VILNIUS UNIVERSITY

INSTITUTE OF CHEMISTRY

Andrius Sazonovas

ESTIMATION OF THE ACUTE TOXICITY AND PREDICTION OF THE METABOLISM SITE FOR ORGANIC MOLECULES USING *GALAS* METHODOLOGY

Summary of doctoral dissertation

Physical Sciences, chemistry (03 P)

Vilnius, 2010

The work presented in this doctoral dissertation has been carried out at the faculty of Chemistry of Vilnius University from 2004 to 2009.

Scientific supervisor:

> Prof. Dr. habil. Eugenijus Butkus (Vilnius University, Physical Sciences, chemistry – 03 P)

The dissertation is defended at the Council of Chemistry sciences of Vilnius University

Chairman:

> Prof. Dr. habil. Sigitas Tumkevičius (Vilnius University, Physical Sciences, chemistry – 03 P)

Members:

> Dr. Daumantas Matulis (Institute of Biotechnology, Physical Sciences, biochemistry – 04 P)

> Doc. Dr. Vytas Martynaitis (Kaunas University of Technology, Physical Sciences, chemistry – 03 P)

> Dr. Viktoras Masevičius (Vilnius University, Physical Sciences, chemistry – 03 P)

> Dr. Olegas Eicher-Lorka (Institute of Chemistry, Physical Sciences, chemistry – 03 P)

Official opponents:

> Prof. Dr. habil. Feliksas Ivanauskas (Vilnius University, Physical Sciences, mathematics – 01 P)

> Doc. Dr. Gintaras Valinčius (Institute of Biochemistry, Physical Sciences, chemistry – 03 P)

The thesis defense will take place at 12 p.m. on May 21st, 2010, at the Auditorium of Inorganic Chemistry at the Faculty of Chemistry of Vilnius University.

> Address: Naugarduko 24, LT-03225, Vilnius, Lithuania

The summary of the thesis was distributed on the 21st of April, 2010.
The thesis is available at the Library of Institute of Chemistry and the Library of Vilnius University.

VILNIAUS UNIVERSITETAS

CHEMIJOS INSTITUTAS

Andrius Sazonovas

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

Vilnius, 2010

Disertacija parengta 2004-2009 metais Vilniaus universitete.

Mokslinis vadovas:

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

Disertacija ginama Vilniaus universiteto chemijos mokslo krypties taryboje

Pirmininkas:

    prof. habil. dr. Sigitas Tumkevičius (Vilniaus universitetas, fiziniai mokslai, chemija – 03 P)

Nariai:

    dr. Daumantas Matulis (Biotechnologijos institutas, fiziniai mokslai, biochemija – 04 P)

    doc. dr. Vytas Martynaitis (Kauno technologijos universitetas, fiziniai mokslai, chemija – 03 P)

    dr. Viktoras Masevičius (Vilniaus universitetas, fiziniai mokslai, chemija – 03 P)

    dr. Olegas Eicher-Lorka (Chemijos institutas, fiziniai mokslai, chemija – 03 P)

Oponentai:

    prof. habil. dr. Feliksas Ivanauskas (Vilniaus universitetas, fiziniai mokslai, matematika – 01 P)

    doc. dr. Gintaras Valinčius (Biochemijos institutas, fiziniai mokslai, chemija – 03 P)

Disertacija bus ginama viešame chemijos mokslo krypties tarybos posėdyje 2010 m. gegužės 21 d. 12 val. Vilniaus universiteto chemijos fakulteto neorganinės chemijos auditorijoje.

    Adresas: Naugarduko 24, LT-03225, Vilnius, Lietuva

Disertacijos santrauka išsiuntinėta 2010 m. balandžio 21 d.
Disertaciją galima peržiūrėti Chemijos instituto ir Vilniaus universiteto bibliotekose.

TABLE OF CONTENTS

INTRODUCTION

One of the most fundamental challenges in modern drug discovery is obtaining satisfactory pharmacokinetic properties of candidate molecules which in turn depend mainly on the ability of the compounds to withstand the metabolism of intestinal and hepatic enzymes. Among these the superfamily of cytochrome P450 enzymes and its two main isoforms CYP3A4 and CYP2D6 play a major role, being responsible for more than 50% of metabolism of marketed drugs [1]. The susceptibility of a new drug candidate to CYP3A4 or CYP2D6 metabolism may compromise its ability to reach the target at a required effective concentration or facilitate the induction of drug-drug interactions if administered with other substrates or inhibitors of these enzymes [2]. With the increasing pressure on the pharmaceutical industry to reduce the time and cost of new drug development there is a growing need to improve the success rate and optimize the whole process by enabling researchers to pick molecules that will successfully pass the various stages of drug development and testing as early as possible in advance [3]. These challenges lie at the origins of computational *in silico* methods aimed at theoretical predictions of various ADME (*Absorption, Distribution, Metabolism, and Excretion*) and toxicological endpoints. The regioselectivity of metabolism enzymes is no exception in this field as numerous methods have been proposed for the computational assessment of this important factor. It is possible to predict metabolism sites of these enzymes using quantum mechanical methods alone or in combination with simulated substrate – enzyme interactions utilizing known crystal structures of CYP3A4 and CYP2D6 [4-6]. However, these methods are usually time consuming and low throughput because of the time required to optimize the 3D structure of the query compound and all subsequent calculations and suffer from well known flaws of the docking methods if they are used. Alternatively the methods and programs also exist that allow identifying all parts of the molecule bearing the features that are known to be a prerequisite of the cytochrome P450 metabolism and therefore obtain a list of all theoretically possible metabolites of multiple enzymes. Quite obviously careful and intelligent application of structure based regioselectivity prediction methods utilizing the information about enzyme's active site can yield very good results, however, they require more experimental data and are best used at later stages of drug development [7]. For cytochrome P450 regioselectivity predictions at the earliest possible stage, including virtual library screening, purely empirical ligand based methods are the most suitable ones. They are fast, require no *a priori* knowledge about compound metabolism, and provide the simplest way to utilize constantly increasing amounts of experimental cytochrome P450 regioselectivity data. Up to date several papers have been published describing the prediction of reaction sites for some cytochrome P450 superfamily enzymes using fingerprints of know substrates or by QSAR models. However, this field of empirical *in silico* enzyme regioselectivity prediction is relatively new and there is still plenty of room for improvement.

Cytochrome P450 enzymes play a major role in the protection of the organism from foreign chemicals (xenobiotics) by facilitating their detoxification and elimination. Ironically, sometimes such transformations lead to the intermediates that are more toxic than the parent compound being metabolized, resulting in the induction of chemical toxicity as well as carcinogenicity. As a result the evaluation of the compound metabolism always goes hand-in-hand with the assessment of the toxicological properties of the identified metabolites. One of the parameters indicating the potential

toxicity of a chemical compound is the median lethal dose (LD_{50}) which indicates the dose that kills 50% of animals within 24 hours of administration. This value has been within the scope of researchers working in the field of QSAR model development for a very long time and a great number of various methods for *in silico* LD_{50} predictions have been proposed over the years. Many alternative ways for the estimation of acute toxicity have been investigated as well, including but not limited to combinations of "expert systems" and individual class QSARs, various correlations with *in vitro* data as well as analysis of interspecies dependencies of acute toxicity. However, each one of them has its own problems and shortcomings significantly limiting their role as a practical solution to the problem of acute toxicity estimation. As a result, despite all the criticism towards QSAR models of LD_{50} and even doubts expressed regarding the feasibility of such predictions in general [8, 9], no other methods exist at the moment allowing fast estimation of LD_{50} values for any larger compound set. Recent reviews of the existing QSAR methods strongly suggest that they must address the problems of high data variability (complex mechanisms and errors) and data gaps (uncharacterized chemical spaces). These requirements conform with the latest OECD principles for QSAR validation, according to which any such models should be associated with a defined "Applicability Domain" (AD) [10], or the "response and chemical structure space in which the model makes predictions with a given reliability" [11].

The ultimate complexity and variability of both CYP450 enzyme regioselectivity and LD_{50}, governed by the huge number of possible mechanisms underlying these phenomena, turns them into a very likely subject to various local non-linear effects. All these facts make those two properties a very special case as well as an excellent challenge for the validation of a methodology capable of QSAR AD assessment, such as GALAS (*Global, Adjusted Locally According to Similarity*) modeling method used in this work. The analysis of the acute toxicity has been included among the objectives of this work as both being a topic very closely related to the compound metabolism in the organism and an intermediate step in complexity between regioselectivity of metabolism enzymes and the simple physicochemical properties for which the successful application of the GALAS methodology has been already reported.

Specific aims:

1. Develop individual acute toxicity (LD_{50}) models using GALAS modeling method for the following systems of rodent species and administration routes: mouse (oral, intraperitoneal, intravenous, and subcutaneous) and rat (oral and intraperitoneal).

2. Assess the possibilities of the GALAS modeling method application for the prediction of atom-centered properties by developing the GALAS regioselectivity models for the main metabolic reactions (N-dealkylation, O-dealkylation, aromatic and aliphatic hydroxylation) mediated by two major CYP450 isoforms (CYP3A4 and CYP2D6).

3. Evaluate the performance of the main GALAS methodology features (local corrections according to similarity, model applicability domain assessment via the estimation of prediction reliability and the ability to adapt to the experimental data for new compounds) in the context of significantly increased complexity of the modeled objects.

Scientific novelty:

Acute toxicity models presented in this work have been developed using one of the largest LD_{50} dataset published up to date and the whole work in this field represents one of the most extensive studies on LD_{50} prediction.

Both the acute toxicity and regioselectivity models derived during this work provide the quantitative estimates of their prediction reliability allowing the effective and accurate assessment of the applicability domain of those models

In addition, the reported regioselectivity models are the first example of the GALAS modeling methodology application for the prediction of atom-centered properties.

Practical value:

All of the reported models conform to one of the main requirements for the regulatory acceptance of QSAR models as alternative research methods – the ability to evaluate their domain of applicability. This work also serves as a demonstration that meeting this requirement of QSAR AD assessment is no longer an ambiguous task. Since the GALAS methodology successfully copes with this task in the analysis of such complicated properties as acute toxicity or enzyme regioselectivity it may be as well adapted to do the same for any properties of similar or lesser complexity.

The presented possibility to subdivide all the results of the GALAS model into several classes that provide information about the expected error of the prediction can be used for compound prioritization before experimental testing and even help reduce the number of such measurements.

The illustrated ability of the GALAS model to adapt itself to the new experimental data allows the expansion of the applicability domain of the reported models to any region of the chemical space. This opens wide possibilities for the practical industry 'in-house' applications of the reported models based on publicly available data.

Findings presented for defense:

Individual models have been created for the estimation of the acute toxicity towards two rodent species following a number of different administration routes as well as for the prediction of CYP3A4 and CYP2D6 regioselectivity in the main metabolic reactions mediated by these enzymes (13 individual models in total).

Each model features the ability of its applicability domain assessment and expansion utilizing user-defined new experimental data.

GALAS modeling methodology has been expanded and adapted for the prediction of binary atom-centered properties, *i.e.* the enzyme regioselectivity.

Scientific approbation and publication of the presented work:

The results of this doctoral study have served as the basis for 2 publications in the journals included in the Thomson Reuters ISI database, as well as 1 oral and 15 poster presentations at the international conferences.

Content of the dissertation:

The dissertation has been written in Lithuanian language on 139 pages including 16 tables, 21 figures, and 144 citations as well as 2 appendices.

RESULTS AND DISCUSSION

1   Data compilations used in the modeling

1.1   Acute toxicity (LD_{50}) dataset

LD_{50} values were taken mainly from the Registry of Toxic Effects of Chemical Substances (RTECS[®]) database [12]. This database was rigorously reviewed and "cleaned" by removing any non-covalent complexes, salts, compounds with incorrect structures (identified automatically), and unusually high deviations in interspecies correlations (animal *vs.* animal, and administration *vs.* administration). Whenever available, the acute toxicity data from the IUCLID Chemical Data Sheets was used to validate, correct or exclude entries of RTECS[®]. The IUCLID database (accessible on line via ESIS, the *European chemical Substances Information System*) provides information on 2,604 high production volume chemicals reported by the European industry in the frame of the existing chemicals risk assessment program [13]. Additionally, this database provided some new compounds that were not available in RTECS[®]. The final database contained nearly 75,000 compounds in several animal/administration systems, as summarized in Table 1. Each data set was randomly dissected into the training (70%) and validation (30%) sets.

Table 1. Numbers of compounds in the analyzed acute toxicity (LD_{50}) data sets

Species	Administration route	*N*		
		Total	**Training Set**	**Test Set**
Rat	Oral (OR)	8,631	6,464	2,167
	Intraperitoneal (IP)	5,002	3,751	1,251
Mouse	Oral (OR)	19,571	14,678	4,893
	Intraperitoneal (IP)	36,031	27,004	9,027
	Intravenous (IV)	19,963	14,972	4,991
	Subcutaneous (SC)	8,577	6,432	2,145

A recent study on the acute toxicity modeling, involving the analysis of an extensive dataset of LD_{50} values for oral exposure in rats [14], emerged as a source of additional valuable data during the course of this work. Upon request, a collection of 7,385 compounds featured in this publication has been kindly provided by the authors in its full format – *i.e.* including all the structures and experimental LD_{50} values[†]. The analysis of this data revealed that it contains 2,718 compounds not present in the dataset of the rat acute oral toxicity model (one of the six models presented here), consequently providing the basis for an external validation study in this particular case.

[†] The author of this dissertation is particularly thankful to Todd M. Martin, Alexander Tropsha and their colleagues for kindly providing the complete dataset, utilized in their recent study of rat acute toxicity by oral exposure [14].

1.2 Metabolism regioselectivity data

The regioselectivity models for the two metabolism enzymes described here were built using experimental CYP3A4 and CYP2D6 metabolism data from scientific publications for 560 and 526 compounds, respectively. In each molecule every carbon atom with at least one hydrogen atom attached was assigned whether it is a site of metabolism mediated by a corresponding enzyme. The datasets of >5,500 marked atoms for the two enzymes were divided into aromatic hydroxylation, N-dealkylation, O-dealkylation and aliphatic hydroxylation subsets according to atom type and position in the molecule [15]. The structure of the obtained subsets is outlined in Table 2. Each subset was further divided into training (70%) and test (30%) sets and individual models were built for all types of reactions.

Table 2. Summary of experimental data used for CYP3A4 and CYP2D6 regioselectivity modeling

Subset	N	No. of metabolism sites	Total No. of marked atoms
CYP3A4			
N-dealkylation	345	199	824
O-dealkylation	310	104	688
Aliphatic hydroxylation	475	172	2,057
Aromatic hydroxylation	462	111	2,087
CYP2D6			
N-dealkylation	357	73	849
O-dealkylation	296	61	610
Aliphatic hydroxylation	456	27	1,901
Aromatic hydroxylation	458	86	2,110

2 Descriptor generation

2.1 Fragmental descriptors in acute toxicity (LD_{50}) modeling

In case of the acute toxicity modeling the training set compounds were characterized in terms of fragments as structural descriptors because of their clear relation to the chemical structure of the molecule, leading to the ease of interpretation of the resulting model. The defined set of 404 fragmental descriptors can be considered as a combination of two parts serving different purposes. The major part of the utilized fragment set was intended for the description of the general chemical constitution of any compound and was comprised of conventional fragmental descriptors, such as atoms, functional groups, molecular 'shape fragments', *etc.* This descriptor set has been obtained by expanding the list of fragments successfully used by J. A. Platts and his co-

workers in their studies [16]. This initial set was expanded with a group of more complex fragments, generally called toxicophores – substructures identified to be responsible for the toxic action of the molecules possessing them. This part of the set was added in order to account for, at least, the most widely-known specific mechanisms and interactions leading to the manifestation of high acute toxicity that have been already identified and confirmed in the previous Outlier-based [17] and Classification-SAR studies. *E.g.*, phosphates, thiophosphates, and carbamates (cholinesterase inhibition), methylene fluorides (Krebs cycle inhibition), mustard derivatives, activated methylene halides, aziridinium and aziridine derivatives (alkylation of macromolecules), activated nitriles (respiratory chain inhibition), activated double bonds (alkylation through the Michael type addition), bicyclophosphates, orthocarboxylates, and silatranes (non competitive GABA receptor inhibition), *etc*.

2.2   Atom-centered layer fragmentation

As it was mentioned previously the regioselectivity of any enzyme is a feature associated with each individual atom in a molecule. In this case a special fragmentation method has to be used that yields different unique molecule representations in the variable matrix depending on the selected central atom as opposed to the traditional fragmentation techniques providing one general "digital image" of the whole molecule. As a result, molecules have been fragmented using atom-centered fragmentation scheme. This method provides information about the atom types present at equidistant positions (called Layers) from the atom marked as a positive or negative metabolism center. Schematically the whole process of molecule fragmentation in this way starting from a selected starting atom is illustrated in Figure 1. The first three layers including the metabolism site itself were described in detail. The merging scheme was applied to the following layers allowing them to overlap to a certain extent, *e.g.* 3-rd and 4-th layers
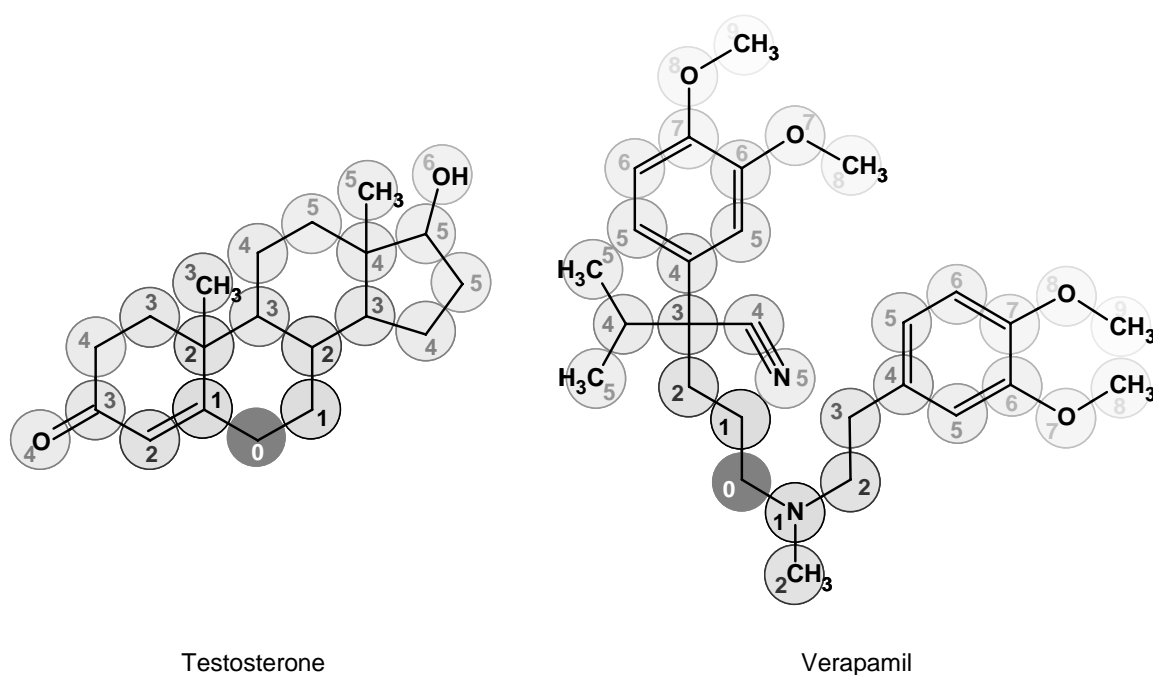

Testosterone   Verapamil

Figure 1. Selected examples of the atom-centered layer fragmentation

were merged and used as a single variable in the statistical analysis followed by merged 4-th, 5-th and 6-th layers, *etc.* In case of the aromatic carbon hydroxylation modeling the exact description was extended to the first five layers in order for it to cover the first layer of all possible substituents in a six-membered aromatic ring. Furthermore, a number of classical representatives of functional groups with strong electronic effects (*e.g.*, -NO_2, -CF_3, *etc.*) were used in the model development. The significance of presence of such groups depending on the distance from the marked atom was assigned in analogy with the known extent of the electronic interaction propagation over the distance in the molecule.

3   *Statistical methods*

3.1   GALAS modeling methodology

Schematically the GALAS model can be viewed as a combination of two systems:
- A structure based QSAR model for the prediction of the property of interest – baseline model;
- A similarity based routine which identifies the most similar compounds contained in the training set and considering their experimental values calculates systematic deviations produced by the baseline QSAR model for any submitted test compound.

The first part is a global model. In this particular instance it is based on either PLS [18] or BPLS – linear, additive methods. It learns general trends and also defines what is "similar" when comparing compounds in terms of a particular analyzed property. On the other hand, the second layer of the methodology can be considered as a local model. It adds non-linearity by making corrections to the global predictions after the analysis of the baseline model performance in a local chemical environment of a query molecule.

3.2   Global QSAR

Fragmental QSARs for the prediction of acute toxicity for every species and route of administration analyzed as well as regioselectivity for each enzyme and reaction type under consideration were developed using PLS [18] and BPLS methods, respectively, in combination with bootstrapping [19] technique. BPLS is a variation of PLS, possessing all the useful features of the latter method in combination with the ability to analyze binary data. The bootstrapping method implies random compound sampling from the initial training set, *i.e.* generation of new "training sub-sets" and derivation of independent model for each sub-set. This procedure is performed 100 times in all cases. Each of the sampled sub-sets is of the same size as the initial training set, however, random manner of their population results in some compounds being selected more than once, others being omitted. Therefore, each global QSAR model in this study actually represents an ensemble of 100 PLS or BPLS models, providing each compound with a vector of 100 corresponding property predictions, each based on a slightly different sub-set of the initial training set. By definition, two compounds with similar trends in the variation patterns of those 100 value vectors predicted by a set of bootstrap PLS (BPLS) models are considered similar in terms of the analyzed property, *i.e.* the differences in

the compound sets used to parameterize each of 100 models, constituting a baseline QSAR, affect estimations for the two compounds in a similar way. Quantitatively the information about the similarity of two compounds is expressed in terms of their individual Similarity Index (SI_i) calculated from the correlation of the corresponding two predicted property value vectors.

3.3 Local corrections (Δ) of the baseline predictions

Global predictions alone may not be very useful, as most of acute effects occur in highly localized chemical spaces that deviate from linear trends. In an attempt to "capture" such effects, each global prediction is further subjected to the local similarity correction procedure. It is based on the analysis of the performance of the global PLS or BPLS QSAR model in the local environment of the query compound, *i.e.* the comparison of the experimental data and baseline predictions of the QSAR model for the most similar compounds from the training set. If baseline predictions for these compounds show any systematic deviations from their reported measured values, the same is expected to happen with the considered similar query structure, *i.e.* its baseline prediction requires local adjustments. The required correction (Δ) is calculated as a weighted average from the differences between global QSAR predictions and experimental data for the most similar compounds in the training set:

$$\Delta = \sum_{i=1}^{n} a^{i-1} \cdot SI_i \cdot \Delta_i / \sum_{i=1}^{n} a^{i-1} \tag{1}$$

where Δ − correction that should be applied for the given prediction from the global model; a − a constant, influencing calculation of the weighted average, the simple average value will be calculated if this constant is set to 1; SI_i − similarity (individual Similarity Index) between given compound and the i-th most similar compound in the training set, calculated as correlation coefficient between corresponding vectors, made of multiple estimated values from baseline bootstrapping models; Δ_i − difference between measured value and value predicted by global model for the i-th most similar compound: $\Delta_i = Y_i - \bar{\hat{Y}}$; n − a constant, that determines how many similar compounds should be taken into consideration while estimating correction.

Obviously, the result depends on the number of similar compounds considered (n in Eq. 1). Empirically it has been observed that optimum result is obtained when considering five most similar compounds. Using a smaller number of similar structures increases the risk of erroneous corrections, whereas using a higher number makes corrections overly "conservative".

3.4 Evaluation of the model applicability domain

As it was mentioned previously, the GALAS modeling methodology allows quantitative estimation of the applicability domain of each created model. This essential feature takes into account the following two aspects:

- Similarity of the tested compound to the training set. No reliable predictions can be made if there are no similar compounds in the training set. Quantitative criterion – Similarity Index (*SI*).
- Consistence of the experimental values for similar compounds. Even if similar compounds are present in the dataset the quality of prediction could be lower if that data is inconsistent with regard to the baseline model. Quantitative criterion – Data-Model Consistency Index (*DMCI*)

The aforementioned Similarity Index is calculated by weighted averaging of all the individual Similarity Indices (*SI$_i$*) for the test molecule and each of the *n* most similar compounds from the training set. On the other hand, the *DMCI* value basically compares the individual differences between experimental and predicted baseline property values (*Δ_i*) for the same most similar compounds from the training set with the overall local correction for the compound of interest calculated by the Equation 1. The more individual differences are scattered around the calculated average (*Δ*), the more inconsistent are the data for the similar compounds with regards to the global baseline model and vice versa:

$$SI = \sum_{i=1}^{n} a^{i-1} \cdot SI_i \Big/ \sum_{i=1}^{n} a^{i-1} \tag{2}$$

where *SI* – Similarity Index to the compounds in the training set;

$$DMCI = e^{-\left(\sum_{i=1}^{n} a^{i-1} \cdot SI_i \cdot (\Delta - \Delta_i)^2 \Big/ \sum_{i=1}^{n} a^{i-1}\right)/b} \tag{3}$$

where *DMCI* – Data Model Consistency Index; *b* – empirical constant, which is used as a threshold of unacceptable data scattering, indicating when *DMCI* should approach 0, *e.g.* $1/10^{\text{th}}$ of the full range of values that the property under consideration obtains for the training set compounds.

Number of the most similar compounds (*n*) considered here is the same as in the calculation of the similarity based correction to the baseline predicted property values (*Δ* in Eq. 1) outlined in the previous section, *i.e.* five.

The final prediction Reliability Index characterizing the applicability domain of the model is calculated in the following manner:

$$RI = SI \cdot DMCI \tag{4}$$

Both *SI* and *DMCI* were scaled to vary from 0 to 1, so the resulting *RI* also varies in this range. If *RI* approaches zero, then a given compound is far from the model applicability domain and the respective prediction is unreliable. This is observed when either *SI* or *DMCI* approach zero, *i.e.*, when either no similar structures are present in the training set or such structures have inconsistent experimental data (respectively to baseline model). If *RI* approaches 1, then a given compound is within the model applicability domain and the respective prediction is highly reliable. This is only observed if both *SI* and *DMCI* approach 1, *i.e.*, very similar compounds are found in the training set and their experimental values are consistent with global predictions.

4   Acute toxicity (LD_{50}) modeling results

4.1   Global *vs.* local models

Table 3 lists the results of predictions for all acute toxicity test sets that were not used in corresponding QSAR development. Compounds with unreliable predictions ($RI < 0.3$) were excluded from considerations (*ca*. 5-10% of each test set), as by definition they fall outside of the model applicability domain and hence provide no meaningful information about the models' performance. "Baseline" indicates global PLS predictions (mean values of 100 bootstraps), whereas "Final" indicates the results after local similarity corrections (Δ).

Table 3. Prediction results for the acute toxicity test sets after excluding compounds outside the applicability domains of corresponding models (with $RI < 0.3$)

Species	Administration route	$N_{RI>0.3}$	Baseline		Final	
			R^2	*RMSE*	R^2	*RMSE*
Rat	Oral (OR)	1,976 (91%)	0.45	0.66	0.56	0.59
	Intraperitoneal (IP)	1,130 (90%)	0.31	0.62	0.42	0.58
Mouse	Oral (OR)	4,545 (93%)	0.30	0.47	0.49	0.40
	Intraperitoneal (IP)	8,568 (95%)	0.35	0.50	0.56	0.41
	Intravenous (IV)	4,754 (95%)	0.47	0.48	0.61	0.41
	Subcutaneous (SC)	2,056 (96%)	0.40	0.59	0.54	0.52

As expected, in all cases local corrections made a notable impact (R^2 increased by >0.1, *RMSE* decreased by >0.05), confirming a significant influence of various non-linear effects that were considered above. Yet, even after such corrections have been made, the "Final" results are still "far from perfect". This could also be anticipated, as there were good reasons why earlier reviewers insisted on any QSAR applicability domain assessment. Thus the primary interest of this work lays not in R^2 or *RMSE* values by themselves, but rather in their dependence on *RI*.

4.2   *RMSE vs. RI*

Figure 2 shows that in all cases *RMSE* continuously decreases as *RI* increases. When *RI* exceeds a certain threshold (0.5 or 0.75) *RMSE* falls well below 0.5 log units, approaching typical accuracy of LD_{50} determination in different laboratories. These two observations alone clearly indicate that the objective of accurate assessment of the applicability domain of the model has been achieved. The only area where *RMSE* doesn't correlate with *RI* is under $RI < 0.3$. The fluctuations of *RMSE* values for these predictions, observed in a few cases in Figure 2 (parts (b), (d), and (f)), can be explained by the fact that sometimes the number of compounds falling into certain *RI* ranges is so low that it even distorts the statistical results. However, such distortions are of minor importance, as they only concern 5-10% compounds (depending on the set) and in any

case corresponding *RMSE* values are large enough to justify the labeling of all such predictions as "unreliable". Another important observation is that when compounds closely approach (or are within) the QSAR AD (*RI* > 0.3), all *RMSE vs. RI* dependences
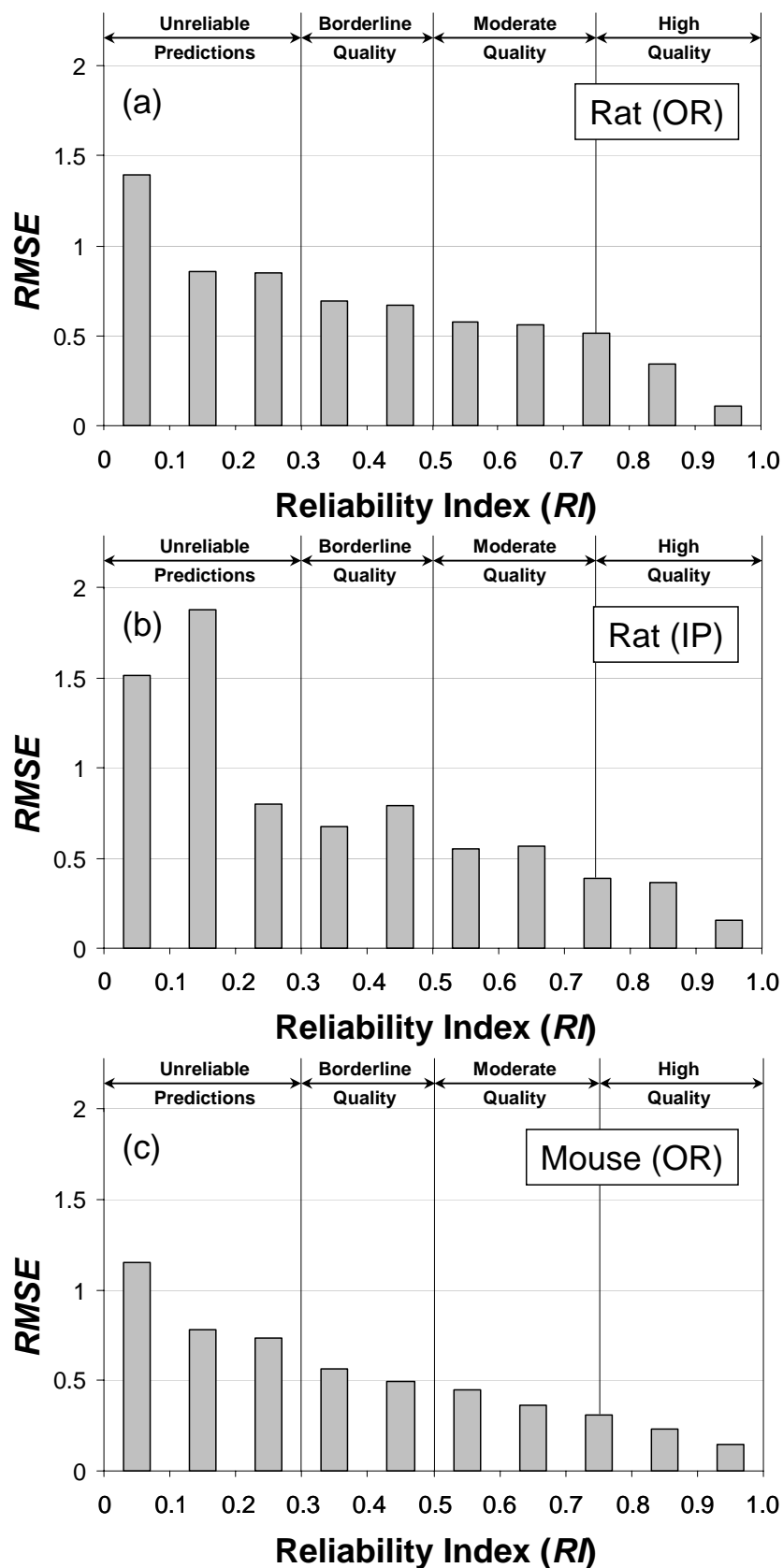

Figure 2. *RMSE vs. RI* dependences obtained for the presented acute toxicity models
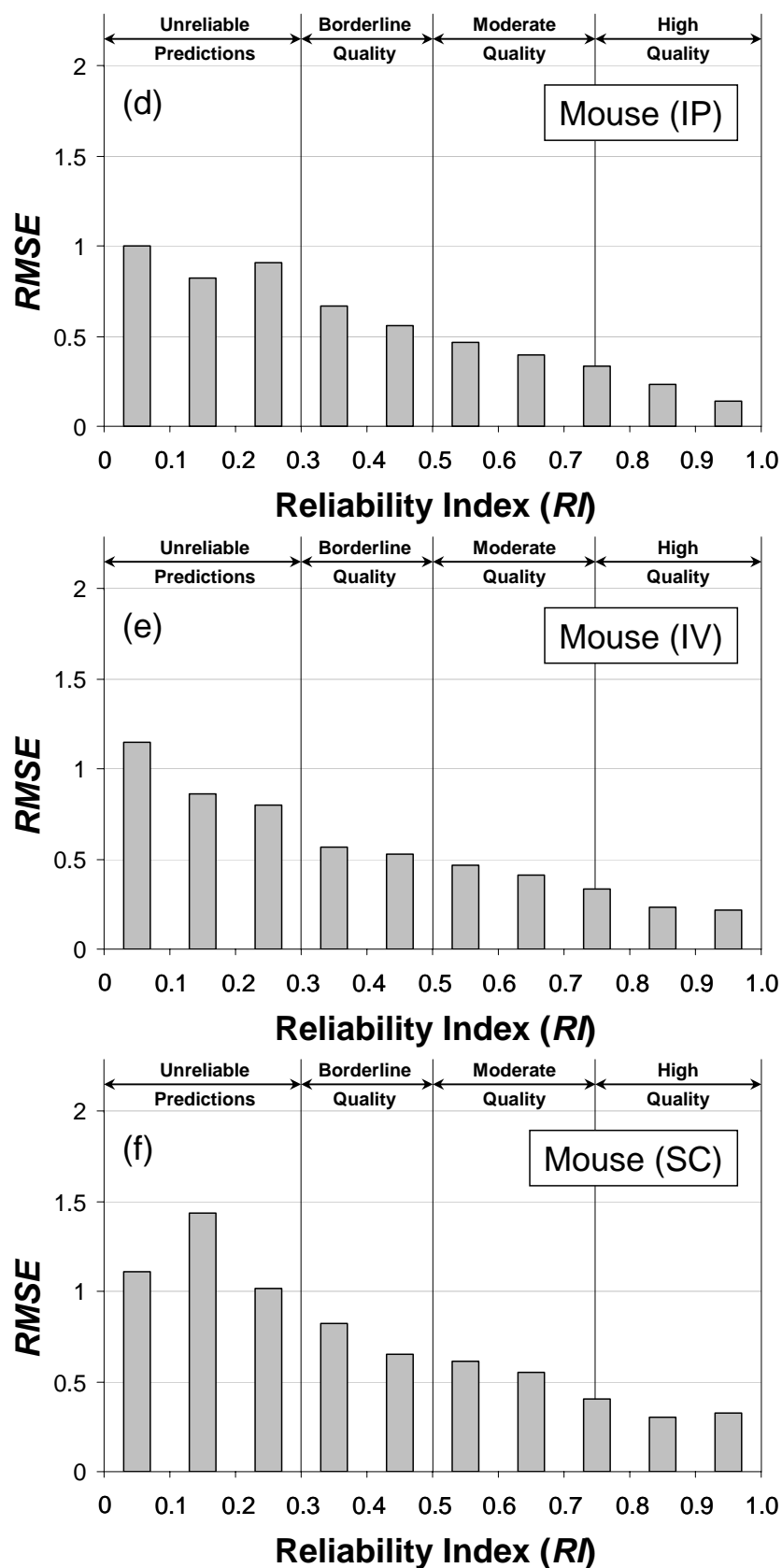
Figure 2 (continued). RMSE *vs.* RI dependences obtained for the presented acute toxicity models

are more or less uniform, providing a good basis for subdividing compounds into reliability categories. The significance of such classification can be seen from Figure 3

that compares observed *vs*. predicted values with $RI > 0.3$ and $RI > 0.75$ (using test set of mouse under IP administration, $N_{RI>0.3} = 8,568$ and $N_{RI>0.75} = 2,537$). While the first plot ($RI > 0.3$) resembles a "distorted cloud" with nearly 5% deviations exceeding 1 log unit, the second plot ($RI > 0.75$) is more like a "typical QSPR" (as if LD_{50} was replaced with $\log P$) with R^2 and *RMSE* values approaching 0.74 and 0.26 respectively, and only ~0.5% deviations exceeding 1 log unit. (Note however, that even under $RI > 0.75$ such deviations still occur, meaning that care should be taken when using it in practice).
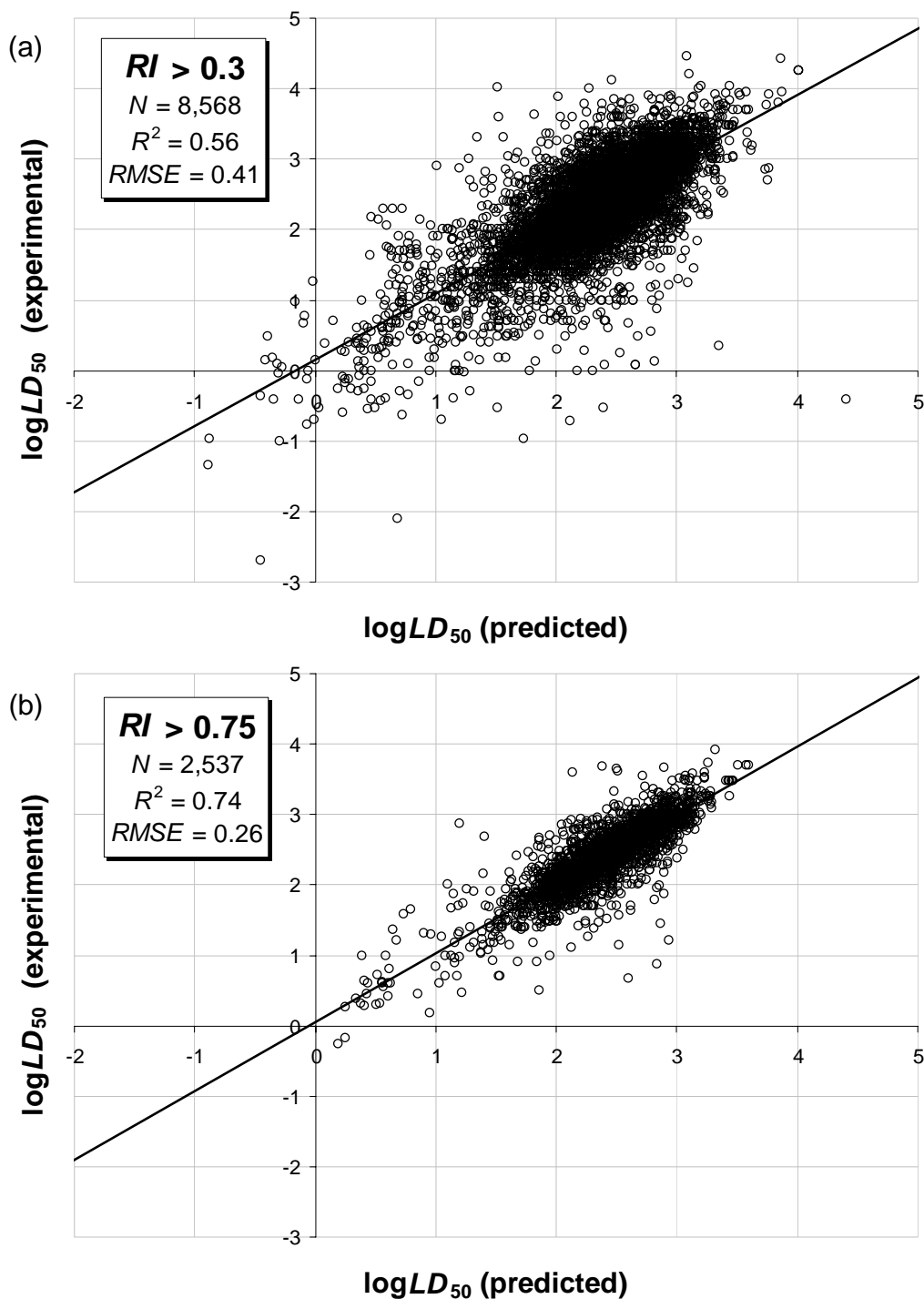

Figure 3. Experimental *vs*. calculated $\log LD_{50}$ for one of internal test sets (mouse, IP) under two *RI* cutoffs

4.3   External model validation

A 7,385 compound dataset with LD_{50} (rat, OR) values received from the authors of a recent publication [14] provided the possibility to subject the corresponding model presented in this work to an external validation procedure. Statistical results obtained for the fraction of this set, containing 2,718 compounds not present in the training set of LD_{50} (rat, OR) model presented here, are provided in the Table 4. Mean Absolute Error (*MAE*) values, being a primary measure of average prediction error in the original study involving this dataset [14], have been also determined in this case and are included in Table 4 along with *RMSE* for the result comparison purposes.

Table 4. Statistical results obtained following the application of the LD_{50} (rat, OR) model on the external validation set of 2,718 compounds

RI range	*N*	R^2	*RMSE*	*MAE*
$RI > 0.3$	2,501 (92%)	0.63	0.60	0.44
$RI > 0.5$	1,804 (66%)	0.70	0.55	0.40
$RI > 0.75$	430 (16%)	0.81	0.44	0.30

Notably, the R^2 and *RMSE* values for the compounds within model AD ($RI > 0.3$) are almost identical to the analogous characteristics of the "Final" model for the acute oral rat toxicity in case of internal test set, reported in Table 3. In their paper [14], H. Zhu and his co-workers attempt to compare five different types of acute oral rat toxicity models to a popular TOPKAT® program, using a 3,913 out of 7,385 compounds that were not used in the development of TOPKAT® predictive algorithms. The best results are reported for the consensus predictor, merging all five individual LD_{50} models, namely R^2 and *MAE* values of correspondingly 0.42 and 0.52 (at 74% coverage), 0.48 and 0.51 (66% coverage), and finally 0.71 and 0.39 (19% coverage). Here, the term "coverage" corresponds to the fraction of the external validation set compounds considered, depending on the model AD definition rules. In this respect it is analogous to the percentages of compounds falling within certain *RI* ranges for the models presented in this work. Obviously this cannot be called a direct comparison, since the size of the external validation set used is different, as is the absolute number of compounds within its fractions considered and their chemical structures, hence no categorical and far reaching conclusions as to which model is better should be made. However, the results from Table 4 can be treated as an indication that the acute oral rat toxicity model presented here is of similar accuracy to the reported consensus LD_{50} (rat, OR) predictor [14]. The latter, in its own turn, has been found consistently superior to the TOPKAT® model, yielding the following R^2 and *MAE* results: 0.35 and 0.59 (74% coverage), 0.25 and 0.70 (66% coverage), 0.54 and 0.52 (19% coverage) [14].

4.4   Reliability classification

The good quality of *RMSE vs. RI* dependencies from Figure 2 suggests the possibility of automated compound screening and/or prioritization before *in vitro* and even *in vivo* animal acute toxicity studies (depending on the stage of drug development).

E.g., compounds with predicted low LD_{50} (presumably toxic) and high *RI* (good reliability) may be discarded from consideration without any *in vitro* testing at the earliest convenience, whereas compounds with high LD_{50} (presumably non-toxic) and low *RI* (questionable prediction) should be tested in the first place. As illustrated in Figure 2, three *RI* cut-offs have been set forth (0.3, 0.5, and 0.75) to subdivide all compounds into four categories (unreliable, borderline, moderate, and reliable). Table 5 shows compound and *RMSE* distributions in all such categories for one of internal test sets (mouse, intraperitoneal administration, $N = 9,025$).

Table 5. Compound distribution with respect to reliability classes assigned according to the prediction Reliability Index. Here, Set 1 – PubChem database ($N = 24,857$) [20], Set 2 – World Drug Index ($N = 4,246$) [21]

Reliability classification	LD_{50} Test Set (mouse, IP)			Set 1	Set 2
	RI	*N*	*RMSE*	*N*	*N*
Not reliable	$0.00 \div 0.30$	459 (5%)	0.91	4,658 (19%)	960 (23%)
Borderline	$0.30 \div 0.50$	1,482 (16%)	0.60	13,144 (53%)	1,482 (35%)
Moderate	$0.50 \div 0.75$	4,549 (50%)	0.41	6,854 (28%)	1,515 (36%)
High	$0.75 \div 1.00$	2,537 (28%)	0.26	201 (1%)	289 (7%)

As expected, *RMSE* gradually decreases as *RI* increases, whereas compound distribution reaches maximum at $RI = 0.5 \div 0.75$. The latter corresponds to $RMSE < 0.5$ that is good enough for the crude estimation of LD_{50} without any *in vitro* testing (in the early stages of drug development). Since $RI > 0.5$ represents 75% of entire data set, only 25% of compounds would require *in vitro* testing (which may be further reduced by setting forth LD_{50} cutoff). Obviously one should use different criteria depending on the goals, *e.g.*, during later stages of drug optimization different *RI* (and LD_{50}) cut-offs may be adopted.

4.5 *RI* distribution in various datasets

The last two columns in Table 5 show compound distribution in *RI* classes for two independent compound sets. Set 1 comes from the PubChem database [20] ($N = 24,857$), whereas Set 2 is a sample ($N = 4,246$) from World Drug Index [21]. The construction of Set 1 was performed by cleaning a part of the complete PubChem collection with compound IDs ranging from 42575001 to 42600000 of any inorganic molecules or mixtures as well as converting any compounds represented in the ionized form to neutral species. Since both sets are not related to LD_{50}, they may provide a more realistic estimate of *RI* performance in "real-life" computations in the datasets of various structural complexities. One can see that $RI > 0.5$ corresponds to *ca.* 35% of all compounds (average of both sets), meaning that in the early stages of drug design up to one third of valid lead candidates could avoid respective *in vitro* testing (recall that *in vitro* tests have limitations, whereas animal tests are only conducted during pre-clinical stage, when substantial investment has already been made and neither developers nor

investors are eager to face any acute effects). On the other hand, $RI > 0.75$ corresponds to as low as 1% of compounds (PubChem), meaning that in pre-clinical stage virtually all compounds should be tested (though it often means only one or two candidates). This is no surprise, keeping in mind the intrinsic complexity of LD_{50} and the fact that similar applications of acute toxicity predictions were not even considered possible so far.

5   CYP3A4 and CYP2D6 regioselectivity modeling results

5.1   Comparison of the baseline and locally corrected GALAS model regioselectivity predictions

As it was mentioned in the Chapter 1.2 the internal validation of the resulting regioselectivity models has been performed using test sets constituting 30% of the initial subsets for each of the four considered metabolic reactions mediated by CYP3A4 and CYP2D6 enzymes. Table 6 below presents detailed results obtained for such test sets in case of the CYP3A4 N-dealkylation and CYP2D6 O-dealkylation which can be considered as the most typical reactions for the corresponding enzymes. Although the predictions obtained from the baseline models can already be considered as satisfactory, the presented results also indicate certain problems associated with them, *e.g.*, a lack of specificity. Indeed, considerable number of atoms is classified as 'False Positives' (14.2% and 6.6% by the corresponding models). Given the number of 'True Positive' predictions this results in a positive prediction accuracy of only around 50%. In other words, for each atom classified as the positive metabolism site by the baseline model there is roughly a 50% possibility that the prediction is wrong – clearly a significant problem. The situation noticeably improves after the application of local adjustments using similarity and the problem becomes virtually non-existent when considering only moderate or above reliability predictions ($RI > 0.5$), which still constitute *ca.* 50% of the test set. Another important observation relates to the amounts of inconclusive predictions (p in the range of $0.4 \div 0.6$). In case of both models these numbers decrease significantly for predictions of moderate or high reliability ($RI > 0.5$ and $RI > 0.7$). These results prove that Reliability Indices allow recognizing most predictions in the intermediate range as unreliable.

Table 6. Regioselectivity prediction results for the test sets of the CYP3A4 N-dealkylation and CYP2D6 O-dealkylation reactions

Considered model and dataset		Calculated probability (*p*)					
		CYP3A4 N-dealkylation			CYP2D6 O-dealkylation		
		<0.4	0.4÷0.6	>0.6	<0.4	0.4÷0.6	>0.6
Baseline predictions Test Set[*]	Negative	**116**	13	30	**128**	11	11
	Positive	10	7	**35**	1	1	**15**
GALAS model Test Set[*]	Negative	**136**	8	15	**141**	7	2
	Positive	8	5	**39**	2	1	**14**

Table 6 (continued). Regioselectivity prediction results for the test sets of the CYP3A4 N-dealkylation and CYP2D6 O-dealkylation reactions

Considered model and dataset		Calculated probability (*p*)					
		CYP3A4 N-dealkylation			**CYP2D6 O-dealkylation**		
		<0.4	0.4÷0.6	>0.6	<0.4	0.4÷0.6	>0.6
GALAS model *RI* > 0.5	Negative	**77**	2	2	**128**	3	1
	Positive	2	2	**27**	2	0	**11**
GALAS model *RI* > 0.7	Negative	**30**	0	0	**96**	0	0
	Positive	1	1	**10**	0	0	**7**

[*] Unreliable predictions (*RI* < 0.3) were not considered in this case which led to the exclusion of 37 marked atoms (10 positive metabolism sites) from the initial test set in case of CYP3A4 N-dealkylation and 10 marked atoms (2 positive sites) in case of the CYP2D6 O-dealkylation.

5.2 Regioselectivity prediction results for the rest of considered metabolic reactions

Regioselectivity of any enzyme (including CYP3A4 and CYP2D6) is related to individual atoms and their environment in the molecule which in its own turn is subject to far greater variability than the chemical constitution of the whole molecule. In other words, the similarity of two molecules possessing identical functional groups arranged differently in the molecule, in most cases will be considerably lower in terms of particular atom related property than it would be in terms of, let's say, log *P*. As a result, this property is expected to be more susceptible to the issues regarding the applicability domain of the predictive models compared to the whole-molecule properties, *i.e.*, the AD of metabolism site models depends more tightly on the training set. That is why it is more unlikely for such a model to perform satisfactory outside of it. In the light of these assumptions, Reliability Index estimation becomes particularly relevant and it is proposed that only predictions of at least moderate reliability should be taken into account. With such considerations in mind, the presentation of the statistical results for the remaining CYP3A4 and CYP2D6 reaction types analyzed has been limited to a condensed summary of moderate or above reliability (*RI* > 0.5) prediction results presented in Table 7. It can be just added that the trends of model results improvement following the application of the similarity based local adjustments for the models presented in Table 7 are the same as for the CYP3A4 N-dealkylation and CYP2D6 O-dealkylation examples that have been previously analyzed in full detail, *i.e.*, the problems with both the number of 'False Positive' and inconclusive predictions are gradually eliminated.

Table 7. Moderate and above reliability predictions (*RI* > 0.5) for the test sets of the remaining considered reactions

Reaction		Calculated probability (*p*)		
		<0.4	0.4÷0.6	>0.6
CYP3A4 aliphatic hydroxylation	Negative	**347**	3	2
	Positive	5	0	**17**
CYP3A4 aromatic hydroxylation	Negative	**394**	8	2
	Positive	4	1	**6**
CYP3A4 O-dealkylation	Negative	**108**	2	0
	Positive	1	0	**4**
CYP2D6 hydroxylation[*]	Negative	**825**	10	2
	Positive	3	1	**6**
CYP2D6 N-dealkylation	Negative	**177**	0	1
	Positive	6	1	**4**

[*] When analyzing experimental data for CYP2D6 aliphatic and aromatic hydroxylation, an assumption was made that requirements for being a CYP2D6 hydroxylation site depend more on binding to enzyme and are less dependent on properties of exact atom, therefore the data were merged and one general model for CYP2D6 hydroxylation was built.

While looking at the results in Table 7, a very small number of identified positive metabolism sites ('True Positives') can hardly go unnoticed in most cases. This observation can be attributed to the fact that for some reaction types the number of experimentally identified metabolism sites is generally very low in the starting datasets (see Table 2). In the field of cytochrome P450 metabolism the scientific literature mostly focuses on major reactions typical for each enzyme (*e.g.*, N-dealkylation for CYP3A4) consequently these reactions are represented best in the compiled training sets. Noticeably lower amounts of experimental data available on identified metabolism sites for less typical (secondary) enzyme reactions result in significantly smaller applicability domains of corresponding models. All this result in a general drawback of the whole presented ensemble of regioselectivity models based on publicly available data – they are much less likely to identify minor metabolites with high reliability. This is clearly illustrated by the total lack of high reliability positive metabolism site predictions (*p* > 0.6, *RI* > 0.7) for some reactions that cannot be called the most typical for that particular enzyme, *e.g.* CYP3A4 aromatic hydroxylation or O-dealkylation. Application of such models, even with the ability to evaluate their applicability domains, would be limited for scientists, focused on characterization of all possible metabolic pathways. However, with the ability to add user-defined data to the model and expand its AD to the desired regions of chemical space provided by the GALAS methodology, presented models have the potential for practical application even in their current state.

5.3   CYP3A4 and CYP2D6 regioselectivity model training

The possibility to train presented GALAS regioselectivity models was investigated using two example scenarios. In each case special regioselectivity models were built after all the atoms belonging to the compounds of a certain chemical class have been excluded from the training sets. Four random compounds were selected from the excluded class. Three are intended for a one-by-one stepwise addition to the local similarity correction routine of the model, whereas the last one is reserved for the testing of resulting four models. The first model has absolutely nothing similar in its training set used both to train the baseline model and to calculate local similarity corrections. All of the subsequent three models use the same baseline predictions derived from the same training set, however each one of them is increasingly aware of the baseline model performance for the compounds of the selected class and can compensate for any errors if necessary via local similarity corrections.

In case of CYP3A4 regioselectivity the initial example models have been derived from the training sets cleaned from atoms belonging to benzodiazepine class compounds or similar. Among the compounds containing excluded atoms, a group of molecules possessing azole ring and two experimental aliphatic hydroxylation sites was chosen [22-25]. Three randomly selected molecules from this group (brotizolam, triazolam, and alprazolam) were reserved for model training while one (a standard CYP3A4 substrate midazolam) was used for testing. Analogous preparation and testing steps in case of the CYP2D6 metabolism regioselectivity models involved removal of any atoms belonging to the propranolol analogues. Out of these the propranolol itself has been chosen to test this initial model as well as three subsequent ones trained with the data for propranolol analogues – all possessing three experimentally determined metabolism sites (one for N-dealkylation and two for aromatic hydroxylation reactions) [26]. For the structures of all the compounds mentioned, see Figure 4. Figures 5 and 6 illustrate the dynamics of the
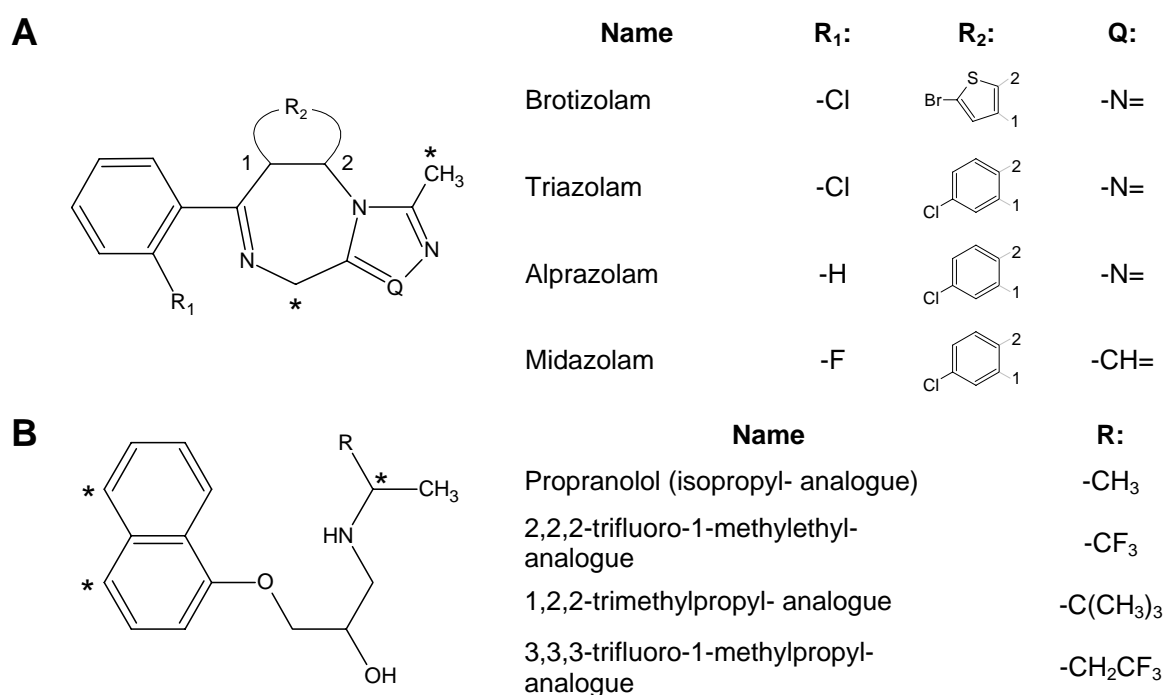
Figure 4. Compounds used for the trainability testing of the (A) CYP3A4 and (B) CYP2D6 regioselectivity models (asterisks note metabolism sites)

predictions for the test compounds (midazolam and propranolol) during the course of this virtual experiment. None of experimental metabolism sites are found within query compounds using the initial example models. A few potentially metabolized atoms ($p > 0.5$ – indicated in larger font) are found in the aromatic system of a propranolol; however the reliability of these results and all the predictions in general is very low. Adding 3 selected analogues one-by-one to the similarity correction part of the model gradually improves the predictions. Following the addition of a second similar compound all experimentally confirmed metabolism sites are correctly identified in both cases and the reliability of most predictions reaches moderate levels ($RI > 0.5$ – indicated in darker ink). Finally, the last analogue added further improves the predictions – the probabilities to be metabolized for almost all atoms are predicted with very high reliability ($RI > 0.7$).
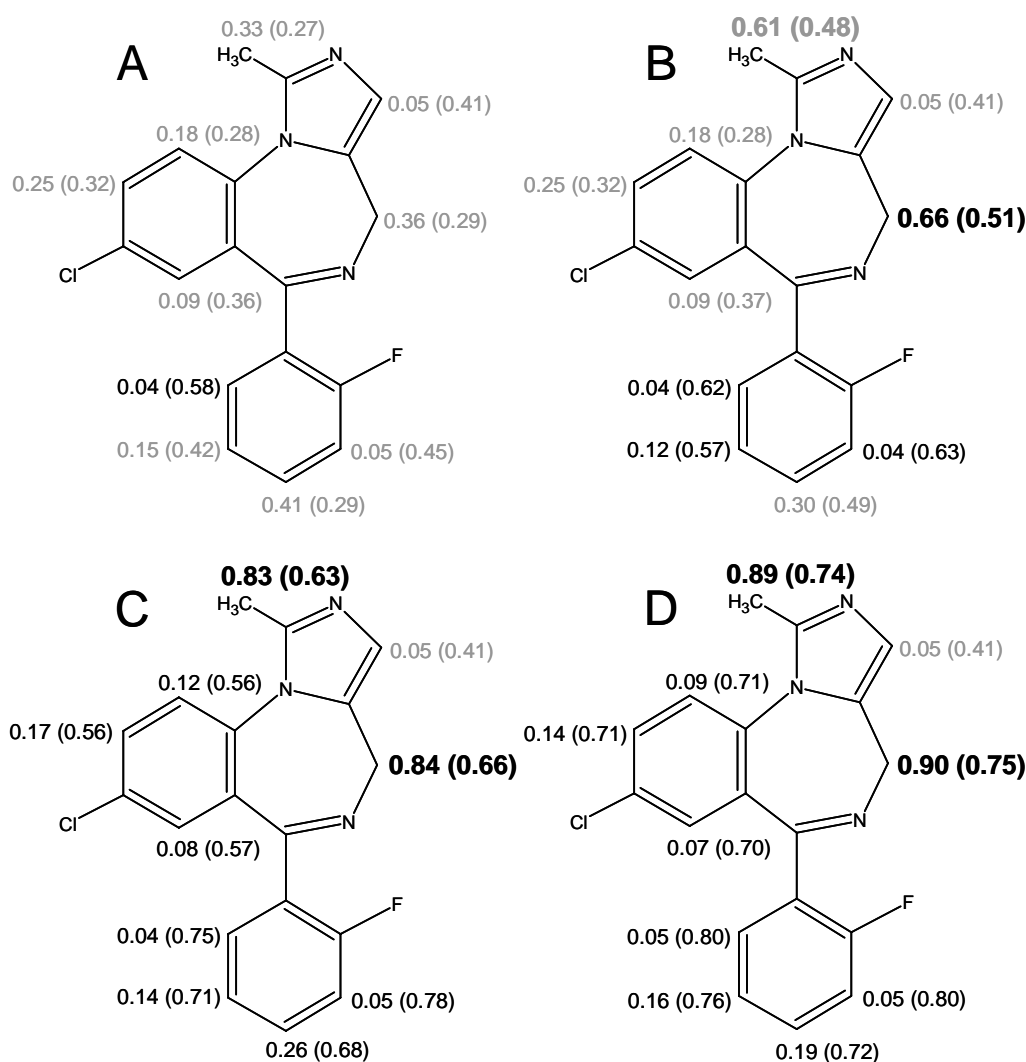

Figure 5. GALAS model predictions of CYP3A4 regioselectivity towards midazolam using initial model (A), after the addition of experimental data for brotizolam (B), brotizolam and triazolam (C), and all three benzodiazepines (D) to the local similarity assessment routine of the model
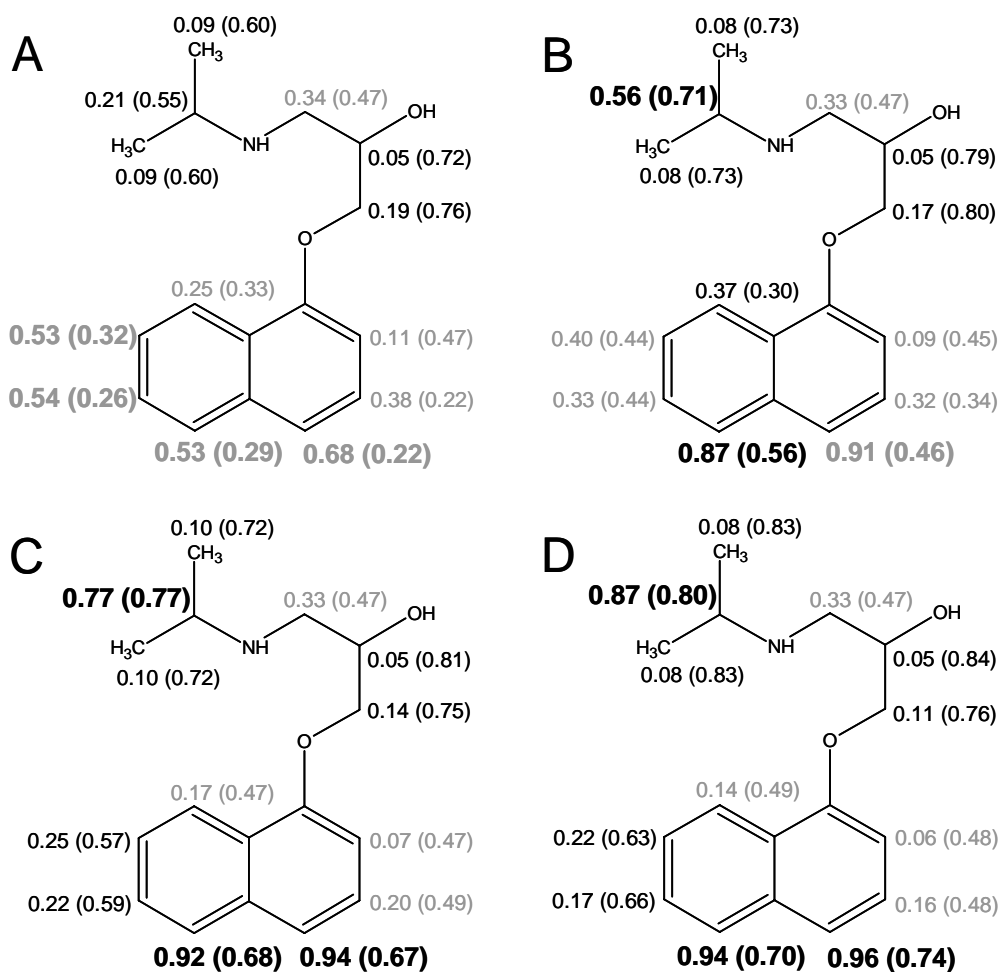
Figure 6. GALAS model predictions of CYP2D6 regioselectivity towards propranolol using initial model (A), after the addition of experimental data for one (B), two (C), and all three structural analogues (D) to the local similarity assessment routine of the model

This example clearly demonstrates the potential for practical applications of such trainable cytochrome P450 regioselectivity models, especially given the fact that the described improvements in predictions following the addition of similar compounds was instant and required no rebuilding of the baseline models.

CONCLUSIONS

1. The statistical models have been created for the estimation of the acute toxicity towards two rodent species following a number of different administration routes as well as for the prediction of CYP3A4 and CYP2D6 regioselectivity in the main metabolic reactions mediated by these enzymes (13 individual models in total). In every case the employed GALAS methodology has been shown to be superior compared to the corresponding linear baseline PLS or BPLS model. This fact confirms the significant influence of various local non-linear effects in case of the considered properties.

2. Results of the acute oral rat toxicity model external validation study suggest that purely accuracy-wise, *i.e.*, disregarding all the additional benefits resulting from the GALAS modeling methodology application, the results obtained are comparable to the recent achievements of other authors in the field and notably are superior to TOPKAT® – one the most popular acute toxicity estimation methods often considered to be the benchmark of toxicology related predictions.

3. A uniform trend of significant reduction in prediction errors or the number of mispredictions with the increase of calculated Reliability Index of the prediction was observed for each of the analyzed models. This fact clearly indicates that the ability of the GALAS methodology to quantitatively assess the applicability domain of the model has been retained despite the significant increase in the complexity of the considered properties. As a result all of the presented models conform to one of the main requirements for the QSAR model acceptance by the EU and other regulatory institutions.

4. Evident correlation between prediction reliability and its accuracy allowed the classification of each model result into one of the several qualitative classes according to the calculated *RI* values. Such information, providing a clue regarding the error that model is most likely to produce in each case, enables compound prioritization before experimental testing and, depending on the stage and objectives of the development, can help achieving a bigger or lesser reduction in the number of necessary measurements.

5. The ability of the obtained GALAS regioselectivity models to adapt themselves to the new experimental data has been demonstrated. This feature allows expanding the applicability domains of the models to any chemical space region being of interest to the research and opens wide possibilities for their practical application.

6. The results obtained in the enzyme regioselectivity modeling show promising perspectives for the utilization of the GALAS modeling technique in the analysis of regioselectivity for other important biotransformation enzymes (*e.g.*, CYP2C9, CYP2C19, CYP1A2, *etc.*).

7. GALAS modeling methodology forming the basis of all the discussed models provides additional opportunities for somewhat unconventional ways of their application, *e.g.* the outlier based analysis indicating interesting compound classes for further investigations potentially leading to the generation of new mechanistic knowledge, automatic identification of experimental errors and other.

LIST OF PUBLICATIONS

Articles in journals:

1. Sazonovas, A., Japertas, P., and Didziapetris, R. *Estimation of reliability of predictions and model applicability domain evaluation in the analysis of acute toxicity (LD_{50})*. 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.

Conference abstracts and materials:

Oral presentations at the international conferences:

1. Japertas, P., Sazonovas, A., Didziapetris, R., Petrauskas, A. *COMP-271 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.
   (The 235th ACS National Meeting, New Orleans, LA, April 6-10, 2008)

Selected poster presentations at the international conferences:

1. Japertas, P., Didziapetris, R., Sazonovas, A., Petrauskas, A. *T-22 Acute toxicity (LD_{50}) modeling utilizing fragmental QSAR, similarity analysis and reliability of prediction*. Toxicol. Lett., 2007, 172 (Supplement 1), S80-S81.
   (44th Congress of the European Societies of Toxicology (EUROTOX 2007), Amsterdam, the Netherlands, October 7-10, 2007)

2. Dapkunas, J., Sazonovas, A., Japertas, P. *MEDI-246 Probabilistic prediction of the human CYP3A4 metabolism sites in a molecule*. Abstr. Paper. Am. Chem. Soc., 2008, 235, Meeting abstract 246-MEDI.
   (The 235th ACS National Meeting, New Orleans, LA, April 6-10, 2008)

3. Japertas, P., Didziapetris, R., Sazonovas, A., Petrauskas, A. *MEDI-247 Similarity based correction for the predictions of compounds physicochemical properties*. Abstr. Paper. Am. Chem. Soc., 2008, 235, Meeting abstract 247-MEDI.
   (The 235th ACS National Meeting, New Orleans, LA, April 6-10, 2008)

4. Dapkunas, J., Sazonovas, A., Japertas, P. *MEDI-438 Probabilistic prediction of the human CYP3A4 and CYP2D6 metabolism sites in a molecule*. Abstr. Paper. Am. Chem. Soc., 2008, 236, Meeting abstract 438-MEDI.
   (The 236th ACS National Meeting, Philadelphia, PA, August 17-21, 2008)

REFERENCES

1. 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.

2. 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.

3. 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.

4. Williams, P. A., Cosme, J., Vinkovic, D. M., Ward, A., Angove, H. C., Day, P. J., Vonrhein, C., Tickle, I. J., and Jhoti, H., *Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone.* Science, 2004, 305, 683-686.

5. 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.

6. 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.

7. 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.

8. 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.

9. 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.

10. 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

11. 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.

12. *The Registry of Toxic Effects of Chemical Substances (RTECS®)*, Symyx Technologies, Inc., 415 Oakmead Parkway, Sunnyvale, CA, 94085, US. www.symyx.com

13. *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 Comission, Ispra (VA), Italy. http://ecb.jrc.ec.europa.eu/esis/index.php?PGM=dat

14. 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.

15. Testa, B., *The Metabolism of Drugs and Other Xenobiotics: Biochemistry of Redox Reactions*. Academic Press, San Diego, 1995.

16. 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.

17. Lipnick, R. L., *Outliers: their origin and use in the classification of molecular mechanisms of toxicity.* Sci. Total Environ., 1991, 109-110, 131-153.

18. 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.

19. Efron, B., *Bootstrap methods: another look at the jackknife*, in *Ann. Statist.* 1979. p. 1-26.

20. *NCBI PubChem database*, National Center for Biotechnology Information of the National Library of Medicine, Bethesda (MD), United States. http://pubchem.ncbi.nlm.nih.gov/

21. *Derwent World Drug Index sample CD*, Derwent Information, London, United Kingdom, www.derwent.com, 2001.

22. 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.

23. 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.

24. Kronbach, T., Mathys, D., Umeno, M., Gonzalez, F. J., and Meyer, U. A., *Oxidation of midazolam and triazolam by human liver cytochrome P450IIIA4.* Mol. Pharmacol., 1989, 36, 89-96.

25. 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.

26. Upthagrove, 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.

CURRICULUM VITAE

Name, Surname:	**Andrius Sazonovas**
Date and place of birth:	8th of March, 1980, Vilnius
Contact details:	
(address)	Vš.Į. Aukštieji Algoritmai
	A. Mickevičiaus g. 29, LT-08117
	Vilnius, Lithuania
	Faculty of Chemistry, Vilnius University
	Naugarduko g. 24, LT-03225
	Vilnius, Lithuania
(phone)	+370 5 262 40 32, +370 5 262 34 08
(fax)	+370 5 262 37 28
(e-mail)	andrius@pharma-algorithms.com

Education:

2004-2009	Vilnius University, Faculty of Chemistry, PhD studies, supervisor prof. dr. habil. E. Butkus.
2002-2004	Vilnius University, Faculty of Chemistry, MSc studies (*Cum laude* diploma). Title of the graduation thesis for master's degree "Prediction of aqueous solubility of organic substances: a problem and its solution"
1998-2002	Vilnius University, Faculty of Chemistry, BSc studies. Title of the graduation thesis for bachelor's degree "The search for optimal prediction algorithms of molecule's polarity/polarizability parameter"
1995-1998	Vilnius Lyceum of exact, nature and technical sciences – higher education
1986-1995	57th (currently Taikos elementary school) and S. Stanevičiaus secondary schools – elementary education

Employment:

2005-present	Vš.Į. Aukštieji Algoritmai, ADME/Tox group project manager.
2001-2004	Main areas of interest include development and applications of various fragmental methods in the QSAR analysis of the main physicochemical and ADME/Tox properties of organic compounds.

Other:

Regular participant and prize winner at the Lithuanian national chemistry olympiads: 2nd, 3rd, and 1st places in 1996, 1997, and 1998, respectively (laureate diploma in 1998). Participant (1997) and 1st prize winner (1998) at the chemistry olympiad of the Baltic states. Bronze medal winner at the 30th International Chemistry Olympiad in 1998.

Grant fellow of the EU 5th Framework project F.A.M.O.U.S. (Fullerene based advanced materials for optoelectronic utilizations – HPRN-CT-2002-00171) during the years 2004 and 2005, including visits to Warsaw and Bologna universities. Main research topic within the project – Monte Carlo simulations of liquid crystalline and other ordered phases within systems of mesogenic fullerene derivatives.

REZIUMĖ

Vienas pagrindinių naujų vaistinių medžiagų paieškos iššūkių yra priimtinų galimo kandidato farmakokinetinių savybių užtikrinimas, didele dalimi nulemiantis būsimojo vaisto efektyvumą. Tarp nagrinėjamų kinetinių veiksnių, nulemiančių junginio pasiskirstymą organizme labai svarbų vaidmenį atlieka metabolinis junginio stabilumas, apsprendžiantis ar vaisto molekulės apskritai turi galimybių pasiekti savo taikinį organizme chemiškai nepakitusios. Savo ruožtu pastaroji savybė daugiausia priklauso nuo to, ar junginys bus metabolizuojamas CYP450 fermentų šeimos ir ypač dviejų pagrindinių jos izoformų – CYP3A4 bei CYP2D6, kartu nulemiančių daugiau kaip 50% šiuo metu rinkoje esančių vaistų metabolizmą. Pastaroji aplinkybė potencialiai gali sukelti visą eilę problemų, įskaitant nesusidarančią ar labai trumpam laikui susidarančią efektyvią junginio koncentraciją ties taikiniu, įvairias vaisto sąveikas su kitais kartu vartojamais tos pačios CYP450 izoformos substratais arba inhibitoriais ir pan. Paradoksalu, kad šie fermentai, visų pirma skirti organizmo apsaugai nuo svetimų cheminių medžiagų poveikio, neretai dalyvauja daugelį kartų už pradinius substratus pavojingesnių metabolitų susidaryme, sukeldami ūmaus ar chroniško apsinuodijimo simptomus, kancerogeniškumo pasireiškimą ir netgi mirtį. Dėl šių priežasčių detalus visų stebimų junginio metabolitų identifikavimas bei jų toksiškumo įvertinimas yra praktiškai neatsiejami ir dažniausiai lygiagrečiai atliekami bet kokio vaistinio junginio kandidato tyrimo žingsniai.

Tiek junginio metabolizmas, tiek toksiškumas patenka į sąrašą esminių savybių, galiausiai nulemiančių, ar junginys sėkmingai įveiks visas vaisto kūrimo, bandymo bei licenzijavimo pakopas. Tuo būdu būtent tokių savybių bei biologinių aktyvumų efektyvus prognozavimas įvairiais kompiuteriniais *in silico* metodais įgalintų perspektyvių molekulių pasirinkimą jau pačiuose pirmuosiuose tyrimo žingsniuose, įskaitant virtualių bibliotekų analizę. Tai savo ruožtu sudarytų bene pačias didžiausias prielaidas optimizuoti naujų vaistų kūrimo procesą, siekiant pastarąjį sutrumpinti bei atpiginti. Galimybė iš anksto numatyti bent jau pagrindinius junginio metabolitus suteikia informaciją apie jo struktūroje esančius lengviausiai metabolizuojamus centrus. Pastarųjų blokavimas arba tiesiog pašalinimas leistų stipriai sulėtinti ir netgi visiškai sustabdyti CYP450 katalizuojamas metabolizmo reakcijas. 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. Selektyvių 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. Taip pat netgi preliminarūs CYP450 fermentų regioselektyvumo modeliai galėtų pagelbėti analizuojant eksperimentinio molekulės metabolitų nustatymo metu gaunamas sudėtingas chromatogramas bei spektrus. Tuo tarpu toksiškumo prognozavimo modeliai leistų iš anksto įvertinti tiek paties pradinio junginio, tiek jo metabolitų keliamą pavojų šiuo požiūriu.

Šiame darbe pristatomi GALAS metodika paremti ūmaus toksiškumo modeliai dviems graužikų rūšims bei visai eilei skirtingų medžiagos patekimo į organizmą būdų. Analogiškas metodas, keleto modifikacijų pagalba pritaikytas kokybinių atomo-centrinių

savybių prognozavimui, buvo panaudotas modeliuojant CYP3A4 bei CYP2D6 fermentų regioselektyvumą pagrindinėse jų katalizuojamose metabolinėse reakcijose. Iš viso darbe pristatomi 13 individualių modelių.

Visi darbe nagrinėjami modeliai kokybiškai išsiskiria iš anksčiau publikuotų analogų dėl gebėjimo kartu su prognozuojama savybės verte pateikti prognozės patikimumo indekso (*RI*) reikšmę. Pademonstruota akivaizdi prognozės paklaidų (ar klaidingų prognozių skaičiaus) priklausomybė nuo *RI* parodė, jog, nepaisant modeliuojamųjų objektų sudėtingumo, šis dydis leidžia efektyviai kiekybiškai įvertinti junginio priklausymą modelio pritaikomumo sričiai bei su tuo susijusią jo prognozės kokybę. Šis ypatumas nulemia gautų modelių atitikimą vienam pagrindinių ES institucijų reikalavimų, keliamų alternatyviems tyrimo metodams, kuris formuluojamas kaip galimybė įvertinti modelio pritaikomumo sritį. Be to, 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 bet kokius eksperimentinius matavimus bei priklausomai nuo vaisto kūrimo pakopos bei keliamų uždavinių daugiau ar mažiau sumažinti pastarųjų skaičių. Pavyzdžiui, išnagrinėtas ūmaus toksiškumo prognozių pasiskirstymas pagal apibrėžtas kokybines tikslumo klases kelioms išorinėms duomenų bazėms rodo, kad bent jau ankstyviausiuose tyrimų žingsniuose, vadovaujantis siūlomais kriterijais galima būtų išvengti iki trečdalio eksperimentų.

Disertacijoje taip pat pademonstruotas gautųjų GALAS modelių gebėjimas greitai bei efektyviai apsimokyti naujais eksperimentiniais duomenimis, išplečiant jų pritaikomumo sritį. Ši savybė suteikia parengtiems modeliams didžiulį praktinio panaudojimo farmacijos pramonėje potencialą, juolab, kad visi jie yra paremti viešai prieinamų duomenų rinkiniais, kaip taisyklė neapimančiais pramonėje nagrinėjamų junginių užimamos srities cheminės įvairovės erdvėje. Kartu su prognoze GALAS modelių pateikiama papildoma informacija (panašumo ir patikimumo indeksai, *Δ* korekcijos ir pan.) taip pat atveria kai kurias ne visai tradicines jų panaudojimo galimybes. Pavyzdžiui, iškrentančių prognozių analizė bei naujų mechanistiniu požiūriu įdomių junginių klasių paieška, eksperimentinių matavimų klaidų identifikavimas ir kt. Apibendrinant galima konstatuoti, kad darbe pristatomų modelių savybių tyrimas bei gauti rezultatai atveria perspektyvą analogišką modeliavimo strategiją taikyti kitų svarbių metabolizmo fermentų (pvz., CYP2C9, CYP2C19, CYP1A2 ir kitų) regioselektyvumo duomenų analizei, tęsiant šios temos nagrinėjimą ateityje.