Realia et naturalia

DISSERTATIONES
CHIMICAE
UNIVERSITATIS
TARTUENSIS
132

IVA BOZHIDAROVA STOYANOVA-SLAVOVA

Validation of QSAR/QSPR for regulatory purposes

DISSERTATIONES CHIMICAE UNIVERSITATIS TARTUENSIS

132

DISSERTATIONES CHIMICAE UNIVERSITATIS TARTUENSIS
132

IVA BOZHIDAROVA STOYANOVA-SLAVOVA

Validation of QSAR/QSPR for regulatory purposes

UNIVERSITY OF TARTU
PRESS
1632

Institute of Chemistry, Faculty of Science and Technology, University of Tartu, Estonia

Dissertation is accepted for the commencement of the Degree of Doctor of Philosophy in Chemistry on November 8th, 2013 by the Doctoral Committee of the Institute of Chemistry, University of Tartu.

Supervisor: Prof. Mati Karelson, Institute of Chemistry,
University of Tartu, Tartu, Estonia

Opponent: Prof. Maria Cristina Menziani,
Dipartimento di Scienze Chimiche e Geologiche,
University of Modena and Reggio Emilia,
Modena, Italy

Commencement: 20th of December 2013 in room 1021, 14A Ravila Street,
Tartu, at 10:00

ISSN 1406–0299
ISBN 978–9949–32–449–1 (print)
ISBN 978–9949–32–450–7 (pdf)

University of Tartu Press
www.tyk.ee

TABLE OF CONTENT

1. LIST OF ORIGINAL PUBLICATIONS

The presented thesis consists of six articles listed below. All papers are denoted in the text by roman numerals I–VI.

I. Katritzky, A. R., Stoyanova-Slavova, I. B., Dobchev, D. A., and Karelson, M. QSPR modelling of flash points: An update J. Mol. Graph. &. Model. 26 (2007) 529–536.

II. Katritzky, A. R., Stoyanova-Slavova, I. B., Tämm, K., Tamm, T., and Karelson, M. Application of the QSPR Approach to the Boiling Points of Azeotropes J. Phys. Chem. A 115 (2011) 3475–3479.

III. Katritzky, A. R., Kuanar, M., Stoyanova-Slavova, I. B., Slavov, S. H., Dobchev, D. A., Karelson, M., Acree, W.E. Jr., Quantitative Structure-Property Relationship Studies on Ostwald Solubility and Partition Coefficients of Organic Solutes in Ionic Liquids J. Chem. Eng. Data. 53 (2008) 1085–1092.

IV. Katritzky, A. R., Slavov, S., Stoyanova-Slavova, I., Kahn, I., Karelson, M. Quantitative Structure-Activity Relationship (QSAR) Modeling of EC_{50} of Aquatic Toxicities for Daphnia magna J. Toxicol. Environ. Sci. 72 (2009) 1181–1190.

V. Katritzky, A. R., Slavov, S. H., Stoyanova-Slavova, I. B., and Karelson, M. Correlation of the Photolysis Half-Lives of Polychlorinated Dibenzo-p-dioxins and Dibenzofurans with Molecular Structure J. Phys. Chem. A 114 (2010) 2684–2688.

VI. Katritzky, A. R., Slavov, S., Radzvilovits, M., Stoyanova-Slavova, I., Karelson, M. Computational Chemistry Approaches for Understanding how Structure Determines Properties Z. Für Naturforsch. B 64 (2009) 773–777.

Author's contribution

Publication I:	The author is responsible for the data sets, calculations, interpretation of the results and preparation of the manuscript.
Publication II:	The author is responsible for the data sets, calculations, interpretation of the results and preparation of the manuscript.
Publication III:	The author is responsible for the calculations, interpretation of the results and writing parts of the manuscript.
Publication IV:	The author is responsible for the calculations, interpretation of the results and writing parts of the manuscript.
Publication V:	The author is responsible for the calculations, interpretation of the results and writing parts of the manuscript.
Publication VI:	The author is responsible for the calculations, interpretation of the results and writing parts of the manuscript.

2. LIST OF ABBREVIATIONS

AM1	Austin Model 1
ANN	Artificial Neural Network
BMLR, (B)MLR	Best Multiple Linear Regression
CODESSA	COmprehensive DEscriptors for Stnictural and Statistical Analysis
LMO	Leave-Many-Out cross-validation
LOO	Leave-One-Out cross-validation
MLR	Multiple Linear Regression
MOPAC	Molecular Orbital PACkage
QSA(P)R	Quantitative Structure – Activity (Property) Relationship(s)
RMSPE	Root-mean Squared Error of Prediction
RMS	Root-mean Squared Error
ECHA	European Chemical Agency
EPA	Environmental Protection Agency
REACH	Registration, Evaluation, and Authorization of Chemicals
FDA	Food and Drug Administration

3. INTRODUCTION

With the ever-growing concerns about the detrimental effect of human activities on the environment, modern societies continuously look at various ways and means to reduce the negative impact of chemicals on human health and the ecosystem. One such way is through the development of legislative strategies establishing a framework for making regulatory decisions aimed at eliminating and/or minimizing the risks that can be posed by chemicals. Today, most developed countries have enacted laws and regulations setting registration and safety evaluation standards as well as have introduced policies for management of hazardous substances. For example, in EU, the European Chemical Agency (ECHA) has established an integrated system for Registration, Evaluation, and Authorization of Chemicals (REACH).

In a recently published annex[1], the REACH legislation suggested the use of QSAR as a tool to reduce or replace animal testing for the purpose of safety assessment of industrial chemicals. Such an official statement acknowledges the role of (Q)SAR as a highly relevant and reliable tool for addressing regulatory tasks, which indeed is routinely used by the U.S. Environmental Protection Agency (EPA) and the Food and Drug Administration (FDA) for these purposes. In Europe, the use of QSA(P)R for regulation and prioritization is strongly encouraged within the current EU chemical policies[2].

As evidenced by the high cost of obtaining even basic physical property and environmental fate data (see Table 1 reprinted from ref[3]) the use of QSA(P)R in regulation is not only desirable, but in most cases practically unavoidable. It has to be emphasized that a basic toxicological profile of a chemical is at least an order of magnitude more expensive (with a cost of up to several million dollars) and may take 4 to 5 years (in the case of carcinogenicity studies) to complete[4]. Having in mind that about 4000 new chemicals enter the market each year[5] the total cost of generating safety data would be in excess of 10 billion dollars per year.

Thus, it is not surprising that the adoption of QSA(P)R in regulation during the past decade has been accelerating. Projects, such as ORCHESTRA, CAESAR, DEMETRA, HAIR and RAINBOW among many others, facilitated the transition of QSAR from a tool used mainly by the pharmaceutical industry into a major driving force behind priority-setting and risk assessment process. Nowadays new chemicals are routinely being assessed and predictions are being made concerning their chemical identity, physical and chemical properties, environmental transport, partitioning, concentration, fate and toxicity. Naturally, to make predictions the regulatory bodies use a consensus of multitude of methods including QSA(P)Rs, nearest analogue analysis, chemical class analogy, mechanisms of toxicity, mode of action, chemical industry survey data, and expert judgment.

Table 1. Cost to obtain physical property and environmental fate data for a single compound/material.

Test	cost (in USD)
Boiling point	420 – 560
Melting point	430 – 565
Vapor pressure	1,360 – 2,520
Water volubility	2,870 – 14,080
Physical property data	5,080 – 17,725
Octanol – water partition coefficient	1,310 – 2,370
Soil sorption Coefficient	6,680 – 9,430
Volatilization	6,180 – 8,800
Partitioning data	14,170 – 20,600
Hydrolysis	5,470 – 7,790
Photolysis	11,880 – 16,910
Aerobic biodegradation	8,250 – 11,800
Anaerobic biodegradation	5,980 – 8,480
Degradability data	31,580 – 44,980
Bioconcentration data	15,580 – 41,690
Total	**66,410 – 124,995**

In addition, the application of QSA(P)R in regulation and the increased demand for cost-effective, relevant, transparent, biologically meaningful and reliable predictions contributed immensely to the improvement of QSA(P)R methodology and its further acceptance as an alternative to animal testing[6]. According to the strict standards imposed by the Chemicals Committee serving the needs of the OECD Member Countries (now known as OECD principles[7]), a QSA(P)R model can be perceived as credible only if all following conditions apply:

i) it models a well defined endpoint;
ii) it uses an unambiguous algorithm;
iii) it has a defined domain of applicability;
iv) it uses appropriate measures of goodness-of-fit, robustness and predictivity;
v) it provides a mechanistic interpretation, if possible.

According to the Technical Guidance Document (TGD) for the assessment of new notified and existing chemicals[8] and[9], QSA(P)R models meeting all quality standards listed above can then be used to:
i) assist data evaluation;
ii) contribute to the decision-making process on whether further testing is necessary to clarify an endpoint of concern;

iii) to supplement the use of experimental data in weight-of-evidence approaches (by filling gaps in the data, by providing mechanistic interpretation, or by supporting the evaluation of existing experimental data);
iv) establish input parameters necessary to conduct exposure assessment;
v) identify effects which may be of potential concern on which test data are not available;
vi) to substitute or if possible to completely replace the need for experimental data especially in cases of animal testing.

In conclusion, during the past decade QSA(P)R has passed a long way from its beginning as a theoretical tool supplementing molecular design to a technology that bridged together computational chemistry and regulatory science. The profound effect of QSA(P)R on regulation can be seen in many legislative documents related to priority settings and risk assessment issued by most of the developed nations. It is worth emphasizing that the opposite trend has also taken place: the need to serve regulatory functions affected the field of QSA(P)R by calling for higher standards in the model development and validation stages.

One such astounding example for the impact of regulation on QSA(P)R concerns the longstanding issue of the reproducibility and "transparency" of models. To address this issue two guideline documents were developed. The QSAR Model Reporting Format (QMRF) documents the details needed to describe the characteristics and validity of a model, while the QSAR Prediction Reporting Format (QPRF) outlines the information needed to justify the adequacy of the QSAR prediction[10]. It is now common, that the majority of the QSAR publications tend to include a statement of compliance of their work with the OECD principles[11, 12, 13, 14].

This PhD thesis provides an overview of the QSA(P)R methodology and outlines some specifics of the QSA(P)Rs used for regulatory purposes. It also summarizes work done on the QSA(P)R modeling of important for the regulatory agencies endpoints such as:

- aquatic toxicity – used for monitoring the health of aquatic ecosystems;
- logP partition coefficient – used to set water quality standards and to regulate ecotoxicants;
- boiling points – used to rank compounds that can be potential hazardous air pollutants;
- flash points – used to determine the fire and air pollution hazards of flammable and combustible materials;
- Ostwald solubilities – used to assess occupational exposures;
- photolysis half-lives – used for assessment of the environmental risk of persistent organic pollutants.

4. LITERATURE OVERVIEW

4.1. QSAR general methodology

The regulatory needs served by QSA(P)R are primarily related to its ability to establish a relationship between toxicological and environmental endpoints and the structure of compounds. The use of QSAR for other purposes (drug discovery, rational design of products such as surfactants, dyes, fragrances, etc.) is out of the scope of the regulatory agencies. As a side fact one of the first applications of QSAR was in toxicology. In his doctoral thesis dating back to 1863 Cros noted that there is a relationship between the toxicity of primary aliphatic alcohols and their water solubility[15]. A few years later his pioneering work was followed by Richardson[16], Richet[17], Meyer[18] and Overton[19] all reporting QSARs in the field of toxicology. A century later, with the introduction of the Linear Free Energy Relationship (LFER) approach[20], QSAR reached maturity and was put on a solid theoretical ground.

The modern QSA(P)R is a complex tool designed to model different types of endpoints (continuous and categorical) using a wide variety of molecular descriptors (indicator variables, scalar quantities (2D-QSAR) and fields (3D-QSAR)) and modeling techniques (linear and nonlinear). Recent trend in QSA(P)Rs used for regulation shows an increased demand for consensus models combining the predictive power of multiple individual approaches: e.g. the consensus between linear and non-linear QSA(P)Rs utilizing different descriptor types is a common practice.

Despite the variability of the endpoints, the descriptor types and the methods, virtually all QSA(P)R follow similar pathway: i) data collection and curation; ii) calculation of molecular descriptors; iii) model building and validation and iv) interpretation. Each of these steps is composed of multiple phases with their own specifics. A typical pipeline of a QSA(P)R modeling process is shown on Figure 1.

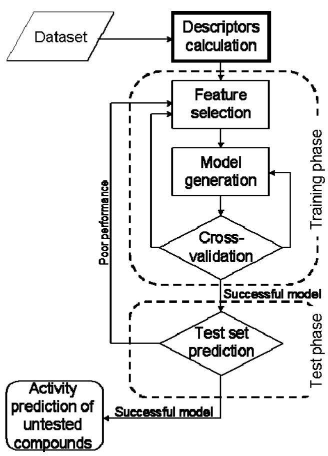

Figure 1. Typical pipeline of a QSA(P)R modeling process

4.2. Data collection and curation

Until recently, this integral part of the QSA(P)R modeling pipeline was mostly associated with a set of requirements generally targeting the size of the dataset, the experimental error and the type and range of the property values. It is commonly accepted that:

i) a dataset suitable for QSA(P)R modeling should consist of at least 25–30 compounds[21];
ii) the relative experimental error should be low (preferably less than 20%);
iii) the property values should cover at least one logarithmic unit[22]
iv) robust biological data such as K_i or IC_{50} should be used.

However, several authors[23, 24, 25] have pointed out the crucial importance of the quality of input data, and the danger of using chemical databases without careful manual curation: it was found that most public and commercial databases contain between 0.1 and 3.4% erroneous structures. It was concluded that the use of erroneous structures represented by erroneous descriptors would affect negatively the models' performance. To ensure an additional level of quality control (QC) on the input data, further steps such as: i) structural conversion and cleaning; ii) normalization of bonds and iii) removal of duplicates would be desirable[25].

Furthermore, QSA(P)Rs intended for regulation impose specific requirements in the initial phase of data collection. Being of a major concern for the regulatory agencies, different modeling priorities are given to the so-called high (chemicals that have been imported or produced in quantities exceeding 1,000 metric tons per year) and low-production-volume (10 to 1,000 metric tons per year) substances. Hence, QSA(P)Rs built for regulation often prioritize the chemicals to be included in the modeling process[26, 27, 28].

The regulatory agencies often categorize the chemicals according to the risk they pose to the environment, the ecosystem or human health. The regulatory QSA(P)R models often deal with the so called "substances of concern", which are likely to be toxic due to structural similarities to compounds already known as such. For example, datasets constructed to model the endocrine disrupting potential of chemicals often start with a core set of compounds that are structurally similar to the estradiol and tamoxifen and are extended if necessary.

4.3. Calculation of molecular descriptors

Ideally and naturally, a QSA(P)R would try to establish a mathematical relationship $\mathbf{E}=f(\mathbf{S})$ between the structures **S** and the endpoints **E** (physico-chemical properties or biological activities) of compounds, depicted by the dashed arrow in Figure 2. Being a real object, the molecule implicitly contains all chemical information – however, only a fraction of this information is accessible through experimental measurements. The development of quantum mechanics (QM) and the chemical graph theory enabled the possibility to capture and convert in a purely theoretical manner a significant portion of the remaining information encoded in the molecular structure (shown by the solid arrows in Figure 2).

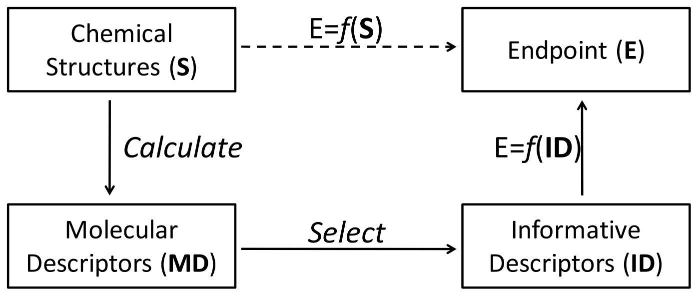

Figure 2. Molecular descriptors bridge chemical structure to biological activity in QSAR.

The extracted information represented as numerical or categorical variables called molecular (or structural) descriptors **MD** can be further used to establish a mathematical relationship between the structure and biological activity. Thereafter, a subset of informative molecular descriptors **ID**, relevant to the modeled endpoint, are selected for constructing a QSAR model (**E**=*f*(**ID**)).

Emphasizing the importance of the molecular descriptors for establishing QSARs, Votano et al. concluded that "one of the critical determining factors for this (modeling) success arises from the nature of the molecular structure representation employed in the model development process" [29]. In a follow up of a previous benchmarking investigation conducted by collaborators from six laboratories, Zhu et al. went even further to state that "the type of chemical descriptors has much greater influence on the prediction performances of QSAR models than the nature of model optimization techniques" [30].

Table 1 lists a classification of the molecular descriptors according to their continuity, complexity, dimensionality, locality, method of derivation and origin. Depending on their origin (e.g. quantum chemical, topological, geometrical, etc), the calculation of certain descriptor types may require QM optimization of the corresponding molecular structures. As QSA(P)Rs usually operate on datasets of tens or hundreds of compounds, the use of *ab-initio* approaches for geometry optimization is still impractical. However, as most of the QSA(P)Rs are built for congeners, for which the errors in the calculated parameters tend to cancel out, less demanding semi-empirical approaches (such as AM1, PM3 and MNDO) can be used.

Table 1. Classification of the molecular descriptors

Classification basis	Class	Definition	Notes	Examples
numerical property	continuous	descriptor that can take on any value on the scale used for its measurement [31]		Wiener index, LogP, Dewar resonance energy
	discrete	descriptor that can only take on a finite number of values [31]	the indicator variables (e.g. 0 coding for absence and 1 for presence) also belong to this group	number of multiple bonds, number of C atoms, number of aromatic rings
complexity	simple	describing a single effect of interatomic interactions [32]		number of hydrophobic domains, average Pauling atomic electronegativity, valence electrons per atom
	composite	describing two or more effects [32]		topographic-electronic index, number of polar non-H atoms, number of heteroatoms
dimensionality	1D	based on the simplest molecular representation – the chemical formula [31]	sometimes this category is split into 0D and 1D	molecular weight, total number of atoms, number of Cl atoms
	2D	considers how the atoms are connected [31]		Zagreb indices, kappa shape indices, molecular walks counts
	3D	derived from the spatial (x, y, z) coordinates of the molecule [31]	some typical 3D descriptors (e.g. molecular surface area) can be calculated from 2D molecular graphs 33.	3D-MoRSE, GETAWAY, CoMFA fields
	4D	same as the 3D, but also takes into account multiple conformational states [34]	some authors [31] define the CoMFA fields as 4D	4D-fingerprints, GRIND, cell occupancy
locality	whole-molecule	based on the properties of the intact molecule [35]		surface area, Kier flexibility index, Randic index
	fragment	calculated for selected molecular substructures [35]		local spectral moments, multilevel neighborhoods, circular fingerprints
method of derivation origin	experimental	use experimentally determined parameters as molecular descriptors [36]	experimental descriptors such as pKa, LogP, LogD are often calculated	molar refractivity, dipole moment, polarizability
	theoretical	derived from a symbolic representation of the molecule [37]	contain no statistical error due to the lack of experimental noise [37]	partial atomic charges, electrotopological index, Balaban index

Classification basis	Class	Definition	Notes	Examples
	physico-chemical	measurable or calculated physico-chemical properties/parameters [36]		refractive indices, chemical shifts, equilibrium constants
	constitutional	represent the chemical composition of a molecule and are independent of molecular connectivity and geometry [38]		number of atoms, numbers of aromatic bonds, number of benzene rings
	geometrical	derived from the geometrical (3D) representation of the molecule [37]	in most cases require geometry optimization by QM methods	bond lengths, valence and torsion angles
	quantum-chemical	derived from models based on quantum chemical calculations [36]		HOMO, LUMO, total energy
	topological	derived from the chemical graph theory, where the atoms are represented by vertices and the bonds by edges [35]	often mixed with QM or geometrical descriptors to produce quantum-topological and graph-weighted indices	harmonic topological index, Harary index, Kier shape index
	information theory	uses the molecular graph as a source of different probability distributions to which the information theory can be applied [31]		information content indices, vertex complexity, graph distance complexity

Since 1960's, more than 10 000 molecular descriptors have been proposed and are currently in use (see Table 2). However, such a tremendous growth has been criticized for introducing more problems to the QSAR field than it solved[39, 40, 41]. It was summarized by Johnson[42]: "With such an infinite array of descriptions ... we can generate so many hypotheses, relating convoluted molecular factors to activity in such complicated ways, that the process of careful hypothesis testing so critical to scientific understanding has been circumvented in favor of blind validation tests with low resulting information content". It was also argued that this creates not only confusion, but reduces the "human input ... to a mere table with structures and activities"[43].

Table 2. List of software packages for descriptors calculation.

Name	Organization/ Institution	Availability	Descriptors	Platform/ Licence
ADAPT	Pennsylvania State University	http://research.chem.psu.edu/ pcjgroup/adapt.html	>260	Unix/Linux (freeware)
ADMET Predictor	Simulations Plus, Inc.	http://www.simulations-plus.com/	297	Windows (commercial)
CODESSA	Semichem	http://www.semichem.com/ codessa/default.php	>600	Windows/Linux (commercial)
DRAGON	Talete SRL	http://www.talete.mi.it/products/ dragon_description.htm	4885	Windows/Linux (commercial)
EPISUITE	EPA	http://www.epa.gov/opptintr/ exposure/pubs/episuite.htm	~20	Windows (freeware)
MOE	Chemical Computing Group	http://www.chemcomp.com/ software-moe2009.htm	>300	Windows/Linux /SGI/MAC/Sun (freeware)
Molconn-Z	EduSoft	http://www.edusoft-lc.com/ molconn/	327	Windows/Unix/ MAC (commercial)
MOLD[2]	NCTR/FDA	http://www.fda.gov/ScienceResearch/ BioinformaticsTools/ Mold2/default.htm	777	Windows (freeware)
MOLGEN	University of Bayreuth	http://www.molgen.de/?src= documents/molgenqspr.html	707	Windows (commercial)
PowerMV	National Institute of Statistical Sciences	http://nisla05.niss.org/PowerMV/ ?q=PowerMV/	> 1000	Windows (freeware)
Sarchitect	Strand Life Sciences	http://www.strandls.com/ sarchitect/index.html	1084	Windows/Linux (commercial)
SciQSAR	SciMatics	http://www.scimatics.com/jsp/ qsar/QSARIS.jsp	>600	Windows (commercial)

As a result of the above criticism, some software products use a more conservative approach including only descriptors with a well defined physico-chemical meaning or structural interpretation. One such product is CODESSA (COmprehensive DEscriptors for Structural and Statistical Analysis), which was used as a main QSA(P)R tool in our research. In addition to its ability to calculate a wide range of physico-chemical, constitutional, geometrical, quantum-chemical and topological descriptors, CODESSA also provides various data processing (MLR, HMPRO, BMLR, PCA and PLS) and visualization modules.

4.4. Model building and validation

4.4.1. Data preprocessing. The data preprocessing step transforms the original dependent and independent variables into a new set of variables suitable for QSA(P)R analysis.

i) Transformation of the dependent variable.
Biological data is often provided in units that are unsuitable for QSAR analysis. For example the values provided in the Material Safety Data Sheets (MSDS) are typically expressed in *mg* of material per *kg* of subject-body-weight. In such cases the endpoint values have to be converted to molar concentrations by dividing by the molecular weight.

Furthermore, the collected experimental data (EC_{50}, LD_{50}, RBA, etc.) may not be normally distributed. As the Person correlation coefficient (frequently used measure of model performance in QSA(P)R)[44] requires normally distributed data, a transformation of the original endpoint values is often necessary. Since all QSA(P)Rs are essentially Linear Free-Energy Relationships (LFERs), logarithmic transformation functions are preferred. Other alternatives applying sinus, tangent or hyperbolic functions are "statistical unicorns – beasts that exist in paper, but not in reality" [45].

ii) Transformation of the independent variables.
Being defined by unique mathematical expressions, all molecular descriptors inevitably cover vastly different ranges of numerical values. For example the molecular volume (expressed as $Å^3$) usually takes values in the range of a hundred to several thousand units, whereas the partial atomic charge of a C-atom may vary from 0.01 to 0.2 *e* units. If these two descriptors are used in a QSA(P)R equation it would be extremely difficult to determine their relative impact on the modeled endpoint. Hence, it is desirable to use normalization or standardization procedures to bring all descriptors in proportion with one another. If the descriptors in a QSA(P)R model are normalized/standardized, the magnitude of their regression coefficients may serve as an indicator for their importance – i.e. descriptors that are more important will be characterized by larger regression coefficients.

To normalize the values of a molecular descriptor, traditionally means to fit its dynamic range on a scale from 0 to 1. To avoid collapse at the value of zero some models may use an arbitrary range (e.g. 0.1–0.9). The following equation is used for descriptor normalization: $X_{i(normalized,0\ldots1)} = \dfrac{X_i - X_{\min}}{X_{\max} - X_{\min}}$. In this equation X_i is the i^{th} descriptor value subjected to normalization, and X_{min} and

X_{max} are the minimum and the maximum of all descriptor values. Furthermore, the descriptor values can be centered (i.e. the predictors have a mean of zero) using $X_{i(centered,-1...1)} = \frac{X_i - (\frac{X_{max} + X_{min}}{2})}{(\frac{X_{max} - X_{min}}{2})}$. However, among these the standardization is the most commonly used transformation. It converts all descriptors to a common scale with an average of zero and standard deviation of one using the following formula: $X_i = \frac{X_i - \overline{X}_S}{\sigma_{X,S}}$. Here, $\overline{X}_S$ represents the mean of the descriptor values and $\sigma_{X,S}$ represent their standard deviation.

4.4.2. Data processing techniques. A plethora of supervised and unsupervised data processing algorithms are currently in use. Although methods, such as SVM[46] (Support Vector Machines), KNN[47] (K-Nearest Neighbors), GA[48] (Genetic Algorithms), DT[49] (Decision Trees), RF[50] (Random Forests), ANN[51] (Artificial Neural Networks) have become increasingly popular during the past decade, classical methods like MLR (Multiple Linear Regression), PCA (Principle Component Analysis) and PLS (Partial Least Squares) are still preferred due to their simplicity and easy interpretability of the generated models. As even a brief overview of each of these methods would be a challenging task, only the theoretical backgrounds of MLR and ANN will be reviewed. In our experience these two methods proved extremely reliable for establishing structure-activity (property) relationships and complemented well each-other by providing improved accuracy (ANN) and easy interpretation (MLR).

Multiple Linear Regression (MLR)[52]
The multiple linear regression analysis tries to establish a correlation between a dependent variable (y), and a set of explanatory variables (x_1, x_2 …x_n), usually denoted by X [53]. The method relies on the following assumptions:

i) The relationship between the independent (x) variables and the dependent (y) variable is linear;
ii) The residuals between the actual and the estimated values of y follow a normal distribution;
iii) The independent variables x_1, x_2 …x_n should be uncorrelated ($R < 0.5$). As the calculation of the regression coefficients is done through matrix inversion if multicollinearity is present the inversion matrix would be unstable.

The following signs can indicate the presence of multicollinearity between the descriptors in a given descriptor pool:

i) The F-test of the QSA(P)R equation as a whole is significant while none of the t-ratios of the coefficients are statistically significant;

ii) The addition of a new descriptor to the equation radically changes either the size or the sign (plus or minus) of the regression coefficients of the remaining descriptors.

The general equation of the MLR is: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_n x_{in} + \varepsilon_i$. In this equation y_i is the response variable (the dependent variable), β_0 is the intercept, $\beta_{i1}...\beta_{in}$ are the coefficients of the predictors ($x_{i1}...x_{in}$), and ε_i is the difference between the predicted $\hat{y}_i$ and the observed value of y_i (residual). In a matrix form this equation can be written as: $y = X\beta + \varepsilon$ The regression function $f\,[y|X]$ can then be given by $f\,[y|X] = X\beta \sum_{i=0}^{n} x_i \tilde{\beta}_i$. If β has to be estimated from the response variable y, its least squares estimator β^* should minimize the $(y - X\beta^*)'(y - X\beta^*)$ function. The solution of this minimization problem would be given by: $b = (X'X)^{-1} X'y$. Respectively, the coefficients b_i of the fitted function $\hat{y} = Xb = \sum_{i=0}^{n} x_i b_i$, can be regarded as an estimate of the change in the expected value of the response variable y, corresponding to a unit change in the independent variable x_i. The solution b would exist only in case if $(X'X)^{-1}$ exists. For $(X'X)^{-1}$ to exist the explanatory variables (the columns of X) must be linearly independent.

Using the above notations several parameters related to the quality of fit can be defined:

Explained Sum of Squares: $$ESS = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

Total Sum of Squares: $$TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Residual Sum of Squares: $$RSS = TSS - ESS$$

Coefficient of determination: $$R^2 = \frac{ESS}{TSS}$$

Fisher's criterion: $$F = \frac{R^2[n - (k+1)]}{k(1 - R^2)}$$

In the last formula n is equal to the number of observations (data points), whereas k represents the number of independent variables in the model. Other parameters related to the quality of a QSA(P)R model such as the adjusted coefficient of determination (R^2_{adj}), squared standard error (s^2), Predictive Residual

Error Sum of Squares (PRESS), Root Mean Square Error (RMSE) and many others are in use.

Most QSA(P)Rs operate on large descriptors pools. However, only a few of the descriptors are relevant to the modeled endpoint. Thus, feature selection algorithms able to extract a small subset of descriptors from a larger pool are often used. Many such feature selection techniques have been proposed: forward stepwise regression, backward elimination, bidirectional elimination, heuristic approach, etc[54].

In our research we used an improved form of the forward stepwise multiple linear regression algorithm known as Best Multiple Linear Regression (BMLR). This method is encoded in CODESSA PRO[55] and processes data in completely automatic fashion. BMLR starts with the elimination of insignificant descriptors that correlate weakly with the endpoint ($R^2 < R^2_{min}$), have zero sample variance, exhibit low t-values in one-parameter models or correlate above a preset threshold with another descriptor which is more significant. After the elimination step, two-parameter equations of orthogonal (within certain limit) descriptor pairs are generated and the best 400 of these are submitted to the next stage. The algorithm keeps adding orthogonal descriptors and tests each equation using R^2, F and t as performance criteria, until no significant improvement in the statistical parameters of the models is observed (or a predefined maximum number of descriptors is reached)[56].

Artificial Neural Networks (ANN)[51]

The artificial neural networks could be regarded as an alternative to the linear approaches (MLR, PCA, PLS, etc) in cases in which nonlinear trends in the data are expected. They are typically organized in layers, which are constructed from a number of interconnected “nodes” containing an “activation function”. The data is fed to the network via the 'input layer”, which transfers signals to one or more “hidden layers” where the actual data processing is performed. The hidden layers are linked to an “output layer” which carries out the generated answer. The majority of the neural networks implement some form of “learning rule” which transforms the weights of the connections.

Many different types of learning rules have been developed and are currently being used. However, the delta rule is the most often used one. It refers to the architecture of a special class of neural networks (backpropagation neural networks or BPNN) which permanently monitor the error at the output and send a feedback which is used to adjust the weights. For ANNs using backpropagation the “learning” is a supervised process repeated multiple times until the difference between the output and the input (the error) reaches some predefined minimum. When a backpropagation ANN is presented with the input data it makes an initial random guess by assigning random weights to the transformation function. When the answer is generated the BPNN compares it against the input and makes appropriate adjustments to its connection weights so that the next cycle will produce an answer that matches better the input. BPNNs perform a gradient descent within the solution’s hyperparaboloid space towards the global minimum. The error at the global minimum is the lowest possible error. As the parametric space of the solution is not necessarily smooth, to prevent the ANN from settling down in a local minimum a large number of individual runs may be required.

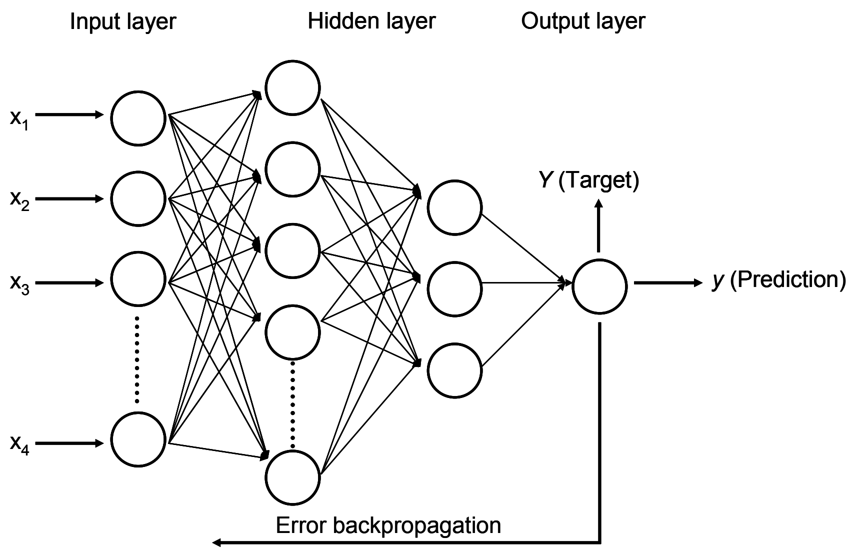

Figure 3. Architecture of a backpropagation neural network.

There are several approached developed with the aim of avoiding local minima. Most ANN algorithms implement routines controlling the "speed" (Beta-coefficient) and the "momentum" of learning[57]. The speed of learning affects the rate of convergence between the current solution and the global minimum, while the momentum helps the network to avoid local minima in the error surface. Once the global minimum is found the already trained network can be used in forward propagation mode to predict the activities of unknowns.

There are two important factors that have to be taken into consideration when using neural networks:

i) The architecture of the network is of critical importance – if too many hidden neurons are used the network will simply memorize the data. If fewer than the optimal number of neurons is used, the network will become unstable, thus unable to generalize.
ii) The number of training cycles must be considered – if excessive number of training cycles is used the network may become over-trained and less likely to predict unknowns correctly.

4.4.3. Validation techniques. Most QSA(P)Rs are built for prediction. Hence, measures based on the statistical quality of fit (R^2, s, F, etc.) are regarded as insufficient for estimating the predictive power of a QSA(P)R model. Therefore several approaches aimed at providing reliable estimates about the predictive performance of QSA(P)R models have been developed. The general principle of validation is to predict the activity of compounds that have not been used to develop the model. Depending on the type of data that is used, two validation procedures are possible: internal and external validation. The internal validation sets aside a given number of compounds to be used as a test set in an iterative procedure that is repeated a certain number of cycles. The external validation ideally uses compounds which activity was reported by a different lab, thus insuring interlab transferability of data and giving a real estimate of the predictive power of a QSA(P)R model.

The simplest form of internal validation is the so called cross-validation [58]. It omits either one (leave-one-out or LOO) or several (leave-many-out or LMO) compounds from the initial set and uses the remaining compounds to build a model. Once the model is built, the activity of the left-out compound(s) is predicted and the cross-validated coefficient of determination (q^2) is calculated. The value of q^2 is calculated using the following formula:

$$q^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \ .$$

In this formula, y_i, $\hat{y}_i$ and $\bar{y}$ are respectively, the measured, predicted, and averaged values of the dependent variable (the endpoint). As a general rule, the difference between the actual R^2 for the QSA(P)R model and q^2 should be less than 0.3[59]. However, high q^2 is not always indicative of the proper behavior of the model when an external set of compounds is being predicted [60].

Bootstrapping [61] is a type of validation in which the initial data set is randomly split multiple times into training and test subsets. Bootstrapping resembles LMO cross-validation. However, in LMO each compound is omitted only once, whereas in bootstraping a compound may be excluded once, multiple times or never. On each bootstrap run a QSA(P)R model is generated, which as in LMO is then used to predict the data for excluded compound. It is recommended that between 20 and 30 percent of the compounds should be excluded to form a test set. As these compounds should be representative of the general population, the $R^2_{bootstrapping}$ should oscillate around the “real” R^2 within reasonable ranges.

A randomization procedure aimed at testing models for potential chance correlations is the so-called Y-scrambling. This procedure randomizes the dependent variable vector, by assigning endpoint values to compounds to which they do not belong. A model relating the permuted dependent variable to the independent variables is then built and R^2_{scr} of this model is compared to the actual R^2 using the original (unpermuted) response data. If the two coefficients of determination are close to each other, the model is a result of chance correlation.

The most stringent form of validation is the external validation. The true external validation uses compounds never seen by the model, which in the most ideal case should have not been a part of the original dataset. The prediction is then carried out and R^2_{ext} for the external test set is compared to the R^2 for the model. In case of data collected from different sources significant differences between these two R^2 should be expected. An important rule when designing external test sets is to make sure that the compounds in the test set are representative of the compounds used to develop the model (i.e. belong to the applicability domain of the model) and to cover approximately the same chemical and activity space. Many algorithms have been developed for this purpose, with most based on similarity [62] or random selection through activity sampling[63].

4.4.4. Identification of outliers. Outliers are atypical compounds with unexpected endpoint values which do not fit well in a QSA(P)R model. This usually happens when: i) a compound acts by a different mechanism, ii) interacts with the target in a different mode[64], iii) interacts via metabolite products, not taken into account[65] or iv) it is affected by a random or a systematic experimental error having little effect on the other chemicals. In statistical terms an outlier is a data point which has a high absolute standardized residual compared to the other compounds in the data set. Because, the coefficients and the intercept of the regression line are highly sensitive to the presence of outliers it is desirable that any such points should be removed. When dealing with outliers it is crucial to: i) make sure that the original data has been entered correctly, ii) generate QSA(P)R models with and without the outliers and estimate their effect on the parameters of the regression line, iii) generate models including explanatory variables that may be essential for explaining the variance attributed to the presence of outliers and iv) if necessary to remove try to find a chemical/biological reasons for the outlier's peculiarity.

Several methods for identification of outliers have gained a widespread application. The most frequently used one is based on the point's deviation from the regression line – if it exceeds 2 standard deviations the point is usually considered an outlier and removed from the model. A plot of the leverage (h) vs the standardized residuals is another approach that can identify not only outliers, but also influential observations (points that may lay on the regression line, but are far away from any of the remaining points and have a huge impact on the parameters of the model if removed). Alternative approach proposed by Tenekedjiev and Radojnova[66] relies on a procedure similar to the cross-validation – each point is excluded once from the model and its Studentized residual is calculated from the regression model of the remaining observations. At the end, all data points with Studentized residuals outside of the confidence interval (95% significance) are considered outliers and removed.

When identifying and removing outliers one has to bear in mind that although the statistical parameters of the model may improve essential information about the data may be lost. The presence of outliers may provide useful insights about important structural characteristics affecting activity, different mechanisms and modes of actions or suggest improvements in the experimental bioassay design.

4.4.5. Domain of Applicability. After a QSA(P)R model is built and properly validated it can then be used to predict the activities/properties of unknowns. However, because the data sets used to build and validate models are finite and necessarily have a limited structural diversity it is to be expected that their applicability to query chemical will be also limited[67]. Hence, reliable predictions are usually confined to chemicals that are structurally similar to the ones used to build the model[68]. The portion of the chemical or activity space in which a model produces reliable predictions is termed Applicability Domain (AD). If an unknown lies outside of the AD of a model it is unlikely that a reliable prediction will be produced. A defined AD provides the following benefits: i) identifies the type of compounds for which reliable predictions can be obtained, ii) determines the degree of generalization of a QSA(P)R model and iii) gives an idea about the interpolation and extrapolation power of a model (the extrapolation of often limited to ±20% beyond the minimum and the maximum values of the endpoint).

Several methods for determining ADs of models have been proposed:

i) Descriptor range based methods – consider either the range of the original descriptors used by the model or their PCA projections;
ii) Geometric method – the AD is defined by the smallest convex area enveloping the training set;
iii) Distance based method – measures the distance (Euclidean, Mahalanobis, Hotelling, Leverage, etc.) between the query chemical and a specified point lying within the cluster formed by the compounds used to build the model;
iv) Probability density function – based on the density distribution of the descriptor values for the compounds used to build the model;
v) Others – based on Tanimoto similarity, KNN, Decision forests, Stepwise approaches, etc.

It should be emphasized that the AD is not intended for resolving problems with models that are not well-founded. For example, the AD can not be used to circumvent the deficiencies of models utilizing collinear descriptors. A better solution would be first to rebuild the models and then to apply ADs.

4.4.6. Model selection. Because the QSA(P)R models are statistical constructs (as opposed to the physical or chemical laws), many possible relationships between the independent and dependent variables can be established. Although, a large number of models may be statistically relevant only a fraction of them would be biologically/chemically sound. Thus, various criteria have to be considered when selecting a final model among the many possible alternatives:

i) *The model should be biologically and/or chemically relevant.* For example if modeling the toxicity of non-polar narcotics, which effect depends on their nonspecific interaction with the cell membranes it can be expected that logP (or any surrogate descriptors related to hydrophobicity) should play a major role. Subsequently, models containing descriptors related to hydrophobicity should be considered more relevant than those containing descriptors that are difficult or impossible to interpret or are unrelated to hydrophobicity.

ii) *Preference for parsimonious models.* The parsimony principle introduced by William of Occam states that among a set of equally good explanations for a given phenomenon, the simplest explanation tends to be the right one. In the context of QSA(P)R themodels should have as few parameters as possible and should be trimmed down until they are minimally adequate. The parsimony principle was extended by Einstein who added that "A model should be as simple as possible. But no simpler" [69].

iii) *Models with superior predictive power.* As the majority of QSA(P)Rs are built for prediction, models able to predict correctly external data should be preferred. Although, an inter-lab data compatibility even if using SOPs is not always granted a model which can predict equally well data coming from several different sources should be preferred.

4.4.7. Interpretation. In general there are two types of QSA(P)R models: 1) models that are built with the sole purpose of predicting an endpoint of interest to fill gaps in the data and 2) models built to pursue a better understanding of the underlying biochemical/chemical/physico-chemical phenomena[70, 71]. When dealing with the second type of QSA(P)R models, the first and major step in their interpretation is the ability to interpret the individual descriptors. Considerable attention has to be paid, since a preconceived notion of the physical interpretation of the descriptors included in the final model may result in misunderstanding of the underlying structure-activity relationship[72].

The problem concerning the descriptors' interpretation is a subject of an increasing attention, due to the high level of abstraction of the theories used for defining new descriptors[73]. For the QM derived descriptors this problem is somewhat mitigated, due to their conceptual proximity to certain physico-chemical concepts such as interaction energies, dipole moments, surface areas, etc.

Although, in most cases the effect of a single molecular descriptor on the endpoint expressed in terms of variation of physicochemical or structural characteristics may be interpretable, the same does not always hold true for models involving multiple descriptors[74]. In such cases, it is the combined effect of all included descriptors that can model the endpoint, and the interpretation of these models cannot be thought of as a simple mechanical unification of the interpretation of the individual descriptors. Moreover, the regression coefficients of the descriptors may not always be straightforward related to their importance and some descriptors may in fact be a surrogate measure for structural features not accurately enough characterized by other more intuitive descriptors.

In most ideal cases the interpretation of a model would include: i) interpretation of the individual descriptors, ii) determination of the importance of each descriptor, iii) understanding how the effects of one descriptor may balance that of another and iv) specific examples from the training and the test sets that highlight the structure-activity relationship encoded by individual descriptors[70].

5. EXAMPLES OF QSAR/QSPR FOR REGULATORY PURPOSES

5.1. QSPR modeling of flash points: An update

Article I proposes MLR and ANN models for the flash points of a diverse set of 758 organic compounds. The flash points are often used in regulation to determine the fire and air pollution hazards of flammable and combustible materials. The Environmental Protection Agency and the US Department of Transportation as well as their analogues in other countries set standards for the storage, handling and transporting of compounds in bulk quantities based on flash point data.

Using the implemented in CODESSA Pro BMLR approach the following relationship between the structure and the property was established:

FP = 0.651(±0.011) x BP + 4948.53(±194.41) x HA dependent HDCA-1/TMSA (Zefirov PC) + 60.704(±5.467) x HASA-1/TMSA (Zefirov PC) (all) + 300.245(±30.508) x Relative number of triple bonds + 30.185(±4.49)

BMLR: n = 758; $R^2 = 0.849$; $R^2_{cv} = 0.846$; $s^2 = 18.9$; avg. $R^2_{training} = 0.861$; avg. $R^2_{hold\text{-}out\ test} = 0.808$

Of the above descriptors the boiling point (BP) was a composite descriptor calculated using the following equation:

$$BP = 65.8 \times \sqrt[3]{G_i} + 18470 \times HDSA(2)$$

In this equation $\sqrt[3]{G_i}$ is the cubic root of the gravitational index and HDSA(2) is the area-weighted surface charge of the hydrogen bonding donor atoms.

The HA dependent HDCA-1/TMSA (Zefirov PC) descriptor determines the compounds' ability to form hydrogen bonds. Due to the positive regression coefficient of this descriptor it was concluded that compounds with smaller hydrogen donor charged surface area would be highly flammable.

The third descriptor (HASA-1/TMSA (Zefirov PC) (all)) encodes the size of the solvent-accessible surface area of H-bonding acceptor atoms as a part of the total molecular surface area. Again, due to its positive regression coefficient compounds with smaller HASA-1 will be easily flammable.

The last descriptor in the QSPR model is the relative number of triple bonds. It plays the role of a correction factor explaining the flammability of compounds having triple bonds in their molecules.

This BMLR model was validated using 2 different strategies: i) cross-validation and ii) a separation of the initial test into training (2/3 of the total) and hold-out test sets (1/3 of the total). The small difference between $R^2_{training}$ and $R^2_{hold\text{-}out\ test}$ indicates excellent stability and predictive power of the model.

The proposed QSPR can be used for the prediction of flash points for a wide range of organic compounds with an average error of 13.9 K.

The use of back-propagation fully-connected artificial neural network with 4 input nodes, 3 neurons in the hidden layer and 1 producing the output resulted in the following statistical parameters:

ANN (4-3-1 architecture): $n_{training} = 600$; $n_{test} = 158$; $R^2_{training} = 0.878$; $R^2_{test} = 0.824$

While 3 of the descriptors used for the ANN modeling overlapped with those used in BMLR, the "relative number of triple bonds" descriptor was displaced by a topological descriptor – the Balaban J index. As the Balaban index depends on the type of bonds connecting the atoms in a molecule such substitution is not surprising. It was found that the ANN model slightly outperforms the BMLR model and resulted in an average error of prediction of 12.6 K.

5.2. Application of the QSPR Approach to the Boiling Points of Azeotropes

Article II presents a model for the boiling points of a large dataset of 426 azeotropic mixtures including alcohols, esters, alkanes, amines, ketones, ethers, and halogeno and nitro compounds. The regulatory agencies use boiling points to rank compounds that can be potential hazardous air pollutants. According to Modified Hazardous Air Pollutant Prioritization System (MHAPPS), compounds of great concern are those produced in quantities exceeding 1 million pounds per year and have a vapor pressure > 100 mm Hg and boiling points < 80°C. Hence methods for estimating of the boiling points are highly sought. This is especially true in the case of the azeotropic mixtures, which for a long time were out of scope of the traditional QSPR, which predominantly deals with the properties of pure compounds.

Two different approaches to calculate the descriptors for the mixtures were used: i) to calculate them as averages of the corresponding molecular descriptors for each individual component (centroid approximation) and ii) to scale each descriptor using weighting factors proportional to the molar fraction of each component in the mixture (weighted-contribution-factor approximation).

X(WF)	X(CA)	X(WFCS)	X(CACS)	Descriptor name
166.8	166.8	170.8	168.9	intercept
0.358	0.3801	0.3541	0.3703	gravitation index (all bonds)
4059	4581	3972	4220	HA-dependent HDCA-1/TMSA (Zefirov PC) (all)
1.131	0.9487	1.153	0.9546	DPSA2 difference in CPSAs (PPSA2-PNSA2) (Zefirov PC)
17.01	11.52	16.73	12.43	PPSA3 atomic-charge-weighted PPSA (Zefirov PC)

WF – weighted contribution factors; **CA** – centroid approximation; **WFCS** – weighted contribution factors for the complete set; **CACS** – centroid approximation for the complete set.

The statistical parameters of the above presented models were as follows:

WF: training set: n = 320, test set: n= 106, $R^2_{training} = 0.750$, $R^2_{cv} = 0.740$, $R^2_{test} = 0.683$, F = 235.7, s = 23.40
CA: training set: n = 320, test set: n= 106, $R^2_{training} = 0.755$, $R^2_{cv} = 0.744$, $R^2_{test} = 0.693$, F = 242.2, s = 23.16.
WFCS: n = 426, $R^2 = 0.732$, $R^2_{cv} = 0.723$, F = 287.0, s = 23.94
CACS: n = 426, $R^2 = 0.738$, $R^2_{cv} = 0.730$, F = 297.1, s = 23.64

As can be seen all models have similar performance and are characterized by stable regression equations, which coefficients vary insignificantly from model to model. A Y-scrambling procedure performed 20 times resulted in $R^2_{scrambling}$ of less than 0.06, which indicates lack of random correlations.

All models are in complete agreement with the Clausius-Clapeyron equation expressing the boiling points as a function of the enthalpy (heat) of vaporization (ΔH_{vap}). The enthalpy of vaporization itself depends on: i) the nonspecific or van der Waals attractions between the molecules composing the mixture, ii) the distribution of the electron density causing induced dipole-dipole interactions, and iii) the ability to form hydrogen bonds.

In terms of ΔH_{vap}, the gravitation index is related to the van der Waals attractions between the molecules.

The second by order of importance descriptor in all models is the "HA-dependent HDCA-1/TMSA (Zefirov PC) (all). It is defined by the hydrogen bond acceptor-dependent hydrogen bond donor charged surface area as a fraction of the total molecular surface area. It is related to the hydrogen-bond formation ability of the components of the azeotropic mixtures, describing the third term of ΔH_{vap}. Both the gravitation index and the HA-dependent HDCA-1/TMSA descriptors have positive regression coefficients, and respectively their increase would lead to an increase in the boiling points. The "DPSA2 difference in CPSAs (PPSA2-PNSA2) (Zefirov PC)" (for which PPSA represents the partial positively charged surface area and PNSA represents the partial negatively charged surface area) and "PPSA3 atomic-charge-weighted PPSA (Zefirov PC)" descriptors are related to the second component of the ΔH_{vap} and are a measure of the charge redistribution within a molecule. The charge redistribution itself determines the ability of molecules to participate in induced dipole-dipole interactions.

5.3. Quantitative Structure-Property Relationship Studies on Ostwald Solubility and Partition Coefficients of Organic Solutes in Ionic Liquids

Article III proposes QSPR models for the correlation of Ostwald solubility coefficients (log*L*) and partition coefficients (log*P*) of organic solutes in eight ionic liquids for datasets ranging in size from 30 to 60 compounds (a total of 92 organic solutes). Both Ostwald solubility and LogP have an immense value for the regulatory agencies. The Ostwald solubility is often used to for assessing occupational exposures, while LogP is used to set water quality standards and to regulate ecotoxicants.

The majority of the investigated ionic liquids were imidazoliums, with the only exception being ionic liquid #8 (with available data for 38 organic solutes) containing a pyridinium ring. For ionic liquids #1 to #5 the anion was bis(trifluoromethylsulfonyl)imide, for #6 and #8 was tetrafluoroborate and lastly for #7 was hexafluorophosphate counterion.

For each of these eight ionic liquids and for all 92 organic solutes, the Ostwald solubility coefficients were calculated using the following equation:

$$\log L = \log(\frac{RT}{\gamma_i^{\infty} pV}).$$

In the above equation γ_i^{∞} is the infinite dilution activity coefficient; p is the vapor pressure of the solute at the system temperature; and V is the molar volume of the ionic liquid solvent.

The partition coefficients log$P$ for the solutes in the eight ionic liquids were calculated by the following equation:

$$\log P = \log L - \log L_W.$$

In this equation $\log L_W$ is the logarithm of the Ostwald coefficient of the solute in water.

Separate QSPR models for logL and logP for each of the 8 ionic liquid solutions were developed (for a total of 16 models). To achieve a level of performance similar to that of log$P$, the log$L$ models needed twice as many descriptors (2 for log$P$ vs 4 for log$L$). Coefficients of determination ranging from 0.913 to 0.992 and s^2 ranging from 0.012 to 0.069 for the training sets were obtained. All models were validated using cross-validation and a hold-out test sets.

The most important descriptors involved in the log$L$ models, namely the "Area weighted surface charge of hydrogen bonding donor atoms" and the "Difference between atomic charge weighted partial positive and negative surface areas" are related to the charge distribution within the molecules. The charge distribution is a major factor describing the electrostatic interactions between the solute and the ionic liquid. The proposed QSPR models also emphasize the role of the hydrogen donor-acceptor interactions in ionic liquid solutions. It was found that in general log$L$ increases with the increase of the hydrogen donor/acceptor ability of the solute. Other descriptors with high frequency of occurrence were "average complementary information content", the "Randic index", and the "Gravitational index" all related to the size and shape of the molecules. Similar charge related descriptors ("DPSA2 difference in CPSAs (PPSA2-PNSA2) Zefirov PC" and "Minimum partial charge (Zefirov) for all atom types") emphasizing the significant role of the solvent-solute interactions were involved in the Log*P* QSPRs.

5.4. Quantitative Structure – Activity Relationship (QSAR) Modeling of EC_{50} of Aquatic Toxicities for *Daphnia Magna*

Article V proposes a QSAR model for the aquatic toxicity of 130 benzoic acids, benzaldehydes, phenylsulfonyl acetates, cycloalkane-carboxylates, benzanilides, and other esters to *Daphnia magna*. *Daphnia magna* is used extensively in bioassays monitoring the health of aquatic ecosystems, because it is both an important primary consumer of algae, bacteria and detritus and itself a major food source in the food chain.

Since the majority of the earlier published models report significant single descriptor correlations with logP, our initial hypothesis was that such a relationship may also exist for the current diverse dataset. For this purpose the KOWWIN logP program, part of the EPI Suite package (v. 3.12) was used to calculate the values of the octanol-water partition coefficients. The calculated logP values for all 130 compounds were then correlated to their corresponding $\log(1/EC_{50})$ values. A statistically insignificant correlation ($R^2 = 0.097$), indicated that factors beyond hydrophobicity play a substantial role in aquatic toxicity to *Daphnia magna* for this dataset. That is, published models reporting significant single parameter correlations with logP may have benefited from the use of highly homogeneous datasets, for which other structural features contributing to toxicity were not varied.

Due to the lack of significant correlation between the $\log(1/EC_{50})$ values and LogP the following model using the implemented in CODESSA BMLR algorithm was developed:

$$\log(1/EC_{50}) = 1.227(\pm0.094) \times \text{count of H-acceptor sites [Zefirov's PC]} - 0.248(\pm0.025) \times \text{Number of aromatic bonds} - 1.929(\pm0.214) \times \text{Balaban index} + 0.332(\pm0.053) \times \text{LogP} - 280.3(\pm46.21) \times \text{HA dependent HDCA-2/TMSA [Semi-MO PC]} + 8.968(\pm0.713)$$

$$n = 130;\ R^2 = 0.712;\ R^2_{cv} = 0.676;\ F = 61.3;\ s^2 = 0.600$$

Two outliers (cyclohexyl 2-(4-nitrophenylsulfonyl)propanoate and isopropyl tetradecanoate) were identified and were subsequently removed from the model. After their removal the following model was obtained:

$$\log(1/EC_{50}) = 1.226(\pm0.086) \times \text{count of H-acceptor sites [Zefirov's PC]} - 0.268(\pm0.023) \times \text{Number of aromatic bonds} - 2.054(\pm0.199) \times \text{Balaban index} + 0.377(\pm0.052) \times \text{LogP} - 273.4(\pm42.24) \times \text{HA dependent HDCA-2/TMSA [Semi-MO PC]} + 9.382(\pm0.664)$$

$$n = 130;\ R^2 = 0.759;\ R^2_{cv} = 0.728;\ F = 77.0;\ s^2 = 0.499$$

These models were validated using internal cross-validation, hold-out test (1/3 of the total) and Y-scrambling. As a result of these procedures an average R^2_{test} of 0.64 and an average $R^2_{scrambling}$ of 0.192 (ranging 0.176 to 0.223) were obtained.

Due to the positive regression coefficient of the "count of H-acceptor sites [Zefirov's PC]" descriptor the higher the number of H-acceptor sites, the greater the toxicity. Usually these acceptor sites are associated with the presence of heteroatoms (mainly N, O, S). Both the "Number of aromatic bonds" and the "Balaban index" descriptors are characterized by negative regression coefficients. The Balaban index is a function of the size of molecule, degree of branching, and the degree of unsaturation. Therefore, bulkier and more branched molecules will be less toxic, probably due to their lower membrane penetration abilities. The "Number of aromatic bonds" descriptor is related to the degree of aromaticity of a compound, and thus to its relative stability. Therefore, the lower degree of saturation will result in compounds which are less toxic. The octanol/water partition coefficient (logP) accounts for hydrophobicity and in agreement with earlier publications its increase would lead to an increased toxicity. The least significant in the above models is the "HA dependent HDCA-2/TMSA" descriptor, which represents the hydrogen-bonding-dependent hydrogen donor charged surface area. This descriptor plays a complementary role to the "Count of H-acceptor sites [Semi-MO PC]" descriptor. Its negative sign suggests that compounds with high toxicity would act mainly as hydrogen bond acceptors, as opposed to compounds with low toxicity, which would predominantly act as hydrogen bond donors.

Furthermore, a meta-analysis of the data suggested a strong relationship between the lethal and sublethal effects on *Daphnia magna.* The $\log(1/LC_{50})$ and $\log(1/EC_{50})$ values were found to correlate with R^2 of 0.98. Hence, the applicability of the carried out model for $\log(1/EC_{50})$ can be extended and used for prediction of the lethal concentrations of unknowns.

5.5. Correlation of the Photolysis Half-Lives of Polychlorinated Dibenzo-p-dioxins and Dibenzofurans with Molecular Structure

Article VI describes a QSPR model for the photolysis half-life of 70 polychlorinated dibenzo-*p*-dioxins (PCDDs) and dibenzofurans (PCDFs). These compounds are byproducts of a variety of industrial processes and their high lipophilicity and chemical stability cause rapid accumulation in organisms, organic phase of soils and sediments making them persistent organic pollutants. The accumulation of PCDDs and PCDFs in vertebrates is known to lead to thymic atrophy, reproductive changes, weight loss, impairment of immune response, and disruption of the endocrine system, etc. Their rate of photodegradation (photolysis half-life ($t_{1/2}$)) is used often for assessment of the environmental risk of these chemicals.

A linear free energy relationship (LFER) was established between the logarithm of the PCDDs and PCDFs half-lives and a set of quantum mechanical descriptors calculated by CODESSA PRO.

$$t_{1/2} = -0.746\ (\pm 0.070) \times \text{total molecular 2-center resonance energy/no. of atoms} + 0.555\ (\pm 0.155) \times \text{min resonance energy for bond C-O} + 0.0633(\pm 0.0213) \times \text{min } e\text{-}n \text{ attraction for bond C-C} - 35.80(\pm 7.11)$$

$$n = 70;\ R^2 = 0.720;\ R^2_{cv} = 0.687;\ F = 56.62;\ s^2 = 0.003$$

This model was validated using cross-validation, hold-out test set (1/3 of the total number of compounds) and Y-scrambling procedures. The average $R^2_{\text{hold-out}}$ was 0.712, whereas $R^2_{\text{scrambling}}$ was 0.194 (ranging from 0.174 to 0.0.207). The significant difference between $R^2_{\text{hold-out}}$ and $R^2_{\text{scrambling}}$ indicates the stability of the model, its statistical significance and the lack of chance correlations.

According to the Student's *t* test the significance of the descriptors in the proposed model decreases in the following order: total molecular two-center resonance energy/ no. of atoms > min resonance energy for bond C-O > minimum *e-n* attraction for bond C-C.

In this model the "total molecular two-center resonance energy/no. of atoms" descriptor emphasizes the importance of resonance energy and bond order for the photolysis process. This descriptor is as a measure of stability of the aromatic system of PCDDs and PCDFs, which depends on the number and orientation of the Cl substituents. The remaining two descriptors, namely the "min resonance energy for bond C-O" and the "min *e-n* attraction for bond C-C" are related, respectively, to the C-O and C-C bond strengths. The positive regression coefficients of these two descriptors are a consequence of the direct dependence of the photolysis half-lives on the stability of the C-O and C-C bonds. The presence of the "minimum resonance energy for bond C-O" descriptor in the proposed QSPR model supports the hypothesis of Dulin et al. (*Environ. Sci. Technol.* 1988, 20, 72) that C-O rather than C-Cl cleavage is the major route for the photodegradation of PCDDs.

The currently proposed has several advantages: i) it is general – describes the photolysis half-lives of both PCDDs and PCDFs; ii) it is simple – involves only 3 quantum chemical descriptors in a multilinear model; iii) it is easy interpretable – the descriptors have clear physicochemical meaning related to the nature of photolysis; (4) it is robust – the model was validated using three different procedures to confirm its stability and that the observed correlation did not arise by chance.

5.6. Computational Chemistry Approaches for Understanding how Structure Determines Properties

Article IV provides a short review of QSAR and QSPR models for a variety of physical, physicochemical and biological endpoints. It emphasizes the significance of the computational modeling approaches for understanding how the molecular structure of compounds determines their properties.

The manuscript gives a brief overview of the QSA(P)R approach and lists some frequently used software products. It discusses the application of validation techniques with an application to QSA(P)Rs derived from large descriptor pools that are aimed at avoiding chance correlations. Further, linear and non-linear QSA(P)R models for various properties and biological activities such as UV spectral absorbance, critical micelle concentrations of non-ionic, anionic and cationic surfactants, flash points, duration of protection of mosquito repellents, milk-to-plasma concentration ratios and others are described. The quality of all models is discussed and a brief interpretation in terms of structural characteristics determining the observed effect is provided.

The manuscript concludes that the QSA(P)R techniques have a huge potential of expansion into areas involving modeling of increasingly complex endpoints and to provide new insights on how molecular structure determines the properties of chemicals.

6. CONCLUSIONS

Linear and non-linear MLR and ANN QSA(P)R approaches were successfully applied to build highly accurate models for biological and physico-chemical endpoints of a great importance for the regulatory agencies. An extension of the standard QSA(P)R approach allowed building of predictive models for mixtures, thus demonstrating the applicability of the chemometric methods for modeling of complex processes. Various methods of validation such as cross-validation, hold-out test set validation and Y-scrambling were utilized to build reliable QSA(P)R models adhering to the OECD principles. Meta-analysis of the data for overlapping sets of compounds allowed transferability of the modeling results from one endpoint to another (e.g. $Log(1/EC_{50})$ and $Log(1/LC_{50})$). Due to the specific nature of the utilized molecular descriptors (most originating from the quantum chemistry) the majority of models (except for those based on ANN) were accompanied by a mechanistic interpretation pointing to specific molecular characteristics affecting the modeled endpoints. It was demonstrated that QSA(P)R approach can be used for making regulatory decisions solely on the basis of computational models, avoiding lengthy and expensive testing.

7. ACKNOWLEDGMENTS

I would like to thank my doctoral advisor Professor Mati Karelson for his excellent guidance throughout my research.

8. REFERENSES

1. REACH: Regulation (EC) No 1907/2006: Consolidated version (June 2012) Annex XI, &1.3
2. Quantitative Structure-Activity Relationships (QSAR) for Pesticide Regulatory Purposes, Ed. Emilio Benfenati. Elsevier, (2011) p.16.
3. http://www.princeton.edu/~ota/disk1/1995/9553/955315.PDF
4. USEPA (U.S. Environmental Protection Agency). (2004) Economic Analysis of the Proposed Change in Data Requirements Rule for Conventional Pesticides. Washington, DC, http://alttox.org/assets/files/usepa-2004.pdf
5. Binetti, R.; Costamagna, F.M.; Marcello, I. Ann. 1st. Super Sanita., 44 (2008) 13–15.
6. Walker, J.D.; Carlsen, L.; Jaworska, J.S. QSAR Comb. Sci. 22 (2003) 346–350.
7. Organization for Economic Co-operation and Development. Guidance document on the validation of (quantitative) structure-activity relationships [(Q)SAR] models. ENV/JM/MONO(2007)2
8. Technical Guidance Document in support of Commission Directive 93/67/EEC on Risk Assessment for New Notified Substances and Commission Regulation (EC) No. 1488/94 on Risk Assessment for Existing Substances. 1996. Luxembourg: Office for Official Publications of the European Communities. Part III: Use of QSAR, Use Categories, Risk Assessment Format, CR-48-96-003-EN-C, ISBN 92-827-8013-9.
9. Recent Advances in QSAR Studies. Challenges and Advances in Computational Chemistry and Physics Volume 8, 2010, pp 367–382.
10. QSAR Model Reporting Format (Version 1.2) EUROPEAN COMMISSION Toxicology and Chemical Substances Unit, Ispra, (2008) Available at: http://ihcp.jrc.ec.europa.eu/our_labs/predictive_toxicology/qsar_tools/qrf/QMRF_version_1.2.pdf
11. Saliner, A. G.; Netzeva, T. I.; Worth, A. P. SAR QSAR Environ. Res. 17 (2006) 195–223.
12. Pavan, M.; Netzeva, T. I.; Worth, A. P. SAR QSAR Environ. Res. 17 (2006) 147–171.
13. Vracko, M.; Bandelj, V.; Barbieri, P.; Benfenati, E.; Chaudhry, Q.; Cronin, M. et al. SAR QSAR Environ. Res. 17 (2006) 265–284.
14. Roberts, D. W.; Aptula, A. O.; Patlewicz, G. Chem. Res. Toxicol. 19 (2006) 1228–1233.
15. Cros, A. F. A. Action de l'alcool amylique sur l'organisme. Thesis, University of Strasbourg, Strasbourg, France. 1863
16. Richardson, J. Medical Times and Gazette 2 (1868) 703.
17. Richet, C., C. R. Seances Soc. Biol. 9 (1893) 775.
18. Meyer, H. Arch. Experim. Pathol. und Pharmakol. 42 (1899) 109.
19. Overton, E. Z. Physik. Chem. 22 (1897) 189
20. Hansch, C., Fujita, T. J. Am. Chem. Soc. 86 (1964) 1616–1626.
21. http://biochem.wustl.edu/wildmans/Methods/QSAR.html
22. QSAR, QSAR: Hansch Analysis and Related Approaches, by Hugo Kubinyi p.96
23. Young, D., Martin, T., Venkatapathy, R., Harten, P. QSAR Comb. Sci. 27 (2008) 1337–1345.
24. Southan, C., Varkonyi, P., Muresan, S. J. Cheminformatics 1 (2009) 1–10.

25. Fourches, D., Muratov, E., and Tropsha, A. J Chem Inf Model. 50 (2010) 1189–1204.
26. Klopman, G., Chakravarti, S. K., Harris, N., Ivanov, J., Saiakhov, R. D. SAR QSAR Environ Res. 14(2) (2003) 165–80.
27. Lozano, S., Lescot, E., Halm, M. P., Lepailleur, A., Bureau, R., Rault, S. J Enzyme Inhib Med Chem. 25(2) (2010) 195–203.
28. Carlsen, L., Walker, J.D. QSARs for Prioritizing PBT Substances to Promote Pollution Prevention. QSAR & Combinatorial Science 22 (2003) 49–57.
29. Votano, J.R.; Parham, M.; Hall, L.H.; Kier, L.B.; Oloff, S.; Tropsha. A.; Xie, Q.; Tong, W. Mutagenesis 19 (2004) 365–377.
30. Zhu, H., Tropsha, A., Fourches, D., Varnek, A., Papa, E., Gramatica, P., Oberg, T., Dao, P., Cherkasov, A., Tetko, I. J. Chem. Inf. Model. 48 (2008) 766–784.
31. Todeschini, R., Consonni, V. Molecular Descriptors for Chemoinformatics, Wiley-VCH (2009) p. 518, p. 815
32. Raevsky, O. Molecular structure descriptors in the computer-aided design of biologically active compounds Rus. Chem. Rev. 1999, 68, 505–524.
33. Kier, L. B. Indexes of molecular shape from chemical graphs, Med. Res. Rev. 7 (1987) 417–440.
34. Andrade, C. H.. Pasqualoto, K. F.. Ferreira, E. I.. Hopfinger, A. J. Molecules 15 (2010) 3281–3294.
35. Winkler, D. A. Briefings in bioinformatics 3(1) (2002) 73–86.
36. Randić, M. J. Mol. Graph. Model. 20 (2001) 19–35.
37. Puzyn, T., Leszczynski, J., Cronin, M. T. D. (Eds.) Recent Advances in QSAR Studies: Methods and Applications, Springer, Dordrecht, The Netherlands, (2010) p.33, p. 68
38. Katritzky, A. R., Kuanar, M., Slavov, S., Hall, C. D., Karelson, M., Kahn, I., Dobchev, D. Chem. Rev. 110 (2010) 5714–5789.
39. Labute, P. J. Mol. Graph. Model. 18 (2000) 464–477.
40. Doweyko, A.M. J. Comput.-Aided Mol. Des. 22 (2008) 81–89.
41. Sprous, D.G., Palmer, R.K., Swanson, J.T., Lawless, M. Curr. Top. Med. Chem. 10 (2010) 619–637.
42. Johnson, S.R. J. Chem. Inf. Model. 48 (2008) 25–26.
43. Olah, M., Bologa, C., Oprea, T.I. J. Comput.-Aided Mol. Des. 18 (2004) 437–449.
44. Rodgers, J. L., W. Nicewander, The American Statistician 42 (1988) 59–66.
45. Kubinyi, H. Hansch analysis and related approaches. Volume 1 of Methods and principles in medicinal chemistry, VCH, Weinheim, 1993.
46. Cortes, C., Vapnik, V. N. Machine Learning, 20 (1995). 273–297.
47. Altman, N. S. The American Statistician 46 (3) (1992) 175–185.
48. Mitchell, M. An Introduction to Genetic Algorithms. Cambridge, MA: MIT Press (1996).
49. Yuan, Y., Shaw, M. J. Induction of fuzzy decision trees. Fuzzy Sets and Systems 69 (1995) 125–139.
50. Liaw, A.,Wiener, M. R News 2(3) (2002) 18–22.
51. Bishop, C.M. Neural Networks for Pattern Recognition, Oxford: Oxford University Press, (1995).
52. Draper, N.R., Smith, H. Applied Regression Analysis (3rd ed.). John Wiley, (1998).
53. Darlington, R. B. Regression and linear models, New York, McGrow-Hill, (1990).
54. Draper, N. and Smith, H. Applied Regression Analysis, 2d Edition, New York: John Wiley & Sons, Inc. (1981).

55. CODESSA PRO Software, University of Florida, 2002. http://www.codessa-pro.com/
56. Katritzky, A., Dobchev, D., Slavov, S., Karelson, M. J. Chem. Inf. Model. 48 (2008) 2207–2213.
57. http://pages.cs.wisc.edu/~bolo/shipyard/neural/local.html
58. Hawkins, D.M. J Chem Inf Comput Sci 44 (2004) 1–12.
59. Leach, A. R. Molecular modeling: Principles and applications. Pearson Education Ltd. Harlow, England, (2001).
60. Golbraikh, A., Tropsha, A., J Mol Graph Model 20(4) (2002) 269–76.
61. Wehrens, R., van der Linden, W. E., J. Chemom. 11 (1997) 157–171.
62. Gramatica, P., Giani, E., Papa, E. J Mol Graph Model 25 (2007) 755–766.
63. Bhhatarai, B., Gramatica, P. Chem Res Toxicol 23 (2010) 528–539.
64. Verma, R. P., Hansch, C. Bioorg Med Chem;13 (2005) 4597–621.
65. Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology, Karcher, W., Devillers, J. Eds. Kluwer Academic Publishers, Dordrecht, The Netherlands (1990), p.132.
66. Tenekedjiev, K., and Radojnova, D. Acta Morphologica et Anthropologica, 6 (2001) 90–97.
67. Dimitrov, S., Dimitrova, G., Pavlov, T., Dimitrova, N., Patlewicz, G., Niemela, J., Mekenyan, O. A. Stepwise approach for defining the applicability domain of SAR and QSAR models. J. Chem. Inf. Model. 45 (2005) 839–849.
68. Nikolova-Jeliazkova, N., Jaworska, J. Altern. Lab. Anim. 33 (2005) 461–470.
69. Crawley, M. J. “Statistics: An Introduction using R”, Wiley, New York, (2005) pp. 7–8.
70. Guha, R. J. Comput. Aided Mol. Des. 22 (2008) 857–871.
71. Livingstone, D. J. J. Chem. Inf. Comput. Sci. 40 (2000) 195–209.
72. Stanton, D. T. J. Chem. Inf. Comput. Sci. 43 (2003) 1423–1433.
73. Faulon, J.L., Brown, W.M., Martin, S. J. Comput. Aided Mol. Des. 19 (2005) 637–650.
74. Katritzky, A. R., Pertukhin,R., Tatham, D., Basak, S., Benfenati, E., Karelson, M., Maran, U. J.Chem. Inf. Comput. Sci 41 (2001) 670–685.

9. SUMMARY IN ESTONIAN

Valideerimine QSAR / QSPR reguleerimise eesmärgil

Lineaarseid ja mittelineaarseid MLR ja ANN QSA(P)R lähenemisi rakendati edukalt täpsete mudelite arendamiseks erinevatele bioloogiliste ja füsiko-keemiliste omadustele, mis omavad suurt tähtsust regulatiivses kasutuses. Traditsioonilise QSA(P)R lähenemise täiendamine võimaldas luua prognoos-mudelid segude jaoks, mis ühtlasi näitas ka kemomeetriliste meetodite rakendatavust keeruliste süsteemide modelleerimiseks. Erinevaid valideerimise meetodeid, nende hulgas ristvalideerimine, välise valideerimisseti kasutamine ja Y-scrambling, rakendati kooskõlas OECD printsiipidega usaldusväärsete QSA(P)R mudelite loomiseks. Kattuvate kemikaalide alamhulkade metaanalüüs võimaldas modelleerimise tulemusi üle kanda ühelt omaduselt teisele, näiteks Log(1/EC_{50}) ja Log(1/LC_{50}) vahel. Kasutatud molekulaardeskriptorite spetsiifiline olemus (enamasti kvantkeemiast tuletatud) võimaldas lisada enamusele mudelitele, välja arvatud ANN meetodil põhinevatele, mehhanistliku interpretatsiooni, mis näitas spetsiifiliste molekulaarsete karakteristikute seoseid modelleeritud omadustega. Näidati, et QSA(P)R lähenemine on rakendatav regulatiivsete otsuste tegemiseks, kusjuures arvutusmudelite kasutamine võimaldab vältida ajamahukaid ja kalleid laboratoorseid katseid.

PUBLICATIONS

CURRICULUM VITAE

Name: Iva Bozhidarova Stoyanova-Slavova
Data of Birth: November 20, 1975
Citizenship: Bulgarian
Address: 13500 Chenal Parkway, Little Rock, USA, ZIP 72211
Phone number: +372 567 91196; +1 352 871 4678;
E-mail: ivabss@yahoo.com

Education:

2013–present PhD student, Institute of Chemistry, University of Tartu, Estonia

1995–2001 M.Sc. in Chemistry, Department of Chemistry, Sofia University "St. Kliment Ohridski", Bulgaria

Language skill:

Bulgarian (native speaker)
English (fluent)
Russian (average in speech and writing)

Professional career:

2011–2013 External contractor at National Center for Toxicological Research, Jefferson, AR

2005–2010 Researcher at the Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, Gainesville, USA

2003–2004 Young Researcher IMAGETOX, 5th EUROPIAN Framework Program, Tartu, Estonia

Main field of research:

Computational Chemistry, QSAR/QSPR and drug design.

List of publications:

Dimitrova, Y.; Slavova, I. Vibrational frequencies and infrared intensities of the hydrogen-bonded complexes of nitrous acid with ethers: ab initio and DFT studies. Spectrochimica Acta, Part A: Molecular and Biomolecular Spectroscopy (2005), 61A, 2095–2102.

Katritzky, A. R.; Kulshyn, O.; Stoyanova-Slavova, I.; Dobchev, D.; Kuanar, M.; Fara, D.; Karelson, M. Antimalarial activity: A QSAR modeling using CODESSA PRO software. Bioorganic & Medicinal Chemistry (2006), 14, 2333–2357.

Katritzky, A. R.; Stoyanova-Slavova, I.; Dobchev, D. A.; Karelson, M. QSPR modeling of flash points: An update. Journal of Molecular Graphics & Modelling (2007), 26, 529–536.

Katritzky, A. R.; Kuanar, M.; Stoyanova-Slavova, I.; Slavov, S.; Dobchev, D.; Karelson, M.; Acree, W. Quantitative Structure-Property Relationship Studies on Ostwald Solubility and Partition Coefficients of Organic Solutes in Ionic Liquids. Journal of Chemical & Engineering Data 2008, 53, 1085–1092.

Katritzky, A. R.; Slavov, S.; Stoyanova-Slavova, I.; Kahn, I.; Karelson, M. QSAR modeling of aquatic toxicity. Journal of Toxicology and Environmental Health, Part A 2009, 72, 1181–1190.

Katritzky, A. R.; Slavov, S.; Radzvilovits, M.; Stoyanova-Slavova, I.; Karelson, M. Computational Chemistry Approaches for Understanding how Structure Determines Properties. Verlag der Zeitschrift für Naturforschung 2009, 64, 773–777.

Slavov, S.; Radzvilovits, M.; LeFrancois, S.; Stoyanova-Slavova, I.; Soti, F.; Kem, W. R.; Katritzky, A. R. A Computational Study of the Binding of 3-(Benzylidene)-Anabaseines to Two Major Brain Nicotinic Acetylcholine Receptors and to the Acetylcholine Binding Protein. European Journal of Medicinal Chemistry 2010, 45, 2433–2446.

Katritzky A. R., Slavov S. H., Stoyanova-Slavova I. B., Karelson M. Correlation of the photolysis half-lives of polychlorinated dibenzo-p-dioxins and dibenzofurans with molecular structure Journal of Physical Chemistry A. 2010, 114, 2684–2688.

Katritzky A. R., Girgis A. S., Slavov S. H., Tala S., Stoyanova-Slavova I. B. QSAR modeling, synthesis and bioassay of diverse leukemia RPMI-8226 cell line active agents. European journal of medicinal chemistry 2010, 45(11) 5183–5199.

Katritzky, A. R.; Stoyanova-Slavova, I.; Tämm, K.; Tämm, T.; Karelson, M. Application of the QSPR approach to the boiling points of azeotropes Journal of physical chemistry. A 2011, 115, 3475–3479.

Conference Participations:

- Poster presented at the Humboldt Research Conference on Computational Chemistry, Veliko Turnovo, Bulgaria, 2002
- First National Conference with International Participation on e-Learning in Higher Education – 14–17 September 2004, Kiten, Bulgaria
- Poster presented at the 6th Annual Florida Heterocyclic and Synthetic Chemistry Conference Gainesville, FL – 2005
- Poster presented at the 7th Annual Florida Heterocyclic and Synthetic Chemistry Conference, Gainesville, FL – 2006
- Poster presented at the 8th Annual Florida Heterocyclic and Synthetic IUPAC-Sponsored Conference, Gainesville, FL – 2007
- Poster presented at the 9th Annual Florida Heterocyclic and Synthetic IUPAC-Sponsored Conference, Gainesville, FL – 2008
- Poster presented at the 10th Annual Florida Heterocyclic and Synthetic IUPAC-Sponsored Conference, Gainesville, FL – 2009
- Poster presented at the 11th Annual Florida Heterocyclic and Synthetic IUPAC-Sponsored Conference, Gainesville, FL – 2010

ELULOOKIRJELDUS

Nimi: Iva Bozhidarova Stoyanova-Slavova
Sünniaeg: 20. november, 1975
Kodakondsus: Bulgaaria
Aadress: 13500 Chenal Parkway, Little Rock, USA, ZIP 72211
Telefon: +372 567 91196; +1 352 871 4678;
E-post: ivabss@yahoo.com

Haridus:
2013–praegu Doktorant – Keemia instituut, Tartu Ülikool, Eesti
1995–2001 Magistri kraad keemias – 2001, Keemia osakond, Sofia Ülikool "St. Kliment Ohridski", Bulgaaria

Keelteoskus:
Bulgaaria (emakeel)
Inglise (kõnes hea, kirjas hea)
Vene (kõnes keskmne, kirjas keskmine)

Teenistuskäik:
2011–2013 Väline töövõtja, Rahvusvaheline Toksikoloogiliste Uuringute Keskus, Jefferson, AR
2005–2010 Teadlane, Heterotsüklilise Ühendite Keskus, Florida Ülikool, Gainesville, USA
2003–2004 Nooremteadur, IMAGETOX, 5th Euroopa Raamprogramm, Tartu, Eesti

Peamised uurimisvaldkonnad:
Arvutuskeemia, QSAR/QSPR ja ravimi disain.

Publikatsioonide loetelu:
Dimitrova, Y.; Slavova, I. Vibrational frequencies and infrared intensities of the hydrogen-bonded complexes of nitrous acid with ethers: ab initio and DFT studies. Spectrochimica Acta, Part A: Molecular and Biomolecular Spectroscopy 2005, 61A, 2095–2102.

Katritzky, A. R.; Kulshyn, O.; Stoyanova-Slavova, I.; Dobchev, D.; Kuanar, M.; Fara, D.; Karelson, M. Antimalarial activity: A QSAR modeling using CODESSA PRO software. Bioorganic & Medicinal Chemistry 2006, 14, 2333–2357.

Katritzky, A. R.; Stoyanova-Slavova, I.; Dobchev, D. A.; Karelson, M. QSPR modeling of flash points: An update. Journal of Molecular Graphics & Modelling 2007, 26, 529–536.

Katritzky, A. R.; Kuanar, M.; Stoyanova-Slavova, I.; Slavov, S.; Dobchev, D.; Karelson, M.; Acree, W. Quantitative Structure-Property Relationship Studies on Ostwald Solubility and Partition Coefficients of Organic Solutes

in Ionic Liquids. Journal of Chemical & Engineering Data 2008, 53, 1085–1092.
Katritzky, A. R.; Slavov, S.; Stoyanova-Slavova, I.; Kahn, I.; Karelson, M. QSAR modeling of aquatic toxicity. Journal of Toxicology and Environmental Health, Part A 2009, 72, 1181–1190.
Katritzky, A. R.; Slavov, S.; Radzvilovits, M.; Stoyanova-Slavova, I.; Karelson, M. Computational Chemistry Approaches for Understanding how Structure Determines Properties. Verlag der Zeitschrift für Naturforschung 2009, 64, 773–777.
Slavov, S.; Radzvilovits, M.; LeFrancois, S.; Stoyanova-Slavova, I.; Soti, F.; Kem, W. R.; Katritzky, A. R. A Computational Study of the Binding of 3-(Benzylidene)-Anabaseines to Two Major Brain Nicotinic Acetylcholine Receptors and to the Acetylcholine Binding Protein. European Journal of Medicinal Chemistry 2010, 45, 2433–2446.
Katritzky A. R., Slavov S. H., Stoyanova-Slavova I. B., Karelson M. Correlation of the photolysis half-lives of polychlorinated dibenzo-p-dioxins and dibenzofurans with molecular structure Journal of Physical Chemistry A. 2010, 114, 2684–2688.
Katritzky A. R., Girgis A. S., Slavov S. H., Tala S., Stoyanova-Slavova I. B. QSAR modeling, synthesis and bioassay of diverse leukemia RPMI-8226 cell line active agents. European journal of medicinal chemistry 2010, 45(11) 5183–5199.
Katritzky, A. R.; Stoyanova-Slavova, I.; Tämm, K.; Tämm, T.; Karelson, M. Application of the QSPR approach to the boiling points of azeotropes Journal of physical chemistry. A 2011, 115, 3475–3479.

Konverentside teesid:

- Osales postriga konverentsil "Humboldt Research Conference on Computational Chemistry", Veliko Turnovo, Bulgaaria, 2002
- Osales konverentsil "First National Conference with International Participation on e-Learning in Higher Education", Kiten, Bulgaaria, 2004
- Osales postriga konveretsil "6th Annual Florida Heterocyclic and Synthetic Chemistry Conference", Gainesville, FL – 2005
- Osales postriga konveretsil "7th Annual Florida Heterocyclic and Synthetic Chemistry Conference", Gainesville, FL – 2006
- Osales postriga konveretsil "8th Annual Florida Heterocyclic and Synthetic IUPAC-Sponsored Conference", Gainesville, FL – 2007
- Osales postriga konveretsil "9th Annual Florida Heterocyclic and Synthetic IUPAC-Sponsored Conference", Gainesville, FL – 2008
- Osales postriga konveretsil "10th Annual Florida Heterocyclic and Synthetic IUPAC-Sponsored Conference", Gainesville, FL – 2009
- Osales postriga konveretsil "11th Annual Florida Heterocyclic and Synthetic IUPAC-Sponsored Conference", Gainesville, FL – 2010

DISSERTATIONES CHIMICAE UNIVERSITATIS TARTUENSIS

1. **Toomas Tamm.** Quantum-chemical simulation of solvent effects. Tartu, 1993, 110 p.
2. **Peeter Burk.** Theoretical study of gas-phase acid-base equilibria. Tartu, 1994, 96 p.
3. **Victor Lobanov.** Quantitative structure-property relationships in large descriptor spaces. Tartu, 1995, 135 p.
4. **Vahur Mäemets.** The ^{17}O and ^{1}H nuclear magnetic resonance study of H_2O in individual solvents and its charged clusters in aqueous solutions of electrolytes. Tartu, 1997, 140 p.
5. **Andrus Metsala.** Microcanonical rate constant in nonequilibrium distribution of vibrational energy and in restricted intramolecular vibrational energy redistribution on the basis of slater's theory of unimolecular reactions. Tartu, 1997, 150 p.
6. **Uko Maran.** Quantum-mechanical study of potential energy surfaces in different environments. Tartu, 1997, 137 p.
7. **Alar Jänes.** Adsorption of organic compounds on antimony, bismuth and cadmium electrodes. Tartu, 1998, 219 p.
8. **Kaido Tammeveski.** Oxygen electroreduction on thin platinum films and the electrochemical detection of superoxide anion. Tartu, 1998, 139 p.
9. **Ivo Leito.** Studies of Brønsted acid-base equilibria in water and non-aqueous media. Tartu, 1998, 101 p.
10. **Jaan Leis.** Conformational dynamics and equilibria in amides. Tartu, 1998, 131 p.
11. **Toonika Rinken.** The modelling of amperometric biosensors based on oxidoreductases. Tartu, 2000, 108 p.
12. **Dmitri Panov.** Partially solvated Grignard reagents. Tartu, 2000, 64 p.
13. **Kaja Orupõld.** Treatment and analysis of phenolic wastewater with microorganisms. Tartu, 2000, 123 p.
14. **Jüri Ivask.** Ion Chromatographic determination of major anions and cations in polar ice core. Tartu, 2000, 85 p.
15. **Lauri Vares.** Stereoselective Synthesis of Tetrahydrofuran and Tetrahydropyran Derivatives by Use of Asymmetric Horner-Wadsworth-Emmons and Ring Closure Reactions. Tartu, 2000, 184 p.
16. **Martin Lepiku.** Kinetic aspects of dopamine D_2 receptor interactions with specific ligands. Tartu, 2000, 81 p.
17. **Katrin Sak.** Some aspects of ligand specificity of P2Y receptors. Tartu, 2000, 106 p.
18. **Vello Pällin.** The role of solvation in the formation of iotsitch complexes. Tartu, 2001, 95 p.

19. **Katrin Kollist.** Interactions between polycyclic aromatic compounds and humic substances. Tartu, 2001, 93 p.
20. **Ivar Koppel.** Quantum chemical study of acidity of strong and superstrong Brønsted acids. Tartu, 2001, 104 p.
21. **Viljar Pihl.** The study of the substituent and solvent effects on the acidity of OH and CH acids. Tartu, 2001, 132 p.
22. **Natalia Palm.** Specification of the minimum, sufficient and significant set of descriptors for general description of solvent effects. Tartu, 2001, 134 p.
23. **Sulev Sild.** QSPR/QSAR approaches for complex molecular systems. Tartu, 2001, 134 p.
24. **Ruslan Petrukhin.** Industrial applications of the quantitative structure-property relationships. Tartu, 2001, 162 p.
25. **Boris V. Rogovoy.** Synthesis of (benzotriazolyl)carboximidamides and their application in relations with *N*- and *S*-nucleophyles. Tartu, 2002, 84 p.
26. **Koit Herodes.** Solvent effects on UV-vis absorption spectra of some solvatochromic substances in binary solvent mixtures: the preferential solvation model. Tartu, 2002, 102 p.
27. **Anti Perkson.** Synthesis and characterisation of nanostructured carbon. Tartu, 2002, 152 p.
28. **Ivari Kaljurand.** Self-consistent acidity scales of neutral and cationic Brønsted acids in acetonitrile and tetrahydrofuran. Tartu, 2003, 108 p.
29. **Karmen Lust.** Adsorption of anions on bismuth single crystal electrodes. Tartu, 2003, 128 p.
30. **Mare Piirsalu.** Substituent, temperature and solvent effects on the alkaline hydrolysis of substituted phenyl and alkyl esters of benzoic acid. Tartu, 2003, 156 p.
31. **Meeri Sassian.** Reactions of partially solvated Grignard reagents. Tartu, 2003, 78 p.
32. **Tarmo Tamm.** Quantum chemical modelling of polypyrrole. Tartu, 2003. 100 p.
33. **Erik Teinemaa.** The environmental fate of the particulate matter and organic pollutants from an oil shale power plant. Tartu, 2003. 102 p.
34. **Jaana Tammiku-Taul.** Quantum chemical study of the properties of Grignard reagents. Tartu, 2003. 120 p.
35. **Andre Lomaka.** Biomedical applications of predictive computational chemistry. Tartu, 2003. 132 p.
36. **Kostyantyn Kirichenko.** Benzotriazole – Mediated Carbon–Carbon Bond Formation. Tartu, 2003. 132 p.
37. **Gunnar Nurk.** Adsorption kinetics of some organic compounds on bismuth single crystal electrodes. Tartu, 2003, 170 p.
38. **Mati Arulepp.** Electrochemical characteristics of porous carbon materials and electrical double layer capacitors. Tartu, 2003, 196 p.

39. **Dan Cornel Fara.** QSPR modeling of complexation and distribution of organic compounds. Tartu, 2004, 126 p.
40. **Riina Mahlapuu.** Signalling of galanin and amyloid precursor protein through adenylate cyclase. Tartu, 2004, 124 p.
41. **Mihkel Kerikmäe.** Some luminescent materials for dosimetric applications and physical research. Tartu, 2004, 143 p.
42. **Jaanus Kruusma.** Determination of some important trace metal ions in human blood. Tartu, 2004, 115 p.
43. **Urmas Johanson.** Investigations of the electrochemical properties of poly-pyrrole modified electrodes. Tartu, 2004, 91 p.
44. **Kaido Sillar.** Computational study of the acid sites in zeolite ZSM-5. Tartu, 2004, 80 p.
45. **Aldo Oras.** Kinetic aspects of dATPαS interaction with P2Y_1 receptor. Tartu, 2004, 75 p.
46. **Erik Mölder.** Measurement of the oxygen mass transfer through the air-water interface. Tartu, 2005, 73 p.
47. **Thomas Thomberg.** The kinetics of electroreduction of peroxodisulfate anion on cadmium (0001) single crystal electrode. Tartu, 2005, 95 p.
48. **Olavi Loog.** Aspects of condensations of carbonyl compounds and their imine analogues. Tartu, 2005, 83 p.
49. **Siim Salmar.** Effect of ultrasound on ester hydrolysis in aqueous ethanol. Tartu, 2006, 73 p.
50. **Ain Uustare.** Modulation of signal transduction of heptahelical receptors by other receptors and G proteins. Tartu, 2006, 121 p.
51. **Sergei Yurchenko.** Determination of some carcinogenic contaminants in food. Tartu, 2006, 143 p.
52. **Kaido Tämm.** QSPR modeling of some properties of organic compounds. Tartu, 2006, 67 p.
53. **Olga Tšubrik.** New methods in the synthesis of multisubstituted hydra-zines. Tartu. 2006, 183 p.
54. **Lilli Sooväli.** Spectrophotometric measurements and their uncertainty in chemical analysis and dissociation constant measurements. Tartu, 2006, 125 p.
55. **Eve Koort.** Uncertainty estimation of potentiometrically measured ph and pK_a values. Tartu, 2006, 139 p.
56. **Sergei Kopanchuk.** Regulation of ligand binding to melanocortin receptor subtypes. Tartu, 2006, 119 p.
57. **Silvar Kallip.** Surface structure of some bismuth and antimony single crystal electrodes. Tartu, 2006, 107 p.
58. **Kristjan Saal.** Surface silanization and its application in biomolecule coupling. Tartu, 2006, 77 p.
59. **Tanel Tätte.** High viscosity $Sn(OBu)_4$ oligomeric concentrates and their applications in technology. Tartu, 2006, 91 p.

60. **Dimitar Atanasov Dobchev**. Robust QSAR methods for the prediction of properties from molecular structure. Tartu, 2006, 118 p.
61. **Hannes Hagu**. Impact of ultrasound on hydrophobic interactions in solutions. Tartu, 2007, 81 p.
62. **Rutha Jäger.** Electroreduction of peroxodisulfate anion on bismuth electrodes. Tartu, 2007, 142 p.
63. **Kaido Viht.** Immobilizable bisubstrate-analogue inhibitors of basophilic protein kinases: development and application in biosensors. Tartu, 2007, 88 p.
64. **Eva-Ingrid Rõõm.** Acid-base equilibria in nonpolar media. Tartu, 2007, 156 p.
65. **Sven Tamp.** DFT study of the cesium cation containing complexes relevant to the cesium cation binding by the humic acids. Tartu, 2007, 102 p.
66. **Jaak Nerut.** Electroreduction of hexacyanoferrate(III) anion on Cadmium (0001) single crystal electrode. Tartu, 2007, 180 p.
67. **Lauri Jalukse.** Measurement uncertainty estimation in amperometric dissolved oxygen concentration measurement. Tartu, 2007, 112 p.
68. **Aime Lust.** Charge state of dopants and ordered clusters formation in CaF_2:Mn and CaF_2:Eu luminophors. Tartu, 2007, 100 p.
69. **Iiris Kahn**. Quantitative Structure-Activity Relationships of environmentally relevant properties. Tartu, 2007, 98 p.
70. **Mari Reinik.** Nitrates, nitrites, N-nitrosamines and polycyclic aromatic hydrocarbons in food: analytical methods, occurrence and dietary intake. Tartu, 2007, 172 p.
71. **Heili Kasuk.** Thermodynamic parameters and adsorption kinetics of organic compounds forming the compact adsorption layer at Bi single crystal electrodes. Tartu, 2007, 212 p.
72. **Erki Enkvist.** Synthesis of adenosine-peptide conjugates for biological applications. Tartu, 2007, 114 p.
73. **Svetoslav Hristov Slavov**. Biomedical applications of the QSAR approach. Tartu, 2007, 146 p.
74. **Eneli Härk.** Electroreduction of complex cations on electrochemically polished Bi(*hkl*) single crystal electrodes. Tartu, 2008, 158 p.
75. **Priit Möller.** Electrochemical characteristics of some cathodes for medium temperature solid oxide fuel cells, synthesized by solid state reaction technique. Tartu, 2008, 90 p.
76. **Signe Viggor.** Impact of biochemical parameters of genetically different pseudomonads at the degradation of phenolic compounds. Tartu, 2008, 122 p.
77. **Ave Sarapuu.** Electrochemical reduction of oxygen on quinone-modified carbon electrodes and on thin films of platinum and gold. Tartu, 2008, 134 p.
78. **Agnes Kütt.** Studies of acid-base equilibria in non-aqueous media. Tartu, 2008, 198 p.

79. **Rouvim Kadis.** Evaluation of measurement uncertainty in analytical chemistry: related concepts and some points of misinterpretation. Tartu, 2008, 118 p.
80. **Valter Reedo.** Elaboration of IVB group metal oxide structures and their possible applications. Tartu, 2008, 98 p.
81. **Aleksei Kuznetsov.** Allosteric effects in reactions catalyzed by the cAMP-dependent protein kinase catalytic subunit. Tartu, 2009, 133 p.
82. **Aleksei Bredihhin.** Use of mono- and polyanions in the synthesis of multisubstituted hydrazine derivatives. Tartu, 2009, 105 p.
83. **Anu Ploom.** Quantitative structure-reactivity analysis in organosilicon chemistry. Tartu, 2009, 99 p.
84. **Argo Vonk.** Determination of adenosine A_{2A}- and dopamine D_1 receptor-specific modulation of adenylate cyclase activity in rat striatum. Tartu, 2009, 129 p.
85. **Indrek Kivi.** Synthesis and electrochemical characterization of porous cathode materials for intermediate temperature solid oxide fuel cells. Tartu, 2009, 177 p.
86. **Jaanus Eskusson.** Synthesis and characterisation of diamond-like carbon thin films prepared by pulsed laser deposition method. Tartu, 2009, 117 p.
87. **Marko Lätt.** Carbide derived microporous carbon and electrical double layer capacitors. Tartu, 2009, 107 p.
88. **Vladimir Stepanov.** Slow conformational changes in dopamine transporter interaction with its ligands. Tartu, 2009, 103 p.
89. **Aleksander Trummal.** Computational Study of Structural and Solvent Effects on Acidities of Some Brønsted Acids. Tartu, 2009, 103 p.
90. **Eerold Vellemäe.** Applications of mischmetal in organic synthesis. Tartu, 2009, 93 p.
91. **Sven Parkel.** Ligand binding to 5-HT_{1A} receptors and its regulation by Mg^{2+} and Mn^{2+}. Tartu, 2010, 99 p.
92. **Signe Vahur.** Expanding the possibilities of ATR-FT-IR spectroscopy in determination of inorganic pigments. Tartu, 2010, 184 p.
93. **Tavo Romann**. Preparation and surface modification of bismuth thin film, porous, and microelectrodes. Tartu, 2010, 155 p.
94. **Nadežda Aleksejeva.** Electrocatalytic reduction of oxygen on carbon nanotube-based nanocomposite materials. Tartu, 2010, 147 p.
95. **Marko Kullapere.** Electrochemical properties of glassy carbon, nickel and gold electrodes modified with aryl groups. Tartu, 2010, 233 p.
96. **Liis Siinor.** Adsorption kinetics of ions at Bi single crystal planes from aqueous electrolyte solutions and room-temperature ionic liquids. Tartu, 2010, 101 p.
97. **Angela Vaasa.** Development of fluorescence-based kinetic and binding assays for characterization of protein kinases and their inhibitors. Tartu 2010, 101 p.

98. **Indrek Tulp.** Multivariate analysis of chemical and biological properties. Tartu 2010, 105 p.
99. **Aare Selberg.** Evaluation of environmental quality in Northern Estonia by the analysis of leachate. Tartu 2010, 117 p.
100. **Darja Lavõgina.** Development of protein kinase inhibitors based on adenosine analogue-oligoarginine conjugates. Tartu 2010, 248 p.
101. **Laura Herm.** Biochemistry of dopamine D_2 receptors and its association with motivated behaviour. Tartu 2010, 156 p.
102. **Terje Raudsepp.** Influence of dopant anions on the electrochemical properties of polypyrrole films. Tartu 2010, 112 p.
103. **Margus Marandi.** Electroformation of Polypyrrole Films: *In-situ* AFM and STM Study. Tartu 2011, 116 p.
104. **Kairi Kivirand.** Diamine oxidase-based biosensors: construction and working principles. Tartu, 2011, 140 p.
105. **Anneli Kruve.** Matrix effects in liquid-chromatography electrospray mass-spectrometry. Tartu, 2011, 156 p.
106. **Gary Urb.** Assessment of environmental impact of oil shale fly ash from PF and CFB combustion. Tartu, 2011, 108 p.
107. **Nikita Oskolkov.** A novel strategy for peptide-mediated cellular delivery and induction of endosomal escape. Tartu, 2011, 106 p.
108. **Dana Martin.** The QSPR/QSAR approach for the prediction of properties of fullerene derivatives. Tartu, 2011, 98 p.
109. **Säde Viirlaid.** Novel glutathione analogues and their antioxidant activity. Tartu, 2011, 106 p.
110. **Ülis Sõukand.** Simultaneous adsorption of Cd^{2+}, Ni^{2+}, and Pb^{2+} on peat. Tartu, 2011, 124 p.
111. **Lauri Lipping.** The acidity of strong and superstrong Brønsted acids, an outreach for the "limits of growth": a quantum chemical study. Tartu, 2011, 124 p.
112. **Heisi Kurig.** Electrical double-layer capacitors based on ionic liquids as electrolytes. Tartu, 2011, 146 p.
113. **Marje Kasari.** Bisubstrate luminescent probes, optical sensors and affinity adsorbents for measurement of active protein kinases in biological samples. Tartu, 2012, 126 p.
114. **Kalev Takkis.** Virtual screening of chemical databases for bioactive molecules. Tartu, 2012, 122 p.
115. **Ksenija Kisseljova**. Synthesis of aza-β^3-amino acid containing peptides and kinetic study of their phosphorylation by protein kinase A. Tartu, 2012, 104 p.
116. **Riin Rebane.** Advanced method development strategy for derivatization LC/ESI/MS. Tartu, 2012, 184 p.

117. **Vladislav Ivaništšev.** Double layer structure and adsorption kinetics of ions at metal electrodes in room temperature ionic liquids. Tartu, 2012, 128 p.
118. **Irja Helm.** High accuracy gravimetric Winkler method for determination of dissolved oxygen. Tartu, 2012, 139 p.
119. **Karin Kipper.** Fluoroalcohols as Components of LC-ESI-MS Eluents: Usage and Applications. Tartu, 2012, 164 p.
120. **Arno Ratas.** Energy storage and transfer in dosimetric luminescent materials. Tartu, 2012, 163 p.
121. **Reet Reinart-Okugbeni**. Assay systems for characterisation of subtype-selective binding and functional activity of ligands on dopamine receptors. Tartu, 2012, 159 p.
122. **Lauri Sikk.** Computational study of the Sonogashira cross-coupling reaction. Tartu, 2012, 81 p.
123. **Karita Raudkivi.** Neurochemical studies on inter-individual differences in affect-related behaviour of the laboratory rat. Tartu, 2012, 161 p.
124. **Indrek Saar.** Design of GalR2 subtype specific ligands: their role in depression-like behavior and feeding regulation. Tartu, 2013, 126 p.
125. **Ann Laheäär.** Electrochemical characterization of alkali metal salt based non-aqueous electrolytes for supercapacitors. Tartu, 2013, 127 p.
126. **Kerli Tõnurist.** Influence of electrospun separator materials properties on electrochemical performance of electrical double-layer capacitors. Tartu, 2013, 147 p.
127. **Kaija Põhako-Esko.** Novel organic and inorganic ionogels: preparation and characterization. Tartu, 2013, 124 p.
128. **Ivar Kruusenberg.** Electroreduction of oxygen on carbon nanomaterial-based catalysts. Tartu, 2013, 191 p.
129. **Sander Piiskop.** Kinetic effects of ultrasound in aqueous acetonitrile solutions. Tartu, 2013, 95 p.
130. **Ilona Faustova**. Regulatory role of L-type pyruvate kinase N-terminal domain. Tartu, 2013, 109 p.
131. **Kadi Tamm.** Synthesis and characterization of the micro-mesoporous anode materials and testing of the medium temperature solid oxide fuel cell single cells. Tartu, 2013, 138 p.