DISSERTATIONES CHIMICAE UNIVERSITATIS TARTUENSIS
98

DISSERTATIONES CHIMICAE UNIVERSITATIS TARTUENSIS
98

INDREK TULP

Multivariate analysis of chemical and biological properties

TARTU UNIVERSITY
PRESS

Institute of Chemistry, University of Tartu, Estonia

This Dissertation is accepted for the commencement of the degree of Doctor of Philosophy in Molecular Design on 21 June, 2010, by the Doctoral Committee of the Department of Chemistry, University of Tartu.

Supervisor: Dr. Uko Maran, University of Tartu

Opponent: Prof. Mark Cronin, School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, England

Commencement: August 25, 2010 at 14A Ravila Str., room 1021, 12:00 h

ISSN 1406–0299
ISBN 978–9949–19–418–6 (trükis)
ISBN 978–9949–19–419–3 (PDF)

Autoriõigus: Indrek Tulp, 2010

Tartu Ülikooli Kirjastus
www.tyk.ee
Tellimus nr. 383

to my family

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

The present thesis consists of four articles listed below. All papers are denoted in the text by Roman numerals I–IV.

I. Katritzky, A.R.; Tulp, I.; Fara, D.C.; Lauria, A.L.; Maran, U.; Acree, W.E. Jr. A General Treatment of Solubility. 3. Principal Component Analysis (PCA) of the Solubilities of Diverse Solutes in Diverse Solvents, *J. Chem. Inf. Model.* **2005**, *45*, 913–923.

II. Tulp, I.; Dobchev, D.A.; Katritzky, A.R.; Acree, W. Jr.; Maran, U. A General Treatment of Solubility. 4. Description and Analysis of a PCA model for Ostwald solubility coefficients, *J. Chem. Inf. Model.* **2010**, (accepted).

III. Tulp, I.; Sild, S.; Maran, U. Relationship between structure and permeability in artificial membranes: theoretical whole molecule descriptors in development of QSAR models, *QSAR & Comb. Sci.* **2009**, *28*, 811–814.

IV. Katritzky, A.R.; Dobchev, D.A.; Tulp, I.; Karelson, M.; Carlson, D.A. QSAR Study of Mosquito Repellents Using Codessa Pro, *Bioorg. Med. Chem. Let.* **2006**, *16*, 2306–2311.

Author's contribution

Publication I, II, III: The author is responsible for the data sets, calculations, interpretation of the results and preparation of the manuscripts.

Publication IV: The author is responsible for the data set, calculations, interpretation of the results and writing of corresponding parts of the manuscript.

LIST OF ABBREVIATIONS

ADME-Tox	Absorption, Distribution, Metabolism, Excretion and Toxicity
ANN	Artificial Neural Network
BMLR	Best Multiple Linear Regression
CPSA	Charged Partial Surface Area
HB	Hydrogen Bonding
HOMO	Highest Occupied Molecular Orbital
log D_{oct}	pH dependent logarithm of the octanol–water partition coefficient
log L	logarithm of the Ostwald solubility coefficient
log P_{oct}	logarithm of the octanol–water partition coefficient
log P_{PAMPA}	logarithm of PAMPA permeability
LUMO	Lowest Unoccupied Molecular Orbital
MLR	Multi-Linear Regression
MO	Molecular Orbital
PAMPA	Parallel Artificial Membrane Permeability Assay
PCA	Principal Component Analysis
PCR	Principal Component Regression
PLS	Partial Least Squares
PT	Protection Time
QSAR	Quantitative Structure–Activity Relationship
QSPR	Quantitative Structure–Property Relationship

INTRODUCTION

The properties of substances are determined by the structure of chemicals. The explanation for and illustrations of this basic knowledge of chemistry is often attributed to the interdisciplinary field of science called computational chemistry where computer-assisted modeling provides means for finding relationships between chemical structure and properties. Usually, this has been established in the form of Quantitative Structure Property/Activity Relationships. More precisely, in QSPR/QSAR studies chemical or biological data are related to structural descriptors through mathematical models. Often a vast number of those descriptors are calculated and they form a multidimensional descriptor space. For the elucidation of useful information from such a multidimensional space several chemometrics approaches exist. For solving multivariate problems, linear methods are often used, such as multilinear regression, but they require effective variable selection. At the same time, principal component analysis can handle large amount of variables simultaneously.

Over the years QSAR/QSPR has served as a valuable tool while nesting mechanistic information from chemical structure and property relationships. In addition to providing valuable scientific information, QSPR/QSAR models have found very practical applications. Properly validated models can be used to predict environmentally and toxicologically important data for new and existing industrial chemicals. The biggest and most recent driving force has been the European legislation on Registration, Evaluation and Authorization of Chemicals (REACH). Also, the U. S. Environmental Protection Agency routinely accepts QSAR-derived values for regulatory purposes. Another even wider use of QSAR models is *in silico* drug design where they are used for screening of new drug candidates.

This Thesis presents research on the multidimensional analysis of some chemical and biological properties of organic compounds using QSPR/QSAR and PCA methodologies. Structurally important characteristics of these properties are revealed and their mechanistic relationships are interpreted. In Chapter 1, an overview of properties and methodologies is given. Chapter 2 summarizes results of original publications in the multivariate modeling of selected properties.

1. LITERATURE OVERVIEW

1.1 Chemical and biological properties of interest

Nowadays, a computational chemist uses chemical structure to explain and describe a wide range of different properties. These properties can be roughly divided into two groups, physico-chemical and chemico-biological. Physico-chemical properties (shortly chemical properties) deal with molecular interactions in clearly defined systems (such as a single solvent) and which usually have an easily definable mechanistic knowledgebase. Chemico-biological properties (shortly biological properties) in turn can be attributed to the interactions of chemicals with complex biological systems, where the mechanistic knowledgebase is not always clearly defined. Biological properties are often *in vivo* and *in vitro* assay endpoints, related to drug design (pharmacology), risk assessment (toxicology), etc. These properties often possess a complex mechanism involving multiple molecular interactions.

Probably the most important chemical properties are lipophilicity and solubility, which characterize the transport and availability of chemicals, respectively. For instance, the octanol–water partition coefficient (log P_{oct}) helps to model the partitioning between water and biological medium, showing the capability of compounds to penetrate a membrane. Solubility in turn, is important for the distribution of chemicals, providing a measure of the availability of a chemical in the environment. Encompassed in the current Thesis is a computational chemistry approach used to explore the solubility of chemicals in different solvents.

Chemical properties are often used in the modeling of biological properties and activities. In the field of drug design, the ADME-Tox profiles of chemical compounds are characterized by multiple values for absorption, distribution, metabolism, excretion and toxicity of chemicals. The absorption process is related to bioavailability which, in addition to solubility, depends directly on the permeability of chemicals. In ADME-Tox profiles, this is characterized by membrane permeability. Included in the present Thesis, an *in silico* model is derived for artificial membrane permeability. Biological properties or model systems can be even more complex. One of these, the activity of repellents is also a subject of the present Thesis. Following, a short description of each endpoint is provided.

Solubility is a fundamental property in almost all fields which are related to chemistry. It is critical in the production of new materials and substances, assessing environmental risk for the sustainability of environment and health, detecting drug-likeness, etc. Solubility can be expressed as concentration, such as molarity, molality, mole fraction, mass percentage, etc., and as a distribution constant or coefficient, such as Henry's law constant, Bunsen coefficient, Ostwald coefficient, etc.[1,2] Most often, solubility in water is assessed, but for an

overall description of solubility, comparative analysis of a variety of solvents and solute–solvent pairs is required. Such data is usually scarce.

A good experimental knowledge base is provided by the Ostwald solubility coefficient (log L) which is a distribution coefficient of a solute distributed between a liquid solvent and gas phase and is related to the solute's free energy of solvation according to eq. 1:[3]

$$\Delta G_{\mathrm{S}} = -2.3RT \log L = -2.3RT \log\left(\frac{c_{\mathrm{l}}}{c_{\mathrm{g}}}\right), \tag{1}$$

where c_{l} and c_{g} are the solute's concentrations in the liquid and gas phases, respectively. This relationship is valid for standard states of unit concentration in the gas phase and in solution, and the dependence is linear with respect to log L at a constant temperature.

The free energy of solvation (ΔG_{S}) is considered to consist of four main components (eq. 2):[4-6] the cavity formation term ($\Delta G_{\mathrm{cavity}}$), dispersion interactions (ΔG_{disp}), free energy of electrostatic interactions (ΔG_{el}), and a term which takes into consideration the formation and reorganization of hydrogen bonds (ΔG_{HB}). The first two terms in eq. 2 are related to the bulk characteristics of the solute, and together they are the major energy contributors to the solvation free energy. This also holds for systems that are normally known to be very polar and strongly hydrogen bonded (HB).[7] Both terms ($\Delta G_{\mathrm{cavity}}$, ΔG_{disp}) can be regarded as characteristics of nonspecific interactions. The term for electrostatic interactions (ΔG_{el}) involves, in addition to the pure electrostatic Coulomb interactions, other interaction forces such as ion–dipole, strong dipole–dipole, and ion–pair formation, etc.[4,8,9] The HB forces are also electrostatic by nature.[10] Hence, it can be concluded that the last two terms comprise the electrostatic-specific interactions.

$$\Delta G_{\mathrm{S}} = \Delta G_{\mathrm{cavity}} + \Delta G_{\mathrm{disp}} + \Delta G_{\mathrm{el}} + \Delta G_{\mathrm{HB}} \tag{2}$$

The described solute–solvent interactions and diverse theories unfolding those interactions have formed the basis for understanding solubility, comprehensively reviewed by Reichardt.[4] Despite more than a century of studies directed toward examining the relationships between chemical structure and solubility, the challenge still remains for improved experimental detection, precise computational prediction, and detailed understanding of interactions between chemicals and the surrounding medium.[11]

Permeability analysis is crucial in estimating drug oral bioavailability in ADME-Tox profile, where the permeability is used to describe the intestinal absorption. *In vitro* experimental studies are often performed with several different cell lines such as Caco-2, MDCK, 2/4/A1, HT29, etc.[12,13] The most

popular of them is the Caco-2 cell monolayer which is derived from human epithelial colon adenocarcinoma and retains many morphological and functional properties of the intestinal enterocytes.[14-16] Thus, the Caco-2 cell monolayer assay provides information about the drug absorption potential at near physiological conditions. However, its use is often limited due to the long membrane growth cycle and high costs.[17,18]

Kansy *et al.*[19] introduced PAMPA as an alternative to cell line permeability studies for high-throughput screenings. The method rapidly became widely used for the estimation of permeabilities. It is easily automated in *in vitro* drug absorption assays and is based on the use of a filter-immobilized artificial lipid (phosphatidylcholine) membrane.[20] Several experimental conditions (different membrane lipid compositions, multiple pH measurements and co-solvent, such as DMSO) have been proposed for the determination of artificial membrane permeability values.[21-24]

Galinis–Luciani *et al.*[25] noted that partition coefficients log P_{oct} or log D_{oct} are as good as PAMPA for estimating drug bioavailability. Alternatively, Avdeef *et al.*[26] discussed the advantages of PAMPA compared to log P_{oct} and the preciseness of PAMPA when the drug is administrated by transporter or efflux processes, or if drugs follow a paracellular diffusion route. They also noted that the Galinis–Luciani *et al.* study had limitations. One of the conclusions by Avdeef *et al.* was that since most of the drugs, and therefore also drug candidates, follow a passive trans-cellular diffusion, then PAMPA provides a correct estimate for those cell membrane permeabilities. In addition, it has been noted that experimental measurements of PAMPA can be easier to measure than log P_{oct}, especially with sparingly-soluble compounds, which a considerable number of drugs are. Avdeef reviewed the historical development of permeability and highlighted current topics in experimental studies of artificial membranes.[12]

Repellent activity. Repellents are chemicals that affect insects and other organisms by disrupting their natural behavior of blood-seeking through biting of humans and animals, and are the first line of defense that can be readily used for this purpose. Insects are believed to detect repellents through receptor uptake of molecules with specific chemical characteristics.[27-30] The best overall standard repellent is *N*,*N*-dimethyl-*m*-toluamide (DEET), systematically named *N*,*N*-diethyl-3-methylbenzamide.[31] The ideal repellent compound would prevent bites from a broad range of arthropod species, remain effective for at least 8 h, causing no irritation to the skin or mucous membranes, possess no systemic toxicity or plasticizing effect, be resistant to abrasion and rub off, and be totally greaseless and odorless.[32]

Repellent activity is expressed in protection time (PT). Different experimental protocols can be followed to measure PT, one example is where PT is determined by applying a test compound at a dose of 1 mg/cm^2 onto the external surface of a human fist followed by exposure to 200 females (5–7 days old) of the day-biter mosquito *Aedes aegypti* for 5 min every 30 min. The PT is defined

as the period of protection offered at given doses until two consecutive bites are obtained at a 30-min interval. The reported protection times represent the average of multiple determinations.[33]

Many authors have noticed that physical properties such as vapor pressure or boiling point are the only parameters that anyone has been able to correlate with repellent activity. It has been well recognized that repellents must be volatile since repellents affects the olfactory chemosensilla of the mosquito. Protection time decreases if repellents are either too volatile or too nonvolatile. If the vapor concentration of the repellent decreases below the minimum repellent concentration, a rapid loss of repellency will result. On the other hand, if a compound is not volatile, it will never come into contact with the olfactory organ.[34-36]

1.2 Molecular descriptors

A comprehensive description of molecular descriptors is given by R. Todeschini: "*The molecular descriptor is the final result of a logical and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into an useful number or the results of some standardized experiment*".[37] According to this, a molecular descriptor can be theoretically or experimentally derived. Experimental descriptors are nowadays not so widely used because they require time consuming and expensive experiments and are not simply available for vast number of chemicals. The use of theoretical molecular descriptors is more popular because they do not need those experiments. They are calculated directly from molecular structure. Thanks to highly improved computer technology, a calculation of such descriptors is generally very fast. These descriptors can even be used for screening of millions compounds within reasonable time.

At current stage, more than 3000 theoretical descriptors[38,39] are available. They cover 1D, 2D, 3D, 4D, quantum chemical, fingerprint-based, field-based, surface area related, chirality related, etc. type of descriptors. These descriptors have different theoretical background starting from simple counts, graph theory and ending with more time-consuming quantum-chemical calculations. Therefore, descriptors can be divided into groups according to their nature. Although the classifications are arbitrary, the following list covers the descriptors that are used in the publications within the current Thesis:[40,41]

- Constitutional – include count of atoms, count of groups, molecular weight, etc.
- Topological – include different graph theory based indices starting from Wiener index up to E-state indices.
- Geometrical – include molecular surface area and volume, gravitational indices, shadow areas, etc.

- Electrostatic – include atomic partial charges, polarity parameters, polarizabilities, etc.
- CPSA – include combinations of charged partial surface areas.
- Quantum-chemical – include total energy of the molecule, particles repulsion and attraction energies, etc.
- Molecular orbital (MO)-related – include HOMO and LUMO energies, reactivity indices, free valences, bond orders, etc.
- Thermodynamic – include vibrational and translational enthalpies and entropies, heat capacities, etc.

The CODESSA family software packages calculate approximately 600 base descriptors, which are extended up to more than 1000, depending on the chemical constitution of compounds. Such a vast number of descriptors need robust statistical and mathematical approaches for modeling. Two of those methods, which were used in current thesis, will be overviewed in the next sections.

1.3 Principal Component Analysis

Principal Component Analysis (PCA) is a multivariate data reduction and exploratory data analysis method. The method was first described by the statistician Karl Pearson.[42] PCA determines dimensions with maximum variation that are orthogonal with each other. These dimensions are called latent variables or components. Figure 1 illustrates the simplest situation with two variables (x1 and x2).

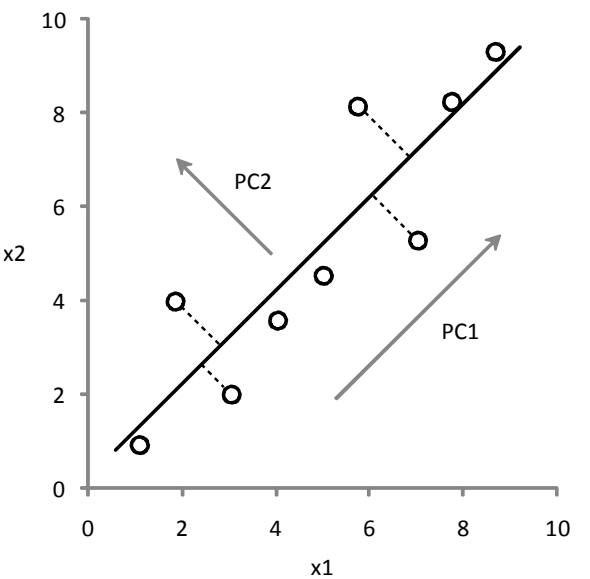

Figure 1. Variables x1 and x2.

The first component (PC1) comprises the maximum common variance. The next component is orthogonal with the first one and comprises the rest of the variation. The projection of new latent variables (components) is given in Figure 2. As can be seen, the variance of PC1 is larger than the variance of the original variables, and controversially, the variance of PC2 is smaller.

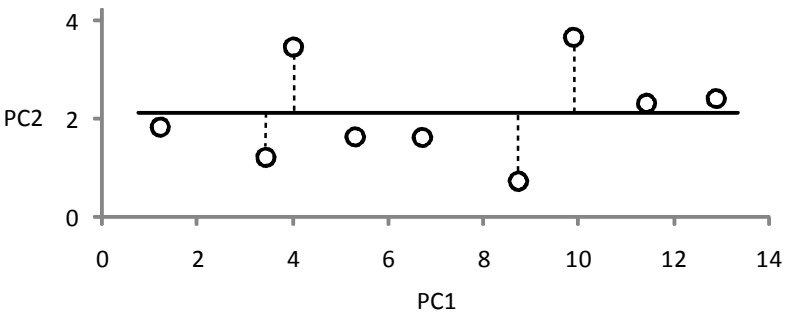

Figure 2. Latent variables PC1 and PC2.

In real situations, where PCA is used, the data matrix consists of a lot more variables. Therefore, such a simple representation cannot be always available. The general form of PCA is given in eq. 3.

$$\mathbf{D} = \mathbf{T} \cdot \mathbf{P} \tag{3}$$

Where D is data matrix, T and P are score and loading matrices, respectively. The given equation is true when the number of components is equal to the number of variables. Generally, fewer components are used and in this case the error term needs to be added. The number of PCs (scores, loadings) existing in the characteristic vector space can be equal to, or less than, the number of variables in the data set. The first principal component is defined as that giving the largest contribution to the respective PCA of linear relationship exhibited in the data. The second component may be considered as the second best linear combination of variables that accounts for the maximum possible of the residual variance after the effect of the first component is removed from the data. Subsequent components are defined similarly until practically all the variance in the data is exhausted. A more detailed explanation of PCA methodology can be found in the following textbooks.[43-46] In PCA, the data is commonly pre-processed to provide all the scales with equal weight, mostly *via* the unit variance scaling method, where the data are standardized, centralized and normalized using their sample standard deviation, variance and mean. This is necessary when the variables differ a lot in their variances. When the variables have the same dimension such preprocessing can be avoided because pre-processing of the data could result in some cases in a loss of information as well as decreasing the sensitivity of the PCA.[45,47,48]

PCA reveals internal relations between characteristics of a class of compounds (objects) and hence enables drastic reduction of the dimensionality of the original raw data. This reduction is achieved by transforming to a new set of variables, the principal components, which are uncorrelated, and which are ordered so that the first few, with descending importance, retain most of the variation in the total set of original variables. PCA can be highly useful for data classification and pattern recognition. In the two-dimensional plotting of score vectors, observations with similar characteristics are clustered. In the two-dimensional plotting of loading vectors, the initial variables reflected in those score loadings are clustered.

Also multilevel PCA techniques – hierarchical and multiblock PCA[49,50] are proposed. In these methods an original data matrix is divided into submatrices according to additional information that allows to group variables and/or objects. Common PCA is applied to the sub-matrices and the resulting components are further used in the PCA which now comprises the relationships between all objects and variables. This methodology is specifically useful for very large and diverse data sets. It simplifies the interpretation of the final PCA model.

PCA is one of the best known multivariate exploratory techniques extensively used in different areas of chemistry[51-55] or other disciplines as well, such as biology, physiology, psychology, color technology, etc.[56-60]

1.4 Quantitative Structure Property/Activity Relationships

The Encyclopedia of Computational Chemistry defines the following: "*Quantitative structure–property relationship or quantitative structure–activity relationship studies probe connections between molecular structure of organic compounds and their chemical or biological properties*".[61] QSPR/QSAR, as it is known currently, has been used nearly 50 years. C. Hansch and co-workers' study where biological activity of pesticides was correlated with the octanol–water partition coefficient[62] is considered to be the first work to introduce this new paradigm. Sometimes such regression is called Hansch analysis. However, the first use of prediction was likely the work of E. J. Mills where the correlation of melting and boiling points in homologous series is presented.[63]

A QSPR/QSAR modeling workflow generally involves three steps (see Figure 3):

(i) collection and preparation of data including collection of property or activity endpoint values and design of training and test sets,
(ii) calculation and collection of descriptive variables (descriptors),
(iii) selection of descriptors that possess relationship with property or biological activity and application of statistical methods that correlate changes in structure with changes among chemical property or activity.[64]

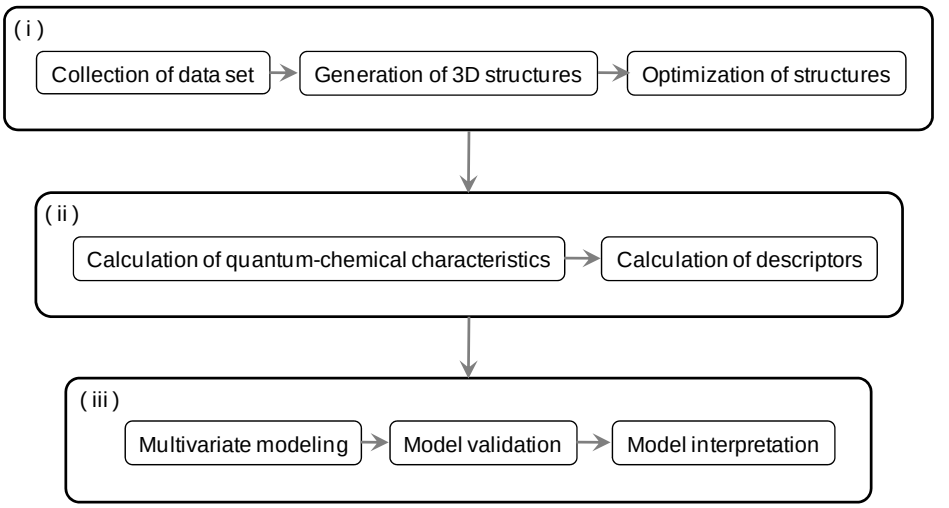

Figure 3. QSPR/QSAR modeling workflow.

The best choice to collect the data is to design a structurally important set of compounds and then to measure their property or activity by one laboratory following standardized methodology. This is possible very seldom and therefore the data sets are often gathered from literature, sometimes merged from different authors. The latter is dangerous because the experimental values for the same compound can differ among different laboratories. This is mostly due to different experimental conditions. For example, different equipment will produce different systematic errors. Even more, different experimental protocols can be followed. Thus, while merging a data from different sources, the values should be carefully analyzed.[65] This is more important for biological activities rather than for chemical properties since the experimental errors of chemical measurements are generally smaller than biological assays.

3D structures can be either drawn from scratch using available software (MDL ISIS Draw,[66] ACD/ChemSketch,[67] ChemDraw,[68] MarvinSketch,[69] etc.) or obtained from available databases (ChemIDPlus,[70] PubChem,[71] etc.). Optimizations of the structures are followed by minimizing the total energy. The procedure is carried out in two steps, where first, a more robust molecular mechanics force field (MM+,[72] MMFFs,[73] AMBER,[74] etc.) is used and in a second step, fine minimization is carried out with semi-empirical parameterizations (AM1,[75] PM3,[76] etc.).

Quantum-chemical characteristics, such as energy and charge distributions, are generally calculated using the same force fields as in the last step of optimization procedure. In the publications of the current Thesis, the MOPAC[77] software package was used with AM1 parameterization. Several other molecular modeling packages also exist, such as AMPAC,[78] HyperChem,[79] Gaussian,[80] Schrödinger Suite,[81] etc.

As discussed previously in the chapter on descriptors, there exist a vast number of theoretical descriptors. Therefore, several calculation packages are available, for example QikProp,[82] MARVIN,[83] DRAGON,[84] TSAR,[85] etc. In publications of the current Thesis, the CODESSA family,[40,41,86] software were used where constitutional, topological and geometrical descriptors are derived from the optimized 3D structure and electrostatic, CPSA, quantum-chemical, MO-related and thermodynamic descriptors are derived from calculated quantum-chemical characteristics.

Such a large number of descriptors (from hundreds up to thousands) need a robust method for variable (descriptor) selection. The most common selection methods in regressions are forward, backward and stepwise selections.[87,88] In publications of the current Thesis, the Best Multiple Linear Regression (BMLR) is used which includes a modified stepwise forward selection method.[89,90] Also, other sophisticated methods (PLS, PCA, Genetic Algorithms) can be used for the selection of descriptors. A detailed explanation of the methods is provided in selected books[88,91] and their comparisons in recent review publications.[92,93]

The most popular regression method in QSPR/QSAR is multivariate linear regression[94,95] for relating the descriptors to the property producing the equation:

$$y_i = a_0 + \sum_k a_k x_{ik} + \varepsilon_i \qquad (4)$$

where y is a property or activity (dependent variable), x is a descriptor (independent variable), a is a regression coefficient, ε is the random error, and subscripts i and k denote the number of objects (compounds) and descriptors, respectively. A variety of other linear and nonlinear regression methods such as PLS, PCR, ANN[44,46,91,96] are also used in the context of QSPR/QSAR.

Commonly in QSPR/QSAR the property under the study is logarithmically transformed. There are two reasons for this. First, raw data mostly are not normally distributed (following Gaussian distribution) and generally, after logarithmic transformation the distribution is closer to normal.[97] Together with data distribution, the error distribution is normalized as well. Another reason is that quantum chemically calculated energy related descriptors (such as HOMO, LUMO, etc.) have exponential relationships with concentrations. Therefore is necessary to take the logarithm of the property to assure linear relationships. It is also important to represent the concentrations in molar units, not in weight units, since the descriptors are calculated based on molecular, not any mass unit.[98]

QSPR/QSAR is successfully applied to chemical properties and biological activities as discussed comprehensively in the following review publications.[89,99-105]

1.5 Model validation

Regression models need always a proper validation. This is important for the estimation of correct predictions. Goodness of fit is determined by the coefficient of determination (R^2). But this does not show much about the capability of prediction. Two types of validations are used in QSPR/QSAR, called internal and external validation.

The most proper validation is external validation where squared correlation coefficient (R^2) is calculated from observations (compounds in QSPR/QSAR) which were not used in model development. A representative way to achieve this is to order observations according to their experimental values and sequentially (every second or every third, for example) moving observations to the external validation set. This assures equal distribution of property variation between training and testing sets. External validation is not always possible. For instance, with very limited data it is usually better to use all available experimental material for the model training in order to obtain more reliable models. This extends also to cases where QSPR/QSAR models are derived rather for analysis of property, not for the prediction. In such cases an internal validation is commonly used.

The Leave One Out (LOO) and Leave Many Out (LMO) internal validations are most commonly used, denoted as R^2_{LOO} and R^2_{LMO}, respectively.[106-108] Sometimes these statistics are called also cross-validation and denoted then as R^2_{CV} or Q^2. In a cross validation procedure, one (LOO) or many (LMO) observations are left out from the set and the model is derived using the same variables. Further, this model is used to predict the values for those observations which were left out. This is repeated several times and a sort of external testing set is derived. Recently, some modifications of cross validation were proposed and discussed.[109,110]

In addition, other internal validation methods are used, such as *Y*-scrambling, bootstrapping, etc. Advantages and disadvantages of these methods along with external and cross-validations are discussed and analyzed in the context of QSPR/QSAR in excellent review publications.[97,106,111-114]

2. SUMMARY OF ORIGINAL PUBLICATIONS

2.1 Prediction, description and analysis of solubility using Ostwald solubility coefficients

Articles I and II are third and forth in the publication series dedicated to predict, describe and analyze solubility. Briefly, the study of solubility has been conducted using a combination of QSPR and PCA methods. A solubility database of about 4,500 experimental data points was used that gathered available experimental data into a matrix of approximately 150 solvents times approximately 390 solutes. Methodology was developed in which QSPR and PCA are combined in order to predict the missing values and to fully fill the data matrix. The solubility is expressed as the logarithm of Ostwald solubility coefficients (log *L*). Article I includes the complete strategy of the research performed divided into in five steps (see Figure 4 for simplified scheme).

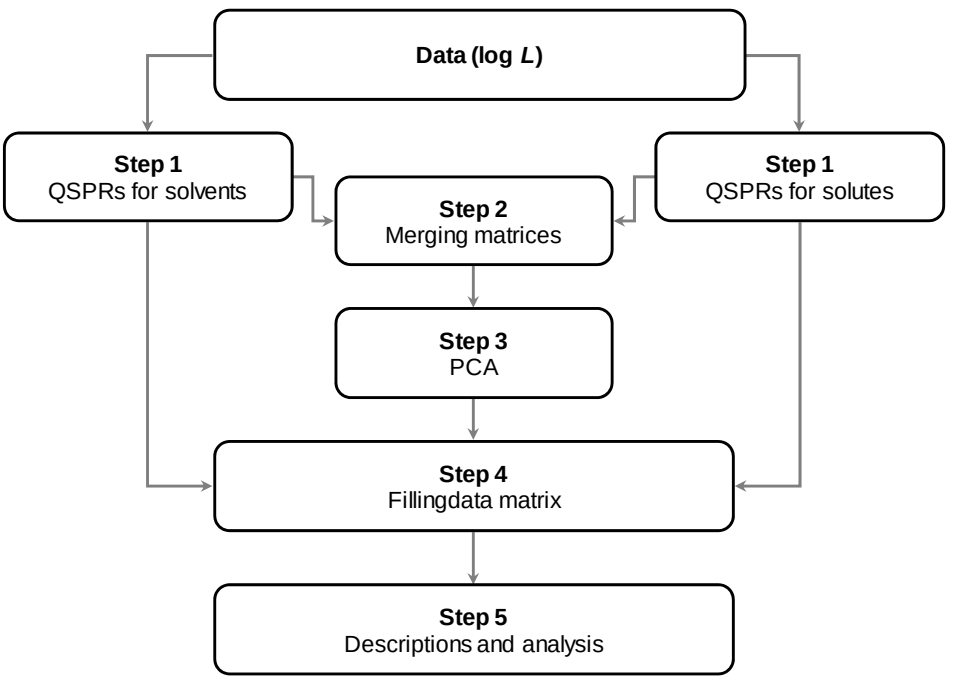

Figure 4. Workflow of solubility study.

QSPR models for solubility (Step 1). Rows and column of data are ordered according to the number of data points that they contain, so the densest area of the data matrix is located in its upper left corner. QSPR models were developed for the densest area of 87 solvents and 91 solutes series for which at least 15

experimental solubility values were available. QSPR models were developed and used to predict the missing values of log *L* for filling empty gaps in the matrix densest area. Considering all the QSPR models of solvent series developed (87), their summarized R^2 is 0.957 with 4,167 data points. The same statistic for solute models (91) is 0.996 with 3,394 data points.

Merging predictions (Step 2). The predictions in Step 1 resulted in two matrices, one based on the solvent series and another based on the solute series. These matrices were merged using several developed rules: (i) the prediction range is defined as 15% of the distribution range of the experimental data points for each model considered, (ii) if the predicted value was in the range in both matrices, a model-weighted average of the two values were calculated, (iii) if the predicted solubility value from the solvent model was out of range of experimental values, then the value predicted from the solute model was taken and *vice versa*, and (iv) if the predicted value was out of the range for both solute and solvent models, then the solvents were ordered according to the ET30 polarity scale and weighted averages of neighboring values were calculated. During the study, an additional 289 experimental solubility values were collected. These values were used for external validation, resulting in an R^2 of 0.882.

PCA–QSPR combined methodology for predicting solubility (Step 3). PCA was applied on the densest area matrix and three components were considered covering 96% of total variation. QSPR models were developed for scores, loadings, standard deviation and mean. These QSPR models were extrapolated for the rest of the solvents and solutes in the whole data matrix and so called "Backward Procedure" of PCA was used to calculate missing log *L* values.

Filling the data matrix (Step 4). The QSPR models developed in Step 1 also allowed predictions of missing log *L* values in the whole matrix. Therefore, the empty gaps in the matrix were filled using either solvent/solutes QSPR models where the prediction was in range, or with the PCA–QSPR "Backward Procedure" combined methodology developed in Step 3. Finally, the data matrix comprises 154 solvents and 397 solutes and consists of 4,540 experimental and 56,598 predicted log *L* values.

Description and analysis of the PCA model (Step 5). In article II, the previously obtained matrix was refined and updated with 1,285 new experimental data points. These new data allowed an external validation of previous predictions which resulted in an R^2_{ext} of 0.59. Statistical outliers were analyzed and after removing 75 outliers, the R^2_{ext} is 0.88. PCA was applied on the data matrix without data preprocessing. Outliers of the PCA model were analyzed and they were excluded. The same outliers appeared also during the prevoius external validation. The finally obtained two-component PCA model describes 98.6% of total variability. The physical meaning of the respective scores and loadings was

analyzed *via* construction of QSPR models. The optimal model for the first score consisted of two molecular descriptors: *Gravitational index (all bonds)* and *HA dependent HDCA-1* and the resulting R^2 is 0.96. These two descriptors are related to cavity formation and hydrogen bonding terms in eq. 2, given in introductory section. The model for the second score comprises mostly electrostatic interaction term related descriptors and resulted in an R^2 of 0.91. QSPR analysis indicates that the principal components describe multiple solubility interactions rather than a single solute–solvent interaction. The first component represents cavity formation and HB interactions which can be codified by the gravitational index and hydrogen donor charged surface area molecular descriptors. The second component covers weaker and more specific electrostatic interaction types. And finally the detailed analysis of the pattern observed in the score plot provides a detailed explanation for each chemical group.

2.2 QSAR for Permeability in Artificial Membranes

Article III presents the QSAR modelling of PAMPA. Permeability studies are used to describe the intestinal absorption of drugs. They are crucial components in an ADME-Tox profile for estimating the oral bioavailability of drugs and are particularly needed to predict compounds' permeability early in the drug design process. The artificial membrane is usually formed from phospholipids (lecithin) and it exhibits only passive permeability, while in the case of natural cell lines, the active transport is also accounted for. PAMPA became an alternative to cell line permeability studies for high-throughput screenings.

A structurally diverse data set comprising 22 peptidic compounds and 38 commercially available drugs was studied. Forward selection (BMLR) of descriptors from a large set of molecular descriptors was used to derive QSAR models. The best 5-parameter model R^2 is 0.71, R^2_{CV} is 0.63 and R^2_{ext} is 0.71. Diagnostics of the model represented by a Williams plot revealed two compounds with high residuals but low leverages and another two compounds with high leverages but low residuals. The descriptors in the model comprise hydrogen bonding ability, which is solely connected with water solubility, hence hydrophilicity. In addition, electrostatic interactions, charge distributions, polarity and polarizability, and the shape of a molecule are described by the descriptors.

It is known that the intestinal permeability mechanism depends on the physicochemical properties of the absorbed compound, such as its stereochemistry, partition into membranes, molecular weight and/or size, molecular volume, pK_a, solubility, chemical stability and charge distribution. Simplifying the system only to passive diffusion gives an opportunity to study permeability in a distinct way. However, most of the drugs permeate passively. A particular QSAR model provides insight into the structural parameters that determine the mechanism of permeability. Based on the model, one can conclude what structural parameters, parts of the molecule, need to be changed in order to enhance a particular compound's property, passive permeability in our case.

2.3 Analysis of Mosquito Repellent Protection Time

In article IV, the mosquito repellent protection time (PT) of 31 benzamide and cyclohexamide derivates was modeled (Figure 5). Two 4-parameter QSAR models were developed for the description of mosquito repellent PT with satisfactory statistical characteristics. The first model is based on theoretical descriptors and the result R^2 is 0.78. The second model includes the squared logarithmic vapor pressure and resulted in an R^2 of 0.80. Descriptors involved in the models were related to repellent activity through three main molecular interactions:

(i) vaporization is connected to the duration of time that a mosquito can have contact with the repellent,
(ii) structural fit on an unknown active receptor center,
(iii) chemical reaction with a receptor resulting in the act of repelling.

Figure 5. Base structures of repellents.

The successful QSAR models of the study suggested that a general QSAR treatment of repellents could be of great benefit in synthetic efforts while discovering better compounds for practical use. The study was developed further and excellent results with new promising repellents were obtained.[115]

2.4 Concluding remarks

Solubility as a fundamental property in chemistry and biology was studied using PCA and QSPR methodologies. Such combination of these methods as a new approach was described and a general scheme was provided. The study revealed important structural characteristics influencing solubility and their relationships with solvation free energy terms was discussed. Pattern analysis revealed two general types of interactions. Also, the pharmacologically important oral bioavailability of drugs was studied. A resulting QSAR model provides insight into the structural parameters that influence and determine the mechanism of permeability. Mosquito repellent activity was studied using QSAR methodology. Three important mechanistic characteristics were discussed.

REFERENCES

1 Fogg, P.G. T.; Bligh, S.W.A.; Derrick, M.E.; Yampolskii, Y.P.; Clever, H.L.; Skrzecz, A.; Young, C.L. IUPAC-NIST Solubility Data Series. 76. Solubility of Ethyne in Liquids. *J. Phys. Chem. Ref. Data.* **2002**, *30*, 1693–1875.

2 Tomkins, R.P.T., Hefter G.T., Eds, *The Experimental Determination of Solubilities, Wiley Series in Solution Chemistry*, John Wiley & Sons: Chichester, England, 2003.

3 Katritzky, A.R.; Oliferenko, A.A.; Oliferenko, P.V.; Petrukhin, P.; Tatham, D.B.; Maran, U.; Lomaka, A.; Acree, W.E. Jr. A General Treatment of Solubility. Part 1. The QSPR Correlation of Solvation Free Energies of Single Solutes in Series Solvents. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1794–1805.

4 Reichardt, C. *Solvents and Solvent Effects in Organic Chemistry*, 3rd ed.; Wiley-VCH: New York, 2003.

5 Tomasi, J.; Persico, M. Molecular Interactions in Solution: An Overview of Methods Based on Continuous Distributions of the Solvent. *Chem. Rev.* **1994**, *94*, 2027–2094.

6 Karelson, M. Quantum Chemical Treatment of Molecules in Condensed Disordered Media. *Adv. Quantum Chem.* **1997**, *28*, 141–157.

7 Vitha, M.; Carr, P.W. The chemical interpretation and practice of linear solvation energy relationships in chromatography. *J. Chrom. A* **2006**, *1–2*, 143–194.

8 Drago, R.S. Solvation. In *Applications of Electrostatic-Covalent Models in Chemistry*, Surfside Scientific Publishers, University of Florida, Gainesville, 1994.

9 Dutt, G.B. Molecular Rotation as a Tool for Exploring Specific Solute–Solvent Interactions. *Chem. Phys. Chem.* **2005**, *6*, 413–418.

10 Desiraju, G.R. Hydrogen Bridges in Crystal Engineering: Interactions without Borders. *Acc. Chem. Res.* **2002**, *35*, 565–573.

11 Dearden, J.C. In Silico Prediction of aqueous solubility. *Expert Opin. Drug Discov.* **2006**, *1*, 31–52.

12 van de Waterbeemd, H.; Lennernäs, H.; Artursson, P. *Drug Bioavailability. Estimation of Solubility, Permeability, Absorption and Bioavailability*, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2003.

13 Ehrhardt, C.; Kim, K-J. *Drug Absorption Studies. In Situ, In Vitro and In Silico Models*, Springer, New York, 2008.

14 Bohets, H.; Annaert, P.; Mannens, G.; van Beijsterveldt, L.; Anciaux, K.; Verboven, P.; Meuldermans, W.; Lavrijsen, K. Strategies for Absorption Screening in Drug Discovery and Development . *Curr. Top. Med. Chem.* **2001**, *1*, 367–383.

15 Hidalgo, I.J.; Raub, T.J.; Borchardt, R.T. Characterization of the human colon carcinoma cell line (Caco-2) as a model system for intestinal epithelial permeability. *Gastroenterology* **1989**, *96*, 736–749.

16 Hilgers, A.R.; Conradi, R.A.; Burton, P.S. Caco-2 Cell Monolayers as a Model for Drug Transport Across the Intestinal Mucosa. *Pharm. Res.* **1990**, *7*, 902–910.

17 Bravo, S.A.; Nielsen, C.U.; Amstrup, J.; Frokjaer S.; Brodin, B. In-depth evaluation of Gly-Sar transport parameters as a function of culture time in the Caco-2 cell model. *Eur. J. Pharm. Sci.* **2004**, *21*, 77–86.

18 Anderle, P.; Niederer, E.; Rubas, W.; Hilgendorf, C.; Spahn-Langguth, H.; Wunderli-allenspach, H.; Merkle, H. P.; Langguth, P. P-glycoprotein (P-gp)

mediated efflux in Caco-2 cell monolayers: The influence of culturing conditions and drug exposure on P-gp expression levels. *J. Pharm. Sci.* **1998**, *87*, 757–762.
19 Kansy, M; Senner, F; Gubermator, K. Physicochemical high throughput screening: parallel artificial membrane permeation assay in the description of passive absorption processes. *J. Med. Chem.* **1998**, *41*, 1007–1010.
20 Thompson, M.; Krull, U.J., Worsfold, P.J. The structure and electrochemical properties of a polymer-supported lipid biosensor. *Anal. Chim. Acta* **1980**, *117*, 133–145.
21 Veber, D.F.; Johnson, S.R.; Cheng, H.Y.; Smith, B.R.; Ward, K.W.; Kopple, K.D. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **2002**, *45*, 2615–2623.
22 Kerns, E.H.; Di, L.; Petuskey, S.; Farris, M.; Ley, R.; Jupp, P. Combined application of parallel artificial membrane permeability assay and Caco-2 permeability assays in drug discovery. *J. Pharm. Sci.* **2004**, *93*, 1440–1453.
23 Wohnsland, F.; Faller, B. High-throughput permeability pH profile and high-throughput alkane/water log P with artificial membranes. *J. Med. Chem.* **2001**, *44*, 923–930.
24 Avdeef, A.; Testa, B. Physicochemical profiling in drug research: a brief survey of the state-of-the-art of experimental techniques. *Cell. Mol. Life Sci.* **2002**, *59*, 1681–1689.
25 Galinis-Luciani, D.; Nguyen, L.; Yazdanian, M. Is PAMPA a useful tool for discovery? *J. Pharm. Sci.* **2007**, *96*, 2886–2892.
26 Avdeef, A.; Bendels, S.; Di, L.; Faller, B.; Kansy, M.; Sugano, K.; Yamauchi, Y. PAMPA — Critical Factors for Better Predictions of Absorption. *J. Pharm. Sci.* **2007**, *96*, 2893–2909.
27 Bowen, M.F.; Davis, E.E.; Romo, J.; Haggart, D. Lactic acid sensitive receptors in the autogenous mosquito Aedes atropalpus. *J. Insect Physiol.* **1994**, *40*, 611–615.
28 Klun, J.A. ; Schmidt, W.F.; Debboun, M. Stereochemical effects in an insect repellent. *J. Med. Entomol.* **2001**, *38*, 809–812.
29 Nolen, J.A.; Bedoukian, R.H.; Maloney, R.E.; Kline, D.L. Method, apparatus and compositions for inhibiting the human scent tracking ability of mosquitoes in environmentally defined three dimensional spaces. 2002, US Patent 6,362,235.
30 Carlson, D.A.; Smith, N.; Gouck, H.K.; Godwin, D.R. Yellow fever mosquitoes: Compounds related to lactic acid that attract females. *J. Econ. Entomol.* **1973**, *66*, 329–331.
31 Plimmer, J. Ed. *Encyclopedia of Agrochemicals*, John Wiley and Sons, New York, 2003.
32 Fradin, M.S. Mosquitoes and mosquito repellents: a clinician's guide. *Ann. Intern. Med.* **1998**, *128*, 931–940.
33 Suryanarayana, M.; Pandey, K.; Prakash, S.; Raghuveeran, C.; Dangi, R.; Swamy, R.; Rao, K. Structure-activity relationship studies with mosquito repellent amides. *J. Pharm. Sci.* **1991**, *80*, 1055–1057.
34 Davis, E.E. Insect repellents: concepts of their mode of action relative to potential sensory mechanisms in mosquitoes (Diptera: Culicidae). *J. Med. Entomol.* **1985**, *22*, 237–243.
35 Ma, D.; Bhattacharjee, A.; Gupta, R.; Karle, J. Predicting mosquito repellent potency of N,N-diethyl-m-toluamide (DEET) analogs from molecular electronic properties. *Am. J. Trop. Med. Hyg.* **1999**, *60*, 1–6.

36 Peterson, C.; Coast, J. Insect repellents – past, present and future. *Pesticide Outlook* **2001**, 154–158.
37 Todeschini, R.; Consonni, V. *Molecular Descriptors for Chemoinformatics*, Wiley-VCH, Weinheim, Germany, 2009.
38 Gramatica, P. A short history of QSAR evolution. September, 2008, www.qsarworld.com/qsar-archives (last accessed on May 2010)
39 Gramatica, P. Chemometric Methods and Theoretical Molecular Descriptors in Predictive QSAR Modeling of the Environmental Behaviour of Organic Pollutants. In *Recent Advances in QSAR Studies: Methods and Applications* eds. Puzyn, T., Leszczynski, J., Cronin, M.T.D., Springer Science+Business Media B.V., 2010, pp. 327–366.
40 Katritzky, A.R.; Lobanov, V.S.; Karelson, M. *CODESSA: Reference manual (version 2.0)*, Gainesville, Florida, 1994.
41 *CodessaPro*; University of Florida, Gainesville, FL, USA (www.codessa-pro.com)
42 Pearson, K. On Lines and Planes of Closest Fit to Systems of Points in Space. *Philos. Mag.* **1901**, *2*, 559–572.
43 Johnson, R.A.; Wichern, D.W. Principal Components. In *Applied Multivariate Statistical Analysis*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1982; pp 361–400.
44 Wold, S.; Albano, C.; Dunn III, W.J.; Edlund, U.; Esbensen, K.; Geladi, P.; Hellberg, S.; Johansson, E.; Lindberg, W.; Sjöström, M. Multivariate Data Analysis in Chemistry. In *Chemometrics: Mathematics and Statistics in Chemistry*, *NATO ASI Series, ser. C Vol. 138*; Kowalski, B. R., Ed., Riedel Publishing, Dordrecht, 1984; pp. 17–95
45 Gempeline, P.J. Principal Component Analysis. In *Practical Guide to Chemometrics*. 2nd ed.; Gempeline Ed.; CRC Press, Taylor & Francis Group, Boca Raton, 2006; pp 69–104.
46 Electronic Statistics Textbook. StatSoft, Inc., Tulsa, 2007, www.statsoft.com/textbook/stathome.html (last accessed June 2010)
47 Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; Wold, S. *Multi- and Megavariate Data Analysis*, Umetrics AB, Umeå, 2001.
48 StatSoft, Inc. Electronic Statistics Textbook. Tulsa, OK: StatSoft. WEB: http://www.statsoft.com/textbook (last accessed June 2010).
49 Wold, S.; Kettaneh, N.; Tjessem, K. Hierarchical multiblock PLS and PC models for easier model interpretation and as an alternative to variable selection. *J. Chemom.* **1996**, *10*, 463–482.
50 Westerhuis, J.A.; Kourti, T.; MacGregor, J.F. Analysis of multiblock and hierarchical PCA and PLS models, *J. Chemom.* **1998**, *12*, 301–321.
51 Dunn, W.J., III; Koehler, M.G.; Grigoras, S. The role of solventaccessible surface area in determining partition coefficients. *J. Med. Chem.* **1987**, *30*, 1121–1126.
52 Castells, C.B.; Reta, M.R. Study of gas-liquid partitioning of alkane solutes in several organic solvents by using principal analysis and linear solvation energy relationships. *Anal. Chim. Acta* **2003**, *488*, 107- 122.
53 Malinowski, E.R.; Howery, D.G. Factor Analysis in Chemistry; Wiley-Interscience: 1980.
54 Strouf, O. Chemical Pattern Recognition; Wiley: New York, 1986.
55 Meloun, M.; Militky, M.; Forina, M. Chemometrics in Analytical Chemistry; Ellis Horwood: New York, 1992.

56 Cunningham, M.J. Genomics and proteomics. The new millennium of drug discovery and development. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 291–300.
57 Tzeng, D.Y.; Berns, R.S. A review of principal component analysis and its applications to color technology, *Color Res. Appl.* **2005**, *30*, 84–98.
58 Kayser, J.; Tenke, C.E.; Kroppmann, C.J.; Fekri, S.; Alschuler, D.M.; Gates, N.A.; Gil, R.; Harkavy-Friedman, J.M.; Jarskog, L.F.; Bruder, Gerard E. Current source density (CSD) old/new effects during recognition memory for words and faces in schizophrenia and in healthy adults. *Int. J. Psychophysiol*. **2010**, *75*, 194–210.
59 Kirby, K.N.; Finch, J.C. The hierarchical structure of self-reported impulsivity, Pers. *Indiv. Differ*. **2010**, *48*, 704–713.
60 Serrano, Á.; de Diego, I.M.; Conde, C.; Cabello, E. Recent advances in face biometrics with Gabor wavelets: A review, *Pattern Recognit. Lett.* **2010**, *31*, 372–381.
61 Schleyer, P.v.R; Allinger, N.L.; Clark, T.; Gasteiger, J.; Kollman, P.A.; Schaefer III, H.F.; Schreiner, P.R., Eds.; *Encyclopedia of Computational Chemistry*. John Wiley & Sons: Chichester, UK, 1998
62 Hansch, C.; Maloney, P.P.; Fujita, T.; Muir, R.M. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature* **1962**, *194*, 178–180.
63 Mills, E.J. On melting point and boiling point as related to composition, *Phil Mag*. 1884, 17, 173–187.
64 Perkins, R.; Fang, H.; Tong, W.; Welsh, W.J. Quantitative structure–activity relationship methods: Perspectives on drug discovery and toxicology. *Environ. Toxicol. Chem.* **2003**, *22*, 1666–1679.
65 Cronin, M.T.D.; Schultz, T.W. Pitfalls in QSAR. *J. Mol. Struct.-Theochem.* **2003**, *622*, 39–51.
66 MDL ISIS Draw; Symyx Technologies, Inc., 3100 Central Expressway, Santa Clara, CA, USA. (www.symyx.com)
67 ACD/ChemSketch; Advanced Chemistry Development, Inc., 33 Richmond St. West, Suite 605, Toronto, ON, Canada. (www.acdlabs.com)
68 ChemDraw; CambridgeSoft Corporation, 100 CambridgePark Drive, Cambridge, MA, USA. (www.cambridgesoft.com)
69 MarvinSketch; ChemAxon Kft., Máramaros köz 3/a, Budapest, Hungary (www.chemaxon.com)
70 U.S. National Library of Medicine, 2009 U.S. National Library of Medicine, 2009. ChemIDplusnext Advanced. http://chem.sis.nlm.nih.gov/chemidplus/ (last accessed in June 2010).
71 Pubchem. http://pubchem.ncbi.nlm.nih.gov (last accessed June 2010).
72 Hocquet, A.; Langgård, M. An Evaluation of the MM+ Force Field, *J. Mol. Model.* **1998**, *4*, 94–112.
73 Halgren, T. A. MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem*. **1999**, *20*, 720–729.
74 Weiner, S.J.; Kollman, P.A.; Case, D.A.; Singh, U.C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.
75 Dewar, M.J.S.; Zoebisch, E.G.; Healy, E.F.; Stewart, J.J.P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc*. **1985**, *107*, 3902–3909.

76 Stewart, J. J. P. Optimization of parameters for semiempirical methods I. Method. *J. Comp. Chem.* **1989**, *10*, 209–220.
77 Stweart, J. J. P. MOPAC Program Package 6.0. QCPE No. 455, 1990.
78 AMPAC; Semichem, Inc. 12456 W, 62nd Terrace, Suite D, Shawnee, KS, USA (www.semichem.com)
79 HyperChem; Hypercube, Inc., 1115 NW 4th St. Gainesville, FL, USA (www.hyper.com)
80 Gaussian; Gaussian, Inc., 340 Quinnipiac St Bldg 40, Wallingford, CT, USA (www.gaussian.com)
81 Schrödinger, LLC, 101 S.W. Main Street, Suite 1300, Portland, OR, USA (www.schrodinger.com)
82 QikProp; Schrödinger, LLC, 101 S.W. Main Street, Suite 1300, Portland, OR, USA (www.schrodinger.com)
83 MARVIN; Calculator Plugins, ChemAxon Kft., Budapest, Hungary, 2008
84 DRAGON; TALETE srl, Via V. Pisani, 13 – 20124, Milano, Italy (www.talete.mi.it)
85 TSAR; Accelrys, Inc., 10188 Telesis Court, Suite 100, San Diego, CA, USA (www.accelrys.com)
86 Katritzky, A.R.; Lobanov, V.S.; Karelson, M. QSPR: the correlation and quantitative prediction of chemical and physical properties from structure. *Chem. Soc. Rev.* **1995**, *24*, 279–287.
87 SYSTAT 12 Statistics I II III IV, SYSTAT Software, Inc., 1735 Technology Drive, Suite 430, San Jose, CA, USA (www.systat.com)
88 Varmuza, K.; Filzmoser, P. Calibration. In *Intoduction to Multivariate Statistical Analysis in Chemometrics*, CRC Press, Taylor & Francis Group, Boca Raton, FL, 2009, pp 103–194.
89 Karelson, M.; Maran, U.; Wang, Y.; Katritzky, A. R. QSPR and QSAR Models Derived Using Large Molecular Descriptor Spaces. A Review of CODESSA Applications. *Collect. Czech. Chem. Commun.* **1999**, *64*, 1551–1571.
90 Karelson, M. *Molecular Descriptors in QSAR/QSPR*, Wiley-Interscience: New York, 2000, pp. 396–400.
91 Kalivas, J.H.; Gemperline, P.J. Calibration. In *Practical Guide to Chemometrics*. 2nd ed.; Gemperline, P.J. Ed.; CRC Press, Taylor & Francis Group, Boca Raton, FL, 2006; pp 105–165.
92 Forina, M.; Lanteri, S.; Cerrato Oliveros, M.C.; Pizarro Millan, C. Selection of useful predictors in multivariate calibration. *Anal. Bioanal. Chem.* **2004**, *380*, 397–418.
93 Gonzalez, M.P.; Teran, C.; Saiz-Urra, L.; Teijeira, M., Variable Selection Methods in QSAR: An Overview. *Curr. Top. Med. Chem.* **2008**, *8*, 1606–1627.
94 Draper, N.R.; Smith, H. *Applied Regression Analysis*, Wiley, New York, 1966.
95 Johnson, R.A.; Wichern, D.W. Multivariate Linear Regression Models. In *Applied Multivariate Statistical Analysis*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1982; pp 291–358.
96 Zupan, J.; Gasteiger, J. *Neural Networks for Chemists: An Introduction*; VCH Verlagsgesellschaft: Weinheim, 1993.
97 Eriksson, L.; Jaworska, J.; Worth, A.P.; Cronin, M.T.D.; McDowell, R.M.; Gramatica, P. Methods for reliability and uncertainty assessment and for applicability

evaluations of classification- and regression-based QSARs. *Environ. Health Perspect.* **2003**, *111*, 1361–1375.
98 Dearden, J.C.; Cronin, M.T.D.; Kaiser, K.L.E. How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR). *SAR QSAR Environ. Res.* **2009**, *20*, 241–266.
99 Katritzky, A.R.; Maran, U.; Lobanov, V.S.; Karelson, M. Structurally diverse quantitative structure-property relationship correlations of technologically relevant physical properties. *J. Chem. Inf. Comput. Sci.* **2000,** *40*, 1–18.
100 Eros, D.; Keri, G.; Kovesdi, I.; Szantai-Kis, C.; Meszaros, G.; Orfi, L. Comparison of predictive ability of water solubility QSPR models generated by MLR, PLS and ANN methods. *Mini Rev. Med. Chem.* **2004**, *4*, 167–177.
101 Katritzky, A.R.; Fara, D.C. How chemical structure determines physical, chemical, and technological properties: An overview illustrating the potential of quantitative structure-property relationships for fuels science. *Energy Fuels* **2005,** *19*, 922–935.
102 Katritzky, A.R.; Fara, D.C.; Petrukhin, R.O.; Tatham, D.B.; Maran, U.; Lomaka, A.; Karelson, M. The present utility and future potential for medicinal chemistry of QSAR/QSPR with whole molecule descriptors. *Curr. Top. Med. Chem.* **2002**, *2*, 1333–1356.
103 Hansch, C.; Hoekman, D.; Gao, H. Comparative QSAR: Toward a Deeper Understanding of Chemicobiological Interactions. *Chem. Rev.* **1996**, *96*, 1045–1075.
104 Hansch, C.; Kurup, A.; Garg, R.; Gao, H. Chem-Bioinformatics and QSAR: A Review of QSAR Lacking Positive Hydrophobic Terms. *Chem. Rev.* **2001**, *101*, 619–672.
105 Schultz T.W.; Cronin M.T.D.; Walker J.D.; Aptula A.O. Quantitative structure-activity relationships (QSARs) in toxicology: a historical perspective. J. Mol. Struct. THEOCHEM **2003**, *622*, 1–22.
106 Stone, M. Cross-Validatory Choice and Assessment of Statistical Predictions. *J. R. Stat. Soc., B* **1974**, *36*, 111–147.
107 Gramatica, P. Evaluation of different statistical approaches for the validation of quantitative structure – activity relationships, ECVAM, Ispra, 2004. (http://ecb.jrc.ec.europa.eu/documents/QSAR/Report_on_QSAR_validation_methods.pdf)
108 Tropsha, A. Variable Selection QSAR Modeling, Model Validation, and Virtual Screening. In *Annual Reports In Computational Chemistry*, Spellmeyer, D.C. Ed.; Elsevier, Amsterdam, 2006; Vol. 2, pp 113–126.
109 Consonni, V.; Ballabio, D.; Todeschini, R. Comments on the definition of the Q^2 parameter for QSAR validation, *J. Chem. Inf. Model.* **2009**, *49*, 1669–1678.
110 Schüürmann, G.; Ebert, R.-U.; Chen, J.; Wang, B.; Kühne, R. External Validation and Prediction Employing the Predictive Squared Correlation Coefficient–Test Set Activity Mean vs Training Set Activity Mean. *J. Chem. Inf. Model.* **2008,** *48*, 2140–2145.
111 Gramatica, P. Principles of QSAR models validation: internal and external. *QSAR Comb. Sci.* **2007**, *26*, 694–701.
112 Tropsha A.; Golbraikh, A. Predictive QSAR Modeling workflow, model applicability domains, and virtual screening. *Curr. Pharm. Des.* **2007**, *13*, 3494–3504.
113 Hawkins, D.M. The Problem of Overfitting. *J. Chem. Inf. Comput. Sci.*, **2004**, *44*, 1–12.

114 Sprous, D.G.; Palmer, R.K.; Swanson, J.T.; Lawless, M. QSAR in the Pharmaceutical Research Setting: QSAR Models for Broad, Large Problems. *Curr. Top. Med. Chem.* **2010**, *10*, 619–637.
115 Katritzky, A.R.; Wang, Z.; Slavov, S.; Tsikolia, M.; Dobchev, D.; Akhmedov, N.G.; Hall, C.D.; Bernier, U.R.; Clark, G.G.; Linthicum, K.J. Synthesis and bioassay of improved mosquito repellents predicted from chemical structure. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 7359–7364.

SUMMARY IN ESTONIAN

Keemiliste ja bioloogiliste omaduste mitmemõõtmeline analüüs

Käesolevas dissertatsioonis uuriti keemilisi ja bioloogilisi omadusi mitmemõõtmeliste analüüsi meetoditega. Analüüsiti orgaaniliste ühendite lahustuvust, kui ainete fundamentaalset omadust, ravimiarenduses olulist ainete läbitavust, mille baasil hinnatakse ravimite suukaudse manustamise efektiivsust, ehk absorbtsiooni, ja sääsetõrje vahendite efektiivsust kasutades tõrje aega.

Dissertatsioon on jagatud kahte ossa. Esimeses osas on antud kirjanduse ülevaade uuritud omadustest. Käsitletakse kasutatud omaduste teoreetilist baasi ja praktilist väljundit. Samuti on kirjanduse ülevaates toodud kasutatud metoodika teoreetilised põhimõtted ning praktilised kasutusnäited. Teine osa võtab kokku dissertatsiooni raames tehtud uurimustöö.

Uurimustöö esimese osas keskenduti lahustunud aine jaotumisele gaasifaasi ja vedelfaasi vahel, väljendatuna Ostwaldi lahustuvuse koefitsiendina. Sellisel kujul väljendatud omadus on vahetult seotud lahustuvuse vabaenergiaga ja võimaldab lahustuvuse protsessi detailselt analüüsi. Töö käigus tuletati mitmed kvantitatiivsed struktuur-omadus sõltuvuse (QSPR) mudelid erinevate solventide ja lahustunud ainete seeriatele. Nii eksperimentaalsed lahustuvuse andmed kui ka saadud QSPR mudelitega ennustatud väärtused koondati ühtsesse maatriksisse. Saadud maatriksi peakomponentanalüüsi (PCA) tulemuseks on terviklik kahekomponente mudel, mis hõlmab 98,6% kogu andmete informatsioonist. Mudel toob välja solventide ja lahustunud ainete sarnased ja erinevad käitumised ühiste variatsioonide näol. Neid variatsioone analüüsiti kahel moel. Esiteks analüüsiti neid kvantitatiivselt, tuletades QSPR mudelid latentsetele muutujatele. Molekulaardeskriptorid saadud mudelites seostati solvatatsiooni vabaenergia liikmetega. Teiseks analüüsiti solventide ja lahustunud ainete jaotumismustreid. Selgelt eristuvad alifaatsed ja aromaatsed lahustunud ained ning samuti eristuvad ained keemiliste funktsionaalrühmade järgi. Kokkuvõtvalt toodi välja kaks peamist interaktsiooni tüüpi. Mittespetsiifilised interaktsioonid, mis hõlmavad endas molekuli suurusest sõltuvaid interaktsioone. Spetsiifilise interaktsioonid, mis koondavad enda alla polaarsusest, elektrostaatikast ja vesiniksidemest tingitud interaktsioone.

Uurimustöö teises osas analüüsiti ravimite läbitavust fosfolipiidsest membraanist. Tuletati viie-parameetriline kvantitatiivne struktuur-aktiivsus sõltuvuse (QSAR) võrrand, mille molekulaardeskriptorid kirjeldavad membraani läbitavuse olulisi karakteristikuid, nagu vesinikside, laengujaotus, polariseeritavus ja molekuli kuju. Mudeli diagnostika tõi välja ka keskpärased kõrvalekaldujad, mis oluliselt mudelit ei mõjuta.

Kolmas osa uurimustööst on seotud sääsetõrje vahendite efektiivsuse modelleerimisega. Saadud nelja-parameetriline QSAR mudel sisaldab molekulaardeskriptoreid, mis on seotud tõrjevahendi efektiivsust mõjutava kolme olulise

karakteristikuga. Esiteks on oluline aine aurustumine, teiseks strukturaalne sobivus retseptoriga ning kolmandaks keemiline interaktsioon retseptoriga.

Kokkuvõtvalt on dissertatsioonis edukalt rakendatud QSPR/QSAR ja PCA mitmemõõtmelisi analüüsi meetodeid. Saadud mudelid, nende valideerimine ja diagnostika omavad häid statistilisi karakteristikuid ning näitavad mudelite rakendatavust. Mudelite mehhanistlikud interpretatsioonid on kooskõlas toimuvate keemiliste ja bioloogiliste protsessidega ning seletavad keemiliste ühendite käitumist keskkonnas antud protsesside raames.

ACKNOWLEDGEMENTS

My biggest gratitude goes to my supervisor, Dr. Uko Maran for his expertise and guidance through the study and research. Many thanks are obliged to all the co-authors of my publications for excellent collaborations and discussions. Great thanks are extended to all the colleagues in the group, especially to Dr. Alfonso T. Garcia-Sosa for sharing office and for fruitful discussions over the years.

I also thank my family who has always believed in me and been very supportive and encouraging. I am thankful to my dear Aime for supporting me during the last efforts of my Ph.D. studies.

This work was supported by the Ministry of Science and Education (SF0182644Bs04 and SF0140031Bs09), Estonian Science Foundation (grants 5805 and 7709), EU 6FP program CardioWorkBench – *"Drug Design for Cardiovascular Diseases: Integration of in Silico and in Vitro Analysis"* (LSHB-CT-2005-018671) and the Doctoral School UTTP.

PUBLICATIONS

CURRICULUM VITAE

INDREK TULP

Born:	29/07/1976, Tallinn, Estonia
Citizenship:	Estonian
Martial Status:	cohabitated
Address:	Pärna 29–60, Tartu, Tartumaa, Estonia
Tel:	+372 737 5270; +372 56 257257
e-mail:	indrek.tulp@ut.ee

Education

2005–present	Ph.D. student of Molecular Engineering, University of Tartu
2002–2005	M.Sc. in Molecular Design, University of Tartu
1994–2000	B.Sc. in Chemistry, University of Tartu

Work Experience

2009–present	researcher, University of Tartu, Institute of Chemistry
2006–2009	extraordinary researcher, University of Tartu, Institute of Chemistry
2003–2004	research scholar, University of Florida, Chemistry Department, USA
2002–2003	junior researcher, Asper Biotech AS
1998–2002	analyst, Health Protection Agency Central Lab

Additional Training

- Theoretical and practical training of molecular modeling in *in silico* drug design and development, Computer Aided Drug Design and Development Society in Turkey (CADD&D), 3 days, 2008, Ankara, Turkey
- Study of Caco-2 and PAMPA experimental permeabilities, Pharmacelsus GmbH, 1 week, 2006, Saarbrücken, Germany
- Training of project writing and management, Tartu Students' Club, 3 months, 2006, Tartu, Estonia
- Heavy metal analysis, EU methods and regulations and method validation according ISO standards, Landesuntersuchungsamt Institut für Lebens-mittelchemie, 2 weeks, 2001, Trier, Germany

Publications

1. Maran, U.; Sild, S.; Tulp, I.; Takkis, K.; Moosus, M. Molecular Descriptors from Two-Dimensional Chemical Structure. In *In Silico Toxicology*, Cronin, M. T. D., Ed., RSC Publishing, Gambridge, UK, 2010, (accepted).
2. Tulp, I.; Dobchev, D.A.; Katritzky, A.R.; Acree, W. Jr.; Maran, U. A General Treatment of Solubility. 4. Description and Analysis of a PCA model for Ostwald solubility coefficients, *J. Chem. Inf. Model.* **2010**, (accepted).
3. Dobchev, D.A.; Mager, I.; Tulp, I.; Karelson, G.; Tamm, T.; Tamm, K.; Janes, J.; Langel, U.; Karelson, M. Prediction of Cell-Penetrating Peptides Using Artificial Neural Networks, *Curr. Comput.-Aided Drug Des.* **2010**, *6*, 79–89.
4. Tulp, I.; Sild, S.; Maran, U. Relationship between structure and permeability in artificial membranes: theoretical whole molecule descriptors in development of QSAR models, *QSAR & Comb. Sci.* **2009**, *28*, 811–814.
5. Karelson, M.; Karelson, G.; Tamm, T.; Tulp, I.; Jänes, J.; Tämm, K.; Lomaka, A.; Savchenko, D.; Dobchev, D. QSAR study of pharmacological permeabilities. *Arkivoc*, **2009**, *Part 2*, 218–238.
6. Karelson, M.; Dobchev, D.; Tamm, T.; Tulp, I.; Jänes, J.; Tämm, K.; Lomaka, A.; Savchenko, D.; Karelson, G. Correlation of Blood-Brain Penetration and Human Serum Albumin Binding with Theoretical Descriptors, *Arkivoc*, **2008**, *xvi*, 38–60.
7. Katritzky, A.R.; Dobchev, D.A.; Tulp, I.; Karelson, M.; Carlson, D.A. QSAR Study of Mosquito Repellents Using Codessa Pro, *Bioorg. Med. Chem. Let.* **2006**, *16*, 2306–2311.
8. Saal, K.; Tätte, T.; Tulp, I.; Kink, I.; Kurg, A.; Mäeorg, U.; Rinken, A.; Lõhmus, A. Sol-gel films for DNA microarray applications, *Mat. Let.* **2006**, *60*, 1833–1838.
9. Katritzky, A.R.; Tulp, I.; Fara, D.C.; Lauria, A.L.; Maran, U.; Acree, W.E. Jr. A General Treatment of Solubility. 3. Principal Component Analysis (PCA) of the Solubilities of Diverse Solutes in Diverse Solvents, *J. Chem. Inf. Model.* **2005**, *45*, 913–923.

ELULOOKIRJELDUS

INDREK TULP

Sündinud:	29/07/1976, Tallinn
Kodakondsus:	eesti
Perekonnaseis:	vabaabielus
Aadress:	Pärna 29–60, Tartu, Tartumaa, Eesti
Tel:	+372 737 5270; +372 56 257257
e-mail:	indrek.tulp@ut.ee

Hariduskäik

2005–praegu	molekulaartehnoloogia doktorant, Tartu Ülikool
2002–2005	M.Sc. molekulaardisain, Tartu Ülikool
1994–2000	B.Sc. keemia,Tartu Ülikool

Teenistuskäik

2009–praegu	teadur, Tartu Ülikool, Keemia Instituut
2006–2009	erakorraline teadur, Tartu Ülikool, Keemia Instituut
2003–2004	külalisteadur, Florida Ülikool, Keemiateadustkond, USA
2002–2003	nooremteadur, Asper Biotech AS
1998–2002	analüütik, Tervisekaitseinspektsiooni Kesklabor

Täienduskoolitus

- *In silico* ravimidisaini ja -arenduse molekulaarse modelleerimise teoreetiline ja praktiline koolitus, Computer Aided Drug Design and Development Society in Turkey (CADD&D), 3 päeva, 2008, Ankara, Türgi
- Caco-2 ja PAMPA membraanläbitavuste teooriad ja eksperimendid, Pharmacelsus GmbH, 1 nädal, 2006, Saarbrücken, Saksamaa
- Projektide kirjutamise ja juhtumise koolitus, Tartu Üliõpilasmaja, 3 kuud, 2006, Tartu, Eesti
- Raskemetalli analüüs, metoodika valideerimine ja laboritöö standardiseerimine vastavalt ISO nõuetele, Landesuntersuchungsamt Institut für Lebensmittelchemie, 2 nädalat, 2001, Trier, Saksamaa

Publikatsioonid

1. Maran, U.; Sild, S.; Tulp, I.; Takkis, K.; Moosus, M. Molecular Descriptors from Two-Dimensional Chemical Structure. In *In Silico Toxicology*, Cronin, M. T. D., Ed., RSC Publishing, Gambridge, UK, 2010, (aktsepteeritud).
2. Tulp, I.; Dobchev, D.A.; Katritzky, A.R.; Acree, W. Jr.; Maran, U. A General Treatment of Solubility. 4. Description and Analysis of a PCA model for Ostwald solubility coefficients, *J. Chem. Inf. Model.* **2010**, (aktsepteeritud).
3. Dobchev, D.A.; Mager, I.; Tulp, I.; Karelson, G.; Tamm, T.; Tamm, K.; Janes, J.; Langel, U.; Karelson, M. Prediction of Cell-Penetrating Peptides Using Artificial Neural Networks, *Curr. Comput.-Aided Drug Des.* **2010**, *6*, 79–89.
4. Tulp, I.; Sild, S.; Maran, U. Relationship between structure and permeability in artificial membranes: theoretical whole molecule descriptors in development of QSAR models, *QSAR & Comb. Sci.* **2009**, *28*, 811–814.
5. Karelson, M.; Karelson, G.; Tamm, T.; Tulp, I.; Jänes, J.; Tämm, K.; Lomaka, A.; Savchenko, D.; Dobchev, D. QSAR study of pharmacological permeabilities. *Arkivoc*, **2009**, *Part 2*, 218–238.
6. Karelson, M.; Dobchev, D.; Tamm, T.; Tulp, I.; Jänes, J.; Tämm, K.; Lomaka, A.; Savchenko, D.; Karelson, G. Correlation of Blood-Brain Penetration and Human Serum Albumin Binding with Theoretical Descriptors, *Arkivoc*, **2008**, *xvi*, 38–60.
7. Katritzky, A.R.; Dobchev, D.A.; Tulp, I.; Karelson, M.; Carlson, D.A. QSAR Study of Mosquito Repellents Using Codessa Pro, *Bioorg. Med. Chem. Let.* **2006**, *16*, 2306–2311.
8. Saal, K.; Tätte, T.; Tulp, I.; Kink, I.; Kurg, A.; Mäeorg, U.; Rinken, A.; Lõhmus, A. Sol-gel films for DNA microarray applications, *Mat. Let.* **2006**, *60*, 1833–1838.
9. Katritzky, A.R.; Tulp, I.; Fara, D.C.; Lauria, A.L.; Maran, U.; Acree, W.E. Jr. A General Treatment of Solubility. 3. Principal Component Analysis (PCA) of the Solubilities of Diverse Solutes in Diverse Solvents, *J. Chem. Inf. Model.* **2005**, *45*, 913–923.

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 polypyrrole 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 hydrazines. 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. **Margo 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.