *docking*) unutar Schrödinger programskog paketa. Njegovom primjenom uz standardnu preciznost (engl. *standard precision*; SP) i visoku preciznost (engl. *extra precision*; XP) provedene su studije vezanja liganda. Protein-ligand kompleksi dodatno su minimizirani, a energija vezanja procijenjena pomoću MM-GBSA (engl. *molecular mechanics energies, generalized Born model and solvent accessibility*) protokola. Korišteno je OPLS3 polje sile s udaljenošću fleksibilnih amino-kiselinskih ostataka od 6 Å od ispitivanog liganda.

Pomoću Cresset računalnog paketa provedena je analiza i preklapanje pripremljenih inhibitora na temelju molekulskih polja. Zatim, razvijeni su kvantitativni QSAR modeli pomoću tri statističke tehnike: parcijalne linearne regresije (engl. Partial least squares method, PLS), multiple linearne regresije uz metodu potpornih vektora (engl. support vector machines, SVM) i klasifikacijskog pristupa uz metodu potpornih vektora. Također, razvijeni su kvalitativni statistički modeli pomoću "*Activity Atlas*" pristupa te "*Activity cliff*".

4. REZULTATI I RASPRAVA

Cilj ovog rada bio je ispitati primjenu 3D-QSAR pristupa temeljenog na molekulskim poljima za razvoj modela za tip I inhibitore Src kinaza. Pri razvoju modela korištene su i dostupne kristalne strukture kompleksa Src kinaze sa inhibitorima, te strukturne informacije o načinima vezanja unutar aktivnog mjesta. Stoga, ovaj je pristup kombinacija LBDD (engl. *ligand-based design*) i SBDD (engl. *structure based design*) pristupa.

Do sada, u literaturi su objavljena dva 3D-QSAR modela za inhibitore Src kinaza (30, 31). Prvi model izrađen je na skupu od 156 derivata anilinokinazolina i kinolinkarbonitrila koristeći CoMFA i CoMSIA metode na temelju fleksibilnih konformacija dobivenih molekulskim uklapanjem. Saracatinib (Slika 16. b)) je korišten kao referentna struktura i rezultati ove studije (Tablica 1.) pokazali su statističku značajnost sa zadovoljavajućom prediktivnošću QSAR modela (CoMFA model, q^2=0.51, r^2=0.70; CoMSIA model, q^2=0.46, r^2=0.73).

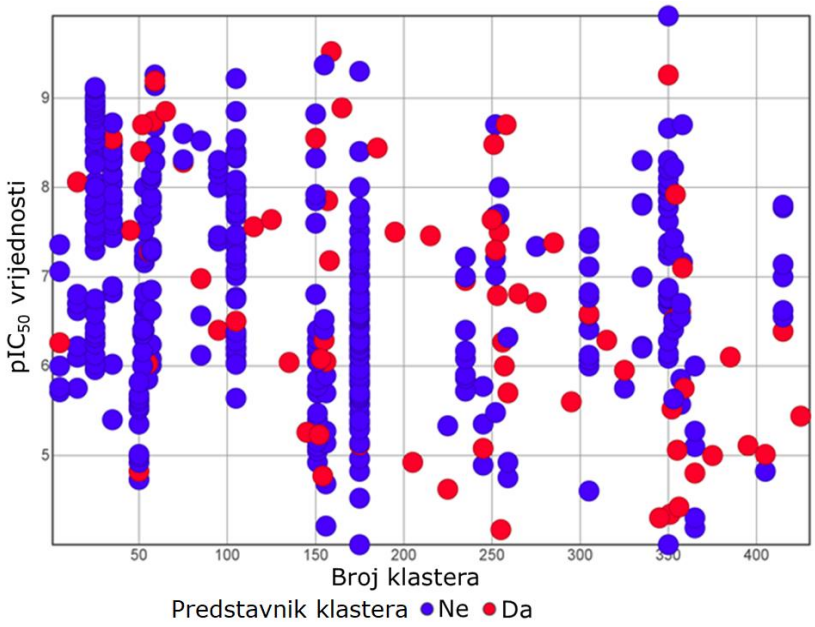
Drugi model razvijen je primjenom WaterMap i Prime MM-GBSA metode na skupu podataka od 49 derivata purina s AP23464 (Slika 16. c)) kao referentnom strukturom. Rezultati studije (Tablica 1.) navode prihvatljivu korelaciju između eksperimentalne i predviđene slobodne energije vezanja (r^2=0.68) za MM-GBSA funkciju bodovanja, dok su vrijednosti za WaterMap (r^2=0.55) i Glide XP (r^2=0.34) funkciju bodovanja ovog skupa podataka pokazale lošiju korelaciju.

Tablica 1. 3D-QSAR modeli za saracatinib (CoMFA i CoMSIA) i AP23464 (MM-GBSA, WaterMap i GlideScore) skupove podataka.

Skup podataka	Broj spojeva	3D model			
Saracatinib	156	CoMFA		CoMSIA	
		Q^2=0.51	R^2=0.70	Q^2=0.46	R^2=0.73
AP23464	49	MM-GBSA	WaterMap	GlideScore	
		R^2=0.68	R^2=0.55	R^2=0.34	

4.1. Kvantitativni 3D-QSAR modeli

U izgradnji QSAR modela, prvi je korak prikupljanje i pročišćavanje eksperimentalnih podataka za svojstvo koje se istražuje, jer je kvalitetni skup podataka preduvjet za izradu kvalitetnih statističkih modela. Eksperimentalni podaci za aktivnosti Src inhibitora prikupljeni su u ChEMBL bazi (IC_{50} vrijednosti) te su pročišćeni na slijedeći način. U probiru od 5020 spojeva uklonjeni su spojevi koji nisu imali definiranu kemijsku strukturu, mali fragmenti (MW<290), peptidi i makrocikli (MW>700), kovalentni inhibitori, polifenoli, PAIN molekule te neselektivni staurosporinski analozi. Također, uklonjeni su spojevi koji nisu imali izmjerene pIC_{50} vrijednosti te spojevi s aktivnostima izmjerenim na staničnim esejima i esejima koji mjere vezanje na domene udaljene od ATP veznog mjesta. Nakon postupka pročišćavanja podataka korišten je generički postupak klasteriranja na temelju strukturne sličnosti koji međutim nije proizveo zadovoljavajuće grupiranje prema određenim kemotipovima, tj. načinima vezanja inhibitora u aktivno mjesto Src kinaze (Slika 20.).

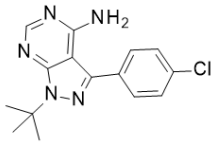


Slika 20. Raspodjela pIC_{50} vrijednost u pojedinim klasterima za skup podataka od 701 inhibitora.

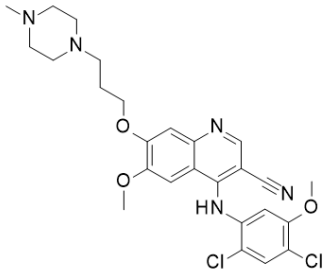
Ispravno preklapanje molekula osnovni je preduvjet za izgradnju 3D-QSAR modela pa su podjela inhibitora u strukturno srodne skupine te odabir referentne strukture iznimno važni.

Preostalih 701 inhibitora grupirano je u tri kategorije prema strukturnoj sličnosti i zajedničkom načinu vezanja na protein pri čemu su za referentne strukture odabrane PP2, saracatinib/botsutinib, dasatinib i AP23464 (Slika 21). Analozi svakog skupa vežu se poput njihovih referentnih struktura.

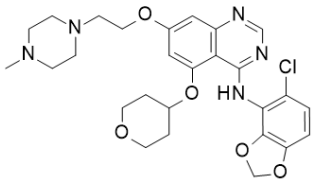
a) PP2



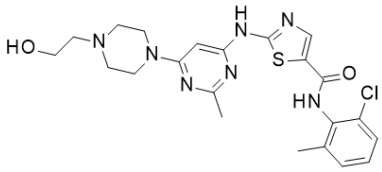
b) Bosutinib (Bosulif, Wyeth)



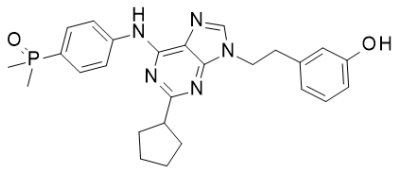
Saracatinib (AstraZeneca)



c) Dasatinib (Sprycel, Bristol-Meyers Squib)



AP23464 (Ariad Pharmaceuticals)

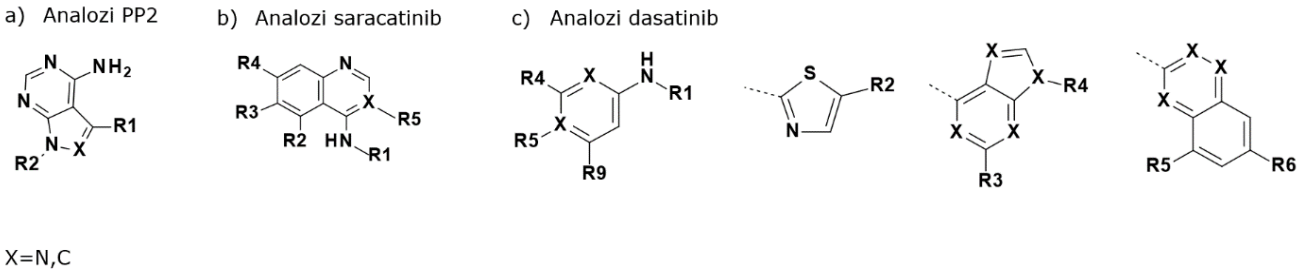


Slika 21. Strukture referentnih molekula za tri skupa analoga Src inhibitora: a) PP2; b) Bosutinib (Bosulif, Wyeth), Saracatinib (AstraZeneca); c) Dasatinib (Sprycel, Bristol-Meyers Squib) i AP23464 (Ariad Pharmaceuticals).

Skup podataka *PP2 analozi* sastoji se od 297 spojeva. Istraženo je nekoliko kombinacija referentnih molekula na koje je preklopljen PP2 skup, uključujući različite PP2 analoge dostupne iz X-ray struktura, kao i referentne strukture drugih istraženih kemotipova poput dasatiniba i

saracatiniba. Najbolji modeli dobiveni su koristeći PP2/dasatinib/saracatinib kao referentne molekule kako bi se obuhvatile strukturne varijacije u PP2 skupu podataka i osiguralo pravilno preklapanje spojeva budući da su PP2 analozi relativno male molekule. Farmakoforna ograničenja korištena su za atome koji se vežu za zglobno (engl. *hinge*) mjesto, a ograničenja hidrofobnog polja za preklapanje aromatskih prstenova u referentnim molekulama.

U drugom skupu podataka, *Saracatinib analozi*, nalazi se 248 spojeva sa saracatinibom i bosutinibom kao referentnim molekulama. *Dasatinib analozi* čine skup od 156 spoja, a dasatinib, saracatinib i AP23464 korišteni su kao referentne molekule. Na Slici 22. prikazane su osnovne heterocikličke strukture (engl. *scaffolds*) za sva tri seta molekula sa označenim vektorima diverzifikacije kojima se povećava strukturna raznolikost.

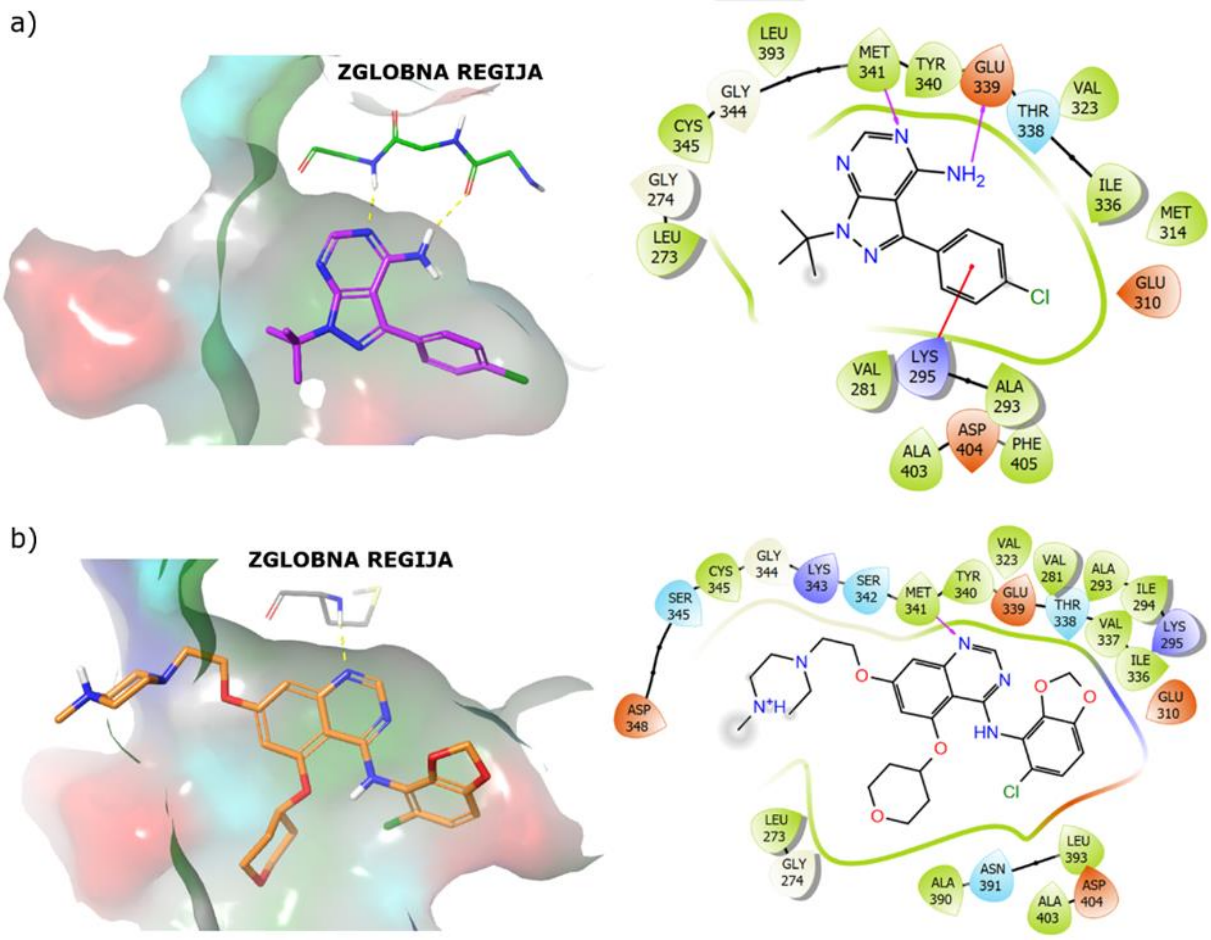


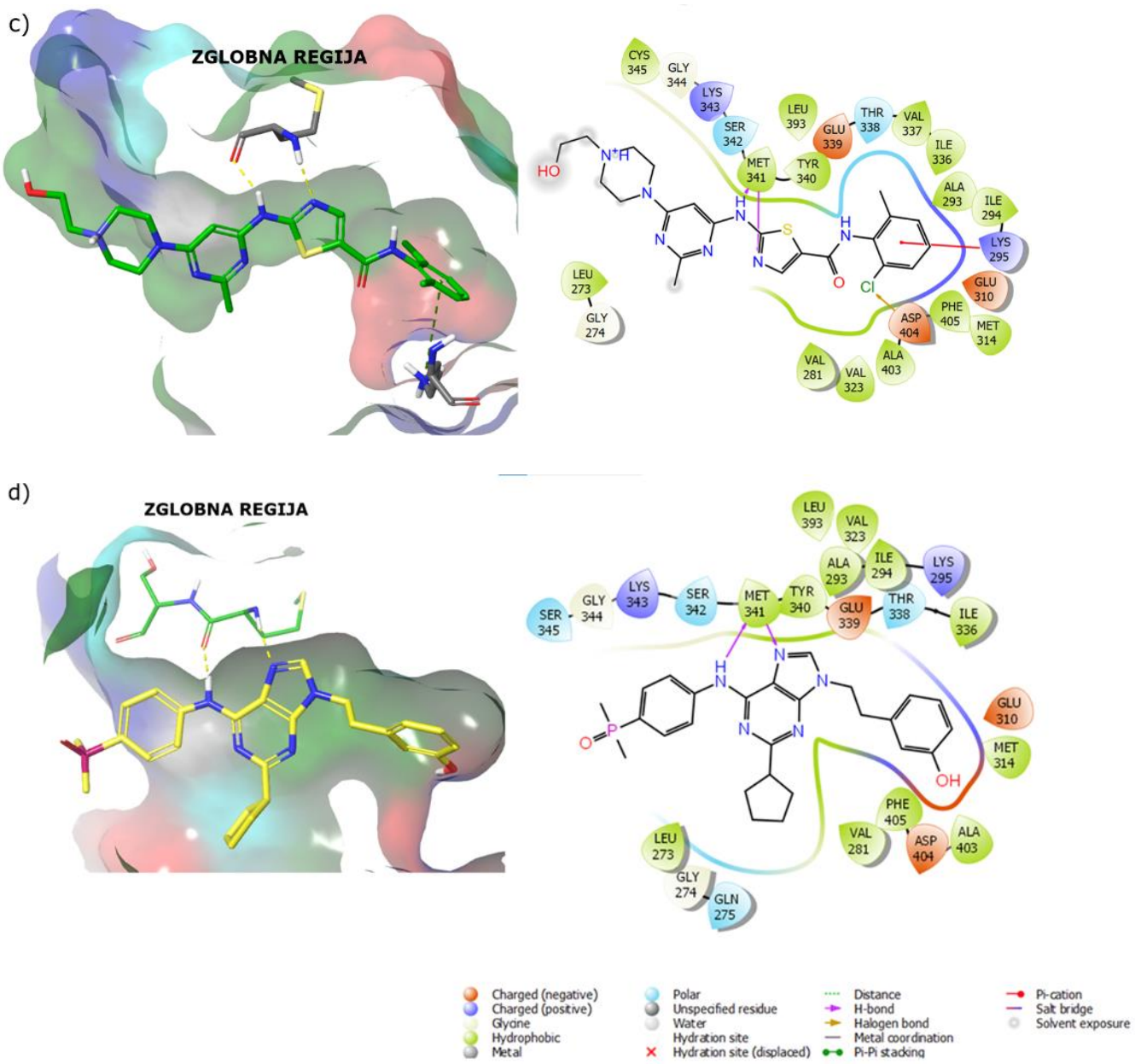
Slika 22. Shematski prikaz strukturne raznolikosti a) PP2 analoga; b) Saracatinib analoga; c) Dasatinib analoga korištenih za razvoj 3D-QSAR modela.

Vezanje referentnih struktura u veznom mjestu Src kinaze prikazano je na slici 23. Budući da su u ovom radu istraživani tip 1 inhibitori Src kinaza, korištena je DFG-"in" konformacija enzima.

Na slici 23. a) vidljivo je da PP2 molekula stvara dvije vodikove veze unutar zglobne (engl. *hinge*) regije s Met341 i Glu339, dok klorfeniln stvara π-π stacking interakcije s Lys295. Saracatinib je jedina molekula

koja stvara samo jednu vodikovu vezu (Met341) (Slika 23. b)) sa zglobnom regijom. Molekule dasatinib (Slika 23. c)) i AP23464 (Slika 23. d)) stvaraju dvije vodikove veze, ali imaju motiv vezanja na HBD-HBA iste aminokiseline (Met341) preko aminoheterocikla. Dasatinib skup, osim direktnih analoga dasatiniba, također sadrži i purinske 5+6 derivate (AP23464 analozi), kinolinske/kinazolinske/benzotriazolinske 6+6 derivate te tiazolonske derivate.





Slika 23. Kristalne strukture referentnih molekula u veznom mjestu Src kinaze, 3D prikaz (lijevo) i 2D prikaz (desno): a) PP2 (PDB ID: 3GEQ); b) Sarcatinib (PDB ID: 2H8H); c) Dasatinib (PDB ID: 3G5D); d) AP23464 (PDB ID: 2BDJ).

Opisane kristalne strukture referentnih molekula korištene su kao predlošci za izradu 3D-QSAR modela u dva osnovna koraka: preklapanje molekula i izrada modela. U prvom koraku preklapanja, zbog fleksibilnosti molekula, korišteno je detaljno pretraživanje konformacijskog prostora s maksimalnim brojem od 1000 mogućih konformacija, 0.1 kcal/mol/A kao granica gradijenta pri minimizaciji konformera, te su sve amidne veze

ograničene na trans izomer. Uzete su u obzir konformacije do 3 kcal/mol nestabilnije od konformacije najniže energije te su preklopljene konformacije ručno pregledane za sva tri skupa molekula. Da bismo osigurali pravilno preklapanja molekula definirana su strukturna ograničenja karakteristična za vezanje na zglobnu regiju, donor i akceptor vodikovih veza te lipofilna polja aromatskih prstenova u referentnim molekulama, te je sam protein korišten za definiranje volumena u kojem se dobivene konformacije mogu smjestiti (*engl. Excluded volume*). U drugom koraku izrađeni su QSAR modeli koristeći tri različite statističke tehnike:

1. Parcijalna linearna regresija (engl. *Partial least squares method*, PLS)
2. Multipla linearna regresija uz metodu potpornih vektora (engl*. support vector machines,* SVM)
3. Klasifikacijski pristup uz metodu potpornih vektora

Također, razvijeni su kvalitativni modeli koristeći "*Activity Atlas*" pristup (str. 33), pri čemu su dobivene informacije o poopćenim farmakofornim svojstvima za svaki skup molekula te strukturnim svojstvima koja uzrokuju neočekivane promjene u aktivnosti molekula (engl*. activity cliffs*). Rezultati svih modela prikazani su u Tablici 2.

Najbolji 3D-QSAR model razvijen je za saracatinib skup podataka koji sadrži 248 spojeva od kojih je 199 pridodano setu za učenje, a preostalih 49 setu za testiranje. PLS metodom koeficijent korelacije između predviđenih i eksperimentalnih pIC_{50} vrijednosti pokazuje dobru korelaciju na setu za učenje (r^2=0.92) i prihvatljivu prediktivnu moć (q^2=0.63). Koeficijent korelacije za test set pokazuje također dobru korelaciju (r^2=0.85). SVM metodom, također, dobiva se dobra korelacija između predviđenih i eksperimentalnih pIC_{50} vrijednosti na setu za učenje (r^2=0.95), ali nešto manja prediktivna moć (q^2=0.60), na što ukazuje i niski koeficijent korelacije za test set (r^2=0.28) (Tablica 2.).

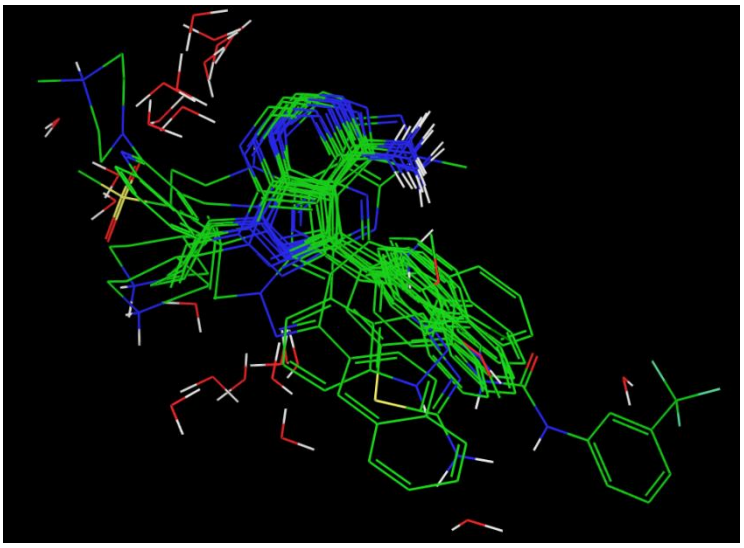
Tablica 2. Statistički parametri 3D-QSAR modela temeljenih na molekulskim poljima za PP2, Saracatinib, Dasatinib i Zajednički set skupova podataka.

Skup podataka	Referentne strukture	Broj spojeva (n)			PLS*				SVM*		
		n	n(set za učenje)	n(test)	R^2(set za učenje)	Q^2(set za učenje)	R^2(test)	n(PC)	R^2(set za učenje)	Q^2(set za učenje)	R^2(test)
PP2 analozi	PP2 Dasatinib Saracatinib	**297**	237	60	0.81	0.25	0.69	8	0.98	0.35	0.56
Saracatinib analozi	Saracatinib Bosutinib	**248**	199	49	0.92	0.63	0.85	7	0.95	0.60	0.28
Dasatinib analozi	Dasatinib Saracatinib AP23464	**156**	130	26	0.99	0.56	0.83	11	0.96	0.49	0.71
Zajednički set		**701**	561	140	0.90	0.47	0.65	13	0.99	0.53	0.56

* parcijalna linearna regresija (engl. *partial least squares method*, *PLS*)

* metoda potpornih vektora (engl*. support vector machines, SVM*)

U PP2 skupu podataka nalazi se 297 spojeva od kojih je 237 pridodano setu za učenje, a 60 u setu za testiranje. Koeficijent korelacije između predviđenih i eksperimentalnih pIC_{50} vrijednosti pokazuje dobru korelaciju (r^2=0.81) na setu za učenje, ali najlošiju prediktivnu moć (q^2=0.25) od svih modela. Koeficijent korelacije za test set također pokazuje prihvatljivu korelaciju (r^2=0.69). Za SVM metodu, koeficijent korelacije između predviđenih i eksperimentalnih pIC_{50} vrijednosti na setu za učenje pokazuje odličnu korelaciju (r^2=0.98), ali najlošiju prediktivnu moć (q^2=0.35). Koeficijent korelacije za set test pokazuje prihvatljivu korelaciju (r^2=0.56) (Tablica 2.). Kako bi se objasnila slabija kvaliteta modela za PP2 skup, dodatno su analizirane dostupne kristalne strukture kompleksa PP2 analoga sa Src kinazom. S obzirom na to da je PP2 molekula male molekulske mase, molekule analozi s različitim supstituentima imaju dovoljno prostora unutar aktivnog mjesta za različite načine vezanja, tj. različite orijentacije kako bi optimirale interakcije s proteinom te stvaraju interakcije preko molekula vode. Na Slici 24. prikazane su različite pozicije suptituenata PP2 analoga u kristalnim strukturama koje onemogućavaju jednoznačno preklapanje struktura i ukazuju na ograničenja računskih pristupa temeljenih na strukturama liganada (LBDD). Također, pri preklapanjima nije uzet u obzir utjecaj molekula vode na način vezanja. Zbog tih razloga pretpostavljamo da je model za PP2 analoge nešto slabije statističke značajnosti.



Slika 24. Preklopljene kristalne strukture PP2 analoga.

Bolji 3D-QSAR model dobiven je za dasatinib skup podataka. U njemu se nalazi 156 spojeva od kojih je 130 u setu za učenje, a preostalih 26 u setu za testiranje. PLS metodom dobivena je odlična korelacija između predviđenih i eksperimentalnih pIC_{50} vrijednosti (r^2=0.99), ali nešto slabija prediktivna moć (q^2=0.56). Koeficijent korelacije za test set pokazuje dobru korelaciju (r^2=0.83). SVM metodom dobivena je također vrlo dobra korelacija između predviđenih i eksperimentalnih pIC_{50} vrijednosti (r^2=0.96), no slabija prediktivna moć (q^2=0.49). Koeficijent korelacije za test set pokazuje dobru korelaciju (r^2=0.71).

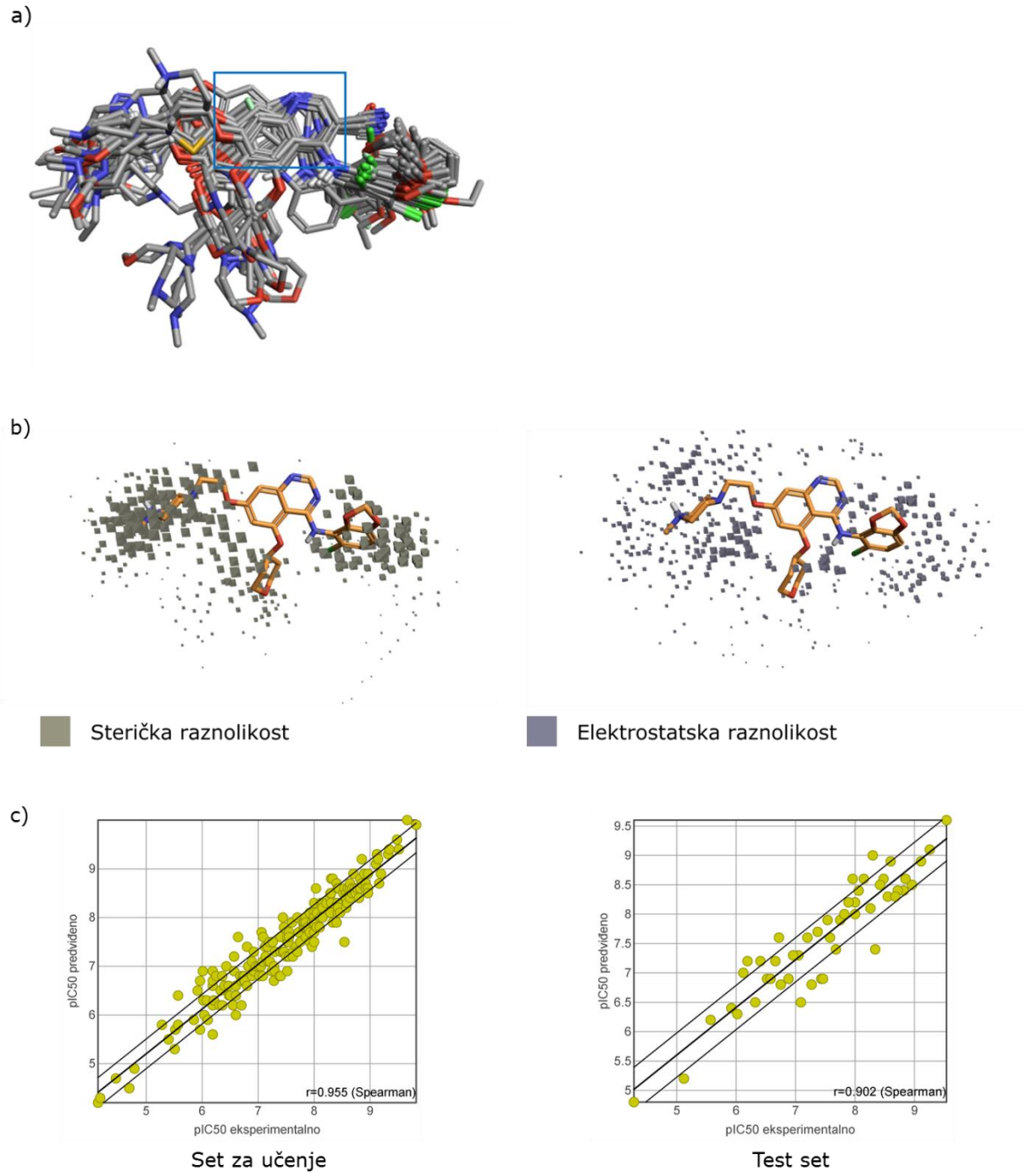
Razvijeni modeli primjenjivi su za bliske analoge pojedinih referentnih molekula, pri čemu se može pouzdano predvidjeti aktivnost novih potencijalnih inhibitora te racionalno dizajnirati molekule sa strukturnim svojstvima koje doprinose povećanju aktivnosti. Primjenjivost razvijenih modela na strukturno različitije spojeve mogla bi se dalje ispitati na neovisnom validacijskom setu. Budući da su razvijeni kvalitetni statistički modeli za pojedine setove podataka, u sljedećem koraku razvijen je zajednički model za sve spojeve, koristeći molekule preklopljene u pojedinačnim modelima. Ovim postepenim postupkom osigurano je

dosljedno preklapanje struktura unutar svakog strukturnog skupa, što je vrlo važno kod izgradnje 3D-QSAR modela.

3D-QSAR model za zajednički skup podataka obuhvaća 701 spoj od kojih je 561 pridodano setu za učenje i preostalih 140 setu za testiranje. PLS metoda pokazuje dobru korelaciju između predviđenih i eksperimentalnih pIC_{50} vrijednosti (r^2=0.90), no slabu prediktivnu moć (q^2=0.47). Koeficijent korelacije za test set pokazuje prihvatljivu korelaciju (r^2=0.65). SVM metodom koeficijent korelacije između predviđenih i eksperimentalnih pIC_{50} vrijednosti pokazuje još bolju korelaciju (r^2=0.99) i zadovoljavajuću prediktivnu moć (q^2=0.53). Koeficijent korelacije za test set također pokazuje prihvatljivu korelaciju (r^2=0.56).

SVM metoda u nekim slučajevima daje točnije predviđanje na setovima za učenje. Međutim, zbog slabije prediktivne moći za test set i unakrsno validacijskog q^2 smatramo da su modeli dobiveni PLS metodom robusniji. Usporedimo li rezultate svih modela (*Saracatinib analozi, PP2 analozi, Dasatinib analozi*) i *zajedničkog skupa* podataka s već navedenim literaturnim vrijednostima, možemo reći da su rezultati značajno bolji nego dosad objavljenih 3D-QSAR modela za Src inhibitore. Prikazani rezultati ukazuju da je za *Saracatinib analoge* razvijen model visoke točnosti, stabilnosti i prediktivnosti te se može pouzdano koristiti za predviđanje aktivnosti virtualnih biblioteka spojeva sličnih saracatinibu. Također, u ovom istraživanju korišten je veći broj strukturno različitih molekula za izradu modela. Stoga očekujemo da su modeli robusniji i da imaju bolju prediktivnu moć od prije objavljenih modela.

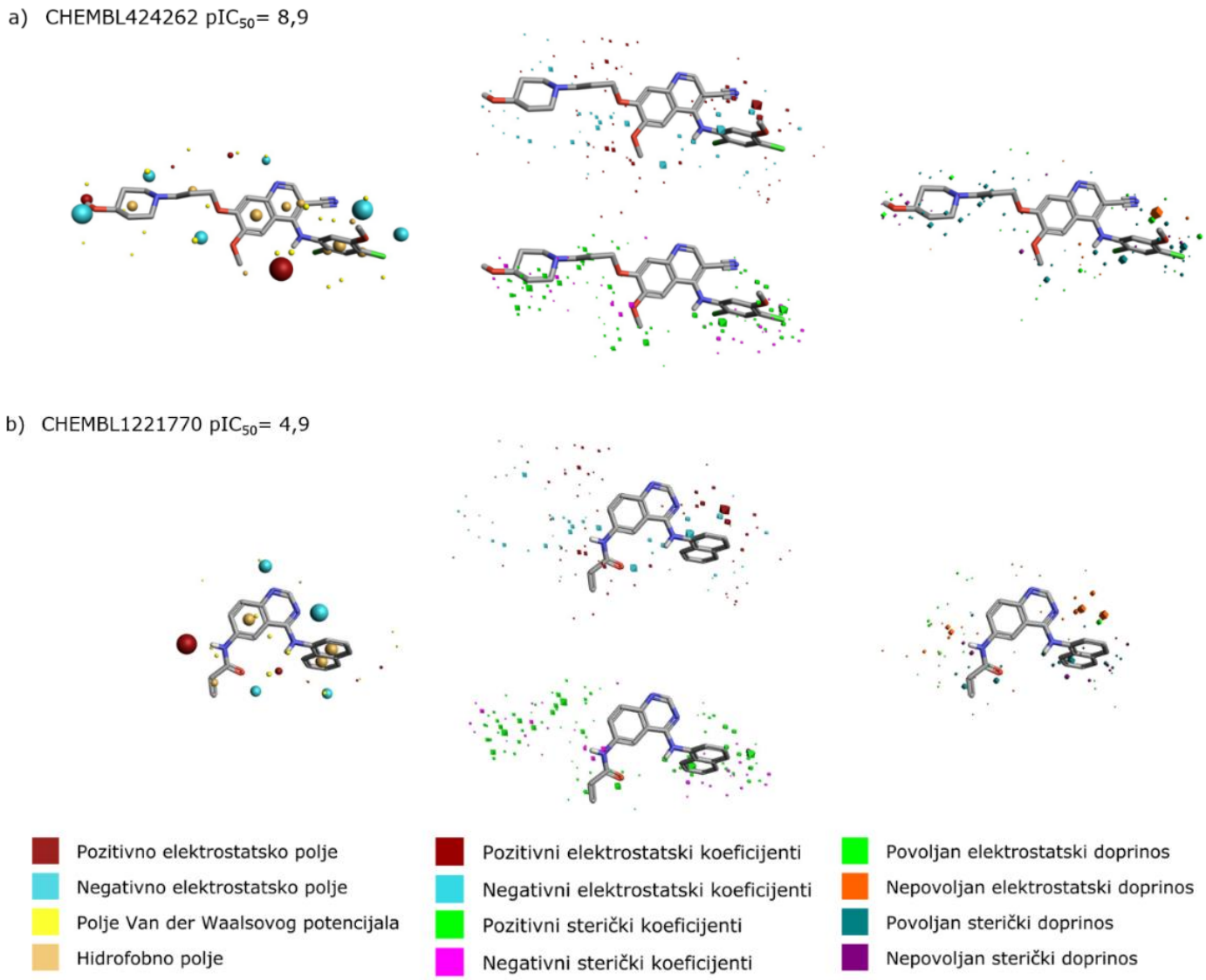
Budući da su vrijednosti saracatinib skupa podataka pokazale najbolje statističke parametre tj. najtočniji model, detaljnije su prikazani rezultati i strukturne informacije koje pruža 3D-QSAR model. Analogna analiza prikazana je i za *zajednički skup* podataka.



Slika 25. 3D-QSAR model za saracatinib skup podataka. a) Preklopljeni saracatinib analozi b) Sterička (lijevo) i elektrostatska (desno) raznolikost molekula korištenih u modelu c) Korelacija eksperimentalnih i predviđenih pIC_{50} vrijednosti za set za učenje (lijevo) i test set (desno).

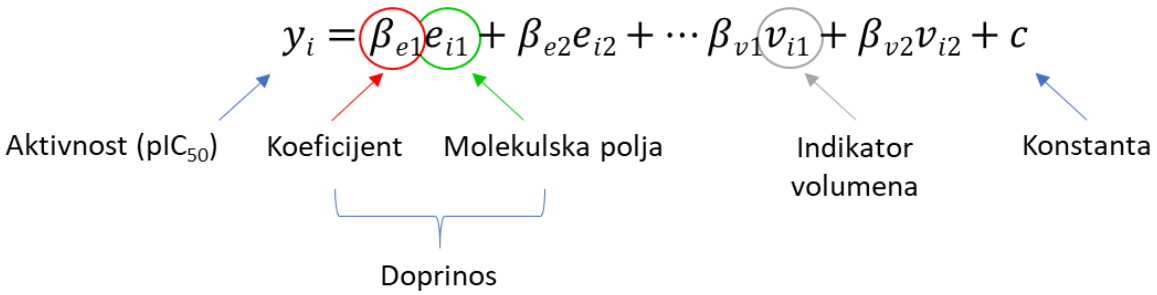
Na slici 25. a) prikazani su preklopljeni svi saracatinib analozi. Sterička i elektrostatska raznolikost skupa podataka za saracatinib analoge

prikazana je na slici 25. b). Vidljivo je da nema strukturnih promjena na osnovnoj benzo-pirimidinskoj strukturi, dok su najveće varijacije na tri mjesta supstitucije kao što je prikazano općom formulom na slici 22. b) Korelacija između eksperimentalnih i predviđenih pIC_{50} vrijednosti za set za učenje i test set prikazane su na slici 25. c).



Slika 26. Molekulska polja (lijevo), elektrostatski i sterički koeficijenti (sredina), te elektrostatski i sterički doprinosi (desno) određeni 3D-QSAR modelom na Saracatinib skupu podataka za a) aktivnu molekulu (CHEMBL424262, pIC_{50}=8,9) i b) neaktivnu molekulu (CHEMBL1221770 pIC_{50}=4,9).

Molekulska polja, elektrostatski i sterički koeficijenti određeni 3D-QSAR modelom na *Saracatinib skupu* podataka, te elektrostatski i sterički doprinosi za CHEMBL424262 kao primjer vrlo aktivne molekule i CHEMBL1221770 kao primjer neaktivne molekule prikazani su na slici 26. Elektrostatski i sterički doprinosi izračunati su pomoću elektrostatskih i steričkih koeficijenata i molekulskih polja za pojedine molekule kao što je prikazano u jednadžbi (Slika 27.), te je vidljivo da aktivna molekula ima veće povoljne elektrostatske i steričke doprinose aktivnosti dok neaktivna molekula ima izraženiji nepovoljan elektrostatski doprinos i manje povoljan sterički doprinos. Dakle, umnožak PLS koeficijenta s odgovarajućim molekulskim poljem generira parcijalni doprinos modelu te njihova suma po svim molekulskim poljima daje predviđanje aktivnosti.

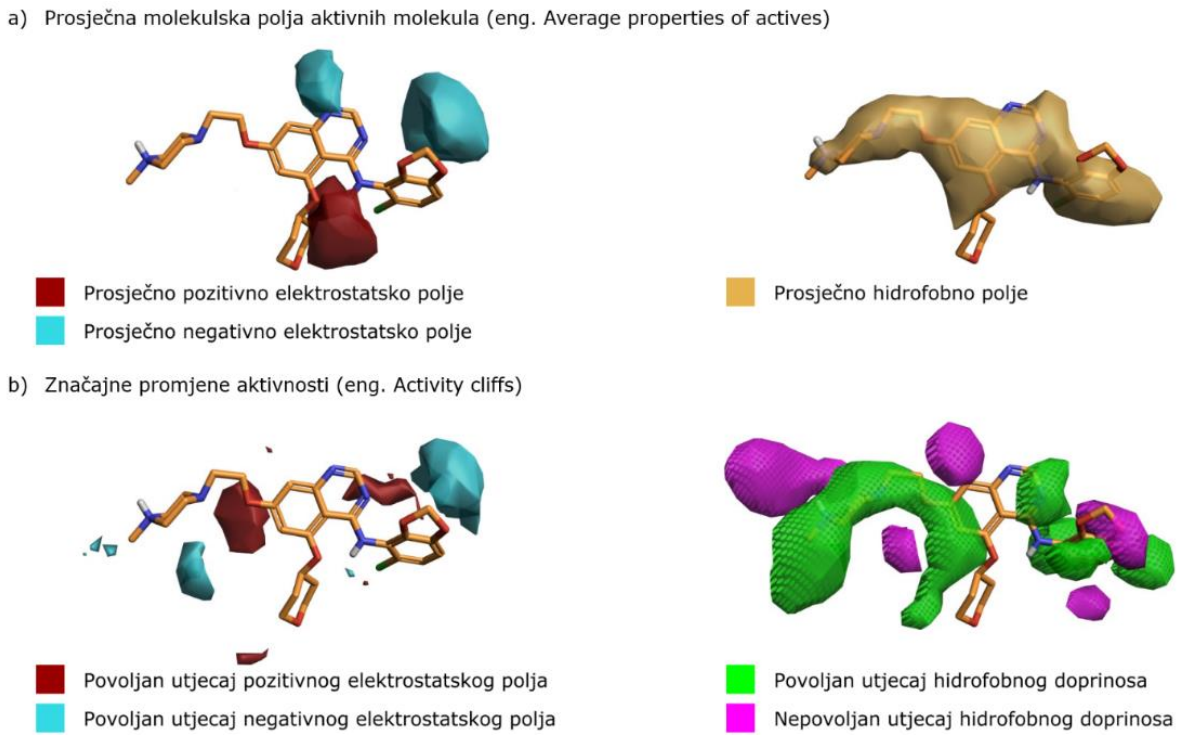


Slika 27. Jednadžba izračuna aktivnosti molekule (y_i).

4.2. Farmakoforni modeli

Osim kvantitativnih 3D-QSAR modela razvijeni su i kvalitativni modeli koji prikazuju strukturna i elektrostatska svojstva prisutna kod svih aktivnih molekula te strukturne promjene koje uzrokuju neočekivanu promjenu aktivnosti (engl. *Activity cliffs*). Tako dobivene rezultate nazivamo generaliziranim farmakofornim modelom, a doprinose boljem razumijevanju čimbenika koji doprinose aktivnosti molekula te diskontinuitetu odnosa između strukture i aktivnosti. Kvalitativni 3D model

za *Saracatinib analoge* određen "*Activity atlas*" modulom prikazan je na slici 28.



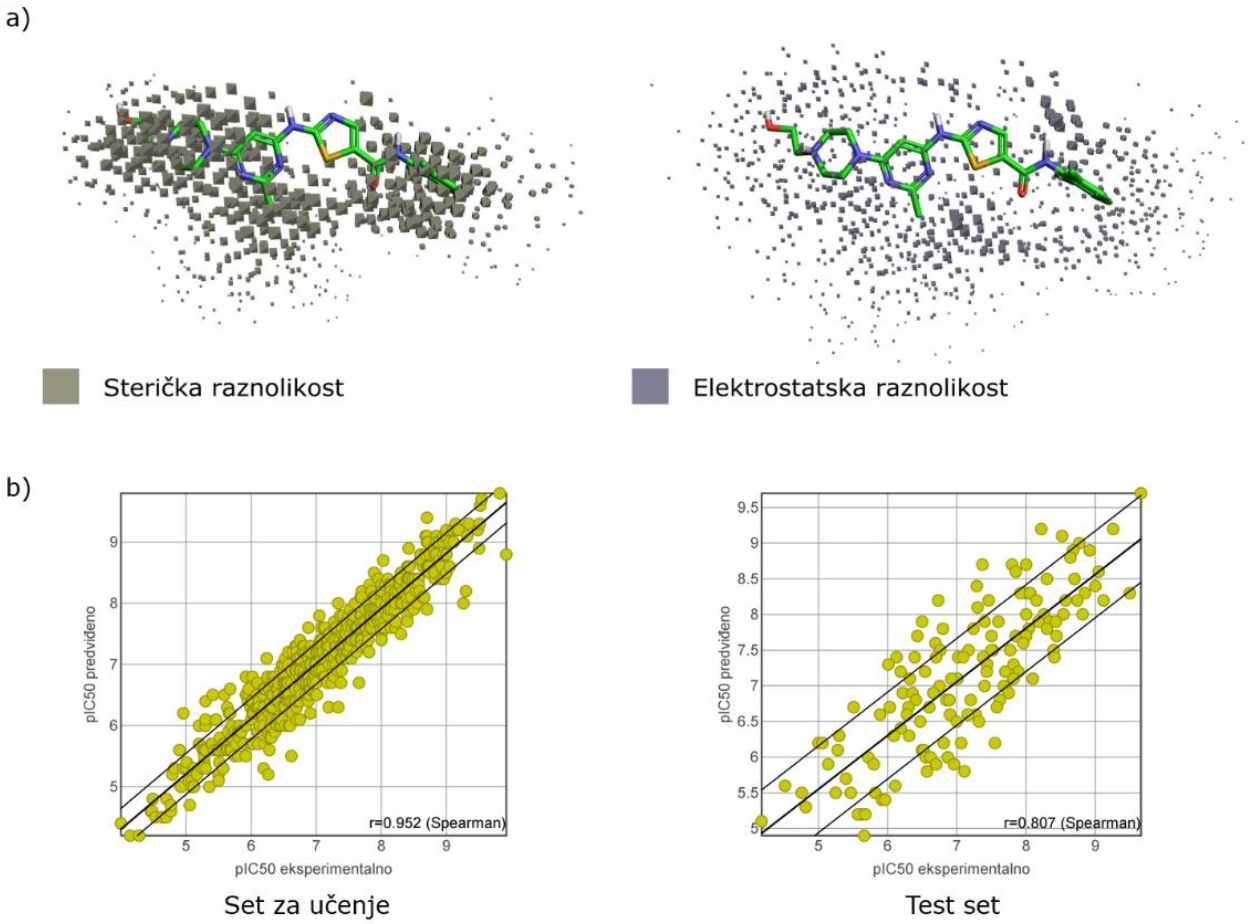
Slika 28. Generalizirana farmakoforna svojstva za saracatinib skup podataka prikazana na referentnoj molekuli saracatiniba određena „Activity atlas" modulom. a) Prosječna molekulska polja aktivnih molekula (engl. Average properties of actives); b) Sažetak neočekivanih promjena aktivnosti u odnosu na promjenu strukture molekula (engl. Activity cliffs).

Prosječno pozitivno (crveno) i negativno (svijetlo plavo) elektrostatsko polje prikazano je na slici 28. a) lijevo, a desno je prikazano prosječno hidrofobno polje (žuto), prisutno kod većine molekula u promatranom skupu molekula. Prosječna molekulska polja aktivnih molekula ukazuju na očekivane elektrostatske i hidrofobne značajke aktivnih molekula za proučavani skup molekula. Negativno elektrostatsko polje je u području interakcije sa zglobnom regijom, jer saracatinib analozi imaju akceptor vodikove veze. Također, aktivni analozi imaju pozitivan elektrostatski potencijal koji potječe od N-H skupine koja nije u interakciji sa Src

kinazom nego utječe na konformacijsku rigidnost stvaranjem intramolekulske vodikove veze. Na slici 28. b) prikazane su promjene u elektrostatskim i hidrofobnim molekulskim poljima koja mogu doprinijeti daljnjem povećanju aktivnosti. Plava područja ukazuju da bi stvaranje negativnijeg (ili manje pozitivnog) područja moglo poboljšati aktivnost, a crvena područja ukazuju da stvaranje pozitivnijeg (ili manje negativnog) područja može doprinjeti povećanju aktivnosti (Slika 28. b), lijevo). Zelena područja na slici 28. b) desno pokazuju mjesta povoljnog utjecaja hidrofobnog i/ili steričkog doprinosa dok ružičasta područja označavaju nepovoljan utjecaj.

Osim za saracatinib skup, u nastavku je prikazana analiza za *zajednički skup* molekula. Sterička i elektrostatska raznolikost skupova podataka za *zajednički skup* podataka prikazana je na slici 29. a) s dasatinibom kao reprezentativnom molekulom. Iako je dasatinib korišten kao reprezentativna molekula, ovaj model podataka pokriva sve molekule preklopljene u pojedinačnim modelima za tri proučavana skupa.

Na području zglobne regije preklapaju se osnovne strukture svih referentnih molekula (engl. *scaffold*), područje je najmanje strukturne raznolikosti i najveće varijacije nalaze se na tri mjesta supstitucije kao što je prikazano općom formulom na slici 22. Korelacija između eskperimentalnih i predviđenih pIC_{50} vrijednosti za set za učenje i test set priikazane su na slici 29. b).

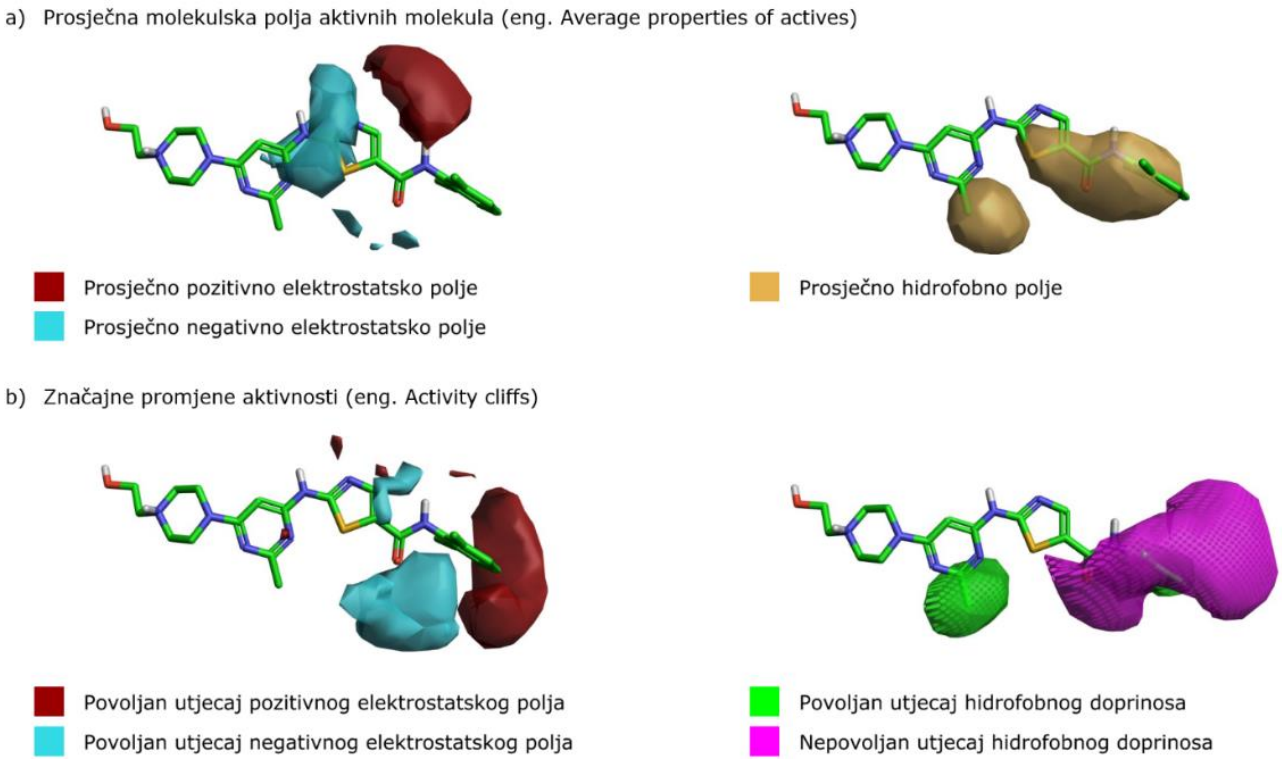


Slika 29. 3D-QSAR model za zajednički skup podataka. a) Sterička (lijevo) i elektrostatska (desno) raznolikost b) Korelaciija eksperimentalnih i predviđenih pIC_{50} vrijednosti za set za učenje (lijevo) i test set (desno).

Dobiveni generalizirani farmakoforni model za *zajednički skup* podataka određen *„Activity atlas"* modulom prikazan je na slici 30.

Kao što je slučaj za *Saracatinib analoge, zajednički skup* podataka također ima negativno elektrostatsko polje u području interakcije sa zglobnom regijom, jer svi inhibitori imaju akceptor vodikove veze. Također, aktivni analozi imaju pozitivan elektrostatski potencijal koji potječe od N-H skupine koja nije u interakciji sa Src kinazom nego utječe na konformacijsku rigidnost stvaranjem intramolekulske vodikove veze. Na slici 30. b) lijevo prikazane su površine pozitivnih i negativnih elektrostatskih potencijala, a desno područja povoljnog i nepovoljnog

hidrofobnog doprinosa. Molekule čija se elektrostatska i hidrofobna polja preklapaju sa generaliziranim farmakofornim svojstvima vrlo vjerojatno će biti aktivne te se na taj način mogu dizajnirati i odabrati nove target molekule.



Slika 30. Generalizirana farmakoforna svojstva za Zajednički skup podataka na referentnoj molekuli dasatiniba određena „Activity atlas" modulom. a) Prosječna molekulska polja aktivnih molekula (engl. Average properties of actives); b) Sažetak značajne promjene aktivnosti (engl. Activity cliffs).

4.3. Klasifikacijski 3D-QSAR modeli

Zbog niže statističke značajnosti modela za *PP2 analoge* i *zajedničkog skupa* podataka, osim kvantitativnih QSAR modela, dodatno su razvijeni klasifikacijski modeli kako bi se provjerilo može li se postići bolje predviđanje za molekule ispitivanih skupova. Za *PP2 analoge* i *Zajednički*

skup podataka klase su definirane tako da su spojevi u kategoriji s pIC_{50} <6.5 neaktivni, spojevi u kategoriji 6.5<pIC_{50}≥7.5 srednje aktivni, a oni koji imaju pIC_{50}≥7.5 su aktivni. Za *Dasatinib* i *Saracatinib analoge* pomaknute su klase prema višim aktivnostima, jer je raspodjela vrijednosti pIC_{50} nagnuta prema većim vrijednostima. U prvoj kategoriji nalaze se neaktivni spojevi s pIC_{50}<7.0, srednje aktivni spojevi su u rasponu aktivnosti od 7.0 do 8 te su aktivni oni čiji je pIC_{50}≥8.0. Najbolji modeli dobiveni su kod podjele skupa podataka na tri klase zato što je potreban približno podjednak broj molekula u svakoj klasi. Dobiveni rezultati prikazani su u Tablicama 1., 2., 3. i 4. Klasifikacijski modeli *PP2* i *Zajednički skup* su usporedive kvalitete i slabiji od klasifikacijskih modela za *Saracatinib* i *Dasatinib skup*.

Klasifikacijski 3D-QSAR model za *Zajednički skup* podataka (Tablica 3.b) lijevo) visoke je preciznosti i osjetljivosti; preko 0.9 na setu za učenje. Od 219 vrlo aktivnih molekula pravilno ih je svrstano u treću klasu 209, dok je 10 molekula raspodijeljeno u klasu srednje aktivnosti. Nijedna molekula nije svrstana u klasu neaktivnih molekula. U drugoj kategoriji je 174 od 182 srednje aktivnih molekula, a 6 molekula je raspodijeljeno pod kategoriju neaktivnih i 2 pod aktivne. U prvoj kategoriji od 160 neaktivnih molekula našlo se 156, dok je 4 molekula raspodijeljeno u klasu srednje aktivnosti i nijedna nije svrstana u klasu aktivnih molekula. Na setu za testiranje dobivena je prihvatljiva preciznost (0.599) i osjetljivost modela (0.603) (Tablica 3. b) desno). Najvjerojatnije je predviđanje za test set u ovom klasifikacijskom modelu slabije nego u kvalitativnom modelu zbog uskog područja aktivnosti u srednjoj klasi.

Tablica 3. Klasifikacijski 3D-QSAR model za Zajednički skup podataka: a) definicija kategorija aktivnosti b) matrice zabune, preciznost i osjetljivost modela na setu za učenje i testnom setu.

a)

Vrijednosti (pIC_{50})	Broj spojeva u kategoriji
<6.5	160
≥6.5 <7.5	182
≥7.5	219

b)

Set za učenje

pIC_{50} eksperimentalno \ pIC_{50} predviđeno		1	2	3
	1	156	4	0
	2	6	174	2
	3	0	10	209
Preciznost		0.960		
Osjetljivost		0.962		

Test set

pIC_{50} eksperimentalno \ pIC_{50} predviđeno		1	2	3
	1	24	13	3
	2	14	20	11
	3	5	8	42
Preciznost		0.599		
Osjetljivost		0.603		

Za *PP2 analoge* (Tablica 4. b) lijevo) dobivena je viša preciznost i osjetljivost modela od preko 0.95 na setu za učenje. Od 97 neaktivnih molekula pravilno ih je svrstano u prvu kategoriju 94, dok je 3 molekule raspodijeljeno u klasu srednje aktivnosti. Nijedna molekula nije svrstana u klasu aktivnih molekula. U drugoj kategoriji je 81 od 86 srednje aktivnih molekula, a 4 molekule su raspodijeljene pod kategoriju neaktivnih i 1 pod aktivne. U trećoj kategoriji od 54 aktivne molekule pravilno ih je raspoređeno 51, dok su 2 molekule raspodijeljene u klasu srednje aktivnosti i jedna u klasu neaktivnih molekula. Na setu za testiranje dobivena je zadovoljavajuća preciznost (0.631) i osjetljivost modela (0.634) (Tablica. 4. b) desno).

Tablica 4. Klasifikacijski 3D-QSAR model za PP2 analoge: a) definicija kategorija aktivnosti b) matrice zabune, preciznost i osjetljivost modela na setu za učenje i testnom setu.

a)

Vrijednosti (pIC_{50})	Broj spojeva u kategoriji
<6.5	97
≥6.5 <7.5	86
≥7.5	54

b)

Set za učenje

pIC_{50} eksperimentalno \ pIC_{50} predviđeno		1	2	3
	1	94	3	0
	2	4	81	1
	3	1	2	51
Preciznost		0.957		
Osjetljivost		0.952		

Test set

pIC_{50} eksperimentalno \ pIC_{50} predviđeno		1	2	3
	1	19	5	1
	2	9	9	3
	3	0	4	10
Preciznost		0.631		
Osjetljivost		0.634		

Najbolji rezultati klasifikacijskog modela dobiveni su za *Saracatinib analoge* s preciznosti i osjetljivosti od 100% na setu za učenje (Tablica 5. b) lijevo). U prvoj kategoriji je 66 molekula koje su neaktivne, srednje aktivnih u drugoj kategoriji je 57 i u trećoj je 76 aktivnih molekula. Također, dobiveni su najbolji rezultati na setu za testiranje s preciznosti od 0.681 i osjetljivosti modela od 0.670 (Tablica 5.b) desno).

Tablica 5. Klasifikacijski 3D-QSAR model za Saracatinib analoge: a) definicija kategorija aktivnosti b) matrice zabune, preciznost i osjetljivost modela na setu za učenje i testnom setu.

a)

Vrijednosti (pIC_{50})	Broj spojeva u kategoriji
<7.0	66
≥7.0 <8.0	57
≥8.0	76

b)

Set za učenje

pIC_{50} eksperimentalno \ pIC_{50} predviđeno	1	2	3
1	63	0	0
2	0	57	0
3	0	0	76
Preciznost	1.000		
Osjetljivost	1.000		

Test set

pIC_{50} eksperimentalno \ pIC_{50} predviđeno	1	2	3
1	11	3	2
2	3	6	5
3	1	1	17
Preciznost	0.681		
Osjetljivost	0.670		

Najlošija preciznost (0.861) i osjetljivost (0.847) modela na setu za učenje dobivena je za *Dasatinib analoge* (Tablica 6. b) lijevo). Od 38 vrlo aktivnih molekula pravilno ih je svrstano u treću klasu 29, dok je 9 molekula raspodijeljeno u klasu srednje aktivnosti. Nijedna molekula nije svrstana u klasu neaktivnih molekula, što ipak ukazuje na relativno dobru klasifikaciju. U drugoj kategoriji je 37 od 43 srednje aktivnih molekula, a 3 molekule je raspodijeljeno pod kategoriju neaktivnih i 3 pod aktivne. U prvoj kategoriji od 49 neaktivnih molekula pravilno je raspoređeno 45, dok je 4 molekula raspodijeljeno u klasu srednje aktivnosti i nijedna nije svrstana u klasu aktivnih molekula. Na setu za testiranje dobivena je zadovoljavajuća preciznost (0.674) i osjetljivost modela (0.633) (Tablica 6. b) desno).

Tablica 6. Klasifikacijski 3D-QSAR model za Dasatnib analoge: a) definicija kategorija aktivnosti b) matrice zabune, preciznost i osjetljivost modela na setu za učenje i testnom setu.

a)

Vrijednosti (pIC_{50})	Broj spojeva u kategoriji
<7.0	49
≥7.0 <8.0	43
≥8.0	38

b)

Set za učenje

pIC_{50} eksperimentalno \ pIC_{50} predviđeno		1	2	3
	1	45	4	0
	2	3	37	3
	3	0	9	29
Preciznost		0.861		
Osjetljivost		0.847		

Test set

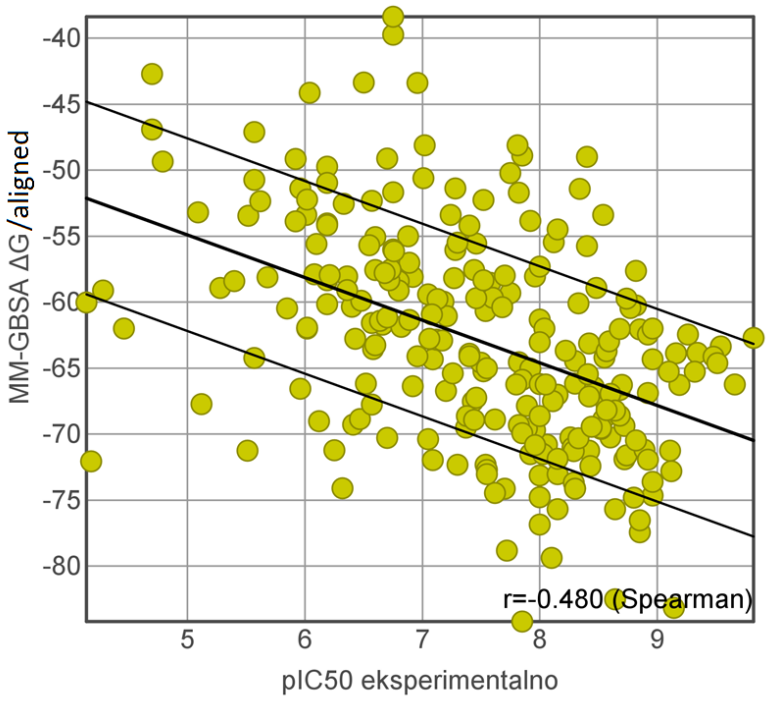
pIC_{50} eksperimentalno \ pIC_{50} predviđeno		1	2	3
	1	9	1	0
	2	2	5	1
	3	0	5	3
Preciznost		0.674		
Osjetljivost		0.633		

Najizazovniji je bio razvoj modela *PP2 analoga*. Sustavno se podcjenjuje aktivnost *PP2 analoga* s visokom učinkovitošću liganda, a precjenjuje se aktivnost spojeva s većim supstituentima. Dodatno, kao što je već komentirano ranije, nije bilo moguće jednoznačno preklapanje zbog različitih vektora diverzifikacije u podskupini *PP2 analoga*.

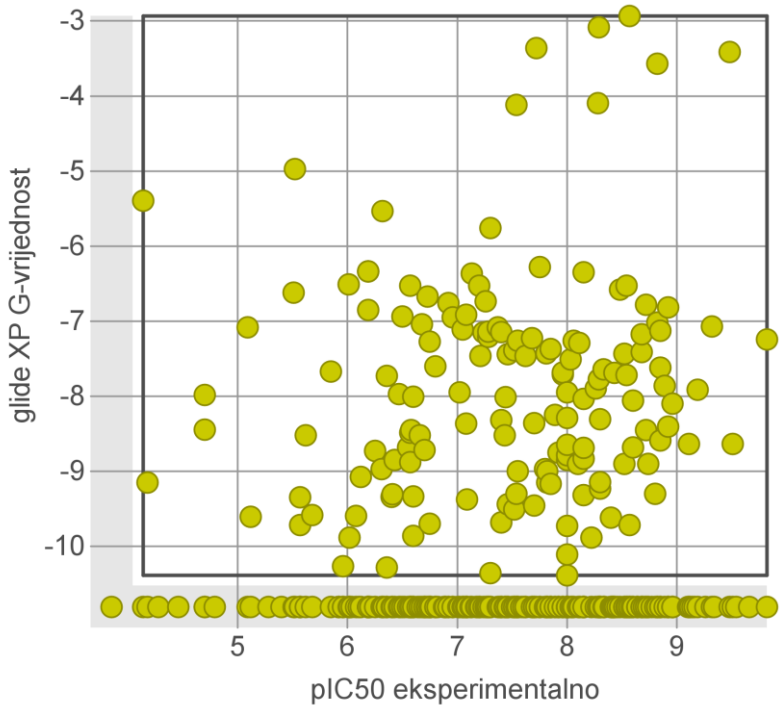
Razvijeni 3D-QSAR modeli uspoređeni su dodatno s predviđanjima aktivnosti pomoću molekulskog uklapanja te minimiziranjem protein-ligand kompleksa metodama molekulske mehanike uz implicitni utjecaj otapala, MM-GBSA metodom. Na slici 31. a) prikazana je korelacija (r=0.48) između eksperimentalno aktivnosti i rezultata MM-GBSA metode, što sugerira da MM-GBSA metoda ne poboljšava predviđanje aktivnosti potencijalnih Src inhibitora. Molekulsko uklapanje provedeno je na SP i XP

razini točnosti te su MM-GBSA energije interakcije izračunate minimiziranjem geometrije kompleksa dobivenih molekulskim uklapanjem. Glide XP metoda predvidjela je manji broj spojeva koji se uklapaju u aktivno mjesto dok je značajan broj spojeva ostao neuklopljen iako su eksperimentalno potvrđeni kao aktivni (Slika 31. b)). Posljedično tome, određivanje korelacije između eksperimentalnih vrijednosti i ΔG vrijednosti izračunatih MM-GBSA metodom na geometrijama protein-ligand kompleksa dobivenih Glide XP metodom provedena je na manjem broju molekula (Slika 31. c)). S druge strane, Glide SP metodom uspješno su uklopljeni svi proučavani spojevi (Slika 31. d)), te je za sve predviđene poze izračunata i ΔG vrijednost MM-GBSA metodom (Slika 31. e)).

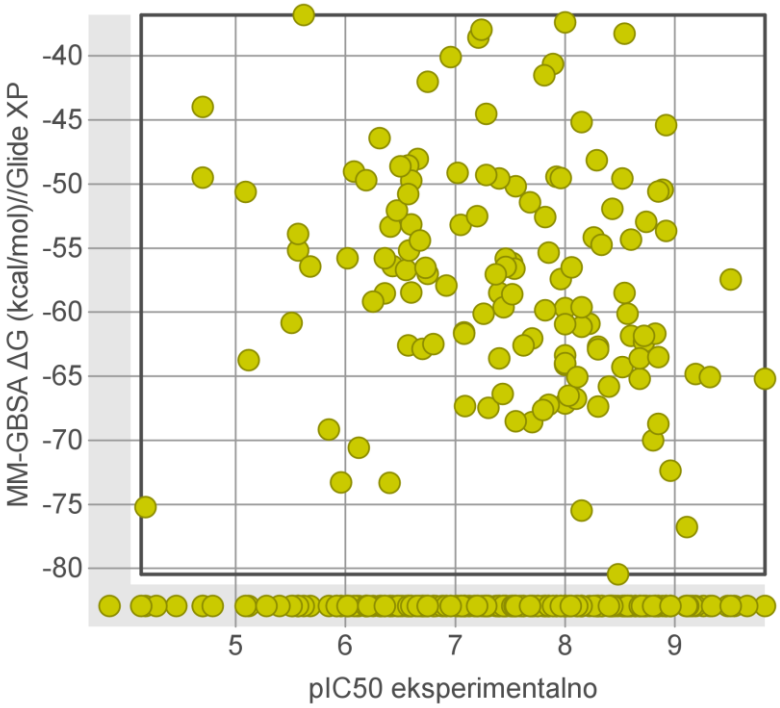
a)



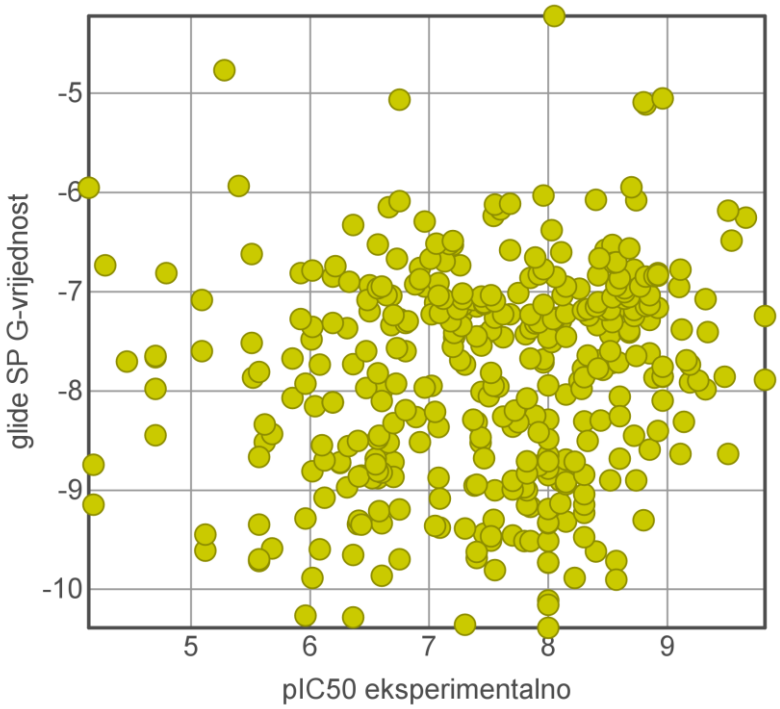
b)


glide XP G-vrijednost
-3
-4
-5
-6
-7
-8
-9
-10
5
6
7
8
9
pIC50 eksperimentalno

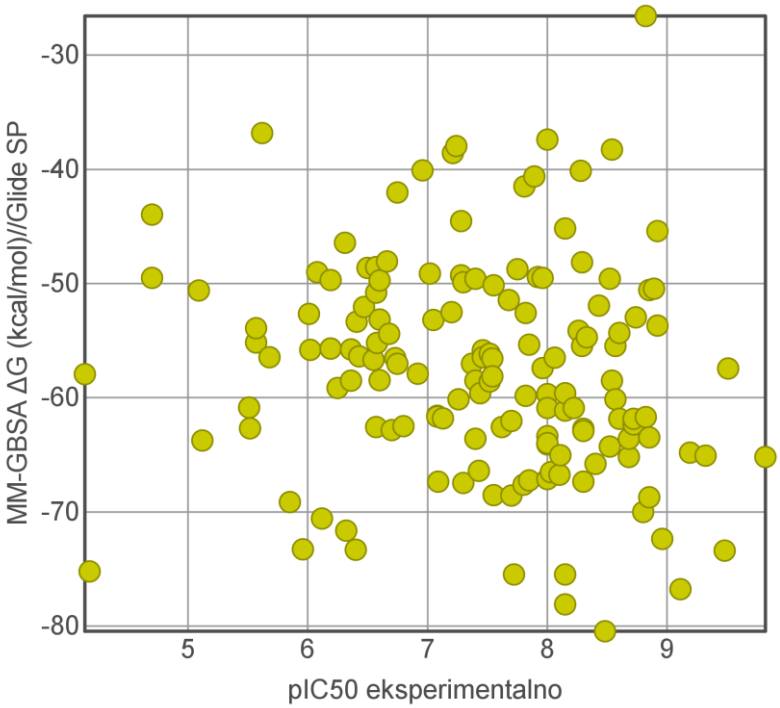
c)


MM-GBSA ΔG (kcal/mol)//Glide XP
-40
-45
-50
-55
-60
-65
-70
-75
-80
5
6
7
8
9
pIC50 eksperimentalno

d)



e)



Slika 31. Korelaciije za Saracatinib analoge između eksperimentalno određenih pIC_{50} vrijednosti i a) ΔG vrijednosti izračunate MM-GBSA metodom na protein-ligand kompleksima dobiveni preklapanjem molekula pomoću molekulskih polja; b) Glide G-vrijednosti izračunate XP metodom molekulskog uklapanja; c) ΔG vrijednosti izračunate

MM-GBSA metodom na protein-ligand kompleksima dobiveni Glide XP metodom molekulskog uklapanja; d) Glide G-vrijednosti izračunate Glide SP metodom molekulskog uklapanja; e) ΔG vrijednosti izračunate MM-GBSA metodom na protein-ligand kompleksima dobiveni Glide SP metodom molekulskog uklapanja.

U skladu s literaturnim podacima prikazanim u Tablici 1. molekulsko uklapanje te MM-GBSA metoda ne mogu se koristiti za pouzdano predviđanje aktivnosti potencijalnih Src inhibitora. Nešto bolje rezultate daje MM-GBSA minimizacija protein-ligand kompleksa dobivenog preklapanjem molekula koristeći molekulska polja uz prisutnost proteina u odnosu na komplekse dobivene molekulskim uklapanjem.

5. ZAKLJUČAK

Ovim istraživanjem razvijeni su 3D-QSAR kvantitativni i kvalitativni modeli temeljeni na molekulskim poljima za tip I Src inhibitore, koji pružaju ključne strukturne informacije. Od četiri razvijena modela (*PP2 analozi, Dasatinib analozi, Saracatinib analozi, Zajednički skup*), najbolji je model razvijen za *Saracatinib* analoge, visoke točnosti, stabilnosti te s prihvatljivom prediktivnom moći (PLS: q^2=0.63) i dobrim koeficijentom korelacije (PLS: r^2=0.92, r_{tes}^2=0.85). Ovaj model može se pouzdano koristiti za predviđanje aktivnosti virtualnih biblioteka spojeva sličnih saracatinibu. Zbog korištenja većeg broja strukturno različitih molekula, modeli korišteni u ovom ispitivanju imaju bolju prediktivnu moć te su robusniji. Rezultati dobiveni iz 3D-QSAR modela pružaju alate za predviđanje modifikacije afiniteta poznatih spojeva i za usmjeravanje sinteze novih inhibitora sa strukturnim svojstvima koje doprinose povećanju aktivnosti. Usporedbom dobivenih rezultata s prethodno objavljenim 3D-QSAR modelima za Src inhibitore, vidljivo je da su rezultati ovog istraživanja značajno bolji. Za modele (*PP2 analozi* i *zajednički skup*) s nižom statističkom značajnosti, klasifikacijskim modelima, potvrđeno je samo zadovoljavajuće predviđanje. Kao najlošiji model pokazao se set *Dasatinib analoga*, a najbolji set *Saracatinib analoga* s preciznosti i osjetljivosti od 100% na setu za učenje. Nadalje, nakon molekulskog uklapanja inhibitora u literaturno dostupne kristalne strukture DFG-"in" konformacije Src kinaze provedena je minimizacija protein-ligand kompleksa MM-GBSA metodom. Korelacija između eksperimentalnih aktivnosti i rezultata MM-GBSA metode (r=0.48) ukazuje da MM-GBSA metoda ne poboljšava predviđanje aktivnosti Src inhibitora. Nadalje, usporedbom dobivenih rezultata s literaturno objavljenim modelima, zaključujemo da se MM-GBSA metoda ne može koristi za pouzdano predviđanje aktivnosti. MM-GBSA metoda je aproksimativna metoda za izračunavanje protein-ligand interakcija, temeljena je na polju sila, a utjecaj otapala uzima u obzir samo

eksplicitno, tako da su entropijski efekti i utjecaj molekula vode slabije opisani. Zbog toga se 3D-QSAR metode pokazuju pouzdanijima za predviđanje aktivnosti novih kemijskih entiteta te je ovim istraživanjem detaljno istražen odnos strukture i aktivnosti inhibitora tip I Src kinaza. Strukturna raznolikost molekula, veličina setova molekula na kojima su razvijeni modeli, postepeni postupak preklapanja molekula te odabir pouzdanih statističkih metoda pridonijeli su stvaranju robusnih modela s dobrom prediktivnom moći. Stoga, dobivene rezultate možemo smatrati korisnim, budući da pridonose selekciji inhibitora, razvoju virtualnih biblioteka analoga, usmjeravaju sintezu novih i aktivnijih inhibitora, te ubrzavaju proces optimizacije novih potencijalnih vodećih molekula.

6. LITERATURA

1. Paul MK, Mukhopadhyay AK. Tyrosine kinase – Role and significance in Cancer. International Journal of Medical Sciences. 2004: 101–115.

2. Siveen KS, Prabhu KS, Achkar IW, Kuttikrishnan S, Shyam S, Khan AQ, Merhi M, Dermime S, Uddin S. Role of Non Receptor Tyrosine Kinases in Hematological Malignances and its Targeting by Natural Products. Molecular Cancer 2018; 17.

3. Wilks AF. Structure and function of the protein tyrosine kinases. Progress in Growth Factor Research 1990; 2: 97–111.

4. Dustin CM, Heppner DE, Lin M-CJ, van der Vliet A. Redox regulation of tyrosine kinase signalling: more than meets the eye. The Journal of Biochemistry 2019; 167: 151–163.

5. Fajer M, Meng Y, Roux B. The Activation of c-Src Tyrosine Kinase: Conformational Transition Pathway and Free Energy Landscape. The Journal of Physical Chemistry B 2016; 121: 3352–3363.

6. Boggon TJ, Eck MJ. Structure and regulation of Src family kinases. Oncogene 2004; 23: 7918–7927.

7. Roskoski R Jr. Src protein-tyrosine kinase structure and regulation. Biochemical and Biophysical Research Communications 2004; 324: 1155-1164.

8. Alvarez RH, Kantarjian HM, Cortes JE. The role of Src in solid and hematologic malignancies. Cancer 2006; 107: 1918–1929.

9. Roskoski R Jr. Src protein-tyrosine kinase structure, mechanism, and small molecule inhibitors. Pharmacological Research 2015; 94: 9-25.

10. Bjorge JD, Jakymiw A, Fujita DJ. Selected glimpses into the activation and function of Src kinase. Oncogene 2000; 19: 5620-5635.

11. Moroco JA, Craigo JK, Iacob RE, Wales TE, Engen JR, Smithgall TE. Differential Sensitivity of Src-Family Kinases to Activation by SH3 Domain Displacement. PLoS ONE 2014; 9.

12. Alba J, Montagna M, D'Abramo M. Modelling the Activation Pathways in Full-Length Src Kinase. Biophysica 2021; 1: 238–248.

13. Jha V, Macchia M, Tuccinardi T, Poli G. Three-Dimensional Interactions Analysis of the Anticancer Target c-Src Kinase with Its Inhibitors. Cancers 2020; 12: 2327.

14. Tsutsui Y, Deredge D, Wintrode PL, Hays FA. Imatinib binding to human c-Src is coupled to inter-domain allostery and suggests a novel kinase inhibition strategy. Scientific Reports 2016; 6.

15. Attwood MM, Fabbro D, Sokolov AV, Knapp S, Schiöth HB. Trends in kinase drug discovery: targets, indications and inhibitor design. Nature Reviews Drug Discovery 2021.

16. Fabbro D, Cowan-Jacob SW, Moebitz H. Ten things you should know about protein kinases: IUPHAR Review 14. British Journal of Pharmacology 2015; 172: 2675–2700.

17. Ward RA, Goldberg FW (ur.). Kinase Drug Discovery. Royal Society of Chemistry, 2018.

18. Rudrapal M, Chetia D. Virtual Screening, Molecular Docking and QSAR Studies in Drug Discovery and Development Programme. Journal of Drug Delivery and Therapeutics 2020; 10: 225–233.

19. Palermo G, De Vivo M. Computational Chemistry for Drug Discovery. Encyclopedia of Nanotechnology. Springer Netherlands 2015; 1–15.

20. Macalino SJY, Gosu V, Hong S, Choi S. Role of computer-aided drug design in modern drug discovery. Archives of Pharmacal Research 2015; 38: 1686–1701.

21. Lima AN, Philot EA, Trossini GHG, Scott LPB, Maltarollo VG, Honorio KM. Use of machine learning approaches for novel drug discovery. Expert Opinion on Drug Discovery 2016; 11: 225–239.

22. Zoete V, Grosdidier A, Michielin O. Docking, virtual high throughput screening and in silico fragment-based drug design. Journal of Cellular and Molecular Medicine 2009; 13: 238–248.

23. Li Q. Application of Fragment-Based Drug Discovery to Versatile Targets. Frontiers in Molecular Biosciences 2020; 7.

24. Makrynitsa GI, Lykouras M, Spyroulias GA, Matsoukas M-T. In silico Drug Design. Encyclopedia of Life Sciences. 2018; : 1–7.

25. Anderson AC. The Process of Structure-Based Drug Design. Chemistry & Biology 2003; 10: 787–797.

26. Du X, Li Y, Xia Y-L, Ai S-M, Liang J, Sang P. Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods. IJMS 2016; 17: 144.

27. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opinion on Drug Discovery 2015; 10: 449–461.

28. Science overview. https://www.cresset-group.com/science/overview/#sci-intro (Accessed: 21-Jul-2021).

29. Tedesco G. Tackling selectivity with Activity Atlas. https://www.cresset-group.com/about/news/tackling-selectivity-with-activity-atlas/ (Accessed: 27-Jul-2021).

30. Cao R, Mi N, Zhang H. 3D-QSAR study of c-Src kinase inhibitors based on docking. Journal of Molecular Modeling 2009; 16: 361–375.

31. Kohlmann A, Zhu X, Dalgarno D. Application of MM-GB/SA and WaterMap to SRC Kinase Inhibitor Potency Prediction. ACS Medicinal Chemistry Letters 2012; 3: 94–99.

7. ŽIVOTOPIS