

VILNIAUS UNIVERSITETAS
CHEMIJOS INSTITUTAS

Andrius Sazonovas

**ORGANINIŲ MEDŽIAGŲ ŪMAUS TOKSIŠKUMO IR
METABOLIZMO VIETOS MOLEKULĖJE
PROGNOZAVIMAS TAIKANT GALAS METODĄ**

Daktaro disertacija
Fiziniai mokslai, chemija (03 P)

Vilnius, 2010

Disertacija parengta 2004-2009 metais Vilniaus universitete.

Mokslinis vadovas:

prof. habil. dr. Eugenijus Butkus (Vilniaus universitetas, fiziniai mokslai, chemija – 03 P)

TURINYS

SANTRUMPOS	5
IVADAS	9
1 LITERATŪROS APŽVALGA	15
1.1 <i>Ūmaus toksiškumo objektas</i>	15
1.1.1 Ūmaus toksiškumo (LD_{50}) modeliavimai	15
1.1.2 Eksperimentinių LD_{50} matavimų kokybė	21
1.2 <i>Citochromų P450 šeimos fermentų apžvalga</i>	25
1.2.1 Žmogaus citochromai P450 ir jų katalizuojamos reakcijos	25
1.2.2 Metabolizmo fermentų regioselektyvumo modeliai	30
1.3 <i>Modelio pritaikomumo sritis (erdvė)</i>	36
2 METODINĖ DALIS	41
2.1 <i>Duomenys ir programinė įranga</i>	41
2.1.1 Toksiškumo (LD_{50}) duomenų bazė	41
2.1.2 Metabolizmo regioselektyvumo duomenys	44
2.1.3 Programinė įranga.....	46
2.2 <i>Deskriptorių generavimas</i>	47
2.2.1 Fragmentiniai deskriptoriai ūmaus toksiškumo (LD_{50}) modeliavimui	47
2.2.2 Sluoksninė atomo-centrinė fragmentacija	50
2.3 <i>Statistiniai metodai</i>	55
2.3.1 GALAS modeliavimo metodika	55
2.3.2 PLS ir BPLS metodai bei bazinis QSAR modelis	56
2.3.3 Lokalus modelis ir bazinių prognozių korekcijos.....	60
2.3.4 Modelio pritaikomumo srities įvertinimas.....	61
3 REZULTATAI IR JŲ APTARIMAS	64
3.1 <i>Ūmaus toksiškumo (LD_{50}) modeliavimų rezultatai</i>	64

3.1.1	Bazinių (globalių) ir galutinių (lokalių) modelių palyginimas	64
3.1.2	<i>RMSE</i> priklausomybė nuo <i>RI</i>	67
3.1.3	Išorinė modelio patikra	72
3.1.4	Prognozių klasifikacija pagal patikimumą.....	76
3.1.5	Apskaičiuotų <i>RI</i> reikšmių pasiskirstymo dėsningumai įvairiuose duomenų rinkiniuose.....	78
3.2	<i>Citochromų P450 šeimos metabolizmo fermentų (CYP3A4 ir CYP2D6) regioselektyvumo modeliavimo rezultatai</i>	81
3.2.1	Bazinių ir lokaliai pakoreguotų GALAS modelio prognozių palyginimas	81
3.2.2	Metabolizmo regioselektyvumo modelių pritaikomumo sritis.....	85
3.2.3	Regioselektyvumo prognozavimo rezultatai likusiems metabolizmo reakcijų tipams	88
3.2.4	CYP3A4 ir CYP2D6 regioselektyvumo modelių apmokymo naujais junginiais bandymas	93
	IŠVADOS	100
	A. SAZONOVO STRAIPSNŲ SĄRAŠAS	102
	LITERATŪROS SĄRAŠAS	103
	PRIEDAS NR.1	120
	PRIEDAS NR.2	128

SANTRUMPOS

ADME (angl. *Absorption, Distribution, Metabolism, Excretion*) – absorbcija, pasiskirstymas, metabolizmas, pašalinimas;

ANN (angl. *Artificial Neural Network*) – dirbtinis neuroninis tinklas;

BPLS (angl. *Binomial Partial Least Squares*) – binominis dalinis mažiausių kvadratų metodas;

CYP2D6 – citochromas P450 2D6, viena iš CYP450 fermentų didšeimės izoformų (citochromas P450, antra šeima, D pošeimis, šeštas polipeptidas);

CYP3A4 – citochromas P450 3A4, viena iš CYP450 fermentų didšeimės izoformų (citochromas P450, trečia šeima, A pošeimis, ketvirtas polipeptidas);

CYP450 (angl. *Cytochrome P450*) – citochromai P450, hemą turintys, oksidoreduktazių klasei priklausantys fermentai, plačiai dalyvaujantys žmogaus organizmo metabolizmo procesuose;

DMCI (angl. *Data-Model Consistency Index*) – duomenų ir modelio neprieštaringumo (nuoseklumo) indeksas;

EC₅₀ (angl. *Effective Concentration Fifty*) – efektyvi cheminio junginio koncentracija sukelianti 50% maksimalaus atsako sistemoje;

ECVAM (angl. *European Centre for the Validation of Alternative Methods*) – Europos alternatyviųjų metodų vertinimo (patikros) centras;

ESIS (angl. *European Chemical Substances Information System*) – Europinė informacijos apie chemines medžiagas sistema;

GABA (angl. *gamma-Aminobutyric acid*) – 4-aminobutano rūgštis;

GALAS (angl. *Global, Adjusted Locally According to Similarity*) – globalus(i), lokaliai pakoreguotas(a) atsižvelgiant į panašumą (kalbant apie modelį ar metodiką);

HPVC (angl. *High Production Volume Chemicals*) – didelė apimtimi gaminamos cheminės medžiagos;

in silico – procesas atliekamas pasitelkiant kompiuterio ir kompiuterinių modeliavimo metodų pagalbą;

in vitro – eksperimentas atliekamas dirbtinės izoliacijos sąlygomis;

in vivo – procesas vykstantis gyvame organizme;

IP (angl. *Intraperitoneal*) – intraperitoninis cheminės medžiagos įvedimo į organizmą būdas (medžiaga įvedama į pilvaplėvės gaubiamą organizmo ertmę);

IUCLID (angl. *International Uniform Chemical Information Database*) – tarptautinė vieningoji cheminės informacijos duomenų bazė;

IV (angl. *Intravenous*) – intraveninis cheminės medžiagos įvedimo į organizmą būdas;

LD_{50} (angl. *Lethal Dose Fifty*) – medžiagos dozė, 24 valandų bėgyje sukelti ūmią mirtį penkiasdešimčiai procentų tiriamosios gyvūnų imties individų;

$\log D_{O/W}$ – visų (įskaitant jonizuotas) junginio formų pasiskirstymo tarp vandens ir *n*-oktanolio koeficientas;

$\log P_{O/W}$ – neutralios junginio formos pasiskirstymo tarp vandens ir *n*-oktanolio koeficientas;

$\log S_w$ – junginio tirpumas vandenyje, išreiškiamas kaip sočiosios koncentracijos S_w (mol/l) logaritmas;

LOO (angl. *Leave-One-Out*) – statistinė kryžminės patikros metodika, kuomet kiekvienos iteracijos metu yra atmetamas vienas stebėjimas, kuriam mėginama prognozuoti savybės vertę naudojantis iš likusių duomenų sukurtu modeliu. Iteracijų skaičius sutampa su stebėjimų skaičiumi duomenų rinkinyje (t.y. kiekvienas stebėjimas yra išmetamas vieną kartą);

LMO (angl. *Leave-Many-Out*) – LOO metodo atmaina, kuomet kiekvienos iteracijos metu yra atmetama iš karto keletas atsitiktinai parinktų stebėjimų. Iteracijų skaičius gali būti parenkamas taip, kad atitiktų pageidaujamą tikimybę, jog kiekvienas konkretus stebėjimas bus atmestas bent vieną kartą;

MAE (angl. *Mean Absolute Error*) – vidutinė absoliutinė paklaida;

MPTP – 1-metil-4-fenil-1,2,3,6-tetrahidropiridinas;

N – junginių skaičius;

$\text{NADP}^+/\text{NADPH}$ (angl. *Nicotinamide Adenine Dinucleotide Phosphate*) – nikotinamido adenino dinukleotido fosfatas, oksidacijos-redukcijos procesuose dalyvaujantis kofermentas; NADP^+ ir NADPH žymi atitinkamai oksiduotą ir redukuotą formas;

NAPQI – N-acetil-*p*-benzochinono iminas;

OECD (angl. *Organization for Economic Cooperation and Development*) – ekonominio bendradarbiavimo ir vystymosi organizacija;

OPS (angl. *Optimal Prediction Space*) – optimalios prognozės erdvė;

OR (angl. *Oral*) – oralinis cheminės medžiagos įvedimo į organizmą būdas;

P-gp (angl. *P-glycoprotein*) – MDR1 membraninis baltymas, atsakingas už ląstelei svetimų cheminių medžiagų pašalinimą;

PLS (angl. *Partial Least Squares/Projection to Latent Structures*) – dalinis mažiausių kvadratų metodas;

QSAR (angl. *Quantitative Structure – Activity Relationship*) – kiekybinis struktūros – aktyvumo sąryšis;

QSPR (angl. *Quantitative Structure – Property Relationship*) – kiekybinis struktūros – savybės sąryšis;

R – koreliacijos koeficientas;

REACH (angl. *Registration, Evaluation and Authorization of Chemicals*) – cheminių junginių registracija, įvertinimas ir sankcionavimas;

RI (angl. *Reliability Index*) – prognozės patikimumo indeksas;

$RMSE$ (angl. *Root Mean Square Error*) – šaknis iš vidutinės kvadratinės paklaidos;

RTECS[®] (angl. *Registry of Toxic Effects of Chemical Substances*) – cheminių medžiagų sukeltamų toksinių efektų registras;

SAR (angl. *Structure Activity Relationship*) – struktūros – aktyvumo sąryšis;

SC (angl. *Subcutaneous*) – poodinis cheminės medžiagos įvedimo į organizmą būdas;

sd (angl. *Standard Deviation*) – standartinis nuokrypis;

SI (angl. *Similarity Index*) – molekulių panašumo indeksas;

T_e (angl. *Excess Toxicity*) – perteklinis toksiškumas;

TOPKAT[®] (*TOxicity Prediction by Komputer Assisted Technology*) – kompiuterinė programa įvairiems cheminių junginių toksiniams poveikiams prognozuoti;

WDI (angl. *World Drug Index*) – pasaulinis vaistų indeksas.

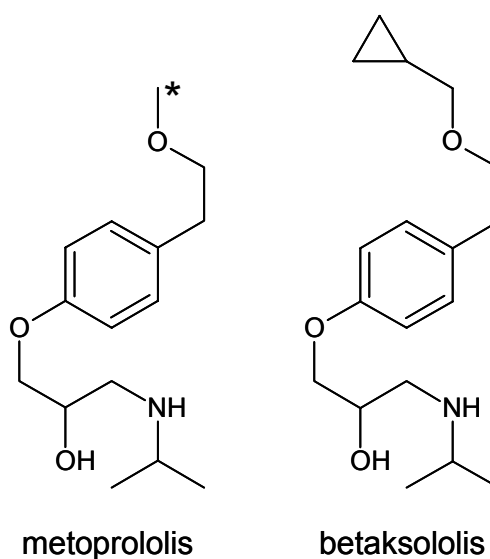
ĮVADAS

Tarp visų pastaruoju metu farmacijos pramonėje stebimų tendencijų, keletas kelia išskirtinai didelį šia veikla užsiimančių kompanijų susirūpinimą. Nuolatos auga naujų vaistinių junginių paieškos bei įdiegimo rinkoje kaštai, daugėja vėlyvų nesėkmių, kuomet kandidato į vaistinius junginius neatitikimas vieniems ar kitiems reikalavimams paaiškėja tik paskutiniuosiose tyrimo ir patvirtinimo proceso pakopose. Pastarieji atvejai yra ypač skaudūs, kadangi prarandamos visos įdėtos pastangos ir investicijos. Tokioje situacijoje, augant spaudimui sutrumpinti ir atpiginti naujo vaisto sukūrimo procesą, atsiranda poreikis ieškoti racionalių sprendimų įgalinančių iš esmės optimizuoti tinkamų junginių paieškos eigą. Kitais žodžiais tariant, siekiama sukurti naujus ar patobulinti jau esamus įrankius ir metodus, leidžiančius kiek įmanoma anksčiau pasirinkti molekules, kurios galų gale sėkmingai praeis visas vaisto kūrimo ir bandymų pakopas [1, 2]. Būtent tokie iššūkiai lėmė kompiuterinių, vadinamųjų *in silico*, metodų atsiradimą, leidžiančių teoriškai prognozuoti pačias įvairiausias medžiagų savybes ir biologinius aktyvumus dar virtualios bibliotekos etape, t.y. net neturint laboratorijoje susintetintų junginių.

Vienas pagrindinių iššūkių naujų vaistinių medžiagų paieškoje yra priimtinių kandidato farmakokinetinių savybių užtikrinimas. Būtent junginio pasiskirstymą organizme nulemiančių kinetinių veiksnių kombinacija su specifiniu farmakologiniu jo veikimo mechanizmu ir nulemia būsimojo vaisto efektyvumą [3]. Šio proceso esmę sudaro vienalaikis visos eilės skirtingų parametrų optimizavimas siekiant geriausios pusiausvyros tarp junginio fiziko-cheminių savybių ir poveikio terapiniam taikiniui stiprumo. Vienas iš tokio optimizavimo aspektų yra pasiekti pageidaujamą junginio metabolinį stabilumą, užtikrinantį, kad junginys pasieks savo taikinį organizme chemiškai nepakitęs. Savo ruožtu, vienas pagrindinių pastarąjį faktą nulemiančių veiksnių yra CYP450 fermentų sistema katalizuojanti absoliučią daugumą organizmui svetimų vaistinių junginių cheminių virsmų. Žvelgiant iš tyrėjo, užsiimančio naujų vaistų paieška, pozicijos, faktas, jog nagrinėjamą kandidatą

metabolizuoja viena iš daugelio citochromų P450 izoformų reiškia visą eilę galimų problemų: nesusidaranti ar labai trumpam laikui susidaranti pakankama junginio koncentracija ties taikiniu, vaisto sąveika su kitais kartu vartojamais tos pačios CYP450 izoformos substratais ar inhibitoriais ir pan. Taigi kiek įmanoma ankstyvesnė informacija apie junginio metabolizmą yra vienas iš svarbiausių naujų vaistų paieškos faktorių. To pasekoje *in silico* metodų kūrimas bei naudojimas neaplenkė tiek CYP450, tiek ir kitų metabolinių fermentų specifiškumo ar regioselektyvumo tyrimų. Vis dėlto, šiuo atveju modeliuojamo objekto sudėtingumas lemia tai, kad pasiekimai šioje konkrečioje srityje kol kas yra palyginti kuklūs ir didžiausio proveržio joje dar tik tikimasi ateityje. Galimybė iš anksto numatyti bent jau pagrindinius junginio metabolitus suteikia informaciją apie jo struktūroje esančius lengviausiai metabolizuojamus centrus. Pastarųjų blokavimas (pvz., pakeičiant vandenilio atomą fluoru, chloru ar kitu pakaitu per daug neįtakojančiu molekulės tūrio) arba tiesiog pašalinimas leistų stipriai sulėtinti ir netgi visiškai sustabdyti CYP450 katalizuojamas metabolizmo reakcijas. Egzistuoja ir kiti ne visai tiesioginiai būdai tam pačiam tikslui pasiekti. Pavyzdžiui, junginio lipofiliškumo sumažinimas ar papildomų grupių įvedimas šalia metabolizuojamo fragmento, siekiant sukelti sterinius trukdžius, apsunkinančius molekulės metabolizmą. Metoprololio molekulės modifikacija įvedant ciklopropilo žiedą [4], pateikiama 1 paveiksle, iliustruoja būtent pastarojo tipo pavyzdį. Su detalia galimybių studija kaip šiuolaikiniai kompiuteriniai metodai gali būti panaudojami tokių pasirinkto pradinio junginio modifikacijų analizėje galima susipažinti M. M. Ahlstrom kartu su kolegomis paskelbtame straipsnyje [5]. Tais atvejais, kuomet neįmanoma panaikinti visų metabolizuojamų centrų molekulėje, tie patys *in silico* modeliai galėtų padėti identifikuoti pagrindinę CYP450 izoformą metabolizuojančią vieną ar kitą konkretų centrą. Selektivių modifikacijų pagalba tokioje situacijoje galima būtų pamėginti užtikrinti, kad už visą junginio metabolizmą būtų atsakinga vienintelė konkreti CYP450 izoforma. Toks pasiekimas taip pat gali turėti didžiulę praktinę naudą. Planuojant kuriamo vaisto vartojimą kartu

su kitais preparatais, skirtingų CYP450 izoformų nulemiamas kiekvieno iš jų metabolizmas minimizuoja galimos vaistų sąveikos pavojų. Be kita ko, netgi preliminarūs CYP450 fermentų regioselektyvumo modeliai gali stipriai pagelbėti analizuojant sudėtingas chromatogramas bei spektrus gaunamus eksperimentinio molekulės metabolitų nustatymo metu. Ši procedūra neišvengiamai būtina atliekama vėlesnėse tyrimo pakopose, jeigu junginys jas pasiekia.



Pašalinimas dėl
pirmojo praėjimo
metabolizmo *in vivo*

50%

15%

V_{\max}
(mikrosomose)

0,46 nM/min

0,07 nM/min

$t_{1/2}$
(žmogaus)

3,5-6 val.

16-22 val.

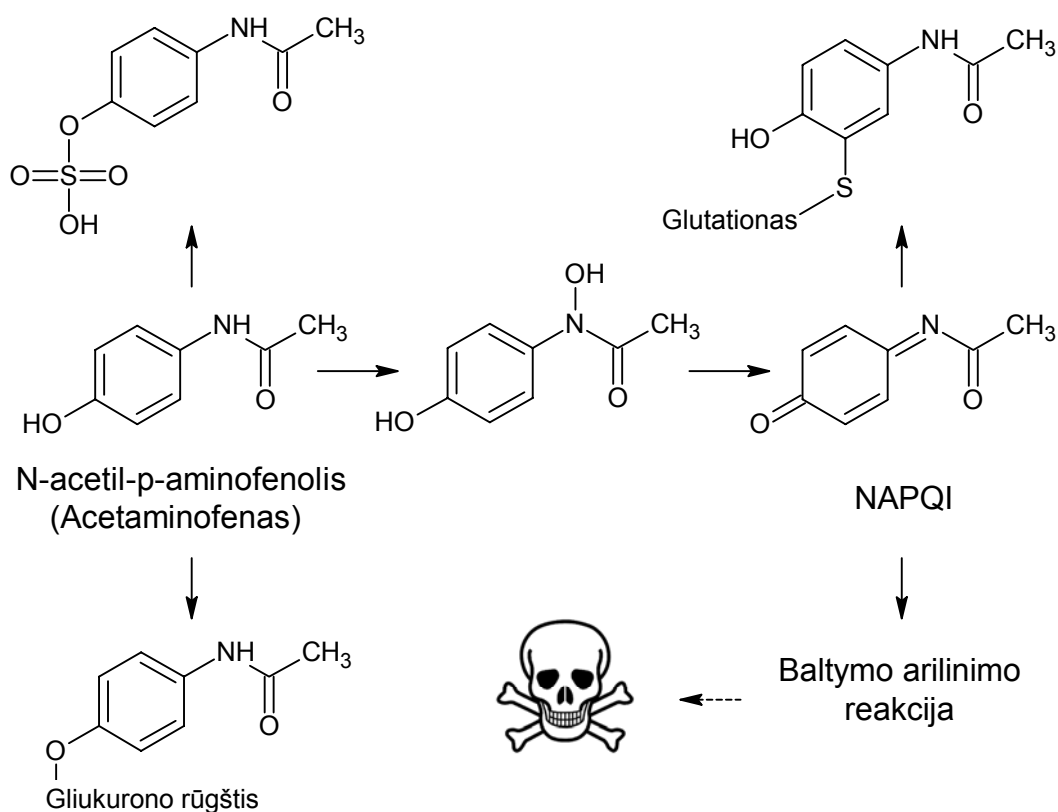
1 pav. Metoprololio molekulės modifikacija, leidžianti keleriopai prailginti junginio egzistavimo žmogaus organizme puslaidį [4]. Žvaigždute pažymėtas metoprololio pagrindinis metabolizmo kelias – CYP2D6 O-dealkilinimas

Šiame darbe CYP450 fermentų regioselektyvumo modelių kūrimui buvo pasirinkta GALAS modeliavimo metodika [6, 7]. Pastarasis metodas, sukurtas spręsti tam tikroms empirinio QSAR modeliavimo problemoms, iš pat pradžių buvo pristatytas „PLS + Korekcija pagal panašumą“ pavadinimu ir tik palyginus neseniai įgijo GALAS vardą, kuriuo yra vadinamas šiame tekste.

Toks pasirinkimas buvo pagrįstas prielaida, jog kai kurios šio metodo savybės turėtų būti ypatingai naudingos sprendžiant pagrindinius iššūkius, kurių galima tikėtis nagrinėjant tokį sudėtingą objektą kaip fermento regioselektyvumas. Tačiau iki šiol GALAS modeliavimo metodas buvo sėkmingai išbandytas tik prognozuojant palyginus paprastas fiziko-chemines organinių junginių savybes, pvz., pasiskirstymo koeficientą tarp vandens ir *n*-oktanolio ($\log P_{O/W}$) ar tirpumą gryname vandenyje ($\log S_w$) [6, 7]. Dėl šios priežasties bandymas iš karto pereiti prie tokio sudėtingo biocheminio objekto kaip citochromų P450 sistema, tiesiog tikintis sėkmės, būtų šiek tiek impulsyvus ir ne visiškai pagrįstas žingsnis. Vietoje to, šiame darbe buvo nuspręsta nagrinėjamų savybių sudėtingumą didinti palaipsniui, visų pirma pamėginant pritaikyti minėtąją GALAS metodiką ūmaus toksiškumo (LD_{50}) duomenų analizei.

Paradoksalu, tačiau CYP450 fermentai, atsakingi už organizmui svetimų cheminių medžiagų nukenksminimą, tuo pačiu yra pagrindiniai klinikinėje praktikoje pasitaikančių su junginio metabolizmu susijusio toksiškumo pasireiškimų kaltininkai. Ko gero, geriausiai šią dviprasmišką situaciją yra apibūdinęs Nobelio premijos medicinos srityje laureatas seras James'as Black'as, kuris yra pasakęs: „Esminis šių fermentų veikimo principas yra selektyvumo nebuvimas: jie atpažįsta paplitusias bendras organinių molekulių chemines ypatybes, o ne pačias molekules. Ši svetimų medžiagų metabolizmo sistema, pritaikyta molekulinei invazijai, yra tiek pat gyvybiškai svarbi žinduolių rūšies išlikimui, kaip ir imunologinis aparatas, kuris buvo pritaikytas susidoroti su mikroorganizmų invazija. Molekuliniame lygmenyje jiems abiem būdinga ta pati bendra ypatybė – selektyvumo nebuvimas. Tačiau įgimtas nešališkumas turi taip pat ir neigiamų aspektų: imunologinė sistema kartais apsigauja savo organizmo pavidale matydama priešą bei pradeda elgtis savižudiškai; tuo tarpu svetimų medžiagų metabolizmo sistema kartas nuo karto pagamina produktų, kurie yra reaktingesni ir biologiškai labiau pražūtingi nei pradinis substratas“ [8]. Kenksmingų metabolitų susidarymo problema aktuali netgi kai kurioms rinkoje gerai žinomoms ir gana paprastoms molekulėms, pavyzdžiui, acetaminofeniui. Šis junginys kaip pagrindinė

veiklioji medžiaga ar viena iš sudėtinių dalių yra sutinkamas tokiuose plačiai žinomuose vaistuose kaip citramonas, paracetamolis, APAP[®], Efferalgan[®] ir kt. Šio vaisto metabolitų medis (žr. 2 pav.) yra iš esmės saugus, išskyrus vieną metabolitą, kuriuo virsta viso labo mažiau nei 15% dozės – NAPQI [9, 10]. Šis cheminis junginys gali reaguoti su ląstelės citozolyje esančiais baltymais ir stipriai sumažinti laisvo glutathiono (GSH) kiekį. Esant nors kiek didesniai acetaminofeno perdozavimui, metabolizmo metu susidarantis NAPQI kiekis gali sukelti kepenų nepakankamumą ir netgi mirtį. Kai kuriose specifinėse žmonių populiacijose, kurioms dėl genetinių ar kitų aplinkos veiksnių būdingas mažesnis nei įprasta GSH kiekis organizme, analogišką poveikį gali sukelti netgi nedidelės acetaminofeno dozės.



2 pav. Reakcijos susijusios su acetaminofeno metabolizmu žmogaus organizme

Taigi ūmaus toksiškumo (LD_{50}) modeliavimo pasirinkimas pradine šios disertacijos tema prieš pereinant prie metabolizmo regioselektyvumo problemos sprendimo, galima sakyti, yra pagrįstas dvejopais sumetimais. Ši

savybė ne tik užima tarpinę padėtį sudėtingumo skalėje tarp fundamentaliai besiskiriančių paprasčiausių fizikocheminių savybių aprašančių visą junginį ir metabolizmo fermentų regioselektyvumo, susijusio su individualiais molekulės atomais, tačiau, kaip matyti iš pateiktų pavyzdžių, taip pat yra labai glaudžiai su pastaruoju susijusi. Iš tiesų detalus visų stebimų junginio metabolitų identifikavimas bei jų toksiškumo įvertinimas yra neatsiejami ir dažniausiai lygiagrečiai atliekami bet kokio vaistinio junginio kandidato tyrimo žingsniai.

Apibendrinant įvadinėje dalyje pateikiamą informaciją, pagrindiniai šios disertacijos tikslai buvo tokie:

- Naudojant GALAS modeliavimo metodiką sukurti individualius ūmaus toksiškumo (LD_{50}) modelius šešioms sistemoms, kiekvieną kurių sudaro konkreti graužikų rūšies ir cheminės medžiagos patekimo į organizmą būdo kombinacija: pelė (OR, IP, IV ir SC) bei žiurkė (OR ir IP);
- GALAS metodo pagalba sukurti svarbiausių CYP450 izoformų (CYP3A4 ir CYP2D6) regioselektyvumo modelius tipiškiausiems šių fermentų katalizuojamų reakcijų tipams (alifatiniam ir aromatiniam hidroksilinimui, bei N- ir O-dealkilinimui), tokiu būdu išnagrinėjant galimybes pastarąją metodiką pritaikyti atomo-centrinių savybių prognozavimui;
- Įvertinti pagrindinių GALAS metodo funkcijų (lokalių korekcijų atsižvelgiant į panašumą, prognozės patikimumo įvertinimo ir apmokymo naujais eksperimentiniais duomenimis) veikimo efektyvumą ženkliai išaugus analizuojamų modeliavimo objektų sudėtingumui.

1 LITERATŪROS APŽVALGA

1.1 Ūmaus toksiškumo objektas

1.1.1 Ūmaus toksiškumo (LD_{50}) modeliavimai

„Letalinė dozė penkiasdešimt“ (LD_{50} , angl. *lethal dose fifty*) yra kiekybinė ūmaus cheminių junginių toksiškumo išraiška, nurodanti medžiagos kiekį (mg/kg), 24 valandų bėgyje nuo patekimo į organizmą, sukeltą mirtį 50% tiriamos gyvūnų populiacijos. Vienas populiariausių cheminių medžiagų toksiškumo eksperimentinio įvertinimo būdų yra jų LD_{50} vertės nustatymas graužikams – konkrečiai pelėms bei žiurkėms. Šiai dienai, mokslinėje literatūroje ir įvairiose viešose bei prieinamose komercinėse duomenų bazėse (pvz., [11-13]) yra sukaupta nesuskaičiuojama daugybė eksperimentinių LD_{50} duomenų. Per visą SAR ir QSAR metodų egzistavimo istoriją, jų taikymo galimybės šioje srityje buvo tyrinėjamos nuolatos, apie ką aiškiai liudija pačių įvairiausių LD_{50} dydžiui prognozuoti pasiūlytų modelių gausa [14-17]. 1 lentelėje pateikiamas sąrašas darbų, kuriuos galima išskirti kaip vienus žinomiausių LD_{50} reikšmių modeliavimo srityje, ir kai kurios esminės juos apibūdinančios detalės.

1 lentelė. Ankstesnių darbų, nagrinėjusių SAR ir QSAR metodų taikymą ūmaus graužikų toksiškumo (LD_{50}) prognozavimui, apžvalga

Metai	Šaltinis	Tyrinėtos junginių klasės	N (apmokomojo rinkinio)
1978	Enslein ir Craig [18]	Įvairios	425
1983	Enslein <i>et al.</i> [19]	Įvairios	1851
1985	Tichy <i>et al.</i> [20]	Alifatiniai alkoholiai (C_1 - C_5)	nuo 5 iki 8
1987	Enslein <i>et al.</i> [21]	Įvairios	147
1989	Enslein <i>et al.</i> [22]	Įvairios	3624
1991	Jaeckel ir Klein. [23]	Aminai ir anilinai	26 ir 33
1991	Lipnick [24]	Alkoholiai ir ketonai	54
1996	Zakarya <i>et al.</i> [25]	Amidai	44

1 lentelė. Ankstesnių darbų, nagrinėjusių SAR ir QSAR metodų taikymą ūmaus graužikų toksiškumo (LD_{50}) prognozavimui, apžvalga (tęsinys)

Metai	Šaltinis	Tyrinėtos junginių klasės	<i>N</i> (apmokomojo rinkinio)
1997	Johnson ir Jurs [26]	Aniliniai	103
1998	Wang ir Bai [27]	Alkoholiai	95
1999	Eldred ir Jurs [28]	Organiniai fosforo junginiai	49
2002	Cronin <i>et al.</i> [29]	Piridino dariniai	21
2003	Hansch ir Kurup [30]	Barbitūratai	11
2004	Devillers [31]	Organiniai fosforo junginiai	51
2006	Jean <i>et al.</i> [32]	Chlorsilantai	10
2006	Guo <i>et al.</i> [33]	Organiniai fosforo junginiai	38
2007	Freidig <i>et al.</i> [34]	Įvairios	49
2007	Toropov <i>et al.</i> [35]	Pakeisti benzeno dariniai	28
2009	Zhu <i>et al.</i> [36]	Įvairios	3472

Be jokios abejonės, kiekvienas iš paminėtų darbų yra savaip įdomus savo idėjomis ir požiūriu į problemą. Juose gaunami rezultatai gali pagilinti suvokimą apie ūmaus toksiškumo efekto prigimtį ir pasireiškimo mechanizmus bei leisti kiekvienu konkrečiu atveju padaryti svarbių išvadų. Pavyzdžiui, Lipnick'as savo analizėje [24] panaudojo paprastą, tačiau ganėtinai efektyvų bazinio QSAR metodą. Pasinaudodamas chemiškai nesudėtingų junginių (monohidroksilių alkoholių ir sočiųjų ketonų) eksperimentinėmis ūmaus toksiškumo vertėmis jisai sukūrė įprastą nuo $\log P_{O/W}$ priklausantį dilinijinį (angl. *bilinear*) LD_{50} (žiurkė, OR) modelį. Tokiu būdu buvo siekiama identifikuoti naujus nežinomus toksikologinius efektus, sukeltus sudėtingesnės struktūros junginių, kurie savo ruožtu identifikuojami kaip iš bazinės koreliacinės analizės iškrentančios prognozės. Šis ir kiti panašūs to paties autoriaus darbai [37, 38] yra svarbūs ūmaus toksiškumo modelių raidos istorijoje, nes padėjo pagrindus mechanistiniam šio reiškimo nagrinėjimui ir leido susieti gautus rezultatus su nagrinėjamųjų junginių reaktingumu. Tačiau bendru atveju, jau pirmasis žvilgsnis į 1 lentelę iš karto atskleidžia vieną ir

pagrindinę problemą, susijusią su absoliučia dauguma praeityje publikuotų ūmaus toksiškumo graužikams QSAR modelių. Tai labai mažas ir, kaip taisyklė, stipriai apribotas cheminės įvairovės atžvilgiu modelio apmokomasis rinkinys. Dar viena labai aktuali problema yra tokių modelių statistinis įvertinimas ir autorių pateikiama informacija apie tai. Neretai publikuojami modelio statistiniai parametrai yra apskaičiuojami sumažintiems junginių rinkiniams (prisiminkime, kad modelių apmokomieji rinkiniai šiuo atveju ir taip yra labai riboti). Kaip taisyklė tokiais atvejais nėra stengiamasi ką nors paaiškinti – ar iš tiesų daliai junginių modelio negalima taikyti, ar vis dėlto modelis tiesiog prastai aprašo į jo pritaikomumo sritį patenkančius junginius. Tiesiog atmetama tam tikra dalis prasčiausiai modelio aprašomų junginių, motyvuojant jog tai iškrentančios prognozės (angl. *outliers*), nors dažniausiai susidarantis išpūdis yra toks, jog autoriai paprasčiausiai stengiasi pagerinti priešingu atveju neypatingai gerus modelio statistinius rodiklius [14]. Taip pat jaučiamas kritiško modelio prognozuojamosios galios įvertinimo trūkumas. Dažnas modelis apsiriboja tik statistiniu turimų duomenų modeliavimu, geriausiu atveju, kartais papildomai pateikiami kryžminės patikros duomenys (angl. *cross validation*) [14]. Ir vos keletas autorių naudoja pilnavertį patikrinamąjį junginių rinkinį, tokiu būdu korektiškai įvertindami tikėtiną savo QSAR modelių veikimą naujiems junginiams [26, 31]. Tiesa sakant, atmetus mokslinę išvardintų darbų reikšmę, labai daugeliu atveju jų praktinė vertė yra geriausiu atveju gana ribota, kadangi atitinkamus modelius be apribojimų galima taikyti tikrai nedideliems tos pačios klasės junginių rinkiniams.

Kita vertus, literatūroje iki šiol yra paskelbti vos keletas sėkmingų QSAR modelių, galinčių prognozuoti LD_{50} reikšmes praktiniu požiūriu pakankamo dydžio bei įvairovės junginių rinkiniams. Tarp tokių darbų yra Enslein'o kartu su kolegomis paskelbti modeliai [18, 19, 22], kurie vėliau tapo pagrindu kuriant vieną žinomiausių įvairių su toksiškumu susijusių efektų prognozavimo programų paketą – TOPKAT[®] [39]. Kitas svarbus pavyzdys yra visiškai neseniai pasirodžiusi Zhu ir bendradarbių publikacija [36] aprašanti tiesioginį jau minėtosios TOPKAT[®] programos palyginimą su keletu pačių

autorių pasiūlytų LD_{50} (žiurkė, OR) modelių. Dėl nagrinėjamų junginių kiekio ir įvairovės pastaruosius darbus derėtų laikyti pačiais aktualiausiais šios disertacijos kontekste. Todėl joje dar bus sugrįžta prie kai kurių paskutiniajame 1 lentelės darbe [36] skelbiamų rezultatų detalesnio aptarimo.

Nepaisant visų šių faktų, liudijančių nemenką įdirbį ūmaus toksiškumo modeliavimo srityje, iki šiol galima aptikti periodiškai įvairių autorių išreiškiamų abejonių ar sėkmingai prognozuoti šį dydį yra apskritai įmanoma [40-42]. Pagrindinius sunkumus sukelia pati nagrinėjamojo objekto prigimtis – ūmus toksiškumas yra nulemiamas tokios gausybės galimų veikimo mechanizmų ir kitų biologinių faktorių, jog iš pirmo žvilgsnio patikimos LD_{50} dydžio prognozės gali iš tiesų atrodyti sunkiai įmanomos.

Viena iš nagrinėjamų potencialių alternatyvų yra QSAR ir vadinamųjų ekspertinių sistemų (angl. *expert systems*) derinimas, kuomet pasireiškiantis ūmus junginio toksiškumas yra skaidomas į sudėtinės dalis. Kiekviena galima biologinio veikimo pakopa ar kelias šiuo atveju yra aprašomi savo individualiais lokaliais QSAR modeliais [41]. Vis dėlto, tokia koncepcija šiuo metu yra gana smarkiai ribojama vis dar nepakankamo supratimo apie galimus toksiškumą nulemiančius mechanizmus. Dėl šių priežasčių, panašu, jog kol kas pilnaverčių tokio tipo šios problemos sprendimų artimiausioje ateityje tikėtis nevertėtų. Taigi, kol kas toks požiūris dažniausiai traktuojamas kaip daug, tačiau tikrai labai ilgalaikėje perspektyvoje žadantis sprendimas, pareikalausiantis daug pastangų jį įgyvendinant [15, 42].

Kita įmanoma strategija yra mėginti prognozuoti LD_{50} pasinaudojant empirinėmis koreliacijomis su įvairiausių tipų *in vitro* toksiškumo eksperimentiniais duomenimis [43-47] ar iš įvairiausių tarprūšinių koreliacijų (angl. *interspecies correlations*). Nors *in vitro* tyrimas yra nepalyginamai pigesnis nei *in vivo* eksperimentas, tačiau kainos atžvilgiu, privalumas ir šiuo atveju vis tiek išlieka praktiškai nieko nekainuojančių *in silico* prognozių pusėje. Be to, net ir palyginus paprastiems *in vitro* tyrimams, kaip ir bendrai paėmus bet kokioms eksperimentinėms studijoms, būdingas mažesnis našumas ir, ko gero svarbiausia, toksiškumo įvertinimo „mėgintuvėlyje“ atveju yra

ignoruojami gyvam organizmui kaip visumai būdingi ADME efektai [15]. Šis trūkumas yra mažiau aktualus vadinamųjų tarprūšinių koreliacijų atveju, kadangi šiuo atveju, priklausomai nuo pasirinkto „modelinio“ gyvojo organizmo, teoriškai yra įmanoma atsižvelgti bent jau į dalį ADME efektų. O kad tokie efektai yra reikšmingi galima įsitikinti panagrinėjus vieną iš būtent tokio tipo darbų, kuriame bandoma susieti graužikams išmatuotas LD_{50} vertes su *Vibrio fischeri* (*V.f.*) bakterijų bioluminescencijos inhibavimo EC_{50} reikšmėmis [48]. Literatūroje publikuojamos tokios šiame darbe išvestos lygtys:

$$\log(1/LD_{50} \text{ žiurkė OR}) = 0,20 \cdot \log(1/EC_{50} V.f.) - 0,96 \quad (1)$$

$N = 471; R = 0,35; sd = 0,74$

$$\log(1/LD_{50} \text{ pelė OR}) = 0,20 \cdot \log(1/EC_{50} V.f.) - 0,86 \quad (2)$$

$N = 344; R = 0,35; sd = 0,72$

$$\log(1/LD_{50} \text{ žiurkė IP}) = 0,29 \cdot \log(1/EC_{50} V.f.) - 0,48 \quad (3)$$

$N = 195; R = 0,48; sd = 0,82$

$$\log(1/LD_{50} \text{ pelė IP}) = 0,25 \cdot \log(1/EC_{50} V.f.) - 0,49 \quad (4)$$

$N = 378; R = 0,43; sd = 0,70$

$$\log(1/LD_{50} \text{ žiurkė IV}) = 0,40 \cdot \log(1/EC_{50} V.f.) - 0,25 \quad (5)$$

$N = 54; R = 0,73; sd = 0,79$

$$\log(1/LD_{50} \text{ pelė IV}) = 0,35 \cdot \log(1/EC_{50} V.f.) - 0,30 \quad (6)$$

$N = 165; R = 0,68; sd = 0,61$

Akivaizdžiai matyti, jog intraveninio junginio įvedimo atveju ((5) ir (6) lygtys), kuomet ADME efektų įtaka yra praktiškai mažiausia iš visų įmanomų įvedimo būdų, kadangi medžiaga įvedama tiesiai į kraujotakos sistemą, gaunami žymiai geresni rezultatai, nei likusiais atvejais. Kita vertus, cheminėms medžiagoms patenkant į organizmą oraliniu keliu ir pereinant visus įmanomus apsaugos nuo pašalinių medžiagų barjerus, nesugebėjimas atsižvelgti į ADME efektus skaudžiai atsiliepia modelio tikslumui ir rezultatai šiam įvedimo būdai yra prasčiausi iš visų pateikiamų ((1) ir (2) lygtys). Problema šiuo atveju kaip ir kituose panašiuose darbuose nagrinėjančiuose

tarprūšines koreliacijas su bakterijomis [49, 50] yra per daug paprastas palyginamasis organizmas (bakterija) nėra iš tolo nepriartėjantis prie sudėtingos žinduolių organizmo sandaros. Be abejo, per visą tyrimų šioje srityje laiką buvo nagrinėjamos tarprūšinės graužikų toksiškumo koreliacijos ir su gerokai artimesniais žinduoliams pagal savo sandaros sudėtingumą organizmais: daugialąščiais bestuburiais [51, 52], kirmėlėmis [53], vėžiagyviais [50-52, 54], žuvimis [55-59]. Žuvų atveju, panašiai kaip ir graužikams, kai kuriuose tyrimuose išskiriamos skirtingos toksiškumo reikšmės priklausomai nuo patekimo į žuvies organizmą būdo (pvz., kieta medžiaga patenka oraliniu keliu, junginio dozė įvedama intraperitoninės injekcijos pagalba ar vandenyje ištirpusi medžiaga pasisavinama per žiaunas ir visu kūno paviršiumi). Tačiau netgi šitoks detalumas dažniausiai neduoda apčiuopiamų teigiamų rezultatų – intraveninis toksiškumas ir toliau dažniausiai prognozuojamas pastebimai geriau, nei likusieji tipai (užuomina, jog netgi žuvies organizmas yra sistema vis dar nepakankamai atkartojanti žinduolių kūne vykstančius biologinius procesus) o patys rezultatai, bendrai paėmus, yra labai vidutiniški. Ir visa tai, turint omenyje, kad modelių gavimui naudojami apmokomieji rinkiniai vėlgi yra palyginti nedideli (daugiausiai keli šimtai junginių) ir nėra vieno iš čia paminėtų interkoreliacinių modelių prognozuojamoji galia net nebuvo pamėginta įvertinti nei naudojant patikrinamąjį junginių rinkinį, nei LOO (LMO) ar panašias metodikas. Visa tai didžiąja dalimi nulemia nepakankamas žinomų eksperimentinių duomenų kiekis, kuris yra neginčijamai pats didžiausias tokio tipo metodų trūkumas. Nesant galimybės literatūroje surasti jau žinomos junginio toksiškumo vertės palyginamojoje sistemoje, kuri koreliuojama su ūmiu pelės ar žiurkės toksiškumu, reikia atlikti eksperimentą. Pastarasis, be abejo, yra tikrai paprastesnis, pigesnis ir ne taip griežtai reglamentuojamas kaip ūmaus toksiškumo testai su žinduoliais, tačiau vis dėlto finansine ir laiko sąnaudų prasme tai yra nepalyginamai sudėtingiau netgi lyginant su minėtaisiais *in vitro* testais, jau nekalbant apie *in silico* prognozes.

Taigi, nors šiai dienai prieinamos LD_{50} prognozės bei metodai naudojami atitinkamiems QSAR modeliams gauti yra gana stipriai

kritikuojami, tačiau iki šiol alternatyvų *in silico* metodams, kokie jie bebūtų, kalbant apie ūmaus toksiškumo įvertinimą dideliems junginių rinkiniams, praktiškai nėra. Na o kalbant apie virtualių bibliotekų analizę, *in silico* skaičiavimai yra apskritai vienintelė teoriškai įmanoma išeitis. Dėl šių priežasčių ūmus toksiškumas kaip tyrimo objektas QSAR modeliavimuose niekada nebuvo dingęs iš tyrėjų akiračio. Pastarojo meto apžvalginuose straipsniuose susijusiuose su šios sudėtingos savybės modeliavimu išskiriama keletas problemų, į kurias tiesiog būtina kažkokiu būdu atsižvelgti tokio tipo darbuose. Visų pirma, tai dažnai pasitaikantis neįprastai didelis eksperimentinių duomenų išsibarstymas panašiems junginiams (nulemiamas visiškai skirtingų veikimo mechanizmų, ar tiesiog paprasčiausių eksperimentinių klaidų) [15, 16] ir aišku „duomenų skylės“ (eksperimentiniais matavimais nepadengtos cheminės junginių erdvės sritys) [16]. Ūmaus toksiškumo (LD_{50}) modeliavimui pritaikytos GALAS metodikos gebėjimas efektyviai identifikuoti šias problemas tarp visų prognozių yra vienas iš laukiamų šio konkretaus darbo rezultatų.

1.1.2 Eksperimentinių LD_{50} matavimų kokybė

Sudėtinga ūmaus toksiškumo reiškinio prigimtis lemia ne tik tai, kad šį dydį yra sunku prognozuoti, tačiau ir tai, jog jį sudėtinga tiksliai išmatuoti eksperimentiškai. Šioje vietoje egzistuoja keletas faktorių įtakojančių eksperimentinius matavimus, kurie yra vertinami labai nevienareikšmiškai. Vienas tokių yra *in vivo* LD_{50} testuose naudojamų pelių ar žiurkių lytis. Plačiai žinoma, kad bet kurios rūšies skirtingos lyties individams būdinga daugybė skirtumų – pradedant išorine išvaizda ir baigiant vidinių medžiagų apykaitos procesų greičiu, reguliavimu ir pan. Taigi natūralu, kad atsparumo cheminių medžiagų toksiniam poveikiui požiūriu, šioje vietoje taip pat galima tikėtis tam tikrų skirtumų. Bėda ta, kad šie skirtumai neretai yra ganėtinai individualūs (lyginant su vyriškosios lyties individais, reakcija į konkrečią cheminę medžiagą gali skirtis tarkime 50% moteriškosios giminės atstovių), taigi esant ribotai gyvūnų imčiai konkretaus eksperimento atveju, jie nebūtinai yra

identifikuojami. Ir apskritai, atskirų lyčių individų skirtingas jautrumas cheminėms medžiagoms, kai kalbama apie ūminį toksiškumą, nėra universalus reiškinys, t.y. šalia tokių junginių, kurių poveikis moteriškai ir vyriškai lytims iš tiesų yra skirtingas, egzistuoja daugybė tokių, kuriems skirtingų lyčių atstovų atsakas yra identiškas. Grįžtant prie 1 lentelės, joje taip pat galima aptikti pavyzdžių, kuomet modelio kūrimo metu yra mėginama atsižvelgti į lyties faktorių, naudojant papildomus kintamuosius statistinėje analizėje [31]. Taip pat jau minėtoje TOPKAT[®] programoje ūminio toksiškumo (LD_{50}) vertės vyriškos ir moteriškos lyties atstovams kiekvienu konkrečiu atveju yra traktuojamos kaip atskiri prognozuojami dydžiai. Tačiau pačiu bendriausiu atveju visuomet reikėtų kelti klausimą, kuriai daliai nagrinėjamųjų junginių skirtumai tarp lyčių jautrumo yra aktualūs bei dar svarbiau, kokio dydžio skirtumų galima potencialiai tikėtis. Vienoje paskutiniųjų pastarojo meto apžvalgų ūmaus toksiškumo tema [14] yra pristatomi 10 pesticidų serijos eksperimentiniai LD_{50} duomenys žiurkei po medžiagos patekimo į organizmą oraliniu keliu [60] (žr. 2 lentelę). Šie junginiai pateikiami kaip labai puikiai žinomas reikšmingų skirtumų tarp atitinkamų LD_{50} reikšmių vyriškai ir moteriškai giminei pavyzdys. Iš to seka natūrali prielaida, kad pateikiami skirtumai ko gero turėtų būti vieni iš didžiausių, su kuriais apskritai galime tikėtis susidurti praktikoje.

2 lentelė. Eksperimentinės pasirinktų pesticidų klasės atstovų ūmaus oralinio toksiškumo vertės suaugusioms vyriškos ir moteriškos lyties Sherman'o žiurkėms [60]

Junginys	Vyriška lytis		Moteriška lytis		log LD_{50} skirtumas
	LD_{50} , mg/kg	log LD_{50}	LD_{50} , mg/kg	log LD_{50}	
Aldrinas	39	1,59	60	1,78	-0,19
Chlordanas	335	2,53	430	2,63	-0,10
Endrinas	17,8	1,25	7,5	0,88	0,37
Delnavas	43	1,63	23	1,36	0,27
Demetonas	6,2	0,79	2,5	0,40	0,39
Diazinonas	108	2,03	76	1,88	0,15

2 lentelė. Eksperimentinės pasirinktų pesticidų klasės atstovų ūmaus oralinio toksiškumo vertės suaugusioms vyriškos ir moteriškos lyties Sherman'o žiurkėms [60] (tęsinys)

Junginys	Vyriška lytis		Moteriška lytis		log LD_{50} skirtumas
	LD_{50} , mg/kg	log LD_{50}	LD_{50} , mg/kg	log LD_{50}	
EPN	36	1,56	7,7	0,89	0,67
Parationas	13	1,11	3,6	0,56	0,55
Fosdrinas	6,1	0,79	3,7	0,57	0,22
Šradanas	9,1	0,96	42	1,62	-0,66

Pirmasis į akis krentantis pastebėjimas yra tas, kad pateiktieji skirtumai yra nevienareikšmiai. Kitaip tariant, nėra taip, kad kaž kurios lyties (pvz., moteriškosios) atstovai būtų visada jautresni cheminių medžiagų poveikiui, ar atvirkščiai. Šiuo atveju tai priklauso nuo konkretaus cheminio junginio. Turint tai omenyje, visiškai logiška yra J. Devillers darbe [31] padaryta išvada, kad nagrinėjant abiejų lyčių ūmaus toksiškumo duomenis sujungtus į vieną rinkinį ir tiesiog analizėje naudojant papildomus kintamuosius identifikuojančius lytį kiekvieno stebėjimo atveju, būtina naudoti netiesinius statistinius metodus (pvz., ANN) norint gauti prasmingą modelį.

Kita vertus iš 2 lentelės duomenų akivaizdu, kad nors individualūs skirtumai LD_{50} reikšmių atveju ir atrodo pakankamai reikšmingi, tačiau konvertavus pastarąsias į modeliavime įprastą log LD_{50} dydį pasirodo, kad vos trijų junginių atveju absoliutūs skirtumai tarp vyriškos ir moteriškos giminės individų ūminio toksiškumo (žiurkė, OR) reikšmių viršija 0,5 logaritminio vieneto (o ir tais trim atvejais skirtumai nėra stipriai didesni už 0,5). Literatūroje nepavyko aptikti aiškios nuorodos į konkrečios LD_{50} vertės eksperimentinio nustatymo paklaidos įvertinimą, tačiau bendrą vaizdą pakankamai nesunku susidaryti iš elementarių samprotavimų. Palyginus nesudėtingų fiziko-cheminių savybių (log $P_{O/W}$, log S_w) matavimų atveju, standartinės paklaidos tarp įvairiose laboratorijose gaunamų rezultatų vertinimai rodo ją esant maždaug 0,5 logaritminio vieneto ribose [61, 62]. Tuo tarpu tokiuose kaip LD_{50} nustatymo eksperimentuose, kai vertinama gyvo

organizmo reakcija, be prieš tai minėtam atvejui būdingos eksperimentinių sąlygų ir žmogiškojo faktoriaus įtakos atsiranda dar visa eilė naujų veiksnių nulemiančių papildomą matavimų išsibarstymą. Tai ir gyvūno amžius bei sveikata, jo sotumo, nuovargio ir netgi streso laipsnis eksperimento metu ir t.t. Be abejo, kiek įmanoma visus šiuos veiksnius yra bandoma standartizuoti ir kontroliuoti, tačiau nereikėtų pamiršti, jog bet kokia tokio pobūdžio kontrolė yra tiek pat patikima, kiek žmogiškasis ją atliekančio kontrolieriaus faktorius. Taigi, gana užtikrintai galima teigti, kad eksperimentinė LD_{50} nustatymo paklaida yra bent jau ne mažesnė nei 0,5 logaritminio vieneto, o labiausiai tikėtina, kad pastebimai didesnė nei šis dydis. Šiame kontekste akivaizdžiai matyti, kad skirtumai pateikiami 2 lentelėje nėra pakankamai statistiškai reikšmingi. Kitaip tariant, kiekvienu konkrečiu atveju praktiškai neįmanoma atskirti ar vyriškos ir moteriškos giminės individams išmatuotos LD_{50} vertės skiriasi dėl skirtingo abiejų lyčių individų jautrumo cheminei medžiagai, ar dėl elementarios eksperimentinės paklaidos. Nesunku numatyti, kad operuojant tūkstančių junginių eilės dydžio duomenų rinkiniais, eksperimentinis stebėjimų išsibarstymas neišvengiamai paprasčiausiai užgoš tuos atvejus, kuomet stebimų LD_{50} reikšmių skirtumai atsiranda iš tiesų dėl skirtingo atskirų lyčių individų jautrumo cheminėms medžiagoms. Taigi šiuo atveju toks stebimo dydžio diferencijavimas ir atskiras ūminio toksiškumo modeliavimas moteriškai ir vyriškai pelių bei žiurkių lytims nėra prasmingas, kadangi esant tokiai eksperimentinių duomenų kokybei, atsižvelgti į atitinkamus skirtumus paprasčiausiai nėra galimybių.

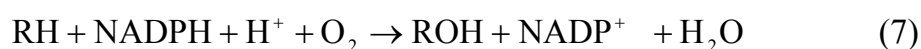
Dar vienas kartais akcentuojamas panašus faktorius yra eksperimentui naudojamos gyvūnų (šiuo atveju pelių ar žiurkių) veislės (angl. *strain*) įtaka [63]. Panašiai kaip ir skirtingų lyčių atveju, egzistuoja vos keletas LD_{50} modelių išvestų pasinaudojant eksperimentiniais rezultatais vien tik konkrečiai pelių ar žiurkių veislei, pvz. [20, 29, 31]. Neatsižvelgiant į tai, kad panašiai kaip ir su skirtingu abiejų lyčių jautrumu, šiuo atveju dideliems duomenų rinkiniams eksperimentinių stebėjimų skirtingoms veislėms apjungimas į bendrą visumą greičiausiai nedaro jokios labiau pastebimos įtakos, egzistuoja

ir kita, gerokai pragmatiškesnė priežastis, kodėl detaliai į abiejų čia aptartų faktorių įtaką atsižvelgiantys modeliai yra tokie reti. Nežinia, ar dėl samprotavimų analogiškų pateiktiesiems šiame darbe, ar dėl kokių nors kitų priežasčių, literatūroje labai retai galima aptikti eksperimentinius LD_{50} duomenis, kuriems be gyvūno rūšies dar būtų papildomai nurodoma detalesnė informacija, pavyzdžiui, veislė arba lytis. Ypač aktualu tai, kad tokia informacija nėra pateikiama didžiosiose su toksiškumu susijusiose duomenų bazėse, tokiose kaip RTECS[®] [12]. Be abejo, kritikuodami šią ir kitas panašias duomenų bazes dėl surinktų duomenų kokybės, mokslininkai tarp kitų atsižvelgia ir į šį faktą [64]. Tačiau nepaisant visos kritikos, dėl bet kokių priežasčių, yra sunku tiesiog ignoruoti ranka pasiekiamą šitokio masto duomenų šaltinį, kuriame bet kuriuo atveju gausu vertingos informacijos.

1.2 Citochromų P450 šeimos fermentų apžvalga

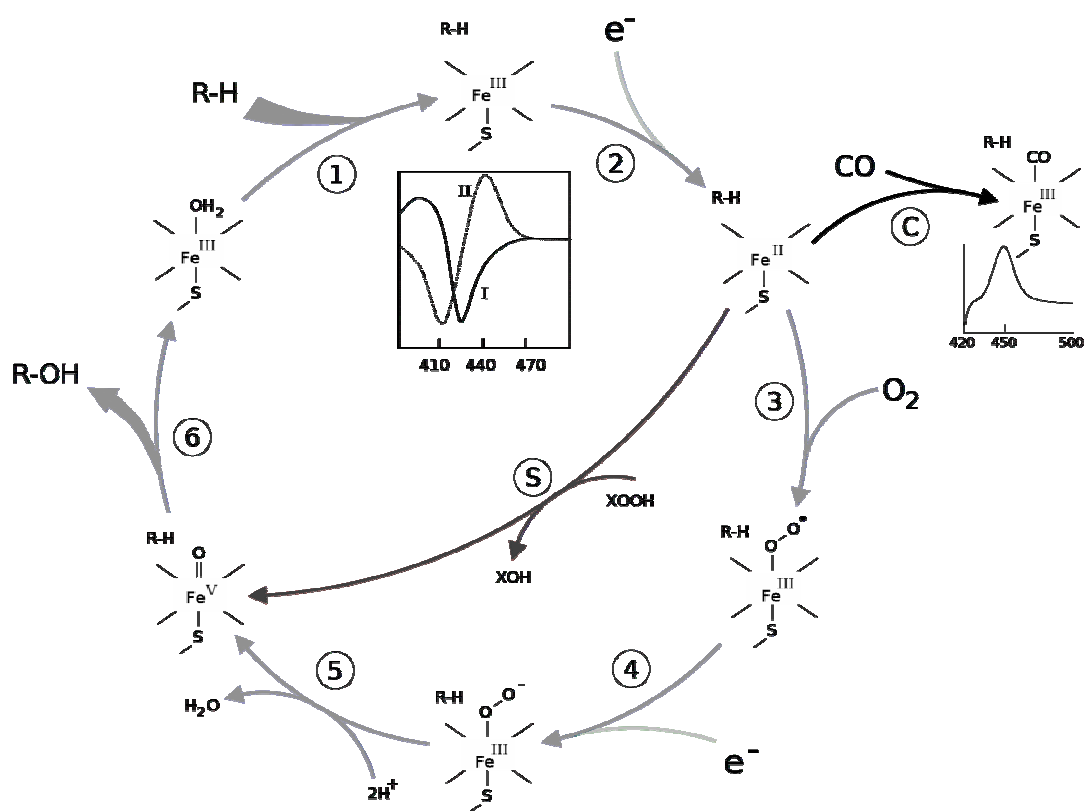
1.2.1 Žmogaus citochromai P450 ir jų katalizuojamos reakcijos

Citochromų P450 (CYP450) didšeimės (angl. *superfamily*) atstovai – tai hemą turintys fermentai, aptinkami visose biologinėse karalystėse [65] ir absoliučioje daugumoje organizmų – tiek prokariotinių, tiek eukariotinių [8]. Bendras fermentų pavadinimas (CYP450) kilęs iš specifinio absorbcijos spektro, kuriame stebimas sugerties maksimumas regimojoje spektro dalyje, ties 450 nm, būdingas redukuotos hemo formos kompleksui su anglies monoksidu [66]. Žinduolių, įskaitant ir žmogaus, organizme citochromų P450 didšeimės fermentai yra aptinkami praktiškai visuose įmanomuose audiniuose, tačiau didžiausias jų kiekis yra sukoncentruotas kepenyse [8]. Pagrindinė CYP450 fermentų katalitinė funkcija yra pačių įvairiausių substratų oksidavimas [8], kuri bendru atveju galima užrašyti tokia reakcijos lygtimi:



čia RH – substratas; ROH – oksiduotas substratas.

Šios reakcijos pasekmė, nagrinėjant eksperimento metu gaunamus metabolitus, gali būti ir dealkilinimas. Šiuo atveju po hidroksilinimo reakcijos susidaro nestabilūs tarpiniai junginiai (O,O-; N,O-acetaliai ir pan.), kuriuose skyla ryšys tarp anglies ir heteroatomo, susidarant galutiniam produktui. Hemo centre esančio geležies atomo kitimai viso (7) lygtimi apibendrinto katalitinio ciklo metu pateikti 3 paveiksle.



3 pav. Katalitinis CYP450 ciklas

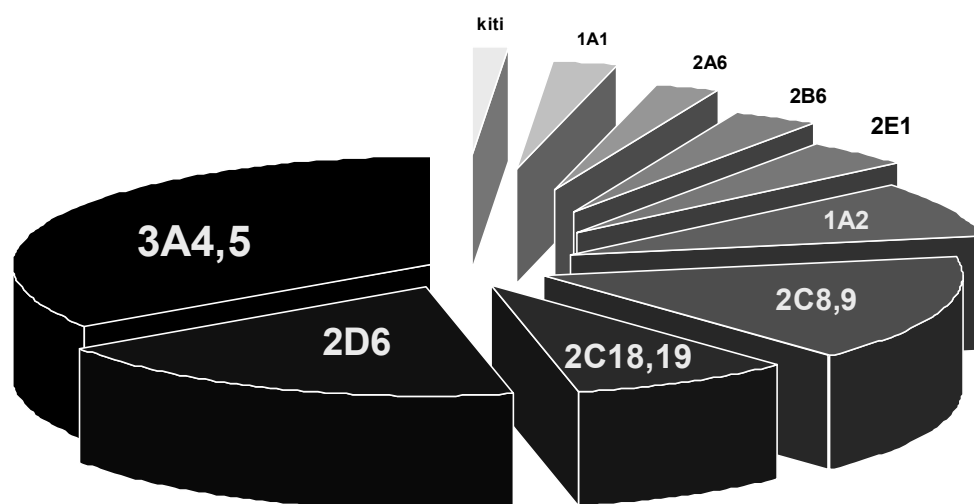
Minimas ciklas yra vienodas visiems citochromų P450 fermentams ir bendru atveju kiekvienas atskirai paimtas substrato bei aktyviojo centro susijungimas gali būti rezultatyvus (ciklas apsisuka ir susidaro oksiduotas substratas) arba ne. Kai kurios katalitinio ciklo pakopos yra grįžtamos, todėl įmanomi atvejai, ypač pačioje ciklo pradžioje, kuomet formaliai žiūrint jis sustoja ir pasuka atgal, galų gale atpalaiduodamas iš aktyviojo centro nepakitusį substratą RH [3]. Ciklo pradžioje substratas patenka į fermento aktyvųjį centrą, sudarydamas fermento-substrato kompleksą. Jame substratas atsiduria šalia hemo,

priešingoje jo plokštumos pusėje, lyginant su polipeptidinės grandinės ir to paties hemo susijungimo vieta. Prisijungusi substrato molekulė sukelia konformacinius aktyviojo centro pakitimus, išstumdamą hemo geležies kompleksą tolimojoje ašinėje pozicijoje koordinuotą vandens molekulę [67] (žr. 3 pav., 1 pakopą). Pastarajame etape geležies atomo sukinių būvis (angl. *spin state*) gali pakisti arba, lygiai taip pat sėkmingai išlikti toks pats dėl tos paprastos priežasties, jog kai kurie citochromai P450 laisvoje formoje egzistuoja žemo sukinių (angl. *low spin*) būvyje, tuo tarpu kiti net ir be prisijungusio substrato jau yra aukšto sukinių (angl. *high spin*) būsenoje [68]. 2 ir 4 pakopose dalyvaujančių elektronų donoras yra NADPH prostetinė grupė, tarpininkaujant citochromo P450 reduktazei, išskyrus retesnius atvejus, kuomet 4 pakopoje šį vaidmenį atlieka citochromas b_5 [69, 70]. Iki divalentės geležies redukuotas citochromo P450 hemas prisijungia molekulinę deguonį (3 pav. 3 pakopa), sudarydamas nestabilų kompleksą, kuriam persigrupuojant gali susidaryti visa eilė deguonį turinčių dalelių. Tačiau pradedant šia pakopa, visų tarpinių procesų ir junginių aprašymas literatūroje yra kur kas mažiau detalus. Galiausiai susidaro junginys, dažniausiai formaliai užrašomas kaip $[\text{FeO}]^{3+}$ arba (kaip aukščiau pateiktame 3 paveiksle) $[\text{Fe}^{\text{v}}=\text{O}]$, tačiau tiksli šios dalelės elektroninė sandara yra nežinoma [68]. Būtent pastarasis kompleksas dažniausiai naudojamas formaliai užrašant citochromų P450 katalizuojamų reakcijų mechanizmus ir aiškinant šių reakcijų stochiometriją, tačiau taip pat galimi ir alternatyvūs mechanizmai.

Priklausomai nuo konkretaus fermento ir nagrinėjamo substrato, CYP450 didšeimės atstovai gali katalizuoti visą eilę cheminių reakcijų. 3 paveiksle kaip pavyzdys pateikiamas vienas iš populiariausių tipų – hidroksilinimas. Po produkto atsipalaidavimo iš aktyviojo centro, fermentas sugrįžta į pradinį būvį, susigražindamas ašinėje geležies kompleksą padėtyje koordinuojamą vandens molekulę. Galimas alternatyvus vieno deguonies atomo įvedimo į substratą kelias, apėinantis 3, 4 bei 5 etapus, yra vadinamasis „peroksidinis šuntas“ (angl. *peroxide shunt*), numatantis tiesioginį geležies okso-dalelės susidarymą, dalyvaujant atominio deguonies donorams, pvz.,

peroksidams [71]. 3 paveiksle tai iliustruojama S pakopa, kurioje dalyvauja hipotetinis organinis peroksidas „XOOH“.

Disertacijoje nagrinėjamų CYP3A4 ir CYP2D6 fermentų metabolizuojamų organizmui svetimų medžiagų dalis tarp visų kepenyse nukenksminamų ksenobiotikų (angl. *xenobiotics*) grafiškai pavaizduota 4 paveiksle [8]. Šios dvi izoformos yra vieni geriausiai ištyrinėtų CYP450



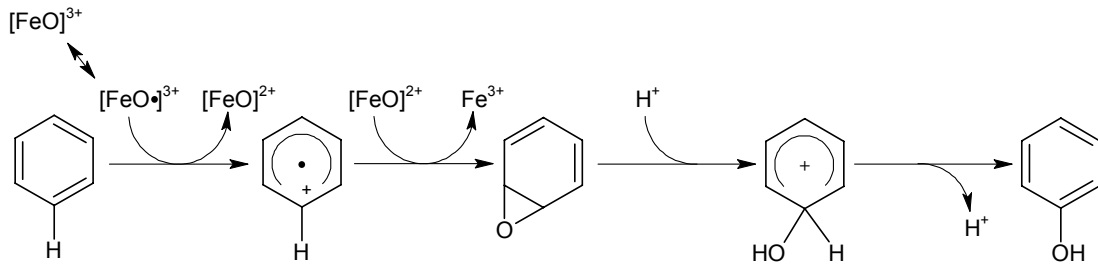
4 pav. Įvairių CYP450 fermentų izoformų nulemiamo kliniškai svarbių vaistų metabolizmo dalies palyginimas

atstovų, kurių tretinės struktūros pakankamai detalai nustatytos [72-75]. CYP3A4 galima sakyti yra klasikinis citochromų P450 atstovas, kuriam būdingas didelis lankstumas. Jo aktyvusis centras yra pakankamai didelis, kad galėtų talpinti didesnes molekules, tačiau tuo pačiu sugeba prisitaikyti ir prie mažesnių junginių. Viso to pasekoje, šio fermento substrato specifiškumas yra be galo platus, struktūrinė metabolizuojamų molekulių įvairovė yra tokia didelė, jog praktiškai yra neįmanoma išskirti kažkokių jas apibendrinančių savybių, t.y. sukurti kompaktiško taisyklių rinkinio, leidžiančio atskirti CYP3A4 substratus nuo nesubstratų. Tai iš esmės sutampa su teorija, kad visa citochromų P450 didšeimė evoliucijos keliu išsivystė kaip mažo giminingumo fermentų grupė, skirta saugotis nuo kuo platesnės ksenobiotikų įvairovės. Kita vertus, CYP2D6 substratų įvairovė yra kur kas mažesnė. Iš tiesų, nors ir šiuo atveju aktyviajame centre yra įmanomi keli molekules prisijungimo būdai,

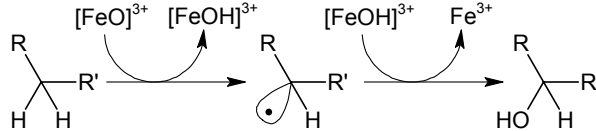
tačiau konformacinis paties centro lankstumas yra kur kas mažesnis, dėl to šis fermentas yra ko gero geriausiai aprašytas galimų substratų farmakoforų (angl. *pharmacophores*) požiūriu. Ilgą laiką bazinio azoto atomo ir nedideliu atstumu nuo jo nutolusios plokščios aromatinės sistemos buvimas molekulėje daugeliu atveju buvo laikomi būtiniais ir pakankamais CYP2D6 fermentų substratų ypatumais [76, 77]. Tai buvo grindžiama šių molekulės sudedamųjų dalių sudaromais vandenilniais ryšiais su atitinkamai Asp-301 ir Glu-216 fermento polipeptidinės grandinės amino rūgštimis, įeinančiomis į šio fermento aktyvųjį centrą [78, 79]. Tačiau su laiku vis daugėjo įrodymų, kad tokio tipo sąveikos nėra būtinos efektyviam kai kurių substratų metabolizavimui, pvz.: MPTP [80-82], testosterono [83] ir pan. Dėl šių priežasčių panašiu metu pradėjo atsirasti alternatyvios teorijos, aiškinančios šių aminorūgščių svarbą ne tik jų vaidmenimi prisijungiant ir orientuojant substratą fermento aktyviajame centre, bet ir dalyvavimu paties aktyviojo centro vientisumo bei topografijos palaikyme [79].

Kaip jau buvo minėta, CYP450 fermentai bendrai, bei CYP3A4 ir CYP2D6 konkrečiai, gali katalizuoti pačias įvairiausias chemines reakcijas žmogaus organizme [8, 68, 84]. Tačiau pagrindinės, dažniausiai sutinkamos ir tyrinėjamos reakcijos, kurioms literatūroje yra pakankamai eksperimentinių duomenų, apibūdinančių atitinkamų fermentų regioselektyvumą, yra keturios: alifatinis bei aromatinis hidroksilinimas ir N- bei O-dealkilinimas. 5 paveikslas schemiškai iliustruoja po vieną iš galimų kiekvieno paminėto reakcijos tipo mechanizmų. Šie keturi reakcijų tipai iš esmės aprėpia ir kai kurias kitas, pagal pavadinimą atrodytų nesusijusias, reakcijas bei leidžia panaudoti joms žinomus eksperimentinius duomenis. Tarkime alkoholių oksidacija tėra atskiras alifatinio hidroksilinimo atvejis, su po to sekančiu vicinalinio diolio persigrupavimu į aldehidą. Pastarojo oksidaciją iki rūgšties stochiometriškai taip pat atitinka vandenilio pakeitimas –OH grupe, t.y. hidroksilinimas. Aromatinių junginių epoksidinimo reakciją galima nagrinėti kaip atskirą atvejį kartu su aromatinio hidroksilinimo duomenimis.

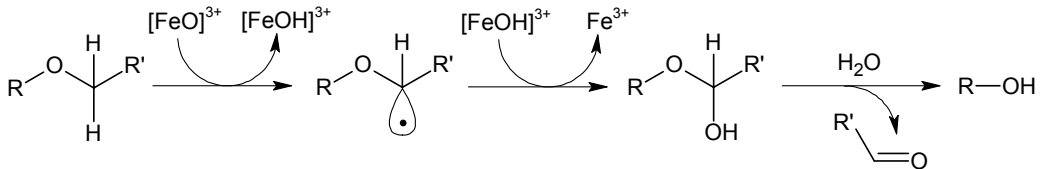
Aromatinis hidroksilinimas



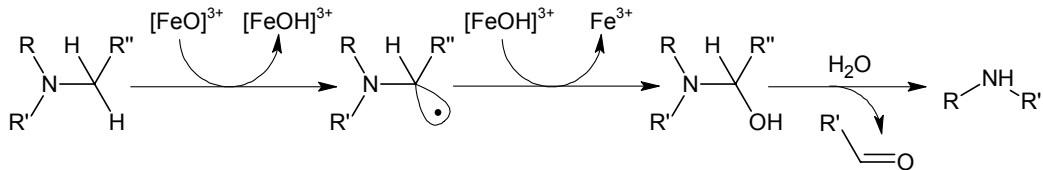
Alifatinis hidroksilinimas



O-dealkilinimas



N-dealkilinimas



5 pav. Pagrindinių CYP3A4 ir CYP2D6 fermentų katalizuojamų reakcijų tipų mechanizmų pavyzdžiai

1.2.2 Metabolizmo fermentų regioselektyvumo modeliai

In silico metodų taikymas tikslios metabolizmo vietos prognozavimui yra žymiai naujesnė sritis, lyginant su ūmaus toksiškumo modeliavimo darbais. Taigi bendras įvairių publikacijų skaičius, kurias galima būtų apžvelgti šioje dalyje taip pat yra ženkliai mažesnis. Tačiau to paties negalima pasakyti apie šiuose darbuose išmėgintų idėjų ir metodų įvairovę. Visus kompiuterinius mėginimus prognozuoti metabolizmo fermentų (absoliučia dauguma atvejų CYP450 didšeimės atstovų) tikslią atakos vietą molekulėje galime suskirstyti į tokius tipus:

- Pagrįsti nagrinėjamojo fermento struktūros analize (angl. *structure based*);
- Pagrįsti žinomų fermento substratų struktūrų ir savybių (reaktingumo, fiziko-cheminių savybių ir t.t.) analize (angl. *ligand based*);
- Pagrįsti įvairiais įmanomais, potencialius substratus aprašančių taisyklių rinkiniais (angl. *rule based*).

Kaip galima suprasti vien iš pavadinimo, fermento struktūros analize pagrįstuose metoduose visų pirma būtina žinoti tretinę nagrinėjamojo fermento ar bent jau jo aktyviojo centro struktūrą. Turint šią informaciją, įvairiais dokavimo (angl. *docking*) metodais pirmiausiai yra mėginama numatyti potencialaus substrato susijungimą su aktyviuoju centru ir jo orientaciją jame (t.y. iš principo yra įvertinama ar duotoji medžiaga gali efektyviai susirišti su konkrečiau nagrinėjamojo fermento aktyviuoju centru). Akivaizdu, jog dokavimo procedūrai atlikti taip pat reikalinga ir trimatė nagrinėjamojo substrato struktūra. Antrame etape jau yra mėginama nustatyti konkrečią metabolinio pakitimo molekulėje vietą. Tai dažniausiai atliekama lyginant skirtingų molekulėje esančių vandenilio atomų radikalinio atplėšimo energijas, paskaičiuojamas naudojant įvairius kvantmechaninius metodus. Labiausiai šiuo požiūriu vandenilio atomą turinti anglis yra laikoma labiausiai tikėtina metabolizmo vieta molekulėje, sekantį pagal radikalinio atplėšimo energijos dydį vandenilį prisijungęs anglies atomas – antra pagal svarbą metabolizmo vieta ir t.t. Tokia dokavimo ir kvantmechaninių metodų kombinacija naudojama keletą darbų [85, 86]. Vis dėlto, visos tiriamojo fermento struktūros analize paremtų metodų variacijos turi nemažai trūkumų. Visų pirma, tai šiuolaikinių dokavimo metodų netobulumas. Net ir preliminari šios srities problematiką nagrinėjančios literatūros analizė akivaizdžiai rodo, kad nuo pat tokių metodų atsiradimo praktiškai iki šių dienų, netyla ginčai dėl dokingo eksperimentuose naudojamų vertinimo funkcijų (angl. *scoring functions*). Šios funkcijos leidžia kiekybiškai įvertinti fermento ir substrato susijungimo kokybę bei palyginti ją su kitiems substratams gaunamais

analogiškais rezultatais. Šių funkcijų yra sukurta be galo daug, tačiau pagrindinė su tuo susijusi problema yra ne šis skaičius, o faktas, kad neretai skirtingos funkcijos taikymo tam pačiam fermentui ir substratui atvejais gaunami labai skirtingi ar net diametraliai priešingi rezultatai [87]. Be viso to, dokingo metodai neatsižvelgia (arba geriausiu atveju atsižvelgia minimaliai ir gana prastai) į fermentui ir jo aktyviajam centrui būdingą lankstumą, t.y. galimybę kisti laike bei susijungimo procese galinčias dalyvauti vandens molekules [88]. Net jeigu konkretaus metodo autoriai savo darbuose sugeba parodyti priimtina jo funkcionavimą, tokios patikros rezultatai dažniausiai nebūna vertingi praktiniu požiūriu ir negali būti laikomi modelio tinkamumo patvirtinimu, kadangi visa procedūra atliekama su junginiais per daug nutolusiais nuo kompanijų vidiniuose duomenų rinkiniuose (angl. *in-house data*) sutinkamų molekulių. Galų gale, netgi reikalingos nagrinėjamojo fermento aktyviojo centro konfigūracijos klausimas nėra visiškai trivialus. Baltymo tretinė struktūra nustatoma rentgeno struktūrinės analizės pagalba, kuriai atlikti reikalingas kristalinio būvio fermentas. Iškristalinto fermento aktyviojo centro konfigūracija jau gali būti skirtinga lyginant su ta pačia konfigūracija tirpale, kuriame aktyvusis centras apskritai yra laike kintanti struktūra, galinti individualiai prisitaikyti prie substrato molekulės. Tokio individualiai pritaikyto aktyviojo centro tretinę struktūrą galima išsiaiškinti analizuojant fermento ir konkretaus substrato komplekso kristalus, tačiau šie duomenys vėliau gali būti panaudoti nebent to paties substrato ir pačių artimiausių jo analogų dokinimui. Visos šios aplinkybės paaiškina kodėl kai kuriais atvejais skirtingi aktyviojo centro konfigūracijos tyrimai pateikia ganėtinai skirtingus duomenis. Bendru atveju, nei viena tokiu būdu gauta tretinė baltymo struktūra negali būti naudojama didele junginių įvairove pasižyminčių duomenų rinkinių analizei naudojant dokinimo metodus.

Vienas iš literatūroje sutinkamų variantų yra prielaida, kad tiksli metabolizmo vieta molekulėje daugiausia priklauso tik nuo skirtingų jos atomų reakingumo, leidžianti palikti tik kvantmechaninę struktūrinių metodų dedamąją ir tokiu būdu išvengti visų su dokinimo metodais susijusių problemų

[89] (šiuo atveju tai jau grynai substratų analize pagrįsti metodai). S. B. Singh su bendradarbiais savo darbe [90] papildomai įveda vandenilio atomo prieinamumo parametras, parodydami, jog iš visų pakankamai energetiškai labilių vandenilio atomų, lengviausiai metabolizuojami tie, kurie turi pakankamą tirpikliui prieinamą paviršiaus plotą (angl. *solvent accessible surface area*). Tokia prielaida yra teisinga tikrai iš dalies ir tai tik patiems nespecifiškiausiems fermentams, pvz., CYP3A4, taigi tokių metodų taikymas jau pagal apibrėžimą ne visada yra korektiškas. Be to, jau anksčiau aptartuose substratų analize pagrįstuose metoduose vis dar plačiai naudojami kvantmechaniniai skaičiavimai, labai stipriai ribojantys tokių metodų greitį.

Visiškai kitas būdas prognozuoti tikslią metabolizmo vietą molekulėje yra tradiciniai (Q)SAR tipo metodai, modeliavime naudojančios vien tikrai empirinius eksperimentinius duomenis. Iki šiol paskelbta keletas tokio tipo darbų, kuriuose metabolizmo reakcijos vieta ganėtinai sėkmingai prognozuojama pasinaudojant žinomų metabolizmo centrų pirštų antspaudais (angl. *fingerprints*) [91, 92]. Į molekulės susijungimui su fermentu būtinas savybes (tam tikri farmakoforai, fiziko-cheminių savybių ribos ir pan.) šiuo atveju atsižvelgiama pasinaudojant pačiais įvairiausiai kintamaisiais (molekulę sudarančiais fragmentais; dydžio, lipofiliškumo ir kt. deskriptoriais). Viename iš tokių darbų pateikiamas modelių tikslumo palyginimas rodo, kad bendru atveju iš empirinių (Q)SAR modelių galima tikėtis didesnio prognozavimo tikslumo nei iš analogiškų substratų analize pagrįstų modelių, naudojančių pusiau empirinius kvantmechaninius parametrus [92].

Atskirai būtina paminėti ir įvairiomis taisyklėmis pagrįstus tikslios metabolizmo vietos prognozavimo metodus [93]. Santykinai tai vienas elementariausių ir paprasčiausių molekulės metabolizmo nagrinėjimo būdų, todėl jais pagrįstos kompiuterinės programos, tokios kaip METEOR[®] [94], pasirodė vienos iš pirmųjų ir sėkmingai naudojamos iki šiol. Tokio tipo taisyklių rinkiniuose aprašyti pagal galimybes visi įmanomi molekulės struktūriniai fragmentai, kuriuose yra stebimos metabolinės reakcijos žinomiems nagrinėjamojo fermento substratams. Kitais žodžiais tariant, tokių

taisyklių taikymas konkrečiam junginiui leidžia gauti visų teoriškai įmanomų jo metabolitų sąrašą be galimybės jį suskirstyti į labiausiai tikėtinus ir visiškai nereikšmingus. Konkrečiai METEOR[®] programoje šiai problemai spręsti papildomai yra įdiegta „ekspertinė sistema“ (angl. *expert system*), kuri remdamasi didžiule į ją įvesta informacijos baze įgalina išrūšiuoti gaunamą visų metabolitų sąrašą pagal pastarųjų aktualumą. Šioje vietoje operuojama keletu anglišku terminų, įvardinančių skirtingą reakcijos tikėtimumo laipsnį. Nors verčiant į lietuvių kalbą, visi trys šie terminai išverčiami visiškai identiška, t.y. tikėtinas (angl. *possible, probable, plausible*), tačiau iš metodo autorių aiškinimų yra įmanoma susidaryti tam tikrą vaizdą apie jų santykinę tarpusavio padėtį tikėtimumo skalėje (*plausible > probable > possible*) [93]. Vis dėlto galima numanyti, jog toks neapibrėžtas klasifikavimas daugeliu atveju stipriai apsunkina tokio tipo prognozių praktinį panaudojimą.

Dar vienas iš nedaugelio šiuo metu komerciškai prieinamų metabolizmo vietos molekulėje prognozavimo programinių paketų yra MetaSite[®] [95, 96]. Teorinį pastarosios programos pagrindą sudaro metodas, pagal savo prigimtį labai artimas modeliams pagrįstiems nagrinėjamojo fermento struktūros analizei. Šiame modelyje taip pat yra įvertinama potencialaus substrato sąveika su jį metabolizuojančiu fermentu. Tačiau autoriai pabrėžia, kad tai nėra dokimimo eksperimentas [95] ir pastarajai analizei nėra reikalinga žinoti eksperimentiškai nustatytos tretinės baltymo struktūros. Taigi šiam metodui teoriškai nebūdingos įvertinimo funkcijos parinkimo ir kitos problemos. Metodas pagrįstas labiliomis molekulių sąveikos jėgomis, apskaičiuojamomis taikant GRID jėgos lauką CYP450 homologinio modeliavimo metu gautoms struktūroms. Taigi iš principo, šis modelis nepatenka į nei vieną aukščiau išvardintų tipų bei, vaizdžiai šnekant, yra vienintelis toks savo klasėje. Nepaisant visko, ir jam būdingi tam tikri trūkumai. Norint atlikti skaičiavimus MetaSite[®] programa yra reikalinga trimatė substrato struktūra. Tai visų pirma susiję su papildomais laiko iš kompiuterio galios reikalavimais, kylančiais iš būtinybės šią struktūrą optimizuoti. Be to, patys autoriai pripažįsta, kad gaunami rezultatai neretai gana stipriai priklauso nuo pradinio pasirinkto

išeities taško (molekulės konformacijos ir pan.) [95]. Akivaizdu, kad tai nėra pageidaujamas bet kokio metodo bruožas. Lygindami savo modelį su MetaSite[®], R. P. Sheridan ir kolegos [92] konstatuoja, kad patys programos autoriai teigia, jog jų pačių surinktame įvairių junginių rinkinyje eksperimentiškai stebima metabolizmo vieta yra tarp pirmų dviejų MetaSite[®] prognozuojamų molekulės atomų 78%, 86% ir 86% atvejų atitinkamai CYP3A4, CYP2D6 ir CYP2C9 fermentams. Kadangi minimas junginių rinkinys niekada nebuvo paskelbtas, straipsnio autoriai negalėjo nei patvirtinti, nei paneigti tokio teiginio. Savo pačių surinktą nepriklausomam patikrinamajam junginių rinkiniui analogiški jų gauti rezultatai buvo atitinkamai 62%, 72% ir 73%. Savo ruožtu darbe pasiūlytų QSAR modelių tikslumas CYP2D6 ir CYP2C9 atvejais buvo palyginamas su šiais MetaSite[®] rezultatais, tuo tarpu CYP3A4 fermentui jie buvo pastebimai geresni. Kita vertus MetaSite[®] naudojama koncepcija turi privalumų, nes jau dabar gali būti pritaikyta fermentams, kuriems literatūroje aptinkamų tikslios metabolizmo vietos eksperimentinių duomenų nepakanka tradiciniam QSAR modeliui sukurti, pvz.: CYP1A2 ar CYP2C19 (iš principo toks modeliavimo būdas gali būti pritaikytas bet kuriam CYP450 fermentui, tereikia sukurti atitinkamą homologinį jo struktūros modelį). Bet kokiu atveju, viską apibendrinanti išvada teigia, jog kadangi nei darbe pasiūlytas modelis, nei MetaSite[®] programa nesugebėjo paaiškinti daugiau nei maždaug 70% iš literatūros surinktų regioselektyvumo duomenų (be abejo, laikant, kad pastarieji yra teisingi), kyla įtarimas, kad visuose bandomuose modeliavimuose yra praleidžiama dalis svarbios informacijos ir šioje srityje dar yra daugybė erdvės darbui bei tobulėjimui.

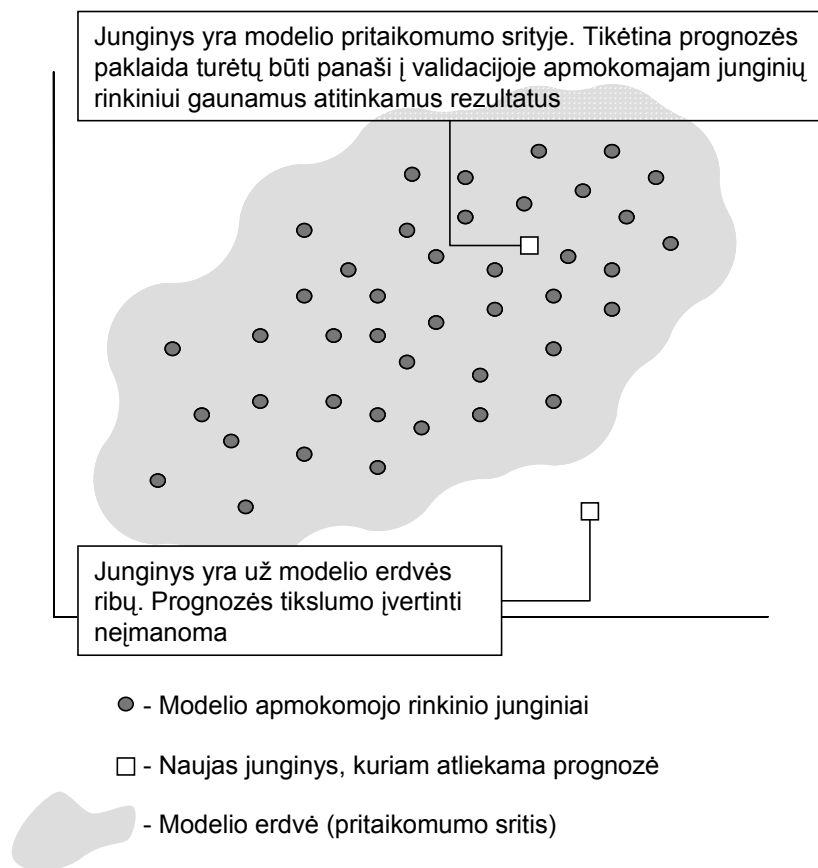
Kitas darbas [3], apžvelgiantis visą eilę įvairių dokingu ir kvantmechaniniais parametrais (kartu ir atskirai) paremtų modelių siūlo tokį požiūrį į literatūroje skelbiamų regioselektyvumo modelių praktinio panaudojimo galimybes. Vėlyvesnėse tyrimo pakopose, kuomet jau yra tiksliai identifikuota konkreti CYP450 izoforma, atsakinga už didžiosios dalies cheminio junginio metabolizmą ir pašalinimą iš organizmo, galima naudoti

lėtesnius, tačiau potencialiai galbūt kiek tikslesnius struktūrinius metodus, naudojančius informaciją apie aktyviojo centro struktūrą. Tačiau pačiuose ankstyviausiuose etapuose didelio našumo atrankai substratų analize paremti modeliai yra neabejotinai pranašesni, kadangi yra greitesni, be to, gali būti naudojami visiškai be jokių išankstinių žinių apie junginio metabolizmą. Tarp pastarųjų, grynai empiriniai metodai leidžia, ko gero paprasčiausiu ir greičiausiu būdu pasinaudoti nuolatos ypač dideliais tempais didėjančiais eksperimentinių citochromų P450 regioselektyvumo duomenų kiekiais. Minėtu aspektu labai gerai pasitarnauti turėtų šiame darbe panaudotos GALAS metodikos sugebėjimas kiekybiškai įvertinti prognozės patikimumą ir greitai praplėsti turimo modelio pritaikomumo erdvę paprasčiausiai pridedant naujus eksperimentinius duomenis prie lokalių korekcijų procedūros (t.y. neapmokant iš naujo bazinio modelio).

1.3 Modelio pritaikomumo sritis (erdvė)

Kiekvienas modelis, nepriklausomai nuo to, koku statistiniu metodu jis yra paremtas, ar koks kiekis ir kokių duomenų buvo panaudoti jo apmokymui, pasižymi tam tikra modelio pritaikomumo sritimi (angl. *Model Applicability Domain*), už kurios ribų bet kokių šio modelio prognozių patikimumas tampa labai abejotinas. Šią sritį sudaro visą įmanomą junginių cheminę įvairovę atspindinčios daugiamatės erdvės dalis, kurioje galima taikyti konkretų modelį, tikintis, kad prognozių paklaidų pasiskirstymas bus artimas apmokomojo rinkinio rezultatams. Visi empiriniai QSAR modeliai, skirtingai nuo mechanistinių, pagal apibrėžimą yra interpoliaciniai modeliai, t.y. skirti duomenų rekonstravimui tam tikro diskrečiai apibrėžto žinomų duomenų rinkinio ribose, ir jų galimybės prognozes ekstrapoliuoti naujiems junginiams yra pakankamai ribotos. Taigi, bet kurio tokio modelio pritaikomumo sritis visuomet yra ganėtinai glaudžiai susijusi su jo apmokomojo rinkinio junginiais. Bendru atveju, visada yra įmanoma atrasti junginių, kuriems riba, iki kurios tokią modelio ekstrapoliaciją dar galima laikyti bent kiek realia ir patikima, bus

peržengta. Vizualiai modelio pritaikomumo erdvės koncepcija pamėginta atvaizduoti 6 paveiksle.



6 pav. Scheminis modelio pritaikomumo erdvės ir į ją patenkančio bei už jos ribų esančio tiriamųjų junginių atvaizdavimas

Modelio veikimas už jo pritaikomumo erdvės ribų gali būti įdomus nebent grynai teoriniu požiūriu, tuo tarpu akivaizdu, kad norint praktiškai pasinaudoti modelio prognozėmis, informacija apie tai, ar tiriamasis junginys patenka į pritaikomumo sritį yra absoliučiai būtina. Būtent dėl šios priežasties, tarp OECD nustatytų QSAR modelių, kaip alternatyvių tyrimo metodų vertinimo kriterijų figūruoja ir reikalavimas, kad bet kuris šiam tikslui siūlomas modelis būtų susietas su aiškiai apibrėžta savo pritaikomumo sritimi [97]. Pastaroji oficialiuose dokumentuose apibrėžiama kaip „atsako ir cheminių struktūrų erdvė, kurioje modelis pateikia atitinkamo tikslumo prognozes“ [98]. Visiškai neseniai Europos Sąjungoje įsigaliojus REACH

direktyvai [99], įpareigojančiai įvertinti visų bendrijoje gaminamų ar į ją importuojamų cheminių medžiagų toksiškumą ir gyvybei keliamą pavojų, buvo rimtai susirūpinta tokių reikalavimų pasekoje drastiškai išaugšiančiu gyvūnų naudojimu *in vivo* toksiškumo testuose. Ši situacija leidžia tikėtis išaugšiančio alternatyvių tyrimo metodų, įskaitant ir *in silico* QSAR modelius, poreikio [41, 43]. To pasekoje buvo netgi įkurta atskira ECVAM institucija, užsiimanti *in vitro* ir *in silico* metodų, galinčių pakeisti eksperimentinius matavimus su gyvūnais, vertinimu, remiantis tame tarpe jau minėtais OECD iškeltais kriterijais. Tai savo ruožtu stipriai padidino susidomėjimą modelio pritaikomumo srities klausimais ir išskėlė šią sritį leidžiančių įvertinti metodų kūrimą kaip vieną aktualiausių problemų šiuolaikinių QSAR modeliavimų srityje. ECVAM veikloje dalyvaujanti J. Jaworska kartu su bendradarbiais, savo darbe siūlo šią problemą spręsti apibrėžiant deskriptorių erdvę, kurioje išsidėstę apmokymo rinkinio junginiai – tam pasitelkiami įvairūs atstumų, rėžių ir tikimybinio pasiskirstymo metodai [100]. S. Dimitrov ir kt. publikacijoje [101] aprašomas pakopinis modelio pritaikomumo erdvės įvertinimas, paeiliui atsižvelgiantis į bendrą junginio savybių erdvę, struktūrinių savybių erdvę, mechanistinę erdvę ir metabolinio modeliavimo erdvę. Tačiau netgi patys pastarosios metodikos autoriai pripažįsta, jog kai kurie iš šių žingsnių yra sunkiai apibrėžiami ir standartizuojami.

Toksiškumo modeliavimuose modelio pritaikomumo erdvės sąvoka taip pat figūruoja ir ji nėra labai nauja. Dar 1995 metais šioje srityje buvo pristatyta OPS metodika [102], prie kurios sukūrimo prisidėjo ir jau minėtasis K. Enslein – ankstyvųjų LD_{50} modelių, tapusių TOPKAT[®] programos pagrindu, autorius. Taigi nenuostabu, jog ši metodika su laiku taip pat buvo įtraukta kaip minėtosios programos dalis, ir OPS įvertinimas joje yra išlikęs iki šių dienų. Vis dėlto OPS metodika nebuvo pilnai paviešinta, o kai kurios jos dalys buvo netgi patentuotos, kas savo ruožtu neleido jai išpopuliarėti ir atrasti savo vietos kituose QSAR modeliuose. Didžiulis dėmesys modelių pritaikomumo erdvės įvertinimui skiriamas ir šioje apžvalgoje jau minėtame tyrime, susijusiame su ūmaus toksiškumo modeliavimais [36]. Kiekvienam iš darbe pristatomų

individualių LD_{50} reikšmių modelių, pagal jo specifiką buvo parinktas būdas, nusakantis to modelio pritaikymo galimybes konkrečiam tiriamajam junginiui. Šioje vietoje operuojama kriterijais varijuojančiais nuo tokių elementarių kaip prognozuojamos LD_{50} vertės patekimas į apmokomojo rinkinio junginiam būdingų ūmaus toksiškumo verčių intervalą iki junginio padėties daugiamačiame apmokomojo rinkinio deskriptorių erdvę atspindinčiame elipsoide. Klasteriavimo (angl. *clustering*) analize paremtuose modeliuose tikrinama ar junginio atstumas nuo jam priskirto klasterio centro nėra didesnis už analogišką atstumą bet kurio kito tam klasteriui priklausančio apmokomojo rinkinio junginio atveju, ir t.t. Visus pastarajame darbe panaudotus, kaip ir iki šiol čia apžvelgtus, modelio pritaikomumo erdvės nustatymo metodus sieja viena savybė – juose vienokiu ar kitokiu būdu yra įvertinamas tiriamojo junginio panašumas į apmokomojo rinkinio molekules (grynai cheminis panašumas, panašumas deskriptorių ar nagrinėjamosios savybės erdvėje). Nors tai, be abejo, yra vienas iš pagrindinių kriterijų, leidžiančių tikėtis patikimos modelio prognozės, tačiau, kaip jau netrukus bus aptarta šiame darbe, jisai nėra vienintelis ir pakankamas, ypač tokių sudėtingų modeliuojamųjų objektų, kaip ūmus toksiškumas ar tuo labiau metabolizmo fermentų regioselektyvumas atveju. Be to, nė vienas iš aprašytų modelio pritaikomumo erdvės įvertinimo metodų nėra neatsiejama integruota atitinkamo modelio dalis. Visos statistinės ar kitokios manipuliacijos, leidžiančios įvertinti konkretaus modelio taikymo tiriamajam junginiui galimybę, atliekamos nepriklausomai nuo modeliavimo proceso ir geriausiu atveju yra susijusios tik su modelio tipu. Tai iš principo leidžia aprašytąsias metodikas naudoti visiems atitinkamo tipo modeliams, ką galima laikyti privalumu, tačiau kartu tai yra ir trūkumas. Viena vertus, turint omenyje milžinišką įvairiausių įmanomų netgi to paties tipo modelių įvairovę, labai sunku tikėtis, kad apskritai yra įmanoma universali pritaikomumo erdvės įvertinimo metodika, patikimai veikianti visai atvejais. Kita vertus, pasakymas „visiems modeliams“ gali nuskambėti apgaulingai, kadangi norint pritaikyti kažkurią iš minimų metodikų yra reikalingos išsamios žinios apie patį modelį –

apmokomojo rinkinio junginiai, naudoti deskriptoriai, kartais netgi eksperimentinės reikšmės, o ši informacija toli gražu ne visuomet yra prieinama, ypač komerciniams modeliams. Na ir galiausiai, visi paminėti modelio pritaikomumo erdvės įvertinimo metodai yra binominiai (kokybiniai), t.y. nubrėžia griežtą ribą tarp į modelio pritaikomumo erdvę patenkančių ir iš jos iškrentančių junginių. Tuo tarpu akivaizdu, kad junginiams esantiems pritaikomumo srities centre ir jos pakraštyje, apsuptis panašiais apmokomojo rinkinio junginiais bei atitinkamai modelio prognozės patikimumas gali gerokai skirtis. Dviprasmiškų situacijų tikimybė egzistuoja ir nagrinėjant junginius, esančius ties modelio pritaikomumo srities riba. Esant griežtai ribai, du pakankamai panašūs junginiai, esantys ant ribos, tokių metodikų gali būti traktuojami kaip visiškai skirtingi. Vieno iš jų atveju modelio taikymas gali būti nurodytas kaip galimas, kito – ne. Tuo tarpu, logiškai maštant, modelio prognozės jiems abiem turėtų būti traktuojamos vienodai, t.y. vienodai mažai patikimos.

Pabaigoje galima pridurti, kad situacija su modelio pritaikomumo erdvės įvertinimu darbuose susijusiuose su ūmaus toksiškumo ir metabolizmo regioselektyvumo modeliavimais yra visiškai skirtinga. Nė vienas iš šioje apžvalgoje aptartų tikslios metabolizmo vietos molekulėje modelių apie šią problemą neužsimena. Šiame kontekste, sėkmingas GALAS modeliavimo metodikos pritaikymas, be abejo, būtų reikšmingas pasiekimas tiek ūmaus toksiškumo, tiek CYP3A4 ir CYP2D6 fermentų regioselektyvumo modeliavimuose, tačiau pastaruoju atveju tai būtų apskritai naujas žingsnis į priekį tokio tipo tyrimuose. GALAS metodikoje įdiegtas modelio pritaikomumo erdvės įvertinimo mechanizmas yra sudedamoji modelio dalis, operuojanti kai kuriais galutinių prognozių apskaičiavimui naudojamais parametrais. Šis metodas įvertina ne tik tiriamojo junginio panašumą į apmokomąjį rinkinį, bet ir duomenų tolydumą modelio atžvilgiu panašiausiems junginiams bei pateikia kiekybinį prognozės patikimumo įvertinimą (*RI*), leidžiantį diferencijuoti junginius pagal jų „atstumą“ nuo modelio pritaikomumo srities centro.

2 METODINĖ DALIS

2.1 Duomenys ir programinė įranga

2.1.1 Toksiškumo (LD_{50}) duomenų bazė

Šiame darbe naudotas ūmaus toksiškumo graužikams LD_{50} verčių rinkinys buvo pagrindinai surinktas iš RTECS[®] duomenų bazės [12]. Pastaroji buvo kruopščiai peržiūrėta siekiant užtikrinti maksimalią pradinių modeliavimo duomenų kokybę. Iš rinkinio buvo pašalinti visi pasitaikę nekovalentiniai kompleksai, organinių junginių mišiniai ar druskos ir junginiai su neteisingomis struktūrinėmis formulėmis (išskyrus tuos atvejus, kuomet buvo įmanomas greitas ir daug pastangų nereikalaujantis jų ištaisymas). Be viso to, taip pat buvo stengtasi pagal galimybes identifikuoti eksperimentines LD_{50} nustatymo klaidas. Akivaizdu, kad to padaryti praktiškai neįmanoma turint tikrai vieną LD_{50} reikšmę, susijusią su konkrečiu junginiu. Tačiau tuo atveju, kuomet tai pačiai molekulei yra žinomi keli ūmaus toksiškumo matavimo rezultatai skirtingoms graužikų rūšims ar įvedimo į organizmą keliams, galima remtis keletu elementarių loginių prielaidų. Pavyzdžiui, cheminio junginio kiekis reikalingas sukelti ūminį apsinuodijimą įvedant jį į organizmą intraveniniu keliu visada yra žymiai mažesnis palyginus su kiekiu sukeliančiu analogišką efektą esant oraliniam patekimui į organizmą. Tai nulemia ta paprasta aplinkybė, kad intraveniniu keliu patekdamas tiesiai į kraujotakos sistemą junginys aplenkia visą eilę organizmo gynybinių barjerų, įtakančių medžiagos oralinį įsisavinimą (angl. *oral bioavailability*) ir tokie faktoriai kaip tirpumas, absorbcija žarnyne, pašalinimas per P-gp baltymus, pirmojo praėjimo metabolizmas (angl. *first-pass metabolism*) ir kiti tampa nebeaktualūs. Įvairūs pasvarstymai galimi ir jei kalba eina apie to paties junginio LD_{50} reikšmes esant identiškiems įvedimo būdams, tačiau skirtingoms graužikų rūšims. Iš vienos pusės akivaizdu, kad šiame darbe nagrinėjamos pelė ir žiurkė yra skirtingos žinduolių rūšys turinčios kiekviena savo ypatumų (pvz., metabolinių fermentų įvairovė, jų kiekiai bei aktyvumas) ir jų jautrumas

cheminėms medžiagoms yra neišvengiamai skirtingas. Kita vertus literatūroje galima aptikti visą eilę tarprūšinių koreliacijų, kurios iš pirmo žvilgsnio rodo netgi stebinantį pelės ir žiurkės ūmaus toksiškumo atsako panašumą esant tam pačiam junginio patekimo į organizmą keliui [48]:

$$\log(1/LD_{50} \text{ žiurkė OR}) = 0,97 \cdot \log(1/LD_{50} \text{ pelė OR}) - 0,04 \quad (8)$$

$N = 330; R = 0,94; R^2 = 0,88; sd = 0,30$

$$\log(1/LD_{50} \text{ žiurkė IP}) = 1,02 \cdot \log(1/LD_{50} \text{ pelė IP}) - 0,02 \quad (9)$$

$N = 162; R = 0,96; R^2 = 0,92; sd = 0,28$

$$\log(1/LD_{50} \text{ žiurkė IV}) = 0,99 \cdot \log(1/LD_{50} \text{ pelė IV}) - 0,10 \quad (10)$$

$N = 41; R = 0,97; R^2 = 0,94; sd = 0,29$

Vis dėlto, derėtų atkreipti dėmesį, kad visos jos gautos labai ribotiems duomenų rinkiniams, todėl stipriai idealizuoja tikrąją situaciją, kadangi žinomas ne vienas junginys, kurio ūmus toksiškumas pelei ir žiurkei skiriasi kelis ar net keliolika kartų.

Apibendrinant galima pažymėti, kad nors šios dvi rūšys be abejo yra skirtingos, jos vis dėlto nėra nutolusios tiek, kad to paties junginio LD_{50} reikšmės pelei ir žiurkei, esant identiškam patekimo į organizmą keliui, besiskiriančios šimtus ar daugiau kartų, nekeltų visiškai jokių abejonių. Taigi iš galutinės duomenų bazės taip pat buvo pašalinti tokiu būdu nustatyti junginiai (lyginant tas pačias LD_{50} reikšmes skirtingoms rūšims ar įvairias LD_{50} reikšmes tai pačiai rūšiai), akivaizdžiai iškrentantys iš bendrų dėsningumų konteksto.

Tais atvejais, kuomet konkretus junginys būdavo taip pat įtrauktas į IUCLID duomenų bazę, jo cheminėje specifikacijoje (angl. *chemical data sheet*) esanti informacija susijusi su ūmiu toksiškumu būdavo lyginama su RTECS[®] duomenimis ir tokiu būdu panaudojama galutinio duomenų rinkinio įrašų patvirtinimui, pataisymui ar pašalinimui. IUCLID duomenų bazė (prieinama internetu per Europinę informacijos apie chemines medžiagas sistemą – ESIS) pateikia 2604-ių dideliais kiekiais gaminamų (HPVC) cheminių junginių duomenis, surinktus iš Europos pramonės įmonių, vadovaujantis šiuo metu egzistuojančiomis cheminio pavojaus įvertinimo (angl. *chemical risk assessment*) procedūromis [11]. Be viso kito, galutinis

LD_{50} verčių rinkinys buvo papildytas iš šios duomenų bazės paimtais keletu junginių, nesančių RTECS®.

Gautas galutinis toksiškumo duomenų rinkinys buvo sudarytas iš daugiau nei 75000 junginių. Kiekvienam iš šių junginių buvo nurodytos LD_{50} reikšmės bent vienai iš nagrinėjamų sistemų (ūmus toksiškumas pelėms esant oraliniam, intraperitoniniam, intraveniniam ar poodiniam įvedimo būdui, ir ūmus toksiškumas žiurkėms esant oraliniam ar intraperitoniniam įvedimo būdui). Prieš pradėdant modeliavimą, surinktosios LD_{50} reikšmės buvo pervestos į logaritminę skalę ($\log LD_{50}$) ir kiekvienas iš šešių atskirų sistemų duomenų rinkinių buvo atsitiktiniu būdu suskaidytas į apmokomąjį (70%) bei patikrinamąjį (30%) rinkinius. Šiame darbe naudoto toksiškumo duomenų rinkinio struktūra ir sudėtis yra apibendrinta 3 lentelėje.

3 lentelė. Darbe naudotų ūmaus toksiškumo (LD_{50}) duomenų rinkinių dydžiai

Rūšis	Įvesties būdas	N		
		Viso	Apmokomasis Rinkinys	Patikrinamasis Rinkinys
Žiurkė	Oralinis (OR)	8631	6464	2167
	Intraperitoninis (IP)	5002	3751	1251
Pelė	Oralinis (OR)	19571	14678	4893
	Intraperitoninis (IP)	36031	27004	9027
	Intraveninis (IV)	19963	14972	4991
	Poodinis (SC)	8577	6432	2145

Pastaruoju metu pasirodęs ūmaus toksiškumo prognozavimo žiurkėms po oralinio medžiagos įvedimo tyrimas [36] tapo dar vienu vertingu šio tipo duomenų šaltiniu. Iš šio straipsnio autorių buvo gauta 7385 junginių duomenų bazė su struktūromis ir eksperimentinėmis LD_{50} (žiurkė, OR) vertėmis†.

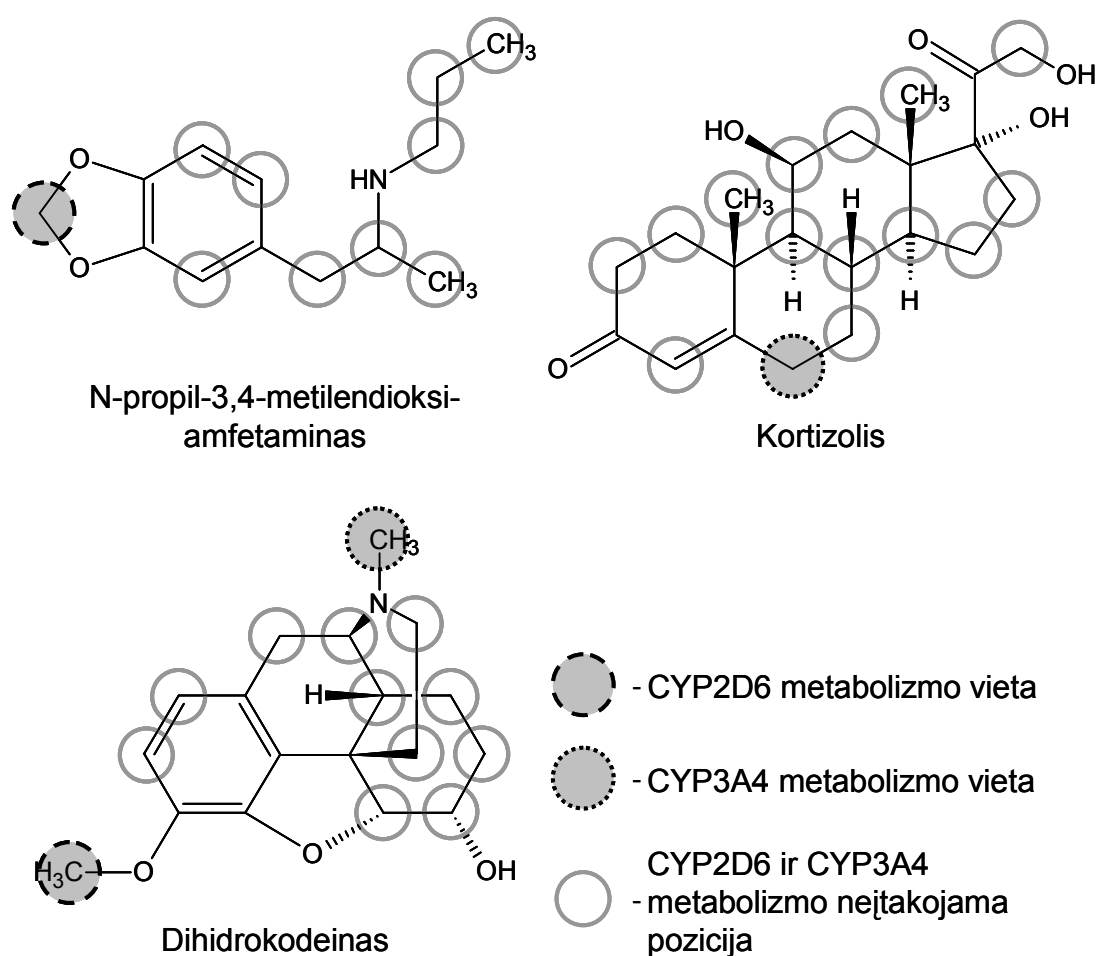
† Disertacijos autorius norėtų išreikšti padėką visiems pastarojo straipsnio [36] autoriams už suteiktą galimybę šiame darbe pasinaudoti jų surinkta ūmaus toksiškumo verčių duomenų baze bei atskirai iš viso jų būrio paminėti Alexander Tropsha bei Todd. M. Martin, su kuriais šiuo klausimu buvo susirašinėjama tiesiogiai.

Palyginus minimą duomenų bazę su šioje disertacijoje pristatomo analogiško modelio apmokomuoju rinkiniu buvo aptikta 2718 naujų junginių. Šie papildomi junginiai sudarė prielaidas atlikti išorinę šio modelio patikrą.

2.1.2 Metabolizmo regioselektyvumo duomenys

Metabolizmo fermentų CYP3A4 ir CYP2D6 regioselektyvumo modeliai buvo kuriami remiantis 681 junginio eksperimentiniais duomenimis, surinktais iš mokslinių publikacijų, kurių absoliučią daugumą sudarė moksliniai straipsniai detaliam nagrinėjantys vieno konkretaus junginio metabolizmo tyrimų rezultatus. Vienas pagrindinių literatūroje sutinkamų junginio metabolitų identifikavimo būdų yra inkubacija su grynu rekombinantiniu fermentu (šiuo konkrečiu atveju CYP3A4 ar CYP2D6) ir tolesnė susidariusio mišinio analizė [103]. Kitas populiarus būdas yra žmogaus kepenų mikrosomose (t.y. veikiant visiems įmanomiems CYP450 šeimos fermentams) susidarančių metabolitų spektro normaliomis sąlygomis ir esant tipinių vieno konkretaus fermento inhibitorių, pavyzdžiui, ketokonazolio [104-106], ritonaviro [106, 107] ir kitų, CYP3A4 ar chinidino [108, 109], fluoksetino [110, 111] ir kitų, CYP2D6 atveju, lyginimas [103]. Analizuojant išvardintus eksperimentinius duomenis metabolizmą aprašantis kintamasis buvo formaliai priskiriamas individualiems molekulių atomams, t.y. šiuo atveju būtent jie, o ne pačios molekulės yra duomenų rinkinio sudėtiniai vienetai. Kiekvienoje molekulyje visi anglies atomai, turintys bent vieną prijungtą vandenilį, buvo pažymėti kaip metabolizmo vietos, jei atitinkamoje pozicijoje buvo stebimas CYP3A4 ar CYP2D6 katalizuojamas hidroksilinimas. Azoto ir deguonies dealkilinimo reakcijų atveju analogiškai buvo sužymėti nueinančiųjų grupių anglies atomai, sudarantys nutrūkstantį ryšį su heteroatomu. Visi kiti anglies atomai, turintys ryšį su vandeniliais, buvo pažymėti kaip nemetabolizuojamos pozicijos. 7 paveiksle pateikiama keletas molekulių atomų žymėjimo, remiantis eksperimentiniais rezultatais, pavyzdžių. N-propil-3,4-metilendioksiamfetamino ir dihidrokodeino molekulėse yra stebimas CYP2D6 katalizuojamas O-dealkilinimas [112, 113]. Pastarajame junginyje taip pat

nustatyta CYP3A4 N-dealkilinimo reakcija [113], tuo tarpu kortizolio molekulėje veikiant šiam fermentui vyksta anglies atomo hidroksilinimas [114, 115]. Kiekvienas tokiu būdu pažymėtas „teigiamas“ ar „neigiamas“ atomas duomenų rinkinyje buvo įtrauktas kaip atskiras įrašas. Pavyzdžiui, CYP3A4 metabolizmo regioselektyvumo duomenų rinkinyje N-propil-3,4-metilendioksiamfetaminas, kortizolis ir dihidrokodeinas yra atspindėti sąrašais iš atitinkamai 10 (visi „neigiami“), 15 (1 „teigiamas“ ir 14 „neigiamų“) ir 13 (1 „teigiamas“ bei 12 „neigiamų“) formaliai sužymėtų atomų.



7 pav. Metabolizmo vietų žymėjimo molekulėje pavyzdžiai

Tokiu būdu formaliai sužymėjus „teigiamas“ ir „neigiamas“ CYP3A4 bei CYP2D6 metabolizmo vietas atitinkamai 560 ir 526 junginiams, abiem atvejais gautos maždaug 5500 atomų duomenų bazės. Atsižvelgiant į pažymėtų anglies atomų tipą ir padėtį molekulėje, šie rinkiniai buvo suskaidyti į atskirus

poaibus, susijusius su atitinkamais metabolizmo reakcijų tipais (N-dealkilinimas, O-dealkilinimas, alifatinis ir aromatinis hidroksilinimas) [84]. Kiekvieno fermento ir reakcijos tipo atveju gautų duomenų rinkinių dydžiai pateikiami 4 lentelėje. Kiekvienas iš jų buvo toliau atsitiktiniu būdu suskaidytas į apmokomąjį (70%) ir patikrinamąjį (30%) rinkinius, kurie buvo panaudoti atitinkamų individualių modelių sudarymui bei patikrai.

4 lentelė. Darbe naudoto CYP3A4 ir CYP2D6 metabolizmo regioselektyvumo duomenų rinkinio struktūra

Rinkinys	<i>N</i>	Metabolizuojamų atomų skaičius	Bendras pažymėtų atomų skaičius
CYP3A4			
N-dealkilinimas	345	199	824
O-dealkilinimas	310	104	688
Alifatinis hidroksilinimas	475	172	2057
Aromatinis hidroksilinimas	462	111	2087
CYP2D6			
N-dealkilinimas	357	73	849
O-dealkilinimas	296	61	610
Alifatinis hidroksilinimas	456	27	1901
Aromatinis hidroksilinimas	458	86	2110

2.1.3 Programinė įranga

Šiame darbe aprašytuose tyrimuose buvo plačiai naudojamos keletas specializuotų kompiuterinių programų. Darbui su duomenų bazėmis, modelių kūrimui reikalingų fragmentinių deskriptorių generavimui (molekulių fragmentavimui) buvo naudojama Algorithm Builder[®] programa [116, 117]. Šios programos statistikos pakete įdiegtais PLS [118], BPLS, ir GALAS [6, 7] metodais buvo naudojamosi atitinkamai ūmaus toksiškumo ir metabolizmo regioselektyvumo duomenų statistinei analizei bei modelių sudarymui. Metabolizmo regioselektyvumo modelių pritaikomumo srities išplėtimo eksperimentuose, sukurtųjų modelių apmokymas naujais junginiais buvo

atliekamas naudojantis jų versijomis įgyvendintomis ACD/ADME Suite[®] programoje [119].

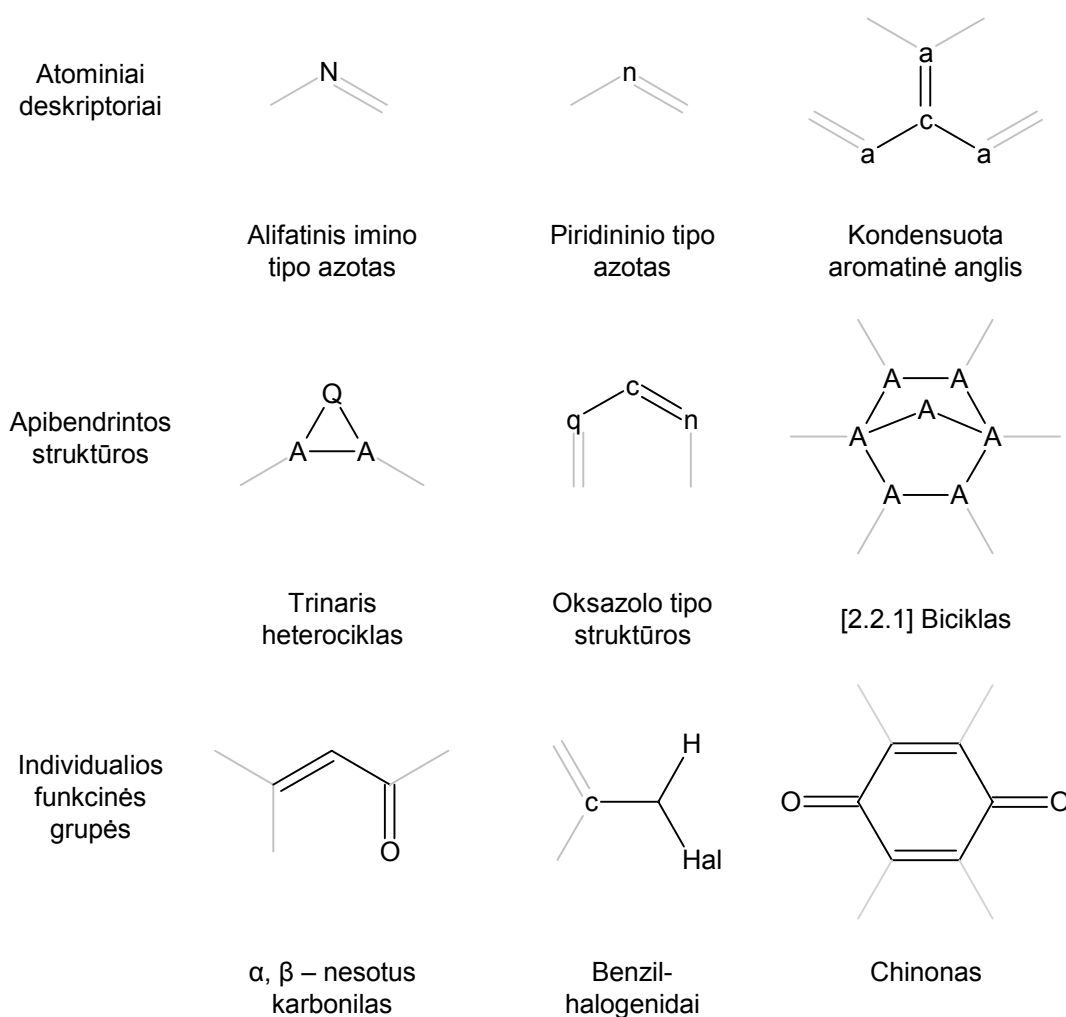
2.2 *Deskriptorių generavimas*

2.2.1 **Fragmentiniai deskriptoriai ūmaus toksiškumo (LD_{50}) modeliavimui**

Ūmaus toksiškumo duomenų bazę sudarančios molekulės buvo aprašytos naudojant fragmentinių deskriptorių rinkinį, kurį galima būtų traktuoti kaip sudarytą iš dviejų esminių dalių, atliekančių skirtingas funkcijas. Pagrindinę dalį sudarė įprastiniai dažname modelyje naudojami fragmentiniai deskriptoriai, tokie kaip atomai, funkcinės grupės, molekulės formą nusakantys fragmentai ir kiti. Šis fragmentų sąrašas buvo gautas išplečiant J. A. Platts'o darbe [120] naudotą analogišką fragmentų rinkinį [6]. Keletas tokio tipo fragmentų pavyzdžių iš šio sąrašo pateikiami 8 paveiksle. Šių fragmentinių deskriptorių naudojimas suteikia modeliui informaciją apie tokias esmines junginio savybes kaip jo dydis (bendras fragmentų skaičius ir aptinkamų fragmentų dydis), šakotumas, priklausymas atskiroms klasėms, pavyzdžiui, tai neutralus ar besijonizuojantis (jeigu pastarasis, tai ar rūgštis, ar bazė, ar amfoterinis), alifatinis ar aromatinis junginys (jeigu pastarasis, tai ar heterociklinis) ir daugybę kitų – trumpiau tariant, bendrąją organinio junginio cheminę sudėtį ir struktūrą.

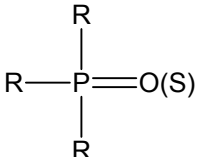
Šis bendrųjų fragmentų sąrašas buvo papildytas tam tikrais specifiniais fragmentais – vadinamaisiais toksikoforais (angl. *toxicophores*). Toksikoforai – tai molekulės struktūros fragmentai, atsakingi už juos turinčio junginio toksinio veikimo mechanizmą ir lemiantys žymų jo LD_{50} reikšmės sumažėjimą lyginant su labai artimomis, tačiau toksikoforo neturinčiomis medžiagomis. Šia fragmentinių deskriptorių rinkinio dalimi buvo siekta atsižvelgti bent jau į pačias žinomiausias ir pavojingiausias sąveikas organizme bei toksinio poveikio mechanizmus, nulemiančius stipraus ūminio apsinuodijimo pasireiškimą. Keletas tipinių toksikoforų pavyzdžių aprašomų literatūroje išvardinta 5 lentelėje. Disertacijoje pristatomų ūmaus toksiškumo modelių

kūrimui naudotų toksikoforų išsamus sąrašas ir papildoma šiuos fragmentus apibūdinanti statistinė informacija, pateikiami prieduose Nr.1 ir Nr.2.

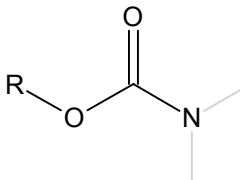
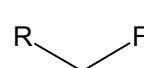
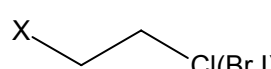
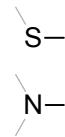
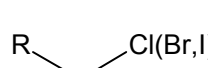
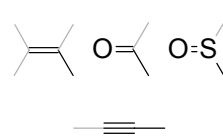

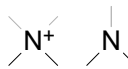
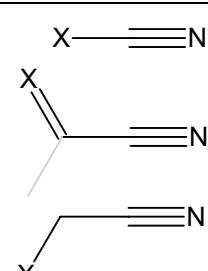
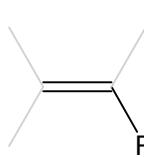
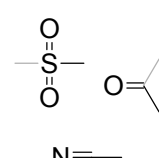


8 pav. Ūmaus toksiškumo modeliavimui naudotų bendrųjų Platts'o tipo fragmentų pavyzdžiai. Čia A – bet koks atomas, a – bet koks aromatinis atomas, Q – bet koks heteroatomas (N, O, S), q – bet koks aromatinis heteroatomas, Hal – bet koks halogenas, n – aromatinis azotas, c – aromatinė anglis

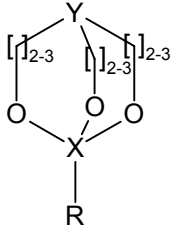
5 lentelė. Ankstesnėse studijose [15, 24, 117, 121-126] identifikuotų ir apibendrintų toksikoforų pavyzdžiai

Junginių klasė	Toksikoforas	Radikalas (R, X ar Y)	Apibūdinimas
(Tio)fosfatai		Išskyrus HO– H ₂ N– HS–	Cholinesterazės inhibicija. Efektas pastebimai sustiprėja, jei R yra teigiamai įkrautas [121]

5 lentelė. Ankstesnėse studijose [15, 24, 117, 121-126] identifikuotų ir apibendrintų toksikoforų pavyzdžiai (tęsinys)

Junginių klasė	Toksikoforas	Radikalas (R, X ar Y)	Apibūdinimas
Karbamatai		Aromatinis arba oksimo tipo	Cholinesterazės inhibicija. Efektas pastebimai sustiprėja, jei pakaitas prie azoto atomo yra -CH ₂ R [121]
Metileno fluoridai		Lyginis anglies atomų skaičius	Krebso ciklo inhibicija, fluorocitrato kaupimasis ir akonitazės inhibicija [122]
Garstyčių dujų dariniai			Makromolekulių alkilinimas [123]
Aktyvuoti metileno halogenidai			Makromolekulių alkilinimas [24]
Aziridino ir aziridinio dariniai			Makromolekulių alkilinimas [24, 124]
Aktyvuoti nitrilai		Bet koks heteroatomas	Cianido išsiskyrimas ir kvėpavimo grandinės slopinimas [24, 125]
Aktyvuoti dvigubieji ryšiai			Alkilinimas pagal Michaelio prisijungimo reakcijos mechanizmą [24]

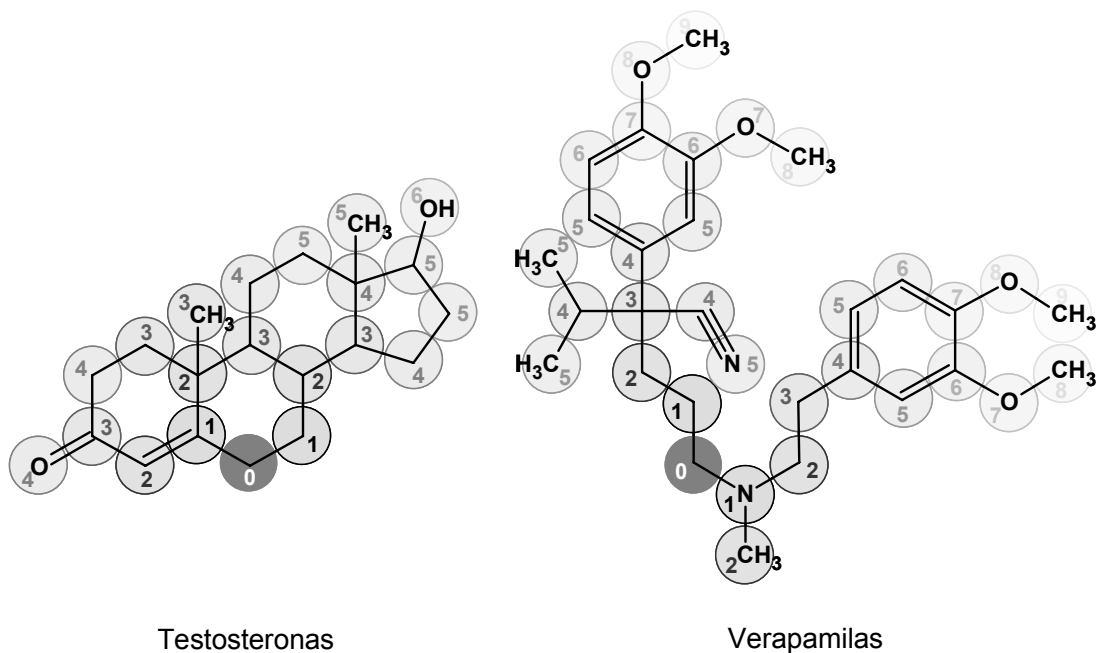
5 lentelė. Ankstesnėse studijose [15, 24, 117, 121-126] identifikuotų ir apibendrintų toksikoforų pavyzdžiai (tęsinys)

Junginių klasė	Toksikoforas	Radikalas (R, X ar Y)	Apibūdinimas
Biciklofosfatai, biciklo- ortokarboksilatai, silatranai		X – P, C, Si; Y – N, C; R – aromatinis žiedas arba alkinas	Nekonkurentinė GABA receptoriaus pikrotoksinino prisijungimo centro inhibicija [126]

2.2.2 Sluoksninė atomo-centrinė fragmentacija

Kaip jau buvo paminėta duomenų rinkinių aprašyme (žr. 2.1.2 skyrių) modeliuojant metabolizmo fermentų regioselektyvumą šiame darbe stebima savybės reikšmė buvo formaliai priskirta kiekvienam atskirai paimtam molekulės atomui. Dėl šios priežasties regioselektyvumo atveju negali būti panaudota nei viena iš daugybės molekulės fragmentacijos schemų sukurtų vadinamųjų „visos molekulės“ (angl. *whole-molecule*) savybių analizei (pvz., $\log P_{O/W}$). Visų šių schemų paskirtis yra generuoti vienintelį ir vienareikšmį molekulės skaitmeninį atvaizdavimą naudojamųjų deskriptorių (kintamųjų) matricoje. Tuo tarpu šioje situacijoje yra reikalingas būdas gauti visą eilę skirtingų ir unikalių struktūros aprašymų, priklausančių nuo pasirinkto konkrečiai mus dominančio atomo molekulėje. Šiam tikslui buvo pasitelktas specialiai sukurtas atomo-centrinės fragmentacijos metodas, suteikiantis informaciją apie pasirinktame sąraše esančių atomų ir fragmentų buvimą pozicijose (vadinamose sluoksniais), vienodai nutolusiose nuo pasirinkto atomo. Pastarasis, savo ruožtu, duomenų bazėje yra pažymėtas kaip teigiamas ar neigiamas metabolizmo centras. Kaip matyti iš scheminės šio fragmentacijos metodo iliustracijos 9 paveiksle, sluoksnių numeracija molekulėje atliekama bangos algoritmo pagalba.

Pasiūlytas sluoksninės fragmentacijos metodas pagrindine savo idėja yra labai panašus į vadinamąjį HOSE kodą (angl. *HOSE code*) paskelbtą dar 1978 metais [127]. Generuojant šį kodą taip pat atsižvelgiama į atomus,



9 pav. Keletas atomo-centrinės fragmentacijos sluoksnių numeravimo pavyzdžių

esančius skirtinguose sluoksniuose nuo pasirinkto fragmentacijos centro. Šiuo atveju kiekvienam tolesniam sluoksniui charakterizuoti yra panaudojama informacija apie visus prieš tai esančius sluoksnius. Pavyzdžiui, šiuo principu aprašant fragmentą $R-CH_2-CH_2-Cl$ (atskaitos tašku pasirinkus radikalą R), gautajame HOSE kode būtų įtraukta tokia informacija: 1 sluoksnis – $-CH_2-$ grupė; 2 sluoksnis – $-CH_2-$ grupė prijungta per pirmajame sluoksnyje esančią kitą $-CH_2-$ grupę; 3 sluoksnis – Cl atomas, sujungtas su kodo atskaitos tašku $-CH_2-CH_2-$ grandine. Analogišku būdu aprašę fragmentą $R-O-CH_2-Cl$ gautume HOSE kodą neturintį visiškai nieko bendro su pirmuoju pavyzdžiu. Kadangi šiuo atveju skiriasi pats pirmasis sluoksnis, automatiškai nebesutampa ir tolesniuose sluoksniuose esančių atomų susijungimo su radikalu R keliai. Išvardintosios savybės lemia HOSE kodų sugebėjimą ypatingai detaliam aprašyti bet kurio pasirinkto molekulės atomo artimiausią cheminę apsuptį. Su šiomis savybėmis susijęs ir bene pagrindinis tokių kodų panaudojimo būdas – spektrų, pavyzdžiui ^{13}C -BMR, prognozavimas. Šiuo atveju ieškoma konkretaus anglies atomo HOSE kodo analogų duomenų bazėje su eksperimentinėmis cheminio poslinkio reikšmėmis. Yra laikoma, kad aptikus HOSE kodų

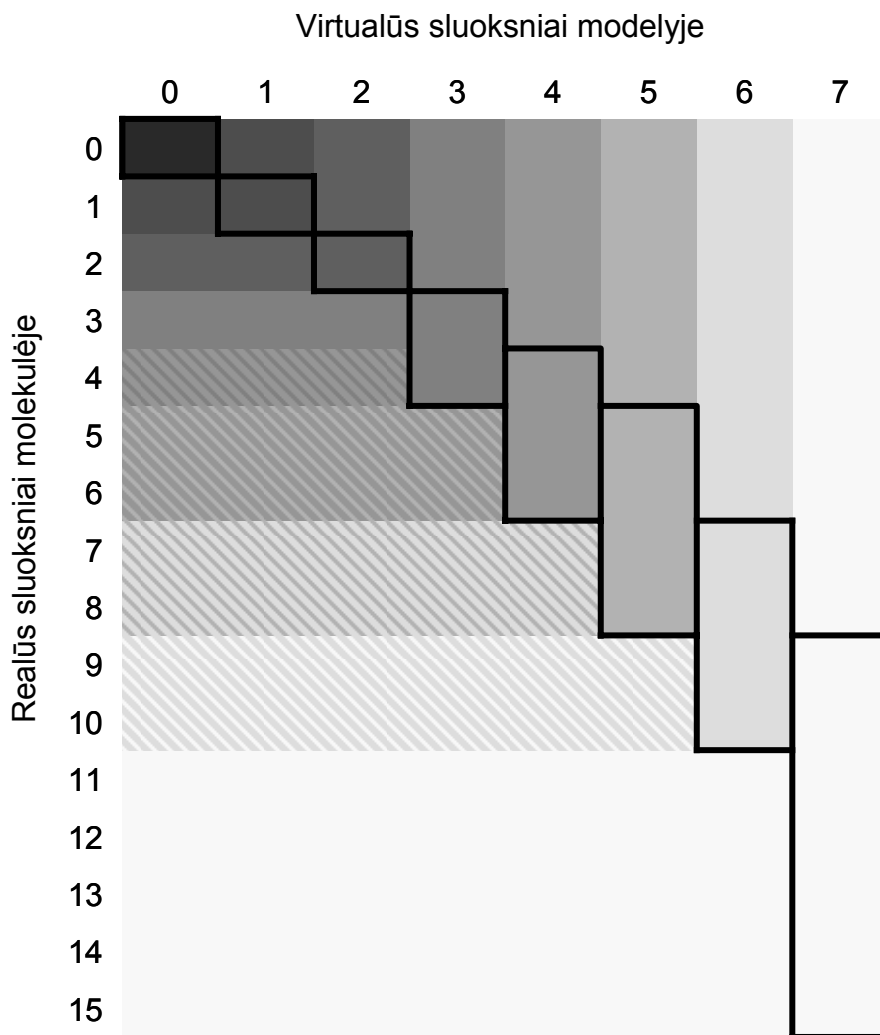
atitikimą iki ketvirto ar aukštesnio lygmens, su didele tikimybe galima prognozuoti, jog atitinkamų anglies atomų cheminiai poslinkiai ^{13}C -BMR spektre bus labai panašūs. Kita vertus, ketvirtas ar penktas sluoksnius yra praktiškai HOSE kodų panaudojimo riba. Dėl ypatingai radikalaus deskriptorių diferencijavimo atsižvelgiant į menkiausius skirtumus, didinant analizuojamų sluoksnių skaičių, generuojamų kintamųjų skaičius labai greitai auga į begalybę.

Nagrinėjant metabolizmo fermentų regioselektyvumą, pirmųjų penkių sluoksnių analizės nepakanka, todėl kuriant sluoksninės fragmentacijos metodą šios disertacijos tikslams, kai kuriose vietose buvo pritaikyti iš esmės nuo HOSE kodų sudarymo taisyklių besiskiriantys principai. Visų pirma, sluoksnių aprašymui šiuo metodu yra naudojama informacija griežtai tik apie tą konkretų sluoksnį. Grįžtant prie pavyzdžio su $\text{R-CH}_2\text{-CH}_2\text{-Cl}$ ir $\text{R-O-CH}_2\text{-Cl}$, antrasis ir trečiasis sluoksniai šiuo atveju visiškai sutaptų bei būtų aprašomi kiekvienas tuo pačiu deskriptoriumi. Taigi šiuo konkrečiu atveju, vietoje šešių deskriptorių, reikalingų naudojant HOSE kodą, galima išsiversti su keturiais. Siekiant išvengti pernelyg didelio molekulių supanašėjimo taikant naują sluoksninės fragmentacijos metodą, sluoksniams buvo suteikti svoriai. Jų pagalba buvo apibrėžta didžiausia pirmųjų sluoksnių įtaka panašumui, palaipsniui mažėjanti tolstant nuo fragmentacijos išeities taško. Tokiu būdu buvo užtikrinta, kad, pavyzdžiui, tie patys fragmentai $\text{R-CH}_2\text{-CH}_2\text{-Cl}$ ir $\text{R-O-CH}_2\text{-Cl}$, nepaisant vienintelio skirtumo pirmajame sluoksnyje, išlaikant cheminę logiką būtų traktuojami kaip pakankamai skirtingi. Antroji pristatomo metodo ypatybė yra ta, kad fragmentacijoje naudojama kintamųjų sąrašą sudaro ne vien atskiri atomai, bet ir ištisi fragmentai. Kitaip tariant, vietoje to, kad aprašant konkretaus atomo aplinką molekulėje minėti šešis skirtingais atstumais nuo jo nutolusius aromatinius anglies atomus, galima viską apibendrinti nurodant tam tikrame sluoksnyje esantį visą benzeno žiedą kaip vieną deskriptorių.

Išvardinti pakeitimai leido ženkliai sumažinti sluoksninės atomocentrinės fragmentacijos metu susidarantį kintamųjų skaičių, lyginant su HOSE

kodų generavimo metodu. Tačiau net ir šiuo atveju generuojamų deskriptorių skaičius išlieka viena opiausių problemų. Tarkime, kad atomo-centrinėje fragmentacijoje norime panaudoti n fragmentų, kurių ieškosime molekulėje, sąrašą. Tuomet, jeigu nagrinėsime k sluoksnių nuo pradinio atomo, galutinėje duomenų matricoje turėsime $n \times k$ kintamųjų (vietoj tiesiog n įprastiniu visos molekulės fragmentacijos atveju), nes bet kuris iš fragmentų gali pasitaikyti bet kuriame iš sluoksnių. Daugeliu atvejų tai vis dar didelis skaičius, taigi siekiant dar labiau sušvelninti tokį kintamųjų skaičiaus padidėjimą lyginant su faktiškai naudojamų fragmentų kiekiu, buvo pritaikyta sluoksnių apjungimo schema, leidžianti jiems šiek tiek persikloti. T.y. pats metabolizmo centras (0-is sluoksniu) bei pirmieji du sluoksniai yra aprašomi individualiai, o toliau 3-ias ir 4-as sluoksniai apjungiami ir naudojami statistinėje analizėje kaip vienas kintamasis. Analogiškai apjungiami 4-as, 5-as bei 6-as sluoksniai ir t.t. (žr. 10 pav.). Ši schema leidžia apsiriboti septyneriopu kintamųjų skaičiaus padidėjimu, nors realiai molekulėje yra nagrinėjami 15 sluoksnių. Aromatinio anglies atomo hidrosilinimo atveju tikslus ir nepersidengiantis sluoksnių aprašymas buvo išplėstas iki pirmųjų penkių sluoksnių įskaitant patį metabolizmo centrą, tuo tikslu, kad detalai būtų aprašomas bent jau pirmasis atomas pakaito, esančio bet kurioje šešianario žiedo pozicijoje metabolizuojamo ar nemetabolizuojamo atomo atžvilgiu (t.y. orto-, meta-, para- padėtyse).

Papildomai kuriant modelį deskriptoriais buvo panaudotas sąrašas klasikinių grupių, pasižyminčių stipriais elektroniniais efektais (pvz., $-\text{NO}_2$, $-\text{CF}_3$ ir pan.). Nagrinėjamosiose metabolizmo reakcijose, katalizuojamose citochromų P450, greitį limituojanti pakopa yra radikaline reakcija (vandenilio radikalo atskėlimas nuo anglies atomo [128-133] arba tiesioginis elektrono atplėšimas nuo heteroatomo kai kuriais N- ir O-dealkilinimo atvejais [134, 135]), taigi bet kokie veiksniai, potencialiai įtakojantys elektronų tankį atomo aplinkoje, gali nulemti galimybę toje vietoje vykti metabolizmo reakcijai. Paminėtųjų grupių buvimo molekulėje įtaka konkrečiam metabolizmo centrui buvo apskaičiuojama pagal analogiją su elektroninių sąveikų silpnėjimu



10 pav. Sluosnių apjungimo schema

priklausomai nuo atstumo tarp sąveikos centrų molekulėje. Pasirinktas gesimo koeficientas – 0,6, t.y., pavyzdžiui, trečiajame sluosnyje esančios grupės įtaka bus lygi $(0,6)^2=0,36$ tokios pačios grupės esančios pirmajame sluosnyje elektroninio efekto, šeštajame – $(0,6)^5=0,078$ ir t.t. Kita vertus, darant prielaidą, kad šio sąrašo grupių įtaka apsiriboja elektroniniais efektais, pasidaro nebe taip svarbu, kiek konkrečių grupių kokių atstumu yra nuo pažymėto centrinio atomo, svarbi tiktai suminė jų įtaka, kuri ir yra naudojama kaip vienintelis deskriptorius kiekvienos konkrečios grupės atveju. Pavyzdžiui, tiek turinčio vieną vienintelę $-NO_2$ grupę antrame sluosnyje, tiek turinčio tris tokias grupes (tarkime dvi trečiame ir vieną penktame) junginio atveju, $-NO_2$

grupės įtakos aprašymui bus naudojamas vienas ir tas pats deskriptorius, skirsis tik jo reikšmės – atitinkamai $(0,6)^1=0,6$ ir $2 \times (0,6)^2 + (0,6)^4=0,85$.

2.3 Statistiniai metodai

2.3.1 GALAS modeliavimo metodika

Schemiškai šiame darbe naudotą GALAS modeliavimo metodą [6, 7] galima išvaizduoti kaip dviejų sistemų kombinaciją:

- Kiekybiniu struktūros-aktyvumo sąryšiu pagrįstas modelis skirtas prognozuoti pasirinktą dominančią savybę ar aktyvumą – pradinis („bazinis“) modelis;
- Panašumo analizės pagrindu atliekama korekcijos procedūra, identifikuojanti panašiausius apmokomojo rinkinio junginius ir atsižvelgiant į jų eksperimentiškai išmatuotas reikšmes nustatanti ar bazinio QSAR modelio prognozė konkrečiam junginiui nėra iškreipta sisteminių paklaidų.

Pirmoji dalis yra globalusis modelis. Šiuo konkrečiu atveju ji yra paremta PLS arba BPLS – tiesiniu, adityviniu metodu. Šis modelis išmoksta globalias tendencijas nagrinėjamojoje savybėje ir dinamiškai apibrėžia panašumo sąvoką. Kitaip tariant „išmoksta“ panašumo raktą, išrenka svarbiausius struktūrinius bruožus, nulemiančius dviejų lyginamų junginių panašumą konkrečios nagrinėjamos savybės atveju. Tuo tarpu antrasis modelio žingsnis turėtų būti traktuojamas kaip lokalusis modelis. Analizuodamas bet kurio junginio panašiausius kaimynus aplinkinėje cheminės erdvės srityje jis atlieka bazinių prognozių korekcijas, tokiu būdu įveddamas netiesiškumo elementą, kuris yra neišvengiamas praktiškai visose, įskaitant pačias paprasčiausias, savybėse ar aktyvumuose.

Be didesnio prognozavimo tikslumo, lyginant su klasikiniiais QSAR metodais, GALAS modeliavimo metodika pasižymi kitomis dviem ne mažiau svarbiomis savybėmis: galimybe kiekybiškai įvertinti sukurto modelio

pritaikomumo sritį ir sugebėjimu greitai apsimokyti naujais, modelio kūrime nenaudotais junginiais.

2.3.2 PLS ir BPLS metodai bei bazinis QSAR modelis

Šiame darbe naudoto PLS [118] metodo istorija yra vienas iš pavyzdžių, kuomet specifiniai cheminių sistemų savitumai įtakoja naujų statistinių metodų, skirtų jų analizei atsiradimą ir vystymąsi. Taip jau atsitiko, kad cheminės specifikos keliamos problemos sutapo su tomis, kurių sprendimui buvo suskurtas šis metodas, todėl jis labai plačiai paplito chemometrijoje ir chemoinformatikoje, tuo tarpu kai kitose statistikos srityse PLS yra naudojamas labai ribotai.

Taigi PLS – tai tiesinės regresijos metodas, skirtas susieti dvi duomenų matricas \mathbf{X} ir \mathbf{Y} . Išskirtinis šio metodo privalumas – galimybė analizuoti dideles, mažai „išsodrintas“ (turinčias daug tuščių įrašų) tarpusavyje koreliuotų kintamųjų matricas. Šis metodas – tai optimizavimo uždavinys, kai siekiama rasti tokius “užslėptus” kintamuosius \mathbf{T} , kurie būtų principiniai \mathbf{X} matricos komponentai ir gerai aprašytų \mathbf{Y} matricą. Ši optimizavimo problema sprendžiama, pasinaudojant žemiau surašytomis lygtimis.

PLS modelyje ieškoma A naujų, latentinių \mathbf{X} matricos kintamųjų (angl. *scores*), žymimų t_a ($a=1, 2, \dots, A$):

$$t_{ia} = \sum_k w_{ka}^* x_{ik} \quad (11)$$

čia x_{ik} – pradiniai kintamieji; w_{ka}^* – pradinių kintamųjų svoriai.

Apskaičiuotiems t_a kintamiesiems keliami užduotis – geras \mathbf{Y} matricos prognozavimas:

$$y_{im} = \sum_a c_{ma} t_{ia} + f_{im} \quad (12)$$

čia c_{ma} – latentinių kintamųjų koeficientai; f_{im} – skirtumai tarp stebimų verčių ir modelio, sudarantys \mathbf{Y} matricos skirtumų matricą \mathbf{F} .

Arba perrašius kaip regresijos modelį:

$$y_{im} = \sum_a c_{ma} \sum_k w_{ka}^* x_{ik} + f_{im} = \sum_k b_{mk} x_{ik} + f_{im} \quad (13)$$

čia $b_{mk} = \sum_a c_{ma} w_{ka}^*$ – PLS regresijos koeficientai.

Latentiniai \mathbf{X} matricos kintamieji pasižymi tokiomis savybėmis:

- Jų yra keletas (A vienetų) ir jie yra ortogonalūs;
- Egzistuoja tokios \mathbf{X} matricos „santraukos“, sudarytos iš latentinių kintamųjų, kad e_{ik} liekanos, gaunamos iš „santraukos“ atgal atvaizduojant matricą \mathbf{X} , būtų mažos:

$$x_{ik} = \sum_a t_{ia} p_{ak} + e_{ik} \quad (14)$$

čia p_{ak} – latentinių kintamųjų pervedimo į pradinius koeficientai; e_{ik} – „santraukos“ atvaizdavimo į pradinius kintamuosius liekana.

Parankiausias būdas PLS modelio apskaičiavimui yra iteracinis, t.y. principiniai komponentai yra apskaičiuojami paeiliui po vieną, panaudojant informaciją, likusią po ankstesnių komponentų ištraukimo. Apskaičiavus pirmąjį latentinį kintamąjį t_{i1} , iš jo, panaudojant koeficientus p_{1k} , gaunama pradinių kintamųjų pavidalo matricinė pirmojo komponento išraiška $t_{i1}p_{1k}$, kuri atimama iš pradinės kintamųjų matricos \mathbf{X} . Po pirmojo komponento „ištraukimo“ matricoje likusi informacija yra panaudojama sekančio latentinio kintamojo (t_{i2}) apskaičiavimui. Taip komponentai traukiami tol, kol matricoje lieka tik vadinamasis triukšmas (angl. *noise*), nebeturintis modeliui vertingos informacijos (e_{ik} 14 lygtyje). Matematiškai iteracinį procesą galima atvaizduoti tokiais lygtimis:

$$t_{ia+1} = \sum_k w_{ka+1} e_{ik,a} \quad (15)$$

$$e_{ik,a} = e_{ik,a-1} - t_{ia} p_{ak} \quad (16)$$

$$e_{ik,0} = x_{ik} \quad (17)$$

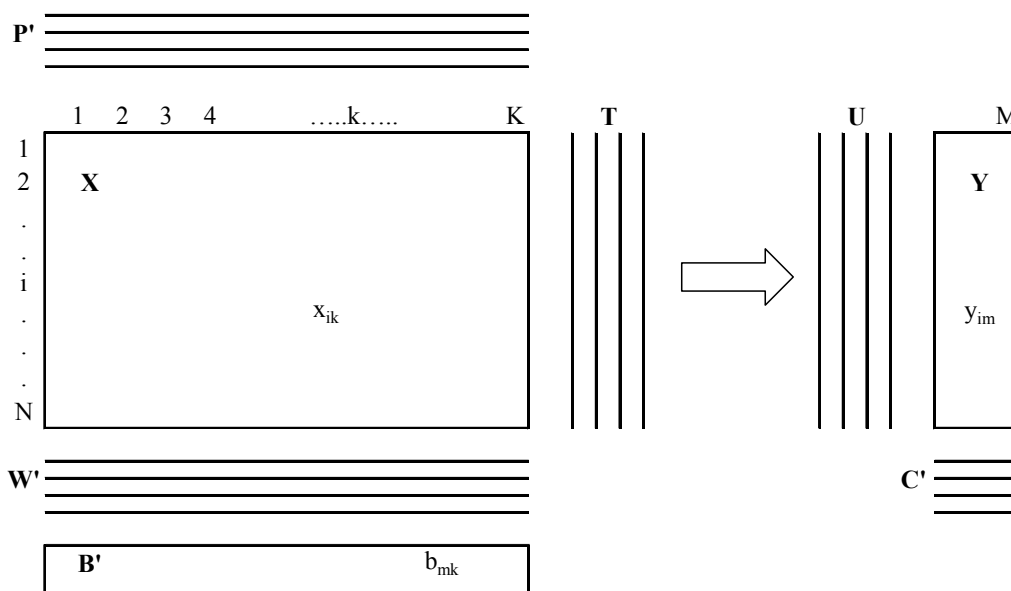
čia $e_{ik,a}$ – liekanų matrica, likusi po a-tojo komponento ištraukimo; w_{ka+1} – liekanų $e_{ik,a}$ svoriai, $e_{ik,a-1}$ – liekanų matrica, likusi po (a-1)-ojo komponento ištraukimo; $t_{ia}p_{ak}$ – matricinė a-tojo komponento išraiška; $e_{ik,0}$ – pradinė matrica, naudojama 1-ojo komponento apskaičiavimui.

Liekanų svorius w , naudojamus iteraciniame procese, su pradinių kintamųjų svoriais w^* sieja ši lygtis:

$$\mathbf{W}^* = \mathbf{W}(\mathbf{P}'\mathbf{W})^{-1} \quad (18)$$

Naudojant iteracinį principinių komponentų traukimo būdą, atsiranda galimybė maksimaliai optimizuoti modelį, t.y. papildomus komponentus traukti iš matricos tik tol, kol gerėja modelio statistiniai parametrai.

Suskaičiavus modelio parametrus, naudojant PLS yra gaunami ne tik lygties koeficientai, bet ir visa eilė matricių (žr. 11 pav.), kurios yra panaudojamos papildomiems statistiniams modelio įvertinimams, negalimiems atlikti įprastą MLR analizę.



11 pav. X ir Y matricių išskaidymas į kitas, PLS procedūroje naudojamas, matricas

BPLS metodas iš esmės yra binominės regresijos ir PLS kombinacija, apjungianti naudingiausias šių metodų savybes. Jis pasižymi binominės regresijos galimybėmis apdoroti binarinius duomenis, atspindinčius molekulės aktyvumą arba priklausymą tam tikrai kokybinei klasei, ir PLS sugebėjimu dirbti su didelėmis, mažai “išsodrintomis” (turinčiomis daug tuščių įrašų) tarpusavyje koreliuotų kintamųjų matricomis. BPLS modelio prognozuojama

reikšmė konkrečios molekulės atveju yra jos aktyvumo arba priklausymo kokybinei klasei tikimybė, išreiškiama tokia lygtimi:

$$p = 1/(1 + e^{-x}), \text{ kur } x = \sum_i a_i F_i \quad (19)$$

čia p – prognozuojama tikimybė; a_i – i-tojo fragmento įtakos koeficientas; F_i – i-tojo fragmento pasitaikymo molekulėje dažnis.

Remiantis aukščiau aprašytais PLS ir BPLS metodais, kombinuojant juos su daugkartinės plėtros procedūra (angl. *bootstrapping*) [136] buvo sukurti atitinkamai 6 baziniai ūmaus toksiškumo modeliai skirtingų gyvūnų rūšių bei patekimo į organizmą kelių duomenų rinkiniams (žr. 3 lentelę) ir 8 baziniai modeliai remiantis duomenimis apie įvairius skirtingų fermentų katalizuojamų metabolizmo reakcijų tipus (žr. 4 lentelę). Kartotinės plėtros procedūra numato atsitiktinę pradinio duomenų rinkinio įrašų parinkimą, formuojant apmokomuosius poaibius ir atskiro PLS ar BPLS modelio parametrizavimą kiekvienam iš jų. Kiekvienas iš tokių poaibių yra tokio paties dydžio kaip ir pradinis duomenų rinkinys, tačiau atsitiktinio užpildymo metu kai kurie pradinio rinkinio junginiai į jį yra įtraukiami daugiau nei vieną kartą, tuo tarpu kiti apskritai į jį nepatenka. Visa atsitiktinio poaibio parinkimo procedūra yra kartojama 100 kartų kiekvieno konkretaus bazinio modelio atveju. Tokiu būdu vietoj vieno statistinių parametrų rinkinio (prognozuotų savybės reikšmių ir lygties koeficientų) gauname 100 analogiškų ir šiek tiek skirtingų statistinių parametrų rinkinių, t.y. bet kuris tokiu būdu sukurtas bazinis modelis iš esmės junginiui pateikia vektorių, susidedantį iš 100 prognozuotų savybės verčių, kurios visos šiek tiek tarpusavyje skiriasi. Galutinė bazinio (globalaus) modelio prognozė yra apskaičiuojama kaip šių 100 verčių aritmetinis vidurkis.

Ši informacija taip pat panaudojama panašumui nustatyti. Lyginant bet kuriuos du junginius, jei jų vektoriuose stebimos panašios prognozuotų verčių kitimo tendencijos, laikoma, kad šie junginiai yra panašūs nagrinėjamos savybės atveju [6, 7]. Kadangi tik tokiu atveju skirtumai apmokomojo rinkinio poaibiuose, naudotuose apmokant 100 PLS ar BPLS modelių, sudarančių bazinį QSAR modelį, galėtų panašiai įtakoti prognozuotas vertes abiem

junginiams. Kiekybiškai ši informacija išreiškiama individualiu panašumo indeksu tarp dviejų junginių (SI_i), apskaičiuojamu kaip koreliacijos koeficientas tarp bazinio modelio šiems junginiams prognozuotų verčių vektorių [6, 7]. Svarbiausias tokio panašumo traktavimo aspektas yra tai, kad jis yra apibrėžiamas dinamiškai, modelio parametrizavimo metu, ir šiame procese yra atsižvelgiama į modeliuojamąją savybę, kadangi bendru atveju tai, kas yra panašu savybės (aktyvumo) A atveju, gali turėti pastebimai menkesnę įtaką panašumui savybės (aktyvumo) B atveju.

2.3.3 Lokalus modelis ir bazinių prognozių korekcijos

Tiesinio modelio globalių prognozių nauda daugeliu atveju gali būti ganėtinai ribota, kadangi didelių ir chemiškai įvairių junginių rinkinių atveju klasikinis QSAR modelis tėra tiesinis priartėjimas, kuris išglotnina nagrinėjamosios savybės ar aktyvumo profilį. Realiai beveik visų savybių atveju yra stebimos lokalsios cheminės erdvės sritys, nukrypstančios nuo bendrų tiesinių dėsningumų. G. M. Maggiora savo straipsnyje pateikia labai vaizdingą šios situacijos apibūdinimą, kuris išvertus skambėtų maždaug taip: „Daugeli metų buvo laikoma, kad panašios molekulės yra linkusios pasižymėti panašiu aktyvumu, ko pasekoje susidarantis savybės landšaftas galėtų būti lyginamas su švelniai banguojančiomis kalvomis Kanzaso prerijoje. Tačiau vis daugėja įrodymų rodančių, jog šis vaizdas nėra toks visuotinis, kaip manyta iki šiol, ir daugeliu atvejų labiau primena raižytą Jutoje esančio Bryce'o kanjono kraštovaizdį“ [137]. Kuo sudėtingesnė yra nagrinėjamoji savybė ar aktyvumas, tuo daugiau panašių nukrypimų galima tikėtis. Iš esmės, nagrinėjant tokias su biocheminėmis sistemomis susijusias savybes ir aktyvumus kaip ūmus toksiškumas ar metabolizmas, šie lokalūs reiškiniai yra ypač svarbūs. Būtent jie dažnai apsprendžia pageidaujamo ar nepageidaujamo efekto atsiradimą, susijusį su konkrečios cheminės erdvės sritys junginiams būdinga jonizacija, tirpumo dėsningumais, hidrofobiniu sulaikymu membranose ar audiniuose, farmakoforiniu „rakto-spynos“ atitikimu, sugebėjimu kovalentiškai prisijungti prie biologinių makromolekulių ir t.t., ir pan.

Dėl šių priežasčių GALAS metodikoje egzistuoja antroji, panašumo analizę naudojanti dalis. Ji paremta globalaus QSAR modelio veikimo mus dominančio junginio kaimyninėje cheminėje aplinkoje įvertinimu, t.y., eksperimentinių duomenų ir bazinių QSAR modelio prognozių palyginimu panašiausiems apmokomojo rinkinio junginiams. Jeigu bazinėms šių junginių prognozėms yra būdingos kokios nors sisteminės paklaidos, lyginant su jiems išmatuotomis savybės ar aktyvumo vertėmis, galima tikėtis, jog tas pats galioja ir konkrečiai analizuojamai molekulei, t.y., jos bazinei PLS ar BPLS modelio prognozei yra reikalinga lokalinė pataisa (Δ). Pastaroji yra apskaičiuojama kaip panašiausiems apmokomojo rinkinio junginiams nustatytų skirtumų tarp globalių QSAR prognozių ir eksperimentinių duomenų svorinis vidurkis [6, 7]:

$$\Delta = \frac{\sum_{i=1}^n a^{i-1} \cdot SI_i \cdot \Delta_i}{\sum_{i=1}^n a^{i-1}} \quad (20)$$

čia Δ – nagrinėjamam junginiui reikalinga korekcija; a – konstanta, įtakojanti svorinio vidurkio skaičiavimą, kai ji lygi 1, gauname paprastą vidurkį; SI_i – panašumas (panašumo indeksas) tarp nagrinėjamo junginio ir i -tojo panašiausio iš apmokomojo rinkinio junginių, apskaičiuojamas kaip koreliacijos koeficientas tarp dviejų prognozuotų reikšmių vektorių iš kartotinės plėtros modelių; Δ_i – skirtumas tarp globalaus modelio prognozuotos ir išmatuotosios savybės reikšmių i -tajam panašiausiam junginiui: $\Delta_i = Y_i - \bar{Y}_i$; n – konstanta, nurodanti kiek panašiausių junginių reikia imti skaičiuojant korekciją.

Akivaizdu, jog (20) lygties rezultatas priklauso nuo panašiausių junginių skaičiaus (n), naudojamo analizėje. Empiriškai nustatyta, kad optimalios modelio charakteristikos pasiekiamos korekcijų skaičiavimui naudojant penkis panašiausius apmokomojo rinkinio junginius, nes tokiu būdu užtikrinama pusiausvyra tarp modelio apmokymo naujais junginiais greičio ir atsitiktinių klaidų tikimybės [6, 7].

2.3.4 Modelio pritaikomumo srities įvertinimas

Kaip jau buvo minėta, GALAS modeliavimo metodika leidžia kiekybiškai įvertinti sukurto modelio pritaikomumo sritį. Jos įvertinimo procedūroje atsižvelgiama į du pagrindinius aspektus:

- Analizuojamojo junginio panašumas į apmokomąjį rinkinį. Patikima prognozė nėra įmanoma, jeigu apmokomajame rinkinyje neturime junginių, panašių į mus dominantį. Kiekybinis kriterijus – panašumo indeksas (SI).
- Eksperimentinių duomenų panašioms junginiams neprieštaringumas modeliui. Netgi turint panašių junginių apmokomojoje duomenų bazėje, prognozės patikimumas gali būti mažas, jei jiems turimi eksperimentiniai duomenys prieštarauja modeliui. Kiekybinis kriterijus – modelio ir duomenų neprieštaringumo indeksas ($DMCI$).

Minimi indeksai apskaičiuojami atitinkamai vidurkinant individualius mus dominančio junginio ir n panašiausių apmokomojo rinkinio junginių panašumų indeksus (SI_i) bei lyginant tų pačių panašiausių junginių skirtumus tarp eksperimentinių duomenų ir bazinės prognozės (Δ_i) su (20) lygtyje apskaičiuojama lokalia korekcija mus dominančiam junginiui (kuo didesnis individualių skirtumų išsibarstymas apie vidurkį Δ , tuo didesnis duomenų prieštaravimas modeliui) [6, 7]:

$$SI = \frac{\sum_{i=1}^n a^{i-1} \cdot SI_i}{\sum_{i=1}^n a^{i-1}} \quad (21)$$

čia SI – panašumo į modelio junginius indeksas.

$$DMCI = e^{-\left(\frac{\sum_{i=1}^n a^{i-1} \cdot SI_i \cdot (\Delta - \Delta_i)^2}{\sum_{i=1}^n a^{i-1}}\right) / b} \quad (22)$$

čia $DMCI$ – duomenų ir modelio neprieštaringumo indeksas; b – empirinė konstanta, kuria apibrėžiama, koks duomenų išsibarstymas nebetoleruotinas ir $DMCI$ tampa artimas nuliui, pvz., dešimtadalis visos įgyjamų reikšmių skalės.

Abiem atvejais naudojamas panašiausių apmokomojo rinkinio junginių skaičius (n) yra tas pats kaip ir apskaičiuojant lokalią korekciją (Δ) pagal (20) lygtį, t.y. penki.

Galutinis prognozės patikimumo indeksas (RI) gaunamas taip [6, 7]:

$$RI = SI \cdot DMCI \quad (23)$$

Patikimumo indeksas (kaip ir abu lygtyje dauginami indeksai) kinta intervale nuo 0 iki 1 – reikšmės mažėjimas rodo, jog junginys tolsta nuo modelio pritaikomumo srities, tuo tarpu didėjančios *RI* reikšmės leidžia vis labiau pasitikėti prognozės patikimumu. Kadangi prognozės patikimumo indeksas apskaičiuojamas kaip dviejų indeksų sandauga, jo vertė nykstamai mažėja bet kuriam iš daugiklių artėjant prie 0. Šis faktas atspindi pagrindinę prielaidą – norint gauti patikimą prognozę reikalingi tiek panašūs junginiai, tiek jų eksperimentiniai duomenys neprieštaraujantys modeliui (tiek *SI*, tiek *DMCI* vertės aukštos). Tuo tarpu jei su bent vienu iš šių kriterijų kyla problemų, patikimas prognozavimas pasidaro problematiškas.

3 REZULTATAI IR JŲ APTARIMAS

3.1 Ūmaus toksiškumo (LD_{50}) modeliavimų rezultatai

3.1.1 Bazinių (globalių) ir galutinių (lokalių) modelių palyginimas

GALAS metodika paremtų ūmaus toksiškumo modelių skirtingoms graužikų rūšims ir patekimo į organizmą keliams veikimas bei jų pritaikomumo srities nustatymo patikimumas buvo vertinami pasitelkiant atitinkamus patikrinamuosius junginių rinkinius aprašytus 3 lentelėje (2.1.1 skyrius). Žemiau esančioje 6 lentelėje pateikiami pagrindiniai statistiniai parametrai, iliustruojantys atitinkamų modelių charakteristikas prieš ir po lokalių, panašumo analize paremtos, korekcijos (Δ) pritaikymą – bazinis ir galutinis modelis atitinkamai.

6 lentelė. Ūmaus toksiškumo (LD_{50}) prognozavimo rezultatai patikrinamųjų rinkinių junginiams, patenkantiems į atitinkamų modelių pritaikomumo sritis ($RI > 0,3$)

Rūšis	Įvesties būdas	$N_{RI>0,3}$	Bazinis		Galutinis	
			R^2	RMSE	R^2	RMSE
Žiurkė	Oralinis (OR)	1976 (91%)	0,45	0,66	0,56	0,59
	Intraperitoninis (IP)	1130 (90%)	0,31	0,62	0,42	0,58
Pelė	Oralinis (OR)	4545 (93%)	0,30	0,47	0,49	0,40
	Intraperitoninis (IP)	8568 (95%)	0,35	0,50	0,56	0,41
	Intraveninis (IV)	4754 (95%)	0,47	0,48	0,61	0,41
	Poodinis (SC)	2056 (96%)	0,40	0,59	0,54	0,52

Pateiktieji rezultatai apima tikrai patenkinamo ar didesnio patikimumo ($RI > 0,3$) prognozes, kadangi junginiai, kuriems modelio prognozės yra nepatikimos ($RI < 0,3$) pagal apibrėžimą nepatenka į modelio pritaikomumo sritį ir jų panaudojimas analizėje nesuteikia jokios prasmingos informacijos apie modelio veikimą. Kaip matyti iš 6 lentelės duomenų, tokių junginių kiekis kiekvienu atskiru atveju sudaro ne daugiau kaip 5-10% viso patikrinamojo

rinkinio dydžio. Visais atvejais korekcijos pagal panašumą taikymas davė pastebimų teigiamų rezultatų (R^2 vidutiniškai išaugo 0,15, $RMSE$ sumažėjo 0,07). Tai rodo, kad esant galimybių apmokomajame rinkinyje aptikti bent kelias į tiriamąjį junginį panašias molekules su nuosekliais eksperimentiniais duomenimis, bazinio modelio veikimo lokatioje cheminėje erdvėje analizė daugeliu atveju leidžia aptikti tam tikrus šio tiesinio modelio nepaaiškinamus svyravimus. Kitais žodžiais tariant, pateiktieji rezultatai patvirtina įvairių netiesinių efektų svarbą ir į tai sugebančios atsižvelgti modeliavimo metodikos naudą tokio biologinio aktyvumo kaip ūmus toksiškumas nagrinėjime. Kaip jau buvo minėta, tokie netiesiškumai kaip taisyklė atsiranda dėl adityvumo principo išimčių, t.y. situacijose, kuomet tam tikra molekulės struktūros dalis dėl kokių nors priežasčių yra atsakinga už žymiai didesnę įnašą junginio ūmiam toksiškumui, nei apskaičiuojamas pagal ją sudarančių struktūrinių deskriptorių koeficientus.

Vienas įdomesnių pastebėjimų šioje vietoje būtų tai, kad šiame darbe panaudotoje GALAS metodikoje galima išvelgti tam tikrų analogijų su klasikine iškrentančių prognozių analize (angl. *outlier based analysis*), pasiūlyta R. L. Lipnick'o [24, 37, 38]. Pastarajame metode naudojamas dilinijinis nuo $\log P_{O/W}$ priklausomas bazinės narkozės modelis, paremtas apmokomuoju rinkiniu sudarytu iš monohidroksilių alkoholių ir sočiųjų monoketonų. Tokio paprasto apmokomojo rinkinio naudojimo pasekoje bet kokie sudėtingesni junginiai buvo blogai aprašomi šio modelio ir jiems prognozuojamos ūmaus toksiškumo (LD_{50}) reikšmės daugiau ar mažiau nukrypavo nuo eksperimentiškai stebimų. Šis skirtumas buvo pavadintas „pertekliniu toksiškumu“ (angl. *excessive toxicity* – T_e):

$$T_e = \frac{(\text{bazinė toksiškumo prognozė})}{(\text{eksperimentiškai stebimas toksiškumas})} \quad (24)$$

T_e reikšmių analizė įgalino suklasifikuoti iškrentančias prognozes į atskiras grupes ir identifikuoti su cheminiu reaktyvumu susijusias specifines sąveikas bei mechanizmus atsakingus už konkrečios junginių klasės toksinio

poveikio pobūdį, pavyzdžiui, makromolekulių alkilinimas ar acilinimas, Michaelio tipo prisijungimas, Schiff'o bazių susidarymas, cianido išsiskyrimas ir kt. Bendru atveju GALAS modeliuose naudojamas žymiai sudėtingesnis ir tikslesnis bazinis modelis, pagal galimybes apimantis kiek įmanoma platesnę junginių įvairovę. Tačiau neišvengiamai modeliui teks nagrinėti junginius, turinčius specifinių struktūrinių ypatybių, atsakingų už stiprų toksinį poveikį, tačiau „nežinomų“ modeliui. Daugeliu atvejų bazinis modelis nesugebės atspindėti realybėje egzistuojančios tokio naujoviško toksikoforo stiprios teigiamos įtakos ūmiam toksiškumui tiesiog sumuodamas jam žinomą toksikoforą sudarančių smulkesnių fragmentinių deskriptorių statistinius koeficientus. To pasekoje tokiam junginiui prognozuojama bazinė LD_{50} vertė bus per maža. Jei modelio apmokomajame rinkinyje ar papildomoje panašių junginių bibliotekoje bus pakankamas kiekis panašių molekulių su baziniam modeliui neprieštariniais eksperimentiniais duomenimis, panašumo korekcijos procedūra suskaičiuotų didelę Δ reikšmę kartu su aukštu prognozės patikimumo indeksu (RI). Tai savo ruožtu reikštų, jog tiriamasis junginys yra naudojamo modelio pritaikomumo srityje (t.y. apmokomajame rinkinyje yra pakankamai informacijos panašioms junginiams, kad gauti geros kokybės prognozė), tačiau bazinio QSAR modelio atveju tai yra iškrentanti prognozė. Šiuo požiūriu Δ reikšmės galėtų būti laikomos Lipnick'o naudoto T_e parametro analogu. Pagrindinis šių idėjų skirtumas tas, kad Δ reikšmė konkrečiam junginiui apskaičiuojama remiantis eksperimentiniais duomenimis panašiausiems junginiams vietoj to paties junginio išmatuotos LD_{50} vertės, naudojamos T_e atveju. Šis konkretus faktas nėra vienareikšmis ir atskirais atvejais tai gali būti metodo trūkumas, tačiau neabejotinas ir toks šio skaičiavimo būdo privalumas, kad jis gali būti pritaikytas visiškai naujiems junginiams, kuriems eksperimentinis ūmaus toksiškumo įvertinimas nėra žinomas. Prognozių klasifikavimas pagal Δ reikšmes potencialiai galėtų atskleisti klases junginių, kurie sistemiškai yra identifikuojami kaip bazinio modelio iškrentančios prognozės ir visi gauna panašiai dideles lokalias panašumo korekcijas. Labiausiai tikėtina tokių rezultatų priežastis yra tam tikri

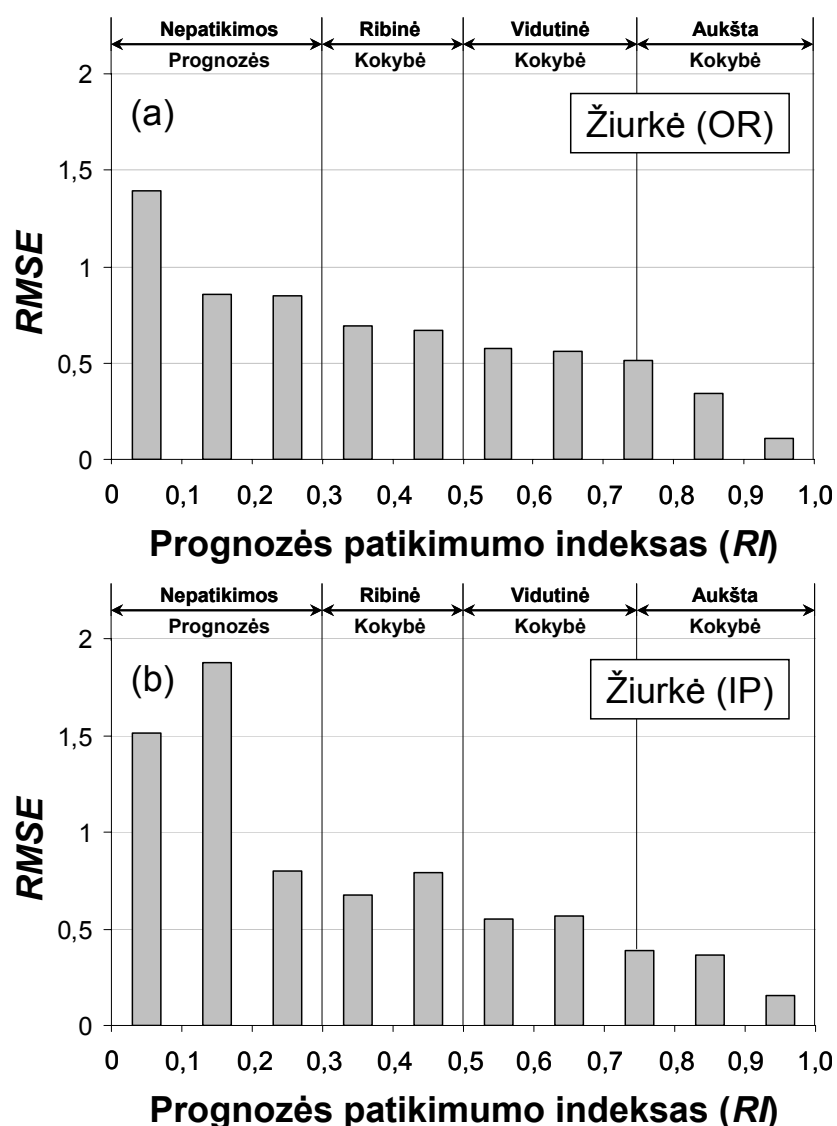
šioms junginiams būdingi struktūros fragmentai, nulemiantys jų toksiinį veikimą per specifines sąveikas ir mechanizmus organizmo viduje, kurie yra nežinomi arba modeliui, arba netgi apskritai. Akivaizdu, kad tokios junginių klasės būtų įdomus objektas tolesnėms detalioms studijoms. Nors tokio tipo analizė nebuvo iškelta tarp šio darbo tikslų, svarbu paminėti, jog papildomos informacijos, GALAS modelio pateikiamos kartu su galutine prognozuota LD_{50} reikšme, gausa sudaro visas prielaidas tokioms naujų mechanistinių išvalgų paieškoms ūmaus toksiškumo duomenų analizėje.

Nepaisant lokalsios panašumo korekcijos procedūros daromos teigiamos įtakos prognozuojamoms LD_{50} vėrtėms, galutiniai rezultatai imant domėn konkretias statistinių parametų reikšmes (ypač R^2) vis dėlto nėra išpūdingi. Iš dalies tai yra rezultatas, patvirtinantis literatūros apžvalgoje minėtas problemas, susijusias su LD_{50} prognozavimu didesniems ir struktūriškai įvairiems junginių rinkiniams bei būtinybę vėrtinti tokių ūmaus toksiškumo modelių pritaikomumo sritį ir rezultatus analizuoti tik tai atsižvelgiant į ją. Todėl visais atvejais pagrindinis šio darbo akcentas yra ne abstrakčios koreliacijos koeficientų, vidutinių paklaidų ar kitų statistinių parametų reikšmės, tačiau jų priklausomybės nuo prognozės patikimumo analizė.

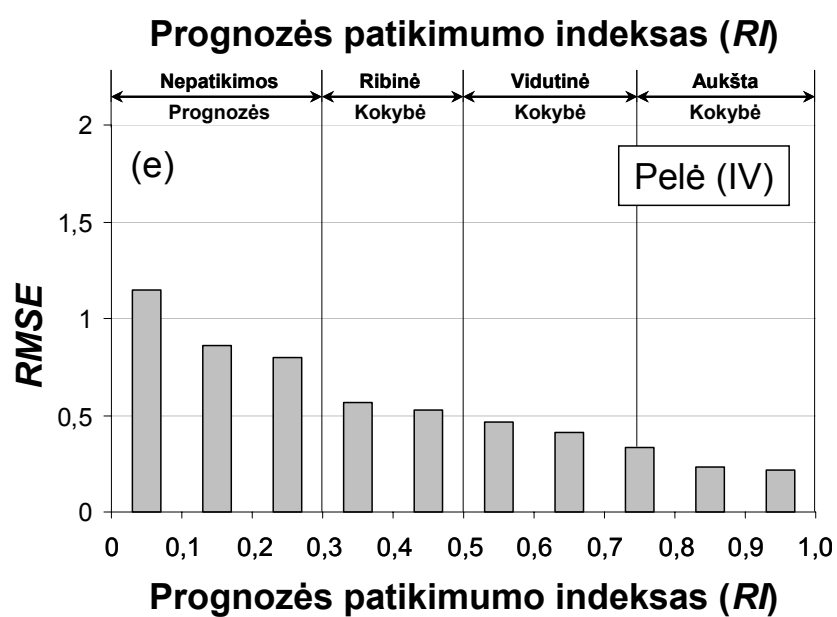
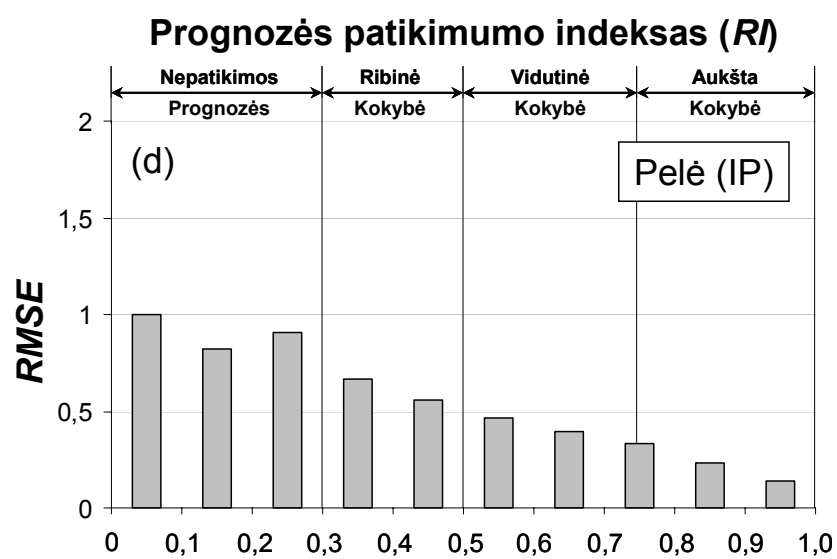
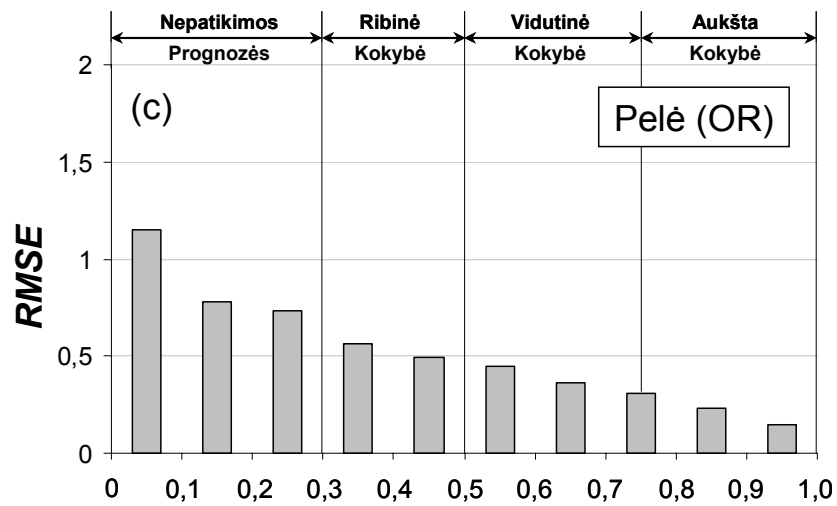
3.1.2 *RMSE* priklausomybė nuo *RI*

12 paveiksle pateikiamos *RMSE* reikšmių pasiskirstymo priklausomai nuo apskaičiuotojo konkretios prognozės patikimumo indekso histogramos visoms šešioms šiame darbe nagrinėjamoms sistemoms, iš kurių matyti, kad didėjant *RI* prognozavimo paklaidos nuosekliai mažėja. Prognozės patikimumo koeficientui viršijus tam tikrą ribą (0,5 ar 0,75, priklausomai nuo atvejo), *RMSE* nukrenta iki 0,5 logaritminio vieneto ir netgi mažiau, kas yra palyginama su LD_{50} nustatymo skirtingose laboratorijose paklaida. Šie rezultatai akivaizdžiai įrodo pasirinktos GALAS metodikos efektyvumą sukurtųjų modelių pritaikomumo srities įvertinime. Šiokios tokios išimty, turint omenyje *RMSE* ir *RI* koreliaciją, yra stebimos srityje kuomet $RI < 0,3$. *RMSE* reikšmių svyravimai šioms prognozėms, matomi 12 paveikslo (b), (d) ir

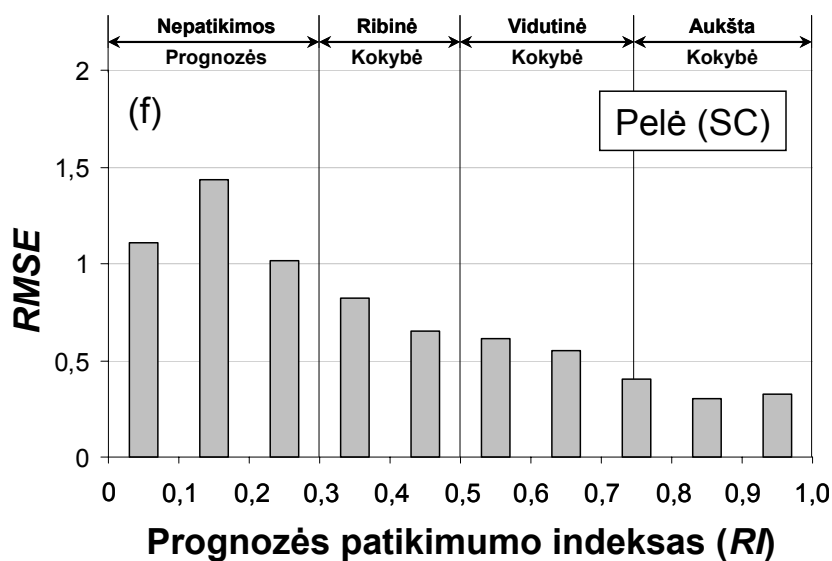
(f) dalyse, paaiškinami tuo paprastu faktu, jog kai kurių *RI* režių atveju statistiniai rezultatai yra iškraipomi dėl labai mažo į juo patenkančių junginių skaičiaus. Tačiau bet kokių atveju šie iškraipymai yra nereikšmingi, kadangi, visų pirma, tai liečia palyginti nedidelę dalį visų junginių (5-10%, priklausomai nuo duomenų rinkinio), o be to – minimos *RMSE* reikšmės vis tiek yra pakankamai didelės (neretai viršija 1 logaritminį vieneta), tad atitinkamų prognozių priskyrimas į nepatikimų klasę nekelia jokių abejonių. LD_{50} prognozėms su didesniu patikimumo indeksu *RMSE* reikšmės yra pastebimai



12 pav. Priklausomai nuo apskaičiuoto prognozės patikimumo indekso stebimų *RMSE* reikšmių pasiskirstymo histogramos atitinkamoms graužikų rūšims ir tiriamųjų junginių patekimo į organizmą keliams



12 pav. Priklausomai nuo apskaičiuoto prognozės patikimumo indekso stebimų $RMSE$ reikšmių pasiskirstymo histogramos atitinkamoms graužikų rūšims ir tiriamųjų junginių patekimo į organizmą keliams (tęsinys)

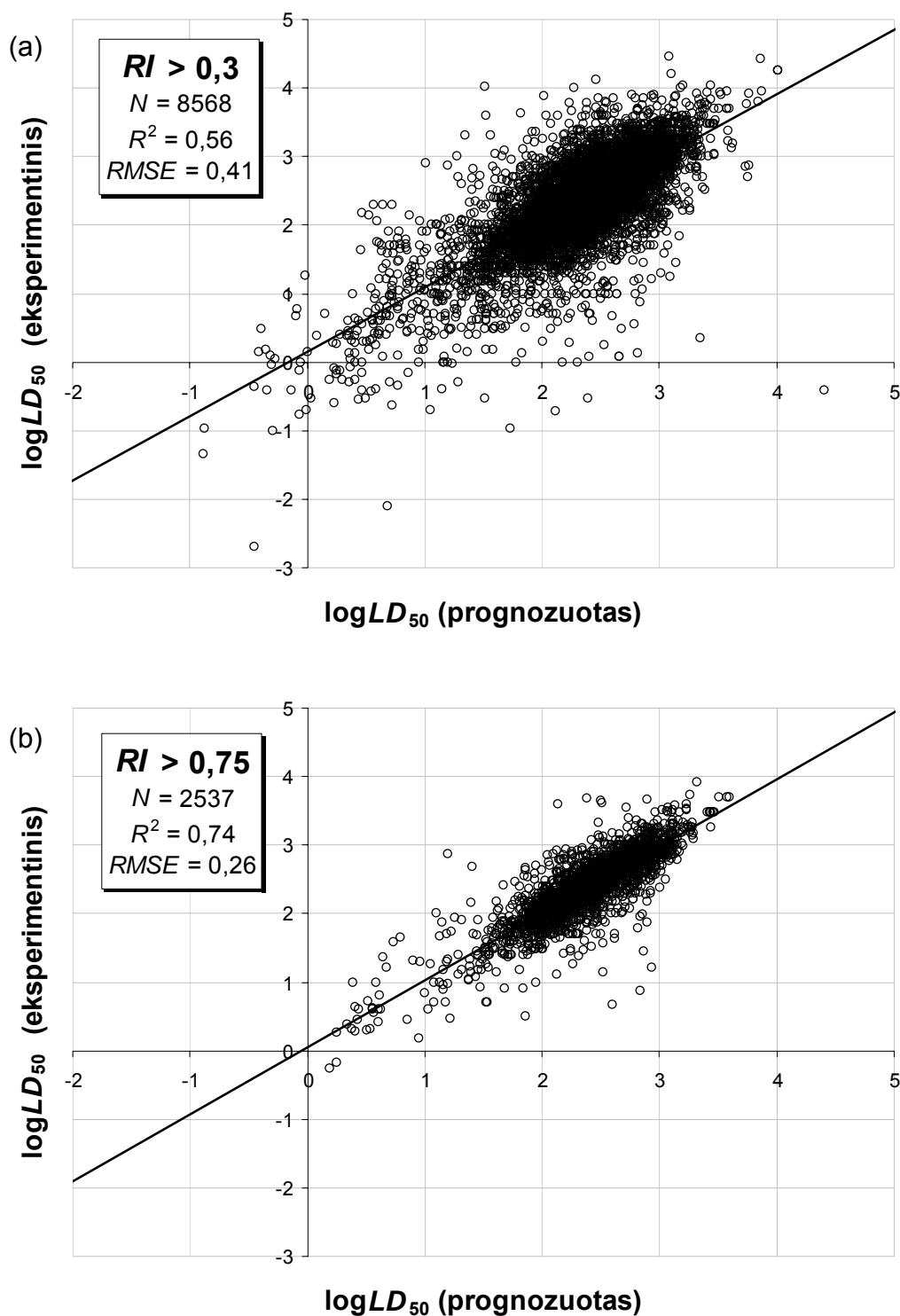


12 pav. Priklausomai nuo apskaičiuoto prognozės patikimumo indekso stebimų $RMSE$ reikšmių pasiskirstymo histogramos atitinkamoms graužikų rūšims ir tiriamųjų junginių patekimo į organizmą keliams (tęsinys)

mažesnės ir, kas svarbiausia, pradedant visų nagrinėjamųjų modelių pritaikomumo srities ribiniais regionais ir toliau ($RI > 0,3$), $RMSE$ ir RI priklausomybė praktiškai visada yra nuosekli. Šiuo faktu, be kita ko, ir yra grindžiamas prognozių suskirstymas į skirtingo patikimumo kategorijas.

Tokios klasifikacijos prasmingumu galima įsitikinti nagrinėjant 13 paveiksle pateiktus duomenis. Kaip pavyzdinė sistema šiai iliustracijai buvo pasirinktas ūmaus toksiškumo pelėms po intraperitoninės medžiagos įvesties modelis, kadangi jo atveju turimas patikrinamasis rinkinys yra pats didžiausias ir įvairiausias iš visų LD_{50} duomenų bazių naudotų šiame darbe. Grafikuose palyginama koreliacija tarp stebimų LD_{50} verčių ir skirtingas patikimumo klases apimančios prognozių dalies – pakankamo patikimumo ($RI > 0,3$; $N_{RI>0,3} = 8568$) bei aukšto patikimumo ($RI > 0,75$; $N_{RI>0,75} = 2537$). Tuo tarpu kai pirmasis grafikas (13 pav. (a)) primena išsibarsčiusių taškų debesį, kuriame beveik 5% prognozuotos LD_{50} reikšmės nuokrypių nuo eksperimentinių duomenų viršija 1 logaritminį vieneta, antruoju atveju (13 pav. (b)) stebimas vaizdas labiau primena tipinį QSPR atvejį su R^2 ir $RMSE$ reikšmėmis (atitinkamai 0,74 ir 0,26) labiau įprastomis ne LD_{50} , o kur kas paprastesnių savybių, pavyzdžiui $\log P_{O/W}$, modeliavimo atvejams. Aukšto patikimumo

prognozių atveju, paklaidų viršijančių 1 logaritminį vienetą skaičius taip pat sumažėja praktiškai dešimteriopai – iki maždaug 0,5%. Bendru atveju



13 pav. Eksperimentinių ir apskaičiuotųjų $\log LD_{50}$ reikšmių priklausomybė pasirinktam patikrinamajam rinkiniui (pelė, IP), esant dviem skirtingoms RI riboms

labiausiai tikėtina, kad tai, jog tokio dydžio paklaidos apskritai išlieka, yra eksperimentinių klaidų pasekmė. Aukštas RI reiškia, kad šiuo atveju apmokomajame rinkinyje ne tik randamas pakankamas kiekis labai artimų junginio analogų, bet ir visų jų eksperimentiniai LD_{50} matavimai yra modelio puikiai atkartojami. Taigi tikimybė stipriai apsirikti įvertinant junginio pavojingumą (toksišką junginį pripažinti saugiu ar atvirkščiai) didelio patikimumo prognozių atveju yra minimali realiame gyvenime, kadangi jį atspindintis eksperimentas, prieštaraujantis tokiai prognozei, greičiausiai „klysta“. Tačiau pats tokių paklaidų egzistavimo faktas reiškia, jog net esant $RI > 0,75$, praktinio prognozės taikymo atveju reikėtų išlaikyti deramą atsargumą. Nors šiuo atveju eksperimentinė klaida ir yra tikėtina tokių nukrypimų priežastis, teoriškai ji nėra vienintelė galima (pvz., taip pat įmanomas klaidingas panašumo nustatymo procedūros suveikimas ar pan.).

3.1.3 Išorinė modelio patikra

Galimybė atlikti pilnavertį išorinį modelio prognozuojamosios galios įvertinimą (angl. *external validation*) vienam iš šiame darbe pristatomų ūmaus toksiškumo modelių atsirado dėka vieno paskutiniųjų pastaruoju metu pasirodžiusio tyrimo, susijusio su LD_{50} reikšmių modeliavimu [36]. Pastarajame straipsnyje minimas 7385 junginių rinkinys, kuriems žinomas medžiagų, patenkančių į organizmą oraliniu keliu, ūmaus toksiškumo žiurkėms įvertinimas. Savo dydžiu šis yra palyginamas su analogiška duomenų baze naudota šiame darbe, todėl galima buvo pagrįstai tikėtis, jog tarp tokios duomenų gausos pavyks atrasti bent kiek didesnę skaičių nežinomų modeliui junginių, kas leistų atlikti pakankamai plataus masto modelio išbandymą su visiškai naujais junginiais. Darbinė minimo duomenų rinkinio versija (su visomis 7385 junginių struktūrinėmis formulėmis ir eksperimentiškai išmatuotomis LD_{50} (žiurkė, OR) vertėmis) buvo gauta susisiekus su šio straipsnio [36] autoriais. Kaip paaiškėjo, šio rinkinio dalis nesikertanti su vidine ūmaus toksiškumo žiurkei, esant oraliniam patekimo į organizmą būdai, duomenų baze yra sudaryta iš 2718 junginių. Statistinių parametrų rezultatai,

gaunami taikant darbe pristatomą oraliniu būdu įvedamų medžiagų toksiškumo žiurkėms modelį šiems junginiams pateikiami 7 lentelėje.

7 lentelė. LD_{50} (žiurkė, OR) modelio statistiniai rezultatai išoriniam patikrinamajam 2718 junginių rinkiniui

<i>RI</i> režis	<i>N</i>	R^2	<i>RMSE</i>	<i>MAE</i>
$RI > 0,3$	2501 (92%)	0,63	0,60	0,44
$RI > 0,5$	1804 (66%)	0,70	0,55	0,40
$RI > 0,75$	430 (16%)	0,81	0,44	0,30

Pirmasis į akis krentantis pastebėjimas yra tas, kad koreliacijos koeficiento kvadrato ir *RMSE* reikšmės junginiams patenkantiems į modelio pritaikomumo sritį ($RI > 0,3$) šiuo atveju yra praktiškai identiškoms analogiškomis galutinio orališkai į organizmą patenkančių medžiagų toksiškumo žiurkei modelio charakteristikoms vidinio patikrinamojo duomenų rinkinio atveju (žr. pirmą eilutę 6 lentelėje). Toks nagrinėjamo modelio veikimo palyginimas vidiniam ir išoriniam patikrinamajam junginių rinkiniui yra vienas iš pagrindinių aspektų nagrinėjamų išorinės patikros metu. Šiuo atveju gauti palyginami rezultatai vienareikšmiškai rodo, kad modelis nėra „persimokęs“, kadangi jo prognozavimo galia išlieka praktiškai nepakitusi nepriklausomai nuo patikrinamojo junginių rinkinio. Vidinis patikrinamasis rinkinys yra sudaromas atsitiktiniu būdu, tačiau vis dėlto jis atrenkamas iš tos pačios pradinės visų turimų junginių imties, kurios didžiąją dalį atspindi ir apmokomasis rinkinys. Taigi nors apylygiai (ar bent jau ne per daug drastiškai besiskiriantys) statistiniai modelio rezultatai apmokomajam ir vidiniam patikrinamajam junginių rinkiniams ir yra pirminis kriterijus, parodantis ar modelis nėra per daug prisitaikęs prie pradinių duomenų, rezultatų nuoseklumas dar ir išoriniam patikrinamajam junginių rinkiniui yra neabejotinai svaresnis argumentas modelio naudai.

7 lentelėje kartu su jau įprastinėmis *RMSE* ir R^2 taip pat pateikiamos vidutinių absoliutinių paklaidų (*MAE*) reikšmės. Straipsnyje iš kurio kilęs

išorinis patikrinamasis duomenų rinkinys [36] šis parametras buvo naudojamas kaip pagrindinis vidutinės modelio prognozės paklaidos matas, taigi palyginimo tikslais *MAE* reikšmės taip pat buvo apskaičiuotos ir šiame darbe. Savo tyrime [36] H. Zhu ir jo kolegos bando objektyviai palyginti penkis skirtingų tipų ūmaus toksiškumo žiurkėms po oralinio medžiagos įvedimo modelius su viena populiariausių kompiuterinių programų ūmaus toksiškumo prognozavimo srityje – TOPKAT[®] [39]. Tarp turimų 7385 junginių su išmatuotomis LD_{50} (žiurkė, OR) reikšmėmis jie identifikavo 3913, kurie nebuvo sutinkami šios programos apmokamajame duomenų rinkinyje (t.y. jos atžvilgiu galėjo būti panaudoti kaip išorinis patikrinamasis rinkinys). Naudodamiesi likusiais 3472 junginiais autoriai apmokė penkis skirtingomis ir gana įvairiomis metodikomis pagrįstus modelius. Kiekviename iš metodų buvo numatyta galimybė įvertinti modelio pritaikomumo sritį, nors tai ir tebuvo vienintelis kokybinis kriterijus – junginys arba patenka į šią sritį, arba ne. Procentinė į pritaikomumo sritį patenkančių junginių skaičiaus išraiška šiame darbe vadinama „apreptimi“ (angl. *coverage*). Stengdamiesi išgauti maksimalų tikslumą ir siekiant šio tikslo norėdami išnaudoti stipriausias kiekvieno modelio puses tuo pačiu apeinant jų trūkumus, jie taip pat sukūrė šeštą modelį, paremtą pirmųjų penkių konsensusu (angl. *consensus model*). T.y. iš esmės kiekvieno junginio atveju yra parenkamas modelis geriausiai tinkamas jo ūmaus toksiškumo prognozei. Nors imant visus penkis modelius atskirai, kiekvieno iš jų pritaikomumo sritis yra mažai įdomi visiems likusiems modeliams, tačiau konsensuso modelio atveju skirtingo dydžio atskirų modelių pritaikomumo sritys, apibrėžiamos skirtingos prigimties ir griežtumo kriterijais, iš esmės atitinka bendrojo modelio pritaikomumo srities dalis, kuriose galima tikėtis skirtingo tikslumo prognozių. Didžiausios „aprepties“ modelio pritaikomumo srityje esantiems junginiams labiau tikėtina situacija, kuomet bendra konsensuso prognozė gaunama tik iš dalies (ar apskritai tik iš vienintelės) visų turimų modelių prognozių, kadangi nemaža dalis tokių junginių tiesiog nepateks į mažesnės „aprepties“ modelių pritaikomumo sritį. Ir atvirkščiai, kuo mažesnė modelio „apreptis“, tuo joje esantiems junginiams

žinomų atskirų modelių prognozių, kurias galima panaudoti bendrame konsensuso modelyje, skaičius tikėtina yra didesnis. Didžiausia tikimybė, kad konsensuso prognozė yra paremta visų penkių turimų modelių individualiomis prognozėmis, be abejo yra mažiausios „aprėpties“ modelio pritaikomumo srityje esantiems junginiams. Bendru atveju, konsensuso modelyje didesnis turimų individualių prognozių skaičius leidžia tikėtis geresnės galutinės bendros prognozės kokybės. Be kita ko, straipsnio autoriai pastebi, jog gautieji rezultatai leidžia daryti išvadą apie egzistuojantį dėsningumą – kuo mažesnė (t.y. griežčiau apibrėžta) yra modelio pritaikomumo sritis, tuo joje gaunamų prognozių paklaidos yra mažesnės. Arba žvelgiant iš priešingos perspektyvos – siekiant geresnių modelio statistinių charakteristikų neišvengiamai tenka aukoti dalį jo pritaikomumo erdvės.

Apibendrinant šiuos samprotavimus, akivaizdu, kad kuo mažesnė bendro 3913 junginių išorinio patikrinamojo rinkinio dalis, atitinkanti kažkurio atskiro konkretaus modelio pritaikomumo sritį, yra imama domėn konsensuso modelio atveju, tuo gaunamos prognozės turėtų būti tikslesnės. Su šiuo modeliu susiję rezultatai, pateikiami 8 lentelės kairiojoje dalyje, akivaizdžiai iliustruoja šį faktą.

8 lentelė. Konsensuso ir TOPKAT[®] modelių rezultatų palyginimas nepriklausomam 3913 junginių patikrinamajam rinkiniui [36]

Aprėptis	Konsensuso modelis		TOPKAT [®]	
	R^2	MAE	R^2	MAE
74%	0,42	0,52	0,35	0,59
66%	0,48	0,51	0,25	0,70
19%	0,71	0,39	0,54	0,52

Pagal savo prigimtį skirtingo dydžio „aprėptys“, nagrinėjamos 8 lentelėje yra visiškai analogiškos junginių, patenkančių į tam tikrą *RI* režį, skaičiui šiame darbe pristatomo ūmaus toksiškumo žiurkei po oralinio medžiagos įvedimo modelio atveju. Šioje vietoje negalima nepastebėti, jog

H. Zhu ir kolegų straipsnyje [36] nagrinėjami konkretūs „apropčių“ dydžiai procentais (74%, 66% bei 19%) yra labai artimi į skirtingas patikimumo kategorijas patenkančių išorinio patikrinamojo rinkinio junginių santykinių dalių dydžiams šiame darbe aprašomos LD_{50} (žiurkė, OR) modelio patikros atveju (92%, 66% ir 16%, esant RI daugiau už atitinkamai 0,3, 0,5 ir 0,75 – žr. 7 lentelę). Be abejo, patikrinamieji duomenų rinkiniai, nuo kurių dydžio skaičiuojami šie procentai abiem atvejais yra pakankamai skirtingi (plg. 3913 ir 2718 junginių), taigi automatiškai skirsis tiek atitinkamas jų dalis sudarančių junginių absoliutus skaičius, tiek jų cheminės struktūros. Tai savo ruožtu neleidžia daryti tiesioginių griežtų išvadų apie tai, jog kažkuris iš šių modelių yra tikslesnis, geresnis, ir apskritai bet kuriuo požiūriu pranašesnis arba atvirkščiai. Tačiau 7 ir 8 lentelėse pateikiamos R^2 bei MAE reikšmės neabejotinai leidžia suprasti, jog šiame darbe aptariamas ūmaus oraliniu keliu į žiurkės organizmą patenkančių medžiagų toksiškumo modelis tikslumo požiūriu galima sakyti priklauso panašiai kategorijai kaip ir neseniai paskelbtas konsensuso modelis skirtas LD_{50} (žiurkė, OR) verčių prognozavimui [36]. Tuo tarpu pastarasis, objektyviai vertinant, visais atvejais buvo pranašesnis nei analogiškas TOPKAT[®] programos, skaitomos vienu iš etalonų ūmaus toksiškumo prognozavime, modelis.

3.1.4 Prognozių klasifikacija pagal patikimumą

Gera $RMSE$ ir RI verčių koreliacija (žr. 12 pav.) sudaro prielaidas gautus modelius panaudoti sėkmingai automatinei junginių atrankai (angl. *automated screening*) bei rūšiavimui pagal svarbą (angl. *prioritization*) prieš *in vitro* testus ar net *in vivo* ūmaus toksiškumo nustatymo eksperimentus su gyvūnais, priklausomai nuo vaisto kūrimo (angl. *drug development*) pakopos. Pavyzdžiui, junginiai kuriems prognozuojama maža LD_{50} reikšmė (t.y. tikėtinais toksiški) ir aukštas prognozės patikimumo indeksas RI (t.y. patikima, geros kokybės prognozė) galėtų būti pašalinami iš tolesnės analizės esant pirmai galimybei, netgi be jokių *in vitro* matavimų. Ir atvirkščiai, junginiai su didelėmis prognozuojamomis LD_{50} (t.y. tikėtinais netoksiški) ir žemomis

apskaičiuotomis *RI* (t.y. abejotina prognozė) vertėmis turėtų būti eksperimentiškai patikrinami pirmiausiai. Kaip pavaizduota 12 paveiksle, buvo pasirinktos trys ribinės *RI* reikšmės (0,3, 0,5 ir 0,75), leidžiančios visas prognozes suskirstyti į keturias kokybines klases (nepatikimos ir ribinės, vidutinės bei aukštos kokybės prognozės). 9 lentelėje pateikiamas junginių skaičiaus ir *RMSE* reikšmių pasiskirstymas, atsižvelgiant į šias keturias kategorijas, ūmaus intraperitoniniu būdu įvedamų medžiagų toksiškumo pelėms modelio patikrinamojo duomenų rinkinio atveju ($N = 9025$).

9 lentelė. Junginių pasiskirstymas pagal prognozių patikimumo klases, priskirtas atsižvelgiant į apskaičiuotas prognozės patikimumo indekso (*RI*) reikšmes pasirinktiems duomenų rinkiniams. Čia „Rinkinys Nr. 1“ – PubChem duomenų bazė ($N = 24857$) [138], „Rinkinys Nr. 2“ – pasaulinis vaistų sąrašas ($N = 4246$) [139]

Prognozės klasifikacija	<i>LD</i> ₅₀ patikrinamasis rinkinys (pelė, IP)			Rinkinys Nr. 1	Rinkinys Nr. 2
	<i>RI</i>	<i>N</i>	<i>RMSE</i>	<i>N</i>	<i>N</i>
Nepatikima	0,00 ÷ 0,30	459 (5%)	0,91	4658 (19%)	960 (23%)
Ribinė kokybė	0,30 ÷ 0,50	1482 (16%)	0,60	13144 (53%)	1482 (35%)
Vidutinė kokybė	0,50 ÷ 0,75	4549 (50%)	0,41	6854 (28%)	1515 (36%)
Aukšta kokybė	0,75 ÷ 1,00	2537 (28%)	0,26	201 (1%)	289 (7%)

Kaip jau buvo stebima visais ankstesniais atvejais, augant prognozės patikimumo indeksui *RMSE* reikšmės nuolat mažėja ir šį kartą. Tuo tarpu junginių skaičiaus pasiskirstymo kreivės maksimumas stebimas *RI* režyje tarp 0,5 ir 0,75, kas atitinka $RMSE < 0,5$. Tokia paklaida yra visiškai pakankama apytiksliam toksiškumo įvertinimui be jokių *in vitro* eksperimentų (pačiuose ankstyviausiuose junginio tyrimų etapuose). Kadangi prognozių, kurioms apskaičiuotas patikimumo indeksas *RI* yra aukštesnis už 0,5, skaičius sudaro 75% visų nagrinėjamojo duomenų rinkinio junginių, tokiu atveju vadovaujantis šia strategija tik 25% junginių būtų reikalingos papildomos išlaidos, susijusios

su *in vitro* eksperimentais. Šis procentas galėtų būti mažinamas ir dar labiau, pavyzdžiui, papildomai kaip kriterijų nustatant nuo kokios (ar atvirkščiai, iki kokios) prognozuotosios LD_{50} vertės junginiui taikyti eksperimentinius tyrimus. Be abejo, kiekvienu konkrečiu atveju, priklausomai nuo norimų pasiekti tikslų, reikėtų vadovautis skirtingais kriterijais. Tarkime vėlesnėse vaisto kandidato optimizavimo pakopose galėtų būti naudojamos jau kitokios RI ar LD_{50} ribinės reikšmės ir vienas iš daugybės įmanomų junginių atrankos eksperimentiniam testavimui scenarijų galėtų, pavyzdžiui, atrodyti taip. Pirmiausiai toksiškumas eksperimentiškai įvertinamas junginiams, kuriems LD_{50} prognozių patikimumo indeksas $RI < 0,3$ (grįžtant prie LD_{50} (pelė, IP) patikrinamojo rinkinio pavyzdžio 9 lentelėje tai būtų apie 5% visų junginių). Toliau ūmaus toksiškumo vertės eksperimentiškai išmatuojamos tam tikro dydžio ribinės kokybės ($RI = 0,30 \div 0,50$) LD_{50} prognozių atveju (konkrečiu junginių 9 lentelėje atveju tai galėtų sudaryti nuo keleto iki visų 16% rinkinio junginių, ribiniu atveju atsisakius prognozuotos LD_{50} vertės dydžio apribojimų) ir t.t.

3.1.5 Apskaičiuotų RI reikšmių pasiskirstymo dėsningumai įvairiuose duomenų rinkiniuose

Paskutiniai du 9 lentelės stulpeliai pateikia junginių skaičiaus pasiskirstymo skirtingose RI klasėse duomenis dviem nepriklausomiems junginių rinkiniams. Pirmasis iš jų surinktas iš PubChem duomenų bazės [138] ($N = 24857$), tuo tarpu antrasis yra pavyzdinė atranka ($N = 4246$) iš pasaulinio vaistų sąrašo [139]. Pirmojo rinkinio sudarymui iš visų PubChem duomenų bazę sudarančių įrašų buvo pasirinkta ta dalis molekulių, kurių identifikaciniai numeriai (ID) yra tarpe tarp 42575001 ir 42600000, t.y. iš viso 25000 junginių. Iš gautojo sąrašo buvo pašalinti visi neorganiniai junginiai ir medžiagų mišiniai. Taip pat visos aptiktos organinių elektrolitų jonizuotos formos (druskos) buvo atstatytos į neutralias (t.y. pakeistos atitinkamų laisvų rūgščių ar bazių struktūromis) tokiu būdu gaunant galutinį pirmojo rinkinio pavidalą.

Abi šios papildomai į analizę įtrauktos duomenų bazės yra niekaip nesusijusios su ūmaus toksiškumo reiškiniu, t.y. tarp šaltinių, iš kurių jos sudarytos, nefigūruoja nei su eksperimentinių LD_{50} reikšmių nustatymu, nei su jų modeliavimu susiję tyrimai. Tai savo ruožtu leidžia tikėtis išvengti potencialaus su toksiškumu susijusių duomenų rinkinių sudėčiai būdingo tendencingumo. Duomenų bazių nevienalytiškumo problema yra aktuali daugumos savybių ar biologinių aktyvumų atveju. Dažniausiai naujo konkrečiam tikslui kuriamo junginio savybės ar biologiniai aktyvumai gali būti traktuojami tik vienareikšmiškai – jie yra arba pageidaujami, arba nepageidaujami. Dėl šios priežasties tyrimuose dėmesys sutelkiamas ties tam tikromis junginių klasėmis, iš kurių iš anksto tikimasi vienokio ar kitokio rezultato. Be to, neigiamiems tyrimo rezultatams, t.y. junginiams, kuriems vis dėlto nustatytas nepakankamas teigiamas aktyvumas ar rastas ryškus nepageidaujamas efektas, skiriamas žymiai mažesnis dėmesys, o kartais jie ir apskritai nepublikuojami. Visa tai neretai priveda prie situacijos, kuomet viešai prieinamų eksperimentinių duomenų pasiskirstymas yra labai aiškiai pasislinkęs į junginių pasižyminčių arba nepasižyminčių konkrečiu biologiniu aktyvumu (analogiškai, turinčių aiškiai išreikštą didesnę ar mažesnę reikšmę tolydžių savybių atveju) pusę. Be jokios abejonės, visi šie faktoriai įtakoja ir struktūrinę tokių duomenų rinkinių įvairovę. Bendrai paėmus, ūmaus toksiškumo atveju galima teigti, jog nenuodingi junginiai praktiškai neegzistuoja – vienintelis klausimas yra reikalinga dozė. Taigi šiuo atveju situacija su prieinamų duomenų šaltinių vienalytiškumu yra pastebimai geresnė, kadangi eksperimentinis toksiškumo įvertinimas yra aktualus vos ne visiems junginiams ir to pasekoje išmatuotos LD_{50} reikšmės yra žinomos pakankamai dideliame skaičiui (pakanka pažvelgti į šiame darbe naudojamų duomenų rinkinių dydžius) pačių įvairiausių struktūrų bei nuodingumo laipsnio cheminių junginių. Tačiau vis dėlto išlieka labai tikėtina, jog pasirinkę netgi gerokai mažesnę, tačiau kitokios kilmės duomenų bazę (pvz., PubChem ar WDI) joje aptiksime ne vieną ir ne du menkai ūmaus toksiškumo požūriū išnagrinėtų struktūrinių klasių atstovus. Todėl modelio išbandymas su tokiais

nepriklausomais junginių rinkiniais gali pagerinti supratimą apie tai, kokios sukurtų modelių pritaikomumo erdvės ir apskaičiuotųjų RI verčių galėtume tikėtis jų taikydami realaus gyvenimo situacijose.

Iš pateiktų rezultatų matyti, jog $RI > 0,5$ atitinka maždaug 35% junginių (vidurkis abiem duomenų bazėms). Visų pirma reikėtų pastebėti, kad šis skaičius, lygiai kaip ir dalis prognozių, peržengiančių kitas nagrinėjamas patikimumo indekso ribines reikšmes, šiuo atveju yra žymiai mažesni nei toje pačioje 9 lentelėje pateikiami analogiški rezultatai LD_{50} (pelė, IP) modelio patikrinamojo rinkinio atveju. Tai patvirtina praeitoje pastraipoje iškelta hipotezę, jog panaudotas dvi nepriklausomas duomenų bazes sudarančių junginių ir darbe naudotų ūmaus toksiškumo duomenų rinkiniuose esančių molekulių cheminė įvairovė pastebimai skiriasi. Be kita ko tai taip pat reiškia, jog ankstyvuosiuose vaisto kūrimo proceso etapuose iki trečdaliao aptinkamų potencialių kandidatų į vaistinio junginio pirmtakus (angl. *lead candidate*) atvejų galima būtų išvengti eksperimentinių *in vitro* matavimų. Šioje vietoje galima prisiminti, kad *in vitro* testams būdingi tam tikri trūkumai [15], tuo tarpu eksperimentų su gyvūnais tokiose ankstyvose tyrimo pakopose niekas neatlieka. *In vivo* testų stengiamasi apskritai kiek įmanoma išvengti, taigi jie nebent atliekami vėlyvuose priešklinikiniuose tyrimuose, tačiau iki to laiko į konkretų junginį jau yra investuota daugybė pinigų, taigi tiek vaisto kūrėjai, tiek investuotojai tikrai nėra nusiteikę šioje vietoje sužinoti apie bet kokius stebimus jų junginio sukeltus ūmius reiškinius. Kita vertus, prognozių su apskaičiuotu $RI > 0,75$ skaičius gali nukristi ir iki vos 1% (PubChem atvejis). Taigi priešklinikiniuose tyrimuose ūmus toksiškumas vis dėlto turėtų būti eksperimentiškai įvertinamas visiems junginiams. Gera žinia šiuo atveju yra ta, kad absoliučia dauguma atvejų tai tėra vienas ar du likę kandidatai. Apskritai toks rezultatas neturėtų stebinti atsižvelgiant į vidinį pačiai LD_{50} savybei būdingą sudėtingumą ir tą faktą, kad panašaus pobūdžio ūmaus toksiškumo prognozių taikymo galimybės iki šiol net nebuvo rimtai vertinamos kaip įmanomos.

3.2 *Citochromų P450 šeimos metabolizmo fermentų (CYP3A4 ir CYP2D6) regioselektyvumo modeliavimo rezultatai*

3.2.1 **Bazinių ir lokaliai pakoreguotų GALAS modelio prognozių palyginimas**

Kaip aprašyta metodinėje dalyje (žr. 2.1.2 skyrių), gautųjų regioselektyvumo modelių charakteristikų vidinė patikra buvo vykdoma pasinaudojant patikrinamaisiais junginių rinkiniais, sudarančiais 30% pradinių duomenų poaibių kiekvieno iš dviejų fermentų ir keturių nagrinėjamų metabolizmo reakcijų tipų atveju. Žemiau esanti 10 lentelė pateikia tokio testavimo detalius rezultatus CYP3A4 N-dealkilinimo ir CYP2D6 O-dealkilinimo reakcijoms, kurios buvo pasirinktos kaip patys tipiškiausi nagrinėjamojo objekto pavyzdžiai. Tais pačiais dviem atvejais gaunamų modelių veikimą apibendrinančių statistinių charakteristikų (jautrio, specifiškumo ir bendro tikslumo) reikšmės išvardytos 11 lentelėje. Bendrai paėmus, jau bazinių prognozių atveju gaunami rezultatai yra visiškai patenkinami atsižvelgiant į sudėtingą nagrinėjamos savybės prigimtį (bendras modelio tikslumas CYP3A4 N-dealkilinimo ir CYP2D6 O-dealkilinimo atveju atitinkamai 72% ir 86%). Tačiau detalus žvilgsnis į pateiktus rezultatus taip pat atskleidžia ir keletą šiems baziniams modeliams būdingų problemų, pavyzdžiui šioki tokį specifiškumo (sugebėjimo teisingai identifikuoti neigiamus eksperimentinius stebėjimus) trūkumą.

11 lentelėje pateikiamos šio statistinio parametro reikšmės baziniams modeliams iš tiesų nėra tokios labai jau mažos, ypač tokio sudėtingumo modeliui. CYP2D6 O-dealkilinimo bazinio modelio specifiškumas yra artimas jautriui, tuo tarpu CYP3A4 N-dealkilinimo atveju jisai yra netgi pastebimai didesnis (plg., 73% su 67%), taigi atrodytų, jog tai tikrai nėra ta sritis, į kurią pirmiausiai reikėtų kreipti dėmesį ir, šiuo atveju atvirkščiai, labiau reikėtų susirūpinti modelio jautriu. Vis dėlto reikėtų nepamiršti keleto faktų. Visų pirma, 11 lentelėje pateikiamos statistinių parametrų vertės yra apskaičiuojamos įtraukiant neapibrėžtas prognozes (prognozuota tikimybė

10 lentelė. CYP3A4 N-dealkilinimo ir CYP2D6 O-dealkilinimo reakcijų regioselektyvumo prognozavimo rezultatai patikrinamiesiems duomenų rinkiniams

Modelis ir duomenų rinkinys		Apskaičiuota tikimybė (<i>p</i>)		
		< 0,4	0,4 ÷ 0,6	> 0,6
CYP3A4 N-dealkilinimas				
Bazinės prognozės Patikrinamasis rinkinys*	Neigiamas	116 (55,0%)	13 (6,2%)	30 (14,2%)
	Teigiamas	10 (4,7%)	7 (3,3%)	35 (16,6%)
GALAS modelis Patikrinamasis rinkinys*	Neigiamas	136 (64,5%)	8 (3,8%)	15 (7,1%)
	Teigiamas	8 (3,8%)	5 (2,4%)	39 (18,5%)
GALAS modelis <i>RI</i> > 0,5	Neigiamas	77 (68,8%)	2 (1,8%)	2 (1,8%)
	Teigiamas	2 (1,8%)	2 (1,8%)	27 (24,1%)
GALAS modelis <i>RI</i> > 0,7	Neigiamas	30 (71,4%)	0 (0,0%)	0 (0,0%)
	Teigiamas	1 (2,4%)	1 (2,4%)	10 (23,8%)
CYP2D6 O-dealkilinimas				
Bazinės prognozės Patikrinamasis rinkinys*	Neigiamas	128 (76,6%)	11 (6,6%)	11 (6,6%)
	Teigiamas	1 (0,6%)	1 (0,6%)	15 (9,0%)
GALAS modelis Patikrinamasis rinkinys*	Neigiamas	141 (84,4%)	7 (4,2%)	2 (1,2%)
	Teigiamas	2 (1,2%)	1 (0,6%)	14 (8,4%)
GALAS modelis <i>RI</i> > 0,5	Neigiamas	128 (88,3%)	3 (2,1%)	1 (0,7%)
	Teigiamas	2 (1,4%)	0 (0,0%)	11 (7,6%)
GALAS modelis <i>RI</i> > 0,7	Neigiamas	96 (93,2%)	0 (0,0%)	0 (0,0%)
	Teigiamas	0 (0,0%)	0 (0,0%)	7 (6,8%)

* Apskaičiuojant pateiktus rezultatus nėra atsižvelgiama į nepatikimas prognozes (*RI* < 0,3). Dėl šios priežasties iš pradinio patikrinamojo duomenų rinkinio CYP3A4 N-dealkilinimo atveju yra atmetami 37 pažymėti atomai (iš jų 10 yra teigiami metabolizmo centrai). CYP2D6 O-dealkilinimo atveju šie skaičiai yra atitinkamai 10 pažymėtų atomų ir 2 teigiami metabolizmo centrai. Detalus junginių pasiskirstymas pagal apskaičiuotą prognozių patikimumą abiem patikrinamiesiems rinkiniams pateikiami 12 lentelėje.

$p = 0,4 \div 0,6$), t.y. atspindinčios blogiausią įmanomą atvejį, kadangi apskaičiuojant visus išvardintus statistinius parametrus tokiu būdu, šios prognozės yra traktuojamos kaip klaidingos. Vertinant jas ne taip griežtai, t.y.

11 lentelė. Pateiktų detalių CYP3A4 N-dealkilinimo ir CYP2D6 O-dealkilinimo reakcijų regioselektyvumo modeliavimo rezultatų apibendrinimas

Modelis ir duomenų rinkinys	Jautris	Specifiškumas	Bendras tikslumas
CYP3A4 N-dealkilinimas			
Bazinės prognozės Patikrinamasis rinkinys *	67%	73%	72%
GALAS modelis Patikrinamasis rinkinys *	75%	86%	83%
GALAS modelis $RI > 0,5$	87%	95%	93%
GALAS modelis $RI > 0,7$	83%	100%	95%
CYP2D6 O-dealkilinimas			
Bazinės prognozės Patikrinamasis rinkinys	88%	85%	86%
GALAS modelis Patikrinamasis rinkinys	82%	94%	93%
GALAS modelis $RI > 0,5$	85%	97%	96%
GALAS modelis $RI > 0,7$	100%	100%	100%

12 lentelė. CYP3A4 N-dealkilinimo ir CYP2D6 O-dealkilinimo reakcijų regioselektyvumo patikrinamųjų duomenų rinkinių atomų pasiskirstymai pagal jiems prognozuotos metabolizmo tikimybės patikimumo indekso (RI) reikšmę

Prognozės klasifikacija	RI	N (teigiamų metabolizmo centrų)	N (pažymėtų atomų)
CYP3A4 N-dealkilinimas			
Nepatikima	0,0 ÷ 0,3	10 (16%)	37 (15 %)
Ribinė kokybė	0,3 ÷ 0,5	21 (34%)	99 (40 %)
Vidutinė kokybė	0,5 ÷ 0,7	19 (31%)	70 (28 %)
Aukšta kokybė	0,7 ÷ 1,0	12 (19%)	42 (17 %)
CYP2D6 O-dealkilinimas			
Nepatikima	0,0 ÷ 0,3	2 (11 %)	10 (6 %)
Ribinė kokybė	0,3 ÷ 0,5	4 (21 %)	22 (12 %)
Vidutinė kokybė	0,5 ÷ 0,7	6 (32 %)	42 (24 %)
Aukšta kokybė	0,7 ÷ 1,0	7 (37 %)	103 (58 %)

laikant tiesiog neapibrėžtomis prognozėmis, apie kurias nieko nežinoma, ir paprasčiausiai pašalinant iš statistinės analizės, apskaičiuojamos modelio veiklos charakteristikos šiek tiek pagerėja. Kadangi CYP3A4 N-dealkilinimo bazinio modelio atveju santykinai tokios neapibrėžtos prognozės yra dažnesnės tarp atomų, prie kurių eksperimentiškai yra stebimos metabolizmo reakcijos (7 iš 52 prieš 13 iš 159), teigiami pokyčiai susiję su jų atmetimu labiau atsiliepia šio modelio jautriui nei specifiškumui. Galiausiai šiuo būdu gaunamos ne tik didesnės, bet ir kiekvieno iš dviejų modelių atveju praktiškai identiškios jautrio bei specifiškumo charakteristikų vertės (atitinkamai 78% ir 79% CYP3A4 N-dealkilinimo bei 94% ir 92% CYP2D6 O-dealkilinimo baziniams modeliams). Kita vertus, tiek jautris, tiek specifiškumas yra santykinės charakteristikos, todėl netgi esant vienodoms jų reikšmėms situacija su klaidingomis prognozėmis abiejuose prognozuojamos tikimybės intervalo galuose gali būti pakankamai skirtinga. Šiame darbe naudojamuose metabolizmo regioselektyvumo duomenų rinkiniuose nemetabolizuojamų atomų skaičius viršija žinomų metabolizmo centrų skaičių nuo mažiausiai maždaug 3 (CYP3A4 N-dealkilinimas) iki daugiau nei 20 (CYP2D6 alifatinis ir aromatinis hidroksilinimas) kartų (žr. 4 lentelę). Automatiškai modelio specifiškumo charakteristikos (apibūdinančios sugebėjimą teisingai identifikuoti neigiamus eksperimentinius rezultatus) sumažėjimas tam tikru fiksuotu dydžiu (tarkime 1%) yra lygiai tiek pat kartų pavojingesnis absoliutaus atsirandančių klaidingų prognozių skaičiaus požiūriu už tokiu pačiu dydžiu sumažėjusį modelio jautrį, nusakantį modelio sugebėjimą aptikti teigiamus stebėjimus. Kitais žodžiais tariant, netgi esant pakankamai aukštomis modelio specifiškumo vertėms šiuo atveju vis tiek susidaro palyginus nemažas absoliutus kiekis klaidingai teigiamų (angl. *false positive*) prognozių – statistikoje vadinamų I-osios rūšies klaidomis. Žiūrint į detalius bazinių prognozių rezultatus patikrinamajam rinkiniui 10 lentelėje (analogiškai kaip ir ūmaus toksiškumo atveju iš rezultatų analizės visais atvejais pašalinamos nepatikimos prognozės – $RI < 0,3$) matyti, jog šių klaidingai teigiamų prognozių skaičius yra toks, kad abiem atvejais svarstant teigiamą modelio prognozę ($p > 0,6$) tėra maždaug 50% tikimybė,

kad ji teisinga (atitinkamai 35 iš 65 CYP3A4 N-dealkilinimo ir 15 iš 26 CYP2D6 O-dealkilinimo atvejais), o tai jau yra rimta problema. Situacija ženkliai pagerėja pritaikius panašumo analize pagrįstas lokalias korekcijas, tuo tarpu imant domėn tikrai vidutinės ir aukštesnės kokybės prognozes ($RI > 0,5$), kurios vis dar sudaro apie 50% patikrinamųjų rinkinių dydžio (žr. 12 lentelę), klaidingai teigiamos (kaip beje ir kelios buvusios klaidingai neigiamos) prognozės praktiškai visiškai išnyksta (žr. 10 lentelę).

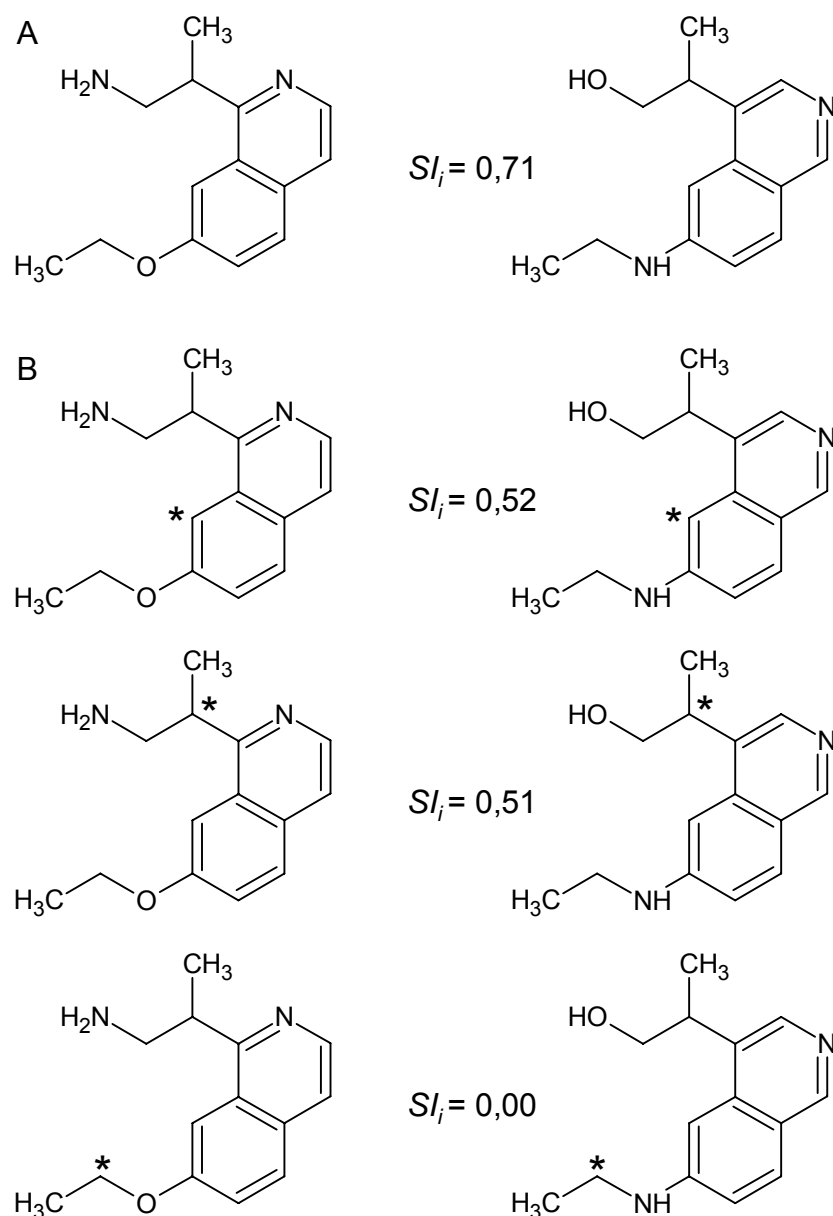
Kitas pastebėjimas susijęs su jau minėtomis neapibrėžtomis prognozėmis (prognozuota tikimybė $p = 0,4 \div 0,6$). Nors negalima sakyti, jog pradinis tokių prognozių skaičius yra nepriimtinais didelis (svyruoja nuo apytiksliai 10% iki 7% priklausomai nuo modelio), tačiau abiejų modelių atveju tokių prognozių skaičius akivaizdžiai mažėja augant nagrinėjamų rezultatų patikimumo indeksui ($RI > 0,5$ ir $RI > 0,7$). Tai savo ruožtu parodo prognozės patikimumo indekso įvertinimo metodikos sugebėjimą daugumą tarpiniame tikimybių intervale esančių prognozių identifikuoti kaip nepatikimas. Apibendrinant, galima teigti, kad panaudota GALAS modeliavimo metodika ir šiuo atveju buvo visapusiškai pranašesnė už paprastu BPLS metodu pagrįstą bazinį modelį.

3.2.2 Metabolizmo regioselektyvumo modelių pritaikomumo sritis

Kaip jau buvo užsiminta, metabolizmo fermentų regioselektyvumas yra savybė susijusi ne su visa molekule kaip vienu vienetu, o su kiekvienu individualiu ją sudarančiu atomu. Ar vyks fermento katalizuojama metabolinė reakcija prie konkretaus atomo priklauso nuo jo cheminės apsupties junginyje, kuriai savo ruožtu būdingas daug didesnis laisvės laipsnių skaičius, nei bendrai paimtos molekulės cheminei sudėčiai. Akivaizdu, kad iš to paties fiksuoto funkcinių grupių rinkinio, skirtingai jas išdėstant galime gauti labai didelę individualių molekules sudarančių atomų įvairovę. Tuo tarpu viso junginio cheminės sudėties požiūriu šios molekulės vis tiek priklausys tai pačiai funkcinei klasei (pvz., aminorūgštims, karboniliniams junginiams,

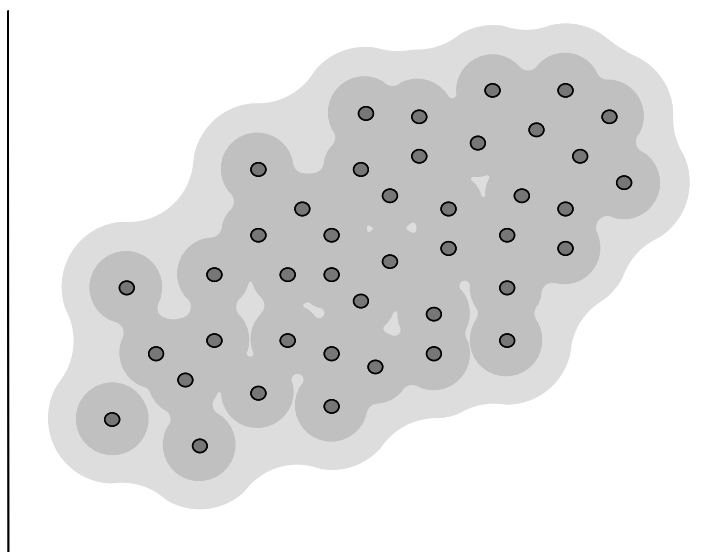
aromatiniais heterocikliniais junginiais, sotiesiems ar nesotiesiems angliavandeniliams ar bet kuriai kitai). Be abejo, didesnių ar mažesnių skirtumų atsirandančių pergrupuojant molekulę sudarančius sudėtinius blokus taip pat pasitaikys (pvz., α , β -nesotus ketonas virsta nekonjuguotu nesočiuoju ketonu, kondensuota žiedų sistema tampa keletu pavienių žiedų rinkiniu ir t.t., ir pan.). Tačiau bendru atveju, dviejų molekulių sudarytų iš skirtingai jų viduje išdėstytų identiškų funkcinių grupių panašumas CYP3A4 ar CYP2D6 metabolizmo vietos požiūriu bus pastebimai mažesnis nei visą molekulę charakterizuojančios savybės atveju (pvz., $\log P_{O/W}$). Akivaizdžiam šio fakto iliustravimui 14 paveiksle pateikiami dviejų vienodos atominės ir praktiškai identiškos funkcinės sudėties molekulių tarpusavio panašumo indeksų (SI_i) palyginimas $\log P_{O/W}$ bei CYP3A4 katalizuojamo metabolizmo prie kelių pasirinktų atomų atvejais. Kaip matyti, 14 paveikslo (B) dalies paskutiniosios atomų poros atveju panašumo koeficientas yra visiškai nulinis. Taip yra dėl to, kad šioje situacijoje jau tenka kalbėti apie skirtingus metabolizmo reakcijų, vykstančių prie šių atomų tipus (O- ir N-dealkilinimas), todėl bet koks bandymas taikyti tą patį modelį, t.y. netgi vertinti šių atomų panašumą vieno modelio kontekste, yra apskritai nekorektiškas. Iš šios situacijos sekančias išvadas galima būtų formuluoti dvejopai. Norint gauti tokią pačią prognozės patikimumo indeksą (RI) reikšmę metabolizmo regioselektyvumo modelio atveju apmokomajame rinkinyje yra reikalingi eksperto akimis žymiai panašesni junginiai. Analogiškai, esant vienodiems panašiausiems apmokomojo rinkinio junginiais, metabolizmo regioselektyvumo modelio atveju gaunama prognozė būtų mažiau patikima. Tačiau nepriklausomai nuo to, iš kurios pusės pažvelgsime į šiuos faktus, visų šių priežasčių pasekmė yra viena – metabolizmo regioselektyvumo QSAR modelių pritaikomumo erdvė visuomet yra žymiai glaudžiau susijusi su apmokomojo rinkinio junginiais, t.y. mažesnė, nei iš tų pačių junginių gauto visos molekulės savybę prognozuojančio modelio. O tai automatiškai dar labiau sumažina tikimybę, kad tokie modeliai gerai veiks cheminės erdvės srityse, esančiose bent kiek

toliau nuo apmokomojo rinkinio junginių. 15 paveiksle pamėginta vaizdžiai pailiustruoti iš tų pačių junginių gaunamų visos molekulės savybės (pvz., jau anksčiau minėtas $\log P_{O/W}$) ir kurios nors iš metabolizmo reakcijų regioselektyvumo modelių pritaikomumo srities dydžius.



14 pav. Molekulių tarpusavio panašumo indeksai $\log P_{O/W}$ (A) ir CYP3A4 fermento metabolizmo prie žvaigždutėmis pažymėtų atomų (B) atvejais

Viena vertus, šioje situacijoje ypatingai svarbus pasidaro GALAS modeliavimo metodikos sugebėjimas įvertinti gaunamų modelių pritaikomumo sritį. Kita vertus, dėl visų išvardintų faktorių, nagrinėjamų metabolizmo



15 pav. Visos molekulos savybės (šviesesnė pilka spalva) ir atomo-centrinės savybės (tamsesnė pilka spalva), pvz. metabolizmo regioselektyvumo, modelių gautų iš tų pačių junginių pritaikomumo srities palyginimas

regioselektyvumo modelių atveju didelės paklaidas lemiantis ekstrapoliavimas prasideda kur kas anksčiau nei tai buvo įprasta, pavyzdžiui, anksčiau aprašyto ūmaus toksiškumo atveju. Todėl šiuo atveju, ko gero, turėtų prasmės nusistatyti griežtesnį prognozės įtraukimo į bendrą rezultatų analizę kriterijų ir atsižvelgti tiktais į bent jau vidutinės ir aukštesnės kokybės prognozuotas tikimybes ($RI > 0,5$).

3.2.3 Regioselektyvumo prognozavimo rezultatai likusiems metabolizmo reakcijų tipams

Turint omenyje praeitame skyriuje išdėstytas mintis, visų tolesnių rezultatų aptarimų metu didžiausias dėmesys bus kreipiamas tik į pakankamo (vidutinio ir didesnio) patikimumo ($RI > 0,5$) prognozes. Atitinkami rezultatai likusiems šių dviejų fermentų metabolinių reakcijų tipams pateikiami 13 ir 14 lentelėse. Trumpai galima būtų tikrai paminėti, jog ir šiais atvejais problemos su atitinkamais baziniais modeliais ir GALAS modelio tikslumo augimo tendencijos atlikus lokalias panašumo analize pagrįstas korekcijas išliko tokios

13 lentelė. Vidutinės ir aukštesnės kokybės ($RI > 0,5$) CYP3A4 ir CYP2D6 fermentų regioselektyvumo modelių prognozės likusioms darbe nagrinėtoms reakcijoms

Reakcija	Apskaičiuota tikimybė (p)			
	< 0,4	0,4 ÷ 0,6	> 0,6	
CYP3A4 alifatinis hidroksilinimas	Neigiamas	347 (92,8%)	3 (0,8%)	2 (0,5%)
	Teigiamas	5 (1,3%)	0 (0,0%)	17 (4,5%)
CYP3A4 aromatinis hidroksilinimas	Neigiamas	394 (94,9%)	8 (1,9%)	2 (0,5%)
	Teigiamas	4 (1,0%)	1 (0,2%)	6 (1,4%)
CYP3A4 O-dealkilinimas	Neigiamas	108 (93,9%)	2 (1,7%)	0 (0,0%)
	Teigiamas	1 (0,9%)	0 (0,0%)	4 (3,5%)
CYP2D6 hidroksilinimas*	Neigiamas	825 (97,4%)	10 (1,2%)	2 (0,2%)
	Teigiamas	3 (0,4%)	1 (0,1%)	6 (0,7%)
CYP2D6 N-dealkilinimas	Neigiamas	177 (93,7%)	0 (0,0%)	1 (0,5%)
	Teigiamas	6 (3,2%)	1 (0,5%)	4 (2,1%)

* Modelis gautas iš jungtinio CYP2D6 alifatinio ir aromatinio hidroksilinimo duomenų rinkinio (daugiau komentarų šiuo klausimu pateikiama pagrindiniame tekste).

14 lentelė. 13 lentelėje pateikiamų likusių disertacijoje nagrinėtų metabolizmo reakcijų regioselektyvumo modeliavimo rezultatų apibendrinimas

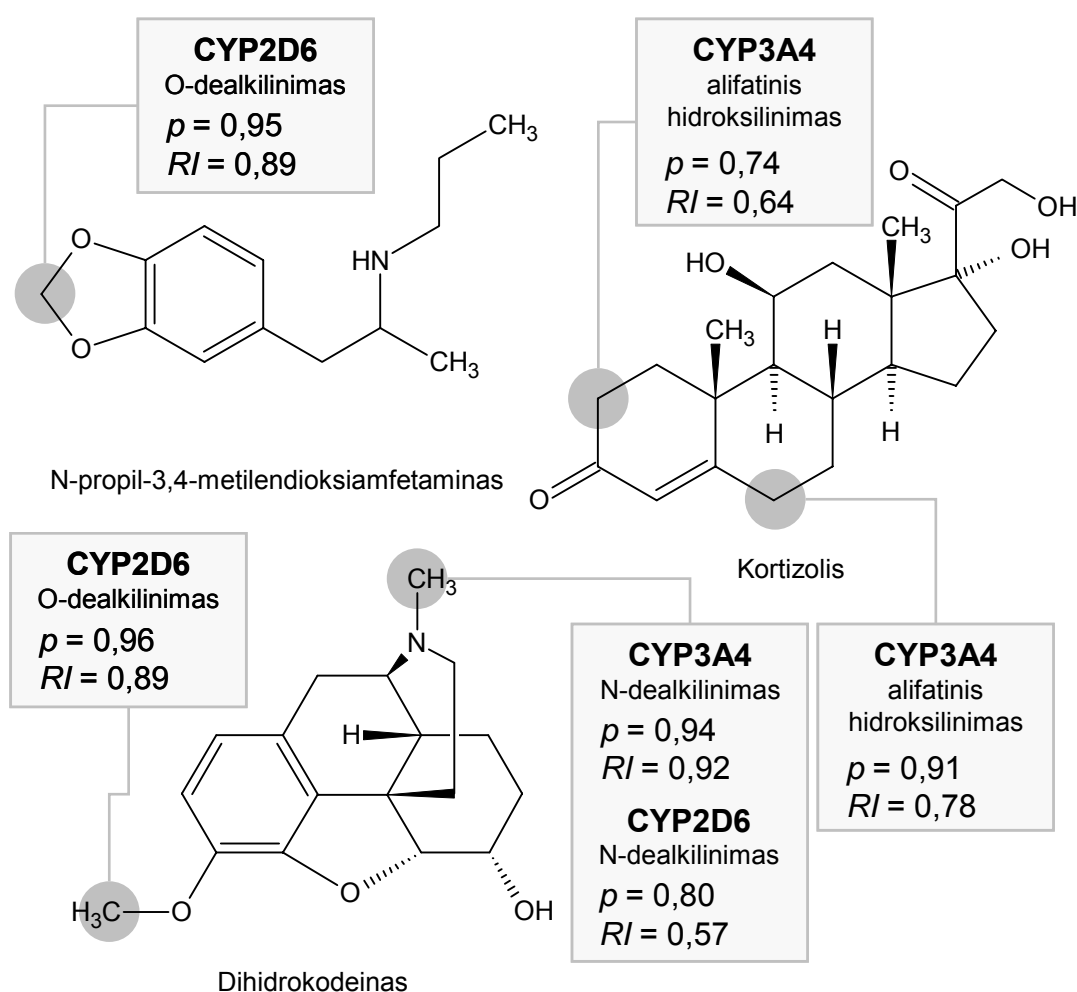
Reakcija	Jautris	Specifiškumas	Bendras tikslumas
CYP3A4 alifatinis hidroksilinimas	77%	99%	97%
CYP3A4 aromatinis hidroksilinimas	55%	98%	96%
CYP3A4 O-dealkilinimas	80%	98%	97%
CYP2D6 hidroksilinimas	60%	99%	98%
CYP2D6 N-dealkilinimas	36%	99%	96%

pačios kaip ir anksčiau detalai išnagrinėtoms CYP3A4 N-dealkilinimo bei CYP2D6 O-dealkilinimo reakcijoms. Kaip matyti CYP2D6 atveju pateikiami

rezultatai tikrai vienam bendram hidroksilinimo modeliui. Deja, CYP2D6 alifatinio hidroksilinimo atveju eksperimentiniai duomenys buvo nepakankami tam kad atskirai gauti prasmingą šio reakcijos tipo modelį. Konkrečiai, teigiamų metabolizmo centrų dalis tarp visų duomenų buvo per maža, o eksperimentiškai identifikuoti substratai buvo per daug skirtingi, kad galima būtų daryti bet kokias statistiškai patikimas išvadas naudojant modeliavimo metodiką, be viso kitko paremtą ir panašumo analize. Siekiant išvengti turimos jau surinktos informacijos nepanaudojimo (t.y. tiesiog praradimo), buvo pamėginta ją prijungti prie CYP2D6 aromatinio hidroksilinimo duomenų rinkinio. Nors pagal pasirinktą modeliavimo strategiją aromatiniai ir alifatiniai anglies atomai traktuojami kaip skirtingi hidroksilinimo reakcijos centrų tipai, tačiau, peržiūrėjus visus duomenų bazėje turimus eksperimentiškai identifikuotus šių dviejų tipų metabolizmo centrus, pasirodė, kad dauguma jų yra struktūriškai labai panašiuose junginiuose. Taigi šiuo atveju galima tikėtis ir daryti prielaidą, kad CYP2D6 katalizuojamos hidroksilinimo reakcijos vyksmas prie konkretaus anglies atomo yra labiau nulemiamas apskritai visos molekulės sugebėjimu prisijungti prie fermento aktyviojo centro ir su jo apsuptimi molekulėje nei su paties atomo chemine prigimtimi. Todėl hidroksilinimo atveju galima mėginti apjungti alifatinių ir aromatinių metabolizmo centrų duomenų bazes bei kurti bendrą modelį. Pats faktas, kad remiantis išsakyta prielaida pavyko sukurti modelį, sėkmingai identifikuojantį abiejų skirtingų tipų CYP2D6 hidroksilinimo centrus, praktiškai įrodo tokių samprotavimų bent jau dalinį pagrįstumą. Tačiau šioje vietoje derėtų paminėti, kad aprašytieji veiksmai tebuvo bandymas kaip nors išspręsti susiklosčiusią padėtį – galima pasakyti, renkantis mažesnę iš dviejų blygybių: jungtinio prielaidomis pagrįsto ir galimai ne tokio kokybiško modelio kūrimo bei visiško dalies duomenų nepanaudojimo. Turint pakankamai duomenų, korektiškas požiūris būtų svarstyti atskirų modelių kūrimą alifatinio ir aromatinio hidroksilinimo reakcijoms.

Disponuojant visų šiame darbe sukurtų žmogaus CYP450 metabolizmo fermentų šeimos atstovų regioselektyvumo modelių skirtingoms reakcijoms

ansambliu jau galime prognozuoti pačius pagrindinius cheminių junginių pirmosios fazės metabolitus (angl. *phase I metabolites*). 16 paveiksle pateikiamos apibendrintos visų modelių prognozės trims pasirinktiems junginiams, kuriuos sudarantys atomai nebuvo įtraukti į nei vieną iš modelių apmokymui naudotą duomenų rinkinį. Šiame pavyzdyje akcentuojamos su dideliu patikimumu ($RI > 0,5$) prognozuojamos metabolizmo vietos junginiuose ($p > 0,6$). Palyginimui galime prisiminti eksperimentinius metabolizmo tyrimų rezultatus tiems patiems trims junginiams, jau pateiktus anksčiau (žr. 7 pav.).



16 pav. Regioselektyvumo modelių prognozuojamų metabolizmo vietų molekulėse pavyzdžiai

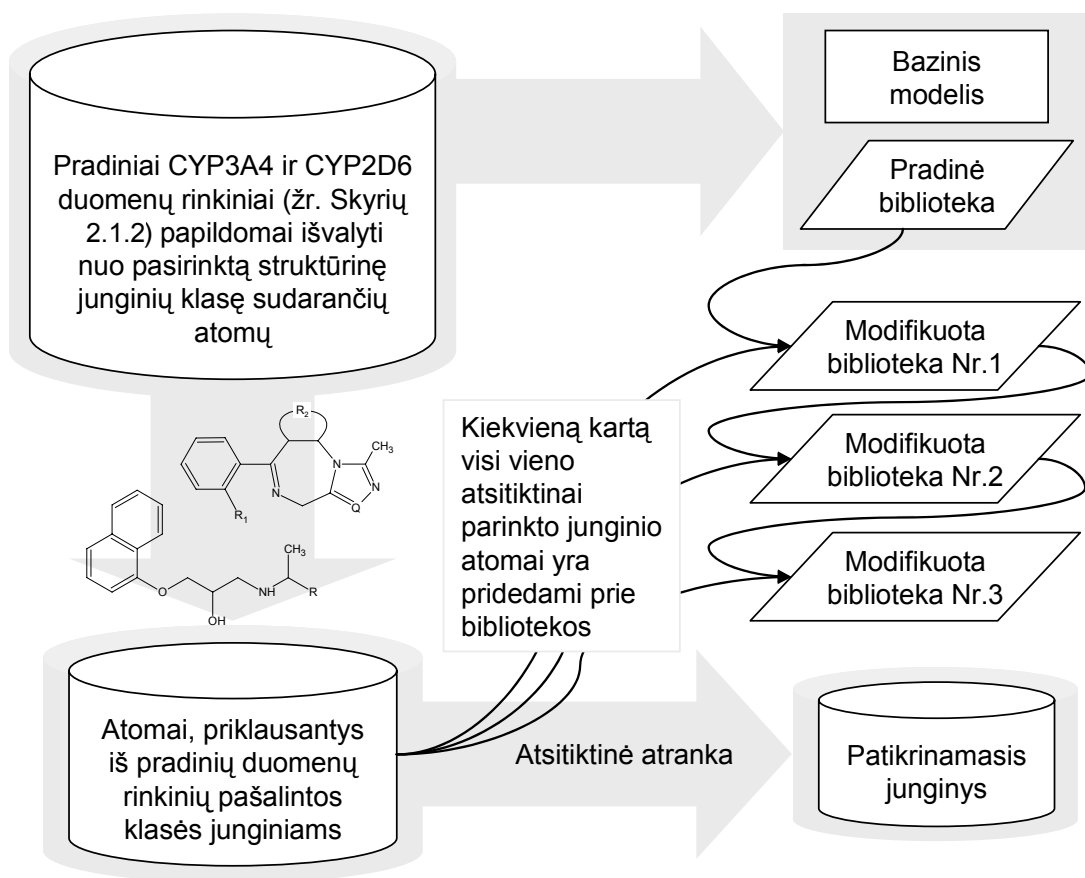
Grįžtant prie rezultatų pateikiamų 13 lentelėje sunku nepastebėti to fakto, kad bendrai paėmus patikimai prognozuojamų teigiamų metabolizmo

centrų skaičius yra palyginti labai mažas (daugeliu atveju mažesnis nei 10). Be abejo, reikia prisiminti, kad ir pradiniuose duomenų rinkiniuose eksperimentiškai identifikuotų metabolizmo vietų skaičius yra nedidelis (o patikrinamuosiuose duomenų rinkiniuose jisai yra atitinkamai dar mažesnis), taigi šiose grafose nereikėtų tikėtis kelių dešimčių eilės skaičių. Tačiau vis dėlto, tokie centrai be abejo egzistuoja, tuo tarpu labai aukšto patikimumo ($RI > 0,7$) prognozių atveju kai kuriems reakcijų tipams (pvz., CYP3A4 aromatiniam hidroksilinimui ar O-dealkilinimui) teigiamų prognozių skaičius yra lygus nuliui. Abu šie reakcijų tipai nėra tarp dažniausiai CYP3A4 fermento katalizuojamų reakcijų, tuo tarpu eksperimentiniuose citochromų P450 metabolizmo tyrimuose ir su jais susijusiuose literatūriniuose šaltiniuose labai dažnai dėmesys sutelkiamas ties pagrindinėmis tipinėmis konkretaus fermento reakcijomis (pvz., CYP3A4 atveju tokia paprastai laikoma N-dealkilinimo reakcija). Dėl mažesnio eksperimentinių duomenų kiekio ir ypač identifikuotų teigiamų metabolizmo centrų skaičiaus trūkumo mažėja ir atitinkamiems reakcijų tipams kuriamų fermento regioselektyvumo modelių pritaikomumo erdvė. Nors galimybė užtikrintai pasakyti, jog fermento katalizuojama konkreti metabolinė reakcija prie pasirinkto atomo nevyksta taip pat yra svarbi, žymiai aktualesnis yra į modelio pritaikomumo sritį patenkančių teigiamų centrų skaičius, kuris šiuo atveju ir nukenčia. Kitaip tariant, galimybė, kad šiame darbe pristatyti CYP450 fermentų izoformų regioselektyvumo modeliai paremti iš viešai prieinamų šaltinių surinktais duomenimis su dideliu patikimumu identifikuos antraeilis metabolitus (angl. *minor metabolites*) vienodai sėkmingai kaip ir pagrindinius, yra mažai tikėtina. Be abejo, šios aplinkybės, nepaisant visų kitų būdingų teigiamų savybių, įskaitant ir galimybę įvertinti pritaikomumo sritį bei konkrečios prognozės patikimumą, rimtai riboja šių modelių taikymo galimybes tais atvejais, kuomet domimasi visais įmanomais junginio metaboliniais kitimais. Šiuo atveju į pagalbą ateina dar viena labai naudinga GALAS metodu sukurtų modelių savybė – galimybė greitai išplėsti modelio pritaikomumo erdvę naujais junginiais.

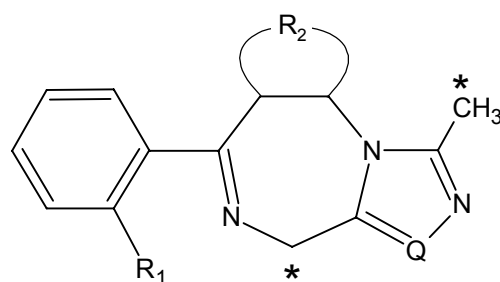
3.2.4 CYP3A4 ir CYP2D6 regioselektyvumo modelių apmokymo naujais junginiais bandymas

Šio darbo metu sukurtų GALAS regioselektyvumo modelių sugebėjimas apsimokyti naujais, modeliui anksčiau nežinomais junginiais buvo išbandytas vadovaujantis dviem testavimo scenarijais. Jų pagrindą sudaro specialiai sukurti atitinkamų CYP450 izoformų modeliai, į kuriuos specialiai neįtraukiami visi vienai kokiam nors struktūrinei junginių klasei bei visiems į ją panašioms analogams priklausantys atomai. Nuo minėtų atomų išgryninti apmokomieji rinkiniai buvo naudoti atitinkamiems baziniams modeliams gauti ir kaip pradinės panašių metabolizmo centrų bibliotekos lokalių korekcijų apskaičiavimo procedūrose. Šių modelių pagalba, iš neįtrauktųjų į modelį junginių tarpo atsitiktiniu būdu išrinktai patikrinamajai molekulei buvo prognozuotos labiausiai tikėtinos metabolizmo vietos. Vienas po kito dar trijų atsitiktinių tos pačios junginių klasės atstovų sudėtiniai atomai paeiliui pridedami prie GALAS modelių panašių metabolizmo centrų bibliotekų, naudojamų korekcijoms pagal panašumą nustatyti. Po kiekvieno naujų duomenų pridėjimo metabolizmo vietų prognozavimo procedūra bandomajam junginiui buvo kartojama, stebint gaunamų tikimybių ir prognozių patikimumo indeksų reikšmių dinamiką. Schemiškai visa tokio testavimo scenarijaus eiga pavaizduota 17 paveiksle.

CYP3A4 regioselektyvumo modelių atveju iš apmokomojo rinkinio buvo eliminuoti visi benzodiazepinų klasės junginiai. Tarp 14 atmestų molekulių buvo išskirta grupė junginių, turinčių savo struktūroje azolo žiedą ir dvi eksperimentiškai patvirtintas CYP3A4 alifatinio hidroksilinimo vietas [140-143]. Atsitiktiniu būdu iš šios grupės atrinktiems trimis junginiams (brotizolamas, triazolamas ir alprazolamas) priklausančius atomus charakterizuojantys eksperimentiniai rezultatai buvo pasilikti atitinkamų CYP3A4 regioselektyvumo modelių apmokymui, tuo tarpu standartiniam CYP3A4 substratui midazolamui teko patikrinamojo junginio vaidmuo (žr. 15 lentelę ir 18 paveikslą).

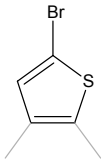
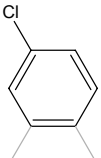
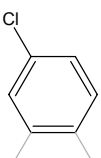
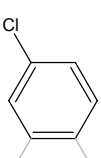


17 pav. GALAS regioselektyvumo modelių pritaikymo prie naujų junginių testavimo eksperimento eigos iliustracija



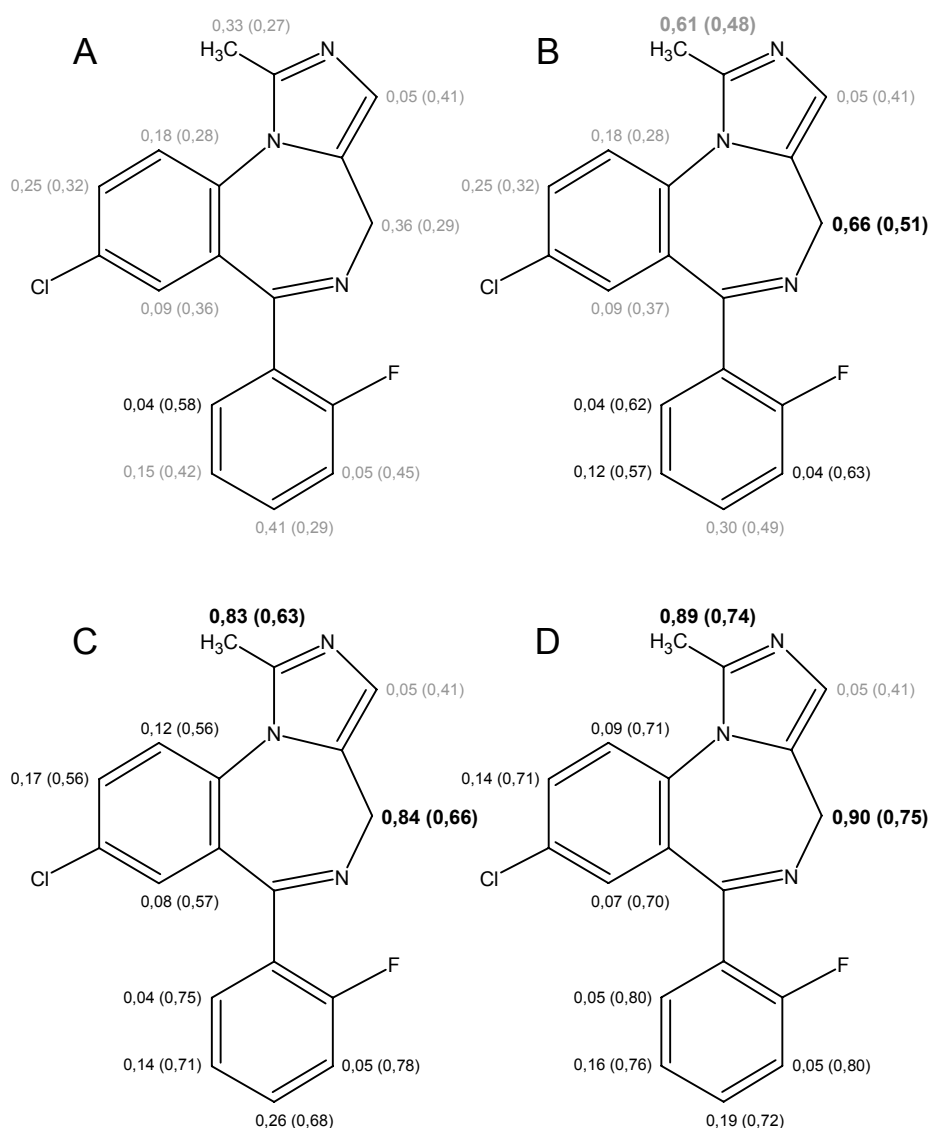
18 pav. Apibendrinta benzodiazepinų klasės junginių, naudotų CYP3A4 regioselektyvumo modelių apmokymo naujais junginiais eksperimente struktūra. Pakaitų reikšmės pateikiamos 15 lentelėje. Žvaigždutėmis pažymėtos eksperimentiškai nustatytos metabolizmo vietos

15 lentelė. Apmokymo naujais duomenimis virtualiame testavime naudotų benzodiazepino klasės junginių sudėtis (apibendrintą struktūrą žr. 18 pav.)

Pavadinimas	R ₁	R ₂	Q
Brotizolamas	-Cl		-N=
Triazolamas	-Cl		-N=
Alprazolamas	-H		-N=
Midazolamas	-F		-CH=

Gautų CYP3A4 regioselektyvumo modelių prognozių midazolamo molekulei rezultatai naudojant skirtingas panašių atomų bibliotekas pateikiami 19 paveiksle. Kaip matyti, nė viena iš eksperimentiškai žinomų CYP3A4 katalizuojamo metabolizmo vietų nėra aptinkama pradinio modelio, naudojančio originalią panašių metabolizmo centrų biblioteką, sudarytą iš apmokomojo rinkinio, į kurį specialiai neįtraukti benzodiazepinams artimų struktūrų atomai. Maža to, vieno vienintelio atomo atveju prognozės patikimumo indeksas viršija 0,5 ribą (juoda spalva). Kitais žodžiais tariant, turėdami 19 paveikslo (A) dalyje matomas prognozes, apie galimą junginio metabolizmo kelią praktiškai nieko neįmanoma spręsti. Situacija pradeda pastebimai keistis panašių atomų bibliotekoje, naudojamoje skaičiuojant lokalias korekcijas, palaipsniui daugėjant informacijos apie metabolizmo centrus, panašius į tiriamojo junginio atomus. Pridėjus prie bibliotekos eksperimentinius duomenis apie vykstančias (ar atvirkščiai nevykstančias) metaboline reakcijas ties brotizolamo ir triazolamo molekulių atomais (19

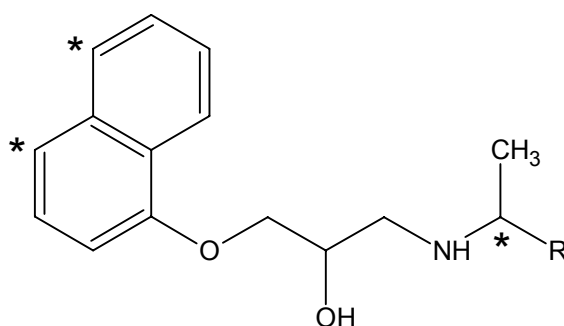
pav. (C)), beveik visų testuojamam junginiui gaunamų prognozių patikimumo indeksas išauga iki $RI > 0,5$. Be to, aiškiai išryškėja dvi molekulės vietos, kuriose, pasak modelio, su labai didele tikimybe galima tikėtis atitinkamų CYP3A4 fermento inicijuojamų cheminių pakitimų (19 paveiksle išskirtos paryškintu didesniu šriftu). Modelį papildžius informacija apie eksperimentiškai stebimas su CYP3A4 susijusias metabolizmo reakcijas alprazolamo molekulėje rezultatai nagrinėjamai midazolamo molekulei dar



19 pav. GALAS CYP3A4 regioselektyvumo modelių prognozės midazolamo molekulei: (A) naudojant modelį su pradine biblioteka, (B) pridėjus brotizolamo, (C) brotizolamo ir triazolamo bei (D) visų trijų benzodiazepinų eksperimentinius duomenis prie lokalsio panašumo korekcijos įvertinimo procedūros

labiau pagerėja – galiausiai absoliučios daugumos atomų atveju metabolizmo tikimybė prognozuojama su labai aukštu patikimumu ($RI > 0,7$).

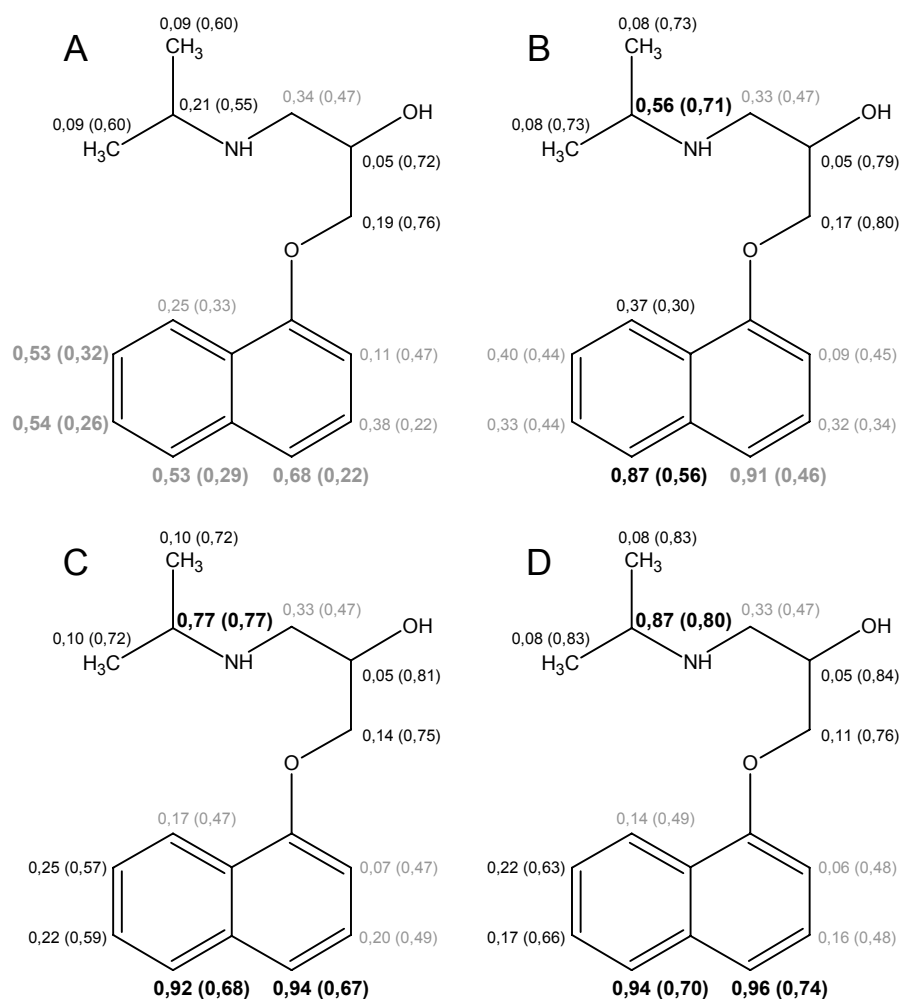
CYP2D6 regioselektyvumo modelių testavimas pagal analogišką scenarijų buvo atliktas pasinaudojant eksperimentiniais duomenimis propranololiui ir jo analogams. Iš viso iš pradinių duomenų rinkinių buvo atmesti 9-ių tokių junginių sudėtiniai atomai. Pagal analogiją su CYP3A4 atveju, iš šios grupės buvo pasirinkti 4 junginiai – vienas (pats propranololis) testavimui ir trys jo analogai modelio apmokymui (žr. 16 lentelę ir 20 paveikslą). Visiems nagrinėjamiems junginiams eksperimentiškai nustatytos trys CYP2D6 metabolizmo vietos (viena N-dealkilinimo ir dvi aromatinio hidrosilinimo) [144]. Naudojant keturias gautas panašių metabolizmo centrų bibliotekas gautos prognozės propranololio molekulės atomams pateikiamos 21 paveiksle. Visiškai analogiškai kaip ir ką tik nagrinėtame CYP3A4 variante, pradine panašių atomų biblioteka, neturinčia jokios informacijos apie propranololiui artimus junginius, paremtas GALAS modelis duoda visiškai neapibrėžtas prognozes (21 pav. (A)), iš kurių praktiškai neįmanoma daryti jokių išvadų. Pakankamo patikimumo prognozių ($RI > 0,5$) šiuo atveju gaunama šiek tiek daugiau, tačiau jokios informacijos apie labiausiai tikėtinas metabolizmo vietas junginyje jos nesuteikia, nes visos yra neigiamos. Turint omenyje eksperimentinius duomenis, šiuo atveju galime pastebėti, jog vienam iš tokių atomų prognozuojama labai maža tikimybė jį netgi klaidingai priskiria



20 pav. Apibendrinta propranololio ir jo analogų, naudotų CYP2D6 regioselektyvumo modelių apmokymo naujais junginiais eksperimente struktūra. Pakaitų reikšmės pateikiamos 16 lentelėje. Žvaigždutėmis pažymėtos eksperimentiškai nustatytos metabolizmo vietos

16 lentelė. Apmokymo naujais duomenimis virtualiame testavime naudotų propranololio ir jo analogų sudėtis (apibendrintą struktūrą žr. 20 pav.)

Pavadinimas	R
Propranololis (izopropil- analogas)	-CH ₃
2,2,2-trifluoro-1-metiletil- analogas	-CF ₃
1,2,2-trimetilpropil- analogas	-C(CH ₃) ₃
3,3,3-trifluoro-1-metilpropil- analogas	-CH ₂ CF ₃



21 pav. GALAS CYP2D6 regioselektyvumo modelių prognozės propranololio molekuli: (A) naudojant modelį su pradine biblioteka, (B) pridėjus vieną, (C) du bei (D) visus tris propranololio analogus ir jų eksperimentinius duomenis prie lokalaus panašumo korekcijos įvertinimo procedūros

prie molekulių vietų, kuriose CYP2D6 įtakojamas metabolizmas nevyksta. Kita vertus, visa eilė potencialių metabolizmo vietų aptinkama kondensuotoje aromatinių žiedų sistemoje, tačiau šių prognozių patikimumas yra žemas. Vienas po kito trijų pasirinktų propranololio analogų eksperimentinių CYP2D6 metabolizmo duomenų pridėjimas prie GALAS modelio lokali korekcijos pagal panašumą procedūroje naudojamos panašių metabolizmo centrų bibliotekos, palaipsniui ženkliai pagerina gaunamų prognozių patikimumą ir tikslumą (21 pav. (B)-(D)). Galutiniame rezultate visos trys eksperimentiškai stebimos CYP2D6 metabolizmo vietos propranololio molekulėje yra su dideliu patikimumu modelio identifikuojamos kaip labiausiai tikėtini metabolizmo centrai. Tuo tarpu du aromatiniai žiedai esantys atomai, pradinio modelio nurodyti kaip galimos metabolizmo vietos, galiausiai yra teisingai perklasifikuojami kaip molekulių vietos, neįtakojamos jokių CYP2D6 katalizuojamų reakcijų.

Išnagrinėti pavyzdžiai vaizdžiai patvirtina, kad nepaisant kai kurių čia aptartų trūkumų, sukurti CYP3A4 ir CYP2D6 fermentų regioselektyvumo modeliai, netgi dabartiniame jų išplėtojimo lygyje, turi potencialo praktiniam taikymui. Tai nulemia tokios naudingos jų savybės kaip galimybė kiekybiškai įvertinti gaunamų prognozių patikimumą bei, visų svarbiausia, galimybė išplėsti modelių pritaikomumo sritį apimant naujas, apmokomajam rinkiniui nežinomas junginių klases, dominančias tyrėją. Ypač turint omenyje, kad aprašytasis modelio veikimo pagerėjimas, pridėjus panašių metabolizmo centrų, pasiekiamas tuoj pat ir nereikalauja ilgai užtrunkančios statistinės bazinio modelio parametrizavimo iš naujo procedūros.

IŠVADOS

1. Sukurti ūmaus toksiškumo dviems graužikų rūšims ir visai eilei skirtingų medžiagos patekimo į organizmą būdų bei CYP3A4 ir CYP2D6 fermentų regioselektyvumo pagrindinėse jų katalizuojamose metabolinėse reakcijose prognozavimo modeliai (iš viso 13 individualių modelių). Absoliučiai visais atvejais buvo pademonstruotas GALAS metodikos pranašumas lyginant su tiesiniu globaliu PLS ar BPLS paremtu modeliu, patvirtinantis reikšmingą įvairių netiesinių efektų lokaliuose cheminės erdvės srityse įtaką nagrinėjamųjų modeliavimo objektų atveju.
2. Vieno iš sukurtųjų ūmaus toksiškumo modelio išorinės patikros metu gauti rezultatai leidžia teigti, jog kalbant vien tik apie prognozių tikslumą, t.y., neatsižvelgiant į papildomus GALAS metodo suteikiamus pranašumus, šio darbo rezultatai yra palyginami su naujausiais kitų autorių pasiekimais ir pastebimai geresni lyginant su vienu populiariausiu ir dažniausiai praktikoje naudojamu metodu – TOPKAT[®].
3. Visiems modeliams stebimas prognozių paklaidų ar klaidingų klasifikacijų ir neapibrėžtų prognozių skaičiaus mažėjimas kylant apskaičiuotam prognozės patikimumo indeksui (*RI*) akivaizdžiai parodo šio GALAS metodikos sudėtinio komponento sugebėjimą sėkmingai aprašyti gautųjų modelių pritaikomumo sritį. Taigi darbe pristatytieji modeliai atitinka vieną pagrindinių alternatyviems tyrimo metodams ES institucijų keliamų reikalavimų.
4. Aiški prognozių tikslumo priklausomybė nuo jų patikimumo įvertinimo suteikia galimybę jas suskaidyti į kokybines klases pagal apskaičiuotąsias *RI* reikšmes. Tai savo ruožtu įgalina junginių prioritetizavimą prieš eksperimentinius matavimus ir priklausomai nuo

vaisto kūrimo pakopos bei keliamų uždavinių sumažinti pastarųjų skaičių.

5. Pademonstruotas GALAS modelio sugebėjimas greitai ir efektyviai apsimokyti naujais eksperimentiniais duomenimis, išplečiant jo pritaikomumo sritį, ženkliai padidina pristatomųjų modelių praktinio panaudojimo potencialą, netgi nepaisant kai kurių jų pradinių trūkumų, pavyzdžiui, ribotos CYP3A4 ir CYP2D6 regioselektyvumo modelių galimybės identifikuoti antraeilius bei nereikšmingus metabolitus.
6. GALAS modeliavimo metodas buvo sėkmingai pritaikytas kokybinių atomo-centrinių savybių prognozavimui, panaudojant pasiūlytą sluoksninės fragmentacijos metodą ir kitus patobulinimus. Šis faktas suteikia vilčių analogišką modeliavimo strategiją pritaikyti kitų svarbių metabolizmo fermentų (pvz., CYP2C9, CYP2C19, CYP1A2 ir kitų) regioselektyvumo duomenų analizei, tęsiant šios temos nagrinėjimą ateityje.
7. Parodyta, jog visų pristatomų modelių pagrindą sudaranti GALAS modeliavimo metodika atveria kai kurias ne visai tradicines jų panaudojimo galimybes. Pavyzdžiui, iškrentančių prognozių analizė ir naujų mechanistiniu požiūriu įdomių junginių klasių paieška, eksperimentinių matavimų klaidų identifikavimas ir kiti.

A. SAZONOVO STRAIPSNIŲ SĄRAŠAS

1. Sazonovas, A., Japertas, P., and Didziapetris, R., *Estimation of reliability of predictions and model applicability domain evaluation in the analysis of acute toxicity (LD₅₀)*. SAR QSAR Environ. Res., 2010, 21, 127-148.
2. Dapkunas, J., Sazonovas, A., and Japertas, P., *Probabilistic prediction of the human CYP3A4 and CYP2D6 metabolism sites*. Chem. Biodiv., 2009, 6, 2101-2106.

LITERATŪROS SĄRAŠAS

1. Lombardo, F., Gifford, E., and Shalaeva, M. Y., *In silico ADME prediction: data, models, facts and myths*. *Mini. Rev. Med. Chem.*, 2003, 3, 861-875.
2. Ekins, S., *In silico approaches to predicting drug metabolism, toxicology and beyond*. *Biochem. Soc. Trans.*, 2003, 31, 611-614.
3. Afzelius, L., Arnby, C. H., Broo, A., Carlson, L., Isaksson, C., Jurva, U., Kjellander, B., Kolmodin, K., Nilsson, K., Raubacher, F., and Weidolf, L., *State-of-the-art tools for computational site of metabolism predictions: comparative analysis, mechanistical insights, and future applications*. *Drug. Metab. Rev.*, 2007, 39, 61-86.
4. Kerns, E. H. and Di, L., *Drug-like Properties: Concepts, Structure Design and Methods: from ADME to Toxicity Optimization*. Academic Press, Boston, MA, 2008.
5. Ahlstrom, M. M., Ridderstrom, M., and Zamora, I., *CYP2C9 structure-metabolism relationships: substrates, inhibitors, and metabolites*. *J. Med. Chem.*, 2007, 50, 5382-5391.
6. Japertas, P., *Fragmentinių ir statistinių metodų naudojimas biologiškai aktyvių junginių struktūros-aktyvumo sąryšio tyrimuose*. Daktaro disertacija, Vilniaus Universitetas (Chemija), Vilnius, 2007.
7. Japertas, P., Sazonovas, A., Didziapetris, R., and Petrauskas, A., *Similarity based assessment of model applicability domain and quantitative evaluation of the reliability of the prediction*. *Abstr. Paper. Am. Chem. Soc.*, 2008, 235, Meeting abstract 271-COMP.
8. Rendic, S. and DiCarlo, F. J., *Human cytochrome P450 enzymes: a status report summarizing their reactions, substrates, inducers and inhibitors*. *Drug. Metab. Rev.*, 1997, 29, 413-580.
9. Kalgutkar, A. S. and Didiuk, M. T., *Structural alerts, reactive metabolites, and protein covalent binding: how reliable are these*

- attributes as predictors of drug toxicity?* Chem. Biodivers., 2009, 6, 2115-2137.
10. Jaeschke, H., Gores, G. J., Cederbaum, A. I., Hinson, J. A., Pessayre, D., and Lemasters, J. J., *Mechanisms of hepatotoxicity*. Toxicol. Sci., 2002, 65, 166-176.
 11. *IUCLID Chemical Data Sheets Information System*, The Consumer Products Safety and Quality Unit of the Institute for Health and Consumer Protection, Joint Research Centre, European Commission, Ispra (VA), Italy. <http://ecb.jrc.ec.europa.eu/esis/index.php?PGM=dat>
 12. *The Registry of Toxic Effects of Chemical Substances (RTECS[®])*, Symyx Technologies, Inc., 415 Oakmead Parkway, Sunnyvale, CA, 94085, US. www.symyx.com
 13. *ChemIDplus database*, Specialized Information Services of the National Library of Medicine, Bethesda (MD), United States. <http://chem.sis.nlm.nih.gov/chemidplus/>
 14. Devillers, J. and Devillers, H., *Prediction of acute mammalian toxicity from QSARs and interspecies correlations*. SAR QSAR Environ. Res., 2009, 20, 467-500.
 15. Zmuidinavicius, D., Japertas, P., Petrauskas, A., and Didziapetris, R., *Progress in toxinformatics: the challenge of predicting acute toxicity*. Curr. Top. Med. Chem., 2003, 3, 1301-1314.
 16. Didziapetris, R., Reynolds, D. P., Japertas, P., Zmuidinavicius, D., and Petrauskas, A., *In silico technology for identification of potentially toxic compounds in drug discovery*. Curr. Comput.-Aided Drug Des., 2006, 2, 95-103.
 17. Tsakovska, I., Lessigiarska, I., Netzeva, T., and Worth, A., *Review of (Q)SARs for mammalian toxicity*, EUR 22486 EN, Institute for Health and Consumer Protection, Joint Research Centre, European Commission, Ispra (VA), Italy, 2006. http://ecb.jrc.ec.europa.eu/documents/QSAR/QSAR_Review_Mammalian_Toxicity.pdf

18. Enslein, K. and Craig, P. N., *A toxicity estimation model*. J. Environ. Pathol. Toxicol., 1978, 2, 115-121.
19. Enslein, K., Lander, T. R., Tomb, M. E., and Craig, P. N., *A Predictive Model for Estimating Rat Oral LD₅₀ Values*. Princeton Scientific Publishers, Princeton, NJ., 1983.
20. Tichy, M., Trecka, V., Roth, Z., and Krivucova, M., *QSAR analysis and data extrapolation among mammals in a series of aliphatic alcohols*. Environ. Health Perspect., 1985, 61, 321-328.
21. Enslein, K., Tuzzeo, T. M., Borgstedt, H. H., Blake, B. W., and Hart, J. B., *Prediction of rat oral LD₅₀ from Daphnia Magna LC₅₀ and chemical structure*. QSAR Environ. Toxicol. Proc. Int. Workshop, 1987, 2, 91-106.
22. Enslein, K., Lander, T. R., Tomb, M. E., and Craig, P. N., *A predictive model for estimating rat oral LD₅₀ values*. Toxicol. Ind. Health., 1989, 5, 261-387.
23. Jaeckel, H. and Klein, W., *Prediction of mammalian toxicity by quantitative structure-activity relationships: Aliphatic amines and anilines*. Quant. Struct.-Act. Relat., 1991, 10, 198-204.
24. Lipnick, R. L., *Outliers: their origin and use in the classification of molecular mechanisms of toxicity*. Sci. Total Environ., 1991, 109-110, 131-153.
25. Zakarya, D., Larfaoui, E. M., Boulaamail, A., and Lakhlifi, T., *Analysis of structure-toxicity relationships for a series of amide herbicides using statistical methods and neural network*. SAR QSAR Environ. Res., 1996, 5, 269-279.
26. Johnson, S. R. and Jurs, P. C., *Prediction of acute mammalian toxicity from molecular structure for a diverse set of substituted anilines using regression analysis and computational neural networks*. In *Computer-Assisted Lead Finding and Optimization. Current Tools for Medicinal Chemistry*, Eds.: van de Waterbeemd, H., Testa, B., and Folkers, G. Wiley-WCH, Basel, 1997. p. 31-48.

27. Wang, G. and Bai, N., *Structure-activity relationships for rat and mouse LD₅₀ of miscellaneous alcohols*. Chemosphere, 1998, 36, 1475-1483.
28. Eldred, D. V. and Jurs, P. C., *Prediction of acute mammalian toxicity of organophosphorus pesticide compounds from molecular structure*. SAR QSAR Environ. Res., 1999, 10, 75-99.
29. Cronin, M. T., Dearden, J. C., Duffy, J. C., Edwards, R., Manga, N., Worth, A. P., and Worgan, A. D., *The importance of hydrophobicity and electrophilicity descriptors in mechanistically-based QSARs for toxicological endpoints*. SAR QSAR Environ. Res., 2002, 13, 167-176.
30. Hansch, C. and Kurup, A., *QSAR of chemical polarizability and nerve toxicity*. 2. J. Chem. Inf. Comput. Sci., 2003, 43, 1647-1651.
31. Devillers, J., *Prediction of mammalian toxicity of organophosphorus pesticides from QSTR modeling*. SAR QSAR Environ. Res., 2004, 15, 501-510.
32. Jean, P. A., Gallavan, R. H., Kolesar, G. B., Siddiqui, W. H., Oxley, J. A., and Meeks, R. G., *Chlorosilane acute inhalation toxicity and development of an LC₅₀ prediction model*. Inhal. Toxicol., 2006, 18, 515-522.
33. Guo, J. X., Wu, J. J., Wright, J. B., and Lushington, G. H., *Mechanistic insight into acetylcholinesterase inhibition and acute toxicity of organophosphorus compounds: A molecular modeling study*. Chem. Res. Toxicol., 2006, 19, 209-216.
34. Freidig, A. P., Dekkers, S., Verwei, M., Zvinavashe, E., Bessems, J. G., and van de Sandt, J. J., *Development of a QSAR for worst case estimates of acute toxicity of chemically reactive compounds*. Toxicol. Lett., 2007, 170, 214-222.
35. Toropov, A. A., Rasulev, B. F., and Leszczynski, J., *QSAR modeling of acute toxicity for nitrobenzene derivatives towards rats: Comparative analysis by MLRA and optimal descriptors*. QSAR Comb. Sci., 2007, 26, 686-693.

36. Zhu, H., Martin, T. M., Ye, L., Sedykh, A., Young, D. M., and Tropsha, A., *Quantitative structure-activity relationship modeling of rat acute toxicity by oral exposure*. Chem. Res. Toxicol., 2009, 22, 1913-1921.
37. Lipnick, R. L., *Correlative and mechanistic QSAR models in toxicology*. SAR QSAR Environ. Res., 1999, 10, 239-248.
38. Lipnick, R. L., *Narcosis: fundamental and baseline toxicity mechanism for non-electrolyte organic chemicals*. In *Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*, Eds.: Karcher, W. and Devillers, J. Kluwer Academic Publishers, Dordrecht, 1990. p. 129-144.
39. *TOPKAT*[®], Accelrys Software Inc., San Diego CA, US, <http://accelrys.com/>, 2007.
40. Jaworska, J. S., Comber, M., Auer, C., and Van Leeuwen, C. J., *Summary of a workshop on regulatory acceptance of (Q)SARs for human health and environmental endpoints*. Environ. Health. Perspect., 2003, 111, 1358-1360.
41. Simon-Hettich, B., Rothfuss, A., and Steger-Hartmann, T., *Use of computer-assisted prediction of toxic effects of chemical substances*. Toxicology, 2006, 224, 156-162.
42. Schultz, T. W. and Seward, J. R., *Health-effects related structure-toxicity relationships: a paradigm for the first decade of the new millennium*. Sci. Total Environ., 2000, 249, 73-84.
43. Lessigiarska, I., Worth, A. P., Netzeva, T. I., Dearden, J. C., and Cronin, M. T. D., *Quantitative structure-activity-activity and quantitative structure-activity investigations of human and rodent toxicity*. Chemosphere, 2006, 65, 1878-1887.
44. Seibert, H., Balls, M., Fentem, J. H., Bianchi, V., Clothier, R. H., Dierickx, P. J., Ekwil, B., Garle, M. J., Gomez-Lechon, M. J., Gribaldo, L., Gulden, M., Liebsch, M., Rasmussen, E., Roguet, R., Shrivastava, R., and Walum, E., *Acute toxicity testing in vitro and the classification and*

- labeling of chemicals. The report and recommendations of ECVAM workshop 16.* Altern. Lab. Anim., 1996, 24, 499-510.
45. Garle, M. J., Fentem, J. H., and Fry, J. R., *In vitro cytotoxicity tests for the prediction of acute toxicity in vivo.* Toxicol. In Vitro, 1994, 8, 1303-1313.
 46. Shrivastava, R., Delomenie, C., Chevalier, A., John, G., Ekwai, B., Walum, E., and Massingham, R., *Comparison of in vivo acute lethal potency and in vitro cytotoxicity of 48 chemicals.* Cell Biol. Toxicol., 1992, 8, 157-170.
 47. Phillips, J. C., Gibson, W. B., Yam, J., Alden, C. L., and Hard, G., *Survey of the QSAR and in vitro approaches for developing non-animal methods to supersede in vivo LD₅₀ test.* Food Chem. Toxicol., 1990, 28, 375-394.
 48. Kaiser, K. L. E., McKinnon, M. B., and Fort, F. L., *Interspecies toxicity correlations of rat, mouse and Photobacterium phosphoreum.* Environ. Toxicol. Chem., 1994, 13, 1599-1606.
 49. Fort, F. L., *Correlation of Microtox EC₅₀ with mouse LD₅₀.* In Vitro Toxicol., 1992, 5, 73-82.
 50. Devillers, J., Pandard, P., Thybaud, E., and Merle, A., *Interspecies correlations for predicting the acute toxicity of xenobiotics.* In *Ecotoxicology Modeling*, Ed., Devillers, J. Springer, New York, 2009. p. 85-115.
 51. Calleja, M. C. and Persoone, G., *Cyst-based toxicity tests. IV. The Potential of ecotoxicological tests for the prediction of acute toxicity in man as evaluated on the first ten chemicals of the MEIC programme.* Altern. Lab. Anim., 1992, 20, 396-405.
 52. Calleja, M. C., Persoone, G., and Geladi, P., *Human acute toxicity prediction of the first 50 MEIC chemicals by a battery of ecotoxicological tests and physicochemical properties.* Food. Chem. Toxicol., 1994, 32, 173-187.

53. Tichy, M., Borek-Dohalsky, V., and Rucki, M., *Worms (Tubifex tubifex) and partition coefficients for estimation of acute toxicity*. In *Progress in the Reduction, Refinement and Replacement of Animal Experimentations*, Eds.: Balls, M., van Zeller, A. M., and Halder, M. E. Elsevier, Amsterdam, 2000. p. 513-520.
54. Khangarot, B. S. and Ray, P. K., *The confirmation of a mammalian poison classification using a water flea (Daphnia magna) screening method*. Arch. Hydrobiol., 1988, 113, 447-455.
55. Janardan, S. K., Olson, C. S., and Schaeffer, D. J., *Quantitative comparisons of acute toxicity of organic chemicals to rat and fish*. Ecotoxicol. Environ. Safety, 1984, 8, 531-539.
56. Kaiser, K. L. E. and Esterby, S. R., *Regression and cluster analysis of the acute toxicity of 267 chemicals to six species of biota and the octanol/water partition coefficient*. In *QSAR in Environmental Toxicology — IV*, Eds.: Hermens, J. L. P. and Opperhuizen, A. Elsevier, Amsterdam, 1991. p. 499-514.
57. Hodson, P. V., *A comparison of the acute toxicity of chemicals to fish, rats and mice*. J. Appl. Toxicol., 1985, 5, 220-226.
58. Delistraty, D., Taylor, B., and Anderson, R., *Comparisons of acute toxicity of selected chemicals to rainbow trout and rats*. Ecotoxicol. Environ. Safety, 1998, 39, 195-200.
59. Delistraty, D., *Acute toxicity to rats and trout with a focus on inhalation and aquatic exposures*. Ecotoxicol. Environ. Safety, 2000, 46, 225-233.
60. Gaines, T. B., *The acute toxicity of pesticides to rats*. Toxicol. Appl. Pharmacol., 1960, 2, 88-99.
61. Kishi, H. and Hashimoto, Y., *Evaluation of the procedures for the measurement of water solubility and n-octanol/water partition coefficient of chemicals results of a ring test in Japan*. Chemosphere, 1989, 18, 1749-1759.

62. Liu, R. and So, S. S., *Development of quantitative structure-property relationship models for early ADME evaluation in drug discovery. 1. Aqueous solubility*. J. Chem. Inf. Comput. Sci., 2001, 41, 1633-1639.
63. Kacew, S., *Role of rat strain in the differential sensitivity to pharmaceutical agents and naturally occurring substances*. J. Toxicol. Environ. Health, Part A, 1996, 47, 1-30.
64. Cronin, M. T. D. and Dearden, J. C., *QSAR in toxicology. 2. Prediction of acute mammalian toxicity and interspecies correlations*. Quant. Struct.-Act. Relat., 1995, 14, 117-120.
65. Lewis, D. F. V., *Cytochromes P450: Structure, Function and Mechanism*. Taylor & Francis, London, 1996.
66. Omura, T. and Sato, R., *The carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature*. J. Biol. Chem., 1964, 239, 2370-2378.
67. Meunier, B., de Visser, S. P., and Shaik, S., *Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes*. Chem. Rev., 2004, 104, 3947-3980.
68. Guengerich, F. P., *Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity*. Chem. Res. Toxicol., 2001, 14, 611-650.
69. Hildebrandt, A. and Estabrook, R. W., *Evidence for the participation of cytochrome b₅ in hepatic microsomal mixedfunction oxidation reactions*. Arch. Biochem. Biophys., 1971, 143, 66-79.
70. Noshiro, M., Ullrich, V., and Omura, T., *Cytochrome b₅ as electron donor for oxy-cytochrome P-450*. Eur. J. Biochem., 1981, 116, 521-526.
71. de Montellano, P. R. O., Ed. *Cytochrome P450: structure, mechanism, and biochemistry*. 2-nd ed., Plenum, New York, 1995.
72. Williams, P. A., Cosme, J., Vinkovic, D. M., Ward, A., Angove, H. C., Day, P. J., Vonnrhein, C., Tickle, I. J., and Jhoti, H., *Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone*. Science, 2004, 305, 683-686.

73. Yano, J. K., Wester, M. R., Schoch, G. A., Griffin, K. J., Stout, C. D., and Johnson, E. F., *The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-Å resolution*. J. Biol. Chem., 2004, 279, 38091-38094.
74. Rowland, P., Blaney, F. E., Smyth, M. G., Jones, J. J., Leydon, V. R., Oxbrow, A. K., Lewis, C. J., Tennant, M. G., Modi, S., Eggleston, D. S., Chenery, R. J., and Bridges, A. M., *Crystal structure of human cytochrome P450 2D6*. J. Biol. Chem., 2006, 281, 7614-7622.
75. Ekroos, M. and Sjogren, T., *Structural basis for ligand promiscuity in cytochrome P450 3A4*. Proc. Natl. Acad. Sci. USA., 2006, 103, 13682-13687.
76. Ekins, S., Bravi, G., Binkley, S., Gillespie, J. S., Ring, B. J., Wikel, J. H., and Wrighton, S. A., *Three and four dimensional-quantitative structure activity relationship (3D/4D-QSAR) analyses of CYP2D6 inhibitors*. Pharmacogenetics, 1999, 9, 477-489.
77. Venhorst, J., ter Laak, A. M., Commandeur, J. N., Funae, Y., Hiroi, T., and Vermeulen, N. P., *Homology modeling of rat and human cytochrome P450 2D (CYP2D) isoforms and computational rationalization of experimental ligand-binding specificities*. J. Med. Chem., 2003, 46, 74-86.
78. Modi, S., Paine, M. J., Sutcliffe, M. J., Lian, L. Y., Primrose, W. U., Wolf, C. R., and Roberts, G. C., *A model for human cytochrome P450 2D6 based on homology modeling and NMR studies of substrate binding*. Biochemistry, 1996, 35, 4540-4550.
79. Ellis, S. W., Hayhurst, G. P., Smith, G., Lightfoot, T., Wong, M. M. S., Simula, A. P., Ackland, M. J., Sternberg, M. J. E., Lennard, M. S., Tucker, G. T., and Wolf, C. R., *Evidence that aspartic acid 301 is a critical substrate-contact residue in the active site of cytochrome P450 2D6*. J. Biol. Chem., 1995, 270, 29055-29058.
80. Modi, S., Gilham, D. E., Sutcliffe, M. J., Lian, L. Y., Primrose, W. U., Wolf, C. R., and Roberts, G. C. K., *1-methyl-4-phenyl-1,2,3,6-*

- tetrahydropyridine as a substrate of cytochrome P450 2D6: allosteric effects of NADPH-cytochrome P450 reductase. Biochemistry, 1997, 36, 4461-4470.*
81. Gilham, D. E., Cairns, W., Paine, M. J. I., Modi, S., Poulson, R., Roberts, G. C. K., and Wolf, C. R., *Metabolism of MPTP by cytochrome P4502D6 and the demonstration of 2D6 mRNA in human foetal and adult brain by in situ hybridization. Xenobiotica, 1997, 27, 111-125.*
 82. Coleman, T., Ellis, S. W., Martin, I. J., Lennard, M. S., and Tucker, G. T., *1-Methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) is N-demethylated by cytochromes P450 2D6, 1A2 and 3A4 - implications for susceptibility to Parkinson's disease. J. Pharmacol. Exp. Ther., 1996, 277, 685-690.*
 83. Smith, G., Modi, S., Pillai, I., Lian, L. Y., Sutcliffe, M. J., Pritchard, M. P., Friedberg, T., Roberts, G. C. K., and Wolf, C. R., *Determinants of the substrate specificity of human cytochrome P-450 CYP2D6: design and construction of a mutant with testosterone hydroxylase activity. Biochem. J., 1998, 331, 783-792.*
 84. Testa, B., *The Metabolism of Drugs and Other Xenobiotics: Biochemistry of Redox Reactions.* Academic Press, San Diego, 1995.
 85. Oh, W. S., Kim, D. N., Jung, J., Cho, K.-H., and No, K. T., *New combined model for the prediction of regioselectivity in cytochrome P450/3A4 mediated metabolism. J. Chem. Inf. Model., 2008, 48, 591-601.*
 86. Zhou, D., Afzelius, L., Grimm, S. W., Andersson, T. B., Zauhar, R. J., and Zamora, I., *Comparison of methods for the prediction of the metabolic sites for CYP3A4-mediated metabolic reactions. Drug. Metab. Dispos., 2006, 34, 976-983.*
 87. 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., Wall, I. D., Woolven, J. M., Peishoff, C. E., and Head, M.

- S., *A critical assessment of docking programs and scoring functions*. J. Med. Chem., 2006, 49, 5912-5931.
88. Moitessier, N., Englebienne, P., Lee, D., Lawandi, J., and Corbeil, C. R., *Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go*. Br. J. Pharmacol., 2008, 153 Suppl. 1, S7-S26.
89. Jones, J. P., Mysinger, M., and Korzekwa, K. R., *Computational models for cytochrome P450: a predictive electronic model for aromatic oxidation and hydrogen atom abstraction*. Drug. Metab. Dispos., 2002, 30, 7-12.
90. Singh, S. B., Shen, L. Q., Walker, M. J., and Sheridan, R. P., *A model for predicting likely sites of CYP3A4-mediated metabolism on drug-like molecules*. J. Med. Chem., 2003, 46, 1330-1336.
91. Boyer, S., Arnby, C. H., Carlson, L., Smith, J., Stein, V., and Glen, R. C., *Reaction site mapping of xenobiotic biotransformations*. J. Chem. Inf. Model., 2007, 47, 583-590.
92. Sheridan, R. P., Korzekwa, K. R., Torres, R. A., and Walker, M. J., *Empirical regioselectivity models for human cytochromes P450 3A4, 2D6, and 2C9*. J. Med. Chem., 2007, 50, 3173-3184.
93. Button, W. G., Judson, P. N., Long, A., and Vessey, J. D., *Using absolute and relative reasoning in the prediction of the potential metabolism of xenobiotics*. J. Chem. Inf. Comput. Sci., 2003, 43, 1371-1377.
94. METEOR, Lhasa Limited, Department of Chemistry, University of Leeds, Leeds, UK, www.lhasalimited.org/meteor/, 2009.
95. Cruciani, G., Carosati, E., De Boeck, B., Ethirajulu, K., Mackie, C., Howe, T., and Vianello, R., *MetaSite: understanding metabolism in human cytochromes from the perspective of the chemist*. J. Med. Chem., 2005, 48, 6970-6979.
96. MetaSite 3.0.4, Molecular Discovery, Ltd., Pinner, Middlesex, UK, www.moldiscovery.com/soft_metasite.php, 2009.

97. Worth, A. P., Bassan, A., Gallegos, A., Netzeva, T. I., Patlewicz, G., Pavan, M., Tsakovska, I., and Vracko, M., *The characterisation of (quantitative) structure-activity relationships: preliminary guidance*, EUR 21866 EN, Institute for Health and Consumer Protection, Joint Research Centre, European Commission, Ispra (VA), Italy, 2005.
http://ecb.jrc.it/documents/QSAR/QSAR_characterisation_EUR_21866_EN.pdf
98. 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-Jeliazkova, N., Patlewicz, G. Y., Perkins, R., Roberts, D., Schultz, T., Stanton, D. W., van de Sandt, J. J., Tong, W., Veith, G., and Yang, C., *Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. The report and recommendations of ECVAM Workshop 52*. *Altern. Lab. Anim.*, 2005, 33, 155-173.
99. *Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC*, The European Parliament and the Council of the European Union, 2006.
http://eur-lex.europa.eu/LexUriServ/site/en/oj/2006/l_396/l_39620061230en00010849.pdf
100. Jaworska, J., Nikolova-Jeliazkova, N., and Aldenberg, T., *QSAR applicability domain estimation by projection of the training set descriptor space: a review*. *Altern. Lab. Anim.*, 2005, 33, 445-459.

101. Dimitrov, S., Dimitrova, G., Pavlov, T., Dimitrova, N., Patlewicz, G., Niemela, J., and Mekenyan, O., *A stepwise approach for defining the applicability domain of SAR and QSAR models*. J. Chem. Inf. Model., 2005, 45, 839-849.
102. Gombar, V. K., Enslein, K., and Blake, B. W., *Assessment of developmental toxicity potential of chemicals by quantitative structure-toxicity relationship models*. Chemosphere, 1995, 31, 2499-2510.
103. Lu, A. Y. H., Wang, R. W., and Lin, J. H., *Cytochrome P450 in vitro reaction phenotyping: a re-evaluation of approaches used for P450 isoform identification*. Drug. Metab. Dispos., 2003, 31, 345-350.
104. von Moltke, L. L., Greenblatt, D. J., Duan, S. X., Harmatz, J. S., Wright, C. E., and Shader, R. I., *Inhibition of terfenadine metabolism in vitro byazole antifungal agents and by selective serotonin reuptake inhibitor antidepressants: relation to pharmacokinetic interactions in vivo*. J. Clin. Psychopharmacol., 1996, 16, 104-112.
105. Moody, D. E., Walsh, S. L., Rollins, D. E., Neff, J. A., and Huang, W., *Ketoconazole, a cytochrome P450 3A4 inhibitor, markedly increases concentrations of levo-acetyl-alpha-methadol in opioid-naive individuals*. Clin. Pharmacol. Ther., 2004, 76, 154-166.
106. Dresser, G. K., Spence, J. D., and Bailey, D. G., *Pharmacokinetic-pharmacodynamic consequences and clinical relevance of cytochrome P450 3A4 inhibition*. Clin. Pharmacokinet., 2000, 38, 41-57.
107. von Moltke, L. L., Greenblatt, D. J., Grassi, J. M., Granda, B. W., Duan, S. X., Fogelman, S. M., Daily, J. P., Harmatz, J. S., and Shader, R. I., *Protease inhibitors as inhibitors of human cytochromes P450: high risk associated with ritonavir*. J. Clin. Pharmacol., 1998, 38, 106-111.
108. Branch, R. A., Adedoyin, A., Frye, R. F., Wilson, J. W., and Romkes, M., *In vivo modulation of CYP enzymes by quinidine and rifampin*. Clin. Pharmacol. Ther., 2000, 68, 401-411.
109. McLaughlin, L. A., Paine, M. J., Kemp, C. A., Marechal, J. D., Flanagan, J. U., Ward, C. J., Sutcliffe, M. J., Roberts, G. C., and Wolf, C. R., *Why*

- is quinidine an inhibitor of cytochrome P450 2D6? The role of key active-site residues in quinidine binding.* J. Biol. Chem., 2005, 280, 38617-38624.
110. Brynne, N., Svanstrom, C., Aberg-Wistedt, A., Hallen, B., and Bertilsson, L., *Fluoxetine inhibits the metabolism of tolterodine - pharmacokinetic implications and proposed clinical relevance.* Br. J. Clin. Pharmacol., 1999, 48, 553-563.
111. Hara, Y., Nakajima, M., Miyamoto, K. I., and Yokoi, T., *Inhibitory effects of psychotropic drugs on mexiletine metabolism in human liver microsomes: prediction of in vivo drug interactions.* Xenobiotica, 2005, 35, 549-560.
112. Keizers, P. H., de Graaf, C., de Kanter, F. J., Oostenbrink, C., Feenstra, K. A., Commandeur, J. N., and Vermeulen, N. P., *Metabolic regio- and stereoselectivity of cytochrome P450 2D6 towards 3,4-methylenedioxy-N-alkylamphetamines: in silico predictions and experimental validation.* J. Med. Chem., 2005, 48, 6117-6127.
113. Kirkwood, L. C., Nation, R. L., and Somogyi, A. A., *Characterization of the human cytochrome P450 enzymes involved in the metabolism of dihydrocodeine.* Br. J. Clin. Pharmacol., 1997, 44, 549-555.
114. Abel, S. M., Maggs, J. L., Back, D. J., and Park, B. K., *Cortisol metabolism by human liver in vitro - I. Metabolite identification and inter-individual variability.* J. Steroid. Biochem. Mol. Biol., 1992, 43, 713-719.
115. Abel, S. M. and Back, D. J., *Cortisol metabolism in vitro - III. Inhibition of microsomal 6 beta-hydroxylase and cytosolic 4-ene-reductase.* J. Steroid. Biochem. Mol. Biol., 1993, 46, 827-832.
116. *Algorithm Builder 1.8*, Pharma Algorithms, Inc., Toronto ON, Canada, <http://www.pharma-algorithms.com>, 2006.
117. Japertas, P., Didziapetris, R., and Petrauskas, A., *Fragmental methods in the design of new compounds. Applications of Advanced Algorithm Builder.* Quant. Struct.-Act. Relat., 2002, 21, 23-37.

118. Wold, S., Sjostrom, M., and Eriksson, L., *PLS in chemistry*. In *The Encyclopedia of Computational Chemistry*, Eds.: Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer III, H. F., and Schreiner, P. R. John Wiley & Sons, Chichester, 1998. p. 2006-2021.
119. *ACD/ADME Suite 4.95*, Advanced Chemistry Development, Inc., Toronto ON, Canada, www.acdlabs.com, 2009.
120. Platts, J. A., Butina, D., Abraham, M. H., and Hersey, A., *Estimation of molecular linear free energy relation descriptors using a group contribution approach*. *J. Chem. Inf. Comput. Sci.*, 1999, 39, 835 - 845.
121. Fukuto, T. R., *Mechanism of action of organophosphorus and carbamate insecticides*. *Environ. Health. Perspect.*, 1990, 87, 245-254.
122. Clarke, D. D., *Fluoroacetate and fluorocitrate: mechanism of action*. *Neurochem. Res.*, 1991, 16, 1055-1058.
123. Povirk, L. F. and Shuker, D. E., *DNA damage and mutagenesis induced by nitrogen mustards*. *Mutat. Res.*, 1994, 318, 205-226.
124. Van Duuren, B. L., *Direct-acting alkylating and acylating agents. DNA adduct formation, structure-activity, and carcinogenesis*. *Ann. N. Y. Acad. Sci.*, 1988, 534, 620-634.
125. Kaplita, P. V. and Smith, R. P., *Pathways for the bioactivation of aliphatic nitriles to free cyanide in mice*. *Toxicol. Appl. Pharmacol.*, 1986, 84, 533-540.
126. Casida, J. E. and Lawrence, L. J., *Structure-activity correlations for interactions of bicyclophosphorus esters and some polychlorocycloalkane and pyrethroid insecticides with the brain-specific t-butylbicyclophosphorothionate receptor*. *Environ. Health. Perspect.*, 1985, 61, 123-132.
127. Bremser, W., *Hose — a novel substructure code*. *Anal. Chim. Acta*, 1978, 103, 355-365.
128. Groves, J. T., McClusky, G. A., White, R. E., and Coon, M. J., *Aliphatic hydroxylation by a highly purified liver microsomal cytochrome P-450*.

- Evidence for a carbon radical intermediate.* Biochem. Biophys. Res. Commun., 1978, 81, 154-160.
129. White, R. E., Miller, J. P., Favreau, L. V., and Bhattacharyya, A., *Stereochemical dynamics of aliphatic hydroxylation by cytochrome P-450.* J. Am. Chem. Soc., 1986, 108, 6024-6031.
130. Lindsay Smith, J. R. and Sleath, P. R., *Model systems for cytochrome P450-dependent monooxygenases. Part 2. Kinetic isotope effects for the oxidative demethylation of anisole and [Me-²H₃]anisole by cytochrome P450-dependent monooxygenases and model systems.* J. Chem. Soc. Perkin Trans. II, 1983, 621-628.
131. Lu, A. Y. H., Harada, N., and Miwa, G. T., *Rate-limiting steps in cytochrome P-450-catalysed reactions: studies on isotope effects in the O-de-ethylation of 7-ethoxycoumarin.* Xenobiotica, 1984, 14, 19-26.
132. Burka, L. T., Guengerich, F. P., Willard, R. J., and Macdonald, T. L., *Mechanism of cytochrome P-450 catalysis. Mechanisms of N-dealkylation and amine oxide deoxygenation.* J. Am. Chem. Soc., 1985, 107, 2549-2551.
133. Heimbrook, D. C., Murray, R. I., Egeberg, K. D., Sligar, S. G., Nee, M. W., and Bruice, T. C., *Demethylation of N,N-dimethylaniline and p-cyano-N,N-dimethylaniline and their N-oxides by cytochrome P450_{LM2} and P450_{CAM}.* J. Am. Chem. Soc., 1984, 106, 1514-1515.
134. Guengerich, F. P., Willard, R. J., Shea, J. P., Richards, L. E., and Macdonald, T. L., *Mechanism-based inactivation of cytochrome P-450 by heteroatom-substituted cyclopropanes and formation of ring-opened products.* J. Am. Chem. Soc., 1984, 106, 6446-6447.
135. Shannon, P. and Bruice, T. C., *A novel P-450 model system for the N-dealkylation reaction.* J. Am. Chem. Soc., 1981, 103, 4580-4582.
136. Efron, B., *Bootstrap methods: another look at the jackknife.* Ann. Statist., 1979, 7, 1-26.
137. Maggiora, G. M., *On outliers and activity cliffs - why QSAR often disappoints.* J. Chem. Inf. Model., 2006, 46, 1535-1535.

138. *NCBI PubChem database*, National Center for Biotechnology Information of the National Library of Medicine, Bethesda (MD), United States. <http://pubchem.ncbi.nlm.nih.gov/>
139. *Derwent World Drug Index sample CD*, Derwent Information, London, United Kingdom, www.derwent.com, 2001.
140. Senda, C., Kishimoto, W., Sakai, K., Nagakura, A., and Igarashi, T., *Identification of human cytochrome P450 isoforms involved in the metabolism of brotizolam*. *Xenobiotica.*, 1997, 27, 913-922.
141. Perloff, M. D., von Moltke, L. L., Court, M. H., Kotegawa, T., Shader, R. I., and Greenblatt, D. J., *Midazolam and triazolam biotransformation in mouse and human liver microsomes: relative contribution of CYP3A4 and CYP2D6 isoforms*. *J. Pharmacol. Exp. Ther.*, 2000, 292, 618-628.
142. Kronbach, T., Mathys, D., Umeno, M., Gonzalez, F. J., and Meyer, U. A., *Oxidation of midazolam and triazolam by human liver cytochrome P450III A4*. *Mol. Pharmacol.*, 1989, 36, 89-96.
143. Gorski, J. C., Jones, D. R., Hamman, M. A., Wrighton, S. A., and Hall, S. D., *Biotransformation of alprazolam by members of the human cytochrome P4503A subfamily*. *Xenobiotica.*, 1999, 29, 931-944.
144. Uptagrove, A. L. and Nelson, W. L., *Importance of amine pKa and distribution coefficient in the metabolism of fluorinated propranolol derivatives. Preparation, identification of metabolite regioisomers, and metabolism by CYP2D6*. *Drug Metab Dispos*, 2001, 29, 1377-1388.