

Univerzitet u Beogradu
Farmaceutski fakultet

Marko D. Kalinić

**Simulacije molekularne dinamike i
računarsko dizajniranje inhibitora
protein lizin metiltransferaze EZH2**

doktorska disertacija

Beograd, 2015.

University of Belgrade
Faculty of Pharmacy

Marko D. Kalinić

**Molecular dynamics simulations and
computational design of inhibitors of
protein lysine methyltransferase EZH2**

Doctoral Dissertation

Belgrade, 2015

MENTOR

Dr sc. Slavica Erić, vanredni profesor
Univerzitet u Beogradu – Farmaceutski fakultet

Potpis

ČLANOVI KOMISIJE

Dr sc. Vladimir Savić, redovni profesor
Univerzitet u Beogradu – Farmaceutski fakultet

Potpis

Dr sc. Mire Zloh, redovni profesor
University of Hertfordshire – Department of Pharmacy

Potpis

U BEOGRADU,

Datum odbrane

Mojim roditeljima

Simulacije molekularne dinamike i računarsko dizajniranje inhibitora protein lizin metiltransferaze EZH2

Rezime

Enhancer of Zeste Homolog 2 (EZH2) je epigenetski enzim koji vrši selektivno metilovanje lizina 27 na histonu H3 (H3K27). Trimetilovani H3K27 je represivan epigenetski signal, čime EZH2 utiče na smanjenje transkripcije ciljnih gena. Prekomerna aktivnost EZH2, izazvana povećanom ekspresijom ili mutacijama njegovog katalitičkog domena, dovedena je u vezu sa većim brojem maligniteta kod ljudi, a inhibicija ovog enzima smatra se perspektivnom strategijom u razvoju novih antitumorskih lekova. Prvi selektivni inhibitori EZH2 otkriveni su 2012. godine i njihova biološka karakterizacija potvrdila je terapijski potencijal inhibicije ovog epigenetskog regulatora. Iako postoji potreba za otkrićem novih inhibitora, značajnu prepreku u njihovom dizajnu do sada je predstavljao nedostatak trodimenzionalne strukture kompleksa EZH2 i okarakterisanih liganada. Cilj istraživanja u okviru ove doktorske disertacije bio je da se primenom računarskih metoda uspostavi i validira model vezivnog mesta EZH2, i da se, zatim, uspostavljeni model upotrebi u dizajnu novih inhibitora kao potencijalnih antitumorskih lekova.

U prvoj fazi istraživanja izvršeno je homologo modeliranje katalitičkog domena EZH2 koji sadrži vezivno mesto za kofaktor, a za koje se kompetitivno vezuju svi poznati inhibitori. Početni model usavršen je primenom simulacija konvencionalne i ubrzane molekularne dinamike (MD), a zatim je daljim simulacijama istraživano kako 13 inhibitora piridonske strukture stupa u interakciju sa vezivnim mestom na enzimu. Korišćenjem prikupljenih trajektorija izvršena je procena slobodne energije vezivanja inhibitora zasnovana na molekularnoj mehanici uz generalizovani Bornov model rastvarača (MM/GBSA). Izračunate energije upoređene su sa dostupnim eksperimentalnim podacima i utvrđeno je dobro slaganje između vrednosti koje ukazuje na validnost uspostavljenog modela. Daljom analizom trajektorija i primenom računarske mutageneze u alanin identifikovane su intermolekulske interakcije koje ključno doprinose aktivnosti i selektivnosti proučavanih inhibitora. Na ovim osnovama, definisan je farmakoforni model inhibitora EZH2 a njegovom validacijom potvrđena je mogućnost uspešne identifikacije i inhibitora baziranih na drugim osnovnim strukturama.

Efikasnost mnogih antitumorskih lekova ograničava razvoj rezistencije tu-

morskih ćelija, pri čemu je opisan veći broj mehanizama koji dovode do pojave rezistencije. Prekomerna ekspresija membranskih transportera, kao što su P-glikoprotein (Pgp) i protein rezistencije karcinoma dojke (BCRP), može usloviti povećan efluks lekova iz tumorskih ćelija i time onemogućiti postizanje terapijskih koncentracija u njima. Iz ovog razloga, u otkriću novih antitumorskih lekova poželjno je težiti dizajnu nesupstrata ovih transportera. U slučaju inhibicije EZH2, ovo je posebno značajno imajući u vidu da više poznatih piridonskih inhibitora jesu supstrati Pgp-a i BCRP-a. Kako bi tokom dizajna novih potencijalnih inhibitora EZH2 rano bila razmatrana mogućnost interakcije sa Pgp i BCRP, jedan od ciljeva ovog istraživanja bio je i razvoj klasifikacionih računarskih modela koji bi omogućili identifikaciju supstrata ovih transportera samo na osnovu hemijske strukture dizajniranih jedinjenja. U tu svrhu, pripremljeni su setovi poznatih supstrata i nesupstrata Pgp-a i BCRP-a, raznovrsnih hemijskih struktura, i to korišćenjem više kriterijuma za procenu pouzdanosti dostupnih eksperimentalnih podataka čime je osigurana njihova uniformnost i tačnost. Klasifikacioni modeli razvijani su primenom veštačkih neuronskih mreža i *support vector machine* algoritma, a njihovom validacijom pokazana je dobra pouzdanost uz prosečnu tačnost predviđanja od približno 80%. Dodatno, uporednom analizom definisani su delovi hemijskog prostora u kojima je najveća verovatnoća pronalazjenja supstrata Pgp-a i/ili BCRP-a.

Kako postojeće kolekcije jedinjenja često sadrže ograničen broj struktura koje su aktivne na ciljnim mestima dejstva čiji se terapijski potencijal tek prepoznaje, u ovom istraživanju identifikacija novih potencijalnih inhibitora EZH2 bila je bazirana na *de novo* pristupu. Koncipiran je i implementiran genetski algoritam za istovremenu optimizaciju više ciljeva dizajna baziran na dobro poznatom NSGA-II (engl. *Nondominated Sorting Genetic Algorithm II*) algoritmu. Ova implementacija, označena akronimom Phabblar (*Pharmacophore-based Building Block Assembler*), omogućava dizajn struktura koje zadovoljavaju definisanu farmakofornu hipotezu, imaju adekvatne fizičko-hemijske karakteristike i nisku sintetsku složenost, i to korišćenjem pristupa koji se oslanja na *in silico* reakcije između molekula veličine fragmenta. Primenom Phabblar algoritma i farmakofornog modela EZH2 inhibitora dizajnirana je kolekcija jedinjenja iz koje je odabrano 11 struktura koje su dalje proučavane simulacijama MD i MM/GBSA proračunima. Kao rezultat, predloženo je 5 struktura sa najvećim potencijalom da predstavljaju nove inhibitore EZH2. Ovih 5 struktura hemijski su različite od postojećih inhibitora, ali zadržavaju interakcije sa vezivnim mestom za koje je utvrđeno da ključno doprinose inhibitornoj aktivnosti.

Predviđanje mogućnosti transporta posredovanog Pgp-om i BCRP-om uka-

zalo je da dizajnirane strukture verovatno predstavljaju supstrate ovih transportera. Stoga je u finalnom delu istraživanja izvršena modifikacija jedne od struktura čiji je cilj bio dizajn Pgp nesupstrata koji će zadržati afinitet za EZH2. Korišćenjem klasifikacionog modela za Pgp transport, analize senzitivnosti molekulskih deskriptora i dodatnim simulacijama MD, kroz nekoliko ciklusa modifikacija dizajnirana je struktura za koju je predviđeno da nije supstrat Pgp-a, a čija je slobodna energija vezivanja, procenjena MM/GBSA metodom, oko 5 kcal mol⁻¹ bolja od polaznog jedinjenja. Time je pokazana korisnost ovog pristupa u racionalnoj modifikaciji struktura sa ciljem eliminacije afiniteta za Pgp. Sinteza i biološka karakterizacija dizajniranih jedinjenja može biti prvi korak u daljem razvoju ovih potencijalnih inhibitora.

Ključne reči: Enhancer of Zeste Homolog 2, epigenetika, antitumorski lekovi, molekularna dinamika, *de novo* dizajn, P-glikoprotein, protein rezistencije karcinoma dojke, BCRP, klasifikacioni računarski modeli.

Naučna oblast: Farmacija

Uža naučna oblast: Farmaceutska–medicinska hemija i strukturna analiza

UDK broj: 615.355:004.9(043.3), 615.277:577.3(043.3)

Molecular dynamics simulations and computational design of inhibitors of protein lysine methyltransferase EZH2

Abstract

Enhancer of Zeste Homolog 2 is an epigenetic enzyme responsible for selective methylation of lysine 27 on histone H3. Trimethylated H3K27 constitutes a repressive epigenetic mark, making EZH2 capable of transcriptional silencing of target genes. Aberrant EZH2 activity, caused by either overexpression or point mutations in the catalytic domain, has been related to a number of human malignancies, making EZH2 inhibition a promising strategy in the development of novel anticancer treatments. First disclosure of selective EZH2 inhibitors in 2012 and their biological characterization has confirmed the therapeutic potential of inhibiting this epigenetic regulator. While there is an outstanding need for the discovery of novel EZH2 inhibitors, the lack of a three-dimensional structure of EZH2 complexed with one of its ligands has been a significant obstacle to this end. The aim of research conducted as part of this doctoral dissertation was to establish and validate a model of EZH2 binding site using computational methods, and to subsequently utilize the established model as a foundation in the design of novel potential EZH2 inhibitors.

In the first stage of our research, homology modeling of the EZH2 catalytic domain was conducted, which contains the cofactor binding site targeted by all known inhibitors. The initial model was refined through conventional and accelerated molecular dynamics (MD) simulations. Next, structures of EZH2 complexed with 13 different pyridone inhibitors were simulated in order to study the key interactions participating in ligand binding. An assessment of inhibitors' binding free energy was performed using the molecular mechanics/generalized Born surface area (MM/GBSA) method. The computed energies were correlated to available experimental data and good agreement was found supporting the validity of the established model. Further analysis of MD trajectories and computational alanine scanning facilitated the definition of key intermolecular interactions that contribute to the observed potency and selectivity of the studied inhibitors. On this basis, a pharmacophore model of EZH2 inhibitors was proposed and validated. It demonstrated good capabilities in identifying truly active molecules, including those not based on a pyridone scaffold, in presence of decoys.

The efficacy of anticancer drugs can become limited with the onset of resi-

stance that can develop in cancer cells through various mechanisms. Often involved are membrane transporters, such as P-glycoprotein (Pgp) and breast cancer resistance protein (BCRP), whose overexpression can lead to increased drug efflux and, consequently, to low drug concentrations in cancer cells. For this reason, novel anti-cancer drug candidates should ideally be non-substrates of these transporters. With respect to targeting EZH2, this can be seen as particularly important given that several existing inhibitors of the enzyme have been characterized as substrates of both Pgp and BCRP. In order to take Pgp- and BCRP-mediated transport into account during the design of novel EZH2 inhibitors, one of the aims of this dissertation was to develop computational classification models which would allow the identification of Pgp/BCRP substrates based on chemical structure alone. To achieve this, sets of chemically diverse Pgp and BCRP substrates and non-substrates were compiled, using previously defined criteria in evaluating the reliability of available experimental data. This assured the resulting sets were highly consistent and contained accurately classified compounds only. Classification models were developed using artificial neural networks and support vector machines. Validation of the models showed they exhibit good reliability with an average classification accuracy of approximately 80%. Furthermore, a comparative analysis allowed us to define regions in chemical space where substrates of Pgp, BCRP, or both transporters, respectively, are most likely to be found.

Given that existing compound collections often contain limited numbers of structures that are active on targets whose therapeutic potential is only starting to become recognized, the identification of novel putative EZH2 inhibitors in this study was based on a *de novo* design approach. A multi-objective optimization genetic algorithm was conceived and implemented, based largely on the well known Non-dominated Sorting Genetic Algorithm II (NSGA-II). This implementation, termed Phabblar (acronym derived from Pharmacophore-based Building Block Assembler), is intended to facilitate the design of structures satisfying a given pharmacophore hypothesis that, at the same time, retain acceptable physico-chemical characteristics and low synthetic complexity. Phabblar-generated solutions evolve through *in silico* reactions between fragment-sized building blocks. Using Phabblar and the previously defined EZH2 inhibitors' pharmacophore, a library of compounds was designed from which 11 structures were selected for further study using MD simulations and MM/GBSA calculations. As a result, 5 structures are proposed which exhibit the greatest potential for being novel putative EZH2 inhibitors. These 5 compounds are structurally distinct from existing inhibitors, yet maintain the key interactions with EZH2 binding site that are thought to be essential for inhibitory potency.

Predictions of Pgp- and BCRP-mediated transport suggested that the designed structures are likely substrates of these transporters. To address this fact, the final stage of the conducted research involved modifying one of the compounds with the aim of designing-out Pgp while retaining EZH2 affinity. Several cycles of modifications, guided by the Pgp classification model, descriptor sensitivity analysis and additional MD simulations, yielded a Pgp non-substrate compound whose binding free energy to EZH2, as evaluated using MM/GBSA calculations, was improved by 5 kcal mol⁻¹. This demonstrated the utility of the approach in guiding rational structural modification aimed at designing-out unwanted interactions with Pgp. Synthesis and biological evaluation of the designed compounds can lay the foundation for further development of these putative inhibitors.

Keywords: Enhancer of Zeste Homolog 2, epigenetics, anticancer drugs, molecular dynamics, *de novo* design, P-glycoprotein, breast cancer resistance protein, BCRP, computational classification models.

Scientific field: Pharmacy

Specific scientific discipline: Pharmaceutical–medicinal chemistry and structural analysis

UDC notation: 615.355:004.9(043.3), 615.277:577.3(043.3)

Sadržaj

Rezime disertacije	i
Sadržaj	vii
Spisak ilustracija	xi
Spisak tabela	xvii
Spisak skraćenica	xix
1 Uvod	1
1.1 Epigenetika normalnih i malignih ćelija	4
1.1.1 Metilovanje DNK	5
1.1.2 Post-translacione modifikacije histona	6
1.1.3 Nekodirajuće RNK	16
1.1.4 Promene epigenetskog pejzaža malignih ćelija	18
1.1.5 Terapijski potencijal epigenetskih ciljnih mesta dejstva	20
1.2 Struktura, funkcija i onkogenost EZH2	21
1.2.1 Struktura PRC2 i EZH2	22
1.2.2 Biološka funkcija PRC2	25

1.2.3	EZH2 kao onkogen	28
1.3	Inhibitori EZH2	31
1.3.1	Neselektivni inhibitori	31
1.3.2	Kofaktor-kompetitivni inhibitori	33
1.3.3	Drugi pristupi inhibiciji	43
1.3.4	Izazovi u dizajnu novih inhibitora	44
1.4	Rezistencija na antitumorske lekove	45
1.4.1	Struktura i funkcija Pgp-a	46
1.4.2	Struktura i funkcija BCRP-a	48
1.4.3	Značaj predviđanja interakcije lekova sa ABC transporterima	49
1.5	Računarske metode u dizajniranju novih molekula	50
1.5.1	Simulacije molekularne dinamike	50
1.5.2	Računarsko <i>de novo</i> dizajniranje	56
2	Cilj rada	66
3	Eksperimentalni deo	68
3.1	Simulacije molekularne dinamike EZH2	68
3.1.1	Homologo modeliranje EZH2	68
3.1.2	Simulacije MD EZH2–SAH kompleksa i apo oblika EZH2	70
3.1.3	Poređenje sličnosti vezivnih mesta PKMT	72
3.1.4	Simulacije MD EZH2 u kompleksu sa piridonskim inhibitorima	73
3.1.5	Opšti parametri MD simulacija i analiza trajektorija	75
3.1.6	MM/GBSA proračuni i analiza efekata mutageneze u alanin	76

3.1.7	Analiza konformacionih preferenci odabranih inhibitora	77
3.1.8	Razvoj farmakofornog modela EZH2 inhibitora	77
3.2	Razvoj modela za predviđanje transporta lekova Pgp-om i BCRP-om	81
3.2.1	Formiranje setova podataka	81
3.2.2	Procena raznovrsnosti formiranih setova	83
3.2.3	Priprema struktura i podela setova	83
3.2.4	Razvoj modela	86
3.2.5	Validacija modela	89
3.2.6	Definisanje domena primenjivosti modela	89
3.3	Računarsko dizajniranje EZH2 inhibitora	90
3.3.1	Implementacija algoritma za <i>de novo</i> dizajn	90
3.3.2	Redukovanje kolekcije dizajniranih jedinjenja	91
3.3.3	Simulacije molekularne dinamike	92
4	Rezultati i diskusija	94
4.1	Simulacije molekularne dinamike EZH2	94
4.1.1	Modeliranje mesta za vezivanje kofaktora	94
4.1.2	Analiza interakcije piridonskih inhibitora i EZH2	99
4.1.3	Analiza efekata mutageneze u alanin	110
4.1.4	Farmakoforni model inhibitora EZH2	114
4.1.5	Uticaj somatskih mutacija EZH2 na vezivanje inhibitora . . .	117
4.1.6	Rezime	120
4.2	Modeli za predviđanje transporta posredovanog Pgp-om i BCRP-om .	122

4.2.1	Modeli za predviđanje transporta posredovanog Pgp-om . . .	122
4.2.2	Modeli za predviđanje transporta posredovanog BCRP-om . . .	125
4.2.3	Analiza preklapanja supstratne specifičnosti Pgp-a i BCRP-a	126
4.2.4	Eksterna validacija modela	129
4.2.5	Rezime	132
4.3	Dizajn novih potencijalnih inhibitora EZH2	134
4.3.1	Implementacija algoritma za <i>de novo</i> dizajn hemijskih struktura	134
4.3.2	<i>De novo</i> dizajn kolekcije potencijalnih EZH2 inhibitora	148
4.3.3	Simulacije molekularne dinamike dizajniranih jedinjenja	149
4.3.4	Modifikacija strukture u cilju eliminisanja afiniteta za Pgp . .	157
4.3.5	Rezime	163
5	Zaključak	164
6	Literatura	168
Prilog A	Rezultati simulacija kompleksa EZH2 i odabranih piridon- skih inhibitora	206
Prilog B	Modeli za predviđanje transporta posredovanog Pgp-om i BCRP-om	211
Prilog C	Karakteristike funkcija procene korišćenih u <i>de novo</i> dizajnu	239
	Biografija autora	241

Spisak ilustracija

1.1	Različiti nivoi organizacije genetskog materijala u ćeliji, od nukleozoma i hromatinske niti do hromozoma. ©Richard Wheeler / Wikimedia Commons / CC-BY-SA 3.0 / GFDL	4
1.2	Trodimenzionalne stukture protein lizin metiltransferaza sa SET domenom, u kompleksu sa histonskim supstratom i proizvodom kofaktora (SAH).	10
1.3	Strukturna organizacija PRC2, sa pojedinačnim komponentama obojenim kao u legendi. ©Ciferri et al. / eLife / CC-BY 3.0	22
1.4	Ilustracija sekvence i kristalografski rešenih delova strukture EZH2.	23
1.5	Prikaz kristalografski rešene strukture SET domena EZH2 (PDB ID: 4MI5) superponiran na strukturu GLP u kompleksu sa histonskim supstratom i SAH (PDB ID: 2RFI). Delovi strukture EZH2 obojeni su žutom, zelenom odnosno plavom bojom, dok je GLP prikazan semitransparentno, u sivoj boji.	24
1.6	Šematski prikaz pretpostavljenih mehanizama kojima PRC2 i trimetilovanje H3K27 dovode do represije gena.	26
1.7	Aberacije EZH2 koje dovode do hipermetilovanja H3K27.	29
1.8	Šematski prikaz mogućih pristupa inhibiciji katalitičke aktivnosti PRC2: (a) inhibicija formiranja kompleksa; (b) inhibicija vezivanja histonskog supstrata za EZH2; (c) inhibicija vezivanja kofaktora za EZH2; (d) inhibicija SAH hidrolaze.	32
1.9	Strukture DZNep-a i sinefungina, neselektivnih inhibitora PKMT	32

1.10	Strukture 13 piridonskih inhibitora EZH2 koje su objavljene u literaturi tokom 2012. godine.	34
1.11	Šematski prikaz dostupnih podataka o odnosu strukture i aktivnosti piridonskih inhibitora (crvene oznake) i pretpostavljenog načina interakcije (plave oznake) ovih jedinjenja sa EZH2	36
1.12	Strukture dodatnih piridonskih inhibitora EZH2.	37
1.13	Strukture tetrametilpiperidinskih inhibitora EZH2	39
1.14	Strukture heksahidroizohinolinskih inhibitora EZH2	42
1.15	Strukture tanšindiola i sintetskog analoga SAH, kao inhibitora EZH2	43
1.16	Kristalografski rešena struktura mišijeg Pgp-a (PDB ID: 4M2T), sa pretpostavljenim položajem ćelijske membrane i različitim bojama istaknutim domenima transportera.	47
1.17	Šematski prikaz mutacije i ukrštanja binarno i decimalno kodiranih gena.	60
1.18	Ilustracija Pareto fronta na primeru optimizacije sa dva cilja, čije vrednosti funkcija procene treba maksimizovati	63
3.1	Superpozicija sekvenci delova EZH2, SETD7 i SETD8 korišćeno u razvoju homologog modela.	69
3.2	Raspodela jedinjenja iz formiranih setova (označena na grafiku kao narandžaste tačke) u hemijskom prostoru definisanom glavnim komponentama ChemGPS-NP sistema. Osenčene oblasti odgovaraju raspodeli jedinjenja iz DrugBank 3.0 kolekcije, pri čemu tamnije regije odražavaju veću, a svetlije manju gustinu referentnih jedinjenja. . . .	84
4.1	Prikaz reprezentativne strukture razvijenog modela EZH2 (obojen plavom) u kompleksu sa proizvodom kofaktora SAH (obojen zelenom). Model je superponiran na početnu kristalografsku strukturu SET domena enzima (obojena roze, PDB ID: 4MI0). Vodonici su uklonjeni iz strukture radi jasnijeg prikaza.	96

4.2	Dinamika apo oblika EZH2, analizirana na osnovu MD trajektorije trajanja 15 ns.	97
4.3	Sličnost vezivnih mesta za kofaktor u PKMT enzimima sa dostupnom kristalnom strukturom. Kao struktura EZH2 upotrebljen je razvijeni model. Stepem sličnosti odgovara obojenosti datog polja, kao u legendi. Dendrogrami grupisanja struktura prikazani su u levom delu svake slike.	98
4.4	Efekat prisustva inhibitora na dinamiku EZH2, ilustrovan kroz smanjenje vrednosti korena srednjeg kvadrata fluktuacija (RMSF) C_{α} atoma u strukturi.	100
4.5	Korelacija između $\Delta G_{MM/GBSA}$ i ΔG_{exp} vrednosti za 13 piridonskih inhibitora EZH2.	101
4.6	Položaj inhibitora G3K (obojen roze) u vezivnom mestu EZH2 (enzim je prikazan polutransparentnom plavom). Tamnije plavom bojom prikazani su bočni lanci aminokiselina koje se nalaze na udaljenosti manjoj od 3 Å od bilo kog atoma inhibitora, i koji formiraju vezivno mesto. Radi jasnijeg prikaza, iz ilustracije su izostavljeni atomi vodonika, aminokiseline F723 i Y728.	103
4.7	Reprezentativne strukture kompleksa EZH2 i odabranih inhibitora koje ilustruju dodatne interakcije ili različite načine vezivanja u odnosu na opšti slučaj prikazan na slici 4.6. Strukture liganada obojene su roze dok su aminokiseline vezivnog mesta obojene plavom. Radi jasnijeg prikaza, atomi vodonika i drugi delovi enzima nisu prikazani na slici. Vodonične veze prikazane su kao isprekidane crne linije. . . .	104
4.8	Evolucija orijentacije G4K u vezivnom mestu EZH2 tokom ekvilibracije sistema trajanja 20 ns.	106
4.9	Reprezentativne strukture kompleksa EZH2 i inhibitora G4K odnosno 126 , za koje je utvrđen alternativni način vezivanja za enzim. . .	107
4.10	Analiza konformacionih preferenci inhibitora 126 i G3K	109
4.11	Rezultati analize efekata mutageneze u alanin.	111

4.12	Ilustracija farmakofornih modela inhibitora EZH2. Farmakoforni elementi prikazani su kao obojene sfere: crvenom su predstavljene lipo-filne grupe, zelenom aromatični centri, plavom donori a roze bojom akseptori vodonične veze. Sive sfere odgovaraju sferama isključenja.	115
4.13	Utica j somatskih mutacija EZH2 na strukturu vezivnog mesta za ko-faktor.	118
4.14	Ilustracija potencijalne strukturne osnove za dizajn mutant-selektivnih inhibitora EZH2 kroz eksploataciju dodatnih šupljina u vezivnom me-stu.	120
4.15	Vrednosti deskriptora izvedenih iz farmakofora tipa I i II (videti tekst) za supstrate i nesupstrate Pgp-a. Donja i gornja granica svakog pra-vougaonika odgovaraju prvom i trećem kvartilu raspodele. Medijana je označena horizontalnom crtom. Krajevi vertikalnih linija označa-vaju minimalne i maksimalne vrednosti.	124
4.16	Distribucija supstrata i nesupstrata Pgp-a i BCRP-a u hemijskom prostoru definisanom vrednostima tri glavne komponente (PC) opi-sane u tekstu.	127
4.17	Preklapanje supstratne specifičnosti Pgp-a i BCRP-a, ilustrovano kroz distribuciju odabranih jedinjenja u hemijskom prostoru definisanom vrednostima tri glavne komponente (PC) opisane u tekstu.	129
4.18	Ilustracija predstavljanja genotipa molekule u Phabbl er algoritmu.	137
4.19	Utica j izbora verovatnoće mutacije na neke aspekte konvergencije Phabbl er algoritma. Različite boje linija odgovaraju različitim ve-rovatnoćama mutacije koje su testirane, kao u legendi.	142
4.20	Utica j izbora verovatnoće ukrštanja na neke aspekte konvergencije Phabbl er algoritma. Različite boje linija odgovaraju različitim ve-rovatnoćama ukrštanja koje su testirane, kao u legendi.	144
4.21	Utica j uklanjanja maksimizacije topološke sličnosti, kao cilja dizajna, na konvergenciju Phabbl er algoritma. Kao mera topološke sličnosti korišćena je Tanimoto sličnost FCFP6 otisaka prstiju (FCFP6 funk-cija procene).	146

4.22	Utica j uklanjanja minimizacije sintetske složenosti, kao cilja dizajna, na konvergenciju Phabblor algoritma. Kao mera sintetske složenosti korišćena je SMCM vrednost molekula (SMCM funkcija procene).	147
4.23	Strukture 11 jedinjenja iz <i>de novo</i> dizajnirane kolekcije čija je interakcija za EZH2 proučavana simulacijama MD.	149
4.24	Reprezentativna struktura kompleksa i rezultati razlaganja energije vezivanja za kompleks PHB-00099 i EZH2.	151
4.25	Reprezentativna struktura kompleksa i rezultati razlaganja energije vezivanja za kompleks PHB-00394 i EZH2.	152
4.26	Reprezentativna struktura kompleksa i rezultati razlaganja energije vezivanja za kompleks PHB-00471 i EZH2.	154
4.27	Reprezentativna struktura kompleksa i rezultati razlaganja energije vezivanja za kompleks PHB-00577 i EZH2.	155
4.28	Reprezentativna struktura kompleksa i rezultati razlaganja energije vezivanja za kompleks PHB-02060 i EZH2.	156
4.29	Grafički prikaz rezultata analize senzitivnosti deskriptora u modelu ANN-24-3 (transport posredovan Pgp-om) za strukturu PHB-00471	158
4.30	Reprezentativna struktura kompleksa i rezultati razlaganja energije vezivanja za kompleks PHB-00471f i EZH2.	159
4.31	Reprezentativna struktura kompleksa i rezultati razlaganja energije vezivanja za kompleks PHB-00471i i EZH2.	161
4.32	Reprezentativna struktura kompleksa i rezultati razlaganja energije vezivanja za kompleks PHB-00471m i EZH2.	162
A.1	Korelacija između $\Delta G_{MM/GBSA}$ i ΔG_{exp} u zavisnosti od dužine evaluirane trajektorije. Nakon 4–5 ns simulacije, primetna je konvergencija rezultata. Ispunjena plava linija odgovara korelaciji bez rezultata za E1Z i 126 , dok isprekidana crvena linija odgovara korelaciji za sve rezultate zajedno.	210

- C.1 Verovatnoća identifikovanja podjednako aktivnog jedinjenja pri datom nivou njegove sličnosti sa referentnom strukturom poznate aktivnosti. Sličnost se odnosi na kombinovani rezultat farmakoforne (Align-it) i podudarnosti oblika (Shape-it), procenjene kao Tversky sličnost (detaljnije opisano u odeljku 3.1.8). Kriva je uspostavljena probabilističkim pristupom opisanom u Swann et al. (2011) *J. Med. Chem.* 54(5): 1223–1232, korišćenjem seta iz navedene publikacije. . . 239
- C.2 Korelacija između izračunate SMCM vrednosti i subjektivne ocene organskih hemičara o sintetskoj složenosti 1731 jedinjenja ($R = 0,74$). 240

Spisak tabela

1.1	Pregled najčešćih post-translacionih modifikacija histona i enzima uključenih u uvođenje, uklanjanje i prepoznavanje ovih modifikacija	9
1.2	Pregled malignih tumora kod kojih je dokazana prekomerna ekspresija EZH2	28
1.3	Konstante inhibicije i fizičko-hemijski parametri 13 piridonskih inhibitora EZH2	35
1.4	Inhibitorne koncentracije i fizičko-hemijski parametri odabranih tetrametilpiperidinskih inhibitora Enhancer of Zeste Homolog 2 (EZH2)	40
1.5	Inhibitorne koncentracije i fizičko-hemijski parametri odabranih heksahidroizohinolinskih inhibitora EZH2	43
3.1	Pregled kriterijuma koji su korišćeni u proceni pouzdanosti klasifikacije jedinjenja prilikom formiranja setova podataka, u zavisnosti od tipa i broja dostupnih eksperimentalnih dokaza	82
4.1	Performanse farmakofornih modela u identifikaciji 28 inhibitora EZH2 u setu sa 2.800 molekula „mamaca”	116
4.2	Performanse konačnih modela za predviđanje transporta posredovanog Pgp-om.	122
4.3	Performanse konačnih modela za predviđanje transporta posredovanog BCRP-om.	125

4.4	Vrednosti jednostavnih molekulskih karakteristika dvojnih supstrata i nesupstrata Pgp-a i BCRP-a, odnosno jedinjenja koje transportuje samo jedan od ova dva transportera.	130
4.5	Rezultati eksterne validacije razvijenih modela.	131
4.6	Rezultati MM/GBSA proračuna za 5 novodizajniranih jedinjenja u kompleksu sa EZH2	150
4.7	Izračunate osobine jedinjenja dizajniranih tokom modifikacije PHB-00471	157
A.1	Rezultati dokinga 13 piridonskih inhibitora u razvijeni model EZH2 .	206
A.2	Rezultati MM/GBSA proračuna sprovedenih na 500 kadrova produkcionih trajektorija trajanja 4 ns. Vrednosti u tabeli date su u kcal mol ⁻¹ kao srednje vrednosti uz standardne greške.	207
A.3	Rezultati MM/GBSA proračuna na stupnjevito dužim delovima dodatne produkcionne trajektorije. Vrednosti u tabeli date su u kcal mol ⁻¹ kao srednje vrednosti uz standardne greške.	208
A.4	Rezultati analize efekata mutagenoze u alanin, prikazani kao srednje vrednosti razlike MM/GBSA energija ($\Delta\Delta G_{MM/GBSA}$ u kcal mol ⁻¹) između kanonske i mutirane forme enzima, uz standardne greške. . . .	209
B.1	Set podataka korišćen u razvoju modela za predviđanje transporta posredovanog Pgp-om.	212
B.2	Set podataka korišćen u razvoju modela za predviđanje transporta posredovanog BCRP-om.	227
B.3	Deskriptori uključeni u finalne modele za predviđanje transporta posredovanog Pgp-om	233
B.4	Deskriptori uključeni u finalne modele za predviđanje transporta posredovanog BCRP-om	236
B.5	Rezultati eksterne validacije modela uz korišćenje dve alternativne definicije domena primenjivosti.	238

Spisak skraćenica

ANN veštačke neuronske mreže (engl. *artificial neural networks*)

ATP adenzin-trifosfat

AUROC površina ispod ROC krive (engl. *area under the ROC curve*)

BCRP protein rezistencije raka dojke (engl. *Breast cancer resistance protein*)

BEDROC engl. *Boltzmann-Enhanced Discrimination of ROC*

CAMD dizajniranje molekula pomoću računara (engl. *computer-aided molecular design*)

CGI CpG ostrvca (engl. *CpG islands*)

CML hronična mijeloidna leukemija (engl. *chronic myeloid leukemia*)

DLBCL difuzni krupnoćelijski limfom B-ćelija (engl. *diffuse large cell B-lymphoma*)

DNK dezoksiribonukleinska kiselina

DNMT DNK metiltransferaza (engl. *DNA methyltransferase*)

DNMTi inhibitori DNK metiltransferaza (engl. *DNA methyltransferase inhibitors*)

EF faktor obogaćenja (engl. *enrichment factor*)

EZH2 Enhancer of Zeste Homolog 2

GA genetski algoritam

HAT histon acetiltransferaza

HDAC histon deacetilaza

HDACi inhibitori histon deacetilaza (engl. *histone deacetylase inhibitors*)

MD molekularna dinamika

MDR engl. *Multiple Drug Resistance*

MM/GBSA metoda molekularne mehanike uz generalizovani Bornov model rastvarača (engl. *Molecular Mechanics/Generalized Born Surface Area*)

PDB Protein Data Bank

PC glavna komponenta (engl. *principal component*)

Pgp P-glikoprotein

PKMT protein lizin metiltransferaza (engl. *protein lysine methyltransferase*)

PRC2 Polycomb Repressive Complex 2

PRMT protein arginin metiltransferaza (engl. *protein arginine methyltransferase*)

PTM post-translaciona modifikacija

RBF funkcija radijalne raspodele (engl. *radial basis function*)

RIE robusno inicijalno obogaćenje (engl. *robust initial enhancement*)

RMSD koren srednje vrednosti kvadrata odstupanja (engl. *root-mean-square deviation*)

RMSF koren srednje vrednosti kvadrata fluktuacija (engl. *root-mean-square fluctuation*)

RNK ribonukleinska kiselina

dnkRNK duga nekodirajuća RNK (engl. *long non-coding RNA, lncRNA*)

miRNK mikro RNK (engl. *microRNA, miRNA*)

ROC engl. *Receiver Operating Characteristic*

SAH *S*-adenozil-L-homocistein

SAM *S*-adenozil-L-metionin

SAR odnos strukture i aktivnosti (engl. *structure activity relationship*)

SMCM sintetska i molekularna složenost (engl. *Synthetic and Molecular Complexity*)

SVM engl. *Support Vector Machines*

Poglavlje 1

Uvod

Maligne bolesti predstavljaju jedan od najvećih zdravstvenih izazova 21. veka. U Evropi se svake godine ustanovi 3,45 miliona novih slučajeva obolelih i 1,75 miliona smrtnih ishoda koji su posledica različitih formi maligniteta [1]. Na nivou svetske populacije, procenjuje se da godišnje od raka umre između 8 i 12,7 miliona ljudi, što čini rak drugom najsmrtonosnijom nezaraznom bolesti na globalnom nivou [2, 3]. Prognoze iz razvijenih zemalja sveta predviđaju da će u narednoj deceniji stopa mortaliteta od malignih bolesti rasti uprkos poboljšanom kvalitetu zdravstvene nege, prevashodno usled uvećanja i starenja populacije [4, 5].

Brz razvoj metoda molekularne biologije u poslednjih 20 godina doprineo je ostvarivanju značajnog napretka u razumevanju biologije kancera, što je ukazalo na brojne mogućnosti razvoja ciljanih terapija. Koncept ciljane terapije podrazumeva farmakološku modulaciju funkcije tačno određenih makromolekula za koje se utvrdi da su od presudnog značaja za preživljavanje datog tipa kancerskih ćelija [6]. Ovakvi lekovi, za razliku od neselektivnih citotoksičnih agenasa, treba da omogućе uništavanje populacije malignih ćelija bez negativnih efekata na zdrave ćelije organizma. Potencijal primene selektivnih lekova u medicinskoj onkologiji najbolje ilustruje primer imatiniba [7, 8]. Imatinib je selektivni inhibitor Bcr-Abl kinaze, koja je konstitutivno aktivna forma tirozin kinaze karakteristično prisutno kod određenih tipova hematoloških maligniteta kao što je hronična mijeloidna leukemija (CML). Do uvođenja imatiniba u terapiju, lečenje CML bilo je ograničeno na usporavanje progresije bolesti, pri čemu većina pacijenata nije živelo duže od 45–55 meseci nakon dijagnoze. Nakon uvođenja imatiniba u terapiju CML, kod više od 90% pacijenata moguće je postići postojanu petogodišnju remisiju bolesti, što jasno ilustruje poten-

cijal ciljanih terapija u suštinskom unapređenju prognoze maligniteta [9].

Imatinib ujedno predstavlja i jedan od prvih uspeha u personalizovanoj terapiji malignih bolesti. Iako postoje brojni izazovi na putu širokog prihvatanja ovog pristupa lečenju, personalizovana terapija kancera može da omogući smanjenje toksičnosti primenjene terapije, smanjenje troškova povezanih sa lečenjem primarne bolesti, može da poveća efikasnost lečenja i tako suštinski unapredi prognozu obolelih [10, 11]. Jedan od najvećih izazova u realizaciji ovog ideala jeste ograničena dostupnost odgovarajućih antitumorskih lekova. Iako je do danas identifikovan veliki broj potencijalnih ciljnih mesta dejstva, broj lekova koji su dostupni u kliničkom okruženju relativno je mali i postoji značajna potreba za otkrićem novih antitumorskih lekova koji bi bili pogodni za primenu u personalizovanoj terapiji [12]. Razvoj ovakvih lekova zahteva validaciju ciljnih mesta dejstva i racionalan pristup otkriću novih jedinjenja, što treba da omogući prevođenje uspeha u razumevanju etiopatogeneze kancera u nove efikasne pristupe lečenju.

Početak ovog veka, rastući broj dokaza počeo je da upućuje na činjenicu da kancer ne nastaje samo kao posledica mutacija na nivou genoma, već da značajnu ulogu u ovoj grupi bolesti imaju i izmenjeni mehanizmi epigenetske regulacije diferencijacije i replikacije ćelija. Termin „epigenetika” prvi je u naučnu literaturu uveo Conrad Waddington sredinom 20. veka [13], ali je definicija onoga što epigenetska regulacija podrazumeva od tada više puta revidirana [14, 15]. U kontekstu savremenih saznanja, epigenetska promena predstavlja „strukturnu adaptaciju regije hromozoma koja omogućava beleženje, signaliziranje ili propagaciju izmenjenog stanja aktivacije gena” [16]. Epigenetika tako pruža okvir za razumevanje promena fenotipa koje nastaju bez promena u genotipu, i to kroz proučavanje efekata koje struktura hromatina ima na transkripcionu aktivnost gena. Danas je široko prihvaćeno da epigenetski mehanizmi igraju ključnu ulogu u održavanju normalnog stanja ćelija ali i da njihova deregulacija ima izuzetno važno mesto u etiopatogenezi kancera [13]. Smatra se da izmene „epigenetskog pejzaža” zapravo predstavljaju jednu od definišućih karakteristika malignih ćelija [17, 18].

Prepoznavanje važne uloge koju epigenetska deregulacija ima u malignim bolestima dovelo je i do intenzivnog interesovanja za razvoj lekova koji moduliraju epigenetske procese. Terapijski potencijal epigenetskih ciljnih mesta dejstva proizlazi iz činjenice da su epigenetske promene, za razliku od genetskih, reverzibilnog karaktera što ostavlja mogućnost ponovnog uspostavljanja normalnog epigenetskog statusa ćelije kroz odgovarajuću farmakološku intervenciju [19]. Dokaz ovog kon-

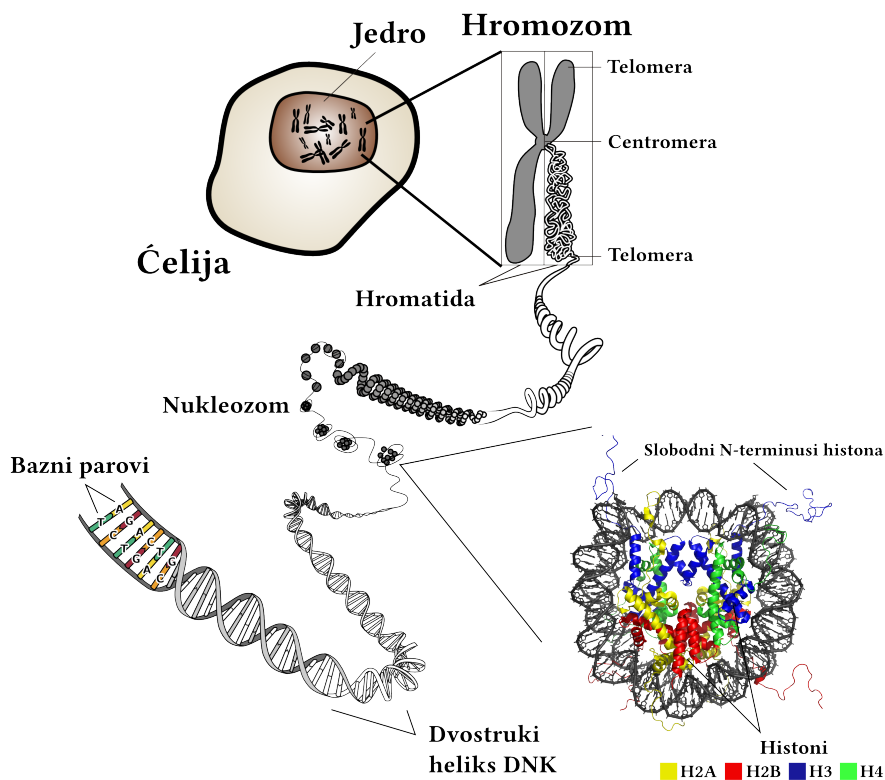
cepta jeste prva generacija antitumorskih lekova koji deluju na epigenetska ciljna mesta dejstva: inhibitori histon deacetilaza ($HDACi$) i inhibitori DNK metiltransferaza ($DNMTi$) danas se koriste u kliničkom okruženju u ograničenom broju indikacija [20, 21]. Iako je prva generacija epigenetskih lekova pokazala potencijal moduliranja epigenetskih promena, ograničeno razumevanje složenosti epigenetskih procesa i mehanizma dejstva ovih lekova tokom perioda njihovog razvoja, uslovia je izostanak selektivnosti za jasno definisane epigenetske regulatore. Stoga je glavni izazov u razvoju druge generacije epigenetskih lekova, sa unapređenom efikasnošću, dizajniranje jedinjenja koja selektivno mogu da utiču na tačno određene epigenetske modifikacije za koje se prethodno utvrdi da su od presudnog značaja za preživljavanje kancerskih ćelija [19, 22, 23].

Osnovna ideja istraživanja sprovedenog u okviru ove doktorske disertacije bilo je upravo davanje doprinosa otkriću novih potencijalnih antitumorskih lekova koji selektivno moduliraju epigenetsko ciljno mesto dejstva kroz racionalno dizajniranje molekula primenom računara. U realizaciji ove ideje, odabran je enzim EZH2, koji pripada grupi protein lizin metiltransferaza (PKMT). PKMT se smatraju veoma perspektivnom klasom novih ciljnih mesta dejstva u razvoju druge generacije epigenetskih lekova [24]. Kao što će u daljem tekstu biti šire opisano, brojni dokazi prikupljeni u poslednjih nekoliko godina ukazali su na značajan terapijski potencijal EZH2, kao i na mogućnost njegove inhibicije primenom malih organskih molekula, na osnovu čega je upravo ovaj enzim izabran kao ciljno mesto dejstva.

Preostali deo uvodnog poglavlja organizovan je na sledeći način: u odeljku 1 dat je kratak pregled epigenetskih regulatornih mehanizama i načina na koji njihova deregulacija može doprineti razvoju kancerskog fenotipa; u odeljku 2 opisani su struktura, funkcija i mehanizmi kojima EZH2 može učestvovati u etiopatogenezi kancera i obrazložena je osnovanost inhibicije ovog enzima u terapijske svrhe; u odeljku 3 prikazan je pregled do sada otkrivenih inhibitora EZH2 i izazova u otkriću novih molekula koje selektivno inhibiraju ovaj enzim; u odeljku 4 dat je kratak pregled mehanizama rezistencije na antitumorske lekove i pristupa kojima se racionalnim dizajniranjem može umanjiti rizik od razvoja rezistencije kod novootkrivenih struktura; konačno, u odeljku 5 prikazane su računarske metode korišćene u ovom istraživanju uz kratku diskusiju o njihovoj ulozi u racionalnom dizajniranju lekova.

1.1 Epigenetika normalnih i malignih ćelija

Genetski materijal eukariota gusto je upakovan u jedru ćelije u vidu hromatina (slika 1.1). Osnovna jedinica građe hromatina jeste nukleozom koji predstavlja asocijaciju približno jednakih masa DNK i histonskih proteina. Jezgro nukleozoma formira oktamer histona, sa po dva heterodimera histona H2A–H2B i H3–H4, oko kojih je spiralno obavijeno 145–147 baznih parova dezoksiribonukleinske kiseline (DNK). C-terminalni krajevi histona formiraju helikse u jezgru nukleozoma, dok su njihovi N-terminalni krajevi intrinzički neuređeni i nalaze se slobodni u spoljašnjosti nukleozoma. Susjedni nukleozomi spojeni su segmentima povezujuće (engl. *linker*) DNK varijabilne dužine, pri čemu inter-nukleozomalne interakcije uslovljavaju formiranje složenih sekundarnih i tercijarnih struktura hromatina: hromatinskih niti, vlakana, hromatida i konačno, hromozoma [25, 26].



Slika 1.1: Različiti nivoi organizacije genetskog materijala u ćeliji, od nukleozoma i hromatinske niti do hromozoma. ©Richard Wheeler / Wikimedia Commons / CC-BY-SA 3.0 / GFDL

Sredinom 60. godina prošlog veka, istraživanja biologije hromatina počela su da ukazuju na činjenicu da hromatin nije inertna potpora genetskog materijala već da njegova struktura u velikoj meri utiče na procese transkripcije, reparacije

i replikacije DNK. Struktura hromatina dinamične je prirode prevashodno usled velikog broja post-translacionih modifikacija histona koje utiču na lokalni stepen dostupnosti genetskog materijala [27]. U ovom kontekstu, razlikuju se dve forme hromatina: euhromatin i heterohromatin. Euhromatin karakteriše niži stepen lokalne kondenzacije što čini zahvaćene gene dostupnim enzimima uključenim u transkripciju i reparaciju. Heterohromatin je kompaktnija forma hromatina, sa većom prostornom koncentracijom nukleozoma, u čijem se sastavu uglavnom nalaze konstitutivno ili fakultativno inaktivni geni. Svi faktori koji utiču na lokalnu strukturu hromatina predstavljaju složen regulatorni nivo koji se ilustrativno naziva „epigenetski pejzaž” ćelije. Različito oblikovanje epigenetskog pejzaža uslovljava nastanak različitih fenotipa u odsustvu promena u sekvenci DNK, čime epigenetske promene igraju ključnu ulogu u procesima normalne diferencijacije ćelija, ali i njihove maligne transformacije. Tri faktora koji u najvećoj meri utiču na epigenetski pejzaž ćelije jesu [28]: metilovanje DNK, kovalentne modifikacije histona i efekti posredovani nekodirajućim ribonukleinskim kiselinama (RNK).

1.1.1 Metilovanje DNK

Metilovanje DNK kod sisara gotovo uvek se odigrava u sklopu CpG dinukleotida, pri čemu je supstrat u reakciji citozin koji se prevodi u 5-metilcitozin aktivnošću enzima DNK metiltransferaza (DNMT) [29]. Do danas su identifikovane tri katalitički kompetentne forme DNMT: DNMT1, DNMT3A i DNMT3B. DNMT1 ima izražen afinitet ka asimetrično (hemimetilovanim) CpG dinukleotidima pa se smatra da je njena glavna uloga održavanje stanja metilovanja DNK nakon replikacije genetskog materijala i deobe ćelije. DNMT3A i DNMT3B su *de novo* metiltransferaze, što podrazumeva da poseduju afinitet za nedomifikovane CpG dinukleotide [30, 31]. Novija saznanja ukazuju na to da aktivnost DNMT1 zavisi od njene lokalizacije u određenim regijama hromatina učešćem DNMT3A ili DNMT3B [32]. Kod eukariota do danas, uprkos aktivnom istraživanju, nije identifikovan enzim koji je specifično uključen u proces aktivnog demetilovanja DNK [33], usled čega je ovo relativno stabilna epigenetska promena. Do pasivnih promena u nivou metilovanja na nivou celokupnog genoma može doći ukoliko je funkcija DNMT1 inhibirana kroz nekoliko ćelijskih generacija.

Funkcionalne posledice metilovanja DNK najviše su proučavane u kontekstu CpG ostrvaca. CpG ostrvca (CGI) su kratki segmenti DNK (1.000–2.000 baznih parova) u kojima je učestalost pojave CpG dinukleotida značajno veća nego u

ostatku genoma. CGI su obično locirana u oblastima transkripcionih start sekvenci, mada se mogu naći i u kodirajućim delovima gena. Metilovanje CGI u sklopu transkripcionih start sekvenci povezano je sa transkripcionom inaktivacijom i tako predstavlja mehanizam epigenetske represije aktivnosti gena. Metilovanje DNK se, međutim, ne može odvijati u odsustvu nukleozoma, u prisustvu H2A.Z varijante histona, kao ni u prisustvu trimetilovanog lizina 4 na histonu 3 (H3K4me3). Kako su ove epigenetske promene karakteristične za aktivno transkribovane gene, metilovanje DNK nije osnovni (inicijatorski) epigenetski mehanizam genske inaktivacije. S druge strane, geni koji su prethodno inaktivirani delovanjem kompleksa Polycomb grupe (odjeljak 1.1.2), imaju značajno veću verovatnoću da postanu metilovani u oblasti transkripcionih start sekvenci. Stoga, metilovanje CGI ovih segmenata DNK verovatno ima ulogu da učvrsti inaktivirano stanje gena koje je prethodno uspostavljeno delovanjem drugih epigenetskih regulatora. Dosledno sa ovim zaključkom, u somatskim ćelijama najveći broj CGI ostaje nemetilovan, dok su izuzetak promotori gena koji podležu dugotrajnoj inaktivaciji (npr. kao kod inaktiviranog X hromozoma) [34].

Za razliku od CGI u okviru transkripcionih start sekvenci, većina drugih CpG nukleotida dominantno se nalazi u metilovanom obliku. Metilovanje DNK u kodirajućim regijama gena nema inhibitorni učinak na elongaciju transkripata niti je dosledno u vezi za inaktivacijom gena. Spontana deaminacija 5-metilcitozina, kojom nastaje timin, jedan je od mehanizama pojave somatskih mutacija. Ovo je ujedno i mehanizam kojim se objašnjava gubitak CGI iz oblasti promotera određenih gena. Primećeno je da ovakvi geni najčešće nisu aktivni, pa se pretpostavlja da je njihovim dugotrajnim metilovanjem evolutivno došlo do gubitka CpG dinukleotida, zamenom citozina timinom. Ova tačkasta mutacija ipak, na određenim mestima u sekvenci DNK, ima i potencijal da doprinese nastanku kancerskog fenotipa [34].

1.1.2 Post-translacione modifikacije histona

Kovalentne modifikacije aminokiselina histona predstavljaju najveću grupu promena koje imaju sposobnost da utiču na lokalnu strukturu hromatina. Do danas je identifikovano preko 60 različitih post-translacionih modifikacija (PTM) histona, od kojih se najveći broj može pronaći na slobodnim, N-terminalnim krajevima ovih proteina [35]. Kako upravo ovi delovi histona stupaju u interakciju sa susednim nukleozomima, kao i sa DNK, prisustvo različitih strukturnih modifikacija histona može direktno da utiče na stepen kondenzacije hromatina. Primera radi, neutrali-

zacija pozitivnog naelektrisanja lizina njegovim acetilovanjem, ili uvođenje negativnog naelektrisanja fosforilacijom serina, utiču na jačinu interakcije između histona i DNK, kao i na blizinu susednih nukleozoma, što se neposredno odražava na lokalnu konformaciju hromatina. Pored ovog direktnog uticaja, važan mehanizam kojim PTM histona regulišu strukturu hromatina jeste regrutovanje drugih signalnih proteina koji se vezuju za hromatin [36, 37]. Ovi proteini sadrže karakteristične domene koji specifično prepoznaju određene modifikacije histona. Njihovo vezivanje za hromatin najčešće inicira formiranje multimernih kompleksa koji su odgovorni za različite funkcionalne ishode – aktivaciju ili inaktivaciju gena – koje PTM histona iniciraju ili regulišu. U samom remodelovanju hromatina, ključnu ulogu imaju i enzimski kompleksi koji koriste energiju oslobođenu hidrolizom adenozin-trifosfata (ATP) za izmeštanje ili uklanjanje nukleozoma, a koji efektivno dovode do prelaska heterohromatina u euhromatin [38].

Još jedna važna karakteristika PTM histona jeste da jedna modifikacija utiče na uvođenje drugih, kao i da funkcionalni ishod jedne modifikacije zavisi od konteksta koji uspostavlja sve ostale modifikacije. Ovaj fenomen međusobne komunikacije PTM histona (engl. *crosstalk*) uslovljava kombinatornu složenost mogućih signala koje dati skup modifikacija može predstavljati. Komunikacija se može odigravati između PTM istog histona („*cis*” efekti) kao i između PTM koje se nalaze na različitim, ali prostorno bliskim histonima („*trans*” efekti) [39, 40]. Sa aspekta prirode međusobne komunikacije, mogu se izdvojiti tri načina na koje PTM utiču jedna na drugu [41, 42]: (i) jedna modifikacija može da deluje stimulatorno ili inhibitorno na aktivnost enzima koji uvodi drugu modifikaciju; (ii) jedna modifikacija može predstavljati vezivno mesto na hromatinu za regrutovanje multimernog kompleksa, u kojem se jedna komponenta vezuje za postojeću modifikaciju a druga poseduje katalitičku aktivnost i modifikuje drugu aminokiselinu histona; (iii) određeni skup modifikacija histona može dovesti do aktivacije enzimskih kompleksa koji ireverzibilno skraćuju N-terminus histona H3 i tako u značajnoj meri menjaju lokalno uređenje hromatina.

Brojnost i raznovrsnost PTM histona, specifičnost enzima koji uvode i uklanjaju ove promene, postojanje proteinskih domena koji specifično prepoznaju određene PTM, kao i složenost njihovih međusobnih komunikacija navele su pojedine autore da predlože postojanje „histonskog kôda” [43] koji, paralelno sa sekvencom DNK, reguliše procese koji uključuju transkripciju, reparaciju i replikaciju DNK. Prema ovoj hipotezi koja je danas u velikoj meri prihvaćena, proteini i enzimi koji uvode, uklanjaju i prepoznaju PTM histona predstavljaju „pisače”, „brisače” i „či-

tače” histonskog kôda, odnosno složenu regulatornu mašineriju koja, modifikujući strukturu hromatina, uslovljava raznovrsne funkcionalne ishode – formiranje različitih fenotipa bez promena u genotipu ćelije [43–45]. Najznačajnije PTM histona i enzimi uključeni u njihovu regulaciju zbirno su prikazani u tabeli 1.1.

Metilovanje histona

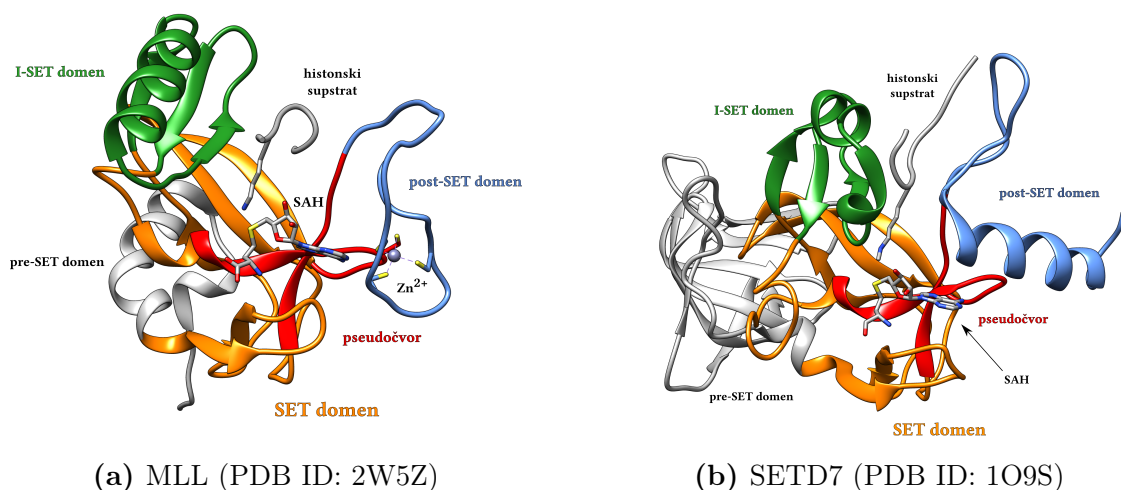
Metilovanje je najraznovrsnija od svih kovalentnih modifikacija histona i ovoj PTM podležu lizin, argininin i, veoma retko, histidin sva četiri osnovna histona. Dodatnu složenost ovoj modifikaciji daje činjenica da lizini mogu biti monometilovani (Kme1), dimetilovani (Kme2) i trimetilovani (Kme3), a arginini mogu biti monometilovani (Rme1), asimetrično dimetilovani (Rme2a) i simetrično dimetilovani (Rme2s). Svaka od ovih promena predstavlja jedinstven epigenetski signal. Postizanje takvog diverziteta signala kroz samo jedan tip PTM histona moguće je zahvaljujući postojanju velikog broja metiltransferaza i demetilaza koje ispoljavaju specifičnost u pogledu supstrata kao i proizvoda reakcija koje katalizuju. U poređenju sa drugim modifikacijama histona, metilovanje se izdvaja po dve karakteristike. Metilovanje predstavlja relativno dugotrajnu promenu, za koju se inicijalno verovalo i da je ireverzibilnog karaktera. Takođe, za razliku od fosforilacije i acetilovanja, metilovanje ne dovodi do promene naelektrisanja N-terminusa histona. Raznovrsni funkcionalni ishodi metilovanja, koje može predstavljati aktivatorni i represivni epigenetski signal, tako nastaju kroz prepoznavanje modifikovanih histona od strane drugih hromatin–vezujućih proteina, a ne kroz direktan uticaj na kompakciju hromatina [36, 52].

Metilovanje lizina Metilovanje histonskih lizina katalizuje familija od preko 30 različitih enzima koji se nazivaju protein lizin metiltransferaze (PKMT). Iako su njihovi primarni supstrati histoni, za više enzima iz ove grupe utvrđen je afinitet i za nehistske proteine [63, 64], usled čega je predloženo da se grupa označava kao protein–, a ne kao histon metiltransferaze [24]. Svaka PKMT ispoljava selektivnost za određene lizine u sekvenci histona, često isključivo jedan, i katalizuje definisan stepen metilovanja svojih supstrata [53]. Zbog značaja ove grupe enzime u istraživanju koje je sprovedeno u okviru ove disertacije, u narednim paragrafima biće kratko izložene najznačajnije karakteristike strukture i mehanizma katalitičke aktivnosti PKMT, pre pregleda funkcionalnih ishoda metilovanja i reverzibilnosti ove modifikacije.

Tabela 1.1: Pregled najčešćih post-translacionih modifikacija histona i enzima uključenih u uvođenje, uklanjanje i prepoznavanje ovih modifikacija

Modifikacija	Značajni supstrati	Enzimi „pisači“	Enzimi „brisači“	Domeni „čitači“	Reference
Acetilovanje	H2AK5	Histon acetiltransferaze (HAT)	Histon deacetilaze (HDAC)	Bromodomen	[46–49]
	H2BK12, H2BK15				
	H3K9, H3K14, H3K18				
	H4K5, H4K8, H4K12, H4K16				
Metilovanje	H1K26	Protein lizin metiltransferaze (PKMT)	Protein lizin demetilaze (PKDM)	Hromodomen Tudor domen PHD prsti MBT domen WD40 domen	[50–53]
	H3K9, H3K27, H4K20				
	H3K4, H3K26, H3K79				
	H3R2, H3R8				
	H3R17, H3R26				
Fosforilacija	H3S10, H3S28 H3T11	Histon kinaze	Histon fosfataze	14-3-3 domen	[54, 55]
	Ubikvitinacija	H2AK119	Ring1B, 2A-HUB	2A-DUB	?
H2BK120		HR6A/B, RNF20	USP22	Cps35	
Sumoilacija	H1-H4K	UBC9	?	?	[57, 58]
Izomerizacija prolina	H3P30, H3P38	Fpr4	/	/	[59]
Citrulinacija	H3R17, H4R3	Protein arginin deiminaza 4 (2?)	?	?	[60, 61]
ADP ribozilacija	H1E2, H1E4, H1E116 H2BE2, H2BK30, H3K27	ADP-riboziltransferaze (ART)	ADP-ribozilhidrolaze (ARH) PAR glikohidrolaze (PARG)	?	[62]

Sa izuzetkom DOT1L, sve PKMT otkrivene do danas sadrže SET domen koji je odgovoran za katalitičku aktivnost ovih enzima. SET domen je sekvenca od oko 130 aminokiselina koje formiraju karakterističnu terciarnu strukturu pseudočvora u kojoj se u neposrednoj blizini nalaze dva očuvana motiva sekvenci RFIN-HXCXPN i ELXFDY, gde je X bilo koja aminokiselina [65]. SET domen formira vezivna mesta za kofaktor *S*-adenozil-L-metionin (SAM) kao i histonski supstrat (slika 1.2). Histonski supstrat vezuje se u usek koji zajedno formiraju I-SET i kratak post-SET domen (20–30 aminokiselina), pri čemu se lizin umeće u kanal koji dominantno formiraju hidrofobni bočni lanci okolnih aminokiselina. Mesto za vezivanje kofaktora locirano je sa suprotne strane proteina i u njegovom oblikovanju, pored SET domena, učestvuje post-SET domen, koji zatvara vezivno mesto i stabilizuje kofaktor. U poređenju sa sekvencama SET domena, post-SET domen je značajno varijabilniji unutar PKMT familije, ali je kod većine enzima karakteristično prisustvo većeg broja cisteina u sekvenci. Dostupne kristalografski rešene strukture pokazale su da kod ovih predstavnika PKMT, tri cisteina iz post-SET domena koordiniraju jon cinka gradeći tetraedarski kompleks koji utiče na finalnu strukturu holo enzima (slika 1.2a). Kod PKMT koje ne sadrže cisteinom-bogat post-SET domen, koordiniranje cinka ne igra ulogu u pozicioniranju i strukturi ovog dela enzima. U tim slučajevima, post-SET domeni uglavnom obrazuju fleksibilne helikse čije su pojedine aminokiseline u kontaktu sa kofaktorom (slika 1.2b) [66]. Važno je naglasiti i da u apo obliku enzima, post-SET domen PKMT jeste delimično ili potpuno neuređen a karakterističnu sekundarnu strukturu i položaj u odnosu na SET domen zauzima tek nakon asocijacije enzima sa kofaktorom i histonskim supstratom.



Slika 1.2: Trodimenzionalne strukture protein lizin metiltransferaza sa SET domenom, u kompleksu sa histonskim supstratom i proizvodom kofaktora (SAH).

Na osnovu dostupnih biohemijskih i strukturnih podataka, predložen je sledeći redosled događaja kojima PKMT vrše metilovanje histona [67]: (i) vezivanje kofaktora za enzim dovodi do pozicioniranja post-SET domena u odnosu na ostatak enzima na način koji dovodi do formiranja useka u koji se uklapa histonski peptid; (ii) prethodno formirani binarni kompleks vezuje se za histon posredstvom elektrostatičkih interakcija dugog dometa između pozitivno naelektrisane površine histona i negativne površine useka na enzimu; (iii) specifično prepoznavanje ciljnog lizina odigrava se na osnovu komplementarnosti sekvence histona i rezidua I-SET i post-SET domena koje formiraju usek; (iv) ulaskom ciljnog lizina u kanal enzima, formirani tercijarni kompleks omogućava deprotonovanje ε -azota lizina i prenos metil grupe sa kofaktora SAM na azot, pri čemu se formira proizvod kofaktora *S*-adenozil-L-homocistein (SAH) i metil-lizin; (v) post-SET domen se zatim ponovo izmešta, što omogućava disocijaciju SAH i naredni krug metilovanja.

Mehanizam transfera metil grupe i strukturna osnova specifičnosti proizvoda PKMT bili su predmet nekoliko računarskih i kristalografskih studija [68–72]. Reakcija se odigrava S_N^2 mehanizmom nukleofilne supstitucije i zahteva deprotonovanje lizina mada nije u celosti izvesno da li u ovom koraku učestvuju aminokiseline aktivnog centra. Pojedine studije pokazale su da ovo jeste slučaj barem kod određenih PKMT i da značajnu ulogu u reakciji transfera ima formiranje vodenog kanala, koji omogućava efikasno deprotonovanje supstrata. Step en metilovanja koji je dati enzim sposoban da katalizuje u najvećoj meri zavisi od zapremine dna kanala u koji se umeće lizin kao i od prisustva čvrsto vezanih molekula vode koje mogu ometati reakciju transfera. Sterno prostraniji aktivni centri daju mogućnost formiranja di- i trimetilovanih lizina. Za specifičnost proizvoda reakcije takođe je veoma značajan i očuvani tirozin aktivnog centra (Y641 u sekvenci EZH2). Nekoliko studija mutageneze ukazalo je na potencijal tirozin-fenilalanin izmene da utiče na specifičnost proizvoda PKMT [67]. Prisustvo tirozina u najvećem broju slučajeva smanjuje efikasnost enzima u katalizovanju trimetilovanja, dok čini mono- i dimetilovanje veoma efikasnim. Na osnovu analize dostupnih struktura, verovatno je da ovaj tirozin učestvuje kao akceptor u formiranju vodonične veze sa nemetilovanim i monometilovanim supstratima, i tako doprinosi orijentaciji slobodnog elektronskog para ε -azota u odnosu na metil grupu SAM, dok kod dimetilovanog supstrata kao donor uspostavlja vodoničnu vezu za slobodnim elektronskim parom, tako sprečavajući njegovu optimalnu orijentaciju. U skladu sa ovom hipotezom, Y/F izmenom najčešće dolazi do pretvaranja efikasne monometilaze u manje efikasne di- ili trimetilaze [73, 74].

Do 2004. godine, metilovanje lizina smatrano je ireverzibilnom prome-

nom. Ove godine je otkrivena prva lizin demetilaza (LSD1), za koju je utvrđeno da koristi flavin adenin dinukleotid kao kofaktor u reakciji uklanjanja metil grupe sa H3K4me1/me2 na izolovanim histonima. Kasnije je pokazano da u sklopu korepresivnih kompleksa, LSD1 pokazuje afinitet i za H3K4me1/me2 u sklopu nukleozoma, čime je potvrđena biološka uloga ovog enzima kao protein lizin demetilaze. Dalja istraživanja dovela su do otkrića dodatne klase lizin demetilaza koje sadrže C-terminalni Jumonji domen, a za reakciju uklanjanja metil grupa koriste Fe^{2+} jon i α -ketoglutarat kao kofaktore. Većina lizin demetilaza aktivna je u sklopu većih hromatin-modifikujućih kompleksa što omogućava prepoznavanje prisutnih PTM histona i uslovljava specifičnost uklanjanja metil grupa [75].

Funkcionalni ishodi metilovanja lizina zavise od specifične aminokiseline u sekvenci koja je modifikovana, kao i od stepena metilovanja odnosno broja uvedenih metil grupa. Uopšteno, može se reći da su H3K9me3, H3K27me3 i H4K20me3 represivni epigenetski signali i da su uključeni u globalne promene nivoa ekspresije gena kroz formiranje konstitutivnog i fakultativnog heterohromatina. Na nivou pojedinačnih gena, najznačajniju represivnu ulogu imaju proteini Polycomb grupe (PRC2 i PRC1, odeljak 1.2.2), koji uvođenjem H3K27me3 i daljom signalnom kaskadom mogu selektivno inaktivirati tačno određene gene. Nasuprot represivnih epigenetskih signala, H3K36me3 i H3K79me3 javljaju se u oblastima euhromatina. Funkcija PTM H3K4 zavisi od nivoa metilovanja: H3K4me3 uvek se nalazi u okviru aktivno transkribovanih gena dok se H3K4me2 može naći i u regijama hromatina koje sadrže inaktivirane gene [36, 50, 51]. Potpuno razumevanje značaja metilovanja histonskih lizina i složenosti interakcija ove PTM sa drugim epigenetskim promenama zahteva još mnogo istraživanja. Izvesno je da je metilovanje dinamična promena od velikog značaja za epigenetski pejzaž ćelije.

Metilovanje arginina Metilovanje arginina kod sisara katalizuje familija od 11 enzima koji su podeljeni u dve grupe. Tip I protein arginin metiltransferaza (PRMT) obuhvata enzime koji katalizuju formiranje Rme1 i Rme2a (PRMT1, PRMT3, CARM1, PRMT6, PRMT8), dok tip II PRMT katalizuje formiranje Rme1 i Rme2s (PRMT5, PRMT7). U poređenju sa metilovanjem lizina, metilovanje arginina je manje specifična modifikacija. Za oko 85% svih modifikacija arginina odgovorna je PRMT1 koja ispoljava veoma široku supstratnu specifičnost i u stanju je da metiluje većinu histonskih arginina koji su locirani neposredno do jednog ili dva glicina [76]. Druge PRMT pokazuju veću specifičnost; primera radi, CARM1 ulazi u sastav SWI/SNF remodelera hromatina u okviru kojeg metiluje isključivo H3R17 i H3R26.

Inicijalno se smatralo da do demetilovanja arginina dolazi samo dejstvom deiminaza, koje prevode monometilovan arginin u citrulin, čije prisustvo takođe predstavlja epigenetski signal. Ova reakcija, međutim, ima relativno mali biološki značaj, jer se retko odigrava, čemu doprinosi i činjenica da je monometilovani arginin veoma slab supstrat za deiminaze. Posledično, monometilovanje arginina zapravo predstavlja promenu koja inhibitorno deluje na citrulinaciju [77]. Reverzibilnost metilovanja arginina u užem smislu potvrđena je kroz otkriće da lizin demetilaza sa Jumonji domenom, JMJD6, takođe pokazuje afinitet za, i demetiluje H3R2 [78]. Postojanje drugih specifičnih PRDM do danas nije ustanovljeno.

Metilovanje arginina, kao i druge PTM histona, može delovati inhibitorno i aktivatorno na uvođenje dodatnih modifikacija. Primera radi, prisustvo H3R2me2a inhibira metilovanje H3K4, dok H3K4me3 onemogućava metilovanje H3R2 [52]. PRMT funkcionišu u okviru kompleksa koji imaju represivne i aktivatorne uloge [77], pa i epigenetski signali koje ove promene predstavljaju zavise od specifičnih arginina koji su modifikovani. O funkcionalnim ishodima metilovanja arginina relativno se malo zna, ali je po analogiji sa modifikacijama lizina verovatno da ove promene prepoznaju određeni efektorski proteini koji u daljoj signalnoj kaskadi vrše represiju ili aktivaciju gena [79].

Acetilovanje histona

Acetilovanje histona katalizuju dve grupe acetiltransferaza za koje je zajedničko da koriste acetyl–koenzim A kao kofaktor u reakciji prenosa acetyl grupe na ϵ –azot lizina. Tip B acetiltransferaza su citoplazmatski enzimi koji acetiluju novosintetisane histone pre njihovog transporta u jedro ćelije, pri čemu je ovaj korak važan za interakciju histona sa proteinima pratiocima (engl. *chaperones*) koji učestvuju u sklapanju nascentnih nukleozoma. Nakon asocijacije histona i DNK, modifikacije uspostavljene od strane tipa B acetiltransferaza se većinom uklanjaju i nemaju većeg značaja u epigenetskoj regulaciji. Glavnu ulogu u regulaciji strukture hromatina acetilovanjem ima tip A acetiltransferaza, koje predstavljaju enzime aktivne u jedru ćelije [37].

Tip A histon acetiltransferaza (HAT) može se podeliti u tri familije prema sličnosti sekvenci i terciarnoj strukturi: GNAT, MYST i CBP/p300. Značajna različitost sekvenci koja je uočena u okviru ove familije enzima, koja deli zajednički katalitički mehanizam, omogućava postizanje jedinstvenih bioloških efekata

različitih HAT. Iako su ovi enzimi samostalno katalitički kompetentni za acetilovanje slobodnih histona, više predstavnika HAT pokazuje afinitet za nukleozome tek nakon asocijacije u multimerne komplekse [80]. Takođe, većina izolovanih HAT ispoljava supstratnu polispecifičnost ali u sklopu različitih HAT kompleksa katalizuje modifikacije tačno određenih histona [48].

Acetilovanje histona je reverzibilna promena, pri čemu u uklanjanju acetil grupa sa lizina učestvuje veći broj histon deacetilaza (HDAC). HDAC se mogu podeliti u 5 klasa: klasu I (HDAC1, HDAC2, HDAC3 i HDAC8), klasu IIa (HDAC4, HDAC5, HDAC7 i HDAC9), klasu IIb (HDAC6 i HDAC10), klasu III (sirtuini) i klasu IV (HDAC11). Katalitička aktivnost klasičnih predstavnika HDAC zavisi od prisustva jona cinka u aktivnom centru, dok je mehanizam deacetilovanja sirtuina drugačiji i podrazumeva učešće nikotinamid adenin dinukleotida kao kofaktora u reakciji [80]. Kao i HAT, HDAC ispoljavaju nizak stepen supstratne specifičnosti, ali sa izuzetkom HDAC8, ovi enzimi stižu katalitičku kompetentnost tek ulaskom u sastav multimernih kompleksa. Ciljno regrutovanje formiranih kompleksa u oblast tačno određene modifikacije histona obezbeđuje da je aktivnost HDAC veoma specifična i strogo regulisana [81].

Složenost regulatorne funkcije HDAC najbolje ilustruje primer enzima klase IIa. Ova grupa enzima ispoljava minimalnu katalitičku aktivnost na acetilovanim histonima, što upućuje na mogućnost da su njihovi glavni supstrati zapravo citoplazmatski proteini. Istovremeno, međutim, HDAC klase IIa predstavljaju važne epigenetske supresore genske aktivnosti pa je verovatno da regulatorna funkcija HDAC nije ograničena na njihovu histon-deacetilaznu, enzimsku aktivnost [49].

Kako acetilovanje dovodi do neutralizacije pozitivnog naelektrisanja lizina, ova PTM histona ima potencijal da direktno utiče na lokalnu konformaciju hromatina. Acetilovanjem se smanjuje jačina interakcije histona i DNK, a posebno značajnu ulogu u ovom kontekstu ima histon H4. N-terminus H4 stupa u interakcije sa negativno naelektrisanim površinama H2A–H2B dimera susednog nukleozoma, pa njegovo acetilovanje uslovljava udaljavanje nukleozoma i favorizuje formiranje euhromatina. *In vitro*, acetilovanje H4K16 dovoljno je da spreči formiranje kompaktnih hromatinskih niti [82], ali značajnost ove ili neke druge modifikacije u *in vivo* uslovima nije jednostavno utvrditi zbog izrazito dinamične prirode acetilovanja. Uopšteno, regije aktivno transkribovanog hromatina karakteriše visok nivo acetilovanja histona, dok je heterohromatin u daleko manjoj meri acetilovan. Stoga se uzima da acetilovanje generalno doprinosi aktivaciji gena a deacetilovanje uzrokuje nji-

hovu represiju. Pored nespecifičnog uticaja na kompakciju hromatina, acetilovanje određenih lizina histona može predstavljati signal za regrutovanje drugih epigenetskih regulatora, slično mehanizmu kojim metilovanje histona ostvaruje regulatorne funkcije [46].

Nivo acetilovanja histona biološki je precizno regulisan interakcijom velikog broja faktora. U regulaciji učestvuju HAT, HDAC, drugi enzimi koji mogu modifikovati histonske lizine kao i enzimi uključeni u metabolizam acetil koenzima A [47]. Nedavno je pokazano i da na nivo acetilovanja u velikoj meri utiče intracelularna pH [83].

Druge modifikacije

Fosforilacija histona odigrava se na serinima, treoninima i tirozinima N-terminalnih krajeva histona, ali i, ređe, u delovima histona koji čine jezgro nukleozoma. Fosforilacija je izuzetno dinamična promena u čijem uspostavljanju može učestvovati preko 10 kinaza. Fosfataze direktno uključene u defosforilaciju histona su slabije okarakterisane i manje se zna o njihovoj epigenetskoj ulozi. Izvesno je da je aktivnost fosfataza važna zbog relativno velike brzine kojom se nivo fosforilacije histona može izmeniti [37]. Biološki efekti fosforilacije histona najviše su proučavani na nivou pojedinih modifikacija histona H3. Koordinisana fosforilacija H3S10, H3S28, H3T3 i H3T11 su promene za koje se smatra da ključno doprinose kondenzaciji hromatina tokom mitoze ćelija. Tačan mehanizam kojim do kondenzacije dolazi, međutim, nije dovoljno proučen. Nasuprot ulozi u kondenzaciji hromatina, fosforilacija za posledicu takođe može imati i aktivaciju gena. Trenutno dostupni dokazi ukazuju na to da fosforilacija H3S10 dovodi do transkripcione aktivacije, ali nije izvesno da li je ovo direktna posledica fosforilacije ili promena koja inicira acetilovanje H3K14 [54, 55].

Ubikvitinacija i sumoilacija se izdvajaju od svih drugih modifikacija histona po veličini strukturne izmene do koje dolazi. Ubikvitin je protein dužine 76 aminokiselina kojim se kovalentno modifikuju najčešće lizini histona sekvencijalnim delovanjem tri enzima. Prvi enzim (E1) vrši ATP-zavisnu aktivaciju ubikvitina. U sledećem koraku, aktivirani ubikvitin se tioestarskom vezom spaja sa ubikvitin-konjugujućim enzimom (E2). U finalnom koraku, ubikvitin se prenosi na ciljni lizin histona delovanjem ubikvitin-protein izopeptidne ligaze (E3). Iako, teorijski, više aminokiselina histona može biti supstrat za konjugaciju sa ubikviti-

nom, dominantno se ova modifikacija kod sisara odigrava kao monoubikvitinacija H2AK119 (H2AK119ub1) i H2BK120 (H2BK120ub1) [56]. Više izopeptidaza je sposobno za uklanjanje ubikvitina sa histona što čini ovu modifikaciju histona neočekivano dinamičnom. Interesantno je da dve navedene promene imaju suprotne biološke efekte. H2BK120ub1 igra važnu ulogu u transkripcionoj inicijaciji i elongaciji, dok H2AK119ub1 inhibira metilovanje H3K4 i dovodi do represije gena [56]. SUMO proteini (engl. *small ubiquitin-related modifier*) se konjuguju sa histonskim proteinim sličnim mehanizmom kao i ubikvitin. Sumoilaciji podležu sva četiri histona a biološki efekti ove modifikacije su većinom represivnog karaktera. Ovo se objašnjava činjenicom da sumoilacija određenih lizina sprečava njihovo acetilovanje i ubikvitinaciju, koje bi inače dovele do aktivacije gena [57, 58].

1.1.3 Nekodirajuće RNK

Sekvence DNK koje kodiraju proteine čine samo oko 1,5–2% genetskog materijala sisara, iako je procenat genoma koji podleže transkripciji veći. Oni RNK transkripti koji ne podležu translaciji u proteinske proizvode nazivaju se nekodirajućim RNK. Biološki i patobiološki značaj ove klase RNK predmet je aktivnog istraživanja i danas je izvesno da nekodirajuće RNK predstavljaju još jedan važan nivo epigenetske kontrole ćelijskih procesa [84]. Nekodirajuće RNK se mogu, po dužini sekvence i funkciji, podeliti u veći broj grupa ali su duge nekodirajuće RNK i mikro RNK najviše proučavane [85].

Duge nekodirajuće RNK

Duge nekodirajuće RNK (*dnk*RNK) sadrže više od 200 nukleotida i nastaju prevashodno aktivnošću RNK polimeraze II. Okarakterisani su i drugi putevi formiranja, uključujući isecanje segmenta transkripta prilikom sinteze malih nukleolarnih RNK kao i put sinteze u kojem učestvuje RNK polimeraza III. Nivoi određenih *dnk*RNK u različitim tipovima somatskih ćelija pokazuju značajnu varijaciju što ukazuje na činjenicu da je ekspresija *dnk*RNK strogo kontrolisana. Biohemijski, *dnk*RNK ne poseduju definišuće karakteristike koje ih izdvajaju od informacione RNK, pa se pripadnost ovoj klasi nukleinskih kiselina utvrđuje na osnovu postojanja određene regulatorne funkcije koju *dnk*RNK obavlja. Procenjuje se da je u ljudskom genomu prisutno barem 5.000–15.000 sekvenci koje kodiraju *dnk*RNK. Brojnost specifičnih funkcija koje ove molekule mogu da obavljaju uslovljena je raz-

novrsnošću sekundarnih i tercijarnih strukturnih motiva koje grade, pri čemu je u tom smislu značajno i da je u okviru *dnk*RNK klase pokazano prisustvo oko 100 vrsta kovalentnih modifikacija nukleotida koje utiču na njihovu strukturu i interakcije sa drugim biomolekulama [86].

Izučavanje regulatornih funkcija *dnk*RNK ukazalo je na četiri moguća mehanizma kojima ovi polinukleotidi učestvuju u epigenetskoj kontroli ćelijskih procesa [87]: (i) kako je njihova transkripcija (epigenetski) strogo regulisan proces, *dnk*RNK mogu se posmatrati kao signalne molekule koje integrišu druge epigenetske signale, delujući kao indikatori transkripciono dostupnog hromatina, ili kao represivni faktori koji sprečavaju transkripciju gena za koje se vezuju; (ii) *dnk*RNK mogu biti „mamci” za transkripcione faktore, čijim kompetitivnim vezivanjem titriraju nivo transkripcije drugih gena; (iii) *dnk*RNK mogu se vezivati za određene sekvence DNK a zatim predstavljati molekularnu platformu za vezivanje drugih hromatin-modifikujućih enzima i time mogu omogućiti ciljano regrutovanje represivnih ili aktivatornih kompleksa u regije tačno određenih gena; (iv) *dnk*RNK mogu biti skeletne molekule za formiranje multimernih nukleoproteinskih kompleksa koji na specifične načine mogu učestvovati u modifikaciji hromatina.

Mikro RNK

Mikro RNK (*mi*RNK) su jednolančani segmenti RNK dužine oko 20–23 nukleotida čija primarna regulatorna uloga jeste represija genske aktivnosti kroz inhibiciju translacije. Ovi kratki polinukleotidi nastaju u jedru kao znatno duži primarni transkripti (*pri-mi*RNK) koje sintetišu RNK polimeraze II i III. Pre aktivnog transporta iz jedra ćelije, *pri-mi*RNK skraćuje Drosha-DGCR8 (Pasha) kompleks, pri čemu nastaju *pre-mi*RNK molekule karakteristične tercijarne strukture „ukosnice” (engl. *hairpin*). Ova karakteristična struktura *pre-mi*RNK formira se tradicionalnim Watson–Crick baznim sparivanjem nukleotida 3' i 5' kraja, pri čemu nesparene baze formiraju kružnu petlju na jednom kraju polinukleotida. Dalja biogeneza krajnjeg oblika *mi*RNK odvija se u citoplazmi, aktivnošću Dicer kompleksa, koji iseca kratak 5' segmenat iz *pre-mi*RNK molekule. Konačno, asocijacijom *mi*RNK sa Argonaute proteinima i ciljnom informacionom RNK nastaje kompleks odgovoran za translacionu represiju (engl. *mRNA-induced silencing complex*, RISC). Iako putevi biogeneze većine *mi*RNK dele sličnosti, regulacija ovog procesa je složena i poznato je da postoje brojni alternativni putevi koji doprinose kontroli nivoa određene *mi*RNK [88]. Na molekularnom nivou, specifičnost interakcije informacione RNK i *mi*RNK prili-

kom formiranja RISC kompleksa uslovljena je specifičnim prepoznavanjem na nivou 5' segmenta *mi*RNK koji najčešće obuhvata 7 nukleotida položaja 2–8. Klasično Watson–Crick sparivanje ovih kratkih segmenata RNK nije jedini mehanizam koji obezbeđuje specifičnost; neklasično bazno sparivanje i interakcije drugih regija nukleotida takođe mogu imati značajnu ulogu [89].

Mikro RNK učestvuju u regulaciji aktivnosti između 30 i 60% gena sisara, sprečavajući translaciju informacione RNK ili doprinoseći njenoj degradaciji u citoplazmi ćelije [90, 91]. Primarni efekat *mi*RNK stoga ima represivan karakter, ali je od skora poznato i više primera koji ukazuju na to da ove regulatorne molekule mogu i stimulatorno delovati na iniciranje translacije [91]. Efekti *mi*RNK regulisani su i zavise od drugih epigenetskih faktora koji utiču na nivoe njihove transkripcije i degradacije, a složene interakcije ovog tipa tek od skora su predmet aktivnog proučavanja [92, 93].

1.1.4 Promene epigenetskog pejzaža malignih ćelija

Proučavanje biologije hromatina normalnih i maligno transformisanih ćelija ukazalo je na činjenicu da epigenetski faktori igraju važnu ulogu u procesu patogeneze kancera. Iako mutacije određenih gena predstavljaju ključan događaj u nastanku kancerskih ćelija, epigenetske promene doprinose uspostavljanju specifičnih transkripcionih programa koji uslovljavaju ekspresiju gena od presudnog značaja za molekulske puteve maligne transformacije [17, 18, 94]. Budući da održavanje visokih nivoa proizvoda gena ključnih za preživljavanje kancerskih ćelija može imati epigenetsku osnovu, izmenjen epigenetski pejzaž smatra se ključnom karakteristikom kancerskog fenotipa [95]. Izmenjeni novi metilovanja DNK, promene u PTM histona kao i poremećena ekspresija *mi*RNK svi su dovedeni u vezu sa nastankom i progresijom različitih formi kancera [96].

Jedan od prvih epigenetskih poremećaja za koje je ustanovljeno da se sa visokom učestalošću javlja u kancerskim ćelijama jeste hipometilacija DNK. U maligno transformisanim ćelijama, nivoi metilacije DNK najčešće su niži u odnosu na zdrave ćelije istog tipa, ali je istovremeno primetna zastupljenost hipermetilovanih oblasti. Delovi genoma zahvaćeni hipermetilacijom često se nalaze u regijama promotera tumor–supresorskih gena, što ukazuje na to da hipermetilacija ovih oblasti doprinosi represiji tumor–supresora. S druge strane, globalna redukcija nivoa metilovanja DNK može doprineti sveukupnoj nestabilnosti genoma, a u okviru spe-

cifičnih sekvenci, može uzrokovati ekspresiju gena koji doprinose imortalizaciji ćelija i gubitku kontrole ćelijskog ciklusa [97]. Iako je relativno jasno kako promene u metilaciji DNK mogu doprineti klonskoj ekspanziji maligno transformisanih ćelija, manje je jasno kako ove promene nastaju, a posebno šta dovodi do ciljnog regrutovanja DNMT u regije određenih gena. Verovatno je da promene DNK metilacije nastaju nishodno od drugih aberacija epigenetske regulacije ili u interakciji DNMT sa onkogenim transkripcionim faktorima [98].

Dalji napredak bioanalitičkih metoda omogućio je prepoznavanje činjenice da u kancerskim ćelijama dolazi i do izraženih promena u nivoima zastupljenosti određenih PTM histona. Dok su neke promene ovog tipa nespecifične i javljaju se u većini kancerskih ćelija (npr. gubitak H4K16ac i H4K20me3), druge su karakteristične za određene forme tumora [98]. Prekomerna ekspresija, izmenjena supstratna specifičnost i aberantno regrutovanje epigenetskih „pisača”, „čitača” i „brisača” su tri osnovna mehanizma koja doprinose promenama epigenetskog pejzaža na nivou strukture hromatina [99]. Do ovih promena najverovatnije dolazi usled mutacija gena koji kodiraju epigenetske regulatore, mada novija istraživanja ukazuju da i bez mutacija na nivou genotipa, izlaganje određenim organskim zagađivačima i teškim metalima može dovesti do izmenjene ekspresije i aktivnosti epigenetskih regulatora [99].

Postoji nekoliko dobro opisanih primera koji ilustruju kako sprega genetskih i epigenetskih promena dovodi do gubitka normalne regulacije ćelijskog ciklusa i patogeneze kancera. Pojedine proteinske himere (proizvodi aberantne fuzije dva gena), karakteristične za neke oblike leukemije, poseduju sposobnost strukturne asocijacije sa HDAC i mogu regrutovati ove enzime u oblast važnih tumor-supresorskih gena. Deacetilovanjem histona i uspostavljanjem heterohromatina, HDAC na ovaj način mogu posredovati u represiji tumor-supresora [100]. Slično, hromozomske translokacije mogu dovesti do formiranja himera PKMT MLL i brojnih hromatin-vezujućih proteina. MLL-AF9 himera ne sadrži funkcionalan metiltransferazni domen (MLL je specifična za H3K4), ali regrutuje DOT1L koji trimetiluje H3K79. Aktivnošću DOT1L na ovaj način dolazi do ekspresije nekoliko onkogeno sa leukemogenim potencijalom [101]. Slični primeri poremećene aktivnosti „čitača” i „brisača” histonskog kôda ukazuju na to da gubitak normalne transkripcione kontrole posredovan epigenetskim faktorima značajno doprinosi nastanku određenih formi kancera [102, 103]. Važno je uočiti da u svim navedenim primerima farmakološka modulacija epigenetskog regulatora potencijalno može da povрати normalnu kontrolu ćelijskog ciklusa. Naime, navedene himere same po sebi ne poseduju sposobnost održavanja izmenje-

nog transkripcionog programa, već zavise od funkcije epigenetskih enzima. Dok je genetska promena (fuzija gena) ireverzibilna, izmenjen epigenetski pejzaž moguće je preoblikovati.

Pored hromatin-modifikujućih enzima, *mi*RNK takođe učestvuju u oblikovanju epigenetskog pejzaža malignih ćelija. Mikro RNK učestvuju u regulaciji nivoa više epigenetskih enzima, poseduju sposobnost ciljnog regrutovanja drugih epigenetskih regulatora, a nivoi ekspresije ovih polinukleotida su i sami pod kontrolom epigenetskih faktora. Uloga *mi*RNK u patogenezi kancera još uvek je nedovoljno razjašnjena, pri čemu trenutno dostupni dokazi ukazuju da ove molekule mogu predstavljati onkogene ali i tumor-supresore [104].

1.1.5 Terapijski potencijal epigenetskih ciljnih mesta dejstva

Kao što je prethodno navedeno, za razliku od genetskih mutacija, epigenetske promene su reverzibilnog karaktera što ukazuje na potencijal da se farmakološkom modulacijom epigenetskih regulatora može ponovno uspostaviti normalan epigenetski pejzaž ćelije. Međutim, treba imati u vidu da se u malignim ćelijama može identifikovati veliki broj epigenetskih aberacija od kojih velika većina predstavlja samo manifestacije značajnih promena koje se u ćeliji odigravaju. Stoga ključan izazov u eksploataciji terapijskog potencijala epigenetskih ciljnih mesta dejstva predstavlja identifikovanje onih promena koje su suštinski neophodne za nastanak i održavanje kancerskog fenotipa (engl. *driver changes*), odnosno njihovo razlikovanje od promena koje samo nastaju kao posledica prethodno izmenjenog epigenetskog pejzaža (engl. *passenger, bystander effects*) [19, 105]. Ovo suštinsko razumevanje biologije malignih ćelija od ključnog je značaja i za identifikaciju pacijenata koji će u najvećoj meri profitirati od personalizovane terapije [106].

U okviru epigenetskih regulatora, mogu se razlikovati dve kategorije potencijalnih ciljnih mesta dejstva [107]. Prvu grupu čine *bona fide* onkogeni, čije somatske mutacije ili prekomerna ekspresija samostalno dovode do maligne transformacije i od čije je funkcije tumorska ćelija apsolutno zavisna (koncept zavisnosti od onkogeno [108, 109]). Drugu grupu čine ne-onkogeni, regulatori koji nisu mutirani niti prekomerno ekspimirani, ali koji kooperativno sa onkogenom omogućavaju preživljavanje maligne ćelije što čini njihovu inhibiciju toksičnom za kancer (koncept sintetske letalnosti [110, 111]). Prethodno izložen primer uloge DOT1L u razvoju leukemije predstavlja primer sintetski letalnog ciljnih mesta dejstva; DOT1L nije

onkogen sam po sebi, ali leukemogeni potencijal MLL–AF9 himere zavisi od enzimske aktivnosti DOT1L. S druge strane, EZH2, diskutovan u narednom odeljku, predstavlja onkogen od čije patološke funkcije zavisi preživljavanje pojedinih tipova limfomskih ćelija. I jedan i drugi koncept pružaju osnovu za razvoj visokoselektivnih terapijika [112].

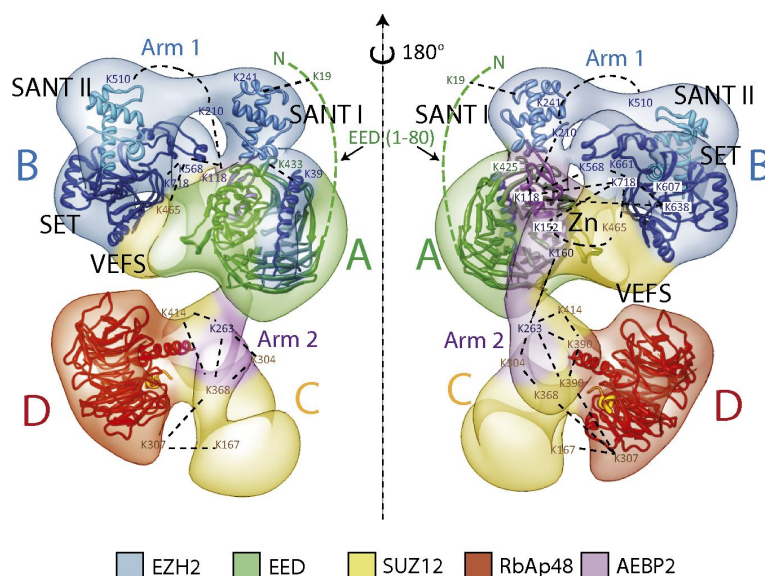
Još jedna važna karakteristika koja čini epigenetska ciljna mesta dejstva perspektivnim jeste mogućnost njihove modulacije malim organskim molekulama. Po analogiji sa onkogenim kinazama, većina epigenetskih enzima sadrži ortosterna mesta i dostupne strukture ovih enzima jasno ukazuju na pogodnost njihove inhibicije racionalno dizajniranim molekulama. Istraživanje hemijske biologije i uspehi u razvoju prve generacije epigenetskih lekova predstavljaju dobru osnovu za dalje istraživanje i razvoj potencijalnih lekova za ovu novu klasu ciljnih mesta dejstva (pregledno prikazano u [19, 20, 22, 113–117]).

1.2 Struktura, funkcija i onkogenost EZH2

Enhancer of Zeste Homolog 2 (EZH2) je protein lizin metiltransferaza sa SET domenom koja pripada grupi Polycomb (PcG) proteina. PcG proteini su evolutivno očuvana grupa proteina koji ulaze u sastav dva multimerna kompleksa koji modifikuju strukturu hromatina i deluju represivno na ekspresiju gena. EZH2 je katalitička podjedinica jednog od ova dva kompleksa, koji se označava kao Polycomb Repressive Complex 2 (PRC2) [118]. PRC2 je iz ćelija prvobitno izolovan kao četvorokomponentni kompleks koji formiraju EZH2, EED (Embryonic Ectoderm Development), SUZ12 (Suppressor of Zeste 12 protein homolog) i RbAp46/48 (Retinoblastoma-binding protein p46/p48) [119]. Kompleks je kasnije prečišćen sa petom komponentom: AEBP2 (Adipocyte enhancer-binding protein 2). PRC2 je jedini hromatin-modifikujući kompleks koji je sposoban za metilovanje H3K27, pri čemu H3K27me3 predstavlja represivni signal i u vezi je sa formiranjem konstitutivnog i fakultativnog heterohromatina. Iako je za metiltransferaznu aktivnost odgovoran EZH2, ova PKMT samostalno nije katalitički kompetentna i zahteva asocijaciju sa EED i SUZ12 koji zajedno čine minimalni funkcionalni trikomponentni kompleks [120, 121], dok dalja asocijacija sa RbAp46/48 i AEBP2 obezbeđuje povećanje afiniteta za histonski supstrat i omogućava interakcije sa velikim brojem drugih regulatornih molekula [122].

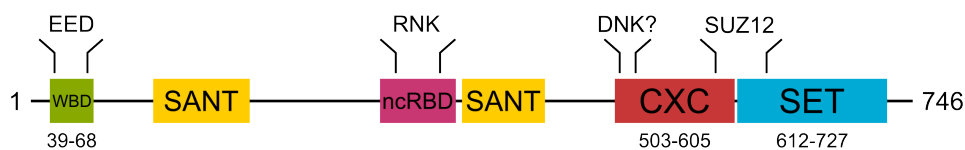
1.2.1 Struktura PRC2 i EZH2

Petokomponentni PRC2 kompleks ima molekulsku masu od približno 275 kDa, usled čega razjašnjavanje njegove strukture biofizičkim metodama predstavlja značajan izazov. Najdetajniji uvid u strukturnu organizaciju PRC2 dobijen je kombinovanjem elektronske mikroskopije, unakrsnog kovalentnog vezivanja lizina i fluorescentnog obeležavanja podjedinica, što je, zajedno sa dostupnim biohemijskim podacima, omogućilo rešavanje arhitekture kompleksa u rezoluciji ≈ 20 Å [123]. Kao što je ilustrovano na slici 1.3, PRC2 čine četiri reznja prečnika približno 55 Å koji su spojeni sa dva uska zgloba. Tri esencijalne komponente kompleksa (EZH2, EED, SUZ12) u bliskom su prostornom kontaktu što je verovatno od značaja za vezivanje za hromatin na način koji optimalno orijentiše supstratni lizin u odnosu na SET domen EZH2. AEBP2 značajno stabilizuje PRC2 i njegovo prisustvo bilo je neophodno za rešavanje strukture kompleksa [123].



Slika 1.3: Strukturna organizacija PRC2, sa pojedinačnim komponentama obojenim kao u legendi. ©Ciferri et al. / eLife / CC-BY 3.0

Kanonska sekvenca EZH2 ima dužinu od 746 aminokiselina, pri čemu je trodimenzionalna struktura EZH2 u velikoj meri nepoznata. Na osnovu sličnosti sekvenci i podataka o interakciji sa drugim molekulama pretpostavljeno je postojanje određenih evolutivno očuvanih domena u strukturi proteina (slika 1.4a) [118, 123]. N-terminalni kraj EZH2 (39–68) sadrži heliksnu strukturu za koju je kristalografski pokazano da stupa u interakciju sa WDR40 domenom EED (slika 1.4b). Ovaj deo EED može da veže metilovane lizine histona što, kao što je detaljnije opisano u sledećem odeljku, omogućava modulaciju katalitičke aktivnosti PRC2 kroz inte-

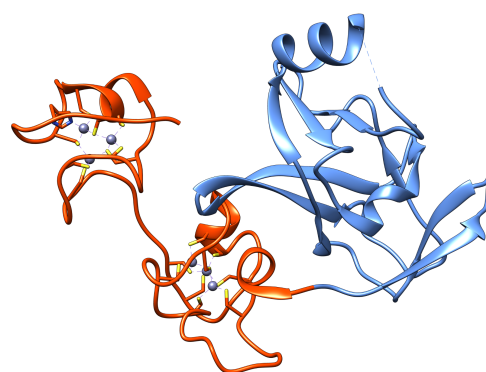


(a) Sekvenca EZH2 sa položajem domena čije je prisustvo dokazano ili pretpostavljeno. Iznad sekvence označene su druge molekule koje stupaju u interakciju sa odgovarajućim segmentom EZH2.



■ EZH2 (39-68) ■ EED (81-441)

(b) EED–EZH2 interakcija (PDB ID: 2QXV)



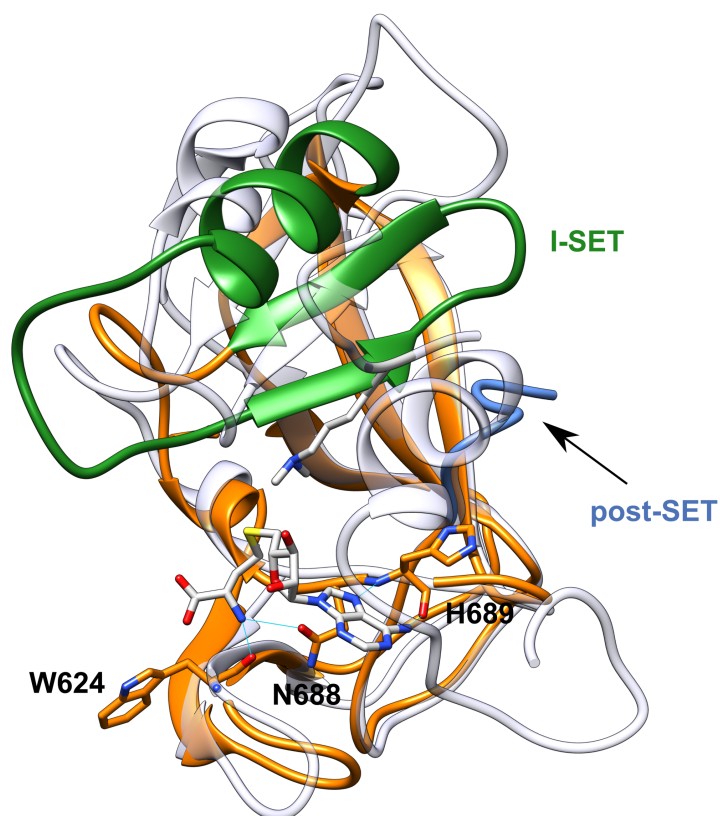
■ SET domen ■ CXC domen

(c) EZH2 SET i CXC domeni (PDB ID: 4MI0)

Slika 1.4: Ilustracija sekvence i kristalografski rešenih delova strukture EZH2.

grisanje signala okolnih PTM histona. Pretpostavlja se da središnji deo sekvence EZH2 sadrži dva evolutivno očuvana SANT domena, koji su ubikvitarno prisutni u hromatin-modifikujućim enzimima i omogućavaju interakciju sa slobodnim krajevima histona [124]. Verovatno je da ovaj deo proteina takođe sadrži i strukturne elemente koji omogućavaju da EZH2 stupi u interakciju sa nekodirajućim RNK. Katalitički SET domen zauzima C-terminalni segmenat sekvence a struktura ovog dela proteina je sve do 2013. godine bila nepoznata.

Tokom 2013. godine, dve grupe objavile su kristalografski rešene strukture SET i susednog, za PKMT atipično prisutnog CXC domena EZH2 (slika 1.4c) [125, 126]. Obe rešene strukture pokazale su da izolovani EZH2 poseduje tipičan SET domen sa motivom pseudočvora i naizgled katalitički-kompetentno aktivno mesto, ali sa nekoliko neuobičajenih karakteristika. Na slici 1.5 superponirane su kristalne strukture SET domena EZH2 i njemu slične PKMT: GLP u kompleksu sa histonskim supstratom i proizvodom kofaktora SAH. Kao što se može uočiti,



Slika 1.5: Prikaz kristalografski rešene strukture SET domena EZH2 (PDB ID: 4MI5) superponiran na strukturu GLP u kompleksu sa histonskim supstratom i SAH (PDB ID: 2RFI). Delovi strukture EZH2 obojeni su žutom, zelenom odnosno plavom bojom, dok je GLP prikazan semitransparentno, u sivoj boji.

vezivno mesto kofaktora dobro je očuvano između struktura tako da SAH, u karakterističnoj U-konformaciji, može da formira ključne vodonične veze sa bočnim lancem N688, kao i peptidnim skeletom H689 i W624. Međutim, vezivno mesto kofaktora samo je delimično kompletno jer se post-SET domen u strukturi EZH2 neočekivano pruža distalno od aktivnog centra. U obe objavljene strukture rešen je položaj samo početnih 5 aminokiselina post-SET domena što upućuje na to da je u izolovanom EZH2 ovaj deo strukture neuređen. U skladu sa nemogućnošću post-SET domena da učestvuje u stabilizaciji kofaktora, iako je SAM bio prisutan u rastvoru iz kojeg je protein kristalisan, ni u jednoj od struktura SAM nije kokristalisao sa EZH2. Antonysamy i saradnici dodatno su pokazali, primenom izotermalne kalorimetrije, da izolovani EZH2 ne vezuje kofaktor niti visokoafinitetni inhibitor **UNC1999** (slika 1.12). Dodatna specifičnost rešene strukture EZH2 ogleda se u neuobičajenom položaju I-SET domena. Dok kod većine PKMT β -ukosnica ovog dela struktura učestvuje u formiranju useka za vezivanje supstrata, položaj I-SET domena i vodonične veze sa post-SET domenom zatvaraju ovaj deo enzima na način

koji onemogućava vezivanje supstrata. Stoga se može zaključiti da izolovani EZH2 zauzima konformaciju koja predstavlja inaktivni ili auto-inhibirani oblik enzima, a da prelazak u katalitički kompetentnu formu verovatno zahteva vezivanje SUZ12 i EED što inicira pokrete I-SET domena, odnosno struktuiranje i adekvatno pozicioniranje post-SET domena [125, 126].

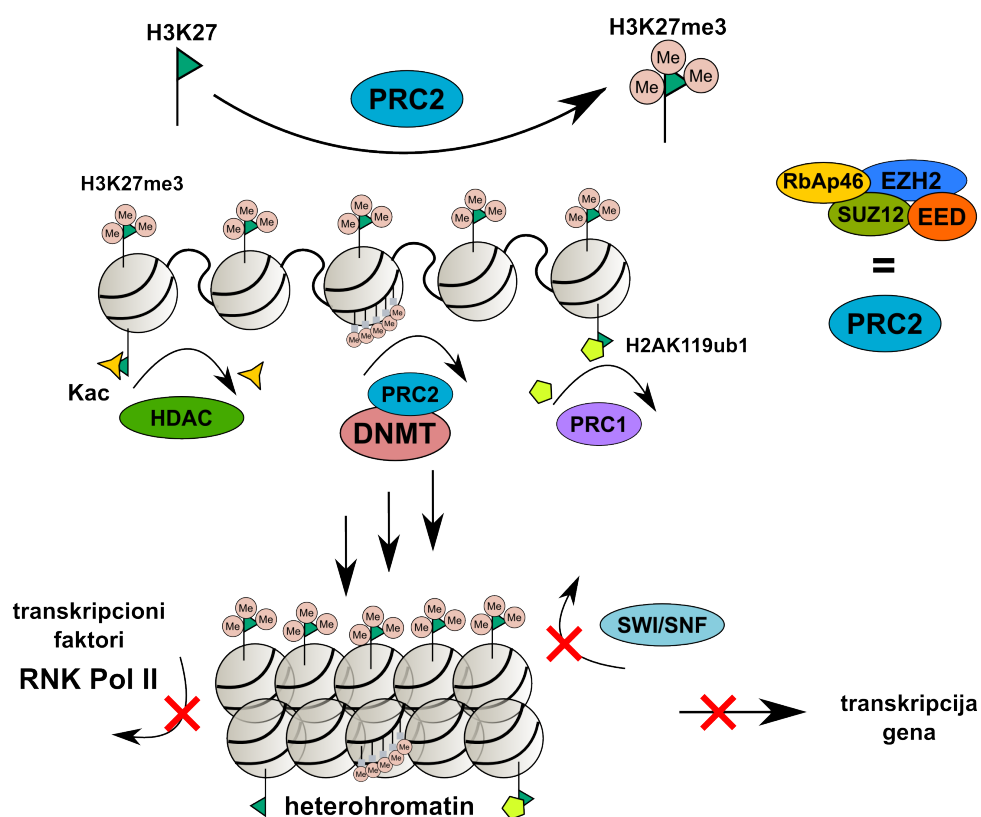
1.2.2 Biološka funkcija PRC2

PRC2 ima ključnu ulogu u inhibiciji transkripcije velikog broja gena, uključujući homeotske gena čija je pravovremena represija ključna u fazi embrionalnog i kasnijeg razvika. Ova funkcija PRC2 zavisna je od metiltransferazne aktivnosti EZH2, pa je genomski gubitak EZH2, EED ili SUZ12 letalan za embrion [118, 127].

Afinitet i efikasnost metiltransferazne aktivnosti EZH2 zavisi od stepena metilovanja H3K27 kao supstrata ali i od alosterne modulacije enzima od strane drugih komponenata PRC2. EZH2 kanonske sekvence najefikasnije vrši monometilovanje dok je efikasnost svakog narednog koraka niža, iako je enzim kompetentan za reakciju trimetilovanja. H3K27me3 je jedini značajan represivni signal koji prepoznaju drugi epigenetski regulatori, ali H3K27me2 predstavlja najzastupljeniji proizvod aktivnosti PRC2. H3K27me2 čini oko 50% ukupnog H3K27 u hromatinu, dok je zastupljenost H3K27me3 između 10 i 20% [128]. Visoka relativna zastupljenost H3K27me2 verovatno ima važnu ulogu u sprečavanju acetilovanja H3K27 i održavanja „spremnog” stanja histona iz kog se brzo može dobiti represivni H3K27me3 signal [129]. Metiltransferaznu aktivnost PRC2 mogu regulisati *cis* i *trans* efektima druge modifikacije histona posredstvom EED. WDR40 domen EED, preko kog se formira kompleks sa EZH2, može sa suprotne strane istovremeno vezati različite trimetilovane lizine histona. Eksperimentalno je pokazano da vezivanje H3K27me3 pospešuje, H1K26me3 inhibira a H3K9me3 ne ispoljava efekte na enzimsku aktivnost PRC2 [130]. Ovim mehanizmom, aktivnost kompleksa postaje zavisna od lokalnog „konteksta” drugih modifikacija hromatina.

Važno je spomenuti da ulogu PKMT u sklopu PRC2 može imati i EZH1. EZH1 je blizak homolog EZH2, i mada dva enzima karakterišu značajne razlike sekvenci u celini (identičnost od 11% i sličnost od 20%), njihovi katalitički SET i post-SET domeni značajno su sličniji (identičnost sekvenci od 92% i sličnost od 96%). PRC2/EZH1 zadržava specifičnost za H3K27 ali ovaj kompleks u biološkim uslovima ispoljava nisku metiltransferaznu aktivnosti i gubitak funkcionalnog EZH1

ne utiče značajnije na ćelijske nivoe H3K27me2/me3. Takođe, tkivna distribucija PRC2/EZH1 i PRC2/EZH2 pokazuje značajne razlike, pri čemu je nivo ekspresije EZH1 viši u ćelijama koja ne proliferišu, dok je EZH2 dominantan u ćelijama koje se brzo dele. Variranje katalitičke podjedinice stoga može omogućiti PRC2 kompleksu da obavlja specifične funkcije u različitim tkivima [131]. S druge strane, noviji eksperimenti na određenim tipovima diferenciranih ćelija (epidermisa i folikula dlake) ukazuju da gubitak funkcionalnog EZH2 ili EZH1 uzrokuje relativno blage defekte fenotipa, dok gubitak oba enzima ozbiljno ugrožava vijabilnost ćelija, sugerišući da su EZH2 i EZH1 potencijalno redundantni enzimi [132].



Slika 1.6: Šematski prikaz pretpostavljenih mehanizama kojima PRC2 i trimetilovanje H3K27 dovode do represije gena.

Način na koji uvođenje H3K27me3 od strane PRC2 dovodi do represije genske aktivnosti nedovoljno je poznat. Prema dugo zastupljenim stanovištima, postoje tri osnovna puta kojima PRC2 inhibira transkripciju (ilustrovano na slici 1.6) [118, 127, 133, 134]. Prvi mehanizam podrazumeva sekvencijalnu aktivnost PRC2 i PRC1. PRC1 je takođe multimerni PcG kompleks koji uključuje, između ostalih, proteine CBX i RING familije. CBX proteini sadrže hromodomen koji specifično prepoznaje H3K27me3, dok RING proteini poseduju aktivnost E3 ubi-

kvitin ligaze. U okviru ovog pretpostavljenog regulatornog puta, PRC2 omogućava vezivanje PRC1 u određenoj regiji hromatina, gde ovaj kompleks vrši monoubikvitinaciju H2AK119, iako ovaj korak nije nužno zavisian od funkcionalnosti PRC2 [127]. Zauzvrat, H2AK119ub1 sprečava elongaciju transkripata koje sintetiše RNK polimeraza II i time inhibira ekspresiju gena. Drugo, poznato je da PRC2 može na molekularnom nivou da stupa u interakciju sa HDAC, DNMT1, DNMT3A i DNMT3B, kao i da je u regijama hromatina sa metilovanom DNK značajno zastupljeniji H3K27me3. Stoga represiji gena posredovanoj PRC2 kompleksom verovatno doprinosi regrutovanje i aktivnost HDAC i DNMT [133, 135]. Konačno, EZH2 može da antagonizuje efekte SWI/SNF kompleksa, koji predstavlja ATP-zavisni remodeler hromatina [136]. Aktivnost PRC2 efektivno sprečava izmeštanje nukleozoma koje vrši SWI/SNF kompleks, usled čega hromatin ostaje u kompaktnom, transkripciono nedostupnom stanju. Molekularni mehanizam EZH2-SWI/SNF antagonizma nedovoljno je proučen. SWI/SNF kompleksi sadrže bromodomene koji vezuju acetilovane lizine histona H3 [38], pa je moguće da PRC2-posredovano regrutovanje HDAC za posledicu ima gubitak signala koji omogućavaju vezivanje SWI/SNF u datoj regiji hromatina. Novija istraživanja ukazala su i na druge potencijalne mehanizme inhibicije transkripcione aktivnosti koju inicira PRC2. Posebno je interesantno prepoznavanje činjenice da PcG kompleksi mogu postojati u različitim varijantama sastava podjedinica što omogućava specijalizaciju i specifičnost njihove represivne aktivnosti [129, 137].

U humanom genomu identifikovano je preko 1.000 gena koji potencijalno podležu represiji posredstvom epigenetskog mehanizma koji uključuje H3K27me3 odnosno PcG proteine [138]. Za daleko manji broj je, međutim, dokazano postojanje funkcionalne veze. Većina ciljnih gena PRC2 kodira proteine značajne u diferencijaciji ćelija. Prostorno i vremenski precizno regulisana aktivnost PRC2 prema ovim genima verovatno utiče na opredeljivanje pluripotentnih ćelija određenoj ćelijskoj lozi. Smatra se da ovim mehanizmom PRC2 ostvaruje važnu ulogu u organogenezi tokom embrionalnog razvića [139]. Istovremeno je aktivnost PRC2 u reverzibilnoj represiji pro-diferencijacijskih gena ključna za održavanje identiteta embrionalnih matičnih ćelija [134]. Sličnu funkciju PRC2 obavlja u okviru hematopoetskog sistema, gde omogućava očuvanje samo-obnavljajućih i multipotentnih karakteristika hematopoetskih matičnih ćelija [140]. Način na koji se PRC2 selektivno usmerava na ciljne gene manje je poznat, izuzev u okviru određenih, dobro proučenih regulatornih puteva. Inaktivacija X hromozoma je takav primer, gde je poznato da *dnkRNK* XIST učestvuje u ciljnom vezivanju PRC2. Nije izvesno da li regrutovanje posredovano *dnkRNK* molekulama predstavlja opšti mehanizam usmeravanja aktivnosti

PRC2. Faktori koji regulišu nivoe ekspresije komponenata PRC2 kompleksa takode su nedovoljno izučeni. Poznato je da *mi*RNK miR-101 deluje kao specifičan represor EZH2 [141, 142].

1.2.3 EZH2 kao onkogen

Onkogene efekte EZH2 može ispoljiti usled povećane ekspresije ili somatskih mutacija katalitičkog domena koje menjaju supstratnu specifičnost enzima (slika 1.7).

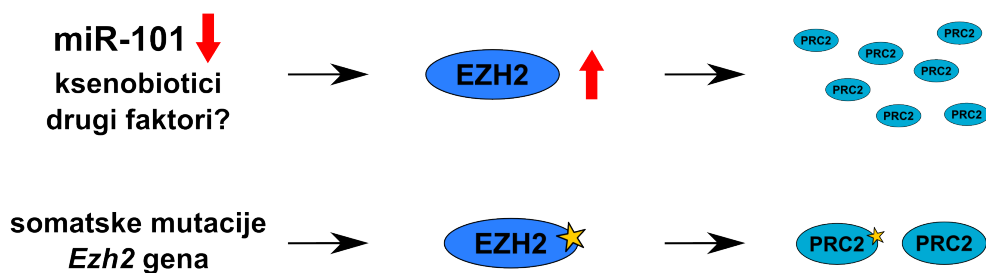
Kao što je navedeno u prethodnom odeljku, PRC2/EZH2 reguliše aktivnost velikog broja gena od kojih čak oko 500 predstavljaju tumor-supresore [143]. Prekomerna aktivnost PRC2/EZH2 može uzrokovati represiju gena koji stimulišu diferencijaciju i inhibiraju proliferaciju. Na ovaj način, izmenjena funkcija PRC2/EZH2 može omogućiti ćeliji sticanje karakteristika sličnih onima koje poseduju matične ćelije, u smislu neograničenog replikativnog potencijala i samo-obnavljanja [143, 144]. Prekomerna ekspresija EZH2 do danas je utvrđena u velikom broju malignih tumora različitih organa (tabela 1.2), pri čemu je ova karakteristika gotovo uvek marker agresivne bolesti, izraženog metastatskog potencijala tumora i loše prognoze [145]. Uzrok povećane ekspresije manje je jasan: više studija pokazalo je da delecije ili mutacije gena za miR-101 mogu biti odgovorne za ovu pojavu [141, 142, 146, 147]. Druge studije na životinjskim modelima pokazale su da je *in utero* izlaganje dietilstilbestrolu i bisfenolu-A povezano sa povećanom ekspresijom EZH2 u tkivu dojke odraslog potomstva, što može predstavljati mehanizam razvoja raka dojke [148].

Tabela 1.2: Pregled malignih tumora kod kojih je dokazana prekomerna ekspresija EZH2

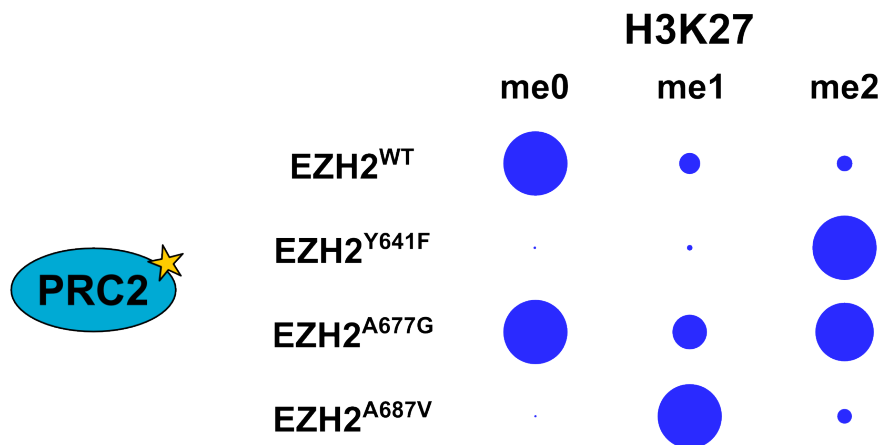
Tip maligniteta	Reference
Rak dojke	[149–153]
Rak prostate	[154–158]
Glioblastom	[159, 160]
Rak mokraćne bešike	[161, 162]
Melanom	[163]
Rak jajnika	[164]
Rak pluća	[165, 166]

Da povišeni nivoi EZH2 igraju aktivnu ulogu u progresiji određenih maligniteta pokazano je u više studija. Nekoliko autora ustanovilo je da je eksperimen-

mentalnim indukovanjem povećane ekspresije EZH2 moguće inicirati nekontrolisan rast i invazivnost ćelijskih kultura [149, 167]. Prateći ove zaključke, drugi autori su pokazali da farmakološka inhibicija ili interferencija sa biosintezom EZH2 selektivno dovede do apoptoze, ili smanjenja invazivnosti i proliferacije različitih tipova tumorskih ćelija koje prekomerno ekspimiraju ovaj enzim [168–171]. Ove studije ujedno su ukazale i na terapijski potencijal farmakološke inihicije EZH2, iako su u vreme njihovog izvođenja dostupni bili samo neselektivni inhibitori (odjeljak 1.3.1). Imajući u vidu reverzibilnost metilovanja histona, inhibicija EZH2 potencijalno može da dovede do reaktivacije tumor-supresora i time doprinese iniciranju programa ćelijske smrti maligno transformisanih ćelija [172].



(a) Prekomerna ekspresija EZH2 i somatske mutacije njegovog katalitičkog domena kao dva puta koji mogu dovesti do hipermetilovanja H3K27.



(b) Efikasnost PRC2 kompleksa u katalizovanju različitih stupnjeva metilovanja H3K27. Površine krugova (desno) odgovaraju relativnoj efikasnosti (izraženoj kao k_{cat}/K_m) kompleksa koji sadrži kanonsku (WT) i različite mutirane forme EZH2.

Slika 1.7: Aberacije EZH2 koje dovode do hipermetilovanja H3K27.

Pored solidnih tumora u kojima je EZH2 prekomerno ekspimiran, u određenim hematološkim malignitetima ustanovljeno je prisustvo somatskih mutacija katalitičkog domena EZH2 za koje se veruje da predstavljaju ključan faktor inicijacije

i razvoja bolesti. Ove mutacije menjaju supstratnu specifičnost enzima (slika 1.7b) usled čega, i u odsustvu promena u nivou njegove ekspresije, dolazi do značajnog porasta nivoa H3K27. Morin i saradnici [173] ustanovili su da je u oko 22% uzoraka B–limfocita poreklom od pacijenata obolelih od difuznog krupnoćelijskog limfoma B–ćelija (DLBCL) i oko 7% uzoraka folikularnog limfoma prisutna Y641F ili Y641N mutacija SET domena EZH2. Kao što je opisano u odeljku 1.1.2, ovaj tirozin očuvan kod većine PKMT sa SET domenom ima ulogu „čuvara kapije” i određuje specifičnost proizvoda reakcije koju dati enzim katalizuje. Phe/Tyr izmena obično dovodi do povećane aktivnosti enzima u katalizovanju višeg stepena metilovanja. U skladu sa ovim pravilom, Y641F i Y641N okarakterisane su kao mutacije koje dovode do izraženog porasta efikasnosti EZH2 u vršenju trimetilovanja, uz gubitak efikasnosti enzima na nemetilovanim supstratima (H3K27me0) [74]. Prisustvo Y641F/N/H/S/C mutacija generalno dovodi do značajnog porasta intracelularnih nivoa H3K27me3. Jedinstvenost mutacije ovog položaja ogleda se u činjenici da onkogenost mutirane forme enzima zavisi od kooperacije sa kanonskim oblikom EZH2 [174]: dok mutirani oblik enzima efikasno dodaje treću metil grupu, neizmenjen enzim je neophodan za inicijalne korake čiji je proizvod H3K27me2. Pored ove, kod pacijenata obolelih od limfoma, identifikovane su i druge mutacije katalitičkog domena EZH2. A687V mutacija ustanovljena je u 1-2% limfoma a dovodi do gubitka efikasnosti enzima prema H3K27me0/1 kao supstratu, višestrukom porastu efikasnosti dimetilovanja i očuvanoj efikasnosti trimetilovanja [175]. Konačno, sa A677G mutacijom, enzim zadržava sposobnost katalizovanja svih koraka u formiranju H3K27me3, pri čemu je u ćelijama sa EZH2^{A677G} formom enzima prisutno izrazito uvećanje nivoa ovog represivnog signala [176]. Iz navedenog može se zaključiti da mutirani EZH2 predstavlja ključan onkogen u razvoju specifičnih hematoloških maligniteta, kao što je DLBCL [135, 140].

Iako je onkogeni potencijal EZH2 temeljno dokumentovan, važno je imati u vidu da postoje primeri hematoloških maligniteta u kojima je ekspresija EZH2 suprimirana ili su prisutne inaktivirajuće mutacije enzima. Ovi nalazi ukazuju na činjenicu da u zavisnosti od konteksta promena u ćeliji, EZH2 može predstavljati i tumor–supresor [135]. Takođe, kao što je već navedeno, u fiziološkim uslovima EZH2 igra značajnu ulogu u očuvanju identiteta hematopoetskih matičnih ćelija. Stoga dvojaka uloga EZH2 zahteva pažljivo razmatranje okolnosti u kojima je inhibicija enzima opravdana i potencijalno terapijski korisna [140]. Dodatno, postoji za sada ograničen broj dokaza da EZH2 može ispoljiti onkogenu aktivnosti nezavisno od PRC2 [177], pa inhibicija katalitičke aktivnosti enzima nije nužno adekvatan pristup u terapiji svih stanja u kojima je EZH2 prekomerno ekspimiran.

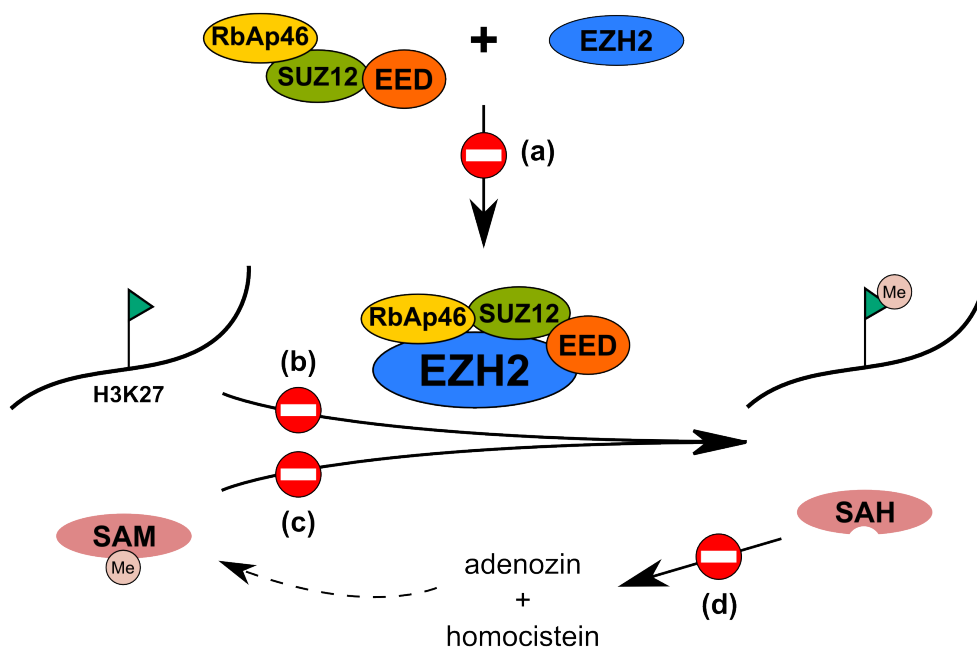
1.3 Inhibitori EZH2

U literaturi je opisano više pristupa za inhibiciju katalitičke aktivnosti PRC2. Kao što je ilustrovano na slici 1.8, prvi mogući pristup jeste inhibicija asocijacije EZH2 sa drugim podjedinicama kompleksa neophodnim za njegovu katalitičku aktivnost (mesto inhibicije označeno sa **a**). Ovaj pristup prate svi izazovi koji se odnose na razvoj inhibitora protein–protein interakcija, i do sada je opisan relativno mali broj kako peptidnih tako i nepeptidnih inhibitora PRC2 koji deluju ovih mehanizmom (opisani u odeljku 1.3.3). Za razliku od vezivnih površina preko kojih proteinske podjedinice stupaju u interakciju, koje su prostrane ali često nedovoljno konkavne, EZH2 sadrži dva dobro definisana vezivna mesta koja se potencijalno mogu iskoristiti u razvoju selektivnih i efikasnih inhibitora i to: mesto za vezivanje supstrata (mesto inhibicije označeno sa **b**) i mesto za vezivanje kofaktora (označeno sa **c**). Kofaktor–kompetitivni inhibitori bili su predmet istraživanja sprovedenog u okviru ove teze i poznata jedinjenja koja deluju ovih mehanizmom inhibicije detaljno su opisana u odeljku 1.3.2, dok su izazovi u dizajnu novih struktura ovog tipa diskutovani u odeljku 1.3.4. Jedini primer inhibitora EZH2 koji deluje kompetitivno sa supstratom opisan je u odeljku 1.3.3. Konačno, neselektivan pristup inhibiciji većine PKMT može biti inhibicija enzima uključenih u metabolizam metionina (mesto inhibicije označeno sa **d**), i ovaj pristup je kraće opisan u odeljku 1.3.1.

1.3.1 Neselektivni inhibitori

Pre otkrića selektivnih inhibitora EZH2, u izučavanju biologije PRC2 korišćena su dominantno dva neselektivna inhibitora: 3-deaza-neplanocin (DZNep) i sinefungin (slika 1.9).

3-deaza-neplanocin (DZNep) je nukleozidni inhibitor SAH–hidrolaze [178], jedinog enzima kod sisara sposobnog za hidrolizu *S*-adenozil-L-homocisteina do homocisteina i adenzina. Kako je SAH inhibitor svih enzima koji koriste SAM kao kofaktor, SAH–hidrolaza je posredni regulator većine bioloških procesa koji uključuju reakcije transmetilacije [179]. Akumulacija SAH, do koje dolazi primenom DZNep, uzrokuje neselektivnu inhibiciju svih enzima uključenih u metilovanje histona. Ovaj efekat je izražen na nivou većine represivnih modifikacija histona i dovodi do reaktivacije velikog broja gena koji su konstitutivno inaktivirani u somatskim ćelijama [180]. Pojedini autori pokazali su da primena DZNep-a dovodi do degradacije svih



Slika 1.8: Šematski prikaz mogućih pristupa inhibiciji katalitičke aktivnosti PRC2: (a) inhibicija formiranja kompleksa; (b) inhibicija vezivanja histonskog supstrata za EZH2; (c) inhibicija vezivanja kofaktora za EZH2; (d) inhibicija SAH hidrolaze.



Slika 1.9: Strukture DZNep-a i sinefungina, neselektivnih inhibitora PKMT

kompenenata PRC2, smanjuje nivoe H3K27me3 i ispoljava selektivan citotoksičan efekat na tumorskim ćelijama [168].

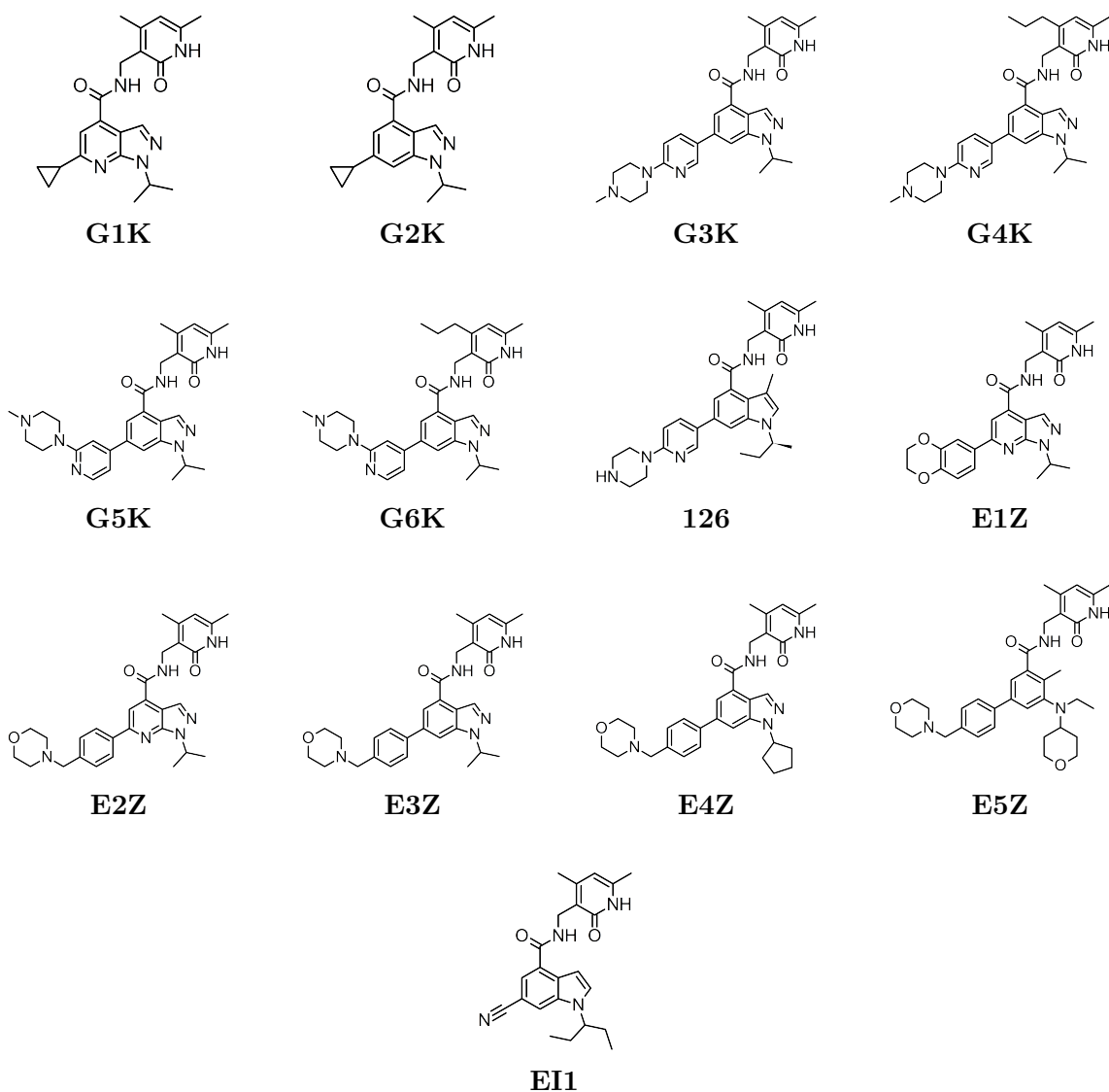
Sinefungin je prvobitno izolovan iz kultura bakterije *Streptomyces griseolus*. Ovo jedinjenje sadrži aminometilensku grupu na mestu metilovane sulfonijum grupe kofaktora SAM i ova strukturna sličnost ga čini kompetitivnim inhibitorom svih SAM-zavisnih metiltransferaza. Efekti sinefungina na inhibiciju metiltransferaza i njegov uticaj na ekspresiju gena u ćelijama sisara prepoznati su pre gotovo 4 decenije [181], od kada se sinefungin koristi u istraživanjima biologije hromatina. Sinefungin i DZNep odigrali su važnu ulogu u prepoznavanju terapijskog potencijala inhibicije više PKMT, uključujući EZH2, ali ih njihova neselektivnost čini

prevashodno jedinjenjima od istorijskog i eksperimentalnog značaja. Koncept kompetitivne inhibicije PKMT nukleozidnim analogima kao što je sinefungin, međutim, i danas je polje aktivnog istraživanja usmerenog ka racionalnom dizajniranju jedinjenja sa unapređenom selektivnošću i inhibitornom aktivnošću prema specifičnim enzimima ove familije [182, 183].

1.3.2 Kofaktor-kompetitivni inhibitori

Tokom 2012. godine, u nekoliko međusobno nezavisnih publikacija prezentovane su piridonske strukture za koje je utvrđeno da veoma selektivno inhibiraju katalitičku aktivnost EZH2. Autori iz kompanije GlaxoSmithKline početkom te godine objavili su studiju validacije biohemijske metode za skrining jedinjenja na inhibiciju EZH2 i prezentovali strukture relativno slabo aktivnih piridonskih derivata koji su identifikovani u korporativnoj kolekciji [184]. Nešto kasnije, ista grupa objavila je primere 6 piridonskih derivata [185] (slika 1.10, **G1K–G6K**) dobijenih kroz nekoliko ciklusa strukturne optimizacije prvobitno identifikovanih jedinjenja, kao i strukturu inhibitora **126** [186], koji predstavlja drugo jedinjenje u klasi sa subnanomolarnom inhibitornom aktivnošću. Isti put pronalaska strukturno veoma bliskih inhibitora objavile su grupe iz kompanija Epizyme (slika 1.10, **E1Z–E5Z**) [187, 188] i Novartis (slika 1.10, **E11**) [189]. Tako je **E1Z** identifikovan eksperimentalnim skriningom kolekcije od preko 175.000 jedinjenja, čijom su strukturnom optimizacijom metodama medicinske hemije dobijena jedinjenja **E2Z–E4Z** značajno unapređene efikasnosti. Činjenica da su sve tri grupe nezavisno identifikovale piridonske strukture ukazuje ili na postojanje ograničene raznovrsnosti postojećih kolekcija ili upućuje na činjenicu da je piridonska struktura od izuzetnog značaja za inhibiciju EZH2.

Biohemijska ispitivanja aktivnosti ovih jedinjenja na rekombinantnom petokomponentnom PRC2 kompleksu u prisustvu rastućih koncentracija histonskog supstrata i kofaktora SAM dala su preliminarni uvid u njihov mehanizam dejstva. Dobijeni rezultati pokazali su da je vezivanje piridonskih inhibitora kompetitivno sa kofaktorom i nekompetitivno sa histonskim peptidima, što sa velikom verovatnoćom ukazuje da se ove strukture vezuju za kofaktor–vezujuće mesto enzima. Dostupni inhibitori svih drugih PKMT sa SET domenom stupaju u interakciju sa alosternim mestima ili vezivnim mestom supstrata, čime EZH2 inhibitori predstavljaju prvu grupu inhibitora PKMT (sa izuzetkom DOT1L) koji deluju kofaktor–kompetitivnim mehanizmom inhibicije [190].



Slika 1.10: Strukture 13 piridonskih inhibitora EZH2 koje su objavljene u literaturi tokom 2012. godine.

Sa aspekta selektivnosti, utvrđeno je da su prezentovane piridonske strukture više od 1.000 puta selektivnije za EZH2 u odnosu na niz drugih PKMT i PRMT, sa izuzetkom EZH1, naspram kog je selektivnost u opsegu između 25 i 142 puta. Efikasnost piridonskih inhibitora ispitivana je i na većem broju ćelijskih linija tumora. U skladu sa snažnom i selektivnom inhibicijom PRC2 katalitičke aktivnosti (inhibitorne konstante sumirane su u tabeli 1.3), pokazano je da ova jedinjenja na dozno- i vremenski-zavisnan način dovode do globalne redukcije nivoa H3K27me3 u ćelijama, bez značajnijih efekata na druge PTM histona. Inhibitor **E5Z** ispitivan je na ćelijskim linijama limfoma sa kanonskom formom EZH2 (EZH2^{WT}: OCI-LY19), kao i na dve linije sa mutiranim formama enzima: WSU-DLCL2 (EZH2^{Y641F}) i Pfeiffer (EZH2^{A677G}). Iako je efikasnost **E5Z** u smanjenju nivoa H3K27me3 bila upore-

diva na svim ćelijskim linijama, pri testiranim koncentracijama **E5Z** nije uticao na proliferaciju OCI-LY19 ćelija, dok je nakon perioda latencije od 4 dana doveo do usporavanja proliferacije i apoptoze WSU-DLCL2 i Pfeiffer ćelija. Utvrđeno je da primena **E5Z** na ove mutirane ćelije limfoma dovodi do reaktivacije velikog broja ciljnih gena PRC2. Sličan biološki profil dobijen je testiranjem inhibitora **126** na širokom panelu ćelijskih linija limfoma DLBCL tipa. Najveću osetljivost na dejstvo **126** ispoljile su Pfeiffer, WSU-DLCL2 i KARPAS-422 (sadrže EZH2^{Y641N}) ćelijske linije, dok su efekti na proliferaciju OCI-LY19 ćelija bili zanemarljivi. Uzimajući u obzir činjenicu da su K_i^{app} vrednosti na izolovanim PRC2 kompleksima sa EZH2^{WT} odnosno EZH2^{Y641F}, EZH2^{A677G} i EZH2^{Y641N} relativno slične (0,57–1 nM), razlike u efektima na vijabilnost ukazuju na činjenicu da u mutiranim ćelijama EZH2 igra presudnu onkogenu ulogu od koje je njihovo preživljavanje zavisno. Vremenska latencija u apoptotskim efektima, primećena i prilikom ispitivanja **126**, upućuje na zaključak da nakon inhibicije PRC2 reaktivacija gena zavisi od aktivnosti protein lizin demetilaza. Važan zaključak ispitivanja **E5Z** i **126** jeste i da u ćelijama neosestljivim na efekte ovih inhibitora, tretman uzrokuje minimalne efekte na ekspresiju gena, što je veoma značajno sa aspekta bezbednosti primene.

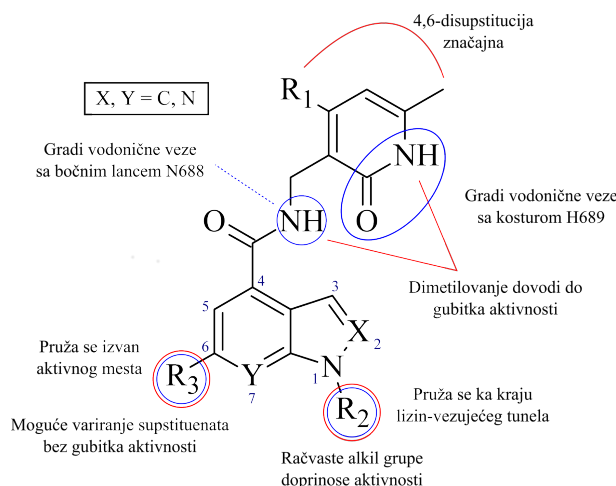
Tabela 1.3: Konstante inhibicije i fizičko-hemijski parametri 13 piridonskih inhibitora EZH2

Inhibitor	K_i^{app} (nM) ^a	Molekulska masa	ClogP ^b	Referenca
G1K	149	379,46	1,42	[185]
G2K	74	378,47	2,14	[185]
G3K	7,9	513,63	2,19	[185]
G4K	0,6	541,69	3,08	[185]
G5K	14	513,63	2,19	[185]
G6K	1,2	541,69	3,08	[185]
126	0,57	526,67	3,87	[186]
E1Z	310	473,52	2,06	[187]
E2Z	180	514,62	2,27	[187]
E3Z	80	513,63	2,73	[187]
E4Z	24	539,67	3,31	[187]
E5Z	2,5	572,74	3,45	[188]
EI1	13	390,48	3,14	[189]

^a Prikazane K_i^{app} vrednosti preuzete su iz citiranih referenci a izračunate su uklapanjem inhibitornih koncentracija u Cheng-Prussif-ljev model kompetitivne inhibicije. Sve inhibitorne koncentracije uniformno su određene u eksperimentima sa petokomponentnim PRC2 kompleksom (EZH2, SUZ12, EED, RBBP4 i AEBP2), u prisustvu visoke koncentracije kofaktora SAM od 7,5 μ M, pri pH 8,0 i na sobnoj temperaturi; ^b Vrednosti izračunate pomoću programa MarvinSketch 15.

Pokušaji da se eksperimentalnim putem reši struktura EZH2 u kompleksu sa nekim od inhibitora do danas nisu bili uspešni, pa je strukturalna osnova efikasne inhibicije EZH2 većinom nerazjašnjena. Dostupni podaci o konstantama inhibicije (tabela 1.3), kao i preliminarne studije molekuskog modeliranja pružaju osnovu

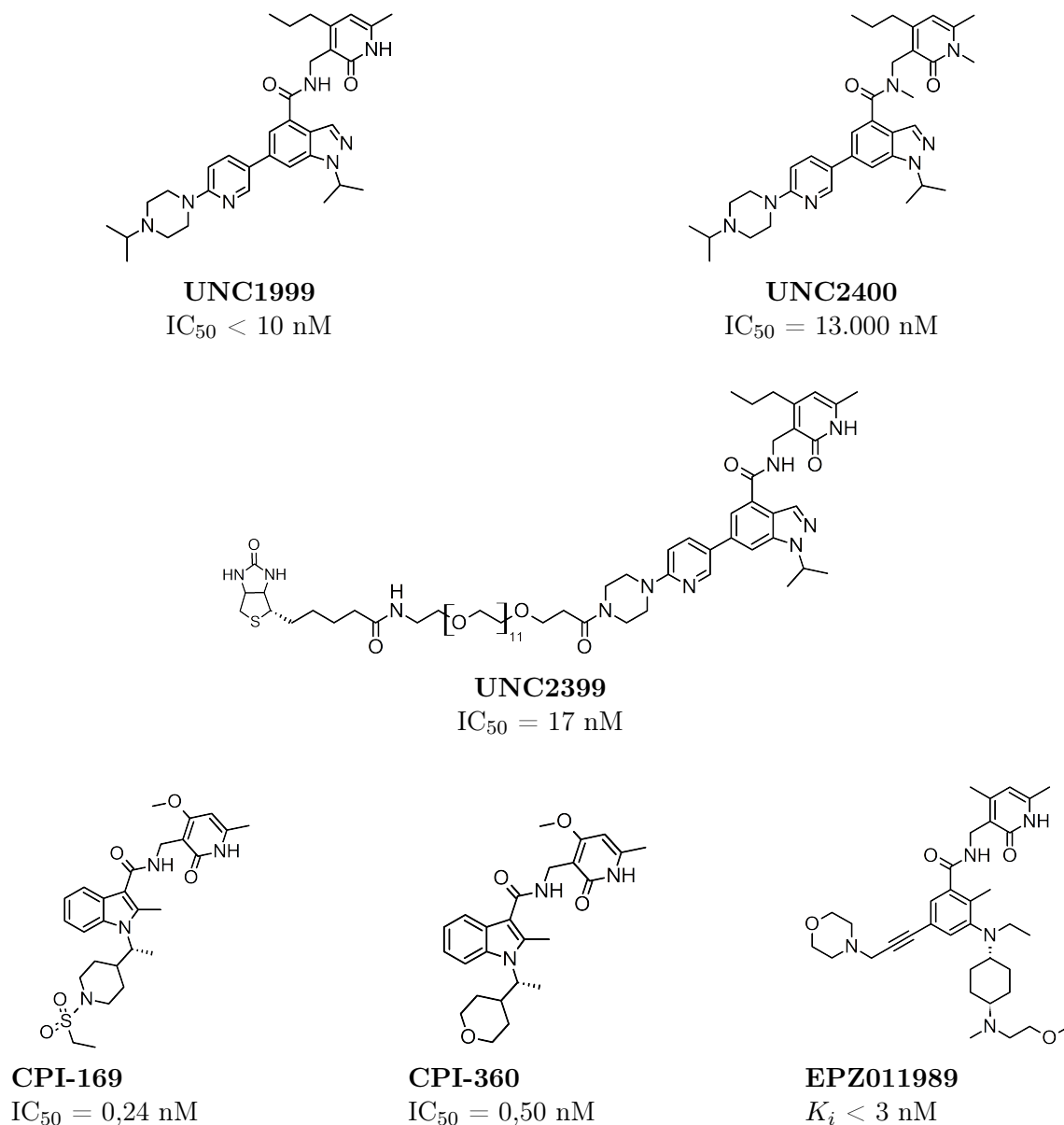
za racionalizaciju odnosa strukture piridonskih derivata i njihove aktivnosti [190]. Razmatranjem objavljenih struktura inhibitora, može se zaključiti da je većinom variran tip centralnog bicikličnog sistema, kao i supstitenti njegovog položaja 1 i 6 (slika 1.11), dok je piridonsko jezgro veoma malo modifikovano i smatra se da je ono ključno za aktivnost. Iako su prvobitne strukture većinom bile bazirane na azaindazolnom i indazolnom jezgru (**G1K**, **E1Z**), eliminacija azota iz bicikličnog sistema ne utiče značajnije na aktivnost, pa i indolne strukture ispoljavaju značajan, često i izraženiji inhibitorni efekat. Kao što će biti diskutovano u narednim paragrafima, bicikličan sistem nije uslov za dejstvo jer i strukture kod kojih je izvršeno otvaranje jednog prstena ispoljavaju snažnu inhibitornu aktivnost. Prisustvo račvastih i cikličnih alkil radikala u položaju 1 povećava aktivnost, pri čemu ona raste srazmerno voluminoznosti hidrofobne grupe. U položaj 6 se mogu uvesti različiti supstituenti, ali se čini da prisustvo π -sistema u ovom položaju značajno doprinosi aktivnosti.



Slika 1.11: Šematski prikaz dostupnih podataka o odnosu strukture i aktivnosti piridonskih inhibitora (crvene oznake) i pretpostavljenog načina interakcije (plave oznake) ovih jedinjenja sa EZH2

Na osnovu homologije kofaktor-vezujućih mesta unutar PKMT familije, više autora je pretpostavilo način orijentacije inhibitora u kofaktor-vezivnom mestu EZH2 [185, 187, 191]. Kao što je ilustrovano na slici 1.11, za piridonsko jezgro i amidski linker se pretpostavlja da uspostavljaju vodonične veze sa H689 očuvanog motiva pseudočvora, dok se alkil supstituent položaja 1 centralnog prstena pruža ka osnovi lizin-vezujućeg tunela, čime se supstituent položaja 6 usmerava izvan vezivnog mesta. Konze i saradnici [191] su delom eksperimentalno potvrdili ovu hipotezu. Autori su sintetisali **UNC2400**, kao N,N-dimetilovani derivat **UNC1999** (slika 1.12), i ustanovili da ima preko 1.000 puta manju inhibitornu aktivnost prema EZH2. Iako gubitak aktivnosti može biti posledica više faktora, ovaj nalaz potkrepljuje

značaj prisustva donora vodonične veze u piridonskom jezgru. Autori su takode sintetisali i **UNC2399**, kao biotinilovani derivat **UNC1999**, za koga je utvrđeno očuvano vezivanje za EZH2, što potvrđuje da se supstituent položaja 6 verovatno pruža izvan vezivnog mesta kofaktora i da je delimično izložen rastvaraču.



Slika 1.12: Strukture dodatnih piridonskih inhibitora EZH2.

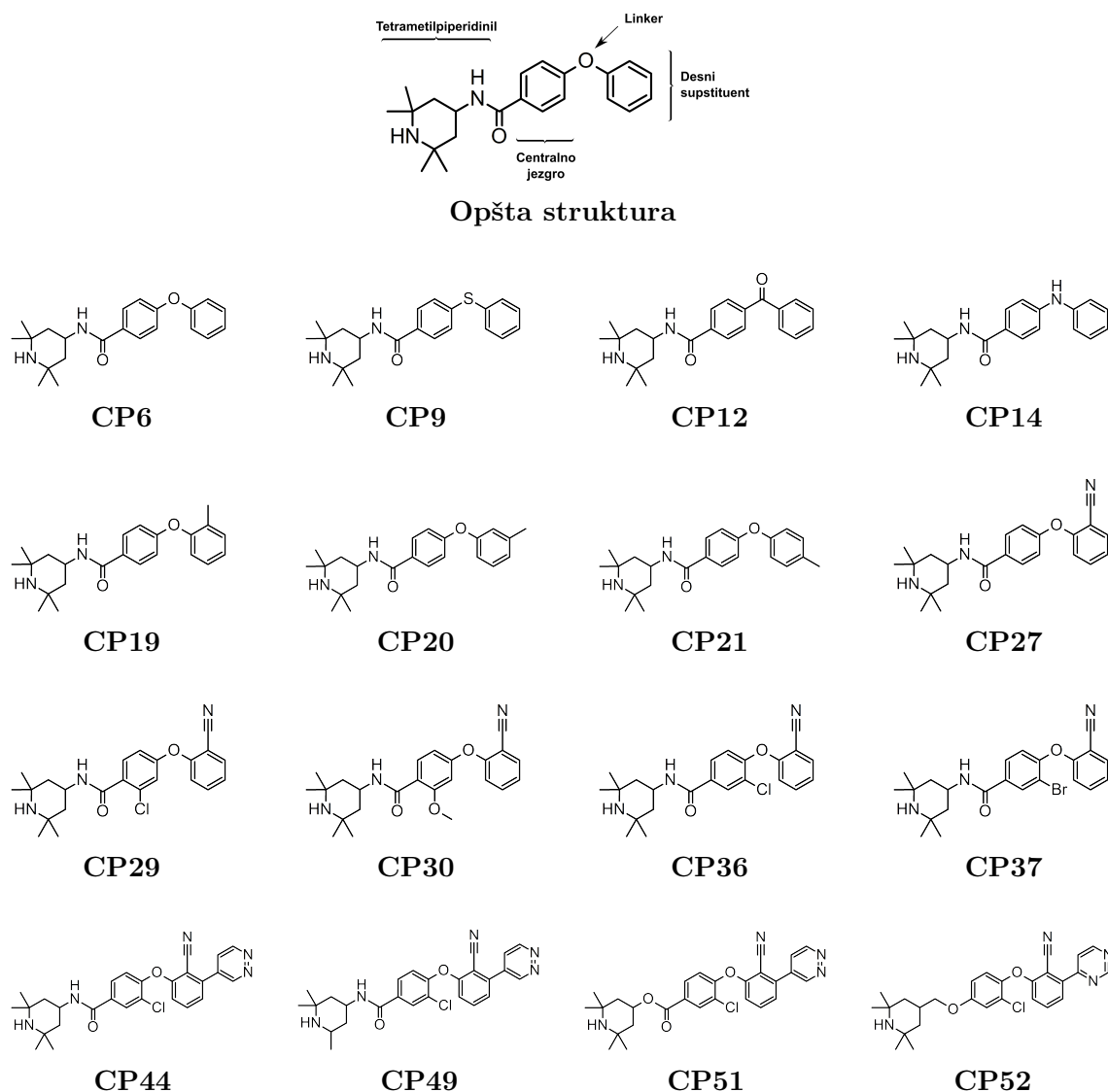
Prvobitno otkrivena piridonska jedinjenja ispoljavaju izražene inhibitorne efekte na izolovanom enzimu kao i na velikom broju tumorskih ćelija, ali su se njihove farmakokinetičke osobine pokazale kao relativno nepovoljne za opsežnije testiranje u *in vivo* uslovima. Prema rezultatima ispitivanja **126** na miševima sa KARPAS-422 tumorskim kalemima, ovo jedinjenje se brzo eliminiše iz organizma usled čega

postizanje inhibicije rasta tumora zahteva režim doziranja od 50 mg/kg jednom dnevno, dok je u cilju postizanja eradikacije tumora bila neophodna doza od 150 mg/kg dnevno. Autori su naveli da ove doze nisu bile toksične. U cilju unapređenja farmakokinetičkih karakteristika, Epizyme je od svoje prvobitne serije jedinjenja (**E1Z–E4Z**) razvio **E5Z**, otvaranjem centralnog bicikličnog sistema i uvođenjem tetrahidropiranskog prstena u strukturu. Prema biohemijskim ispitivanjima, **E5Z** zadržava isti mehanizam dejstva i selektivnost ranijih jedinjenja, ali poseduje izraženije inhibitorne efekte na samom enzimu, oralno je bioraspoloživ i ima unapređene farmakokinetičke osobine [188]. Jedinjenje **E5Z** (EPZ-6438, E-7438, tazemetostat) trenutno je jedan od tri inhibitora EZH2 koji se nalaze u kliničkim ispitivanjima. Pretklinička ispitivanja na miševima i pacovima pokazala su dobru distribuciju u tumorska tkiva ali i relativno brz klirens kod pacova. Ova *in vivo* ispitivanja ukazala su i na razlike u osetljivosti različitih tipova limfoma na dejstvo **E5Z**, koja nisu utvrđena u *in vitro* studijama. Kod kalema tipa WSU-DLCL2 limfoma, i u najvišoj ispitivanoj dozi (160 mg/kg, dva puta dnevno, 28 dana) **E5Z** dovodi samo do delimične inhibicije rasta tumora, dok u modelu limfoma Pfeiffer ćelija, doza od 114 mg/kg, jednom dnevno, dovodi do potpune eradikacije tumora tokom jednomesečnog doziranja [192].

Nekoliko novijih piridonskih struktura (slika 1.12) takođe je razvijeno sa ciljem da se unaprede farmakokinetičke osobine ove klase jedinjenja, uz zadržavanje selektivnosti i aktivnosti. Campbell i saradnici [193] pokušali su zamenu pirana u strukturi **E5Z** *trans* N,N-dimetilcikloheksilaminom, čime je dobijeno jedinjenje očuvane aktivnosti. Prisustvo dodatnog baznog centra, međutim, uslovalo je izostanak aktivnosti na ćelijama, usled čega je baznost novouvedenog tercijarnog amina modifikovana dodatkom metoksietil grupe. U cilju smanjenja molekulske mase rezultujućeg jedinjenja, benzenov prsten vezan za morfolin zamenjen je acetilenskom grupom čime je dobijeno visoko-aktivno i oralno bioraspoloživo jedinjenje **EPZ011989**. Bradley i saradnici [194] opisali su biološke efekte dva nova predstavnika grupe piridonskih inhibitora, razvijena optimizacijom strukture nisko-aktivnog piridonskog jedinjenja identifikovanog eksperimentalnim skriningom korporativne kolekcije. Jedinjenja **CPI-169** i **CPI-360** inhibiraju EZH2 u niskim nanomolarnim koncentracijama, ali tačne strukturne razlike između ovih piridona i prethodno objavljenih inhibitora koje doprinose njihovoj aktivnosti nisu u celosti izvesne.

Prvu grupu ne-piridonskih inhibitora EZH2 prezentovala je grupa iz kompanije Constellation Pharmaceuticals. Ova grupa izvršila je eksperimentalni skrining kolekcije od 150.000 jedinjenja i identifikovala 288 jedinjenja koja u visokim koncen-

tracijama (preko 20 μM) inhibiraju katalitičku aktivnost petokomponentnog PRC2 kompleksa [195]. Daljim biohemijskim ispitivanjima, eliminisana su promiskuitetna jedinjenja i ona sa kovalentnim mehanizmom dejstva, a zatim je za preostala utvrđivan mehanizam inhibicije PRC2. Za dalju optimizaciju izabrano je tetrametilpiperidinil benzamidno jedinjenje, za koje je potvrđen kofaktor–kompetitivan mehanizam dejstva. Autori su zatim sintetisali niz derivata u cilju ispitivanja odnosa strukture i aktivnosti (SAR) u okviru ovog hemotipa (slika 1.13) [196].



Slika 1.13: Strukture tetrametilpiperidinskih inhibitora EZH2

Nalazi SAR studije tetrametilpiperidina [196] ukratko se mogu sumirati na sledeći način. Inicijalno variranje strukture linkera (videti opštu strukturu ilustrovanu na slici 1.13 i tabelu 1.4) pokazalo je da zamena bisfenil etra (**CP6**) tioetrom (**CP9**) ne dovodi do gubitka aktivnosti. Oksidacija **CP9** do sulfona ili sulfoksida,

Tabela 1.4: Inhibitorne koncentracije i fizičko-hemijski parametri odabranih tetra-metilpiperidinskih inhibitora EZH2

Inhibitor	IC ₅₀ (μM) ^a	Molekulska masa	ClogP ^b	Referenca
CP6	51	352,47	3,59	[196]
CP9	33	368,54	4,27	[196]
CP12	57	364,48	3,54	[196]
CP14	>80	351,48	3,53	[196]
CP19	48	366,50	4,10	[196]
CP20	61	366,50	4,10	[196]
CP21	>80	366,50	4,10	[196]
CP27	3,4	377,48	3,44	[196]
CP29	37	411,92	4,05	[196]
CP30	>80	407,50	3,28	[196]
CP36	0,45	411,92	4,05	[196]
CP37	0,32	456,37	4,21	[196]
CP44	0,032	490,00	3,50	[196]
CP49	0,871	475,97	3,22	[196]
CP51	0,172	490,98	4,42	[196]
CP52	0,189	477,00	5,38	[196]

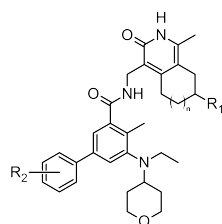
^a Prikazane IC₅₀ vrednosti preuzete su iz citirane reference. Inhibitorne koncentracije uniformno su određene u eksperimentima sa petokomponentnim PRC2 kompleksom (EZH2, SUZ12, EED, RBBP4 i AEBP2), u prisustvu aktivatornog H3K27me3 peptida, biotinirovanih nukleozomalnih supstrata i kofaktora SAM u koncentraciji od 200 nM, pri pH 8,5 i na sobnoj temperaturi; ^b Vrednosti izračunate pomoću programa MarvinSketch 15.

međutim, rezultuje potpunim gubitkom aktivnosti, dok zamena etarskog linkera ketonom (**CP12**) ne utiče značajnije na inhibitornu koncentraciju (u odnosu na **CP6**). Uvođenje donora vodonične veze u položaj linkera, kao kod jedinjenja **CP14**, ima za posledicu gubitak aktivnost. Na osnovu variranja ovog dela strukture, zaključeno je da specifična zakrivljena konformacija difeniletra, uz prisustvo akceptora vodonične veze u ovom položaju, predstavlja optimalan izbor linkera. Bez daljih modifikacija ovog dela strukture, zatim je istražen uticaj supstituenata aromatičnog jezgra desne strane. Uvođenje supstituenata u *m*- i *p*- položaje (**CP20**, **CP21**) izrazito nepovoljno utiče na aktivnost, dok *o*- položaj toleriše supstituente različite veličine, od metil (**CP19**) do *tert*-butil grupe, ali je optimalno poboljšanje aktivnosti moguće postići uvođenjem cijano grupe (**CP27**). Polazeći od strukture **CP27**, u sledećem koraku varirani su supstituenti centralnog aromatičnog jezgra. Supstitucija *o*- položaja u odnosu na centralni amid dovela je do značajnih gubitaka aktivnosti (**CP29**, **CP30**), dok je za derivate dobijene uvođenjem niza različitih grupa (halogeni, metil, ciklopropil, cijano) u *m*- položaj utvrđena poboljšana aktivnost. U skladu sa hipotezom o značaju specifične konformacije difeniletra, prisustvo grupa u ovom položaju verovatno stabilizuje konformaciju koja odstupa od koplanarnosti što dovodi do višetrnog povećanja aktivnosti (**CP36**, **CP37**). Daljim modifikacijama supstituenta desne strane, dobijeno je jedinjenje **CP44** koje je odabrano za detaljnu biološku

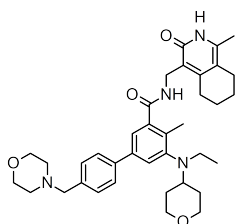
karakterizaciju. Utvrđeno je da **CP44** poseduje nisku pasivnu permeabilnost kao i da u Caco-2 ćelijama podleže aktivnom efliksu, što ograničava njegovu efikasnost na ćelijama. Pokušaji da se smanji baznost tetrametilpiperidinskog azota, u cilju poboljšanja pasivne permeabilnosti, nisu bili uspešni i ukazali su na značaj prisustva baznog centra u ovom položaju. Interesantno je da i veoma male promene ovog dela strukture (npr. uklanjanje jedne metil grupe, **CP49**) dovode do značajnih gubitaka aktivnosti. S druge strane, centralni amid nije uslov za dejstvo i može biti zamenjen kako estrom (**CP51**) tako i etoksi etrom (**CP52**). Poboljšana permeabilnost **CP51** i **CP52**, međutim, nije uticala na poboljšanje biološkog efekta inhibitora na testiranim KARPAS-422 limfomskim ćelijama.

Nedavno, u literaturi je objavljena serija heksahidroizohinolinskih inhibitora EZH2 [197]. Ove strukture (slika 1.14) veoma su slične jedinjenju **E5Z**, sa ključnom modifikacijom dobijenom kondenzacijom dodatnog alifatičnog prstena sa piridonskim jezgrom preko položaja 4 i 5. Inhibitorna aktivnost ovih jedinjenja na izolovanom PRC2 kompleksu, međutim, nije veća u odnosu na **E5Z** (tabela 1.5). Ipak, ova studija je interesantna jer predstavlja prvi značajni pokušaj variranja strukture piridonskog prstena. SAR sintetisanih derivata ukazuje na to da EZH2 toleriše prisustvo bicikličnih sistema u položaju piridona; kondenzacija sa cikloheksanom (**ZS1**) daje najaktivnija jedinjenja, dok su strukture sa sedmočlanim i osmočlanim prstenovima u ovom položaju nekoliko desetina puta manje aktivne. Uvođenje dodatnih alkil radikala u strukturu **ZS1** pokazalo je da supstitucija položaja 7 doprinosi aktivnosti (**ZS2**), ali da je efekat najizraženiji sa metil grupom jer voluminozniji supstituenti u ovom položaju (R1 = Et, *t*-Bu, di-Me) uslovljavaju manje izraženo povećanje aktivnosti. Nasuprot supstitucije položaja 7, uvođenje metil grupe u susedni položaj 6 (**ZS3**) dovodi do manjeg pada aktivnosti u odnosu na nesupstituisani **ZS1**. Variranje supstituenata centralnog aromatičnog jezgra (**ZS1**, **ZS4–ZS6**) uslovlja slične promene u aktivnosti kao i kod prethodno poznatih piridonskih struktura, što ukazuje na to da je interakcija heksahidroizohinolinskih jedinjenja sa kofaktor-vezivnim mestom EZH2 verovatno u velikoj meri slična interakciji piridona.

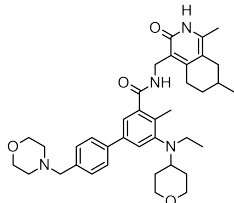
Za dodatne dve grupe jedinjenja utvrđena je sposobnost kofaktor-kompetitivne inhibicije EZH2, ali je selektivnost ovih inhibitora nepotpuna ili nedovoljno okarakterisana. Istražujući aktivnost tanšinona, prirodnih proizvoda izolovanih iz korena biljke *Salvia miltiorrhiza* (Danshen) koja se koristi u kineskoj tradicionalnoj medicini, Woo i saradnici [199] pokazali su da nekoliko predstavnika ove klase jedinjenja inhibira EZH2. Za tanšindiol B (slika 1.15) i njegov enantiomer



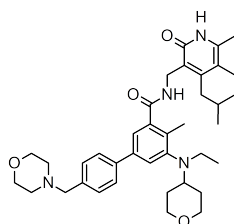
Opšta struktura



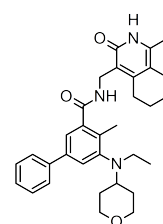
ZS1



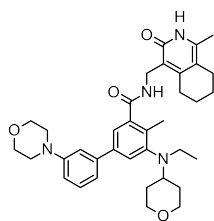
ZS2



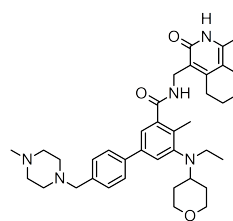
ZS3



ZS4



ZS5



ZS6

Slika 1.14: Strukture heksahidroizohinolinskih inhibitora EZH2

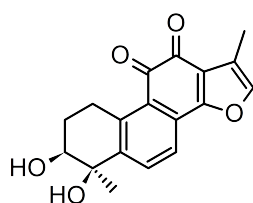
tanšindiol C utvrđena je submikromolarna inhibitorna aktivnost. Dalja biohemijska karakterizacija ukazala je da se ova jedinjenja vezuju za EZH2 kompetitivno sa kofaktorom, ali i da na njihovo vezivanje utiču rastuće koncentracije histonskog supstrata, što ukazuje na mešoviti mehanizam inhibicije.

Kung i saradnici [200] pokušali su da iskoriste afinitet nukleozidnih analoga ka inhibiciji većeg broja PKMT kao polaznu tačku u dizajnu derivata SAH sa povećanim afinitetom i selektivnošću za EZH2. Autori su sintetisali preko 20 analoga, od kojih je najpovoljniji profil aktivnosti i selektivnosti pokazalo jedinjenje 27 (**SA27**, slika 1.15). Pored submikromolarne aktivnosti, **SA27** ima značajno unapređenu selektivnost u poređenju sa većinom nukleozidnih jedinjenja, pri čemu je ovaj inhibitor oko 2 puta selektivniji za EZH2 naspram EZH1, oko 35 puta u odnosu na SUV39H1, i oko 70 puta u odnosu na MLL. Sintetisani inhibitori, međutim, verovatno usled neadekvatnih fizičko-hemijskih karakteristika, nisu pokazali aktivnost na ćelijama.

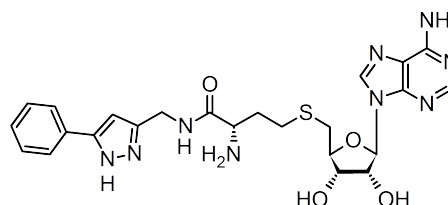
Tabela 1.5: Inhibitorne koncentracije i fizičko-hemijski parametri odabranih heksahidroizohinolinskih inhibitora EZH2

Inhibitor	IC ₅₀ (nM) ^a	Molekulska masa	ClogP ^b	Referenca
ZS1	14	611,81	5,16	[197]
ZS2	0,66	625,84	5,45	[197]
ZS3	20	625,84	5,45	[197]
ZS4	75	512,68	5,44	[197]
ZS5	10	597,79	5,33	[197]
ZS6	7,3	624,86	5,22	[197]

^a Prikazane IC₅₀ vrednosti preuzete su iz citirane reference. Inhibitorne koncentracije određivane su AlphaLISA metodom [198]; ^b Vrednosti izračunate pomoću programa MarvinSketch 15.



tanšindiol B
IC₅₀ = 0,52 μM



SA27
IC₅₀ = 0,27 μM

Slika 1.15: Strukture tanšindiola i sintetskog analoga SAH, kao inhibitora EZH2

1.3.3 Drugi pristupi inhibiciji

U literaturi postoji veći broj primera malih organskih molekula koje selektivno inhibiraju određene PKMT mehanizmom koji podrazumeva vezivanje za enzim koje je kompetitivno sa histonskim supstratima [201–204]. Analize dostupnih struktura PKMT ukazuju da je mesto za vezivanje supstrata, prevashodno kanal u koji se umeće lizin i okruženje na ulazu u ovaj kanal, pogodno za dizajn selektivnih inhibitora ove klase enzima [205]. Prvi, i za sada jedini, pokušaj da se izvrši inhibicija EZH2 korišćenjem histon–kompetitivnih inhibitora, opisali su Smadbeck i saradnici [206]. Primenom računarskog *de novo* dizajniranja peptida i korišćenjem homologog modela EZH2 SET domena, autori su predložili strukture 17 peptidnih inhibitora EZH2 od kojih je 10 eksperimentalno ispitano. Najaktivniji od ispitanih peptida (SQ037) imao je IC₅₀ vrednost od 13,57 μM i pokazao je očekivani mehanizam dejstva, kompetitivan sa histonskim supstratima. Za SQ037 takođe je potvrđena i sposobnost redukcije H3K27me3 u ćelijama, uz minimalne efekte na druge modifikacije histona. Relativno niska aktivnost ovog peptida ograničava njegovu korisnost, ali uspeh u dizajnu selektivnog, supstrat–kompetitivnog inhibitora predstavlja važan pokazatelj mogućnosti za dalje istraživanje ovog pristupa inhibiciji EZH2. Na efikasnost dobijenih peptida verovatno je uticala i činjenica da model

strukture SET domena koji su autori uspostavili u velikoj meri odstupa od kristalne strukture koja je kasnije postala dostupna.

Alternativna strategija za inhibiciju enzimske aktivnosti PRC2 jeste dizajn inhibitora protein–protein interakcija koji poseduju sposobnost da poremete formiranje kompleksa i na taj način inhibiraju katalitičku aktivnost EZH2. Kim i saradnici [207] su potvrdili potencijal ovog pristupa pripremom nekoliko peptida koji inhibiraju interakciju EED i EZH2 podjedinica PRC2 kompleksa. Polazeći od sekvence EZH2 heliksa za koji je poznato da stupa u interakciju sa vezivnom površinom EED, autori su sintetisali nekoliko peptida kod kojih je svaka četvrta aminokiselina bila zamenjena sintetskom aminokiselinom sa olefinskim bočnim lancem. Olefinskom metatezom zatim su dobijeni stabilizovani heliksi. Njihova biohemijska i biološka karakterizacija pokazale su da ovi peptidi u visokim nanomolarnim koncentracijama inhibiraju EED–EZH2 interakciju, sprečavaju formiranje katalitički kompetentnog PRC2 i ispoljavaju antitumorsku efikasnost na MLL–AF9 ćelijskim linijama leukemije, smanjujući nivoe H3K27me3. Mogućnost postizanja sličnih efekata primenom male organske molekule nedavno su potvrdili Kong i saradnici [208]. Korišćenjem dostupne kristalne strukture EED–EZH2 kompleksa (slika 1.4b) [209], autori su virtuelnim skriningom kolekcije komercijalno dostupnih jedinjenja identifikovali astemizol kao potencijalni inhibitor ove protein–protein interakcije. Eksperimentalnom karakterizacijom potvrđeno je da astemizol inhibira EED–EZH2 interakciju sa IC_{50} od 93,8 μ M i ispoljava citotoksične efekte na EZH2–zavisnim GCB–DLBCL ćelijama različitih podtipova. Time je ova studija pružila dodatnu osnovu za istraživanje mogućnosti inhibicije PRC2 remećenjem formiranja samog kompleksa. Pored terapijskog potencijala, proučavanje razlika u biološkim efektima različitih pristupa inhibiciji PRC2 može imati važnu ulogu u produbljivanju razumevanja biologije i patobiologije ovog epigenetskog regulatora.

1.3.4 Izazovi u dizajnu novih inhibitora

Iako su do sada otkriveni kofaktor–kompetitivni inhibitori dokazali izvodljivost i terapijski potencijal selektivne inhibicije EZH2, otkrivene strukture su relativno ograničenog hemijskog diverziteta, neke ispoljavaju brz klirens *in vivo* dok druge poseduju nisku ćelijsku aktivnost usled loše permeabilnosti. Dodatno, novija istraživanja (videti odeljak 1.4.3) pokazala su da više predstavnika ove grupe jedinjenja predstavljaju supstrate mebranskih transportera koji utiču na njihovu raspodelu u organizmu i potencijalno mogu dovesti do uspostavljanja rezistencije tumorskih

ćelija na ove inhibitore. Stoga postoji potreba za otkrićem novih struktura koje poseduju potencijal selektivne inhibicije EZH2, a istovremeno imaju poboljšane biofarmaceutske karakteristike.

Jedna od osnovnih prepreka u ostvarenju ovog cilja jeste nedostatak razumevanja strukturne osnove aktivnosti i selektivnosti poznatih inhibitora EZH2 na molekulskom nivou. Kao što je opisano u odeljku 1.2.1, dostupne kristalne strukture EZH2 su nekompletne i nije u celosti jasno kako inhibitori stupaju u interakciju sa kofaktor-vezivnim mestom enzima. Campagna-Slater i saradnici [210] ispitivali su sličnost kofaktor-vezivnih mesta različitih PKMT čije su kristalne strukture bile dostupne 2011. godine i pokazali da postoji visok stepen očuvanosti ovog dela strukture SET domena, pri čemu su kontakti koje SAM odnosno SAH ostvaruju sa enzimom gotovo identični u svim rešenim strukturama. Iako ova činjenica ukazuje na potencijalne prepreke u dizajnu selektivnih inhibitora, autori su, međutim, takođe zaključili da raznovrsnost aminokiselina koje okružuju vezivno mesto, a posebno delovi post-SET domena, pružaju osnovu za dizajn takvih selektivnih jedinjenja. Kako su inhibitori EZH2 za sada jedina klasa selektivnih, kofaktor-kompetitivnih inhibitora bilo koje PKMT sa SET domenom, razumevanje strukturne osnove njihove aktivnosti može imati značaja za eksploatisanje terapijskog potencijala cele klase enzima. Šta više, kako su za onkogene efekte EZH2 prevashodno odgovorne izmenjene forme enzima koje sadrže mutacije SET domena, poznavanje strukture EZH2 u kompleksu sa inhibitorima može predstavljati polaznu tačku u dizajnu inhibitora selektivnih za mutirane forme enzima.

1.4 Rezistencija na antitumorske lekove

Inovativni pristupi ciljanoj farmakološkoj terapiji kancera imaju značajan potencijal da unaprede standard lečenja i prognozu obolelih. Međutim, jedan od ključnih faktora koji ograničava kliničku efikasnost hemoterapije jeste razvoj rezistencije tumora na primenjeni lek [211]. Primer imatiniba, istaknutog na početku ovog poglavlja kao jednog od prvih uspešnih primera ciljane terapije kancera, ilustruje i kako racionalno dizajnirani lekovi mogu postati potpuno neefikasni usled promena u kancerskim ćelijama. Nakon samo godinu dana od njegovog puštanja na tržište, ustanovljen je niz mutacija Bcr-Abl kinaze koje dovode do gubitka afiniteta imatiniba za enzim i kliničkog neuspeha terapije ovim lekom [212].

Stečena rezistencija na antitumorske lekove može nastati brojnim meha-

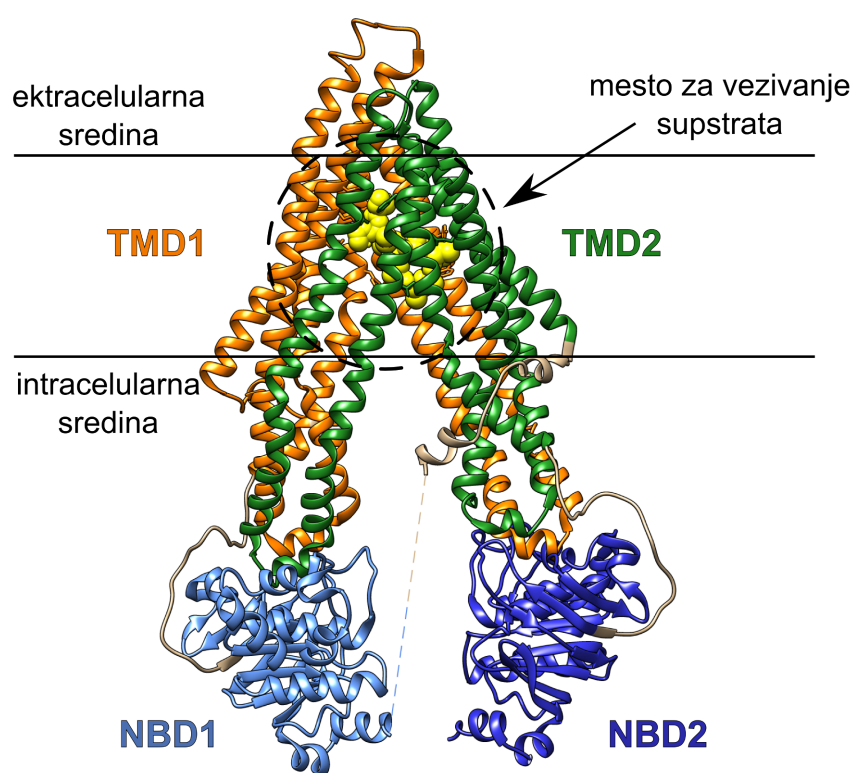
nizmima. Nestabilnost genoma dovodi do pojave čestih mutacija i uspostavljanja heterogenosti populacije ćelija tumora, pri čemu istovremena primena antitumorskih agenasa utiče na klonalnu selekciju onih kancerskih ćelija koje imaju rezistentan fenotip [213, 214]. Na molekularnom nivou, rezistencija može biti posledica [215]: (i) mutacija ciljnog mesta dejstva koje utiču na sposobnost leka da se za isto veže; (ii) aktivacije paralelnih signalnih puteva koji čine inhibirani put redundantnim za preživljavanje kancerskih ćelija; (iii) promena u nishodnim ili ushodnim regulatorima koji kompenzuju efekte inhibicije ciljnog mesta dejstva. Dizajniranje lekova čiji su efekti *a priori* robusni na ovakve promene nije uvek moguće, jer navedeni putevi razvoja rezistencije suštinski menjaju zavisnost kancera od ciljnog mesta dejstva koje se farmakološki modulira. Drugi vidovi razvoja rezistencije, međutim, ne menjaju pojedinačne signalne puteve već dovode do promena koje rezultuju rezistencijom na veći broj antitumorskih lekova (engl. *multiple drug resistance*, MDR fenotip) [216]. Ovaj fenomen može biti posledica promena u mikrookruženju solidnih tumora koje uslovljavaju neadekvatnu raspodelu leka u tumorsko tkivo [217], može nastati usled izmenjenog ćelijskog metabolizma, povećane degradacije ili sekvestracije leka, kao i aktiviranja anti-apoptotskih signala [214, 218, 219]. Često, međutim, ključnu ulogu u razvoju MDR fenotipa ima prekomerna ekspresija membranskih transportera [220, 221].

Većina ćelija eksprimira na svojoj površini transmembranske proteine koji poseduju sposobnost aktivnog transporta supstrata, nasuprot gradijenta koncentracije, iz ćelije u ekstracelularnu sredinu korišćenjem energije oslobođene hidrolizom ATP-a. Ovi proteini, označeni kao ABC transporteri (od engleskog *ATP-binding cassette*) mogu postati prekomerno ekspimirani u kancerskim ćelijama, što omogućava aktivan efluks antitumorskih lekova iz ćelija i dovodi do izostanka njihovog efekta na nivou ciljnog mesta dejstva. Uloga ABC transportera u razvoju MDR fenotipa najopsežnije je istražena u kontekstu prekomerne ekspresije i funkcije tri predstavnika ove grupe [222]; to su: P-glikoprotein (Pgp/ABCB1), protein rezistencije raka dojke (BCRP/ABCG2) i protein multiple rezistencije na lekove (MRP1/ABCC1).

1.4.1 Struktura i funkcija Pgp-a

P-glikoprotein (Pgp) je otkriven 1976. godine kao površinski glikoprotein koji utiče na permeaciju lekova kroz ćelijske membrane [223]. Nivoi ekspresije Pgp-a fiziološki su konstitutivno visoki u određenim tkivima i organima: na apikalnoj površini enterocita, kanalikularnim površinama hepatocita kao i kapilarnom endotelu

krvno-moždane i krvno-testisne barijere. Njegova specifična tkivna distribucija sugerira da Pgp, zajedno sa citohromalnim enzimima, predstavlja važnu komponentu odbrane tkiva od ksenobiotika. U prilog ovoj činjenici ide i izražena supstratna polispecifičnost transportera koja se ogleda u sposobnosti Pgp-a da transportuje veliki broj strukturno raznovrsnih organskih jedinjenja. Na ovaj način, Pgp može da utiče na oralnu bioraspoloživost lekova, može ograničiti njihovu raspodelu u određena tkiva i posredovati u interakcijama između lekova [224–226]. Na primeru više tipova tumorskih ćelija, pokazano je da prekomerna ekspresija Pgp-a dobro koreliše sa postojanjem MDR fenotipa [227–229], što ukazuje da efluks lekova posredovan ovim transporterom igra ključnu ulogu u razvoju rezistencije na antitumorske lekove.



Slika 1.16: Kristalografski rešena struktura mišijeg Pgp-a (PDB ID: 4M2T), sa pretpostavljenim položajem ćelijske membrane i različitim bojama istaknutim domenima transportera.

Tokom 2009. godine, kristalografski je rešena struktura prvog ABC transportera eukariota i to mišijeg Pgp-a [230], čija je sekvenca veoma slična sekvenci humanog Pgp-a (identičnost od 87%). Ovaj protein mase približno 170 kDa odlikuje visoka sličnost C- i N-terminalne polovine koje se međusobno prepliću formirajući dva transmembranska domena (TMD) od po 6 transmembranskih heliksa (slika 1.16). Sa citoplazmatske strane locirana su dva nukleotid-vezujuća domena

(NBD) koji učestvuju u prevođenju energije oslobođene hidrolizom ATP-a u pokrete proteina kojima se unutrašnja šupljina otvara ka ekstracelularnoj sredini prilikom transporta supstrata. Unutrašnja šupljina proteina, koja vezuje supstrate, ima zapreminu od približno 6.000 \AA^3 . U ranijim studijama, baziranim na biohemijskim podacima, pretpostavljeno je da u okviru centralne šupljine postoji između 2 i 7 različitih vezivnih mesta za supstrate koja stupaju u složene alosterne interakcije [231–234]. Na osnovu dostupnih strukturnih informacija, izvesnije je da Pgp ne poseduje prostorno razdvojena, jasno definisana vezivna mesta, već dinamične i prilagodljive vezivne regije u okviru kojih postoje oblasti za koje pojedinačni supstrati pokazuju izraženiji afinitet [235, 236]. Struktura Pgp-a je time pružila dodatnu osnovu za razumevanje njegove supstratne polispecifičnosti.

1.4.2 Struktura i funkcija BCRP-a

Protein rezistencije raka dojke (BCRP) otkriven je više od dve decenije nakon Pgp-a, i to kao ATP-zavisan transporter koji je odgovoran za rezistenciju MCF-7 ćelija raka dojke na mitoksantron, doksorubicin i daunorubicin, po čemu je i dobio naziv [237]. Najviši nivo ekspresije BCRP dostiže u placentalom tkivu, gde se čini da igra važnu ulogu u zaštiti fetusa od ksenobiotika. Visoki nivoi BCRP-a takođe se mogu naći u tkivu dojke, testisa, u hematopoetskim matičnim ćelijama kao i na krvno-moždanoj barijeri, što ukazuje na sličnu zaštitnu ulogu ovog transportera. Obzirom na relativno skoro otkriće, uloga BCRP-a u rezistenciji tumorskih ćelija na antitumorske lekove nije tako opsežno proučena kao što je slučaj sa Pgp-om. Trenutno dostupni dokazi ukazuju na potencijalno važnu ulogu ovog transportera u rezistenciji nekoliko oblika hematoloških maligniteta na hemoterapiju, uključujući DLBCL [238].

Na nivou primarne strukture, BCRP je protein mase približno 72 kDa čija sekvencija sadrži 665 aminokiselina. Na osnovu sličnosti sekvenci, može se pretpostaviti da BCRP sadrži jedan N-terminalni NBD i jedan C-terminalni TMD, odnosno samo polovinu strukture uobičajene za ABC transportere. Stoga je verovatno da funkcionalni transporter nastaje homodimerizacijom dva proteinska monomera, mada postoje dokazi da BCRP u membranama gradi tetramerne strukture koje čine četiri homodimera. Iako je malo poznato o strukturi BCRP-a, za veći broj lekova raznovrsnih hemijskih struktura je ustanovljeno da predstavljaju supstrate ovog transportera što ukazuje na njegovu supstratnu polispecifičnost i potencijalni uticaj na farmakokinetiku lekova [239].

1.4.3 Značaj predviđanja interakcije lekova sa ABC transporterima

Uzimajući u obzir njihov uticaj na farmakokinetiku lekova i značaj u razvoju MDR fenotipa, mogućnost interakcije novodizajniranih lekova sa membranskim transporterima neophodno je ispitivati u ranim fazama otkrića i razvoja [240]. Eksperimentalnu karakterizaciju interakcije leka–kandidata sa ABC transporterima preporučuju različite regulatorne agencije i savetodavni organi [241], i u tu svrhu su razvijene pouzdane *in vitro* eksperimentalne metode [242, 243]. Međutim, u fazi razvoja, strukturne modifikacije leka–kandidata materijalno su i vremenski veoma skupe, usled čega bi mogućnost interakcije sa ABC transporterima trebalo procenjivati značajno ranije, tokom otkrića same vodeće molekule [244]. Računarski prediktivni modeli za predviđanje transporta posredovanog ABC transporterima imaju prednost u tome što ne zahtevaju dostupnost sintetisane supstance, te se mogu inkorporirati u savremene protokole dizajniranja lekova pomoću računara [245]. Njihova predviđanja mogu se iskoristiti u eliminaciji jednog dela kolekcije virtuelnih jedinjenja i mogu poslužiti u izboru manjeg broja jedinjenja za prioritarno biološko testiranje. Time računarski modeli za predviđanje interakcije sa membranskim transporterima, ukoliko je njihova pouzdanost validirana, mogu značajno doprineti racionalizaciji procesa otkrića kroz izbor jedinjenja sa željenim karakteristikama [246–248].

U otkriću novih potencijalnih antitumorskih jedinjenja, predviđanje mogućnosti da novodizajnirana struktura bude supstrat za Pgp i/ili BCRP ima dvojak značaj. Iako svaki od ova dva transportera pojedinačno može značajno uticati na apsorpciju, distribuciju i efikasnost antitumorskih lekova, niz novijih istraživanja ukazalo je na postojanje značajnog preklapanja supstratne specifičnosti Pgp-a i BCRP-a, kao i na potencijalno sinergističke efekte koje zajedno mogu ostvariti [249–253]. Stoga je rano prepoznavanje mogućnosti da je novo jedinjenje dualni supstrat oba transportera važno u donošenju odluka o njegovom isključivanju iz daljeg razvoja ili pak planiranju strukturnih modifikacija koje bi izmenile profil interakcije sa Pgp-om i BCRP-om. Dodatan značaj pravovremenom predviđanju interakcije sa ABC transporterima daje činjenica da svi do sada dizajnirani inhibitori Pgp-a nisu pokazali uspeh u kliničkim ispitivanjima [254, 255]. Za ova jedinjenja se očekivalo da mogu da povrate efikasnost standardnih antineoplastika u malignitetima MDR fenotipa kroz inhibiciju transportera odgovornog za aktivan efluks leka. Iako postoje različita istraživanja alternativnih pristupa prevazilaženju rezistencije, u dizajnu no-

vih antitumorskih lekova treba težiti minimizaciji rizika da lek uopšte bude supstrat ovih transportera [256]. U kontekstu otkrića novih inhibitora EZH2, ovo može biti posebno važno imajući u vidu da su studije na pojedinim ćelijskim linijama pokazale da prekomerna ekspresija EZH2 može doprineti razvoju MDR fenotipa kroz prekomernu ekspresiju Pgp-a [257]. Nedavno objavljeni rezultati ispitivanja transporta inhibitora **126**, **E4Z** i **E5Z** pokazali su da sva tri jedinjenja predstavljaju supstrate za Pgp i BCRP, te da ovi transporteri značajno mogu ograničiti njihovu raspodelu u moždano tkivo [258].

1.5 Računarske metode u dizajniranju novih molekula

Računarske metode nalaze primenu u svim fazama otkrića novog leka, a ova oblast istraživanja često se označava kao dizajn molekula pomoću računara (engl. *computer-aided molecular design*, CAMD). CAMD metode se u najširem smislu mogu podeliti na one koje su zasnovane na proučavanju struktura i osobina poznatih liganada ciljne makromolekule (engl. *ligand-based*) i na one gde osnovu dizajna predstavlja trodimenzionalna struktura same ciljne makromolekule (engl. *structure-based*). Iako je u sklopu istraživanja obavljenog u okviru ove teze korišćen širok spektar CAMD metoda, u daljem tekstu će šire biti opisani principi i ograničenja dve metode: molekularne dinamike i proračuna slobodne energije baziranih na trajektorijama molekularne dinamike, odnosno računarskog *de novo* dizajna baziranog na primeni genetskih algoritama. Kao što će biti diskutovano u delu Rezultati i diskusija, dok su simulacije molekularne dinamike bile ključne za uspostavljanje strukturne osnove za dizajn novih inhibitora EZH2, implementirani algoritam za *de novo* dizajn omogućio je kreiranje nekoliko novih struktura potencijalnih EZH2 inhibitora.

1.5.1 Simulacije molekularne dinamike

Na temeljima velikog broja kristalografskih studija i rezultata prikupljenih drugim biofizičkim metodama, danas je izvesno da prilikom interakcije liganda i ciljne makromolekule često dolazi do značajnih konformacionih promena obe molekule. Brze i aproksimativne metode za predviđanje položaja i energije vezivanja malih organskih molekula za vezivna mesta na proteinima, kao što je doking, najčešće

ne uzimaju u obzir ovu činjenicu. Najveći broj algoritama za doking tretira receptor kao rigidnu strukturu, dozvoljavajući samo fleksibilnost liganda. Dodatno, efekti rastvarača se najčešće zanemaruju, iako molekule vode mogu igrati važnu ulogu u procesu vezivanja. Simulacije molekularne dinamike (MD) omogućavaju detaljno praćenje ponašanja molekulskih sistema u funkciji vremena, istovremeno uzimajući eksplicitno u obzir fleksibilnost struktura i efekte rastvarača, čime ove simulacije mogu pružiti jedinstven uvid u interakcije liganada i njihovih ciljnih makromolekula. Takav uvid može predstavljati racionalnu osnovu za dizajn novih molekula i strukturnu optimizaciju postojećih [259, 260].

Simulacija MD najčešće započinje od sve-atomske strukture sistema koji se simulira, koja može biti dobijena kristalografski, NMR spektroskopijom ili modeliranjem zasnovanom na homologiji. Iako je osobine i dinamiku takvog sistema precizno moguće opisati samo kvantno-mehanički (QM), metode ovog nivoa formalizma najčešće nisu primenjive ukoliko sistem ima više od nekoliko stotina atoma. Usled toga se za opis potencijala u MD simulacijama koriste empirijski razvijena, molekularno mehanička polja sila pomoću kojih je aproksimativno moguće odrediti silu koja deluje na svaki od atoma u datom trenutku vremena, i to u teorijskom okviru klasične mehanike [261]. Pretpostavka u razvoju ovih modela jeste da su sile u sistemu strogo aditivne, efekti polarizacije se zanemaruju, pa se potencijalna energija sistema date konfiguracije $V(\mathbf{r}_1, \mathbf{r}_2 \dots \mathbf{r}_n)$ najčešće izražava jednačinom oblika [262]:

$$\begin{aligned}
 V(\mathbf{r}_1, \mathbf{r}_2 \dots \mathbf{r}_n) = & \sum_{bonds} K_b (b - b_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 \\
 + & \sum_{dihedrals} K_\chi (1 + \cos(n\chi - \delta)) + \sum_{impropers} K_\varphi (\varphi - \varphi_0)^2 \\
 & + \sum_{nonbonded} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{\epsilon r_{ij}} \right) \quad (1.1)
 \end{aligned}$$

Kao što se može uočiti iz jednačine 1.1, potencijalna energija sistema parametrizovanog molekularno mehaničkim poljem sila zavisi, redom, od: (i) odstupanja dužina veza od ravnotežnih vrednosti; (ii) odstupanja uglova veza od ravnotežnih vrednosti; (iii) odstupanja torzionih uglova od ravnotežnih vrednosti; (iv) nedozvoljenih odstupanja od koplanarnosti; (v) van der Waals-ovih i elektrostatiskih interakcija, pri čemu se van der Waals-ove interakcije najčešće predstavljaju

Lennard–Jones-ovim (12–6) potencijalom.

Sila koja deluje na i -ti atom sistema, može se izraziti kao gradijent potencijalne energije:

$$F_i = -\nabla_{r_i} V(\mathbf{r}_1, \mathbf{r}_2 \dots \mathbf{r}_n) \quad (1.2)$$

U simulacijama MD, uzima se da se sistem ponaša u skladu sa zakonima klasične mehanike, pa se korišćenjem Njutnovog drugog zakona (jednačina 1.3) može izraziti odnos koji povezuje promenu položaja i -tog atoma u sistemu sa silom koja deluje na taj atom:

$$F_i = m_i a_i \quad (1.3)$$

$$F_i = m_i \frac{d^2 \mathbf{r}_i}{dt^2} \quad (1.4)$$

Integracijom ovih jednačina kretanja uz korišćenje malih vremenskih koraka (dt) moguće je pratiti promene konformacije i pokrete sistema u funkciji vremena. Vremenski koraci moraju biti srazmerni najbržim fluktuacijama u sistemu što su tipično vibracije kovalentnih veza između vodonika i drugih atoma (talasni broj 3.000–3.500 cm^{-1}). Treba napomenuti da je opis ovih fluktuacija na nivou harmonijskog oscilatora klasične mehanike (jednačina 1.1) veoma neprecizan, pa je ukupnu potencijalnu energiju sistema najčešće neophodno analitički korigovati. Alternativno, moguće je nametnuti ograničenje na fluktuacije dužine ovih veza što omogućava korišćenje nešto većih vremenskih koraka (2 ili 4, umesto 1 fs). Interesantno je da bi, na osnovu gore izloženog, tok MD simulacije trebalo da bude deterministički i da zavisi samo od početnih položaja i brzina atoma u sistemu. U praksi, međutim, numerička nepreciznost simulacije i upotreba stohastičkih algoritama najčešće uslovljavaju odstupanja od determinističkog toka.

Konačno, treba imati u vidu da MD, kao metoda statističke mehanike, za cilj ima razumevanje makroskopskih (ili termodinamičkih) svojstava sistema kroz posmatranje jednog njegovog izolovanog dela. Na primer, simulacija MD kompleksa jedne molekule liganda i proteina treba da omogući razumevanje konstante afiniteta tog liganda za dati protein, izmerene u *in vitro* uslovima gde se u jednoj epruveti može nalaziti reda veličine 10^{23} molekula, od kojih se svaka, u svakom trenutku

vremena, nalazi u veoma različitoj konformaciji i ima različitu energiju. MD stoga počiva na pretpostavci da prosečna vrednost nekog merljivog svojstva sistema tokom dugog vremenskog perioda posmatranja jednog njegovog izolovanog dela odgovara makroskopski određenoj vrednosti tog svojstva. Simulacija MD tako ima za cilj kompletno istraživanje faznog prostora koji je dostupan u skladu sa nekom funkcijom gustine verovatnoće, validnom u datom statističkom ansamblu. U okvirima ergodske hipoteze i u limitu veoma dugog vremena posmatranja, pretpostavlja se da će sistem istražiti sva dostupna mikrostanja u faznom prostoru, na osnovu čega se mogu donositi zaključci o makroskopskim svojstvima sistema. Ergodičnost većine MD trajektorija je upitna, imajući u vidu njihovo ograničeno trajanje. Rezultate ovih simulacija je zato uvek neophodno pažljivo interpretirati u kontekstu dostupnih eksperimentalnih podataka [263, 264].

I pored velikog broja aproksimacija koje su inherentni deo MD kao metode, simulacije MD danas igraju ključnu ulogu u više aspekata dizajniranja lekova pomoću računara. Ovim simulacijama može se poboljšati kvalitet struktura dobijenih biofizičkim metodama što ih čini komplementarnim sa eksperimentalnim pristupima razjašnjavanju strukture biomolekula. Istovremeno, simulacije MD mogu pružiti uvid u dinamiku biomolekula koji nije dostupan biofizičkim metodama. Primera radi, simulacijama MD mogu se identifikovati tranzitorno prisutna („kratkoživeća“) alosterna mesta na proteinima i razjasniti putevi alosternih komunikacija. Generisanje alternativnih konformacija proteina i vezivnih mesta primenom MD može biti korisno u identifikaciji strukturno raznovrsnih, potencijalno bioaktivnih jedinjenja u virtuelnom skriningu kolekcija. Simulacije MD mogu doprineti razumevanju uloge molekula rastvarača u procesu vezivanja liganada, i to kroz identifikaciju snažno vezanih molekula vode ili onih čije istiskivanje može doprineti energiji vezivanja. Procesi asocijacije liganada i proteina koji podrazumevaju značajne konformacione promene takođe se mogu okarakterisati pomoću simulacija MD. Konačno, simulacije MD omogućavaju proračune slobodne energije vezivanja (kao što je detaljnije opisano u narednom odeljku) pa mogu biti veoma korisne u razumevanju ključnih interakcija između poznatih liganada i proteina, kao i u identifikaciji novih, potencijalno bioaktivnih jedinjenja [259, 265].

Proračuni slobodne energije vezivanja

Jedna od ključnih primena simulacija molekularne dinamike jeste precizna procena slobodne energije vezivanja liganada za ciljnu makromolekulu (ΔG_{bind}). U

principu, ovakav proračun bi zahtevao simuliranje celokupnog procesa vezivanja, počevši od stanja gde su ligand i protein slobodni i veoma udaljeni u rastvoru, do stanja u kojem je došlo do njihove asocijacije i formiranja kompleksa:



Međutim, proces vezivanja liganda za receptor može trajati između nekoliko mikrosekundi i nekoliko stotina sekundi, dok je sa trenutno raspoloživim računarskim resursima najčešće moguće simulirati sisteme veličine oko 100.000 atoma u trajanju od nekoliko stotina nanosekundi do par mikrosekundi. Dodatno, statistički pouzdana procena ΔG_{bind} zahteva nekoliko ponavljanja događaja asocijacije i disocijacije tokom trajanja simulacije. Kako je ovo praktično neizvodivo, računarske metode za procenu slobodne energije vezivanja počivaju na određenim aproksimacijama ili primeni metoda koje ubrzavaju istraživanje energetski teško dostupnih stanja sistema. U najširem smislu, sve metode za procenu slobodne energije vezivanja mogu se podeliti u metode „putanje” (engl. *pathway*) i metode „krajnjeg stanja” (engl. *end-state*) [266, 267].

Rigorozniju od ove dve grupe čine metode „putanje”. Kako je slobodna energija funkcija stanja sistema, „putanja” koja se simulira ne mora nužno biti fizički moguća. Ovu činjenicu koriste metode „alhemijske transformacije” pomoću kojih se prate promene energije prilikom anihilacije ili dekuplovanja liganda iz kompleksa sa receptorom, i njegovog uklanjanja iz rastvora, pri čemu se iz ovako koncipiranog termodinamskog ciklusa može precizno proceniti ΔG_{bind} [263]. Druge metode iz ove grupe zasnivaju se na modifikaciji potencijala uvođenjem biasa koji pospešuje istraživanje energetski teško dostupnih stanja sistema (engl. *enhanced sampling*). Modifikovan potencijal najčešće se odnosi na jednu ili više kolektivnih varijabli sistema, koje se nalaze u funkciji trodimenzionalnih koordinata odabranih atoma. Dobar izbor kolektivnih varijabli omogućava efikasno istraživanje „putanje” vezivanja, daje uvid u relevantna prelazna stanja sistema a potencijalno i preciznu procenu relativne energije vezivanja. U ovu grupu metoda spadaju različiti vidovi metadinamike, *adaptive biasing force* algoritam, kao i ciljane (engl. *targeted*) odnosno upravljane (engl. *steered*) molekulska dinamika, pod pretpostavkom da zadovoljavaju jednakost Jarzinskog. Treba imati u vidu da većina ovih metoda ima za cilj da učini istraživanje faznog prostora efikasnijim, dok je proračun slobodne energije moguć ukoliko je zadovoljeno više kriterijuma. Detaljniji pregled dostupan je u referencama [268–270].

Konvergencija rezultata proračuna ΔG_{bind} pomoću metoda „putanje” zah-teva veoma dugačke simulacije, što čini njihovu rutinsku primenu u dizajniranju novih molekula nepraktičnom. Računarski efektivnija alternativa jesu metode kraj-njeg stanja, koje zanemaruju intermedijerna stanja (bilo fizička, bilo „alhemijaska”) i prilikom procene energije vezivanja koriste najčešće samo konfiguracije sistema iz kratkih simulacija kompleksa liganda i receptora. Dodatna aproksimacija ovih metoda ogleda se u zanemarivanju pojedinačnih doprinosa koje interakcije sa mo-lekulama rastvora (vode) daju slobodnoj energiji, i to kroz uvođenje modela efe-kata rastvarača (engl. *implicit solvent model*). Tretiranjem rastvarača implicitno i razmatranjem samo krajnjih stanja sistema, postiže se kompromis između tač-nosti proračuna i njegove računarske efikasnosti [271]. Dve najznačajnije metode ovog tipa razlikuju se po modelu kojim se aproksimiraju efekti rastvarača, i to su metoda molekularne mehanike uz generalizovani Bornov model (Molecular Mecha-nics/Generalized Born Surface Area, MM/GBSA) i metod molekularne mehanike uz Poisson-Boltzmann-ov model (Molecular Mechanics/Poisson-Boltzmann Surface Area, MM/PBSA) rastvarača. Teorijski okvir ovih metoda može se ukratko pred-staviti na sledeći način [272].

Slobodna energija vezivanja liganda za receptor može se izraziti kao razlika slobodnih energija kompleksa (LR) i receptora (R) odnosno liganda (L) u rastvoru:

$$\Delta G_{bind} = G_{LR_{sol}} - G_{R_{sol}} - G_{L_{sol}} \quad (1.6)$$

Slobodne energije kompleksa, receptora odnosno liganda mogu se izraziti kao suma više komponenata:

$$G = E_{MM} + G_{sol} - TS \quad (1.7)$$

gde je E_{MM} molekularno mehanička energija data poljem sila kojim je si-tem parametrizovan, G_{sol} slobodna energija solvatacije, S entropija sistema i T apsolutna temperatura. Ukupna molekularno mehanička energija dalje se može raš-članiti na doprinose koji rezultuju interakcijama kroz hemijske veze ($E_{internal}$), i na doprinose van der Waals-ovih (E_{vdW}) i elektrostatskih interakcija (E_{ele}):

$$E_{MM} = E_{internal} + E_{vdW} + E_{ele} \quad (1.8)$$

Konačno, slobodna energija solvatacije računa se uzimajući u obzir doprinos polarnih interakcija, procenjen pomoću odgovarajućeg modela efekata rastvarača, i nepolarnih interakcije za koje se uzima da su proporcionalne površini dostupnoj rastvaraču:

$$G_{sol} = G_{GB/PB} + G_{SASA} \quad (1.9)$$

MM/GBSA i MM/PBSA su aproksimativne metode čiji rezultati najčešće ne odgovaraju apsolutnoj energiji vezivanja. Međutim, ove metode se smatraju veoma korisnim u računarskom dizajniranju novih molekula jer je na velikom broju primera pokazano da se rezultati dobijeni ovim putem dobro slažu sa relativnim energijama vezivanja odnosno redosledom afiniteta liganada za dato ciljno mesto dejstva. Stoga, MM/GBSA i MM/PBSA mogu pružiti okvir za testiranje slaganja eksperimentalnih i računarskih rezultata, mogu biti korisne u raščlanjavanju doprinosa pojedinih interakcija ukupnoj energiji vezivanja i značajno mogu poboljšati uspešnost računarske identifikacije novih bioaktivnih molekula [271, 272].

1.5.2 Računarsko *de novo* dizajniranje

Pod *de novo* dizajnom se obično podrazumeva dizajniranje molekula uz pomoć računara koje za cilj ima kreiranje hemijskih struktura koje ispoljavaju afinitet za definisano ciljno mesto dejstva, a koje istovremeno, sa aspekta hemijske sličnosti, predstavljaju nove strukture različite od poznatih liganada date makromolekule. Osnovni problem koji se javlja prilikom *de novo* dizajna jeste dimenzionalnost hemijskog prostora: pretpostavlja se da između 10^{15} i 10^{18} jedinjenja potencijalno može predstavljati lekove [273]. Čak i veoma brze računarske metode za formiranje kombinatornih kolekcija i virtuelni skrining jedinjenja ne omogućavaju iscrpnu pretragu celokupnog hemijskog prostora u cilju identifikovanja željenih struktura. Stoga računarsko *de novo* dizajniranje molekula podrazumeva korišćenje heurističkih i metaheurističkih metoda. Ove metode predstavljaju grupu metoda stohastičke optimizacije koje odlikuje efikasnost pretrage prostora potencijalnih rešenja, ali i činjenica da njihova primena ne garantuje pronalaženje pravog optimuma [274].

Računarsko *de novo* dizajniranje predmet je aktivnog istraživanja od ranih 90-tih godina prošlog veka. Jedan od glavnih nedostataka prvih algoritama ovog tipa ogledao se u činjenici da predložene strukture najčešće nisu bile sintetski dostupne

[275], što je značajno ograničilo primenljivost *de novo* dizajna u racionalnom otkriću novih lekova. Prepoznavanje ovog nedostatka uticalo je na razvoj novih algoritama, pojavom tendencije da se sintetska dostupnost struktura inkorporira u same algoritme kroz empirijska pravila o mogućim vezama između različitih tipova atoma ili kroz upotrebu definisanih reakcija medicinske hemije za spajanje manjih fragmenata [276]. Ovakav pristup je u značajnoj meri unapredio kvalitet kolekcija jedinjenja koje se mogu dobiti *de novo* dizajnom.

Glavna motivacija za primenu *de novo* dizajna ogleda se u činjenici da ova metoda omogućava pronalazak bioaktivnih struktura u novim (neistraženim) delovima hemijskog prostora. Virtuelni ili eksperimentalni skrining komercijalno dostupnih kolekcija jedinjenja uvek je ograničen na relativno mali deo hemijskog prostora, koji je često već patentno zaštićen ili prethodno prepoznat u literaturi. Takođe, modulatori novih ciljnih mesta dejstva često su nedovoljno zastupljeni u postojećim kolekcijama jedinjenja [277], što dodatno govori u prilog primeni *de novo* dizajna. Ipak, važno je imati u vidu da računarski algoritmi za *de novo* dizajn i sami imaju značajna ograničenja. Primena ove metode gotovo nikada ne rezultuje otkrićem jedinjenja nanomolarne aktivnosti, izražene selektivnosti i optimalnog farmakokinetičkog profila. Ovakve algoritme, stoga, treba posmatrati kao metode za generisanje hemijski validnih ideja, odnosno sintetski dostupnih struktura koje imaju veću verovatnoću ispoljavanja aktivnosti na datom ciljnom mestu dejstva od onih struktura koje su zastupljene u postojećim, nefokusiranim kolekcijama. Algoritmi za *de novo* dizajn prevashodno omogućavaju automatizovano kreiranje početnih, inovativnih struktura čija dalja optimizacija može dati jedinjenja željenog profila [273, 275, 278].

Kao što je prethodno navedeno, kombinatorna eksplozija mogućih struktura koje se *in silico* mogu dobiti spajanjem manjih gradivnih jedinica (engl. *building blocks*) nameće izbor odgovarajuće heuristike koja će učiniti algoritam računarski efikasnim. Primera radi, jedan od novijih algoritama za *de novo* dizajn iz literature – DOGS (*Design of Genuine Structures*) [279] – zasnovan je na „pohlepnom algoritmu” (engl. *greedy algorithm*). DOGS koristi preko 50 reakcija čestih u medicinskoj hemiji [280] za iscrpno spajanje fragmenata sa svim drugim, kompatibilnim fragmentima iz kolekcije. Nakon formiranja kombinatorne kolekcije „intermedijera”, strategija „pohlepnog algoritma” podrazumeva rangiranje svih potencijalnih rešenja i izbor najboljeg seta koji se uključuje u drugi korak reakcija, ponovo sa svim kompatibilnim fragmentima iz kolekcije. Kako ova strategija zahteva veliki broj evaluacija kvaliteta generisanih rešenja, DOGS se oslanja na procenu topološke sličnosti novodizajnira-

nih struktura i struktura jedinjenja za koja se zna da su aktivna na datom ciljnom mestu dejstva. Iako se ovaj pristup pokazao kao uspešan na većem broju primera [279, 281, 282], ovako koncipiran algoritam ne omogućava korišćenje informacija o trodimenzionalnoj strukturi, odnosno bioaktivnim konformacijama jedinjenja i njihovim ključnim farmakofornim elementima. Naime, procena farmakoforne sličnosti ili sličnosti oblika jedinjenja podrazumeva generisanje reprezentativnih konformera novodizajniranih struktura, pa je ovaj vid procene kvaliteta potencijalnih rešenja računarski značajno zahtevniji. Kompromis između računarske zahtevnosti i efikasnosti pretrage prostora potencijalnih rešenja nameće alternativni izbor heuristike na kojoj bi takav algoritam bio zasnovan. U ovom kontekstu, genetski algoritmi su se pokazali kao veoma primenjivi u *de novo* dizajnu. Kako je algoritam razvijen u okviru ovog istraživanja takođe baziran na genetskom algoritmu, u narednim odeljcima detaljnije su opisani principi ove metaheuristike i neke od primena u *de novo* dizajnu.

Principi genetskih algoritama

Genetski algoritmi (GA) su jedna od najšire korišćenih metaheuristika koja primenu nalazi u raznovrsnim oblastima nauke od inženjerstva i ekonomije do hemoinformatike. GA pripada grupi tzv. algoritama inspirisanih prirodom i prvi ga je predložio Holland sredinom 70-tih godina prošlog veka [283]. Ovaj algoritam predstavlja apstrakciju Darwinove teorije evolucije osmišljen tako da potencijalna rešenja problema koji se optimizuje predstavljaju populaciju jedinki koja evoluira kroz generacije, uz veću verovatnoću opstanka onih jedinki koje su prilagođenije, odnosno onih rešenja koja su bolja. U osnovi većine genetskih algoritama stoje tri genetska operatora: mutacija, ukrštanje ili rekombinacija (engl. *cross-over*) i selekcija. Mutacijama i ukrštanjem populacije roditeljskih rešenja dobija se nova generacija jedinki, pri čemu se primenom selekcije obezbeđuje da bolja rešenja imaju veću verovatnoću da se prenesu u naredne generacije (elitizam). Vršenje selekcije zahteva prethodno definisanje jedne ili više funkcija procene (engl. *objective function*) pomoću kojih je moguće kvantifikovati određena svojstva svakog rešenja u populaciji (npr. složenost, tačnost). Funkcija prilagođenosti ili spremnosti (engl. *fitness function*) prevodi vrednosti objektivnih funkcija u meru prilagođenosti jedinke. Ukoliko se problem optimizacije postavi kao problem maksimizacije, funkcija prilagođenosti ima veće vrednosti za više vrednosti objektivne funkcije. Obratno, ukoliko je cilj minimizacija vrednosti objektivne funkcije, veće vrednosti prilagođenosti imaju rešenja sa manjom vrednošću objektivne funkcije. Primer genetskog algoritma prikazan je

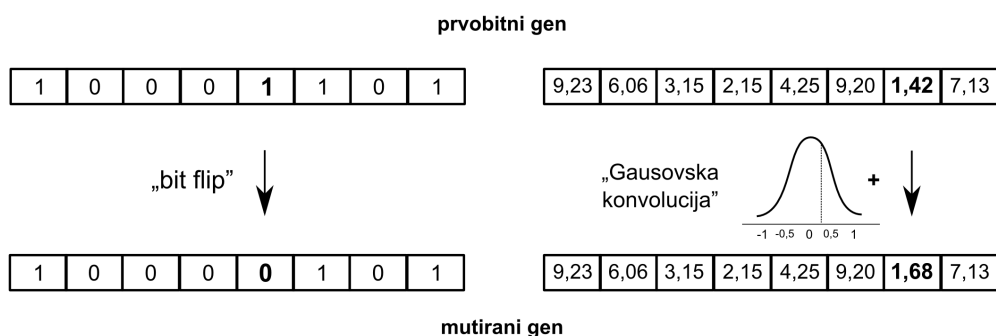
u pseudokodu kao algoritam 1 [284]:

Algoritam 1 Genetski algoritam

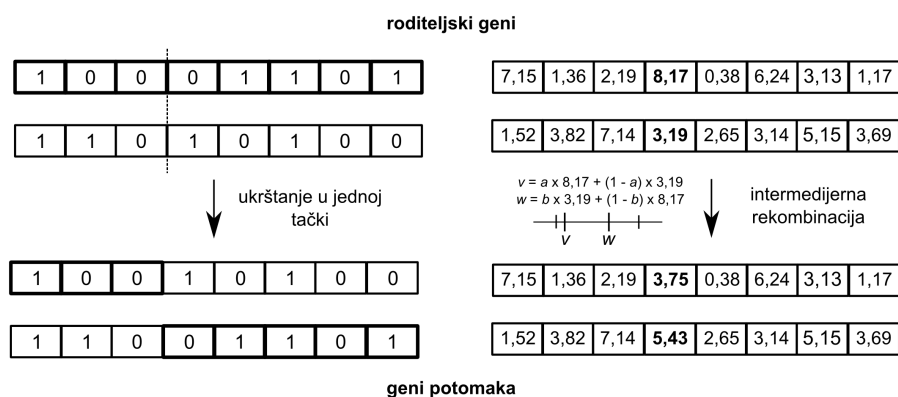
```
1: Definisati funkcije procene  $f(x)$ ,  $x = (x_1 \dots x_i)$ 
2: Definisati funkciju prilagođenosti  $F(x)$  ( $F(x) \propto f(x)$  za problem maksimizacije)
3:  $N_P \leftarrow$  veličina populacije
4:  $p_m \leftarrow$  verovatnoća mutacije
5:  $p_c \leftarrow$  verovatnoća ukrštanja
6:  $P \leftarrow \{\}$ 
7: for  $i \leftarrow 0, i < N_P, i++$  do ▷ Slučajni odabir početne populacije
8:    $P \leftarrow P \cup \{ \text{nasumično generisana jedinka} \}$ 
9: end for
10: repeat
11:   for  $P_i \in P$  do ▷ Procena prilagođenosti jedinki u populaciji
12:     Izračunati  $F(P_i)$ 
13:   end for
14:    $Q \leftarrow \{\}$ 
15:   while  $|Q| < N_P$  do ▷ Formiranje sledeće generacije
16:      $R_a \leftarrow$  roditelj izabran iz  $P$ 
17:      $R_b \leftarrow$  roditelj izabran iz  $P$ 
18:     if slučajan decimalni broj iz  $[0, 1] > p_c$  then
19:        $D_a, D_b \leftarrow$  Ukrštanje( $R_a, R_b$ )
20:     else
21:        $D_a, D_b \leftarrow$  kopija( $R_a$ ), kopija( $R_b$ )
22:     end if
23:     if slučajan decimalni broj iz  $[0, 1] > p_m$  then
24:        $D_a \leftarrow$  Mutacija( $D_a$ )
25:        $D_b \leftarrow$  Mutacija( $D_b$ )
26:     end if
27:      $Q \leftarrow Q \cup \{D_a, D_b\}$ 
28:   end while
29:    $P \leftarrow Q$ 
30: until dostignut kriterijum zaustavljanja ili maksimalan broj generacija
31: return finalna populacija  $P$ 
```

Iako je GA konceptualno jednostavan, njegova primena na specifičan problem, kao što je *de novo* dizajn hemijskih struktura, najčešće zahteva razvoj odgovarajućeg načina predstavljanja individua (hromozoma) kao i pogodnih formi operatora mutacije i ukrštanja. Tradicionalno, hromozomi u GA su predstavljeni kao nizovi fiksne dužine koji sadrže binarne ili decimalne brojeve. Operatori mutacije i ukrštanja moraju se prilagoditi načinu kodiranja hromozoma. Primera radi, mutacija niza bitova može se ostvariti jednostavnim „obrtnjem” jednog bita, dok se na niz decimalnih vrednosti može primeniti tzv. Gausovska konvolucija (slika 1.17a). Ukrštanje dva niza bitova moguće je izvesti na različite načine, najjednostavniji od kojih

je ukrštanje u jednoj tački (engl. *single point cross-over*); slično, dva niza decimalnih brojeva mogu se ukrstiti intermedijernom rekombinacijom (slika 1.17b) [274]. Mogućnost različitog kodiranja hromozoma i dizajn specifično prilagođenih genetskih operatora daje veliku fleksibilnost genetskim algoritmima i omogućava njihovu primenu u širokom spektru problema optimizacije. Ipak, važno je razumeti način na koji svaki od genetskih operatora utiče na evoluciju rešenja, kao što je dalje opisano.



(a) Obrtanje jednog bita (levo) i Gausovska konvolucija (desno) kao primeri mutacije binarno odnosno decimalno kodiranog gena.



(b) Ukrštanje u jednoj tački (levo) i intermedijerna rekombinacija u jednoj slučajno izabranoj tački (desno) kao primeri ukrštanja binarno odnosno decimalno kodiranih gena.

Slika 1.17: Šematski prikaz mutacije i ukrštanja binarno i decimalno kodiranih gena.

Mutacija Genetski algoritam treba da omogući istraživanje celokupnog prostora mogućih rešenja (eksplorativnost) ali i da obezbedi fokusiranu potragu u određenim delovima prostora u cilju pronalaska optimuma (eksploatacija). Mutacija stohastički dovodi do povećanja diverziteta populacije i time uslovljava istraživanje novih delova prostora mogućih rešenja. Mutacije osiguravaju da GA ima eksplorativan karakter, pomažu napuštanju lokalnih optimuma i sprečavaju preranu konvergenciju

algoritma. S druge strane, preučestale mutacije mogu dovesti do proizvoljne pretrage prostora mogućih rešenja, što smanjuje verovatnoću pronalaska optimuma i odlaže konvergenciju algoritma. Empirijski je pokazano da optimalan izbor verovatnoće mutacije uglavnom podrazumeva male vrednosti (0,01–0,05), mada u zavisnosti od specifične implementacije GA, adekvatan izbor može biti i značajno drugačiji [274, 284, 285].

Ukrštanje Nasuprot mutacije, ukrštanje je genetski operator koji obezbeđuje istraživanje ograničenog podprostora potencijalnih rešenja. Ukrštanje omogućava mešanje karakteristika postojećih rešenja i time favorizuje konvergenciju do lokalnog optimuma. Ukrštanjem najčešće nije moguće postići eksploraciju celokupnog prostora rešenja; primera radi, ukrštanjem rešenja $S_1 = \{1, 0, 1, 1, 0\}$ i $S_2 = \{0, 0, 1, 0, 1\}$ uvek će se dobiti rešenje u podprostoru $\{*, 0, 1, *, *\}$. Dok je stepen mutacije u GA najčešće nizak, verovatnoća ukrštanja se obično podešava na relativno visoke vrednosti (0,8–0,9). Međutim, u kontekstu interakcije niske učestalosti mutacija i visokog elitizma algoritma, visoka učestalost ukrštanja može dovesti do prerane konvergencije. Konceptualno, ukrštanje treba da omogući širenje povoljnih karakteristika kroz populaciju razmenom gradivnih jedinica koje doprinose izraženoj prilagođenosti najboljih rešenja. U ranim generacijama algoritma, međutim, ovo može značajno smanjiti diverzitet populacije, ograničiti potragu na lokalni optimum i time sprečiti konvergenciju ka globalnom optimumu u prostoru rešenja [274, 284, 285].

Selekcija Selekcija najboljih rešenja je ključan element GA jer se ovom operacijom evolucija populacije prilagođava pejzažu funkcija procene koje određuju kvalitet rešenja. Selekcija može biti deo različitih delova algoritma. Obično, prilikom formiranja naredne generacije individua, izbor roditelja (algoritam 1, linije 16 i 17) nije nasumičan, već je, u većoj ili manjoj meri, verovatnoća izbora proporcionalna prilagođenosti roditelja. Ukoliko je verovatnoća izbora direktno proporcionalna prilagođenosti roditelja, ovaj vid selekcije se označava kao selekcija točka ruleta (engl. *roulette wheel selection*). Manje elitističan karakter ima selekcija pomoću turnira (engl. *tournament selection*) gde se iz populacije slučajno bira n individua, od kojih najprilagođenija postaje roditelj. Često se koriste binarni turniri ($n = 2$). Pored izbora roditelja, selekcija može biti implementirana u delu algoritma gde se vrši smena generacija (algoritam 1, linija 27). Pored jednostavne zamene celokupne roditeljske generacije – generacijom potomaka, moguće je u narednoj generaciji zadržati n najboljih roditeljskih rešenja (tzv. elita) ili celokupnu populaciju roditelja i potomaka

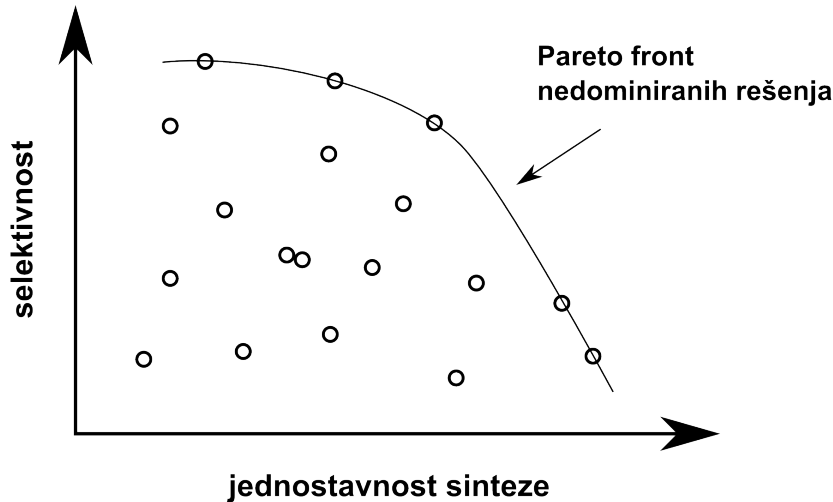
rangirati po prilagođenosti, uz zadržavanje najprilagođenije polovine. Verzija GA algoritma označena kao GA stacionarnog stanja (engl. *steady-state GA*) podrazumeva da se nakon generisanja dva nova rešenja proverava da li su nove individue prilagođenije od najlošijeg rešenja u postojećoj populaciji. Ukoliko je to slučaj, lošija rešenja se uklanjaju iz populacije a nove individue dodaju. U poređenju sa generacijskim, GA stacionarnog stanja ima izraženiji karakter eksploatacije jer najprilagođenije individue ostaju u populaciji veoma dugo. Zauzvrat, ovo može biti uzrok prerane konvergencije [274, 284, 285].

Genetski algoritmi za optimizaciju sa više ciljeva

De novo dizajniranje hemijskih struktura najčešće podrazumeva istovremeno zadovoljenje više ciljeva, koji često mogu biti suprotstavljeni. Primera radi, pored aktivnosti na određenom ciljnom mestu dejstva, dizajnirane strukture treba da imaju odgovarajuće fizičko-hemijske karakteristike, da budu sintetski dostupne i hemijski raznovrsne. Iz ovog razloga je u oblasti *de novo* dizajna postala ustaljena praksa da razvijeni algoritam bude prilagođen za optimizaciju sa više ciljeva [275, 286].

Opisan je veći broj metoda kojima se GA mogu prilagoditi optimizaciji sa više ciljeva (engl. *multi-objective optimization*, MOO) [284]. Jedan broj ovih metoda podrazumeva prevođenje vrednosti više funkcija prilagođenosti u jednu vrednost (kroz zbir, prosečnu vrednost ili prosečnu vrednost dobijenu korišćenjem različitih koeficijenata uz različite metrike prilagođenosti) nakon čega problem optimizacije postaje problem maksimizacije samo jedne funkcije prilagođenosti. Ovi pristupi najčešće zahtevaju arbitraran ili empirijski izbor koeficijenata kojima će svakom od ciljeva biti dodeljena određena „težina”. Alternativni, i češće korišćeni pristup, jeste pronalažanje Pareto-optimalnih ili nedominiranih rešenja. Koncept Pareto-optimalnosti direktno je povezan sa konceptom Pareto-dominacije. Za rešenje A se može reći da Pareto-dominira nad rešenjem B, ukoliko je rešenje A barem podjednako dobro kao rešenje B i bolje od rešenja B po vrednosti barem jedne od funkcija prilagođenosti [274]. Sva rešenja za koja ne postoji Pareto-dominantno rešenje čine Pareto-optimalna rešenja ili tzv. Pareto front (slika 1.18). Iako korišćenje koncepta Pareto-optimalnosti u GA-MOO eliminiše potrebu za dodeljivanjem relativne „težine” svakom od ciljeva optimizacije, treba imati u vidu da efikasnost optimizacije značajno opada sa brojem ciljeva. Dodatno, Pareto-optimalna rešenja nisu nužno ravnomerno raspodeljena na Pareto frontu (ili hiperravni). Moguće je da Pareto-

optimalna rešenja sva budu relativno loša u kontekstu jednog od ciljeva optimizacije, usled čega većina implementacija MOO algoritama nastoji i da generiše skup rešenja koja su dobro raspoređena duž Pareto fronta [274].



Slika 1.18: Ilustracija Pareto fronta na primeru optimizacije sa dva cilja, čije vrednosti funkcija procene treba maksimizovati

Jedna od najpopularnijih varijacija GA prilagođena za MOO jeste algoritam pod nazivom *Nondominated Sorting Genetic Algorithm II* (NSGA-II) koji su predložili Deb i saradnici [287]. Prilagođenost individua u populaciji kvantifikuje se kroz Pareto rang (broj rešenja koja Pareto-dominiraju nad datim) i kroz udaljenost datog rešenja od drugih rešenja na istom Pareto frontu (engl. *crowding distance*). NSGA-II ima elitistički karakter: nakon formiranja populacije potomaka – kroz mutacije, ukrštanje i selekcijom roditelja kroz binarne turnire – roditeljska populacija i populacija potomaka zbirno se rangiraju i gornja polovina populacije prenosi u sledeću generaciju. Elitistička priroda NSGA-II trebalo bi da obezbedi relativno brzu konvergenciju.

Primene genetskih algoritama u *de novo* dizajnu

U literaturi je opisan veliki broj algoritama za *de novo* dizajn [288–297]. Ovde će biti izloženo samo nekoliko primera zasnovanih na GA, kao ilustracija raznovrsnih pristupa kodiranju genotipa dizajniranih molekula, različitih implementacija operatora mutacije i ukrštanja kao i načina za procenu prilagođenosti dizajniranih struktura u kontekstu definisanih ciljeva.

Jedan od najpopularnijih GA za *de novo* dizajn je GANDI (*Genetic Algo-*

rithm-based de Novo Design of Inhibitors) koji su opisali Dey i Caffisch [288]. GANDI zahteva dostupnost strukture ciljnog mesta dejstva, a opciono prihvata i strukture poznatih inhibitora. Optimizacija započinje dokingom kolekcije fragmenata u definisano vezivno mesto, pri čemu se hromozomi kodiraju kao nizovi gena, gde je svaki gen predstavljen celobrojnomo vrednosti koja odgovara nekoj od poza kolekcije fragmenata. Ukrštanje hromozoma implementirano je kao ukrštanje u jednoj tački, a mutacije se vrše slučajnim izborom nove celobrojne vrednosti iz normalne raspodele. Fragmenti kodirani hromozomom zatim se povezuju korišćenjem dostupnih linkera. Prilagođenost svakog rešenja kvantifikuje se procenom energije vezivanja i unutrašnje energije svake dizajnirane strukture, a opciono se uzima u obzir i topološka ili trodimenzionalna sličnost sa poznatim inhibitorima.

Durrant i saradnici [296] opisali su algoritam, AutoGrow 3.0, koji se takođe zasniva na dokingu struktura u definisano vezivno mesto, ali koji direktno uzima u obzir sintetsku dostupnost kreiranih struktura. Dizajn započinje od kolekcije fragmenata, pri čemu je mutacija individua u populaciji implementirana kao slučajno izabrana reakcija klik–hemije postojećeg i njemu kompatibilnog fragmenta. Za mutacije AutoGrow koristi bazu reakcija klik–hemije implementiranu u AutoClickChem programu [298]. Ukrštanje hromozoma vrši se razmenom bočnih lanaca dva roditelja, nakon identifikovanja najveće zajedničke podstrukture korišćenjem LigMerge algoritma [299]. Proizvodi ukrštanja, stoga, nisu nužno sintetski dostupni reakcijama klik–hemije. Prilagođenost dobijenih struktura procenjuje se fleksibilnim dokingom.

Pfeffer [291] i saradnici razvili su GARLig (*Genetic Algorithm using Reagents to compose Ligands*) algoritam inspirisan principima kombinatorne sinteze. Dizajn biblioteke započinje od osnovne strukture, koja se ne menja tokom optimizacije, u okviru koje se označavaju mesta na kojima je moguća varijacija supstituenata. U cilju kodiranja genotipa rešenja u populaciji, reaktanti se grupišu po hemijskoj sličnosti i svakom se dodeljuje celobrojni identifikator, na način koji obezbeđuje da slični reaktanti imaju slične vrednosti ovog jedinstvenog broja. Sam GA je implementiran uz elitističke elemente koji dozvoljavaju prenos roditeljske elite u narednu generaciju bez promena. Mutacija i ukrštanje implementirani su slično Gausovskoj konvoluciji i intermedijernoj rekombinaciji, samo u prostoru celobrojnih vrednosti. Prilagođenost rešenja procenjuje se dokingom.

Jedan od prvih algoritama za *de novo* dizajn uz optimizaciju više ciljeva opisali su Nicolaou i saradnici [289]. MEGA (*Multiobjective Evolutionary Graph Al-*

gorithm) kodira genetsku reprezentaciju struktura kao grafove, gde su čvorovi grafa gradivne jedinice, a grane njihova povezanost. Gradivne jedinice mogu biti atomi, veze ili fragmenti, pri čemu se njihova mogućnost povezivanja utvrđuje pomoću RECAP pravila [300]. Mutacija podrazumeva promene slučajno izabranih čvorova grafa, dok se ukrštanje vrši razmenom čvorova roditeljskih grafova, pri čemu se izvodivost izmene procenjuje takođe RECAP pravilima. MEGA optimizuje strukture sa tri cilja: (i) povećanje afiniteta dizajnirane strukture za ciljno mesto, što se procenjuje dokingom; (ii) povećanje sličnosti ili različitosti novih struktura u poređenju sa strukturom poznatog bioaktivnog liganda; (iii) smanjenje broja kršenja pravila Lipinskog. Prilagođenost rešenja procenjuje se prema Pareto rangu, pri čemu algoritam istovremeno koristi koncept niša (engl. *nicning*) u cilju očuvanja diverziteta populacije.

Poglavlje 2

Cilj rada

Ciljevi istraživanja u okviru ove doktorske disertacije bili su:

1. Uspostavljanje osnove za racionalno dizajniranje novih inhibitora EZH2, zasnovano na poznavanju strukture ciljnog vezivnog mesta na enzimu. Specifični ciljevi u okviru ovog dela istraživanja su:
 - (a) Uspostavljanje i validacija modela trodimenzionalne strukture vezivnog mesta za kofaktor EZH2, u konformaciji relevantnoj za proučavanje interakcije liganada sa ovim vezivnim mestom.
 - (b) Izučavanje interakcije poznatih inhibitora EZH2 sa kofaktor-vezivnim mestom primenom simulacija molekularne dinamike i proračuna slobodne energije vezivanja, kao i definisanje intermolekulskih interakcija koje ključno doprinose aktivnosti i selektivnosti proučavanih inhibitora.
 - (c) Uspostavljanje i validacija farmakofornog modela inhibitora EZH2 na osnovama rezultata simulacija MD odgovarajućih kompleksa.
2. Razvoj pouzdanih modela za računarsku klasifikaciju malih organskih molekula kao supstrata i nesupstrata Pgp-a i BCRP-a. Uz analizu strukturalnih odrednica transporta posredovanog ovim transporterima, razvijeni modeli treba da podrže dizajn novih potencijalnih inhibitora EZH2 koji neće biti njihovi supstrati. Dizajn takvih inhibitora ima za cilj smanjenje rizika od izostanka ćelijske efikasnosti usled prekomerne ekspresije Pgp-a i/ili BCRP-a u tumorskim ćelijama. Specifični ciljevi u okviru ovog dela istraživanja su:

- (a) Formiranje visoko-konzistentnih setova hemijski raznovrsnih supstrata i nesupstrata Pgp-a, odnosno BCRP-a, primenom sistematičnog pristupa u proceni pouzdanosti dostupnih eksperimentalnih podataka o transportu posredovanom ovim transporterima.
 - (b) Razvoj i validacija prediktivnih klasifikacionih modela primenom dve metode mašinskog učenja: veštačkih neuronskih mreža i *support vector machine* algoritma.
 - (c) Definisane delova hemijskog prostora u kojem je najveća verovatnoća pronalaska supstrata Pgp-a, BCRP-a, odnosno dualnih supstrata ova dva transportera.
3. Računarski dizajn novih potencijalnih inhibitora EZH2 primenom *de novo* pristupa. Specifični ciljevi u okviru ovog dela istraživanja su:
- (a) Implementacija algoritma za *de novo* dizajn hemijskih struktura uz istovremenu optimizaciju više ciljeva. Data implementacija treba da obezbedi delimično automatizovan dizajn jedinjenja koja zadovoljavaju prethodno uspostavljenu farmakofornu hipotezu inhibitora EZH2, poseduju adekvatne fizičko-hemijske karakteristike i imaju nisku sintetsku složenost.
 - (b) Formiranje fokusirane, virtuelne kolekcije potencijalnih EZH2 inhibitora korišćenjem implementiranog algoritma, i pretraživanje ove kolekcije u cilju identifikacije manjeg broja struktura za dalje proučavanje.
 - (c) Identifikacija optimalnih struktura potencijalnih EZH2 inhibitora primenom simulacija MD i proračuna slobodne energije vezivanja.
 - (d) Modifikacije strukture predloženih EZH2 inhibitora u cilju eliminisanja eventualne podložnosti transportu posredovanom Pgp-om i/ili BCRP-om.

Poglavlje 3

Eksperimentalni deo

3.1 Simulacije molekularne dinamike EZH2

3.1.1 Homologo modeliranje EZH2

U razvoju modela kompletnog kofaktor–vezivnog mesta EZH2 iskorišćena je dostupna kristalna struktura njegovog SET domena, dok je deo strukture koji formiraju aminokiseline post-SET domena modeliran na osnovu homologije sa drugim PKMT.

U razvoju modela post-SET domena, kao strukturne matrice iskorišćene su kristalne strukture enzima SETD7 i SETD8, koji, kao i EZH2, ne poseduju post-SET domen bogat cisteinima. Za modeliranje su upotrebljene strukture tercijarnih kompleksa, u kojima su enzimi kokristalisani sa proizvodom kofaktora SAH i histonskim supstratom, jer je u ovim formama post-SET domen potpuno struktuiran i pozicioniran tako da zatvara vezivno mesto, stabilizujući SAH. Trodimenzionalne strukture preuzete su iz internet baze Protein Data Bank (PDB). Korišćene su kristalografski rešene strukture kompleksa SETD7 rezolucije 1,75 Å iz PDB unosa 1O9S [68] i SETD8 rezolucije 1,45 Å iz PDB unosa 1ZKK [301]. Kanonske sekvence humanih EZH2 (segmenat 611–746), SETD7 (segmenat 214–366) i SETD8 (segmenat 257–393) preuzete su iz internet baze UniProt Protein Knowledgebase (unosni Q15910, Q8WTS6 i Q9NQR1, redom). Poravnanje sekvenci izvršeno je pomoću T-Coffee Servera [302, 303] i ilustrovano je na slici 3.1.

Kristalne strukture kompleksa SETD7 i SETD8 obrađene su u programu

```

EZH2_(UniProtKB_Q15910) 611 K K H L L L - - A P S D V A G W G I F I K D P V Q K N E F I S E Y C G E I I S Q D E A D R R G K V Y
SETD7_(UniProtKB_Q8WTS6) 214 E R - V Y V A E S L I S S A G E G L F S K V A V G P N T V M S F Y N G V R I T H Q E V D S R D W A L
SETD8_(UniProtKB_Q9NQR1) 257 E G - M K I - - D L I D G K G R G V I A T K Q F S R G D F V V E Y H G D L I E I T D A K K R E A L Y

EZH2_(UniProtKB_Q15910) 659 D K - - - - - Y M C S F L F N L N N D F V V D A T R K G N K - - - - - I - R F A N H S V N P N
SETD7_(UniProtKB_Q8WTS6) 263 N G - - - - - N - - - - - T L S L D E E T V I D V P E P Y N H V S K Y C A S L G H K A N H S F T P N
SETD8_(UniProtKB_Q9NQR1) 304 A Q D P S T G C Y M Y Y - F Q Y L S K T Y C V D A T R E T N R - - - - - L G R L I N H S K C G N

EZH2_(UniProtKB_Q15910) 695 C Y A K V M M V N G - D H R I G I F A K R A I Q T G E E L F F D Y R Y S Q A D A L K Y V G I E - R E
SETD7_(UniProtKB_Q8WTS6) 303 C I Y D M F V H P R F G P I K C I R T L R A V E A D E E L T V A Y G Y D H S P P G K S - G P E A P E
SETD8_(UniProtKB_Q9NQR1) 346 C Q T K L H D I D G - V P H L I L I A S R D I A A G E E L L Y D Y G D - - - - R S K A - S I E A H P

EZH2_(UniProtKB_Q15910) 743 - - - - - - - - - - - M E I P
SETD7_(UniProtKB_Q8WTS6) 352 W Y Q V E L K A F Q A T Q Q K
SETD8_(UniProtKB_Q9NQR1) 390 - - - - - - - - - - - W L K H

```

Slika 3.1: Superpozicija sekvenci delova EZH2, SETD7 i SETD8 korišćeno u razvoju homologog modela.

Chimera 1.8 [304], uklanjanjem svih osim jednog proteinskog lanca, histonskih supstrata i drugih kokristalisanih molekula. Kokristalisanu SAH, naprotiv, ostavljen je u strukturama koje su korišćene u razvoju homologog modela EZH2. Modeliranje je izvršeno pomoću programa Modeller 9.11 [305, 306], korišćenjem Chimera-e kao grafičkog interfejsa. Generisano je ukupno 1.000 homologih modela EZH2 od kojih je sačuvano 50 sa najboljim vrednostima Modeller funkcije procene.

Najveći deo strukture finalnog modela EZH2 preuzet je iz kristalografski rešene strukture njegovih SET i CXC domena rezolucije 2 Å (PDB unos 4MI0) [125]. Ova struktura skraćena je na deo koji obuhvata SET domen (segment 611–726), pri čemu su aminokiseline Y658 i D659, čiji položaj nije kristalografski rešen, date pomoću protokola za modeliranje petlji programa Modeller [307]. Na ovako pripremljenu strukturu SET domena EZH2, redom je superponirano 50 najboljih homologih modela u cilju odabira najslabijeg, uzimajući u obzir koren srednje vrednosti kvadrata odstupanja (engl. *root-mean-square deviation*, RMSD) položaja C_{α} atoma. Post-SET domen odabranog modela (segment 727–746) zatim je pripojen kristalografski rešenoj strukturi SET domena, čime je dobijen konačni model dužine 136 aminokiselina, sa potpuno formiranim mestom za vezivanje kofaktora. U cilju daljeg unapređenja modela, izvršene su simulacije molekularne dinamike (MD). U vezivno mesto prvo je dodat proizvod kofaktora SAH u konformaciji za koju je poznato da je visoko očuvana u binarnim i tercijarnim strukturama svih PKMT. Ovaj korak je izveden superponiranjem strukture modela sa kristalografski rešenom strukturom binarnog kompleksa GLP (PDB unos 3HNA), pri čemu su drugi autori ranije ukazali na sličnost vezivnih mesta ova dva enzima [125]. Dinamika pripremljenog kompleksa istraživana je pomoću simulacija klasične i ubrzane MD.

3.1.2 Simulacije MD EZH2–SAH kompleksa i apo oblika EZH2

Priprema EZH2–SAH kompleksa za simulacije MD izvršena je na sledeći način. N–terminus EZH2 neutralisan je dodavanjem acetil grupe. Nedostajući vodonici dodati su u strukturu, pri čemu je stanje jonizacije bočnih lanaca aminokiselina prilagođeno njihovim dominantnim oblicima na pH 8,0 u skladu sa pKa predviđanjima dobijenim sa PROPKA3 [308] servera. Preciznije, svi aspartati i glutamati u strukturi su deprotonovani, arginini i lizini protonovani a histidini monoprotovani. Na osnovu vizuelne analize okruženja histidina, u smislu potencijala formiranja vodoničnih veza, H613 je protonovan na δ -azotu, dok su H689 i H706 ε -protonovani. Kompleks je zatim solvatisan, neutralisan i parametrizovan pomoću Tleap modula AmberTools 12 [309] paketa programa. Upotrebljen je kvadar prethodno ekvilibrisane vode dimenzija odabranih tako da sa svake strane kompleksa bude zastupljen sloj rastvarača od najmanje 10 Å. Sistem je neutralisan dodatkom jednog Na⁺ jona. Proteinski deo strukture parametrizovan je Amber ff99SB poljem sila [310], za vodu su korišćeni parametri TIP3P modela [311], dok su parametri za SAH preuzeti iz literature [312].

Pripremljeni kompleks zatim je minimizovan metodom konjugovanih pravaca do konvergencije ukupne potencijalne energije sistema u dve faze. U prvoj fazi, 5.000 koraka minimizacije izvršeno je fiksiranjem položaja proteina i proizvoda kofaktora, tako da su samo molekule vode i jona mogle da menjaju položaje. U drugoj fazi, tokom narednih 5.000 koraka minimizacije, položaji teških atoma proteina i proizvoda kofaktora harmonijski su bili vezani za svoje početne položaje korišćenjem konstante sile od 10 kcal mol⁻¹ Å⁻². Sistem je zatim zagrevan od 0 do 298 K, podizanjem temperature simulacije za 1 K svakih 0,3 ps, pri čemu je harmonijsko ograničavanje fluktuacije teških atoma bilo aktivno i tokom faze zagrevanja. Nakon postizanja temperature od 298 K, sistem je ekvilibrisan kratkom simulacijom u trajanju od 2 ns tokom kojih su pokreti post-SET domena bili slobodni, dok je teškim atomima ostatka proteina i proizvodu kofaktora nametnuto blago harmonijsko ograničenje, korišćenjem konstante sile od 1 kcal mol⁻¹ Å⁻². Nakon ekvilibracije, kratka simulacija od 1 ns sprovedena je u cilju prikupljanja podataka neophodnih za definisanje parametara simulacije ubrzane MD (aMD).

Algoritam ubrzane MD (aMD) [313], implementiran u programu NAMD [314], iskorišćen je u cilju efikasnog istraživanja faznog prostora dostupnog post-SET domenu. Tokom aMD simulacija, strukture SET domena i kofaktora održavane su blizu početnim, eksperimentalno utvrđenim koordinatama kroz upotrebu blagog

harmonijskog ograničenja konstante sile od $1 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. Simulacija aMD sprovedena je u *dual boost* režimu [315], a parametri ubrzanja odabrani su na osnovu preporuka iz literature za solvatisane proteine parametrizovane Amber poljem sila [316, 317]. Korišćene su sledeće formule:

$$E_{dih} = \langle V_{dih} \rangle + 4 \cdot n_{res} \quad (3.1)$$

$$\alpha_{dih} = \frac{4}{5} \cdot n_{res} \quad (3.2)$$

$$E_{total} = \langle V_{total} \rangle + \frac{n_{res}}{5} \quad (3.3)$$

$$\alpha_{total} = \frac{n_{atom}}{5} \quad (3.4)$$

gde su E_{dih} i α_{dih} parametri ubrzanja diedralnih uglova a E_{total} i α_{total} parametri ubrzanja celog sistema, određeni na osnovu srednjih vrednosti potencijalne energije diedralnih uglova $\langle V_{dih} \rangle$ i ukupne potencijalne energije sistema $\langle V_{total} \rangle$ koji sadrži n_{res} proteinskih rezidua, odnosno ukupno n_{atom} atoma. Srednje vrednosti potencijalne energije procenjene su iz kratke simulacije (1 ns) ekvilibrisanog sistema. Primenom navedenih formula, parametri ubrzanja podešeni su na sledeće vrednosti: $E_{dih} = 1.954 \text{ kcal mol}^{-1}$, $\alpha_{dih} = 109 \text{ kcal mol}^{-1}$, $E_{total} = -60.509 \text{ kcal mol}^{-1}$ i $\alpha_{total} = 5.300 \text{ kcal mol}^{-1}$.

Simulacija ubrzane molekularne dinamike ostvarena je kroz 12,5 miliona koraka, uz korišćenje vremenskog koraka od 2 fs u integraljenju jednačina kretanja. Treba imati u vidu da zbog modifikovanog oblika potencijala, istraživanje faznog prostora tokom ove simulacije ne odgovara onom koje bi bilo moguće uočiti za 25 ns konvencionalne MD i da je određivanje vremenskog ekvivalenta generalno složeno [318]. Iz trajektorije aMD simulacije zatim su izdvojene konfiguracije sistema koje predstavljaju energetski najstabilnija stanja, uzimajući u obzir 30% trajektorije u kojem je modifikacija potencijala bila najizraženija. Ovaj podset trajektorije grupisan je prema RMSD vrednostima kostura proteina i klaster sa najvećim brojem članova u kojem je post-SET domen bio u kontaktu sa SAH uzet je kao reprezentativna struktura za dalje istraživanje.

U cilju generisanja većeg broja konformacija vezivnog mesta kofaktora, koje potencijalno predstavljaju konformacije relevantne za vezivanje inhibitora, reprezentativna struktura iz aMD simulacije dalje je simulirana konvencionalnom MD, slično metodi relaksiranog kompleksa [319]. Iz reprezentativne strukture uklonjen je SAH i

sprovedena je simulacija od 15 ns apo oblika enzima. Kako bi se izdvojile strukture EZH2 pogodne za doking inhibitora u pretpostavljeno vezivno mesto, izvršeno je grupisanje kadrova trajektorije: prvo na osnovu RMSD vrednosti kostura proteina, a zatim na osnovu vrednosti RMSD teških atoma aminokiselina koje formiraju vezivno mesto kofaktora. U reprezentativne strukture svakog klastera sa više od 150 članova izvršen je doking piridonskih inhibitora da bi konačno bila odabrana ona struktura pomoću koje je bilo moguće najbolje reprodukovati poze inhibitora koje bi bile očekivane na osnovu dostupnih eksperimentalnih podataka. Stereohemijska validnost odabrane reprezentativne strukture, nakon energetske minimizacije, proverena je pomoću PROCHECK programa [320] na SAVS v3 serveru.

3.1.3 Poređenje sličnosti vezivnih mesta PKMT

Struktura kofaktor-vezivnog mesta u razvijenom modelu EZH2 upoređena je sa drugim PKMT koje imaju SET domen a čije su strukture dostupne u PDB-u. U poređenje su uključeni sledeći enzimi: G9a (PDB unos 2O8J), GLP (PDB unos 3HNA), MLL (PDB unos 2W5Z), SETD7 (PDB unos 3M53), SETD8 (PDB unos 1ZKK), SETD2 (PDB unos 4H12), SETMAR (PDB unos 3BO5) i Suv39H2 (PDB unos 2R3A). Kofaktor-vezivno mesto svake od struktura definisano je prema sc-PDB definiciji [321] kao skup svih aminokiselina koje se nalaze na udaljenosti $\leq 6,5$ Å od bilo kog teškog atoma SAM ili SAH kokristalisanog sa datim enzimom. Poređenje sličnosti vezivnih mesta izvršeno je primenom FuzCav metode [322]. FuzCav je metoda koja ne zavisi od superponiranja poređenih struktura, već se hemijsko okruženje svakog vezivnog mesta opisuje „otiskom prsta” u vidu binarnog niza dužine 4.833 bitova. Otisak prsta kodira se tako što se svakoj aminokiselini pridružuje jedna ili više farmakofornih karakteristika (npr. aromatičan centar fenilalaninu, akceptor vodonične veze i negativan centar aspartatu itd.), nakon čega se u fiksnim udaljenostima između $C\alpha$ atoma utvrđuje prisustvo ili odsustvo jedinstvenih farmakofornih tripleta, svaki od kojih definišu tri farmakoforne tačke i tri rastojanja između njih. Pored poređenja sličnosti celovitih vezivnih mesta, pripremljen je i dodatan set struktura iz kojih su uklonjene rezidue koje pripadaju post-SET domenu (sve aminokiseline C-terminalno od očuvanog tirozina Y726 u strukturi EZH2). Ovo je učinjeno sa ciljem utvrđivanja doprinosa post-SET domena diverzitetu vezivnog mesta kofaktora. Važno je naglasiti da sama veličina vezivnog mesta ne utiče na procenu sličnosti baziranu na FuzCav otiscima prstiju [322].

3.1.4 Simulacije MD EZH2 u kompleksu sa piridonskim inhibitorima

Molekulski doking

Početne strukture EZH2 u kompleksu sa 13 piridonskih inhibitora koji su obuhvaćeni ovim delom studije (slika 1.10 i tabela 1.3) pripremljene su molekulskim dokingom korišćenjem programa AutoDock Vina [323]. Stanje jonizacije svakog od inhibitora odabrano je prema očekivanom dominantnom obliku pri pH 8,0 u skladu sa predviđanjima Calculator Plugin-a Marvin 6.0.3 paketa [324]. Preciznije, uzeto je da se svi inhibitori dominantno nalaze u neutralnom obliku izuzev jedinjenja **126** koje je protonovano na piperidinskom azotu. Oblast pretrage doking algoritma (engl. *search box*) centrirana je na pretpostavljeno mesto kofaktora u reprezentativnoj strukturi EZH2 i dimenzionisana tako da se proteže 20 Å u svakom pravcu. Iscrpnost pretrage (engl. *exhaustiveness*) podešena je na 16, pri čemu je sačuvano 15 najboljih poza u cilju vizuelne identifikacije optimalne. Optimalnom pozom je smatrana ona koja je u najvećoj meri bila u skladu sa očekivanom orijentacijom inhibitora (diskutovano u uvodnom odeljku i ilustrovano na slici 1.11), bez obzira na energiju vezivanja procenjenu dokingom.

Parametrizacija liganada

Ligandi su parametrizovani za dalje simulacije GAFF (engl. *general Amber forcefield*) poljem sila [325], verzija 1.4. GAFF-kompatibilni tipovi dodeljeni su atomima liganada automatski, korišćenjem Antechamber programa [326], dok su parcijalna atomska naelektrisanja izvedena iz elektrostatskog potencijala izračunatog na HF/6-31G* nivou teorije, primenom RESP postupka uklapanja [327]. Korišćena je RESP implementacija programa R.E.D. Tools IV [328] na R.E.D. Serveru [329]. U dobijanju parcijalnih atomskih naelektrisanja korišćena su tri koformera svakog inhibitora, čije su ravnotežne geometrije optimizovane na HF/6-31G* nivou teorije pomoću programa Orca 2.9 [330], pre slanja na R.E.D. Server. Finalni kvantno-mehanički proračuni na serveru obavljani su u programu Gaussian09 [331].

Simulacije MD

Svaki od 13 kompleksa pripremljen je za simulacije sledećim postupkom. Sistem je solvatisan korišćenjem kvadra TIP3P vode, dimenzija odabranih tako da sa svake strane kompleksa bude prisutan sloj od barem 10 Å rastvarača. Neutralizacija svih kompleksa izvršena je dodatkom jednog Na⁺ jona, izuzev EZH2–126 kompleksa, gde neutralizacija nije bila neophodna. Proteinski deo kompleksa parametrizovan je Amber ff99SB poljem sila. Pripremljeni sistemi u proseku su sadržali 24.500 atoma i imali su dimenzije 68 × 60 × 68 Å.

Minimizacija energije sistema obavljena je kroz 10.000 koraka algoritma konjugovanih pravaca, pri čemu su tokom prvih 5.000 koraka protein i ligand bili fiksirani, dok su tokom narednih 5.000 koraka njihovi položaji bili harmonijski vezani za početne koordinate uz konstantu sile od 10 kcal mol⁻¹ Å⁻². Zagrevanje sistema izvršeno je postepeno, povećanjem temperature simulacije od 0 do 298 K, tempom od 1 K na svakih 0,3 ps. Tokom zagrevanja, teški atomi proteina i liganda harmonijski su bili vezani za početne položaje uz konstantu sile od 1 kcal mol⁻¹ Å⁻². Nakon zagrevanja, ovo blago harmonijsko ograničenje primenjivano je na atome SET domena dodatnih 1 ns, dok su post-SET domen i ligand bili slobodni. Početnih 1 ns ekvibracije nastavljeno je bez ikakvih ograničenja na pokrete sistema barem 4 ns u izotermalno-izobarskom ansamblu, pri čemu je ekvibracija sistema praćena kroz RMSD vrednosti koordinata kostura proteina, teških atoma vezivnog mesta odnosno teških atoma liganada. Ukoliko bi sistem postigao zadovoljavajuću ekvilibrisanost nakon ove 4 ns, simulacija bi bila nastavljena dodatne 4 ns (produkciona faza) u izotermalno-izobarskom ansamblu, pri čemu je ova trajektorija korišćenja u MM/GBSA proračunima. U suprotnom, ekvibracija je nastavljena do stabilizovanja fluktuacija RMSD vrednosti pre započinjanja simulacije od 4 ns koja je bila upotrebljena za procenu slobodne energije vezivanja. U slučaju svakog od kompleksa, nakon završetka prve produkcione simulacije od 4 ns, brzine atoma u sistemu ponovo su inicijalizovane i sprovedeno je dodatnih 10 ns simulacije u cilju procene efekata dužeg trajanja na proračun slobodne energije vezivanja i stabilnost poze inhibitora u vezivnom mestu.

Dodatan set simulacija sproveden je sa modelom strukture EZH2 u kojem su prisutne mutacije određenih aminokiselina. Simulirani su EZH2^{A677G} u kompleksu sa inhibitorima **E4Z** i **E5Z**, EZH2^{Y641F} u kompleksu sa **E4Z** i **E5Z**, EZH2^{A687V} u kompleksu sa **E5Z**, kao i apo oblici EZH2^{Y641N} i EZH2^{Y641C}. Strukture svakog od mutiranih oblika EZH2 pripremljene su pomoću Mutagenesis Wizard funkcije

programa PyMOL [332] korišćenjem Dunbrack biblioteke rotamera [333] u izboru konformacije bočnog lanca mutirane aminokiseline. Prilikom zamene bočnog lanca, izbor rotamera izvršen je na način da izmenjena rezidua ne stupa u nedozvoljene kontakte sa okolnim aminokiselinama.

3.1.5 Opšti parametri MD simulacija i analiza trajektorija

Simulacije molekularne dinamike sprovedene su pomoću programa NAMD 2.8b3 [334]. Uslovi periodičnih granica (engl. *periodic boundary conditions*) postojali su u sve tri prostorne koordinate. Kretanje vodonika kovalentno vezanih za teške atome ograničavano je SHAKE algoritmom, izuzev u okviru molekula TIP3P vode, za koje je korišćen brži SETTLE algoritam. U integraljenju jednačina kretanja korišćen je vremenski korak od 2 fs, a isti vremenski korak korišćen je i u proceni van der Waals-ovih i elektrostatskih sila. Van der Waals-ove interakcije između atoma računane su do udaljenosti od 12 Å uz korišćenje analitičke korekcije u kompenzovanju efekata „skraćanja” Lennard-Jones-ovog potencijala. Elektrostatske interakcije procenjivane su pomoću *particle-mesh Ewald* metode, računanjem doprinosa u realnom prostoru do rastojanja od 12 Å i korišćenjem mreže jedinične dimenzije od 1 Å u recipročnom prostoru. Temperatura u simulacijama održavana je konstantnom pomoću Langevin-ove dinamike uz koeficijent prigušenja (engl. *damping coefficient*) od 2 ps⁻¹. U simulacijama sa konstantnim pritiskom, korišćen je Nose-Hoover-ov klip sa periodom oscilacije od 500 fs i prigušenjem tokom 200 fs, podešen na ciljnu vrednost pritiska od 1,01 bar.

Vizuelizacija i analiza trajektorija izvršena je pomoću programa VMD [335], gcrarma [336] i MDpocket [337]. MDpocket iskorišćen je u proceni uticaja mutacija EZH2 na veličinu i druge osobine vezivnog mesta za kofaktor. U ovoj analizi korišćeno je 500 podjednako vremenski udaljenih kadrova iz trajektorija trajanja 4 ns, korišćenih i u MM/GBSA proračunima. Pre analize, strukture EZH2 iz ovih kadrova superponirane su minimizacijom RMSD koordinata kostura proteina pomoću ptraj programa iz AmberTools paketa. Detekcija džepova u proteinu koju MDpocket vrši zasniva se na proceduri prvobitno opisanoj u sklopu Fpocket metode [338]. Ova procedura zasniva se na postavljanju alfa-sfera u protein, čiji minimalni i maksimalni poluprečnik definiše korisnik. Površina alfa-sfere istovremeno mora da dodiruje četiri atoma proteina, bez uključivanja ijednog atoma proteina u svojoj unutrašnjosti. Nakon postavljanja, svaka alfa-sfera pridružuje se najbližoj tački preseka virtuelne rešetke jedinične dimenzije 1 Å, koja je superponirana na prvi kadar

analizirane trajektorije. Broj alfa-sfera pridruženih svakoj tački preseka broji se za svaki kadar trajektorije i konačno normalizuje brojem kadrova čime se dobija mapa gustina alfa-sfera. U ovoj studiji, minimalni poluprečnik sfere od 3 Å, maksimalan prečnik od 6 Å i mapa izogustine sa izovrednošću od 8 korišćeni su za definisanje džepova jer je ovaj izbor parametara omogućio dobro razlikovanje mesta za vezivanje kofaktora od drugih šupljina u proteinu. Izovrednost od 8 podrazumeva da svaka tačka preseka virtuelne rešetke u okolnom prostoru od 8 Å³ sadrži centre barem 3 izosfere. Spajanjem mreže takvih tačaka, dobijena je kontura vezivnog mesta u proteinu.

3.1.6 MM/GBSA proračuni i analiza efekata mutageneze u alanin

MM/GBSA metoda [339] korišćen je za procenu relativnih energija vezivanja ($\Delta G_{MM/GBSA}$) 13 piridonskih inhibitora za EZH2. Proračuni su izvršeni pomoću MMPBSA.py programa [340] iz AmberTools 12 paketa, pristupom koji koristi samo trajektoriju simulacije kompleksa. Slobodna energija procenjena je iz 500 vremenski podjednako udaljenih kadrova trajektorije trajanja 4 ns, pri čemu je entropijska komponenta $\Delta G_{MM/GBSA}$ zanemarena, pod pretpostavkom da je ovaj doprinos slobodnoj energiji vezivanja približno isti za sve analizirane ligande. U izračunavanju slobodne energije solvatacije korišćen je modifikovani generalizovani Bornov model GB^{OBCI} [341], dok je, u proceni doprinosa nepolarnih interakcija, za izračunavanje površine dostupne rastvaraču korišćen LCPO metod [342].

Doprinos pojedinačnih aminokiselina stabilizaciji liganada u vezivnom mestu procenjen je pomoću računarskog „alaninskog skeniranja” (engl. *alanine scanning*) implementiranog u MMPBSA.py programu. U okviru ove metode, odabrane aminokiseline prevode se u alanin a zatim se ponovo vrše MM/GBSA proračuni u cilju kvantifikovanja energetskog doprinosa datog bočnog lanca energiji vezivanja liganada. Kako bi ova analiza mogla da koristi postojeće MD trajektorije, pretpostavlja se da mutacija veće aminokiseline u alanin ne bi trebalo značajnije da utiče na celokupnu konformaciju proteina [340]. U literaturi je pokazano da je ova pretpostavka validna na većem broju sistema, odnosno da računarsko „alaninsko skeniranje” može kvalitativno da reprodukuje eksperimentalne rezultate ovakve mutacije [343]. U sklopu analize mutirane su sledeće aminokiseline vezivnog mesta: V621, W624, Y661, C663, F665, N688, S690, F723, Y726, Y728, I739, E742 i M743. Promena u slobodnoj energiji vezivanja liganada nakon mutacije definisana je izrazom:

$$\Delta\Delta G_{MM/GBSA} = \Delta G_{MM/GBSA}^{mutant} - \Delta G_{MM/GBSA}^{WT} \quad (3.5)$$

Vrednosti $\Delta\Delta G_{MM/GBSA}$ su relativnog karaktera i međusobno uporedive samo u okviru istog kompleksa, pa su prevashodno korisne u identifikovanju amino-kiselina koje daju ključan doprinos stabilizaciji liganada.

3.1.7 Analiza konformacionih preferenci odabranih inhibitora

Analiza konformacionih preferenci inhibitora **126** i **G3K** izvršena je korišćenjem *well-tempered* metadinamike [344, 345], implementirane u PLUMED kôdu [346] kojim je modifikovan izvorni kôd NAMD programa. Definisane su dve kolektivne varijable (engl. *collective variable*, CV) na koje je delovao metadinamikom izmenjen potencijal: odabrana su dva diedralna ugla koji odgovaraju rotaciji oko dve veze koje spajaju piridonski prsten za centralnim amidom. Modifikacija potencijala u prostoru obe CV bila je Gausovskog oblika, širine 0,35 rad, i dodavana je početnim tempom od 0,3 kcal mol⁻¹ ps⁻¹ koji je postepeno smanjivan tako da ukupan faktor otklona (engl. *bias factor*) bude 6 ($\Delta T = 1.490$ K).

Simulirani sistemi sastojali su se od izolovanog liganda solvatisanog u kocki TIP3P vode dimenzija prilagođenih tako da sa svake strane liganda bude bar 18 Å sloja rastvarača. Sistemi su postepeno zagrevani do 298 K, ekvilibrisani 100 ps u izotermalno-izobarskom ansamblu, i još 100 ps u kanonskom ansamblu. Simulacije metadinamike obavljene su u kanonskom ansamblu. Konvergencija simulacije praćena je kroz visinu deponovane modifikacije potencijala i razlike u energijama na rekonstruisanoj površini slobodne energije. Na osnovu ovih parametara, utvrđeno je da su simulacije konvergirale nakon 20 ns, kada su i okončane.

3.1.8 Razvoj farmakofornog modela EZH2 inhibitora

Farmakoforni model inhibitora EZH2 razvijen je na osnovu rezultata simulacija molekularne dinamike, uzimajući u obzir interakcije za koje je ustanovljeno da ključno doprinose energiji vezivanja poznatih piridonskih inhibitora. Za definisanje farmakofore i poređenje farmakoforne sličnosti korišćeni su Silicos-it Align-it [347] i Shape-it [348] programi. Align-it omogućava percepciju farmakofornih tačaka

u organskim molekulama baziranu na definicijama predloženim od strane Greene i saradnika [349], superponiranje struktura sa definisanom farmakofornom hipotezom i izračunavanje farmakoforne sličnosti [350]. Shape-it omogućava procenu sličnost oblika dve molekule, korišćenjem Gausovskih funkcija kao deskriptora oblika [351].

Razvoj farmakoforne hipoteze bio je baziran na reprezentativnim strukturama kompleksa EZH2 i inhibitora **G5K** odnosno **E5Z**. Iz svake strukture je u Pymol programu, pomoć Align-it dodatka, izdvojen skup farmakofornih tačaka liganda i sfera isključenja (engl. *exclusion spheres*) koje odgovaraju delovima vezivnog mesta koji su u neposrednoj blizini liganda. Farmakoforne tačke koje su korišćene uključuju: aromatične i hidrofobne centre, donore i akceptore vodonične veze, kao i pozitivno odnosno negativno nalektrisane funkcionalne grupe. Iz početnog skupa farmakofornih tačaka uklonjene su one za koje je, tokom simulacija inhibitora, ustanovljeno da nisu od ključnog značaja za energiju vezivanja inhibitora. Dodatno, iz reprezentativne strukture kompleksa, izdvojena je i struktura samog liganda, u odgovarajućoj konformaciji, koja je korišćena kao referentni oblik inhibitora u Shape-it programu.

Procena farmakoforne sličnosti ispitivane molekule sa datom farmakofornom hipotezom uključivala je nekoliko koraka. U prvom koraku, za svaku molekulu generisan je ansambl konformacija pomoću RDKit biblioteka korišćenjem pristupa sličnom protokolu ranije objavljenom u literaturi [352]. Ukratko, za molekule sa ≤ 7 rotabilnih veza generiše se 50, za 8 do 12 – 200, i za molekule sa > 12 rotabilnih veza 300 konformera. Početne geometrije konformera stohastički se uspostavljaju, uz poštovanje ograničenja u smislu međuatomskih rastojanja, što uslovljava raznovrsnost dobijenog ansambla. Geometrija svakog od konformera zatim je optimizovana u MMFF94s polju sila, implementiranom u odgovarajućem RDKit modulu [353], do konvergencije energije uz maksimalno 500 iteracija optimizacije. Konformeri iz tako generisanog ansambla prvo su superponirani na farmakofornu hipotezu i deset konformera sa najboljim rezultatom sačuvano je u zaseban fajl. Bez daljeg superponiranja, sličnost oblika sačuvanih konformera poređena je sa bioaktivnom konformacijom referentnog liganda. Srednja vrednost farmakoforne i sličnosti oblika uzeta je kao kombinovani rezultat, pri čemu je kao konačna mera farmakoforne sličnosti uzet najbolji rezultat bilo kog od konformera iz ansambla. Farmakoforna sličnost kvantifikovana je Tversky sličnošću kroz izraz:

$$Tversky_{pharmacophore} = \frac{V_O}{V_{ref}} \quad (3.6)$$

gde je V_O zapremina preseka farmakofornih tačaka date molekule i hipoteze, a V_{ref} zapremina farmakofornih tačaka same hipoteze. Sličnost oblika takođe je kvantifikovana Tversky sličnošću, kroz izraz:

$$Tversky_{shape} = \frac{V_O}{0,95V_{ref} + 0,05V_{query}} \quad (3.7)$$

gde su V_O , V_{ref} i V_{query} redom, zapremine preseka, referentne i poredene molekule, izražene Gausovskim funkcijama.

Validacija uspostavljene farmakoforne hipoteze izvršena je korišćenjem seta poznatih piridonskih inhibitora i kolekcije molekula „mamaca” (engl. *decoys*). Set aktivnih jedinjenja obuhvatio je 13 piridonskih inhibitora (slika 1.10 i tabela 1.3) proučavanih molekularnom dinamikom, piridonske inhibitore **CPI-169**, **CPI-360** i **EPZ-011989** (slika 1.12), 6 tetrametilpiperidina aktivnih u submikromolarnom opsegu (**CP36**, **CP37**, **CP44**, **CP49**, **CP51**, **CP52**, slika 1.13 i tabela 1.4), kao i 6 heksahidroizohinolinskih inhibitora (**ZS1–ZS6**, slika 1.14 i tabela 1.5). Za svako aktivno jedinjenje, pronađeno je 100 „mamaca” iz ZINC baze podataka [354] koji imaju slične fizičko-hemijske karakteristike kao dato aktivno jedinjenje, ali su topološki različita od njega. Fizičko-hemijska sličnost definisana je kroz molekulska masu (MW), izračunatu lipofilnost (ClogP), broj akceptora protona (nHBA) i broj donora (nHBD) vodonične veze, kao i broj rotabilnih veza u molekuli (nRB). Da bi jedinjenje bilo prihvaćeno kao „mamaca” moralo je da ima $MW \pm 35$, $ClogP \pm 0,5$, $nHBA \pm 1$, $nHBD \pm 1$ i $nRB \pm 2$ u odnosu na odgovarajuće aktivno jedinjenje. Istovremeno, Tanimoto sličnost MACCS otisaka prstiju [355] (implementiranih u RDKit bibliotekama) potencijalnog „mamaca” i aktivne molekule nije smela da bude veća od 0,75. Generisanje kolekcije „mamaca” izvršeno je pomoću Python skripta uz korišćenje ZINC *all purchasable* baze podataka, verzija iz februara 2015. godine.

Performanse definisanog farmakofornog modela i opisanog pristupa proceni farmakoforne sličnosti u identifikovanju 28 aktivnih jedinjenja u prisustvu 2.800 „mamaca” procenjivane su kroz 3 mere. Kriva ROC (engl. *Receiver Operating Characteristic*) dobija se praćenjem udela stvarno pozitivnih aktivnih jedinjenja u funkciji udela lažno pozitivnih jedinjenja na odabranim intervalima liste svih jedinjenja rangiranih po sličnosti u odnosu na farmakofornu hipotezu. Površina ispod ROC krive

(AUROC) je neparametarska metrika koja uzima vrednosti u intervalu $[0, 1]$, pri čemu vrednost od 0,5 odražava performanse nasumično rangirane liste jedinjenja. AUROC vrednost računata je iz 100 odabranih pozicija na rangiranoj listi pomoću formule:

$$AUROC = \frac{1}{n \cdot N} \sum_{i=2}^N A_i(I_i - I_{i-1}) \quad (3.8)$$

gde je n broj aktivnih jedinjenja u setu; N ukupan broj jedinjenja u setu; i pozicija na rangiranoj listi; A ukupan broj aktivnih jedinjenja na listi, a I broj neaktivnih jedinjenja („mamaca”) na listi – iznad date pozicije i .

Kako AUROC vrednost može biti visoka i onda kada većina aktivnih jedinjenja nije visoko na rangiranoj listi, u proceni tzv. „ranog obogaćenja” (engl. *early enrichment*) korišćeni su faktor obogaćenja (engl. *Enrichment Factor*, EF) i robustno početno obogaćenje (engl. *Robust Initial Enhancement*, RIE). EF predstavlja odnos broja aktivnih jedinjenja u određenom gornjem delu rangirane liste i broja aktivnih jedinjenja koja bi se tu pronalazila nasumičnim izborom. EF je parametarska metrika u smislu definisanja procenta liste (χ) u kom se faktor obogaćenja računa. Uobičajeni izbori su 1% i 5%. Minimalna vrednost EF je uvek nula, dok maksimalna zavisi od izbora vrednosti χ . U ovoj studiji računat je faktor obogaćenja na 5% liste ($\chi = 0,05 < n/N$), pa je maksimalna vrednost EF $1/\chi = 20$. EF računat je formulom:

$$EF(\chi) = \frac{\sum_{i=1}^n \delta(r_i)}{\chi \cdot n} \quad (3.9)$$

gde $\delta(r_i)$ ima vrednost 1 ako je rang i -tog aktivnog jedinjenja iznad odabrane granice na listi ($r_i \leq \chi \cdot N$), odnosno 0 u suprotnom.

RIE ima sličnosti sa EF, ali se, umesto vrednosti od 1 za svako aktivno jedinjenje rangirano iznad crte, koristi faktor skaliranja α , koji uslovljava eksponencijalno smanjenje doprinosa niže rangiranih aktivnih jedinjenja ukupnoj RIE vrednosti. RIE je računat formulom:

$$RIE(\alpha) = \frac{N \sum_{i=1}^n e^{-\alpha r_i/N}}{n \frac{1-e^{-\alpha}}{e^{\alpha/N}-1}} \quad (3.10)$$

Minimalne i maksimalne vrednosti RIE zavise od izbora parametra α , kao i broja aktivnih jedinjenja i ukupnog broja jedinjenja u setu. Vrednost RIE se može skalirati pomoću RIE_{min} i RIE_{max} tako da uzima vrednosti iz intervala $[0, 1]$, čime se dobija BEDROC vrednost (engl. *Boltzmann-Enhanced Discrimination of ROC*). U ovom radu, BEDROC je računat uz $\alpha = 20$, korišćenjem formula:

$$RIE_{min}(\alpha) = \frac{N}{n} \frac{1 - e^{\alpha n/N}}{1 - e^{\alpha}} \quad (3.11)$$

$$RIE_{max}(\alpha) = \frac{N}{n} \frac{1 - e^{-\alpha n/N}}{1 - e^{-\alpha}} \quad (3.12)$$

$$BEDROC(\alpha) = \frac{RIE(\alpha) - RIE_{min}(\alpha)}{RIE_{max}(\alpha) - RIE_{min}(\alpha)} \quad (3.13)$$

Izrazi za računanje AUROC, EF i BEDROC preuzeti su iz reference [356].

3.2 Razvoj modela za predviđanje transporta lekova Pgp-om i BCRP-om

3.2.1 Formiranje setova podataka

Prilikom pripreme setova podataka koji su upotrebljeni u uspostavljanju računarskih modela za predviđanje interakcije sa Pgp-om i BCRP-om, usvojene su sledeće definicije supstrata i nesupstrata ova dva transportera. Supstratima su smatrana jedinjenja za koje postoje eksperimentalni dokazi o transcelularnom transportu specifično posredovanom datim transporterom. Nesupstratima su smatrana kako jedinjenja koja uopšte ne interaguju sa datim transporterom, tako i ona jedinjenja koja se vezuju za transporter ali nisu efikasno transportovana (ne postoji njihov značajan ćelijski efluks).

Kako bi se na uniforman način postupalo sa potencijalno suprotstavljenim

podacima iz literature, pre prikupljanja podataka uspostavljen je skup kriterijuma koji su korišćeni u proceni pouzdanosti dostupnih eksperimentalnih rezultata. Ovi kriterijumi zbirno su prikazani u tabeli 3.1. Za procenu pouzdanosti klasifikacije jedinjenja u Pgp supstrate i nesupstrate, kriterijumi prikazani u tabeli 3.1 usvojeni su objedinjavanjem ranije predloženih kriterijuma iz literature [242, 357, 358]. Za BCRP klasifikaciju, kriterijumi su uspostavljeni po analogiji sa Pgp-om, uzimajući u obzir poznata ograničenja testova koji se koriste za ispitivanje transporta ovim proteinom [359].

Tabela 3.1: Pregled kriterijuma koji su korišćeni u proceni pouzdanosti klasifikacije jedinjenja prilikom formiranja setova podataka, u zavisnosti od tipa i broja dostupnih eksperimentalnih dokaza

Nivo dokaza	Pgp testovi	BCRP testovi
A (najviši)	<p><i>Konsenzus rezultata:</i></p> <ol style="list-style-type: none"> 1. <i>in vitro</i> studije transcelularnog transporta sprovedene na Pgp-om transficiranim ćelijskim linijama (MDCK, LLC-PK1) i <i>in vivo</i> studije distribucije na transgenim životinjama bez i sa funkcionalnim Pgp-om; ili 2. tri studije: <i>in vitro</i> studije transcelularnog transporta, stimulacije ATP-azne aktivnosti i testa akumulacije sa kalcein-AM probom. 	<p><i>Konsenzus rezultata:</i></p> <ol style="list-style-type: none"> 1. <i>in vitro</i> studije transcelularnog transporta sprovedene na BCRP-om transficiranim ćelijskim linijama (MDCK, HEK293) i <i>in vivo</i> studije distribucije na transgenim životinjama bez i sa funkcionalnim BCRP-om; ili 2. tri studije: <i>in vitro</i> studije transcelularnog transporta, stimulacije ATP-azne aktivnosti i testa akumulacije sa BODIPY-prazosin probom.
B	<ol style="list-style-type: none"> 1. jedna ili više <i>in vitro</i> studija transcelularnog transporta sprovedena na Pgp-om transficiranim ćelijskim linijama, uz netransfektovane ćelije kao kontrolu; ili 2. konsenzus rezultata dve studije iz kategorije C, koje ne daju nivo dokaza A (npr. transport kroz Caco-2 ćelije i studija na transgenim miševima); ili 3. konsenzus rezultata dva <i>in vitro</i> testa iz kategorije A2. 	<ol style="list-style-type: none"> 1. jedna ili više <i>in vitro</i> studija transcelularnog transporta sprovedena na BCRP-om transficiranim ćelijskim linijama, uz netransfektovane ćelije kao kontrolu; ili 2. konsenzus rezultata dve studije iz kategorije C, koje ne daju nivo dokaza A (npr. transport kroz Caco-2 ćelije i studija na transgenim miševima).
C	<ol style="list-style-type: none"> 1. polarizovan transport kroz ćelijske linije koje prekomerno ekspiriraju Pgp, ali i druge transportere (npr. Caco-2); ili 2. stimulacija Pgp ATP-azne aktivnosti; ili 3. <i>in vivo</i> studija distribucije na transgenim životinjama bez i sa funkcionalnim Pgp-om; ili 4. studije akumulacije ili efluksa sprovedene korišćenjem membranskih vezikula sa ekspimiranim Pgp-om; ili 5. studije rezistencije na citotoksične efekte na ćelijskim linijama MDR fenotipa. 	<ol style="list-style-type: none"> 1. polarizovan transport kroz ćelijske linije koje prekomerno ekspiriraju BCRP, ali i druge transportere; ili 2. stimulacija BCRP ATP-azne aktivnosti; ili 3. <i>in vivo</i> studija distribucije na transgenim životinjama bez i sa funkcionalnim BCRP-om; ili 4. studije akumulacije ili efluksa sprovedene korišćenjem membranskih vezikula sa ekspimiranim BCRP-om; ili 5. testovi sa Pheophorbide A (PhA) probom; ili 6. studije rezistencije na citotoksične efekte na ćelijskim linijama MDR fenotipa.
D (najniži)	Podaci iz preglednih i stručnih radova, gde originalne studije nisu navedene ili dostupne.	

Prilikom razmatranja rezultata eksperimenata transcelularnog transporta, usled velikih razlika u literaturi nije bilo moguće usvojiti jedinstvenu graničnu vrednost za stepen efluksa (engl. *efflux ratio*) koja se smatra značajnom. Stoga su prihvatani zaključci navedeni u originalnim studijama. Ukoliko su za dato jedinjenje u literaturi pronađeni dokazi istog nivoa pouzdanosti koji upućuju na suprotne zaključke u pogledu klasifikacije jedinjenja kao supstrata ili nesupstrata – takva jedinjenja nisu uključena u finalni set podataka u cilju minimizacije rizika od korišćenja pogrešne klasifikacije.

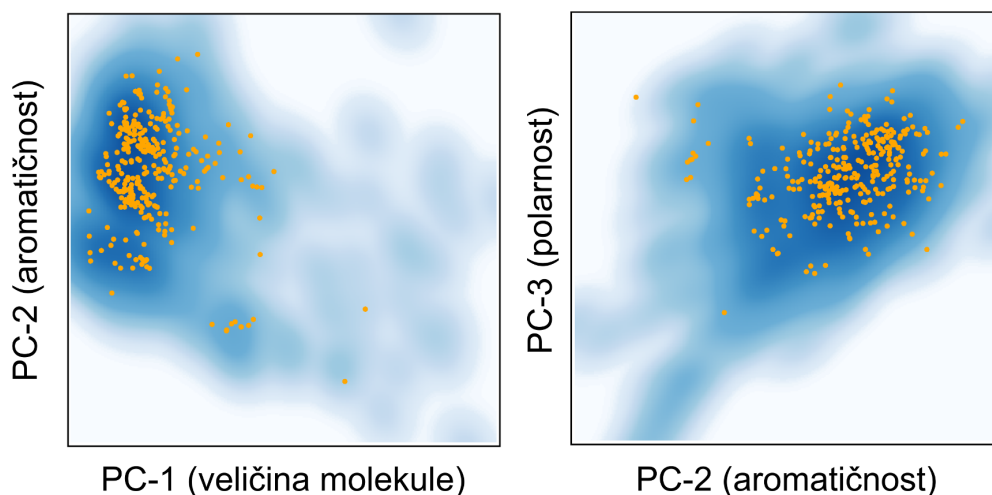
Prikupljanje podataka izvršeno je pretragom primarne i sekundarne literature posredstvom PubMed-a. Finalni setovi podataka sadržali su 310 jedinjenja za Pgp (159 supstrata i 151 nesupstrat) i 127 jedinjenja za BCRP (64 supstrata i 63 nesupstrata). Spisak jedinjenja, njihovi InChI ključevi, klasifikacija i PubMed identifikatori relevantnih referenci oba seta podataka prikazani su u Prilogu B, tabele B.1 i B.2.

3.2.2 Procena raznovrsnosti formiranih setova

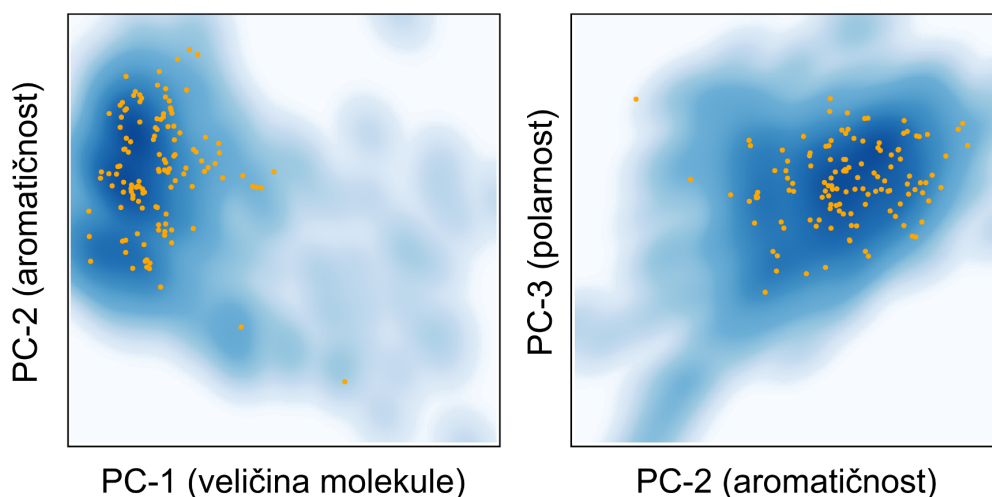
Pokrivenost relevantnog hemijskog prostora jedinjenjima koja su ušla u sastav oformljenih setova procenjena je korišćenjem DrugBank 3.0 [360] baze podataka kao referentne kolekcije *drug-like* jedinjenja. Iz polazne kolekcije uklonjena su jedinjenja neorganskog karaktera, a za preostale strukture izračunate su vrednosti glavnih komponentni (engl. *principal components*, PC) ChemGPS-NP sistema [361] posredstvom javno dostupne internet aplikacije. Distribucija jedinjenja procenjena je u prostoru definisanom pomoću prve tri glavne komponente ChemGPS-NP sistema, koje, redom, odražavaju veličinu, aromatičnost i polarnost jedinjenja. Raspodela jedinjenja u ovako definisanom hemijskom prostoru ilustrovana je na slici 3.2 i potvrđuje da pripremljeni setovi poseduju strukturnu raznovrsnost koja ih čini pogodnim za razvoj globalno–primenjivih modela.

3.2.3 Priprema struktura i podela setova

Strukture svih jedinjenja preuzete su iz PubChem baze i njihove geometrije su optimizovane AM1 semi-empirijskom metodom u programu Spartan [362]. Optimizovane trodimenzionalne strukture objedinjene su u jedan SDF fajl u koji su dodate i oznake pripadnosti jedinjenja supstratima (oznaka S), odnosno nesup-



(a) Set podataka za Pgp.



(b) Set podataka za BCRP.

Slika 3.2: Raspodela jedinjenja iz formiranih setova (označena na grafiku kao narandžaste tačke) u hemijskom prostoru definisanom glavnim komponentama ChemGPS-NP sistema. Osenčene oblasti odgovaraju raspodeli jedinjenja iz Drug-Bank 3.0 kolekcije, pri čemu tamnije regije odražavaju veću, a svetlije manju gustinu referentnih jedinjenja.

stratima (oznaka NS) datog transportera. Pripremljeni SDF fajlovi korišćeni su u programu ADMET PredictorTM 5.5 [363] za sve dalje proračune. U prvom koraku, ADMET ModelerTM modul programa ADMET Predictor upotrebljen je za izračunavanje preko 300 molekularskih deskriptora koji opisuju različita svojstva jedinjenja. U izračunavanju deskriptora čije vrednosti zavise od stanja jonizacije (pH-zavisni deskriptori), korišćena je vrednost pH od 7,4 i konstante jonizacije jedinjenja koje ADMET Predictor predviđa automatski.

Kako bi se obezbedilo da na performanse razvijenih modela minimalno utiče metoda podele seta u set za uspostavljanje (training set) i set za testiranje modela (test set), korišćena su tri različita pristupa podeli: manuelna podela, Kohonenove samoorganizujuće mape [364] (SOM) i *k-means clustering* [365]. Dok raspodela pomoću druge dve metode uzima u obzir samo hemijsku sličnost između jedinjenja, manuelnom podelom bilo je moguće osigurati i ravnomernu zastupljenost kako supstrata tako i nesupstrata u oba seta dobijena podelom. Korišćenjem svake od navedenih metoda, podela je izvršena tako da 75% jedinjenja bude u training setu a 25% u setu za testiranje modela. Poseban deo training seta bio je set za verifikaciju, koji je korišćen prilikom treniranja veštačkih neuronskih mreža u cilju pravovremenog zaustavljanja algoritma i izbegavanja gubitka sposobnosti generalizacije uspostavljenog modela. Veličina seta za verifikaciju bila je 25% od ukupne veličine seta.

Za manuelnu podelu setova korišćen je pristup sličan procedurama koje su prethodno opisane u literaturi [366, 367]. U prvom koraku, sa dostupnim vrednostima molekulskih deksriptora izvršena je analiza glavnih komponenti i odabrane su tri najznačajnije PC, svaka iz različite grupe tako da su glavne komponente redom odražavale: oblik i veličinu molekula, potencijal formiranja vodoničnih veza i jonizacione karakteristike jedinjenja. Nakon mapiranja jedinjenja u ovako definisanom trodimenzionalnom hemijskom prostoru pomoću Miner3DTM modula ADMET Predictor-a, izbor jedinjenja izvršen je vizuelnom selekcijom sa ciljem ravnomernog uzorkovanja u hemijskom prostoru pokrivenom setom u celini.

Podela zasnovana na SOM algoritmu izvršena je pomoću mape dimenzija 17×17 za Pgp set, odnosno 11×11 neurona za BCRP set. Mapiranje je izvršeno u dve faze: u prvoj fazi od 150 epoha, koeficijenti neurona podešavani su u okruženju od 100 neurona u susjedstvu pobedničkog, dok je u drugoj fazi od 100 epoha podešavanje vršeno samo u sklopu pobedničkog i njemu najbližeg neurona u prostoru vrednosti koeficijenata. Na kraju mapiranja, iz svakog od neurona mreže na koji je mapirano više od jednog jedinjenja izdvojene su komponente u training set. Ovaj postupak je ponavljan sve dok na mreži nije preostalo 25% početnog broja jedinjenja, koja su uzeta kao set za testiranje modela.

Podela pomoću *k-means clustering* algoritma izvršena je takođe u dve faze. U prvoj fazi, broj klastera odgovarao je broju jedinjenja training seta, umanjen za veličinu seta za verifikaciju. Iz svakog od klastera, centroid je zatim premešten u training set. U drugoj fazi, broj klastera podešen je da odgovara broju jedinjenja

u setu za verifikaciju, pri čemu su centriodi premešteni u ovaj set, a sva preostala jedinjenja uzeta su kao set za testiranje modela.

3.2.4 Razvoj modela

Izbor najznačajnijih deskriptora

Prvobitan broj izračunatih deskriptora smanjen je: (i) uklanjanjem deskriptora koji su imali nedostajuće ili vrednosti jednake nuli za preko 10% seta; (ii) uklanjanjem deskriptora čiji je koeficijent varijacije bio $< 3\%$; (iii) uklanjanjem deskriptora koji su bili visoko korelisani sa drugim deskriptorima ($R > 0,8$), tako što je zadržan samo onaj deskriptor sa najvišom vrednošću Spearmanov-og koeficijenta korelacije u odnosu na zavisno promenjivu. Nakon inicijalne redukcije, u izboru najznačajnijih deskriptora korišćena su dva pristupa. U okviru prvog pristupa, na osnovu prethodnih studija koje su ukazale na značajnost određenih molekulskih karakteristika za predviđanje transporta Pgp-om i BCRP-om, manuelno je odabran manji broj deskriptora. Odabrani deskriptori opisivali su veličinu molekula, njihovu fleksibilnost, polarnost, lipofilnost, jonizacione karakteristike i potencijal formiranja vodoničnih veza, i njihov eventualni značaj jednostavno je moguće interpretirati u kontekstu fenomena transporta Pgp-om odnosno BCRP-om. U okviru drugog pristupa, svi deskriptori, bez prethodne manuelne selekcije, razmatrani su tokom izbora najznačajnijih.

Sam izbor najznačajnijih deskriptora izvršen je pomoću genetskog algoritma (GA) implementiranog u ADMET Predictor-u uz korišćenje metode najmanjih kvadrata sa kernel funkcijom (engl. *kernel partial least squares*, KPLS) kao regresione metode. Funkcija radijalne raspodele korišćena je kao KPLS kernel. Veličina populacije bila je dvostruko veća od ukupnog broja deskriptora iz kojih je vršena selekcija. Broj deskriptora (gena) u okviru jedne individue odgovarao je traženom broju deskriptora za dati model, pri čemu je GA-KPLS selekcija vršena iznova za svaki trenirani model. GA implementiran u ADMET Modeler-u predstavlja GA ravnotežnog stanja, tako da se nakon svake genetske operacije najlošije rešenje postojeće populacije zamenjuje novim rešenjem samo ukoliko je to novo rešenje prilagođenije i nije već zastupljeno u populaciji. Verovatnoća mutacije je u okviru implementacije podešena na 0,9 a verovatnoća ukrštanja na 0,1, i ovi parametri nisu varirani. Prilagođenost jedinki u populaciji procenjivana je unakrsnom validacijom nakon uspostavljanja KPLS modela za jedinjenja trening seta korišće-

njem date kombinacije deskriptora. Selekcija je vršena kao selekcija točka ruleta. Maksimalan broj GA koraka (korak predstavlja jednu genetsku operaciju) podešen je na 30.000, pri čemu bi selekcija bila ranije prekinuta u slučaju 100 uzastopnih neuspelih koraka (genetska operacija koja ne dovodi do nove individue), 500 odbačenih koraka (genetska operacija koja ne dovodi do prilagođenije jedinke) ili 1.000 koraka bez unapređenja prilagođenosti najprilagođenije jedinke u populaciji.

Uspostavljanje modela

Razvijene su dve grupe modela koji omogućavaju klasifikaciju jedinjenja kao supstrata ili nesupstrata Pgp-a, odnosno BCRP-a: jednu grupu činili su ansambli modela zasnovani na veštačkim neuronskim mrežama (engl. *artificial neural networks*, ANN) a drugu grupu ansambli modela zasnovani na *support vector machine* (SVM) klasifikacionim modelima. Svaki razvijeni ansambl karakterišu sledeći parametri: (i) broj pojedinačnih klasifikacionih modela objedinjenih u ansambl; (ii) broj deskriptora koji se nalaze u sklopu modela; i (iii) kod ANN-baziranih modela, broj neurona u skrivenom sloju mreže. Zarad jednostavnosti, u daljem tekstu će ansambli modela biti označavani samo kao modeli. Tokom njihovog razvoja, optimalan izbor parametara modela identifikovan je kroz dve faze njihovog variranja. U prvoj, eksploratornoj fazi, korišćeni su veliki intervali: broj deskriptora variran je u opsegu od 5 do 35 u razmacima od 5, a broj modela u ansamblu od 20 do 50 u koracima od 10. Najbolji izbor parametara iz prve faze dalje je optimizovan u drugoj fazi gde su vrednosti parametara varirane u opsegu ± 5 od prvobitne vrednosti, sa razmacima od 1. Iz dobijene matrice potencijalnih modela, odabran je najbolji za dalju validaciju. Specifični parametri korišćeni u uspostavljanju ANN i SVM modela detaljnije su opisani u narednim odeljcima.

ANN modeli Veštačke neuronske mreže implementirane u ADMET Modeler-u predstavljaju *backpropagation* tip ANN-a sa hiperboličnom tangencijalnom funkcijom aktivacije. Tokom razvoja testirane su različite arhitekture i broj ANN mreža u modelu. Broj deskriptora variran je između 5 i maksimalnog broja koji je dobijen deljenjem broja jedinjenja u trening setu faktorom šest. Broj neurona u ulaznom sloju mreže uvek je odgovarao odabranom broju deskriptora. Sve trenirane mreže sadržale su samo jedan skriveni sloj, a broj neurona u ovom sloju biran je iz opsega od 1 do 5. Optimalan broj ANN modela po ansamblu određen je kao što je prethodno opisano, s tim što je broj treniranih mreža uvek bio 5 puta veći od traženog

broja. Na ovaj način su u ansambl svaki put objedinjeni najbolji ANN modeli izabrani iz redundantnog broja ukupno treniranih mreža. One arhitekture mreža gde bi broj koeficijenata neurona ($(\text{broj neurona ulaznog sloja} \times \text{broj neurona skrivenog sloja}) + \text{broj neurona skrivenog sloja}$) prevazilazio 90% od ukupnog broja jedinjenja u trening setu – nisu razmatrane. Inicijalizovanje početnih vrednosti koeficijenata vršeno je pojedinačnim Monte Carlo simulacijama. Pre treniranja svake od mreža, izvršena je slučajna razmena delova trening seta i seta za verifikaciju modela, pri čemu set za verifikaciju nije korišćen u treniranju mreže već samo u praćenju njenih prediktivnih sposobnosti. Treniranje mreže bilo bi okončano u slučaju da njene prediktivne sposobnosti pokažu opadanje kroz 15 uzastopnih epoha algoritma. U izboru najboljih mreža koje su objedinjene u konačni ansambl korišćen je Youden indeks ($\text{senzitivnost} + \text{specifičnost} - 1$). Predviđanje klase jedinjenja pomoću oformljenog ansambla dobija se „glasanjem”: klasa (S/NS) koju predvidi većina ANN modela uzima se kao rezultat predviđanja ansambla.

SVM modeli ADMET Modeler koristi implementaciju ν -SVM algoritma opisanu u literaturi [368] uz funkciju radijalne raspodele (engl. *radial basis function*, RBF) kao kernel. Kao i prilikom uspostavljanja ANN modela, pre faze treninga delovi trening seta i seta za verifikaciju modela nasumično su razmenjivani. Optimalan izbor SVM hiperparametara utvrđivan je iscrpnim variranjem vrednosti konstante „mekane margine” C (engl. *soft-margin constant*), γ parametra RBF funkcije i ν parametra ν -SVM algoritma. Optimalan izbor hiperparametara identifikovan je prema performansama uspostavljenog modela u klasifikaciji jedinjenja seta za verifikaciju. Prilikom treniranja SVM modela, broj modela po ansamblu variran je između 20 i 40, a svaki put je uspostavljan redundantan broj (između 30 i 50) iz kog bi bili izabrani najbolji SVM modeli. Ovde treba naglasiti da, za razliku od ANN gde inicijalizacija početnih vrednosti koeficijenata neurona daje delom stohastički karakter algoritmu, SVM modeli uspostavljeni sa istim izborom hiperparametara uvek daju ista predviđanja, što čini uspostavljanje ansambla suvišnim. Međutim, kako se pre treninga svakog od SVM modela vrši preraspodela jedinjenja seta za verifikaciju i trening seta, manje razlike u izboru SVM hiperparametara i sastavu trening seta uslovljavaju diverzitet modela u ansamblu. Kao i kod ANN modela, finalno predviđanje ansambla dobija se „glasanjem”, pri čemu predviđanje svakog od pojedinačnih modela ima istu težinu.

3.2.5 Validacija modela

Najbolji ANN i SVM modeli izabrani su na osnovu vrednosti Youden indeksa za jedinjenja test seta, uzimajući u obzir i uravnoteženost performansi na jedinjenjima trening i test seta, kao i jednostavnost modela. Odabrani modeli dalje su validirani korišćenjem alternativno podeljenih setova, dobijenih korišćenjem SOM odnosno *k-means clustering* raspodele.

U eksternoj validaciji korišćeni su drugi setovi dostupni u literaturi. Za Pgp modele korišćen je set koji su objavili Levatić i saradnici [369] koji sadrži 934 jedinjenja prikupljenih iz NCI-DTP baze. Od ovih 934 jedinjenja, 23 su već bila zastupljena u setu oformljenom u sklopu ovog istraživanja. Dodatnih 50 jedinjenja uklonjeno je iz originalnog seta jer su njihove strukture sadržale nedefinisane R-grupe, retke elemente ili neuobičajena valentna stanja. Nakon uklanjanja ovih jedinjenja, konačan set korišćen u validaciji sadržao je 861 jedinjenje.

U validaciji BCRP transportnih modela korišćen je set podataka koji su objavili Zhong i saradnici [370], a koji je nakon uklanjanja već zastupljenih jedinjenja sadržao 94 komponente.

3.2.6 Definisanje domena primenjivosti modela

Domeni primenjivosti razvijenih modela definisani su u prostoru vrednosti deskriptora, na način koji je konzistentan sa domenima primenjivosti integrisanih modela ADMET Predictor-a. Korišćenjem ovog pristupa, za svaki deskriptor u datom modelu definiše se opseg vrednost od X_{min} do X_{max} , koji obuhvataju jedinjenja trening seta. Prilikom primene modela u klasifikaciji novih jedinjenja, ukoliko vrednost bilo kog od relevantnih deskriptora odstupa za više od 10% od definisanog intervala – smatra se da je takvo jedinjenje izvan domena primenjivosti modela.

3.3 Računarsko dizajniranje EZH2 inhibitora

3.3.1 Implementacija algoritma za *de novo* dizajn

U cilju računarskog dizajniranja potencijalnih inhibitora EZH2, razvijen je i implementiran algoritam koji će u daljem tekstu biti označavan akronimom Phabblen (engl. *Pharmacophore-based Building Block Assembler*). Phabblen predstavlja specijalizovanu implementaciju NSGA-II algoritma, opisanog u odeljku 1.5.2, koji omogućava generisanje struktura iz manjih fragmenata (engl. *building blocks*) uz istovremenu optimizaciju više ciljeva dizajna. Kako bi se NSGA-II prilagodio ovoj specifičnoj nameni, razvijena je odgovarajuća genetska reprezentacija molekula, koncipirani su genetski operatori mutacije i ukrštanja, i definisano više funkcija prilagodivosti, što je detaljnije opisano u odeljku 4.3.1. Ovde su navedeni samo tehnički detalji implementacije.

Osim gde to nije drugačije naglašeno, celokupna implementacija algoritma izvršena je u programskom jeziku Python 2, prevashodno korišćenjem funkcionalnosti RDKit biblioteka. Reakcije preuzete iz DOGS [279, 280] i AutoGrow [296, 298] algoritama (ukupno 59) predstavljene su SMARTS sintaksom hemijskih transformacija. Kao kolekcija fragmenata korišćena je Enamine Building Blocks kolekcija preuzeta iz ZINC baze podataka u februaru 2015. Iz početne kolekcije od 108.115 jedinjenja pomoću Silicos-it Filter-it [371] aplikacije uklonjena su sva jedinjenja koja (i) sadrže elemente koji nisu H, C, O, N, F, Cl, Br, I, S, P ili B; (ii) koja sadrže više od 7 rotabilnih veza; (iii) koja sadrže više od jednog cikličnog sistema odnosno više od dva prstena; (iv) koja sadrže prstene sa više od 7 atoma; (v) čija je molekulska masa manja od 50 ili veća od 300. Iz kolekcije su zatim uklonjene duplirane strukture kao i ona jedinjenja koje ne mogu da učestvuju ni u jednoj od 59 reakcija implementiranih u Phabblen-u. Konačna kolekcija korišćena u *de novo* dizajnu sadržala je 37.743 jedinstvena fragmenta. Farmakoforni otisak prsta izračunat je za po jedan konformer svakog fragmenta korišćenjem Align-it aplikacije za percepciju farmakofornih tačaka. Reaktivnost fragmenta, zapremina datog konformera i Balabanov J indeks izračunati su pomoću RDKit-a. Ove informacije za celu kolekciju sačuvane su u binarnom formatu i korišćene su u svakom sledećem izvršenju skripta.

Procena farmakoforne sličnosti generisanih molekula sa datom farmakofornom hipotezom vršena je u svakoj generaciji algoritma na isti način kao što je opisano u odeljku 3.1.8, procedurom korišćenom za validaciju uspostavljenog farma-

kofornog modela. Pre generisanja konformernog ansambla, stanje jonizacije svakog jedinjenja podešeno je korišćenjem skupa jednostavnih pravila koja su definisana SMARTS sintaksom transformacija (alifatični amini su protonovani, karboksilne kiseline deprotonovane, i sl.). Align-it i Shape-it pozivani su kao eksterne aplikacije iz Python procesa. Kako generisanje konformera i procena farmakoforne sličnosti predstavljaju računarski najzahtevnije delove algoritma koji se mogu nezavisnosti izvršavati za svako od rešenja u populaciji, ovaj deo skripta trivijalno je paralelizovan korišćenjem *concurrent* Python biblioteke.

U proceni topološke sličnosti između dizajniranih molekula i poznatih inhibitora EZH2 korišćena je RDKit implementacija Morganovih cirkularnih otisaka prstiju molekule [372]. Razmatrano je okruženje atoma poluprečnika do 3 susedne veze, pri čemu su u generisanju otisaka korišćena farmakoforna svojstva susednih atoma, a ne podstrukture direktno. Ovakav otisak prsta može se smatrati analognim sa FCFP6 otiscima prstiju [373]. Hešovana verzija otiska pretvorena je u binarni vektor dužine 2048 bitova, a kao metrika sličnosti korišćena je Tanimoto sličnost.

Za procenu sintetske dostupnosti dizajniranih molekula korišćena su pravila koja su predložili Allu i Oprea [374]. U okvirima ovog pristupa, svakoj strukturi se na osnovu njene veličine, topologije i prisustva određenih podstrukture može dodeliti SMCM vrednost (engl. *Synthetic and Molecular Complexity*), pri čemu rastuće SMCM vrednosti odgovaraju složenijim strukturama. Odgovarajuće podstrukture predstavljene su kao SMARTS nizovi.

3.3.2 Redukovanje kolekcije dizajniranih jedinjenja

Nakon generisanja kolekcije novodizajniranih jedinjenja kroz nekoliko izvršenja Phabblor algoritma, primenjeno je nekoliko koraka filtriranja u cilju identifikovanja manjeg broja jedinjenja za dalje proučavanje. Iz kolekcije su, pomoću Silicos-it Filter-it aplikacije prvo uklonjena jedinjenja čija je molekulska masa bila manja od 200 ili veća od 500, koja su imala više od 10 rotabilnih veza, više od 5 donora ili više od 10 akceptora vodonične veze. Nakon ovog koraka, iz kolekcije su uklonjena sva jedinjenja koja sadrže reaktivne funkcionalne grupe koje mogu biti uzrok nespecifičnog formiranja adukta sa proteinima [375]. Konačno, na kolekciju je primenjen skup podstrukturnih filtera za jedinjenja koja su često promiskuitetna i pojavljuju se kao aktivna u velikom broju biohemijskih testova (engl. *Pan Assay Interference Compounds*, PAINS) [376].

Nakon opisanih koraka filtriranja, izvršen je doking preostalih jedinjenja u model EZH2 i to u dve konformacije: onoj koja odgovara reprezentativnoj strukturi kompleksa sa **G5K** i onoj sa **E5Z**. Najbolji doking rezultat je zadržan i odgovarajući kompleksi su vizuelno proučavani. Na osnovu rezultata dokinga, predviđenog načina vezivanja i fizičko-hemijskih karakteristika, odabran je manji broj jedinjenja koja su zatim proučavana simulacijama molekularne dinamike i primenom proračuna slobodne energije vezivanja. Mogućnost da ove molekule predstavljaju supstrate za Pgp i/ili BCRP zatim je predviđena korišćenjem modela čiji je razvoj opisan u odeljku 3.2. U slučaju novodizajniranih molekula gde je pokušana optimizacija strukture u cilju smanjenja verovatnoće za interakciju sa ovim transporterima, korišćena je analiza senzitivnosti deskriptora, implementirana u ADMET Predictor-u. Analiza senzitivnosti deskriptora moguća je kod ANN klasifikacionih modela, i olakšava identifikovanje molekularnih karakteristika čija varijacija ima najizraženiji potencijal da izmeni predviđeni afinitet molekule za dati transporter.

3.3.3 Simulacije molekularne dinamike

Odabrane dizajnirane molekule simulirane su u kompleksu sa EZH2 u cilju pouzdanije procene mogućnosti njihove interakcije sa kofaktor-vezivnim mestom enzima. Početne strukture kompleksa dobijene su molekularnim dokingom, kao što je opisano u prethodnom odeljku. Za simulacije kompleksa korišćeno je Amber ff99SB-ILDN polje sila [377]. Parametrizacija liganada izvršena je na način opisan u odeljku 3.1.4, izuzev što je korišćena verzija 1.7 GAFF polja sila.

Za simulacije je korišćen program Gromacs [378], verzija 4.6.6. Jedinična ćelija simulacije imala je oblik dodekaedra, dimenzija definisanih tako da sa svake strane kompleksa bude sloj od barem 10 Å rastvarača. Minimizacija energije sistema izvršena je u dve faze metodom najstrmijeg spusta (engl. *steepest descent*). U prvoj fazi, koordinate svih teških atoma proteina i liganda bile su fiksirane, dok je u drugoj fazi celokupan sistem bio slobodan. Minimizacija je vršena do konvergencije sila ispod vrednosti od 100 KJ mol⁻¹ nm⁻². Nakon minimizacije, brzine čestica u sistemu inicijalizovane su u skladu sa Maksvelovom raspodelom na temperaturi simulacije od 298 K. Ekvilibracija sistema započeta je kratkom simulacijom u kanonskom ansamblu trajanja 100 ps, nakon čega je simulacija nastavljena u izotermalno-izobarskom ansamblu dodatnih 400 ps. Tokom termalizacije sistema za kontrolu temperature korišćen je Berendsenov termostat sa konstantom kuplovanja od 0,2 ps, kao i Berendsenov barostat sa izotropnim skaliranjem dimenzija sistema,

uz ciljni pritisak od 1 bara i konstantu kuplovanja od 1 ps. Atomi kompleksa i atomi rastvarača tretirani su kao dve odvojene grupe za potrebe kontrole temperature i pritiska. Tokom termalizacije, položaji atoma kostura proteina i liganda bili su vezani za svoje početne koordinate korišćenjem konstante sile od $1.000 \text{ KJ mol}^{-1} \text{ nm}^{-2}$. Nakon prvih 0,5 ns termalizacije, ekvilibracija sistema nastavljena je dodatnih 10 ns. Tokom ekvilibracije i u daljim simulacijama, za kontrolu temperature korišćen je algoritam skaliranja brzina (engl. *velocity rescale*), a za kontrolu pritiska Parrinello-Rahman-ov barostat. Nakon ekvilibracije, simulacije su nastavljene dodatnih 5 ns koje su tretirane kao produkciona faza; trajektorije ovih simulacija korišćene su u MM/GBSA proračunima.

Za integraljenje jednačina kretanja korišćen je vremenski korak od 2 fs, uz korišćenje LINCS algoritma za ograničavanje fluktuacija vodonikovih atoma, odnosno SETTLE algoritma za molekule TIP3P vode. Izračunavanje van der Waals-ovih i elektrostatičkih sila između parova atoma vršeno je do udaljenosti od 11 Å, pri čemu je potencijal „izmešten” tako da bude integral sile čija je vrednost jednaka nuli na odsečnoj udaljenosti. Energija sistema i doprinos pritisku analitički su korigovani u cilju kompenzacije skraćanja Lennard-Jones-ovog potencijala. Elektrostatičke interakcije procenjujane su pomoću *particle-mesh Ewald* metode, pri čemu je u recipročnom prostoru korišćena mreža jedinične dimenzije od 1,2 Å.

Procena energije vezivanja novodizajniranih molekula za EZH2 izvršena je MM/GBSA metodom pomoću MMPBSA.py programa iz AmberTools 14 paketa, koristeći isti pristup koji opisan u odeljku 3.1.6, izuzev što su proračuni izvršeni na 500 kadrova produkcionih trajektorija, koji su međusobno vremenski bili udaljeni 10 ps. Dodatno, umesto analize efekata mutageneze u alanin, u proceni pojedinačnog doprinosa aminokiselina vezivnog mesta stabilizaciji liganda – korišćena je metoda razlaganja energije po parovima (engl. *pairwise free energy decomposition*).

Poglavlje 4

Rezultati i diskusija

4.1 Simulacije molekularne dinamike EZH2

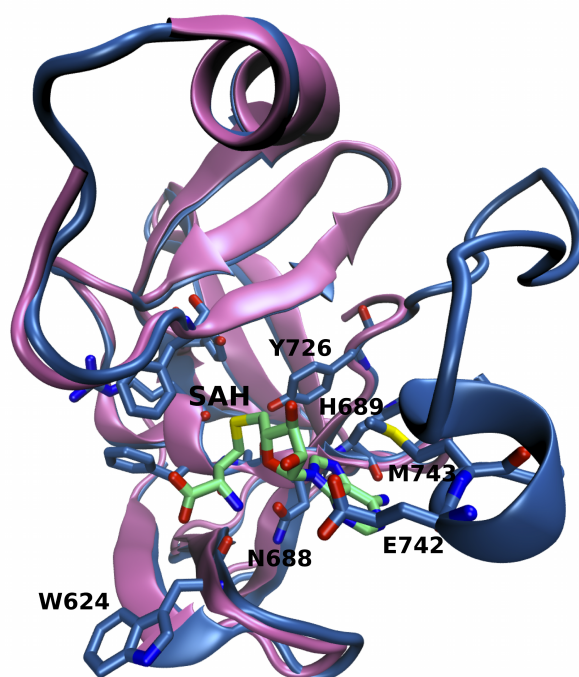
4.1.1 Modeliranje mesta za vezivanje kofaktora

EZH2 postaje katalitički kompetentna metiltransferaza tek nakon asocijacije sa SUZ12, EED i RBBP4/7 proteinima, koji zajedno formiraju funkcionalni, četvorokomponentni PRC2 kompleks (odjeljak 1.2.1). Kristalografske studije izolovane forme EZH2, skraćene na katalitički SET domen i susedni CXC domen, pokazale su da enzim zauzima konformaciju uobičajenu za PKMT sa SET domenom i da obrazuje vezivno mesto u koje se može uklopiti kofaktor SAM, u karakterističnoj konformaciji poznatoj iz velikog broja struktura srodnih enzima [125, 126]. Međutim, izotermalnom titracijom kalorimetrijom potvrđeno je da izolovana forma EZH2 ne vezuje kofaktor SAM niti poznate piridonske inhibitore enzima. Kao potencijalno objašnjenje ovog rezultata, pretpostavljeno je da vezivanje kofaktora i inhibitora zahteva struktuiranje i pravilno pozicioniranje post-SET domena. U dostupnim kristalnim strukturama EZH2, post-SET domen je neuređen a njegov početni deo pruža se u pravcu koji je suprotan od očekivanog položaja u kom bi ovaj deo strukture mogao da doprinese stabilizovanju kofaktora i supstrata (slika 1.5). Posledično, dostupne kristalne strukture ne mogu se direktno iskoristiti u proučavanju interakcije inhibitora sa EZH2. Ipak, najveći deo mesta za vezivanje kofaktora potpuno je formiran, a poređenje apo, binarnih i tercijarnih kompleksa drugih PKMT ukazuje da ovaj deo strukture ne podleže značajnijim promenama tokom vezivanja kofaktora i supstrata [67]. Stoga, dostupne kristalne strukture EZH2 mogu predstavljati

pouzdanu osnovu za razvoj modela kompletnog kofaktor-vezivnog mesta enzima.

U cilju dobijanja takvog modela, u ovom radu prvo je izvršeno modeliranje strukture i položaja post-SET domena EZH2, zasnovano na homologiji sa odgovarajućim delom strukture enzima SETD7 i SETD8, kao matricama. Post-SET domeni ova dva enzima, kao i EZH2, ne sadrže veći broj cisteina koji bi mogli da koordiniraju jon cinka tokom uspostavljanja tercijarne strukture. Uprkos činjenici da je post-SET domen EZH2 dugačak samo 19 aminokiselina, niska sličnost njegove sekvence sa korišćenim matricama, kao i inherentna fleksibilnost ovog dela strukture, ukazali su na potrebu da se početni model dalje unapredi. Ovo je postignuto primenom simulacija molekularne dinamike kroz razmatranje većeg broja fizički dostupnih struktura u cilju identifikovanja najverovatnije konformacije i položaja post-SET domena. Pri tome, usavršavanje modela bilo je bazirano na poznatim eksperimentalnim činjenicama, pre svega na tome da je konformacija kofaktora SAM, odnosno proizvoda kofaktora SAH, gotovo identična u svim strukturama PKMT. Fiksiranjem položaja i konformacije kofaktora u vezivnom mestu tokom simulacija molekularne dinamike, post-SET domenu je omogućeno da se prilagodi preostalom delu strukture bez udaljavanja SET domena od eksperimentalno određenih koordinata. Ubrzana molekularna dinamika (aMD) u *dual boost* režimu upotrebljena je kako bi se poboljšalo istraživanje faznog prostora, i to kroz istovremeno i nezavisno pospešivanje promena konformacije post-SET domena i njegovih difuzionih pokreta u rastvoru. Pri tome treba naglasiti da primena aMD i dalje obezbeđuje pravilno istraživanje svih minimuma slobodne energije sistema [313].

Kao rezultat simulacija aMD dobijena je verovatna struktura EZH2-SAH kompleksa. Na slici 4.1 prikazani su razvijeni model i početna kristalna struktura SET domena EZH2 (PDB unos 4MI0). Kao što se može uočiti, najveći deo vezivnog mesta razlikuje se veoma malo između dve strukture, ali se u razvijenom modelu post-SET domen nalazi u položaju gde zatvara vezivno mesto i doprinosi stabilizaciji SAH preko rezidua E742 i M743. U obe strukture, SAH može da formira mrežu vodoničnih veza između svog aminokiselinskog dela, W624 i N688, odnosno između nukleozidnog dela svoje strukture i peptidnog kostura H689. Ovaj skup vodoničnih veza karakteristično je očuvan kod gotovo svih PKMT sa SET domenom za koje su dostupne strukturne informacije [67]. Treba imati u vidu da dobijeni model i dalje ne odražava katalitički kompetentnan enzim, budući da je položaj I-SET domena i dalje takav da sprečava vezivanje histonskog supstrata. Ipak, uzimajući u obzir uočene interakcije između kofaktora i EZH2, kao i činjenicu da kofaktor-vezivno mesto ne podleže značajnijim promenama tokom vezivanja supstrata, može se zaključiti da

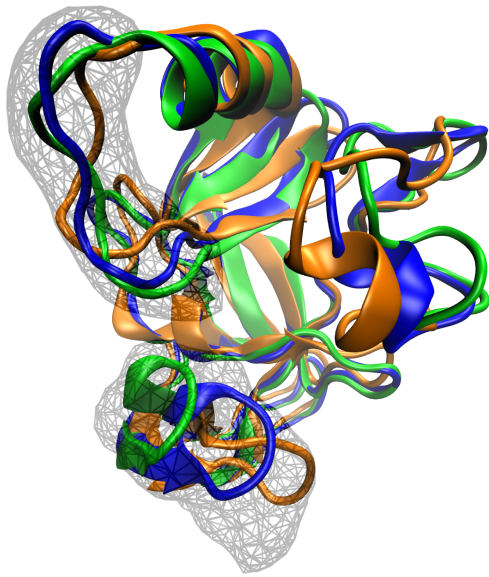


Slika 4.1: Prikaz reprezentativne strukture razvijenog modela EZH2 (obojen plavom) u kompleksu sa proizvodom kofaktora SAH (obojen zelenom). Model je superponiran na početnu kristalografsku strukturu SET domena enzima (obojena roze, PDB ID: 4MI0). Vodonici su uklonjeni iz strukture radi jasnijeg prikaza.

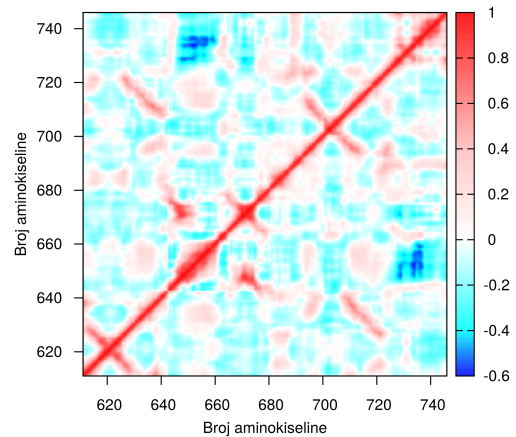
razvijeni model predstavlja dobru osnovu za izučavanje molekulske osnove inhibicije EZH2. U ovom kontekstu, važno je naglasiti i podatak da svi dostupni biohemijski dokazi upućuju na to da otkriveni piridonski inhibitori jesu kompetitivni samo sa kofaktorom, ne i sa supstratom enzima, te da je za razumevanje njihove interakcije sa EZH2 prevashodno važna struktura mesta za vezivanje kofaktora.

Počevši od reprezentativne strukture prikazane na slici 4.1, u daljem toku istraživanja sprovedena je konvencionalna simulacija MD apo oblika enzima trajanja 15 ns. Cilj ove simulacije nije bilo iscrpno istraživanje faznog prostora dostupnog apo formi enzima, jer bi time struktura značajno počela da odstupa od one koje odgovara binarnom EZH2–SAH kompleksu. Ovaj efekat nije bio poželjan, budući da je pretpostavljeno da se inhibitori vezuju za EZH2 delimično reprodukujući određene interakcije prisutne u binarnom kompleksu. Međutim, izvestan stepen relaksacije strukture bio je neophodan kako bi omogućio doking inhibitora u vezivno mesto. Dodatno, analiza dobijene trajektorije dala je uvid u stabilnost i dinamiku razvijenog modela.

Tokom trajanja simulacije, apo oblik enzima zadržao je stabilnu strukturu pri čemu su njegovi glavni pokreti bili ograničeni na otvaranje i zatvaranje ulaza



(a) Tri strukture EZH2 ilustruju varijabilnost njegovih konformacija tokom trajanja simulacije: narandžastom bojom prikazana je početna struktura, zelenom je obojena struktura sa najvećim odstupanjem od početne, dok je plavom bojom prikazana struktura reprezentativna za celu simulaciju, uzimajući u obzir koordinate C_α atoma. Mrežasto osenčeno područje predstavlja volumetrijsku mapu okupiranosti fleksibilnih aminokiselina na ulazu u kofaktor-vezivno mesto.

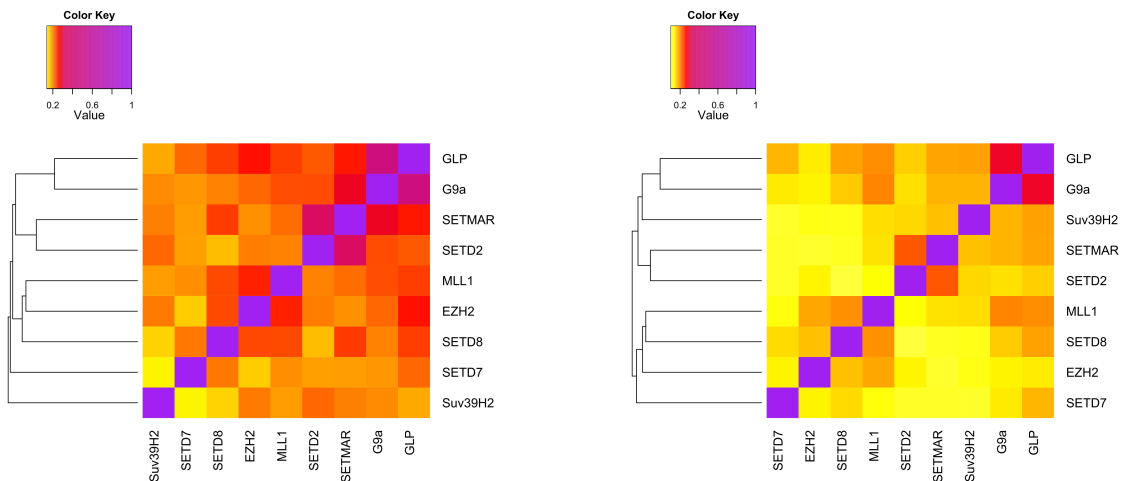


(b) Mapa dinamičke unakrsne korelacije fluktuacija C_α atoma u strukturi. Opseg boja odgovara stepenu korelacije: crvena područja odgovaraju parovima atoma čiji su pokreti visoko korelisani, plava odražavaju anti-korelisane fluktuacije, dok bela boja ukazuje na odsustvo korelacije u fluktuacijama odgovarajućeg para atoma.

Slika 4.2: Dinamika apo oblika EZH2, analizirana na osnovu MD trajektorije trajanja 15 ns.

u kofaktor-vezivno mesto. Varijabilnost konformacija EZH2 tokom trajanja simulacije ilustrovana je na slici 4.2a. Uočeno je da su najpokretljiviji delovi strukture post-SET domen i kratka petlja koja spaja heliks i antiparalelnu beta-ploču I-SET domena. Dodatno, značajnu pokretljivost je pokazao i deo enzima pri vrhu beta-ukosnice koja formira jednu stranu mesta za vezivanje kofaktora. Analiza unakrsne korelacije pokreta delova proteina (slika 4.2b) ukazala je na određene pravilnosti u dinamici enzima. Fluktuacije vrha spomenute beta-ukosnice (aminokiseline 620-623) delimično su anti-korelisane sa fluktuacijama C-terminalnih delova post-SET domena. Dodatno, nešto izraženiji stepen anti-korelisanih pokreta uočen je i između delova I-SET domena i delova post-SET domena koji su naspramno pozicionirani. Ovi pokreti verovatno odražavaju dinamiku enzima od značaja za vezivanje kofaktora i histonskog supstrata. U celini, analiza dinamike modela ovog dela strukture

EZH2 ukazuje na izraženu plastičnost mesta za vezivanje kofaktora i potvrđuje postojanje strukturne osnove za vezivanje malih organskih molekula. Konačno, treba naglasiti da bi u limitu mnogo duže simulacije esencijalna dinamika sistema verovatno odgovarala gubitku sekundarne strukture post-SET domena i njegovom udaljavanju od vezivnog mesta; u odsustvu kofaktora i histona, malo je verovatno da bi struktura zadržala ovaj stepen uredenosti.



(a) Strukture PKMT skraćene na SET domen.

(b) Strukture intaktnih PKMT, uključujući post-SET domen.

Slika 4.3: Sličnost vezivnih mesta za kofaktor u PKMT enzimima sa dostupnom kristalnom strukturom. Kao struktura EZH2 upotrebljen je razvijeni model. Stepenn sličnosti odgovara obojenosti datog polja, kao u legendi. Dendrogrami grupisanja struktura prikazani su u levom delu svake slike.

Takođe je analizirana i sličnost vezivnog mesta za kofaktor u razvijenom modelu EZH2 i kristalografski rešenim strukturama drugih PKMT sa SET domonom. Rezultate slične analize u okviru familije PKMT enzima prethodno su objavili Campagna-Slater i saradnici [210], ali u trenutku njihove studije, nijedna kristalna struktura EZH2 nije bila dostupna. Visok stepen sličnosti ortosternih mesta enzima iz iste familije može preliminarno da ukaže na rizik od promiskuitetnosti jedinjenja koja se vezuju za takva mesta u jednom od članova familije. U tom smislu, relevantan nivo sličnosti u velikoj meri zavisi od primenjene metode poređenja struktura vezivnih mesta. U okviru FuzCav metode, koja je korišćena za procenu sličnosti u ovom radu, rezultat $\geq 0,16$ ukazuje na značajan nivo sličnosti dva vezivna mesta [322]. Naša analiza obuhvatila je dva seta struktura: set PKMT skraćenih na SET domen i set PKMT sa celovitim vezivnim mestom, koje uključuje i pravilno pozicioniran post-SET domen. Ukoliko se razmotre samo delovi kofaktor-vezivnog mesta koje formira SET domen (slika 4.3a), izvesno je da je ovaj deo strukture izuzetno

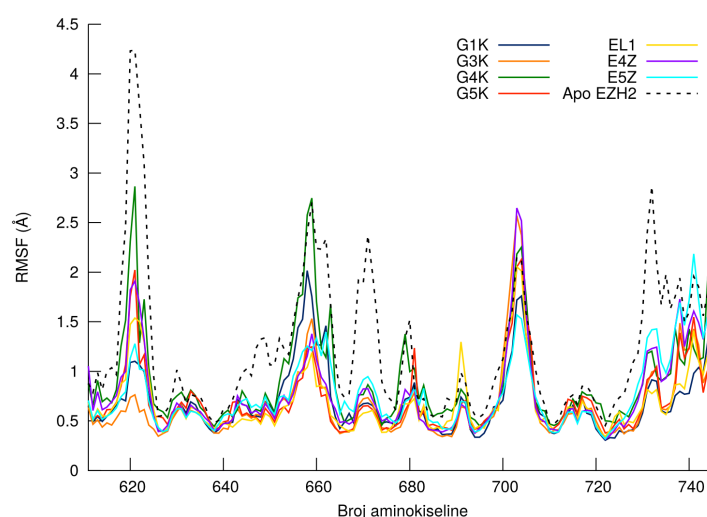
očuvan u okviru cele PKMT familije, pri čemu je FuzCav sličnost daleko iznad 0,16 za sve parove enzime, izuzev SETD7–Suv39H2 para. Međutim, kada se razmotre kompletna vezivna mesta, nivo sličnosti između PKMT je daleko niži (slika 4.3b), što ukazuje na to da post-SET domen ključno doprinosi njihovoj raznovrsnosti. Pri tome, interesantno je da dobijeni dendrogrami sličnosti vezivnih mesta dobro oslikavaju filogenetske veze između pojedinih članova PKMT familije, pretpostavljene ranije samo na osnovu analize višestrukog poravnanja njihovih primarnih sekvenci [67]. Kada je reč o EZH2, njemu dva najbližija vezivna mesta nalaze se u okviru MLL (FuzCav sličnost 0,18) i SETD8 (sličnost 0,16) enzima, dok je sličnost sa svim ostalim PKMT značajno niža. Iz ovih rezultata može se zaključiti da je raznovrsnost analiziranih vezivnih mesta u okviru PKMT familije verovatno dovoljno velika da omogući dizajn jedinjenja specifičnih za pojedinačne enzime, što je ujedno i potvrđeno profilom selektivnosti piridonskih inhibitora EZH2. Kako ključan faktor raznovrsnosti jesu aminokiseline post-SET domena, verovatno je da način vezivanja selektivnih inhibitora mora da oslikava ovu činjenicu kroz odgovarajuće, energetski povoljne kontakte sa ovim delom strukture.

4.1.2 Analiza interakcije piridonskih inhibitora i EZH2

Nakon simulacija apo oblika EZH2, kadrovi dobijene trajektorije grupisani su i izdvojeno je nekoliko raznovrsnih, reprezentativnih struktura koje su upotreb- ljene u molekulskom dokingu piridonskih inhibitora, koji je za cilj imao pripremu početnih struktura kompleksa za dalje simulacije MD. Ovakav pristup suštinski je sličan „metodi relaksiranog kompleksa” (engl. *relaxed complex scheme*) [319] i treba da omogući uključivanje dinamične prirode receptora u korak dokinga koji ga inače tretira kao rigidnu strukturu. Sve dostupne informacije o odnosima strukture i aktivnosti piridonskih inhibitora (odeljak 1.3.2, slika 1.11) uzete su u obzir tako da su iz većeg seta potencijalnih struktura kompleksa dobijenih ovim pristupom izdvojene one koje su pokazivale najbolje slaganje sa dostupnim eksperimentalnim podacima. Specifično, odabrane su poze inhibitora u kojima su piridonski prsten i centralni amid mogli da ostvaruju vodonične veze za enzimom, i gde je orijentacija supsti- tuenta položaja 6 indazolnog jezgra bila takva da se oni pružaju ka spoljašnosti vezivnog mesta. Odabrane poze uglavnom su bile među energetski najpovoljnijim (procenjeno Vina doking energijom) što je prikazano u Prilogu A, tabela A.1.

Ekvilibracija početnih struktura kompleksa zahtevala je 5-16 ns simula- cije MD. Poređenje dinamike apo oblika EZH2 i EZH2 u kompleksu sa inhibito-

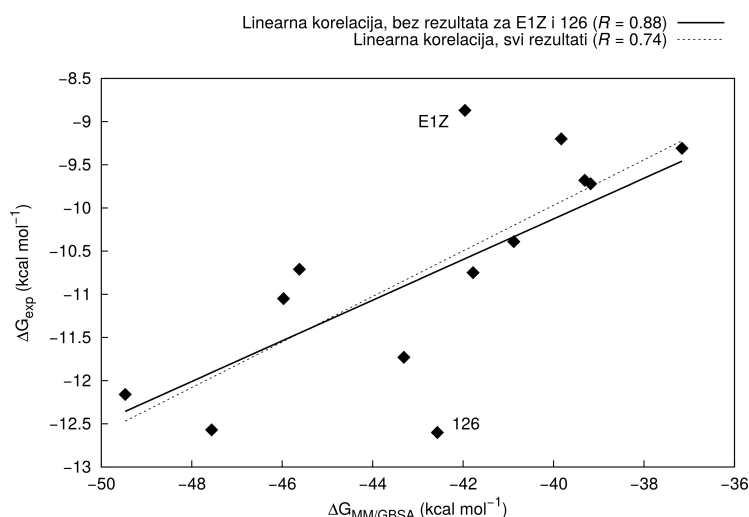
rima ukazalo je da vezivanje inhibitora dovodi do smanjenja fluktuacija strukture, što je u skladu sa čvrstim vezivanjem ovih jedinjenja za kofaktor-vezivno mesto. Na slici 4.4 ilustrovan je efekat prisustva nekoliko različitih inhibitora na vrednosti korena srednje vrednosti kvadrata fluktuacija (engl. *root-mean-square fluctuation*, RMSF) C_{α} atoma u EZH2. RMSF vrednosti izračunate su iz simulacija ekvilibrisanog sistema trajanja 4 ns. Uočljivo je izraženo smanjenje RMSF vrednosti kod većine aminokiselina koja formiraju vezivno mesto, dok je ovo smanjenje neznatnije u slučaju C-terminalnih rezidua post-SET domena koje ostaju slobodne i difuzno pokretljive u rastvaraču. Aminokiseline broja 700-710 u sekvenci EZH2 uobičajeno su u kontaktu sa CXC domenom enzima, pa je nešto viši stepen njihove pokretljivosti verovatno posledica skraćivanja modela na SET i post-SET domene. Kako su ove rezidue prostorno udaljene od kofaktor-vezivnog mesta i ne stupaju u interakcije sa inhibitorima, njihova dinamika nije uticala na sveukupnu stabilnost simuliranih kompleksa.



Slika 4.4: Efekat prisustva inhibitora na dinamiku EZH2, ilustrovan kroz smanjenje vrednosti korena srednjeg kvadrata fluktuacija (RMSF) C_{α} atoma u strukturi.

Produkcione simulacije trajanja 4 ns, koje su sprovedene nakon ekvilibracije kompleksa, upotrebljene su u MM/GBSA proračunima kojima je računarskim putem procenjena slobodna energija vezivanja proučavanih inhibitora za EZH2. Rezultati MM/GBSA proračuna zatim su korelisani sa energijama vezivanja izračunatim iz eksperimentalno određenih konstanti inhibicije u cilju procene koliko dobro rezultati simulacija odgovaraju eksperimentalno određenim afinitetima piridonskih inhibitora za EZH2. Svi rezultati MM/GBSA proračuna sumirani su u Prilogu A, tabela A.2, a korelacija sa eksperimentalnim vrednostima ilustrovana je na slici 4.5. Kada se uzmu u obzir rezultati za svih 13 inhibitora, postoji dobra linearna korelacija

između $\Delta G_{MM/GBSA}$ i ΔG_{exp} čiji koeficijent korelacije iznosi $R = 0,74$ (Spearman-ov koeficijent korelacije $\rho = 0,72$). Kao što je prikazano na slici 4.5, rezultati za dva inhibitora značajnije odstupaju od opšteg uočenog trenda: $\Delta G_{MM/GBSA}$ za najpotentniji od proučavanih inhibitora, jedinjenje **126**, je u rangu desetostruko manje aktivnih jedinjenja, dok je izračunata vrednost energije vezivanja za najslabiji od proučavanih inhibitora, jedinjenje **E1Z**, u rangu sa desetostruko aktivnijim jedinjenjima. Ukoliko se ova dva rezultata isključe iz korelacije, koeficijent korelacije ima vrednost $R = 0,88$ ($\rho = 0,88$). U daljem tekstu će detaljnije biti analizirani potencijalni razlozi za odstupanje ova dva rezultata od opšteg trenda. Ovde će samo kratko biti zaključeno da je uočena korelacija između eksperimentalnih i računarskih rezultata uporediva sa rezultatima test studija pouzdanosti MM/GBSA proračuna. Naime, u studijama gde su MM/GBSA proračuni bili bazirani na kristalografski rešenim početnim strukturama kompleksa biomolekula i liganada, stepen korelacije sa eksperimentalnim energijama vezivanja bio je u opsegu R između 0,5 i 0,7 [379]. Time je R vrednost dobijena u ovoj studiji u nivou pouzdanosti same metode. Dobro slaganje između eksperimentalnih rezultata i izračunatih vrednosti slobodne energije vezivanja daje podršku zaključku da način interakcije inhibitora i EZH2, koji je uočen tokom simulacija, odgovara interakcijama relevantnim u biološkom okruženju. Stoga je bilo validno pristupiti detaljnijoj analizi prikupljenih trajektorija u cilju identifikovanja interakcija koje ključno doprinose aktivnosti i selektivnosti proučavanih jedinjenja.



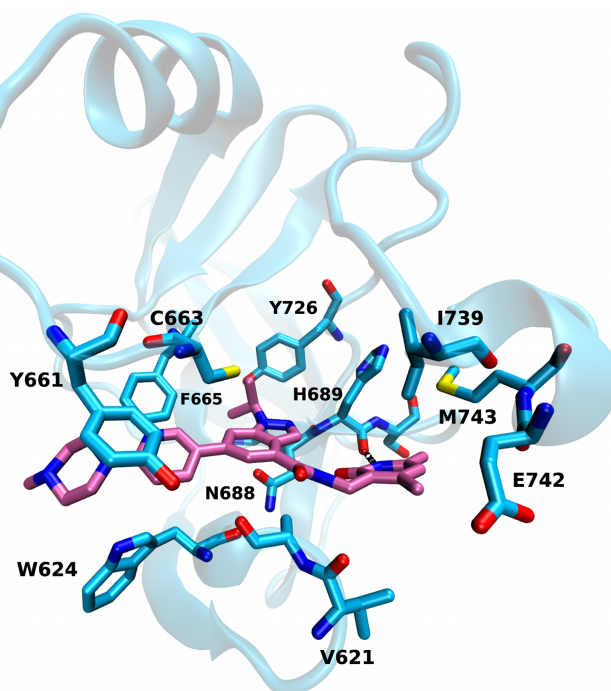
Slika 4.5: Korelacija između $\Delta G_{MM/GBSA}$ i ΔG_{exp} vrednosti za 13 piridonskih inhibitora EZH2.

Takođe je važno naglasiti i da produžavanje produkcionih trajektorija u režimu konvencionalne MD nije dovelo do značajnijih promena u stabilnosti simuliranih kompleksa niti u $\Delta G_{MM/GBSA}$ vrednostima. Procene slobodne energije ve-

zivanja tokom dodatnih 10 ns simulacija zbirno su prikazane u Prilogu A, tabela A.3. Na slici A.1 u Prilogu A takođe je prikazana zavisnost koficijenta korelacije između $\Delta G_{MM/GBSA}$ i ΔG_{exp} u funkciji dužine analizirane trajektorije. Ovim simulacijama potvrđeno je da se, bar kada je reč o metodi „krajnje tačke” kao što je MM/GBSA, konvergirani rezultati za proučavani sistem mogu dobiti iz 4–5 ns produkcionih simulacija, nakon ekvibracije kompleksa dobijenog dokingom.

Tipična poza piridonskog inhibitora u mestu za vezivanje kofaktora u EZH2 može se ilustrovati na primeru jedinjenja **G3K** (slika 4.6). Piridonski deo strukture inhibitora stupa u interakciju sa peptidnim kosturom proteina u delu koji odgovara residui H689, i to prevashodno kroz obrazovanje vodonične veze između piridonskog azota i karbonilnog kiseonika peptidne veze. Ukoliko se kao geometrijski kriterijumi za postojanje vodonične veze uzmu donor–akceptor rastojanje od 3 Å i odstupanje od planarnosti ne veće od 25°, ova vodonična veza zastupljena je u oko 35% kadrova produkcione trajektorije. Druga vodonična veza, između karbonila piridona i azota peptidne veze, ima daleko manju zastupljenost. Kod struktura inhibitora sa azotom u položaju 2 indazolnog jezgra, ovaj azot može biti alternativni akceptor, pri čemu formiranje takve vodonične veze deluje geometrijski povoljnije. Imajući u vidu da su u okviru molekularno mehaničkog polja sila zastupljena samo parcijalna atomska naelektrisanja centrirana na koordinatama atoma, precizniji opis geometrije vodoničnih veza u ovom delu strukture ograničen je korišćenom metodologijom. Ipak, malo je verovatan položaj inhibitora gde bi on bio izmešten ka I-SET domenu, jer bi time bili izgubljeni povoljni kontakti sa post-SET domenom. Naime, post-SET domen zatvara značajan deo piridonskog dela inhibitora u hidrofobnoj šupljini koju formiraju aminokiseline I739, E742 i M743, dok V621 doprinosi donjem delu vezivnog mesta. Pored piridona, i drugi delovi strukture analiziranih inhibitora takođe stupaju u relativno uniformne interakcije sa vezivnim mestom. Izopropil grupa položaja 1 indazolnog jezgra pruža se ka osnovi kanala za vezivanje lizina gde stupa u povoljne disperzivne interakcije sa Y726 i F665. Centralni indazol stupa u interakcije sa više aminokiselina I-SET domena, prevashodno sa C663, Y661 i F665. Bočni lanac C663 pozicioniran je tako da se atom sumpora nalazi približno 3,7 Å iznad ravni aromatičnog jezgra inhibitora, tako da njihov relativni položaj odgovara geometriji sumpor-aren interakcije [380]. Orjentacija supstituenta položaja 6, u skladu sa pretpostavkama baziranim na eksperimentalnim podacima, uslovljava pružanje ovog dela strukture ka spoljašnjosti vezivnog mesta. Piridinski prsten ostvaruje značajnije interakcije sa enzimom, pre svega stupanjem u aromatičnu interakciju sa Y661, u „paralelno izmeštenoj” konfiguraciji. S druge strane, piperazinski deo strukture zadržava značajniju pokretljivost usled delimične izloženosti rastvaraču, uz tek

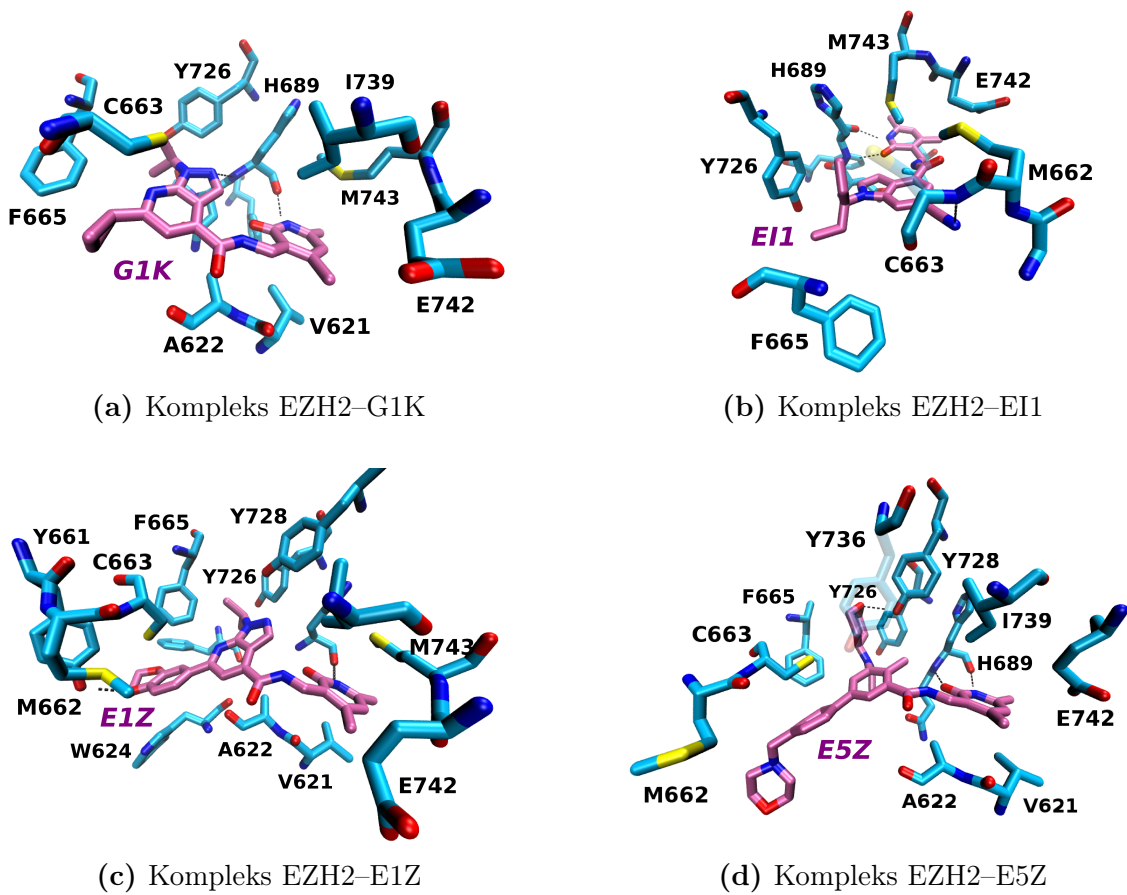
prolazne kontakte prevashodno sa bočnim lancem W624. Opisani profil interakcija sa vezivnim mestom odražen je i u energiji vezivanja procenjenoj MM/GBSA metodom, gde je primetno da dominantna entalpijska komponenta energije vezivanja **G3K** i sličnih inhibitora (**G5K**, **E2Z**, **E3Z** i **E4Z**) potiče od van der Waals-ovih interakcija.



Slika 4.6: Položaj inhibitora **G3K** (obojen roze) u vezivnom mestu EZH2 (enzim je prikazan polutransparentnom plavom). Tamnije plavom bojom prikazani su bočni lanci aminokiseline koje se nalaze na udaljenosti manjoj od 3 Å od bilo kog atoma inhibitora, i koji formiraju vezivno mesto. Radi jasnijeg prikaza, iz ilustracije su izostavljeni atomi vodonika, aminokiseline F723 i Y728.

Iako su interakcije opisane za **G3K** većinom prisutne i kod drugih piridon-skih struktura, nekoliko proučavanih inhibitora pokazalo je dodatne specifičnosti interakcije sa EZH2, kao što je ilustrovano na slici 4.7. Inhibitori **G1K** i **G2K** su relativno slabo aktivna jedinjenja što se može objasniti ograničenim interakcijama sa aminokiselinama I-SET domena. Kao što je ilustrovano na slici 4.7a, ciklopropil grupa kao supstituent u položaju 6 indazolnog jezgra obezbeđuje veoma ograničene kontakte sa F665 i C663, dok aromatične interakcije sa Y661 izostaju u potpunosti. Ova činjenica se jasno odražava na energiju vezivanja ova dva inhibitora.

Za inhibitor **EI1** (slika 4.7b) utvrđen je jedinstven način vezivanja koga prevashodno karakteriše prisustvo dodatnih vodoničnih veza za peptidnim kosturom I-SET domena. Specifično, **EI1** naizmenično obrazuje vodonične veze sa peptidnim



Slika 4.7: Reprezentativne strukture kompleksa EZH2 i odabranih inhibitora koje ilustruju dodatne interakcije ili različite načine vezivanja u odnosu na opšti slučaj prikazan na slici 4.6. Strukture liganada obojene su roze dok su aminokiseline vezivnog mesta obojene plavom. Radi jasnijeg prikaza, atomi vodonika i drugi delovi enzima nisu prikazani na slici. Vodonične veze prikazane su kao isprekidane crne linije.

azotima C663 i M662 preko nitrilne grupe kao akceptora, pri čemu je svaka od ove dve veze zastupljena u oko 15% kadrova produkcionih trajektorija. Usled mogućnosti obrazovanja ovih specifičnih interakcija, struktura **EI1** je blago izmeštena u pravcu I-SET domena, pri čemu centralno aromatično jezgro ima nešto veći nagib u odnosu na ravan piridona. U ovom položaju, bočni lanci M743 i V621 mogu da obrazuju „sendvič” oko piridona, značajno redukujući njegove fluktuacije, što se odražava i na vodonične veze koje **EI1** obrazuje sa peptidnim kosturom H689. Naspram uobičajenih 35%, zastupljenost obe vodonične veze koje obrazuje piridonski deo strukture iznosi približno 54%. Konačno, dodatna voluminoznost pentanskog supstituenta u položaju 1 indazola doprinosi hidrofobnim interakcijama sa odgovarajućim džepom u strukturi enzima. Zahvaljujući opisanim interakcijama, iako

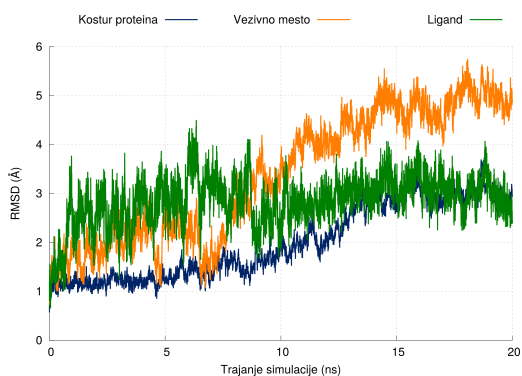
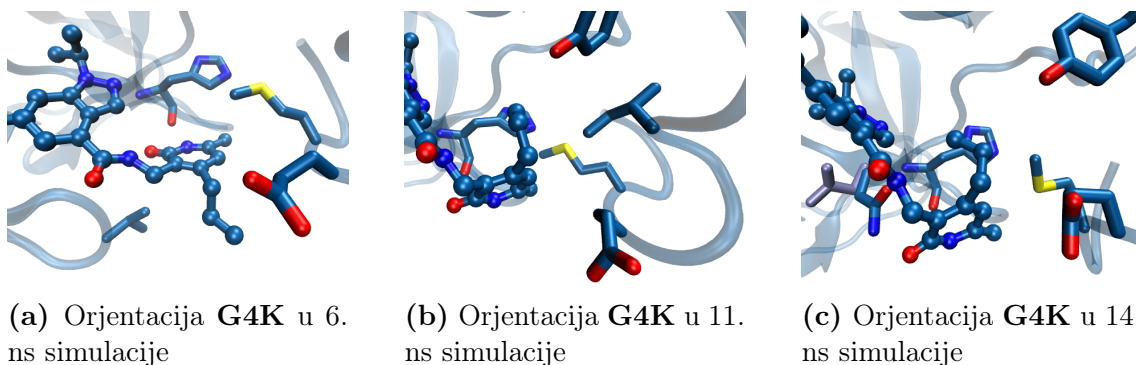
relativno mala molekula, **EI1** ispoljava izraženu efikasnost kao ligand.

Prisustvo 1,4-benzodioksana na položaju 6 pirazolopiridinskog sistema u inhibitoru **E1Z** (slika 4.7c) uslovljava uspostavljanje specifične konformacije fleksibilne petlje I-SET domena u kojoj u stabilizaciji ovog dela strukture učestvuju C663, Y661, M662 i W624. Iako su uspostavljeni kontakti entalpijski veoma povoljni, što je primetno iz niske vrednosti $\Delta G_{MM/GBSA}$, gubitak fleksibilnosti ovog dela strukture EZH2 entropijski je nepovoljan, što je međutim zanemareno u našim proračunima. Ovo činjenica pruža verovatno obrazloženje za neslaganje $\Delta G_{MM/GBSA}$ sa eksperimentalno određenim afinitetom vezivanja.

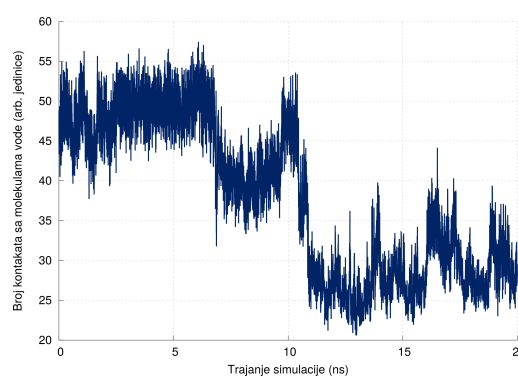
Jedinjenje **E5Z** (slika 4.7d) je jedini od proučavanih inhibitora koji ne sadrži centralni biciklični sistem. Iako su interakcije ove strukture sa EZH2 u najvećem delu slične sa opštim slučajem opisanim za **G3K**, značajnije razlike uočene su u regiji vezivnog mesta pri osnovi kanala za vezivanje supstratnog lizina. Otvaranje centralnog bicikličnog sistema uslovljava da etil grupa tercijarnog amina zauzima položaj u hidrofobnom džepu vezivnog mesta gde se kod ostalih inhibitora pozicionira izopropil, dok se tetrahidropiranski prsten pruža paralelno sa pretpostavljenim kanalom za vezivanje supstratnog lizina. Tetrahidropiran značajno doprinosi energiji vezivanja kroz kontakte sa Y726, Y728 i Y736. Y728 obrazuje vodoničnu vezu sa kiseonikom tetrahidropirana koja je prisutna u približno 50% kadrova produkcione trajektorije, dok Y726 i Y736 okružuju prsten sa dve strane, zatvarajući ga u hidrofobnom kanalu. Kao i kod **EI1**, položaj **E5Z** je blago izmešten u smeru I-SET domena, što doprinosi stabilizaciji piridona kroz kontakte sa I739 i V621 kao i kroz povoljniju geometriju za uspostavljanje vodoničnih veza sa H689. Drugi benzenov prsten, vezan za centralni fenil, ostvaruje interakcije prevashodno sa C663, slično indazolnom jezgru **G3K**. Morfolinski deo strukture je veoma mobilan, i uspostavlja prolazne kontakte sa M662 i W624. Interesantno je da je položaj tetrahidropirana takav da dalje produžavanje strukture može da omogući dodatne kontakte sa aminokiselinama I-SET i naspramnih delova post-SET domena. U prilog ovoj mogućnosti govore i strukture kasnije objavljenih inhibitora **CPI-169** i **CPI-360** (slika 1.12), za koje se čini da eksploatišu upravo ovu mogućnost.

Konačno, za inhibitore **G4K**, **G6K** i **126** utvrđeno je da stupaju u interakciju za vezivnim mestom na način koji je značajno drugačiji od preostalih inhibitora. Inicijalno su se, tokom ekvibracije kompleksa ova tri inhibitora, pojavile poteškoće u stabilizovanju fluktuacija sistema usled čega je početno vreme ekvibracije od 4 ns produženo. Počevši od uobičajene strukture kompleksa, nakon oko 10 ns simula-

cije, rotacijom oko veza koje spajaju piridonski prsten sa centralnim amidom, došlo je do reorijentacije piridonskog prstena. Reorijentacija piridona praćena je promenom rotamernog oblika N688 pri čemu je dobijeni kompleks ostao stabilan do kraja simulacije trajanja 20 ns kao i tokom produkcionih simulacija. Opisana evolucija trajektorije na primeru inhibitora **G4K** ilustrovana je na slici 4.8.



(d) Fluktuacije kompleksa tokom trajanja simulacije

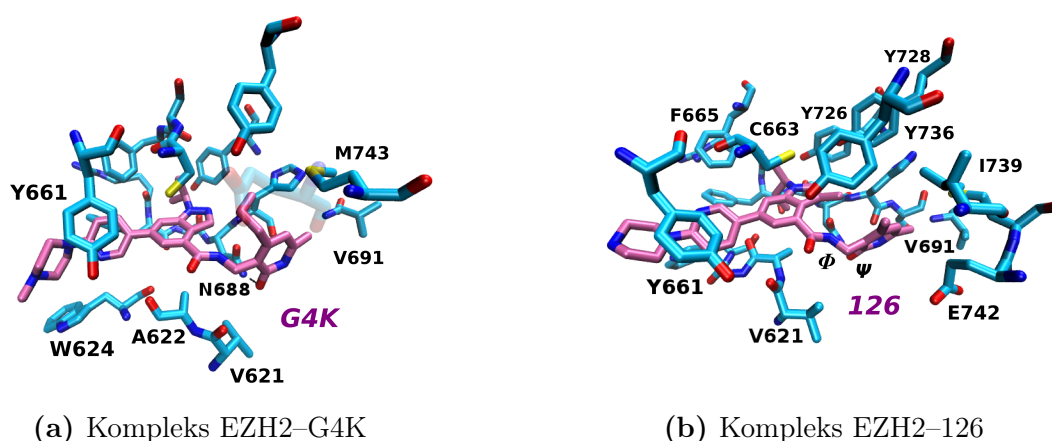


(e) Izloženost propil grupe **G4K** rastvaraču

Slika 4.8: Evolucija orjentacije **G4K** u vezivnom mestu EZH2 tokom ekvibracije sistema trajanja 20 ns.

Reprezentativne strukture kompleksa ova tri inhibitora i EZH2 razlikuju se od prethodno opisanih pre svega po orijentaciji i kontaktima piridonskog prstena. Kao što je ilustrovano na slici 4.9, supstituent položaja 4 piridona usmeren je ka post-SET domenu i stupa u interakcije sa bočnim lancima E742, M743 i Y736. Dominantno zastupljena vodonična veza postaje ona između karbonilnog kiseonika piridona i azota amida u reorijentisanom bočnom lancu N688. Dodatno, azot piridona može da uspostavi kontakt sa peptidnim kosturom proteina posredstvom molekule vode, ali je upitno koliko je ova interakcija značajna. Budući da nakon karakterizacije ovog alternativnog načina vezivanja nije bilo u celosti izvesno da li je opisana poza energetski povoljnija i za druge inhibitore, sprovedene su dodatne simulacije. Simulacije svih kompleksa produžene su dodatnih 10–20 ns, ali slična promena poze

nije ustanovljena ni kod jednog drugog inhibitora osim **G4K**, **G6K** i **126**. Početne konfiguracije kompleksa ova tri inhibitora zatim su ponovo pripremljene korišćenjem konformacije EZH2 iz kompleksa sa **G3K**, tako da budu što sličnije „uobičajenoj” pozi ostalih inhibitora i simulacije su ponovljene. Tokom ponovljenih simulacija, nijedan od ova tri inhibitora nije ostao stabilno vezan a rezultujuće energije vezivanja, procenjene MM/GBSA proračunima, bile su do 9 kcal mol⁻¹ manje povoljne od prethodno dobijenih. Sličan pristup primenjen je i na prethodno proučavane inhibitora ali sa alternativnom konformacijom enzima; npr. inhibitor **G5K** dokovan je u EZH2 tako da reprodukuje način vezivanja ustanovljen za **G6K**. Iako je **G5K** ostao stabilno vezan u ovoj pozi, MM/GBSA proračuni ukazali su da je ovaj alternativni način vezivanja za oko 7 kcal mol⁻¹ manje povoljan. Kod manjih inhibitora, kao što je **G1K**, simulacije započete iz alternativne poze inhibitora dovode do njezove disocijacije. Ovo je prevashodno uzrokovano činjenicom da promena rotamera N688 dovodi do povećane fleksibilnost naspramne beta-ukosnice, koja dalje destabilizuje manje inhibitora čiji supstituent položaja 6 indazola nije dovoljno veliki za „premošćavanje” vezivnog mesta.

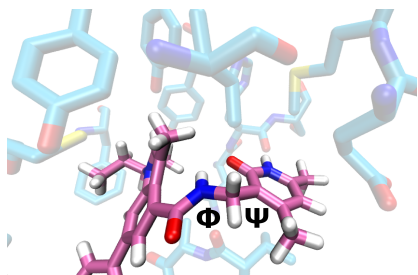


Slika 4.9: Reprezentativne strukture kompleksa EZH2 i inhibitora **G4K** odnosno **126**, za koje je utvrđen alternativni način vezivanja za enzim.

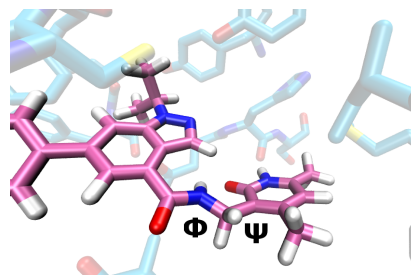
Nakon što je dodatnim simulacijama potvrđeno da alternativni način vezivanja ustanovljen za **G4K**, **G6K** i **126** nije artefakt stohastičke prirode simulacija, pristupljeno je analizi faktora koji uslovljavaju uočene promene. Analiza trajektorije sa **G4K** kao inhibitorom, tokom njegovog premeštanja iz „uobičajenog” u „alternativni” način vezivanja, ukazala je da minimizacija kontakata propil grupe položaja 4 piridona sa vodom verovatno ključno doprinosi ovoj reorijentaciji (slika 4.8e). Međutim, iako u slučaju **G4K** i **G6K**, optimalno pozicioniranje propil grupe u odnosu na post-SET domen može biti razlog za reorijentaciju piridona, kod inhibitora **126**

to izvesno nije slučaj. Inhibitor **126** je strukturno najbliži jedinjenju **G3K** (slika 1.10), ali se ova dva inhibitora značajno razlikuju po svom inhibitornom potencijalu, a pokazano je i da zauzimaju potencijalno različite poze u kompleksu sa enzimom. U racionalizaciji ovih činjenica, pažnja je inicijalno fokusirana na dodatnu metil grupu prisutnu u položaju 3 indolnog jezgra **126**. Leung i saradnici [381] analizirali su veliki broj primera gde uvođenje metil grupe u osnovnu strukturu dovodi do izraženog poboljšanja aktivnosti. Povoljne interakcije dodatne metil grupe sa džepom vezivnog mesta i izmenjene konformacione preference liganda su dva osnovna faktora koja su autori identifikovali kao odgovorne za povećanje afiniteta. Kako u slučaju jedinjenja **126** dodatna metil grupa sterno može da utiče na orijentaciju piridona, izvršeno je ispitivanje konformacionih preferenci **G3K** i **126** u rastvoru korišćenjem metadinamike kao metode za ubrzavanje prevazilaženja energetske barijere i rekonstruisanje profila slobodne energije (ilustrovani na slikama 4.10e i 4.10f). Ukoliko se razmotre poze inhibitora sa „uobičajenom” orijentacijom piridona (slike 4.10a i 4.10b), primetno je da se osnovna razlika ogleda u činjenici da prisustvo metil grupe u **126** uslovljava da ravan centralnog amida zaklapa ugao sa ravni indola koji je za oko 20° veći u odnosu na **G3K**. Vrednosti obeleženih torzionih uglova Φ i Ψ variraju u funkciji prilagođavanja piridona optimalnim kontaktima sa post-SET domenom, N688 i H689. U „uobičajenim” pozama, ove dve vrednosti se mogu mapirati na rekonstruisanu površinu slobodne energije (engl. *free energy surface*, FES) kao tačke A na slikama 4.10e i 4.10f. Kao što se može uočiti, konformacija **126** prikazana na slici 4.10a je za oko 3 kcal mol⁻¹ iznad globalnog minimuma energije. Nasuprot ovoj „uobičajenoj” pozi, u alternativnoj orijentaciji piridona (slike 4.10c i 4.10d), vrednosti Φ i Ψ mapiraju se kao tačke B na FES-u. Na slici 4.10e je očigledno da alternativna poza **126** odgovara minimumu slobodne energije. Dok je kod **126** razlika slobodnih energija oko 3 kcal mol⁻¹ u korist alternativne poze, kod **G3K** je razlika zanemarljiva i iznosi oko 0,8 kcal mol⁻¹, pri čemu je „uobičajeni” način vezivanja energetski ukupno daleko povoljniji.

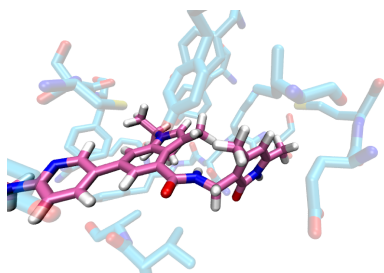
Stoga se može zaključiti da alternativna orijentacija koju **126** zauzima u vezivnom mestu prevashodno predstavlja posledicu relaksacije konformacionog naboja, pri čemu je alternativna bioaktivna konformacija dodatno stabilizovana prilagođavanjem post-SET domena, i to kroz hidrofobne kontakte piridona i dodatne metil grupe sa bočnim lancima aminokiselina I739, Y728 i Y736. Iz površina slobodne energije može se primetiti i da prisustvo dodatne metil grupe dovodi do produbljenja minimuma slobodne energije, što smanjuje verovatnoću prelaska između različitih konformacija, dovodeći do „preorganizacije” inhibitora u bioaktivnu konformaciju. Ovaj efekat takođe verovatno doprinosi povećanoj energiji vezivanja **126**



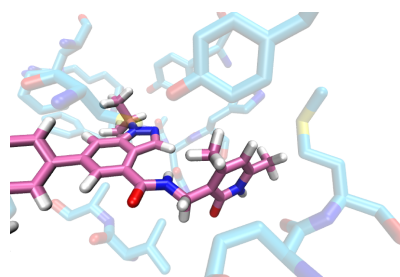
(a) Piridonski deo **126** u „uobičajenoj” orijentaciji, označenoj sa A na panelu (e)



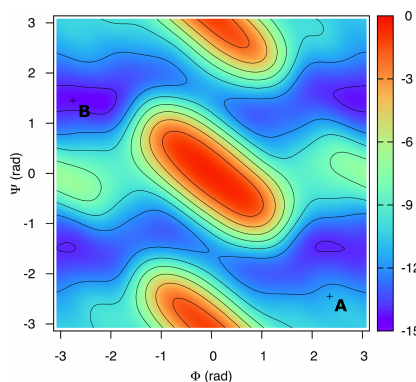
(b) Piridonski deo **G3K** u „uobičajenoj” orijentaciji, označenoj sa A na panelu (f)



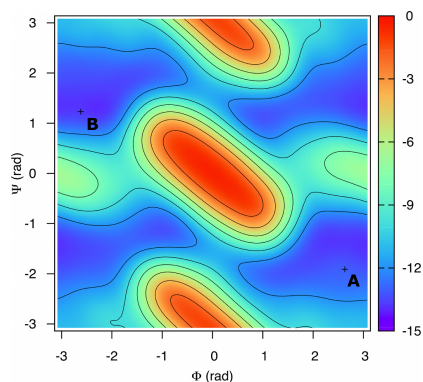
(c) Piridonski deo **126** u „alternativnoj” orijentaciji, označenoj sa B na panelu (e)



(d) Piridonski deo **G3K** u „alternativnoj” orijentaciji, označenoj sa B na panelu (f)



(e) Površina slobodne energije **126** u rastvoru, rekonstruisana u funkciji dva diedralna ugla Φ i Ψ , koji odgovaraju vezama između piridona i centralnog amida.



(f) Površina slobodne energije **G3K** u rastvoru, rekonstruisana u funkciji dva diedralna ugla Φ i Ψ , koji odgovaraju vezama između piridona i centralnog amida.

Slika 4.10: Analiza konformacionih preferenci inhibitora **126** i **G3K**.

za EZH2. Ipak, treba imati u vidu da značajni konformacioni naboji liganada vezanih za proteine nisu neuobičajeni [382, 383], te da **126** verovatno može, barem prelazno, da zauzima i pozu uobičajenu za najveći broj piridonskih inhibitora.

Dodatni dokazi u prilog hipotezi da **G4K**, **G6K** i **126** stupaju u interakciju

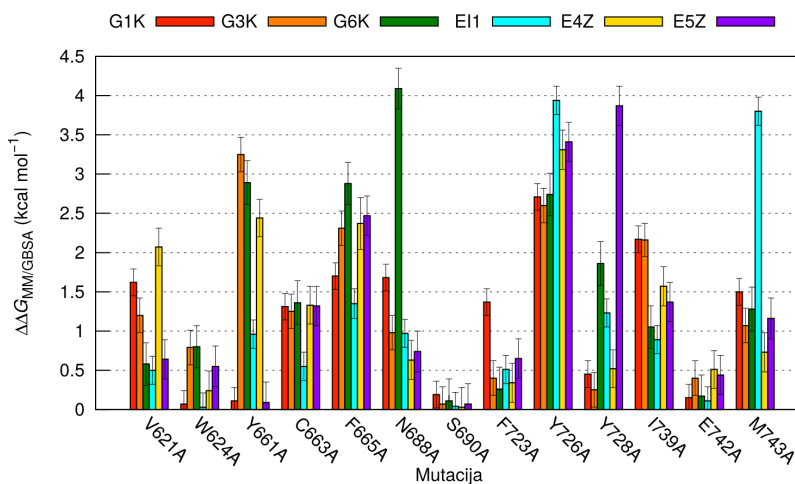
sa vezivnim mestom EZH2 na način različit od drugih piridonskih inhibitora mogu se pronaći u studiji kinetike vezivanja **G5K**, **G6K** i **126** za aktiviranu i neaktiviranu formu PRC2/EZH2 kompleksa [384]. U ovoj studiji pokazano je da je inhibicija alosterno aktivirane forme PRC2/EZH2 pomoću **G5K** koji sadrži 4,6-dimetilpiridonsko jezgro – brzo reverzibilna – dok je za **G6K** i **126** ustanovljeno dugotrajno vezivanje za enzim, i to od 43 i 62 minuta, redom. Ovi rezultati ukazuju na činjenicu da relativno male varijacije strukture (produženje alkil supstituenta na piridonu, uvođenje dodatne metil grupe u indolno jezgro) mogu značajno uticati kako na aktivnost, tako i na kinetiku vezivanja, što potencijalno može biti objašnjeno jedinstvenim načinima interakcije sa vezivnim mestom koji su uočeni tokom naših simulacija. Ipak, treba naglasiti da se navedeni eksperimentalni rezultati odnose na formu PRC2/EZH2 koja je alosterno aktivirana prisustvom histona sa trimetilovanim H3K27, te da se stoga ne mogu smatrati direktno u vezi sa rezultatima naših simulacija. Specifično, naše simulacije ne mogu pouzdano da predvide sve strukturne promene SET i post-SET domena koje se odigravaju usled alosternih interakcija EZH2 sa EED, i ovaj aspekt dinamike vezivnog mesta nije razmatran u sklopu trenutne studije.

U ovom kontekstu, međutim, interesantno je primetiti i razlike u dinamici EZH2 kada se uporede kompleksi sa **G3K** i **G4K** (slika 4.4). Rezidualna mobilnost enzima nakon vezivanja inhibitora, izražena kroz RMSF vrednosti C_{α} atoma, primetno je veća u slučaju **G4K** (kao i **G6K** i **126**; rezultati nisu ilustrovani). Ovaj rezultat može da sugeriše da niske energije vezivanja tri najaktivnija inhibitora delom jesu posledica i relativno niske entropijske „kazne” formiranja kompleksa. Poređenjem struktura apo enzima i kompleksa, uočljivo je da većina inhibitora stabilizuje EZH2 u konformaciji koja je veoma slična strukturi njegovog binarnog kompleksa sa proizvodom kofaktora SAH. Nasuprot većine inhibitora, **G4K**, **G6K** i **126** pokazuju afinitet za konformaciju EZH2 u kojoj je vezivno mesto nešto otvorenije i poseduje izraženiju fleksibilnost, odnosno konformaciju koja deli sličnost sa verovatnim prelaznim stanjem enzima tokom razmene proizvoda kofaktora SAH i dodatne molekule SAM.

4.1.3 Analiza efekata mutageneze u alanin

U cilju uspostavljanja mere relativnog doprinosa svake od aminokiselina vezivnog mesta energiji vezivanja piridonskih inhibitora, sprovedena je analiza efekata mutageneze u alanin. Promena u slobodnoj energiji vezivanja inhibitora nakon uklanjanja datog bočnog lanca („mutacije” u alanin) ukazuje, pre svega, na relati-

van značaj datog bočnog lanca u stabilizaciji inhibitora. Numerički rezultati analize za sve inhibitore prikazani su u tabeli A.4 priloga A, a za manji podset inhibitora ilustrovani su na slici 4.11. Na slici 4.11 može se uočiti da pojedine aminokiseline daju konstantan doprinos stabilizaciji većine inhibitora, dok se kod drugih uočavaju razlike koje i kvantitativno ilustruju razlike u načinu vezivanja inhibitora opisane u prethodnom odeljku.



Slika 4.11: Rezultati analize efekata mutageneze u alanin.

Bočni lanci aminokiselina uključenih u interakcije sa piridonskim delom strukture inhibitora, V621, N688, I739, E742 i M743, daju sličan doprinos stabilizaciji gotovo svih inhibitora, sa izuzetkom **EI1** i **G6K** (kao i sličnih **G4K** i **126**). Kao što je prethodno opisano, M743 pozicioniran je direktno iznad ravni piridona u kompleksu EZH2 i **EI1**, usled čega daje izražen doprinos energiji vezivanja ovog inhibitora. Slično, relativno veliko učešće N688 u energiji vezivanja **G6K** odgovara dodatnim vodoničnim vezama koje se formiraju nakon reorijentacije piridonskog dela strukture i promene rotamera samog asparagina. Ova reorijentacija ujedno uslovljava i manje izražene efekte I739A mutacije, budući da propil grupa **G4K** i **G6K** ostvaruje opsežnije kontakte sa bočnim lancem Y736. E742A mutacija ima zanemarljive efekte što odgovara činjenici da je prevashodno C_β atom uključen u interakciju sa inhibitorima.

Nekoliko aromatičnih aminokiselina SET domena ključno doprinosi vezivanju inhibitora. Y726, F665 i Y661 pojedinačno daju najizraženije doprinose: Y726 i F665 stupaju u hidrofobne interakcije sa centralnim aromatičnim sistemom i njegovim alkil supstituentom položaja 1, dok je Y661 uključen u aromatične interakcije sa piridonskim delom supstituenta u položaju 6. Važna uloga Y661 u stabilizaciji piridonskih inhibitora nedavno je potvrđena i eksperimentalno, u studiji koja je po-

kazala da Y661D mutacija dovodi do drastičnog gubitka afiniteta inhibitora za EZH2 [385]. Dok Y728 umereno doprinosi energiji vezivanja većine inhibitora, značajan pad energije vezivanja **E5Z** za Y728A mutant odražava značaj vodonične veze ove rezidue i tetrahidropiranskog dela strukture inhibitora.

Efekti mutacija C663A i M743A su posebno interesantni budući da se aminokiseline na ovim položajima razlikuju između sekvenci EZH2 i EZH1, pa potencijalno mogu ukazati na strukturnu osnovu delimične selektivnosti inhibitora za EZH2 naspram EZH1. U EZH1, ove dve aminokiseline zamenjuju S664 i T744, pri čemu je T744 prva od četiri C-terminalne aminokiseline post-SET domena koje se razlikuju između inhibitora, dok je C663–S664 izmena izolovana unutar inače identičnog I-SET domena. Ukoliko se pretpostavi da je sekundarna struktura post-SET domena slična između EZH2 i EZH1, jasno je kako zamena relativno hidrofobnog metionina – treoninom značajno može uticati na vezivanje inhibitora. Manji i polarniji bočni lanac treonina verovatno uslovljava suboptimalnu stabilizaciju piridonskog jezgra i posledično niži afinitet za EZH1. Ovo je posebno očigledno na primeru **EI1**, koji govori u prilog ovoj hipotezi. Naime, **EI1** je najselektivniji od svih poznatih inhibitora, sa $IC_{50}(\text{EZH2})$ od $0,009 \mu\text{M}$ i $IC_{50}(\text{EZH1})$ od $1,34 \mu\text{M}$, što odgovara razlici u energiji vezivanja od približno 3 kcal mol^{-1} . Kako rezultati analize efekata mutageneze u alanin ukazuju da M743 daje ključan doprinos energiji vezivanja **EI1**, jasno je da prisustvo treonina u ovom položaju može predstavljati strukturnu osnovu opisane selektivnosti. Stoga se može zaključiti da su interakcije sa ovim delom strukture post-SET domena ključna odrednica selektivnosti za EZH2 naspram EZH1.

S druge strane, razlika između cisteina (C663) i serina (S664) je značajno manja. C663A mutacija negativno se odražava na energiju vezivanja svih inhibitora, što ukazuje na učešće sumporovog atoma u interakcijama sa centralnim aromatičnim biciklom inhibitora. Iako bi serin na istom položaju hipotetički mogao da bude uključen u uspostavljanje slabe vodonične veze sa π -sistemom inhibitora, naša je pretpostavka da ovo nije energetski povoljno iz jednog od dva razloga. Ukoliko se uzme u obzir orijentacija C663 u kristalnim strukturama EZH2, S664 u istom položaju efektivno bi predstavljao mimetik prolina kroz uspostavljanje vodonične veze sa karbonilnim kiseonikom peptidne veze susednog S665. Raskidanje ove veze je energetski nepovoljno, što čini potencijalnu reorijentaciju serina manje verovatnom. Dodatno, energetska cena desolvatacije serina je veća nego što je to slučaj sa cisteinom, što takođe potencijalno doprinosi manjoj verovatnoći uspostavljanja odgovarajuće interakcije ove aminokiseline sa indazolnim delom strukture inhibitora. Sveukupno uzevši, rezultati obavljenih simulacija pružaju dobru osnovu za razu-

mevanje selektivnosti piridonskih inhibitora za EZH2 naspram EZH1, što dodatno potkrepljuje validnost predloženog načina vezivanja.

Strukturna racionalizacija SAR

Na osnovu rezultata simulacija MD i analize efekata računarske mutageneze u alanin, odnos između strukture piridonskih inhibitora i njihove inhibitorne aktivnosti moguće je razjasniti iz perspektive strukture samog vezivnog mesta odnosno kompleksa EZH2 i proučavanih inhibitora, kroz sledeća opažanja:

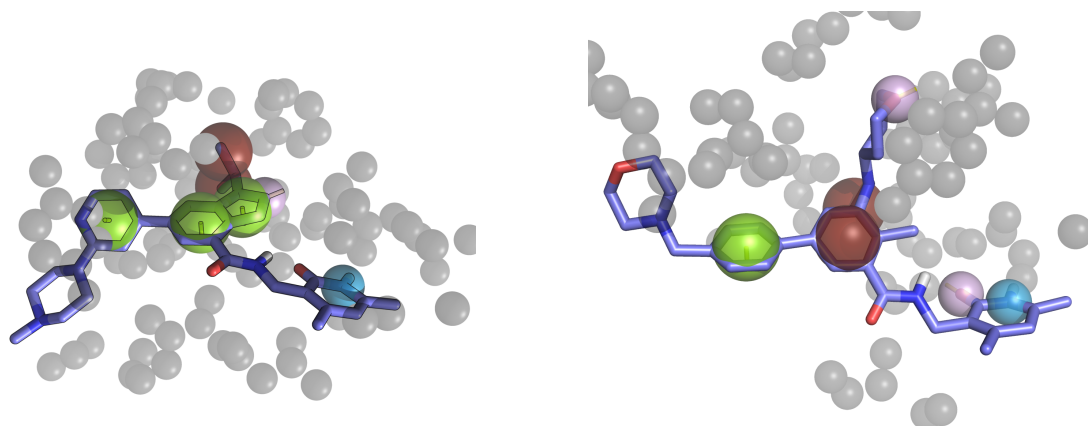
1. Piridonsko jezgro predstavlja deo strukture ključan za formiranje stabilnog kompleksa kroz obrazovanje vodoničnih veza između azota piridona i karbonilnog kiseonika H689. Iako je moguće formiranje vodonične veze između karbonilnog kiseonika piridona i azota H689, u formiranom kompleksu ova vodonična veza ne daje značajan doprinos energiji vezivanja. Moguće je da se obe veze formiraju tokom inicijalnih faza asocijacije, ali da prilagođavanjem enzima inhibitoru samo jedna ostaje dominantno zastupljena. Može se zaključiti da prisustvo donora vodonične veze u ovom položaju predstavlja ključan deo farmakofore piridonskih inhibitora EZH2.
2. Alkil supstituenti piridona položaja 4 i 6 važni su za aktivnost usled kontakata koje uspostavljaju sa post-SET domenom, prevashodno sa E742 i M743. Povoljne hidrofobne interakcije sa post-SET domenom pomažu stabilizaciju delimično spregnute konformacije inhibitora koja nastaje tokom uspostavljanja specifične orijentacije piridona u odnosu na centralni amid i aromatično jezgro. Zamena metil grupe položaja 4 *n*-propil supstituentom favorizuje reorijentaciju piridonskog jezgra a dodatni kontakti sa post-SET domenom doprinose aktivnosti ovih inhibitora.
3. Odstupanje amidskog linkera od koplanarnosti sa ravni centralnog aromatičnog jezgra (indazola ili indola) pruža optimalnu geometriju ovom delu strukture, pri čemu amid doprinosi elektrostatskim interakcijama vezivanju inhibitora, a delimično i kroz jasno usmerene interakcije (vodonične veze) sa bočnim lancem N688. Gubitak aktivnost dimetil-supstituisanih inhibitora stoga je jasna posledica gubitka ključnih donora vodoničnih veza, uvođenja sterno ometajućih grupa i promena oblika rezultujućih molekula.

4. Alkil supstituenti u položaju 1 centralnog aromatičnog jezgra stupaju u hidrofobne interakcije sa džepom koji formiraju bočni lanci F665 i Y726, usled čega aktivnost inhibitora dosledno raste sa porastom broja ugljenika ovog supstituenta. No, kako je ovaj deo strukture vezivnog mesta najpodložniji izmenama uzrokovanim somatskim mutacijama enzima (videti odeljak 4.1.5), variranje ovog supstituenta može potencijalno biti iskorišćeno u optimizaciji selektivnosti inhibitora EZH2 za određene mutirane forme enzima.
5. Centralno aromatično jezgro uključeno je u kontakte sa C663 i F665, i uklapa se u dominantno hidrofoban deo vezivnog mesta. Azot položaja 7 u azaindazolnim strukturama (**G1K**, **E1Z**, **E2Z**) ne može da učestvuje u bilo kakvim polarnim interakcijama, a doprinosi energiji desolvatacije, usled čega indazolne strukture imaju veću aktivnost. Iako azot položaja 2 indazola potencijalno može da obrazuje vodonične veze sa H689, indolne strukture nemaju smanjenu aktivnost verovatno usled njihovog π -suficitarnog karaktera.
6. Aromatični prsten u položaju 6 centralnog bicikličnog sistema značajno doprinosi aktivnosti kroz aromatične interakcije sa bočnim lancem Y661. Biciklični supstituenti u ovom položaju (**E1Z**) nisu optimalan izbor za povećanje aktivnosti jer zahtevaju uspostavljanje specifičnog položaja i gubitak fleksibilnosti petlje I-SET domena, što je sveukupno nepovoljno za njihovu energiju vezivanja.

4.1.4 Farmakoforni model inhibitora EZH2

Kao što je u prethodnim odeljcima detaljno opisano, analiza struktura kompleksa EZH2 i piridonskih inhibitora ukazala je na postojanje značajnih razlika u načinu interakcije različitih inhibitora sa vezivnim mestom enzima. U razvoju farmakofornog modela zasnovanog na strukturi kompleksa moguće je pokušati identifikaciju najmanjeg zajedničkog sadržaoa superponiranih farmakofornih elemenata većeg broja aktivnih jedinjenja. Međutim, u situaciji kada postoje značajne razlike u bioaktivnim konformacijama i tipu ključnih interakcija sa vezivnim mestom, ovaj pristup nije nužno optimalan. Kao što je pokazano u retrospektivnim studijama, uspešnost farmakoforne pretrage može se značajno poboljšati ukoliko se istovremeno koristi više farmakofornih hipoteza koje kodiraju različite mogućnosti interakcije sa ciljnim mestom dejstva [386]. Šta više, kombinovanje farmakofornog modela sa oblikom molekule kao kriterijumom sličnosti daje rezultate bolje od svakog od ovih

pristupa ponaosob, kao i uspešnost pretrage veću od računarski zahtevnijih metoda kao što je molekularni doking [387].



(a) Farmakoforni model izveden iz **G5K**

(b) Farmakoforni model izveden iz **E5Z**

Slika 4.12: Ilustracija farmakofornih modela inhibitora EZH2. Farmakoforni elementi prikazani su kao obojene sfere: crvenom su predstavljene lipofilne grupe, zelenom aromatični centri, plavom donori a roze bojom akceptori vodonične veze. Sive sfere odgovaraju sferama isključenja.

U skladu sa ovim činjenicama, odabrano je da se u daljem računarskom dizajnu koriste dve farmakoforne hipoteze. Kao predstavnik piridonskih inhibitora sa bicikličnim centralnim sistemom odabrana je struktura **G5K**, dok je za definisanje drugog farmakofornog modela korišćena struktura **E5Z**. **E5Z** se, kako po svojoj konformaciji i obliku, tako i po dodatnim interakcijama tetrahidropiranskog dela strukture izdvaja od ostalih izučavanih inhibitora. Interakcije sa delom vezivnog mesta koji se pruža ka useku za vezivanje supstrata predstavljaju jedinstvenu karakteristiku ovog inhibitora, koja, kao što primeri inhibitora **CPI-169**, **CPI-360** (slika 1.12) i heksahidroizohinolinskih struktura (slika 1.14) ilustruju – može doprineti razvoju visoko-aktivnih inhibitora. Da bi farmakoforni model bio upotrebljiv u identifikaciji novih bioaktivnih molekula, uobičajeno je da broj farmakofornih tačaka u modelu bude ograničen na 3–7 [388]. U definisanju ključnih farmakofornih tačaka u **G5K** i **E5Z** korišćeni su prethodno izloženi rezultati simulacija molekularne dinamike, pa su zadržane farmakoforne tačke liganda koje: (i) učestvuju u obrazovanju vodoničnih veza sa kosturom proteina, odnosno H689; (ii) učestvuju u hidrofobnim interakcijama sa odgovarajućim džepom koji formiraju Y726 i F665; (iii) odgovaraju centralnom jezgru strukture. Dodatno, u farmakoforu izvedenu iz **G5K** uključen je aromatični centar koji stupa u interakciju sa Y661, dok je kod **E5Z** zadržan i akceptor koji obrazuje vodoničnu vezu sa Y728. Kako bi se u potpu-

nosi iskoristila dostupnost trodimenzionalne strukture kompleksa, u farmakoforne modele uključene su sfere isključenja koje odgovaraju obliku vezivnog mesta EZH2. Rezultujući farmakoforni modeli ilustrovani su na slici 4.12.

Rezultati validacije farmakofornih modela na setu od 2.800 molekula „mamac” prikazani su u tabeli 4.1. Oba modela, u kombinaciji sa srednjim rezultatom Tversky sličnosti farmakofora i oblika referentnih molekula, pokazala su veoma dobru sposobnost identifikacije poznatih bioaktivnih molekula. Dobijene su visoke vrednosti metrika ranog obogaćenja, kao i postojeane performanse modela u *bootstrap* validaciji. Model izveden iz jedinjenja **G5K** ima nešto bolje performanse, ali oba farmakoforna modela pokazuju veoma dobru sposobnost generalizacije što ukazuje na njihov potencijal u identifikaciji novih osnovnih struktura. U ovom kontekstu, važno je naglasiti da iako su modeli izvedeni iz piridonskih struktura, u 28 aktivnih molekula seta korišćenog u validaciji bile su zastupljene sve poznate osnovne strukture kofaktor-kompetitivnih inhibitora (odjeljak 1.3.2), uključujući tetrametilperidinske i heksahidroizohinolinske strukture. Očekivano, korišćenjem oba modela u pretrazi dobijaju se rezultati bolji od svakog modela ponaosob.

Tabela 4.1: Performanse farmakofornih modela u identifikaciji 28 inhibitora EZH2 u setu sa 2.800 molekula „mamac”

Farmakoforni model	Metrika	Ceo set	<i>Bootstrap</i> validacija ^a
G5K	EF _{5%}	13,57	13,48 ± 1,65
	AUROC	0,913	0,902 ± 0,027
	BEDROC _{α=20}	0,692	0,671 ± 0,067
E5Z	EF _{5%}	10,71	10,68 ± 1,70
	AUROC	0,874	0,857 ± 0,034
	BEDROC _{α=20}	0,620	0,596 ± 0,076
G5K + E5Z	EF _{5%}	14,29	14,22 ± 1,63
	AUROC	0,926	0,911 ± 0,030
	BEDROC _{α=20}	0,738	0,718 ± 0,076

^a Prikazane su srednje vrednosti uz standardnu devijaciju za 100 ponavljanja

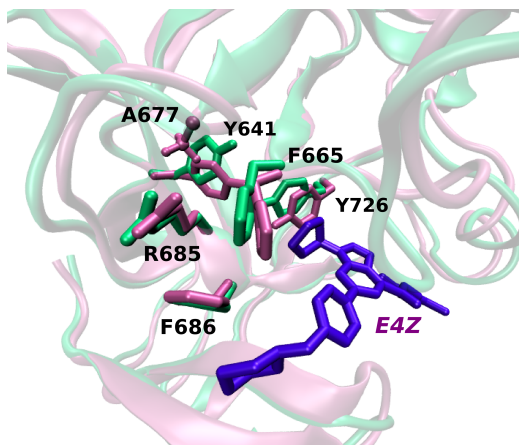
Ovako definisani i validirani farmakoforni modeli korišćeni su u algoritmu za *de novo* dizajn kao osnovne matrice u identifikaciji novih potencijalnih inhibitora EZH2 (odjeljak 4.3).

4.1.5 Uticaj somatskih mutacija EZH2 na vezivanje inhibitora

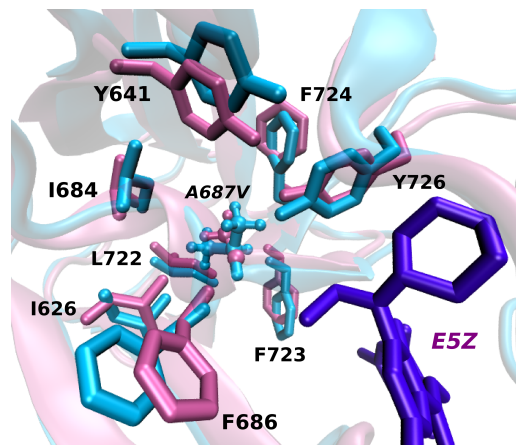
Kao što je detaljnije izloženo u uvodnom odeljku 1.2.3, do danas je opisan niz somatskih mutacija EZH2 koje menjaju supstratnu specifičnost enzima i tako ključno doprinose etiopatogenezi specifičnih formi kancera. Iako pojedine mutirane forme enzima (EZH2^{Y641F}, EZH2^{A687V}) gube afinitet prema nemetilovanom H3K27 usled čega zavise od funkcije neizmenjenog EZH2 (EZH2^{WT}), adekvatna aktivnost inhibitora EZH2 i prema ovim mutiranim formama od velikog je terapijskog značaja. Za **E4Z** i **E5Z** pokazano je da ispoljavaju i do pet puta veći afinitet prema A677G i A687V mutantima naspram EZH2^{WT}, uz približno istu ili neznatno umanjenu aktivnost prema Y641(F/C/S/H/N) mutantima [187, 188].

U cilju razumevanja kako ove mutacije utiču na strukturu vezivnog mesta i afinitet inhibitora prema EZH2, sprovedene su simulacije sledećih kompleksa: EZH2^{A677G}-E4Z, EZH2^{A677G}-E5Z, EZH2^{A687V}-E5Z, EZH2^{Y641F}-E4Z i EZH2^{Y641F}-E5Z. Svaki od kompleksa sa mutiranom formom enzima analiziran je nasuprot odgovarajućeg kompleksa sa EZH2^{WT}. Većina promena vezivnog mesta koje su uočene odgovaraju predviđanjima koja su ranije predložena na osnovu samih kristalnih struktura SET domena [125, 126]. Za mutaciju A677 u glicin se veruje da dovodi do uvećanja vezivnog mesta koje je dovoljno da omogući reorijentaciju H3K27me2 u cilju prenosa treće metil grupe. Našim simulacijama je ustanovljeno da ova mutacija uzrokuje preuređenje niza aminokiselina u blizini pretpostavljenog dna kanala za vezivanje lizina. Kao što je ilustrovano na slici 4.13a, u odsustvu sterne smetnje alaninovog C_β atoma, bočni lanac Y641 se izmešta ka gore, što zauzvrat omogućava bočnim lancima F665 i Y726 da se pomere ka unutrašnjosti enzima, time značajno uvećavajući zapreminu vezivnog mesta (slika 4.13c). U kontekstu vezivanja inhibitora, proširenje hidrofobnog džepa pri osnovi kanala za lizin omogućava alkil grupi položaja 1 centralnog jezgra da se veže dublje.

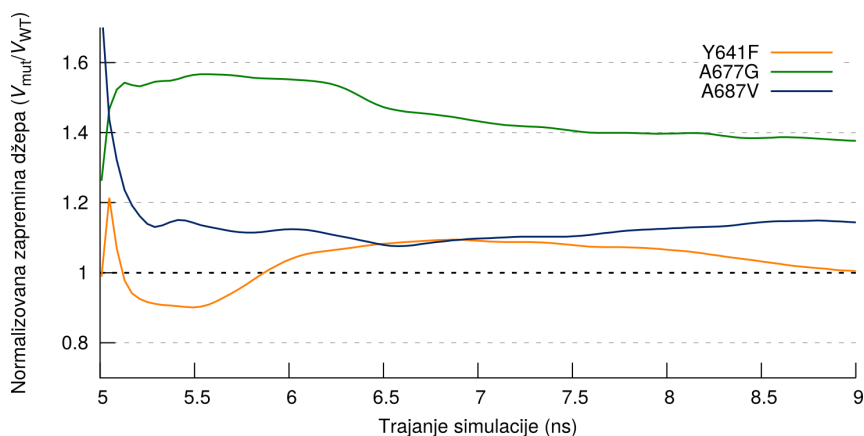
Mutacija A687V dovodi do izmeštanja većeg broja aminokiselina, kao što je ilustrovano na slici 4.13b. Iako su položaji većeg broja hidrofobnih aminokiselina koje okružuju ovaj položaj (L722, I684, I626, F724, F723 i F686) izmenjeni usled prisustva voluminoznijeg valina, najuočljivija promena je i kod ovog mutanta izmeštanje Y641, što je dosledno sa njegovom predloženom ulogom „čuvara kapije” i glavne odrednice supstratne specifičnosti PKMT sa SET domenom. Kao i kod A677G mutacije, neto efekat jeste uvećanje zapremine vezivnog mesta, mada je sa A687V mutacijom ova promena manje izražena (slika 4.13c). Sa aspekta vezivanja



(a) **E4Z** (obojen ljubičasto) u kompleksu sa neizmenjenim EZH2 (prikazan roze) odnosno A677G mutantom (prikazan pastelno zelenom). Strukture su preklapljene po koordinatama C_{α} atoma.



(b) **E5Z** (obojen ljubičasto) u kompleksu sa neizmenjenim EZH2 (prikazan roze) odnosno A687V mutantom (prikazan plavom). Strukture su preklapljene po koordinatama C_{α} atoma.



(c) Srednja vrednost zapremine vezivnog mesta za kofaktor, prikazana u funkciji trajanja simulacije. Boje linija odgovaraju različitim mutiranim formama EZH2, kao u legendi. Zapremine su normalizovane srednjom vrednošću zapremine vezivnog mesta kod neizmenjenog enzima, tako da vrednosti veće od 1 odgovaraju uvećanju, a vrednosti manje od 1 smanjenju zapremine u odnosu na neizmenjeni enzim.

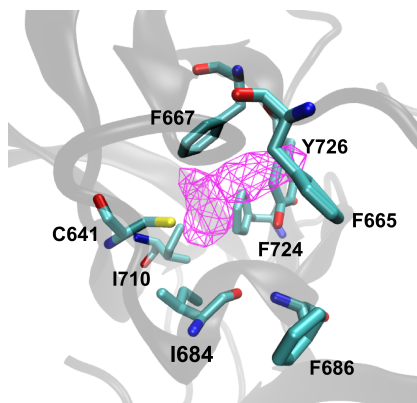
Slika 4.13: Uticaj somatskih mutacija EZH2 na strukturu vezivnog mesta za kofaktor.

inhibitora, promene su suštinski identične onima ustanovljenim kod A677G mutanta. Konačno, mutacija Y641F dovodi do najmanje izraženih promena, izuzev blagog pomeranja bočnog lanca fenilalanina naviše. Ova blaga promena položaja ne dovodi do značajnog proširenja vezivnog mesta (slika 4.13c), što nije ni očekivano budući da se pretpostavlja da ova mutacija pre svega utiče na mogućnost formiranja

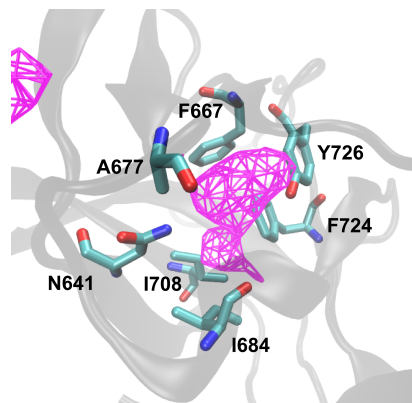
vodonične veze sa slobodnim elektronskim parom ε -azota supstratnog lizina.

Promene strukture vezivnog mesta uočene tokom trajanja simulacija nisu dovele do značajnijih promena u pozici inibitora **E4Z** i **E5Z**, niti u načinu njihove interakcije sa enzimom. Korišćenjem MM/GBSA proračuna, nisu ustanovljene razlike u energiji vezivanja inibitora veće od nivoa statističke neizvesnosti same metode. Eksperimentalni podaci sugerišu da bi energija vezivanja **E4Z** za EZH2^{A677G}, odnosno **E5Z** za EZH2^{A677G} i EZH2^{A687V}, trebalo da bude oko 1 kcal mol⁻¹ negativnija u poređenju sa EZH2^{WT}. Na osnovu uočenih promena, naša je hipoteza da je za razumevanje povećanog afiniteta inibitora za ove mutirane forme važno adekvatno uzeti u obzir razlike u hidrataciji vezivnog mesta. Preciznije, u kristalnim strukturama, Y641 uspostavlja vodoničnu vezu sa molekulom vode, koja je dodatno stabilizovana i drugim vezama sa karbonilnim kiseonicima kostura proteina. Tokom vezivanja liganda, ova voda mora biti istisnuta iz vezivnog mesta, pri čemu bi promena orijentacije Y641 uočena tokom svih simulacija mutanata olakšala ovaj proces. Međutim, isto objašnjenje bi bilo validno i za Y641F mutant (koji nema sposobnost uspostavljanja vodonične veze sa spomenutom molekulom vode), ali je, naprotiv, slobodna energija vezivanja **E4Z** i **E5Z** za ovako izmenjen enzim oko 0,2 kcal mol⁻¹ pozitivnija u odnosu na EZH2^{WT}. Stoga na osnovu trenutno dostupnih rezultata nije moguće dati celokupno objašnjenje promena u afinitetu inibitora prema mutiranim formama enzima. Dodatne simulacije koje bi ispitale celokupan proces asocijacije liganda i enzima, uzimajući u obzir atomske detalje procesa desolvatacije vezivnog mesta i liganada, mogle bi da pruže potpunije objašnjenje ustanovljenih razlika u afinitetu.

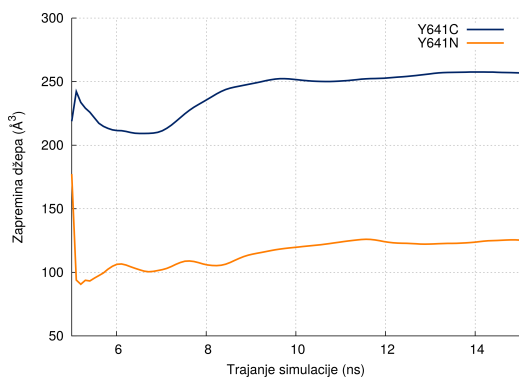
Interesantno je uočiti i da različite mutacije EZH2 dovode do dobro definisanih promena vezivnog mesta, što može predstavljati strukturnu osnovu za dizajn inibitora koji će ispoljavati veću selektivnost za ove patološki izmenjene forme EZH2. Pored već opisanih, simulacije apo oblika EZH2^{Y641C} i EZH2^{Y641N} dodatno ilustruju ovu mogućnost. Kao što je prikazano na slikama 4.14a i 4.14b, mutacije Y641 mogu dovesti do otvaranja novih džepova u vezivnom mestu, čija je zapremina dovoljno velika (slika 4.14c) za prihvatanje adekvatno odabranih supstituenata inibitora. Razlike u polarnosti ovih džepova (slika 4.14d) ilustruju da postoji potencijal za dizajn mutant-specifičnih inibitora koji bi predstavljali interesantno oruđe u ispitivanjima hemijske biologije EZH2. Variranje supstituenta položaja 1 centralnog jezgra poznatih piridonskih inibitora bio bi logičan prvi korak u istraživanju mogućnosti dizajna takvih molekula.



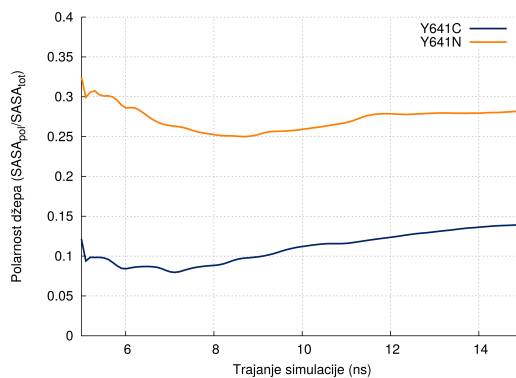
(a) Promene vezivnog mesta kod Y641C mutanta. Mrežasti roze prikaz odgovara obrisima džepa u vezivnom mestu.



(b) Promene vezivnog mesta kod Y641N mutanta. Mrežasti roze prikaz odgovara obrisima džepa u vezivnom mestu.



(c) Prosečna zapremina formiranih džepova u funkciji trajanja simulacije. Boje linija odgovaraju različitim mutantima, kao u legendi.



(d) Prosečna polarnost formiranih džepova u funkciji trajanja simulacije. Boje linija odgovaraju različitim mutantima, kao u legendi.

Slika 4.14: Ilustracija potencijalne strukturne osnove za dizajn mutant-selektivnih inhibitora EZH2 kroz eksploataciju dodatnih šupljina u vezivnom mestu.

4.1.6 Rezime

U ovom segmentu istraživanja, razvijen je model EZH2 sa celovitim vezivnim mestom za kofaktor koji je usavršen simulacijama molekularne dinamike. Ovaj model je zatim upotrebljen za istraživanje strukturne osnove aktivnosti i selektivnosti poznatih piridonskih inhibitora EZH2. Iako metodologija korišćena u proceni slobodne energije vezivanja ima više ograničenja, dobro slaganje između svih dostupnih eksperimentalnih podataka i rezultata izvršenih simulacija ukazuju da razvijeni model pruža dobru osnovu za razumevanje načina na koji inhibitori stupaju u interakciju sa svojim vezivnim mestom na enzimu. Uz stečeno razumevanje odrednica

aktivnosti poznatih inhibitora, bilo je moguće razviti i validirati farmakoforni model inhibitora EZH2 koji je upotrebljen u dizajnu novih potencijalnih inhibitora ovog enzima (rezultati izloženi u odeljku 4.3). Nadalje, validirani model strukture EZH2 omogućio je istraživanje dinamike kompleksa enzima i novodizajniranih molekula.

4.2 Modeli za predviđanje transporta posredovanog Pgp-om i BCRP-om

4.2.1 Modeli za predviđanje transporta posredovanog Pgp-om

Primenom različitih pristupa opisanih u Eksperimentalnom delu razvijena su tri konačna modela za predviđanje transporta posredovanog Pgp-om. Od toga su dva zasnovana na ansamblima veštačkih neuronskih mreža i to jedan dobijen korišćenjem svih dostupnih deskriptora (u daljem tekstu ANN-24-3), i jedan razvijen nakon ručnog odabira manjeg broja deskriptora za koje je pretpostavljeno da dobro opisuju molekulske karakteristike od značaja za transport posredovan Pgp-om (u daljem tekstu ANN-8-1). Treći model bio je zasnovan na ansamblu SVM modela i u njegovom razvoju korišćeni su svi dostupni deskriptori, bez prethodnog ručnog odabira. Prediktivne sposobnosti razvijenih modela zbirno su prikazane u tabeli 4.2.

Tabela 4.2: Performanse konačnih modela za predviđanje transporta posredovanog Pgp-om.

Model	Parametri	Set	Senzitivnost ^a	Specifičnost ^b	Youden indeks ^c	Tačnost ^d
ANN-24-3	50 ANN	trening _{prosek}	0,89	0,90	0,78	89%
	3 neurona ^e	test _{manuelni}	0,83	0,84	0,67	83%
	24 deskriptora	test _{SOM} ^f	0,82	0,80	0,62	81%
	$N_{trening} = 232$	test _{K-means} ^f	0,81	0,82	0,63	81%
	$N_{test} = 78$	test _{prosek}	0,82	0,82	0,64	82%
ANN-8-1	50 ANN	trening _{prosek}	0,87	0,78	0,64	82%
	1 neuron	test _{manuelni}	0,85	0,76	0,61	81%
	8 deskriptora	test _{SOM}	0,82	0,89	0,71	86%
	$N_{trening} = 232$	test _{K-means}	0,80	0,79	0,59	79%
	$N_{test} = 78$	test _{prosek}	0,82	0,81	0,63	82%
SVM	33 SVM	trening _{prosek}	0,81	0,84	0,65	82%
		test _{manuelni}	0,78	0,84	0,62	81%
	29 deskriptora	test _{SOM}	0,80	0,78	0,58	79%
	$N_{trening} = 232$	test _{K-means}	0,83	0,73	0,56	77%
	$N_{test} = 78$	test _{prosek}	0,80	0,78	0,59	79%

^a Udeo tačno predviđenih supstrata u ukupnom broju supstrata

^b Udeo tačno predviđenih nesupstrata u ukupnom broju nesupstrata

^c Izračunat kao: Senzitivnost + Specifičnost - 1

^d Procenat tačnih predviđanja

^e Broj neurona skrivenog sloja

^f Srednja vrednost 10 nezavisnih raspodela seta

Kao što se može uočiti iz tabele 4.2, sva tri modela pokazuju zadovoljavajuće prediktivne sposobnosti, pri čemu ANN modeli imaju nešto bolje performanse. Interesantno je primetiti da najjednostavniji model, ANN-8-1, razvijen korišćenjem

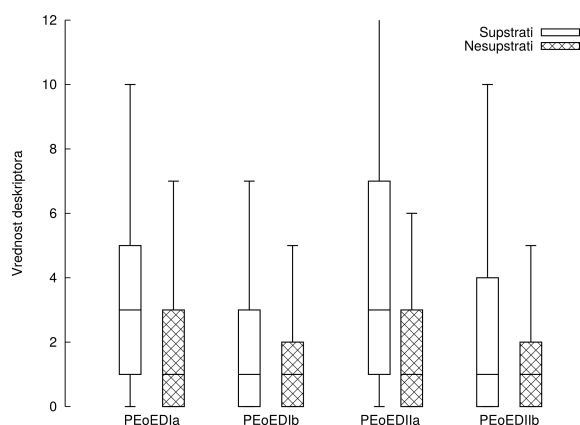
pristupa zasnovanog na prethodnom manuelnom izboru deskriptora, daje sličnu tačnost predviđanja kao i složeniji ANN model sa 24 deskriptora. Međutim, ANN-24-3 ima bolji odnos senzitivnosti i specifičnosti, budući da se primenom ANN-8-1 nesupstrati mogu identifikovati sa nešto manjom pouzdanošću. Svi deskriptori koji su uključeni u ove modele navedeni su u prilogu B, tabela B.3.

Iako tačan mehanizam transporta posredovanog Pgp-om nije u celosti razjašnjen, u literaturi postoji konsenzus da je preduslov za transport supstrata njihova raspodela i zadržavanje u unutrašnjoj ovojnici lipidnog dvosloja membrane, gde mogu stupiti u interakciju za vezivnom šupljinom Pgp-a, u njegovoj otvorenoj konformaciji [389–391]. Samo uklapanje u vezivnu šupljinu zavisi od specifičnih intermolekulskih interakcija, čija raznovrsnost uslovljava i raznovrsnost struktura potencijalnih supstrata. Deskriptori koji čine konačne modele razvijene u ovoj studiji opisuju molekulske karakteristike od značaja za oba koraka transporta. Deskriptori lipofilnosti, polarnosti, veličine i fleksibilnosti odražavaju sposobnost molekule da se preraspodeli u lipidni dvosloj. Slično, deskriptori lipofilnost i polarnost, deskriptori koji opisuju sposobnost molekule da formira vodonične veze, deskriptori polarnih površina i raspodele naelektrisanja – kvantifikuju karakteristike od ključnog značaja za interakciju molekule sa aminokiselinama vezivne šupljine transportera.

Glavne razlike između deskriptora uključenih u model ANN-24-3 naspram ANN-8-1 ogledaju se u prisustvu deskriptora na čiju vrednost utiču jonizacione karakteristike jedinjenja. U sastav ANN-24-3 ulaze dva deskriptora (QAvgNeg i QAvgPos) koji kvantifikuju prosečno naelektrisanje molekule pri pH od 7,4, što odgovara intracelularnoj sredini. Ustanovljeno je da postoji jasna i statistički značajna razlika ($p < 0.005$) u neto naelektrisanju supstrata i nesupstrata u našem setu, pri čemu molekule koje su pri fiziološkoj pH značajnije zastupljene u jonizovanom obliku imaju daleko manju verovatnoću da budu supstrati Pgp-a. Ovo opažanje ukazuje na mogućnost da dužina zadržavanja u membrani utiče na povećanje verovatnoće transporta posredovanog Pgp-om. Takav zaključak je u skladu sa eksperimentalnim rezultatima koji ukazuju na veću efikasnost transporta neutralnih oblika molekula [392]. Daljom analizom jonizacionih karakteristika jedinjenja u našem setu takođe je uočena pravilnost u pogledu kiselinsko-baznih karakteristika supstrata i nesupstrata. U skladu sa stanovištem da je Pgp transporter prevashodno neutralnih jedinjenja ili katjona, neto naelektrisanje analiziranih jedinjenja bilo je pozitivno za 102 supstrata, nula za 17 i negativno za samo 40 supstrata. Od jedinjenja koja su negativno naelektrisana pri fiziološkoj pH, prosečno neto naelektrisanje svih oblika iznosilo je 0,01 kod supstrata, odnosno 0,20 kod nesupstrata. Izražen diskriminativni karakter

deskriptora jonizacije verovatno doprinosi tačnosti predviđanja modela ANN-24-3, naspram ANN-8-1, i ukazuje na korisnost upotrebe pH-senzitivnih deskriptora u modelovanju transporta posredovanog Pgp-om.

Zajednička karakteristika ANN-8-1 i SVM modela ogleda se u prisustvu deskriptora izvedenih iz rada Ane Seelig o farmakofornim karakteristikama Pgp supstrata [393]. Iako je ovaj rad bio baziran na relativno malom setu od 100 jedinjenja, rezultati naših modela sugerišu da ključne molekulske karakteristike supstrata, identifikovane u tadašnjem radu, značajno doprinose pouzdanosti predviđanja i u daleko širem hemijskom prostoru. U referenci [393] opisane su dve farmakofore: tip I koji uključuje dva elektron–donora na rastojanju od $2,5 \pm 0,3 \text{ \AA}$; i tip II koji čine dva elektron–donora na rastojanju od $4,6 \pm 0,6 \text{ \AA}$ ili tri elektron–donora čiji je međusobni prostorni raspored takav da se dva najudaljenija nalaze na rastojanju od $4,6 \pm 0,6 \text{ \AA}$. U ovom radu dalje je zaključeno da jedinjenja koja poseduju barem jednu farmakoforu tipa I ili jednu farmakoforu tipa II imaju veliku verovatnoću da budu supstrati Pgp-a. Naši rezultati ukazuju da se ovi kriterijumi mogu smatrati minimalnim zahtevom u identifikaciji supstrata, ali i da bi predviđanja bazirana samo na ovim farmakoforima imala veoma nisku specifičnost. Naime, kao što je ilustrovano na slici 4.15, iako je izvesno da supstrati sadrže daleko veći broj grupa tipa I i II, nesupstrati u proseku takođe sadrže barem jednu od ove dve farmakofore u strukturi. Stoga, iako samostalno ovi deskriptori ne mogu biti osnova za pouzdana predviđanja, njihovo uključivanje u složenije modele značajno doprinosi ukupnoj tačnosti predviđanja i istovremeno omogućava razvoj modela sa manjim brojem deskriptora.



Slika 4.15: Vrednosti deskriptora izvedenih iz farmakofora tipa I i II (videti tekst) za supstrate i nesupstrate Pgp-a. Donja i gornja granica svakog pravougaonika odgovaraju prvom i trećem kvartilu raspodele. Medijana je označena horizontalnom crtom. Krajevi vertikalnih linija označavaju minimalne i maksimalne vrednosti.

4.2.2 Modeli za predviđanje transporta posredovanog BCRP-om

Performanse modela razvijenih za predviđanje transporta posredovanog BCRP-om prikazane su u tabeli 4.3. Za razliku od Pgp-a, gde su ANN modeli bili nešto pouzdaniji, SVM model je u slučaju transporta BCRP-om pokazao najveću pouzdanost predviđanja. Najjednostavniji od tri konačna modela, ANN-5-1, pokazao je performanse koje su uporedive sa složenijim ANN-15-1 i SVM modelima, koji su razvijeni bez prethodnog ručnog odabira manjeg broja deskriptora. Ipak, kao i kod analognih Pgp modela, ANN-5-1 se pokazao nešto manje pouzdanim u identifikovanju nesupstrata, tako da ANN-15-1 i SVM modeli imaju bolji odnos senzitivnosti i specifičnosti. Deskriptori uključeni u tri konačna modela navedeni su u prilogu B, tabela B.4.

Tabela 4.3: Performanse konačnih modela za predviđanje transporta posredovanog BCRP-om.

Model	Parametri	Set	Senzitivnost ^a	Specifičnost ^b	Youden indeks ^c	Tačnost ^d
ANN-15-1	30 ANN	trening _{prosek}	0,94	0,94	0,87	94%
	1 neuron ^e	test _{manuelni}	0,75	0,88	0,63	81%
	15 deskriptora	test _{SOM} ^f	0,88	0,86	0,75	87%
	$N_{trening} = 232$	test _{K-means} ^f	0,90	0,92	0,82	91%
	$N_{test} = 78$	test _{prosek}	0,84	0,89	0,73	86%
ANN-5-1	30 ANN	trening _{prosek}	0,95	0,86	0,81	91%
	1 neuron	test _{manuelni}	0,88	0,81	0,69	84%
	5 deskriptora	test _{SOM}	0,96	0,77	0,73	86%
	$N_{trening} = 232$	test _{K-means}	0,92	0,78	0,70	85%
	$N_{test} = 78$	test _{prosek}	0,92	0,79	0,71	85%
SVM	20 SVM	trening _{prosek}	0,94	0,92	0,85	93%
		test _{manuelni}	0,88	0,81	0,69	84%
	12 deskriptora	test _{SOM}	0,89	0,91	0,80	89%
	$N_{trening} = 232$	test _{K-means}	0,95	0,88	0,83	92%
	$N_{test} = 78$	test _{prosek}	0,91	0,87	0,77	88%

^a Udeo tačno predviđenih supstrata u ukupnom broju supstrata

^b Udeo tačno predviđenih nesupstrata u ukupnom broju nesupstrata

^c Izračunat kao: Senzitivnost + Specifičnost - 1

^d Procenat tačnih predviđanja

^e Broj neurona skrivenog sloja

^f Srednja vrednost 10 nezavisnih raspodela seta

Za ABC transportere se pretpostavlja da dele zajednički mehanizam transporta supstrata [394, 395]. Iako su detalji transporta posredovanog BCRP-om na molekularnom nivou relativno nepoznati, verovatno je da između Pgp-a i BCRP-a postoje značajne sličnosti. Moguća razlika se ogleda u činjenici da BCRP potencijalno preuzima supstrate iz regije između dve ovojnice lipidnog dvosloja membrane [396], ali je najveći deo molekularnih odrednica transporta posredovanog Pgp-om

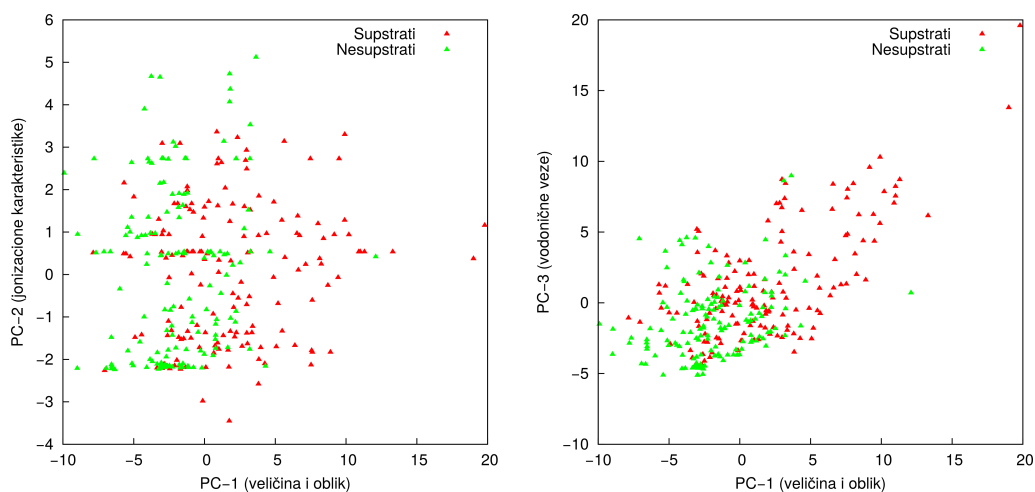
primenjiv i na BCRP [366]. Ipak, poređenje deskriptora iz uspostavljenih modela istovremeno ukazuje da pored sličnosti, postoje i određene razlike u najznačajnijim strukturnim karakteristikama supstrata, što je u skladu sa tek delimičnim preklapanjem supstratne specifičnosti ova dva transportera.

4.2.3 Analiza preklapanja supstratne specifičnosti Pgp-a i BCRP-a

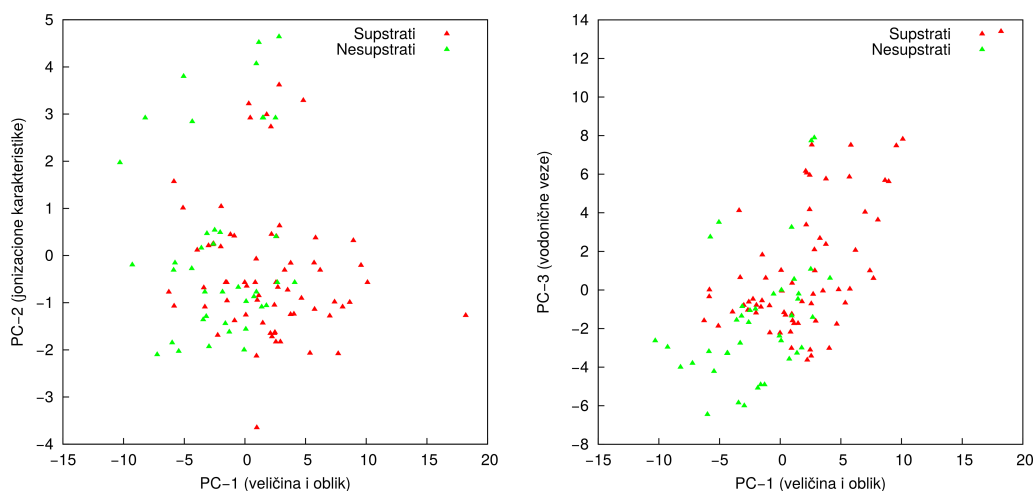
Do danas je u literaturi dokumentovan veliki broj primera jedinjenja koja istovremeno predstavljaju supstrate Pgp-a i BCRP-a. Ovakva jedinjenja češće imaju neadekvatan farmakokinetički profil, stupaju u veći broj interakcija sa drugim lekovima i podložnija su razvoju rezistencije. Ipak, veoma malo je poznato o molekularnoj osnovi supstratnog preklapanja, kao i o tome da li postoje određene strukturne osobine koje su karakteristično prisutne kod dvojnih supstrata ova dva transportera. Primera radi, poređenje ANN-8-1 Pgp modela i ANN-5-1 BCRP modela ukazuje na značajnost prisustva farmakofore tipa I, definisane u referenci [393], za transport posredovan i jednim i drugim transporterom. Ovo zapažanje može ukazati na postojanje zajedničke molekularne osnove za prepoznavanje supstrata, barem u nekom delu hemijskog prostora. Takva sličnost je iznenađujuća, međutim, imajući u vidu da transmembranski delovi ova dva proteina, u okviru kojih se nalazi vezivna šupljina, dele minimalnu identičnost sekvenci [397].

U cilju daljeg istraživanja preklapanja supstratne specifičnosti Pgp-a i BCRP-a, izdvojen je skup jedinjenja koja su bila prisutna u oba seta korišćena za uspostavljanje prediktivnih modela; tačnije, identifikovana su ona jedinjenja čiji je transport okarakterisan na oba transportera. Takvih jedinjenja bilo je ukupno 107 i to: 38 supstrata oba transportera (dvojni supstrati), 25 jedinjenja koja transportuje Pgp ali ne i BCRP, 11 jedinjenja koje transportuje BCRP ali ne i Pgp, i 33 jedinjenja koja nisu supstrati ni Pgp-a ni BCRP-a. Korišćenjem dostupnih molekularnih deskriptora, izvršena je analiza glavnih komponenti i rezultati su prikazani grafički (slike 4.16 i 4.17). Distribucija ovih jedinjenja analizirana je u hemijskom prostoru definisanom sa tri najznačajnije komponente. PC-1 opisuje oblik i veličinu molekule, pri čemu vrednosti PC-1 rastu sa porastom broja atoma u molekuli, njenom površinom i zapreminom. PC-2 opisuje jonizacione karakteristike, pri čemu veće vrednosti ove glavne komponente imaju jedinjenja sa većim brojem kiselih funkcionalnih grupa i ona jedinjenja koja su u većoj meri jonizovana pri fiziološkoj pH. PC-3 opisuje sposobnost molekule da formira vodonične veze, pri čemu veće vrednosti odgovaraju

većem broju atoma koji su akceptori ili donori vodonične veze.



(a) Transport posredovan Pgp-om.



(b) Transport posredovan BCRP-om.

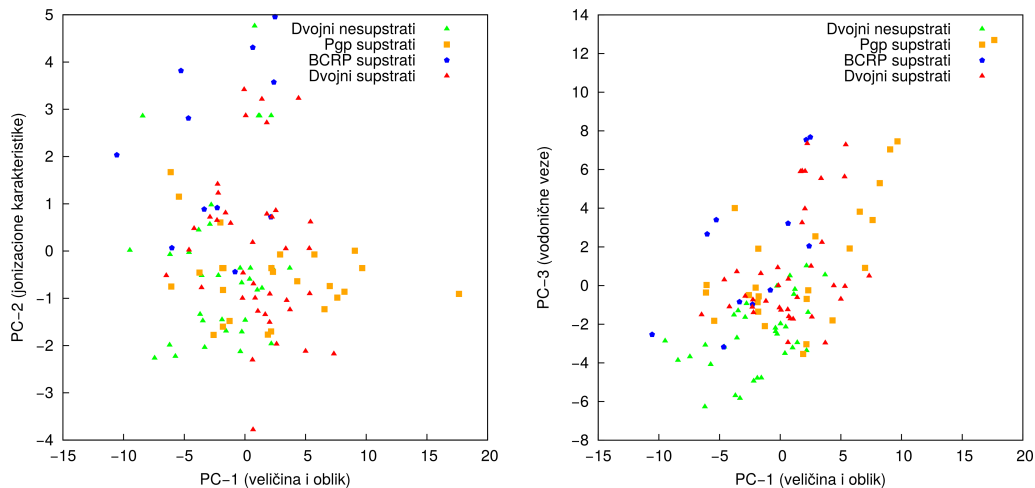
Slika 4.16: Distribucija supstrata i nesupstrata Pgp-a i BCRP-a u hemijskom prostoru definisanom vrednostima tri glavne komponente (PC) opisane u tekstu.

Na slikama 4.16 i 4.17 uočljiva je široka distribucija supstrata u hemijskom prostoru, kao i odsustvo jasne linije razdvajanja supstrata i nesupstrata, što zajedno ilustruje izraženu supstratnu polispecifičnost oba transportera. Ipak, neke pravilnosti se mogu definisati. Pgp je promiskuitetniji transporter od BCRP-a, barem u delu hemijskog prostora obuhvaćenim formiranim setom. Dok Pgp preferencijalno transportuje umereno do izrazito velike molekule, BCRP ispoljava afinitet prema kompaktnijim supstratima. Takođe, BCRP će verovatnije transportovati visoko naelektrisane, anjonske supstrate, bez obzira na njihovu veličinu, dok ovakva jedinjenja nisu dobra supstrata za Pgp. Uopšteno se čini da kiselinobazna priroda jedinjenja

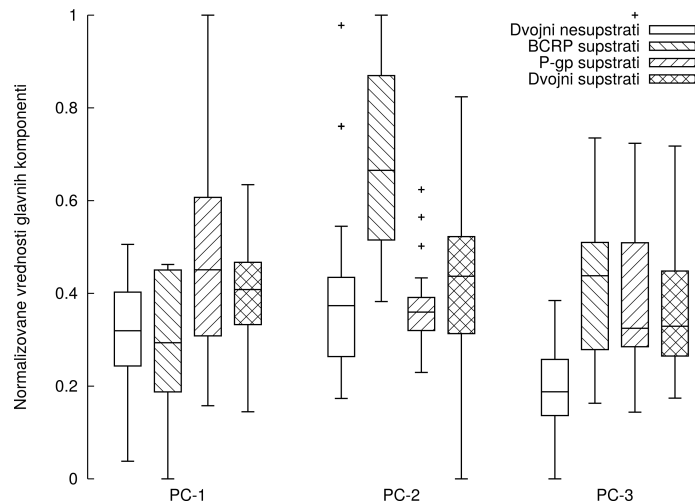
dozvoljava razlikovanje Pgp i BCRP supstrata: Pgp je efikasan transporter baza, dok je BCRP efikasan transporter kiselina. Nijedan od ova dva transportera, naprotiv, nije efikasan u transportu malih, hidrofobnih molekula sa ograničenim brojem donora i akceptora vodoničnih veza. Verovatno je da ovakve molekule ne stupaju u visokoafinitetne interakcije sa vezivnom šupljinom, a pored toga, poseduju i visoku pasivnu permeabilnost. Jedan podset ovakvih molekula posebno je interesantan jer mogu predstavljati inhibitore Pgp-a čija karakterizacija pomoću testova transcelularnog transporta daje nepouzdan rezultate. Primera radi, blokatori kalcijumovih kanala, kao što su verapamil, nifedipin i nitrendipin, pravilno se mogu klasifikovati samo kao „netransportovani supstrati” [242]. Iako stupaju u interakcije sa vezivnom šupljinom transportera, izražena pasivna permeabilnost ovih jedinjenja u visokim koncentracijama dovodi do njegove saturacije, usled čega je stepen njihov transcelularnog transporta prividno nizak. Neslaganje eksperimentalnog rezultata i predviđanja modela u takvim slučajevima mogu poslužiti kao indikacija za testiranje jedinjenja u nižoj koncentraciji.

Iako se uopšteno može reći da jedinjenja niske permeabilnosti (visoke vrednosti PC-2 i PC-3) uglavnom predstavljaju supstrate Pgp-a i BCRP-a, jedan broj dvojnih nesupstrata ukazuje na postojanje izuzetaka od ovog pravila. Sterno kompaktna, polarna jedinjenja kao što su albendazol, etosukcimid, valproat i 2-fluoro-A-85380 mogu u celini da izbegnu transport membranskim transporterima što je verovatno posledica njihovog veoma kratkog zadržavanja u lipidnom dvosloju membrane.

Gore izložena diskusija dobro se ogleda i u vrednostima jednostavnih molekularnih karakteristika 107 analiziranih jedinjenja, zbirno prikazanih u tabeli 4.4. Sveukupno se može zaključiti da Pgp i BCRP ispoljavaju sinergističke i komplementarne uloge u zaštiti tkiva od ksenobiotika, i da unija njihovih supstratnih specifičnosti pokriva izuzetno veliki deo hemijskog prostora. Dvojni nesupstrati ograničeni su na molekule niske molekulske mase i male polarne površine. S druge strane, dvojni supstrati najčešće se nalaze u prostoru umereno velikih molekula koje su neutralne pri fiziološkoj pH. Precizne molekulske karakteristike dvojnih supstrata, kao što je specifičan prostorni raspored donora i akceptora vodoničnih veza, zahtevaju dodatno proučavanje.



(a) Distribucija jedinjenja u hemijskom prostoru.



(b) Distribucija normalizovanih vrednosti glavnih komponenti. Donja i gornja granica svakog pravougaonika odgovaraju prvom i trećem kvartilu raspodele. Medijana je označena horizontalnom crtom. Krajevi vertikalnih linija označavaju minimalne i maksimalne vrednosti. Tačke predstavljaju odstupajuće vrednosti.

Slika 4.17: Preklapanje supstratne specifičnosti Pgp-a i BCRP-a, ilustrovano kroz distribuciju odabranih jedinjenja u hemijskom prostoru definisanom vrednostima tri glavne komponente (PC) opisane u tekstu.

4.2.4 Eksterna validacija modela

Rigorozan pristup prikupljanju i proceni eksperimentalnih podataka prilikom formiranja setova korišćenih u ovoj studiji predstavlja vremenski-zahtevan proces i u velikoj meri ograničava veličinu seta koji je moguće pripremiti. Nedavno je

Tabela 4.4: Vrednosti jednostavnih molekularnih karakteristika dvojnih supstrata i nesupstrata Pgp-a i BCRP-a, odnosno jedinjenja koje transportuje samo jedan od ova dva transportera.

Supstrat?		Osobina (srednja vrednost \pm standardna devijacija; opseg dat u zagradi)						
Pgp	BCRP	MW	MlogP	PSA	HBD	HBA	QAvgNeg	QAvgPos
ne	ne	345 \pm 100 (141–501)	2,60 \pm 0,95 (0,39–4,32)	61 \pm 31 (3–130)	1,4 \pm 1,4 (0–4)	4,3 \pm 1,6 (1–7)	0,18 \pm 0,46 (0–1,93)	0,30 \pm 0,40 (0–0,98)
da	ne	523 \pm 99 (232–1203)	1,74 \pm 1,36 (-0,85–4,36)	115 \pm 57 (44–279)	3 \pm 1,3 (0–6)	8,2 \pm 4,5 (4–23)	0,03 \pm 0,10 (0–1)	0,24 \pm 0,32 (0–1,98)
ne	da	345 \pm 120 (130–544)	1,03 \pm 1,54 (0,39–4,32)	115 \pm 54 (49–211)	2,8 \pm 1,5 (1–6)	7,2 \pm 3,5 (2–13)	0,66 \pm 0,70 (0–1,98)	0,01 \pm 0,03 (0–0,78)
da	da	440 \pm 235 (234–657)	1,43 \pm 1,24 (0,39–4,32)	108 \pm 40 (48–206)	2,7 \pm 1,5 (1–8)	7,5 \pm 2,8 (3–13)	0,15 \pm 0,33 (0–1)	0,29 \pm 0,45 (0–1)

MW – molekulska težina; MlogP – logP izračunat metodom po Moriguchi-ju;

PSA – polarna površina; HBD – broj donora vodonične veze; HBA – broj akceptora vodonične veze;

QAvgNeg – srednja vrednost formalno negativnog naelektrisanja svih jonizovanih oblika molekule;

QAvgPos – srednja vrednost formalno pozitivnog naelektrisanja svih jonizovanih oblika molekule;

opisan alternativni pristup koji se oslanja na veliku količinu bioloških podataka koji su javno dostupni a predstavljaju rezultate testiranja raznovrsnih eksperimentalnih jedinjenja na panelu ćelijskih linija koje je sproveo američki Nacionalni institut za kancer u sklopu Programa razvoja novih terapeutika (NCI-DTP). Korišćenjem podataka o citotoksičnosti testiranih jedinjenja iz NCI-DTP programa i uzimajući u obzir nivoe ekspresije Pgp-a u tretiranim ćelijskim linijama, Levatić i saradnici su nedavno objavili do sada najveći set podataka o transportu posredovanom Pgp-om koji, dodatno, pokazuje minimalno preklapanje sa drugim setovima iz literature [369]. U eksternoj validaciji razvijenih Pgp modela, odabran je upravo ovaj set budući da pruža uvid u performanse modela u klasifikaciji veoma raznovrsnih jedinjenja. Pored toga, korišćenje ovog seta i interpretacija dobijenih rezultata pružili su uvid u to kako klasifikacija jedinjenja bazirana na različitim eksperimentalnim podacima utiče na predviđanja računarskih modela. U validaciji BCRP modela, takođe je korišćen najraznovrsniji set koji je u datom trenutku bio dostupan u literaturi. Rezultati eksterne validacije sumirani su u tabeli 4.5.

Sa aspekta performansi Pgp modela, primetno je da senzitivnost predviđanja ostaje na nivou koji je ustanovljen prilikom internog testiranja (tabela 4.2), ali da je broj jedinjenja koja su pogrešno klasifikovana kao Pgp supstrati izrazito povećan. Povećanje udela lažno pozitivnih rezultata može biti posledica načina na koji su

Tabela 4.5: Rezultati eksterne validacije razvijenih modela.

Model	U domenu primenjivosti	TP	TN	FP	FN	Se	Sp	Y	Ac
Pgp modeli (broj jedinjenja seta za validaciju: 861)									
ANN-24-3	744 (86%)	254	257	161	72	0,78	0,61	0,39	69%
ANN-8-1	780 (90%)	317	204	225	34	0,90	0,48	0,38	68%
SVM	623 (72%)	213	208	155	47	0,82	0,58	0,40	68%
BCRP modeli (broj jedinjenja seta za validaciju: 94)									
ANN-15-1	80 (85%)	35	23	4	18	0,66	0,85	0,51	72%
ANN-5-1	93 (99%)	39	22	5	14	0,74	0,81	0,55	76%
SVM	80 (85%)	36	21	6	16	0,69	0,78	0,47	72%

TP – stvarno pozitivni; TN – stvarno negativni; FP – lažno pozitivni; FN – lažno negativni; Se – senzitivnost; Sp – specifičnost; Y – Youden indeks; Ac – tačnost.

jedinjenja klasifikovana u supstrate i nesupstrate. Naime, korišćenjem eksperimentalnih podataka o citotoksičnosti suštinski se ne uzima direktno u obzir mogućnost interakcije sa transporterom, već samo sposobnost jedinjenja da ostvari značajnu koncentraciju u ćelijama koje ekspimiraju Pgp. Stoga jedinjenja sa visokom pasivnom permeabilnošću, koja mogu da prezasiť transporter, mogu dati ćelijski odgovor koji je konzistentan sa efektima nesupstrata, iako potencijalno stupaju u interakcije sa Pgp-om. Budući da je u ovoj studiji usvojena definicija supstrata (odjeljak 3.2.1) koja odgovara nameri da se izvrši modelovanje interakcije sa transporterom, a ne samo neto efekat na intracelularnu koncentraciju jedinjenja, razlike u tome kako se jedinjenje klasifikuje kao nesupstrat mogu biti uzrok povećanog broja lažno pozitivnih predviđanja.

U cilju daljeg istraživanja ovog potencijalnog uzroka smanjene specifičnosti predviđanja modela, izvršena je analiza distribucije izračunatih $\log D$ vrednosti svih jedinjenja pri pH 7,4 ($\log D_{7,4}$). Ustanovljeno je da postoje statistički značajne razlike u vrednostima $\log D_{7,4}$ (jednofaktorska analiza varijanse (ANOVA), $p < 0,001$) u grupama tačno i pogrešno klasifikovanih supstrata i nesupstrata iz seta za validaciju, dok takve razlike nisu prisutne u samom setu korišćenom za uspostavljanje modela ($p = 0,38$). Preciznije, lažni supstrati imaju prosečnu $\log D_{7,4}$ vrednost od 3,29 što je više od prosečnih vrednosti ovog parametra u ostalim grupama (stvarni supstrati, srednja vrednost $\log D_{7,4}$ od 2,16, $p < 0,001$; stvarni nesupstrati, srednja vrednost $\log D_{7,4}$ od 2,97, $p = 0,09$; lažni nesupstrati, srednja vrednost $\log D_{7,4}$ od 1,97, $p < 0,001$). Uloga pasivne permeabilnosti u ograničavanju efekata Pgp-a na transcelularni transport jedinjenja dobro je poznata i veruje se da doprinosi gubitku pouzdanosti većine modela dostupnih u literaturi [398]. U ovom kontekstu, pojedini autori su predložili i mogućnost uključivanja netransportovanih supstrata (jedinje-

nja koja stupaju u interakciju sa transporterom ali prevazilaze njegov transportni kapacitet) kao zasebne kategorije u prediktivne modele [398] jer bi na ovaj način potencijalno bilo moguće postići bolji odnos između parametara modela koji opisuju molekularne interakcije i onih koji oslikavaju pasivnu permeabilnost jedinjenja.

Dodatan faktor čiji je uticaj ispitivan bila je i definicija domena primenjivosti. Naime, neadekvatna definicija može usloviti pogoršanje performansi ukoliko jedinjenja iz seta za validaciju pripadaju delu hemijskog prostora gde testirani modeli nisu primenjivi. Pristup baziran na opsezima vrednosti deskriptora, korišćen za razvijene modele, ima slabost koja postaje očigledna ukoliko u okvirima definisanih opsega postoje značajni „prazni prostori”. Jedinjenje koje pripada ovom „praznom prostoru” naizgled ulazi u domen primenjivosti modela, iako zapravo predviđanja njegovih karakteristika jesu daleko manje pouzdana [399, 400]. Iz ovog razloga, pokušano je definisanje domena primenjivosti na dva alternativna načina, bazirana na hemijskoj sličnosti [401] i Euklidskim udaljenostima u hemijskom prostoru [402]. Korišćenje ovih pristupa, međutim, nije dovelo do poboljšanja pouzdanosti predviđanja u redefinisanim domenima primenjivosti (rezultati prikazani u prilogu B, tabela B.5), usled čega su zadržane jednostavnije, prethodno opisane definicije.

Sveukupno, može se zaključiti da je eksterna validacija modela potvrdila njihove prediktivne sposobnosti. Očekivano, utvrđena pouzdanost je nešto niža u poređenju sa internim setom za testiranje. Ipak, poređenje modela iz literature (pregledno prikazani u referenci [403]) sa računarskim modelima razvijenim u ovoj studiji ukazuje da je dostignuti nivo pouzdanosti predviđanja u celosti uporediv sa većinom publikovanih modela globalne primenjivosti. Posebno u slučaju BCRP modela, koji su zasnovani na relativno malom setu podataka, unapređenje pouzdanosti se može očekivati njihovim ponovnim uspostavljanjem uz korišćenje dodatnih eksperimentalnih podataka, kada oni postanu dostupni. Na osnovu uporednih performansi modela razvijenih u ovoj studiji, može se zaključiti da su za dalju primenu optimalan izbor modeli ANN-24-3, za transport posredovan Pgp-om, odnosno ANN-5-1, za transport posredovan BCRP-om.

4.2.5 Rezime

U ovom delu istraživanja, razvijeni su i validirani modeli za klasifikaciju jedinjenja u supstrate i nesupstrate Pgp-a i BCRP-a, kao dva membranska transportera koji mogu ključno uticati na farmakokinetiku i efikasnost antitumorskih

lekova. Ispitivanjem pouzdanosti modela na eksternim setovima za validaciju, identifikovana su dva modela sa optimalnim prediktivnim karakteristikama koja će biti korišćena u daljem istraživanju, za predviđanje mogućnosti transporta novodizajniranih jedinjenja. Analiza supstratne specifičnosti oba transportera, međutim, još jednom je ukazala na njihovu izraženu promiskuitetnost kao i na preklapanje njihovih supstratnih specifičnosti u značajnom delu hemijskog prostora. Ovaj nalaz upućuje da je optimizacija strukture u cilju eliminisanja afiniteta za Pgp i BCRP veliki izazov. Kao što su drugi autori prethodno zaključili, onda kada se farmakofora od značaja za aktivnost preklapa sa pretpostavljenim strukturnim odrednicama transporta posredovanog Pgp-om, najpraktičniji pristup optimizaciji može biti povećanje pasivne permeabilnosti jedinjenja [404]. I pored mogućnosti interakcije sa transporterom, na jedinjenja koja poseduju izraženu pasivnu permeabilnost i koja su aktivna u niskim koncentracijama, membranski transporteri ne moraju ispoljiti značajne efekte. Ipak, interpretabilnost modela razvijenih u ovoj studiji ostavlja mogućnost racionalne i specifične modifikacije strukture u cilju smanjenja rizika od značajnije interakcije sa Pgp-om i BCRP-om. Primer ovog pristupa i primenjivosti razvijenih modela biće ilustrovan na novodizajniranim strukturama u odeljku 4.3.4.

4.3 Dizajn novih potencijalnih inhibitora EZH2

4.3.1 Implementacija algoritma za *de novo* dizajn hemijskih struktura

Osnovni koncept

De novo dizajn postaje sve popularniji pristup u računarskom dizajniranju jedinjenja jer omogućava formiranje fokusiranih kolekcija uz istraživanje delova hemijskog prostora koji najčešće nisu zastupljeni u postojećim kolekcijama komercijalno dostupnih jedinjenja. Ovaj pristup dizajnu takođe može omogućiti prevazilaženje patentnih barijera u otkriću novih lekova [273]. Implementacija algoritma za *de novo* dizajn mora da obezbedi adekvatan odgovor na nekoliko glavnih izazova u primeni ove metode [276]: (i) kako pristupiti „kombinatorijskoj eksploziji” potencijalnih rešenja u hemijskom prostoru; (ii) kako obezbediti sintetsku dostupnost dizajniranih struktura; (iii) kako pristupiti optimizaciji u kontekstu činjenice da dizajn potencijalno bioaktivnih struktura najčešće podrazumeva zadovoljenje međusobno suprotstavljenih ciljeva. U literaturi su opisane desetine algoritama za *de novo* dizajn, od kojih svaki na specifičan način odgovara na ove izazove, pa je implementacija algoritma u ovoj studiji izvršena na osnovama pojedinih elemenata prethodno opisanih algoritama. Osnovni zahtevi koje je novi algoritam trebalo da ispuni bili su: (i) da razmatranje sintetske dostupnosti jedinjenja bude ključan element generisanja novih molekula; (ii) da implementirana funkcija procene koristi podatke o ključnim odrednicama aktivnosti poznatih inhibitora, koji su prikupljeni simulacijama MD u prvom delu ove studije.

U skladu sa svojim osnovnim karakteristikama, algoritam implementiran u ovoj studiji u daljem tekstu će biti označavan akronimom Phabblor (engl. *Pharmacophore-based Building Block Assembler*).

Dva algoritma opisana u literaturi koriste *in silico* reakcije kao metod generisanja novih molekula, što značajno uvećava sintetsku dostupnost uobičajenog rešenja: DOGS [279] i AutoGrow [296]. Kao funkcija procene, međutim, u DOGS algoritmu se koristi mera topološke sličnosti poznatih inhibitora i generisanih molekula, usled čega nije moguće korišćenje dostupnih informacija o trodimenzionalnog strukturi kompleksa ciljnog mesta i liganda. Dok je DOGS-u implementirano preko 50 reakcija, AutoGrow metoda je bazirana isključivo na reakcijama „klik hemije”

što potencijalno ograničava raznovrsnost generisanih struktura. S druge strane, AutoGrow koristi rezultate dokinga kao funkciju procene, što omogućava korišćenje trodimenzionalne strukture receptora. Ipak, više retrospektivnih studija pokazalo je da korišćenje dokinga u identifikaciji aktivnih molekula virtuelnim skriningom daje lošije rezultate i od molekulskih otisaka prstiju i od pristupa baziranih na farmakoforama [387, 405, 406]. Stoga Phabblar implementira samo pristup generisanju novih molekula iz ova dva algoritma i koristi *in silico* reakcije u spajanju fragmenata kao metodu generisanja struktura koje bi sintetski trebalo da budu dostupne.

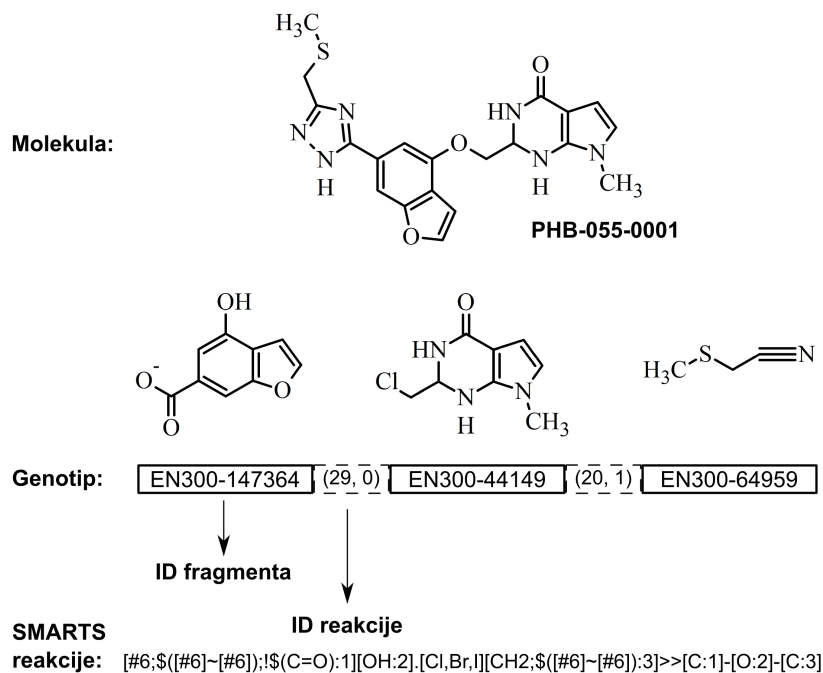
Imajući u vidu prethodno navedenu činjenicu da metode zasnovane na poređenju oblika i farmakofora imaju veću uspešnost od molekulskog dokinga, osnovna funkcija procene Phabblar algoritma upravo predstavlja kombinovanu meru farmakoforne i sličnosti oblika generisane i referentne molekule. Ovakav izbor osnovne funkcije procene omogućava poptuno iskorišćenje prethodno uspostavljene strukturne osnove aktivnosti poznatih inhibitora. Ipak, ovaj metod zahteva generisanje konformnog ansambla svake novogenerisane molekule, a uključuje i računarski relativno zahtevan korak superponiranja svakog od konformera sa referentnom farmakoforom. Ova činjenica je značajna jer računarska zahtevnost procene prilagođenosti rešenja utiče na izbor metaheuristike koja će biti upotrebljena u optimizaciji. Primera radi, „pohlepni algoritam” koji koristi DOGS podrazumeva veliki broj evaluacija prilagođenosti, što je praktično samo ukoliko je funkcija procene računarski efikasna.

Dodatan faktor u izboru metaheuristike predstavljala je i potreba za istovremenom optimizacijom više ciljeva dizajna. Primera radi, sa porastom molekulske mase, broja donora i akceptora vodoničnih veza – raste i prosečan broj farmakofornih tačaka, pa i verovatnoća boljeg slaganja sa referentnom farmakofornom hipotezom. S druge strane, ovakve molekule sintetski su složenije i češće poseduju neadekvatne biofarmaceutske karakteristike. Stoga je odlučeno da algoritam treba da obezbedi podršku za optimizaciju više ciljeva, a konačan izbor metaheuristike predstavljao je genetski algoritam. Preciznije, Phabblar predstavlja specijalizovanu implementaciju dobro poznatog NSGA-II algoritma. Jedinstvenost svake implementacije predstavljaju način kodiranja genotipa rešenja i implementacija genetskih operatora koji su u daljem tekstu detaljnije opisani.

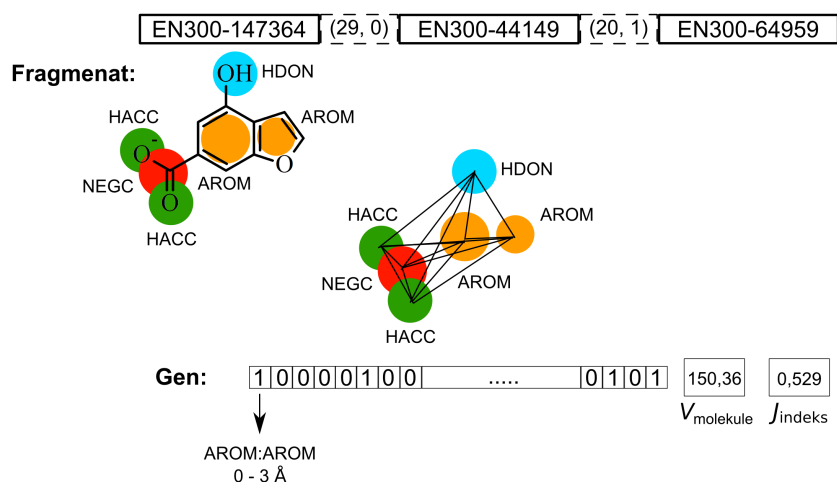
Genetska reprezentacija molekula

Kao što je prethodno istaknuto, generisanje novih molekula u okviru Phab-
bler-a odigrava se *in silico* reakcijama spajanja odabranih fragmenata iz odgovara-
juće kolekcije komercijalno dostupnih jedinjenja. Način kodiranja genotipa rešenja
prilagođen je ovoj osnovnoj ideji algoritma, pa se svako potencijalno rešenje pred-
stavlja kao niz jedinstvenih identifikacionih brojeva fragmenata iz kolekcije i reakcija
upotrebljenih za njihovo *in silico* spajanje (slika 4.18a).

Kako bi se omogućilo ukrštanje rešenja i njihova mutacija, svaki gen (frag-
menat) u genotipu poseduje i odgovarajuću numeričku reprezentaciju kojom su ko-
dirane njegove ključne molekulske karakteristike (slika 4.18b). Prvi deo gena čini
binarni vektor dužine 108 bitova koji predstavlja farmakoforni otisak prsta frag-
menta. Otisak prsta generiše se u dve faze. U prvoj fazi, u trodimenzionalnoj
strukturi fragmenta identifikuju se farmakoforne tačke koje mogu biti aromatični
centri, hidrofobne grupe, donori i akceptori vodonične veze, odnosno pozitivno i
negativno naelektrisane funkcionalne grupe. U drugoj fazi, procenjuje se prisustvo
farmakofornih dubleta na fiksnim rastojanjima od 0 do 3, od 3 do 5 i od 5 do 7,5 Å, i
odgovarajući bit se podešava u skladu sa odsustvom ili prisustvom tog dubleta. Iako
je, u principu, bilo moguće koristiti i farmakoforne triplete odnosno kvartete, kao
i veći broj odabranih rastojanja, dobijeni otisci prsta su za molekule veličine frag-
menta slabo popunjeni usled čega je u trenutnoj implementaciji korišćen relativno
mali otisak prsta. Preostale dve numeričke vrednosti koje čine gen imaju decimalne
vrednosti i predstavljaju molekulsku zapreminu i Balabanov J indeks. Korišćenjem
ovakve genetske reprezentacije, gen sadrži informacije o farmakofornim karakteristi-
kama fragmenata, njihovoj veličini i obliku, i to u formatu pogodnom za primenu
uobičajenih genetskih operatora. Nadalje, odabrana reprezentacija omogućava da
se mutacijama i ukrštanjem istražuje prostor koji u najvećoj meri odgovara funk-
ciji procene zasnovane na sličnosti dizajnirane molekule definisanoj farmakofornoj
hipotezi. Stoga je evolucija rešenja prevashodno orjentisana na maksimizaciju far-
makoforne sličnosti dizajniranih molekula sa datom hipotezom. Međutim, budući da
NSGA-II počiva na konceptu Pareto optimalnosti, rešenja koja su bolja u kontekstu
drugih funkcija prilagođenosti ostaju u populaciji i doprinose njenoj raznovrsnosti.



(a) Genotip molekule.



(b) Numerička reprezentacija gena.

Slika 4.18: Ilustracija predstavljanja genotipa molekule u Phabblers algoritmu.

Mutacija

Mutacija je u Phabblers-u implementirana kao genetska operacija kojom se stohastički menja postojeći genotip: (i) bez promene njegove dužine (klasična mutacija); (ii) kroz uvođenje dodatnog gena (duplikacija). U slučaju da se genotip odabrane individue podvrgava duplikaciji, prvo se generiše skup svih reakcija u kojima data molekula može da učestvuje. Iz tog skupa, nasumično se bira jedna

reakcija i definiše reaktivnost kompatibilnog fragmenta koji se može dodati na postojeću molekulu. Celokupna kolekcija fragmenata se zatim pretražuje i iz skupa svih kompatibilnih fragmenata nasumično se bira jedan. Kako se reaktivnost svih fragmenata u kolekciji unapred procenjuje, ova operacija traje relativno kratko. Nakon izbora kompatibilnog fragmenta, gen se dodaje u postojeći genotip i vrši se *in silico* spajanje.

Klasični operator mutacije implementiran je tako da menja samo jedan gen u genotipu. Nad binarnim vektorom slučajno odabranog gena vrši se obrtanje nasumično odabranih bitova (maksimalno jedne desetine celokupnog vektora) dok se decimalne vrednosti u genu podvrgavaju Gausovskoj konvoluciji. Mutirani gen predstavlja prototip jer ne odgovara nužno ni jednom fragmentu dostupnom u kolekciji. Dodatno, mutacija može da izmeni reaktivnost fragmenta uslovljavajući nemogućnost njegovog spajanja sa drugim genima u genotipu roditeljskog rešenja. Stoga u Phabblu-u mutacija može imati konzervativan ili segregativan karakter. U slučaju konzervativne mutacije, prvo se identifikuje skup fragmenata iz kolekcije čija je reaktivnost kompatibilna sa ostalim fragmentima roditeljskog rešenja. Iz ovog skupa se zatim bira fragmenat koji je najbliži mutiranom genu (prototipu fragmenta) i on zamenjuje dati gen u genotipu mutirane individue. Udaljenost između fragmenata se računa sabiranjem Tanimoto rastojanja binarnih vektora sa razlikama vrednosti molekulskih zapremina i J indeksa, normalizovanih tako da uzimaju vrednosti iz intervala $[0, 1]$ u datom skupu gena. Efektivno, konzervativna mutacija dovodi samo do zamene jednog fragmenta u molekuli drugim. U slučaju segregativne mutacije, iz celokupne kolekcije fragmenta se identifikuje gen najbliži prototipu, sa jednim ograničenjem da odabrani fragmenat mora sadržati broj reaktivnih centara koji nije manji od ovog broja u prvobitnom fragmentu. Ovim uslovom se obezbeđuje da mutirana individua može sadržati isti broj gena kao i roditeljsko rešenje. Nakon segregativne mutacije, jedan ili više gena roditeljskog rešenja može biti nekompatibilno sa mutiranim genom (nema reakcije kojom se mogu spojiti). U tom slučaju, iz kolekcije fragmenata se identifikuje najsličnija zamena (udaljenost se računa kao što je prethodno opisano) koja poseduje neophodnu reaktivnost. Nova molekula se zatim rekonstruiše praćenjem sekvence reakcija kodirane mutiranim genotipom. Na ovaj način, nakon segregativne mutacije, u novoj molekuli je zamenjen barem jedan a potencijalno i svi fragmenti.

Relativna verovatnoća da operator mutacije dovede do duplikacije, konzervativne ili segregativne mutacije može se podešavati. U trenutnoj implementaciji, odnos verovatnoća je 0,9:0,67:0,5, imajući u vidu da segregativna mutacija dovodi

do najvećih promena rezultujuće molekule i da je računarski najmanje efikasna.

Ukrštanje

Ukrštanje genotipa dva rešenja započinje sparivanjem gena između kojih će doći do razmene. Kako položaj gena u genotipu (lokus), na način na koji je on predstavljen u Phabblor-u, ne odražava sličnost između gena, sparivanje se vrši Mađarskom metodom [407] pri čemu se udaljenost gena računa kao što je opisano u prethodnom odeljku. Budući da genotipovi mogu imati različitu dužinu, nesporeni geni ostaju nepromenjeni. U okviru svakog para gena, binarni vektori se ukrštaju uniformno, sa verovatnoćom od 0,5 za razmenu bitova. Decimalni deo gena ukršta se intermedijernom rekombinacijom. Rekombinantni geni, slično mutiranim, predstavljaju prototipove. U slučaju ukrštanja, najbliži fragmenat se, međutim, uvek traži tako da poseduje istu (ili širu) reaktivnost u odnosu na odgovarajući gen roditeljskog rešenja. Nakon ukrštanja svih sporenih gena, molekule potomaka rekonstruišu se praćenjem sekvence reakcija kodirane ukrštenim genotipovima.

Funkcije procene prilagođenosti

NSGA-II omogućava upotrebu proizvoljnog broja funkcija procene, ali je poznato da uključivanje više od 4 faktora dovodi do značajnog smanjenja efikasnosti optimizacije. Pored osnovnog cilja optimizacije – maksimizacije farmakoforne sličnosti – optimizacija u Phabblor-u može da podrazumeva i maksimizaciju topološke sličnosti sa poznatom aktivnom molekulom (ili molekulama), minimizaciju sintetske složenosti i minimizaciju broja molekulskih karakteristika koje su često udružene sa neadekvatnim biofarmaceutskih svojstvima.

Uključivanje funkcije procene topološke sličnosti u optimizaciju može imati različite efekte. S jedne strane, ovakva funkcija može dovesti do generisanja rešenja koja su isuviše slična poznatim inhibitorima odnosno matrici dizajna, što nije poželjno. S druge strane, više studija je pokazalo da molekulski otisci prstiju zasnovani na topologiji molekule imaju značajan i često potcenjen potencijal za identifikaciju bioizosternih zamena [408]. Imajući ovu činjenicu u vidu, uključivanje ovakve funkcije procene u optimizaciju može doprineti ranoj identifikaciji bioizostera privilegovane osnovne strukture i time može ubrzati konvergenciju algoritma. U okviru ovog istraživanja, pokušana su izvršenja Phabblor-a sa i bez ove funkcije, a efekti su

diskutovani u odeljku 4.3.1.

Iako broj gena pojedinačnog rešenja posredno odražava njegovu složenost i broj reakcija između fragmenata neophodnih za *in silico* „sintezu”, ovaj broj je relativno loša mera stvarne sintetske složenosti date molekule. Između ostalog, čak i ukoliko je sintetski put upotrebljen u Phabblor-u direktno izvodiv, veliki broj specifičnih gradivnih jedinica (fragmenata) koje bi morale da budu na raspolaganju čini ga nepraktičnim i ekonomski skupim. Iz ovog razloga je u praćenju sintetske kompleksnosti rešenja upotrebljena funkcija bazirana na SMCM proceni koju su predložili Allu i Oprea [374].

Konačno, broj kršenja pravila Lipinskog treba da obezbedi eliminaciju onih rešenja koja imaju rizik od posedovanja neadekvatnih biofarmaceutskih karakteristika.

Druge izmene NSGA-II

U okviru poređenja prilagođenosti dve individue u populaciji, NSGA-II pored Pareto ranga uzima u obzir i *crowding distance* parametar koji odgovara udaljenosti date individue od drugih individua na istom Pareto frontu, u prostoru svih ciljeva optimizacije. Upotreba *crowding distance* parametra treba da obezbedi ravnomernu raspoređenost optimalnih rešenja duž Pareto fronta, odnosno posredno favorizuje očuvanje diverziteta populacije. U dizajnu optimalnog seta struktura, raznovrsnost rešenja je svrsishodnije definisati u hemijskom a ne prostoru vrednosti funkcija procene. Iz ovog razloga, u Phabblor-u je dodeljivanje *crowding distance* vrednosti implementirano kao mera strukturne različitosti dva rešenja, procenjene poređenjem FCFP6 molekulskih otisaka prstiju. Sličan pristup prvobitno su opisali Kruisselbrink i saradnici [409]. Na ovaj način, *crowding distance* vrednost odgovara Tanimoto udaljenosti datog rešenja od njemu najbližijeg rešenja na istom Pareto frontu. Time se dodatno favorizuje očuvanje raznovrsnosti populacije tokom evolucije rešenja.

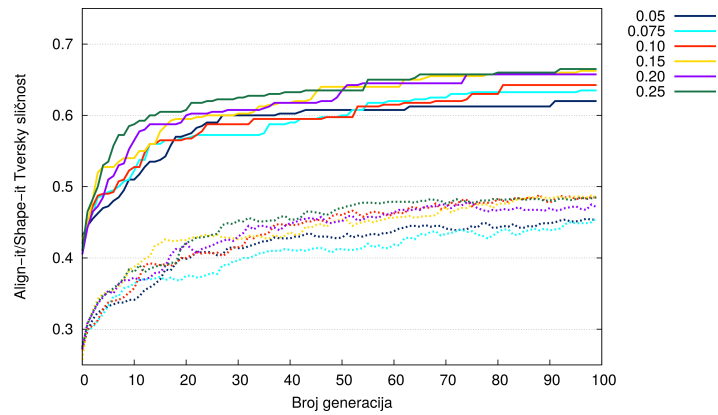
Parametri algoritma

U literaturi je poznato da nije moguće analitički obezbediti dokaz da primena NSGA-II, sa bilo kojim izborom parametara i/ili funkcija procene, obezbeđuje konvergenciju do Pareto-optimalnog skupa rešenja [410]. Međutim, kako iz-

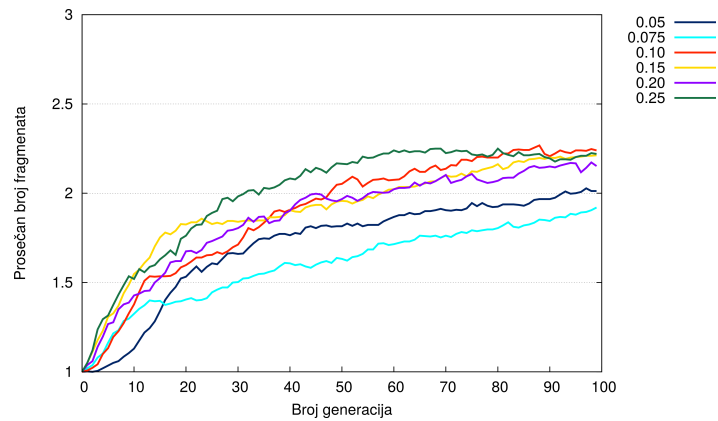
bor verovatnoće mutacije i ukrštanja nesumnjivo utiče na konvergenciju algoritma i kvalitet finalne populacije, kroz veliki broj preliminarnih izvršenja pokušana je empirijska identifikacija optimalnih vrednosti ovih parametara. Takođe je ispitivano kako različite funkcije procene utiču na evoluciju rešenja. Suštinski, izbor parametara Phabblers-a jeste složen problem hiperoptimizacije, čiji je rezultat osetljiv i na faktore čiji uticaj nije ispitivan, kao što su: sastav i veličina kolekcije fragmenata, variranje referentne farmakofore, interakcije između različitih faktora i slično. Stoga su prikupljeni rezultati pre svega bili korisni u definisanju vrednosti koje će biti korišćene u izvršenjima Phabblers-a u ovoj studiji, i ne treba ih posmatrati kao generalne zaključke o optimalnom setu parametara Phabblers algoritma.

U prvom koraku, kao faktor je varirana verovatnoća mutacije, dok je verovatnoća ukrštanja fiksirana na vrednost od 0,80, broj jedinki u populaciji bio je 100, optimizacija je trajala 100 generacija, pri čemu su korišćene sve četiri funkcije prilagođenosti, odnosno sva 4 cilja optimizacije. Verovatnoća mutacije varirana je u opsegu od 0,05 do 0,25, a za svaki izbor parametara izvršenje algoritma ponovljeno je 4 puta i analizirane su medijane. Evolucija prosečne i maksimalne farmakoforne sličnosti rešenja u populaciji sa referentnom farmakoforom ilustrovana je na slici 4.19a. Primetan je uravnotežen trend konvergencije, bez skokova, koji je prisutan nezavisno od izbora verovatnoće mutacije i oslikava elitističku prirodu algoritma. Iako se može zaključiti da su sva izvršenja algoritma konvergirala ka rešenjima približno iste prilagođenosti, treba imati u vidu da je odnos između farmakoforne sličnosti i verovatnoće da je identifikovana molekula podjednako aktivna kao referentna – zapravo eksponencijalan (slika C.1 u prilogu C). Stoga male promene u vrednosti farmakoforne sličnosti iznad 0,6 odgovaraju značajnijem porastu verovatnoće identifikovanja aktivne molekule. Sa ovom činjenicom u vidu, može se zaključiti da nešto veća vrednost verovatnoće mutacije doprinose bržoj konvergenciji algoritma kao boljim rešenjima. Ovo je verovatan efekat bržeg uvećanja prosečne veličine molekule duplikacijama (slika 4.19b). Međutim, učestalost mutacija utiče pre svega na brzinu konvergencije, a ne nužno i na optimalnost konačnog skupa rešenja. Tako trajanje optimizacije od 200 generacija omogućava dostizanje farmakoforne sličnosti od $\approx 0,7$ i uz najnižu verovatnoću mutacije od 0,05. Ipak, kao što je prikazano na slici 4.19c, veoma visoke i veoma niske vrednosti verovatnoće mutacije povezane su sa najizraženijim smanjenjem raznovrsnosti populacije. U prvom slučaju, verovatno je da značajan deo potomaka nastalih mutacijama biva odbačen pre ukrštanja sa postojećim članovima populacije, dok je u drugom slučaju učestalost mutacija nedovoljna za uvođenje dodatne raznovrsnosti u populaciju. Na temeljima ove analize, može se zaključiti da vrednost učestalosti mutacija od 0,10 predstavlja dobar kom-

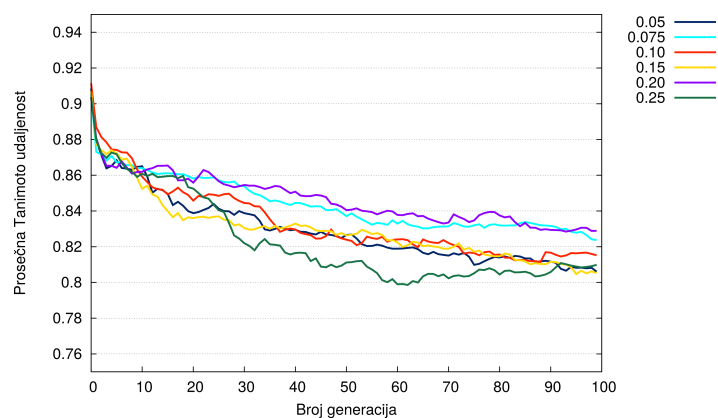
promis između brzine konvergencije, održanja raznovrsnosti populacije i dostizanja željene farmakoforne podudarnosti sa matricom dizajna.



(a) Prosečna (isprekidane linije) i maksimalna (pune linije) sličnost između rešenja u populaciji i zadate farmakoforne hipoteze, u funkciji trajanja optimizacije.



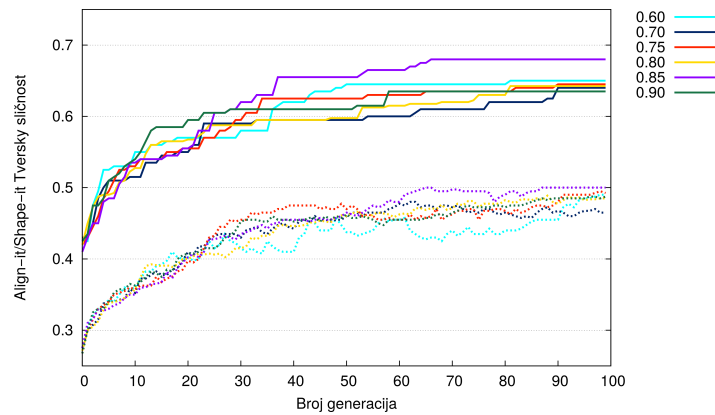
(b) Prosečan broj fragmenata u molekuli, u funkciji trajanja optimizacije.



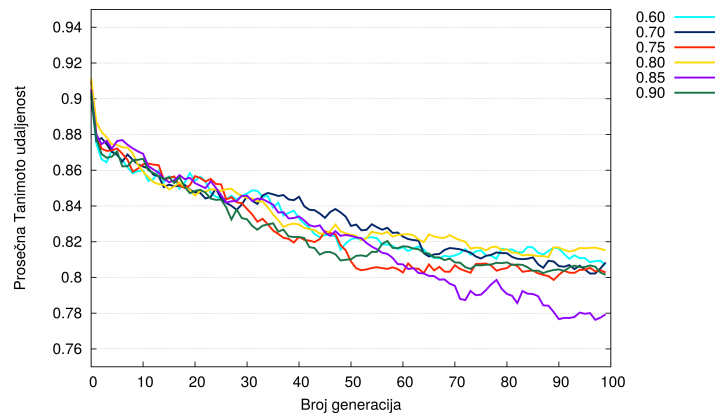
(c) Prosečna Tanimoto udaljenost između rešenja u populaciji, u funkciji trajanja optimizacije.

Slika 4.19: Uticaj izbora verovatnoće mutacije na neke aspekte konvergencije Phab-
bler algoritma. Različite boje linija odgovaraju različitim verovatnoćama mutacije
koje su testirane, kao u legendi.

U drugom koraku, kao faktor je varirana verovatnoća ukrštanja, pri čemu je verovatnoća mutacije fiksirana na vrednost od 0,10 a svi ostali parametri bili su isti kao u prethodnom setu izvršenja Phabblers-a. Testirane vrednosti verovatnoće ukrštanja kretale su se u opsegu od 0,60 do 0,90. Kao što je ilustrovano na slici 4.20, izbor verovatnoće ukrštanja ima manje izražen uticaj na konvergenciju algoritma od prethodno razmatranog uticaja verovatnoće mutacije. Ovakvom rezultatu doprinosi činjenica da je trenutna implementacija algoritma takva da se potomci roditelja odbacuju ukoliko su istovetni roditeljskim rešenjima. Preciznije, ukoliko se nakon selekcije roditelja potomci ne podvrgnu niti mutaciji niti ukrštanju, oni ne postaju deo nove populacije. Iako je ovo rešenje bilo neophodno u cilju sprečavanja gubitka raznovrsnosti populacije u okvirima elitističke prirode NSGA-II, ono istovremeno podrazumeva da izbor verovatnoće mutacije i ukrštanja pre svega određuje relativnu učestalost jedne odnosno druge genetske operacije, a ne nužno njihovu apsolutnu verovatnoću. Uzimajući u obzir dato ograničenje, interesantno je primetiti da se populacije rešenja dobijene sa verovatnoćom ukrštanja od 0,85 jasno izdvajaju od ostalih. Ova činjenica može ukazati na postojanje optimalnog odnosa između verovatnoće ukrštanja i mutacije, koji bi, prema dobijenim rezultatima, trebalo da bude oko 8–10:1.



(a) Prosečna (isprekidane linije) i maksimalna (pune linije) sličnost između rešenja u populaciji i zadate farmakoforne hipoteze, u funkciji trajanja optimizacije.

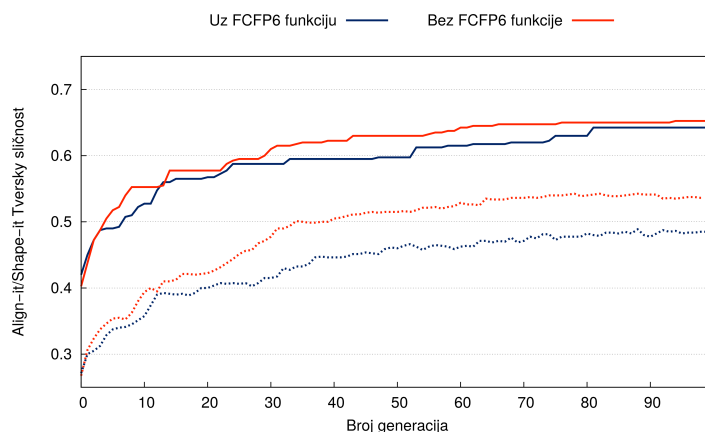


(b) Prosečna Tanimoto udaljenost između rešenja u populaciji, u funkciji trajanja optimizacije.

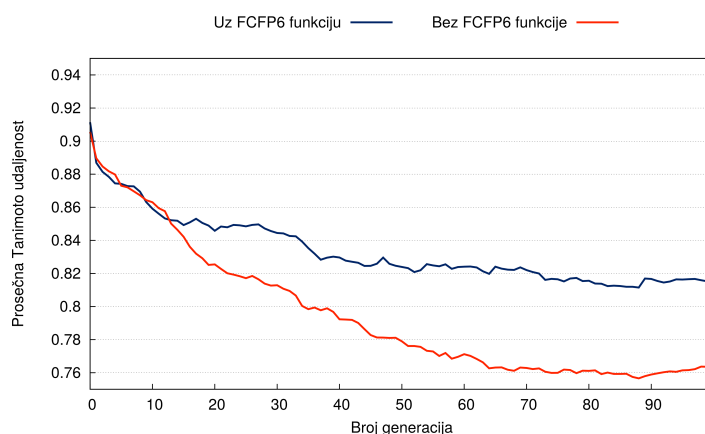
Slika 4.20: Uticaj izbora verovatnoće ukrštanja na neke aspekte konvergencije Phabblar algoritma. Različite boje linija odgovaraju različitim verovatnoćama ukrštanja koje su testirane, kao u legendi.

U sledećim koracima, ispitivani su efekti eliminacije jednog od 4 ciljeva dizajna, pri čemu su verovatnoće ukrštanja i mutacije fiksirane na vrednosti od 0,80 i 0,10, redom. Efekti isključivanja FCFP6 funkcije procene iz procesa optimizacije ilustrovani su na slici 4.21. Uklanjanje jednog cilja dizajna očekivano čini algoritam nešto efikasnijim, odnosno uslovljava njegovu bržu konvergenciju. Međutim, ukoliko se uporede farmakoforne sličnosti optimalnih rešenja dobijenih sa i bez FCFP6 funkcije, primetno je da su razlike gotovo neznatne (slika 4.21a). S druge strane, strukturalna raznovrsnost populacije značajno brže opada kada se FCFP6 ne koristi (slika 4.21b). Iako ovo opažanje deluje kontraintuitivno, ono je efektivno posledica činjenice da Pareto optimalna rešenja često mogu imati izraženu prilagođenost u prostoru različitih funkcija prilagođenosti. Stoga dodavanje mere topološke sličnosti u ciljeve optimizacije može biti korisno kada se želi diverzifikacija populacije rešenja, ali pre svega u slučajevima kada je poželjno da generisane molekule budu derivati ili bioizosteri određene privilegovane osnovne strukture prisutne u referentnom jedinjenju. U drugim slučajevima, eliminisanje ovog cilja dizajna može ubrzati konvergenciju algoritma.

Analiza efekata SMCM funkcije na evoluciju rešenja ukazala je da, uopšteno, njena primena u algoritmu favorizuje rešenja sa nižom molekulskom masom (slika 4.22a) i većim udelom heteroatoma (slika 4.22b). Kako ova činjenica potencijalno ukazuje da SMCM procena nije dovoljno osetljiva za datu namenu, odlučeno je da je potrebno dodatno ispitati da li ova mera adekvatno oslikava sintetsku složenost struktura, na način koji bi odgovarao percepciji medicinskog hemičara. U tu svrhu, iskorišćene su subjektivne procene sintetske složenosti hemičara kompanije Merck čiji su rezultati nedavno objavljeni [411]. U citiranoj studiji, velikom broju medicinskih hemičara randomizovano su prezentovane hemijske strukture koje su oni ocenjivali vrednostima od 1 do 5, gde ocena 1 odgovara sintetski najjednostavnijim, a ocena 5 najsloženijim strukturama. Srednje vrednosti ovih ocena za 1731 jedinjenje upoređene su sa izračunatim SMCM vrednostima. Utvrđeno je da postoji dobro linearno slaganje ($R = 0,74$, slika C.2) između izračunatih SMCM vrednosti i prosečne ocene sintetske složenosti koju su dali medicinski hemičari. Pri tome, treba naglasiti da je prosečna standardna devijacija srednje subjektivne ocene bila ≈ 1 [411]. Na osnovu ovih rezultata, može se zaključiti da SMCM rezultati nisu nespecifični već da adekvatno kvantifikuju sintetsku složenost. Ključno je istaći da optimizacija SMCM tokom izvršenja algoritma uslovljava generisanje molekula koje su značajno sintetski dostupnije i manje složene, ali da pri tome smanjuje farmakofornu sličnost hipotezi za oko 0,05 u poređenju sa optimalnim rešenjima koja nisu ograničavana SMCM vrednošću (slika 4.22c). Prosečno rešenje dobijeno uz optimizaciju



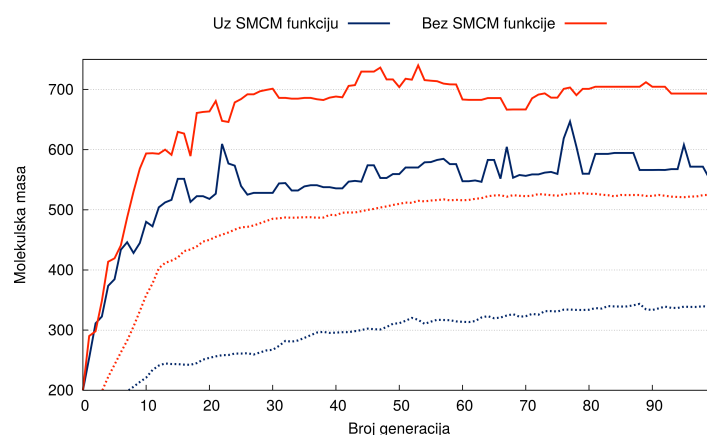
(a) Prosečna (isprekidane linije) i maksimalna (pune linije) sličnost između rešenja u populaciji i zadate farmakoforne hipoteze, u funkciji trajanja optimizacije.



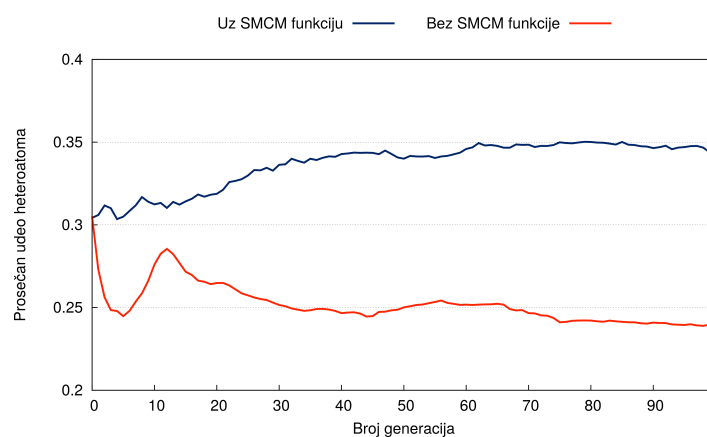
(b) Prosečna Tanimoto udaljenost između rešenja u populaciji, u funkciji trajanja optimizacije.

Slika 4.21: Uticaj uklanjanja maksimizacije topološke sličnosti, kao cilja dizajna, na konvergenciju Phabblers algoritma. Kao mera topološke sličnosti korišćena je Tanimoto sličnost FCFP6 otisaka prstiju (FCFP6 funkcija procene).

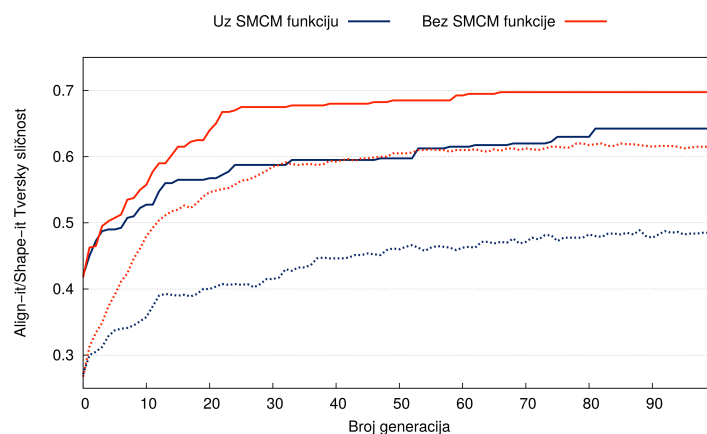
SMCM tako u najvećoj meri odgovara manjoj molekuli pogodnoj za dalje strukturne modifikacije, dok su prosečna rešenja neograničena SMCM vrednošću sličnija finalnim proizvodima dizajna. Međutim, skup Pareto-optimalnih rešenja u oba slučaja uključuje i relativno složene molekule, pod uslovom da je njihova farmakoforna sličnost hipotezi visoka. Iz ovog razloga je prednost optimizacije SMCM najčešće daleko veća u odnosu na relativno mali gubitak farmakoforne sličnosti optimalnih rešenja. Tokom evolucije duže od 100 generacija, SMCM je neophodna za adekvatno ograničavanje „rasta” molekula, jer se broj kršenja pravila Lipinskog nije pokazao kao dovoljan za ovu namenu.



(a) Prosečna (isprekidane linije) i maksimalna (pune linije) molekulska masa rešenja u populaciji, u funkciji trajanja optimizacije.



(b) Prosečan udeo heteroatoma u teškim atomima molekula u populaciji, u funkciji trajanja optimizacije.



(c) Prosečna (isprekidane linije) i maksimalna (pune linije) sličnost između rešenja u populaciji i zadate farmakoforne hipoteze, u funkciji trajanja optimizacije.

Slika 4.22: Uticaj uklanjanja minimizacije sintetske složenosti, kao cilja dizajna, na konvergenciju Phabblor algoritma. Kao mera sintetske složenosti korišćena je SMCM vrednost molekula (SMCM funkcija procene).

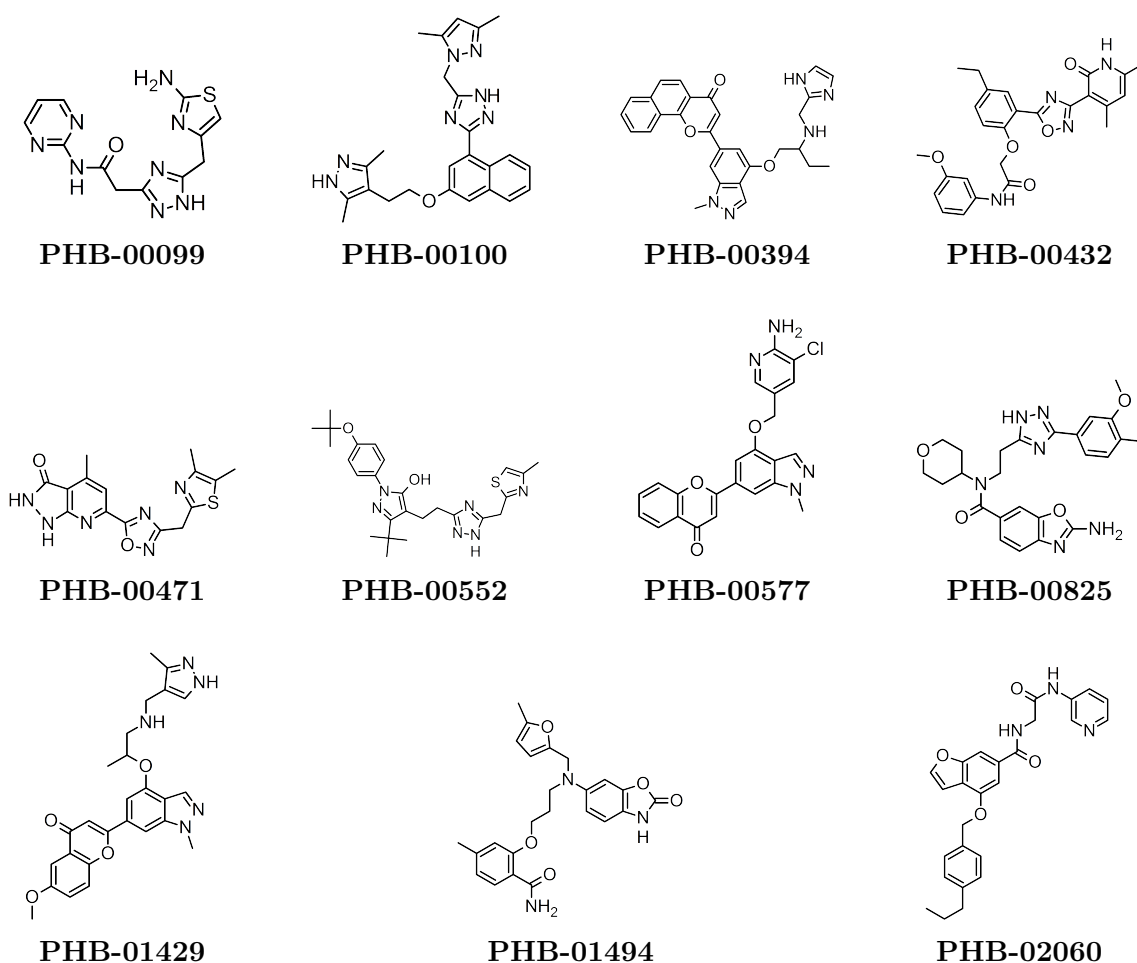
Na osnovu izložene analize uticaja izbora parametara na efikasnost Phab-
bler algoritma, u *de novo* dizajnu kolekcije potencijalnih EZH2 inhibitora, korišćena
je verovatnoća mutacija od 0,10, verovatnoća ukrštanja od 0,80–0,85 kao i sva 4 cilja
optimizacije. Radi dodatne diverzifikacije formirane kolekcije, u nekoliko izvršenja
bili su isključeni maksimizacija topološke sličnosti, odnosno minimizacija SMCM
vrednosti, kao ciljevi dizajna.

4.3.2 *De novo* dizajn kolekcije potencijalnih EZH2 inhibitora

Kroz nekoliko izvršenja Phab-
bler algoritma sa odabranim parametrima, uz
naizmenično korišćenje jedne, druge ili obe razvijene farmakoforne hipoteze, dizajni-
rana je virtuelna kolekcija koje je sadržala 2.237 strukturno jedinstvenih molekula.
Ova početna kolekcija zatim je filtrirana u cilju ukljanjanja zaostalih fragmenata
kao i jedinjenja veoma visoke molekulske mase, a zatim i onih jedinjenja koja sadrže
podstrukture koje su često odgovorne za promiskuitetno ponašanje u biohemijским
testovima, i kao takve imaju mali potencijal za dalji razvoj u selektivne inhibitore.
Time je iz kolekcije eliminisano 460 jedinjenja. Sa preostalih 1.777 jedinjenja izvr-
šen je molekularni doking u dve konformacije EZH2, koje odgovaraju kompleksima
sa **G5K** i **E5Z**, čime su dobijene dve liste jedinjenja rangiranih po Vina doking
energijama. Iz svake liste odabrano je 100 najbolje rangiranih jedinjenja za ponovni
doking, s tim što je ovaj put sačuvano po 5 najboljih poza za vizuelno istraživanje.
Ovde treba naglasiti da je u okviru najužeg kruga jedinjenja bilo zastupljeno više
strukturnih analoga svake jedinstvene osnovne strukture, što ilustruje sposobnost
Phab-
bler algoritma da kroz ukrštanje rešenja temeljno istražuje odabrane delove
hemijškog prostora. Raznovrsnost osnovnih struktura u kolekciji, s druge strane,
ukazuje da algoritam nije bio ograničen na mali deo hemijškog prostora, već da je
istraživanje imalo i globalni karakter. Vizuelna analiza struktura kompleksa EZH2
i dizajniranih jedinjenja potvrdila je da je primenom Phab-
bler algoritma moguće
generisati relevantne početne strukture jedinjenja koja zadovoljavaju definisanu far-
makofornu hipotezu a strukturno su različita od poznatih inhibitora datog ciljnog
mesta dejstva.

Kako bi detaljnije istražili interakciju više različitih osnovnih struktura sa
EZH2, iz rangiranih lista odabrano je 11 strukturno raznovrsnih jedinjenja (prika-
zana na slici 4.23) čija je interakcija sa enzimom ispitivana simulacijama molekularne
dinamike. Ovaj korak je takođe bio važan imajući u vidu da je u više studija po-
kazano da je simulacijama MD i MM/GBSA ili MM/PBSA proračunima značajno

moguće unaprediti verovatnoću identifikacije stvarno aktivnih jedinjenja u virtuelnom skriningu [412, 413].



Slika 4.23: Strukture 11 jedinjenja iz *de novo* dizajnirane kolekcije čija je interakcija za EZH2 proučavana simulacijama MD.

4.3.3 Simulacije molekularne dinamike dizajniranih jedinjenja

Od 11 jedinjenja čija je interakcija sa EZH2 istraživana molekularnom dinamikom, 5 jedinjenja je tokom ekvibracije i produkcionih simulacija ostalo stabilno vezano u vezivnom mestu za kofaktor. Rezultati MM/GBSA proračuna za ova jedinjenja zbirno su prikazani u tabeli 4.6, a način njihove interakcije sa EZH2 detaljnije je opisan u daljem tekstu.

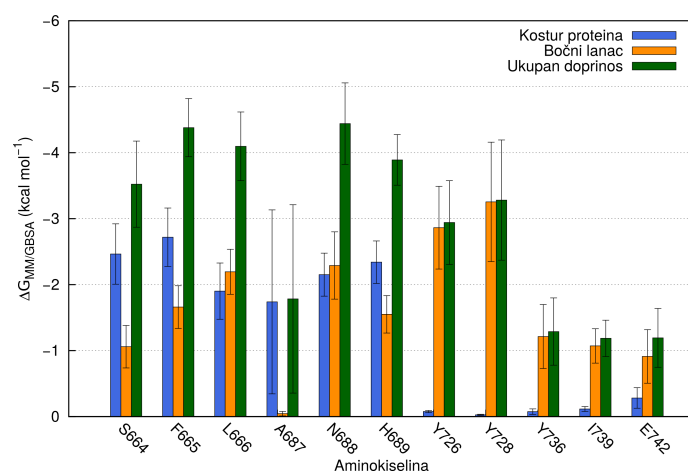
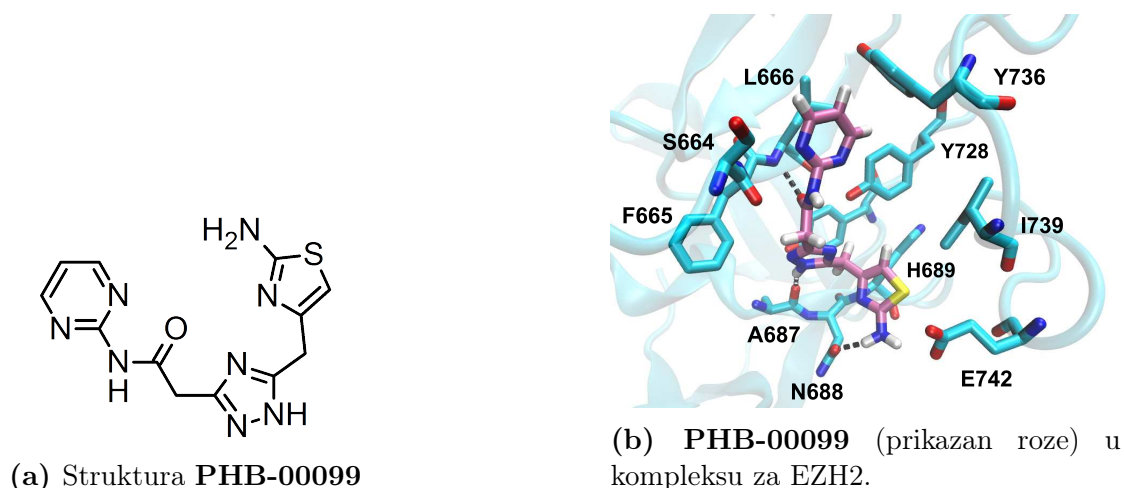
Tabela 4.6: Rezultati MM/GBSA proračuna za 5 novodizajniranih jedinjenja u kompleksu sa EZH2

Jedinjenje	Molekularna masa	ClogP	$\Delta G_{MM/GBSA}^a$
PHB-00099	316,34	1,01	-26,48 ± 0,14
PHB-00394	493,57	3,50	-42,34 ± 0,16
PHB-00471	342,38	3,76	-29,68 ± 0,14
PHB-00577	432,86	3,13	-37,34 ± 0,13
PHB-02060	443,50	3,87	-29,84 ± 0,18

^a Prikazane su srednje vrednosti uz standardnu grešku u kcal mol⁻¹

PHB-00099 dizajniran je na osnovu farmakofore izvedene iz **E5Z**, što se jasno odražava na način vezivanja ove molekule za EZH2, ustanovljen u simulacijama MD. Kao što je ilustrovano na slici 4.24b, veliki deo molekule stupa u interakcije sa I-SET domenom u položaju koji je upravan na tipičnu orijentaciju inhibitora u vezivnom mestu. Ovakva orijentacija je interesantna jer primeri **E5Z**, **CPI-169** i **CPI-360** ukazuju da dodatni akceptor vodonične veze, čiji očekivan položaj odgovara pirimidinskom delu strukture **PHB-00099**, mogu značajno doprineti aktivnosti. Analiza trajektorije simulacije **PHB-00099** u kompleksu za EZH2 ukazuje na prisustvo više vodoničnih veza i polarnih interakcija koje dominantno doprinose stabilizaciji liganda. Amino grupa položaja 2 tiazola uspostavlja vodonične veze sa bočnim lancem N688 i E742, dok sam tiazol ostvaruje kontakte sa H689 i I739 iz post-SET domena. Triazolni deo strukture prevashodno stabilizuje vodonična veza sa karbonilnim kiseonikom A687. Najveći broj polarnih interakcija obrazuje amid koji povezuje triazol sa pirimidinom, tako naizmenično interagujući sa peptidnim skeletom S664, F665 i L666. Piridin je zatvoren u hidrofobnom džepu koji se obrazuje prilagođavanjem post-SET domena ligandu i u čijem formiranju učestvuju bočni lanci L666, Y728 i Y736. Razlaganje energije vezivanja **PHB-00099** po parovima interakcija sa aminokiselinama vezivnog mesta i kvantitativno pokazuje značaj opisanih interakcija (slika 4.24c).

Kako je **PHB-00099** relativno mala i polarna molekula za koju simulacije MD ukazuju da se potencijalno vezuje za EZH2 sa visokom specifičnošću, ovo dizajnirano jedinjenje poseduje značajan potencijal dalje optimizacije. Iako je izračunata vrednost $\Delta G_{MM/GBSA}$ relativno visoka (tabela 4.6), efikasnost ovog liganda u kontekstu njegove lipofilnosti i molekulske mase je veoma povoljna. Dodatkom aromatičnog supstituenta, koji bi se granao od dela strukture između triazola i amida, bilo bi moguće ostvariti povoljne interakcije sa Y661, za koje je poznato da značajno doprinose energiji vezivanja. Dodatno, uvođenjem račvastog alkil supstituenta u po-



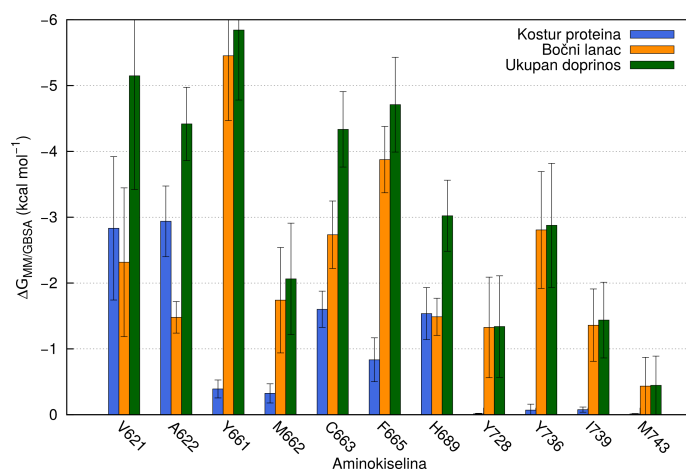
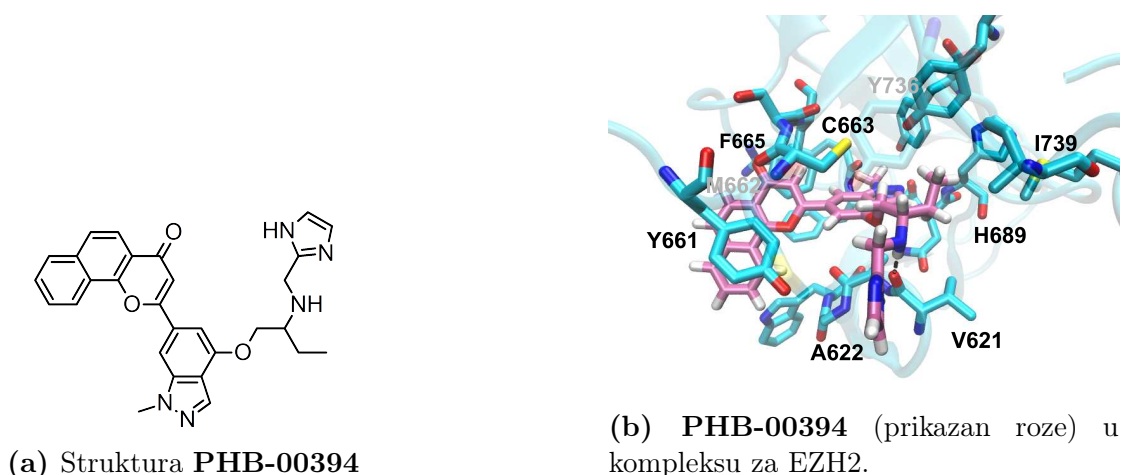
(c) Doprinos različitih aminokiselina vezivnog mesta energiji vezivanja **PHB-00099**

Slika 4.24: Reprezentativna struktura kompleksa i rezultati razlaganja energije vezivanja za kompleks **PHB-00099** i EZH2.

ložaj 5 tiazola bilo bi moguće ostvariti dodatne interakcije sa hidrofobnim bočnim lancima post-SET domena.

Najveće od 5 jedinjenja koja su pokazala afinitet za EZH2 u MD simulacijama je **PHB-00394**, čija je reprezentativna struktura kompleksa sa enzimom prikazana na slici 4.25b. Prisustvo 4H-benzo[h]hromen-4-ona u strukturi uslovljava veoma povoljne hidrofobne i aromatične interakcije ove molekule sa Y661 i F665, što odgovara ulozi supstituenta položaja 6 centralnog jezgra u poznatim piridon-skim inhibitorima EZH2. Sličnost **PHB-00394** sa poznatim inhibitorima očigledna je i centralnom delu strukture, gde je prisutan indazol koji ostvaruje očekivane interakcije sa C663, H689, Y726 i A622. Kod ove molekule, međutim, značajan deo

strukture uspostavlja interakcije sa post-SET domenom i vrhom beta-ukosnice koja se nalazi nasuprot motiva pseudo-čvora. Ovaj deo strukture povezan je sa centralnim jezgrom etarskom vezom, pri čemu njegov ugljovodonični deo ostvaruje brojne hidrofobne kontakte sa I739, Y728, Y736 i bočnim lancem H689. Sekundarni amin u strukturi uključen je u formiranje vodonične veze sa peptidnim skeletom beta-ukosnice, preko V621. Ovu vodoničnu vezu može da formira i indazol, ali on zadržava značajno veći stepen fleksibilnosti i naizmenično stupa u interakcije sa bočnim lancem V621 i aminokiselinama post-SET domena.



(c) Doprinos različitih aminokiselina vezivnog mesta energiji vezivanja **PHB-00394**

Slika 4.25: Reprezentativna struktura kompleksa i rezultati razlaganja energije vezivanja za kompleks **PHB-00394** i EZH2.

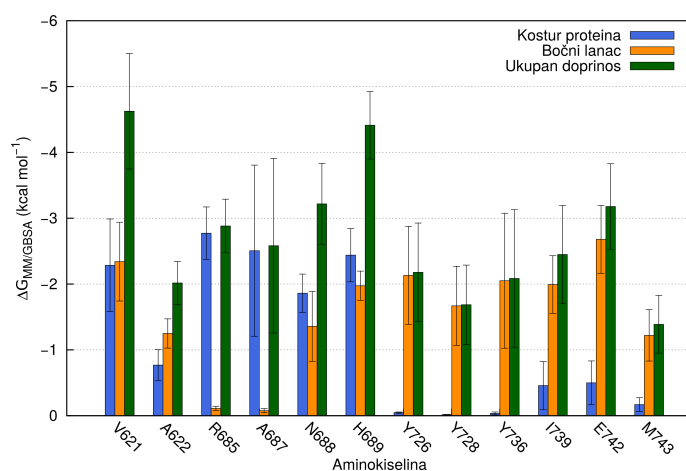
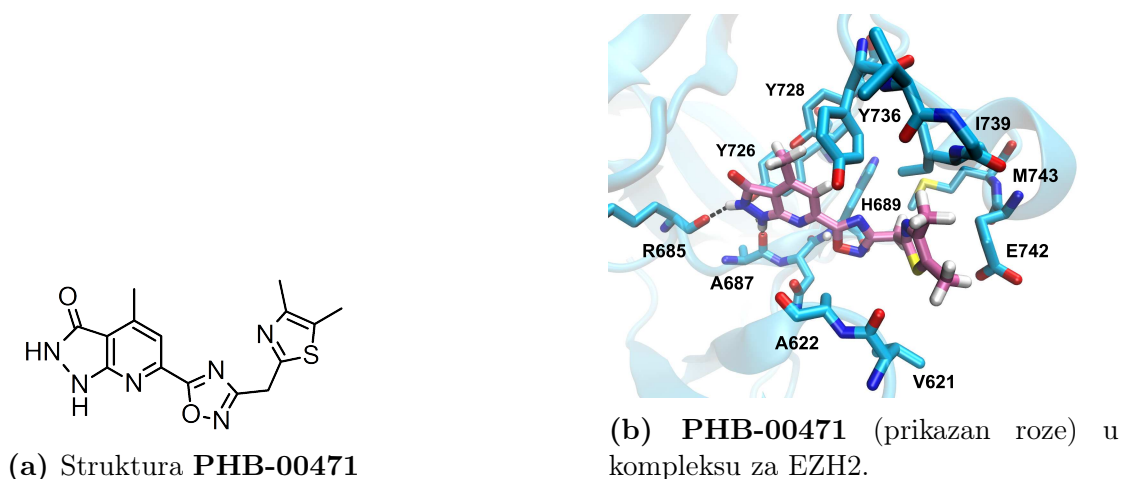
Na osnovu analize razložene energije vezivanja **PHB-00394** (ilustrovano na slici 4.25c), može se zaključiti da najizraženiji doprinos vrednosti $\Delta G_{MM/GBSA}$ daju interakcije benzohromenskog dela strukture. Kao što je ilustrovano na primeru

poznatog inhibitora **E1Z**, kondenzovani supstituenti položaja 6 centralnog aromatičnog jezgra daju značajan entalpijski doprinos energiji vezivanja ali se ovo ne prevodi u značajan afinitet za enzim, verovatno usled entropijske kompenzacije vezane za gubitak fleksibilnosti petlje I-SET domena. Stoga dobijenu energiju vezivanja, najnižu u seriji proučavanih jedinjenja (tabela 4.6) treba razmatrati u kontekstu ovog ograničenja.

PHB-00471 ima najmanje sličnosti, u pogledu načina interakcije sa vezivnim mestom, u odnosu na poznate inhibitore EZH2. Ova molekula sadrži u strukturi nesupstituisani pirazolo[3,4-b]piridin-3-on koji može postojati u većem broju tautomernih oblika. Za simulacije je odabran tautomerni oblik za koji je Calculator Plugin-om programa Marvin 15.1.19 predviđeno da predstavlja dominantnu formu na pH 8. U ovom tautomernom obliku, pirazolonski deo strukture uspostavlja dve vodonične veze sa peptidnim kosturom EZH2 preko A687 i R685, koje značajno doprinose energiji vezivanja (slika 4.26). Azot piridinskog dela strukture, kao i središnji 1,2,4-oksadiazol, stupaju u polarne interakcije sa peptidnim kosturom H689 i bočnim lancem N688, pri čemu elektrostatička komplementarnost ovih delova molekule i enzima uslovljavaju dodatno sniženje $\Delta G_{MM/GBSA}$. 4,5-dimetil supstituisani tiazol je, s druge strane, dominantno uključen u hidrofobne kontakte sa aminokiselinama post-SET domena kao i bočnim lancem V621 sa vrha beta-ukosnice.

I pored brojnih energetski povoljnih interakcija u koje **PHB-00471** stupa sa aminokiselinama vezivnog mesta, visoka energija desolvatacije ovako polarne strukture negativno se odražava na apsolutnu vrednost $\Delta G_{MM/GBSA}$. Iz reprezentativne poze molekule (slika 4.26b) očigledno je da postoji mogućnost za više strukturnih izmena koje bi kroz dodatne hidrofobne interakcije mogle da doprinesu energiji vezivanja. Jedinstven način interakcije sa EZH2 čini **PHB-00471** veoma interesantnim hemotipom za dalje istraživanje, uz činjenicu da njegova relativno niska molekulaska masa pruža značajan prostor za dalje modifikacije strukture.

PHB-00577 je strukturno sličan **PHB-00394** ali je benzohromenon zamenjen hromen-4-onom, a deo strukture koji je u kontaktu sa post-SET domenom i NHS motivom – značajno je uprošćen. Iako aromatične interakcije hromenonskog dela strukture sa Y661 i F665 i dalje daju izražen doprinos ukupnoj izračunatoj energiji vezivanja (slika 4.27c), glavni doprinos $\Delta G_{MM/GBSA}$ daju vodonične veze koje aminopiridinski deo strukture obrazuje sa N688 i H689. Kao što je ilustrovano na slici 4.27b, N688 je u alternativnom rotamernom obliku, kao kod kompleksa EZH2 sa **G6K** ili **126**, pri čemu je i orijentacija piridinskog prstena slična „alterna-

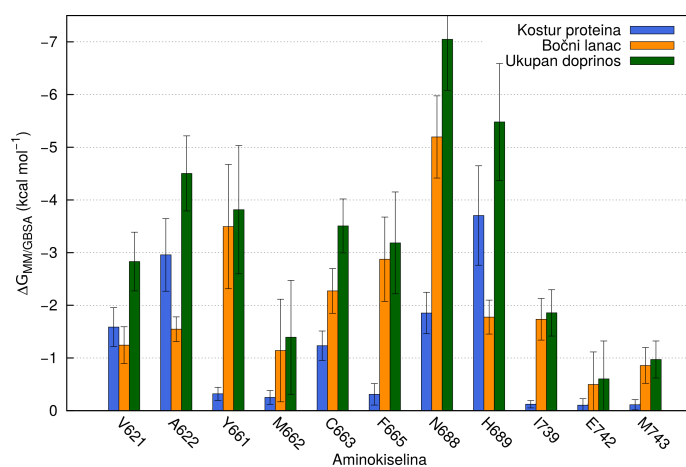
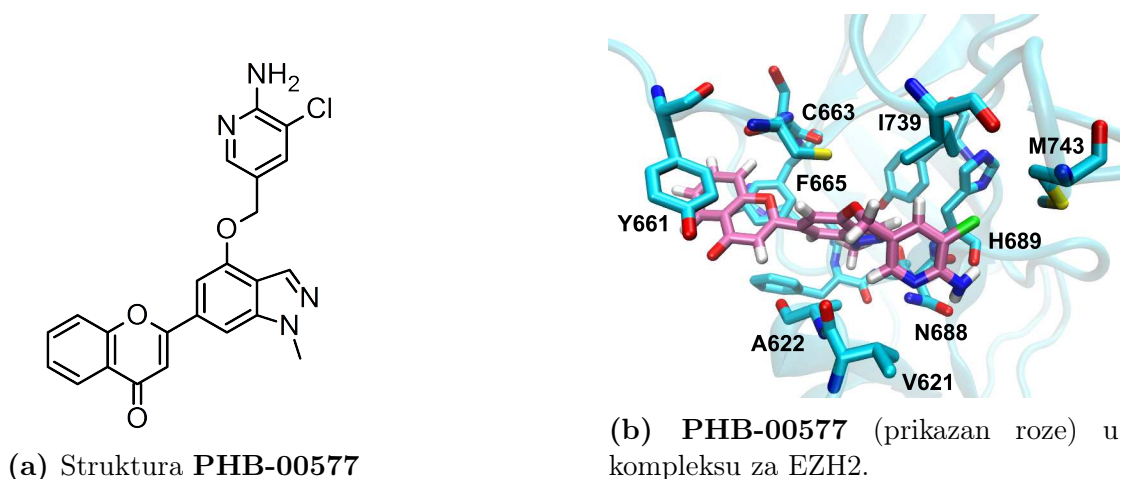


(c) Doprinos različitih aminokiselina vezivnog mesta energiji vezivanja **PHB-00471**

Slika 4.26: Reprezentativna struktura kompleksa i rezultati razlaganja energije vezivanja za kompleks **PHB-00471** i EZH2.

ativnoj” orijentaciji piridona kod ovih inhibitora. Dodatnu stabilizaciju piridinu daju interakcije sa V621 i A622, kao i hidrofobne interakcije sa M743 i I739 post-SET domena.

Kao što je prethodno diskutovano, kondenzovani biciklični sistemi nisu nužno najbolji izbor u eksploataciji interakcije sa Y661 i F665, pa bi jedna od mogućnosti optimizacije aktivnosti ovog jedinjenja bila zamena hromenona drugim aromatičnim supstituentom. Značajnije, supstituisanje piridina dodatnim alkil supstituentima u položaju 3 i 4 moglo bi da omogući ostvarivanje opsežnijih hidrofobnih kontakata sa post-SET domenom. Zamena aminopiridinske strukture aminohinolin-skom, ili aminotetrahidrohino-linskom strukturom takođe su potencijalni načini za poboljšanje energije vezivanja ove molekule.

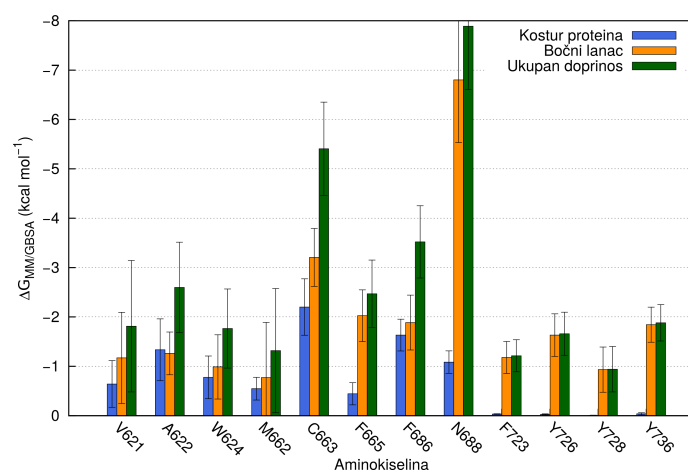
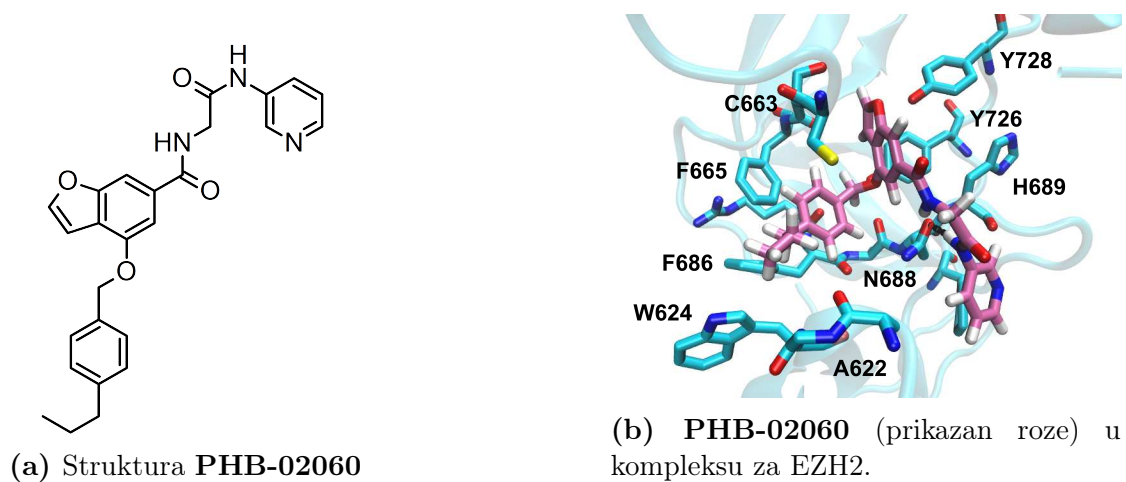


(c) Doprinos različitih aminokiselina vezivnog mesta energiji vezivanja **PHB-00577**

Slika 4.27: Reprezentativna struktura kompleksa i rezultati razlaganja energije vezivanja za kompleks **PHB-00577** i EZH2.

Konačno, simulacije MD **PHB-02060** u kompleksu sa EZH2 ukazale su da dipeptidna struktura ove molekule potencijalno može da učestvuje u formiranju snažnih polarnih interakcija sa NHS motivom enzima, prevashodno kroz uspostavljanje vodoničnih veza sa bočnim lancem N688 (slika 4.28). Specifičnosti interakcije ove molekule sa EZH2 mogu doprineti hidrofobni kontakti koje benzenov prsten i propil grupa uspostavljaju sa F665, F686 i C663, a u manjoj meri i sa W624. Benzofuranski deo strukture pozicioniran je slično tetrahidropiranu u **E5Z**, ali slabe akceptorske karakteristike kiseonika u ovom heterociklu uz neodgovarajuću orijentaciju onemogućavaju uspostavljanje jakih vodoničnih veza sa Y728 ili peptidnim kosturom I-SET domena.

Piridinski deo strukture **PHB-02060** ne stupa u značajnije interakcije



(c) Doprinos različitih aminokiselina vezivnog mesta energiji vezivanja **PHB-02060**

Slika 4.28: Reprezentativna struktura kompleksa i rezultati razlaganja energije vezivanja za kompleks **PHB-02060** i EZH2.

sa vezivnim mestom, izuzev prolaznih kontakata sa beta-ukosnicom naspramnoj NHS motivu. Kod ove molekule izostaju kontakti sa post-SET domenom, što je karakteristika koja nije uočena kod poznatih inhibitora EZH2. Stoga je, u poređenju sa ostalim opisanim molekulama koje su rezultat *de novo* dizajna, potencijal **PHB-02060** za dalji razvoj najmanje izvestan.

4.3.4 Modifikacija strukture u cilju eliminisanja afiniteta za Pgp

Primenom razvijenih modela za predviđanje transporta posredovanog Pgp-om i BCRP-om, ustanovljeno je da svih 5 struktura koje su izabrane za simulacije MD predstavljaju supstrate ova dva transportera. Iako je bilo moguće modele iskoristiti ranije, u filtriranju početne kolekcije dizajniranih struktura, čime bi izbor manje kolekcije jedinjenja bio ograničen samo na nesupstrate, odlučeno je da ovaj pristup nije optimalan. Prevažodno, široka supstratna specifičnost Pgp-a i BCRP-a uslovlila bi eliminaciju potencijalno interesantnih struktura inhibitora i time bi ih preuranjeno eliminisala iz daljeg istraživanja.

Ipak, rano definisan cilj dizajna novih potencijalnih inhibitora u ovoj studiji bila je minimizacija rizika za interakciju sa membranskim transporterima. U tom kontekstu, razvijeni prediktivni modeli za transport posredovan Pgp-om i BCRP-om, s jedne strane, i značajna korisnost modela EZH2 u kombinaciji sa simulacijama MD, s druge strane, omogućavaju racionalnu modifikaciju dizajniranih struktura. Za optimizaciju odabran je **PHB-00471** zbog svoje dobre efikasnosti u smislu odnosa procenjene energije vezivanja i broja atoma u molekuli. Prevažodan cilj modifikacije strukture bila je eliminacija afiniteta za Pgp, zbog njegovog izraženijeg uticaja u poređenju sa BCRP-om, uz očuvanje ili poboljšanje afiniteta za EZH2. Izračunate osobine jedinjenja dizajniranih tokom optimizacije, uključujući predviđen afinitet za Pgp i BCRP, odnosno izračunatu energiju vezivanja za EZH2, zbirno su prikazane u tabeli 4.7.

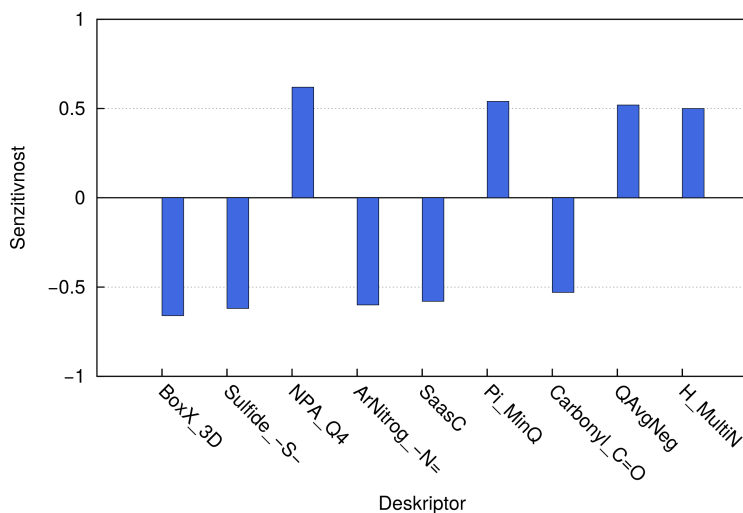
Tabela 4.7: Izračunate osobine jedinjenja dizajniranih tokom modifikacije **PHB-00471**

Jedinjenje	Supstrat za Pgp	Supstrat za BCRP	$\Delta G_{MM/GBSA}^a$	MW	ClogP
PHB-00471	da	da	$-29,68 \pm 0,14$	342,38	3,76
PHB-00471f	ne	da	$-36,72 \pm 0,14$	354,41	3,39
PHB-00471i	ne	ne	$-28,40 \pm 0,16$	348,44	2,53
PHB-00471m	ne	da	$-34,21 \pm 0,14$	384,44	2,02

^a Prikazane su srednje vrednosti uz standardnu grešku u kcal mol⁻¹

Kao početna tačka u modifikaciji strukture **PHB-00471** upotrebljena je analiza senzitivnosti deskriptora Pgp modela ANN-24-3. Analiza senzitivnosti deskriptora klasifikacionih modela u ADMET Predictor-u omogućava identifikaciju onih deskriptora čije promene vrednosti imaju najizraženiji potencijal da „preokrenu”

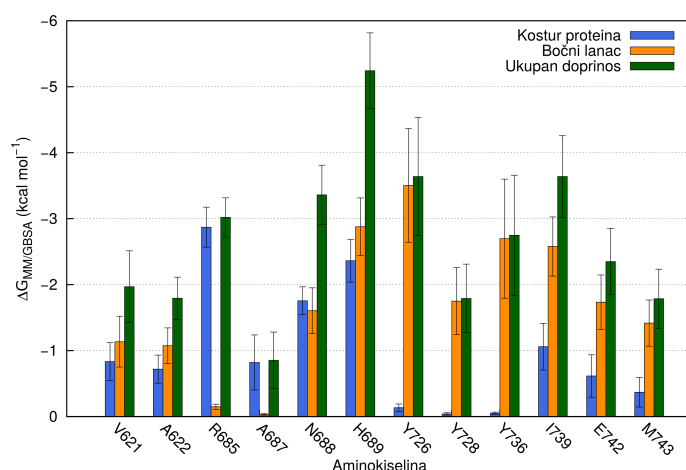
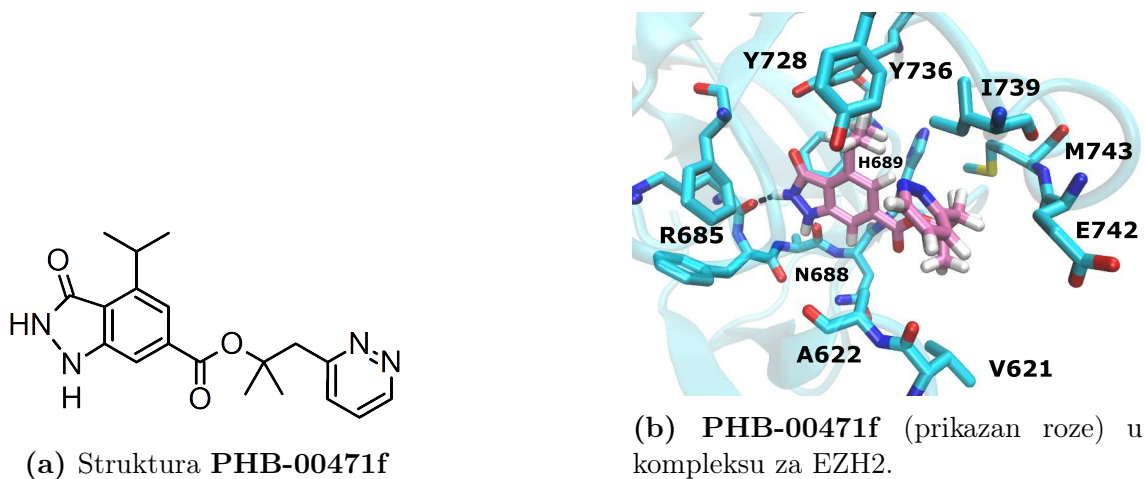
predviđanje modela. Tačnije, za svaki od deskriptora moguće je dobiti meru relativne uticajnosti na predviđanje modela, kao i smer promene (povećanje ili smanjenje vrednosti, ili oba) kojim je moguće uticati na promenu predviđanja. Rezultat analize senzitivnosti deskriptora za **PHB-00471** grafički je prikazan na slici 4.29.



Slika 4.29: Grafički prikaz rezultata analize senzitivnosti deskriptora u modelu ANN-24-3 (transport posredovan Pgp-om) za strukturu **PHB-00471**.

Na osnovu rezultata analize senzitivnosti deskriptora, zaključeno je da naj-dostupniji put za smanjenje verovatnoće za interakciju sa Pgp-om predstavlja smanjenje broja aromatičnih azota u strukturi, kao i uklanjanje sumpora kao heteroatoma u aromatičnoj strukturi koja stupa u interakcije sa post-SET domenom. Kako bi se ovo postiglo, inicijalno je dizajnirano nekoliko jedinjenja koja su, umesto 1,2,4-oksadiazola, imala amidski linker u strukturi i kod kojih je tiazol bio zamenjen jednostavnim benzenovim prstenom. Dobijene strukture po predviđanjima modela, međutim, i dalje su predstavljale supstrate za Pgp, pri čemu je broj aromatičnih azota i dalje bio jedan od najosetljivijih deskriptora predviđanja. Iz ovog razloga, pirazolopiridinonska struktura zamenjena je benzopirazolonskom. Ponovljena analiza senzitivnosti deskriptora zatim je ukazala na neophodnost uklanjanja amida iz strukture, pri čemu je povećanje broja estarskih funkcionalnih grupa bilo udruženo sa visokom senzitivnošću „obrtnja” klase jedinjenja. Stoga je centralni amid zamenjen estrom, čime je dobijen prvi analog **PHB-00471** za koji je predviđeno da ne predstavlja supstrat za Pgp. Ipak, uvođenje hemijski labilne funkcionalne grupe u strukturu nije smatrano dobrim rešenjem, ali drugi linkeri (keto, etarski, tioetarski) davali su strukture za koje je predviđeno da predstavljaju supstrate Pgp-a. Susedni ugljenikov atom estra kvaternizovan je u cilju pružanja dodatne sterne zaštite, ali i u cilju ostvarivanja dodatnih hidrofobnih kontakata sa post-SET domenom. Dodatno, metil supstituent benzopirazolonske strukture zamenjen je izopropil grupom, u cilju

ostvarivanja izraženijih interakcija sa Y726, Y728 i Y736. Konačno, u cilju smanjenja lipofilnosti, benzen koji je prvobitno iskorišćen kao zamena tiazola, zamenjen je piridazinom. Struktura dizajniranog jedinjenja **PHB-00471f** prikazana je na slici 4.30a. Dizajnirana molekula zatim je simulirana u kompleksu sa EZH2.



(c) Doprinos različitih aminokiselina vezivnog mesta energiji vezivanja **PHB-00471f**

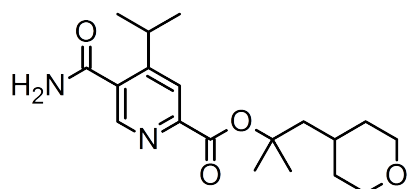
Slika 4.30: Reprezentativna struktura kompleksa i rezultati razlaganja energije vezivanja za kompleks **PHB-00471f** i EZH2.

Kao što se može uočiti na slici 4.30b, orijentacija **PHB-00471f** u vezivnom mestu slična je onoj koja je ustanovljena za polazno jedinjenje, ali je cela struktura delimično izmeštena ka post-SET domenu. Ovo je prevashodno posledica zamene pirazolopiridinonske benzopirazolonskom strukturom, odnosno efekat odsustvo azota kao akceptora i dodatne voluminoznosti nove strukture. Razlaganje energije vezivanja po parovima interakcija (slika 4.30c) ukazuje da vodonična veza sa R685 nastavlja da daje značajan doprinos energiji vezivanja, dok je analogna interakcija sa A687 marginalno zastupljena usled izmeštanja strukture. S druge strane, kar-

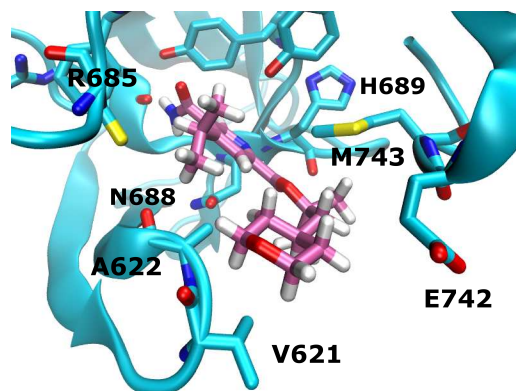
bonilni kiseonik estra uspostavlja povoljne interakcije sa peptidnim azotom H689. Ipak, dominantan entalpijski doprinos energiji vezivanja daju hidrofobne interakcije izopropil grupe benzopirazolona sa tri tirozina vezivnog mesta i I739, kao i interakcije piridazina sa post-SET domenom, V621 i A622. Ove dodatne interakcije, uz očuvanje ključnih vodoničnih veza sa H689 i R685, uslovljavaju poboljšanje energije vezivanja od oko 7 kcal mol⁻¹, u odnosu na polazno jedinjenje.

Kao što je prethodno navedeno, u nastojanju da se estar iz strukture **PHB-00471f** zameni stabilnijom funkcionalnom grupom pokušane su različite zamene, ali predviđanja modela nisu ukazala na mogućnost dizajna nesupstrata korišćenjem početne osnovne strukture i drugih linkera. Stoga je razmotrena mogućnost zamene benzopirazolona drugom strukturom koja bi zadržala istu sposobnost formiranja ključnih vodoničnih veza sa vezivnim mestom EZH2. Kako vodonična veza sa A687 daje relativno mali doprinos energiji vezivanja, dok elektrostatičke interakcije sa NHS motivom, kao i vodonična veza sa kosturom proteina u regiji R685 imaju izraženiji uticaj, pokušano je otvaranje pirazolopiridinonske strukture. Dizajniran je piridinski derivat, pri čemu je u strukturi privremeno zadržan estar, čime je dobijena struktura **PHB-00471i** (slika 4.31a), za koju je takođe predviđeno da ne predstavlja supstrat za Pgp. Prilikom dizajna **PHB-00471i** istovremeno je pokušana i optimizacija interakcija sa post-SET domenom, V621 i A622 zamenom piridazina hidrofobnijim tetrahidropiranom.

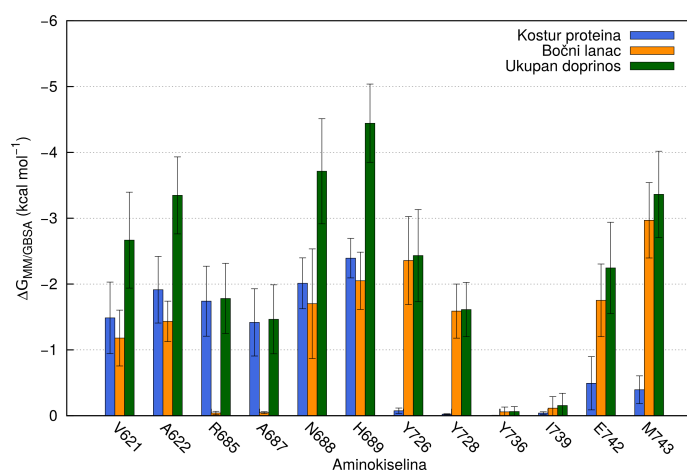
Simulacije MD kompleksa **PHB-00471i** i EZH2 ukazale su da otvaranje pirazolopiridinonskog prstena ne uslovljava potpuni gubitak afiniteta potencijalnog inhibitora za enzim, iako uslovljava značajnije promene u njegovoj orijentaciji u vezivnom mestu (slika 4.31b). Piridonsko jezgro uspostavlja povoljne elektrostatičke interakcije sa peptidnim azotom H689, a amid položaja 5 naizmenično uspostavlja vodoničnu vezu sa karbonilnim kiseonicima R685 i A686. U ovom položaju piridinskog jezgra, međutim, gube se entalpijski značajne hidrofobne interakcije izopropil grupe sa bočnim lancima Y728, Y736 i I739 (slika 4.31c), usled čega je $\Delta G_{MM/GBSA}$ vrednost za **PHB-00471i** oko 8 kcal mol⁻¹ veća u odnosu na **PHB-00471f**, mada veoma bliska vrednosti polaznog jedinjenja (tabela 4.7). S druge strane, uvođenje tetrahidropirana u strukturu imalo je željeni efekat izraženijih interakcija sa post-SET domenom i beta-ukosnicom koje je naspramna HNS motivu. Ipak, najznačajniji rezultat dizajna **PHB-00471i** ogleda se u činjenici da je nakon modifikacije benzopirazolonskog dela strukture utvrđena mogućnost zamene estarskog linkera amidom, pri čemu je za rezultujuću strukturu predviđeno da ne predstavlja supstrat Pgp-a. Na ovaj način bilo je moguće ukloniti hemijski nestabilnu funkcionalnu



(a) Struktura **PHB-00471i**



(b) **PHB-00471i** (prikazan roze) u kompleksu za EZH2.

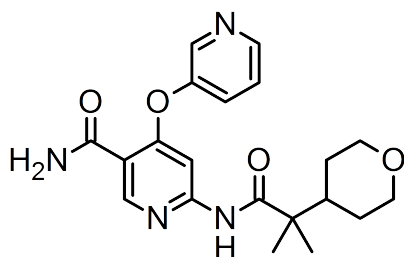


(c) Doprinos različitih aminokiselina vezivnog mesta energiji vezivanja **PHB-00471i**

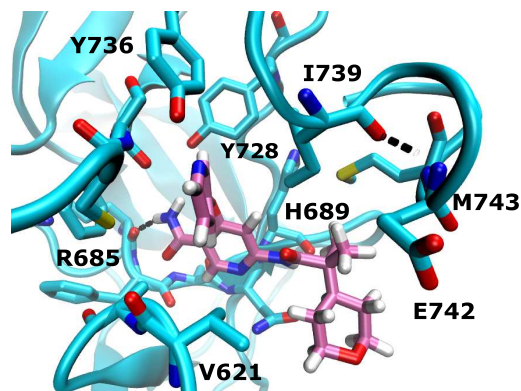
Slika 4.31: Reprezentativna struktura kompleksa i rezultati razlaganja energije vezivanja za kompleks **PHB-00471i** i EZH2.

grupu. Istovremeno, nova orijentacija inhibitora u vezivnom mestu iskorišćena je u cilju ostvarivanja dodatnih kontakata sa I-SET domenom, zamenom izopropil grupe dodatnim aromatičnim prstenom, za koji se očekivalo da će istovremeno ponovo omogućiti uspostavljanje interakcija inhibitora sa tirozinima vezivnog mesta i I739. Kroz ove modifikacije strukture dizajniran je **PHB-00471m** (slika 4.32a) kao finalni rezultat optimizacije **PHB-00471**.

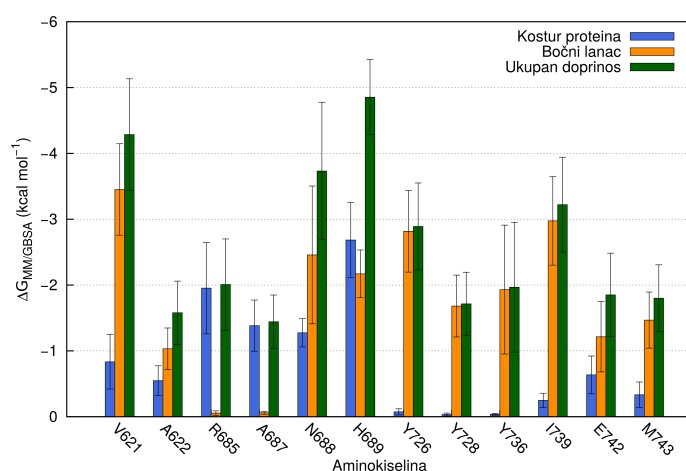
Način vezivanja **PHB-00471m** za vezivno mesto na enzimu (slika 4.32b), utvrđen simulacijama MD kompleksa, ukazuje da piridinski azot i amid položaja 1 uspostavljaju povoljne elektrostatičke interakcije sa N688 i H689, dok amid položaja 5 nastavlja da obrazuje vodoničnu vezu sa karbolnilnim kiseonikom R685. Tetrahidropiranski deo strukture, kao i dve metil grupe kvaternernog ugljenika, stupaju



(a) Struktura **PHB-00471m**



(b) **PHB-00471m** (prikazan roze) u kompleksu za EZH2.



(c) Doprinos različitih aminokiselina vezivnog mesta energiji vezivanja **PHB-00471m**

Slika 4.32: Reprezentativna struktura kompleksa i rezultati razlaganja energije vezivanja za kompleks **PHB-00471m** i EZH2.

u brojne hidrofobne interakcije sa bočnim lancima aminokiselina post-SET domena, kao i V621, pri čemu ovaj deo strukture zadržava značajnu mobilnost. Uklanjanje jednog ugljenika iz niza koji spaja amid i tetrahidropiran, u odnosu na strukturu **PHB-00471i**, omogućilo je optimizaciju ovih interakcija. Dodatni piridin u strukturi, kao aromatičan prsten povezan etrom za centralni deo inhibitora, pokazuje sposobnost uspostavljanja hidrofobnih interakcija sa Y728, Y736 i I739, koje su bile odsutne kod **PHB-00471i**. Dodatno, azot piridina kao akceptor protona, može prolazno da formira vodonične veze sa kosturom I-SET domena. Ovi kontakti, pored ključnih elektrostatskih interakcija sa NHS motivom i R685 (slika 4.32c), uslovljavaju poboljšanje procenjene energije vezivanja: $\Delta G_{MM/GBSA}$ za **PHB-00471m** je oko 5 kcal mol⁻¹ niža od vrednosti za polazno jedinjenje. U tabeli 4.7 se takođe može uočiti da je eliminisanje predviđenog afiniteta za Pgp izvršeno bez značajnijeg

povećanja molekulske mase i lipofilnosti jedinjenja, i uz istovremeno poboljšanje izračunate energije vezivanja za EZH2. Interesantno je i da orijentacija **PHB-00471m** u vezivnom mestu otvara dodatne mogućnosti optimizacije interakcija sa I-SET domenom, koje nisu bile jednostavne sa strukturom polaznog jedinjenja. Takođe, variranjem linkera između dva aromatična prstena moguće je favorizovati reorijentaciju amida položaja 5 piridina, čime se potencijalno dalje može optimizovati jačina polarnih interakcija sa kosturom R685 i A687.

4.3.5 Rezime

U ovom delu istraživanja, implementiran je algoritam za *de novo* dizajn hemijskih struktura koje zadovoljavaju prethodno definisani farmakoforni model inhibitora EZH2. Primenom algoritma dizajnirana je fokusirana kolekcija jedinjenja iz koje je, kroz nekoliko koraka filtriranja, izdvojeno 11 struktura sa izučavanjem primenom MD. Pet struktura sa najvećim procenjenim afinitetom predložene su kao potencijalni novi inhibitori EZH2. Ova jedinjenja detaljnije su analizirana u kontekstu sličnosti sa poznatim inhibitorima i mogućnostima za dalje unapređenje energije vezivanja. Ustanovljeno je da dizajnirana jedinjenja poseduju željene farmakoforne sličnosti za poznatim inhibitorima iako predstavljaju nove hemotipove. Pokazana je izvodljivost racionalne modifikacije dizajniranih struktura, vođene razvijenim modelima za predviđanje transporta posredovanog Pgp-om i dodatnim simulacijama MD, koja može da doprinese dizajnu novih inhibitora EZH2 koji nisu supstrati ovog membranskog transportera. Rezultati ove studije, potvrđeni sintezom i eksperimentalnom karakterizacijom predloženih molekula, mogu da predstavljaju polaznu tačku u razvoju novog hemotipa selektivnih inhibitora EZH2.

Poglavlje 5

Zaključak

Biološka ispitivanja, rezultati pretkliničkih studija, kao i preliminarni rezultati kliničkih ispitivanja ukazuju da je inhibicija EZH2 validan nov pristup lečenju specifičnih formi hematoloških maligniteta koji su inače refraktarni na druge vidove terapije. U okviru ove doktorske disertacije dat je doprinos otkriću novih inhibitora EZH2, koji treba da omoguće dalje ispitivanje hemijske biologije ovog epigenetskog regulatora i posluže kao osnova u razvoju novih potencijalnih antitumorskih lekova:

1. Razvijen je trodimenzionalni model kofaktor-vezivnog mesta EZH2, koji je u sprezi sa simulacijama molekularne dinamike i MM/GBSA proračunima omogućio sticanje razumevanja strukturnih odrednica aktivnosti i selektivnosti piridonskih inhibitora ovog enzima. Utvrđeno je dobro slaganje između MM/GBSA rezultata i eksperimentalno određenih energija vezivanja ($R = 0,74 - 0,88$) što podržava validnost uspostavljenog modela. Katalitički segmenat EZH2 ispoljava značajnu fleksibilnost a njegov post-SET domen igra ključnu ulogu u stabilizaciji liganada i važna je odrednica njihove selektivnosti. Najznačajniji entalpijski doprinos energiji vezivanja piridonskih inhibitora daju: hidrofobne interakcije piridonskog dela strukture sa bočnim lancima aminokiselina I739, E742 i M743 post-SET domena; elektrostatičke interakcije piridona i amidskog linkera sa N688 i H689; hidrofobne interakcije centralnog bicikličnog sistema sa Y726, F665 i C663; kao i aromatične interakcije sa Y661. Istovremeno, utvrđeno je da male varijacije u strukturi inhibitora značajno utiču na njihov način interakcije sa EZH2. Na osnovama ove analize i korišćenjem struktura kompleksa, uspostavljeni su farmakoforni modeli EZH2 inhibitora. Primena ovih farmakofornih modela omogućava uspešnu identi-

fikaciju inhibitora različitih osnovnih struktura, što je potvrđeno njihovom validacijom (AUROC 0,874–0,926; EF_{5%} 10,71–14,29). Zbirno, ovi rezultati predstavljaju racionalnu osnovu za dizajn novih i optimizaciju postojećih inhibitora EZH2.

2. Razvijeni su klasifikacioni modeli za predviđanje transporta posredovanog Pgp-om i BCRP-om koji su primenjeni u dizajnu novih inhibitora, koji neće biti supstrati ovih transportera. Konzistentnost oformljenih setova podataka doprinosi razvoju pouzdanih modela, a strukturna raznovrsnost obuhvaćenih jedinjenja obezbeđuje njihovu primenjivost u značajnom delu hemijskog prostora. Validacija je sprovedena korišćenjem internih i eksternih setova za testiranje i pokazala je prosečnu tačnost predviđanja modela od približno 80%, što potvrđuje njihovu pouzdanost i primenjivost. Analiza supstratnih preferenci Pgp-a i BCRP-a ukazuje na postojanje veoma ograničenog dela hemijskog prostora u kome se nalaze nesupstrati oba transportera. Ograničena relevantnost ovog dela hemijskog prostora za medicinsku hemiju navodi na zaključak da je u dizajnu novih lekova neophodno napraviti kompromis prilikom optimizacije interakcija sa membranskim transporterima. U daljem dizajnu inhibitora EZH2, prioritet je stoga dat minimizaciji verovatnoće za transport posredovan Pgp-om.
3. Izvršena je implementacija NSGA-II algoritma koja omogućava *de novo* dizajn hemijskih struktura koje zadovoljavaju definisanu farmakofornu hipotezu, imaju prihvatljive fizičko-hemijske karakteristike i minimalnu sintetsku složenost. Ovom implementacijom, koja je označena akronimom Phabblar, formirana je fokusirana kolekcija virtuelnih jedinjenja koja predstavljaju potencijalne inhibitore EZH2. Phabblar uspešno generiše hemijski validne strukture, koje su suštinski različite od struktura poznatih inhibitora, ali sadrže ključne farmakoforne elemente matrice dizajna. Razvijeni algoritam generalno je primenjiv i može se koristiti u *de novo* dizajnu liganada drugih ciljnih mesta, pri čemu je moguće i korišćenje drugih funkcija procene.
4. Primenom dodatnih simulacija MD i MM/GBSA proračuna, iz prvobitne kolekcije identifikovano je 5 jedinjenja koja poseduju najveći potencijal efikasne inhibicije EZH2. Ove strukture imaju očuvan šablon ključnih interakcija sa vezivnim mestom EZH2 i dobar odnos procenjene energije vezivanja i molekulske mase usled čega predstavljaju dobru polaznu tačku u razvoju novog hemotipa EZH2 inhibitora. Modifikacijom jedne od struktura eliminisan je predviđeni afinitet za Pgp a istovremeno je poboljšana energija vezivanja za

EZH2. Model vezivnog mesta EZH2, simulacije molekularne dinamike, prediktivni modeli za Pgp i analiza senzitivnosti deskriptora predstavljaju izuzetno efikasan set komplementarnih metoda kojima se racionalno može izvršiti optimizacija EZH2 inhibitora.

Objavljeni rezultati iz doktorske disertacije

M. Kalinić, M. Zloh, S. Erić (2014) Structural insights into binding of small molecule inhibitors to Enhancer of Zeste Homolog 2. *J. Comput. Aided Mol. Des.* 28(11): 1109–1128.

(IF2013: 2,782; **M21**)

S. Erić, **M. Kalinić**, K. Ilić, M. Zloh (2014) Computational classification models for predicting the interaction of drugs with P-glycoprotein and breast cancer resistance protein. *SAR QSAR Environ. Res.* 25(12): 939–966.

(IF2013: 1,924; **M21**)

S. Erić, **M. Kalinić** (2015) Računarski modeli za predviđanje transporta lekova posredovanog P-glikoproteinom. *Arh. Farm.* 65(2): 89–114.

(**M53**)

Poglavlje 6

Literatura

- [1] J. Ferlay, E. Steliarova-Foucher, J. Lortet-Tieulent, S. Rosso, J. Coebergh, H. Comber, et al. (2013) Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur. J. Cancer* 49(6): 1374–1403.
- [2] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, D. Forman (2011) Global cancer statistics. *CA. Cancer J. Clin.* 61(2): 69–90.
- [3] R. Lozano, M. Naghavi, K. Foreman, S. Lim, K. Shibuya, V. Aboyans, et al. (2013) Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 380(9859): 2095–2128.
- [4] M. Mistry, D. Parkin, A. S. Ahmad, P. Sasieni (2011) Cancer incidence in the United Kingdom: projections to the year 2030. *Br. J. Cancer* 105(11): 1795–1803.
- [5] H. K. Weir, T. D. Thompson, A. Soman, B. Møller, S. Leadbetter (2015) The past, present, and future of cancer incidence in the United States: 1975 through 2020. *Cancer* 121(11): 1827–1837.
- [6] J. De Bono, A. Ashworth (2010) Translating cancer research into targeted therapeutics. *Nature* 467(7315): 543–549.
- [7] B. J. Druker, C. L. Sawyers, H. Kantarjian, D. J. Resta, S. F. Reese, J. M. Ford, et al. (2001) Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome. *N. Engl. J. Med.* 344(14): 1038–1042.

- [8] R. Capdeville, E. Buchdunger, J. Zimmermann, A. Matter (2002) Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nat. Rev. Drug Discovery* 1(7): 493–502.
- [9] C. Fausel (2007) Targeted chronic myeloid leukemia therapy: seeking a cure. *Am. J. Health. Syst. Pharm.* 64(24 Supplement 15): S9–S15.
- [10] F. Ciardiello, D. Arnold, P. Casali, A. Cervantes, J.-Y. Douillard, A. Eggermont, et al. (2014) Delivering precision medicine in oncology today and in future – the promise and challenges of personalised cancer medicine: a position paper by the European Society for Medical Oncology (ESMO). *Ann. Oncol.* 25(9): 1673–1678.
- [11] M. Arnedos, P. Vielh, J.-C. Soria, F. Andre (2014) The genetic complexity of common cancers and the promise of personalized medicine: is there any hope? *J. Pathol.* 232(2): 274–282.
- [12] R. L. Schilsky (2010) Personalized medicine in oncology: the future is now. *Nat. Rev. Drug Discovery* 9(5): 363–366.
- [13] L. Van Speybroeck (2002) From Epigenesis to Epigenetics. *Ann. N. Y. Acad. Sci.* 981(1): 61–81.
- [14] R. Holliday (2006) Epigenetics: a historical overview. *Epigenetics* 1(2): 76–80.
- [15] S. L. Berger, T. Kouzarides, R. Shiekhattar, A. Shilatifard (2009) An operational definition of epigenetics. *Genes Dev.* 23(7): 781–783.
- [16] A. Bird (2007) Perceptions of epigenetics. *Nature* 447(7143): 396–398.
- [17] M. Berdasco, M. Esteller (2010) Aberrant epigenetic landscape in cancer: how cellular identity goes awry. *Dev. Cell* 19(5): 698–711.
- [18] D. Hanahan, R. A. Weinberg (2011) Hallmarks of cancer: the next generation. *Cell* 144(5): 646–674.
- [19] T. K. Kelly, D. D. De Carvalho, P. A. Jones (2010) Epigenetic modifications as therapeutic targets. *Nat. Biotechnol.* 28(10): 1069–1078.
- [20] A. Mai, L. Altucci (2009) Epi-drugs to fight cancer: from chemistry to cancer treatment, the road ahead. *Int. J. Biochem. Cell Biol.* 41(1): 199–213.

- [21] K. Helin, D. Dhanak (**2013**) Chromatin proteins and modifications as drug targets. *Nature* 502(7472): 480–488.
- [22] P. Jones (**2012**) Development of second generation epigenetic agents. *MedChemComm* 3(2): 135–161.
- [23] D. Dhanak, P. Jackson (**2014**) Development and classes of epigenetic drugs for cancer. *Biochem. Biophys. Res. Commun.* 455(1): 58–69.
- [24] R. A. Copeland, M. E. Solomon, V. M. Richon (**2009**) Protein methyltransferases as a target class for drug discovery. *Nat. Rev. Drug Discovery* 8(9): 724–732.
- [25] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, T. J. Richmond (**1997**) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389(6648): 251–260.
- [26] K. Luger, M. L. Dechassa, D. J. Tremethick (**2012**) New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat. Rev. Mol. Cell Biol.* 13(7): 436–447.
- [27] S. Chakravarthy, Y.-J. Park, J. Chodaparambil, R. S. Edayathumangalam, K. Luger (**2005**) Structure and dynamic properties of nucleosome core particles. *FEBS Lett.* 579(4): 895–898.
- [28] A. D. Goldberg, C. D. Allis, E. Bernstein (**2007**) Epigenetics: a landscape takes shape. *Cell* 128(4): 635–638.
- [29] J. A. Law, S. E. Jacobsen (**2010**) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11(3): 204–220.
- [30] R. J. Klose, A. P. Bird (**2006**) Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* 31(2): 89–97.
- [31] F. Chédin (**2011**) The DNMT3 Family of Mammalian De Novo DNA Methyltransferases. In: *Progress in Molecular Biology and Translational Science*, volume 101, pp. 255–285, Elsevier, ISBN 9780123876850.
- [32] P. A. Jones, G. Liang (**2009**) Rethinking how DNA methylation patterns are maintained. *Nat. Rev. Genet.* 10(11): 805–811.

- [33] S. C. Wu, Y. Zhang (**2010**) Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol.* 11(9): 607–620.
- [34] P. A. Jones (**2012**) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* 13(7): 484–492.
- [35] M. Tan, H. Luo, S. Lee, F. Jin, J. S. Yang, E. Montellier, et al. (**2011**) Identification of 67 Histone Marks and Histone Lysine Crotonylation as a New Type of Histone Modification. *Cell* 146(6): 1016–1028.
- [36] T. Kouzarides (**2007**) Chromatin Modifications and Their Function. *Cell* 128(4): 693–705.
- [37] A. J. Bannister, T. Kouzarides (**2011**) Regulation of chromatin by histone modifications. *Cell Res.* 21(3): 381–395.
- [38] L. Manelyte, G. Langst (**2013**) Chromatin Remodelers and Their Way of Action. In: *Chromatin Remodelling*, (ed.) D. Radzioch, InTech, ISBN 978-953-51-1087-3.
- [39] W. Fischle, Y. Wang, C. D. Allis (**2003**) Histone and chromatin cross-talk. *Curr. Opin. Cell Biol.* 15(2): 172–183.
- [40] J. A. Latham, S. Y. R. Dent (**2007**) Cross-regulation of histone modifications. *Nat. Struct. Mol. Biol.* 14(11): 1017–1024.
- [41] T. Suganuma, J. L. Workman (**2008**) Crosstalk among Histone Modifications. *Cell* 135(4): 604–607.
- [42] J.-S. Lee, E. Smith, A. Shilatifard (**2010**) The Language of Histone Crosstalk. *Cell* 142(5): 682–685.
- [43] B. D. Strahl, C. D. Allis (**2000**) The language of covalent histone modifications. *Nature* 403(6765): 41–45.
- [44] T. Jenuwein, C. D. Allis (**2001**) Translating the Histone Code. *Science* 293(5532): 1074–1080.
- [45] K. E. Gardner, C. D. Allis, B. D. Strahl (**2011**) OPERating ON Chromatin, a Colorful Language where Context Matters. *J. Mol. Biol.* 409(1): 36–46.
- [46] M. D. Shahbazian, M. Grunstein (**2007**) Functions of Site-Specific Histone Acetylation and Deacetylation. *Annu. Rev. Biochem.* 76(1): 75–100.

- [47] X.-J. Yang, E. Seto (2007) HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. *Oncogene* 26(37): 5310–5318.
- [48] K. K. Lee, J. L. Workman (2007) Histone acetyltransferase complexes: one size doesn't fit all. *Nat. Rev. Mol. Cell Biol.* 8(4): 284–295.
- [49] M. Haberland, R. L. Montgomery, E. N. Olson (2009) The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nat. Rev. Genet.* 10(1): 32–42.
- [50] M. Iizuka, M. M. Smith (2003) Functional consequences of histone modifications. *Curr. Opin. Genet. Dev.* 13(2): 154–160.
- [51] R. J. Sims III, K. Nishioka, D. Reinberg (2003) Histone lysine methylation: a signature for chromatin function. *Trends Genet.* 19(11): 629–639.
- [52] E. L. Greer, Y. Shi (2012) Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* 13(5): 343–357.
- [53] J. C. Black, C. Van Rechem, J. R. Whetstine (2012) Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. *Mol. Cell* 48(4): 491–507.
- [54] S. J. Nowak, V. G. Corces (2004) Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation. *Trends Genet.* 20(4): 214–220.
- [55] A. Sawicka, C. Seiser (2012) Histone H3 phosphorylation – A versatile chromatin modification for different occasions. *Biochimie* 94(11): 2193–2201.
- [56] V. M. Weake, J. L. Workman (2008) Histone Ubiquitination: Triggering Gene Activity. *Mol. Cell* 29(6): 653–663.
- [57] Y. Shiio, R. N. Eisenman (2003) Histone sumoylation is associated with transcriptional repression. *Proc. Natl. Acad. Sci. U. S. A.* 100(23): 13225–13230.
- [58] J. A. Iñiguez-Lluhí (2006) For a Healthy Histone Code, a Little SUMO in the Tail Keeps the Acetyl Away. *ACS Chem. Biol.* 1(4): 204–206.
- [59] C. J. Nelson, H. Santos-Rosa, T. Kouzarides (2006) Proline Isomerization of Histone H3 Regulates Lysine Methylation and Gene Expression. *Cell* 126(5): 905–916.

- [60] J. Jones, C. Causey, B. Knuckley, J. L. Slack-Noyes, P. R. Thompson (2009) Protein arginine deiminase 4 (PAD4): current understanding and future therapeutic potential. *Curr. Opin. Drug Discovery Dev.* 12(5): 616–627.
- [61] K. L. Bicker, P. R. Thompson (2013) The protein arginine deiminases: Structure, function, inhibition, and disease. *Biopolymers* 99(2): 155–163.
- [62] S. Messner, M. O. Hottiger (2011) Histone ADP-ribosylation in DNA repair, replication and transcription. *Trends Cell Biol.* 21(9): 534–542.
- [63] A. Morgunkova, N. A. Barlev (2006) Lysine Methylation Goes Global. *Cell Cycle* 5(12): 1308–1312.
- [64] J. Huang, S. L. Berger (2008) The emerging field of dynamic lysine methylation of non-histone proteins. *Curr. Opin. Genet. Dev.* 18(2): 152–158.
- [65] S. C. Dillon, X. Zhang, R. C. Trievel, X. Cheng (2005) The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol.* 6(8): 227.
- [66] C. Qian, M.-M. Zhou (2006) SET domain protein lysine methyltransferases: Structure, specificity and catalysis. *Cell. Mol. Life Sci.* 63(23): 2755–2763.
- [67] M. Schapira (2011) Structural Chemistry of Human SET Domain Protein Methyltransferases. *Curr. Chem. Genomics* 5(Suppl 1-M5): 85–94.
- [68] B. Xiao, C. Jing, J. R. Wilson, P. A. Walker, N. Vasisht, G. Kelly, et al. (2003) Structure and catalytic mechanism of the human histone methyltransferase SET7/9. *Nature* 421(6923): 652–656.
- [69] P. Hu, Y. Zhang (2006) Catalytic mechanism and product specificity of the histone lysine methyltransferase SET7/9: An ab initio QM/MM-FE study with multiple initial structures. *J. Am. Chem. Soc.* 128(4): 1272–1278.
- [70] H.-B. Guo, H. Guo (2007) Mechanism of histone methylation catalyzed by protein lysine methyltransferase SET7/9 and origin of product specificity. *Proc. Natl. Acad. Sci. U. S. A.* 104(21): 8797–8802.
- [71] J.-F. Couture, L. M. A. Dirk, J. S. Brunzelle, R. L. Houtz, R. C. Trievel (2008) Structural origins for the product specificity of SET domain protein methyltransferases. *Proc. Natl. Acad. Sci. U. S. A.* 105(52): 20659–20664.

- [72] X. Zhang, T. C. Bruice (**2008**) Enzymatic mechanism and product specificity of SET-domain protein lysine methyltransferases. *Proc. Natl. Acad. Sci. U. S. A.* 105(15): 5728–5732.
- [73] X. Cheng, X. Zhang (**2007**) Structural dynamics of protein lysine methylation and demethylation. *Mutat. Res., Fundam. Mol. Mech. Mutagen.* 618(1–2): 102–115.
- [74] D. B. Yap, J. Chu, T. Berg, M. Schapira, S.-W. G. Cheng, A. Moradian, et al. (**2011**) Somatic mutations at EZH2 Y641 act dominantly through a mechanism of selectively altered PRC2 catalytic activity, to increase H3K27 trimethylation. *Blood* 117(8): 2451–2459.
- [75] H. Hou, H. Yu (**2010**) Structural insights into histone lysine demethylation. *Curr. Opin. Struct. Biol.* 20(6): 739–748.
- [76] M. T. Bedford, S. Richard (**2005**) Arginine Methylation: An Emerging Regulator of Protein Function. *Mol. Cell* 18(3): 263–272.
- [77] M. T. Bedford, S. G. Clarke (**2009**) Protein Arginine Methylation in Mammals: Who, What, and Why. *Mol. Cell* 33(1): 1–13.
- [78] B. Chang, Y. Chen, Y. Zhao, R. K. Bruick (**2007**) JMJD6 Is a Histone Arginine Demethylase. *Science* 318(5849): 444–447.
- [79] A. Di Lorenzo, M. T. Bedford (**2011**) Histone arginine methylation. *FEBS Lett.* 585(13): 2024–2031.
- [80] S. C. Hodawadekar, R. Marmorstein (**2007**) Chemistry of acetyl transfer by histone modifying enzymes: structure, mechanism and implications for effector design. *Oncogene* 26(37): 5528–5540.
- [81] P. Gallinari, S. D. Marco, P. Jones, M. Pallaoro, C. Steinkühler (**2007**) HDACs, histone deacetylation and gene transcription: from molecular biology to cancer therapeutics. *Cell Res.* 17(3): 195–211.
- [82] M. Shogren-Knaak, H. Ishii, J.-M. Sun, M. J. Pazin, J. R. Davie, C. L. Peterson (**2006**) Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. *Science* 311(5762): 844–847.
- [83] M. A. McBrian, I. S. Behbahan, R. Ferrari, T. Su, T.-W. Huang, K. Li, et al. (**2013**) Histone Acetylation Regulates Intracellular pH. *Mol. Cell* 49(2): 310–321.

- [84] D. Holoch, D. Moazed (**2015**) RNA-mediated epigenetic regulation of gene expression. *Nat. Rev. Genet.* 16(2): 71–84.
- [85] M. Esteller (**2011**) Non-coding RNAs in human disease. *Nat. Rev. Genet.* 12(12): 861–874.
- [86] T. R. Mercer, J. S. Mattick (**2013**) Structure and function of long noncoding RNAs in epigenetic regulation. *Nat. Struct. Mol. Biol.* 20(3): 300–307.
- [87] K. C. Wang, H. Y. Chang (**2011**) Molecular Mechanisms of Long Noncoding RNAs. *Mol. Cell* 43(6): 904–914.
- [88] J. Winter, S. Jung, S. Keller, R. I. Gregory, S. Diederichs (**2009**) Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat. Cell Biol.* 11(3): 228–234.
- [89] D. P. Bartel (**2009**) MicroRNAs: Target Recognition and Regulatory Functions. *Cell* 136(2): 215–233.
- [90] W. Filipowicz, S. N. Bhattacharyya, N. Sonenberg (**2008**) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Rev. Genet.* 2008(2): 102–114.
- [91] M. R. Fabian, N. Sonenberg, W. Filipowicz (**2010**) Regulation of mRNA Translation and Stability by microRNAs. *Annu. Rev. Biochem.* 79(1): 351–379.
- [92] J. Krol, I. Loedige, W. Filipowicz (**2010**) The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.* 11: 597–610.
- [93] E. Dai, X. Yu, Y. Zhang, F. Meng, S. Wang, X. Liu, et al. (**2014**) EpimiR: a database of curated mutual regulation between miRNAs and epigenetic modifications. *Database* 2014: bau023.
- [94] S. K. Zaidi, A. J. Van Wijnen, J. B. Lian, J. L. Stein, G. S. Stein (**2013**) Targeting deregulated epigenetic control in cancer. *J. Cell. Physiol.* 228(11): 2103–2108.
- [95] H. Shen, P. W. Laird (**2013**) Interplay between the Cancer Genome and Epigenome. *Cell* 153(1): 38–55.
- [96] M. Hatziapostolou, D. Iliopoulos (**2011**) Epigenetic aberrations during oncogenesis. *Cell. Mol. Life Sci.* 68(10): 1681–1702.

- [97] M. Ehrlich (**2002**) DNA methylation in cancer: too much, but also too little. *Oncogene* 21: 5400–5413.
- [98] S. Sharma, T. K. Kelly, P. A. Jones (**2010**) Epigenetics in cancer. *Carcinogenesis* 31(1): 27–36.
- [99] T. Waldmann, R. Schneider (**2013**) Targeting histone modifications — epigenetics in cancer. *Curr. Opin. Cell Biol.* 25(2): 184–189.
- [100] D. Mottet, V. Castronovo (**2007**) Histone deacetylases: target enzymes for cancer therapy. *Clin. Exp. Metastasis* 25(2): 183–189.
- [101] K. M. Bernt, N. Zhu, A. U. Sinha, S. Vempati, J. Faber, A. V. Krivtsov, et al. (**2011**) MLL-Rearranged Leukemia Is Dependent on Aberrant H3K79 Methylation by DOT1L. *Cancer Cell* 20(1): 66–78.
- [102] P. Chi, C. D. Allis, G. G. Wang (**2010**) Covalent histone modifications — miswritten, misinterpreted and mis-erased in human cancers. *Nat. Rev. Cancer* 10(7): 457–469.
- [103] G. G. Wang, C. D. Allis, P. Chi (**2007**) Chromatin remodeling and cancer, part I: covalent histone modifications. *Trends Mol. Med.* 13(9): 363–372.
- [104] M. Malumbres (**2013**) miRNAs and cancer: An epigenetics view. *Mol. Aspects Med.* 34(4): 863–874.
- [105] S. B. Baylin, P. A. Jones (**2011**) A decade of exploring the cancer epigenome—biological and translational implications. *Nat. Rev. Cancer* 11(10): 726–734.
- [106] I. I. Wistuba, J. G. Gelovani, J. J. Jacoby, S. E. Davis, R. S. Herbst (**2011**) Methodological and practical challenges for personalized cancer therapies. *Nat. Rev. Clin. Oncol.* 8(3): 135–141.
- [107] B. Mair, S. Kubicek, S. M. B. Nijman (**2014**) Exploiting epigenetic vulnerabilities for cancer therapeutics. *Trends Pharmacol. Sci.* 35(3): 136–145.
- [108] I. B. Weinstein, A. Joe (**2008**) Oncogene Addiction. *Cancer Res.* 68(9): 3077–3080.
- [109] J. Luo, N. L. Solimini, S. J. Elledge (**2009**) Principles of Cancer Therapy: Oncogene and Non-oncogene Addiction. *Cell* 136(5): 823–837.
- [110] W. G. Kaelin Jr. (**2005**) The Concept of Synthetic Lethality in the Context of Anticancer Therapy. *Nat. Rev. Cancer* 5(9): 689–698.

- [111] W. G. Kaelin Jr. (2009) Synthetic lethality: a framework for the development of wiser cancer therapeutics. *Genome Med.* 1(10): 1–6.
- [112] R. M. Campbell, P. J. Tummino (2014) Cancer epigenetics drug discovery and development: the challenge of hitting the mark. *J. Clin. Invest.* 124(1): 64–69.
- [113] A. Ganesan, L. Nolan, S. J. Crabb, G. Packham (2009) Epigenetic Therapy: Histone Acetylation, DNA Methylation and Anti-Cancer Drug Discovery. *Curr. Cancer Drug Targets* 9(8): 963–981.
- [114] A. Spannhoff, W. Sippl, M. Jung (2009) Cancer treatment of the future: Inhibitors of histone methyltransferases. *Int. J. Biochem. Cell Biol.* 41(1): 4–11.
- [115] E. M. Green, O. Gozani (2012) Everybody’s welcome: The big tent approach to epigenetic drug discovery. *Drug Discovery Today: Ther. Strategies* 9(2-3): e75–e81.
- [116] C. H. Arrowsmith, C. Bountra, P. V. Fish, K. Lee, M. Schapira (2012) Epigenetic protein families: a new frontier for drug discovery. *Nat. Rev. Drug Discovery* 11(5): 384–400.
- [117] D. DeCarlo, M. K. Hadden (2012) Oncoepigenomics: Making histone lysine methylation count. *Eur. J. Med. Chem.* 56: 179–194.
- [118] R. Margueron, D. Reinberg (2011) The Polycomb complex PRC2 and its mark in life. *Nature* 469(7330): 343–349.
- [119] A. Kuzmichev, K. Nishioka, H. Erdjument-Bromage, P. Tempst, D. Reinberg (2002) Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev.* 16(22): 2893–2905.
- [120] R. Cao, Y. Zhang (2004) SUZ12 Is Required for Both the Histone Methyltransferase Activity and the Silencing Function of the EED-EZH2 Complex. *Mol. Cell* 15(1): 57–67.
- [121] D. Pasini, A. P. Bracken, M. R. Jensen, E. L. Denchi, K. Helin (2004) Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J.* 23(20): 4061–4071.

- [122] H. Kim, K. Kang, J. Kim (2009) AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2. *Nucleic Acids Res.* 37(9): 2940–2950.
- [123] C. Ciferri, G. C. Lander, A. Maiolica, F. Herzog, R. Aebersold, E. Nogales (2012) Molecular architecture of human polycomb repressive complex 2. *eLife* 1: e00005.
- [124] L. A. Boyer, R. R. Latek, C. L. Peterson (2004) The SANT domain: a unique histone-tail-binding module? *Nat. Rev. Mol. Cell Biol.* 5(2): 158–163.
- [125] H. Wu, H. Zeng, A. Dong, F. Li, H. He, G. Senisterra, et al. (2013) Structure of the Catalytic Domain of EZH2 Reveals Conformational Plasticity in Cofactor and Substrate Binding Sites and Explains Oncogenic Mutations. *PLoS ONE* 8(12): e83737.
- [126] S. Antonysamy, B. Condon, Z. Druzina, J. B. Bonanno, T. Gheyi, F. Zhang, et al. (2013) Structural Context of Disease-Associated Mutations and Putative Mechanism of Autoinhibition Revealed by X-Ray Crystallographic Analysis of the EZH2-SET Domain. *PLoS ONE* 8(12): e84147.
- [127] A. Laugesen, K. Helin (2014) Chromatin Repressive Complexes in Stem Cells, Development, and Cancer. *Cell Stem Cell* 14(6): 735–751.
- [128] A. H. F. M. Peters, S. Kubicek, K. Mechtler, R. J. O’Sullivan, A. A. H. A. Derijck, L. Perez-Burgos, et al. (2003) Partitioning and Plasticity of Repressive Histone Methylation States in Mammalian Chromatin. *Mol. Cell* 12(6): 1577–1589.
- [129] Y. B. Schwartz, V. Pirrotta (2013) A new world of Polycombs: unexpected partnerships and emerging functions. *Nat. Rev. Genet.* 14(12): 853–864.
- [130] C. Xu, C. Bian, W. Yang, M. Galka, H. Ouyang, C. Chen, et al. (2010) Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). *Proc. Natl. Acad. Sci. U. S. A.* 107(45): 19266–19271.
- [131] R. Margueron, G. Li, K. Sarma, A. Blais, J. Zavadil, C. L. Woodcock, et al. (2008) Ezh1 and Ezh2 Maintain Repressive Chromatin through Different Mechanisms. *Mol. Cell* 32(4): 503–518.
- [132] E. Ezhkova, W.-H. Lien, N. Stokes, H. A. Pasolli, J. M. Silva, E. Fuchs (2011) EZH1 and EZH2 cogovern histone H3K27 trimethylation and are essential for hair follicle homeostasis and wound repair. *Genes Dev.* 25(5): 485–498.

- [133] A. Sparmann, M. van Lohuizen (**2006**) Polycomb silencers control cell fate, development and cancer. *Nat. Rev. Cancer* 6(11): 846–856.
- [134] Y.-H. Chen, M.-C. Hung, L.-Y. Li (**2012**) EZH2: a pivotal regulator in controlling cell differentiation. *Am. J. Transl. Res.* 4(4): 364–375.
- [135] A. Chase, N. C. P. Cross (**2011**) Aberrations of EZH2 in Cancer. *Clin. Cancer Res.* 17(9): 2613–2618.
- [136] B. G. Wilson, X. Wang, X. Shen, E. S. McKenna, M. E. Lemieux, Y.-J. Cho, et al. (**2010**) Epigenetic Antagonism between Polycomb and SWI/SNF Complexes during Oncogenic Transformation. *Cancer Cell* 18(4): 316–328.
- [137] C. Beisel, R. Paro (**2011**) Silencing chromatin: comparing modes and mechanisms. *Nat. Rev. Genet.* 12(2): 123–135.
- [138] A. P. Bracken, N. Dietrich, D. Pasini, K. H. Hansen, K. Helin (**2006**) Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev.* 20(9): 1123–1136.
- [139] I. Aldiri, M. L. Vetter (**2012**) PRC2 during vertebrate organogenesis: A complex in transition. *Dev. Biol.* 367(2): 91–99.
- [140] K. Lund, P. D. Adams, M. Copland (**2014**) EZH2 in normal and malignant hematopoiesis. *Leukemia* 28(1): 44–49.
- [141] S. Varambally, Q. Cao, R.-S. Mani, S. Shankar, X. Wang, B. Ateeq, et al. (**2008**) Genomic Loss of microRNA-101 Leads to Overexpression of Histone Methyltransferase EZH2 in Cancer. *Science* 322(5908): 1695–1699.
- [142] J. M. Friedman, G. Liang, C.-C. Liu, E. M. Wolff, Y. C. Tsai, W. Ye, et al. (**2009**) The Putative Tumor Suppressor microRNA-101 Modulates the Cancer Epigenome by Repressing the Polycomb Group Protein EZH2. *Cancer Res.* 69(6): 2623–2629.
- [143] J. A. Simon, C. A. Lange (**2008**) Roles of the EZH2 histone methyltransferase in cancer epigenetics. *Mutat. Res., Fundam. Mol. Mech. Mutagen.* 647(1–2): 21–29.
- [144] M. Widschwendter, H. Fieggl, D. Egle, E. Mueller-Holzner, G. Spizzo, C. Marth, et al. (**2007**) Epigenetic stem cell signature in cancer. *Nat. Genet.* 39(2): 157–158.

- [145] J.-z. Tan, Y. Yan, X.-x. Wang, Y. Jiang, H. E. Xu (2014) EZH2: biology, disease, and structure-based drug discovery. *Acta Pharmacol. Sin.* 35(2): 161–174.
- [146] M. Smits, S. E. Mir, R. J. A. Nilsson, P. M. van der Stoop, J. M. Niers, V. E. Marquez, et al. (2011) Down-Regulation of miR-101 in Endothelial Cells Promotes Blood Vessel Formation through Reduced Repression of EZH2. *PLoS ONE* 6(1): e16282.
- [147] J. Carvalho, N. C. van Grieken, P. M. Pereira, S. Sousa, M. Tijssen, T. E. Buffart, et al. (2012) Lack of microRNA-101 causes E-cadherin functional deregulation through EZH2 up-regulation in intestinal gastric cancer. *J. Pathol.* 228(1): 31–44.
- [148] L. F. Doherty, J. G. Bromer, Y. Zhou, T. S. Aldad, H. S. Taylor (2010) In Utero Exposure to Diethylstilbestrol (DES) or Bisphenol-A (BPA) Increases EZH2 Expression in the Mammary Gland: An Epigenetic Mechanism Linking Endocrine Disruptors to Breast Cancer. *Horm. Cancer* 1(3): 146–155.
- [149] C. G. Kleer, Q. Cao, S. Varambally, R. Shen, I. Ota, S. A. Tomlins, et al. (2003) EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc. Natl. Acad. Sci. U. S. A.* 100(20): 11606–11611.
- [150] K. Collett, G. E. Eide, J. Arnes, I. M. Stefansson, J. Eide, A. Braaten, et al. (2006) Expression of enhancer of zeste homologue 2 is significantly associated with increased tumor cell proliferation and is a marker of aggressive breast cancer. *Clin. Cancer Res.* 12(4): 1168–1174.
- [151] M. E. Gonzalez, X. Li, K. Toy, M. DuPrie, A. C. Ventura, M. Banerjee, et al. (2008) Downregulation of EZH2 decreases growth of estrogen receptor-negative invasive breast carcinoma and requires BRCA1. *Oncogene* 28(6): 843–853.
- [152] C.-J. Chang, J.-Y. Yang, W. Xia, C.-T. Chen, X. Xie, C.-H. Chao, et al. (2011) EZH2 promotes expansion of breast tumor initiating cells through activation of RAF1- β -catenin signaling. *Cancer Cell* 19(1): 86–100.
- [153] K. H. Yoo, L. Hennighausen (2011) EZH2 Methyltransferase and H3K27 Methylation in Breast Cancer. *Int. J. Biol. Sci.* 8(1): 59–65.

- [154] S. Varambally, S. M. Dhanasekaran, M. Zhou, T. R. Barrette, C. Kumar-Sinha, M. G. Sanda, et al. (2002) The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* 419(6907): 624–629.
- [155] R. Bryant, N. Cross, C. Eaton, F. Hamdy, V. Cunliffe (2007) EZH2 promotes proliferation and invasiveness of prostate cancer cells. *Prostate* 67(5): 547–556.
- [156] X.-S. Ke, Y. Qu, K. Rostad, W.-C. Li, B. Lin, O. J. Halvorsen, et al. (2009) Genome-Wide Profiling of Histone H3 Lysine 4 and Lysine 27 Trimethylation Reveals an Epigenetic Signature in Prostate Carcinogenesis. *PLoS ONE* 4(3): e4687.
- [157] B. D. Karanikolas, M. L. Figueiredo, L. Wu (2010) Comprehensive evaluation of the role of EZH2 in the growth, invasion, and aggression of a panel of prostate cancer cell lines. *Prostate* 70(6): 675–688.
- [158] Y. J. Shin, J.-H. Kim (2012) The Role of EZH2 in the Regulation of the Activity of Matrix Metalloproteinases in Prostate Cancer Cells. *PLoS ONE* 7(1): e30393.
- [159] M.-L. Suvà, N. Riggi, M. Janiszewska, I. Radovanovic, P. Provero, J.-C. Stehle, et al. (2009) EZH2 Is Essential for Glioblastoma Cancer Stem Cell Maintenance. *Cancer Res.* 69(24): 9211–9218.
- [160] M. Smits, J. Nilsson, S. E. Mir, P. M. van der Stoop, E. Hulleman, J. M. Niers, et al. (2011) miR-101 is down-regulated in glioblastoma resulting in EZH2-induced proliferation, migration, and angiogenesis. *Oncotarget* 1(8): 710–720.
- [161] J. D. Raman, N. P. Mongan, S. K. Tickoo, S. A. Boorjian, D. S. Scherr, L. J. Gudas (2005) Increased Expression of the Polycomb Group Gene, EZH2, in Transitional Cell Carcinoma of the Bladder. *Clin. Cancer Res.* 11(24): 8570–8576.
- [162] S. Weikert, F. Christoph, J. Köllermann, M. Müller, M. Schrader, K. Miller, et al. (2005) Expression levels of the EZH2 polycomb transcriptional repressor correlate with aggressiveness and invasive potential of bladder carcinomas. *Int. J. Mol. Med.* 16(2): 349–353.
- [163] J. Tiffen, S. J. Gallagher, P. Hersey (2015) EZH2: an emerging role in melanoma biology and strategies for targeted therapy. *Pigm. Cell Melanoma Res.* 28(1): 21–30.

- [164] H. Li, R. Zhang (2013) Role of EZH2 in Epithelial Ovarian Cancer: From Biological Insights to Therapeutic Target. *Front. Oncol.* 3(47): 1–5.
- [165] R. H. Breuer, P. J. Snijders, E. F. Smit, T. G. Sutedja, R. G. Sewalt, A. P. Otte, et al. (2004) Increased expression of the EZH2 polycomb group gene in BMI-1-positive neoplastic cells during bronchial carcinogenesis. *Neoplasia* 6(6): 736–743.
- [166] J. Kikuchi, I. Kinoshita, Y. Shimizu, E. Kikuchi, J. Konishi, S. Oizumi, et al. (2010) Distinctive expression of the polycomb group proteins Bmi1 polycomb ring finger oncogene and enhancer of zeste homolog 2 in nonsmall cell lung cancers and their clinical and clinicopathologic significance. *Cancer* 116(12): 3015–3024.
- [167] A. P. Bracken, D. Pasini, M. Capra, E. Prosperini, E. Colli, K. Helin (2003) EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *EMBO J.* 22(20): 5323–5335.
- [168] J. Tan, X. Yang, L. Zhuang, X. Jiang, W. Chen, P. L. Lee, et al. (2007) Pharmacologic disruption of Polycomb-repressive complex 2-mediated gene repression selectively induces apoptosis in cancer cells. *Genes Dev.* 21(9): 1050–1063.
- [169] A. Hayden, P. W. M. Johnson, G. Packham, S. J. Crabb (2010) S-adenosylhomocysteine hydrolase inhibition by 3-deazaneplanocin A analogues induces anti-cancer effects in breast cancer cell lines and synergy with both histone deacetylase and HER2 inhibition. *Breast Cancer Res. Treat.* 127(1): 109–119.
- [170] Z.-Y. Rao, M.-Y. Cai, G.-F. Yang, L.-R. He, S.-J. Mai, W.-F. Hua, et al. (2010) EZH2 supports ovarian carcinoma cell invasion and/or metastasis via regulation of TGF- β 1 and is a predictor of outcome in ovarian carcinoma patients. *Carcinogenesis* 31(9): 1576–1583.
- [171] J. Kikuchi, T. Takashina, I. Kinoshita, E. Kikuchi, Y. Shimizu, J. Sakakibara-Konishi, et al. (2012) Epigenetic therapy with 3-deazaneplanocin A, an inhibitor of the histone methyltransferase EZH2, inhibits growth of non-small cell lung cancer cells. *Lung Cancer* 78(2): 138–143.
- [172] Y. Xiao (2011) Enhancer of zeste homolog 2: A potential target for tumor therapy. *Int. J. Biochem. Cell Biol.* 43(4): 474–477.

- [173] R. D. Morin, N. A. Johnson, T. M. Severson, A. J. Mungall, J. An, R. Goya, et al. (2010) Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat. Genet.* 42(2): 181–185.
- [174] C. J. Sneeringer, M. P. Scott, K. W. Kuntz, S. K. Knutson, R. M. Pollock, V. M. Richon, et al. (2010) Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. *Proc. Natl. Acad. Sci. U. S. A.* 107(49): 20980–20985.
- [175] C. R. Majer, L. Jin, M. P. Scott, S. K. Knutson, K. W. Kuntz, H. Keilhack, et al. (2012) A687V EZH2 is a gain-of-function mutation found in lymphoma patients. *FEBS Lett.* 586(19): 3448–3451.
- [176] M. T. McCabe, A. P. Graves, G. Ganji, E. Diaz, W. S. Halsey, Y. Jiang, et al. (2012) Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). *Proc. Natl. Acad. Sci. U. S. A.* 109(8): 2989–2994.
- [177] K. Xu, Z. J. Wu, A. C. Groner, H. H. He, C. Cai, R. T. Lis, et al. (2012) EZH2 Oncogenic Activity in Castration-Resistant Prostate Cancer Cells Is Polycomb-Independent. *Science* 338(6113): 1465–1469.
- [178] P. K. Chiang (1998) Biological Effects of Inhibitors of S-Adenosylhomocysteine Hydrolase. *Pharmacol. Ther.* 77(2): 115–134.
- [179] M. A. Turner, X. Yang, D. Yin, K. Kuczera, R. T. Borchardt, P. L. Howell (2000) Structure and function of S-adenosylhomocysteine hydrolase. *Cell Biochem. Biophys.* 33(2): 101–125.
- [180] T. B. Miranda, C. C. Cortez, C. B. Yoo, G. Liang, M. Abe, T. K. Kelly, et al. (2009) DZNep is a global histone methylation inhibitor that reactivates developmental genes not silenced by DNA methylation. *Mol. Cancer Ther.* 8(6): 1579–1588.
- [181] M. Vedel, F. Lawrence, M. Robert-Gero, E. Lederer (1978) The antifungal antibiotic sinefungin as a very active inhibitor of methyltransferases and of the transformation of chick embryo fibroblasts by rous sarcoma virus. *Biochem. Biophys. Res. Commun.* 85(1): 371–376.

- [182] A. Basavapathruni, L. Jin, S. R. Daigle, C. R. A. Majer, C. A. Therkelsen, T. J. Wigle, et al. (2012) Conformational Adaptation Drives Potent, Selective and Durable Inhibition of the Human Protein Methyltransferase DOT1L. *Chem. Biol. Drug Des.* 80(6): 971–980.
- [183] K. Devkota, B. Lohse, Q. Liu, M.-W. Wang, D. Stärk, J. Berthelsen, et al. (2014) Analogues of the Natural Product Sinefungin as Inhibitors of EHMT1 and EHMT2. *ACS Med. Chem. Lett.* 5(4): 293–297.
- [184] E. Diaz, C. A. Machutta, S. Chen, Y. Jiang, C. Nixon, G. Hofmann, et al. (2012) Development and Validation of Reagents and Assays for EZH2 Peptide and Nucleosome High-Throughput Screens. *J. Biomol. Screen.* 17(10): 1279–1292.
- [185] S. K. Verma, X. Tian, L. V. LaFrance, C. Duquenne, D. P. Suarez, K. A. Newlander, et al. (2012) Identification of Potent, Selective, Cell-Active Inhibitors of the Histone Lysine Methyltransferase EZH2. *ACS Med. Chem. Lett.* 3(12): 1091–1096.
- [186] M. T. McCabe, H. M. Ott, G. Ganji, S. Korenchuk, C. Thompson, G. S. Van Aller, et al. (2012) EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* 492(7427): 108–112.
- [187] S. K. Knutson, T. J. Wigle, N. M. Warholic, C. J. Sneeringer, C. J. Allain, C. R. Klaus, et al. (2012) A selective inhibitor of EZH2 blocks H3K27 methylation and kills mutant lymphoma cells. *Nat. Chem. Biol.* 8(11): 890–896.
- [188] S. K. Knutson, N. M. Warholic, T. J. Wigle, C. R. Klaus, C. J. Allain, A. Raimondi, et al. (2013) Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. *Proc. Natl. Acad. Sci. U. S. A.* 110(19): 7922–7927.
- [189] W. Qi, H. Chan, L. Teng, L. Li, S. Chuai, R. Zhang, et al. (2012) Selective inhibition of Ezh2 by a small molecule inhibitor blocks tumor cells proliferation. *Proc. Natl. Acad. Sci. U. S. A.* 109(52): 21360–21365.
- [190] H. Ü. Kaniskan, K. D. Konze, J. Jin (2015) Selective Inhibitors of Protein Methyltransferases. *J. Med. Chem.* 58(4): 1596–1629.
- [191] K. D. Konze, A. Ma, F. Li, D. Barsyte-Lovejoy, T. Parton, C. J. MacNevin, et al. (2013) An Orally Bioavailable Chemical Probe of the Lysine Methyltransferases EZH2 and EZH1. *ACS Chem. Biol.* 8(6): 1324–1334.

- [192] S. K. Knutson, S. Kawano, Y. Minoshima, N. M. Warholc, K.-C. Huang, Y. Xiao, et al. (2014) Selective Inhibition of EZH2 by EPZ-6438 Leads to Potent Antitumor Activity in EZH2-Mutant Non-Hodgkin Lymphoma. *Mol. Cancer Ther.* 13(4): 842–854.
- [193] J. E. Campbell, K. W. Kuntz, S. K. Knutson, N. M. Warholc, H. Keilhack, T. J. Wigle, et al. (2015) EPZ011989, A Potent, Orally-Available EZH2 Inhibitor with Robust In Vivo Activity. *ACS Med. Chem. Lett.* 6(5): 491–495.
- [194] W. D. Bradley, S. Arora, J. Busby, S. Balasubramanian, V. S. Gehling, C. G. Nasveschuk, et al. (2015) EZH2 Inhibitor Efficacy in Non-Hodgkin’s Lymphoma Does Not Require Suppression of H3K27 Monomethylation. *Chem. Biol.* 21(11): 1463–1475.
- [195] S. Garapaty-Rao, C. Nasveschuk, A. Gagnon, E. Y. Chan, P. Sandy, J. Busby, et al. (2013) Identification of EZH2 and EZH1 Small Molecule Inhibitors with Selective Impact on Diffuse Large B Cell Lymphoma Cell Growth. *Chem. Biol.* 20(11): 1329–1339.
- [196] C. G. Nasveschuk, A. Gagnon, S. Garapaty-Rao, S. Balasubramanian, R. Campbell, C. Lee, et al. (2014) Discovery and Optimization of Tetramethylpiperidinyl Benzamides as Inhibitors of EZH2. *ACS Med. Chem. Lett.* 5(4): 378–383.
- [197] L. Zhang, X. Song, N. Wang, L. Zhao, Q. Feng, X. You, et al. (2015) Design, synthesis and biological evaluation of novel 1-methyl-3-oxo-2,3,5,6,7,8-hexahydroisoquinolins as potential EZH2 inhibitors. *RSC Adv.* 5: 25967–25978.
- [198] J. R. Simard, M. Plant, R. Emkey, V. Yu (2013) Development and Implementation of a High-Throughput AlphaLISA Assay for Identifying Inhibitors of EZH2 Methyltransferase. *Assay Drug Dev. Technol.* 11(3): 152–162.
- [199] J. Woo, H.-Y. Kim, B. J. Byun, C.-H. Chae, J. Y. Lee, S. Y. Ryu, et al. (2014) Biological evaluation of tanshindiols as EZH2 histone methyltransferase inhibitors. *Bioorg. Med. Chem. Lett.* 24(11): 2486–2492.
- [200] P.-P. Kung, B. Huang, L. Zehnder, J. Tatlock, P. Bingham, C. Krivacic, et al. (2015) SAH derived potent and selective EZH2 inhibitors. *Bioorg. Med. Chem. Lett.* 25(7): 1532–1537.

- [201] M. Vedadi, D. Barsyte-Lovejoy, F. Liu, S. Rival-Gervier, A. Allali-Hassani, V. Labrie, et al. (2011) A chemical probe selectively inhibits G9a and GLP methyltransferase activity in cells. *Nat. Chem. Biol.* 7(8): 566–574.
- [202] A. D. Ferguson, N. A. Larsen, T. Howard, H. Pollard, I. Green, C. Grande, et al. (2011) Structural Basis of Substrate Methylation and Inhibition of SMYD2. *Structure* 19(9): 1262–1273.
- [203] D. Barsyte-Lovejoy, F. Li, M. J. Oudhoff, J. H. Tatlock, A. Dong, H. Zeng, et al. (2014) (R)-PFI-2 is a potent and selective inhibitor of SETD7 methyltransferase activity in cells. *Proc. Natl. Acad. Sci. U. S. A.* 111(35): 12853–12858.
- [204] A. Ma, W. Yu, F. Li, R. M. Bleich, J. M. Herold, K. V. Butler, et al. (2014) Discovery of a Selective, Substrate-Competitive Inhibitor of the Lysine Methyltransferase SETD8. *J. Med. Chem.* 57(15): 6822–6833.
- [205] K. T. Nguyen, F. Li, G. Poda, D. Smil, M. Vedadi, M. Schapira (2013) Strategy to Target the Substrate Binding site of SET Domain Protein Methyltransferases. *J. Chem. Inf. Model.* 53(3): 681–691.
- [206] J. Smadbeck, M. B. Peterson, B. M. Zee, S. Garapaty, A. Mago, C. Lee, et al. (2014) De Novo Peptide Design and Experimental Validation of Histone Methyltransferase Inhibitors. *PLoS ONE* 9(2): e90095.
- [207] W. Kim, G. H. Bird, T. Neff, G. Guo, M. A. Kerényi, L. D. Walensky, et al. (2013) Targeted disruption of the EZH2–EED complex inhibits EZH2-dependent cancer. *Nat. Chem. Biol.* 9(10): 643–650.
- [208] X. Kong, L. Chen, L. Jiao, X. Jiang, F. Lian, J. Lu, et al. (2014) Astemizole Arrests the Proliferation of Cancer Cells by Disrupting the EZH2-EED Interaction of Polycomb Repressive Complex 2. *J. Med. Chem.* 57(22): 9512–9521.
- [209] Z. Han, X. Xing, M. Hu, Y. Zhang, P. Liu, J. Chai (2007) Structural Basis of EZH2 Recognition by EED. *Structure* 15(10): 1306–1315.
- [210] V. Campagna-Slater, M. W. Mok, K. T. Nguyen, M. Feher, R. Najmanovich, M. Schapira (2011) Structural Chemistry of the Histone Methyltransferases Cofactor Binding Site. *J. Chem. Inf. Model.* 51(3): 612–623.
- [211] R. Brown, M. Links (1999) Clinical relevance of the molecular mechanisms of resistance to anti-cancer drugs. *Expert Rev. Mol. Med.* pp. 1–21.

- [212] N. P. Shah, J. M. Nicoll, B. Nagar, M. E. Gorre, R. L. Paquette, J. Kuriyan, et al. (2002) Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. *Cancer Cell* 2(2): 117–125.
- [213] N. C. Turner, J. S. Reis-Filho (2012) Genetic heterogeneity and cancer drug resistance. *Lancet Oncol.* 13(4): e178–e185.
- [214] C. Holohan, S. Van Schaeybroeck, D. B. Longley, P. G. Johnston (2013) Cancer drug resistance: an evolving paradigm. *Nat. Rev. Cancer* 13(10): 714–726.
- [215] L. A. Garraway, P. A. Jänne (2012) Circumventing Cancer Drug Resistance in the Era of Personalized Medicine. *Cancer Discovery* 2(3): 214–226.
- [216] T. Ozben (2006) Mechanisms and strategies to overcome multiple drug resistance in cancer. *FEBS Lett.* 580(12): 2903–2909.
- [217] O. Trédan, C. M. Galmarini, K. Patel, I. F. Tannock (2007) Drug Resistance and the Solid Tumor Microenvironment. *J. Natl. Cancer Inst.* 99(19): 1441–1454.
- [218] D. Longley, P. Johnston (2005) Molecular mechanisms of drug resistance. *J. Pathol.* 205(2): 275–292.
- [219] J.-P. Gillet, M. M. Gottesman (2010) Mechanisms of Multidrug Resistance in Cancer. In: *Multi-Drug Resistance in Cancer*, (ed.) J. Zhou, number 596 in *Methods in Molecular Biology*, chapter 4, pp. 47–66, Humana Press, ISBN 978-1-60761-415-9, 978-1-60761-416-6.
- [220] M. M. Gottesman, T. Fojo, S. E. Bates (2002) Multidrug Resistance In Cancer: Role Of ATP-Dependent Transporters. *Nat. Rev. Cancer* 2(1): 48–58.
- [221] M. Mimeault, R. Hauke, S. Batra (2008) Recent Advances on the Molecular Mechanisms Involved in the Drug Resistance of Cancer Cells and Novel Targeting Therapies. *Clin. Pharmacol. Ther.* 83(5): 673–691.
- [222] P. D. W. Eckford, F. J. Sharom (2009) ABC Efflux Pump-Based Resistance to Chemotherapy Drugs. *Chem. Rev.* 109(7): 2989–3011.
- [223] R. L. Juliano, V. Ling (1976) A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. *Biochim. Biophys. Acta, Biomembr.* 455(1): 152–162.

- [224] L. M. S. Chan, S. Lowes, B. H. Hirst (2004) The ABCs of drug transport in intestine and liver: efflux proteins limiting drug absorption and bioavailability. *Eur. J. Pharm. Sci.* 21(1): 25–51.
- [225] T. N. Thompson (2011) The Clinical Significance of Drug Transporters in Drug Disposition and Drug Interactions. In: *Pharmacokinetics in Drug Development*, (eds.) P. L. Bonate, D. R. Howard, pp. 285–313, Springer US, ISBN 978-1-4419-7936-0, 978-1-4419-7937-7.
- [226] D. J. H. Lin, M. Yamazaki (2012) Role of P-Glycoprotein in Pharmacokinetics. *Clin. Pharmacokinet.* 42(1): 59–98.
- [227] K. G. Chen, J. C. Valencia, J.-P. Gillet, V. J. Hearing, M. M. Gottesman (2009) Involvement of ABC transporters in melanogenesis and the development of multidrug resistance of melanoma. *Pigm. Cell Melanoma Res.* 22(6): 740–749.
- [228] C. Sánchez, P. Mendoza, H. R. Contreras, J. Vergara, J. A. McCubrey, C. Hudobro, et al. (2009) Expression of multidrug resistance proteins in prostate cancer is related with cell sensitivity to chemotherapeutic drugs. *Prostate* 69(13): 1448–1459.
- [229] N. Walsh, A. Larkin, S. Kennedy, L. Connolly, J. Ballot, W. Ooi, et al. (2009) Expression of multidrug resistance markers ABCB1 (MDR-1/P-gp) and ABCC1 (MRP-1) in renal cell carcinoma. *BMC Urol.* 9(1): 6.
- [230] S. G. Aller, J. Yu, A. Ward, Y. Weng, S. Chittaboina, R. Zhuo, et al. (2009) Structure of P-Glycoprotein Reveals a Molecular Basis for Poly-Specific Drug Binding. *Science* 323(5922): 1718–1722.
- [231] A. B. Shapiro, V. Ling (1997) Positively Cooperative Sites for Drug Transport by P-Glycoprotein with Distinct Drug Specificities. *Eur. J. Biochem.* 250(1): 130–137.
- [232] A. B. Shapiro, K. Fox, P. Lam, V. Ling (1999) Stimulation of P-glycoprotein-mediated drug transport by prazosin and progesterone. *Eur. J. Biochem.* 259(3): 841–850.
- [233] C. Martin, G. Berridge, C. F. Higgins, P. Mistry, P. Charlton, R. Callaghan (2000) Communication between Multiple Drug Binding Sites on P-glycoprotein. *Mol. Pharmacol.* 58(3): 624–632.

- [234] A. R. Safa (**2004**) Identification and Characterization of the Binding Sites of P-Glycoprotein for Multidrug Resistance-Related Drugs and Modulators. *Curr. Med. Chem.: Anti-Cancer Agents* 4(1): 1–17.
- [235] R. Ernst, P. Kueppers, J. Stindt, K. Kuchler, L. Schmitt (**2010**) Multidrug efflux pumps: Substrate selection in ATP-binding cassette multidrug efflux pumps – first come, first served? *FEBS J.* 277(3): 540–549.
- [236] K. Wong, J. Ma, A. Rothnie, P. C. Biggin, I. D. Kerr (**2014**) Towards understanding promiscuity in multidrug efflux pumps. *Trends Biochem. Sci.* 39(1): 8–16.
- [237] L. A. Doyle, W. Yang, L. V. Abruzzo, T. Krogmann, Y. Gao, A. K. Rishi, et al. (**1998**) A multidrug resistance transporter from human MCF-7 breast cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* 95(26): 15665–15670.
- [238] K. Natarajan, Y. Xie, M. R. Baer, D. D. Ross (**2012**) Role of breast cancer resistance protein (BCRP/ABCG2) in cancer drug resistance. *Biochem. Pharmacol.* 83(8): 1084–1103.
- [239] R. W. Robey, K. K. K. To, O. Polgar, M. Dohse, P. Fetsch, M. Dean, et al. (**2009**) ABCG2: A perspective. *Adv. Drug Delivery Rev.* 61(1): 3–13.
- [240] K. M. Giacomini, S.-M. Huang, D. J. Tweedie, L. Z. Benet, K. L. R. Brouwer, X. Chu, et al. (**2010**) Membrane transporters in drug development. *Nat. Rev. Drug Discovery* 9(3): 215–236.
- [241] K. Maeda, Y. Sugiyama (**2013**) Transporter biology in drug approval: Regulatory aspects. *Mol. Aspects Med.* 34(2–3): 711–718.
- [242] J. W. Polli, S. A. Wring, J. E. Humphreys, L. Huang, J. B. Morgan, L. O. Webster, et al. (**2001**) Rational Use of in Vitro P-glycoprotein Assays in Drug Discovery. *J. Pharmacol. Exp. Ther.* 299(2): 620–628.
- [243] C. Chang, P. M. Bahadduri, J. E. Polli, P. W. Swaan, S. Ekins (**2006**) Rapid Identification of P-glycoprotein Substrates and Inhibitors. *Drug Metab. Dispos.* 34(12): 1976–1984.
- [244] N. Mizuno, T. Niwa, Y. Yotsumoto, Y. Sugiyama (**2003**) Impact of Drug Transporter Studies on Drug Discovery and Development. *Pharmacol. Rev.* 55(3): 425–461.

- [245] J. Shimada, S. Ekins, C. Elkin, E. I. Shakhnovich, J.-P. Wery (**2002**) Integrating computer-based de novo drug design and multidimensional filtering for desirable drugs. *Targets* 1(6): 196–205.
- [246] S. Ekins (**2004**) Predicting undesirable drug interactions with promiscuous proteins in silico. *Drug Discovery Today* 9(6): 276–285.
- [247] G. F. Ecker, T. Stockner, P. Chiba (**2008**) Computational models for prediction of interactions with ABC-transporters. *Drug Discovery Today* 13(7–8): 311–317.
- [248] M. Pinto, D. Digles, G. F. Ecker (**2014**) Computational models for predicting the interaction with ABC transporters. *Drug Discovery Today: Technol.* 12: e69–e77.
- [249] J. W. Polli, K. L. Olson, J. P. Chism, L. S. John-Williams, R. L. Yeager, S. M. Woodard, et al. (**2009**) An unexpected synergist role of P-glycoprotein and breast cancer resistance protein on the central nervous system penetration of the tyrosine kinase inhibitor lapatinib (N-{3-chloro-4-[(3-fluorobenzyl)oxy]phenyl}-6-[5-({2-(methylsulfonyl)ethyl}amino)methyl]-2-furyl]-4-quinazolinamine; GW572016). *Drug Metab. Dispos.* 37(2): 439–442.
- [250] H. Kodaira, H. Kusuhara, J. Ushiki, E. Fuse, Y. Sugiyama (**2010**) Kinetic analysis of the cooperation of P-glycoprotein (P-gp/Abcb1) and breast cancer resistance protein (Bcrp/Abcg2) in limiting the brain and testis penetration of erlotinib, flavopiridol, and mitoxantrone. *J. Pharmacol. Exp. Ther.* 333(3): 788–796.
- [251] J. S. Lagas, R. A. B. van Waterschoot, R. W. Sparidans, E. Wagenaar, J. H. Beijnen, A. H. Schinkel (**2010**) Breast cancer resistance protein and P-glycoprotein limit sorafenib brain accumulation. *Mol. Cancer Ther.* 9(2): 319–326.
- [252] S. C. Tang, N. A. G. Lankheet, B. Poller, E. Wagenaar, J. H. Beijnen, A. H. Schinkel (**2012**) P-glycoprotein (ABCB1) and breast cancer resistance protein (ABCG2) restrict brain accumulation of the active sunitinib metabolite N-desethyl sunitinib. *J. Pharmacol. Exp. Ther.* 341(1): 164–173.
- [253] S. Agarwal, A. M. Hartz, W. F. Elmquist, B. Bauer (**2011**) Breast Cancer Resistance Protein and P-glycoprotein in Brain Cancer: Two Gatekeepers Team Up. *Curr. Pharm. Des.* 17: 2793–2802.

- [254] H. M. Coley (**2010**) Overcoming Multidrug Resistance in Cancer: Clinical Studies of P-Glycoprotein Inhibitors. In: *Multi-drug Resistance in Cancer*, (ed.) J. Zhou, number 596 in Methods in Molecular Biology, pp. 341–358, Humana Press, Totowa, NJ, ISBN 978-1-60761-415-9, 978-1-60761-416-6.
- [255] M. Yu, A. Ocana, I. F. Tannock (**2012**) Reversal of ATP-binding cassette drug transporter activity to modulate chemoresistance: why has it failed to provide clinical benefit? *Cancer Metastasis Rev.* 32(1-2): 211–227.
- [256] M. Saraswathy, S. Gong (**2013**) Different strategies to overcome multidrug resistance in cancer. *Biotechnol. Adv.* 31(8): 1397–1407.
- [257] L. Wang (**2013**) RNAi-mediated EZH2 depletion decreases MDR1 expression and sensitizes multidrug-resistant hepatocellular carcinoma cells to chemotherapy. *Oncol. Rep.* 29: 1037–1042.
- [258] P. Zhang, M. C. de Gooijer, L. C. Buil, J. H. Beijnen, G. Li, O. van Tellingen (**2015**) ABCB1 and ABCG2 restrict the brain penetration of a panel of novel EZH2-inhibitors. *Int. J. Cancer* 137(8): 2007–2018.
- [259] J. D. Durrant, J. A. McCammon (**2011**) Molecular dynamics simulations and drug discovery. *BMC Biol.* 9(1): 71.
- [260] D. W. Borhani, D. E. Shaw (**2011**) The future of molecular dynamics simulations in drug discovery. *J. Comput. Aided Mol. Des.* 26(1): 15–26.
- [261] M. W. van der Kamp, K. E. Shaw, C. J. Woods, A. J. Mulholland (**2008**) Biomolecular simulation and modelling: status, progress and prospects. *J. R. Soc. Interface* 5(Suppl 3): 173–190.
- [262] A. D. Mackerell (**2004**) Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem.* 25(13): 1584–1604.
- [263] J. Wereszczynski, J. A. McCammon (**2012**) Statistical mechanics and molecular dynamics in evaluating thermodynamic properties of biomolecular recognition. *Q. Rev. Biophys.* 45(01): 1–25.
- [264] C. Mura, C. E. McAnany (**2014**) An introduction to biomolecular simulations and docking. *Mol. Simul.* 40(10-11): 732–764.
- [265] J. Michel (**2014**) Current and emerging opportunities for molecular simulations in structure-based drug design. *Phys. Chem. Chem. Phys.* 16(10): 4465–4477.

- [266] M. K. Gilson, H.-X. Zhou (2007) Calculation of Protein-Ligand Binding Affinities. *Annu. Rev. Biophys. Biomol. Struct.* 36(1): 21–42.
- [267] Y. Deng, B. Roux (2009) Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *J. Phys. Chem. B* 113(8): 2234–2246.
- [268] C. Abrams, G. Bussi (2013) Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. *Entropy* 16(1): 163–199.
- [269] A. Barducci, M. Bonomi, M. Parrinello (2011) Metadynamics. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* 1(5): 826–843.
- [270] C. Dellago, G. Hummer (2013) Computing Equilibrium Free Energies Using Non-Equilibrium Molecular Dynamics. *Entropy* 16(1): 41–61.
- [271] N. Homeyer, H. Gohlke (2012) Free Energy Calculations by the Molecular Mechanics Poisson–Boltzmann Surface Area Method. *Mol. Inform.* 31(2): 114–122.
- [272] J. M. Hayes, G. Archontis (2012) MM-GB(PB)SA Calculations of Protein-Ligand Binding Free Energies. In: *Molecular Dynamics – Studies of Synthetic and Biological Macromolecules*, (ed.) L. Wang, pp. 171–190, InTech, ISBN 978-953-51-0444-5.
- [273] G. Schneider (2013) De novo design – hop(p)ing against hope. *Drug Discovery Today: Technol.* 10(4): e453–e460.
- [274] L. Sean (2013) *Essentials of Metaheuristics*. 2nd edition, Lulu. URL [http://cs.gmu.edu/~protect\\$\relax\sim\\$sean/book/metaheuristics](http://cs.gmu.edu/~protect$\relax\sim$sean/book/metaheuristics).
- [275] C. A. Nicolaou, N. Brown (2013) Multi-objective optimization methods in drug design. *Drug Discovery Today: Technol.* 10(3): e427–e435.
- [276] G. Schneider, K.-H. Baringhaus (2013) De Novo Design: From Models to Molecules. In: *De novo Molecular Design*, (ed.) G. Schneider, pp. 1–55, Wiley-VCH Verlag GmbH & Co. KGaA, ISBN 9783527677016.
- [277] S. Dandapani, L. A. Marcaurelle (2010) Grand Challenge Commentary: Accessing new chemical space for ‘undruggable’ targets. *Nat. Chem. Biol.* 6(12): 861–863.

- [278] M. Hartenfeller, G. Schneider (**2011**) De Novo Drug Design. In: *Cheminformatics and Computational Chemical Biology*, (ed.) J. Bajorath, *Methods in Molecular Biology*, volume 672, pp. 299–324, Humana Press, ISBN 978-1-60761-839-3.
- [279] M. Hartenfeller, H. Zettl, M. Walter, M. Rupp, F. Reisen, E. Proschak, et al. (**2012**) DOGS: Reaction-Driven de novo Design of Bioactive Compounds. *PLoS Comput. Biol.* 8(2): e1002380.
- [280] M. Hartenfeller, M. Eberle, P. Meier, C. Nieto-Oberhuber, K.-H. Altmann, G. Schneider, et al. (**2011**) A Collection of Robust Organic Synthesis Reactions for In Silico Molecule Design. *J. Chem. Inf. Model.* 51(12): 3093–3098.
- [281] G. Schneider, T. Geppert, M. Hartenfeller, F. Reisen, A. Klenner, M. Reutlinger, et al. (**2011**) Reaction-driven de novo design, synthesis and testing of potential type II kinase inhibitors. *Future Med. Chem.* 3(4): 415–424.
- [282] B. Spänkuch, S. Keppner, L. Lange, T. Rodrigues, H. Zettl, C. P. Koch, et al. (**2013**) Drugs by Numbers: Reaction-Driven De Novo Design of Potent and Selective Anticancer Leads. *Angew. Chem. Int. Ed.* 52(17): 4676–4681.
- [283] J. Holland (**1975**) *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI, USA.
- [284] X.-S. Yang (**2014**) *Nature-Inspired Optimization Algorithms*. 1st edition, Elsevier, London, UK, ISBN 978-0-12-416743-8.
- [285] R. R. Sharapov (**2007**) Genetic Algorithms: Basic Ideas, Variants and Analysis. In: *Vision Systems: Segmentation and Pattern Recognition*, (eds.) G. Obinata, A. Dutta, pp. 407–422, InTech, Vienna, Austria, ISBN 987-3-902613-05-9.
- [286] C. A. Nicolaou, C. Kannas, E. Loizidou (**2012**) Multi-Objective Optimization Methods in De Novo Drug Design. *Mini Rev. Med. Chem.* 12(10): 979–987.
- [287] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan (**2002**) A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evolut. Comput.* 6(2): 182–197.
- [288] F. Dey, A. Caffisch (**2008**) Fragment-Based de Novo Ligand Design by Multiobjective Evolutionary Optimization. *J. Chem. Inf. Model.* 48(3): 679–690.

- [289] C. A. Nicolaou, J. Apostolakis, C. S. Pattichis (2009) De Novo Drug Design Using Multiobjective Evolutionary Graphs. *J. Chem. Inf. Model.* 49(2): 295–307.
- [290] P. S. Kutchukian, D. Lou, E. I. Shakhnovich (2009) FOG: Fragment Optimized Growth Algorithm for the de Novo Generation of Molecules Occupying Druglike Chemical Space. *J. Chem. Inf. Model.* 49(7): 1630–1642.
- [291] P. Pfeffer, T. Fober, E. Hüllermeier, G. Klebe (2010) GARLig: A Fully Automated Tool for Subset Selection of Large Fragment Spaces via a Self-Adaptive Genetic Algorithm. *J. Chem. Inf. Model.* 50(9): 1644–1659.
- [292] D. White, R. C. Wilson (2010) Generative Models for Chemical Structures. *J. Chem. Inf. Model.* 50(7): 1257–1274.
- [293] J. R. Damewood, C. L. Lerman, B. B. Masek (2010) NovoFLAP: A Ligand-Based De Novo Design Approach for the Generation of Medicinally Relevant Ideas. *J. Chem. Inf. Model.* 50(7): 1296–1303.
- [294] Q. Huang, L.-L. Li, S.-Y. Yang (2010) PhDD: A new pharmacophore-based de novo design method of drug-like molecules combined with assessment of synthetic accessibility. *J. Mol. Graph. Model.* 28(8): 775–787.
- [295] T. Lippert, T. Schulz-Gasch, O. Roche, W. Guba, M. Rarey (2011) De novo design by pharmacophore-based searches in fragment spaces. *J. Comput. Aided Mol. Des.* 25(10): 931–945.
- [296] J. D. Durrant, S. Lindert, J. A. McCammon (2013) AutoGrow 3.0: An improved algorithm for chemically tractable, semi-automated protein inhibitor design. *J. Mol. Graph. Model.* 44: 104–112.
- [297] K. Kawai, N. Nagata, Y. Takahashi (2014) De Novo Design of Drug-Like Molecules by a Fragment-Based Molecular Evolutionary Approach. *J. Chem. Inf. Model.* 54(1): 49–56.
- [298] J. D. Durrant, J. A. McCammon (2012) AutoClickChem: Click Chemistry in Silico. *PLoS Comput. Biol.* 8(3): e1002397.
- [299] S. Lindert, J. D. Durrant, J. A. McCammon (2012) LigMerge: A Fast Algorithm to Generate Models of Novel Potential Ligands from Sets of Known Binders. *Chem. Biol. Drug Des.* 80(3): 358–365.

- [300] X. Q. Lewell, D. B. Judd, S. P. Watson, M. M. Hann (**1998**) RECAP – Retrosynthetic Combinatorial Analysis Procedure: A Powerful New Technique for Identifying Privileged Molecular Fragments with Useful Applications in Combinatorial Chemistry. *J. Chem. Inf. Comput. Sci.* 38(3): 511–522.
- [301] J.-F. Couture, E. Collazo, J. S. Brunzelle, R. C. Trievel (**2005**) Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase. *Genes Dev.* 19(12): 1455–1465.
- [302] C. Notredame, D. G. Higgins, J. Heringa (**2000**) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302(1): 205–217.
- [303] P. D. Tommaso, S. Moretti, I. Xenarios, M. Orobitg, A. Montanyola, J.-M. Chang, et al. (**2011**) T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* 39(suppl 2): W13–W17.
- [304] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, et al. (**2004**) UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* 25(13): 1605–1612.
- [305] A. Šali, T. L. Blundell (**1993**) Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* 234(3): 779–815.
- [306] N. Eswar, B. Webb, M. A. Marti-Renom, M. Madhusudhan, D. Eramian, M.-y. Shen, et al. (**2006**) Comparative Protein Structure Modeling Using Modeller. In: *Current Protocols in Bioinformatics*, volume Supplement 15, pp. 5.6.1–5.6.30, John Wiley & Sons, Inc., ISBN 9780471250951.
- [307] A. Fiser, R. K. Do, A. Sali (**2000**) Modeling of loops in protein structures. *Protein Sci.* 9(9): 1753–1773.
- [308] M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, J. H. Jensen (**2011**) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* 7(2): 525–537.
- [309] D. A. Case, T. A. Darden, T. E. I. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, et al., AMBER 12.
- [310] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling (**2006**) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* 65(3): 712–725.

- [311] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79(2): 926–935.
- [312] W. Stacklies, F. Xia, F. Gräter (2009) Dynamic Allostery in the Methionine Repressor Revealed by Force Distribution Analysis. *PLoS Comput. Biol.* 5(11): e1000574.
- [313] D. Hamelberg, J. Mongan, J. A. McCammon (2004) Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys.* 120(24): 11919–11929.
- [314] Y. Wang, C. B. Harrison, K. Schulten, J. A. McCammon (2011) Implementation of accelerated molecular dynamics in NAMD. *Comput. Sci. Discovery* 4(1): 015002.
- [315] D. Hamelberg, C. A. F. d. Oliveira, J. A. McCammon (2007) Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. *J. Chem. Phys.* 127(15): 155102.
- [316] C. A. F. de Oliveira, B. J. Grant, M. Zhou, J. A. McCammon (2011) Large-Scale Conformational Changes of Trypanosoma cruzi Proline Racemase Predicted by Accelerated Molecular Dynamics Simulation. *PLoS Comput. Biol.* 7(10): e1002178.
- [317] J. Wereszczynski, A. J. McCammon (2012) Accelerated Molecular Dynamics in Computational Drug Design. In: *Computational Drug Discovery and Design*, (ed.) R. Baron, number 819 in Methods in Molecular Biology, pp. 515–524, Humana Press, New York.
- [318] Y. Xin, U. Doshi, D. Hamelberg (2010) Examining the limits of time reweighting and Kramers’ rate theory to obtain correct kinetics from accelerated molecular dynamics. *J. Chem. Phys.* 132(22): 224101.
- [319] R. E. Amaro, R. Baron, J. A. McCammon (2008) An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J. Comput. Aided Mol. Des.* 22(9): 693–705.
- [320] R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26(2): 283–291.

- [321] E. Kellenberger, P. Muller, C. Schalon, G. Bret, N. Foata, D. Rognan (**2006**) sc-PDB: an Annotated Database of Druggable Binding Sites from the Protein Data Bank. *J. Chem. Inf. Model.* 46(2): 717–727.
- [322] N. Weill, D. Rognan (**2010**) Alignment-Free Ultra-High-Throughput Comparison of Druggable Protein–Ligand Binding Sites. *J. Chem. Inf. Model.* 50(1): 123–135.
- [323] O. Trott, A. J. Olson (**2010**) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31(2): 455–461.
- [324] ChemAxon (**2013**), Marvin 6.0.3. URL <https://www.chemaxon.com/products/marvin/>.
- [325] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case (**2004**) Development and testing of a general amber force field. *J. Comput. Chem.* 25(9): 1157–1174.
- [326] J. Wang, W. Wang, P. A. Kollman, D. A. Case (**2006**) Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* 25(2): 247–260.
- [327] C. I. Bayly, P. Cieplak, W. Cornell, P. A. Kollman (**1993**) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* 97(40): 10269–10280.
- [328] F.-Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, et al. (**2010**) The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.* 12(28): 7821–7839.
- [329] E. Vanquelef, S. Simon, G. Marquant, E. Garcia, G. Klimerak, J. C. Delepine, et al. (**2011**) R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.* 39(Suppl 2): W511–W517.
- [330] F. Neese (**2012**), ORCA, version 2.9, an ab initio, density functional theory and semiempirical program package.
- [331] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, et al., Gaussian~09 Revision D.01. Gaussian Inc. Wallingford CT 2009.

- [332] Schrödinger, LLC (**2010**), The PyMOL Molecular Graphics System, version 1.3.
- [333] R. L. Dunbrack, F. E. Cohen (**1997**) Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* 6(8): 1661–1681.
- [334] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, et al. (**2005**) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26(16): 1781–1802.
- [335] W. Humphrey, A. Dalke, K. Schulten (**1996**) VMD: Visual molecular dynamics. *J. Mol. Graph.* 14(1): 33–38.
- [336] P. I. Koukos, N. M. Glykos (**2013**) Grcarma: A fully automated task-oriented interface for the analysis of molecular dynamics trajectories. *J. Comput. Chem.* 34(26): 2310–2312.
- [337] P. Schmidtke, A. Bidon-Chanal, F. J. Luque, X. Barril (**2011**) MDpocket : Open Source Cavity Detection and Characterization on Molecular Dynamics Trajectories. *Bioinformatics* 27(23): 3276–3285.
- [338] V. L. Guilloux, P. Schmidtke, P. Tuffery (**2009**) Fpocket: An open source platform for ligand pocket detection. *BMC Bioinform.* 10(1): 168.
- [339] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, et al. (**2000**) Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* 33(12): 889–897.
- [340] B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke, A. E. Roitberg (**2012**) MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* 8(9): 3314–3321.
- [341] A. Onufriev, D. Bashford, D. A. Case (**2004**) Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Struct., Funct., Bioinf.* 55(2): 383–394.
- [342] J. Weiser, P. S. Shenkin, W. C. Still (**1999**) Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comput. Chem.* 20(2): 217–230.

- [343] I. Massova, P. A. Kollman (**1999**) Computational Alanine Scanning To Probe Protein–Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. *J. Am. Chem. Soc.* 121(36): 8133–8143.
- [344] A. Laio, M. Parrinello (**2002**) Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* 99(20): 12562–12566.
- [345] A. Barducci, G. Bussi, M. Parrinello (**2008**) Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* 100(2): 020603.
- [346] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, et al. (**2009**) PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* 180(10): 1961–1972.
- [347] Silicos-it (**2013**), Align-it, version 1.0.4. URL <http://silicos-it.be.s3-website-eu-west-1.amazonaws.com/software/align-it/1.0.4/align-it.html>.
- [348] Silicos-it (**2013**), Shape-it, version 1.0.1. URL <http://silicos-it.be.s3-website-eu-west-1.amazonaws.com/software/shape-it/1.0.1/shape-it.html>.
- [349] J. Greene, S. Kahn, H. Savoj, P. Sprague, S. Teig (**1994**) Chemical Function Queries for 3D Database Search. *J. Chem. Inf. Comput. Sci.* 34(6): 1297–1308.
- [350] J. Taminau, G. Thijs, H. De Winter (**2008**) Pharao: Pharmacophore alignment and optimization. *J. Mol. Graph. Model.* 27(2): 161–169.
- [351] J. A. Grant, M. A. Gallardo, B. T. Pickup (**1996**) A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape. *J. Comput. Chem.* 17(14): 1653–1666.
- [352] J.-P. Ebejer, G. M. Morris, C. M. Deane (**2012**) Freely Available Conformer Generation Methods: How Good Are They? *J. Chem. Inf. Model.* 52(5): 1146–1158.
- [353] P. Tosco, N. Stiefl, G. Landrum (**2014**) Bringing the MMFF force field to the RDKit: implementation and validation. *J. Cheminform.* 6(1): 37.
- [354] J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad, R. G. Coleman (**2012**) ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.* 52(7): 1757–1768.

- [355] J. L. Durant, B. A. Leland, D. R. Henry, J. G. Nourse (**2002**) Reoptimization of MDL Keys for Use in Drug Discovery. *J. Chem. Inf. Comput. Sci.* 42(6): 1273–1280.
- [356] S. Riniker, G. A. Landrum (**2013**) Open-source platform to benchmark fingerprints for ligand-based virtual screening. *J. Cheminform.* 5(1): 26.
- [357] R. Didziapetris, P. Japertas, A. Avdeef, A. Petrauskas (**2003**) Classification Analysis of P-Glycoprotein Substrate Specificity. *J. Drug Target.* 11(7): 391–406.
- [358] Y. Zhang, C. Bachmeier, D. W. Miller (**2003**) In vitro and in vivo models for assessing drug efflux transporter activity. *Adv. Drug Delivery Rev.* 55(1): 31–51.
- [359] C. Hegedűs, G. Szakács, L. Homolya, T. I. Orbán, Á. Telbisz, M. Jani, et al. (**2009**) Ins and outs of the ABCG2 multidrug transporter: An update on in vitro functional assays. *Adv. Drug Delivery Rev.* 61(1): 47–56.
- [360] C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, A. Frolkis, et al. (**2011**) DrugBank 3.0: a comprehensive resource for ‘Omics’ research on drugs. *Nucleic Acids Res.* 39(Suppl 1): D1035–D1041.
- [361] J. Larsson, J. Gottfries, S. Muresan, A. Backlund (**2007**) ChemGPS-NP: Tuned for Navigation in Biologically Relevant Chemical Space. *J. Nat. Prod.* 70(5): 789–794.
- [362] Wavefunction, Inc. (**2002**), Spartan ’02 for Linux. URL <http://www.wavefun.com/products/spartan.html>.
- [363] Simulations Plus Inc. (**2011**), ADMET Predictor.
- [364] T. Kohonen (**2001**) *Self-Organizing Maps*. Springer Science & Business Media.
- [365] S. Lloyd (**1982**) Least squares quantization in PCM. *IEEE Trans. Inform. Theory* 28(2): 129–137.
- [366] P. Matsson, G. Englund, G. Ahlin, C. A. S. Bergström, U. Norinder, P. Artursson (**2007**) A Global Drug Inhibition Pattern for the Human ATP-Binding Cassette Transporter Breast Cancer Resistance Protein (ABCG2). *J. Pharmacol. Exp. Ther.* 323(1): 19–30.

- [367] E. Furusjö, A. Svenson, M. Rahmberg, M. Andersson (2006) The importance of outlier detection and training set selection for reliable environmental QSAR predictions. *Chemosphere* 63(1): 99–108.
- [368] C.-C. Chang, C.-J. Lin (2011) LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.* 2(3): 27:1–27:27.
- [369] J. Levatić, J. Ćurak, M. Kralj, T. Šmuc, M. Osmak, F. Supek (2013) Accurate Models for P-gp Drug Recognition Induced from a Cancer Cell Line Cytotoxicity Screen. *J. Med. Chem.* 56(14): 5691–5708.
- [370] L. Zhong, C.-Y. Ma, H. Zhang, L.-J. Yang, H.-L. Wan, Q.-Q. Xie, et al. (2011) A prediction model of substrates and non-substrates of breast cancer resistance protein (BCRP) developed by GA-CG-SVM method. *Comput. Biol. Med.* 41(11): 1006–1013.
- [371] Silicos-it (2013), Filter-it, version 1.0.2. URL <http://silicos-it.be.s3-website-eu-west-1.amazonaws.com/software/filter-it/1.0.2/filter-it.html>.
- [372] H. L. Morgan (1965) The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* 5(2): 107–113.
- [373] D. Rogers, M. Hahn (2010) Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* 50(5): 742–754.
- [374] T. K. Allu, T. I. Oprea (2005) Rapid Evaluation of Synthetic and Molecular Complexity for in Silico Chemistry. *J. Chem. Inf. Model.* 45(5): 1237–1243.
- [375] G. M. Rishton (2003) Nonleadlikeness and leadlikeness in biochemical screening. *Drug Discovery Today* 8(2): 86–96.
- [376] J. B. Baell, G. A. Holloway (2010) New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *J. Med. Chem.* 53(7): 2719–2740.
- [377] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, et al. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Bioinf.* 78(8): 1950–1958.

- [378] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, et al. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29(7): 845–854.
- [379] T. Hou, J. Wang, Y. Li, W. Wang (2011) Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* 51(1): 69–82.
- [380] L. M. Salonen, M. Ellermann, F. Diederich (2011) Aromatic Rings in Chemical and Biological Recognition: Energetics and Structures. *Angew. Chem. Int. Ed.* 50(21): 4808–4842.
- [381] C. S. Leung, S. S. F. Leung, J. Tirado-Rives, W. L. Jorgensen (2012) Methyl Effects on Protein–Ligand Binding. *J. Med. Chem.* 55(9): 4489–4500.
- [382] E. Perola, P. S. Charifson (2004) Conformational Analysis of Drug-Like Molecules Bound to Proteins: An Extensive Study of Ligand Reorganization upon Binding. *J. Med. Chem.* 47(10): 2499–2510.
- [383] M.-H. Hao, O. Haq, I. Muegge (2007) Torsion Angle Preference and Energetics of Small-Molecule Ligands Bound to Proteins. *J. Chem. Inf. Model.* 47(6): 2242–2252.
- [384] G. S. Van Aller, M. B. Pappalardi, H. M. Ott, E. Diaz, M. Brandt, B. J. Schwartz, et al. (2014) Long Residence Time Inhibition of EZH2 in Activated Polycomb Repressive Complex 2. *ACS Chem. Biol.* 9(3): 622–629.
- [385] V. Gibaja, F. Shen, J. Harari, J. Korn, D. Ruddy, V. Saenz-Vash, et al. (2015) Development of secondary mutations in wild-type and mutant EZH2 alleles cooperates to confer resistance to EZH2 inhibitors. *Oncogene* URL <http://dx.doi.org/10.1038/onc.2015.114>.
- [386] J. Kirchmair, S. Distinto, P. Markt, D. Schuster, G. M. Spitzer, K. R. Liedl, et al. (2009) How To Optimize Shape-Based Virtual Screening: Choosing the Right Query and Including Chemical Information. *J. Chem. Inf. Model.* 49(3): 678–692.
- [387] G. B. McGaughey, R. P. Sheridan, C. I. Bayly, J. C. Culberson, C. Kreatsoulas, S. Lindsley, et al. (2007) Comparison of Topological, Shape, and Docking Methods in Virtual Screening. *J. Chem. Inf. Model.* 47(4): 1504–1519.

- [388] S.-Y. Yang (2010) Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discovery Today* 15(11–12): 444–450.
- [389] M. Hennessy, J. P. Spiers (2007) A primer on the mechanics of P-glycoprotein the multidrug transporter. *Pharmacol. Res.* 55(1): 1–15.
- [390] F. J. Sharom (2007) ABC multidrug transporters: structure, function and role in chemoresistance. *Pharmacogenomics* 9(1): 105–127.
- [391] K. P. Locher (2009) Structure and mechanism of ATP-binding cassette transporters. *Phil. Trans. R. Soc. B: Biol. Sci.* 364(1514): 239–245.
- [392] F. Frézard, E. Pereira-Maia, P. Quidu, W. Priebe, A. Garnier-Suillerot (2001) P-Glycoprotein preferentially effluxes anthracyclines containing free basic versus charged amine. *Eur. J. Biochem.* 268(6): 1561–1567.
- [393] A. Seelig (1998) A general pattern for substrate recognition by P-glycoprotein. *Eur. J. Biochem.* 251(1-2): 252–261.
- [394] D. A. P. Gutmann, A. Ward, I. L. Urbatsch, G. Chang, H. W. van Veen (2010) Understanding polyspecificity of multidrug ABC transporters: closing in on the gaps in ABCB1. *Trends Biochem. Sci.* 35(1): 36–42.
- [395] D. C. Rees, E. Johnson, O. Lewinson (2009) ABC transporters: the power to change. *Nat. Rev. Mol. Cell Biol.* 10(3): 218–227.
- [396] C. Lemos, G. Jansen, G. J. Peters (2008) Drug transporters: recent advances concerning BCRP and tyrosine kinase inhibitors. *Br. J. Cancer* 98(5): 857–862.
- [397] Z. Ni, Z. Bikadi, M. F. Rosenberg, Q. Mao (2010) Structure and Function of the Human Breast Cancer Resistance Protein (BCRP/ABCG2). *Curr. Drug Metab.* 11(7): 603–617.
- [398] F. Broccatelli (2012) QSAR Models for P-Glycoprotein Transport Based on a Highly Consistent Data Set. *J. Chem. Inf. Model.* 52(9): 2462–2470.
- [399] T. I. Netzeva, A. G. Saliner, A. P. Worth (2006) Comparison of the applicability domain of a quantitative structure-activity relationship for estrogenicity with a large chemical inventory. *Environ. Toxicol. Chem.* 25(5): 1223–1230.

- [400] J. Jaworska, N. Nikolova-Jeliazkova, T. Aldenberg (2005) QSAR applicability domain estimation by projection of the training set descriptor space: a review. *ATLA, Altern. Lab. Anim.* 33(5): 445–459.
- [401] R. P. Sheridan, B. P. Feuston, V. N. Maiorov, S. K. Kearsley (2004) Similarity to Molecules in the Training Set Is a Good Discriminator for Prediction Accuracy in QSAR. *J. Chem. Inf. Comput. Sci.* 44(6): 1912–1928.
- [402] F. Sahigara, K. Mansouri, D. Ballabio, A. Mauri, V. Consonni, R. Todeschini (2012) Comparison of Different Approaches to Define the Applicability Domain of QSAR Models. *Molecules* 17(5): 4791–4810.
- [403] S. Erić, M. Kalinić (2015) Računarski modeli za predviđanje transporta lekova posredovanog P-glikoproteinom. *Arh. Farm.* 65(2): 89–114.
- [404] T. J. Raub (2006) P-Glycoprotein Recognition of Substrates and Circumvention through Rational Drug Design. *Mol. Pharm.* 3(1): 3–25.
- [405] P. C. D. Hawkins, A. G. Skillman, A. Nicholls (2007) Comparison of Shape-Matching and Docking as Virtual Screening Tools. *J. Med. Chem.* 50(1): 74–82.
- [406] G. Hu, G. Kuang, W. Xiao, W. Li, G. Liu, Y. Tang (2012) Performance Evaluation of 2D Fingerprint and 3D Shape Similarity Methods in Virtual Screening. *J. Chem. Inf. Model.* 52(5): 1103–1113.
- [407] H. W. Kuhn (1955) The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly* 2(1-2): 83–97.
- [408] J. Bajorath (2013) Molecular crime scene investigation – dusting for fingerprints. *Drug Discovery Today: Technol.* 10(4): e491–e498.
- [409] J. W. Kruisselbrink, A. Aleman, M. T. Emmerich, A. P. IJzerman, A. Bender, T. Baeck, et al. (2009) Enhancing Search Space Diversity in Multi-objective Evolutionary Drug Molecule Design Using Niching. In: *Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation, GECCO '09*, pp. 217–224, ACM, New York, NY, USA, ISBN 978-1-60558-325-9.
- [410] M. Laumanns, L. Thiele, K. Deb, E. Zitzler (2002) Combining convergence and diversity in evolutionary multiobjective optimization. *Evol. Comput.* 10(3): 263–282.

- [411] R. P. Sheridan, N. Zorn, E. C. Sherer, L.-C. Campeau, C. Z. Chang, J. Cumming, et al. (2014) Modeling a Crowdsourced Definition of Molecular Complexity. *J. Chem. Inf. Model.* 54(6): 1604–1616.
- [412] G. Rastelli, G. Degliesposti, A. Del Rio, M. Sgobba (2009) Binding Estimation after Refinement, a New Automated Procedure for the Refinement and Rescoring of Docked Ligands in Virtual Screening. *Chem. Biol. Drug Des.* 73(3): 283–286.
- [413] R. Cao, N. Huang, Y. Wang (2014) Evaluation and Application of MD-PB/SA in Structure-Based Hierarchical Virtual Screening. *J. Chem. Inf. Model.* 54(7): 1987–1996.

Prilog A

Rezultati simulacija kompleksa EZH2 i odabranih piridonskih inhibitora

Tabela A.1: Rezultati dokinga 13 piridonskih inhibitora u razvijeni model EZH2

Inhibitor	Najniža Vina energija	Vina energija odabrane poze	RMSD odabrane od poze za najnižom energijom (Å)
G1K	-7,8	-7,4	1,255
G2K	-8,2	-8,2	5,687
G3K	-9,2	-8,9	1,932
G4K	-9,0	-8,8	1,543
G5K	-9,3	-9,3	0,000
G6K	-9,0	-9,0	0,000
126	-9,2	-8,6	2,379
EI1	-7,8	-7,2	2,378
E1Z	-9,2	-9,2	0,000
E2Z	-9,2	-9,2	0,000
E3Z	-9,1	-8,9	1,674
E4Z	-8,5	-8,2	6,976
E5Z	-8,2	-8,2	2,227

Tabela A.2: Rezultati MM/GBSA proračuna sprovedenih na 500 kadrova produkcionih trajektorija trajanja 4 ns. Vrednosti u tabeli date su u kcal mol⁻¹ kao srednje vrednosti uz standardne greške.

Inhibitor	ΔE_{vdW}	ΔE_{ele}	ΔG_{GB}	ΔG_{SA}	ΔE_{MM}	ΔG_{sol}	$\Delta G_{MM/GBSA}$	ΔG_{exp}^a
G1K	-46,17 ± 0,12	-13,73 ± 0,12	27,60 ± 0,12	-4,86 ± 0,01	-59,90 ± 0,17	22,74 ± 0,12	-37,16 ± 0,12	-9,31
G2K	-48,49 ± 0,13	-13,47 ± 0,14	27,92 ± 0,12	-5,15 ± 0,01	-61,96 ± 0,19	22,78 ± 0,12	-39,18 ± 0,13	-9,72
G3K	-61,84 ± 0,16	-13,75 ± 0,16	36,04 ± 0,15	-6,41 ± 0,02	-75,60 ± 0,25	29,62 ± 0,14	-45,97 ± 0,16	-11,05
G4K	-61,50 ± 0,24	-10,04 ± 0,26	30,67 ± 0,19	-6,69 ± 0,02	-71,53 ± 0,26	23,98 ± 0,19	-47,56 ± 0,21	-12,57
G5K	-61,80 ± 0,18	-12,89 ± 0,33	35,52 ± 0,30	-6,44 ± 0,02	-74,70 ± 0,45	29,07 ± 0,29	-45,62 ± 0,20	-10,71
G6K	-64,22 ± 0,20	-8,33 ± 0,27	29,92 ± 0,27	-6,85 ± 0,02	-72,54 ± 0,36	23,08 ± 0,26	-49,47 ± 0,20	-12,16
126	-57,93 ± 0,16	-31,32 ± 0,40	53,18 ± 0,37	-6,49 ± 0,01	-89,25 ± 0,41	46,68 ± 0,37	-42,57 ± 0,14	-12,60
EL1	-46,74 ± 0,14	-30,78 ± 0,18	41,18 ± 0,16	-5,44 ± 0,01	-77,52 ± 0,21	35,74 ± 0,16	-41,78 ± 0,13	-10,75
E1Z	-53,06 ± 0,15	-17,56 ± 0,14	34,03 ± 0,12	-5,37 ± 0,01	-70,62 ± 0,19	28,66 ± 0,12	-41,96 ± 0,15	-8,87
E2Z	-53,03 ± 0,14	-18,22 ± 0,20	37,04 ± 0,18	-5,62 ± 0,01	-71,26 ± 0,26	31,42 ± 0,18	-39,83 ± 0,14	-9,20
E3Z	-53,71 ± 0,14	-14,60 ± 0,15	34,79 ± 0,14	-5,79 ± 0,01	-68,31 ± 0,21	29,00 ± 0,13	-39,31 ± 0,14	-9,68
E4Z	-56,35 ± 0,23	-12,90 ± 0,17	34,58 ± 0,20	-6,20 ± 0,02	-69,25 ± 0,32	28,37 ± 0,19	-40,88 ± 0,17	-10,39
E5Z	-55,58 ± 0,18	-16,07 ± 0,27	34,88 ± 0,21	-6,54 ± 0,02	-71,65 ± 0,33	28,34 ± 0,21	-43,31 ± 0,18	-11,73

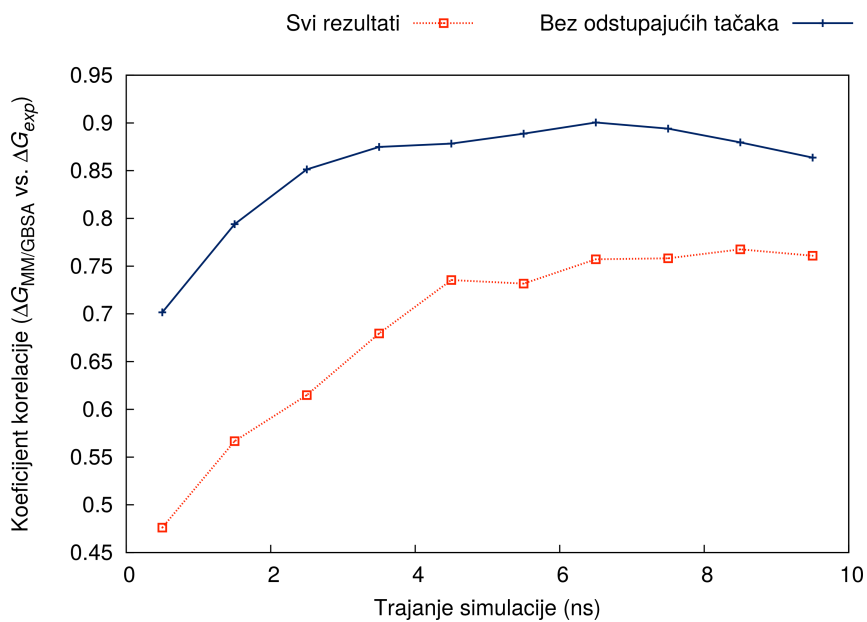
^a Slobodne energije vezivanja izračunate su iz K_i vrednosti korišćenjem odnosa $\Delta G^0 = RT \ln K_i$, uz $T = 298$ K što odgovara temperaturi na kojoj su eksperimenti izvođeni.

Tabela A.3: Rezultati MM/GBSA proračuna na stupnjevito dužim delovima dodatne produkcijske trajektorije. Vrednosti u tabeli date su u kcal mol⁻¹ kao srednje vrednosti uz standardne greške.

Inhibitor	0–1 ns	0–2 ns	0–3 ns	0–4 ns	0–5 ns	0–6 ns	0–7 ns	0–8 ns	0–9 ns	0–10 ns
G1K	-35.54 ± 0.31	-36.45 ± 0.21	-36.98 ± 0.17	-36.95 ± 0.15	-36.83 ± 0.14	-36.86 ± 0.12	-36.85 ± 0.11	-36.76 ± 0.10	-36.52 ± 0.10	-36.10 ± 0.11
G2K	-37.32 ± 0.26	-37.21 ± 0.18	-37.56 ± 0.15	-37.39 ± 0.13	-37.70 ± 0.12	-37.96 ± 0.11	-38.16 ± 0.10	-38.28 ± 0.10	-38.36 ± 0.10	-38.41 ± 0.09
G3K	-48.05 ± 0.30	-47.75 ± 0.22	-46.95 ± 0.18	-46.31 ± 0.17	-45.95 ± 0.15	-45.48 ± 0.14	-45.61 ± 0.13	-45.24 ± 0.12	-44.54 ± 0.12	-44.81 ± 0.12
G4K	-46.00 ± 0.38	-45.97 ± 0.25	-46.96 ± 0.21	-46.95 ± 0.18	-46.75 ± 0.17	-46.17 ± 0.16	-45.69 ± 0.15	-44.96 ± 0.15	-44.42 ± 0.15	-43.99 ± 0.15
G5K	-49.73 ± 0.30	-47.96 ± 0.25	-47.68 ± 0.19	-47.40 ± 0.17	-47.24 ± 0.14	-46.19 ± 0.17	-45.19 ± 0.18	-44.70 ± 0.17	-44.64 ± 0.16	-44.47 ± 0.15
G6K	-42.51 ± 0.35	-45.41 ± 0.30	-46.89 ± 0.26	-47.37 ± 0.22	-47.63 ± 0.18	-47.77 ± 0.16	-47.99 ± 0.15	-48.22 ± 0.13	-48.46 ± 0.13	-48.71 ± 0.12
126	-43.00 ± 0.27	-43.15 ± 0.22	-42.39 ± 0.19	-42.85 ± 0.17	-43.26 ± 0.16	-42.55 ± 0.16	-42.42 ± 0.14	-42.35 ± 0.14	-42.66 ± 0.13	-42.73 ± 0.12
EL1	-42.87 ± 0.26	-41.62 ± 0.20	-41.11 ± 0.16	-41.33 ± 0.14	-41.37 ± 0.12	-41.48 ± 0.11	-41.30 ± 0.10	-41.20 ± 0.10	-41.09 ± 0.10	-41.09 ± 0.09
E1Z	-46.25 ± 0.34	-45.78 ± 0.25	-44.92 ± 0.20	-43.80 ± 0.21	-43.32 ± 0.22	-42.09 ± 0.19	-41.43 ± 0.18	-41.16 ± 0.17	-40.76 ± 0.16	-40.57 ± 0.15
E2Z	-39.10 ± 0.31	-37.76 ± 0.22	-37.45 ± 0.20	-37.12 ± 0.17	-37.38 ± 0.16	-37.63 ± 0.14	-37.88 ± 0.13	-38.08 ± 0.12	-38.12 ± 0.11	-38.32 ± 0.11
E3Z	-38.91 ± 0.35	-38.96 ± 0.24	-39.81 ± 0.21	-39.69 ± 0.17	-39.12 ± 0.16	-38.95 ± 0.15	-38.89 ± 0.13	-38.77 ± 0.12	-39.21 ± 0.12	-39.32 ± 0.12
E4Z	-40.02 ± 0.30	-38.00 ± 0.25	-38.52 ± 0.21	-38.54 ± 0.18	-38.11 ± 0.16	-38.06 ± 0.14	-38.47 ± 0.14	-38.64 ± 0.13	-38.48 ± 0.12	-38.51 ± 0.11
E5Z	-50.30 ± 0.41	-50.57 ± 0.25	-49.39 ± 0.22	-48.29 ± 0.21	-47.79 ± 0.20	-46.83 ± 0.20	-46.57 ± 0.18	-46.79 ± 0.16	-46.91 ± 0.15	-47.01 ± 0.14

Tabela A.4: Rezultati analize efekata mutageneze u alanin, prikazani kao srednje vrednosti razlike MM/GBSA energija ($\Delta\Delta G_{MM/GBSA}$ u kcal mol⁻¹) između kanonske i mutirane forme enzima, uz standardne greške.

Inhibitor	V621A	W624A	Y661A	C663A	F665A	N688A	S690A	F723A	Y726A	Y728A	I739A	E742A	M743A
G1K	1.62 ± 0.17	0.07 ± 0.17	0.11 ± 0.17	1.31 ± 0.17	1.70 ± 0.17	1.68 ± 0.17	0.19 ± 0.17	1.37 ± 0.17	2.71 ± 0.17	0.45 ± 0.17	2.17 ± 0.17	0.15 ± 0.17	1.50 ± 0.17
G2K	0.97 ± 0.18	0.05 ± 0.19	0.78 ± 0.19	1.48 ± 0.19	2.14 ± 0.19	1.25 ± 0.19	0.09 ± 0.19	0.43 ± 0.19	2.44 ± 0.19	0.77 ± 0.19	2.40 ± 0.19	0.24 ± 0.19	0.77 ± 0.19
G3K	1.20 ± 0.22	0.79 ± 0.22	3.25 ± 0.22	1.25 ± 0.22	2.31 ± 0.22	0.98 ± 0.22	0.07 ± 0.22	0.40 ± 0.22	2.60 ± 0.22	0.25 ± 0.22	2.16 ± 0.21	0.40 ± 0.22	1.07 ± 0.22
G4K	0.21 ± 0.29	1.43 ± 0.29	2.79 ± 0.31	1.07 ± 0.29	2.26 ± 0.29	4.37 ± 0.28	0.08 ± 0.29	0.29 ± 0.29	2.22 ± 0.29	1.26 ± 0.29	0.22 ± 0.29	0.00 ± 0.29	1.51 ± 0.29
G5K	1.42 ± 0.23	0.78 ± 0.23	3.63 ± 0.23	1.60 ± 0.23	2.72 ± 0.23	1.37 ± 0.23	0.07 ± 0.23	0.68 ± 0.23	2.74 ± 0.23	0.29 ± 0.23	2.25 ± 0.23	0.33 ± 0.22	1.07 ± 0.23
G6K	0.58 ± 0.27	0.80 ± 0.27	2.89 ± 0.28	1.36 ± 0.28	2.88 ± 0.27	4.09 ± 0.26	0.11 ± 0.28	0.26 ± 0.28	2.74 ± 0.27	1.86 ± 0.28	1.05 ± 0.27	0.17 ± 0.27	1.28 ± 0.28
I26	1.12 ± 0.20	0.80 ± 0.21	2.93 ± 0.20	1.10 ± 0.20	2.81 ± 0.20	4.14 ± 0.20	0.08 ± 0.20	0.25 ± 0.20	3.07 ± 0.20	0.87 ± 0.20	1.27 ± 0.21	0.19 ± 0.20	0.91 ± 0.20
EL1	0.50 ± 0.18	0.03 ± 0.18	0.96 ± 0.18	0.55 ± 0.18	1.35 ± 0.19	0.97 ± 0.18	0.04 ± 0.18	0.51 ± 0.18	3.94 ± 0.18	1.23 ± 0.18	0.89 ± 0.18	0.11 ± 0.18	3.80 ± 0.18
E1Z	1.62 ± 0.23	1.38 ± 0.21	1.97 ± 0.21	1.14 ± 0.22	2.56 ± 0.22	0.15 ± 0.22	0.06 ± 0.22	0.31 ± 0.22	2.27 ± 0.23	2.74 ± 0.22	2.33 ± 0.22	0.14 ± 0.22	1.85 ± 0.22
E2Z	1.18 ± 0.19	0.12 ± 0.20	2.25 ± 0.20	1.40 ± 0.19	3.09 ± 0.19	1.69 ± 0.19	0.08 ± 0.20	0.51 ± 0.20	3.04 ± 0.20	0.42 ± 0.20	1.16 ± 0.20	0.19 ± 0.20	1.11 ± 0.19
E3Z	1.00 ± 0.19	0.07 ± 0.19	2.22 ± 0.19	1.73 ± 0.19	3.36 ± 0.19	1.15 ± 0.19	0.07 ± 0.19	0.30 ± 0.19	2.65 ± 0.19	0.31 ± 0.19	1.31 ± 0.19	0.28 ± 0.20	0.88 ± 0.19
E4Z	2.07 ± 0.24	0.24 ± 0.25	2.44 ± 0.24	1.33 ± 0.24	2.37 ± 0.33	0.63 ± 0.25	0.03 ± 0.25	0.34 ± 0.25	3.31 ± 0.25	0.52 ± 0.24	1.57 ± 0.25	0.51 ± 0.24	0.73 ± 0.25
E5Z	0.64 ± 0.25	0.55 ± 0.26	0.09 ± 0.26	1.32 ± 0.25	2.47 ± 0.25	0.74 ± 0.26	0.07 ± 0.26	0.65 ± 0.25	3.41 ± 0.25	3.87 ± 0.25	1.37 ± 0.25	0.44 ± 0.25	1.16 ± 0.26



Slika A.1: Korelacija između $\Delta G_{MM/GBSA}$ i ΔG_{exp} u zavisnosti od dužine evaluirane trajektorije. Nakon 4–5 ns simulacije, primetna je konvergencija rezultata. Ispunjena plava linija odgovara korelaciji bez rezultata za **E1Z** i **126**, dok isprekidana crvena linija odgovara korelaciji za sve rezultate zajedno.

Prilog B

Modeli za predviđanje transporta posredovanog Pgp-om i BCRP-om

Tabela B.1: Set podataka korišćen u razvoju modela za predviđanje transporta posredovanog Pgp-om.

Jedinjenje	InChI ključ	Klasa	PMID referenci
2-fluoro-A-85380	GVOOYVOZPRROMP-ZETCQYMHSA-N	NS	21321274
5-fluorouracil	GHASVSINZRGABV-UHFFFAOYSA-N	NS	12081145
5-hydroxymethyl tol- terodine	DUXZAXCGJSBGDW-HXUWFJFHSA-N	S	21392072
9-hydroxyrisperidone	PMXMIIMHBWHSKN-LJQANCHMSA-N	S	15502009, 16810505, 15683552, 16936711, 16118767
abacavir	MCGSCOLBFJQGHM-SCZZXKLOSA-N	S	17709369
acrivastine	PWACSDKDOHSSQD-IUTFFREVSA-N	S	12438524
actinomycin D	RJURFGZVJUQBHK-IIXSONLDSA-N	S	11602674, 11707575
adefovir	SUPKOOSCJHTBAH-UHFFFAOYSA-N	NS	14762102, 9804626, 11602668
albendazole	HXHWSAZORRCQMX-UHFFFAOYSA-N	NS	11901088, 15703302
albendazole sulfoxide	VXTGHWHFYNYFFV-LJQANCHMSA-N	NS	15703302
aldosterone	PQSUYGKTWSAVDQ-ZVIOFETBSA-N	S	12213137, 1360010, 10837354
alfentanil	IDBPHNDTYPBSNI-UHFFFAOYSA-N	NS	11964599
alfuzosin	WNMJYKCGWZFFKR-AWEZLNQCLSA-N	S	19273529
alpha-methyldigoxin	IYJMSDVSVDVGT-AVUZACQCSA-N	S	11284449
alprenolol	PAZJSJFMUHDSTF-AWEZLNQCLSA-N	NS	12438524
amantadine	DKNWSYNQZKUICI-CHIWXEEVSA-N	NS	11602674, 12438524, 15804173
amiodarone	IYIKLHRQXLHMJQ-UHFFFAOYSA-N	S	11231118, 15502009
amisulpride	NTJOBXMMWNYJFB-GFCCVEGCSEA-N	S	16252067, 15285840, 12865899, 17683917

Jedinjenje	InChI ključ	Klasa	PMID referenci
amitriptyline	KRMDCWKBEZIMAB-UHFFFAOYSA-N	NS	12438524
amonafide	UPALIKSFLSVKIS-UHFFFAOYSA-N	NS	17826829
amprenavir	YMARZQAQMVYCKC-OEMFJLHTSA-N	S	11602674, 12438524, 11242145
antipyrine	VEQOALNAAJBPNY-UHFFFAOYSA-N	NS	12438524, 15804173
astemizole	GXDALQBWZGODGZ-UHFFFAOYSA-N	S	12438524
atenolol	METKIMKYRPQLGS-LBPRGKRZSA-N	NS	12438524
atorvastatin	XUKUURHRXDUEBC-KAYWLYCHSA-N	S	11474784, 21447733, 10751037
axitinib	RITAVMQDGBJQJZ-FMIVXFBMSA-N	S	21282407
befloxatone	IALVDLPLCLFBCF-CHWSQXEVSA-N	NS	21321274
belotecan	LNHWXBUNXOXMRL-VWLOTQADSA-N	S	18654741
berberine	YBHILYKTIRIUTE-UHFFFAOYSA-N	S	17213652
beta-acetyldigoxin	NREAGDHHMSOWKZ-OOVFFRPTSA-N	S	11284449
bezafibrate	IIBYAHWJQTYFKB-UHFFFAOYSA-N	NS	14698041
bimatoprost	AQOKCDNYWBIDND-FTOWTWDKSA-N	S	20028257
bimatoprost acid	YFHHIZGZVLHBQZ-IIVAHVFHSA-N	NS	20028257
biperiden	YSXKPIUOCJLQIE-YXYSEUPNSA-N	NS	12438524
bisantrene	NJSMWLQOCQIOPE-OCHFTUDZSA-N	S	11497241
bufuralol	SSEBTPPFLCUMN-ZDUSSCGKSA-N	NS	12438524
bunitrolol	VCVQSRCYSKKPBA-LBPRGKRZSA-N	S	10206323
buprenorphine	RMRJXGBAOAMLHD-IHFGGWKQSA-N	NS	19887017, 19370547
buspirone	QWCRAEMEVGPNT-UHFFFAOYSA-N	NS	17962372, 12438524
caffeine	RYYVLZVUVIJVGH-UHFFFAOYSA-N	NS	16686367, 17962372

Jedinjenje	InChI ključ	Klasa	PMID referenci
carbamazepine	FFGPTBGBLEHEPO-UHFFFAOYSA-N	NS	18824002, 11318771, 12438524, 12438524, 20080116, 17962372
carisoprodol	OFZCIYFFPZCNJE-GFCCVEGCSA-N	NS	17962372
carvedilol	OGHNVEJMJSYVRP-KRWDZBQOSA-N	S	16671962, 17910813, 15951832
catharanthine	CMKFQVZJOWHHDV-DYHNYNMBSA-N	S	16686367
cediranib	XXJWYDDUDKYVKI-UHFFFAOYSA-N	S	22323823
celiprolol	JOATXPAWOHTVSZ-MRXNPFEDSA-N	S	19067419, 7905337
cerivastatin	SEERZIQQUAZTOL-ANMDKAQQSA-N	S	15901800
cetirizine	ZKLPARSLTMPFCP-OAQYLSRUSA-N	S	12438524, 17180728, 14502547, 12584158
CGP 12177	UMQUQWCJKFOUGV-UHFFFAOYSA-N	NS	21321274
chloroquine	WHTVZRBIWZFKQO-CQSZACIVSA-N	S	11602674
chlorothiazide	JBMKAUGHUNFTOL-UHFFFAOYSA-N	NS	22123130
chlorpheniramine	SOYKEARSMXGVTM-OAHLLOKOSA-N	NS	11602674, 12438524
chlorpromazine	ZPEIMTDSQAKGNT-UHFFFAOYSA-N	NS	11602674
chlorprothixene	WSPOMRSOLSGNFJ-AUWJEWJLSA-N	NS	12438524
cimetidine	AQIXAKUUQRKLND-UHFFFAOYSA-N	S	11602674, 12438524, 9862768
citalopram	WSEQXVZVJXJVFP-FQEVSTJZSA-N	NS	17962372
clarithromycin	AGOYDEPGAOXOCK-LIIZIJDPSA-N	S	11602674, 20724802, 20621639
clomipramine	GDLIGKIOYRNHDA-UHFFFAOYSA-N	NS	12438524
clonidine	GJSURZIOUXUGAL-UHFFFAOYSA-N	NS	12438524

Jedinjenje	InChI ključ	Klasa	PMID referenci
clorgyline	BTFHLQRNAMSRLC-UHFFFAOYSA-N	NS	21321274
clozapine	QZUDBNBUXVUHMW-UHFFFAOYSA-N	NS	10831014, 8647944, 16810505, 15961125
colchicine	IAKHMKGGTNLKSZ-INIZCTEOSA-N	S	11602674, 11284688
corticosterone	OMFXVFTZEKFBZ-HJTSIMOOSA-N	NS	11356720, 18556350
cortisol	JYGXADMDFJGBT-VWUMJDOOSA-N	S	11966513, 12213137, 10586385
cyclobenzaprime	JURKNVYFZMSNLP-UHFFFAOYSA-N	NS	12438524, 17962372
cyclosporin A	PMATZTZNYRCHOR-CGLBZJNRSA-N	S	11602674, 7560060, 11966513, 7681059
danofloxacin	QMLVECGLEOSESV-RYUDHWBXSAN	S	17211460
darifenacin	HXGBXQDTNZMWGS-RUZDIDTESAN	S	21826715, 21392072
dasatinib	ZBNZXTGUTAYRHI-UHFFFAOYSA-N	S	19276246, 19491323
daunorubicin	STQGQHZAVUOBTE-VGBVRHCVSAN	S	11602674, 7575681
debrisoquine	JWPGJSVJDAJRLW-UHFFFAOYSA-N	S	10213372
desipramine	HCYAFALTSJYZDH-UHFFFAOYSA-N	NS	12438524
desloratadine	NINPQHBCPTUOFE-UHFFFAOYSA-N	S	12584158
dexamethasone	UREBDLICKHMUKA-CXSFZGCWSAN	S	11602674, 7560060
diazepam	AAOVKJBEBIDNHE-UHFFFAOYSA-N	NS	17962372, 17916018
dibucaine	PUFQVTATUTYEAL-UHFFFAOYSA-N	S	16686367, 7913604
diclofenac	DCOPUUMXTXDBNB-UHFFFAOYSA-N	NS	18845662
digitoxin	WDJUZGPOPHTGOT-QLHMFLTHSAN	S	11284449
digoxin	LTMHDMANZUZIPE-ALVQDVFTSAN	S	11181899, 7560060, 11284449

Jedinjenje	InChI ključ	Klasa	PMID referenci
diltiazem	HSUGRBWQSSZJOP-RTWAWAEBSA-N	S	11602674, 12438524
diphenhydramine	ZZVUWRFHKOJYTH-UHFFFAOYSA-N	NS	12584158, 10762610
diprenorphine	OIJXLIIMXHRJJH-KNLIKEYSA-N	S	21321274
dipyridamole	IZEKFCXSFNUWAM-UHFFFAOYSA-N	S	11602674
docetaxel	ROQZUJKOMJECOM-CJQOFVABSA-N	S	7945455, 19920203, 21351274
domperidone	FGXWKSZQVQUSTL-UHFFFAOYSA-N	S	12438524, 8647944, 17683917
doxapram	XFDJYSQDBULQSI-JOCHJYFZSA-N	NS	12438524
doxepin	ODQWQRRAPPTVAG-GZTJUZNOSA-N	NS	12438524
doxorubicin	AOJJSUZBOXZQNB-TZSSRYMLSA-N	NS	11602674
doxylamine	HCFDWZZGGLSKEP-KRWDZBQOSA-N	NS	12438524
DPA-714	FLZZFWBNYJNHMY-UHFFFAOYSA-N	NS	21321274
efavirenz	XPOQHMRABVBWPR-ZDUSSCGKSA-N	NS	19748977, 16779703
eletriptan	PWVXXGRKLHYWKM-LJQANCHMSA-N	S	11602674, 12438524
elvitegravir	JUZYLCPPVHEVSV-LJQANCHMSA-N	S	21393174
emetine	AUVVAXYIELKVAI-CKBKHPWSA-N	S	11602674
endoxifen	MHJBZVSGOZTKRH-IZHYLOQSSA-N	S	21148080, 21378205
epirubicin	AOJJSUZBOXZQNB-VTZDEGQISA-N	S	7575681
erlotinib	AAKJLRGGTJKAMG-UHFFFAOYSA-N	S	18723475
erythromycin	ULGZDMOVFRHVEP-RMEHMVEFSA-N	S	11602674, 10213372
estriol	PROQIPRRNZUXQM-ZXXIGWHRSA-N	S	15290871
estrone	DNXHEGUUPJUMQT-CBZIJGRNSA-N	S	15290871
ethosuximide	HAPOVYFOVVWLRZ-ZETCQYMNSA-N	NS	20417647

Jedinjenje	InChI ključ	Klasa	PMID referenci
ethynylestradiol	BFPYWIDHMRZLRN-SLHNCBLASA-N	S	15290871
etoposide	VJJPUSNTGOMMGY-LCBFZJDDSA-N	S	11602674, 8968083
etravirine	PYGWGZALEOIKDF-UHFFFAOYSA-N	NS	21189339
EXP3174	ZEUXAIYYDDCIRX-UHFFFAOYSA-N	NS	10725273
fallypride	OABRYNHZQBZDMG-INIZCTEOSA-N	NS	21321274
famciclovir	GGXKVVWZWMLJEH-UHFFFAOYSA-N	S	12438524
felodipine	RZTAMFZIAATZDJ-OAHLLOKOSA-N	NS	11861813, 10227700
fenbendazole	HDDSHPAODJUKPD-UHFFFAOYSA-N	NS	15703302
fenofibrate	YMTINGFKWWXKFG-UHFFFAOYSA-N	NS	14698041
fenofibric acid	MQOBSOSZFYZQOK-UHFFFAOYSA-N	NS	14698041
fentanyl	PJMPHNIQZUBGLI-UHFFFAOYSA-N	NS	11964599
fexofenadine	RWTNPBWLIMQHL-LJAQVGFWSA-N	S	17180728
flavopiridol	BIIVYFLTOXDAOV-YVEFUNNKSA-N	S	20304939, 19225039, 11355953
fluconazole	RFHAOTPXVQNOHP-UHFFFAOYSA-N	NS	12235267, 11751127
flumazenil	OFBIFZUFASYRE-UHFFFAOYSA-N	NS	12438524, 21321274, 15804173
flunitrazepam	YGLJJRRNHBOWJN-UHFFFAOYSA-N	NS	10831014, 8647944
fluoxetine	RTHCYVBBDHJXIQ-INIZCTEOSA-N	NS	12438524, 17962372
flurazepam	SAADBVGJQAEFS-UHFFFAOYSA-N	NS	12438524
fluvastatin	FJLGEFLZQAZZCD-JUFISIKESA-N	S	21447733
fluvoxamine	CJOFXWAVKWHTFT-XSFVSMFZSA-N	NS	12438524, 17962372
furosemide	ZZUFCTLCJUWOSV-UHFFFAOYSA-N	NS	11834888
gatifloxacin	XUBOMFCQGDBHMK-UHFFFAOYSA-N	S	20573570

Jedinjenje	InChI ključ	Klasa	PMID referenci
gefitinib	XGALLCVXEZPNRQ-UHFFFAOYSA-N	S	20421331
gemfibrozil	HEMJJKBWTPKOJG-UHFFFAOYSA-N	NS	14698041
gemifloxacin	ZRCVYEYHRGVLOC-HYARGMPZSA-N	S	21864659
gimatecan	UIVFUQKYVFCEKJ-OPTOVBNMSA-N	NS	18089724
glyburide	ZNNLBTZKUZBEKO-UHFFFAOYSA-N	NS	16460798
grepafloxacin	AIJTTZAVMXIJGM-UHFFFAOYSA-N	S	11877335, 10992002, 15131241, 11370709
hoechst 33342	PRDFBSVERLRMY-UHFFFAOYSA-N	S	11602674
hydrocodone	LLPOLZWFMWNKH-CMKMFDUSA-N	NS	17962372
hydroxy-bupropion	NDPTTXIBLSWNSF-SKDRFNHKSAN	NS	18239278, 20599802
hydroxyrubicin	HEQRYQONNHFDHG-GPUIEZTRSA-N	S	7805856, 16686367
hydroxyzine	ZQDWXGKKHFNSQK-OAQYLSRUSA-N	NS	17962372, 22188412
idarubicin	XDXDZDZNSLXDNA-TZNDIEGXSA-N	S	7575681, 14641819
imatinib	KTUFNOKKBVMGRW-UHFFFAOYSA-N	S	12975485, 18449471
imipramine	BCGWQEUPMDMJNV-UHFFFAOYSA-N	NS	12438524
indinavir	CBVCZFGXHXORBI-PXQQMZJSSA-N	S	11602674, 12438524, 11181899, 11242145, 9435299
indomethacin	CGIGDMFJXJATDK-UHFFFAOYSA-N	NS	12438524
irinotecan	UWKQSNNFCGGAFS-XIFFEERXSA-N	S	10101137, 20583968
isoxicam	YYUAYBYLJSNDCX-UHFFFAOYSA-N	S	16686367
itraconazole	VHVPQPYKVGDNFY-ZPGVKDDISA-N	NS	11602674
ivermectin	AZSNMRSAGSSBNP-XPNUAGNSA-N	S	7560060, 7910522, 20513441

Jedinjenje	InChI ključ	Klasa	PMID referenci
ixabepilone	FABUFPQFXZVHFB-PVYNADRNSA-N	S	21262849
ketamine	YQEZLKZALYSWHR-CYBMUJFWSA-N	NS	12438524
ketobemidone	ALFGKMXHOUSVAD-UHFFFAOYSA-N	NS	16278191
ketoconazole	XMAYWYJOQHxEEK-OZXSUGGESA-N	S	18197554
LAAM	XBMIVRRWGCYBTQ-AVRDEDQJSA-N	S	17690563
labetalol	SGUAFYQXFOLMHL-FZKQIMNGSA-N	S	12438524
lamivudine	JTEGQNOMFQHVDC-NKWVEPMBSA-N	S	19472342
lamotrigine	PYZRQGJRPPTADH-UHFFFAOYSA-N	S	18824002
lansoprazole	MJIHNNLFOKEZEW-VWLOTQADSA-N	S	11770010
lapatinib	BCFGMOOMADDAQU-UHFFFAOYSA-N	S	18216274
latanoprost	GGXICVAJURFBLW-CEYXHVGTSA-N	NS	20028257
latanoprost acid	HNPFPERDNWXAGS-NFVOFSAMSA-N	NS	20028257
LBT-999	XZWRXMFAFBSAJC-BFKYMBBCJSA-N	NS	21321274
levetiracetam	HPHUVLMMVZITSG-LURJTMIESA-N	S	18824002
levofloxacin	GSDSWSVVBLHKDQ-JTQLQIEISA-N	S	9262363, 10992002
levomeprazine	VRQVVMDWGGWHTJ-AWEZLNQCLSA-N	S	12438524
lidocaine	NNJVILVZKWQKPM-UHFFFAOYSA-N	NS	11602674, 12438524, 15804173
lomefloxacin	ZEKZLJVOYLTDKK-UHFFFAOYSA-N	NS	16686367, 11102735
loperamide	RDOIQA HITMMDAJ-UHFFFAOYSA-N	S	11602674, 12438524, 15244080, 8647944
lopinavir	KJHKTHWMRKYKJE-SUGCFTRWSA-N	S	17890284, 17451894, 20551216
loratadine	JCCNYMKQOSZNPW-UHFFFAOYSA-N	S	11602674, 17180728, 12584158

Jedinjenje	InChI ključ	Klasa	PMID referenci
lorcainide	XHOJAWVAWFHGHU-UHFFFAOYSA-N	NS	12438524
losartan	LXOVSDHESOKIRD-UHFFFAOYSA-N	S	10725273
lovastatin	PCZOHLXUXFIOCF-BXMDZJJMSA-N	NS	11474784, 21447733
mannitol	FBPFZTCFMRRESA-GUCUJZIJSA-N	NS	11602674, 12438524
maprotiline	QSLMDECMDJKHMQ-GSXCWMCISA-N	NS	16686367
mebendazole	OPXLLQIJSORQAM-UHFFFAOYSA-N	NS	16686367, 20513441
melphalan	SGDBTWWUNNDEQ-LBPRGKRZSA-N	S	16686367, 19447222
mephentermine	RXQCGGRTAILOIN-UHFFFAOYSA-N	NS	12438524
mephobarbital	ALARQZQTBTVLJV-CYBMUJFWSA-N	NS	22342565
meprobamate	NPPQSCRMBWNHMHU-UHFFFAOYSA-N	NS	12438524
mequitazine	HOKDBMAJZXIPGC-MRXNPFEDSA-N	S	12438524
metergoline	WZHJKEUHNJHDLS-QTGUNEKASA-N	NS	12438524
methadone	USSIQXCVUWKGNF-QGZVFWFLSA-N	S	17690563, 11408360, 15876424, 19887017, 14712343
methotrexate	FBOZXECLQNJBKD-ZDUSSCGKSA-N	NS	11602674
methylphenidate	DUGOZIWVEXMGBE-STQMWFEEESA-N	NS	17962372
methylprednisolone	VHRSUDSXCMTMA-PJHHCJLFSAN	S	22464980, 19883743, 10586385
methysergide	KPJZHOPZRAFDTN-ZRGWGRIASAN	S	12438524
metoclopramide	TTWJBBZEZQICBI-UHFFFAOYSA-N	NS	17962372
metoprolol	IUBSYMUCCVWXPE-AWEZLNQCLSAN	NS	12438524
mexiletine	VLPIATFUUWWMKC-SNVBAGLBSAN	NS	12438524
mibefradil	HBNPJJILLOYFJU-VMPREFPWSAN	S	10901697

Jedinjenje	InChI ključ	Klasa	PMID referenci
midazolam	DDLIGBOFAVUZHB-UHFFFAOYSA-N	NS	10213372, 12438524
mitomycin C	NWIBSHFKIJFRCO-WUDYKRTCSEA-N	S	12406646
mitoxantrone	KKZJGLLVHKMTCM-UHFFFAOYSA-N	S	11602674
monensin	GAOZTHIDHYLHMS-GDMSFIFLSA-N	S	11602674
morphine	BQJCRHHNABKAKU-KBQPJGBKSA-N	S	21589798, 16686367, 19481690
N-desmethyl- loperamide	ZMOPTLXEYOVARP-UHFFFAOYSA-N	S	20212014
nalbuphine	NETZHAKZCGBWSS-CEDHKZHLSA-N	S	12438524
naloxone	UZHSEJADLWPNE-GRGSLBFTSA-N	NS	12438524, 19530072
naltrexone	DQCKKXVULJGBQN-XFWGSAIBSA-N	NS	12438524
nelfinavir	QAGYKUNXZHXMNR-HKWSIXNMSA-N	S	11602674, 12438524, 9435299
neostigmine	ALWKGYPQUAPLQC-UHFFFAOYSA-N	S	11602674, 12438524
nevirapine	NQDJXKOVJZTUJA-UHFFFAOYSA-N	NS	19748977, 17328866, 12180537
nicardipine	XAJFUWAWPIDHAI-DEOSSOPVSA-N	NS	12128170, 11145223
nifedipine	MRPWJGQWTDENNT-UHFFFAOYSA-N	NS	10213372, 11602674, 1352973
nitrazepam	YNIRJEBNUCUES-UHFFFAOYSA-N	NS	12438524
nitrendipine	PFOADBWKUBBKRY-MRXNPFEDSA-N	NS	11602674
nitrofurantoin	RRMUNAVQJLIUIA-YCRREMRBSA-N	NS	22004608
nizatidine	MZDZEYRHRKDOOP-IZZDOVSWSA-N	S	19319690
norbuprenorphine	YOYLLRBMGQRFTN-IOMBULRVSA-N	S	19887017
nordazepam	AKPLHCDWDRPJGD-UHFFFAOYSA-N	NS	12438524
nortriptyline	PHVGLTMQBUIQQ-UHFFFAOYSA-N	S	10700657, 16841511

Jedinjenje	InChI ključ	Klasa	PMID referenci
norverapamil	UPKQNCPKPOLASS-SANMLTNESA-N	NS	10773005
noscapine	AKNNEGZIBPJZJG-MSOLQXFVSA-N	NS	12438524
olanzapine	KVWDHTXUZHCGIO-UHFFFAOYSA-N	S	14702023, 12031686
olmesartan	VTRAEEXHOVJFV-UHFFFAOYSA-N	NS	17823233, 20222053
ondansetron	FELGMEQIXOGIFQ-CYBMUJFWSA-N	S	8647944
oseltamivir	VSZGPKBBMSAYNT-RRFJBIMHSA-N	S	19721074, 18039806
oxfendazole	BEZZFPOZAYTVHN-JOCHJYFZSA-N	NS	15703302
oxprenolol	CEMAWMOMDPGJMB-ZDUSSCGKSA-N	NS	12438524
oxybutynin	XIQVNETUBQGFHX-JOCHJYFZSA-N	NS	21392072
p-MPPF	YJZYDPRMWYWYCG-UHFFFAOYSA-N	NS	21321274
paclitaxel	RCINICONZNJXQF-MZXODVADSA-N	S	11602674, 9050899, 22359351
pafenolol	PKWZWSXSCKVUJB-MRXNPFEDSA-N	S	16686367, 12477351
PE2I	YYCOWMQMXNLRHZ-DXWKFQSLSA-N	NS	21321274
pentazocine	VOKSWYLNZZRQPF-CCKFTAQKSA-N	S	15848954, 8101846
perphenazine	RGCVKNLCSQQDEP-UHFFFAOYSA-N	NS	12438524
pheniramine	IJHNSHDBIRRJRN-OAHLLOKOSA-N	NS	12438524
phenobarbital	DDBREPKUVSBGFI-UHFFFAOYSA-N	S	18824002, 20417647
phenytoin	CXOFVDLJLONNDW-UHFFFAOYSA-N	S	18824002, 20417647
PIB	ZQAQXZBSGZUUNL-UHFFFAOYSA-N	NS	21321274, 17204702
pirarubicin	KMSKQZKKOZQFFG-HSUXVGOQSA-N	S	15350632, 12489924
pirenzepine	RMHMFHUVIITRHF-UHFFFAOYSA-N	S	12438524
piroxicam	QYSPLQLAKJAUJT-UHFFFAOYSA-N	NS	16686367

Jedinjenje	InChI ključ	Klasa	PMID referenci
pitavastatin	VGYFMXBACGZSIL-MCBHFWOFSA-N	S	21206133
PK11195	RAVIZVQZGXBOQO-CQSZACIVSA-N	NS	21321274, 17204702
podophyllotoxin	YJGVMLPVUAXIQN-XVVVDYKMHSA-N	S	16686367
practolol	DURULFYMVIFBIR-ZDUSSCGKSA-N	NS	11602674
prazosin	IENZQIKPVFGBNW-UHFFFAOYSA-N	S	11602674, 21321274
prednisolone	OIGNJSKKLXVSLV-VWUMJDOOSA-N	S	12379510, 10586385
probenecid	DBABZHXTCFAPX-UHFFFAOYSA-N	NS	16686367
procyclidine	WYDUSKDSKCASEF-IBGZPJMESA-N	NS	12438524
progabide	DWEQWXS KOHNBNT-SAPNQHFASA-N	NS	21991360
promazine	ZGUGWUXLJSTTMA-UHFFFAOYSA-N	NS	12438524
promethazine	PWWVAXIEGOYWEE-CYBMUJFWSA-N	NS	12438524
propoxyphene	XLMALTXPSGQGBX-GCJKJVERSA-N	NS	17962372
propranolol	AQHHDHDLHHXJYJD-CQSZACIVSA-N	NS	11602674, 12438524
protriptyline	BWPIARFWQZKAIA-UHFFFAOYSA-N	S	12438524
puromycin	RXWNC PJZOCPEPQ-ZSRCTWEP SA-N	S	11602674, 11707575
pyridostigmine	RVOLLAQWKVFTGE-UHFFFAOYSA-N	NS	11602674
pyrilamine	YECBIJXISLIIDS-UHFFFAOYSA-N	S	17180728
quinidine	LOUPRKONTZGTKE-LHHVKLHASA-N	S	11602674
raclopride	WAOQONBSWFLFPE-VIFPVBQESA-N	NS	21321274
raltegravir	ZEJGORPKXHJSER-SQFISAMPSA-N	S	21393174, 21078936
ranitidine	HSKYAXOF SOYCML-LCYFTJDESA-N	NS	11602674
reserpine	QEVHRUUCFGRFIF-MDEJGZGSSA-N	S	11602674

Jedinjenje	InChI ključ	Klasa	PMID referenci
rhodamine 123	FCGVBHISQBBIQF-UHFFFAOYSA-N	S	11602674
rifampicin	FZYOVNIOYYPUPY-RIEJT TDMSA-N	S	8633005
riluzole	FTALBRSUTC GOEG-UHFFFAOYSA-N	S	17635670
risperidone	RAPZEAPATHNIPO-UHFFFAOYSA-N	S	12438524, 16118767, 12031686
ritonavir	NCDNCNXC DXHOMX-XGKFQTDJSA-N	S	11602674, 11181899, 11242145
rivaroxaban	KGFYHTZWP PHNLQ-AWEZ NQCLSA-N	S	21515813
roscovitine	BTIHMVBBUGXLCJ-OAHLLOKOSA-N	S	20699410
rosiglitazone	YASAKCUCGLMORW-OAHLLOKOSA-N	S	20350646, 19776663
rosuvastatin	BPRHUIZQVSMCRT-VEUZHWNKSA-N	S	21447733
roxindole	HGEYJZMMUGWEOT-UHFFFAOYSA-N	NS	12438524
salinomycin	KQXDH UJYNAXLNZ-XQSDOZFQSA-N	S	18195061
saquinavir	QWAXKHKRTORLEM-UGJKXSETSA-N	S	11602674, 12438524, 9435299
scopolamine	STECJAGHUSJQJN-FWXGHANASA-N	NS	12438524
selegiline	MEZLKOACVSPNER-GFCCVEGCSA-N	NS	21321274
simvastatin	RYMZZMVNJRMUDD-HGQWONQESA-N	NS	11474784, 21447733
SN-38	FJHBVJOVLFPMQE-QFIPXVFZSA-N	S	20583968
sorafenib	MLDQJTXFUGDVEO-UHFFFAOYSA-N	S	20103600, 21309545
sparfloxacin	DZZWHBIBMUVIIW-DTORHVGOSA-N	S	11102735
sufentanil	GGCSSNBKKAUURC-UHFFFAOYSA-N	NS	11964599
sulfamethoxazole	JLKIGFTWXXRPMT-UHFFFAOYSA-N	NS	21991360
sulfasalazine	NCEXYHBECQHGNR-QZQOTICOSA-N	NS	12438524, 19541926
sulpiride	BGRJTUBHPOOWDU-NSHDSACASA-N	NS	17962372

Jedinjenje	InChI ključ	Klasa	PMID referenci
sumatriptan	KQKPFRRSPSRPDEB-UHFFFAOYSA-N	NS	11602674, 12438524
sunitinib	WINHZLLDWRZWRT-ATVHPVEESA-N	S	21351087, 21309545
tacrine	YLJREFDVOIBQDA-UHFFFAOYSA-N	NS	12438524
talinolol	MXFWWQICDIZSOA-KRWDZBQOSA-N	S	11231102
tandutinib	UXXQOJXBIDBUAC-UHFFFAOYSA-N	S	20670210
teniposide	NRUKOCRGYNPUPR-QBPJDGROSA-N	S	12649196, 16686367
tenofovir	SGOIRFVFHAKUTI-ZCFIWIBFSA-N	NS	17005808, 19075072
tenoxicam	LZNWYQJJBLGYLT-UHFFFAOYSA-N	S	16686367
terfenadine	GUGOEXESWIERI-PMERELPUSA-N	S	11602674, 12438524, 17180728, 10213372
thalidomide	UEJJHQNACJXSKW-SECBINFHSA-N	NS	16136308
thiopental	IUJDSEJGGMCXSG-ZETCQYMHSA-N	NS	17962372
topiramate	KJADKKWYZYXHBB-XBWDGYHZSA-N	S	19730994
topotecan	UCFGDBYHRUNTLO-QHCPKHFHSA-N	S	17975156, 19567673, 18654741
topovale	UVXCXZBZPFCAAJ-UHFFFAOYSA-N	NS	14679002
trabectedin	BANJOPLRVCTHHC-MCWJFWOWSA-N	S	16633714
travoprost	MKPLKVHSHYCHOC-AHTXBMBWSA-N	NS	20028257
travoprost acid	WWSWYXNVCBLWNZ-QIZQQNKQSA-N	NS	20028257
trazodone	PHLBKPHSAVXXEF-UHFFFAOYSA-N	NS	17962372
triamterene	FNYLWPVRPXGIIP-UHFFFAOYSA-N	NS	11602674
trimethoprim	IEDVJHCEMCRBQM-UHFFFAOYSA-N	S	11602674, 12438524
trimipramine	ZSCDBOWYZJWBIY-MRXNPFEDSA-N	NS	12438524

Jedinjenje	InChI ključ	Klasa	PMID referenci
triprolidine	CBEQULMOCCWAQT-WOJGMQOQSA-N	NS	12584158
tropium	OYYDSUSKLWTMMQ-NUNAXRQHSA-N	S	19389858, 21392072
ulifloxacin	SUXQDLLXIBLQHW-UHFFFAOYSA-N	S	15618759
valproic acid	NIJYAXOARWZEE-UHFFFAOYSA-N	NS	17043155, 20080116
verapamil	SGTNSNPWRIOYBX-HHHXNRCGSA-N	NS	11602674
vicriviroc	CNPVJJQCETWNEU-CYFREDJKSA-N	S	21393174
vinblastine	JXLYSJRDCGARV-XQKSVPLYSAN	S	11602674, 11181899, 7945455
vincristine	OGWKCGZFUXNPDA-PVMMVQJQSA-N	S	11602674
vindesine	HHJUWIANJFBDHT-KOTLKJBCSA-N	S	16686367
vinorelbine	GBABOYUKABKIAF-YNWCSYGCSA-N	S	11602674
warfarin	PJVWKTQMONHTI-OAHLLOKOSA-N	NS	12438524
WAY-100635	SBPRIAGPYFYCRT-UHFFFAOYSA-N	NS	21321274
yohimbine	BLGXFZZNTVWLAY-SCYLSFHTSA-N	NS	11602674
zidovudine	HBOMLICNUCNMMY-XLPZGREQSA-N	S	19472342
zimeldine	OYPPVKRFBIWMSX-SXGWCWSVSA-N	NS	12438524
zolmitriptan	ULSDMUVEXKOYBU-ZDUSSCGKSA-N	S	12438524
zolpidem	ZAFYATHCZYHLPB-UHFFFAOYSA-N	NS	17962372

Tabela B.2: Set podataka korišćen u razvoju modela za predviđanje transporta posredovanog BCRP-om.

Jedinjenje	InChI ključ	Klasa	PMID referenci
2-fluoro-A-85380	GVOOYVOZPRROMP-ZETCQYMHSA-N	NS	21321274
5-fluorouracil	GHASVSINZRGABV-UHFFFAOYSA-N	S	18820913
abacavir	MCGSCOLBFJQGHM-SCZZXKLOSA-N	S	17437964, 17202245
acyclovir	MKUXAQIIEYXACX-UHFFFAOYSA-N	S	21859328
albendazole	HXHWSAZORRCQMX-UHFFFAOYSA-N	NS	15703302
albendazole sulfoxide	VXTGHWHFYNYFFV-LJQANCHMSA-N	S	15703302
amprenavir	YMARZQAQMVYCKC-OEMFJLHTSA-N	NS	15007102, 19491037
atazanavir	AXRYRYVKAWYZBR-GASGPIRDSA-N	NS	20551216, 17202245
atorvastatin	XUKUURHRXDUEBC-KAYWLYCHSA-N	S	21447733
axitinib	RITAVMQDGBJQJZ-FMIVXFBMSA-N	S	21282407
beclomethasone	NBMKJKDGKREAPL-DVTGEIKXSA-N	NS	15365089
befloxatone	IALVDLPLCLFBCF-CHWSQXEVSA-N	S	21321274
belotecan	LNHWXBUNXOXMRL-VWLOTQADSA-N	S	18654741
bimatoprost	AQOKCDNYWBIDND-FTOWTWDKSA-N	NS	20028257
bimatoprost acid	YFHHIZGZVLHBQZ-IIVAHVFHSA-N	NS	20028257
bisantrene	NJSMWLQOCQIOPE-OCHFTUDZSA-N	S	10806112
buprenorphine	RMRJXGBAOAMLHD-IHFGGWKQSA-N	NS	19887017
canertinib	OMZCMEYTWSXEPZ-UHFFFAOYSA-N	S	16686368
carbamazepine	FFGPTBGBLSHEPO-UHFFFAOYSA-N	NS	16529607
cediranib	XXJWYDDUDKYVKI-UHFFFAOYSA-N	S	22323823
cerivastatin	SEERZIQQUAZTOL-ANMDKAQQSA-N	S	15901800

Jedinjenje	InChI ključ	Klasa	PMID referenci
CGP 12177	UMQUQWCJKFOUGV-VIFPVBQESA-N	NS	21321274
chlorothiazide	JBMKAUGHUNFTOL-UHFFFAOYSA-N	S	22123130
cimetidine	AQIXAKUUQRKLND-UHFFFAOYSA-N	S	15365089
ciprofloxacin	MYSWGUAQZAJKOK-UHFFFAOYSA-N	S	16434544, 21930826
clonazepam	QKPRGUMFCHDCOK-UHFFFAOYSA-N	NS	16529607
clorgyline	BTFHLQRNAMSRLC-UHFFFAOYSA-N	NS	21321274
cyclosporin A	PMATZTZNYRCHOR-CGLBZJNRSA-N	NS	17220244, 16404634, 17031644
cytarabine	UHDGCWIWMRVCDJ-CCXZUQQUSA-N	NS	14574327
danofloxacin	QMLVECGLEOSESV-RYUDHWBXSAN	S	20950350
dasatinib	ZBNZXTGUTAYRHI-UHFFFAOYSA-N	S	19491323
daunorubicin	STQGQHZAVUOBTE-VGBVRHCVSAN	S	16686368, 9861027
dexamethasone	UREBDLICKHMUKA-CXSFZGCWSAN	NS	15365089
diclofenac	DCOPUUMXTXDBNB-UHFFFAOYSA-N	S	18845662
digoxin	LTMHDMANZUZIPE-ALVQDVFTSAN	NS	15365089
diprenorphine	OIJXLIIMXHRJJH-KNLIKEYSAN	NS	21321274
dipyridamole	IZEKFCXSFNUWAM-UHFFFAOYSA-N	S	16247709
doxorubicin	AOJJSUZBOXZQNB-TZSSRYMLSAN	S	18820913
DPA-714	FLZZFWBNYJNHMY-UHFFFAOYSA-N	NS	21321274
enrofloxacin	SPFYMRJSYKXGV-UHFFFAOYSA-N	S	16846465
epirubicin	AOJJSUZBOXZQNB-VTZDEGQISAN	S	16619483
erlotinib	AAKJLRGGTJKAMG-UHFFFAOYSA-N	S	21088257, 20963470, 18723475
estrone-3-sulfate	JKKFKPJIXZFSSB-CBZIJGRNSAN	S	22359351

Jedinjenje	InChI ključ	Klasa	PMID referenci
ethosuximide	HAPOVYFOVVWLRZ-ZETCQYMHSA-N	NS	16529607
etoposide	VJJPUSNTGOMMGY-LCBFZJDDSA-N	S	21935580, 12649196
etravirine	PYGWGZALEOIKDF-UHFFFAOYSA-N	NS	21189339
fallypride	OABRYNHZQBZDMG-INIZCTEOSA-N	NS	21321274
fenbendazole	HDDSHPAODJUKPD-UHFFFAOYSA-N	NS	15703302
flavopiridol	BIIVYFLTOXDAOV-YVEFUNNKSA-N	S	19225039
flumazenil	OFBIFZUFASYRE-UHFFFAOYSA-N	NS	21321274
fluvastatin	FJLGEFLZQAZZCD-JUFISIKESA-N	S	21447733
gefitinib	XGALLCVXEZPNRQ-UHFFFAOYSA-N	S	20421331
genistein	TZBJGXHYKVUXJN-UHFFFAOYSA-N	S	22359351
gimatecan	UIVFUQKYVFCEKJ-OPTOVBNMSA-N	S	18089724
glyburide	ZNNLBTZKUZBEKO-UHFFFAOYSA-N	S	16460798
grepafloxacin	AIJTTZAVMXIJGM-JTQLQIEISA-N	S	17639028
hoechst 33342	PRDFBSVERLRMY-UHFFFAOYSA-N	S	11781231, 11801536
hydroxy-bupropion	NDPTTXIBLSWNSF-SKDRFNHKSAN	NS	20599802
imatinib	KTUFNOKKBVMGRW-UHFFFAOYSA-N	S	15251980, 18523872
indinavir	CBVCZFGXHXORBI-PXQQMZJSSA-N	NS	15007102
ixabepilone	FABUFPQFXZVHFB-PVYNADRNSA-N	NS	21262849
kaempferol	IYRMWMYZSQPJKC-UHFFFAOYSA-N	S	21139040
ketoconazole	XMAYWYJOQHxEEK-OZXSUGGESAN	NS	17518356
lamotrigine	PYZRQGJRPPTADH-UHFFFAOYSA-N	NS	16529607
lapatinib	BCFGMOOMADDAQU-UHFFFAOYSA-N	S	18216274

Jedinjenje	InChI ključ	Klasa	PMID referenci
latanoprost	GGXICVAJURFBLW-CEYXHVGTSAN	NS	20028257
latanoprost acid	HNPFPERDNWXAGS-NFVOFSAMSA-N	NS	20028257
LBT-999	XZWRXMFAFBSAJC-BFKYMBBCJSA-N	NS	21321274
leflunomide	VHOGYURTWQBHIL-UHFFFAOYSA-N	S	18397960
levofloxacin	GSDSWSVVBLHKDQ-JTQLQIEISA-N	S	16434544
loperamide	RDOIQA HITMMDAJ-UHFFFAOYSA-N	NS	21321274
lopinavir	KJHKTHWMRKYKJE-SUGCFTRWSA-N	NS	20551216, 17202245
lovastatin	PCZOHLXUXFIOCF-BXMDZJJMSA-N	NS	21447733
methotrexate	FBOZXECLQNJBKD-ZDUSSCGKSA-N	S	12208758, 12874005, 18820913
methylprednisolone	VHRSUDSXCMQTMA-PJHHCJLFSAN	NS	15365089
mitomycin C	NWIBSHFKIJFRCO-WUDYKRTCASAN	NS	18820913
mitoxantrone	KKZJGLLVHKMTCM-UHFFFAOYSA-N	S	19572416, 16686368, 18820913
N-desmethyl- loperamide	ZMOPTLXEYOVARP-UHFFFAOYSA-N	NS	20212014
nilotinib	HHZIURLSWUIHRB-UHFFFAOYSA-N	S	19427995, 19785662
nitrofurantoin	RRMUNAVQJLIUIA-YCRREMRBSAN	S	22004608
nizatidine	MZDZEYRHRKDOOP-IZZDOVSWSAN	NS	19319690
norbuprenorphine	YOYLLRBMGQRFTN-IOMBULRVSAN	NS	19887017
norfloxacin	OGJPXUAPXNRGGI-UHFFFAOYSA-N	S	16434544
olmesartan	VTRAEW XHOVJFV-UHFFFAOYSA-N	NS	17823233, 20222053
oseltamivir	VSZGPKBBMSAYNT-RRFJBIMHSAN	NS	19721074, 18039806
oxfendazole	BEZZFPOZAYTVHN-JOCHJYFZSAN	S	15703302

Jedinjenje	InChI ključ	Klasa	PMID referenci
p-MPPF	YJZYDPRMWWYCYG-UHFFFAOYSA-N	NS	21321274
paclitaxel	RCINICONZNJXQF-MZXODVADSA-N	NS	18820913, 22359351
PE2I	YYCOWMQMXNLRHZ-DXWKQFQSLA-N	NS	21321274
phenobarbital	DDBREPKUVSBGFI-UHFFFAOYSA-N	NS	16529607
phenytoin	CXOFVDLJLONNDW-UHFFFAOYSA-N	NS	16529607
PhIP	UQVKZNNCIHJZLS-UHFFFAOYSA-N	S	15365089
PIB	ZQAQXZBSGZUUNL-UHFFFAOYSA-N	NS	21321274
pirarubicin	KMSKQZKKOZQFFG-HSUXVGOQSA-N	S	18820913
pitavastatin	VGYFMXBACGZSIL-MCBHFWOFSAN	S	15955871
PK11195	RAVIZVQZGXBOQO-CQSZACIVSA-N	NS	21321274
pravastatin	TUZYXOIXSAXUGO-PZAWKZKUSA-N	S	15901800
prazosin	IENZQIKPVFGBNW-UHFFFAOYSA-N	S	18523872, 21321274, 22355323
primidone	DQMZLTXERSFNPB-UHFFFAOYSA-N	NS	16529607
puromycin	RXWNCPJZOCPEPQ-ZSRCTWEPSA-N	S	17638908
raclopride	WAOQONBSWFLFPE-VIFPVBQESA-N	NS	21321274
riluzole	FTALBRSUTCGOEG-UHFFFAOYSA-N	S	19146924
rosuvastatin	BPRHUIZQVSMCRT-VEUZHWNKSA-N	S	18617601, 16415124, 21447733
selegiline	MEZLKOACVSPNER-GFCCVEGCSA-N	NS	21321274
simvastatin	RYMZZMVNJRMUDD-HGQWONQESA-N	NS	21447733
SN-38	FJHBVJOVLFPMQE-QFIPXVFZSA-N	S	16686368
sorafenib	MLDQJTXFUGDVEO-UHFFFAOYSA-N	S	20952483, 21309545, 20103600
sulfasalazine	NCEXYHBECQHGNR-QZQOTICOSA-N	S	19541926

Jedinjenje	InChI ključ	Klasa	PMID referenci
sunitinib	WINHZLLDWRZWRT-ATVHPVEESA-N	S	21309545
tandutinib	UXXQOJXBIDBUAC-UHFFFAOYSA-N	S	20670210
teniposide	NRUKOCRGYNPUPR-QBPJDGROSA-N	S	12649196
teriflunomide	UTNUDOFZCWSZMS-YFHOOEESVSA-N	S	18397960
topotecan	UCFGDBYHRUNTLO-QHCPKHFHSA-N	S	17975156, 18523872, 16686368, 18654741
topovale	UVXCXZBZPFCAAJ-UHFFFAOYSA-N	NS	14679002
travoprost	MKPLKVHSHYCHOC-AHTXBMBWSA-N	NS	20028257
travoprost acid	WWSWYXNVCBLWNZ-QIZQQNKQSA-N	NS	20028257
triamcinolone	GFNANZIMVAIWHM-OBYCQNJPSA-N	NS	15365089
triprolidine	CBEQULMOCCWAQT-WOJGMQOQSA-N	NS	16686368
ulifloxacin	SUXQDLLXIBLQHW-MRVPVSSYSA-N	S	17639028
URB937	CMEQHOXCIGFZNJ-UHFFFAOYSA-N	S	21767647
valproic acid	NIJYAXOARWZEE-UHFFFAOYSA-N	NS	16529607
verapamil	SGTNSNPWRIOYBX-HHHXNRCGSA-N	NS	21321274, 22004608, 22200670
vinblastine	JXLYSJRDCGARV-XQKSVPLYSA-N	NS	10806112
vincristine	OGWKCGZFUXNPDA-PVMMVQJQSA-N	NS	18820913
vindesine	HHJUWIANJFDHT-KOTLKJBCSA-N	NS	18820913
WAY-100635	SBPRIAGPYFYCRT-UHFFFAOYSA-N	NS	21321274
zidovudine	HBOMLICNUCNMMY-XLPZGREQSA-N	S	17437964

Tabela B.3: Deskriptori uključeni u finalne modele za predviđanje transporta posredovanog Pgp-om

Deskriptor	Opis	Domen primenjivosti
ANN-24-3 model		
BoxX_3D	Najmanja dimenzija najmanjeg kvadra koji ograničava molekulu u Å	0,0 – 9,08
Prop2_3D	Simulations Plus zaštićeni deskriptor	0,0671 – 0,598
QAvgPos	Srednja vrednost formalno pozitivnog naelektrisanja svih jonizovanih oblika molekule	0,0 – 1,9859
QAvgNeg	Srednja vrednost formalno negativnog naelektrisanja svih jonizovanih oblika molekule	0,0 – 1,9902
FZwitter	Udeo cviterjona u svim formalno neutralnim molekulskim vrstama	0,0 – 1,00
Pi_FMi4	Četvrta komponenta vektora autokorelacije pi Fukui(-) indeksa	0,0 – 104,9
Pi_MinQ	Najmanja vrednost pi Hückel-ovih atomskih naelektrisanja	-0,589 – 0,0
Pi_AFMic	Zbir apsolutnih vrednosti pi Fukui(-) indeksa C atoma	0,0 – 18,1
Pi_Q5	Peta komponenta vektora autokorelacije Hückel-ovih pi atomskih naelektrisanja	-0.3746 – 1.7038
EEM_XFh	Najveći sigma Fukui indeks H atoma	0,23 – 1,9
EEM_MaxF	Najveći sigma Fukui indeks	1,49 – 7,74
EEM_NFnp	Najmanji sigma Fukui indeks polarnih atoma	0,3 – 1,86
NPA_Q2	Druga komponenta vektora autokorelacije procenjenih NPA atomskih naelektrisanja	-3,36 – 12,6
NPA_Q4	Četvrta komponenta vektora autokorelacije procenjenih NPA atomskih naelektrisanja	-7,38 – 8,58
M_POL	Broj aromatičnih polarnih supstituenata (Moriguchi-jev deskriptor)	0,00 – 4,00
T_RDmtr	Relativni topološki dijametar	0,25 – 0,74
SaasC	Atomski indeks elektrotopološkog stanja -C(a)a grupa	-3,1999 – 8,9322
H_MultiN	Prisustvo Multi-N grupe (Meylan-ov deskriptor)	0,0 – 1,00
Sulfide_-S-	Broj sulfidnih grupa	0,0 – 2,0
ArNitrog_=N-	Broj aromatičnih azota	0,0 – 4,0
Ester_C(=O)OC	Broj estarskih funkcionalnih grupa	0,0 – 4,0
PriAmine_-NH2	Broj primarnih amina	0,0 – 3,0
Carbonyl_C=O	Broj karbonila	0,0 – 3,0
Amide_C(=O)N<	Broj amidskih funkcionalnih grupa	0,0 – 11,0

Deskriptor	Opis	Domen primenjivosti
ANN-8-1 model		
MlogP	logP izračunat metodom po Moriguchi-ju	-2,5 – 4,94
Prop2_3D	Simulations Plus zaštićeni deskriptor	0,0671 – 0,598
Dipole_3D	Moment molekuskog dipola procenjen iz NPA atomskih naelektrisanja	5,24e-16 – 5,42
PEoED1a_3D	Broj farmakofora tipa I [393] (uključujući elektron donore sa H atomima)	0,0 – 16,0
F_NLP	Prosečan broj slobodnih elektronskih parova na N, O, S i P kod svih jonizovanih oblika, podeljen ukupnim brojem atom	0,0001 – 1,0
F_HBP	Prosečan broj protona koji mogu formirati H-veze u svim jonizovanim oblicima, podeljen ukupnim brojem atom	0,0 – 0,5
PolarizM	Polarizabilnost u Å ³ izračunata metodom po Miller-u	10,53 – 130,88
N_FrRotB	Broj rotabilnih veza	0,0 – 22,0
SVM model		
BoxX_3D	Najmanja dimenzija najmanjeg kvadra koji ograničava molekulu u Å	0,0 – 9,08
SolvEMt_3D	Dužina vektora momenta energije solvatacije u Å kcal mol ⁻¹	2,84e-15 – 122,0
PEoEDI1b_3D	Broj farmakofora tipa II [393] (isključujući elektron donore sa H atomima)	0,0 – 26,0
QAvgNeg	Srednja vrednost formalno negativnog naelektrisanja svih jonizovanih oblika molekule	0,0 – 1,9902
HBDnch	Zbir atomskih naelektrisanja H atoma vezanih za N koji je donor H-veze	0,0 – 2,4
PHB	Deskriptor prostorne bliskosti donora i akceptora H-veze	0,0 – 6,0
Pi_FMi5	Peta komponenta vektora autokorelacije pi Fukui(-) indeksa	0,0 – 69,57
Pi_MaxFMi	Najveća vrednost pi Fukui(-) indeksa	0,0 – 25,13
Pi_Q6	Šesta komponenta vektora autokorelacije Hückel-ovih pi atomskih naelektrisanja	-0,4237 – 0,3414
Pi_MinQ	Najmanja vrednost pi Hückel-ovih atomskih naelektrisanja	-0,589 – 0,0
Pi_FPl6	Šesta komponenta vektora autokorelacije pi Fukui(+) indeksa	0,0 – 77,16
EEM_XFh	Najveći sigma Fukui indeks H atoma	0,23 – 1,9
EEM_XFnp	Najveći sigma Fukui indeks nepolarnih atoma	1,17 – 7,74
EEM_NFnp	Najmanji sigma Fukui indeks nepolarnih atoma	0,3 – 1,86
EEM_NFpl	Najveći sigma Fukui indeks polarnih atoma	-0,44 – 1,71

Deskriptor	Opis	Domen primenjivosti
EEM_MaxF	Najveći sigma Fukui indeks	1.49 – 7.74
NPA_Q4	Četvrta komponenta vektora autokorelacije procesnih NPA atomskih naelektrisanja	-7,38 – 8,58
NPA_MinQ	Najmanje NPA atomsko naelektrisanje	-1,08 – -0,56
T_RDmtr	Relativni topološki dijametar	0,25 – 0,74
Blbn_J	Balabanov J indeks	0,81 – 4,27
H_MultiN	Prisustvo Multi-N grupe (Meylan-ov deskriptor)	0,0 – 1,00
SaaS	Atomski indeks elektrotopološkog stanja -C(a)a grupa	-3,1999 – 8,9322
SaaaC	Atomski indeks elektrotopološkog stanja -C(a)(a)a grupa	-0,9965 – 5,2542
SsCH3	Atomski indeks elektrotopološkog stanja -[CH3] grupa	0,0 – 39,7804
SaaS	Atomski indeks elektrotopološkog stanja -N(a)a grupa	0,0 – 4,7884
ArNitrog_=N-	Broj aromatičnih azota	0,0 – 4,0
Carbonyl_C=O	Broj karbonila	0,0 – 3,0
Ester_C(=O)OC	Broj estarskih funkcionalnih grupa	0,0 – 4,0
Amide_C(=O)N<	Broj amidskih funkcionalnih grupa	0,0 – 11,0

Tabela B.4: Deskriptori uključeni u finalne modele za predviđanje transporta posredovanog BCRP-om

Deskriptor	Opis	Domen primenjivosti
ANN-15-1 model		
SM2xx_3D	Najmanji glavni statički moment drugog reda	0,0 – 0,204
NegASA_3D	Površina dostupna rastvaraču okupirana NPA parcijalno negativnim atomima	11,1 – 359,0
FUnion	Udeo formalno neutralnih oblika u svim jonizovanim oblicima pri datoj pH	0,0 – 1,0
Pi_MaxFPI	Najveća vrednost pi Fukui(+) indeksa	1,64 – 7,39
Pi_MinQ	Najmanja vrednost pi Hückel-ovih atomskih naelektrisanja	-0,569 – -0,006
Pi_FMi3	Treća komponenta vektora autokorelacije pi Fukui(-) indeksa	0,0 – 78,27
NPA_MinQ	Najmanje NPA atomsko naelektrisanje	-1,06 – -0,56
NPA_Q4	Četvrta komponenta vektora autokorelacije procenjenih NPA atomskih naelektrisanja	-7,38 – 8,58
Blbn_J	Balabanov J indeks	0,83 – 4,27
SsssC	Atomski indeks elektrotopološkog stanja [CX4] grupa	-7,8769 – 0,0
SaaaC	Atomski indeks elektrotopološkog stanja -C(a)(a)a grupa	-0,272 – 4,5936
SssCH2	Atomski indeks elektrotopološkog stanja -[CH2]- grupa	-2,0433 – 11,5061
SsF	Atomski indeks elektrotopološkog stanja -F grupa	0,0 – 40,8275
F_AFRBWF	Prosečna vrednost faktora ponderacije rotabilnih veza	0,0 – 0,38
PriAmAro_-NH2	Broj primarnih aromatičnih amina	0,0 – 2,0
ANN-5-1 model		
BoxX_3D	Najmanja dimenzija najmanjeg kvadra koji ograničava molekulu u Å	0,0 – 8,92
PolASA_3D	Površina dostupna rastvaraču okupirana NPA parcijalno pozitivnim atomima	0,0 – 318,0
PEoEDia_3D	Broj farmakofora tipa I [393] (uključujući elektron donore sa H atomima)	0,0 – 13,0
DStokes_3D	Stokes-Einstein-ov koeficijent difuzije (cm ² /s · 10 ⁵)	0,258 – 1,14
F_AromB	Udeo aromatičnih veza u ukupnom broju veza	0,0 – 0,84
SVM model		
BoxX_3D	Najmanja dimenzija najmanjeg kvadra koji ograničava molekulu u Å	0,0 – 8,92

Deskriptor	Opis	Domen primenjivosti
SM2zz_3D	Najveći glavni statički moment drugog reda	0,469 – 0,964
PEoEDla_3D	Broj farmakofora tipa I [393] (uključujući elektron donore sa H atomima)	0,0 – 13,0
MinQ	Najmanje PEOE atomsko naelektrisanje	-0,77 – -0,27
Pi_AFMic	Zbir apsolutnih vrednosti pi Fukui(-) indeksa C atoma	0,01 – 18,1
Pi_FMi3	Treća komponenta vektora autokorelacije pi Fukui(-) indeksa	0,0 – 78,27
Pi_AQn	Zbir apsolutnih vrednosti Hückel-ovih pi atomskih naelektrisanja N atoma	0,0 – 2,899
EEM_XFh	Najveći sigma Fukui indeks H atoma	0,23 – 1,86
EEM_NFnp	Najmanji sigma Fukui indeks nepolarnih atoma	0,23 – 1,7
T_RDmtr	Relativni topološki dijametar	0,25 – 0,71
F_AromB	Udeo aromatičnih veza u ukupnom broju veza	0,0 – 0,84
F_AFRBWF	Prosečna vrednost faktora ponderacije rotabilnih veza	0,0 – 0,38

Tabela B.5: Rezultati eksterne validacije modela uz korišćenje dve alternativne definicije domena primenjivosti.

Metod najbližih suseda ^a										
		Pgp modeli								
		ANN-24-3			ANN-8-1			SVM		
Broj NS ^b	Deo seta u DP ^c	Se	Sp	Ac	Se	Sp	Ac	Se	Sp	Ac
0	100%	0.78	0.63	70%	0.89	0.47	67%	0.80	0.59	69%
1	14%	0.93	0.30	79%	0.98	0.30	82%	0.96	0.33	81%
2	11%	0.95	0.26	80%	1.00	0.16	83%	0.96	0.21	80%
3	10%	0.94	0.26	79%	1.00	0.16	81%	0.95	0.21	79%
		BCRP modeli								
		ANN-15-1			ANN-5-1			SVM		
0	100%	0.68	0.84	73%	0.73	0.78	74%	0.62	0.78	68%
1	23%	0.89	0.75	87%	0.95	0.75	91%	0.83	0.75	82%
2	13%	0.91	0.50	84%	0.91	0.50	84%	0.90	0.50	83%
3	11%	1.00	0.50	90%	1.00	0.50	90%	1.00	0.50	90%
Metod udaljenosti ^d										
		Pgp modeli								
		ANN-24-3			ANN-8-1			SVM		
Udaljenost ^e	Deo seta u DP	Se	Sp	Ac	Se	Sp	Ac	Se	Sp	Ac
<i>max</i>	98%	0.77	0.63	69%	0.89	0.47	66%	0.79	0.59	68%
3μ	94%	0.76	0.63	69%	0.88	0.48	66%	0.78	0.60	68%
2μ	84%	0.73	0.64	68%	0.86	0.49	64%	0.73	0.61	66%
		BCRP modeli								
		ANN-15-1			ANN-5-1			SVM		
<i>max</i>	99%	0.64	0.82	71%	0.78	0.79	78%	0.68	0.79	72%
3μ	99%	0.64	0.82	71%	0.78	0.79	78%	0.68	0.79	72%
2μ	94%	0.64	0.80	70%	0.78	0.76	77%	0.68	0.76	71%

^a Korišćen je pristup sličan metodi opisanoj u referenci [401]. Molekulska sličnost između jedinjenja seta za validaciju i jedinjenja korišćenih u uspostavljanju modela procenjivana je pomoću Daylight molekulskih otisaka prstiju, kroz njihovu implementaciju u RDKit-u. Jedinjenja čija je Tanimoto sličnost bila $\geq 0,7$ smatrana su najbližim susedima.

^b Minimalan broj najbližih suseda (NS) u setu korišćenom za uspostavljanje modela, neophodan da bi se smatralo da je jedinjenje u domenu primenjivosti modela.

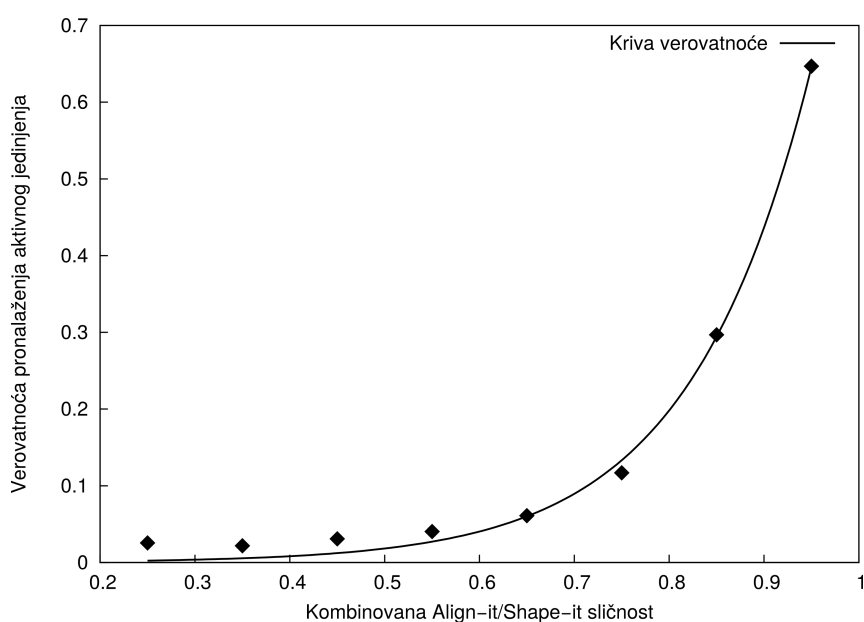
^c Procenat seta za validaciju koji je u domenu primenjivosti (DP) modela.

^d Metod Euklidske udaljenosti, opisan u referenci [402]. U izračunavanju udaljenosti korišćene su vrednosti svih glavnih komponenti ChemGPS-NP sistema [361]. Na osnovu izračunatih udaljenosti, definisana je maksimalna (*max*) i prosečna (μ) udaljenost između jedinjenja u setu za uspostavljanje modela. Ukoliko je udaljenost jedinjenja iz seta za validaciju od najbližeg jedinjenja iz seta za uspostavljanje modela bila veća od definisane odsečne vrednosti, smatrano je da takvo jedinjenje nije u domenu primenjivosti modela.

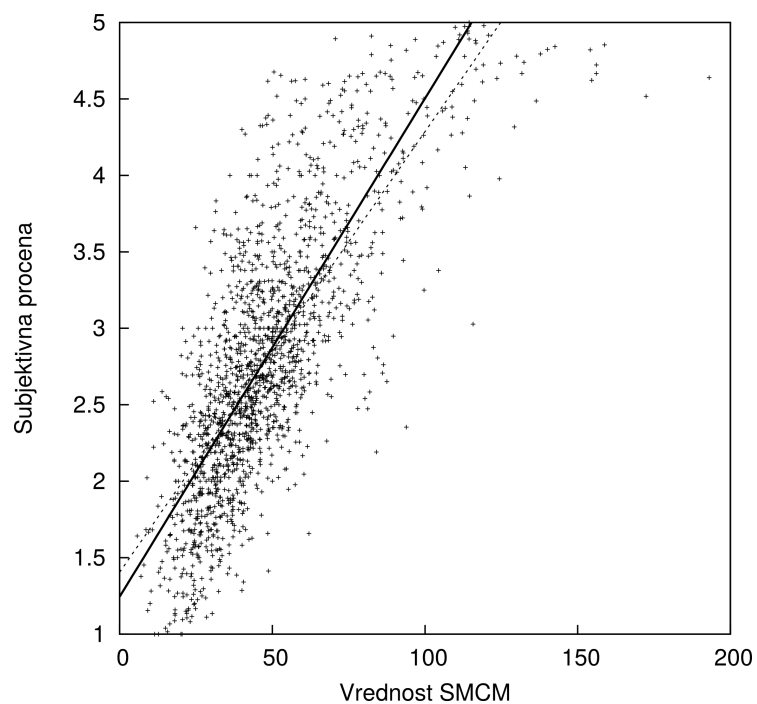
^e Različite odsečne vrednosti korišćene u definisanu domena primenjivosti, pri čemu je značenje *max* i μ opisano u prethodnoj fusnoti.

Prilog C

Karakteristike funkcija procene korišćenih u *de novo* dizajnu



Slika C.1: Verovatnoća identifikovanja podjednako aktivnog jedinjenja pri datom nivou njegove sličnosti sa referentnom strukturom poznate aktivnosti. Sličnost se odnosi na kombinovani rezultat farmakoforne (Align-it) i podudarnosti oblika (Shape-it), procenjene kao Tversky sličnost (detaljnije opisano u odeljku 3.1.8). Kriva je uspostavljena probalističkim pristupom opisanom u Swann et al. (2011) *J. Med. Chem.* 54(5): 1223–1232, korišćenjem seta iz navedene publikacije.



Slika C.2: Korelacija između izračunate SMCM vrednosti i subjektivne ocene organskih hemičara o sintetskoj složenosti 1731 jedinjenja ($R = 0,74$).

Biografija autora

Marko Kalinić rođen je 15. februara 1988. godine u Sarajevu (BiH, SFRJ). Osnovnu školu i gimnaziju završio je u Zemunu. Farmaceutski fakultet Univerziteta u Beogradu upisao je 2006. godine kao prva generacija studenata po Bolonjskom programu. Tokom studija bio je aktivan član Studentskog parlamenta; školske 2010/2011 obavljao je funkciju predsednika Studentskog parlamenta, bio je član Saveta Farmaceutskog fakulteta i član Saveta Univerziteta u Beogradu. U istraživanja u oblasti računarske medicinske hemije uključen je od 3. godine osnovnih studija kroz aktivnosti Centra za naučno-istraživački rad studenata. Bio je stipendista Fondacije za razvoj naučnog i umetničkog podmlatka i Fonda za mlade talente Republike Srbije.

Integrisane akademske studije završio je 2011. godine i zvanje magistra farmacije stekao sa prosečnom ocenom 9,53. Iste godine upisao je doktorske akademske studije na modulu Farmaceutska hemija. Od 2012. godine zaposlen je na Farmaceutskom fakultetu Univerziteta u Beogradu kao istraživač pripravnik na projektu osnovnih istraživanja u oblasti hemije. U zvanje istraživača saradnika izabran je 2014. godine. Tokom doktorskih akademskih studija bio je uključen u realizaciju praktične nastave na predmetima Farmaceutska hemija 1, Farmaceutska hemija 3 i Dizajn i sinteza lekova.

Koautor je 5 radova objavljenih u međunarodnim i 3 rada objavljena u domaćim časopisima, kao i 9 saopštenja sa naučnih skupova štampana u izvodu.

Govori engleski jezik.

Прилог 1.

Изјава о ауторству

Потписани-а Марко Калинић
број индекса 1/2011

Изјављујем

да је докторска дисертација под насловом

**Симулације молекуларне динамике и рачунарско дизајнирање инхибитора
протеин лизин метилтрансферазе EZH2**

- резултат сопственог истраживачког рада,
- да предложена дисертација у целини ни у деловима није била предложена за добијање било које дипломе према студијским програмима других високошколских установа,
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио интелектуалну својину других лица.

Потпис докторанда

У Београду, 7.9.2015.

Марко Калинић

Прилог 2.

Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора

Марко Калинић

Број индекса

1/2011

Студијски програм докторске академске студије фармацеутских наука – модул фармацеутска хемија

Наслов рада Симулације молекуларне динамике и рачунарско дизајнирање инхибитора протеин лизин метилтрансферазе EZH2

Ментор

др сц. Славица Ерић, ванредни професор

Потписани/а

Марко Калинић

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла за објављивање на порталу **Дигиталног репозиторијума Универзитета у Београду**.

Дозвољавам да се објаве моји лични подаци везани за добијање академског звања доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис докторанда

У Београду, 7. 9. 2015.

Марко Калинић

Прилог 3.

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

Симулације молекуларне динамике и рачунарско дизајнирање инхибитора протеин лизин метилтрансферазе EZH2

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигитални репозиторијум Универзитета у Београду могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство
2. Ауторство - некомерцијално
3. Ауторство – некомерцијално – без прераде
4. Ауторство – некомерцијално – делити под истим условима
5. Ауторство – без прераде
6. Ауторство – делити под истим условима

(Молимо да заокружите само једну од шест понуђених лиценци, кратак опис лиценци дат је на полеђини листа).

Потпис докторанда

У Београду, 7.3.2015.

Марко Јамнић