

VILNIAUS UNIVERSITETAS

Vytautas Raškevičius

BALTYMŲ IR MAŽŲ MOLEKULIŲ SĄVEIKOS
KOMPIUTERINIS MODELIAVIMAS

Daktaro disertacija

Technologijos mokslai, chemijos inžinerija (05T)

Vilnius 2017

Disertacija rengta 2012–2017 metais Vilniaus universitete

Mokslinis vadovas:

dr. Visvaldas Kairys (Vilniaus universitetas, technologijos mokslai, chemijos inžinerija – 05T).

Turinys

Turinys.....	3
Naudojamų sutrumpinimų sąrašas	5
Įvadas	6
Darbo tikslas	9
Uždaviniai.....	9
Mokslinį naujumą pagrindžiantys darbo rezultatai	9
Ginamieji disertacijos teiginiai	10
1. Literatūros apžvalga	11
1.1. Karboanhidrazės: šeimos ir izoformos	11
1.2. Dokinimo metodai ir programos	16
1.3. Ligandų jungimosi laisvosios energijos skaičiavimai tiesinės sąveikos energijos metodu.....	21
1.4. Ligandų laisvosios energijos skaičiavimai metadinamikos metodu.....	22
1.5. Kiekybinio struktūros ir aktyvumo ryšio samprata ir metodai	23
1.5.1. QSAR – kas tai?	23
1.5.2. QSAR istorija	26
1.5.3. Šiuolaikinis QSAR.....	29
1.5.4. QSAR modelių vidinis ir išorinis testavimas, tinkamumo patvirtinimas .	33
1.5.5. QSAR modelių statistinis vertinimas ir vizualizavimas	39
1.5.6. QSAR deskriptoriai	45
1.5.7. Trimatis, keturmatas ir penkiamatis QSAR.....	49
1.5.8. QSAR taikymas.....	50
1.5.9. QSAR metodo kritika	54
1.6. Kiti metodai.....	58
2. Metodai.....	59
2.1. α -Pakeistų α,β -nesočiųjų ketonų antiproliferacinio aktyvumo žmogaus hematologinėse ir kietose vėžinių ląstelių linijose QSAR modelių kūrimas.....	59
2.2. Benzensulfonamidų rinkinio dokinimas	59
2.3. Tiesinės sąveikos energijos skaičiavimai.....	60
2.4. Metadinamikos skaičiavimai.....	61
2.5. Karboanhidrazės slopinimo ir selektyvumo QSAR benzensulfonamidams .	62
2.6. PHASE atomais paremtas trimatis QSAR	64
3. Rezultatai ir jų aptarimas	66
3.1. α -Pakeistų α,β -nesočiųjų ketonų antiproliferacinio aktyvumo žmogaus vėžinių ląstelių linijose tyrimas QSAR metodu.....	66
3.1.1. Duomenų rinkinys	66

3.1.2. QSAR modeliai	70
3.2. CA slopinimo benzensulfonamidais prognozavimas įvairiais metodais	74
3.2.1. Benzensulfonamidų rinkinys skaičiavimams	74
3.2.2. Dokinimas į CA II receptorių	77
3.2.3. Tiesinės sąveikos energijos skaičiavimai	79
3.2.4. Metadinamikos skaičiavimai	82
3.2.5. CA II QSAR modeliai	82
3.2.6. Dokinimo, LIE, metadinamikos ir QSAR metodų tinkamumo benzensulfonamidų slopinimo savybėms prognozuoti palyginimas	84
3.2.7. QSAR įvairioms CA izoformoms	85
3.2.8. PHASE atomais paremtas trimatis QSAR, skirtas CA I ir CA XII	91
3.2.9. Selektyvumo iš atskirų afiniškumo QSAR modelių skaičiavimai palyginti su selektyvumo QSAR	96
3.2.10. Benzensulfonamidų QSAR apibendrinimas įvairioms CA izoformoms	98
3.3. Naujų panašių į metabolitus slopiklių paieška naudojant Tanimoto įvertį	99
4. Išvados	103
Literatūros sąrašas	104
Disertacijos tema paskelbtų publikacijų sąrašas	137
Priedas	138

Naudojamų sutrumpinimų sąrašas

CA – karboanhidrazė arba karbonato dehidratazė;

hCA – žmogaus karboanhidrazė arba karbonato dehidratazė;

ITC – izoterminio titravimo kalorimetrija (angl. *Isothermal Titration Calorimetry*);

TSA – terminio poslinkio analizė (angl. *thermal shift assay*);

SAR – struktūros ir savybių ryšys;

KSAR (QSAR) – kiekybinis struktūros ir aktyvumo ryšys;

QSPR – kiekybinis struktūros ir savybių ryšys;

PDB – baltymų duomenų bankas (<http://www.rcsb.org/>) (angl. *Protein Data Bank*);

PDB ID – struktūrą žymintis kodas naudojamas baltymų duomenų banke;

RMSD (RMSE) – šaknis vidutinis kvadratinis nuokrypis (paklaida);

VdW – van der Valso;

LOO, LOOCV – Kryžminis patvirtinimas išmetant vieną (angl. *Leave-one-out cross-validation*);

LIE – tiesinė sąveikos energija;

MD – molekulinė dinamika;

MLR – daugybinė tiesinė regresija;

CV – molekulinės dinamikos sistemos kolektyvinio kintamojo vertė;

QM/MM – kvantinė mechanika ir molekulinė mechanika;

Aminorūgščių pavadinimai trumpinami standartiniu trijų raidžių trumpiniu. Po to trumpinio gali būti nurodytas numeris baltymo aminorūgščių sekoje.

Įvadas

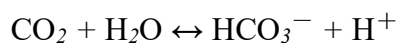
Vienas iš svarbiausių biochemijos mokslo praktinių pritaikymų ir tikslų yra kova su ligomis. Spartus technologijų ir mokslinių metodų vystymasis (pvz., funkcinės genomikos, proteomikos metodai) padeda spręsti daugelį iššūkių. Vienas iš jų yra baltymų – vaistų taikinių, nustatymas, kuriuos paveikus, būtų išgydyta liga arba bent jau palengvinta jos eiga. Ne mažiau svarbus iššūkis yra vaistų prieš ligas paieška, kurių dauguma yra mažos molekulinės masės junginiai, kurie slopina arba aktyvuoja baltymus-taikinius. Naudinga slopinti tik ligas sukeliančių virusų ar bakterijų baltymus, neliečiant svarbių žmogaus organizmo baltymų. Kovoje su vėžiu, dažnai yra taikomasi į baltymus, kurie svarbūs vėžinių ląstelių išlikimui.

Ligandų (mažos molekulinės masės cheminių junginių), darančių įtaką baltymų veiklai, aptikimas ir jų savybių tobulinimas yra pagrindinės užduotys šiuolaikiniame vaistų kūrime. Kyla poreikis lanksčių, greitų ir efektyvių paieškos strategijų skirtų rasti ligandus, besijungiančius į baltymus-taikinius. Rastieji cheminiai junginiai gerinami, kol pasiekia svarbų etapą – pakliūna į klinikinius tyrimus. Juose išsamiai patikrinti junginiai jau gali būti naudojami gydymui, kaip vaistai. Geresni slopikliai randami, jei jų paieškai yra naudojami keli metodai [1]. Skaičiuojamieji metodai *in silico* (lotyniškai „silicyje“, kas reiškia, kad modeliavimas atliktas, panaudojant kompiuterius) modeliuoja mažų molekulių jungimąsi į makromolekulių (pvz., baltymų-taikinių) struktūras, susidarant tarpmolekuliniams ryšiams, bei įvertina jų komplementarumą taikinių atžvilgiu. Jie yra plačiai taikomi ligandų atradime ir tobulinime. Nemažo skaičiaus vaistų, pavyzdžiui, ŽIV proteazių slopiklių, kūrimas buvo plačiai paremtas struktūriniais metodais ir informatinėmis paieškos strategijomis [2]. Nepaisant to, skaičiuojamieji slopiklių paieškos metodai nėra iki galo ištobulinti. Įvairios kompiuterinės ligandų vertinimo schemas nėra visiškai tikslios, nes teisingai nepaskaičiuoja ligandų jungimosi energijos.

Vienas iš didžiųjų iššūkių naujų vaistų kūrime yra junginių selektyvumo

tobulinimas [3]. Jei selektyvumas mažas, slopiklis jungsis į kelis taikinius, sukeldamas šalutinius poveikius. Daug slopiklių, jau naudojamų kaip vaistai, nėra ypatingai selektyvūs kuriai nors vienai fermento izoformai, bet slopina kelias [4]. Tam dažnai pasitelkiamas kiekybinio struktūros ir aktyvumo ryšio (KSAR, angl. QSAR) metodas. Praktikoje QSAR yra matematiniai modeliai, naudojami cheminių medžiagų biologiniams aktyvumams prognozuoti pagal molekulinis deskriptorius suskaičiuotus iš jų cheminės struktūros.

Karboanhidrazės arba karbonato dehidratazės (CA; EC 4.2.1.1) įeina į metalofermentų šeimą. Šie fermentai katalizuoja grįžtamą anglies dioksido ir vandens reakciją, susidarant bikarbonato jonams ir protonams [5]:



CA II yra vienas iš aktyviausių žinomų katalizatorių (su k_{cat} (katalitinė konstanta)/ K_m (Michaelio konstanta) = $1.5 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$) [4]. CA jungiasi prie maždaug trisdešimties vaistų ar preparatų klinikinių tyrimų stadijoje. Žmogaus CA slopikliai jau kelis dešimtmečius naudojami kaip diuretikai ar vaistai nuo glaukomos [6].

Darbe aprašomas tyrimas vyko trimis etapais.

Pirmoje darbo dalyje buvo atliekamas QSAR modeliavimas su 62 naujai susintetintais ketonais, kurių dauguma yra antiproliferaciškai aktyvūs prieš vėžines ląsteles-taikinius. Duomenys apie tuos junginius ir jų antiproliferacinį aktyvumą buvo gauti, bendradarbiaujant su VU Chemijos fakulteto ir VU Biochemijos instituto mokslininkais. Buvo sukurti QSAR modeliai, kiekybiškai aprašantys minėtų junginių antiproliferacinius aktyvumus. Generuotieji QSAR modeliai yra naudingi naujų, dar nesusintetintų, cheminių junginių, panašių į nagrinėjamą ketonų seriją, antiproliferaciniam aktyvumui prognozuoti, ir tai gali pasitarnauti, kuriant priešvėžinius vaistus.

Antroje darbo dalyje karboanhidrazės buvo pasirinktos skaičiuojamųjų metodų palyginimo tyrimui. Buvo pritaikytas QSAR metodas tam, kad būtų pasiūlyti pakeitimai sulfonamidų struktūrose, kurie pagerintų jų selektyvumą.

Buvo atlikta dvimačio QSAR ir trimačio QSAR analizė įvairioms CA izoformoms, naudojant keturiasdešimties slopiklių duomenų rinkinį. Naudojantis QSAR, buvo nagrinėti selektyvumai (aktyvumų santykiai) CA XII atžvilgiu. Taip pat buvo pritaikytas ir labiau tradicinis QSAR protokolas, kai modeliuojami slopiklių aktyvumai prieš atskiras CA izoformas. Šie du QSAR atlikimo protokolai buvo palyginti, ir su naujai sukurtu QSAR deskriptoriumi buvo pagerinta CA XII QSAR modelių statistika. Sukurtasis deskriptorius gali būti dar modifikuojamas, sukuriant daug kitų panašių į jį molekulinį deskriptorių.

Naudojant įvairius skaičiuojamuosius metodus, buvo atlikta palyginamoji cheminės struktūros ir biologinio aktyvumo studija, nagrinėjant 40 karboanhidazės slopiklių seriją. Darbui buvo pasirinkti dokimo, tiesinės sąveikos energijos (LIE), metadinamikos ir QSAR metodai. Paskaičiuoti jungimosi afiniškumai buvo palyginti su eksperimentinėmis vertėmis, siekiant išnagrinėti minėtų metodų privalumus ir trūkumus, kai jie taikomi tiriamai karboanhidrazės slopiklio sistemai. Iš išbandytų metodų, tik QSAR rodė ganėtinai teigiamus rezultatus. To priežastys išnagrinėtos ir aptartos. Taip pat buvo pasiūlyta LIE metodo modifikacija, kuri galutiniams skaičiavimams naudoja alternatyvią „LIE-like“ lygtį. Dėl to pavyko gauti gerokai geresnius rezultatus nei naudojant originalų LIE metodą.

Trečioje darbo dalyje buvo ieškomi nauji vėžio augimo slopikliai, naudojantis teorija, kad galima rasti naujų vėžio augimo slopiklių, ieškant panašių junginių į metabolitus, kurie yra substratai fermentų, aktyviai ekspresuojamų vėžinėse ląstelėse. Norimo cheminio junginio paieškai plačios apimties prieinamų cheminių junginių duomenų bazėse buvo pasitelktas Tanimoto cheminių struktūrų panašumo įvertis. Buvo įvertinta tikimybė rasti slopiklius, naudojantis Tanimoto įverčiu, nagrinėjant žinomų vaistų cheminės struktūros panašumus į žmogaus metabolitus. Galiausiai buvo pasiūlyti keli junginiai, galimai vėžio augimo slopiklių, poros iš kurių aktyvumas buvo vėliau patvirtintas eksperimentiniu keliu.

Darbo tikslas

Darbo tikslas buvo skaičiavimais paremta naujų biologiškai aktyvių junginių paieška ir tai paieškai skirtų metodų palyginimas, pritaikymas ir patobulinimas, ypatingą dėmesį skiriant QSAR.

Uždaviniai

1. Sukurti QSAR modelius serijai 62 α -pakeistų α,β -nesočiųjų ketonų, kurių dauguma antiproliferaciškai aktyvūs prieš NB4, MCF-7 ir A549 ląsteles-taikinius.
2. Palyginti dokavimo, tiesinės sąveikos energijos, metadinamikos ir QSAR metodus. Nustatyti tinkamiausią iš jų CA II slopiklių paieškai.
3. Patobulinti tiesinės sąveikos energijos metodą, kad jis būtų geriau pritaikytas CA II slopiklių paieškai.
4. Naudojant kelias metodikas, sukurti QSAR modelius prognozuoti 40 benzonsulfonamidinių slopiklių jungimosi konstantas ir selektyvumus įvairioms karboanhidrazės izoformoms, bei palyginti naudotų metodų efektyvumą.

Mokslinį naujumą pagrindžiantys darbo rezultatai

40 CA II slopiklių, benzonsulfonamidų darinių su prijungtu pirimidinio žiedu, struktūros ir pK_d verčių palyginamieji skaičiavimai buvo atlikti, naudojant dokavimo, LIE, metadinamikos ir QSAR metodais. Paskaičiuotosios pK_d vertės buvo palygintos su eksperimentiniais duomenimis, siekiant išnagrinėti visų panaudotų metodų privalumus ir trūkumus, juos taikant slopiklių, prisijungusių prie CA II, tyrimui. Naudojant QSAR, buvo gauti teigiami rezultatai (R^2 nuo 0,83 iki 0,89 tarp eksperimentiškai išmatuotų ir prognozuotų pK_d verčių). Buvo išnagrinėtos galimos metodų sėkmingumo ar nesėkmingumo priežastys. Buvo pasiūlytas LIE metodo patobulinimas. Taikant pasiūlytą LIE energijos skaičiavimo lygtį, metodo rezultatai žymiai pagerėjo (R^2 tarp eksperimentiškai išmatuotų ir prognozuotų ΔG_{bind} verčių padidėjo nuo

0,24 iki 0,50).

Naudojant gautus α -pakeistų α,β -nesočiujų ketonų antiproliferacinio aktyvumo duomenis prieš NB4, MCF-7 ir A549 ląsteles-taikinius, buvo sukurti įvairūs 5 deskriptorių QSAR modeliai tų junginių antiproliferaciniam aktyvumui paaiškinti ir prognozuoti.

Buvo įvertinta tikimybė rasti naujus slopiklius su pasiūlyta panašių į metabolitus slopiklių paieškos strategija, naudojant Tanimoto struktūrinio panašumo įvertį, ir toji strategija sėkmingai pritaikyta slopiklių paieškai.

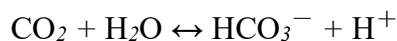
Ginamieji disertacijos teiginiai

1. Sukurtieji šeši α -pakeistų α,β -nesočiujų ketonų QSAR modeliai trims ląstelių linijoms (NB4, MCF-7 ir A549) gali atskirti aktyvius junginius nuo neaktyvių.
2. Iš penkių išbandytų metodų QSAR rezultatai geriausiai koreliuoja su eksperimentiniais duomenimis (R^2 intervale nuo 0,83 iki 0,89), taigi QSAR metodas geriausiai tinka prognozuoti benzensulfonamidų jungimosi prie CA II konstantas.
3. Įvedus papildomą kintamąjį (ligando van der Valso sąveiką su savimi vidurkį per visą molekulinės dinamikos simuliaciją), buvo patobulintas LIE metodas CA atvejui (R^2 padidėjo nuo 0,24 iki 0,50).
4. Kuriant specializuotus selektyvumo QSAR modelius, gaunami aukštesni modelių statistiniai įverčiai, nei prognozuojant slopiklių afiniškumus porai CA izoformų atskirai, ir paskui paskaičiuojant jų selektyvumą iš turimų prognozių.

1. Literatūros apžvalga

1.1. Karboanhidrazės: šeimos ir izoformos

CA įeina į metalofermentų šeimą [7] su cinko (II), geležies (II), kadmio (II) arba kobalto (II) jonu [8]. Cinko (II) jonas jose dažniausiai yra koordinuojamas trejomis baltymo histidinių imidazolų grupėmis ir vandens molekule ar hidrokso jonu. Karboanhidrazės jungiasi prie maždaug 30 vaistų ar preparatų klinikinių tyrimų stadijoje [6]. Šie fermentai katalizuoja grįžtamą anglies dioksido ir vandens reakciją, susidarant bikarbonato jonams ir protonams [5]:



Remiantis šia fermentine reakcija, buvo pasiūlytas karboanhidrazių pritaikymas, mažinant globalinio atšilimo padarinius bei gaminant naudingas medžiagas iš CO₂ [9]. CA II yra vienas iš aktyviausių žinomų katalizatorių (su $k_{\text{cat}}/K_m = 1.5 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$) [4].

Karbonato dehidratazės skirstomos į septynias šeimas, žymimas graikiškomis raidėmis α , β , γ , δ , ϵ , η , ζ , tarp kurių nėra nei struktūros, nei sekos panašumų [10]. Tai reiškia, kad dėl aukštos šio fermento fiziologinės svarbos jo katalizinis aktyvumas nepriklausomai kelis kartus susikūrė iš naujo įvairiuose gyvuose organizmuose. Nors yra daug CA formų, dabartinės žinios ir supratimas apie CA daugiausia remiasi keliais labiausiai tyrinėtiais α ir β šeimų atstovais, gryninamais iš žinduolių ir augalų. Neseniai buvo atrasta nauja šeštoji CA šeima, žymima η [11]. Galima rasti literatūrinių duomenų ir apie ζ -CA [12]. Priklausomai nuo CA šeimos, Zn(II), Cd(II) arba Fe(II) jonai yra būtini CA kataliziniam aktyvumui kofaktoriai, esantys CA fermentų aktyviuosiuose centruose. Iš bakterijų yra išgryninamos α , β ir γ CA [13]. Tos trys pagrindinės CA šeimos (α , β ir γ) plačiai paplitusios gyvuose organizmuose [14].

Iš visų CA šeimų α -CA yra labiausiai tyrinėta, nes ji yra plačiai paplitusi tarp gyvų organizmų, įskaitant žinduolius ir žmones [15]. Žmogaus CA

(žymima hCA) tyrinėjama dėl jos svarbos medicinoje ir sąveikos su vartojamais vaistais. hCA aktyvumas yra optimalus, esant fiziologiniam rūgštingumui (pH 7,34) ir fiziologinei temperatūrai (~37 °C) [16]. hCA slopikliai yra naudojami kaip vaistai jau kelis dešimtmečius. CA, priklausančios α šeimai, pasižymi nemažu kataliziniu lankstumu. Jos gali katalizuoti fermentines reakcijas ir kaip esterazės, naudojančios karboksilinių rūgščių, sulfoninių rūgščių ir fosforo rūgščių esterius kaip substratus. Kinetiniais, spektroskopiniais ir masių spektrometrijos tyrimais parodyta, kad jos dar pasižymi ir tioesteraziniu aktyvumu, veikdamos ditiokarbamato esterį (PhSO₂NHCSSMe) [17]. CA, priklausančių α -CA šeimai, aktyvusis centras yra kūgio formos įdubos, kuri tęsiasi nuo baltymo paviršiaus iki pat centro, gilumoje. Toji įduba yra maždaug 15 Å gylio ir 15 Å pločio savo plačiausioje vietoje [18]. CA, priklausančios β -CA šeimai, randamos prokariotuose, dumbliuose ir augaluose [19]. β -CA randamos daugumoje *Archaea* ir *Bacteria* domenams priklausančių organizmų rūšių ir tikriausiai visose grybų ir augalų rūšyse [20]. β -CA neegzistuoja žmoguje ar stuburiniuose gyvūnuose [21]. Pirmiausiai β -CA buvo išgrynintos ir ištirtos iš dviejų daugialąsčių organizmų. Syrjänen su bendraautoriais išgrynino vieną β -CA, CAH β (dar kitaip vadinama DmBCA, <http://flybase.org/reports/FBgn0037646.html>), iš vaisinės muselės *Drosophila melanogaster* [21]. Fasseas su bendraautoriais išgrynino β -CA iš nematodo *Caenorhabditis elegans*, kuris turi pora β -CA, iš kurių viena yra kataliziškai aktyvi [10]. β -CA tyrinėjamos dėl galimo praktinio pritaikymo kaip baltymai-taikiniai naujiems, stuburinių gyvūnų parazitus naikinantiesiems vaistiniams preparatams, kadangi β -CA paplitusios įvairiuose parazituose, bet jų nėra stuburiniuose gyvūnuose [11]. Selektyvūs β -CA slopikliai sutrikdytų tik parazitinių organizmų veiklą. Šitam teiginiui patvirtinti buvo dedami β -CA slopikliai į augimo terpę su gyvais *Leishmania* parazitais, kas iššaukė jų viduląstelines pažaidas ir žūtį [24]. Yra žinoma, kad β -CA yra reikalingos *Drosophila melanogaster* patelių vaisingumui [14]. CA, priklausančios γ -CA šeimai, randamos *Archaea* domenui priklausančių organizmų rūšyse [25]. Pagal savo katalizuojamos fermentinės reakcijos mechanizmą γ šeimos CA

panašios į α šeimos CA. Šių dviejų šeimų baltymų erdvinės struktūros visai nepanašios, neskaitant to, kad abejose šeimose cinko joną koordinuoja histidino liekanos [26]. Cam baltymas, gryninamas iš *Archaea* domeno organizmų, tyrinėjamas kaip modelinis γ -CA šeimos atstovas. Tai homotrimeras, kurio monomerų erdvinė struktūra yra nustatyta. Ji susideda iš paralelių spiralių, kurios sudarytos iš β -klosčių. Histidino liekanos iš gretimų monomerų koordinuoja tris metalo jonus, taip sudarydamos fermento aktyviųjų centrų pagrindus. Arti minėtų histidino liekanų yra kitos kataliziškai svarbios aminorūgščių liekanos. Jos formuoja vandenilinių ryšių tinklą, reikalingą reakcijos mechanizmui. Cam aktyviausia su geležies jonu, kuris, matyt, ten yra, esant fiziologinėms sąlygomis gyvame organizme [27]. CA, priklausančios δ ir ζ CA šeimoms, randamos titnaginiuose dumbliuose. Jos neabejotinai svarbios CO₂ įsisavinime, bet galbūt dalyvauja ir SiO₂ įsisavinimo cikle [28]. Pirmoji CA, priklausanti δ -CA šeimai, buvo atrasta Robert ir bendraautorių 1997 metais titnaginiuose dumbliuose, *Thalassiosira weissflogii* [29]. Kaip ir α bei γ šeimų CA, δ -CA aktyviajame centre esantis metalo jonas prisijungęs prie trijų histidino liekanų. Nustatyta δ -CA baltymo erdvinė struktūra visai kitokia, nei kitų šeimų CA. δ -CA aktyvusis centras yra kūgio formos įduboje, kuri apytiksliai tęsiasi nuo baltymo paviršiaus iki jo centro. Jame Cd(II) metalo jonas koordinuotas apytiksliai tetraedrine geometrija su dviem cisteino liekanomis, histidino liekana ir vandens molekule [30, 9]. ζ -CA gali veikti su Cd(II) ir Zn(II) jonais, prijungtais aktyviajame centre, visai neprarasdama savo katalizinio aktyvumo. Kai kurioms kitoms CA klasėms Cd(II) jonas yra veiklus slopiklis [30]. Kokią fiziologinę funkciją įvairių CA šeimų baltymai atlieka prokariotuose, nėra daug žinoma. Norint visiškai suprasti jų funkcijas, reikalingi papildomi tyrimai [9]. ϵ -CA randamos daugiausiai bakterijose, kai kuriuose chemolitotrofiniuose organizmuose, jūrų cianobakterijose, turinčiose karboksisomas, ir kai kuriuose titnaginiuose dumbliuose [31]. Anthony su bendradarbiais atrado, kad karboksisomų apvalkalinis baltymas, vadinamas CsoS3, kuris gryninamas iš *Halothiobacillus neapolitanus*, yra naujos ϵ šeimos CA, kuri labai skiriasi nuo kitų šeimų CA [32]. CA, priklausančių

nesenai atrastai η -CA šeimai, kaip α , γ ir δ šeimų CA, aktyviajame centre yra Zn(II) jonas, koordinuotas trimis histidino liekanomis ir vandens molekule. Jos rastos tik įvairiose maliariją sukeliančių patogenų *Plasmodium* rūšyse. η -CA yra gana artimai susijusios su α -CA šeima, bet yra ir skirtumų tarp jų. η -CA neturi protonų perdavimo tarpininko His64 ir poros prie aktyviojo centro esančių aminorūgščių (Glu106 ir Thr199), kurios visados būna α -CA šeimos baltymuose [11].

Žmonėse ekspresuojama α -CA šeimos 15 izoformų, iš kurių 12 yra kataliziškai aktyvios. Kataliziškai aktyvios žmogaus CA izoformos žymimos romėniškais skaičiais. Žmogaus kataliziškai neaktyvios CA vadinamos CARP (angl. carbonic anhydrase related proteins): CARP VIII, CARP X ir CARP XI [4]. Visos žmogaus CA (hCA) skiriasi savo aminorūgščių seka ir fermentiniu aktyvumu. Žinduolių CA izofermentai turi skirtingas kinetines savybes ir randami skirtinguose kūno audiniuose [33]. 16 katalizinių CA izoformų (žymimų simboliais: I, II, III, IV, VA, VB, VI, VII, VIII, IX, X, XI, XII, XIII, XIV, XV), atrastų žinduoliuose, atlieka svarbių funkcijų fiziologiniuose procesuose, tokiuose kaip kvėpavimas, CO₂/HCO₃⁻ transportas tarp metabolizuojančių audinių ir plaučių [34-36], elektrolitų sekrecija epiteliniuose audiniuose, pavyzdžiui, ašarų formavimosi procesas [37, 38], cerebrospinalinio skysčio sekrecija, kasos [34] ir skrandžio sulčių formavimasis [34, 39], šlapimo rūgšties gamyba ir sekrecija inkstų ląstelėse [34, 40] ir daug kitų [34]. CA I, II, III, VII ir XIII randamos ląstelių citozolyje, CA VA ir VB egzistuoja ląstelių mitochondrijose, CA VI sekretuojama ir įeina į žinduolių seilių sudėtį, CA IV, IX, XII, XIV ir XV aptinkamos membranose arba susijungusios su membrana [41].

CA slopikliai plačiai tyrinėjami, nes jie gali būti taikomi eilės ligų prevencijai ir gydymui [42-44, 6]. Įvairių sulfonamidų, susijungusių su CA, kristalinės struktūros parodo pagrindines tos klasės junginių sąveikas su metalofermentu. Cinko (II) jonas, esminis katalizinio aktyvumo komponentas, yra pačiame kūginės formos kišenės dugne. Sulfonamidai savo anijoninėje

formoje ($R-SO_2NH^-$) jungiasi su CA izoformomis, pakeisdami prie cinko katijono prisijungusią vandens molekulę, ir sutrikdo fermento katalizės reakcijos ciklą [45-49].

CA yra taikiny s keliems glaukomai gydyti plačiai naudojamiems vaistams, tokiems kaip acetazolamidas (N-(5-sulfamoil-1,3,4-tiadiazol-2-il)acetamidas), metazolamidas (N-[5-(aminosulfonil)-3-metil-1,3,4-tiadiazol-2(3H)-iliden]acetamidas) ir dichlorfenamidas (4,5-dichlorbenzen-1,3-disulfonamidas) [50]. Vienas iš didžiausių iššūkių, kuriant naujus medicininius preparatus, yra slopiklių, selektyvių svarbioms fermento izoformoms, sukūrimas [3]. Tokie slopikliai kaip vaistai tada būtų saviti ir tam tikrai ligos eigai. Sukurti efektyvūs žmogaus CA XII slopikliai galėtų būti panaudoti hipoksinių auglių gydyme [51]. Dėl CA svarbos daug skaičiuojamųjų cheminės informatikos metodų taikomi jos slopiklių jungimuisi paaiškinti ir prognozuoti [52-54]. Vienas iš plačiausiai atliekamų ir efektyviausių metodų CA slopiklių aktyvumų prognozavimui yra QSAR [55]. Supuran su bendraautorais parašytos apžvalgos ir knygos apibendrina QSAR tyrimus su įvairiais sulfonamidais, slopinančiais įvairias CA izoformas [56, 57]. Be sulfonamidinių CA slopiklių, dar yra žinomi nesulfonamidiniai CA slopikliai, kaip fenoliai, poliaminai, karboksirūgštys, kumarinai, sulfokumarinai, 2-tioksokumarinai, fulerenai [58]. CA II yra svarbus taikiny s skaičiuojamiesiems metodams bei geras skaičiavimų objektas tų metodų vystymui ir palyginimui, nes literatūroje ir duomenų bazėse yra daug eksperimentinių duomenų: kinetinių matavimų, jungimosi konstantų, kristalografinių struktūrų su prisijungusiais ligandais. Karboanhidrazės yra geros modelinės sistemos tirti baltymo-ligando sąveikai [59]. Mūsų grupė aktyviai dalyvauja naujų CA slopiklių paieškoje ir jų skaičiuojamojoje analizėje [60-62]. CA XII yra svarbus taikiny s vėžio gydyme [63]. Daugybėje rūšių hipoksinių auglių CA IX ir CA XII gerokai labiau ekspresuojamos, nei įprastai kituose audiniuose [64]. Šie du izofermentai sukelia auglio rūgštėjimą, to pasekoje metastazes, invaziją. Jų slopinimas sutrikdo vėžio vystymąsi, sukelia priešvėžinį efektą [65]. Prieš CA IX buvo

sukurtas „Girentuximab“ antikūnas, su kuriuo atliekami klinikiniai tyrimai inkstų ląstelių vėžiui gydyti [66]. Kaip galimi priešvėžiniai vaistai, reikalingi efektyvūs CA IX arba CA XII slopikliai, kurie kuo mažiau slopintų kitas CA izoformas, ypač CA I ir CA II, tam, kad būtų sumažintas nepageidaujamas šalutinis poveikis. Dėl šių priežasčių šiam tyrimui CA II ir CA XII buvo pasirinktos kaip pagrindiniai taikiniai, išbandant ir lyginant įvairius skaičiavimo metodus.

1.2 Dokinimo metodai ir programos

Mažų molekulių prisijungimo prie didelių baltymų-taikinių modeliavimas yra svarbus, įvertinant aktyvių junginių (ligandų) įtaką svarbiems biologiniams ir biocheminiams procesams. Molekulių dokinimas yra skaičiuojamosios chemijos metodas, kuris nagrinėja mažų molekulių konformacijas jų prisijungimo prie baltymų-taikinių vietose. Tikslus ligando prisijungimo prie baltymo prognozavimas, kuris yra dokinimo uždavinio sprendimas, yra didelės svarbos šiuolaikiniame, struktūra paremtame, biologiškai aktyvių junginių, vaistų kūrime [67]. Jei ligandas turi daug lanksčių cheminių ryšių, jis gali turėti daug įvairių konformacijų, iš kurių viena (retais atvejais kelios) nustatoma rentgenostruktūrinės analizės metodu, kaip eksperimentinė jungimosi konformacija. Dokinimo vertinimo funkcijos naudojamos dokinime, siekiant įvertinti, kurios iš nagrinėjamų konformacijų geriausiai atitinka prisijungimo vietą [68], bei įvertinant ligando ir taikinio susirišimo stiprumą [69]. Įvairiose molekulinio dokinimo programose ligando konformacijų generavimo, atrankos ir optimizavimo procedūra atliekama, pasitelkiant įvairius algoritmus. Tai gali būti genetinis algoritmas, „Monte Carlo“ simuliacija, simuluotas grūdinimas, atstumų geometrija paremti algoritmai (angl. *distance geometry*) ir kiti skaičiuojamieji metodai. Kiekvieną iš tokių konformacijų generavimo procedūrų link optimalaus rezultato turi vesti vertinimo funkcija. Norint teisingai išspręsti dokinimo uždavinį, būtini du dalykai: efektyvi konformacijų paieškos procedūra ir kokybiška vertinimo funkcija. Svarbi vertinimo funkcijos ir konformacijų paieškos procedūros

savybė yra jos greitis: jos abi turi būti pakankamai greitos, kad jas būtų galima taikyti, siekiant įvertinti didžiulį kiekį galimai teisingų ligando konformacijų ir padėčių baltymo atžvilgiu [70]. Jei dokinami keli ligandai, vertinimo funkcija ir ja paskaičiuoti dokinimo įverčiai, gali būti panaudojami ir reitinguoti tuos ligandus tarpusavyje pagal tai, kaip gerai jie turėtų jungtis prie baltymo. Minėti įverčiai ligandų rinkiniui turi koreliuoti su eksperimentiškai išmatuotu ligando afiniškumu baltymui [71].

Molekulių dokinimo sritis egzistuoja ir plėtojasi nuo savo atsiradimo jau kelis dešimtmečius. 1988 metais buvo sukurta pirmoji kompiuterinė dokinimo programa DOCK [72]. Su šiuolaikinėmis populiariomis dokinimo programomis galima pakankamai greitai ir tiksliai prognozuoti ligandų-baltymų kompleksų struktūras [68]. Plačiai naudojamos dokinimo programos yra DOCK, [73] AUTODOCK, [74] FLEXX, [75] GOLD, [76] POSIT, [77] FRED [78-80] ir daug kitų. DOCK (<http://dock.compbio.ucsf.edu/>) dokinimo programoje ligando konformacijos paieškos strategija yra laipsniškas nagrinėjamos ligando struktūros rekonstravimas iš funkcinių grupių baltymo aktyviajame centre bei atsitiktinė konformacijų paieška. Joje paprastai skaičiuojama vertinimo funkcija, paremta Kulono ir Lenardo-Džonso potencialais bei kai kuriomis kitomis papildomomis funkcijoms. Pavyzdžiui, BUMP FILTER vertinimo funkcija skaičiuoja persidengiančius Van der Valso tūrius tarp ligando ir baltymo ir tokiu būdu leidžia iš karto atmesti prastus rezultatus, o CONTACT SCORE skaičiuoja ligando ir baltymo atomus, kurie yra arčiau nei tam tikras ribinis atstumas, t.y., liečiasi vienas su kitu [73, 81]. AUTODOCK (<http://autodock.scripps.edu/>) kūrėjai ligando konformacijos paieškai taiko Lamarko genetinį modelį, kuriame individo fenotipo adaptacijos prie aplinkos sąlygų yra atgaliniu ryšiu įtraukiamos į genotipą ir tampa paveldimomis savybėmis. Simuliuojamas genotipas šiuo atveju yra ligando konformacija, kuri turi „evoliucionuoti“ kompiuteriniame modeliavime į optimalią. Be Lamarko genetinėmis teorijomis paremto ligando konformacijos paieškos algoritmo, AUTODOCK gali naudoti tradicinį genetinį algoritmą bei

pagal „Monte Carlo“ teoriją simuluotą grūdinimą. AUTODOCK laisvosios energijos vertinimo funkcija yra empiriškai parametrizuota, kalibruojant trisdešimt eksperimentiškai nustatytų ligando-baltymo susijungimo erdvinių struktūrų su išmatuotomis afiniškumo konstantomis tiems kompleksams [74]. FLEXX (<https://www.biosolveit.de/LeadIT/>) dokimo programa (LEADIT programinio paketo dalis, BioSolveIT GmbH) suderina dokimamų ligandų atitinkamą fizikinių ir cheminių savybių modelį su efektyviais jų konformacijų paieškos metodais [75]. GOLD (<https://www.ccdc.cam.ac.uk/solutions/csd-discovery/components/gold/>) (angl. *Genetic Optimisation for Ligand Docking*) yra automatizuota ligandų dokimo programa, kuri prisijungimo konformacijų paieškai taiko tradicinį genetinį algoritimą. Atliekama ir ribota baltymo aktyviojo centro pasirinktų fragmentų konformacinė paieška. Priešingai nei daugumoje kitų dokimo programų, yra atsižvelgiama į tai, kad ligandas turi pakeisti vandenį savo prisijungimo vietoje [76]. Išsamus baltymo struktūrinių pakitimų, kai prie jo jungiasi ligandas, nagrinėjimas pareikalautų daug skaičiuojamųjų resursų, tad dauguma dokimo programų laikosi prielaidos, kad baltymo struktūra nekinta, prie jos prijungiant mažos molekulinės masės molekulei. OEDOCKING programinio paketo dalis (<http://www.eyesopen.com/oedocking>) dokimo įrankis FRED (OpenEye Scientific Software, Inc.), turi ne tik tradicinį dokimo režimą, bet ir hibridinį režimą, kuriame be baltymo erdvinės struktūros, kaip įprastai, nagrinėja ir kitas eksperimentiškai nustatytas prisijungusio ligando struktūras dokimo uždaviniui tiksliau išspręsti [79]. FRED 70% ligandų struktūrų dokina 2 Å tikslumu atliktame dokimo kokybės teste [82]. POSIT yra kitas OEDOCKING programinio paketo dokimo įrankis. Jis sukurtas, laikantis principo, kad dokimo prognozės kokybė yra susijusi su dokimo programos vertinimo funkcijai parametrizuoti naudotų eksperimentinių duomenų apie prie baltymų prisijungusius ligandus kokybę ir prieinamumu. POSIT dar įtraukia informaciją apie geometrines formas ir cheminius panašumus, kad geriau išnaudotų informaciją apie kristalografines ligandų konformacijas. Taikinio struktūra ir eksperimentiškai nustatytos ligando konformacijos taikinio atžvilgiu POSIT

dokinime naudojamos minimizuojant ligandą. Tai vadinama formos valdomu ligando minizavimu (angl. *shape-guided ligand minimization*). Visa tai leidžia geriau prognozuoti ligando prisijungimo konformaciją [77].

Vienas iš dažnai naudojamų dokinimo metodų yra virtualioji junginių atranka (angl. *virtual screening*) [83]. Dokinimo pagalba iš egzistuojančių arba planuojamų sintetinti junginių duomenų bazių yra atrenkami ligandai su geriausiu prognozuojamu afiniškumu baltymo-taikinio atžvilgiu. Eksperimentiškai išbandžius atrinktuosius junginius, lengviau ir pigiau randami perspektyvūs slopikliai, turintys aukštą afiniškumą norimam biologiniam taikiniui. Dokinimo programų tinkamumas virtualiajai junginių atrakai remiasi dokinimo greičiu ir tikslumu. Kadangi dokinimo programų tikslumas visada yra ribotas, paprastai eksperimentiniam aktyvumo patikrinimui yra paimamas didesnis kiekis molekulių-kandidatų, siekiant, kad nebūtų per klaidą išmetami teisingai teigiami (angl. *true positive*) ligandai, nes kai kada jie pasižymi žemesnėmis vertinimo funkcijos reikšmėmis, palyginti su geriausius įvertinimus gavusiais klaidingai teigiamais (angl. *false positive*) junginiais. Yra nemažai apžvalgų, kuriose lyginamos įvairios dokinimo programos, jų versijos, ir dokinimo algoritmai [84-86] bei lyginama, kaip su dokinimo programomis atkartojami eksperimentiniai rezultatai [87-89]. Kontoyianni ir bendraautoriai išbandė FLEXX, DOCK, GOLD, LIGANDFIT ir GLIDE programas, iš kurių su GOLD ir GLIDE geriausiai prognozavo ligandų prisijungimo konformacijas ir padėtis baltymo atžvilgiu [84]. 2004 metais Kellenberger ir bendraautoriai išbandė kelias dokinimo programas. Geriausiai iš jų savo uždavinį atliko GLIDE, GOLD ir SURFLEX programos [85]. 2005 metais ta pati Kontoyianni ir bendraautorių grupė vėl bandė dokinimo programas ir, išnagrinėję rezultatus, priėjo išvados, kad geriausia programa yra LIGANDFIT, joje naudojant LIGSCORE1 arba GOLD vertinimo funkcijas. Minėtos programos ir vertinimo funkcijos sugebėjo atrinkti daugiausiai biologiškai aktyvių junginių iš aktyvių ir neaktyvių junginių mišinio [86]. 2015 metais Jug ir bendraautoriai palygino AUTODOCK, FRED, CDOCKER, FLEXX

ir GOLD programos ir kritiškai jas įvertino. Iš jų FRED pasirodė esanti geriausia [90]. Palyginimo rezultatų įvairovė paaiškintina tuo, kad skirtingos programos skirtingai gerai veikia su skirtingais taikiniai ir įvairiais ligandais. 2015 metais parašyta mokslinių straipsnių apžvalga pateikia medžiagos apie tai, kad molekulinio dokavimo rezultatų pritaikomumas gali būti abejotinas, nepaisant aukštų pačios dokavimo programos pateiktų įverčių, nes net jei ligandas jungiasi į taikinį, nėra aišku ar jis bus slopiklis, ar aktyvatorius [91].

Mažų organinių junginių dokavimas buvo taip pat aliekamas ir į karboanhidrazės receptorių, bet neretai buvo susiduriama su įvairiais sunkumais, kurie didele dalimi susiję su daugelio dokavimo programų nesugebėjimu teisingai sumodeliuoti donorių-akceptorinių ryšių tarp cinko ir ligando bei dėl to, kad nėra teisingų torsinių kampų parametrų tarp sulfonamido grupės ir likusio ligando [92, 93]. Dėl to gali tekti naudoti ligando suvaržymus [60, 61]. Be to, Xiang ir bendraautorai atliko skaičiuojamuosius tyrimus, kuriuose rankiniu būdu išrinko teisingas sudokintų ligandų konformacijas tolesniems tyrimams, kad pataisytų AUTODOCK VINA programoje esančios automatinės ligandų konformacijų vertinimo funkcijos trūkumus [94].

Su dokavimo programomis paprastai bandoma suskaičiuoti bendrą indėlį visų energetinių efektų, kurie atsiranda dėl suformavimo ir nutraukimo įvairių cheminių sąveikų tarp baltymo, ligando ir tirpiklio (įprastai vandens). Kita vertus, įvairiose programose su tirpikliu susiję efektai dažnai lieka ignoruojami. Pirmieji dokavimo metodą pasiūlė Blaney ir Kuntz [95]. Jų metodo versija apsiribojo tik fiksuotomis baltymo ir ligando konformacijomis. Nei ligando, nei baltymo struktūrų optimalių konformacijų nebuvo ieškoma. Vėliau šis metodo trūkumas buvo pašalintas, realizuota optimalesnės ligando konformacijos paieška tarp jo įvairių konformacijų [96-98]. Programiškai buvo įgyvendinta galimybė atlikti dalinę taikinio konformacinę paiešką, randant konformaciją, kuri geriausiai prisijungia ligandą [99]. Tokia paieška apsiriboja keliomis aminorūgščių liekanomis, kurios gali laisvai judėti, keisti

konformaciją dokinimo metu. Tas aminorūgščių liekanas iš visų esamų pasirenka dokinimo programos naudotojas kaip galimai keičiančias konformaciją ligando prisijungimo metu. Dokinimas į visiškai konformaciškai judrų taikinį vis dar niekad neatliekamas [96]. Tai skaičiuojamuoju požiūriu sudėtinga problema, mat taikynys-baltymas gali turėti šimtus ar net tūkstančius konformacijas galinčių keisti cheminių ryšių.

Dokinimas buvo atliekamas su CA II ligandais ir ankščiau [100-102], bet arba palyginimas su eksperimentiniais ligandų sąveikos energijos nebuvo atliekamas [103], arba dokinimo rezultatai buvo atidėti lentelėje kartu su eksperimentiniais duomenimis, bet jų statistinė analizė nebuvo atlikta [104], arba buvo publikuota prasta koreliacija tarp paskaičiuotos ligandų jungimosi laisvosios energijos (ΔG) ir jų eksperimentinių CA II slopinimo konstantų [105].

1.3 Ligandų jungimosi laisvosios energijos skaičiavimai tiesinės sąveikos energijos metodu

Tiesinė sąveikos energija (angl. *linear interaction energy*, LIE) yra alternatyvus metodas prognozuoti ΔG vertes [106]. Šiame metode, yra atliekami du atskiri molekulinės dinamikos (MD) skaičiavimai: vienas su ligandu jo prisijungimo vietoje taikinyje-baltyme, ir kitas su atskiru ligandu tirpiklyje (vandenyje). Tada ΔG_{bind} aproksimuojama kaip vidutinis skirtumas tarp dviejų sąveikų dviem ligando būsenoms:

$$\Delta G_{bind} = \alpha(\langle V_{el} \rangle_{bond} - \langle V_{el} \rangle_{free}) + \beta(\langle V_{vdw} \rangle_{bond} - \langle V_{vdw} \rangle_{free}) + \gamma. \quad (1.3-1)$$

Šioje lygtyje $\langle V_{el} \rangle$ ir $\langle V_{vdw} \rangle$ yra vidurkiai elektrostatiinių ir van der Valso sąveikų tarp ligando ir jo aplinkos (vandens arba taikinio ir vandens, priklausomai nuo skaičiavimo) per visą molekulinės dinamikos simuliaciją. Apatinis indeksas „free“ žymi sąveiką, kai ligandas yra tirpiklyje, o apatinis indeksas „bond“ žymi sąveiką, kai ligandas susijungęs su taikiniu. Šie skirtumai dauginami iš empirinių koeficientų α ir β , ir dar pridedamas pastovius laisvosios energijos dedamasis γ . Literatūroje buvo paskelbti įvairių šių trijų koeficientų rinkiniai LIE metodui. Galiausiai prieita išvados, kad tie

koeficientai priklauso nuo MD simuliacijoje naudojamo jėgos lauko ir nuo nagrinėjamos baltymo-ligando sistemos ypatybių [107-112]. Iki šiol literatūroje nebuvo duomenų apie sėkmingą LIE metodo pritaikymą CA. Kosugi et al. pastebėjo, kad tik keli LIE metodo taikymai aprašyti metalofermentams, nes nėra sukurto bendro jėgos lauko, kuris aprašytų tiek baltymą, tiek jo koordinacinius ryšius su metalo jonais [113]. Kita vertus, LIE yra kiek sėkmingai naudojama kitoms ligando-baltymo sistemoms tirti [114-118]. Pagerintas hibridinis kvantinės mechanikos ir molekulinės mechanikos (QM/MM) LIE modelis yra pasiūlytas slopiklių afiniškumo prognozavimui cinko joną turinčiuose metalofermentuose [115].

1.4. Ligandų laisvosios energijos skaičiavimai metadinamikos metodu

Viena iš molekulinės dinamikos problemų yra pernelyg trumpa simuliacijos trukmė. Dauguma biologiškai svarbių biomakromolekulių konformacijų pokyčių trunka kur kas ilgiau, negu išgalima modeliuoti su esamais kompiuteriniais resursais. Viena išiečių būtų naudoti nepusiausvyrinius molekulinės dinamikos metodu (angl. *non-equilibrium molecular dynamics*). Tarp nepusiausvyrinių MD metodų yra paminėtini pagreitinta molekulinė dinamika (angl. *accelerated MD*) [119], metadinamika [120], piltuvo metadinamika (angl. *funnel metadynamics*) [121], λ -metadinamika [122], altruistinė metadinamika [123] ir kiti. Iš jų, metadinamika buvo pasirinkta kaip dar vienas metodas ligandų prisijungimo afiniškumui prie CA prognozuoti. Metadinamika įgalina sumažinti kompiuterinių resursų reikalavimus iki praktiškai prieinamų [124]. Metadinamika leidžia rekonstruoti laisvosios energijos paviršių kaip funkciją nuo pasirinktų simuliuojamos sistemos kolektyvinių kintamųjų (CV) verčių. Standartinėje metadinamikoje yra įtraukiamas papildomas (angl. *bias*) potencialas - nedidelis Gauso tipo funkcijos kalnelis, kuris yra pastoviai pridedamas simuliacijos metu aplankytuose taškuose CV erdvėje. Sistema nuolat skatinama keliauti po naujus aukštos laisvosios energijos regionus, nes likti toje pačioje žemos energijos būsenoje jai trukdo vis augantis papildomas potencialas [124]. Tokiu

būdu, ilgainiui žemiausios potencialinės energijos taškai CV erdvėje yra užpildomi papildomo potencialo funkcijoms, kas leidžia nustatyti tikslų potencialinės energijos paviršių. Šiuo papildomu potencialu pokyčiai MD simuliacijoje pagreitinami, ir reti įvykiai įvyksta per palyginti trumpą laiką, nereikalaujant gausių kompiuterinių resursų. Metadinamika, be kitų taikymų, naudojama tirti ligandų prisijungimą ir indukuotus konformacinius pokyčius [125]. Kita vertus, metadinamika gana sėkmingai taikoma kitoms ligando-baltymo sistemoms tirti [126-129]. Iki šiol nebuvo aprašytas sėkmingas metadinamikos atlikimas su CA.

1.5. Kiekybinio struktūros ir aktyvumo ryšio samprata ir metodai

1.5.1. QSAR – kas tai?

Kitas metodas, tinkamas ligandų jungimosi prie baltymų afiniškumui ir savitumui prognozuoti – QSAR. QSAR tyrimų tikslas – atskleisti, kaip nustatyti cheminės medžiagos gebėjimą sukelti tam tikrą poveikį pagal jos cheminę struktūrą, taikant analogiją su kitomis jau žinomų struktūrų ir aktyvumų medžiagomis [130, 131]. Praktikoje QSAR yra matematiniai modeliai, naudojami cheminių medžiagų biologiniams aktyvumams prognozuoti pagal jų molekulinę struktūrą. Tai klasikinis, plačiai taikomas cheminės informatikos metodas, dažnai naudojamas kaip vienas iš pagrindinių skaičiuojamosios chemijos metodų ligandų jungimosi ar slopinimo aktyvumo eksperimentiniams duomenims racionalizuoti ir prognozuoti [132]. Tarkime, norint nuodugniai išnagrinėti cheminių junginių seriją, kurioje yra keturios galimos pakaitų vietos, kiekvienoje iš jų – dešimt galimų funkcinių grupių, reikia žinoti, jog tokia serija iš viso susidarys iš dešimties tūkstančių cheminių junginių. Visus tuos junginius susintetinti ir išmatuoti jų norimas savybes reiktų daug laiko ir išteklių. Išėitis – susintetinti mažą dalį galimų junginių, ir pagal gautus duomenis nustatyti empirines taisykles, kurias galima būtų panaudoti likusių tos serijos cheminių junginių savybėms prognozuoti. Paprastai tam pasitelkiamas QSAR. Kiekybinis struktūros ir aktyvumo ryšys, visų pirma, grindžiamas bendraisiais medicininės chemijos principais, pagal

kuriuos teigiama, kad skirtumai cheminio junginio struktūros bruožuose nulemia to cheminio junginio biologinį aktyvumą. Biologinis cheminio junginio aktyvumas susijęs su to junginio chemine struktūra ar fizikinėmis-cheminėmis savybėmis, be to, panašios cheminės struktūros molekulės dažnokai linkusios jungtis prie savo biologinio taikinio panašiai, todėl gali pasižymėti panašiu biologiniu aktyvumu [133]. Laikomasi ir principo, kad ne kiekvienas cheminės struktūros fragmentas vienodai svarbus galutiniam, bendram biologiniam aktyvumui: kai kurie jos fragmentai gali turėti mažą, artimą nuliui arba neigiamą įtaką. Reikia atsižvelgti į tai, kad kai kurie cheminės struktūros fragmentai gali daryti skirtingą įtaką biologiniam aktyvumui priklausomai nuo to, ar yra kitų svarbių fragmentų toje struktūroje, ir kur jie yra. Tam tikras į cheminę struktūrą įtrauktas fragmentas gali sukelti konformacinius pokyčius, po kurių kardinaliai keičiasi seniau nagrinėtų fragmentų įtaka, mat po konformacinių pokyčių jie jau sąveikauja su visai kitomis receptoriaus (taikinio) vietomis. Žinant, kad nagrinėjamoje serijoje skirtingi junginiai biologiškai aktyvūs dėl kiek skirtingų struktūrinių savybių, jas išsiaiškinus ir išnagrinėjus, galiausiai galima sukombinuoti naują, aktyvesnę cheminę struktūrą su keliais biologiniam aktyvumui teigiamą įtaką darančiais fragmentais ir pašalinti vien neigiamą įtaką darančius fragmentus.

Atliekant QSAR, informacija apie molekulės cheminę struktūrą pirmiausia matematiškai išreiškiama molekuliniais deskriptoriais. Junginio fizikinės ir cheminės savybės irgi gali būti panaudojamos kaip molekuliniai deskriptoriai. Įvedus į QSAR deskriptorių skaičiavimo programą junginio cheminę struktūrą, įvairių algoritmų pagalba sugeneruojamas rinkinys cheminių ir fizikinių deskriptorių. Po to iš visos molekulinų deskriptorių aibės paprastai išrenkami tie, kurie gerai reprezentuoja nagrinėjamą junginių biologinį aktyvumą. Kuriant šiuolaikinį QSAR modelį, išbandomi šimtai ir tūkstančiai deskriptorių. Po to nustatoma statistinė koreliacija, kuri randama tarp deskriptorių ir eksperimentinių duomenų, pavyzdžiui, tų junginių jungimosi konstantų. Galiausiai sudaromas QSAR modelis ar QSAR lygtis,

apibrėžianti matematinį sąryšį tarp molekulinės struktūros, užkoduotos molekuliniiais deskriptoriais, ir žinomų cheminių junginių biologinių aktyvumų. Paprastai priimamas daugybinės tiesinės regresijos (MLR) matematinis modelis tarp deskriptorių ir junginių aktyvumo logaritmo. Išmėginti visus įmanomus QSAR modelius, naudojant po kelis iš tūkstančių galbūt tinkančių deskriptorių ir iš jų išrinkti geriausią, yra labai neefektyvus ir nepraktiškas QSAR atlikimo būdas, nors tai ir būtų idealus QSAR modelis pagal turimus pradinis duomenis. Norint rasti pakankamai gerą QSAR modelį, kurį praktiškai įmanoma sukonstruoti su turimais kompiuteriniais ištekliais, naudojami įvairūs artiniai.

QSAR modelis yra statistiškai reikšmingas matematinis modelis, chemometriniu būdu apibūdinantis cheminių junginių afiniškumo biologiniam taikiniui, slopinimo konstantų, kinetinių konstantų ar kitaip išreikštų biologinių sąveikų verčių ir jų struktūrinių-geometrinių ir fizikinių-cheminių savybių koreliaciją. Pati bendriausia QSAR modelio formulė gali būti užrašoma taip:

biologinis aktyvumas = funkcija (*molekulės ar jos fragmentų savybės*).

(1.5.1-1)

Tokio ryšio nustatymas gali būti naudingas pats savaime, aprašant junginių funkcines grupes ir kitus struktūrinius bruožus, nulemiančius tų junginių biologinį aktyvumą, arba gali būti naudojamas naujų cheminių junginių nežinomų biologinių savybių prognozėms. Be biologinio cheminių junginio aktyvumo prognozavimo, QSAR modeliai plačiai taikomi cheminių junginių, kompozitinių medžiagų ir pan. farmakokinetikai, toksiškumui, biotransformacijai, poveikiui ekologinėms sistemoms ir žmogaus sveikatai prognozuoti. Cheminėje informatikoje plačiai naudojamas kitas analogiškas metodas – kiekybinis struktūros ir savybių ryšys (QSPR). Pagal jį iš nagrinėjamų cheminių junginių molekulinį deskriptorių modeliuojamos tų junginių cheminės savybės, tokios kaip tirpumas vandenyje ar kituose

tirpikliuose, stabilumas, rūgštingumo konstanta, adsorbcija ant įvairių paviršių, ir t.t. Jei QSAR modeliavimas taikomas nebiologinėms sistemoms, tuomet jis vadinamas QSPR, nors visomis kitomis savybėmis metodas išlieka panašus arba identiškas. QSAR arba QSPR modeliavimui naudojami regresijos, klasifikacijos matematiniai modeliai ar sistemos mokymosi (angl. *machine learning*) metodai. QSAR (QSPR) modeliavimas plačiai taikomas akademinėse, pramoninės gamybos ir administravimo srityse visame pasaulyje [134]. Vis dažniau pripažįstama, kad šis metodas padeda efektyviai optimizuoti turimą cheminių junginių, veikiančių norimą biologinį taikinį [135], nors negalima visiškai ignoruoti bandymų ir klaidų faktoriaus. QSAR metodo taikymas sumažina skaičių cheminių junginių kandidatų, kuriuos reikia susintetinti ir išmatuoti, norint padidinti reikiamą biologinį aktyvumą [136].

1.5.2. QSAR istorija

Crum-Brown ir Faster dar 1868 metais rašė, kad cheminių junginių struktūra ir elementinė sudėtis susijusi su jų fiziologiniu aktyvumu, tačiau neįvedė jokio matematinio sąryšio tarp šių dalykų [137]. Richardson tais pačiais metais išreiškė cheminių junginių struktūros modelį kaip tų junginių tirpumo matematinę funkciją [138]. Mills 1884 metais sukūrė QSPR modelį, pagal kurį gana tiksliai prognozavo savo nagrinėjamų cheminių junginių serijų lydymosi ir virimo temperatūras [139]. Po beveik dešimtmečio, 1893 metais, Richet pasirinko nagrinėti biologines sistemas ir alkoholių, eterių, ketonų seriją. Richet matematiniais skaičiavimais parodė, kad jo nagrinėjamos serijos junginių toksiškumas yra susijęs su jų tirpumu vandenyje: kuo tos serijos junginių tirpumas vandenyje mažesnis, tuo jų toksiškumas didesnis [140]. Vėliau Overton ir Mayer nepriklausomai vienas nuo kito pastebėjo koreliacijas tarp cheminių junginių pasiskirstymo koeficientų ir jų anestetinių savybių stiprumo. Jie susiejo nagrinėjamų junginių anestetines savybes su tų junginių lipofiliškumu [141, 142]. Tokie buvo QSAR (QSPR) vystymuisi svarbūs tyrimai XIX a.

1935 ir 1937 metais Hammett paskelbė savo garsiąsias lygtis, kurios

aprašo tiesinį laisvosios energijos sąryšį tarp serijos benzenkarboksirūgšties darinių pusiausvyros konstantos ar reakcijos greičių ir σ ir ρ (Hammett konstantų) sandaugos, kur ρ nurodo reakcijos tipą, o σ – pakaito tipą ir jo vietą [143, 144]. Hammett konstantos toliau buvo naudojamos QSAR modeliavime. 1942 metais Bell ir Roblin nustatė, kad antibakterinis sulfonamidų aktyvumas yra susijęs su jų jonizacija [145]. Albert 1948 metais tyrinėjo erdvinio prieinamumo cheminėje struktūroje, jonizacijos ir elektronų pasiskirstymo efektus tam tikrai aminoakridinų serijai [146]. Taft 1952 metais pasiūlė metodą, kaip atskirti polinius, rezonansinius ir sterinius efektus. Jis įvedė pirmąjį sterinį parametą, dažnai vadinamą Taft vardu [147]. Nepaisant šių fundamentalių tyrimų, dažniausiai teigiama, kad pirmasis mokslinis straipsnis apie QSAR metodą pasirodė tik 1962 ar 1964 metais. QSAR metodo pradininku laikomas Hansch, kuris 1962 metais kartu su bendradarbiais susiejo fenoksiacto rūgščių biologinį aktyvumą su Hammett konstantomis ir pasiskirstymo koeficientais [148]. 1964 metais Hansch kartu su Fujita susiejo augalų augimo reguliatorių aktyvumą (C), Hammett konstantas (σ), pasiskirstymo koeficientą (π) ir sterinį faktorių (k) viena sumine lygtimi [148]:

$$\log 1/C = a\sigma + b\pi + ck. \quad (1.5.2-1)$$

Tais pačiais metais Hansch su Fujita pademonstravo, kad panašių cheminių junginių serijos biologinis aktyvumas gali būti prognozuojamas lygtimi iš Hammett (ϵ), Taft (S) ir hidrofobiškumo parametru (π) [149]:

$$\log 1/C_{50} = a\pi + b\epsilon + cS + d. \quad (1.5.2-2)$$

Tiesinės su keliais kintamaisiais matematinės funkcijos kaip ši, pasiūlyta dar 1964 metais, tebėra klasikiniai QSAR modeliai. Šiais laikais kuriamo QSAR modelio kintamuosius galima rinktis iš didelio skaičiaus deskriptorių, o ne vien iš kelių vardinių parametru. Nuo to laiko, kai QSAR metodas buvo išplėtotas į šiuolaikinį metodą Hansch ir jo bendradarbių, praėjo daugiau kaip penkiasdešimt metų. QSAR dar kartais vadinamas Hansch-Fujita

analize. Tuo metu vyravo struktūros ir savybių ryšio (SAR) metodas, o prognozavimas neturėjo didelės svarbos. Pirmosiose QSAR koreliacijose vietoje deskriptorių buvo naudojami vos keli fizikiniai ir cheminiai parametrai, tokie kaip $\log(P)$ ar π arba kitaip išreikštas lipofiliškumas, be to, junginio elektroninė sandara, išreikšta kaip σ , sterinės savybės, poliarizuojamumas ir panašiai. Beveik tuo pačiu metu, 1964 metais, buvo pasiūlyta labai panaši Free-Wilson analizė. Wilson su Free, kiek kitaip nei Hansch su bendradarbiais, siūlė tam tikrų funkcinių grupių junginyje buvimą arba nebuvimą paprasčiausiai vertinti vienetu arba nuliu, tada dauginti iš empirinių kontribucijos koeficientų, sumuoti, galiausiai pridėti bazinį (vidutinį) aktyvumą ir taip apskaičiuoti biologinį aktyvumą iš cheminio junginio struktūros [150]:

$$\text{Biologinis aktyvumas} = \sum a_i x_i + b, \quad (1.5.2-3)$$

kur b yra vidutinis serijos junginių indėlis į biologinį aktyvumą, x_i – tam tikrų funkcinių grupių ar struktūrinių fragmentų buvimas arba nebuvimas, įvertintas nuliu arba vienetu, a_i – empirinių parametru, parodančių funkcinės grupės indėlio svarbą biologiniam aktyvumui, aibė.

1969 metais Hansch pasiūlė parabolinę lygtį, kurioje naudojamas kvadratu pakeltas pasiskirstymo koeficientas, turintis neigiamą įtaką biologiniam aktyvumui [151]. Toks biologinio aktyvumo modeliavimas iš fizikinių-cheminių junginio parametru vis labiau priminė dabartinę QSAR, kuriame kartu naudojami įvairiais būdais perskaičiuoti molekuliniai deskriptoriai. 1971 metais Fujita su Ban postulavo Fujita-Ban lygtį, panašią į Free-Wilson lygtį, tik joje naudojo biologinio aktyvumo logaritmą vietoje biologinio aktyvumo [152]. Tai leido sujungti Free-Wilson ir Hansch-Fujita metodus į vieną metodą su bendra deskriptorių aibe – tai ir yra klasikinė QSAR metodologija. Pagal Hansch ir kitus pirmuosius QSAR metodo autorius ir taikytojus, deskriptorius reikėtų parinkti iš gana nedidelio jų rinkinio rankine atranka, jau žinant, kurie struktūriniai cheminio junginio bruožai svarbūs, pagal

biologinio aktyvumo mechanizmą priskiriant aiškia prasmę cheminio junginio deskriptoriams, kurie bus toliau naudojami kuriamame QSAR modelyje. Vėliau buvo sukurtas sprendimas iš didelio deskriptorių rinkinio geriausias deskriptorių kombinacijas parinkti statistiniais metodais ir paskui statistiškai ir automatiškai parinktoje deskriptorių aibėje ieškoti prasmės, kaip priešprieša rankinei deskriptorių atrankai. Iki šiol naudojami abu deskriptorių atrankos protokolai, bet, ko gero, vyrauja antrasis.

Visuose tradiciniuose, pirmuosiuose dvimačiuose QSAR modeliuose, tokiuose kaip Free-Wilson ir Hansch-Fujita, kiekybinėms prognozėms apie nagrinėjamus junginius atlikti buvo naudojami dvimačiai molekulių fragmentai ar pakaitai ir jų fizikinės-cheminės savybės. Nuo to laiko, kai 1962–1964 metais Free, Wilson, Hansch ir Fujita [148, 153, 150] išpopuliarino ir išplėtojo originalią QSAR koncepciją, buvo pasiūlytas platus spektras įvairių QSAR metodologijų [154, 155]. QSAR plačiai buvo taikytas kaip vertinga priemonė cheminių junginių savybių prognozėms, ypač kuriant vaistus ir ūkinės reikšmės chemines medžiagas – pesticidus, herbicidus, fungicidus ir kitas panašias. Metodui palaipsniui vystantis, XX amžiaus paskutiniame dešimtmetyje mokslinių straipsnių apie QSAR (QSPR) skaičius pasiekė šimtus per metus. Dabar tokių straipsnių parašoma daugiau kaip tūkstantis per metus. Metodo populiarumo vertinimai pagal paieškas duomenų bazėse nėra tikslūs, mat QSAR pavirto ištisa metodų su daug skirtingų pavadinimų grupe. Vien pagal tikslią metodo taikymo paskirtį, be jau minėtų QSAR atšakų, egzistuoja dar toksiškumo QSAR (QSTR) [156], metabolizmo QSAR (QSMR) [157], reakcingumo QSAR (QSRR), sulaikymo QSAR (QSRR) [158], pereinamumo QSAR (QSPR), farmakokinetikos QSAR (QSPR) [159], bioprieinamumo QSAR (QSPR), prisijungimo QSAR (QSPR) [160] ir kitos atšakos.

1.5.3. Šiuolaikinis QSAR

QSAR (QSPR) šiuo metu yra vienas iš dažniausiai taikomų metodų, modeliuojant fizikines ir chemines įvairių cheminių junginių savybes ir jų biologinį aktyvumą. QSAR tyrimai laikomi viena iš pagrindinių vaistų kūrimo

srities priemonių, taikomų siekiant kiekybiškai išnagrinėti ligando ir receptoriaus (taikinio) sąveikų stiprumą. QSAR yra nepakeičiamas metodas, jei nežinoma biologinio taikinio erdvinė struktūra, nes jis grindžiamas vien ligandų struktūra. QSAR šiais laikais apima visus statistinius metodus, kuriais cheminių junginių biologinis aktyvumas, dažniausiai išreiškiamas kaip jų logaritmas, siejamas su cheminių junginių struktūrinėmis, fizikinėmis-cheminėmis ar kvantinėmis savybėmis arba palyginamąja molekulių laukų analize [161, 162]. QSAR modeliavimas išaugo, įgavo daug krypčių ir evoliucionavo nuo taikymo mažoms serijoms cheminių junginių su paprastais matematinės regresijos analizės metodais ryšiui su biologiniu aktyvumu nustatyti į ištisų duomenų bazių analizę. Ryšiui su biologiniu aktyvumu nustatyti naudojama daug įvairių statistinių ir sistemos mokymosi metodų [163]. Dabartiniame QSAR sujungiama Hansch ir Wilson analizė, o modelių atranka vykdoma sistemingai, naudojant šimtus ir tūkstančius įvairių parametrų, kurie galėtų tikti galutinei lygčiai, ir atrenkant, derinant pačius tinkamiausius iš jų. QSAR modeliavimo nustatytas sąryšis tarp biologinio aktyvumo ir lengvai apskaičiuojamų molekulių savybių toliau gali būti naudojamas naujų junginių biologiniam aktyvumui prognozuoti. Dažniausiai QSAR modeliavime naudojami cheminių junginių parametrai yra junginio ar jo fragmentų, funkcinių grupių hidrofobiškumas, aromatiškumas, teigiama ar neigiama jonizacija solvatacijos metu, galimybė sudaryti donorinius/akceptorinius vandenilinius ryšius, ir kitokie. Šie parametrai yra išreiškiami skaitinėmis vertėmis ir tolesniame etape naudojami QSAR lygčiai (modeliui) kurti, koreliacijai tarp kai kurių iš jų derinio ir nagrinėjamo biologinio aktyvumo nustatyti. Skaičiavimai automatizuojami ir įgalinami, naudojant kompiuterius.

QSAR analizės uždavinio pradinė sąlyga aprašoma taip: yra rinkinys (p) junginių, kurių kiekvienas turi išmatuojamą savybę (K) (pavyzdžiui, pusiausvyros konstantą) ir daug apskaičiuojamų deskriptorių (x_N) (pavyzdžiui, molekulinė masė, hidrofobiškumas ir t.t.). Pradinę sąlygą sudaro konstantų

vektorius K_p ir deskriptorių matrica $x_{p,N}$, tarp kurių ieškoma matematinio sąryšio. QSAR įvesties pavyzdys:

Junginiai	Biologinis aktyvumas	Deskriptorius 1	Deskriptorius 2	... Deskriptorius N
1 junginys	0,1	1,1	2,3	... 3,3
2 junginys	0,2	3,4	4,3	... 7,7
...
M junginys	0,4	2,0	3,0	... 4,8

Dažniausiai, sprendžiant šį uždavinį iš visų N deskriptorių išrenkami keli reikšmingiausi n galutinei lygčiai sudaryti. Tai daroma ne visuomet. Norint sukurti kokybišką QSAR modelį, svarbiausia atrinkti tinkamus deskriptorius, kurie skaičiais geriausiai užkoduotų ir atspindėtų informaciją apie turimus cheminius junginius, jų struktūrą ir įvairovę. Tam sukurta daug įvairių metodų. Patys populiariausi: visų galimų variantų sistemingas išbandymas, evoliuciniai algoritmai, pavyzdžiui, genetinis algoritmas, ir sistemos mokymosi algoritmai. Visi jie gali būti naudojami cheminių junginių parametrų (deskriptorių) – fizikinėms, cheminėms, kvantinėms, matematiškai aprašytos cheminės struktūros savybėms – išrūšiuoti į tinkamus ir netinkamus biologiniam aktyvumui prognozuoti QSAR modelio pagalba. Tokiai parametru pasirinkimo problemai spręsti yra kelios metodų grupės. Visų pirma paminėtini klasikiniai metodai, grindžiami MLR [164]. Gerokai mažiau populiarus yra prognoze paremtas deskriptorių išrinkimo ir QSAR modeliavimo metodas VSMP (angl. *variable selection and modeling method based on the prediction*) [165]. Tarp naujesnių paminėtini dirbtiniu intelektu grindžiami metodai, pavyzdžiui, dirbtinių neuronų tinklai, nekonkrečios logikos metodai ir panašūs [166-168]. Dirbtiniu intelektu grindžiami metodai ypač gerai tinka siekiant nustatyti netiesinę parametru rinkinio ir dominančio biologinio aktyvumo priklausomybę, ir turi eilę kitų privalumų, palyginti su klasikiniais QSAR parametru parinkimo metodais [169].

Pasirinkus tinkamus molekulinis deskriptorius, būtina parinkti tinkamą matematinį ar statistinį modelį priklausomybei tarp tų deskriptorių ir biologinio aktyvumo aprašyti. Šis etapas glaudžiai susijęs su ankstesniu, nes paprastai,

prieš ieškant tinkamų deskriptorių, iškeliamą hipotezę apie tam tikros rūšies priklausomybę tarp jų ir biologinio aktyvumo. Todėl kartais abu minėti etapai vykdomi kartu. Yra įvairių pasirenkamos QSAR priklausomybės funkcijos pavyzdžių. Dažniausiai, kaip jau minėta, pasirenkama MLR. Kiti metodai – klasikinė tiesinė regresija, dalinių mažiausių kvadratų metodas (PLS), pagrindinių komponentų analizė (PCA), pagrindinių komponentų regresija (PCR), k artimiausių kaimynų metodas (kNN) ir kiti [136]. QSAR modeliavimą atliekant su PLS metodu, pagal $x_{p,N}$ ir K_p nustatomi PLS vektoriai (paslėptos vertės), kurie toliau naudojami, prognozuojant matuojamą junginių savybę (K). Taip išvengiama parametrų matricos supaprastinimo iki n stulpelių [170]. QSAR metode tradiciškai nagrinėjama tiesinė priklausomybė tarp deskriptorių ir biologinio aktyvumo logaritmo, bet kai kurie autoriai atlieka QSAR ir su netiesinėmis priklausomybėmis [171-173]. Gana populiari tą patį deskriptorių, pakeltą kvadratu, panaudoti antrą kartą toje pačioje QSAR lygtyje ($\dots x+x^2 \dots$). Yra sudėtingesnių netiesinių metodų, pavyzdžiui, dirbtinių neuronų tinklai (ANNs) ir atraminių vektorių mašinos (SVM) Visa tai naudojama koreliacijos funkcijai ar ryšiui aprašyti QSAR metode. Pasirinkus tinkamą priklausomybės funkcijos aprašymo būdą, QSAR priklausomybės funkcija parametrizuojama su visų turimų junginių, kurių biologinis aktyvumas išmatuotas, atsitiktinai parinkta dalimi, vadinama apmokymo rinkiniu (angl. *training set*). Likę junginiai panaudojami vėliau. Parametrizavimo procesas su apmokymo rinkiniu kartojamas, kol pasiekiamas patenkinamas rezultatas.

Jei pasirenkama MLR priklausomybės funkcija, tada tarp x_n ir K (ar $\log(K)$) vykdoma tiesinė regresija su keliais nepriklausomais kintamaisiais. K ir x_n tarpusavio priklausomybė aprašoma lygtimi, kuri bendrai gali būti išreiškiama taip [174]:

$$K = a_1x_1 + a_2x_2 + \dots + a_nx_n + b. \quad (1.5.3-1)$$

Pagal šią lygtį prognozuojamas testavimo rinkinio ir dar neegzistuojančių junginių K , ir pati lygtis parodo, kokius jos kintamuosius,

aprašančius nagrinėjamų cheminių junginių struktūrą, reikėtų keisti, norint sukurti junginį su didesne ar mažesne konstanta K . Nubraižyti grafikus tiesiogiai QSAR lygčiai dažniausiai neįmanoma dėl per didelio kintamųjų skaičiaus, todėl paprastai QSAR modelių grafikai braižomi tarp prognozuojamų biologinių aktyvumų, apskaičiuotų QSAR lygtimi, ir eksperimentiškai išmatuotų jų verčių.

Sugalvota ir suformuluota daug taisyklių, kaip sukurti QSAR modelius, kai kurios jų būna nevienareikšmiškai naudingos ir yra diskusijų objektas mokslinėje bendruomenėje [175, 176]. Bendra modelio kokybė ir kiekvieno parametro reikalingumas, prasmė įtraukti jį, turi būti įvertinti pagal daug statistinių parametrų (F testas, determinacijos koeficientai pagal įvairias procedūras ir pan.). Galutinis QSAR modelis turi būti su minimaliu skaičiumi modeliavimo parametrų ir molekulinį deskriptorių. Siūloma neieškoti tinkamų parametrų per dideles pradines jų aibes, nes galimas grynai atsitiktinis tinkamumas [177]. Vieni iš jų tapo populiarūs QSAR modeliavimo srityje, kiti po pradinių darbų paplito mažai [178, 179].

2015 metais Ruusmann su bendradarbiais sukūrė QSAR modelių duomenų bazę (QSARDB – <http://qsar.db.org/>) [180]. Autoriai tikisi, kad QSARDB bus praktiškai naudinga priemonė QSAR metodo taikytojų bendruomenei ir dėl jos QSAR modeliavimas ir jo rezultatai taps skaidresni, atkartojami ir lengviau prieinami visiems [181].

1.5.4. QSAR modelių vidinis ir išorinis testavimas, tinkamumo patvirtinimas

QSAR modeliavimui paimti visi pradiniai nagrinėjamų cheminių junginių biologinio aktyvumo duomenys paprastai pačioje modeliavimo pradžioje išskirstomi į tokius rinkinius:

Apmokymo rinkinys (angl. *training set*) – visiems QSAR modeliams sukurti reikia taip vadinamo apmokymo rinkinio. Šis rinkinys naudojamas generuoti ir parametrizuoti naujiems QSAR modeliams. Dažniausiai apmokymo rinkinys būna didesnės apimties, susideda iš mažiausiai keliolikos

junginių, kurių bent keli turi vienodas modeliuojamas struktūrinės savybes, cheminės struktūros bruožus. Nėra gerai, jei kiekvienas junginys rinkinyje turi keletą originalių cheminės struktūros bruožų, nes tuomet vieno struktūros fragmento įtaka nėra aiški.

Testavimo rinkinys (angl. *test set*) yra duomenys, kurie panaudojami įvertinti jau sukurto QSAR modelio galimybes prognozuoti biologinį aktyvumą. Jeigu yra pakankamai duomenų, dalis jų atskiriama ir panaudojama kaip testavimo rinkinys tam, kad būtų pademonstruotos naujai sukurto QSAR modelio galimybės prognozuoti. Kaip gerai tas modelis aprašo apmokymo rinkinį, kuriam jis parametrizuotas, irgi yra svarbu. Pagal testavimo rinkinio duomenis nėra nei parenkami QSAR deskriptoriai, nei parametrizuojami jokie QSAR modeliai, nei vykdoma kokia nors jų atranka. Kadangi testavimo rinkinys dažniausiai parenkamas atsitiktinai iš visų pradinių duomenų, gerai, kai pradinių duomenų rinkinys yra pakankamai didelis. Rečiau, be atsitiktinio parinkimo, dar gali būti naudojamas koks nors kitas metodas testavimo rinkiniui parinkti, pavyzdžiui, klasterinė analizė. Koks bebūtų testavimo rinkinio sudarymo metodas, galutinis testavimo rinkinys turi reprezentatyviai atspindėti visą turimą duomenų aibę. Pavyzdžiui, jame negali atsirasti eilė biologiškai aktyviausių ar neaktyviausių cheminių junginių ar eilė cheminių junginių su tuo pačiu struktūriniu bruožu, kurio nelieka apmokymo rinkinyje. Visi tie dalykai neigiamai iškreipia, pablogina testavimo rezultatus [182-184].

Tokio junginių skirstymo gali ir nebūti, tuomet visi imti duomenys būna panaudoti kaip apmokymo rinkinys. Kartais pritaikomi įvairūs kitokie skirstymo būdai.

QSAR modelio „apmokymo“ su apmokymo duomenų rinkiniu metu sukurtas QSAR modelis naudojamas testavimo rinkinio junginių aktyvumui prognozuoti. Testavimo junginių rinkinio, kartais vadinamo išoriniu rinkiniu, biologinio aktyvumo prognozavimo kokybė priimama, kaip QSAR modelio galimybių prognozuoti nežinomų junginių savybes vertinimas. Naudojant jau parametrizuotą QSAR modelį, papildomai prognozuojamas biologinis

aktyvumas likusiems testavimo rinkinio junginiams. Taip imituojama situacija, kai jis būtų prognozuojamas junginiams su kol kas nežinomais biologiniais aktyvumais ar visai nesusintetintiems junginiams. Statistiškai įvertinus, kaip pavyko prognozuoti biologinį aktyvumą, testavimo rinkinio junginiams galima gana realistiškai įvertinti QSAR modelio tinkamumą prognozėms. Testavimo rinkinys atskiriamas ne visuomet. Kai QSAR modelių kūrimui, tinkamų deskriptorių paieškai jiems, pasirenkami neuronų tinklų metodai, tuomet teigiama, kad nėra reikalo papildomai naudoti testavimo rinkinį sukurto QSAR modelio teisingumui patvirtinti [185]. Nepriklausomo testavimo rinkinio panaudojimas laikomas standartu visame QSAR modeliavime, vertinant naujai sukurto QSAR modelio prognozavimo kokybę. Tai griežčiausias kriterijus QSAR modeliavimo kokybei vertinti. Ne viename moksliniame QSAR metodologijos straipsnyje aprašyti argumentai, dėl kurių reiktų visuomet naudoti testavimo rinkinį sukurtiems QSAR modeliams testuoti [186, 187].

Idealu, jei testavimas atliekamas, kiek įmanoma atkartojant praktinį QSAR modelio taikymą. Kai atliekamas toks testavimas, iš sukurto QSAR modelio prognozuojami keli iki tol eksperimentiškai neišmatuoti, labiau biologiškai aktyvūs cheminiai junginiai. Po to jie visi susintetinami laboratorijoje, ir eksperimentiškai yra patvirtinamas arba paneigiamas jų aktyvumas. Praktiškai labai retai kada daromas toks eksperimentinis QSAR modelių testavimas, ir skirstymas viso turimo junginių rinkinio į apmokymo ir testavimo rinkinius yra tokios praktikos imitacija. QSAR modelių testavimas skirstomas į vidinį ir išorinį testavimą. Literatūroje nėra konkretaus šių testavimo rūšių apibrėžimo. Cheminėje informatikoje laikoma, kad testavimas išorinis, jei testavimo rinkinys panaudotas tik vieną kartą. Tai yra tuomet, jei į jį visai nežiūrima QSAR modelio kūrimo ir parinkimo procese [188]. Atitinkamai vidinis testavimas yra, kai testavimo rinkinys vienaip ar kitaip dalyvauja naujo QSAR modelio kūrime, testuojant potencialius QSAR modelius daug kartų, bet jų neparimetrizuoju. Parametrizavimui naudojamas išimtinai tik apmokymo rinkinys. Išorinis testavimas turi trūkumų, nes dalis

serijos junginių lieka nepanaudoti QSAR modelių tobulinime. Egzistuoja atsitiktinumo faktorius: priklausomai nuo to, kaip pradinis duomenų rinkinys išskaidomas į apmokymo ir testavimo rinkinius, testavimo rezultatai keičiasi. Griežtai teigiant, tik vidiniu testavimu testuotas QSAR modelis lieka be nepriklausomo įvertinimo arba tinkamumo patvirtinimo (angl. *validation*).

QSAR atlikimo variantai, nenaudojantys testavimo rinkinio, ima visus pradinio rinkinio junginius apmokymo rinkiniui. Dažniau tai daroma tada, kai pradinis junginių rinkinys mažas, pavyzdžiui, vos dešimt junginių, ir jį išskirsčius į du rinkinius, vienam jų liktų gal net mažiau kaip penki junginiai, ir tuomet būtų mažai prasmės nagrinėti statistines koreliacijas ar skaičiuoti bet kokią statistiką su tokiomis mažomis imtimis. Abu šie rinkiniai turi išlikti pakankamai dideli ir tinkamos įvairovės, reprezentatyvūs. Plačios apimties duomenų rinkiniai yra reta „prabanga“ QSAR modeliavime. Jei apimtys yra mažos, vietoje testavimo rinkimo pasitelkiami kitokie metodai QSAR modelio prognozavimo galimybės vertinti, tam modeliui patikrinti. Kiti dažnai naudojami statistiniai metodai QSAR modelių prognozinėms galimybės įvertinti yra kryžminis patikrinimas (angl. *cross-validation*) [189] ir jo variantai, tokie kaip dvigubas kryžminis patikrinimas, įkrovimas (angl. *bootstrapping*) ir atsitiktinis Y sumaišymas (angl. *Y scrambling*).

Atliekant kryžminį patikrinimą, dalis pradinio duomenų rinkinio narių išmetama, ir QSAR modelis iš naujo generuojamas su likusiais duomenų rinkinio nariais. Dažniausiai išmetamas vienas pradinio duomenų rinkinio narys ir tuomet metodas vadinamas kryžminių patikrinimu išmetant vieną (angl. *leave one out cross validation*, LOOCV). Jei išmetamas daugiau kaip vienas pradinio duomenų rinkinio narys, tuomet metodas vadinamas kryžminių patikrinimu išmetant kelis (angl. *leave multiple out cross-validation*, LMO-CV). Naujai generuotas QSAR modelis naudojamas išmestoms duomenų vertėms prognozuoti. Visa tai kartojama daug kartų ar net išėikvojant visus įmanomus išmetimo variantus. Kuo daugiau tokių pakartojimų atliekama, tuo geriau, mat pašalinami svyravimai tarp skirtingų rezultatų, išmetant skirtingus

aibės narius. Galiausiai paskaičiuojamas statistinis įvertis, nusakantis, kaip gerai tokie QSAR/QSPR modeliai prognozuoja išmestas vertes. Paprastai tokiu būdu paskaičiuotas R^2 yra žymimas kaip Q^2 . Pagal jį yra vertinamas pradinis QSAR/QSPR modelis. Golbraikh ir Tropsha aprašė kryžminio patikrinimo išmetant vieną, skaičiuoto su apmokymo rinkiniu, ir taikyto QSAR modelių prognozavimo galimybėms vertinti trūkumus [186], šiuo statistiniu metodu gautos Q^2 vertės išnagrinėtos, ir pasiūlyti griežtesni kriterijai, kaip vertinti QSAR modelių prognozavimo galimybes (aukštas Q^2 , R_0^2 paskaičiuoti per nulį koordinatų tašką turi būti artimi R^2 ir 1). Kryžminis patikrinimas išmetant vieną linkę vesti prie perdėtai teigiamų QSAR modelio vertinimo rezultatų [190].

Kartais būna atliekamas dvigubas kryžminis patikrinimas (angl. *double cross-validation*). Tuomet išoriniame dvigubo kryžminio patikrinimo cikle visi duomenys atsitiktinai išskirstomi į apmokymo ir testavimo rinkinius, kaip įprastai. Šis išorinis testavimo rinkinys naudojamas išimtinai tik galutinio modelio statistiniam įverčiui skaičiuoti. Išorinis apmokymo rinkinys papildomai atsitiktine tvarka yra skirstomas į vidinį apmokymo rinkinį ir vidinį testavimo rinkinį, kurie atitinkamai naudojami modelio sukūrimui ir įvertinimui. Po pirmo modelio sukūrimo ir įvertinimo, vidiniai rinkiniai atsitiktinai performuojami iš naujo. Tuomet lyginant visus modelius tarpusavyje, geriausiojo iš jų išrinkimui turima ne vien statistika su apmokymo rinkiniu, bet ir su testavimo rinkiniu. Pagal šias vidinių rinkinių statistikas, geriausias modelis pasirenkamas efektyviau. Galiausiai atliekamas galutinis geriausio modelio statistinis įvertinimas su išoriniais duomenų rinkiniais [191]. Tai būtina, mat pakartotinai perskaičiuojama ir galiausiai išrenkama geriausia statistika su vidiniais testavimo duomenų rinkiniais tampa iškreipta, šališka [192], ir tai nebėra visiškai korektiškas QSAR modelio testavimas. Vienas iš šios statistinės metodologijos trūkumų yra tas, kad tokiai jai atlikti reikia dar daugiau pradinių, išmatuotų junginių serijoje nei įprastai, mat jie suskirstomi į daugiau duomenų rinkinių įvairiais pavadinimais. Šitokia QSAR modelių

testavimo ir pripažinimo tinkamais praktika gana sena, bet tik paskutiniaisiais metais naudojama vis plačiau įvairiuose QSAR darbuose, nepaisant to, kad jai atlikti reikia papildomų junginių [193-195].

Krstajic su bendradarbiais 2014 metais pasiūlė naudoti sudėtingesnes, imlesnes skaičiavimo resursams, bet efektyvesnes QSAR modelių tikrinimo procedūras, kurias įgalina debesų kompiuterijos atsiradimas [196], tokiais kaip penkiagubas kryžminis patikrinimas (angl. *fivefold*) ir dešimtubas kryžminis patikrinimas (angl. *tenfold*). Šių metodologijų principai išlieka tie patys, kaip ir dvigubame kryžminiame patikrinime, tik, kaip pavadinimas sako, procedūra nėra dvigubinama. Tokių kryžminių patikrinimų metodai sudėtingėja, nes kartojami keleriopai ar keliais sluoksniais. Parodyta šių metodologijų nauda ir svarba, lyginant su senesniais būdais, bet atitinkamai išauga poreikis skaičiuojamiesiems resursams.

Įkrovimas (angl. *bootstrapping*) yra dar vienas statistinis metodas, naudojamas QSAR/QSPR modeliavime. Jo atlikimo metu pakartotinai daroma prielaida, kad modelio paskaičiuotos biologinio aktyvumo ar kt. vertės yra gautos eksperimentiškai, ir iš jų atsitiktiniu parinkimu sudaroma duomenų imtis, tik su naujomis biologinio aktyvumo vertėmis. Taip gautai naujai duomenų imčiai (įkrovos imčiai) iš naujo sukuriamas QSAR modelis, ir tai kartojama daug (keliasdešimt ar net šimtus) kartų. Galiausiai visi gautieji QSAR modeliai lyginami tarpusavyje, ir pastebimas didesnis ar mažesnis statistinio įverčio, tokio kaip R^2 , nuokrypis nuo pradinio, vertinamojo QSAR modelio. Jei nuokrypis artimas nuliui, tai pradinis QSAR modelis yra kokybiškas.

Atsitiktinis Y sumaišymas yra rečiau naudojamas statistinis metodas QSAR/QSPR modeliavime. Jo metu, kaip įvardija pavadinimas, apmokymo junginių rinkinio aktyvumai atsitiktinai sumaišomi, ir mėginama modeliuoti tokius nerealistiškus duomenis. Tikrinant, kaip modeliavimas pavyksta su realistiškais eksperimentiniais duomenimis, ir kaip su nerealistiškais arba atsitiktiniais skaičiais, irgi galima įvertinti realistiško QSAR/QSPR modelio

kokybę, pastebėti, kuo jis geresnis ar identiškias lyginant su beprasmiu „triukšmo“ modeliu.

Kryžminio patikrinimo išmetant vieną metodas, įkrovimo metodas, ir visi kiti panašūs statistiniai metodai linkę pateikti gana perdėtus, perdėm optimistiškus rezultatus apie QSAR modelio galimybes prognozuoti naujų, dar nežinomų, junginių biologinį aktyvumą: kaip bebūtų perdėliojamas turimas, jau panaudotas modelio apmokymui, junginių rinkinys, tai nėra visai nauja imtis, kurios savybės bus prognozuojamos. Tai yra tik vienoks ar kitoks artinys prie naujos imties situacijos. Gali būti taip, kad QSAR modelis puikiai aprašo ir paaiškina apmokymo rinkinio junginių savybes, bet netinkamas kitiems modeliavime nenaudotiems junginiams prognozuoti, arba kad jo tinkamumas tam lieka neišnagrinėtas ir statistiškai neįvertintas [186, 197].

Sukurtas klasikinio QSAR modelio vizualizavimo įrankis, su kuriuo galima skaičiuoti molekulinis deskriptorius ir vizualizuoti QSAR modelių tinkamumo patvirtinimo vertinimus, jų prognozes, vadinamas CHES-MAPPER 2.0 [198]. Jis veikia tiesiog interneto puslapyje (<http://ches-mapper.org/>). Autoriai tokį sprendimą vadina vizualiu įvertinimu, kuris nepakeičia esamo modelių statistinio vertinimo, bet gali padėti pastebėti ir lengviau išspręsti turimo duomenų rinkinio trūkumus, interpretuoti ir patobulinti modelį. Su programa generuojamas paveikslėlis su viso duomenų rinkinio cheminėmis struktūromis, ir jos pažymimos dviem sferomis, kurių mažesnė rodo realų biologinį aktyvumą spalvinėje skalėje nuo raudonos iki mėlynos. Didesnė sfera rodo prognozuojamą biologinį aktyvumą toje pačioje skalėje. Iš karto puikiai matyti klaidingos prognozės, pažymėtos dvejomis skirtingomis spalvomis ant tos pačios cheminės struktūros.

1.5.5. QSAR modelių statistinis vertinimas ir vizualizavimas

Taikant QSAR metodą, būtina atlikti sukurtų QSAR modelių atranką ir vertinimą, siekiant išsiaiškinti, kuris iš gautų QSAR modelių yra geriausias. Galutiniai rasti QSAR modeliai turi būti geresni už aibę ne tokių gerų pagal statistinius įverčius ir todėl atmestų modelių. Praktikoje QSAR modeliai

paprastai nėra absoliučiai geriausi iš visų galimų pagal turimus pradinius duomenis apie chemines struktūras ir jų biologinį aktyvumą. Taikant QSAR metodologiją, biologinis aktyvumas suskaičiuojamas turimai junginių serijai, ir tuomet nagrinėjama paskaičiuotų verčių koreliacija su eksperimentiškai išmatuotu biologiniu aktyvumu, taikant tinkamą statistinį metodą – regresinę analizę. Dažniausiai QSAR modelio kokybei vertinti naudojamas determinacijos koeficientas R^2 :

$$R^2 = (TSS - RSS) / TSS = 1 - RSS / TSS, \quad (1.5.5-1)$$

kur liekamųjų paklaidų kvadratų suma (RSS) yra išreiškiama taip:

$$RSS = \sum (y_i - f_i)^2, \quad (1.5.5-2)$$

o visa kvadratų suma (TSS) yra

$$TSS = \sum (y_i - \bar{y})^2. \quad (1.5.5-3)$$

R^2 yra standartinis QSAR modelio kokybės matas, paskaičiuojamas iš išmatuotų duomenų verčių (y_i), jų vidurkio (\bar{y}) ir su jomis susijusių paskaičiuotų duomenų verčių (f_i) [199]. TSS parodo visą y dispersiją, o RSS parodo y dispersiją, kuri liko nepaaiškinta per x atlikus regresiją. TSS – RSS parodo y dispersiją, kuri paaiškinta per x turimu modeliu. R^2 gali būti suprantamas, kaip santykis tarp modeliu paaiškintos ir visos y dispersijos [200]. Kitas būdas paskaičiuoti determinacijos koeficientą tiesinėje regresijoje yra tiesiog pakelti Pirsono koreliacijos koeficientą (r) kvadratu. Koeficientas r yra paskaičiuojamas taip:

$$r = \text{cov}(x, y) / (\sigma_x \sigma_y), \quad (1.5.5-4)$$

kur duomenų rinkinio su n elementų kovariacija ($\text{cov}(x, y)$) yra

$$\text{cov}(x, y) = \sum (y_i - \bar{y})(x_i - \bar{x}) / n. \quad (1.5.5-5)$$

Čia σ_x ir σ_y yra atitinkamai x ir y standartiniai nuokrypiai.

Determinacijos koeficientas QSAR modeliavime parodo, kaip gerai QSAR modelis aprašo eksperimentiškai išmatuotas biologinio aktyvumo

vertes. Jei determinacijos koeficientas lygus nuliui, tai QSAR modelis visai neaprašo eksperimentiškai išmatuotų biologinio aktyvumo verčių. Jei determinacijos koeficientas lygus vienetui, tai QSAR modelis idealiai atkartoja eksperimentiškai išmatuotas biologinio aktyvumo vertes ($y_i = f_i$). Abu šie atvejai yra kraštutiniai, ir realiai QSAR modeliavime determinacijos koeficientas būna intervale tarp nulio ir vieneto. Determinacijos koeficientas, paskaičiuotas apmokymo rinkiniui, yra perdėtas QSAR modelio kokybės matas, mat visados deskriptoriai yra dauginami iš papildomų QSAR modelio parametrizavimo koeficientų (angl. *fitting coefficient*) taip, kad būtų pasiektas maksimalus determinacijos koeficientas, taigi jis yra maksimizuotas apmokymo rinkiniui. Determinacijos koeficientas, paskaičiuotas testavimo rinkiniui, jau gerokai realistiškiau parodo QSAR modelio galimybes prognozuoti kitus junginius, kurie nebuvo naudoti to modelio parametrizavime.

Čia minėtas determinacijos koeficientas neatsižvelgia į tai, kiek deskriptorių ir modeliavimo parametrų yra naudojama QSAR modelyje. Kai imama vis daugiau kintamųjų į QSAR modelį, determinacijos koeficientas linkęs atitinkamai didėti. Vadinasi, determinacijos koeficientą galima padidinti tiesiog pridėdant daugiau kintamųjų. Nėra korektiška tarpusavyje palyginti visai skirtingus QSAR modelius su skirtingais kintamųjų kiekiais vien pagal determinacijos koeficientą. Tam specialiai sukurtas redukuotas determinacijos koeficientas (R^2_{ADJ}), kuris paskaičiuojamas taip [201]:

$$R^2_{ADJ} = (R^2(n - 1) - p) / (n - p - 1), \quad (1.5.5-6)$$

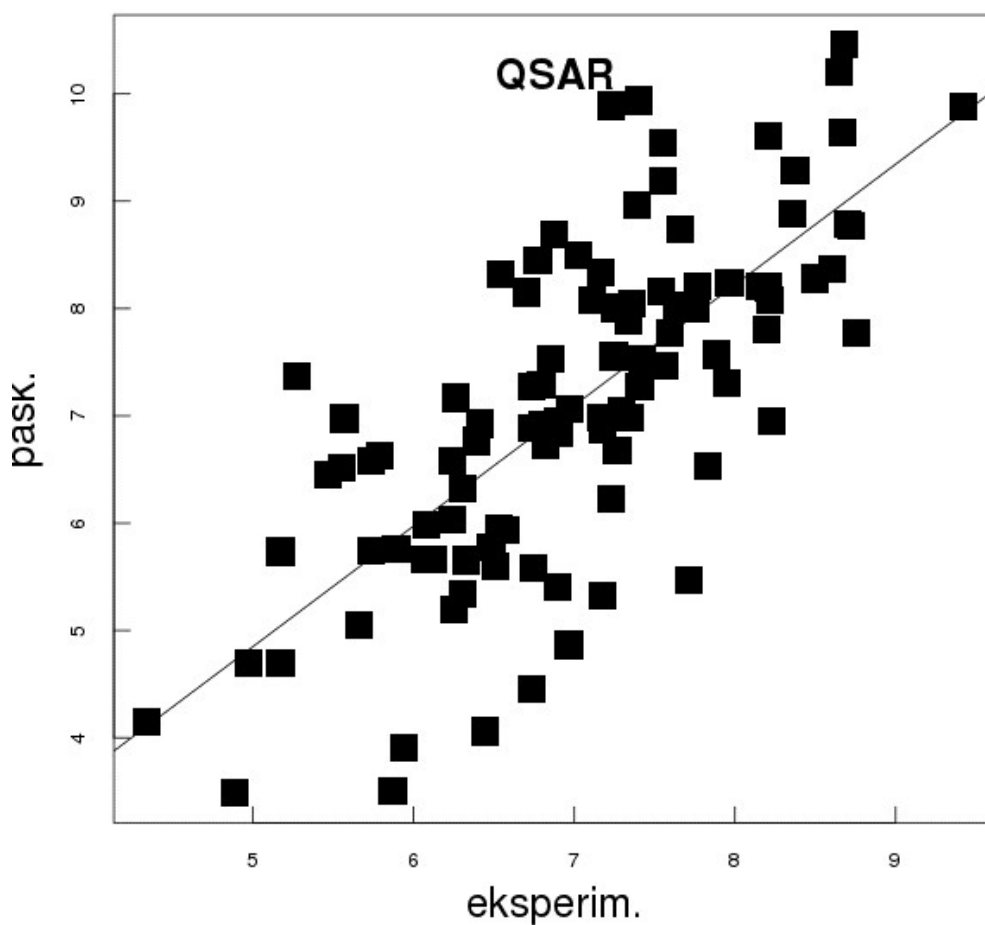
kur n yra imties dydis, o p – deskriptorių skaičius panaudotas tame QSAR modelyje. Redukuotas determinacijos koeficientas ne tik ne didėja, bet gali ir mažėti, jei imama daugiau kintamųjų.

Du dažniausiai QSAR modeliavime pateikiami paklaida paremti įverčiai yra RMSE (angl. *root mean square error*) ir MAE (angl. *mean absolute error*). Jie skaičiuojami taip:

$$RMSE = \sqrt{(1/n) \sum (Y_{obs} - Y_{pred})^2}, \quad (1.5.5-7)$$

$$\text{MAE} = (1/n)\Sigma(|Y_{\text{obs}} - Y_{\text{pred}}|), \quad (1.5.5-8)$$

kur n – imties dydis, Y_{obs} – eksperimentinės stebėjimų vertės, o Y_{pred} yra su jomis susiję prognozuojami dydžiai. Iš formulių 1.5.5-7 ir 1.5.5-8 matosi, kad abu minėti įverčiai parodo vidutines prognozavimo paklaidas per visą nagrinėjamą prognozuojamų duomenų aibę. Yra teigiama, kad RMSE interpretacija labiau komplikauta, negu MAE [202]. MAE parodo vidutinį prognozės nuokrypį nuo matavimo tais pačiais vienetais, kuriais atlikti matavimai.



Paveikslas 1.5.5-1. Tipinis QSAR grafikas. Pagal sumodeliuotą QSAR lygtį suskaičiuojamos ir atidedamos paskaičiuotos tam tikrų išmatuojamų cheminių

junginių savybių vertės apmokymo rinkinio junginiams X ašyje ir Y ašyje atidedamos eksperimentiškai išmatuotos atitinkamų savybių vertės tiems patiems junginiams.

Paprastai nubraižomas QSAR grafikas su modelio paskaičiuotomis biologinio ir eksperimentinėmis biologinio aktyvumo vertėmis, atidėtomis kaip taškai (paveikslas 1.5.5-1). Kuo visi taškai arčiau tiesės, tuo geresnis QSAR modelis, nes tada prognozuojamos biologinio aktyvumo vertės yra artimos eksperimentinėmis. Be to, grafike galima matyti taškus, kurie blogai atitinka modelį, kitaip vadinamus išskirtimis (angl. *outliers*). Vadinasi, tais taškais pažymėtus junginius prastai aprašo turimas QSAR modelis. Iš to, kokie junginiai prastai aprašomi turimu QSAR modeliu, galima apytiksliai matyti, kokių juos aprašančių QSAR deskriptorių trūksta tame modelyje. Tuos deskriptorius pridėjus, dažnai gali pavykti sukurti kitą, dar geresnį, QSAR modelį.

QSAR modeliavimo tikslas yra keleriopas: pirma, QSAR modelis turi paaiškinti esamus duomenis, t.y. prognozuoti, kas vyksta tarp jų (interpoliuoti), antra, papildomai optimizuoti turimus serijos junginius ir pasiūlyti smulkius pakeitimus juose, ir trečia, prognozuoti visai nežinomus junginius (ekstrapoliuoti). Pavyzdžiui, jei nagrinėjame junginius, kurie jau buvo naudoti modelio kūrime, tai atliekame interpoliavimą. Jei nagrinėjame naujus junginius, kurie nebuvo naudoti modelio kūrime, tai tada atliekame ekstrapoliavimą. Šie trys dalykai praktiniame kontekste gali būti nevienodai svarbūs.

Didžiausia QSAR modelio vertė slypi tame, kaip sėkmingai jis paaiškina biologinių aktyvumų skirtumus turimame duomenų rinkinyje, ir, svarbiausia, kaip gerai jis tinka prognozėms. Pasitaiko, jog keli modelio deskriptoriai su biologiniu aktyvumu susiję tik visi drauge, bet po vieną atskirai su juo koreliuoja labai prastai ar net visai nekoreliuoja, todėl automatiškai deskriptorių reikšmių perkelti į QSAR modelio interpretaciją negalima. Tai nereiškia, kad visas QSAR modelis yra blogas, tiesiog tada jį

sunku paaiškinti per deskriptorius po vieną. Dažniausiai tokiais atvejais matematinis sąryšis yra per daug sudėtingas, kad būtų išreikštas tekstu, ir yra pateikiama vien jo matematinė lygtis. Neretai QSAR modeliavimui keliamas uždavinys rasti naujausius ir geriausius dar nežinomus cheminius junginius, ir nėra svarbu tiksliai iš anksto žinoti, koks bus tų junginių biologinis aktyvumas [203]. Zefirov su bendraautoriumi skirsto QSAR į prognozinį ir aprašomąjį. Kai QSAR modelį lengva interpretuoti, ir galima tą interpretaciją, bet ne iš modelio skaičiuotas biologinio aktyvumo vertes naudoti prognozėms tai yra aprašomasis QSAR [204].

Viena iš nesenujų sąvokų moksle apie QSAR yra pritaikomumo domenas (angl. *applicability domain*). Pritaikomumo domenas yra fizikocheminė, struktūrinė, biologinė erdvė, žinojimas ar informacija, kuria remiantis buvo sukurtas QSAR apmokymo duomenų rinkinys, ir iš kurios galima daryti tinkamas prognozes apie naujus junginius [205]. QSAR pritaikomumo domenas turi būti aprašomas svarbiausiais duomenimis, pavyzdžiui, molekuliniais deskriptoriais, kurie naudojami QSAR modelyje. Idealu, jei QSAR yra naudojamas prognozėms tame pritaikomumo domene tik su interpoliacija ir ne su ekstrapoliacija [206]. QSAR pritaikomumo domenas dar yra apibrėžtas kaip atsako ir cheminių struktūrų erdvė, kurioje QSAR modeliu atliekamos prognozės su duotu patikimumu. QSAR pritaikomumo domenas plačiai ir paprastai suprantamas tiesiog kaip modelio prognozinė galimybių riba, pavyzdžiui, visas rinkinys cheminių struktūrų, kurioms tas QSAR modelis gali būti sėkmingai pritaikomas [205]. Panaudojus QSAR modelius, gaunamos patikimos prognozės, jei yra išlaikomos tų QSAR modelių kūrimo metu padarytos prielaidos. Cheminė erdvė, kurią užima apmokymo rinkinio junginiai, yra pagrindas, pagal kurį galima įvertinti, kada prognozės bus teisingos, nes interpoliacija yra labiau patikima, nei ekstrapoliacija. QSAR apmokymo rinkinys gali būti analizuojamas tiek tiesiogiai pagal struktūrinius panašumus, tiek pagal iš jo suskaičiuotų molekulinė deskriptorių intervalus, kuriuos tas rinkinys užima [207]. Netzeva su bendraautoriais [205] ir Jaworska

su bendraautoriais [207] parašė apžvalgas apie QSAR pritaikomumo domenų nagrinėjimo metodus ir pateikė daug rekomendacijų apie jų nustatymą ir panaudojimą. Jie rekomendavo, kad QSAR pritaikomumo domeno apibrėžimas konkreitiems QSAR modeliams turi būti atliekamas to modelio kūrėjo, o ne naudotojo. Dar jie rekomendavo, kad pats paprasčiausias būdas tai padaryti, tai bent jau publikuoti visas apmokymo rinkinio junginių chemines struktūras ir visas iš jų suskaičiuotas deskriptorių vertes, kurios naudojamos galutiniuose QSAR modeliuose. Labai mažai QSAR/QSPR modeliavimus aprašančių mokslinių straipsnių tiksliai laikosi šios rekomendacijos, tuomet publikuotų QSAR modelių pritaikomumo domenai lieka nežinomas [208]. QSAR pritaikomumo domeno koncepcija realizuota keliose komercinėse QSAR modeliavimo programose, tokiose kaip MCASE, TOPKAT ir CHEM-TOX [205]. Konkretaus QSAR modelio pritaikomumo domenui apibrėžti dažniausiai naudojamos tam tikros struktūrinės taisyklės, tolygūs deskriptorių verčių intervalai arba abu šie dalykai kartu [209]. Jei naudojami tolygūs deskriptorių verčių intervalai, tada galima apibrėžti pritaikomumo domeną pagal tai, kaip jie dengia apmokymo duomenų rinkinį QSAR modelio deskriptorių aibėje [210].

1.5.6. QSAR deskriptoriai

Molekuliniai deskriptoriai yra skaitinės vertės, kurios charakterizuoja įvairias tiek išmatuojamas, tiek paskaičiuojamas tos molekulės savybes ir aprašo jos struktūrą. Moderniuose QSAR protokoluose įprasta naudoti platų rinkinį įvairių molekulinį deskriptorių, kurie gali išreikšti skaičiais skirtingus nagrinėjamos cheminės struktūros aspektus. Atliekant QSAR modeliavimą, visados pirma suskaičiuojami molekuliniai deskriptoriai visai nagrinėjamai cheminių junginių serijai. QSAR literatūroje ir QSAR programose minimi įvairūs jų tipai, pvz.: konstituciniai, topologiniai, elektrostatiniai, geometriniai, steriniai, fizikocheminiai ir kvantiniai. Galima rasti ir daugiau QSAR deskriptorių, kurie netelpa nei į vieną iš šių pagrindinių kategorijų. Kiekviena QSAR modeliavimo ar deskriptorių skaičiavimo programa savo turimų QSAR deskriptorių aibę klasifikuoja kiek savaip. Šiuo metu yra per tris tūkstančius

molekulinių deskriptorių, kurie gali būti suskaičiuoti vien iš cheminio junginio struktūros, jie apibendrinti ir aprašyti knygose [211-213].

Konstituciniai QSAR deskriptoriai skaitmeniniu būdu išreiškia modeliuojamų junginių cheminę sudėtį. Tai paprastai iš cheminio junginio struktūros suskaičiuojami deskriptoriai. Kaip paprasčiausias konstitucinis deskriptorius gali būti panaudojama cheminio junginio molekulinė masė. Paskaičiuoti įvairių atomų skaičiai molekulėje, pavyzdžiui, anglies, azoto atomų skaičius, taip pat dažnai būna naudojami kaip konstituciniai deskriptoriai. Elektrostatiniai, arba kitaip vadinami elektroniniai deskriptoriai, yra dar vienas galimas molekulinių deskriptorių tipas. Elektrostatiniai ar elektroniniai deskriptoriai yra tokie kaip poliškumas, poliarizuojamumas, dielektrinė energija, atomų daliniai krūviai, dipolio momentas. Elektronų pasiskirstymas cheminiame junginyje turi didžiulę įtaką to cheminio junginio biologiniam aktyvumui. Vien tam, kad cheminis junginys pasiektų savo biologinį taikinį, jis paprastai turi pirmiausiai pereiti per ląstelės membraną. Napoliniai ir besijonizuojantys cheminiai junginiai savo neutraliose formose paprastai lengviau pereina per membraną, ir beveik visi plačiai pardavinėjami ir naudojami vaistai yra tokie cheminiai junginiai. Atitinkamai, poliniai cheminiai junginiai ir besijonizuojantys cheminiai junginiai savo jonizuotose formose dažniausiai sulaikomi membranų. Kai cheminis junginys pasiekia savo biologinį taikinį, tos pačios jo elektrostatinės savybės jau kitaip svarbios antrą kartą: jos nulemia, kokios elektrostatinės tarpmolekulinės sąveikos susiformuos tarp to cheminio junginio ir biologinio taikinio. Tai daro įtaką jungimosi sąveikos energijai, cheminio junginio biologiniam aktyvumui.

Kitas QSAR deskriptorių tipas yra geometriniai deskriptoriai, dažnai vadinami steriniais arba erdviniais deskriptoriais. Šie deskriptoriai įvairiai aprašo nagrinėjamos molekulės struktūros geometrines savybes. Tai gali būti, pavyzdžiui, molekulės ilgis, plotis, paviršiaus plotas, tūris ir kiti panašūs dalykai, bei įvairiausios jų kombinacijos. Tam, kad nagrinėjama molekulė gerai jungtųsi su savo biologiniu taikiniu, jos išmatavimai turi būti komplementarūs

su taikinio aktyviojo centro geometriniais parametrais, todėl ir geometriniai QSAR deskriptoriai yra svarbūs QSAR modeliavime. Pirmasis garsus bandymas sukurti tokį deskriptorių buvo Tafto sterinis parametras, kuris matematiškai siejo ligando dydį ir formą su biologinio taikinio dydžiu ir forma ir šiuos dydžius su ligando/taikinio sąveikos stiprumu. Yra žymiai daugiau įvairių sterinių deskriptorių, pavyzdžiui, Chartono sterinis parametras ar Verloopo sterinis parametras [214].

Fizikocheminiai QSAR deskriptoriai aprašo įvairias molekulės fizikochemines savybes, kurios gali būti paskaičiuojamos iš struktūros, pavyzdžiui, lipofiliškumas, tirpumas vandenyje ar (rečiau) kituose tirpikliuose, absorbcija įvairiose biologinėse sistemose, ir kitos savybės [215]. Vėliau QSAR modeliavime buvo pradėti naudoti kvantiniai deskriptoriai. Tai gali būti HOMO (angl. *highest occupied molecular orbital*) ir LUMO (angl. *lowest unoccupied molecular orbital*) energijos, vibraciniai dažniai, ryšių eilės, bendra struktūros energija ir kiti iš kvantinių skaičiavimų gaunami dydžiai, aprašantys nagrinėjamus cheminius junginius.

Topologiniai deskriptoriai išvedami iš sudarytos molekulėje esančių cheminių ryšių lentelės. Paprasčiausių topologinių deskriptorių pavyzdys yra suminiai atstumai tarp įvairių atomų porų, išreikšti ryšių skaičiumi. E-DRAGON programoje tokie deskriptoriai yra žymimi T(X..Y), kur X ir Y tai nagrinėjamų atomų tipai. Be to, tai įvairūs molekulės simetrijos aprašai ir įvairūs vardiniai indeksai (Chi, Randic, Weiner ir t.t.).

Kai kurie deskriptoriai sunkiai priskiriami kokiam nors vienam tipui, nes apjungia kelias molekulės savybes į vieną visumą, todėl kartais vadinami hibridiniais deskriptoriais. Tai, pavyzdžiui, molekulės polinio paviršiaus plotas, tik tam tikrų jos dalių tūris, molekulinio lauko deskriptoriai (angl. *molecular field descriptors*) [216], skaitmeniškai išreikšti farmakoforai ir pan. Vėliau buvo kuriami ir naudojami elektrotopologiniai deskriptoriai, WHIM indeksai, BCUT [217], GETAWAY deskriptoriai ir kiti [218, 219].

Pagal savo matematinės savybes deskriptoriai rūšiuojami į tolygius,

kurie gali įgyti įvairias vertes, pavyzdžiui, molekulos tūris, ir diskrečiuosius, kurie gali įgyti tik tam tikras fiksuotas vertes ir neretai lygūs tik vienetui arba nuliui priklausomai nuo junginio struktūros bruožų, pavyzdžiui, *orto* pakaitų deskriptorius: jei yra *orto* pakaitų, jis lygus vienetui, kai jų nėra, tai jis lygus nuliui. Pasitaiko, kad diskretieji deskriptoriai lygūs nuliui visiems turimiems serijos junginiams ir tuomet jų naudojimas yra beprasmiškas. Kai kurie deskriptoriai būna išreiškiami ir kompleksiniais skaičiais. Tai pritaikoma, pavyzdžiui, molekulos chiralizmui aprašyti dvimačio QSAR modeliuose [220].

Pavieniai deskriptoriai rankiniu būdu jau beveik niekad neskaičiuojami. Sukurta daug programų, laisvai prieinamų internete, su kuriomis lengvai galima paskaičiuoti šimtus ir net tūkstančius deskriptorių turimoms cheminių junginių struktūrinėms formulėms, užduodamoms kaip įvesties failas (paminėtinos iš jų: E-DRAGON, CDK TOOL, POWER MV, MOLD2, PADEL-DESCRIPTOR). Yra įvairios komercinės programinės įrangos tiems tikslams (paminėtinos iš jų: PHASE, SYBYL, SYBYL-X, BIOVIA QSAR WORKBENCH, COSMOSAR3D, CODESSA). Minėtose komercinėse programose visi reikiami QSAR modeliavimo įrankiai yra dažniausiai integruoti į vieną programinį paketą, jos retai apsiriboja vien deskriptorių skaičiavimais. Kita vertus, cheminės informatikos internetinių serverių nėra tiek daug, kaip bioinformatinių, nes didžiausios cheminių junginių duomenų bazės, tokios kaip IRESEARCH, su keliasdešimt milijonų junginių užima vos keliasdešimt ar kelis šimtus megabaitų SMILES failo formate (<http://www.chemnavigator.com/cnc/products/iRL.asp>). Cheminių informatinių skaičiavimų apimtys nėra didelės [221].

Viena iš naujausių deskriptorių skaičiavimo programų yra CHEMDES, publikuota Dong ir bendraautorių 2014 ir 2015 metais (scbdd.com/chemdes/) [222, 223]. Šis internetinis serveris kartu skaičiuoja ir seniau literatūroje aprašytus deskriptorius, ir deskriptorius iš kitų programų – iš viso 3679 skirtingus molekulinis deskriptorius, surūšiuotus į 61 kategoriją.

Įvairiose QSAR metodologijose deskriptoriai yra dažnai skaliuojami (normalizuojami) taip, kad visi turėtų vienodą įtaką tolesnėje analizėje. Viena iš tokių technikų yra vidurkio centravimas: iš deskriptoriaus verčių visiems junginiams atimamas visų jų vidurkis. Tuomet skaliuotos deskriptoriaus vertės turi vidurkį, lygų nuliui. Paprastai QSAR modeliavimo pradžioje išmetami deskriptoriai, kurie poromis koreliuoja tarpusavyje. Laikoma, kad du tarpusavyje koreliuojantys deskriptoriai kiek skirtingai perteikia tą pačią informaciją apie cheminę struktūrą, todėl juos kartu naudoti nėra prasmės. Nėra vienareikšmio būdo, kaip tai padaryti: jei du deskriptoriai koreliuoja tarpusavyje, vienas iš jų parenkamas išmetimui atsitiktiniu būdu.

1.5.7. Trimatis, keturmatis ir penkiamatis QSAR

Cramer su bendradarbiais 1988 metais sukūrė pirmą trimačio QSAR metodą, vadinamą palyginamąja molekulių laukų analize (CoMFA) [224]. CoMFA metodas tapo pradžia visai eilei įvairių QSAR metodų, kurie dabar bendrai vadinami trimačio QSAR metodais. Trimačio QSAR (dar vadinamo 3D-QSAR) metodai – tai natūrali įprasto QSAR metodo tęsia. Jie atsižvelgia į tai, kad nagrinėjamos molekulės yra erdvinės, trimatės struktūros ir pagal erdvinės struktūros bruožus prognozuoja cheminių junginių biologinį aktyvumą. Tai yra privalumas, bet sukelia ir sunkumų metodo taikyme. Trimačio QSAR atveju visos molekulės užklojamos viena ant kitos, ir yra skaičiuojami trimatėje erdvėje jas aprašantys deskriptoriai. Dažniausiai tie deskriptoriai remiasi funkcinių grupių padėtimi trimatėje gardelėje, paversta į skaičius. Taip matematiškai perteikiami junginio stereocheminiai elementai, ir nagrinėjamas jų santykis su junginio savybėmis [225]. Tas svarbiausia, prognozuojant ligandų sąveiką su baltymais. Be to, trimačių QSAR metodų rezultatai paprastai susideda ne vien iš statistinio vertinimo ir prognozuojamų biologinio aktyvumo verčių, bet ir pateikiami kaip trimatė grafinė iliustracija, užklota ant nagrinėjamų junginių erdvių struktūrų, kas tyrinėtojiui padeda lengviau suprasti, kokios junginio grupės, kokią įtaką daro jo biologiniam aktyvumui. Be jau minėto CoMFA, yra ir kitų trimačio QSAR metodų, pavyzdžiui, LISA [228]. Trimačio QSAR atsiradimas skatino žengti toliau, ir

buvo sukurti dar aukštesnių matavimų QSAR. Gerindami molekulių formų analizės (MSA) metodą, 1997 metais Hopfinger su kolegomis pasiūlė keturmatį QSAR metodą [226], kuris, be trimačių molekulių struktūrų, dar atsižvelgia ir į nagrinėjamų cheminių struktūrų konformacinę laisvę. Tai yra keturmačio QSAR metodo „ketvirtasis išmatavimas“. Su apmokymo rinkinio junginių struktūromis yra atliekama molekulinės dinamikos simuliacija, ar kitaip generuojamas konformacinių ansamblių profilis, ir iš jo matematiškai išreiškiamas jų konformacinių kitimų dažnis, kuris paskui konvertuojamas į keturmačio QSAR deskriptorius [227]. Penkiamačio QSAR metode, be konformacinių ansamblių profilio, bandoma žengti toliau ir sukurti indukuoto atitikimo modelį. Šešiamatį QSAR metode, be viso, kas yra penkiamačio QSAR metode, dar atsižvelgiama į nagrinėjamų junginių paskaičiuotąją solvatacijos energiją. Vienas iš naujausių trimačio QSAR metodų yra LISA, kuriame, lygiai kaip ir trimačiame, QSAR deskriptoriai naudojami nagrinėjamų cheminių struktūrų lokalaus panašumo įverčiai (LSI), bet LISA metode dar galima generuoti pseudoreceptorių, jei nėra jokios informacijos apie nagrinėjamą baltymą-taikinį [228]. Visi vėliau atsiradę QSAR metodai bandė išspręsti esmines pirmųjų trimačių QSAR metodų problemas. Skirtingų QSAR metodų autoriai išvelgė kiek skirtingas trimačio QSAR problemas ir pasiūlė skirtingus patobulinimus.

1.5.8. QSAR taikymas

QSAR metodai, kurie iš esmės liko artimi pirmiesiems Free, Wilson, Hansch ir Fujita darbams, yra vadinami klasikinais QSAR metodais. Jie nenagrinėja ir nenaudoja turimos cheminių junginių serijos erdvių struktūrų, erdvių konformacijų ar iš jų išvestų sudėtingesnių matematinių erdvinės cheminės struktūros aprašų, pavyzdžiui, hologramų. Kita vertus, jie gali turėti įvairių kitokių patobulinimų. Klasikiniai QSAR metodai greitesni, tinkamesni biologiškai aktyvių junginių paieškai didelėse duomenų bazėse, jie yra paprastesni ir lengviau automatizuojami, nei trimatis QSAR ar panašūs metodai. Klasikiniai QSAR metodai kartais skirstomi į vienmatį QSAR, jei turimame QSAR modelyje kaip deskriptoriai taikomos vien tik globalios,

paskaičiuojamos cheminio junginio savybės, ir dvimatį QSAR, jei turimame QSAR modelyje naudojami įvairūs deskriptoriai, paskaičiuoti remiantis dvimate junginio struktūrine formule. Įprasta suvienyti vienmatį QSAR ir dvimatį QSAR į vieną visumą ir įtraukti kiek deskriptorių apie pakaitų išsidėstymą erdvėje ar kitų panašių dalykų, todėl tokie metodai vadinami tiesiog QSAR.

Trimatis QSAR yra tuomet, jei QSAR modelyje naudojamos cheminių junginių erdvinės struktūros formulės. Dažnai pasitaiko, kad cheminių junginių trimatė konformacija, kuri nulemia biologinį aktyvumą, nėra žinoma. Jei biologiškai aktyvios konformacijos negalima sužinoti iš kristalografijos ar branduolių magnetinio rezonanso eksperimentinių duomenų bazės, nagrinėjamų junginių erdvinė konformacija generuojama molekulinės mechanikos, pusiau empiriniais kvantiniais metodais arba *ab initio* kvantinės chemijos metodais. Taip pat ji gali būti prognozuojama, naudojant molekulių fragmentų bibliotekas (<http://www.lifechemicals.com/services/fragments>). Jei būtų panaudojama kokia nors kita, ne biologiškai aktyvi junginių erdvinė konformacija, tai vestų link klaidingų trimačio QSAR metodo taikymo rezultatų. Prieš atliekant trimatį QSAR, visa nagrinėjamų junginių serija perklojama taip, kad kuo daugiau bendrų struktūrinių elementų atsidurtų toje pat vietoje erdvėje. Ne toje pačioje vietoje erdvėje likę junginių fragmentai yra laikomi skirtingais. Neretai net ir mažų cheminių junginių perklojimo uždavinys neturi vienareikšmio sprendimo. Problema žymiai sudėtingėja, kai nagrinėjamose molekulėse daugėja lanksčių fragmentų. Skirtingos programos ir skirtingi metodai gali junginius perkloti skirtingai. Kai sutapatinami vienodi fragmentai, neretai lieka įvairių galimų variantų, kurie likę struktūrų pakaitai, tada turi būti panašiai toje pat erdvės vietoje. Galiausiai atliekama molekulių laukų analizė. Erdvėje aplinkui molekulės struktūrą programose nustatyti žingsniu, kuris dažniausiai lygus 2 Å, yra stumdoma cheminė grupė-zondas ir skaičiuojama kiekvieno nagrinėjamo cheminio junginio struktūros fragmento sąveika su ja. Kaip grupė-zondas yra naudojami vandens molekulė, vandenilio

jonas (protonas), hipotetinis atomas su tam tikrais parametrais ir panašiai. Dažniausiai yra skaičiuojama tarpmolekulinė, elektrostatinė ir van der Valso sąveika. CoMFA metode van der Valso sąveikai skaičiuoti pasitelkiamas standartinis (6–12) Lenardo-Džonso potencialas arba kiti analogiški potencialai. Tarpmolekulinė sąveika čia skaičiuojama su ribinėmis vertėmis, antraip būtų gaunami nepageidaujami rezultatai: judančiai grupei-zondui susidūrus su nagrinėjamo cheminio junginio struktūra, gali būti paskaičiuotos aukštos sąveikos vertės, o nežymiai pakeitus to junginio konformaciją – visai kitokios. Keblu tai, kad galutinis rezultatas priklauso nuo to, kaip erdvėje, molekulės struktūros atžvilgiu, išdėstomi minėti žingsniai, ir nuo kurio taško jie pradedami skaičiuoti. Tam spręsti yra naudojamos įvairios strategijos ar išbandomi įvairūs variantai, kol pasiekiamas geriausias rezultatas. Turimos paskaičiuotos tarpmolekulinės sąveikos su grupe-zondu vertės apdorojamos, centruojamos ir skaliuojamos vienu ar kitu metodu (iš kiekvienos vertės atimamas vidurkis, ji padalinama iš standartinio nuokrypio ir panašiai), kad būtų pasiskirsčiusios vienodame intervale. Galiausiai, statistiniu metodu ieškoma koreliacija tarp turimų paskaičiuotų duomenų ir eksperimentiškai nustatytų biologinių sąveikų konstantų tai pačiai junginių serijai [229]. Naudojami visi tie patys tiesinės ar netiesinės priklausomybės modeliavimo metodai, kaip ir tradiciniame QSAR: MLA, PCA, PCR, neuronų tinklai ir kiti. Po statistinės analizės, rezultatai yra pateikiami grafiškai: šalia nagrinėjamos serijos cheminio junginio struktūros ar jų rinkinio kubeliais arba erdviniu paviršiumi yra vaizduojamos statistiškai reikšmingos, koreliuojančios su eksperimentiniais duomenimis, tarpmolekulinės sąveikos su grupe-zondu. Atvaizduoti elementai yra nuspalvinami sutartinėmis spalvomis pagal koreliacijos ženklą ir dydį. Be to, pateikiamas grafikas koreliacijos tarp eksperimentinių biologinės sąveikos verčių ir paskaičiuotų iš trimačio QSAR modelio. Yra galimybė prognozuoti ir kitų dar nenagrinėtų cheminių junginių biologinį aktyvumą.

Klasikinis QSAR, kuris atliekamas nenaudojant erdvinių cheminių

junginių struktūrų modelių, išskyrus tik kokius nors stereocheminius deskriptorius, neprarado svarbos. Iki šių dienų QSAR įprasta atlikti tiek su dvimačiais, tiek su trimačiais molekulių modeliais. Klasikinis QSAR modeliavimas gali būti atliekamas, net jei nežinoma biologinio taikinio struktūra ar net pats biologinis taikynys. Pasiūlyti kiti patobulinti metodai, besiremiantys seniau sukurtais dvimačio QSAR metodais, tokie kaip molekulės hologramos QSAR (HQSAR), fragmentais besiremiantis QSAR (FB-QSAR), fragmentų panašumais besiremiantis QSAR (FS-QSAR), aukščiausios svarbos fragmento QSAR (angl. *Top Priority Fragment QSAR*) ir kiti fragmentais paremti QSAR metodai [230]. Taikant HQSAR metodą, kiekviena molekulė, esanti apmokymo rinkinyje, išardoma į fragmentus. Iš visų tų fragmentų ir jų pasikartojimo dažnio sudaromos molekulių hologramos. Atlikus molekulių hologramų sąryšio su biologiniu aktyvumu statistinę analizę, sudaromi HQSAR modeliai [231]. Daugybė sėkmingų QSAR metodologijų taikymų parodė, kad klasikinio QSAR ir daugiamačio QSAR metodai yra tinkamas įrankis vaistų kūrime. Neretai sukurtas QSAR metodas pripažįstamas tinkamu, tik papildomai išbandžius jo kokybę įvairiais kitais skaičiuojamaisiais metodais. QSAR metodai naudojami ne vien vaistų kūrime, bet ir kituose biologiniuose, cheminiuose ir toksikologiniuose tyrimuose. QSAR yra atliekamas visur, kur reikia prognozuoti dar nesusintetintų cheminių junginių savybes arba rasti cheminių junginių duomenų bazėse egzituojančius junginius su norimomis savybėmis [232-236]. Galimybė pasitelkti QSAR modelius, kuriant naujoviškus cheminius junginius, yra viena iš svarbiausių strategijų vaistų kūrime, kurios dėka galima parinkti iš esamų ar atrasti daug žadančius junginius ankstyvoje vaistų kūrimo proceso stadijoje, pagreitinti pradinių junginių tobulinimą į vaistus-kandidatus [231]. QSAR tyrimuose atrandant perspektyviausius junginius *in silico*, galima sumažinti brangių nesėkmių skaičių, kurios kyla sintetinant mažai žadančius, neperspektyvius cheminius junginius ir matuojant tų junginių biologinį aktyvumą. Apibendrinant, QSAR modeliavimo metodai nuo pradinių į QSAR panašių darbų XIX amžiaus pabaigoje iki dabar patyrė evoliuciją, QSAR metodo taikymas gausiai išplito, ir

buvo sukurta daug QSAR metodo variantų. QSAR taip pat buvo taikomas įvairių ligandų, besijungiančių prie CA II, jungimuisi aprašyti [237-240].

1.5.9. QSAR metodo kritika

Norint sukurti aukštos kokybės, naudingą QSAR modelį, pirmiausia yra svarbu atkreipti dėmesį į naudojamų eksperimentinių duomenų kokybę. Akivaizdu, kad iš prastų eksperimentinių duomenų rinkinio, kuriame daug paklaidų arba yra klaidų, bus sudarytas nekokybiškas QSAR modelis. Populiarus „GIGO“ (angl. *garbage in, garbage out*) pasakymas puikiai tinka ir bet kokiam QSAR modeliavimui. Visi eksperimentiniai duomenys turėtų būti gauti, naudojant tuos pačius eksperimentinius metodus, tomis pačiomis sąlygomis, pavyzdžiui, buferinio tirpalo sudėtis, pH, temperatūra ir t.t. Dar geriau, jei visi tam pačiam QSAR modeliavimui imami duomenys būtų gauti toje pačioje laboratorijoje, laikantis vienos, kiek galima labiau identiškos, eksperimentinės praktikos. Taip išvengiama papildomų nepageidaujamų fizikocheminių įtakų, kurios būtų nevienodos visų aprašytų junginių biologiniam aktyvumui. Didelė painiava QSAR modeliavime yra tai, kad literatūroje esančios cheminių junginių biologinių aktyvumų vertės gali labai skirtis tam pačiam junginiui [241]. Pavyzdžiui, duomenų bazėse, tokiose kaip PDB ar BINDINGDB, pateikiama acetazolamido, besijungiančio prie CA II, išmatuota K_i vertė nuo 3,3 nM [242] iki 36 μ M [243]. Dar blogesnė situacija gali būti su IC_{50} vertėmis, kurios, be kitų veiksnių ir slopiklio cheminės prigimties, priklauso ir nuo naudojamo substrato koncentracijos. Negana to, duomenų rinkinyje turi būti užtektinai junginių, kurių aktyvumas turi ženkliai skirtis [244], tūkstantį ar daugiau kartų. Jei duomenų rinkinyje yra keli junginiai, kurie visi skiriasi keliais struktūriniais bruožais, bet turi panašų biologinį aktyvumą, vargu ar pavyks rasti struktūrinius bruožus, darančius įtaką biologiniam aktyvumui. Biologinio aktyvumo dydžiai junginių serijoje turi būti tolygiai pasiskirstę: jei serijoje bus tik vienas išskirtinai aktyvus junginys, tai QSAR modeliavimo rezultatai tiek ir pasakys, kad visi jo struktūriniai bruožai, išskiriantys jį iš kitų junginių, yra svarbūs. Tai nebūtinai gali būti teisinga. Netolygumas gali būti bent iš dalies pašalintas, turimas

eksperimentines biologinio aktyvumo konstantas K logaritmuojant. Būtų idealu, jei tiesiog iš naujo sudaromas duomenų rinkinys su tolygiau pasiskirsčiusiomis biologinio aktyvumo vertėmis. Tą galima padaryti atlikus papildomus matavimus su kitais junginiais. Geriausi rezultatai gaunami, jeigu apmokymo serijos junginiai yra giminingi tarpusavyje, jungiasi į to paties biologinio receptoriaus tą pačią vietą, bei turi vienodą erdvinę konformaciją. Tai ypač svarbu trimačio QSAR metoduose [245].

Jei į galutinį QSAR modelį įtraukiama labai daug parametrizavimo koeficientų ir deskriptorių, tai sukurtasis modelis gali dirbtinai sujungti parametrizavime naudotus eksperimentinius taškus, ir turėti įvairius aukštus statistinius įverčius, bet faktiškai iš tokio QSAR modelio mažai naudos. Jo tinkamumas prognozėms bus niekinis, mat tame modelyje slypi tik ir taip žinomi, apmokyme naudoti, eksperimentiniai duomenys išreikšti kitaip, per deskriptorius ir parametrizavimo koeficientus. Kuriant QSAR modelius, reikia laikytis „Okamo skustuvo“ principo, ir tai šiuo atveju reiškia, kad, pirma, reikia kurti kiek įmanoma paprastesnį matematinį QSAR modelį, ir antra, jau sukūrus QSAR modelį, išnagrinėti, kas jame nėra tikrai svarbu ir pašalinti. Sudėtingas modelis adaptuojasi prie triukšmo, esančio eksperimentiniuose duomenyse ir gali tapti apgaulingai, perdėtai optimistinis.

Toks reiškinys vadinamas QSAR modelio permokymu (angl. *overfitting*). Jis yra gerai žinomas ir plačiai aprašytas mokslinėje literatūroje [246-248]. Siekiant to išvengti QSAR modeliavime, dažnai laikomasi empirinės taisyklės, kad QSAR apmokymo rinkinyje turi būti bent penkis kartus daugiau cheminių junginių, nei iš jo sukurtame modelyje deskriptorių [249]. Juk du taškai tiksliai apibrėžia liniją, trys taškai tiksliai apibrėžia plokštumą ir t.t. Grubiai pažeidus minėtą taisyklę, gauta koreliacija bus beprasmė ir niekam nenaudinga [250]. Dar ir dabar literatūroje pasitaiko tokių mažai prasmingų, atsitiktinių matematinių sąryšių, kylančių iš permokymo. QSAR modeliavimo metodai jau naudojami daugiau kaip penkis dešimtmečius, bet dar nėra vieningos nuomonės, koku būdu ar koku

statistiniu vertinimu reiktų nuspręsti, kurie QSAR modeliai yra geriausi. Skirtingi QSAR metodo taikytojai naudoja kiek skirtingus statistinius kriterijus ir empirines taisykles nuspręsti, kurie jų QSAR modeliai geriausi [251-252]. Egzistuoja ir modelių parinkimo šališkumo reiškinys. Idealiu atveju tik pats geriausias QSAR modelis turi mažiausias paklaidas eksperimentinių duomenų atžvilgiu. Deja, gali atsitikti, ypač su mažesniais duomenų rinkiniais, kad dėl permokymo (atsitiktinio sutapimo) keli eksperimentiniai taškai yra arti paskaičiuotųjų. Tada jau nebe tinkamiausias prognozėms QSAR modelis turės geriausią statistinį įvertį [253]. To pasekmėje tinkamiausias prognozėms QSAR modelis lieka atmestas. Tai galima pastebėti tik su nepriklausomu ir modeliui kurti ar testuoti iki tol nepanaudotu testavimo rinkiniu. Įvairios, dažnai naudojamos, be kitų papildomų priemonių, QSAR statistinio tinkamumo patvirtinimo metodologijos visgi minimizuoja tam tikrą statistinį parametą ir taip daro iškreiptai teigiamą įtaką galutiniam to parametro dydžiui su apmokymo rinkiniu. Šis nepriklausomo vertinimo, testavimo rinkinio, nebuvimas daro statistinio vertinimo rezultatus nepatikimais ir iškreiptais. Kai testavimas atliekamas su nepriklausomu testavimo rinkiniu, ši problema yra išsprendžiama.

Kita problema QSAR modeliavime yra ta, kad koreliacija nebūtinai reiškia priežastinį ryšį tarp koreliuojančių duomenų. Imtame junginių rinkinyje gali būti gana aiški koreliacija tarp kokios nors cheminio junginio struktūrinės savybės ir biologinio aktyvumo, bet tai nebūtinai reiškia, kad keičiant tą savybę, bus proporcingai pakeistas junginų biologinis aktyvumas. Priežastinis ryšys tarp dviejų dydžių yra tik tada, kai pademonstruojama, jog keičiant vieną iš jų, kartu keičiasi ir kitas. QSAR modeliai tik pademonstruoja, kaip struktūra koreliuoja su biologiniu aktyvumu.

QSAR metodų galutinių rezultatų tinkamumas prognozėms priklauso nuo pradinės informacijos ir įvairių išorinių faktorių, o ne vien tik nuo pačio modeliavimo metodo ir eigos. Visų pirma, jis priklauso nuo modeliuojamo biologinio aktyvumo pobūdžio, prigimties, jo išmatavimo tikslumo. Antra, nuo

nagrinėjamų junginių rinkinių dydžio. Trečia, jei tie junginiai sintetunami ir išbandomi sėkmingai, tai kartu ir modeliavimo metodo efektyvumas bus papildomai patvirtintas. Jeigu tie junginiai sintetunami ir išbandomi nesėkmingai, tai ir metodo efektyvumas liks abejotinas. Su tam tikrų serijų junginiais tai padaryti gali būti sudėtinga ir brangu. Dažnai QSAR modeliai, jau patvirtinti su iš anksto turimu testavimo rinkiniu, būna teisingi, bet kartais būna, kad jie taip ir nepasiteisina, nepaisant visos eilės teigiamai juos aprašančių statistinių įverčių apmokymo rinkiniui. Kiti esami skaičiuojamieji metodai nepasirodo efektyvesni už QSAR [254]. Aiškinamasi, kodėl kartais QSAR nepasiteisina. Huang rašo, kad siekiant išvengti nesėkmių, reikia sukurti gerokai griežtesnes ir efektyvesnes QSAR modelių vertinimo sistemas, ir kad modeliai, sukurti naudojant deskriptorius, kurie linkę kartotis ekvivalentiškuose modeliuose, sukurtuose lygiai tai pačiai biologinei sistemai nagrinėti, yra daug daugiau žadantys [255].

Ne visi QSAR modeliai vienodai lengvai interpretuojami. Tiesines lygtis interpretuoti lengviau, nei sudėtingus ir modernius algoritmus, tokius kaip neuronų tinklai ir kiti. Neklasikinis QSAR metodas gali pateikti lengvai įsisavinamus įrankius, kuriais galima prognozuoti nežinomų junginių savybės, biologinį aktyvumą, bet jei nėra aišku, kaip tas įrankis veikia, ir kodėl pateikia vienokius ar kitokius rezultatus, tai krenta pasitikėjimas juo, nepaisant to, kad svarbiausia yra tai, kaip tas modelis tinka prognozėms. Iš sunkiai interpretuojamų QSAR modelių lieka neatsakytai klausimai, kaip galima didinti turimų junginių aktyvumą, kas jį lemia, kodėl junginiui prognozuojamas aktyvumas [256].

Grynai tiesiniai QSAR modeliai, apmokyti naudojant labai siaurą cheminių junginių seriją, kuriuose visi parametrai yra tiesiniai, vargu ar patenkinamai aprašys cheminių junginių serijos aktyvumus, kai yra reikšmingas skaičius junginių, kurių biologinis aktyvumas netikėtai ir gerokai didesnis ar mažesnis už kitų junginių aktyvumus. Tokių junginių biologinis aktyvumas vadinamas išskirtimis (angl. *outliers*). Jos gali atsirasti dėl įvairių

priežasčių, pavyzdžiui, statistinių fluktuacijų, matavimo klaidų ar taip vadinamų „aktyvumų skardžių“. Aktyvumo skardžiai atsiranda tada, kai vieno junginio aktyvumas netikėtai ir labai smarkiai skiriasi nuo kitų į jį panašių junginių. Net visiškai teisingi duomenys gali turėti išskirčių. Tuomet labai svarbu išnagrinėti bent kelis junginius, kuo panašesnius į tą vieną, kuris yra „aktyvumo skardyje“, kad būtų tinkamai įtraukti visi „aktyvumo skardžio“ aprašai į vėliau sudaromą QSAR modelį. Taip sudarytas QSAR modelis tada neturės netikslumų dėl netinkamai aprašytų išskirčių ir tinkamai aprašys tokius struktūros-aktyvumo sąryšio duomenis [257].

1.6. Kiti metodai

Tarp kitų metodų, kurie nebuvo naudoti šiame darbe, bet kuriuos verta paminėti, yra MMGBSA (Molekulinė mechanika su generalizuota Borno elektrostatika ir paviršiaus ploto hidrofobine pataisa) ir su juo susijęs MMPBSA (Molekulinė mechanika su Puasono-Bolcmano elektrostatika ir paviršiaus ploto hidrofobine pataisa). MMGBSA paremtos ΔG_{bind} vertės ligandams gerai koreliuoja su eksperimentiškai nustatytomis [258, 259]. Srivastava su bendraautorais ištyrinėjo MMPBSA ir MMGBSA tinkamumą prognozuoti ΔG_{bind} vertėms, kurios paskui galėtų būti naudojamos naujų ligandų atradime [260]. Pašalinus chemines struktūras su peptidiniais, polisacharidiniais, turinčiais kelis fosforo atomus ligandais (daugiausiai kofaktoriais), metalo jonais taikinio aktyviajame centre bei kitais klaidų šaltiniais, MMGBSA metodu gauta gana nebloga koreliacija ($R^2 = 0,63$) su ligando ir receptoriaus jungimosi pK , imtomis iš PDBBIND (<http://sw16.im.med.umich.edu/databases/pdbbind/index.jsp>) duomenų bazės [261]. MMGBSA ir MMPBSA buvo taikomas slopiklių jungimuisi prie CA aprašyti [262, 263]. Dar paminėtina pora neempirinių, fizikocheminių metodų, kurie yra taikomi paskaičiuoti jungimosi energijas iš MD simuliacijų: termodinaminė integracija (TI) [264] ir laisvosios energijos perturbacija (FEP) [265].

2. Metodai

2.1. α -Pakeistų α,β -nesočiujų ketonų antiproliferacinio aktyvumo žmogaus hematologinėse ir kietose vėžinių ląstelių linijose QSAR modelių kūrimas

Trimačiai junginių struktūrų modeliai buvo surinkti ir optimizuoti molekulinės mechanikos metodu su AVOGADRO programa (versija 1.1.1) (http://avogadro.cc/wiki/Main_Page) [266]. Junginių rinkinyje 1 kai kurių junginių aktyvumai buvo nustatyti raceminiam mišiniui, su nežinomais aktyvumais atskiriems enantiomerams. Visoms toms molekulių struktūroms priskirtas tas pats chiraliskumas ir tas pats pakaitų išsidėstymas erdvėje, siekiant išvengti papildomo triukšmo. Generuoti molekulių struktūrų failai buvo su OPEN BABEL programa (versija 2.3.1) [267] apjungti į vieną *sdf* formato failą, kuris buvo įkeltas į programą E-DRAGON molekuliniais deskriptoriams skaičiuoti [268]. E-DRAGON skaičiuoja daug rūšių, įvairių, dvimačių ar trimačių, deskriptorių [213]. Pašalinus deskriptorius, kurių vertės visiems imtiems serijos junginiams buvo vienodos, liko galutinis 1666 struktūrinių deskriptorių rinkinys. Tinkami junginių rinkinio pIC_{50} priklausomybei nuo tų junginių cheminių struktūrų aprašyti deskriptoriai buvo atrinkti ir daugybinės tiesinės regresijos QSAR modeliai buvo sukurti su R programinės įrangos paketu (<https://www.r-project.org/>), naudojant LEAPS programą. LEAPS perspėjo apie deskriptorių tarpusavio tiesines koreliacijas. Deskriptoriai, kurie koreliuoja tarpusavyje, ir modeliai, sudaryti su tarpusavyje koreliuojančiais deskriptoriais, buvo automatiškai atmesti. Galutinės QSAR lygtys buvo sudarytos, naudojant po penkis deskriptorius pagal bendrai priimtą taisyklę, kad pradiniam junginių rinkinyje turi būti bent penkis kartus daugiau junginių, negu iš tų junginių sukurtame QSAR modelyje deskriptorių [249].

2.2. Benzensulfonamidų rinkinio dokinimas

3.2.1 skyrelyje aprašytas keturiasdešimties benzensulfonamidų struktūrų rinkinys buvo naudojamas dokinimui su programa ARGUSLAB atlikti (<http://www.arguslab.com/arguslab.com/ArgusLab.html>) [269]. CA II struktūra (PDB ID: 3MYQ), paruošta su programos UCSF CHIMERA (versija 1.8)

(<https://www.cgl.ucsf.edu/chimera/>) paruošimo dokinimui procedūra [270], buvo pasirinkta kaip dokinimo taikiny. ARGUSLAB programoje stačiakampiu gretasieniu buvo pažymėta apytikslė ligando prisijungimo vieta. Programoje buvo pasirinkti aukšto dokinimo tikslumo ir lankstaus dokinimo nustatymai. Tas pats ligandų rinkinys buvo naudojamas ir dokinimui su GLIDE [271] atlikti. GLIDE yra dalis MAESTRO (versija 2014.2) programinio paketo. Ta pati CA II struktūra (PDB ID: 3MYQ), ir visas ligandų rinkinys buvo iš naujo paruoštas dokinimui, naudojant MAESTRO (versija 2014.2) programinio paketo įrankius. Dokinimas buvo atliktas su priimtais pagal nutylėjimą nustatymais, išskyrus tai, kad dokinimo tikslumas nustatytas į aukščiausią („extra“), ir kad kaip galimi buvo nustatyti donoriniai-akceptoriniai ryšiai tarp ligandų ir cinko jono, esančio CA II aktyviajame centre. GLIDE programoje donoriniai-akceptoriniai ryšiai nesudaromi priverstinai, taigi ir jungimasis be jų taip pat buvo prognozuojamas, kaip vienas iš galimų variantų.

2.3. Tiesinės sąveikos energijos skaičiavimai

Visi MD skaičiavimai, reikalingi tolesniems LIE skaičiavimams buvo atlikti su programa GROMACS (versija 4.6.2) [272]. CA II struktūra (PDB ID: 3MYQ) buvo paruošta su UCSF CHIMERA programos (versija 1.8) paruošimo dokinimui įrankiu [270]. AMBER *FF99SB* [273] ir GAFF (GENERAL AMBER FORCE FIELD) [274] jėgų laukai buvo naudojami atitinkamai baltymui ir ligandams. Tam, kad ligandas neiškeliautų iš savo prisijungimo vietos per MD simuliaciją, koordinacinis ryšys buvo suformuotas tarp cinko jono, esančio CA II aktyviajame centre, ir ligandų sulfonamido grupių azoto atomų. Programos algoritme nebuvo galimybės jo nutraukti jau paleistoje simuliacijoje. Tarpmolekulinės sąveikos parametrai R ir ϵ cinko jonui aprašyti buvo paimti iš literatūros [275]. Formalus cinko jono krūvis buvo nustatytas kaip +1 pagal literatūroje aprašytus kvantinės chemijos skaičiavimus [275]. Parametrai cheminiams ryšiams, jų kampams ir diedriniams kampams tarp cinko jono ir aplinkinių trijų histidino liekanų ir prisijungusio sulfonamido azoto atomo buvo paimti iš literatūros [276] ir [277]. Taip sudaryta ligando – taikinio (CA II) sistema buvo įdėta į stačiakampio gretasienio formos

simuliacijos zoną su vandens molekulėmis likusiai erdvei užpildyti. Buvo pridėta papildomų natrio ar chloro jonų, siekiant neutralizuoti visą sistemą. Visiems keturiasdešimčiai ligandų buvo imti tie patys simuliacijos parametrai. Minimalus simuliacijos laiko intervalas (žingsnis) buvo pasirinktas 2 fs. Simuliacija buvo vykdoma esant 300 K temperatūrai. Prieš MD simuliacijas buvo atliktos 800 žingsnių greičiausio nusileidimo minimizacija ir trumpa paruošiamoji 80 ps MD simuliacija su įtvirtintais baltymo sunkesniais atomais, kad vandens molekulės bei vandenilio atomai, kurių nesimatė rentgeno struktūroje, optimizuotų vandenilinių ryšių tinklą. Du 1 ns trukmės MD skaičiavimai buvo atlikti kiekvienam iš keturiasdešimties ligandų: vienas baltymo-ligando kompleksui vandenyje, antras laisvam ligandui vandenyje. Yra žinoma, kad CA receptorius nepasižymi plastiškumu [18], todėl buvo laikoma, kad 1 ns skaičiavimas yra adekvatus ilgio. Per darbinius MD skaičiavimus slėgio sąvarža buvo atlikta pagal Parinelo-Ramano algoritmą [278, 279]. Temperatūros sąvarža atlikta pagal „V-rescale“ algoritmą [280]. Evaldo dalelių tinklo (PME) [281] algoritmas buvo taikomas ilgo nuotolio elektostatinei sąveikai skaičiuoti su maksimaliu pasirinktu atstumu 9 Å. Tas pats maksimalus 9 Å atstumas pasirinktas ir van der Waals sąveikoms skaičiuoti. Galiausiai ΔG_{bind} buvo aproksimuota pagal lygtį 1.3-1. Įeinantys į lygtį energijų komponentai, buvo skaičiuojami kaip visų MD žingsnių vidurkis, naudojantis GROMACS įrankiais.

2.4. Metadinamikos skaičiavimai

Sistema buvo paruošta metadinamikos skaičiavimams, naudojant tą pačią procedūrą, kaip ir tiesinės sąveikos energijos skaičiavimams. Metadinamikos skaičiavimai buvo atlikti, naudojant PLUMED įskiepį (versija 2.1.0) [282], sukompiliuotą kartu su GROMACS programos versija 4.6.7. Metadinamikos simuliacijoms donorinis-acceptorinis ryšys tarp cinko jono ir ligando nebuvo suformuotas, nes metadinamikos procedūrai reikia, kad ligandas galėtų judėti laisvai ir palikti savo prisijungimo vietą. Buvo padaryta

prielaida, kad minėto ryšio pašalinimas neturėtų būti didelė problema, nes visiems keturiasdešimčiai ligandų tas ryšys turėtų būti apytiksliai vienodo stiprumo, ir todėl ligandų reitingavimas nuo geriausiai iki blogiausiai besijungiančio neturėtų pasikeisti. Pasikeistų tik absoliuti jungimosi laisvosios energijos vertė, ir tai apytiksliai tuo pačiu dydžiu, kurį galima paskaičiuoti, palyginus visas žinomas ligandų jungimosi laisvosios energijos vertes su paskaičiuotomis iš metadinamikos.

Metadinamikos potencialas buvo nustatytas veikti per vieną CV vertę: atstumą tarp cinko jono baltymo aktyviajame centre ir įkrauto azoto atomo ligando sulfonamido grupėje. Papildomo (angl. *bias*) potencialo Gausinė funkcija, kurios aukštis 2 kJ/mol ir plotis 0,04 nm, buvo pridedama kas 500 molekulinės dinamikos simuliacijos žingsnių. Buvo atliekamos 4 ns metadinamikos simuliacijos, kad užtektų laiko papildomam potencialui visiškai užpildyti ligando prisijungimo vietą baltyme. Kartu su PLUMED įskiepiu platinama programa SUM_HILLS buvo naudota iš metadinamikos papildomo potencialo apytiksliai paskaičiuoti laisvąją ligando prisijungimo energiją, kaip funkciją nuo metadinamikos CV.

2.5 Karboanhidrazės slopinimo ir selektyvumo QSAR benzensulfonamidams

3.2.1 skyrelyje aprašytas keturiasdešimties benzensulfonamidų ligandų cheminių struktūrų duomenų rinkinys *sdf* formato faile buvo panaudotas, kaip įvestis E-DRAGON molekulinį deskriptorių skaičiavimo programai [268]. E-DRAGON skaičiuoja daug deskriptorių tipų, tokių kaip topologiniai aprašai, kelio ir ėjimo skaičiavimas, jungiamumo indeksai, dvimatės autokoreliacijos, geometriniai aprašai, RDF aprašai, trimačiai MoRSE deskriptoriai, WHIM deskriptoriai, GETAWAY deskriptoriai ir t.t [213]. Galutiniame paskaičiuotų deskriptorių rinkinyje, atmetus deskriptorius, kurie buvo lygūs nuliui arba turėjo tą pačią reikšmę visiems keturiasdešimčiai ligandų, liko 1666 deskriptoriai. Tinkami galutiniams modeliams deskriptoriai buvo atrinkti ir daugybinės tiesinės regresijos QSAR modeliai buvo sukurti su R programiniu

paketu (versija 3.2.3) (<https://www.r-project.org/>) ir LEAPS programa dviem žingsniais. Pirmiausia, deskriptoriai su panašiomis ar vienodomis vertėmis visiems junginiams ar stipriai koreliuojantys tarpusavyje buvo pašalinti. Jei deskriptorių poroje $R^2 > 0,90$, vienas iš poros deskriptorių pašalintas atsitiktinio parinkimo būdu. Naudojant šį sumažintą perspektyviausių 404 deskriptorių rinkinį galutinė deskriptorių atranka QSAR modeliams atrasti atlikta su LEAPS programa, kuri yra papildoma R programos dalis. LEAPS perspėdavo apie tai, kad yra tiesinė priklausomybė tarp deskriptorių ir tokie QSAR deskriptoriai buvo atmesti kaip netinkami. Ta pati procedūra buvo pakartota kiekvienam E-DRAGON programos deskriptoriais paremtam QSAR modeliui atrasti. Galutinės QSAR lygtys buvo sudarytos, imant trisdešimties slopiklių duomenų apmokymo rinkinį, ir dešimt likusių slopiklių sudarė testavimo rinkinį, skirtą sukurtų modelių prognozavimo kokybei testuoti (šio rinkinio junginiai visų keturiasdešimties benzensulfonamidų rinkinio lentelėje 2.2.1-1 pažymėti žvaigždute).

Geriausias galimas QSAR modelis iš LEAPS programos rezultatų buvo atrinktas pagal aukščiausias R^2 (determinacijos koeficiento) vertes. Modelių statistinis reikšmingumas buvo nustatytas paskaičiuojant p vertę su F -testu. Modelių tinkamumo prognozėms įvertinimas buvo atliktas su LOO metodu, suskaičiuojant Q^2_{LOO} (kryžminio patikrinimo R^2 vertė, suskaičiuota iš prognozių gautų LOO būdu), ir dar papildomai suskaičiuojant R^2_{TEST} (determinacijos koeficiento vertę prognozuojamiems aktyvumams dešimties junginių testavimo rinkinio duomenims, kurie nebuvo naudoti, parametrizuojant QSAR modelius). Trys ar daugiau deskriptorių buvo įtraukiami į galutinį QSAR modelį, laikantis visuotinai pripažintos taisyklės, kad bent penki cheminiai junginiai su eksperimentiniais duomenimis turi būti paimti kiekvienam imtam deskriptoriui [249]. Q^2_{TEST} buvo paskaičiuotas su tokia pačia skaičiavimo procedūra, kaip ir PHASE Q^2 . MAE buvo paskaičiuotas pagal procedūrą, aprašytą literatūroje [283].

Pritaikomumo domeno nustatymas yra modernus įrankis QSAR

modeliavime. Pritaikomumo domenas nustato cheminę erdvę, iš kurios yra išvestas QSAR modelis, ir kurioje prognozės yra laikomos patikimomis [284]. Šame darbe buvo naudota paprasta, standartizavimu paremta, metodika, sukurta Roy ir bendradarbių (http://teqip.jdvu.ac.in/QSAR_Tools/) [285]. Ši metodika yra taikoma jau sukurtiems QSAR modeliams, kai reikia rasti išskirtis apmokymo duomenų rinkinyje ir už pritaikomumo domenų ribų esančius junginius testavimo duomenų rinkinyje.

2.6. PHASE atomais paremtas trimatis QSAR

PHASE trimačio QSAR skaičiavimai įvairių CA afiniškumams ir CA XII selektyvumams buvo atlikti su programa PHASE (versija 4.4) MAESTRO (versija 10.3) molekulių modeliavimo programiniame pakete, sukurtame Schrödinger, LLC [286, 287]. PHASE QSAR modeliai gali būti sudaromi atomais paremti (angl. *atom-based*), lauku paremti (angl. *field-based*) arba savybėmis paremti (angl. *property-based*). Skirtumas tarp jų yra tas, kad į QSAR modeliavimą įtraukiami arba visi nagrinėjamų cheminių junginių atomai, arba tik skirtingi farmakoforai, kurie atitinka hipotezę apie tai, kas nulemia biologinį aktyvumą. Visų nagrinėtų cheminių junginių struktūros buvo iš naujo perklotos, naudojant „*common scaffold alignment*“ funkciją MAESTRO programoje. Atomais paremto trimačio QSAR modeliai buvo paskaičiuoti, nagrinėjant koreliacijas tarp išmatuotų ir prognozuojamų pK_d arba selektyvumų verčių trisdešimties junginių apmokymo rinkiniui, atliekant PLS analizę. Atrastų modelių tinkamumo patvirtinimas buvo atliktas, naudojant LOO metodą bei dešimties junginių testavimo rinkinį. Testavimo rinkinys buvo visada išlaikytas vienodas tam, kad būtų galima vėliau palyginti įvairius taikytus QSAR atlikimo protokolus. PLS regresija buvo atlikta su maksimaliu skaičiumi ($n/5$) PLS faktorių (n = ligandų skaičius apmokymo rinkinyje) ir trimačio QSAR parametrų skaičiavimo grotelių langelio dydžiu 1.0 Å. Visų atrastų modelių įvertinimas buvo atliktas, prognozuojant pK_d arba selektyvumą visiems testavimo rinkinio junginiams. Su programa apskaičiuoti šie QSAR modelių statistiniai parametrai: SD (regresijos standartinis nuokrypis), R^2 (regresijos determinacijos koeficientas), $R^2 CV$ (kryžminio patikrinimo

determinacijos koeficientas, suskaičiuotas iš prognozių, gautų naudojant LOO; šis statistinis parametras buvo identiškas Q^2_{LOO} skaičiavimams E-DRAGON deskriptoriais paremto QSAR modeliavimo protokoluose), $R^2_{scrambled}$ (vidutinė determinacijos koeficiento vertė iš serijos modelių, sukurtų, naudojant sumaišytus biologinius aktyvumus apmokymo rinkinio junginiams; ji parodo, kiek gerai atliekamas modeliavimas aprašo tik faktiškai prasmingus duomenis; $R^2_{scrambled}$ parametro vertė turi būti kuo mažesnė), *stability* (modelių prognozių stabilumą keičiant apmokymo rinkinio sudėtį; ši statistika turi maksimalią vertę, lygią vienetui (tai reiškia visišką stabilumą)), F (statistikos F reikšmė; didelės F vertės parodo labiau statistiškai reikšmingą regresiją), P (p -reikšmė; mažesnės vertės parodo mažesnę tikimybę, kad modelio parametrai yra lygūs nuliui), $RMSE$ (vidutinė kvadratinė prognozės paklaidą testavimo duomenų rinkiniui), Q^2 (determinacijos koeficientą testavimo rinkiniui, skaičiuotą pagal apmokymo rinkinio regresijos modelį; identišką Q^2_{TEST} E-DRAGON deskriptoriais paremtame QSAR modeliavime)), *Pearson-r* (Pearsono koreliacijos koeficiento vertę tarp testavimo rinkinio išmatuotų ir prognozuojamų duomenų). Papildomai PHASE rezultatų patikrai ir atrastų QSAR modelių tinkamumo analizei buvo panaudotos rentgenostruktūrinės analizės struktūros (PDB ID: 4WR7 ir 4KP8), parodančios CA I ir CA XII aktyviuosius centrus ir juose esančius ligandus.

3. Rezultatai ir jų aptarimas

3.1. α -Pakeistų α,β -nesočiųjų ketonų antiproliferacinio aktyvumo žmogaus vėžinių ląstelių linijose tyrimas QSAR metodu

3.1.1. Duomenų rinkinys

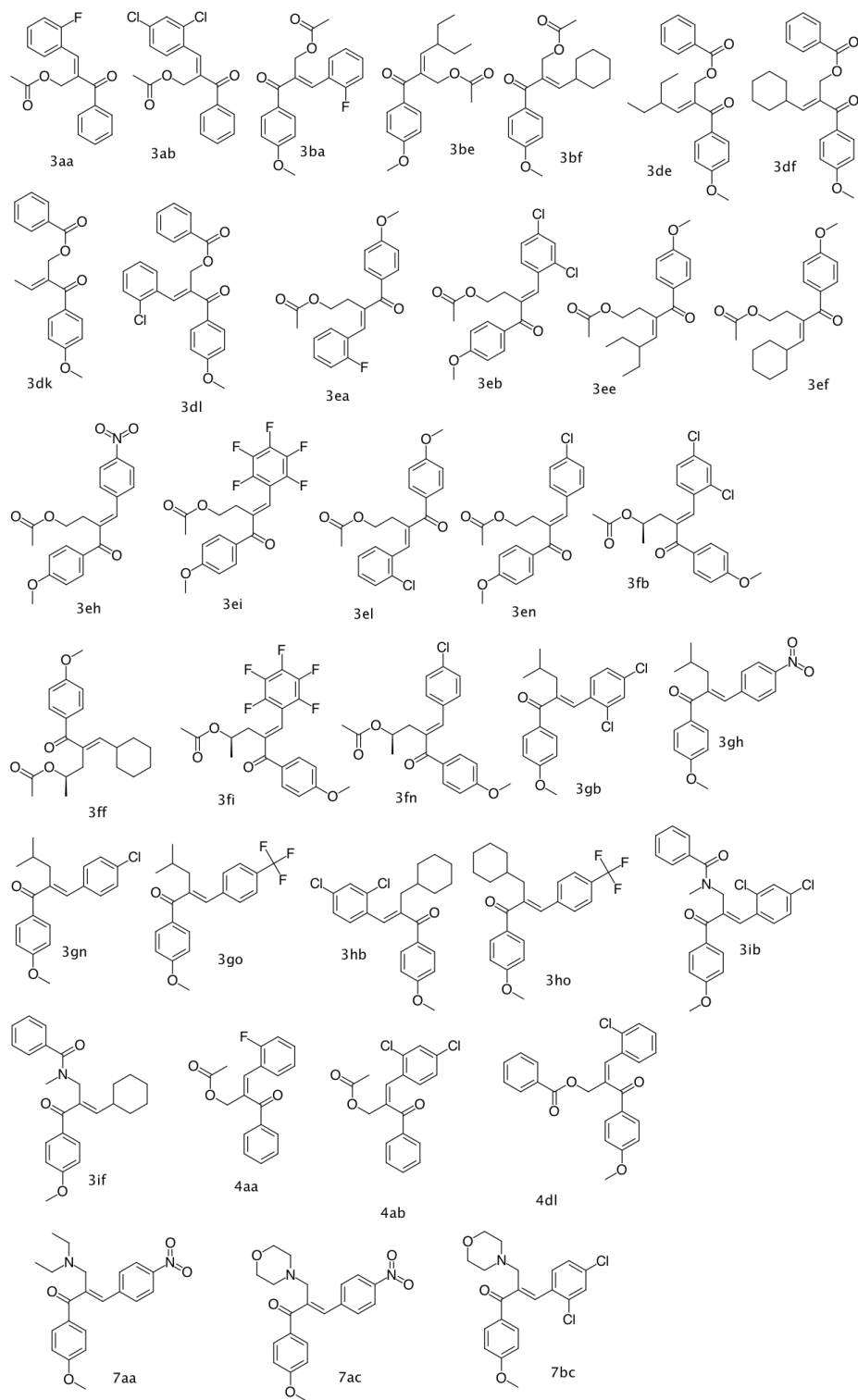
Buvo gautas 62 molekulių struktūrų, α -pakeistų α,β -nesočiųjų ketonų, duomenų rinkinys su išmatuotu antiproliferaciniu aktyvumu prieš NB4 (ūminės promielocitinės leukemijos), MCF-7 (plaučių vėžio) ir A549 (krūties vėžio) ląstelių linijas kiekvienam junginiui. Jis buvo išdalintas į du cheminių junginių rinkinius pagal tai, ar nagrinėjama molekulė turi terminalinę alkeno grupę ar ne: junginių rinkinys 1 yra be terminalinės alkeno grupės (35 molekulių struktūros) (paveikslas 3.1.1-1), junginių rinkinys 2 yra su terminaline alkeno grupe (27 molekulių struktūros) (paveikslas 3.1.1-2). Paprastumo dėlei, visos IC_{50} vertės buvo konvertuotos į pIC_{50} ($pIC_{50} = -\log(IC_{50})$). Tam, kad su visais junginiais būtų galima atlikti tolimesnius skaičiavimus, buvo padaryta prielaida, kad junginių, kurių IC_{50} pradinuose eksperimentiniuose duomenyse nurodyta kaip „ $> 100 \mu M$ “, pIC_{50} lygi 4. Visų junginių pIC_{50} vertės pateiktos lentelėje 3.1.1-1.

Lentelė 3.1.1-1. Nagrinėtų α -pakeistų α,β -nesočiųjų ketonų duomenų rinkinys su išmatuotu antiproliferaciniu aktyvumu (pateikta kaip pIC_{50} vertės) prieš NB4, MCF-7 ir A549 ląstelių linijas kiekvienam junginiui.

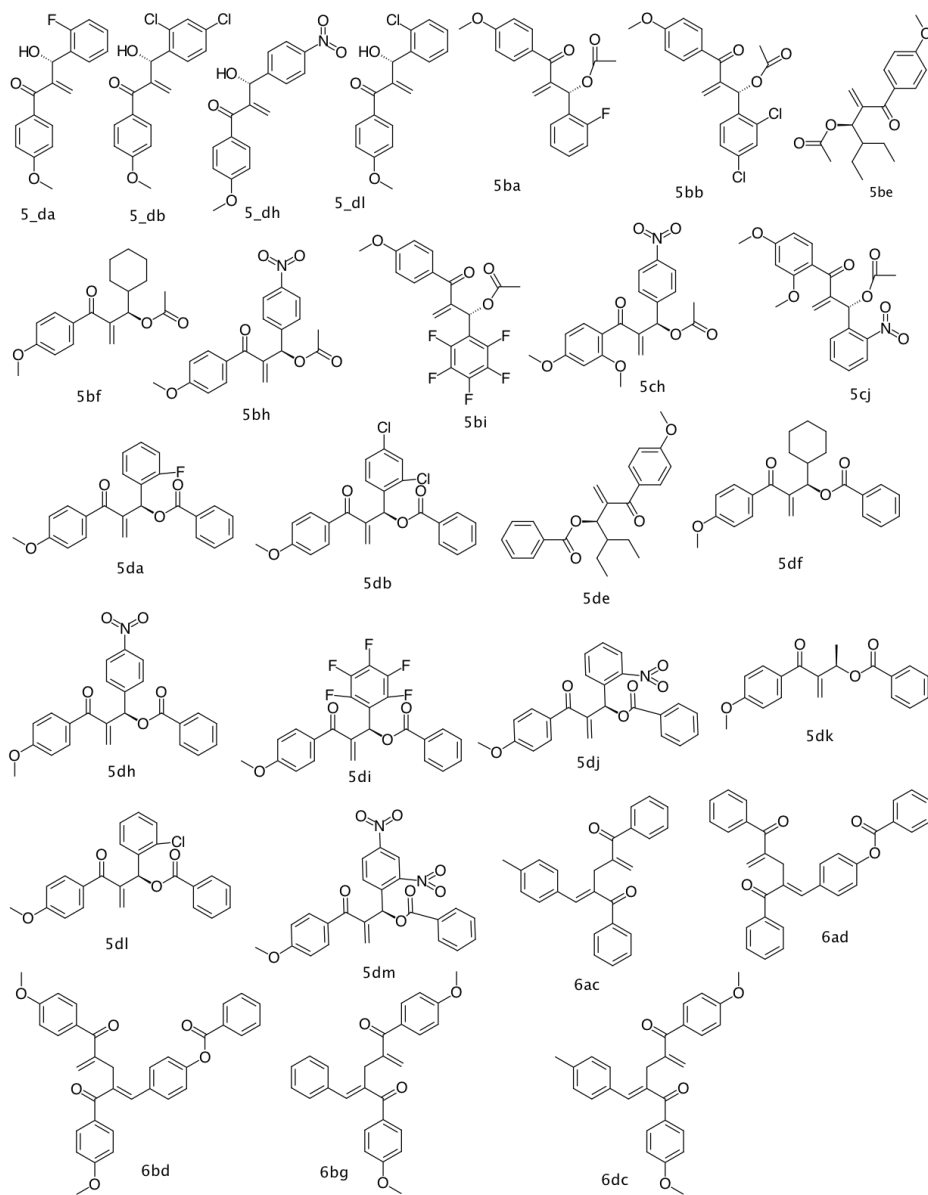
Junginys	pIC_{50}		
	NB4	A549	MCF7
3aa	5,2	4,2	4,8
3ab	5,2	4,0	5,1
3ba	5,1	4,6	4,5
3be	5,0	4,0	4,9
3bf	4,6	4,0	4,6
3de	4,6	4,0	4,6
3df	5,0	4,4	4,9
3dk	4,8	4,0	4,5
3dl	5,2	4,7	5,0
3ea	4,0	4,0	4,0

3eb	4,5	4,2	4,1
3ee	4,0	4,0	4,0
3ef	4,0	4,0	4,0
3eh	4,0	4,0	4,0
3ei	4,5	4,2	4,0
3el	4,3	4,0	4,0
3en	4,4	4,1	4,0
3fb	4,4	4,4	4,0
3ff	4,5	4,0	4,0
3fi	4,3	4,0	4,0
3fn	4,2	4,1	4,0
3gb	4,8	4,1	4,0
3gh	6,2	4,0	4,0
3gn	4,9	4,0	4,0
3go	4,9	4,0	4,0
3hb	4,3	4,0	4,0
3ho	4,7	4,0	4,0
3ib	4,6	4,5	4,0
3if	4,4	4,2	4,5
4aa	5,1	4,7	4,4
4ab	4,9	4,1	4,5
4dl	4,7	4,3	4,4
5ba	5,0	4,2	4,7
5bb	5,3	4,8	5,1
5be	4,9	4,5	4,4
5bf	5,0	4,0	4,9
5bh	5,0	4,7	5,2
5bi	4,9	4,8	5,2
5ch	5,3	4,4	5,0
5cj	5,1	4,0	4,2
5_da	5,2	4,8	4,5
5da	4,8	4,1	4,5
5_db	4,6	4,2	5,0
5db	5,0	4,1	4,6
5de	4,9	4,5	4,7
5df	5,0	4,2	4,5
5_dh	5,3	4,6	5,0
5dh	4,9	4,2	4,3
5di	4,9	4,5	4,3
5dj	4,9	4,3	4,9
5dk	4,8	4,0	4,8
5_dl	5,1	4,5	4,5
5dl	5,1	4,5	4,4
5dm	5,0	4,9	4,0

6ac	5,1	4,7	4,8
6ad	5,1	4,4	4,2
6bd	5,2	4,0	4,0
6bg	5,1	4,6	4,6
6dc	4,9	4,4	4,6
7aa	5,9	4,7	5,4
7ac	5,3	4,2	5,2
7bc	5,1	4,0	4,4



Paveikslas 3.1.1-1. Junginių rinkinys 1 (35 molekulių struktūros).



Paveikslas 3.1.1-2. Junginių rinkinys 2 (27 molekulių struktūros).

3.1.2. QSAR modeliai

Keli QSAR modeliai buvo sukurti junginių rinkiniams 1 ir 2, naudojant junginių antiproliferacinius aktyvumus prieš NB4, MCF-7 ir A549 ląsteles-

taikinius. Šeši geriausi modeliai pateikiami apačioje.

$$\text{pIC}_{50} (\text{NB4; Set 1}) = 5,111 \text{ MATS1m} - 1,83 \text{ MATS5p} - 2,47 \text{ HOMA} + 14,1 \text{ E2u} - 10,4 \text{ E2e} + 4,27,$$

(3.1.2-1)

$$R^2 = 0,86, R^2_{ADJ} = 0,83, Q^2 = 0,75, F(5,29) = 35,1, p < 10^{-5}$$

$$\text{pIC}_{50} (\text{A549; Set 1}) = 0,968 \text{ ATS6e} + 1,55 \text{ MATS8m} + 0,370 \text{ Mor05v} - 0,484 \text{ Mor11p} + 0,765 \text{ Mor20p} + 1,59,$$

(3.1.2-2)

$$R^2 = 0,81, R^2_{ADJ} = 0,78, Q^2 = 0,75, F(5,29) = 25,3, p < 10^{-5}$$

$$\text{pIC}_{50} (\text{MCF-7; Set 1}) = 0,0457 \text{ RDF050u} - 0,119 \text{ RDF035p} - 0,620 \text{ Mor20m} - 3,20 \text{ H7u} - 9,44 \text{ HATS2u} + 8,73,$$

(3.1.2-3)

$$R^2 = 0,92, R^2_{ADJ} = 0,90, Q^2 = 0,88, F(5,29) = 64,8, p < 10^{-5}$$

$$\text{pIC}_{50} (\text{NB4; Set 2}) = -5,17 \text{ MATS2m} + 5,25 \text{ MATS2e} - 3,90 \text{ E3u} + 8,17 \text{ G2s} - 1,29 \text{ R4v} + 5,69,$$

(3.1.2-4)

$$R^2 = 0,86, R^2_{ADJ} = 0,82, Q^2 = 0,77, F(5,21) = 24,8, p < 10^{-5}$$

$$\text{pIC}_{50} (\text{A549; Set 2}) = -25,0 \text{ X3Av} + 1,34 \text{ Mor16u} + 7,19 \text{ G3u} - 8,85 \cdot \text{G3p} - 0,140 \text{ Htv} + 7,97,$$

(3.1.2-5)

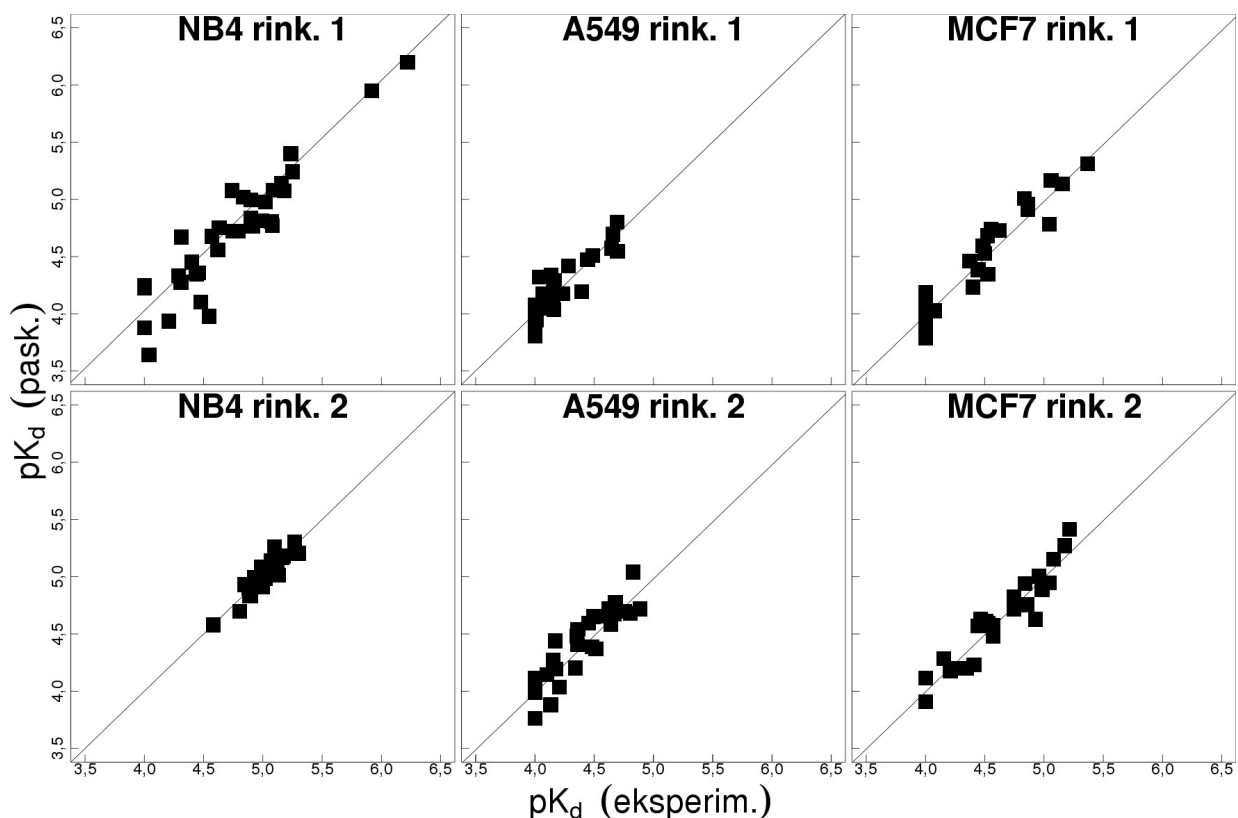
$$R^2 = 0,80, R^2_{ADJ} = 0,75, Q^2 = 0,67, F(5,21) = 16,4, p < 10^{-5}$$

$$\text{pIC}_{50} (\text{MCF-7; Set 2}) = 0,0149 \text{ CSI} - 0,109 \text{ UNIP} - 0,0475 \text{ MPC10} + 0,815 \text{ X2} - 0,909 \text{ GATS1v} + 1,11.$$

(3.1.2-6)

$$R^2 = 0,89, R^2_{ADJ} = 0,87, Q^2 = 0,83, F(5,21) = 34,7, p < 10^{-5}$$

Modeliai kokybiškai aprašyti keliais statistiniais parametrais, kurie pateikiami žemiau kiekvienos QSAR lygties. Modeliai pripažinti priimtinais, jei tiek determinacijos koeficientas R^2 , tiek suderintas determinacijos koeficientas R^2_{ADJ} yra aukščiausi tam duomenų rinkiniui, ir F–testo vertės didesnės nei kritinės. Visų sukurtų modelių tinkamumai prognozėms buvo įvertinti, naudojant kryžminį patvirtinimą išmetant vieną (LOOCV) su įverčiu Q^2 . Q^2 įverčio skalė yra tokia pati kaip ir R^2 : jei Q^2 lygus vienetui, tada modelis prognozėms tinkamas idealiai, o jei Q^2 mažesnis nei 0,40, tada modelis laikomas netinkamu prognozėms. Molekuliniai deskriptoriai, esantys QSAR lygtyse 3.1.2-1 – 3.1.2-6, plačiau aprašyti priede, deskriptorių lentelėje A. Statistiniai parametrai esantys QSAR lygtyse, plačiau aprašyti kitur darbo tekste, o nuorodos ir pavadinimai yra priede, lentelėje D. Grafike 3.1.1-1 pateiktos koreliacijos tarp eksperimentiškai nustatytų ir QSAR lygtimis apskaičiuotų pIC_{50} verčių. Tiesi linija žymi idealų atitikimą tarp jų.



Grafikas 3.1.1–1. Koreliacijos tarp eksperimentiškai nustatytų ir QSAR lygtimis apskaičiuotų pIC_{50} verčių. Tiesi linija žymi visišką atitikimą. Visų x ir y ašių skalių masteliai sutapatinti palyginimui.

Įdomu, kad R^2 panašūs tai pačiai ląstelių linijai, nesvarbu, kuris junginių rinkinys naudojamas, nors QSAR lygtys skirtingos. Ląstelių linijai A549 išmatuotų pIC_{50} duomenų intervalas junginių rinkiniams 1 ir 2 yra atitinkamai tik 0,7 ir 0,9. Tai iš dalies paaiškina kiek blogesnę modelių kokybę ląstelių linijai A549, lyginant su kitomis. Įdomu panagrinėti, kaip gerai modeliai klasifikuoja junginius į aktyvius ir neaktyvius. Aibeį taškų grafike 3.1.1-1, kurie žymi neaktyvius junginius ir kurių IC_{50} pradiniuose eksperimentiniuose duomenyse nurodyta kaip „ $> 100 \mu M$ “ (pIC_{50} modeliuojant yra prilyginama 4), QSAR modeliai prognozuoja pIC_{50} vertes intervale nuo 3,76 iki 4,25, t.y. eksperimentiškai nustatyti neaktyvūs junginiai yra gana tiksliai prognozuojami kaip neaktyvūs sukurtaisiais QSAR modeliais. Nagrinėtuose junginių rinkiniuose 1 ir 2 yra tik du junginiai 3ei ir 3fn, kurie klaidingai priskirti neaktyviems, nors jiems eksperimentiškai aptiktas aktyvumas. Abiems jiems

QSAR prognozuojamos pIC_{50} vertės yra mažesnės, negu 4,0, nors eksperimentiškai išmatuotos pIC_{50} vertės yra didesnės, negu 4,2, t.y. prognozės paklaida buvo nedidelė. Tie junginiai priklauso junginių rinkiniui 1, ir minėtos pIC_{50} vertės yra NB4 ląstelių linijai. Visa tai parodo, kad QSAR modeliai gali atskirti aktyvius junginius nuo neaktyvių, ir kad jie gali būti panaudoti, kuriant naujas molekules su pagerintomis antiproliferacinėmis savybėmis. Visi šie šeši modeliai buvo įkelti į QSAR modelių duomenų bazę QSARDB (<http://dx.doi.org/10.15152/QDB.178>).

3.2. CA slopinimo benzensulfonamidais prognozavimas įvairiais metodais

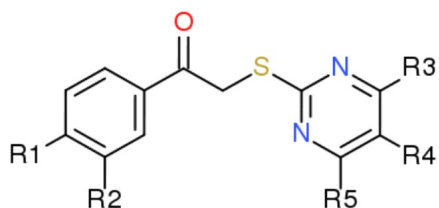
3.2.1. Benzensulfonamidų rinkinys skaičiavimams

Skaičiavimams buvo naudotas keturiasdešimties benzensulfonamidų duomenų rinkinys su anksčiau publikuotomis eksperimentiškai išmatuotomis jungimosi su karboanhidrazėmis konstantomis [288]. Junginių žymėjimas raidėmis ir skaičiais (1-4) (a-j) buvo išlaikytas toks pat, kaip originaliame straipsnyje (paveikslas 3.2.1-1). Jų afiniškumo duomenys pateikti lentelėje 3.2.1-1. Aštuonių ligandų erdvinės struktūros (1a, 1f, 1i, 1j, 2f, 2j, 4f ir 4g) buvo paimtos iš PDB duomenų bazės struktūrų, kuriose eksperimentiškai nustatyta, kaip šie junginiai prisijungę prie CA II (PDB ID: 3S8X, 3S9T, 3SAP, 3SAX, 3SBH, 3SBI, 4KNI ir 4KNJ), pirmiau atlikus visų minėtų baltymų struktūrų perklojimą su programa PYMOL (versija 1.7.0.0; <https://www.pymol.org/>). Trūkstami vandenilio atomai buvo pridėti su programa AVOGADRO (versija 1.1.1) [266]. Likusių junginių struktūros buvo surinktos ir optimizuotos trimatėje erdvėje su ta pačia programine įranga. Erdvinės struktūros optimizacijai buvo taikyti molekulinės mechanikos metodai, siekiant išlaikyti tą patį perklojimą su pirmiau paruoštomis ligandų erdvinėmis struktūromis. Atskiri keturiasdešimties junginių rinkinio molekulių erdvinių struktūrų failai buvo sujungti į vieną molekulių struktūrų *sdf* formato failą su programa OPEN BABEL (versija 2.3.1) [267].

Visas duomenų rinkinys QSAR modeliavimams buvo padalintas į

apmokymo rinkinį su trisdešimčia slopiklių, ir testavimo rinkinį su likusiais dešimt slopiklių (1b, 1c, 1h, 1i, 1j, 2f, 2h, 2i, 3i, 4e; lentelėje QSAR testavimo rinkinio junginiai pažymėti žvaigždute). Kad būtų galima objektyviai palyginti QSAR modeliavimo rezultatus, tie patys apmokymo ir testavimo rinkiniai buvo naudoti visuose QSAR protokoluose.

K_d vertės buvo perskaičiuotos į pK_d vertes pagal formulę $pK_d = -\log(K_d)$ visiems tolimesniems skaičiavimams. Selektivumas ($pK_d, diff$) CA XII buvo apibrėžtas kaip skirtumas tarp CA XII pK_d ir kitos CA pK_d verčių. Suminis selektivumas CA XII taikiniui ($\Sigma CA XII$) buvo apibrėžtas kaip visų selektivumų suma tam junginiui. Jei junginys labai selektyvus CA XII, lyginant su tik viena CA izoforma, ir visai nėra selektyvus likusių izoformų atžvilgiu, pagal taip apibrėžtą suminį selektivumą jis laikomas kaip selektyvus CA XII.



- | | |
|---------------------------------------|---|
| a: R3=OH, R4=H, R5=Me | 1: R1=SO ₂ NH ₂ , R2=H |
| b: R3=OH, R4=Bn, R5=Me | 2: R1=SO ₂ NH ₂ , R2=Cl |
| c: R3=OH, R4=H, R5=Pr | 3: R1=H, R2=SO ₂ NH ₂ |
| d: R3=OH, R4=H, R5=tert-Bu | 4: R1=Cl, R2=SO ₂ NH ₂ |
| e: R3=OH, R4=CO ₂ Et, R5=H | |
| f: R3=Me, R4=H, R5=Me | |
| g: R3=H, R4=Et, R5=H | |
| h: R3=H, R4=Pr, R5=H | |
| i: R3=H, R4=Bu, R5=H | |
| j: R3=H, R4=H, R5=H | |

Paveikslas 3.2.1-1. Benzensulfonamidų rinkinio, S-alkilintų pirimidino darinių, struktūros su ankščiau eksperimentiniais publikuotais afiniškumo duomenimis [288]. Junginiai buvo sugrupuoti pagal keturis galimus

benzensulfonamido žiedo pakaitų ir jų išsidėstymo variantus (1–4) ir dešimt pirimidino žiedo pakaitų ir jų išsidėstymo variantų (a–j). Iš viso čia pavaizduota keturiasdešimt junginių.

Lentelė 3.2.1-1. Eksperimentinės skaičiavimuose naudotų benzensulfonamidų jungimosi su įvairiomis CA izoformomis disociacijos konstantos [288].

Junginys	Disociacijos konstanta K_d (nM)					
	CA I	CA II	CA VI	CA VII	CA XII	CA XIII
1a	66,7	66,7	3700	167	556	357
2a	4,00	21,3	2500	35,7	333	33,3
3a	833	1250	7100	4000	3700	1570
4a	833	200	900	133	500	200
1b*	33,3	25	1300	143	714	13,0
2b	20,0	66,7	1900	100	1110	25,0
3b	2220	1110	3300	12500	8330	1430
4b	455	33,3	290	143	294	42,0
1c*	17,2	83,3	1250	250	90,9	62,5
2c	10,0	111	4500	50,0	250	83,3
3c	500	500	13000	3330	2000	833
4c	667	66,7	830	55,6	100	55,6
1d	1,9	29	5000	180	143	3,50
2d	0,50	40	9100	11,1	333	5,00
3d	100	500	40000	500	1250	167
4d	100	50	2000	20,0	43,5	25,0
1e	17,0	18	2500	200	370	286
2e	14,3	20	2800	25,0	250	76,9
3e	1250	667	2100	1000	1250	2220
4e*	1250	50	200	25,0	200	125
1f	22,2	33,3	2500	125	909	83,3
2f*	5,00	16,7	2900	12,5	909	10,0
3f	370	1000	13000	6670	3330	1000
4f	1000	222	1000	333	1430	333
1g	30,3	12,5	5000	55,6	1000	28,6
2g	6,67	7,14	5000	10,0	286	2,08
3g	500	667	8300	2500	3850	556
4g	1820	143	670	333	1250	111
1h*	33,3	12,5	5000	66,7	1430	22,7
2h*	5,26	7,14	3300	16,7	345	1,25
3h	370	556	5000	3330	4170	333
4h	1000	100	1100	500	1110	67,0
1i*	41,7	14,3	5000	100	1430	28,6
2i*	6,25	7,69	3330	11,1	357	1,82
3i*	500	313	4300	2220	1820	278
4i	1430	66,7	500	200	1110	66,7
1j*	66,7	16,7	910	66,7	556	125

2j	10,0	11,1	2500	5,00	833	16,7
3j	769	1110	1900	4000	3330	1330
4j	4550	333	560	714	1670	333

*Žvaigždute pažymėti junginiai buvo naudoti kaip QSAR modelių testavimo rinkinys.

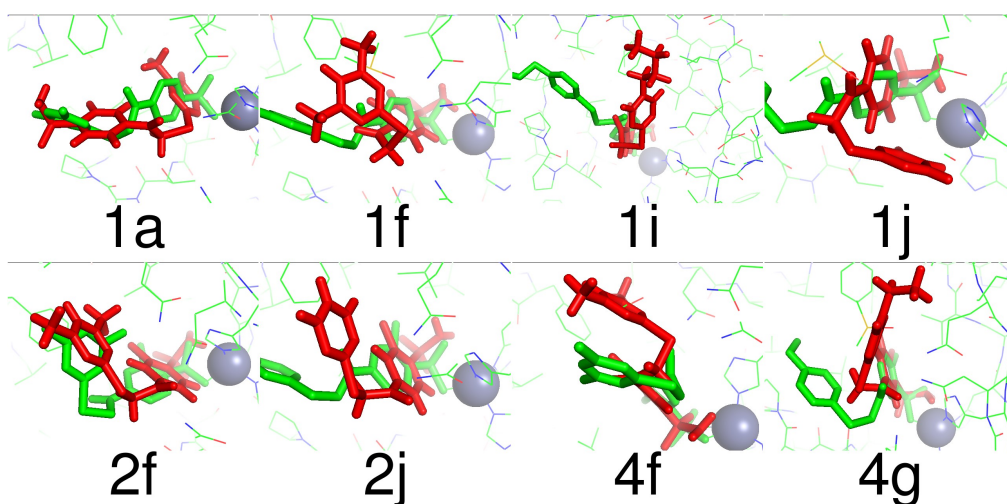
3.2.2. Dokinimas į CA II receptorių

Buvo atlikta tiesinės regresijos analizė tarp eksperimentiškai nustatytų pK_d verčių ir dokinimo skaičiavimais prognozuotų jungimosi energijos verčių. Lyginant eksperimentines pK_d vertes su ARGUSLAB ir GLIDE programomis paskaičiuotomis energijų vertėmis, buvo gautos determinacijos koeficiento (R^2) vertės, lygios atitinkamai 0,00 ir 0,07. Taigi, iš esmės nėra koreliacijos tarp su šiomis dviem programomis paskaičiuotų energijos verčių ir eksperimentinių pK_d verčių.

Buvo paskaičiuota vidutinio kvadratinio nuokrypio kvadratinė šaknis (angl. *root-mean-square deviation*, RMSD) tarp kristalografiškai nustatytų ligandų prisijungimo konformacijų ir tarp tų pačių ligandų, dokintų su GLIDE ir ARGUSLAB (lentelė 3.2.2-1). ARGUSLAB nepateikia dokintų konformacijų RMSD junginių duomenų bazės dokinimo režime, todėl dokinimas buvo atliktas, pakartotinai dokinant ligandus po vieną į atitinkamas kristalografines struktūras. Su GLIDE programa nebuvo atkartota nei viena iš aštuonių žinomų ligandų 1a, 1f, 1i, 1j, 2f, 2j, 4f ir 4g kristalografinių prisijungimo prie CA II konformacijų (paveikslas 3.2.2-1). Ligandas 1a dokinimo rezultatuose su programa GLIDE išsidėstė visiškai apsuktoje konformacijoje (RMSD = 7,8 Å). Šešioms kitiems ligandams pirimidino žiedas dokinimo rezultatuose su programa GLIDE išsidėstė visiškai klaidingai. Tik ligando 2f dokinta konformacija su GLIDE programa gautuose rezultatuose atrodė kiek panašesnė į jo kristalografinę prisijungimo prie CA II konformaciją (RMSD = 2,2 Å). Panašūs rezultatai buvo gauti ir dokinant su ARGUSLAB (lentelė 3.2.2-1).

Lentelė 3.2.2-1. RMSD tarp ligandų konformacijų, prognozuotų su GLIDE ir ARGUSLAB programomis, ir kristalografinių prisijungimo prie CA II receptoriaus konformacijų.

Ligandas	RMSD, Å	
	GLIDE	ARGUSLAB
1a	7,8	4,6
1f	3,5	8,8
1i	5,9	10,6
1j	5,1	6,2
2f	2,2	2,8
2j	3,6	4,9
4f	2,8	3,8
4g	3,9	3,2



Paveikslas 3.2.2-1. Kristalografinės ligandų prisijungimo konformacijos prie CA II ligandams 1a, 1f, 1i, 1j, 2f, 2j, 4f ir 4g, pavaizduotos žaliai, ir tie patys ligandai, dokinti su GLIDE, pavaizduoti raudonai. Cinko jonas pavaizduotas kaip sfera.

Tiek su GLIDE, tiek su ARGUSLAB dokinimo programomis prastai pavyko prognozuoti eksperimentines minėtų benzensulfonamidų prisijungimo konformacijas ir padėtis prie CA II. Siekiant ištirti galimą tautomerų susidarymo įtaką, su ARGUSLAB programa buvo sudokinti du alternatyvūs ligando 1a, kurio kristalografinė konformacija yra žinoma, tautomerai, su dar prastesniais RMSD, negu lentelėje 3.2.2-1 (8,3 ir 8,5 Å). Panašių rezultatų reikėtų tikėtis, dokinant ir su GLIDE programa.

Jau ankščiau buvo aprašyti sunkumai, mėginant prognozuoti ligandų

prisijungimo konformacijas prie CA II su VDOCK programa, ir tuomet suvaržytas dokinimas buvo naudojamas problemoms spręsti [60].

Vienas pasiūlymas, kaip pagerinti dokinimo procedūrą šitokiais atvejais, yra toks: leisti judėti tik ligandų atomams, kurie yra skirtingi ligandų serijoje. Visus kitus atomus tada reiktų fiksuoti jau žinomoje kristalografinėje prisijungimo konformacijoje. Tai galėtų gerokai supaprastinti užduotį. Su dabartinėmis ARGUSLAB ir GLIDE versijomis nėra galimybės to išmėginti.

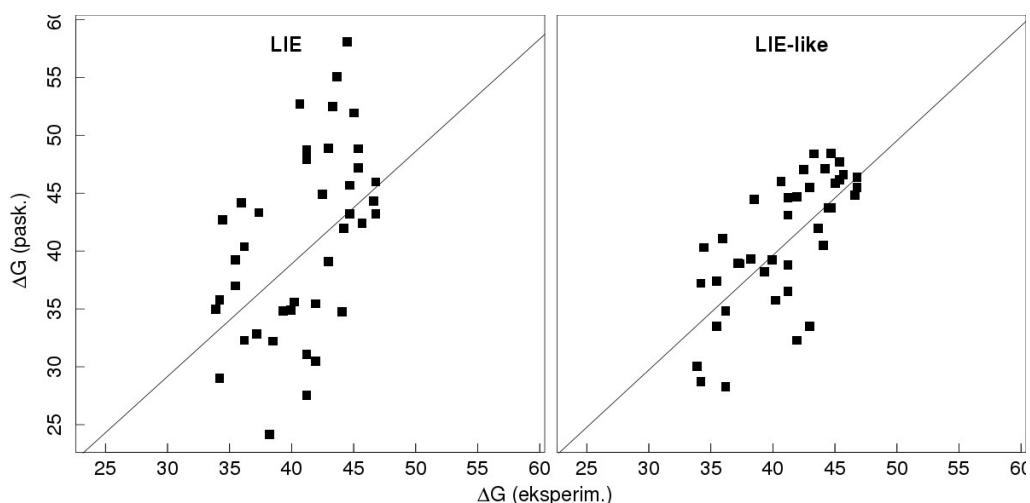
Net pusė iš nagrinėtų kristalografinių struktūrų turi dimetilsulfoksido molekulę šalia ligando (PDB ID: 3SBH, 3SBI, 3S9T, 3SAX), ir tai gali kelti trukdžius. Tokios molekulės-artefaktai paprastai ištrinamos, ruošiant taikinį dokinimui, tad ligandas dokinimo konformacijoje gali užimti tą vietą, kur anksčiau buvo artefaktas. Taip gali kilti neatsikartojamumas tarp dokinimo ir kristalografijos.

3.2.3. Tiesinės sąveikos energijos skaičiavimai

Buvo atlikti LIE skaičiavimai kiekvienam iš keturiasdešimties ligandų, prisijungusių prie CA II receptoriaus. Naudojant LIE lygtį (1.3-1) su įprastais LIE parametrais ($\alpha = 0,5$, $\beta = 0,162$, $\gamma = 0$), buvo rasta labai prasta koreliacija ($R^2 = 0,21$) tarp LIE metodu paskaičiuotų jungimosi energijų ir eksperimentinių ligandų jungimosi verčių. Koreliacija galėjo būti bloga dėl to, kad tarp besijungiančio ligando ir cinko jono, esančio CA II aktyviajame centre, susidaro koordinacinis ryšys, kuriam greičiausiai prastai tinka įprastiniai LIE metodo parametrai. Todėl buvo išvesti nauji parametrai, pritaikyti jungimuisi prie konkretaus CA II receptoriaus prognozuoti. Šios LIE parametrų vertės yra pateiktos lentelėje 3.2.3-1. Paveiksle 3.2.3-1 kairėje yra grafiškai pavaizduota koreliacija taip eksperimentiškai nustatytų ΔG_{bind} ir paskaičiuotų LIE metodu.

Lentelė 3.2.3-1. LIE metodui išvesti parametrai, taikant lygtį 1.3-1.

Parametras	Vertė
α	0,197
β	0,462
γ	10,4



Paveikslas 3.2.3-1. Grafiškai pavaizduota koreliacija taip ΔG_{bind} , nustatytų eksperimentiškai ir paskaičiuotų LIE arba nauju „LIE-like“ metodu.

Parametras laisvajai energijai prognozuoti (γ) turėtų būti apytiksliai lygus koordinacinio ryšio indėliui į kiekvieno ligando jungimąsi. Tačiau netgi LIE lygtis 1.3-1 su naujais parametrais, pateiktais lentelėje 3.2.3-1, nepadėjo žymiai padidinti R^2 : gauta determinacijos koeficiento vertė, lygi 0,24. Ieškant priežasčių, kodėl LIE metodologijų tinkamumas šiai užduočiai toks žemas, pastebėjome, kad intramolekulinės ligando pačio su savimi sąveikos vaidina svarbų vaidmenį nagrinėjamoje CA II ir ligando sistemoje, todėl negali būti ignoruojamos.

Remiantis šiais pastebėjimais, mes pasiūlėme naują būdą, kaip pagerinti ΔG_{bind} prognozę. Naujoji schema yra labai panaši į LIE, todėl ji buvo pavadinta „LIE-like“. Dvi Kulono ir Lenardo-Džonso sąveikos čia naudojamos panašiu principu, kaip ir LIE metodologijoje, remiantis pastebėjimu, kad jos nekoreliuoja tarpusavyje, bet kiekviena atskirai šiek tiek koreliuoja su ΔG_{bind} :

$$\Delta G_{bind} = \alpha \langle V^{el} \rangle_{lig-prot} + \beta \langle V^{vdw} \rangle_{lig-lig} + \gamma. \quad (3.2.3-1)$$

Šioje lygtyje kintamieji $\langle V^{el} \rangle$ ir $\langle V^{vdw} \rangle$ yra elektrostatinių ir van der Valso sąveikų vidurkiai per visą molekulinės dinamikos simuliaciją. Žemutiniai indeksai „*lig-lig*“ nurodo, kad imama ligando sąveika pačio su savimi baltymo aplinkoje, o indeksas „*lig-prot*“ žymi ligando sąveiką su taikiniu-baltymu. Naujosios lygties (3.2.3-1) parametrizacija yra žymiai sėkmingesnė, nei lygties 1.3-1 – gauta determinacijos koeficiento vertė, lygi 0,50. Išvestų parametru vertės nurodytos lentelėje 3.2.3-2. Paveiksle 3.2.3-1 dešinėje grafiškai pavaizduota koreliacija taip eksperimentiškai nustatytų ir paskaičiuotų „LIE-like“ metodu ΔG_{bind} .

Lentelė 3.2.3-2. „LIE-like“ metodui išvesti parametrai, taikant lygtį 3.2.3-1.

Parametras	Vertė
α	0,0884
β	0,689
γ	51,1

Norint pagerinti jungimosi energijų prognozavimą įprastiniu LIE metodu, buvo bandyta fiksuoti ligando konformaciją į tiksliai tokią, kokia ji matoma kristalografinėse struktūrose aštuoniems ligandams (PDB ID: 3S8X, 3S9T, 3SAP, 3SAX, 3SBH, 3SBI, 4KNI ir 4KNJ). LIE metodo rezultatai ligandams su fiksuotomis konformacijomis buvo ne geresni, lyginant su to paties metodo rezultatais, kai ligandas galėjo laisvai judėti baltymo aktyviajame centre ar visai iš jo išeiti. Įdomu tai, kad šiam aštuonių ligandų rinkiniui įprastinio LIE metodo (be fiksuotų ligandų konformacijų) rezultatai koreliuoja su eksperimentiniais duomenimis apie tų ligandų jungimąsi su CA II: gauta determinacijos koeficiento vertė, lygi 0,54. Tai parodo, kad teisinga pradinė ligando konformacija baltymo aktyviajame centre naudinga LIE skaičiavimams, nes pagerina rezultatus. Gali būti, kad per visą MD modeliavimo laiką ligandas neapsilanko teisingose konformacijose, arba, jei ligandas ir pereina į teisingą konformaciją, tai MD programa dėl įvairių priežasčių (pvz., neteisingų jėgų lauko parametru) tas konformacijas įvertina kaip turinčias pernelyg aukštą energiją. Be to, ligandų konformacijų fiksavimas

galimai sutrikdo teisingą natūraliai egzistuojančių konformacijų atranką (angl. *sampling*) iš įvairių konformacijų ansamblio, esančio dinaminėje pusiausvyroje.

3.2.4. Metadinamikos skaičiavimai

Buvo atliekami metadinamikos, kaip vieno iš alternatyvių jungimosi energijų skaičiavimo būdų, skaičiavimai visiems keturiasdešimčiai ligandų, prisijungusių prie CA II receptoriaus. 4 ns metadinaminių molekulinės dinamikos simuliacijų metu buvo stebima, kaip ligandai išsina iš CA II aktyviojo centro į aplinkinę vandens terpę, pridėjus papildomą metadinamikos potencialą. Tokiu būdu buvo gauti laisvosios energijos įvertinimai, kaip funkcijos nuo atstumo tarp cinko jono CA II aktyviajame centre iki sulfonamido grupės azoto ligande. Paskaičiuotas laisvosios energijos skirtumas tarp laisvo ir prisijungusio ligando būsenų visiškai nekoreliavo su eksperimentiniais duomenimis ($R^2=0,00$). Be to, laisvosios energijos minimumų padėtis nesutapo su kristalografinėmis aštuonių ligandų prisijungimo konformacijomis, o septyniolikai ligandų laisvosios energijos minimumas buvo rastas net ne baltymo aktyviajame centre, o vandenyje, nors iš eksperimentinių duomenų akivaizdu, kad jie jungiasi prie baltymo. Ligando konformacija gerokai pasikeičia lyginant su eksperimentine per paruošiamąsias MD simuliacijas (minimizacijas) dar prieš metadinamikos skaičiavimų pradžią. Tai gali sukelti problemų tiek metadinamikos, tiek LIE skaičiavimams. MD jėgos laukas, naudojamas simuliacijose, turėtų būti pagerintas ar modifikuotas, norint gauti teigiamus rezultatus. Teoriškai ligandas turėtų didelę dalį MD simuliacijos laiko išbūti konformacijose, artimose kristalografinėms, ir ligandas turėtų išeiti iš savo prisijungimo konformacijos tik per metadinamikos simuliacijas.

3.2.5. CA II QSAR modeliai

QSAR buvo vienas iš išbandytų skaičiuojamųjų jungimosi konstantų prognozės metodų. Buvo sukurti keli QSAR modeliai su trisdešimties benzonsulfonamidų junginių jungimosi prie CA II apmokymo konstantų

rinkiniu. Trys geriausi QSAR modeliai pateikti žemiau.

$$pK_d(\text{CA II}) = 2,05 \text{ MATS7m} + 1,02 \text{ H5u} + 4,49 \text{ HATS8m} + 4,86, \quad (3.2.5.1)$$

$$R^2 = 0,89, R^2_{ADJ} = 0,87, Q^2 = 0,87, F(3,26) = 68, p < 10^{-9}, R^2_{TEST} = 0,57$$

$$pK_d(\text{CA II}) = 0,0867 \text{ G(N..S)} + 1,23 \text{ H5u} + 68,7 \text{ (R7p+)} + 2,19, \quad (3.2.5.2)$$

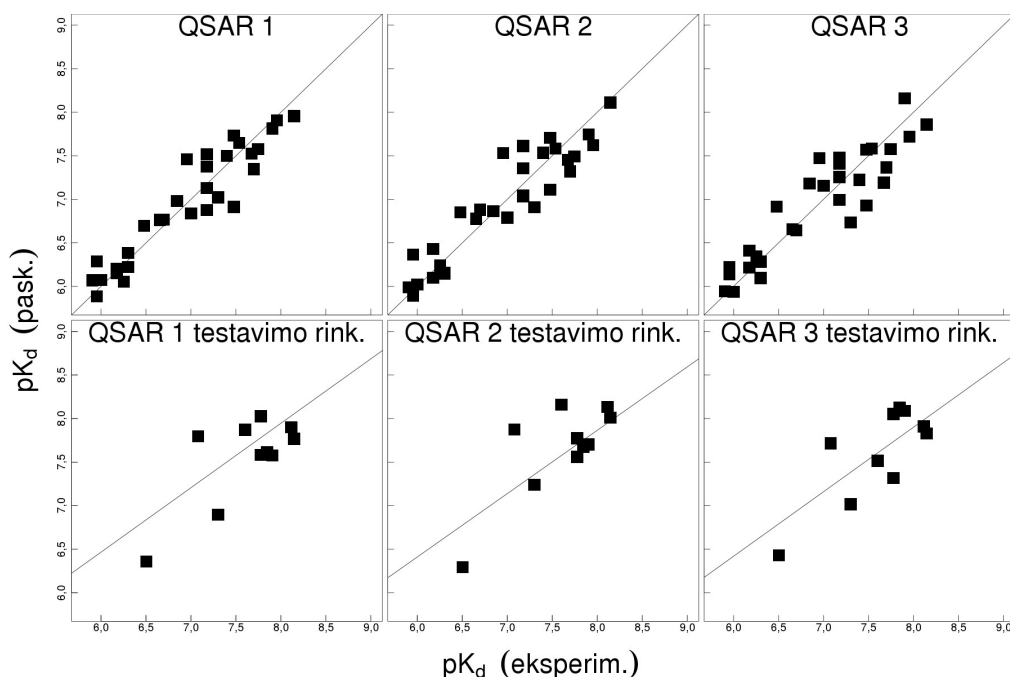
$$R^2 = 0,86, R^2_{ADJ} = 0,84, Q^2 = 0,83, F(3,26) = 53, p < 10^{-9}, R^2_{TEST} = 0,61$$

$$pK_d(\text{CA II}) = 0,0106 \text{ VAR} + 2,26 \text{ MATS7m} - 2,65 \text{ MATS8p} + 4,30. \quad (3.2.5.3)$$

$$R^2 = 0,83, R^2_{ADJ} = 0,81, Q^2 = 0,79, F(3,26) = 43, p < 10^{-9}, R^2_{TEST} = 0,63$$

Modelių statistiniai parametrai yra pateikti po kiekviena iš lygčių. Modeliai buvo laikomi priimtinais, jeigu determinacijos koeficientas R^2 ir suderintas determinacijos koeficientas R^2_{ADJ} buvo aukščiausi tam duomenų rinkiniui, ir jei F testo vertė didesnė, negu kritinė (~ 5 , pagal tam skirtą lentelę). Modelių prognozavimo pajėgumas buvo išbandytas trimis metodais: LOO metodu, paskaičiuojant Q^2 , bei su dešimties ligandų, kurie nebuvo naudoti modelių kūrime, testavimo rinkiniu.

Šiam testavimo rinkiniui buvo paskaičiuota determinacijos koeficiento R^2_{TEST} vertė, lyginant eksperimentinius pK_d duomenis su iš QSAR modelių gautomis vertėmis. Įvairių tipų molekuliniai deskriptoriai, įtraukti į lygtis 3.2.5.1, 3.2.5.2 ir 3.2.5.3, yra plačiau paaiškinti lentelėje C priede. Koreliacija taip eksperimentiškai nustatytų ir prognozuojamų visais trimis QSAR modeliais pK_d verčių, naudojant benzensulfonamidų apmokymo ir testavimo junginių rinkinius, grafiškai pavaizduota paveiksle 3.2.5-1. Visi šie trys modeliai buvo įkelti į QSAR modelių duomenų bazę QSARDB (<http://dx.doi.org/10.15152/QDB.176>).



Paveikslas 3.2.5-1. Grafiškai pavaizduota koreliacija taip eksperimentiškai nustatytų pK_d ir prognozuojamų su QSAR.

3.2.6. Dokinimo, LIE, metadinamikos ir QSAR metodų tinkamumo benzensulfonamidų slopinimo savybėms prognozuoti palyginimas

Penki metodai (ARGUSLAB dokinimas, GLIDE dokinimas, LIE, metadinamika ir QSAR) buvo taikyti, siekiant prognozuoti keturiasdešimties benzensulfonamidų junginių rinkinio su eksperimentiškai nustatytomis pK_d vertėmis prisijungimą prie CA II. Lentelėje 3.2.6-1 parodyti palyginimo su eksperimentinėmis pK_d vertėmis determinacijos koeficientai. Iš visų išbandytų metodų, tik naudojant QSAR buvo gauta gera koreliacija tarp teorinės prognozės ir eksperimentinių jungimosi konstantų.

Lentelė 3.2.6-1. Skaičiuojamųjų metodų tinkamumas ligandų jungimuisi prie CA II prognozuoti palygintas pagal R^2 tarp paskaičiuotų ir eksperimentiškai nustatytų pK_d .

Metodas	R^2
LIE	0,24
Naujas „LIE-like“	0,50
QSAR	0,83-0,89

ARGUSLAB dokinimas	0,00
GLIDE dokinimas	0,07
Metadinamika	0,00

Prasta tiek dokinimo, tiek molekulinė dinamika pagrįstų metodų koreliacija su eksperimentu greičiausiai liudija apie trūkumus, teisingai įvertinant sąveikas tarp atomų, taigi galimai neteisingus jėgų laukų parametrus. Iš nagrinėtųjų metodų tuo nepasižymi tik QSAR metodas, kuris paprastai nagrinėja dvimates ligandų struktūras, neatsižvelgdamas į sąveikas su receptoriumi.

Buvo pasiūlyta nauja modifikuota LIE metodo lygtis 3.2.3-1, modifikuojanti klasikinę LIE metodo išraišką (lygtis 1.3-1). Šis LIE protokolo patobulinimas buvo pavadintas „LIE-like“. Jis leido žymiai pagerinti jungimosi konstantų prognozes. Potencialus pasiūlymas dokinimui pagerinti su fiksuotomis ligando konformacijos dalimis pagal rentgenostruktūrinės analizės duomenis liko neišbandytas, nes naudotose dokinimo programų versijose tam nėra galimybių.

QSAR metodas efektyviausiai padėjo kiekybiškai įvertinti subtilius empirinius sąryšius tarp ligandų struktūros ir jų aktyvumo CA II atžvilgiu. Pateiktieji QSAR modeliai benzensulfonamidų jungimosi su CA II konstantoms prognozuoti gali būti panaudoti, kuriant naujus CA II slopiklius.

3.2.7. QSAR įvairioms CA izoformoms

Kaip buvo aprašyta skyrelyje 3.2.1, suskaičiavus benzensulfonamidų selektyvumus iš lentelės 3.2.1-1 pagal formulę pateiktą skyrelyje 3.2.1, buvo sukurti QSAR modeliai, skirti benzensulfonamidų selektyvumams CA XII atžvilgiu prognozuoti, kurie yra pateikiami žemiau:

$$pK_{d,diff}(CA\ XII - CA\ I) = -0,566\ T(N..N) + 0,0733\ EPS0 + 21,0\ (R7m+) + 7,14, \quad (3.2.7-1)$$

$$R^2 = 0,84, R^2_{ADJ} = 0,83, Q^2_{LOO} = 0,82, F(3,26) = 47, p < 10^{-9}, R^2_{TEST} = 0,79,$$

$$Q^2_{TEST} = 0,67, MAE = 0,33$$

$$pK_{d,diff}(CA XII - CA II) = 8,76 MATS1p - 1,41 GATS8m - 1,03 GATS3v + 4,57, \quad (3.2.7-2)$$

$$R^2 = 0,81, R^2_{ADJ} = 0,79, Q^2_{LOO} = 0,77, F(3,26) = 38, p < 10^{-8}, R^2_{TEST} = 0,58, Q^2_{TEST} = 0,53, MAE = 0,28$$

$$pK_{d,diff}(CA XII - CA VI) = -12,1 GNar + 5,01 MATS7v - 2,74 MATS7p + 23,5, \quad (3.2.7-3)$$

$$R^2 = 0,82, R^2_{ADJ} = 0,80, Q^2_{LOO} = 0,78, F(3,26) = 39, p < 10^{-9}, R^2_{TEST} = 0,42, Q^2_{TEST} = 0,22, MAE = 0,19$$

$$pK_{d,diff}(CAXII - CA VII) = 2,04 MATS8p + 53,3 BELe5 - 0,214 Mor04m - 107, \quad (3.2.7-4)$$

$$R^2 = 0,83, R^2_{ADJ} = 0,81, Q^2_{LOO} = 0,78, F(3,26) = 42, p < 10^{-9}, R^2_{TEST} = 0,49, Q^2_{TEST} = 0,46, MAE = 0,45$$

$$pK_{d,diff}(CA XII - CA XIII) = -12,7 MATS2v + 2,29 MATS8p - 0,866 H4p + 3,43, \quad (3.2.7-5)$$

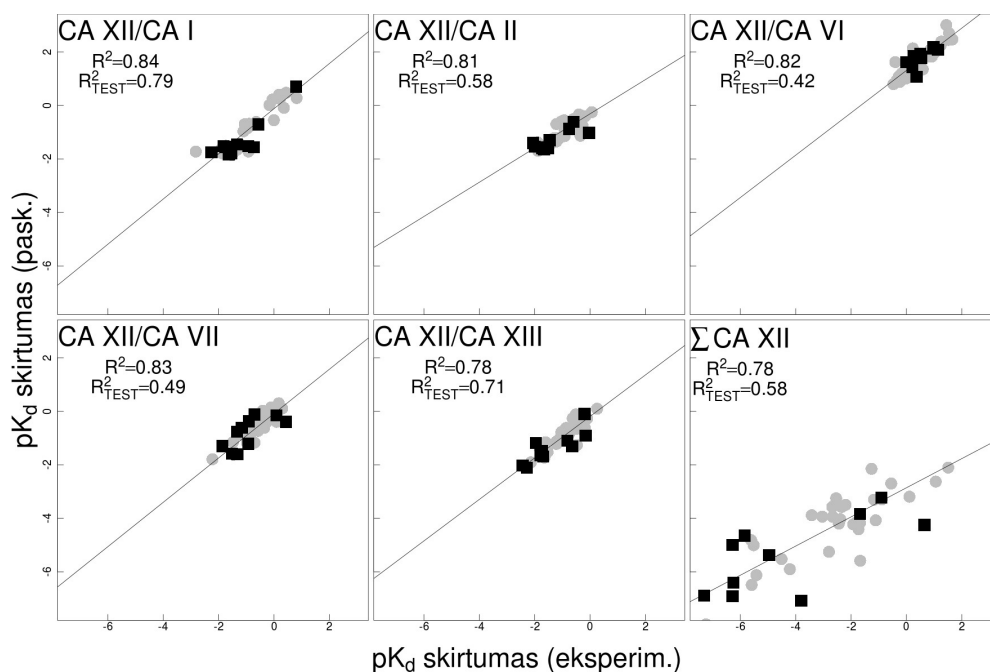
$$R^2 = 0,78, R^2_{ADJ} = 0,75, Q^2_{LOO} = 0,74, F(3,26) = 30, p < 10^{-7}, R^2_{TEST} = 0,71, Q^2_{TEST} = 0,68, MAE = 0,36$$

$$pK_{d,diff}(\sum CA XII) = -5,40 MATS8e + 0,308 RDF055m - 41,9 HATS8p - 0,0513. \quad (3.2.7-6)$$

$$R^2 = 0,78, R^2_{ADJ} = 0,75, Q^2_{LOO} = 0,74, F(3,26) = 31, p < 10^{-7}, R^2_{TEST} = 0,58, Q^2_{TEST} = 0,56, MAE = 1,34$$

Statistiniai parametrai selektyvumo QSAR modeliams yra pateikti po kiekviena QSAR lygtimi. Modeliai buvo laikomi priimtinais, jeigu jų determinacijos koeficientas R^2 ir redukuotas determinacijos koeficientas R^2_{ADJ} buvo aukščiausi tam duomenų rinkiniui. Modelių tinkamumas prognozėms buvo išnagrinėtas, tiek taikant statistinį LOO metodą su įverčiu Q^2 , tiek pagal suskaičiuotas R^2_{TEST} , Q^2_{TEST} ir MAE statistinių parametru vertes dešimties

junginių testavimo rinkiniui. Molekuliniai deskriptoriai, esantys lygtyse 3.2.7-1 – 3.2.7-6, plačiau aprašyti priede esančioje lentelėje B. Grafikai su atidėtomis eksperimentinėmis ir prognozuojamomis iš sukurtų QSAR modelių selektyvumų (pK_d skirtumų) vertėmis, naudojant apmokymo (pažymėta pilka spalva) ir testavimo (pažymėta juoda spalva) rinkinius, pateikti paveiksle 3.2.7-1.



Paveikslas 3.2.7-1. Selektivityvumų QSAR grafikai, naudojant E-Dragon deskriptorius apmokymo (pilki skrituliukai) ir testavimo (juodi kvadratėliai) rinkinių junginiams. X ir Y ašyse atidėti atitinkamai eksperimentiniai ir prognozuojami selektyvumai.

Po to buvo sukurti QSAR afiniškumo modeliai, imant išmatuotus afiniškumus kiekvienai atskirai CA izoformai:

$$pK_d(\text{CA I}) = 1,14 \text{ MATS8e} + 5,47 \text{ GATS5v} - 2,98 \text{ GATS7p} + 4,47, \quad (3.2.7-7)$$

$$R^2 = 0,94, R^2_{ADJ} = 0,93, Q^2_{LOO} = 0,93, F(3,26) = 138, p < 10^{-15}, R^2_{TEST} = 0,89,$$

$$Q^2_{TEST} = 0,87, MAE = 0,21$$

$$pK_d(\text{CA II}) = 2,05 \text{ MATS7m} + 1,02 \text{ H5u} + 4,49 \text{ HATS8m} + 4,86, \\ (3.2.7-8)$$

$$R^2 = 0,89, R^2_{ADJ} = 0,87, Q^2_{LOO} = 0,87, F(3,26) = 68, p < 10^{-11}, R^2_{TEST} = 0,57, \\ Q^2_{TEST} = 0,48, MAE = 0,31$$

$$pK_d(\text{CA VI}) = -0,883 \text{ GATS5m} + 0,264 \text{ H2m} + 13,1 \text{ (R7m+)} + 5,04, \\ (3.2.7-9)$$

$$R^2 = 0,79, R^2_{ADJ} = 0,76, Q^2_{LOO} = 0,72, F(3,26) = 32, p < 10^{-8}, R^2_{TEST} = 0,77, \\ Q^2_{TEST} = 0,56, MAE = 0,20$$

$$pK_d(\text{CA VII}) = 0,104 \text{ T(S..Cl)} - 6,54 \text{ PCR} + 1,59 \text{ MATS7v} + 15,2, \\ (3.2.7-10)$$

$$R^2 = 0,89, R^2_{ADJ} = 0,88, Q^2_{LOO} = 0,86, F(3,26) = 70, p < 10^{-11}, R^2_{TEST} = 0,87, \\ Q^2_{TEST} = 0,74, MAE = 0,26$$

$$pK_d(\text{CA XII}) = 1,47 \text{ T(O..Cl)} + 49,0 \text{ MATS7m} + 0,0473 \text{ H6e} + 4,24, \\ (3.2.7-11a)$$

$$R^2 = 0,82, R^2_{ADJ} = 0,80, Q^2_{LOO} = 0,78, F(3,26) = 40, p < 10^{-9}, R^2_{TEST} = 0,43, \\ Q^2_{TEST} = 0,16, MAE = 0,44$$

$$pK_d(\text{CA XII}) = 1,47 \text{ H6e} + 49,0 \text{ (R7p+)} + 0,0473 \text{ T(OH..Cl)} + 4,24, \\ (3.2.7-11b)$$

$$R^2 = 0,82, R^2_{ADJ} = 0,80, Q^2_{LOO} = 0,77, F(3,26) = 40, p < 10^{-9}, R^2_{TEST} = 0,60, \\ Q^2_{TEST} = 0,51, MAE = 0,24$$

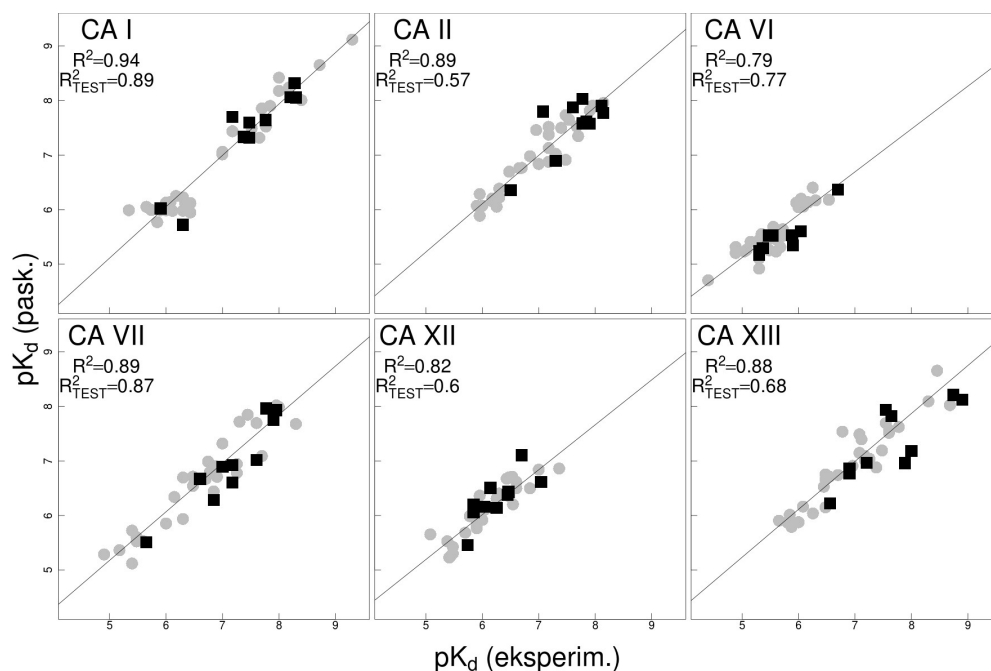
$$pK_d(\text{CA XIII}) = 18,0 \text{ EEig01r} + 0,219 \text{ G(N..S)} + 0,0380 \text{ Mor02u} - 76,5. \\ (3.2.7-12)$$

$$R^2 = 0,88, R^2_{ADJ} = 0,86, Q^2_{LOO} = 0,85, F(3,26) = 63, p < 10^{-11}, R^2_{TEST} = 0,68, \\ Q^2_{TEST} = 0,49, MAE = 0,32$$

Statistiniai QSAR modelių parametrai, pateikti žemiau kiekvienos

QSAR lygties, yra tokie pat, kaip prieš tai prie selektyvumo QSAR modelių (lygtys 3.2.7-1 – 3.2.7-6). Molekuliniai deskriptoriai, esantys šiose lygtyse, plačiau aprašyti priede, lentelėje B. Grafikai, kuriuose palygintos eksperimentiškai nustatytos afiniškumo pK_d vertės su prognozuotomis apmokymo (pažymėta pilka spalva) ir testavimo (pažymėta juoda spalva) rinkiniams, pateikti paveiksle 3.2.7-2. Priimtinos kokybės afiniškumo QSAR modeliai su nagrinėtais junginiais, taikant E-DRAGON deskriptorius, buvo sukurti visoms CA, išskyrus CA XII (lygtis 3.2.7-11a). Dėl nepriimtinos itin žemos R^2_{TEST} vertės QSAR modelyje 3.2.7-11a mes sukūrėme naują deskriptorių T(OH..Cl), kuris skirtas pagerinti CA XII afiniškumo QSAR modelio kokybę. Pagerintas QSAR modelis CA XII izoformai yra pateiktas 3.2.7-11b lygtyje. Šis molekulinis deskriptorius apibrėžtas, kaip topologinių atstumų suma tarp hidroksilo grupių ir chloro atomų molekulėje. Naujasis deskriptorius yra analogiškas E-DRAGON naudojamam deskriptoriui T(O..Cl), įtrauktam į lygtį 3.2.7-11a, išskyrus tai, kad T(O..Cl) deskriptoriuje naudojamas atstumas tarp bet kurio deguonies atomo ir chloro atomo. Dėl didelio deguonies atomų skaičiaus nagrinėjamos cheminės struktūros (paveikslas 3.2.1-1) (lentelė 3.2.1-1) nėra nieko keisto, kad, pakeitus T(O..Cl) į T(OH..Cl), žymiai pagerėja R^2_{TEST} ir Q^2_{TEST} statistiniai rodikliai (Lygtys 3.2.7-11a ir 3.2.7-11b). T(OH..Cl) išskiria sąveikai su receptoriumi konkretesnę įtaką darančias hidroksilo grupes iš daug įvairesnę įtaką darančių visų deguonies atomų, įskaičiuotų į T(O..Cl) molekulinį deskriptorių. Toks rezultatas galėtų pateisinti platesnį naudojimą tokio tipo, farmakoforą primenančių, deskriptorių, kurie skirtų įvairius to paties elemento atomų tipus (pvz. sp^2 hibridizuotas deguonis, sp^3 hibridizuotas deguonis, sujungtas su vandeniliu, ir t.t.), atliekant QSAR modeliavimus. Naujasis QSAR deskriptorius nepagerino kitų QSAR modelių kokybės, gal todėl, kad jie ir taip turėjo gana aukštus statistinius įverčius. Dar pastebėtina, kad šeši junginiai (1d, 1g, 2g, 1h, 3h, 1i) turi tokią pačią CA VI K_d vertę (5000 nM). QSAR modeliavimo procedūra atrado deskriptorių $R7m+$, kurio dydis irgi apytiksliai vienodas šiems šešiams junginiams (intervale nuo 0,058 iki 0,072). Šito deskriptoriaus paskaičiuotos

vertės varijuoja gerokai labiau kitiems nagrinėtiems junginiams (intervale nuo 0,045 iki 0,120).



Paveikslas 3.2.7-2. Afiniškumo įvairioms CA izoformoms QSAR grafikai, naudojant E-DRAGON deskriptorius apmokymo (pilki skrituliukai) ir testavimo (juodi kvadratėliai) rinkinių junginiams. X ir Y ašyse atidėti atitinkamai eksperimentiniai ir prognozuojami selektyvumai.

Taip pat buvo nustatyti pritaikomumo domenai ir išnagrinėtos prognozuojamų duomenų išskirtys. Beveik visų apmokymo rinkiniui priklausančių junginių prognozuotos jungimosi konstantos nepriskirtos išskirtims visuose šiame skyriuje minėtuose QSAR modeliuose, išskyrus vienintelį junginį 2j, kurio prognozuota disociacijos konstantos vertė visgi priskirta išskirčiai vien tik CA II slopinimo QSAR modelyje (lygtis 3.2.7-8). Pagal eksperimentinėje dalyje aprašytą metodą, visos prognozuotos testavimo rinkiniui priklausančių junginių jungimosi konstantos buvo priskirtos pritaikomumo domeno intervalui visiems šiame skyriuje minėtiems QSAR

modeliams. Σ CAXII QSAR modelis yra labai blogas pagal MAE kriterijų, o CA I slopinimo modelis yra labai geras pagal MAE kriterijų.

3.2.8. PHASE atomais paremtas trimatis QSAR, skirtas CA I ir CA XII

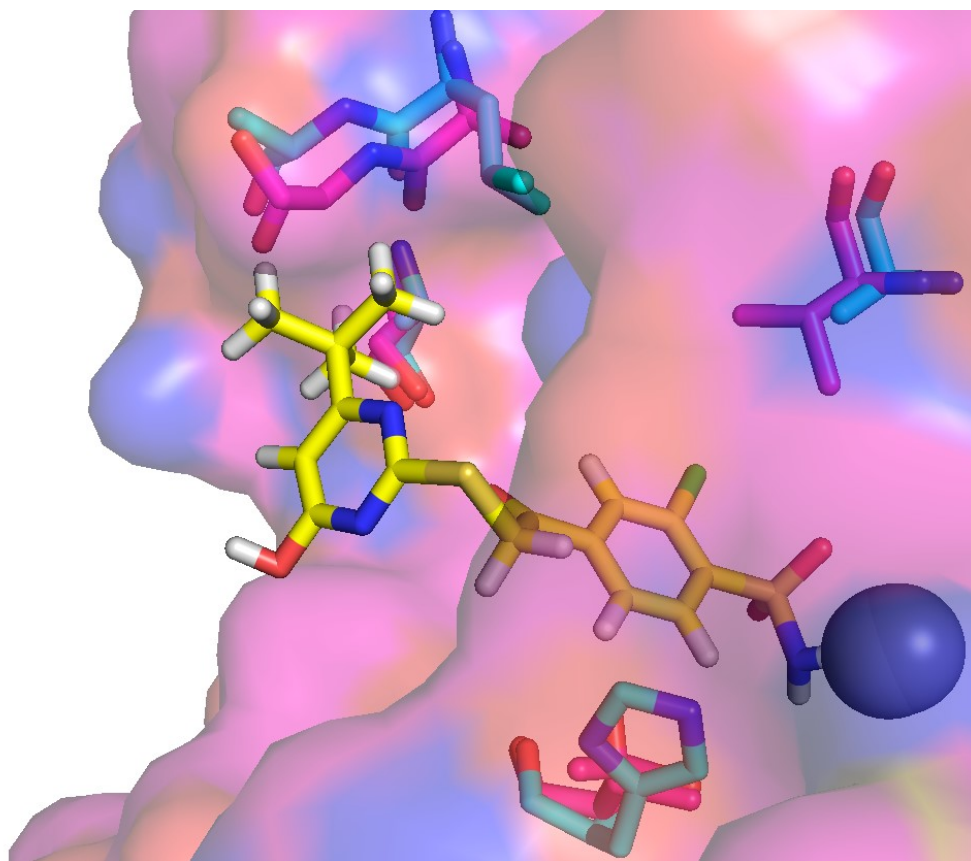
PHASE trimačio QSAR modeliai buvo ieškomi, analizuojant cheminių junginių selektyvumus CA XII, lyginant su kitomis CA izoformomis (afiniškumų santykius $pK_{d,diff}$), bei tų cheminių junginių afiniškumo duomenis (pK_d) visoms nagrinėtoms CA.

Skirtingai nei E-DRAGON molekuliniais deskriptoriais paremtame QSAR, su PHASE buvo atrasti tik du priimtinos kokybės QSAR modeliai, vertinant pagal pačios PHASE statistinio įvertinimo kriterijus – vienas iš jų CA XII/CA I selektyvumo, o kitas CA I afiniškumo.

PHASE QSAR statistiniai parametrai paskaičiuoti CA XII/CA I selektyvumo modeliui buvo tokie: $SD = 0,45$, $R^2 = 0,75$, $R^2_{CV} = 0,69$, $R^2_{Scrambled} = 0,27$, $Stability = 0,99$, $F = 86$, $P < 10^{-9}$, $RMSE = 0,52$, $Q^2 = 0,61$, $Pearson-r = 0,82$. CA XII/CA I selektyvumo modelis buvo prastas pagal MAE kriterijų ($MAE = 0,39$), kuris buvo paskaičiuotas atskirai, naudojantis R programa. Šiame modelyje testavimo rinkinio junginiai 1c ir 4e turėjo absoliučias prognozavimo paklaidas, lygias atitinkamai 0,99 ir 0,83 logaritminio vieneto. Analogiški statistiniai parametrai paskaičiuoti CA I afiniškumo modeliui buvo tokie: $SD = 0,40$, $R^2 = 0,86$, $R^2_{CV} = 0,82$, $R^2_{Scrambled} = 0,21$, $Stability = 0,99$, $F = 168$, $P < 10^{-12}$, $RMSE = 0,38$, $Q^2 = 0,75$, $Pearson-r = 0,89$. Šis modelis buvo geras ir pagal MAE kriterijų ($MAE = 0,35$).

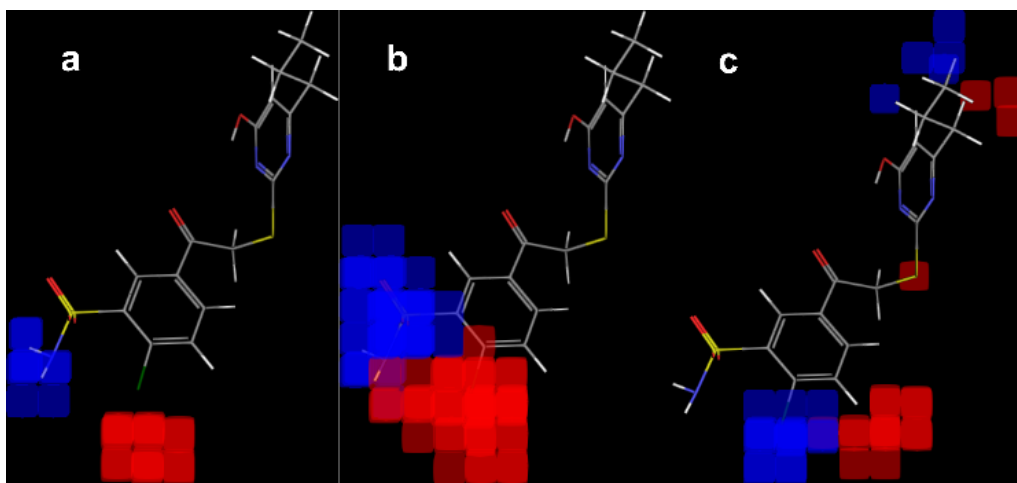
Siekiant geriau suprasti PHASE CA I ir CA XII selektyvumo modeliavimo rezultatus, mes išnagrinėjome ligandų sąveikas su dviejų CA izoformų aktyviaisiais centrais, atlikdami rankinio dokavimo procedūrą su PYMOL programa (versija 1.7.4.0) (Schrödinger, LLC). CA izofermentų CA I ir CA XII PDB struktūros (PDB ID: 4WR7 ir 4KP8) buvo perklotos tarpusavyje, ir aktyviausias CA I atžvilgiu junginys 2d buvo rankiniu būdu įdėtas į perklotą

CA aktyvųjį centrą. Tret-butilo pakaitas, esantis ant 2d junginio pirimidino žiedo, sąveikauja su CA I hidrofobinėmis Ala132, Ala135 ir Leu131 šoninėmis grandinėmis, o hidroksilo grupė lieka orientuota palankiai sąveikai su vandeniu. Ji yra atsukta į išorę, į vandenį (paveikslas 3.2.8-1). Palyginimui, CA XII turi Ser130, Ser133 ir Ala129 šonines grandines homologinėse pozicijose ir negali palankiai sąveikauti su nagrinėjamo ligando hidrofobine tret-butilo grupe. Pirimidino žiedas galėtų apsisukti, ir tada hidroksilo grupė sudarytų geresnius kontaktus su minėtomis trejomis CA XII šoninėmis grandinėmis, bet tada hidrofobinė tret-butilo grupė būtų nepalankiai atsukta į aktyviojo centro išorę, į vandenį.



Paveikslas 3.2.8-1. Aktyviausias CA I atžvilgiu junginys 2d (geltoni pagaliukai), rankiniu būdu įdokintas į perklotas CA I (žydra, PDB ID: 4WR7) ir CA XII (violetinė, PDB ID: 4KP8) PDB struktūras. Didelė mėlva sfera vaizduoja cinko joną.

PHASE programa kartu su MAESTRO programiniu paketu suteikia galimybę vizualizuoti įvairių ligando funkcinių grupių įtaką biologiniam aktyvumui pagal atrastą trimačio QSAR modelį, šiuo atveju ligando afiniškumui ar jo jungimosi selektyvumui. PHASE trimačio QSAR modelis CA XII/CA I selektyvumui parodytas paveiksle 3.2.8-2, naudojant labiausiai šioje poroje selektyvų junginį 4c, kaip pavyzdį. Pagal paveikslą 3.2.8-2, trys pagrindiniai faktoriai, kurie veikia CA XII/CA I selektyvumą 4c junginiui, yra: chloro atomas, prijungtas prie benzeno žiedo *para* padėtyje jungties tarp dviejų junginio žiedų atžvilgiu (mėlyna zona apatinėje paveikslo 3.2.8-2 c dalyje), šalia to chloro atomo esanti sulfonamido grupė *meta* padėtyje (mėlyna zona apatinėje dalyje paveikslų 3.2.8-2 a ir 3.2.8-2 b), ir, galiausiai, kiek mažiau svarbus selektyvumui hidrofobinis pakaitas pirimidino žiede, *meta* padėtyje skaičiuojant nuo jungties tarp dviejų junginio žiedų (viršutinė paveikslo 3.2.8-2 c dalis).

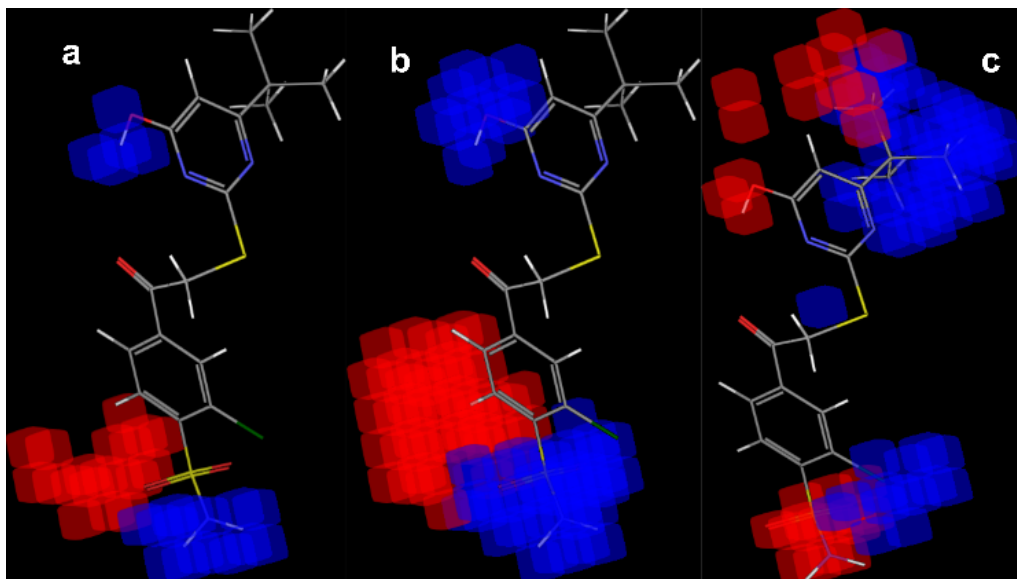


Paveikslas 3.2.8-2. PHASE atomais paremto trimačio QSAR CA XII/CA I selektyvumo modelis, vizualizuotas, kaip pavyzdį imant selektyviausią šioje poroje junginį 4c. Mėlyni ir raudoni kubeliai vaizduoja atitinkamai palankius ir nepalankius selektyvumui regionus. Trijose paveikslo dalyse pavaizduoti: (a) vandenilinio ryšio donoriai, (b) elektronus ištraukiantys atomai (įskaitant ir vandenilinio ryšio acceptorius), (c) hidrofobinės/nepolinės grupės. Atvaizduojant kubelius, slenkstinis grupės svarbos koeficientas parinktas

$\pm 2,5 \cdot 10^{-3}$.

PHASE/MAESTRO atrasto afiniškumo QSAR modelio CA I vizualizacija, naudojant labiausiai aktyvų junginį 2d, kaip pavyzdį, pateikta paveiksle 3.2.8-3. Įdomu, kad paveikslas 3.2.8-3 daugeliu atžvilgių yra atvirkščias paveiksliui 3.2.8-2: dažnai viename paveiksle mėlyna zona yra raudona kitame paveiksle, ir atvirkščiai.

Mėlynos zonos prie benzeno žiedo abiejuose PHASE QSAR paveiksluose 3.2.8-2 (a, b) ir 3.2.8-3 (a, b) parodo sulfonamido grupės padėtis geriausiose nagrinėjamos serijos subserijose 4 ir 2 atitinkamuose QSAR modeliuose. Lygiai kaip ir rankinio dokavimo paveikslas 3.2.8-1, taip ir PHASE afiniškumo QSAR modelio paveikslas 3.2.8-3 rodo hidrofobinių ir hidrofilinių pakaitų svarbą ant pirimidino žiedo nagrinėjamų junginių jungimuisi su CA I. Vienas iš QSAR metodo, kuriame paprastai naudojami vien tik ligandai, trūkumų yra tas, kad šis metodas neatsižvelgia į galimus ligando prisijungimus visai kitoje konformacijoje. Šiame pavyzdyje PHASE perkloja visų junginių benzeno ir pirimidino žiedus panašiai, kaip nagrinėjamos serijos junginių struktūras aprašančiame paveiksle 3.2.1-1. Iš esamų kristalografinės analizės duomenų galima spręsti, kad labiausiai ligando prisijungimo vietoje suvaržyta yra jo sulfonamido grupė, kadangi ji yra prijungta prie cinko, tad tinka perkloti junginius taip, kad sulfonamido grupės būtų užklotos viena ant kitos. Jungtukai tarp abiejų žiedų, bei pirimidino žiedai taip pat gali turėti kelias gana skirtingas pozicijas erdvėje, t.y. tikrovėje nebūtinai gali būti persikloję [288].



Paveikslas 3.2.8-3. PHASE atomais paremta trimačio QSAR CA I afiniškumo modelis vizualizuotas, kaip pavyzdį imant aktyviausią CA I atžvilgiu junginį 2d. Mėlyni ir raudoni kubeliai vaizduoja atitinkamai palankius ir nepalankius aktyvumui regionus. Trijose paveikslų dalyse pavaizduoti: (a) vandenilinio ryšio donoriai, (b) elektronus-ištraukiantys atomai (įskaitant ir vandenilinio ryšio akceptorius), (c) hidrofobinės/nepolinės grupės. Atvaizduojant kubelius, slenkstinis grupės svarbos koeficientas parinktas $\pm 2,5 \cdot 10^{-3}$.

PHASE QSAR ne visada gali tiksliai parodyti visas priežastis, nulemiančias selektyvumą. Pavyzdžiui, suvaržyta sulfonamido pozicija ketvirtojoje subserijoje priverčia jos junginių pirimidino žiedus pereiti į kitą prisijungimo zoną, nei įprastai, lyginant su kitomis subserijomis (neparodyta). CA I aktyviajame centre žiedus jungianti grandis ligande erdviškai sąveikauja su gana didele His200 šonine grandine, o CA XII turi Thr199 homologiškoje padėtyje. Visa tai veda prie pagerinto CA XII/CA I selektyvumo ketvirtajai subserijai. Reikšmingų įtakų aplinkui pirimidino žiedą CA XII/CA I selektyvumo QSAR modelyje trūkumas, pavaizduotas 3.2.8-2 paveikslų viršutinėje dalyje, irgi parodo, kad selektyvumą daugiausiai lemia įvairūs benzeno žiedo pakaitai. Kita vertus, galimybės sukurti selektyvius CA XII/CA I junginius, parenkant pirimidino žiedo pakaitus, tada ne tokios didelės. Jei ligandai jungiasi skirtingose konformacijose, ligandų perklojimas nebėra tiek

prasmingas: neteisingai perkloti ligandai gali sukelti ligandų funkcines grupes į vieną padėtį erdviėje, o tos grupės ligando jungimosi konformacijose gali būti visai skirtingose taikinio vietose.

Siekiant patobulinti trimačio QSAR modelius, erdviškai suvaržytos ligando funkcinės grupės ("farmakoforai") turetų būti perklotos kartu, bet PHASE programa nepritaikyta tokiam ligandų paruošimo restruktūravimui. Nesant jokios struktūrinės informacijos apie ligando sąveiką su taikiniu, viena praktinė išeitis būtų naudoti kiek įmanoma panašesnius struktūrinius fragmentus.

Su PHASE programa pavyko generuoti geros kokybės trimatį QSAR modelį CA XII/CA I selektyvumui. Kita vertus, priimtino CA XII slopinimo QSAR modelio su PHASE programa sukurti nepavyko, kaip ir kitoms CA izoformoms, išskyrus CA I. Gero CA XII su PHASE programa kurto trimačio QSAR modelio nebuvo sukurta, bet kaip alternatyvą galima pamėginti naudoti CAXII/CA I selektyvumo modelį ir iš jo tobulinti ne tik selektyvumą, bet ir CA XII slopinimą įvairiais slopikliais. E-DRAGON deskriptoriais paremtas QSAR padeda sukurti daug geresnius modelius, mat E-DRAGON turi žymiai daugiau ir įvairių molekulinį deskriptorių ir kriterijų, iš kurių galima generuoti modelius. Skirtingos ligandų prisijungimo konformacijos serijoje daro ženklų įtaką trimatį PHASE modelių kokybei, o dvimačių E-DRAGON deskriptoriais paremtų modelių gerokai mažesnę, nes E-DRAGON tik kai kurie pavieniai molekuliniai deskriptoriai skaičiuojami iš trimatės struktūros. Didelis PHASE programos pranašumas yra tas, kad įvairių faktorių įtaką afiniškumui ar selektyvumui lengva vizualizuoti. E-DRAGON molekulinis deskriptorius dažnai nėra trivialu interpretuoti.

3.2.9. Selektivumo iš atskirų afiniškumo QSAR modelių skaičiavimai palyginti su selektivumo QSAR

Prieš tai buvusiuose skyriuose apie selektivumo QSAR, selektivumas buvo skaičiuojamas iš afiniškumo duomenų ir tuomet įkeliamas į QSAR modeliavimo programą. Gauti optimizuoti selektivumo QSAR modeliai buvo

taikomi selektyvumui prognozuoti. Mes išbandėme alternatyvų selektyvumo prognozavimo kelią, pirmiausiai paskaičiuojant afiniškumą iš dviejų afiniškumo QSAR modelių nagrinėjamos CA izoformoms, ir tik po to skaičiuojant tų prognozuojamų afiniškumų santykį, taigi selektyvumą. Lentelėje 3.2.9-1 pateiktas R^2_{TEST} verčių palyginimas selektyvumams, skaičiuotiems abiem minėtais būdais iš E-DRAGON molekuliniais deskriptoriais paremtų QSAR modelių.

Lentelė 3.2.9-1. E-DRAGON deskriptoriais paremto dvimačio QSAR testavimo rinkinio R^2_{TEST} statistinio parametro vertė visiems CA XII selektyvumo modeliams bei CA XII selektyvumams, paskaičiuotiems iš atskirų afiniškumo QSAR modelių. $R^2_{TEST} > 0,40$ vertės pateiktos paryškintu juodu šriftu.

	E-DRAGON deskriptoriais paremto QSAR R^2_{TEST} vertės: selektyvumas, paskaičiuotas iš atskirų slopinimo QSAR/selektyvumo QSAR
CA XII/CA I	0,85/0,79
CA XII/CA II	0,34/ 0,58
CA XII/CA VI	0,28/ 0,42
CA XII/CA VII	0,58/0,49
CA XII/CA XIII	0,65/0,71
Σ CA XII	0,63/0,58

Naudojant E-DRAGON molekuliniais deskriptoriais paremtą selektyvumo QSAR, priimtinos R^2_{TEST} vertės buvo gautos visais atvejais. Dviem atvejais, CA XII/CA I ir CA XII/CA XIII, R^2_{TEST} buvo didesnis, negu 0,70. Kai selektyvumai buvo paskaičiuoti iš atskirų afiniškumo QSAR modelių, priimtinos R^2_{TEST} vertės buvo gautos tik keturiais atvejais iš šešių nagrinėtų. CA XII/CA II ir CA XII/CA VI selektyvumų iš atskirų afiniškumo QSAR modelių R^2_{TEST} buvo mažesnis, negu 0,40. R^2_{TEST} parametro verčių aritmetinis vidurkis yra geresnis selektyvumo QSAR modeliams, nei selektyvumams, paskaičiuotiems iš afiniškumo QSAR prognozių (0,595 ir 0,555 atitinkamai). Tai dar kartą parodo selektyvumo QSAR privalumus.

3.2.10. Benzensulfonamidų QSAR apibendrinimas įvairioms CA izoformoms

Dvimačiai ir trimačiai, afiniškumo ir selektyvumo QSAR modeliai buvo sukurti, naudojant trisdešimties junginių apmokymo rinkinį bei du skirtingus protokolus šešioms nagrinėtoms CA izoformoms (CA I, CA II, CA VI, CA VII, CA XII ir CA XIII). Pirmasis protokolus buvo paremtas E-DRAGON ir R programine įranga, antrasis buvo atliktas, naudojant Schrödinger programinį paketą MAESTRO (versija 2015-3 (10.3)) [289]. Buvo pritaikytos ir palygintos dvi skirtingos strategijos prognozuoti CA XII selektyvumui kitų CA izoformų atžvilgiu: arba sukuriant specializuotus selektyvumo modelius, arba prognozuojant selektyvumus iš dviejų atskirų afiniškumo QSAR modelių. Keturiolika iš visų benzensulfonamidams sukurtų QSAR modelių pagal įvairius statistinius tinkamumo patvirtinimo įverčius buvo pripažinti tinkamais prognozėms. Modeliai buvo pripažinti tinkamais tiek su LOO metodu, tiek testuojant su dešimties junginių testavimo rinkiniu. E-DRAGON molekuliniais deskriptoriais paremtu QSAR protokolu buvo generuoti daugiau ar mažiau prognozėms tinkami modeliai visų šešių nagrinėtų CA izoformų slopiklių afiniškumui, ir visiems CA XII slopiklių selektyvumams kitų penkių izoformų atžvilgiu prognozuoti. SCHRÖDINGER trimatis QSAR (PHASE programa) suteikė grafinius atvaizdavimus apie analizuotą duomenų rinkinį, bet daugumos PHASE QSAR modelių bendra statistinė kokybė buvo prastesnė, lyginant su E-DRAGON ir R programomis paremtu protokolu. Dar buvo parodyta, kad, kuriant specializuotus selektyvumo QSAR modelius, gaunami aukštesni tų modelių statistiniai įverčiai, nei prognozuojant slopiklių afiniškumus porai CA izoformų atskirai, ir tik paskui paskaičiuojant jų selektyvumą iš turimų prognozių. Dar mes parodėme, kad imant mūsų naują molekulinį deskriptorių T(OH..Cl), kuris aprašo topologinį atstumą tarp junginių funkcinių grupių OH ir Cl, pavyko pagerinti E-DRAGON ir R programine įranga paremto CA XII afiniškumo QSAR modelio kokybę, lyginant su originaliu QSAR modeliu, į kurį buvo įtrauktas molekulinis deskriptorius T(O..Cl), aprašantis topologinį atstumą tarp O ir Cl elementų

atomų.

Šiame darbe naudotos 40 benzensulfonamidų sąveikos su įvairiomis CA izoformomis K_d vertės buvo publikuotos 2013 metais [288]. Jos eksperimentiškai buvo nustatytos terminio poslinkio analizės (TSA) metodu. 2014 metais 16 iš šių benzensulfonamidų sąveikos su įvairiomis CA izoformomis buvo tirtos izoterminio titravimo kalorimetrijos (ITC) metodu [290]. Tiek TSA, tiek ITC metodais nustatytos benzensulfonamidų K_d vertės laikomos stebimosiomis. Iš jungimosi konstantų pK_d , junginių disociacijos konstantų pK_a ir CA aktyviajame centre esančio cinko jono rūgštingumo pK_a galima paskaičiuoti tikrines (angl. *intrinsic*) junginių disociacijos konstantas, kurios nepriklauso nuo tirpalo pH [290]. Įdomu būtų patikrinti, kaip tiksliai šiame darbe naudoti metodai prognozuotų tikrines disociacijos konstantas. Kadangi šiame darbe modelių ir metodų parametrizavimui buvo imtos tik TSA stebimosios K_d vertės, atitinkamai ir visas šiame darbe prognozuojamas K_d vertes reiktų laikyti *prognozuojamomis stebimosiomis*, kurios lygintinos tik su analogiškais TSA rezultatais. Iš jų ir tų junginių pK_a galima skaičiuoti *prognozuojamas tikrines* K_d vertes lygiai taip pat, kaip ir iš eksperimentinių duomenų, kurios tada lygintinos tik su analogiškai paskaičiuotomis tikrinėmis vertėmis.

3.3. Naujų panašių į metabolitus slopiklių paieška naudojant Tanimoto įvertį

Naujų slopiklių paieškos būdų yra įvairių. Tradiciškai išbandomas didelio organinių junginių rinkinio jungimasis su nagrinėjamu taikiniu. Dažnai tam naudojami *in silico* metodai. Kai tapo prieinamos didelės metabolitų, vaistų ir kitų organinių junginių duomenų bazės, palyginamosios slopiklių paieškos strategijos gali būti taikomos naujų slopiklių, panašių į metabolitus, paieškai. Šiame darbe sukurta ir naudota nauja slopiklių paieškos strategija, paremta Tanimoto molekulių struktūrų panašumo įverčiu. Pagal ją buvo palyginamos skirtingos molekulių struktūros trijose duomenų bazėse (KEGG, DRUGBANK ir ZINC) [291-296], siekiant rasti naujus galimus slopiklius. Buvo

paskaičiuota tikimybė rasti naujus slopiklius su šia paieškos strategija. Sėkmingai taikant sukurtąją strategiją praktikoje, buvo surastas naujas slopiklis, panašus į vieną iš metabolitų, besijungiantis į tą taikinį, kurio substratas yra minėtasis metabolitas.

Naujų slopiklių paieška paprastai yra atliekama mažoms organinėms molekulėms, kurios jungiasi prie biologinio taikinio ir slopina jo biologinę funkciją. Labai pageidautina, kad atrastaisiais junginiais būtų prekiaujama rinkoje, antraip tektų pirmiausiai tuos junginius susintetinti, o tik tada būtų galima eksperimentiškai parodyti jų biologinį aktyvumą. Iš to seka akivaizdi nauda strategijos, kuri pagrįsta slopiklių paieška prieinamų organinių junginių duomenų bazėse. Iš enzimologijos gerai žinoma, kad slopiklis dažnai panašus į substratą. Panašumui tarp junginių aprašyti kiekybiškai gali būti panaudotas Tanimoto įvertis. Tanimoto įvertis – tai skaičiumi išreikštas panašumas tarp vadinamųjų molekulių pirštų antspaudų, išreikštų kaip dvejetainės skaitmenų (bitų) serijos. Tanimoto įvertis gali svyruoti nuo nulio (jokio panašumo) iki vieneto (aukštas panašumas arba identiškumas) [297]. Molekulių struktūros gali būti konvertuojamos į molekulių pirštų antspaudus, naudojantis įvairiomis programomis.

Buvo parsisiųstos 1475 žmogaus metabolitų cheminės struktūros iš KEGG duomenų bazės, kurioje yra išsamūs duomenys apie žmogaus metabolitus. Visoms galimoms metabolitų ir DRUGBANK (<http://www.drugbank.ca/>) duomenų bazės cheminių struktūrų poroms buvo paskaičiuotas Tanimoto įvertis naudojant „FP2” molekulių pirštų antspaudus su OPEN BABEL (<http://openbabel.org>) programa [267].

Iš visų gautų duomenų tolesnei analizei buvo išrinktos tik metabolitų ir DRUGBANK junginių poros, kurių tarpusavio Tanimoto įverčiai buvo didesni, negu 0,9. Kiekvienai išrinktai DRUGBANK molekulės struktūrai jos taikinių EC numeriai buvo paimti iš DRUGBANK internetinio puslapio. Tokia pati procedūra buvo pakartota kiekvienai išrinktai žmogaus metabolito struktūrai, tik naudojant duomenis iš KEGG duomenų bazės. 2817 junginių porose abi

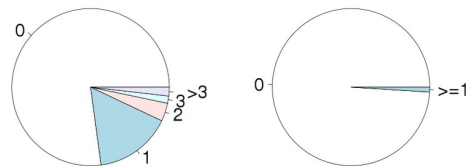
nagrinėtos struktūros turėjo žinomus taikinius. Kitos poros buvo atmestos. Galiausiai buvo išrinktos 644 žmogaus metabolitų ir DRUGBANK esančių junginių poros su Tanimoto įverčiu, didesniu kaip 0,9, ir su bent vienu sutampančiu taikiniu.

Iš šių skaičiavimų buvo padaryta išvada, kad, kai analizuojama molekulės struktūra yra panaši į metabolito struktūrą (Tanimoto panašumo įvertis didesnis kaip 0,9), tada yra maždaug 23% ($644/2817 \cdot 100\%$) tikimybė, kad analizuojamoji molekulė jungsis prie to paties taikinio, kaip ir panašus metabolitas (paveikslas 3.3-1, viršuje ir kairėje). Tai tik apytikris įvertinimas, nes nebuvo nagrinėti literatūroje neaprašyti taikiniai bei taikiniai, kurie neįkelti į duomenų bazes, taikiniai be EC numerių ir t.t.

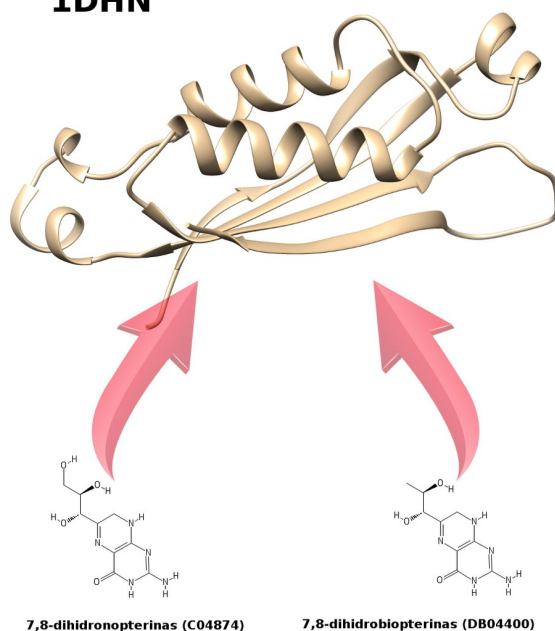
Galimas variantas, kad taikiniai tiesiog sutampa atsitiktinai. Siekiant įvertinti tokią tikimybę, buvo generuota keturi tūkstančiai atsitiktinių žmogaus metabolitų ir DRUGBANK junginių porų. Kiekvienai atsitiktinai parinktai DRUGBANK molekulei jos taikinių EC numeriai buvo paimti iš DRUGBANK internetinio puslapio. Tokia pat procedūra pakartota kiekvienai atsitiktinai parinktai žmogaus metabolito struktūrai, tik naudojant KEGG duomenų bazės duomenis. Išnagrinėjus atvejus, kai taikiniai sutampa abiem struktūroms, paskaičiuota apytiksliai tik 1% tikimybė, kad atsitiktinai parinkus žmogaus metabolito ir DRUGBANK molekulės struktūros porą, abu tos poros nariai turės tiksliai tą patį taikinį (paveikslas 3.3-1, viršuje ir dešinėje).

Pasiūlyta metodologija buvo pritaikyta penkiasdešimties su vėžio augimu susijusių metabolitų sąrašui. Procedūra leido pasiūlyti keliolika galimų vėžio augimo slopiklių, iš kurių buvo tokių, kurių aktyvumas prieš vėžines ląsteles vėliau pasitvirtino eksperimentiškai. Šie radiniai yra toliau tyrinėjami, ir jau yra išsiųstas straipsnis į aukšto citavimo indekso mokslinį žurnalą.

Tanimoto > 0,90 Atsitiktinės poros



1DHN



Paveikslas 3.3-1. Du panašūs junginiai 7,8-dihidronopterinas ir 7,8-dihidrobiopterinas (Tanimoto > 0,90) turi tą patį taikinį, 7,8-dihidronopterino aldolazę (PDB ID: 1DHN). Tikimybė, kad junginių, kurių Tanimoto panašumo įvertis yra didesnis, negu 0,90, taikinis sutaps, yra ~23%. Tikimybė, kad sutaps keli taikiniai, yra gerokai mažesnė. Sutampančių taikinių skaičius pažymėtas šalia viršutinio grafiko.

4. Išvados

1. Sukurti šeši α -pakeistų α,β -nesočiųjų ketonų QSAR modeliai trims ląstelių linijoms (NB4, MCF-7 ir A549) gali atskirti aktyvius junginius nuo neaktyvių.
2. Iš penkių išbandytų metodų QSAR rezultatai geriausiai koreliuoja su eksperimentiniais duomenimis (R^2 nuo 0,83-0,89). Su QSAR metodu galima geriausiai prognozuoti benzensulfonamidų jungimosi prie CA II konstantas.
3. Taikant naujai sukurtą „LIE-like“ metodą koreliacija tarp eksperimentinių duomenų ir prognozuotų žymiai pagerėjo lyginant su originaliu LIE metodu (R^2 padidėjo nuo 0,24 iki 0,50).
4. Parodyta, kad, kuriant specializuotus selektyvumo QSAR modelius, gaunami aukštesni tų modelių statistiniai įverčiai, nei prognozuojant slopiklių afiniškumus porai CA izoformų atskirai, ir tik paskui paskaičiuojant jų selektyvumą iš turimų prognozių.

Literatūros sąrašas

[1] Bleicher, K.H.; Böhm, H.J.; Müller, K.; Alanine, A.I. Hit and lead generation: beyond high-throughput screening. *Nat. Rev. Drug. Discov.* **2003**, *2*, 369-378.

[2] Kitchen, D.B.; Decornez, H.; Furr, J.R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug. Discov.* **2004**, *3*, 935-949.

[3] Supuran, C.T. Carbonic anhydrases: novel therapeutic applications for inhibitors and activators. *Nat. Rev. Drug Discov.* **2008**, *7*, 168-181.

[4] Supuran, C.T.; Scozzafava, A. Carbonic anhydrases as targets for medicinal chemistry. *Bioorg. Med. Chem.* **2007**, *15*, 4336.

[5] Badger, M.R.; Price, G.D. The role of carbonic anhydrase in photosynthesis. *Annu. Rev. Biol.* **1994**, *45*, 369-392.

[6] Supuran, C.T. Carbonic anhydrases: novel therapeutic applications for inhibitors and activators. *Nat. Rev. Drug Discov.* **2008**, *7*, 168-81.

[7] Dedeoglu, N.; DeLuca, V.; Isik, S.; Yildirim, H.; Kockar, F.; Capasso, C.; Supuran, C.T. Sulfonamide inhibition study of the β -class carbonic anhydrase from the caries producing pathogen *Streptococcus mutans*. *Bioorg. Med. Chem. Lett.* **2015**, *25*, 2291-2297.

[8] Coleman, J.E. Mechanism of action of carbonic anhydrase. Substrate, sulfonamide, and anion binding. *J. Biol. Chem.* **1967**, *242*, 5212-5219.

[9] Faridi, S.; Satyanarayana, T. Applicability of Carbonic Anhydrases in Mitigating Global Warming and Development of Useful Products from CO₂. *CCES.* **2015**, *3*, 77-92.

[10] Ferry, J.G. Carbonic Anhydrases of environmentally and medically relevant anaerobic prokaryotes. Elsevier BV: Berlin, Germany; 2015 Jan 14.

[11] Supuran, C.T.; Capasso, C. The η -class carbonic anhydrases as drug targets for antimalarial agents. *Expert Opin. Ther. Targets* **2015**, *19*, 551-563.

[12] Innocenti, A.; Supuran, C. T. Paraoxon, 4-nitrophenyl phosphate and acetate are substrates of α - but not of β -, γ - and ζ -carbonic anhydrases. *Bioorg. Med. Chem. Lett.*, **2010**, *20*, 6208-6212.

[13] Dedeoglu, N.; DeLuca, V.; Isik, S.; Yildirim, H.; Kockar, F.; Capasso, C.; Supuran, C.T. Sulfonamide inhibition study of the β -class carbonic anhydrase from the caries producing pathogen *Streptococcus mutans*. *Bioorg. Med. Chem. Lett.* **2015**, *25*, 2291-2297.

[14] Syrjänen, L.; Valanne, S.; Kuuslahti, M.; Tuomela, T.; Sriram, A.; Sanz, A.; Jacobs, H.T.; Rämetsä, M.; Parkkila, S. β carbonic anhydrase is required for female fertility in *Drosophila melanogaster*. *Front. Zool.* **2015**, *12*, 1.

[15] Aggarwal, M.; Boone, C.D.; Kondeti, B.; McKenna, R. Structural annotation of human carbonic anhydrases. *J. Enzyme Inhib. Med. Chem.* **2013**, *28*, 267-277.

[16] Mahon, B.P.; Bhatt, A.; Vullo, D.; Supuran, C.T.; McKenna, R. Exploration of anionic inhibition of the α -carbonic anhydrase from *Thiomicrospira crunogena* XCL-2 gamma proteobacterium: a potential biocatalytic agent for industrial CO₂ removal. *Chem. Eng. Sci.* **2015**, *138*, 575-580.

[17] Tanc, M.; Carta, F.; Scozzafava, A.; Supuran, C.T. α -Carbonic Anhydrases Possess Thioesterase Activity. *ACS Med Chem Lett.* **2015**, *6*, 292-295.

[18] Krishnamurthy, V.M.; Kaufman, G.K.; Urbach, A.R.; Gitlin, I.; Gudiksen, K.L.; Weibel, D.B.; Whitesides, G.M. Carbonic anhydrase as a model for biophysical and physical-organic studies of proteins and protein-ligand binding. *Chem. Rev.* **2008**, *108*, 946-1051.

[19] Rowlett, R.S. Structure and catalytic mechanism of the beta-carbonic anhydrases. *Biochim. Biophys. Acta, Proteins Proteomics* **2010**, *1804*, 362–373.

[20] Hewett-Emmett, D. Evolution and distribution of the carbonic anhydrase gene families. *EXS*. **2000**, *90*, 29-76.

[21] Syrjänen, L.; Tolvanen, M.; Hilvo, M.; Olatubosun, A.; Innocenti, A.; Scozzafava, A.; Leppiniemi, J.; Niederhauser, B.; Hytönen, V.P.; Gorr, T.A.; Parkkila, S. Characterization of the first beta-class carbonic anhydrase from an arthropod (*Drosophila melanogaster*) and phylogenetic analysis of beta-class carbonic anhydrases in invertebrates. *BMC Biochem*. **2010**, *11*, 1.

[22] Fasseas, M.K.; Tsikou, D.; Flemetakis, E.; Katinakis, P. Molecular and biochemical analysis of the α class carbonic anhydrases in *Caenorhabditis elegans*. *Mol. Biol. Rep.* **2009**, *38*, 1777-1785.

[23] Emaeh, R.Z.; Kuuslahti, M.; Vullo, D.; Barker, H.R.; Supuran, C.T.; Parkkila, S. *Ascaris lumbricoides* β carbonic anhydrase: a potential target enzyme for treatment of ascariasis. *Parasit. Vectors* **2015**, *8*, 1.

[24] Syrjänen, L.; Vermelho, A.B.; de Almeida Rodrigues, I.; Corte-Real, S.; Salonen, T.; Pan, P.; Vullo, D.; Parkkila, S.; Capasso, C.; Supuran, C.T. Cloning, characterization, and inhibition studies of a β -carbonic anhydrase from *Leishmania donovani chagasi*, the protozoan parasite responsible for leishmaniasis. *J. Med. Chem.* **2013**, *56*, 7372-7381.

[25] Smith, K.S.; Ferry, J.G. Prokaryotic carbonic anhydrases. *FEMS Microbiol. Rev.* **2000**, *24*, 335-366.

[26] Iverson, T.M.; Alber, B.E.; Kisker, C.; Ferry, J.G.; Rees, D.C. A closer look at the active site of gamma-class carbonic anhydrases: high-resolution crystallographic studies of the carbonic anhydrase from *Methanosarcina thermophila*. *Biochemistry*. **2000**, *39*, 9222-9231.

[27] Ferry, J.G. The γ class of carbonic anhydrases. *Biochim. Biophys.*

Acta, Proteins Proteomics. **2010**, *1804*, 374-381.

[28] del Prete, S.; Vullo, D.; de Luca, V.; Supuran, C.T.; Capasso, C. Biochemical characterization of the δ -carbonic anhydrase from the marine diatom *Thalassiosira weissflogii*, TweCA. *J. Enzyme Inhib. Med. Chem.*, **2014**, *29*, 906-911.

[29] Robert, S.B.; Lane, T.W.; Morel, F.M.M. Carbonic anhydrase in the marine diatom *thalassiosira weissflogii* (bacillariophyceae) 1. *J. Phycol.* **1997**, *33*, 845-850.

[30] Xu, Y.; Feng, L.; Jeffrey, P.D.; Shi, Y.; Morel, F.M. Structure and metal exchange in the cadmium carbonic anhydrase of marine diatoms. *Nature*. **2008**, *452*, 56-61.

[31] Mondal, M.; Khanra, S.; Tiwari, O.N.; Gayen, K.; Halder, G.N. Role of carbonic anhydrase on the way to biological carbon capture through microalgae—A mini review. *Environ. Prog. Sustain. Energy*. **2016**.

[32] So, A.K.C.; Espie, G.S.; Williams, E.B.; Shively, J.M.; Heinhorst, S.; Cannon, G.C. A novel evolutionary lineage of carbonic anhydrase (ϵ class) is a component of the carboxysome shell. *J. Bacteriol.* **2004**, *186*, 623-630.

[33] Sly, W.S.; Hu, P.Y. Human carbonic anhydrases and carbonic anhydrase deficiencies. *Annu. Rev. Biochem.* **1995**, *64*, 375-401.

[34] Maren, T.H. Carbonic anhydrase: chemistry, physiology, and inhibition. *Physiol. Rev.* **1967**, *47*, 595-782.

[35] Supuran, C.T. In: Carbonic anhydrase and modulation of physiologic and pathologic processes in the organism; Puscas, I. Ed., Helicon: Timisoara, 1994, 29-111.

[36] Waheed, A.; Zhu, X.L.; Sly, W.S. Membrane-associated carbonic anhydrase from rat lung. Purification, characterization, tissue distribution, and comparison with carbonic anhydrase IVs of other mammals. *J. Biol. Chem.* **1992**, *267*, 3308-11.

[37] Friedenwald, J.S. The formation of the intraocular fluid. *Am. J. Ophthalmol.* **1949**, *32*, 9-27.

[38] Supuran, C.T.; Mincione, F.; Scozzafava, A.; Briganti, F.; Mincione, G.; Ilies, M.A. Carbonic anhydrase inhibitors—part 52. Metal complexes of heterocyclic sulfonamides: a new class of strong topical intraocular pressure-lowering agents in rabbits. *Eur. J. Med. Chem.* **1998**, *33*, 247-254.

[39] Supuran, C.T.; Mincione, F.; Scozzafava, A.; Briganti, F.; Mincione, G.; Ilies, M.A. Carbonic anhydrase inhibitors—part 52. Metal complexes of heterocyclic sulfonamides: a new class of strong topical intraocular pressure-lowering agents in rabbits. *Eur. J. Med. Chem.* **1998**, *33*, 247-54.

[40] Supuran, C.T.; Scozzafava, A.; Ilies, M.A.; Iorga, B.; Cristea, T.; Briganti, F.; Chiraleu, F.; Banciu, M.D. Carbonic anhydrase inhibitors—Part 53. Synthesis of substituted-pyridinium derivatives of aromatic sulfonamides: The first non-polymeric membrane-impermeable inhibitors with selectivity for isozyme IV. *Eur. J. Med. Chem.* **1998**, *33*, 577-594.

[41] Breton, S. The cellular physiology of carbonic anhydrases. *JOP.* **2001**, *2*, 159-164.

[42] Maren, T.H.; Wynns, G.C.; Wistrand, P.J. Chemical properties of carbonic anhydrase IV, the membrane-bound enzyme. *Mol. Pharmacol.* **1993**, *44*, 901.

[43] Supuran, C.T.; Scozzafava, A.; Casini, A. Carbonic anhydrase inhibitors. *Med. Res. Rev.* **2003**, *23*, 146.

[44] Supuran, C.T.; Scozzafava, A. Applications of carbonic anhydrase inhibitors and activators in therapy. *Exp. Opin. Ther. Pat.* **2002**, *12*, 217.

[45] Maren, T.H. Relations between structure and biological activity of sulfonamides. *Annu. Rev. Pharmacol. Toxicol.* **1976**, *16*, 309-327.

[46] Supuran, C.T.; Scozzafava, A. Applications of carbonic anhydrase inhibitors and activators in therapy. *Expert Opin. Ther. Pat.* **2002**, *12*, 217-242.

[47] Supuran, C.T.; Scozzafava, A.; Casini, A. Carbonic anhydrase inhibitors. *Med. Res. Rev.* **2003**, *23*, 146-189.

[48] Liljas, A.; Håkansson, K.; Jonsson, B.H.; Xue, Y. Inhibition and catalysis of carbonic anhydrase. *Eur. J. Biochem.* **1994**, *219*, 1-10.

[49] Lindskog, S. Structure and mechanism of carbonic anhydrase. *Pharmacol. Ther.* **1997**, *74*, 1-20.

[50] Francis, S.A.; Taylor-Wells, J.; Gross, A.D.; Bloomquist, J.R. Toxicity and Physiological Actions of Carbonic Anhydrase Inhibitors to *Aedes aegypti* and *Drosophila melanogaster*. *Insects.* **2016**, *8*, 2.

[51] Vullo, D.; Innocenti, A.; Nishimori, I.; Pastorek, J.; Scozzafava, A.; Pastoreková, S.; Supuran, C.T. Carbonic anhydrase inhibitors. Inhibition of the transmembrane isozyme XII with sulfonamides-a new target for the design of antitumor and antiglaucoma drugs? *Bioorg. Med. Chem. Lett.* **2005**, *15*, 963-969.

[52] Durdagi, S.; Şentürk, M.; Ekinci, D.; Balaydın, H.T.; Göksu, S.; Küfrevioğlu, Ö.İ.; Innocenti, A.; Scozzafava, A.; Supuran, C.T. Kinetic and docking studies of phenol-based inhibitors of carbonic anhydrase isoforms I, II, IX and XII evidence a new binding mode within the enzyme active site. *Bioorg. Med. Chem.* **2011**, *19*, 1381-1389.

[53] Melagraki, G.; Afantitis, A.; Sarimveis, H.; Igglessi-Markopoulou, O.; Supuran, C.T. QSAR study on para-substituted aromatic sulfonamides as carbonic anhydrase II inhibitors using topological information indices. *Bioorg. Med. Chem.* **2006**, *14*, 1108-1114.

[54] Grüneberg, S.; Wendt, B.; Klebe, G. Subnanomolar Inhibitors from Computer Screening: A Model Study Using Human Carbonic Anhydrase II. *Angew. Chemie Int. Ed.* **2001**, *40*, 389-393.

[55] Raškevičius, V.; Kairys, V. Comparison of Performance of Docking, LIE, Metadynamics and QSAR in Predicting Binding Affinity of Benzenesulfonamides. *Curr. Comput. Aided Drug Des.* **2015**, *11*, 237-244.

[56] Supuran, C.T.; Scozzafava, A.; Conway, J. Carbonic anhydrase: its inhibitors and activators. CRC Press: London, New York, 2004, 149-182.

[57] Clare, B. W.; Supuran, C. T. A perspective on quantitative structure-activity relationships and carbonic anhydrase inhibitors. *Expert Opin. Drug Metab. Toxicol.* **2006**, *2*, 113-137.

[58] Lomelino, C.L.; Supuran, C.T.; McKenna, R. Non-classical inhibition of carbonic anhydrase. *Int. J. Mol. Sci.* **2016**, *17*, 1150.

[59] Krishnamurthy, V.M.; Kaufman, G.K.; Urbach, A.R.; Gitlin, I.; Gudiksen, K.L.; Weibel, D.B.; Whitesides, G.M. Carbonic anhydrase as a model for biophysical and physical-organic studies of proteins and protein-ligand binding. *Chem. Rev.* **2008**, *108*, 946-1051.

[60] Čapkauskaitė, E.; Zubrienė, A.; Baranauskienė, L.; Tamulaitienė, G.; Manakova, E.; Kairys, V.; Gražulis, S.; Tumkevičius, S.; Matulis, D. Design of [(2-pyrimidinylthio)acetyl]benzenesulfonamides as inhibitors of human carbonic anhydrases. *Eur. J. Med. Chem.* **2012**, *51*, 259-270.

[61] Morkūnaitė, V.; Baranauskienė, L.; Zubrienė, A.; Kairys, V.; Ivanova, J.; Trapencieris, P.; Matulis, D. Saccharin sulfonamides as inhibitors of carbonic anhydrases I, II, VII, XII, and XIII. *BioMed Res. Int.* **2014**, *2014*.

[62] Dudutienė, V.; Matulienė, J.; Smirnov, A.; Timm, D.D.; Zubrienė, A.; Baranauskienė, L.; Morkūnaitė, V.; Smirnovienė, J.; Michailovienė, V.; Juozapaitienė, V.; Mickevičiūtė, A.; Kazokaitė, J.; Bakšytė, S.; Kasiliauskaitė, A.; Jachno, J.; Revuckienė, J.; Kišonaitė, M.; Pilipuitytė, V.; Ivanauskaitė, E.; Milinavičiūtė, G.; Smirnovas, V.; Petrikaitė, V.; Kairys, V.; Petrauskas, V.; Norvaišas, P.; Lingė, D.; Gibieža, P.; Čapkauskaitė, E.; Zakšauskas, A.; Kazlauskas, E.; Manakova, E.; Gražulis, S.; Ladbury, J.E.; Matulis, D. Discovery and characterization of novel selective inhibitors of carbonic

anhydrase IX. *J. Med. Chem.* **2014**, *57*, 9435-9446.

[63] Alafeefy, A.M.; Ahmad, R.; Abdulla, M.; Eldehna, W.M.; Al-Tamimi, A.M.; Abdel-Aziz, H.A.; Al-Obaid, O.; Carta, F.; Al-Kahtani, A.A.; Supuran, C.T. Development of certain new 2-substituted-quinazolin-4-yl-aminobenzenesulfonamide as potential antitumor agents. *Eur. J. Med. Chem.* **2016**, *109*, 247-253.

[64] Boztaş, M.; Çetinkaya, Y.; Topal, M.; Gülçin, I.; Menzek, A.; Şahin, E.; Tanc, M.; Supuran, C.T. Synthesis and carbonic anhydrase isoenzymes I, II, IX, and XII inhibitory effects of dimethoxybromophenol derivatives incorporating cyclopropane moieties. *J. Med. Chem.* **2015**, *22*, 640–650.

[65] Thiry, A.; Supuran, C.T.; Masereel, B.; Dogné, J.M. Recent developments of carbonic anhydrase inhibitors as potential anticancer drugs. *J. Med. Chem.* **2008**, *12*, 3051-3056.

[66] Bedke, J.; Stenzl, A. Immunotherapeutic strategies for the treatment of renal cell carcinoma: where are we now? *Expert Rev. Anticancer Ther.* **2013**, *13*, 1399-1408.

[67] Taylor, R.D.; Jewsbury, P.J.; Essex, J.W. A review of protein-small molecule docking methods. *J. Comput. Aided Mol. Des.* **2002**, *16*, 151-166.

[68] Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287-2303.

[69] Warren, G.L.; Andrews, C.W.; Capelli, A.M.; Clarke, B.; LaLonde, J.; Lambert, M.H.; Lindvall, M.; Nevins, N.; Semus, S.F.; Senger, S.; Tedesco, G. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912-5931.

[70] Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins: Struct., Funct., Bioinf.* **2002**, *47*, 409-443.

[71] Yuriev, E.; Agostino, M.; Ramsland, P.A. Challenges and advances in computational docking: 2009 in review. *J. Mol. Recognit.* **2011**, *24*, 149-164.

[72] DesJarlais, R.L.; Sheridan, R.P.; Seibel, G.L.; Dixon, J.S.; Kuntz, I.D.; Venkataraghavan, R. Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. *J. Med. Chem.* **1988**, *31*, 722-729.

[73] Ewing, T.J.A.; Makino, S.; Skillman, A.G.; Kuntz, I.D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided Mol. Des.* **2001**, *15*, 411-428.

[74] Morris, G.M.; Goodsell, D.S.; Halliday, R.S.; Huey, R.; Hart, W.E.; Belew, R.K.; Olson, A.J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639-1662.

[75] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470-489.

[76] Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727-748.

[77] Kelley, B.P.; Brown, S.P.; Warren, G.L.; Muchmore, S.W. POSIT: Flexible Shape-Guided Docking For Pose Prediction. *J. Chem. Inf. Model.* **2015**, *55*, 1771-1780.

[78] McGaughey, G.B.; Sheridan, R.P.; Bayly, C.I.; Culberson, J.C.; Kreatsoulas, C.; Lindsley, S.; Maiorov, V.; Truchon, J.F.; Cornell, W.D. Comparison of topological, shape, and docking methods in virtual screening. *J. Chem. Inf. Model.* **2007**, *47*, 1504-1519.

[79] McGann, M. FRED pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **2011**, *51*, 578-596.

[80] McGann, M.R.; Almond, H.R.; Nicholls, A.; Grant, J.A.; Brown, F.K. Gaussian docking functions. *Biopolymers*. **2003**, *68*, 76–90.

[81] Allen, W.J.; Balius, T.E.; Mukherjee, S.; Brozell, S.R.; Moustakas, D.T.; Lang, P.T.; Case, D.A.; Kuntz, I.D.; Rizzo, R.C. DOCK 6: Impact of new features and current docking performance. *J. Comput. Chem.* **2015**, *36*, 1132-1156.

[82] McGann, M. FRED and HYBRID docking performance on standardized datasets. *J. Comp. Aid. Mol. Design*. **2012**, *26*, 897-906.

[83] Walters, W.P.; Stahl, M.T.; Murcko, M.A. Virtual screening—an overview. *Drug Discov. Today*. **1998**, *3*, 160-178.

[84] Kontoyianni, M.; McClellan, L.M.; Sokol, G.S. Evaluation of docking performance: comparative data on docking algorithms. *J. Med. Chem.* **2004**, *47*, 558-565.

[85] Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins: Struct., Funct., Bioinf.* **2004**, *57*, 225-242.

[86] Kontoyianni, M.; Sokol, G.S.; McClellan, L.M. Evaluation of library ranking efficacy in virtual screening. *J. Comput. Chem.* **2005**, *26*, 11-22.

[87] Gohlke, H.; Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem., Int. Ed.* **2002**, *41*, 2644-2676.

[88] Olsen, L.; Pettersson, I.; Hemmingsen, L.; Adolph, H.W.; Jørgensen, F.S. Docking and scoring of metallo-beta-lactamases inhibitors. *J. Comput. Aided Mol. Des.* **2004**, *18*, 287-302.

[89] Wang, R.; Lu, Y.; Fang, X.; Wang, S. An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein-ligand complexes. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2114-2125.

[90] Jug, G.; Anderluh, M.; Tomašič, T. Comparative evaluation of several docking tools for docking small molecule ligands to DC-SIGN. *J. Mol. Model.* **2015**, *21*, 164.

[91] Chen, Y.C. Beware of docking! *Trends Pharmacol. Sci.* **2015**, *36*, 78-95.

[92] Senger, S.; Convery, M.A.; Chan, C.; Watson, N.S. Arylsulfonamides: a study of the relationship between activity and conformational preferences for a series of factor Xa inhibitors. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5731-5735.

[93] Senger, S.; Chan, C.; Convery, M.A.; Hubbard, J.A.; Shah, G.P.; Watson, N.S.; Young, R.J. Sulfonamide-related conformational effects and their importance in structure-based design. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 2931-2934.

[94] Xiang, F.; Xiang, J.; Fang, Y.; Zhang, M.; Li, M. Discovering isozyme-selective inhibitor scaffolds of human carbonic anhydrases using structural alignment and de novo drug design approaches. *Chem. Biol. Drug Des.* **2014**, *83*, 247-258.

[95] Kuntz, I.D.; Blaney, J.M.; Oatley, S.J.; Langridge, R.; Ferrin, T.E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269-288.

[96] May, A.; Zacharias, M. Accounting for global protein deformability during protein-protein and protein-ligand docking. *Biochim. Biophys. Acta.* **2005**, *1754*, 225-231.

[97] Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727-748.

[98] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.*

1996, 261, 470-489.

[99] Wei, B.Q.; Weaver, L.H.; Ferrari, A.M.; Matthews, B.W.; Shoichet, B.K. Testing a flexible-receptor docking algorithm in a model binding site. *J. Mol. Biol.* **2004**, 337, 1161-1182.

[100] Matulis, D.; Capkauskaitė, E.; Zubriene, A.; Baranauskiene, L.; Tamulaitiene, G.; Manakova, E.; Kairys, V.; Grazulis, S.; Tumkevicius, S. Design of [(2-pyrimidinylthio)acetyl]benzenesulfonamides as inhibitors of human carbonic anhydrases. *Biophys. J.* **2013**, 104, 558a-9a.

[101] Pinard, M.A.; Boone, C.D.; Rife, B.D.; Supuran, C.T.; McKenna, R. Structural study of interaction between brinzolamide and dorzolamide inhibition of human carbonic anhydrases. *Bioorg. Med. Chem.* **2013**, 21, 7210-7215.

[102] Vullo, D.; Durante, M.; Di Leva, F.S.; Cosconati, S.; Masini, E.; Scozzafava, A.; Novellino, E.; Supuran, C.T.; Carta, F. Monothiocarbamates Strongly Inhibit Carbonic Anhydrases in Vitro and Possess Intraocular Pressure Lowering Activity in an Animal Model of Glaucoma. *J. Med. Chem.* **2016**.

[103] Durdagi, S.; Şentürk, M.; Ekinçi, D.; Balaydın, H.T.; Göksu, S.; Küfrevioğlu, Ö.İ.; Innocenti, A.; Scozzafava, A.; Supuran, C.T. Kinetic and docking studies of phenol-based inhibitors of carbonic anhydrase isoforms I, II, IX and XII evidence a new binding mode within the enzyme active site. *Bioorg. Med. Chem.* **2011**, 19, 1381-1389.

[104] Vicker, N.; Ho, Y.; Robinson, J.; Woo, L.L.; Purohit, A.; Reed, M.J.; Potter, B.V. Docking studies of sulphamate inhibitors of estrone sulphatase in human carbonic anhydrase II. *Bioorg. Med. Chem. Lett.* **2003**, 13, 863-865.

[105] Abdel-Hamid, M. K.; Abdel-Hafez, A. A.; El-Koussi, N. A.; Mahfouz, N. M.; Innocenti, A.; Supuran, C. T. Design, synthesis, and docking studies of new 1,3,4-thiadiazole-2-thione derivatives with carbonic anhydrase inhibitory activity. *Bioorg. Med. Chem.* **2007**, 15, 6975-6984.

[106] Aqvist, J.; Medina, C.; Samuelsson, J. A new method for predicting binding affinity in computer-aided drug design. *Protein. Eng.* **1994**, *7*, 385-391.

[107] Paulsen, M. D.; Ornstein, R. L. Binding free energy calculations for P450cam-substrate complexes. *Protein. Eng.* **1996**, *9*, 567-571.

[108] Jones–Hertzog, D. K.; Jorgensen, W. L. Binding affinities for sulfonamide inhibitors with human thrombin using Monte Carlo simulations with a linear response method. *J. Med. Chem.* **1997**, *40*, 1539-1549.

[109] Lamb, M. L.; Tirado–Rives, J.; Jorgensen, W. L. Estimation of the binding affinities of FKBP12 inhibitors using a linear response method. *Bioorg. Med. Chem.* **1999**, *7*, 851-860.

[110] Wall, I. D.; Leach, A. R.; Salt, D. W.; Ford, M. G.; Essex, J. W. Binding constants of neuraminidase inhibitors: an investigation of the linear interaction energy method. *J. Med. Chem.* **1999**, *42*, 5142-5152.

[111] Wang, W.; Wang, J.; Kollman, P. A. What determines the van der waals coefficient β in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations? *Proteins.* **1999**, *34*, 395-402.

[112] Chen, J.; Wang, R.; Taussig, M.; Houk, K. N. Quantitative Calculations of Antibody– Antigen Binding: Steroid– DB3 Binding Energies by the Linear Interaction Energy Method. *J. Org. Chem.* **2001**, *66*, 3021-3026.

[113] Kosugi, T.; Nakanishi, I.; Kitaura, K. Binding free energy calculations of adenosine deaminase inhibitor and the effect of methyl substitution in inhibitors. *J. Chem. Inf. Model.* **2009**, *49*, 615-622.

[114] Poongavanam, V.; Kongsted, J. Binding affinity models for Falcipain inhibition based on the Linear Interaction Energy method. *J. Mol. Graph. Model.* **2016**.

[115] Miranda, W.E.; Ngo, V.A.; Valiente, P.A.; Noskov, S.Y. Improved

QM/MM Linear-Interaction Energy Model for Substrate Recognition in Zinc-Containing Metalloenzymes. *J. Phys. Chem. B.* **2016**, *120*, 7824-7835.

[116] Perdih, A.; Wolber, G.; Solmajer, T. Molecular dynamics simulation and linear interaction energy study of D-Glu-based inhibitors of the MurD ligase. *J. Comput. Aided Mol. Des.* **2013**, *27*, 723-738.

[117] Kjellgren, E.R.; Glue, O.E.; Reinholdt, P.; Meyer, J.E.; Kongsted, J.; Poongavanam, V. A comparative study of binding affinities for 6,7-dimethoxy-4-pyrrolidylquinazolines as phosphodiesterase 10A inhibitors using the linear interaction energy method. *J. Mol. Graph. Model.* **2015**, *61*, 44-52.

[118] Mikulskis, P.; Genheden, S.; Rydberg, P.; Sandberg, L.; Olsen, L.; Ryde, U. Binding affinities in the SAMPL3 trypsin and host-guest blind tests estimated with the MM/PBSA and LIE methods. *J. Comput. Aided Mol. Des.* **2012**, *26*, 527-541.

[119] Wang, Y.; Harrison, C.B.; Schulten, K.; McCammon, J.A. Implementation of Accelerated Molecular Dynamics in NAMD. *Comput. Sci. Discov.* **2011**, *4*, 015002.

[120] Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 826-843.

[121] Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel metadynamics as accurate binding free-energy method. *Proc. Natl. Acad. Sci.* **2013**, *110*, 6358-6363.

[122] Wu, P.; Hu, X.; Yang, W. λ -Meta Dynamics Approach To Compute Absolute Solvation Free Energy. *J. Phys. Chem. Lett.* **2011**, *2*, 2099-2103.

[123] Hosek, P.; Toulcová, D.; Bortolato, A.; Spiwok, V. Altruistic metadynamics: Multisystem biased simulation. *J. Phys. Chem. B.* **2016**, *120*, 2209-2215.

[124] Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley*

Interdiscip. Rev. Comput. Mol. Sci. **2011**, *1*, 826-843.

[125] Bešker, N.; Gervasio, F.L. Using metadynamics and path collective variables to study ligand binding and induced conformational transitions. *Methods Mol. Biol.* **2012**, *819*, 501-513.

[126] Clark, A.J.; Tiwary, P.; Borrelli, K.; Feng, S.; Miller, E.B.; Abel, R.; Friesner, R.A.; Berne, B.J. Prediction of Protein-Ligand Binding Poses via a Combination of Induced Fit Docking and Metadynamics Simulations. *J. Chem. Theory Comput.* **2016**.

[127] Paloncýová, M.; Navrátilová, V.; Berka, K.; Laio, A.; Otyepka, M. Role of Enzyme Flexibility in Ligand Access and Egress to Active Site: Bias-Exchange Metadynamics Study of 1,3,7-Trimethyluric Acid in Cytochrome P450 3A4. *J. Chem. Theory Comput.* **2016**, *12*, 2101-2109.

[128] Cavalli, A.; Spitaleri, A.; Saladino, G.; Gervasio, F.L. Investigating drug-target association and dissociation mechanisms using metadynamics-based algorithms. *Acc. Chem. Res.* **2014**, *48*, 277-285.

[129] De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59*, 4035-4061.

[130] Dearden, J. C. The history and development of quantitative structure-activity relationships (QSARs). *IJQSPR*, **2016**, *1*, 1-44.

[131] Roy, K.; Kar, S.; Das, R. N. Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment. Academic press, 2015.

[132] Nantasenamat, C.; Isarankura-Na-Ayudhya, C.; Naenna, T.; Prachayasittikul, V. A practical overview of quantitative structure-activity relationship. *EXCLI J.* **2009**, *8*, 74-88.

[133] Allen, C.H.; Koutsoukas, A.; Cortés-Ciriano, I.; Murrell, D.S.; Malliavin, T.E.; Glen, R.C.; Bender, A. Improving the prediction of organism-

level toxicity through integration of chemical, protein target and cytotoxicity qHTS data. *Toxicol.Res.* **2016**, *5*, 883-94.

[134] Le, T.; Epa, V.C.; Burden, F.R.; Winkler, D.A. Quantitative structure–property relationship modeling of diverse materials properties. *Chem. Rev.* **2012**; *112*, 2889-2919.

[135] Cramer, R.D. The inevitable QSAR renaissance. *J. Comput. Aided. Mol. Des.* **2012**, *26*, 35-38.

[136] Verma, J; Khedkar, V.M.; Coutinho, E.C. 3D-QSAR in drug design-a review. *Curr. Top. Med. Chem.* **2010**, *10*, 95-115.

[137] Crum-Brown, A.; Fraser, T.R. On the connection between chemical constitution and physiological action. Part 1. On the physiological action of the ammonium bases, derived from Strychia, Brucia, Thebaia, Codeia, Morphia and Nicotia. *Trans. R. Soc. Edinburgh.* **1868**, *25*, 151-203.

[138] Richardson, B.J. Physiological research on alcohols. *Med. Times Gaz.* **1868**, *2*, 703-706.

[139] Mills, E.J. On melting point and boiling point as related to composition. *Philos. Mag.* **1884**, *17*, 173-187.

[140] Richet, C. On the relationship between the toxicity and the physical properties of substances. *Compt. Rendus Seances Soc. Biol.* **1893**, *9*, 775-776.

[141] Overton, E. Osmotic properties of cells in the bearing on toxicology and pharmacology. *Z. Physik. Chem.* **1897**, *22*, 189-209.

[142] Lifnick, R.L. Hans Horst Meyer and the lipid theory of narcosis. *Trends Pharmacol. Sci.* **1989**, *10*, 265-269.

[143] Hammett, L.P. Some relations between reaction rates and equilibrium constants. *Chem. Rev.* **1935**, *17*, 125-136.

[144] Hammett, L.P. The effect of structure upon the reactions of

organic compounds. Benzene derivatives. *J. Am. Chem. Soc.* **1937**, *59*, 96-103.

[145] Bell, P.H.; Roblin, R.O. Studies in Chemotherapy. VII. A Theory of the Relation of Structure to Activity of Sulfanilamide Type Compounds1. *J. Am. Chem. Soc.* **1942**, *64*, 2905-2917.

[146] Albert, A.; Goldacre, R.; Phillips, J. 455. The strength of heterocyclic bases. *J. Chem. Soc.* **1948**, 2240-2249.

[147] Taft, R.W. Polar and steric substituent constants for aliphatic and o-Benzoate groups from rates of esterification and hydrolysis of esters1. *J. Am. Chem. Soc.* **1952**, *74*, 3120-3128.

[148] Hansch, C.; Maloney, P.; Fujita, T.; Muir, R. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature.* **1962**, *194*, 178-180.

[149] Fujita T.; Iwasa J.; Hansch, C. A new substituent constant, π , derived from partition coefficients. *J. Am. Chem. Soc.* **1964**, *86*, 5175-5180.

[150] Free, S.J.; Wilson, J. A mathematical contribution to structure-activity studies. *J. Med. Chem.* **1964**, *7*, 395-399.

[151] Hansch, C. Quantitative approach to biochemical structure-activity relationships. *Acc. Chem. Res.* **1969**, *2*, 232-239.

[152] Fujita, T.; Ban, T. Structure-activity relation. 3. Structure-activity study of phenethylamines as substrates of biosynthetic enzymes of sympathetic transmitters. *J. Med. Chem.* **1971**, *14*, 148-152.

[153] Hansch, C.; Fujita, T. Additions and Corrections- ρ - σ - π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *J. Am. Chem. Soc.* **1964**, *86*, 1616-1626.

[154] Avram, S.; Milac, A.L.; Flonta, M.L. Computer-Aided Drug Design for Typical and Atypical Antipsychotic Drugs: A Review of Application of QSAR and Combinatorial Chemistry Methods-Tools for New Antipsychotics Design. *Curr. Comput. Aided Drug. Des.* **2005**, *1*, 347-364.

[155] Dudek, A.Z.; Arodz, T.; Galvez, J. Computational methods in developing quantitative structure-activity relationships (QSAR): a review. *Comb. Chem. High Throughput Screen.* **2006**, *9*, 213-228.

[156] Kar, S.; Das, R.N.; Roy, K.; Leszczynski, J. Can Toxicity for Different Species be Correlated?: The Concept and Emerging Applications of Interspecies Quantitative Structure-Toxicity Relationship (i-QSTR) Modeling. *IJQSPR.* **2016**, *1*, 23-51.

[157] Athersuch, T.J.; Wilson, I.D.; Keun, H.C.; Lindon, J.C. Development of quantitative structure-metabolism (QSMR) relationships for substituted anilines based on computational chemistry. *Xenobiotica.* **2013**, *43*, 792-802.

[158] Aalizadeh, R.; Thomaidis, N.S.; Bletsou, A.A.; Gago Ferrero, P. Quantitative Structure–Retention Relationship Models To Support Nontarget High-Resolution Mass Spectrometric Screening of Emerging Contaminants in Environmental Samples. *J. Chem. Inf. Model.* **2016**, *56*, 1384-1398.

[159] Watkins, M.; Sizochenko, N.; Rasulev, B.; Leszczynski, J. Estimation of melting points of large set of persistent organic pollutants utilizing QSPR approach. *J. Mol. Model.* **2016**, *22*, 1-4.

[160] González-Durruthy, M.; Werhli, A.V.; Cornetet, L.; Machado, K.S.; González-Díaz, H.; Wasiliesky, W.; Ruas, C.P.; Gelesky, M.A.; Monserrat, J.M. Predicting the binding properties of single walled carbon nanotubes (SWCNT) with an ADP/ATP mitochondrial carrier using molecular docking, chemoinformatics, and nano-QSBR perturbation theory. *RSC Adv.* **2016**, *6*, 58680-58693.

[161] Kubinyi, H. QSAR and 3D QSAR in drug design Part 2: applications and problems. *Drug Discov. Today.* **1997**, *2*, 538-546.

[162] Benfenati, E.; Pardoe, S.; Martin, T.; Gonella Diaza, R.; Lombardo, A.; Manganaro, A. Using toxicological evidence from QSAR models in practice. *Altex.* **2013**, *30*, 19-40.

[163] Bowen, J.P.; Zhong, H.A. Perspectives in Medicinal Chemistry: The Evolution of Medicinal Chemistry. *Curr. Top. Med. Chem.* **2016**, *16*, 897-898.

[164] Yang, S.P.; Song, S.T.; Tang, Z.M.; Song, H.F. Optimization of antisense drug design against conservative local motif in simulant secondary structures of HER-2 mRNA and QSAR analysis. *Acta Pharmacol. Sin.* **2003**, *24*, 897-902.

[165] Liu, S.S.; Liu, H.L.; Yin, C.S.; Wang, L.S. VSMP: a novel variable selection and modeling method based on the prediction. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 964-969.

[166] Soto, A.J.; Cecchini, R.L.; Vazquez, G.E.; Ponzoni, I. Multi-Objective Feature Selection in QSAR Using a Machine Learning Approach. *QSAR Comb. Sci.* **2009**, *28*, 1509-1523.

[167] Shahlaei, M.; Madadkar-Sobhani, A.; Saghaie, L.; Fassihi, A. Application of an expert system based on Genetic Algorithm–Adaptive Neuro-Fuzzy Inference System (GA–ANFIS) in QSAR of cathepsin K inhibitors. *Expert. Syst. Appl.* **2012**, *39*, 6182-6191.

[168] Teixeira, A.L.; Leal, J.P.; Falcao, A.O. Random forests for feature selection in QSPR Models-an application for predicting standard enthalpy of formation of hydrocarbons. *J. Cheminform.* **2013**, *5*, 9.

[169] Shahlaei, M. Descriptor selection methods in quantitative structure–activity relationship studies: a review study. *Chem. Rev.* **2013**, *113*, 8093-8103.

[170] Kubinyi, H. QSAR and 3D QSAR in drug design Part 1: methodology. *Drug Des.* **1997**, *2*, 457-467.

[171] Liu, R.; Jiang, W.; Walkey, C.D.; Chan, W.C.; Cohen, Y. Prediction of nanoparticles-cell association based on corona proteins and physicochemical properties. *Nanoscale.* **2015**, *7*, 9664-9675.

[172] Raevsky, O.A.; Liplavskaya, E.A.; Yarkov, A.V.; Raevskaya, O.E.; Worth, A.P. Linear and nonlinear QSAR models of acute intravenous toxicity of organic chemicals for mice. *Biochemistry (Moscow) Supplement Series B: Biomedical Chemistry*. **2011**, *5*, 213-225.

[173] Saihi, Y.; Kraim, K.; Ferkous, F.; Djeghaba, Z.; Azzou, A.; Benouis, S. Nonlinear QSAR Study of Xanthone and Curcuminoid Derivatives as α -Glucosidase Inhibitors. *Bull. Korean Chem. Soc.* **2013**, *34*, 1.

[174] Putz, M.V.; Lazea, M.; Putz, A.M.; Duda-Seiman C. Introducing catastrophe-QSAR. Application on modeling molecular mechanisms of pyridinone derivative-type HIV non-nucleoside reverse transcriptase inhibitors. *Int. J. Mol. Sci.* **2011**, *12*, 9533-9569.

[175] Wold, S. Validation of QSAR's. *Quant. Struct.-Act. Relat.* **1991**, *10*, 191-193.

[176] Mager, H.; Mager, P.P. Validation of QSARs: some reflections. *Quant. Struct.-Act. Relat.* **1992**, *11*, 518-521.

[177] Unger, S.H.; Hansch, C. Model building in structure-activity relations. Reexamination of adrenergic blocking activity of. beta.-halo-. beta.-arylalkylamines. *J. Med. Chem.* **1973**, *16*, 745-749.

[178] Tropsha, A. Best practices for QSAR model development, validation, and exploitation. *Mol. Inform.* **2010**, *29*, 476-488.

[179] Lubinski, L.; Urbaszek, P.; Gajewicz, A.; Cronin, M.T.; Enoch, S.J.; Madden, J.C.; Leszczynska, D.; Leszczynski, J.; Puzyn, T. Evaluation criteria for the quality of published experimental data on nanomaterials and their usefulness for QSAR modelling. *SAR QSAR Environ. Res.* **2013**, *24*, 995-1008.

[180] Ruusmann, V.; Sild, S.; Maran, U. QSAR DataBank - an approach for the digital organization and archiving of QSAR model information. *J. Cheminform.* **2014**, *6*, 1-17.

[181] Ruusmann, V.; Sild, S.; Maran, U. QSAR DataBank repository: open and linked qualitative and quantitative structure-activity relationship models. *J. Cheminform.* **2015**, *7*, 1-11.

[182] Rácz, A.; Bajusz, D.; Héberger, K. Consistency of QSAR models: Correct split of training and test sets, ranking of models and performance parameters. *SAR QSAR Environ. Res.* **2015**, *26*, 683-700.

[183] Burden, F.R.; Winkler, D.A. Robust QSAR models using Bayesian regularized neural networks. *J. Med. Chem.* **1999**, *42*, 3183–3187.

[184] Burden, F.R.; Winkler, D.A. Optimal sparse descriptor selection for QSAR using Bayesian methods. *QSAR Comb. Sci.* **2009**, *28*, 645–653.

[185] Burden, F.R.; Winkler, D.A. New QSAR methods applied to structure– activity mapping and combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 236-242.

[186] Golbraikh, A.; Tropsha, A. Beware of q^2 ! *J. Mol. Graph. Model.* **2002**, *20*, 269–276.

[187] Tropsha, A.; Gramatica, P.; Gombar, V.K. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci.* **2003**, *22*, 69–77.

[188] Hawkins, D.M.; Basak, S.C.; Mills, D. Assessing model fit by cross-validation. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 579-586.

[189] Gramatica, P.; Sangion, A. A Historical Excursus on the Statistical Validation Parameters for QSAR Models: A Clarification Concerning Metrics and Terminology. *J. Chem. Inf. Model.* **2016**, *56*, 1127-1131.

[190] Shao, J. Linear model selection by cross-validation. *J. Am. Stat. Assoc.* **1993**, *88*, 486-494.

[191] Baumann, D.; Baumann, K. Reliable estimation of prediction errors for QSAR models under model uncertainty using double cross-validation. *J. Cheminform.* **2014**, *6*, 47.

[192] Soeria-Atmadja, D.; Wallman, M.; Björklund, A.K.; Isaksson, A.; Hammerling, U.; Gustafsson, M.G. External cross-validation for unbiased evaluation of protein family detectors: Application to allergens. *Proteins*. **2005**, *61*, 918-925.

[193] Statnikov, A.; Wang, L.; Aliferis, C.F. A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. *BMC Bioinformatics*. **2008**, *9*, 1.

[194] Lise, S.; Buchan, D.; Pontil, M.; Jones, D.T. Predictions of hot spot residues at protein-protein interfaces using support vector machines. *PLoS ONE*. **2011**, *6*, e16774.

[195] Singh, K.P.; Gupta, S.; Basant, N. QSTR modeling for predicting aquatic toxicity of pharmacological active compounds in multiple test species for regulatory purpose. *Chemosphere*. **2015**, *120*, 680-689.

[196] Krstajic, D.; Buturovic, L.J.; Leahy, D.E.; Thomas, S. Cross-validation pitfalls when selecting and assessing regression and classification models. *J. Cheminform*. **2014**, *6*, 1-15.

[197] Mansourian, M.; Fassihi, A.; Saghaie, L.; Madadkar-Sobhani, A.; Mahnam, K.; Abbasi, M. QSAR and docking analysis of A2B adenosine receptor antagonists based on non-xanthine scaffold. *Med. Chem. Res*. **2015**, *24*, 394-407.

[198] Gütlein, M.; Karwath, A.; Kramer, S. CheS-Mapper 2.0 for visual validation of (Q) SAR models. *J. Cheminform*. **2014**, *6*, 41.

[199] Alexander, D.L.; Tropsha, A.; Winkler, D.A. Beware of R²: simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models. *J. Chem. Inf. Model*. **2015**, *55*, 1316-1322.

[200] James, G.; Witten, D.; Hastie, T.; Tibshirani, R. An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics), Springer-Verlag New York, 2013.

[201] Roy, K.; Ghosh, G. Exploring QSARs with Extended Topochemical Atom (ETA) indices for modeling chemical and drug toxicity. *Curr. Pharm. Des.* **2010**, *16*, 2625-2639.

[202] Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247-1250.

[203] Pearlman, D.A.; Charifson, P.S. Are free energy calculations useful in practice? A comparison with rapid scoring functions for the p38 MAP kinase protein system. *J. Med. Chem.* **2001**, *44*, 3417-3423.

[204] Zefirov, N.S.; Palyulin, V.A. QSAR for boiling points of “small” sulfides. Are the “high-quality structure-property-activity regressions” the real high quality QSAR models? *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1022-1027.

[205] Netzeva, T.I.; Worth, A.; Aldenberg, T.; Benigni, R.; Cronin, M.T.; Gramatica, P.; Jaworska, J.S.; Kahn, S.; Klopman, G.; Marchant, C.A.; Myatt, G.; Nikolova-Jeliaskova, N.; Patlewicz, G.Y.; Perkins, R.; Roberts, D.; Schultz, T.; Stanton, D.W.; van de Sandt, J.J.; Tong, W.; Veith, G.; Yang, C. Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. The report and recommendations of ECVAM Workshop 52. *ATLA*, **2005**, *33*, 155–173.

[206] Jaworska, J.; Comber, M.; Van Leeuwen, C.; Auer, C. Summary of a workshop on regulatory acceptance of (Q)SARs for human health and environmental endpoints. *Environ. Health Perspect.* **2003**, *111*, 1358-1360.

[207] Jaworska, J.; Nikolova-Jeliaskova, N.; Aldenberg, T. QSAR applicability domain estimation by projection of the training set descriptor space: a review. *ATLA*. **2005**, *33*, 445–459.

[208] Dearden, J.C.; Cronin, M.T.; Kaiser, K.L. How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR). *SAR QSAR Environ. Res.* **2009**, *20*, 241-266.

[209] Schultz, T.W.; Cronin, M.T.D.; Netzeva, T.I.; Aptula, A.O. Structure-toxicity relationships for aliphatic chemicals evaluated with *Tetrahymena pyriformis*. *Chem. Res. Toxicol.* **2002**, *15*, 1602-1609.

[210] Sahigara, F.; Mansouri, K.; Ballabio, D.; Mauri, A.; Consonni, V.; Todeschini, R. Comparison of different approaches to define the applicability domain of QSAR models. *Molecules.* **2012**, *17*, 4791-4810.

[211] Devillers, J.; Balaban, A.T.; Eds. *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon Breach Sci. Pub., Amsterdam, 1999.

[212] Karelson, M. *Molecular Descriptors in QSAR/QSPR*, Wiley-InterScience, New York, 2000.

[213] Todeschini, R.; Consonni, V. *Molecular Descriptors for Chemoinformatics (2 volumes)*, Wiley-VCH, Weinheim, 2009.

[214] Falivene, L.; Credendino, R.; Poater, A.; Petta, A.; Serra, L.; Oliva, R.; Scarano, V.; Cavallo, L. SambVca 2. A Web Tool for Analyzing Catalytic Pockets with Topographic Steric Maps. *Organometallics.* **2016**, *35*, 2286-2293..

[215] Raevsky, O.A. Physicochemical descriptors in property-based drug design. *Mini. Rev. Med. Chem.* **2004**, *4*, 1041-1052.

[216] Schuffenhauer, A.; Gillet, V. J.; Willett, P. Similarity searching in files of three-dimensional chemical structures: analysis of the BIOSTER database using two-dimensional fingerprints and molecular field descriptors. *J. Chem. Inf. Model.*, **2000**, *40*, 295-307.

[217] Gao, H. Application of BCUT metrics and genetic algorithm in binary QSAR analysis. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 402-407.

[218] Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors (Methods and Principles in Medicinal Chemistry, Volume 11*, Mannhold, R.; Kubinyi, H.; Timmerman, H.; Eds.), Wiley-VCH, Weinheim, 2000.

[219] Zanni, R.; Galvez-Llompart, M.; García-Domenech, R.; Galvez, J.

Latest advances in molecular topology applications for drug discovery. *Expert. Opin. Drug Discov.* **2015**, *10*, 945-957.

[220] Golbraikh, A.; Tropsha, A. QSAR modeling using chirality descriptors derived from molecular topology. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 144-154.

[221] Tetko, I. V. Computing chemistry on the web. *Drug Discov. Today* **2005**, *10*, 1497-500.

[222] Dong, J.; Cao, D.S.; Miao, H.Y.; Liu, S.; Deng, B.C.; Yun, Y.H.; Wang, N.N.; Lu, A.P.; Zeng, W.B.; Chen, A. ChemDes: an integrated web-based platform for molecular descriptor and fingerprint computation. *J. Cheminform.* **2015**, *7*, 60.

[223] Dong, J.; Liu, S.; Xu, X. B.; Zeng, W.; Cao, D. S. ChemDes: a free web-based tool for computing molecular descriptors. (<http://cpfd.cnki.com.cn/Article/CPFDTOTAL-YYWS201406001039.htm>) **2014**.

[224] Cramer, R.D.; Patterson, D.E.; Bunce, J.D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959-5967.

[225] Suh, M.E.; Park, S.Y.; Lee, H.J. Comparison of QSAR methods (CoMFA, CoMSIA, HQSAR) of anticancer 1-N-substituted imidazoquinoline-4, 9-dione derivatives. *Bull. Korean Chem. Soc.* **2002**, *23*, 417-422.

[226] Hopfinger, A.J.; Wang, S.; Tokarski, J.S.; Jin, B.; Albuquerque, M.; Madhav, P.J.; Duraiswami, C. Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *J. Am. Chem. Soc.* **1997**, *119*, 10509-10524.

[227] Andrade, C.H.; Pasqualoto, K.F.; Ferreira, E.I.; Hopfinger, A.J. 4D-QSAR: perspectives in drug design. *Molecules.* **2010**, *15*, 3281-3294.

[228] Verma, J.; Malde, A.; Khedkar, S.; Iyer, R.; Coutinho, E. Local Indices for Similarity Analysis (LISA) □ A 3D-QSAR Formalism Based on

Local Molecular Similarity. *J. Chem. Inf. Model.* **2009**, *49*, 2695-2707.

[229] Begam, B.F.; Kumar, J.S. Computer assisted QSAR/QSPR approaches—a review. *Indian J. Sci. Technol.* **2016**, *9*.

[230] Myint, K.Z.; Xie, X.Q. Recent advances in fragment-based QSAR and multi-dimensional QSAR methods. *Int. J. Mol. Sci.* **2010**, *11*, 3846-3866.

[231] Salum, L.; Andricopulo, A. Fragment-based QSAR: perspectives in drug design. *Mol. Divers.* **2009**, *13*, 277-285.

[232] Bradbury, S.P. Quantitative structure-activity relationships and ecological risk assessment: an overview of predictive aquatic toxicology research. *Toxicol. Lett.* **1995**, *79*, 229-237.

[233] Bradbury, S.P.; Russom, C.L.; Schmieder, P.K.; Schultz, T.W.; Diderich, R.; Auer, C.M. Advancing computational toxicology in a regulatory setting: A selected review of the accomplishments of Gilman D. Veith (1944–2013). *Appl. In Vitro Toxicol.* **2015**, *1*, 16-25.

[234] Hansch, C.; Telzer, B.R.; Zhang, L. Comparative QSAR in toxicology: examples from teratology and cancer chemotherapy of aniline mustards. *Crit. Rev. Toxicol.* **1995**, *25*, 67-89.

[235] Perkins, R.; Fang, H.; Tong, W.; Welsh, W. Quantitative structure-activity relationship methods: Perspectives on drug discovery and toxicology. *Environ. Toxicol. Chem.* **2003**, *22*, 1666-1679.

[236] Chen, J.Z.; Han, X.W.; Liu, Q.; Makriyannis, A.; Wang, J.; Xie, X.Q. 3D-QSAR studies of arylpyrazole antagonists of cannabinoid receptor subtypes CB1 and CB2. A combined NMR and CoMFA approach. *J. Med. Chem.* **2006**, *49*, 625-636.

[237] Melagraki, G.; Afantitis, A.; Sarimveis, H.; Igglessi-Markopoulou, O.; Supuran, C.T. QSAR study on para-substituted aromatic sulfonamides as carbonic anhydrase II inhibitors using topological information indices. *Bioorg. Med. Chem.* **2006**, *14*, 1108-1114.

[238] Supuran, C.T.; Clare, B.W. Carbonic anhydrase inhibitors–Part 57: Quantum chemical QSAR of a group of 1, 3, 4-thiadiazole-and 1, 3, 4-thiadiazoline disulfonamides with carbonic anhydrase inhibitory properties. *Eur. J. Med. Chem.* **1999**, *34*, 41-50.

[239] Clare, B.W.; Supuran, C.T. Carbonic anhydrase inhibitors. Part 61. Quantum chemical QSAR of a group of benzenedisulfonamides. *Eur. J. Med. Chem.* **1999**, *34*, 463-474.

[240] Thakur, A.; Thakur, M.; Khadikar, P.V.; Supuran, C.T.; Sudele, P. QSAR study on benzenesulphonamide carbonic anhydrase inhibitors: topological approach using Balaban index. *Bioorg. Med. Chem.* **2004**, *12*, 789-793.

[241] Renner, R. The KOW controversy. *Environ. Sci. Technol.* **2002**, *36*, 410A-413A.

[242] Maryanoff, B.E.; McComsey, D.F.; Costanzo, M.J.; Hochman, C.; Smith-Swintosky, V.; Shank, R.P. Comparison of sulfamate and sulfamide groups for the inhibition of carbonic anhydrase-II by using topiramate as a structural platform. *J. Med. Chem.* **2005**, *48*, 1941-1947.

[243] Alp, C.; Ekinci, D.; Gültekin, M.S.; Sentürk, M.; Sahin, E.; Küfrevioğlu, O.I. A novel and one-pot synthesis of new 1-tosyl pyrrol-2-one derivatives and analysis of carbonic anhydrase inhibitory potencies. *Bioorg. Med. Chem.* **2010**, *18*, 4468-4474.

[244] Cherkasov, A.; Muratov, E.N.; Fourches, D.; Varnek, A.; Baskin, I.I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y.C.; Todeschini, R.; Consonni, V.; Kuz'min, V.E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terfloth, L.; Gasteiger, J.; Richard, A.; Tropsha, A. QSAR modeling: where have you been? Where are you going to? *J. Med. Chem.* **2014**, *57*, 4977-5010.

[245] Leemans, E.; Mahasenani, K.V.; Kumarasiri, M.; Spink, E.; Ding, D.; O'Daniel, P.I.; Boudreau, M.A.; Lastochkin, E.; Testero, S.A.; Yamaguchi, T.; Lee, M. Three-dimensional QSAR analysis and design of new 1, 2, 4-

oxadiazole antibacterials. *Bioorg. Med. Chem. Lett.* **2015**, *26*, 1011-1015.

[246] Reunanen, J. Overfitting in making comparisons between variable selection methods. *J. Mach. Learn. Res.* **2003**, *3*, 1371-1382.

[247] Hawkins, D.M. The problem of overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1-12.

[248] Cawley, G.C.; Talbot, N.L.C. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.* **2010**, *11*, 2079-2107.

[249] Karelson, M.; Karelson, G.; Tamm, T.; Tulp, I.; Jänes, J.; Tämm, K.; Lomaka, A.; Savchenko, D.; Dobchev, D. QSAR study of pharmacological permeabilities. *ARKIVOC.* **2009**, *2*, 218-238.

[250] Topliss, J.G.; Costello, R.J. Chance correlations in structure-activity studies using multiple regression analysis. *J. Med. Chem.* **1972**, *15*, 1066-1068.

[251] Eriksson L.; Jaworska J.; Worth A.P.; Cronin M.T.D.; McDowell R.M.; Gramatica P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based QSARs. *Environ. Health Perspect.* **2003**, *111*, 1361-1375.

[252] Chirico, N.; Gramatica, P. Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. *J. Chem. Inf. Model.* **2011**, *51*, 2320-2335.

[253] Zucchini, W. An introduction to model selection. *J. Math. Psychol.* **2000**, *44*, 41-61.

[254] Raškevičius, V.; Kairys, V. Comparison of performance of docking, LIE, metadynamics and QSAR in predicting binding affinity of benzenesulfonamides. *Curr. Comput. Aided Drug Des.* **2015**, *11*, 237-244.

[255] Huang, J.; Fan, X. Why QSAR fails: an empirical evaluation

using conventional computational approach. *Mol. Pharm.* **2011**, *8*, 600-608.

[256] Sushko, Y.; Novotarskyi, S.; Körner, R.; Vogt, J.; Abdelaziz, A.; Tetko, I.V. Prediction-driven matched molecular pairs to interpret QSARs and aid the molecular optimization process. *J. Cheminform.* **2014**, *6*, 1-18.

[257] Maggiora, G.M. On outliers and activity cliffs--why QSAR often disappoints. *J. Chem. Inf. Model.* **2006**, *46*, 1535-1535.

[258] Srivastava, H. K.; Sastry, G. N. Efficient estimation of MMGBSA-based BEs for DNA and aromatic furan amidino derivatives. *J. Biomol. Struct. Dyn.* **2013**, *31*, 522-537.

[259] Srivastava, H. K.; Chourasia, M.; Kumar, D.; Sastry, G. N. Comparison of computational methods to model DNA minor groove binders. *J. Chem. Inf. Model.* **2011**, *51*, 558-571.

[260] Srivastava, H. K.; Sastry, G. N. Molecular dynamics investigation on a series of HIV protease inhibitors: assessing the performance of MM-PBSA and MM-GBSA approaches. *J. Chem. Inf. Model.* **2012**, *52*, 3088-3098.

[261] Greenidge, P. A.; Kramer, C.; Mozziconacci, J. C.; Wolf, R. M. MM/GBSA binding energy prediction on the PDBbind data set: successes, failures, and directions for further improvement. *J. Chem. Inf. Model.* **2012**, *53*, 201-209.

[262] Guimarães, C. R. A Direct Comparison of the MM-GB/SA Scoring Procedure and Free-Energy Perturbation Calculations Using Carbonic Anhydrase as a Test Case: Strengths and Pitfalls of Each Approach. *J. Chem. Theory Comput.*, **2011**, *7*, 2296-2306.

[263] Zhang, J. L.; Zheng, Q. C.; Zhang, H. X. Theoretical improvement of the specific inhibitor of human carbonic anhydrase VII. *Comp. Biol. Chem.*, **2011**, *35*, 50-56.

[264] Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935**, *3*, 300-313.

[265] Zwanzig, R. W. High-temperature equation of state by a perturbation method. I. nonpolar gases. *J. Chem. Phys.* **1954**, *22*, 1420-1426.

[266] Hanwell, M.D.; Curtis, D.E.; Lonie, D.C.; Vandermeersch, T.; Zurek, E.; Hutchison, G.R. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **2012**, *4*, 17.

[267] O'Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, 33.

[268] Tetko, I. V.; Gasteiger, J.; Todeschini, R.; Mauri, A.; Livingstone, D.; Ertl, P.; Palyulin, V.A.; Radchenko, E.V.; Zefirov, N.S.; Makarenko, A.S.; Tanchuk, V.Y.; Prokopenko, V. V. Virtual computational chemistry laboratory—design and description. *J. Comput. Aided Mol. Des.* **2005**, *19*, 453-463.

[269] ArgusLab, Mark A. Thompson, Planaria Software LLC, Seattle, <http://www.ArgusLab.com>.

[270] Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605-1612.

[271] Liu, J.; He, X.; Zhang, J. Z. Improving the scoring of protein–ligand binding affinity by including the effects of structural water and electronic polarization. *J. Chem. Inf. Model.* **2013**, *53*, 1306-1314.

[272] Berendsen, H.J.C.; van der Spoel, D.; van Drunen, R. GROMACS: a message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm.* **1995**, *91*, 43-56.

[273] Salomon-Ferrer, R.; Case, D.A.; Walker, R.C. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3*, 198-210.

[274] Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A.

Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157-1174.

[275] Hoops, S.C.; Anderson, K.W.; Merz Jr., K.M. Force field design for metalloproteins. *J. Am. Chem. Soc.* **1991**, *113*, 8262-8270.

[276] Peters, M.B.; Yang, Y.; Wang, B.; Füsti-Molnár, L.; Weaver, M.N.; Merz Jr., K.M. Structural survey of zinc-containing proteins and development of the zinc AMBER force field (ZAFF). *J. Chem. Theory Comput.* **2010**, *6*, 2935-2947.

[277] Lin, F.; Wang, R. Systematic derivation of AMBER force field parameters applicable to zinc-containing systems. *J. Chem. Theory Comput.* **2010**, *6*, 1852-1870.

[278] Nosé, S.; Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983**, *50*, 1055-1076.

[279] Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. phys.* **1981**, *52*, 7182-7190.

[280] Bussi, G.; Parrinello, M. Accurate sampling using Langevin dynamics. *Phys. Rev. E Stat. Nonlin. Soft Matter. Phys.* **2007**, *75*, 056707.

[281] Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. phys.* **1993**, *98*, 10089-10092.

[282] Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, R.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R.A.; Parrinello, M. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961-1972.

[283] Roy, K.; Das, R. N.; Ambure, P.; Aher, R. B. Be aware of error measures. Further studies on validation of predictive QSAR models. *Chemometr. Intell. Lab.* **2016**, *152*, 18-33.

[284] Nicolotti, O.; Gadaleta, D.; Mangiatordi, G. F.; Catto, M.; Carotti, A. Applicability domain for QSAR models: where theory meets reality. *IJQSPR*. **2016**, *1*, 45-63.

[285] Roy, K.; Kar, S.; Ambure, P. On a simple approach for determining applicability domain of QSAR models. *Chemometr. Intell. Lab.* **2015**, *145*, 22-29.

[286] Dixon, S.L.; Smondyrev, A.M.; Knoll, E.H.; Rao, S.N.; Shaw, D.E.; Friesner, R.A. PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J. Comput. Aided Mol. Des.* **2006**, *20*, 647-671.

[287] Dixon, S.L.; Smondyrev, A.M.; Rao S.N. PHASE: a novel approach to pharmacophore modeling and 3D database searching. *Chem. Biol. Drug. Des.* **2006**, *67*, 370-372.

[288] Čapkauskaitė, E.; Zubrienė, A.; Smirnov, A.; Torresan, J.; Kišonaitė, M.; Kazokaitė, J.; Gylytė, J.; Michailovienė, V.; Jogaitė, V.; Manakova, E.; Gražulis, S.; Tumkevičius, S.; Matulis, D. Benzenesulfonamides with pyrimidine moiety as inhibitors of human carbonic anhydrases I, II, VI, VII, XII, and XIII. *Bioorg. Med. Chem.* **2013**, *21*, 6937-6947.

[289] **Schrödinger Release 2015-3**: Maestro, version 10.3, Schrödinger, LLC, New York, NY, 2015.

[290] Kišonaitė, M.; Zubrienė, A.; Čapkauskaitė, E.; Smirnov, A.; Smirnovienė, J.; Kairys, V.; Michailovienė, V.; Manakova, E.; Gražulis, S.; Matulis, D. Intrinsic thermodynamics and structure correlation of benzenesulfonamides with a pyrimidine moiety binding to carbonic anhydrases I, II, VII, XII, and XIII. *PLOS ONE*. **2014**, *9*, e114106.

[291] Kanehisa, M.; Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids res.* **2000**, *28*, 27-30.

[292] Wishart, D.S.; Knox, C.; Guo, A.C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z. Woolsey, J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic acids res.* **2006**, *34*, D668-D672.

[293] Wishart, D.S.; Knox, C.; Guo, A.C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic acids res.* **2008**, *36*, D901-D906.

[294] Knox, C.; Law, V.; Jewison, T.; Liu, P.; Ly, S.; Frolkis, A.; Pon, A.; Banco, K.; Mak, C.; Neveu, V.; Djoumbou, Y. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic acids res.* **2011**, *39*, D1035-D1041.

[295] Law, V.; Knox, C.; Djoumbou, Y.; Jewison, T.; Guo, A.C.; Liu, Y.; Maciejewski, A.; Arndt, D.; Wilson, M.; Neveu, V.; Tang, A. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic acids res.* **2014**, *42*, D1091-D1097.

[296] Irwin, J.J.; Shoichet, B.K. ZINC--a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177-182.

[297] Willett, P.; Barnard, J.M.; Downs G. M. Chemical similarity searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983-996.

Disertacijos tema paskelbtų publikacijų sąrašas

Straipsniai ISI sąrašo žurnaluose:

1. Ieva Karpavičienė, Giedrė Valiulienė, Vytautas Raškevičius, Indrė Lebedytė, Algirdas Brukštus, Visvaldas Kairys, Rūta Navakauskienė, Inga Čikotienė. (2015) Synthesis and Antiproliferative Activity α -Branched α , β -Unsaturated Ketones in Human Hematological and Solid Cancer Cell Lines. *European journal medicinal chemistry*, **98**, 30-48.
2. Vytautas Raškevičius, Visvaldas Kairys. (2015) Comparison of performance of docking, LIE, metadynamics and QSAR in predicting binding affinity of benzenesulfonamides. *Current computer-aided drug design*, **11**, 237-244.
3. Vytautas Raškevičius, Visvaldas Kairys. (2017) Predicting isoform-specific binding selectivities of benzenesulfonamides using QSAR and 3D-QSAR. *Current computer-aided drug design*, **13**, 75-83.

Tarptautinių konferencijų pranešimų medžiaga:

1. "Comparison performance docking, LIE, metadynamics and QSAR in predicting binding affinity benzenesulfonamides." Coins Conference Natural and Life Sciences (Vilnius, Lietuva / kovo 3–7, **2015**).
2. "QSAR studies α -Branched α,β -Unsaturated Ketones Antiproliferative Activity in Human Hematological and Solid Cancer Cell Lines" 10, European Conference on Computational Chemistry – EuCO–CC 2015 (Fulda, Vokietija / rugpjūčio 31 – rugsėjo 3, **2015**).
3. „Lead compound similar to metabolite discovery using Tanimoto score“ Vita Scientia 2016 (Vilnius, Lietuva / sausio 4, **2016**).

Nacionalinių konferencijų pranešimų medžiaga:

1. „Benzensulfonamidų prisijungimo selektyvumo prie karboanhidrazės izoformų prognozavimas taikant QSAR ir 3D–QSAR metodus“ Fizinių ir technologijos mokslų tarpdalykiniai tyrimai (Vilnius, Lietuva / vasario 10, **2016**).

Priedas

Lentelė A. QSAR deskriptorių, naudotų α -pakeistų, α,β -nesočiųjų ketonų, QSAR modeliuose pavadinimai, tipai ir trumpi aprašymai.

Molekulės deskriptorius	Tipas	Aprašymas
CSI	Topologiniai aprašai	ekscentriško ryšio indeksas
UNIP		vienpoliškumas
MPC10	Kelio ir ėjimo skaičiavimas	molekulinis kelio skaičiavimas 10 eilės
X2	Jungiamumo indeksai	jungiamumo indeksas χ_2
X3Av		vidutinio valentiškumo jungiamumo indeksas χ_3
ATS6e	Dviamtės autokoreliacijos	Broto–Moreau topologinės struktūros autokoreliacijos – inertiškumas 6 / priderintas pagal atomų Sanderson elektroneigiamumus
MATS1m		Moran autokoreliacijos – inertiškumas 1 / priderintas pagal atomų mases
MATS2m		Moran autokoreliacijos – inertiškumas 2 / priderintas pagal atomų mases
MATS8m		Moran autokoreliacijos – inertiškumas 8 / priderintas pagal atomų mases
MATS2e		Moran autokoreliacijos – inertiškumas 2 / priderintas pagal atomų Sanderson elektroneigiamumus
MATS5p		Moran autokoreliacijos – inertiškumas 5 / priderintas pagal atomų poliarizuojamumą
GATS1v		Geary autokoreliacijos – inertiškumas 1 /

		priderintas pagal atomų van der Valso tūrius
HOMA	Geometriniai aprašai	Harmoninio osciliatoriaus modelio aromatiškumo indeksas
RDF050u	RDF aprašai	Radialinio pasiskirstymo funkcija – 5.0 / nepriderintas
RDF035p		Radialinio pasiskirstymo funkcija – 3.5 / priderintas pagal atomų poliarizuojamumą
Mor16u	3D–MoRSE deskriptoriai	3D–MoRSE – signalas 16 / nepriderintas
Mor20m		3D–MoRSE – signalas 20 / priderintas pagal atomų mases
Mor05v		3D–MoRSE – signalas 05 / priderintas pagal atomų van der Valso tūrius
Mor11p		3D–MoRSE – signalas 15 / priderintas pagal atomų poliarizuojamumą
Mor20p		3D–MoRSE – signalas 20 / priderintas pagal atomų poliarizuojamumą
G3u	WHIM deskriptoriai	Trečio komponento simetrijos kryptinis WHIM indeksas / nepriderintas
E2u		Antro komponento prieinamumo kryptinis WHIM indeksas / nepriderintas
E3u		3rd komponento prieinamumo kryptinis WHIM indeksas / nepriderintas
E2e		Antro komponento prieinamumo kryptinis WHIM indeksas / priderintas pagal atomų Sanderson elektroneigiamumus
G3p		Trečio komponento simetrijos kryptinis WHIM indeksas / priderintas pagal atomų poliarizuojamumą
G2s		Antro komponento simetrijos kryptinis WHIM indeksas / priderintas pagal atomų elektrotopologines būsenas

H7u	GETAWAY deskriptoriai	H autokoreliacijos inertiškumas 7 / nepriderintas
HATS2u		Svertu pasvertas autokoreliacijos inertiškumas 2 / nepriderintas
HTv		H bendras indeksas / priderintas pagal atomų van der Valso tūrius
R4v		R autokoreliacijos inertiškumas 4 / priderintas pagal atomų van der Valso tūrius

Lentelė B. Molekulinių deskriptorių, naudotų QSAR modeliuose selektyvumams, pavadinimai ir trumpi aprašymai.

Molekulės deskriptorius	Aprašymas	Kiek kartų panaudotas
BELe5	žemiausia tikrinė vertė n. 5 Burden matricos / priderinta pagal atominius Sandersono elektroneigiamumus	1
DISPm	perkėlimo vertė / priderinta pagal masę	1
EPS0	kraštų jungiamumo indeksas nulinės eilės	1
G(N..S)	suma geometrinių atstumų tarp N ir S	1
G1e	pirmo komponento simetrijos kryptinis WHIM indeksas / priderintas pagal Sandersono elektroneigiamumą	1
GATS3v	Geary autokoreliacijos inertiškumas 3 priderintas pagal van der Valso tūrį	1
GATS5m	Geary autokoreliacijos inertiškumas 5 priderintas pagal masę	1
GATS7e	Geary autokoreliacijos inertiškumas 7 priderintas pagal Sandersono elektroneigiamumą	1
GATS8m	Geary autokoreliacijos inertiškumas 8 priderintas pagal masę	1
GATS8v	Geary autokoreliacijos inertiškumas 8 priderintas pagal van der Valso tūrį	1
H2m	H autokoreliacijos inertiškumas 2 / priderintas pagal masę	1
H4p	H autokoreliacijos inertiškumas 4 / priderintas pagal poliarizuojamumą	1

H5u	H autokoreliacijos inetiškumas 5 / nepriderintas	1
H6e	H autokoreliacijos inetiškumas 6 / priderintas pagal Sandersono elektroneigiamumą	2
HATS8m	Svertu pasvertas autokoreliacijos inetiškumas 8 / priderintas pagal masę	1
HATS8p	Svertu pasvertas autokoreliacijos inetiškumas 8 / priderintas pagal poliarizuojamumą	1
JGI2	vidutinis topologinių krūvių indeksas antro eilės	1
MATS1p	Moran autokoreliacijos inetiškumas 1 priderintas pagal poliarizuojamumą	1
MATS2v	Moran autokoreliacijos inetiškumas 2 priderintas pagal van der Valso tūrį	1
MATS4p	Moran autokoreliacijos inetiškumas 4 priderintas pagal poliarizuojamumą	1
MATS6m	Moran autokoreliacijos inetiškumas 6 priderintas pagal masę	1
MATS7m	Moran autokoreliacijos inetiškumas 7 priderintas pagal masę	2
MATS7v	Moran autokoreliacijos inetiškumas 7 priderintas pagal van der Valso tūrį	1
MATS8e	Moran autokoreliacijos inetiškumas 8 priderintas pagal Sandersono elektroneigiamumą	1
MATS8p	Moran autokoreliacijos inetiškumas 8 priderintas pagal poliarizuojamumą	2
Mor04m	3D-MoRSE - signalas 04 / priderintas pagal masę	1
Mor19v	3D-MoRSE - signalas 19 / priderintas pagal van der Valso tūrį	1
Mor20v	3D-MoRSE - signalas 20 / priderintas pagal van der Valso tūrį	1
PCR	santykis tarp kartotinių kelių skaičiaus ir kelių skaičiaus	1
R4e	R autokoreliacijos inetiškumas 4 / priderintas pagal Sandersono elektroneigiamumą	1
R7m+	R maksimalios autokoreliacijos inetiškumas 7 / priderintas pagal masę	2
R7p+	R maksimalios autokoreliacijos inetiškumas 7 / priderintas pagal poliarizuojamumą	1
RDF055m	radialinio pasiskirstymo funkcija - 055 / priderinta pagal masę	1

T(N..N)	suma topologinių atstumų tarp N ir N	1
T(O..Cl)	suma topologinių atstumų tarp O ir Cl	1
T(S..Cl)	suma topologinių atstumų tarp S ir Cl	1
T(S..S)	suma topologinių atstumų tarp S ir S	1

Lentelė C. QSAR deskriptorių naudotų benzensulfonamidų QSAR modeliuose pavadinimai, tipai ir trumpi aprašymai.

Molekulės deskriptorius	Tipas	Aprašymas
VAR	Topologiniai aprašai	Variacija
MATS7m	Dvimatės autokoreliacijos	Moran autokoreliacijos – inertiškumas 7 / priderintas pagal atomų mases
MATS8p		Moran autokoreliacijos – inertiškumas 8 / priderintas pagal atomų poliarizuojamumą
G(N..S)	3D atomų poros	geometrinių atstumų tarp N ir S atomų suma
H5u	GETAWAY deskriptoriai	H autokoreliacija 5 / nepriderinta
HATS8m		Svertu pasvertas autokoreliacijos inertiškumas 8 / priderintas pagal masę
R7p+		R maksimali autokoreliacija 7 / priderinta pagal atomų poliarizuojamumą

Lentelė D. Statistinių įverčių naudotų QSAR modeliuose pavadinimai, trumpi aprašymai ir nuorodos į formules.

F	F testo vertė, kai nulinė hipotezė, kad koeficientai lygūs nuliui
-----	--

MAE	(angl. mean absolute error) žr. 1.5.5-8
p	Tikimybė, kad koeficientai lygūs nuliui
Pearson-r	Pirsono koreliacijos koeficientas žr. 1.5.5-4
$Q^2 (Q^2_{LOO})$	R^2 paskaičiuotas kryžminiu patikrinimu išmetant vieną žr. 1.5.4
$Q^2_{TEST} (R^2 CV)$	R^2 paskaičiuotas kryžminiu patikrinimu išmetant vieną su testavimo duomenų rinkiniu žr. 1.5.4
R^2	Determinacijos koeficientas žr. 1.5.5-1
$R^2_{scrambled}$	Vidutinė determinacijos koeficiento vertė iš serijos modelių, sukurtų, naudojant sumaišytus biologinius aktyvumus apmokymo rinkinio junginiams
R^2_{ADJ}	Redukuotas determinacijos koeficientas žr. 1.5.5-6
R^2_{TEST}	Determinacijos koeficientas paskaičiuotas su su testavimo duomenų rinkiniu.
RMSE	(angl. root mean square error) žr. 1.5.5-7
SD	Regresijos standartinis nuokrypis
<i>stability</i>	Modelių prognozių stabilumas keičiant apmokymo rinkinio sudėtį