

VILNIAUS UNIVERSITETAS

Justas Dapkūnas

CITOCHROMŲ P450 KATALIZUOJAMO VAISTŲ METABOLIZMO  
KOMPIUTERINIS MODELIAVIMAS

Daktaro disertacijos santrauka  
Fiziniai mokslai, biochemija (04P)

Vilnius, 2011

Disertacija rengta 2007 - 2011 metais Vilniaus universitete bendradarbiaujant su VŠĮ „Aukštieji algoritmai“.

Mokslinis vadovas:

dr. Remigijus Didžiapetris (VŠĮ „Aukštieji algoritmai“, fiziniai mokslai, biochemija – 04P).

**Disertacija ginama Vilniaus universiteto Biochemijos mokslo krypties taryboje:**

Pirmininkė

prof. dr. Vida Kirvelienė (Vilniaus universitetas, fiziniai mokslai, biochemija – 04P).

Nariai:

prof. dr. Rimantas Daugelavičius (Vytauto Didžiojo universitetas, fiziniai mokslai, biochemija – 04P);

dr. Česlovas Venclovas (Vilniaus universitetas, fiziniai mokslai, biochemija – 04P);

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

doc. dr. Saulius Šatkauskas (Vytauto Didžiojo universitetas, biomedicinos mokslai, biofizika – 02B).

Oponentai:

dr. Daumantas Matulis (Vilniaus universitetas, fiziniai mokslai, biochemija – 04P);

prof. habil. dr. Rimvydas Simutis (Kauno technologijos universitetas, technologijos mokslai, informatikos inžinerija - 07T).

Disertacija bus ginama viešame Biochemijos mokslo krypties tarybos posėdyje 2011 m. rugsėjo mėn. 30 d. 14 val. Vilniaus universiteto Gamtos mokslų fakulteto Didžiojoje auditorijoje. Adresas: M.K. Čiurlionio 21/27 – 214, LT-03101, Vilnius, Lietuva.

Disertacijos santrauka išsiuntinėta 2011 m. rugpjūčio \_\_\_ d.

Disertaciją galima peržiūrėti Vilniaus universiteto ir Biochemijos instituto bibliotekose.

VILNIUS UNIVERSITY

Justas Dapkūnas

COMPUTATIONAL MODELING OF CYTOCHROME P450-MEDIATED DRUG  
METABOLISM

Summary of doctoral dissertation  
Physical sciences, biochemistry (04P)

Vilnius, 2011

The dissertation work was carried out at Vilnius University from 2007 to 2011 in cooperation with researchers at VŠĮ “Aukštieji algoritmai”.

Scientific supervisor:

dr. Remigijus Didžiapetris (VŠĮ “Aukštieji algoritmai”, physical sciences, biochemistry – 04P).

### **The defence council:**

Chair

prof. dr. Vida Kirvelienė (Vilnius University, physical sciences, biochemistry – 04P).

Members:

prof. dr. Rimantas Daugelavičius (Vytautas Magnus University, physical sciences, biochemistry – 04P);

dr. Česlovas Venclovas (Vilnius University, physical sciences, biochemistry – 04P);

prof. habil. dr. Eugenijus Butkus (Vilnius University, physical sciences, chemistry – 03P);

doc. dr. Saulius Šatkauskas (Vytautas Magnus University, biomedical sciences, biophysics – 02B).

Opponents:

dr. Daumantas Matulis (Vilnius University, physical sciences, biochemistry – 04P);

prof. habil. dr. Rimvydas Simutis (Kaunas University of Technology, technological sciences, informatics engineering – 07T).

The dissertation will be defended at the public meeting of the council on the September 30th, 2011 at 14:00 at the Auditorium No 214 at the Faculty of Natural Sciences of Vilnius University. Address: M.K. Čiurlionio 21/27 – 214, LT-03101, Vilnius, Lithuania.

The summary of the dissertation was distributed on August \_\_, 2011.

The dissertation is available at the libraries of Vilnius University and Institute of Biochemistry.

# Turinys

<b>Turinys</b>	<b>1</b>
<b>1 Įvadas</b>	<b>2</b>
<b>2 Duomenys ir metodai</b>	<b>6</b>
2.1 CYP3A4 slopinimo duomenys . . . . .	6
2.2 Regioselektyvumo duomenys . . . . .	6
2.3 Modeliavimo metodai . . . . .	8
<b>3 Rezultatai ir jų aptarimas</b>	<b>9</b>
3.1 CYP3A4 slopinimo modelis . . . . .	9
3.2 Metabolizmo regioselektyvumo modelis . . . . .	14
<b>4 Išvados</b>	<b>21</b>
<b>5 Publikacijos disertacijos tema</b>	<b>22</b>
<b>6 Curriculum Vitae</b>	<b>24</b>
<b>7 Summary</b>	<b>25</b>
<b>Literatūros šaltiniai</b>	<b>27</b>

# 1 Įvadas

Kepenyse vykstančios citochromų P450 katalizuojamos vaistų metabolizmo reakcijos yra pagrindinis daugelio vaistų ir kitų ksenobiotikų pašalinimo iš organizmo būdas. Susidare metabolitai paprastai yra mažiau nuodingi ir gali būti greičiau pašalinami, nes po biotransformacijos lipofiliškos molekulės tampa labiau polinės ir geriau tirpsta vandenyje. Visgi per greitas vaisto pašalinimas iš organizmo yra nepageidaujamas, todėl farmacinės kompanijos stengiasi padidinti vaistų kandidatų metabolinį stabilumą. Tai dažniausiai daroma blokuojant ar pakeičiant žinomas molekulės vietas, kuriose vyksta oksidacija [1]. Be to, tam tikrais atvejais gali susidaryti toksiški vaistų metabolitai, kurie sukelia pašalinį poveikį. Šiuo metu vaistų pramonę reguliuojančios institucijos reikalauja patvirtinti ne tik paties vaisto, bet ir jo biotransformacijos produktų saugumą [2]. Dėl šių priežasčių vėlesnėse vaisto kūrimo stadijose metabolitai visada yra nustatinėjami eksperimentiškai, norint sužinoti galimą jų farmakologinį ar toksinį poveikį.

Ksenobiotikų biotransformacijos fermentų slopinimas yra pagrindinė vaistų tarpusavio sąveikų priežastis. Nuslopinus vaisto metabolizmą, sulėtėja jo pašalinimas ir padidėja koncentracija kraujyje. Dėl to taip pat gali pasireikšti nepageidaujamas pašalinis poveikis. Keletas vaistų netgi buvo pašalinti iš rinkos dėl CYP3A4 slopinimo sukeltų sąveikų su kitais vaistais [3]. Gebėjimas slopinti CYP3A4 turi būti nustatomas visiems vaistų kandidatams [1], nes CYP3A4 yra pagrindinis vaistų metabolizmo fermentas, lemiantis daugiau nei 50% vaistų metabolizmą. Šis baltymas yra plataus specifiskumo monooksigenazė, oksiduojanti labai įvairius substratus [4].

Norint kuo anksčiau išsiaiškinti riziką, susijusią su vaistų metabolizmu ar jo slopinimu, dažniausiai naudojami įvairūs *in vitro* metodai. Juos paprastai papildo ir kompiuterinis modeliavimas [1]. Nepaisant mažesnio tikslumo, *in silico* metodai turi daug privalumų: jie yra greiti, pigūs ir gali būti taikomi ankstyviausiose vaistų paieškos stadijose, netgi prieš naujų cheminių junginių sintezę. Nors vaistų metabolizmas yra sudėtingas procesas, jo prognozavimas turint tik medžiagos struktūrinę formulę yra įmanomas [5], o sukurti modeliai yra plačiai taikomi pramonėje. Metabolizmo vietų prognozės buvo naudojamos ciklooksigenazės-2 slopiklių metabolinio stabilumo optimizavimui [6], o metabolitų identifikavimas masių spektrometrijos metodais yra greitesnis, naudojant kompiuteriu prognozuotas jų struktūras [2, 7, 8]. CYP3A4 slopinimo prognozavimas leidžia iš anksto sužinoti galimas vaistų sąveikas ir palengvina eksperimentinių tyrimų planavimą.

ADME savybių, tarp jų ir metabolizmo bei citochromų P450 slopinimo, prognozavimui dažnai naudojami kiekybinio struktūros ir aktyvumo ryšio (angl. quantitative structure-activity relationship, QSAR) modeliai. Cheminių junginių struktūros aprašomos fizikocheminiais, kvantinės chemijos, topologiniais bei dvimatės ar trimatės struktūros deskriptoriais. Modelių apmokymui naudojami *in vitro* ar *in vivo* eksperimentų duomenys ir statistiniai regresijos ir klasifikacijos metodai, pvz., dalinis mažiausių kvadratų (angl. Partial Least Squares, PLS) ar atraminių vektorių (angl. Support Vector Machines, SVM) metodai.

Pagrindinis struktūros ir aktyvumo ryšio modelių trūkumas yra tai, kad juos galima taikyti tik junginiams, struktūriškai panašiems į naudotus modelio apmokymui. Jų apibrėžiama cheminės erdvės dalis yra vadinama modelio pritaikymo sritimi. Pagal OECD (Ekonominio bendradarbiavimo ir plėtros organizacijos) reikalavimus QSAR modeliams, kiekvienas iš jų turi prognozuoti konkretų eksperimentiškai matuojamą dydį, būti patikrintas, naudojant duomenis, kurie nenaudoti modeliavimui, ir, svarbiausia, susietas su tam tikra pritaikymo sritimi [9]. Deja, dauguma anksčiau publikuotų vaistų metabolizmą ar jo slopinimą prognozuojančių modelių yra sukurti naudojant metodus, kurie neleidžia įvertinti prognozių patikimumo, dėl to kyla klausimų dėl jų praktinio panaudojimo galimybių.

Norint išspręsti aukščiau paminėtas problemas, neseniai buvo sukurtas GALAS (angl. Global, Adjusted Locally According to Similarity – Globalus, lokaliai pakoreguotas pagal panašumą) modeliavimo metodas. Šis metodas visų pirma buvo sėkmingai panaudotas prognozuojant tolydžias savybes ( $\text{Log}P$  ir ūmų toksiškumą) [10, 11]. Jo esmė yra dviejų metodų kombinavimas: globalaus modelio, prognozuojančio tiriamą savybę, ir lokalių korekcijų pagal panašių cheminių junginių eksperimentinius duomenis. Pastarosios taip pat leidžia įvertinti prognozės patikimumą.

Be modelio pritaikymo srities įvertinimo, kita svarbi GALAS modeliavimo metodikos savybė yra galimybė išplėsti šią sritį ir pritaikyti modelį naujų junginių savybių prognozavimui. Ši adaptacija yra labai paprasta: naujų cheminių junginių struktūros ir eksperimentiniai duomenys tiesiog pridedami į lokalią modelio dalį, ir tuomet jiems suskaičiuojama reikiama korekcija. Ji yra nesvarbi, jei globalaus modelio prognozės teisingos, tačiau neteisingos prognozės paprastai ženkliai pakeičiamos, remiantis naujais eksperimentiniais duomenimis.

## Darbo tikslai

Šio tyrimo pagrindiniai tikslai buvo:

- pritaikyti GALAS modeliavimo metodą su vaistų metabolizmu susijusių savybių prognozavimui;
- sukurti bei patikrinti modelį, prognozuojantį metabolizmo regioselektyvumą.

Norint pasiekti šiuos tikslus, buvo iškelti konkretūs uždaviniai:

- sukurti modelį, prognozuojantį tikimybę, kad organinis junginys slopins CYP3A4 (ši savybė buvo pasirinkta kaip su vaistų metabolizmu susijusios savybės pavyzdys);
- sukurti metabolizmo regioselektyvumo modelį, prognozuojantį oksidacijos tikimybę kiekvienam organinio junginio atomui;
- parodyti, kaip šiuose modeliuose veikia pagrindinės GALAS modeliavimo metodo savybės (korekcija pagal panašumą, prognozės patikimumo indeksas, galimybė greitai apmokyti modelius);
- patikrinti metabolizmo regioselektyvumo prognozes naudojant naujų junginių duomenis;
- palyginti metabolizmo regioselektyvumo prognozes su anksčiau publikuoto modelio *SMARTCyp* prognozėmis.

## Disertacijos turinys

Disertacija parašyta anglų kalba. Ją sudaro šios dalys: Įvadas, Literatūros apžvalga, Duomenys ir metodai, CYP3A4 slopinimo modeliavimo rezultatai ir jų aptarimas, Metabolizmo regioselektyvumo modeliavimo rezultatai ir jų aptarimas, Išvados. Disertacijoje yra 134 puslapiai, 26 paveikslai ir 17 lentelių. Pateikiamas 1 priedas, cituojami 182 literatūros šaltiniai.



## Rezultatų mokslinis naujumas

- GALAS modeliavimo metodas buvo sėkmingai pritaikytas su vaistų metabolizmu susijusių savybių prognozavimui.
- Sukurtas žmogaus kepenų mikrosomų katalizuojamo metabolizmo vietas prognozuojantis modelis yra pirmasis regioselektyvumo QSAR modelis, kuris apibrėžia savo pritaikymo sritį.
- Sukurti CYP3A4 slopinimą ir metabolizmo regioselektyvumą prognozuojantys modeliai gali būti apmokomi naudojant eksperimentinius naujų junginių duomenis tiesiog pridėdant juos į lokalią modelio dalį, o pilnas permodeliavimas yra nereikalingas.

## Praktinis taikymas

Šiame darbe sukurti GALAS modeliai gali būti naudojami farmacijos pramonėje. *In silico* metodais galima tirti naujų vaistų kandidatų savybes dar prieš jų sintezę. Prognozavus CYP3A4 slopinimą galima nuspėti tikėtinas vaistų sąveikas ankstyviausiose jų paieškos stadijose. Regioselektyvumo modeliavimo rezultatai leidžia tikėtis praktinio šio modelio pritaikymo metabolinio stabilumo optimizavimui ir metabolitų struktūros nustatymui. Pateikiamas prognozės patikimumo įvertinimas palengvina junginių prioritetizavimą prieš eksperimentinius tyrimus, o naudojantis paprastu GALAS modelio apmokymu galima jį pritaikyti junginiams, tiriamiems bet kokiame vaistų kūrimo projekte.

## Ginami teiginiai

- Struktūros ir aktyvumo ryšio modeliai, prognozuojantys CYP3A4 slopinimą ir žmogaus kepenų mikrosomų katalizuojamo metabolizmo regioselektyvumą buvo sukurti naudojant GALAS modeliavimo metodą.
- Sukurti modeliai geba apibrėžti savo pritaikymo sritį, suskaičiuodami prognozės patikimumo indeksą, kurio didesnės reikšmės atitinka didesnę prognozių tikslumą.

- Sukurtų GALAS modelių pritaikymo sritis yra lengvai išplečiama, pridėdant naujų junginių į lokalią modelio dalį.
- Regioselektyvumo prognozių, gautų naudojant sukurtą modelį, kokybė yra palyginama su programos *SMARTCyp* prognozėmis, kurios apskaičiuojamos kvantinės chemijos metodais.

## 2 Duomenys ir metodai

### 2.1 CYP3A4 slopinimo duomenys

Iš literatūros buvo surinkti duomenys apie CYP3A4 slopinimą 907 junginiams. Tie junginiai, kurie turėjo  $IC_{50} < 40 \mu M$ , buvo klasifikuoti kaip CYP3A4 slopikliai, turintys  $IC_{50} > 60 \mu M$  – kaip neslopikliai, o turintys tarpines  $IC_{50}$  reikšmes ( $40-60 \mu M$ ) ar prieštarīgus rezultatus skirtinguose eksperimentuose buvo pažymėti kaip neapibrėžti duomenys. Kitas CYP3A4 slopinimo duomenų rinkinys buvo parsijstas iš PubChem duomenų bazės (assay ID 884) 2008 m. birželio 10 d. [12]. Jis buvo suklasifikuotas dvejopai. Pirmoji klasifikacija atitinka literatūros duomenų klasifikaciją. Kitu atveju CYP3A4 slopikliai buvo tik efektyviai ši fermentą slopinantys junginiai. Junginių skaičius, klasifikavimo kriterijai ir pasiskirstymas pagal CYP3A4 slopinimą visuose duomenų rinkiniuose yra pateikti 1 lentelėje.

### 2.2 Regioselektyvumo duomenys

873 junginių metabolizmo duomenys buvo surinkti iš publikacijų, kuriose aprašyti metabolitai, susidarę inkubuojant organinius junginius su žmogaus kepenų mikrosomomis ar rekombinantiniais žmogaus citochromais P450. Anglies atomai, turintys bent vieną prijungtą vandenilio atomą, ir sieros atomai buvo pažymėti kaip metabolizmo vietos, jei buvo stebima jų oksidacija. Visas atomų rinkinys buvo sudalintas pagal atomams būdingus reakcijų tipus, jo santrauka yra pateikta 2 lentelėje. Taip pat, jeigu buvo įmanoma, buvo sužymėta, kurios metabolizmo vietos yra būdingos CYP2D6 katalizuojamoms biotransformacijoms.

1 lentelė: Junginių pasiskirstymas pagal aktyvumą CYP3A4 slopinimo duomenų rinkiniuose.

Duomenų rinkinys	Junginių skaičius	Slopikliai	Neslopikliai	Neapibrėžti
Literatūros duomenys <sup>a</sup>	907	335 (36,9%)	497 (54,8%)	75 (8,3%)
PubChem duomenys (bendras slopinimas) <sup>b</sup>	11.060	303 2 (27,4%)	5.496 (49,7%)	2.532 (22,8%)
PubChem duomenys (efektyvus slopinimas) <sup>c</sup>	11.060	1.238 (11,2%)	6.401 (57,9%)	3.421 (30,9%)

<sup>a</sup> Slopikliai:  $IC_{50} < 40 \mu M$ ; neslopikliai:  $IC_{50} > 60 \mu M$ ; neapibrėžti junginiai:  $IC_{50}$  40-60  $\mu M$  arba prieštaringi skirtingų eksperimentų rezultatai.

<sup>b</sup> Slopikliai: PubChem aktyvumo įvertis  $> 40$ ; neslopikliai: PubChem aktyvumo įvertis 0; neapibrėžti junginiai: kiti (taip pat ir prieštaringų rezultatų kartojant eksperimentus atvejai).

<sup>c</sup> Slopikliai:  $IC_{50} < 5 \mu M$  ir didžiausias efektyvumas  $> 70\%$ ; neslopikliai:  $IC_{50} > 30 \mu M$ ; neapibrėžti junginiai: kiti.

2 lentelė: Regioselektyvumo modeliavimui naudoti duomenų rinkiniai.

Duomenų rinkinio poaibis	Aprašymas	Junginių sk.	Metabolizmo vietų sk.	Žymėtų atomų sk.
N-dealkilinimas	Alifatinis CH prie N	511	333	1.173
O-dealkilinimas	Alifatinis CH prie O	488	260	1.033
Alifatinis	Alifatinis CH (likę)	723	318	2.904
hidroksilinimas				
Aromatinis	Aromatinis CH	739	358	3.341
hidroksilinimas				
S-oksidacija	S, turinti $<4$ kaimynus	135	57	157
Iš viso:		873	1.326	8.608

Vėliau, jau sukūrus modelį, iš naujausių publikacijų buvo surinkti 42 junginių metabolizmo duomenys, kurie buvo naudojami kaip išorinio patikrinimo rinkinys, ir taip pat prognozavimo palyginimui su kitomis programomis. Metabolizmo vietos juose buvo sužymėtos remiantis tais pačiais aukščiau paminėtais kriterijais.

## 2.3 Modeliavimo metodai

Junginių struktūros buvo aprašytos fragmentiniais deskriptoriais. CYP3A4 slopinimo atveju buvo naudojamas 379 fragmentų rinkinys, kurio didžioji dalis yra sudaryta iš funkcinų grupių, molekulės formą aprašančių fragmentų, ir atspindi bendrą junginio struktūrą. Šis standartinis rinkinys buvo papildytas CYP3A4 specifiškumą apibūdinančiais fragmentais (azotiniais heterociklais, metilenodioksibenzenu ir kt.). Kadangi metabolizmo regioselektyvumo modelis prognozuoja ne visos molekulės savybę, o oksidacijos tikimybę kiekvienam atomui, šiuo atveju buvo sukurta fragmentacija nuo centrinio atomo, kuri randa skirtingus fragmentų rinkinius skirtingiems molekulėje esantiems atomams.

Prognozuojantiems modeliams kurti buvo naudojamas GALAS modeliavimo metodas [10, 11]. Kiekvienas GALAS modelis susideda iš dviejų dalių: globalaus (bazinio) modelio, šiuo atveju kuriamo naudojant logistinę PLS regresiją, ir lokalių korekcijų, apskaičiuojamų pagal panašių apmokymo duomenų rinkinio junginių eksperimentinius duomenis:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \sum_i a_i f_i + c + \Delta, \quad (1)$$

čia  $p$  yra tikimybė, kad junginys yra aktyvus;  $\sum_i a_i f_i + c$  yra globalaus regresijos modelio prognozė ( $f_i$  – fragmento pasikartojimų molekulėje skaičius;  $a_i$  – fragmento statistinis koeficientas;  $c$  – laisvasis narys),  $\Delta$  – lokalaus modelio korekcija pagal panašių junginių duomenis.

Lokali korekcija pagal panašių junginių eksperimentinius duomenis ne tik pagerina modelio prognozes, bet ir leidžia įvertinti prognozės patikimumą. Skaičiuojant prognozės patikimumo indeksą (angl. reliability index,  $RI$ ), yra atsižvelgiama į du dalykus, lemiančius konkrečios prognozės tikslumą:

- Panašių junginių buvimą modelio apmokymo rinkinyje;
- Eksperimentinių duomenų neprieštarumą, lyginant juos su modelio prognozėmis.

Be to, pridant naujus junginius tiesiog į korekcijos pagal panašumą duomenų rinkinį (apmokymo biblioteką), galimas modelio pritaikymo srities išplėtimas ir adaptacija naujų cheminių molekulių klasių prognozavimui.

Norint patikrinti, kaip veikia šios GALAS modeliavimo metodo savybės, naudojant literatūros duomenis buvo sukurtas CYP3A4 slopinimą prognozuojantis modelis. Šis modelis prognozuoja tikimybę, kad junginys bus CYP3A4 slopiklis, turintis  $IC_{50} < 40 \mu M$ . Modelis buvo patikrintas naudojant vidinį duomenų rinkinį, sudarantį maždaug 20% literatūros duomenų (190 junginių), ir išorinį rinkinį – PubChem duomenis. Junginys buvo laikomas slopikliu, jei prognozuota tikimybė buvo  $> 0,5$ . Vėliau buvo išbandyta galimybė apmokyti sukurtą modelį, pridant naujus junginius į korekcijos pagal panašumą biblioteką.

Kuriant žmogaus kepenų mikrosomų katalizuojamo metabolizmo regioselektyvumo modelį, buvo sukurti atskiri prognozuojantys modeliai kiekvienam 2 lentelėje pavaizduotam reakcijos tipui. Kiekvienas duomenų rinkinio poaibis buvo sudalintas į apmokymo (70%) ir patikrinimo rinkinius (30%). Po vidinio patikrinimo naudojant šiuos duomenis, oksidacijos tikimybė buvo suskaičiuota kiekvienam 42 naujų junginių atomui, ir prognozės šiems junginiams buvo suskirstytos pagal jų kokybę į „puikias“, „geras“, „patenkinamas“ ir „nepatenkinamas“. Norint palyginti regioselektyvumo rezultatus su kitais modeliais, metabolizmo vietos šiuos junginiuose taip pat buvo prognozuotos, naudojant programą *SMARTCyp* [13, 14], ir šios prognozės taip pat suskirstytos pagal panašius kokybės kriterijus. Taip pat buvo bandoma galimybė apmokyti regioselektyvumo modelį.

## 3 Rezultatai ir jų aptarimas

### 3.1 CYP3A4 slopinimo modelis

Norint pritaikyti GALAS modeliavimo metodą savybių, susijusių su vaistų metabolizmu, prognozavimui, visų pirma buvo sukurtas modelis, prognozuojantis CYP3A4 slopinimą.

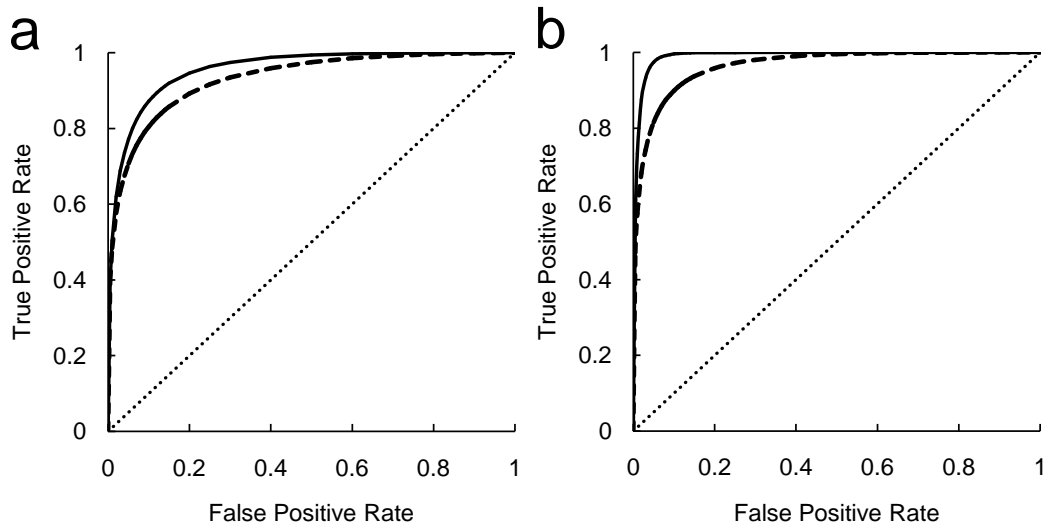
3 lentelė: Globalaus CYP3A4 slopinimo modelio prognozė vidiniam ir išoriniam patikrinimo duomenų rinkiniams.

Vidinis patikrinimo rinkinys	Eksp. +	Progn. +	Progn. -	Jautris: 75,4% Specifiškumas: 86,0% Tikslumas: 82,1%
	Eksp. -	52	17	
Išorinis patikrinimo rinkinys	Eksp. +	Progn. +	Progn. -	Jautris: 60,8% Specifiškumas: 84,4% Tikslumas: 76,0%
	Eksp. -	1.843	1.189	
		855	4.641	

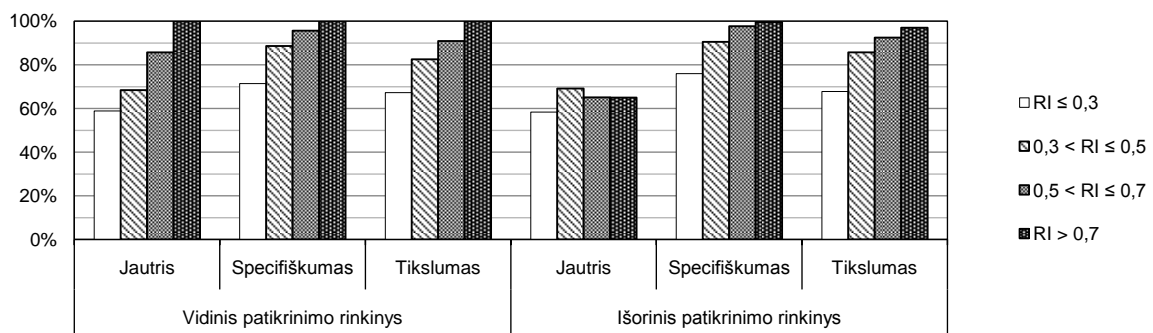
Literatūroje publikuoti duomenys buvo naudojami apmokymui. Gautas modelis atsižvelgia į bendras susijungimo su šiuo fermentu tendencijas ir prognozuoja tikimybę, kad junginys slopina CYP3A4, turėdamas  $IC_{50} < 40 \mu M$ . Vėliau šis modelis buvo apmokomas PubChem duomenimis. Parodyta, kad šis modelis gali prisitaikyti prie CYP3A4 slopinimo duomenų netgi tais atvejais, kai klasifikacijai naudojami skirtingi kriterijai, t. y. tik efektyvūs slopikliai ( $IC_{50} < 5 \mu M$ ) buvo laikomi aktyviais junginiais.

Globalus CYP3A4 slopinimo modelis buvo sukurtas naudojant logistinę PLS. Prognozavimo patikrinimo rinkiniams tikslumas yra apie 80% (3 lentelė). Tai, kad adityvus modelis gerai prognozuoja tikimybę būti CYP3A4 slopikliu, neprieštarauja žinioms apie šį baltymą. CYP3A4 aktyvusis centras yra pakankamai didelis, gali išsiplėsti prisijungiant ligandui ar kelioms ligandų molekulėms [15, 16]. Šį fermentą gali slopinti netgi junginiai, turintys molinę masę  $> 750 Da$ , nors pasiekus šią slopiklių masę kitų žmogaus citochromų slopinimo efektyvumas sumažėja [17].

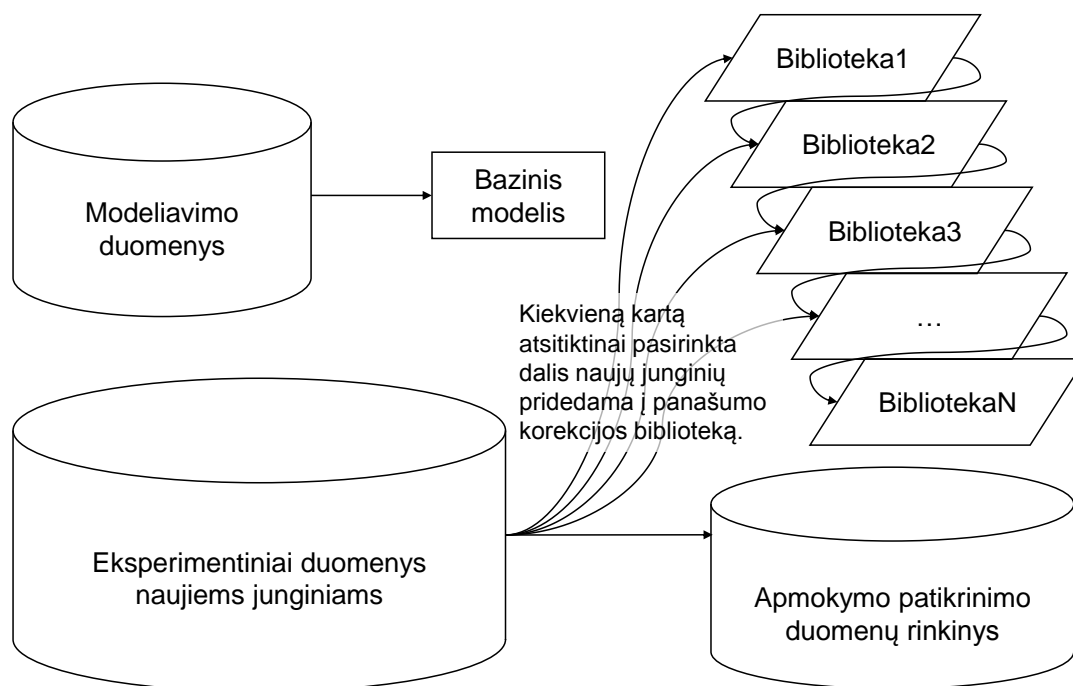
Visgi prognozavimo jautris, kuris yra žemesnis už specifiškumą, rodo, kad kai kurios CYP3A4 slopiklių savybės nebuvo aprašytos tiesiniame modelyje. Lokalios korekcijos pagal labiausiai panašių junginių eksperimentinius duomenis padeda išspręsti šią problemą. Be to, jos leidžia nuspėti prognozės patikimumą, skaičiuojant prognozės patikimumo indeksą ( $RI$ ), ir taip pat greitai apmokyti modelį. 1 paveiksle yra pateikiamos ROC kreivės, kurios po korekcijos pagal panašių junginių duomenis artėja link idealios klasifikacijos (taško (1;0)). Tai rodo prognozių pagerėjimą po lokalaus modeliavimo. Prognozių tikslumas koreliuoja su suskaičiuotomis patikimumo indekso reikšmėmis (2 pav.): visi statistiniai parametrai yra geresni esant didesnėms  $RI$  reikšmėms. Laikoma, kad junginiai, kurių  $RI < 0.3$ , nepatenka į modelio pritaikymo sritį [11]. Į ją patenka didžioji dalis (73%) vidinio patikrinimo rinkinio junginių, tačiau tik mažiau nei pusė (41%)



1 pav.: ROC kreivės, skirtos globalaus ir koreguoto CYP3A4 slopinimo modelio prognozavimo vidiniam patikrinimo rinkiniui palyginimui: a – junginiai, patenkantys į modelio pritaikymo sritį ( $RI > 0.3$ ); b – junginiai, kuriems prognozuojama su aukštu patikimumu ( $RI > 0.5$ ). Taškinė linija rodo atsitiktinį klasifikavimą, punktyrinė linija – bazinį modelį, ištisinė linija – modelį po panašumo korekcijos.



2 pav.: CYP3A4 slopinimą prognozuojančio GALAS modelio statistiniai parametrai skirstant prognozes pagal patikimumo indekso reikšmes.



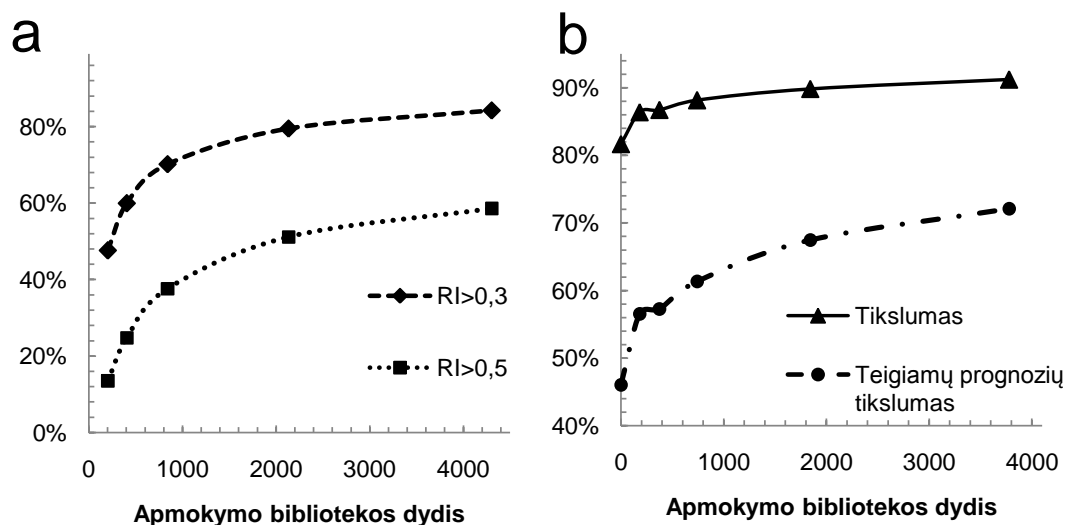
3 pav.: GALAS modelio apmokymo naujais junginiais testavimas.

PubChem duomenų. Akivaizdu, kad šiuo atveju yra reikalingas modelio apmokymas ir pritaikymo srities išplėtimas.

Išbandant galimybes apmokyti modelį, PubChem duomenys buvo naudojami kaip naujų eksperimentinių duomenų pavyzdys. Pusė šio rinkinio duomenų buvo naudojami testavimui, o kiti buvo po truputį pridedami į korekcijos pagal panašumą biblioteką, kaip parodyta 3 paveiksle. Visų pirma, modelio pritaikymo srities išplėtimas buvo parodytas pridedant šį rinkinį, klasifikuotą pagal tuos pačius kriterijus, kaip apmokymo duomenis (4a pav.). Mažiau nei pusė prognozių patenka į modelio pritaikymo sritį ( $RI > 0.3$ ) pridėjus apie 200 PubChem junginių į apmokymo biblioteką, o pridėjus dar 200 junginių (iš viso 400), ši sritis jau apima apie 60% prognozių. Toliau didinant apmokymo biblioteką, modelio apmokymo sritis didėja ir apima 89% patikrinimo rinkinio pridėjus > 4000 junginių (t. y. išsiplečia apie du kartus). Aukšto patikimumo prognozių ( $RI > 0.5$ ) skaičius didėja dar greičiau: nuo maždaug 14% pridėjus 200 junginių iki 59% pridėjus > 4000 junginių; iš viso apie keturis kartus. Šie rezultatai aiškiai parodo, kad modelis, sukurtas naudojant literatūros duomenis, prisitaikė prie cheminės erdvės dalies, kurioje yra PubChem duomenų rinkinio junginiai.

Vėliau tas pats 3 paveiksle pavaizduotas eksperimentas buvo pakartotas, naudojant Pub-

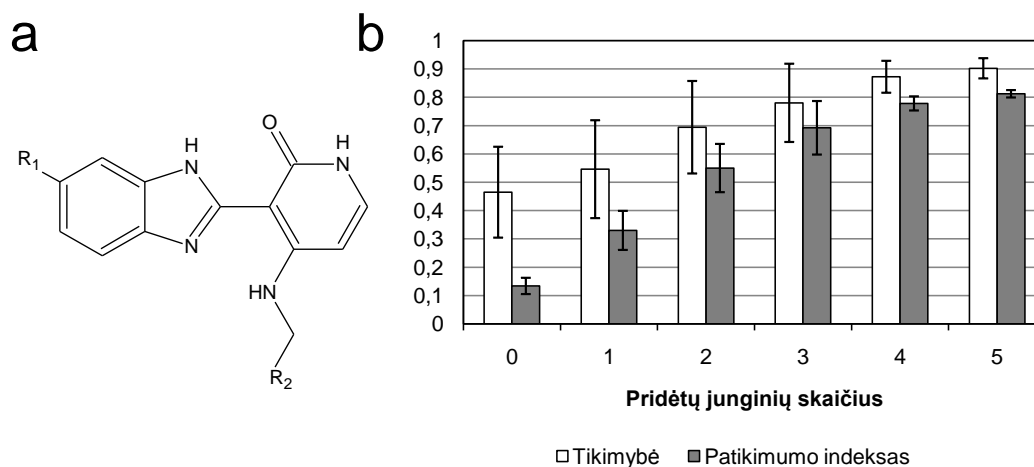




4 pav.: GALAS modelio apmokymo naujais junginiais testavimas: a – prognozių, patenkančių į modelio pritaikymo sritį, ir aukšto patikimumo prognozių skaičiaus augimas, pridedant PubChem duomenis, klasifikuotus taip pat, kaip modelio apmokymo rinkinys; b – prognozių tikslumo augimas, pridedant PubChem duomenis, imituojančius skirtingą eksperimentą.

Chem duomenų rinkinį, kuriame tik efektyvūs CYP3A4 slopikliai buvo laikomi aktyviais slopikliais. Tokiu atveju buvo imituojamas apmokymas duomenimis, gautais naudojant visiškai skirtingą eksperimentinį metodą. Svarbiausias apmokymo rezultatas pateiktas 4b paveiksle. Slopiklių atskyrimui naudojama kita  $IC_{50}$  riba ( $5 \mu M$ , o ne  $40 \mu M$ ), dėl to daug bazinio modelio prognozių yra klaidingai teigiamos. Teigiamų prognozių tikslumas yra tik apie 46%, t. y. tik apie pusę prognozuotų slopiklių yra efektyvūs CYP3A4 slopikliai. Klaidingai teigiamų prognozių skaičius mažėja ir teigiamų prognozių tikslumas auga, pridedant junginių į apmokymo biblioteką. Svarbu ir tai, kad patikimų prognozių skaičius didėja panašiai, kaip ir pirmajame eksperimente. Šie rezultatai rodo, kad GALAS modeliai gali būti apmokomi duomenimis iš pačių įvairiausių CYP3A4 slopinimo eksperimentų.

Trečiasis modelio apmokymo eksperimentas buvo daromas naudojant junginius, turinčius naują cheminį karkasą. Neseniai buvo publikuoti CYP3A4 slopinimo duomenys 10 naujų vaistinių junginių – IGF-1R (angl. insulin-like growth factor-1 receptor) slopiklių [18]. Visų šių junginių, kurių bendras karkasas pateiktas 5a paveiksle,  $IC_{50} < 40 \mu M$ , dėl to jie buvo klasifikuoti kaip aktyvūs slopikliai. Penki junginiai buvo pridėti į apmokymo



5 pav.: GALAS modelio apmokymas naudojant naujos cheminės klasės junginius: a – bendras junginių struktūrinis karkasas ( $R_1$  – alifatinis heterociklas,  $R_2$  – aromatinis pakaitas); b – vidutinių tikimybių ir  $RI$  reikšmių didėjimas (pasikliautiniai intervalai rodo standartinę nuokrypį).

biblioteką, o kitiems po kiekvieno pridėjimo buvo skaičiuojama tikimybė ir prognozės patikimumo indeksas. Pradinis modelis visiems junginiams pateikia tikimybes mažiau nei 0,5, o prognozių patikimumas labai mažas ( $RI$  tarp 0,09 ir 0,16). Kaip ir tikėtasi, pridėdant panašių junginių į panašumo korekcijos biblioteką, auga ir suskaičiuotos tikimybės būti CYP3A4 slopikliais, ir prognozės patikimumo indekso vertės. Pridėjus tris junginius, visiems IGF-1R slopikliams skaičiuojamos tikimybės ir  $RI$  reikšmės buvo  $> 0,5$  – modelis prisitaikė prie junginių turinčių naują cheminį karkasą.

### 3.2 Metabolizmo regioselektyvumo modelis

Regioselektyvumo modelio pirminio patikrinimo rezultatai yra pateikiami 4 lentelėje. Joje visų reakcijų patikrinimo rinkiniai yra sujungti į vieną duomenų rinkinį. Globalaus modelio prognozės yra neblogos, 68% metabolizmo vietų prognozuojama tikimybė  $> 0,5$ , o iš viso 79% atomų yra teisingai klasifikuojami. Panagrinėjus fragmentų koeficientus, galima interpretuoti šį modelį kaip aprašantį metabolizmo vietų aktyvumą citochromų P450 katalizuojamose reakcijose.

Visgi globalus modelis randa daug klaidingai teigiamų metabolizmo vietų (apie 16% visų prognozių). Lokalios korekcijos pagal panašumą keičia šią situaciją. Nors jautris

4 lentelė: Regioselektyvumo prognozavimo vidiniam patikrinimo rinkiniui rezultatai.

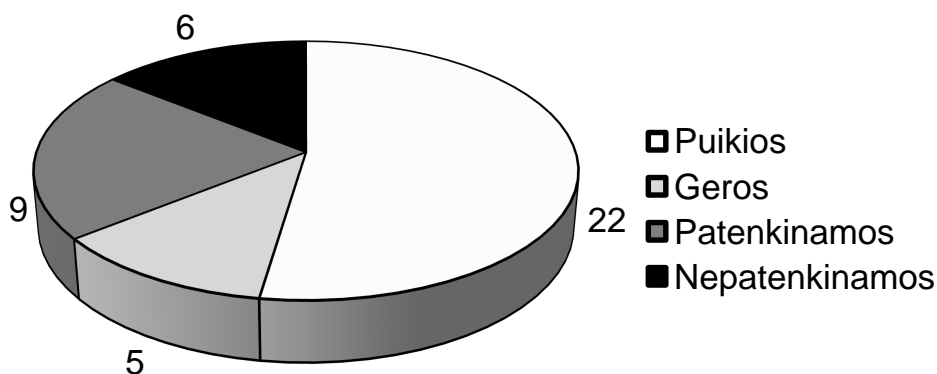
Globalus modelis	Steb. +	Progn. + 273	Progn. - 128	Jautris: 68,1% Specifiškumas: 80,7% Tikslumas: 78,8%
	Steb. -	420	1759	
Po lokalsios korekcijos	Steb. +	Progn. + 260	Progn. - 141	Jautris: 64,8% Specifiškumas: 93,9% Tikslumas: 89,4%
	Steb. -	133	2046	
Po lokalsios korekcijos, $RI > 0,3$	Steb. +	Progn. + 203	Progn. - 97	Jautris: 67,7% Specifiškumas: 96,3% Tikslumas: 92,5%
	Steb. -	71	1864	
Po lokalsios korekcijos, $RI > 0,5$	Steb. +	Progn. + 119	Progn. - 34	Jautris: 77,8% Specifiškumas: 98,4% Tikslumas: 96,2%
	Steb. -	22	1310	

šiek tiek sumažėja iki 65%, stebimas klaidingai teigiamų prognozių skaičiaus mažėjimas ir specifiškumo padidėjimas iki 94%. Bendras prognozavimo tikslumas išauga iki 89%. Panašūs ir stabilūs rezultatai buvo stebimi visų 5 reakcijų atveju. Be to, prognozės patikimumo indeksas efektyviai randa teisingas prognozes, kaip ir CYP3A4 slopinimo modelyje.

Vėliau modelis buvo patikrintas naudojant 42 naujų junginių duomenis. 30 junginių (71%) didžiausia tikimybė buvo prognozuota eksperimentiškai nustatytai metabolizmo vietai. Bent viena tokia vieta tarp trijų labiausiai tikėtinų atomų buvo rasta 35 junginiams (83%). Iš viso 43 iš 73 metabolizmo vietų buvo suskaičiuota tikimybė  $> 0,5$  (jautris 59%).

Dauguma ankstesnių regioelektyvumo modelių įvertinimų apsiriboja tokiu kelių metabolizmo vietų su didžiausiu įverčiu nagrinėjimu [19–23], tačiau šis požiūris nėra tinkamas. Atomai, kuriam prognozuotas didžiausias metabolizmo įvertis, žinoma, yra labiausiai tikėtina oksidacijos vieta molekulėje, bet pati suskaičiuota įverčio vertė (tikimybė, aktyvacijos energija ar kt.) gali būti nereikšminga ir rodyti, kad ši metabolizmo vieta nereikšminga, o cheminis junginys apskritai nėra biotransformuojamas. Dėl to vertinant sukurto modelio prognozių kokybę buvo stengiamasi atkreipti dėmesį ne tik į atomo rangą, išrikiavus juos pagal prognozuotą tikimybę, bet ir į pačias skaičiuotas tikimybes.

6 paveiksle yra pateiktas prognozių išorinio patikrinimo rinkinio junginiams pasiskirsty-



6 pav.: Regioselektyvumo modelio prognozių išorinio patikrinimo rinkinio junginiams pasiskirstymas pagal jų kokybę.

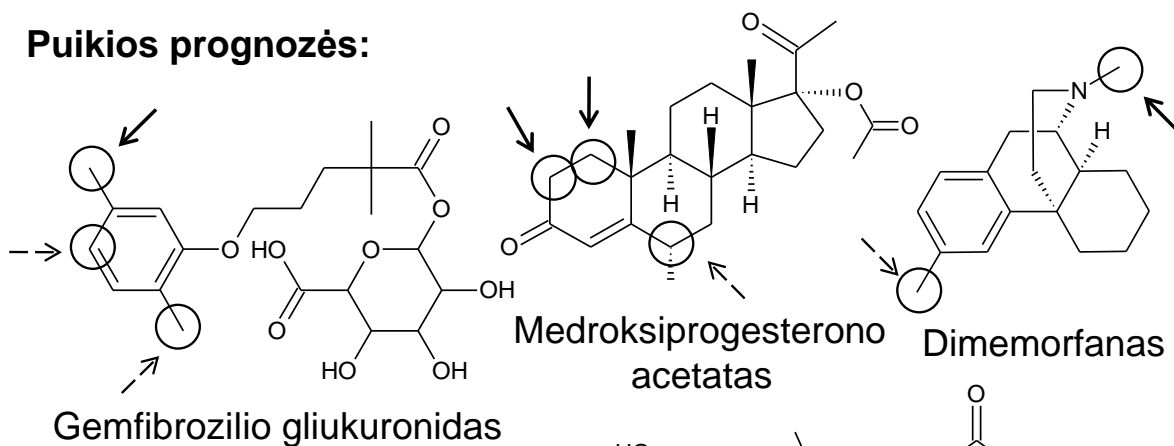
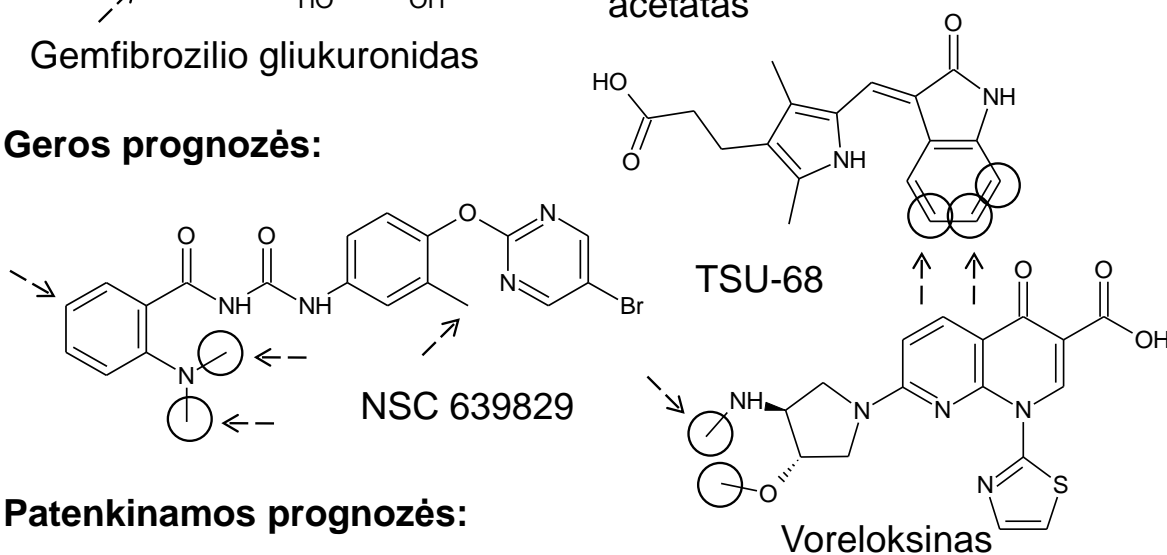
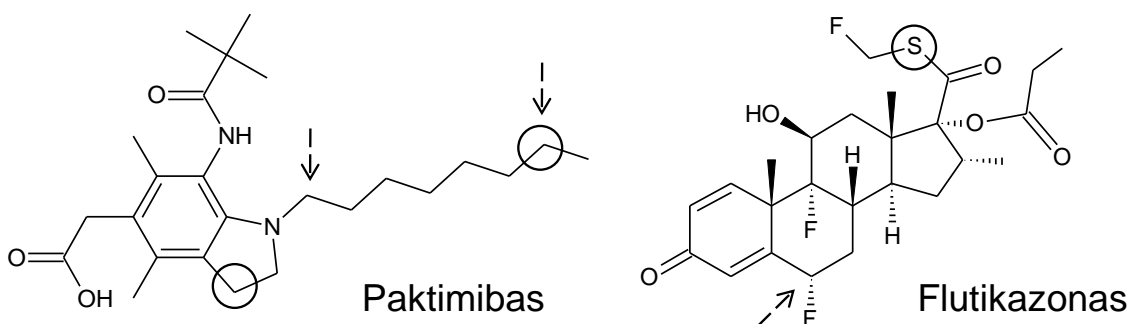
mas pagal priskirtas kokybės klases. 7 paveiksle yra pateikiami prognozių pavyzdžiai.

Regioselektyvumo prognozė junginiui buvo pažymėta kaip „puiki“, jei visoms eksperimentiškai nustatytoms metabolizmo vietoms buvo skaičiuojama tikimybė  $> 0,5$ , o atomas su didžiausia tikimybe buvo oksiduojamas. Tais atvejais, kai daugumai metabolizmo vietų buvo prognozuotos tikimybės  $> 0,5$ , prognozė buvo laikoma „gera“, nors spėjamos labiausiai tikėtinos metabolizmo vietos atomo oksidacija eksperimentiškai ne visada būdavo stebima.

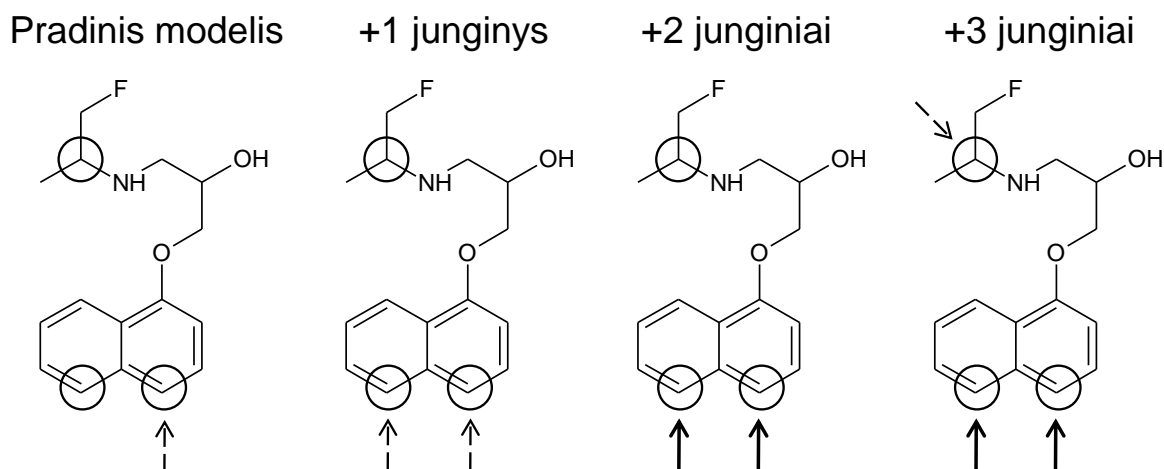
„Puikios“ ir „geros“ prognozės buvo gautos daugiau nei 60% junginių (22 „puikios“ ir 5 „geros“). Tarp jų yra įvairių vaistų klasių junginių. Akivaizdu, kad gerai žinomoms vaistų klasėms, tokioms kaip steroidai, opioidai, fosfodiesterazės-5 slopikliai, regioselektyvumo prognozės buvo „puikios“, nes daug šių junginių analogų yra modelio apmokymo rinkinyje. Tačiau svarbu ir tai, kad tarp junginių, kurių metabolizmo vietos prognozuotos teisingai, buvo pakankamai daug naujų vaistinių medžiagų.

Tais atvejais, kai mažiau nei pusei metabolizmo vietų suskaičiuota tikimybė  $> 0,5$ , prognozės buvo žymimos kaip „patenkinamos“. Taip pat šiai kokybės klasei buvo priskiriami junginiai, kurių vienintelė eksperimentiškai nustatyta metabolizmo vieta buvo pripažįstama kaip labiausiai tikėtina, bet tikimybė nesiekdavo 0,5. Nors daugumos „patenkinamų“ prognozių patikimumo indeksas buvo žemas, visgi bent vienas galimas metabolitas būdavo prognozuojamas.

Tik 6 junginiams nepavyko identifikuoti nei vienos metabolizmo vietos nei pagal su-

**Puikios prognozės:****Geros prognozės:****Patenkinamos prognozės:**

7 pav.: Regioselektyvumo prognozių pavyzdžiai. Eksperimentiškai nustatytos metabolizmo vietas yra apibrėžtos [24–31], ryškios rodyklės rodo patikimai prognozuotas metabolizmo vietas (tikimybė ir  $RI > 0,5$ ), punktyrinės – prognozuotas vietas su  $RI < 0,5$ .

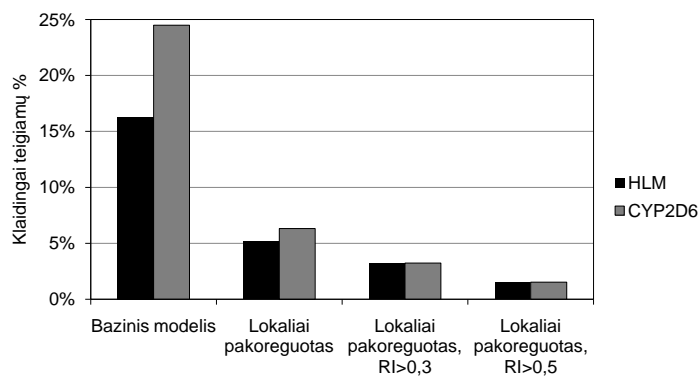


8 pav.: Fluoropropranololio metabolizmo vietų prognozavimo pokyčiai, apmokant regioselektyvumo modelį.

skaičiuotą tikimybę, nei pagal rangą („nepatenkinamos“ prognozės). 5 iš jų yra gamtiniai junginiai – tulžies rūgštys, antriniai metabolitai, kurių analogų nebuvo modelio apmokymo rinkinyje, sudarytame daugiausiai iš vaistų. Tokių gamtinių medžiagų biotransformacijos reakcijos paprastai yra nestandartinės, ir dėl to sunkiai prognozuojamos. Dėl to galima teigti, kad modelio prognozės gerai atspindi vaistinių junginių reaktingumą žmogaus citochromų P450 katalizuojamose reakcijose. Be to, lyginant šiame darbe sukurtą modelio prognozes su programos *SMARTCyp*, kuri remiasi kvantinės chemijos skaičiavimais, prognozėmis, buvo pastebėta, kad abi programos randa vienodai „puikių“ ir „gerų“ prognozių.

Viena iš pagrindinių problemų, pastebėtų naudojant modelį naujų junginių metabolizmo vietų spėjimui, yra tai, kad dauguma prognozių turėjo žemą patikimumo indeksą, t. y. naujų vaistinių medžiagų atomai nepriklauso modelio pritaikymo sričiai. Buvo nuspręsta pabandyti modelį apmokyti taip pat, kaip jau buvo padaryta CYP3A4 slopinimo atveju. Regioselektyvumo modelio adaptacijai tirti buvo pasirinkti propranololio analogai, kurie yra oksiduojami CYP2D6 ir CYP1A2 ir turi po tris metabolizmo vietas [32, 33].

Visų pirma buvo sukurtas specialus modelis, naudojant apmokymo rinkinį, iš kurio buvo pašalinti visi propranololio analogai ir metabolitai. Tuomet į korekcijos pagal panašumą biblioteką buvo po vieną pridėti trys junginiai. Po kiekvieno pridėjimo buvo skaičiuojamos metabolizmo tikimybės kiekvienam propranololio analogų atomui. Fluoropropranololio metabolizmo vietų spėjimo pavyzdys yra pateiktas 8 paveiksle.



9 pav.: Klaidingai teigiamų prognozių dalis žmogaus kepenų mikrosomų katalizuojamo metabolizmo regioselektyvumo modelyje ir modelyje, apmokytame naudojant CYP2D6 duomenis.

Pradinis modelis, sukurtas be propranololio darinių, prognozuoja tik vieną metabolizmo vietą iš trijų, t. y. tik „patenkinamai“. Po pirmo pridėto panašaus junginio, abi aromatinio hidroksilinimo vietos yra randamos, o pridėjus du analogus, jos yra prognozuojamos su dideliu patikimumu. Pridėjus trečią junginį, randama ir N-dealkilinimo vieta – gaunama „puiki“ prognozė. Svarbu yra tai, kad auga ne tik metabolizmo tikimybė, bet ir visų atomų prognozės patikimumo indeksas, o tai aiškiai parodo, kad naujos cheminės klasės junginiai tampa modelio pritaikymo srities dalimi.

Sėkmingai pritaikius regioselektyvumo GALAS modelį naujų junginių klasių prognozavimui, iškilo klausimas, ar galima šį modelį pritaikyti konkrečių fermentų katalizuojamų reakcijų regioselektyvumui prognozuoti? Juk tolimesnėse vaisto kandidato metabolizmo tyrimo stadijose yra visada nustatinėjama, kurie fermentai lemia konkrečių metabolitų susidarymą. Šiuo atveju jų katalizuojamo metabolizmo regioselektyvumo prognozavimas būtų labai naudingas. Jei tarsime, kad globalus regioselektyvumo modelis atspindi bendras reaktivity tendencijas, o lokaliai korekcijos – prisijungimo baltymo aktyviajame centre įtaką, tai naudojant tik atskirų fermentų duomenis kaip panašumo korekcijų biblioteką, turėtų būti įmanoma pritaikyti modelį jų specifiškumui.

Tokio pritaikymo išbandymui buvo pasirinktas CYP2D6, kuris pasižymi aiškiau išreikštu specifiškumu nei kiti žmogaus vaistų metabolizmo fermentai. CYP2D6 substratai yra aromatiniai junginiai, turintys bazinę grupę, nutolusią 5-10 Å nuo aromatinio žiedo [34, 35]. Kaip ir tikėtasi, didelė dalis bazinio modelio prognozių yra klaidingai teigiamos (apie 24% visų prognozių, lyginant su 16% bendrame žmogaus mikrosomų modelyje, 9

pav.). Po korekcijos naudojant CYP2D6 katalizuojamo metabolizmo duomenis, situacija keičiasi, ir abiejų modelių klaidingai teigiamų prognozių dalis yra panaši. Be to, filtruojant prognozes pagal patikimumo indeksą, matomos panašios klaidingų spėjimų skaičiaus sumažėjimo tendencijos, kaip ir žmogaus kepenų mikrosomų metabolizmo modelyje. Tai leidžia teigti, kad panašumo korekcija pagal CYP2D6 specifiškumo duomenis leido pritaikyti prognozavimą šio fermento specifiškumui. Manome, kad ši strategija turėtų sėkmingai veikti ir kitų mikrosominių fermentų atveju.



## 4 Išvados

1. Struktūros ir aktyvumo ryšio modelis prognozuojantis CYP3A4 slopinimą buvo sukurtas naudojant GALAS modeliavimo metodą. Jo prognozės gerai atitinka eksperimentų rezultatus: junginiams, priklausantiems modelio pritaikymo sričiai, prognozių tikslumas yra 89%.
2. Sukurtas modelis, prognozuojantis žmogaus kepenų mikrosomų katalizuojamo metabolizmo regioselektyvumą, skaičiuoja oksidacijos tikimybę kiekvienam molekulės atomui, ir taip pat pateikia prognozės patikimumo įvertinimą, taip apibrėždamas modelio pritaikymo sritį. Teisingai identifikuota 78% metabolizmo vietų iš tų, kurios buvo stebėtos eksperimentiškai ir prognozuotos su dideliu patikimumu.
3. Visuose sukurtuose modeliuose lokali korekcijos pagal panašumą pagerina klasifikacijos tikslumą, o suskaičiuotos prognozės patikimumo indekso vertės koreliuoja su prognozių kokybe.
4. Galimybė apmokyti GALAS modelį buvo parodyta pridėdant naujų cheminių junginių eksperimentinius duomenis į lokalią CYP3A4 slopinimo modelio dalį. Po šios procedūros modelio pritaikymo sritis išsiplėtė ir buvo prognozuojami nauji fermento slopikliai.
5. Taip pat pademonstruota regioselektyvumo modelio apmokymo galimybė. Modelis buvo pritaikytas anksčiau nerastų metabolizmo vietų prognozavimui molekulėse, turinčiose naują struktūrinį karkasą. Be to, modelis buvo pritaikytas konkretaus fermento – CYP2D6 – specifiškumo prognozavimui.
6. Regioselektyvumą prognozuojantis modelis buvo patikrintas naudojant 42 naujų junginių duomenis. 27 iš jų dauguma metabolizmo vietų buvo teisingai prognozuotos, o 36 molekulėse buvo rasta bent viena eksperimentiškai nustatyta metabolizmo vieta. Šie rezultatai yra palyginami su programos *SMARTCyp* prognozėmis, apskaičiuotomis kvantinės chemijos metodais.

## 5 Publikacijos disertacijos tema

### Straipsniai

1. Dapkunas, J.; Sazonovas, A.; Japertas, P. Probabilistic prediction of the human CYP3A4 and CYP2D6 metabolism sites. *Chem Biodivers* **2009**, *6*, 2101-2106.
2. Didziapetris, R.; Dapkunas, J.; Sazonovas, A.; Japertas, P. Trainable structure-activity relationship model for virtual screening of CYP3A4 inhibition. *J Comput Aided Mol Des* **2010**, *24*, 891-906.

### Žodiniai pranešimai konferencijose

1. Dapkunas, J., Sazonovas, A., Japertas, P. Probabilistic prediction of the human CYP3A4 and CYP2D6 metabolism sites. LogP2009, Zürich, Switzerland, February 8-11, 2009.
2. Dapkunas, J., Sazonovas, A., Japertas, P. Probabilistic model of regioselectivity of metabolism in human liver microsomes. ADMET Europe, Munich, Germany, March 28-29, 2011.

### Stendiniai pranešimai konferencijose

1. Dapkunas, J., Sazonovas, A., Japertas, P. Probabilistic prediction of the human CYP3A4 metabolism sites in a molecule. 235th ACS National Meeting, New Orleans, LA, USA, April 6-10, 2008.
2. Dapkunas, J., Sazonovas, A., Japertas, P. Probabilistic model of regioselectivity of metabolism in human liver microsomes. 16th North American ISSX Meeting, Baltimore, MD, USA, October 18-22, 2009.
3. Sazonovas, A., Didziapetris, R., Dapkunas, J., Juska, L., Japertas, P. GALAS modeling methodology applications in the prediction of the drug safety related properties. 16th North American ISSX Meeting, Baltimore, MD, USA, October 18-22, 2009.

4. Dapkunas, J., Didziapetris, R., Sazonovas, A., Japertas, P. Evaluation of ACD/ADME Suite 5.0 regioselectivity predictions. 9th International ISSX meeting, Istanbul, Turkey, September 4-8, 2010.
5. Didziapetris, R., Dapkunas, J., Sazonovas, A., Japertas, P. Probabilistic GALAS models for the prediction of the human cytochrome P450 inhibition. 9th International ISSX meeting, Istanbul, Turkey, September 4-8, 2010.
6. Dapkunas, J., Sazonovas, A., Didziapetris, R., Japertas, P. In silico identification of metabolic soft spots: case study using ACD/ADME Suite software. 241st ACS National Meeting & Exposition, Anaheim, CA, USA, March 27-31, 2011.
7. Dapkunas, J., Sazonovas, A., Japertas, P. QSAR model of regioselectivity of metabolism in human liver microsomes: development, validation, comparison and adaptation to novel compounds. 241st ACS National Meeting & Exposition, Anaheim, CA, USA, March 27-31, 2011.

## 6 Curriculum Vitae

### Justas Dapkūnas

Gimimo data: 1983-12-25.

El. paštas: justas.dapkunas@gmail.com

### Išsilavinimas:

Nuo 2007-10 – Vilniaus universiteto biochemijos krypties doktorantas.

2007-06 – Vilniaus universitetas, biochemijos magistras.

Magistrinio darbo tema: „Žmogaus citochromų P450 substratams būdingų fizikocheminių ir struktūrinių savybių identifikavimas“.

2005-06 – Vilniaus universitetas, biochemijos bakalauras.

Bakalaurinio darbo tema: „Citochromo P450 3A4 specifiškumo tyrimas in silico“.

2001-06 – Vilniaus Užupio gimnazija, abitūra.

### Darbo patirtis:

Nuo 2003-06 iki dabar: VšĮ „Aukštieji Algoritmai“ ADME/Tox grupės bendradarbis.

### Mokslinio domėjimosi sritys:

Su vaistų metabolizmu susijusių savybių modeliavimas: metabolizmo kepenyse regio-selektyvumo prognozavimas ir galimų metabolitų spėjimas, žmogaus citochromų P450 slopinimo QSAR analizė.

Kitų ADME savybių (nešiklių transporto žmogaus žarnyne, hematoencefalinės užtvaros perėjimo) ir toksiškumo (sąveikos su estrogenų receptoriais  $\alpha$ , genotoksiškumo) modeliavimas.

### Kita:

Šeimyninė padėtis – vedęs, žmona Rūta, dukros Ieva Marija ir Austėja.

Pomėgiai – muzika (ypač sunkusis metalas), Protmušis (intelektualus komandinis žaidimas), kelionės po Lietuvą.

Apdovanojimai – Lietuvos chemijos olimpiados antrosios vietos laimėtojas (2001), Baltijos chemijos olimpiados dalyvis (2001), Protmušio čempionas (2007).

## 7 Summary

The use of computational tools for estimation of drug metabolism-related properties is highly encouraged as these properties are desired to be known in the earliest stages of drug discovery. Prediction of possible sites of metabolism allows forecasting the major metabolites which facilitates experimental elucidation of their structures and optimization of metabolic stability. Moreover, *in silico* analysis of inhibition of xenobiotic metabolizing enzymes can be useful in the identification of compounds possessing the risk to cause drug-drug interactions.

The ADME properties of drug candidates, including hepatic metabolism, can be predicted using quantitative structure-activity relationship (QSAR) models. A novel QSAR modeling method, called GALAS (Global, Adjusted Locally According to Similarity), has been introduced recently. This method is a combination of a baseline global QSAR model and local similarity based corrections. It allows forecasting the reliability of prediction thus defining the model applicability domain. This domain can be later easily expanded by simply adding new data to the local part of the model. The application of the GALAS modeling method for drug metabolism-related properties, namely human cytochrome P450 3A4 (CYP3A4) inhibition and regioselectivity of metabolism in human liver microsomes, was the main objective of this study.

A structure-activity relationship model predicting the probability for a compound to inhibit CYP3A4 has been developed using data for > 800 compounds from various literature sources and tested on PubChem screening data. For compounds within the model applicability domain the statistical results approach the data consistency between experimental data from literature and PubChem datasets with the overall accuracy of 89%. However, the original model is applicable only for less than a half of PubChem database. Since the similarity correction procedure of GALAS modeling method allows straightforward model training, the possibility to expand the applicability domain has been investigated. Experimental data from PubChem dataset served as an example of in-house data. The model successfully adapted itself to both data classified using the same and different  $IC_{50}$  threshold compared with the training set. In addition, adjustment of the CYP3A4 inhibition model to compounds with a novel chemical scaffold has been demonstrated. The reported GALAS model is proposed as a useful tool for virtual screening of compounds for possible drug-drug interactions even prior to the actual synthesis.

Later a model predicting sites of human liver microsomal metabolism was developed. It calculates the probability to be metabolized by human cytochrome P450 enzymes for any atom in a molecule and allows forecasting the most probable phase I metabolites. Experimental data for 873 compounds were used for modeling. Every atom in the molecule has been marked whether it is a metabolism site or not, resulting in dataset with > 8500 atoms (> 1300 different metabolism sites). The whole dataset of atoms has been split into subsets according to reaction types and five different models have been developed for major reaction types (aromatic and aliphatic hydroxylation, N-dealkylation, O-dealkylation, S-oxidation).

The performance of the model was initially validated on internal test sets, which contained 30% of initial atoms. Baseline models already showed acceptable results, and the overall accuracy of predictions increased after the similarity based corrections. Moreover, the numbers of mispredictions reduced significantly when only results of higher reliability are taken into account.

The regioselectivity predictions have been evaluated using newly published data for 42 drug-like compounds. The predictions were divided into four groups according to their quality ("Excellent", "Good", "Satisfactory", and "Unsatisfactory"). At least one metabolism site was found for more than 80% of compounds, and more than 60% of them had most sites predicted ("Excellent" or "Good" results). Some compounds obtaining such predictions belong to well-known drug classes, which are well represented in the training sets, but there are some novel drugs as well. While most of the "Satisfactory" predictions had low reliability, the experimentally determined metabolism sites were still predicted correctly by the model. No metabolism sites were predicted for only 6 compounds ("unsatisfactory" results). Most of them are natural compounds which contain sites not typical for cytochrome P450-mediated metabolism. It is also important to note that the overall evaluation results are in good agreement with the predictions of *SMARTCyp* software, which is based on different computational methods.

The regioselectivity model was trained using propranolol analogues as an example of a new compound class. New metabolism sites could be identified after the training procedure, and the reliability of predictions increased for all atoms. This shows good promise for this model to be successfully adapted to in-house data in pharmaceutical industry. Analogously the developed model was adapted for cytochrome P450 isoform profiling by training with data for individual enzymes.

## Literatūros šaltiniai

- [1] Kerns, E.; Di, L. *Drug-like Properties: Concepts, Structure Design and Methods: from ADME to Toxicity Optimization*, 1st ed.; Academic Press: San Diego, 2008.
- [2] Nedderman, A. N. R. *Biopharm Drug Dispos* **2009**, *30*, 153–162.
- [3] Wienkers, L. C.; Heath, T. G. *Nat Rev Drug Discov* **2005**, *4*, 825–833.
- [4] Rendic, S.; Di Carlo, F. J. *Drug Metab Rev* **1997**, *29*, 413–580.
- [5] Tarcsay, A.; Keseru, G. M. *Expert Opin Drug Metab Toxicol* **2011**, *7*, 299–312.
- [6] Boyer, D.; Bauman, J. N.; Walker, D. P.; Kapinos, B.; Karki, K.; Kalgutkar, A. S. *Drug Metab Dispos* **2009**, *37*, 999–1008.
- [7] Anari, M. R.; Baillie, T. A. *Drug Discov Today* **2005**, *10*, 711–717.
- [8] Bonn, B.; Leandersson, C.; Fontaine, F.; Zamora, I. *Rapid Commun Mass Spectrom* **2010**, *24*, 3127–3138.
- [9] Worth, A. P.; Hartung, T.; Van Leeuwen, C. J. *SAR QSAR Environ Res* **2004**, *15*, 345–358.
- [10] Japertas, P. Fragmentinių ir statistinių metodų naudojimas biologiškai aktyvių junginių struktūros-aktyvumo sąryšio tyrimuose. Ph.D. thesis, Vilniaus universitetas, Vilnius, 2007.
- [11] Sazonovas, A.; Japertas, P.; Didziapetris, R. *SAR QSAR Environ Res* **2010**, *21*, 127–148.
- [12] The PubChem Project. <http://pubchem.ncbi.nlm.nih.gov/>.
- [13] Rydberg, P.; Gloriam, D. E.; Zaretski, J.; Breneman, C.; Olsen, L. *ACS Med Chem Lett* **2010**, *1*, 96–100.
- [14] Rydberg, P.; Gloriam, D. E.; Olsen, L. *Bioinformatics* **2010**, *26*, 2988–2989.
- [15] Ekroos, M.; Sjögren, T. *Proc Natl Acad Sci USA* **2006**, *103*, 13682–13687.
- [16] Yano, J. K.; Wester, M. R.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. *J Biol Chem* **2004**, *279*, 38091–38094.

- [17] Gleeson, M. P. *J Med Chem* **2008**, *51*, 817–834.
- [18] Velaparthi, U. et al. *J Med Chem* **2008**, *51*, 5897–5900.
- [19] Zhou, D.; Afzelius, L.; Grimm, S. W.; Andersson, T. B.; Zauhar, R. J.; Zamora, I. *Drug Metab Dispos* **2006**, *34*, 976–983.
- [20] Afzelius, L.; Arnby, C. H.; Broo, A.; Carlsson, L.; Isaksson, C.; Jurva, U.; Kjellander, B.; Kolmodin, K.; Nilsson, K.; Raubacher, F.; Weidolf, L. *Drug Metab Rev* **2007**, *39*, 61–86.
- [21] Trunzer, M.; Faller, B.; Zimmerlin, A. *J Med Chem* **2009**, *52*, 329–335.
- [22] Hennemann, M.; Friedl, A.; Lobell, M.; Keldenich, J.; Hillisch, A.; Clark, T.; Göller, A. H. *ChemMedChem* **2009**, *4*, 657–669.
- [23] Zheng, M.; Luo, X.; Shen, Q.; Wang, Y.; Du, Y.; Zhu, W.; Jiang, H. *Bioinformatics* **2009**, *25*, 1251–1258.
- [24] Baer, B. R.; DeLisle, R. K.; Allen, A. *Chem Res Toxicol* **2009**, *22*, 1298–1309.
- [25] Zhang, J.; Liu, Y.; Zhao, J.; Wang, L.; Ge, G.; Gao, Y.; Li, W.; Liu, H.; Liu, H.; Zhang, Y.; Sun, J.; Yang, L. *Drug Metab Dispos* **2008**, *36*, 2292–2298.
- [26] Chou, Y.; Chung, Y.; Liu, T.; Wang, S.; Chau, G.; Chi, C.; Soucek, P.; Krausz, K. W.; Gelboin, H. V.; Lee, C.; Ueng, Y. *J Pharm Sci* **2010**, *99*, 1063–1077.
- [27] Rudek, M. A.; Zhao, M.; Smith, N. F.; Robey, R. W.; He, P.; Hallur, G.; Khan, S.; Hidalgo, M.; Jimeno, A.; Colevas, A. D.; Messersmith, W. A.; Wolff, A. C.; Baker, S. D. *Clin Cancer Res* **2005**, *11*, 8503–8511.
- [28] Kitamura, R.; Asanoma, H.; Nagayama, S.; Otagiri, M. *Drug Metab Dispos* **2008**, *36*, 1003–1009.
- [29] Evanchik, M. J.; Allen, D.; Yoburn, J. C.; Silverman, J. A.; Hoch, U. *Drug Metab Dispos* **2009**, *37*, 594–601.
- [30] Kotsuma, M.; Tokui, T.; Ishizuka-Ozeki, T.; Honda, T.; Iwabuchi, H.; Murai, T.; Ikeda, T.; Saji, H. *Drug Metab Dispos* **2008**, *36*, 529–534.
- [31] Pearce, R. E.; Leeder, J. S.; Kearns, G. L. *Drug Metab Dispos* **2006**, *34*, 1035–1040.



- [32] Uptagrove, A. L.; Nelson, W. L. *Drug Metab Dispos* **2001**, *29*, 1377–1388.
- [33] Uptagrove, A. L.; Nelson, W. L. *Drug Metab Dispos* **2001**, *29*, 1389–1395.
- [34] Chohan, K. K.; Paine, S. W.; Waters, N. J. *Curr Top Med Chem* **2006**, *6*, 1569–1578.
- [35] 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.; Bridges, A. M. *J Biol Chem* **2006**, *281*, 7614–7622.