DISSERTATIONES CHIMICAE UNIVERSITATIS TARTUENSIS
73

DISSERTATIONES CHIMICAE UNIVERSITATIS TARTUENSIS
73

BIOMEDICAL APPLICATIONS OF THE QSAR APPROACH

SVETOSLAV HRISTOV SLAVOV

TARTU UNIVERSITY
PRESS

Institute of Chemistry, University of Tartu, Estonia

Dissertation is accepted for the commencement of the degree of Doctor of Phylosophy in Chemistry on 12 November, 2007, by the Doctoral Committee of the Department of Chemistry, University of Tartu.

Supervisor: Professor Mati Karelson

Opponents: Professor Peeter Burk, Department of Chemistry, University of Tartu, Estonia
Professor Mart Saarma, University of Helsinki, Institute of Biotechnology, Helsinki, Finland

Commencement: 20 December, 2007 at 18 Ülikooli St., room 204, 11:00 h

ISSN 1406–0299
ISBN 978–9949–11–783–3 (trükis)
ISBN 978–9949–11–784–0 (PDF)

Tartu Ülikooli Kirjastus
www.tyk.ee
Tellimus nr 510

CONTENTS

2

LIST OF ORIGINAL PUBLICATIONS

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

I. Slavov, Svetoslav; Atanassova, Mariyana; Galabov, Boris. **QSAR analysis of the anticancer activity of 2,5-disubstituted 9-aza-anthrapyrazoles.** QSAR & Combinatorial Science (2007), 26(2), 173–181.

II. Katritzky, Alan R.; Slavov, Svetoslav H.; Dobchev, Dimitar A.; Karelson, Mati. **Comparison between 2D and 3D-QSAR approaches to correlate inhibitor activity for a series of indole amide hydroxamic acids.** QSAR & Combinatorial Science (2007), 26(3), 333–345.

III. Katritzky, Alan R.; Pacureanu, Liliana M.; Slavov, Svetoslav; Dobchev, Dimitar A.; Karelson, Mati. **QSAR study of antiplatelet agents.** Bioorganic & Medicinal Chemistry (2006), 14(22), 7490–7500.

IV. Katritzky, Alan R.; Kuanar, Minati; Slavov, Svetoslav; Dobchev, Dimitar A.; Fara, Dan C.; Karelson, Mati; Acree, William E.; Solov'ev, Vitaly P.; Varnek, Alexandre. **Correlation of blood-brain penetration using structural descriptors.** Bioorganic & Medicinal Chemistry (2006), 14(14), 4888–4917.

V. Katritzky, Alan R.; Slavov, Svetoslav; Dobchev, Dimitar A.; Karelson, Mati. **QSAR modeling of the antifungal activity against Candida albicans for a diverse set of organic compounds.** Submitted for publication to *Journal of Computer-Aided Molecular Design.*

Author's contribution

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

Publication III, IV: The author is responsible for the calculations, preparation and writing parts of the manuscript and interpretation of the results.

LIST OF ABBREVIATIONS

AM1	Austin Model 1
ANN	Artificial Neural Network
BMLR, (B)MLR	Best Multilinear Regression
CODESSA	COmprehensive DEscriptors for Structural and Statistical Analysis
CoMFA	Comparative Molecular Field Analysis
LMO	Leave-Many-Out cross-validation
LOO	Leave-One-Out cross-validation
MLR	Multilinear Regression
MOPAC	Molecular Orbital PACkage
PCA	Principal Component Analysis
PLS	Partial Least Squares
QSAR	Quantitative Structure – Activity Relationship(s)
RMSPE	Root-mean Squared Error of Prediction
RMS	Root-mean Squared Error
SVD	Singular Value Decomposition
WLS	Weighted Least Squares

INTRODUCTION

Many useful medicines still in use today were discovered millennia ago by providing a mixture of naturally occurring compounds to the sick people. They include: morphine (~ 4000 BC), reserpine (< 1000 BC), aspirin (< 200 BC) and ephedrine (~ 1 AD). These mixtures were usually derived from parts of plants (*e.g.* bark, root or seeds), fungi, insects or animals, or from extracts of these organisms. As a result of the development of science and medicine, changes in the perceptions toward human life value and commercial pressure, this strategy for drug discovery was abandoned.

The process of studying the curare-like paralysing properties of a set of quaternised strychnines led to the proposal of the Crum-Brown and Frazer equation:

$$\Phi = f(C) \tag{1}$$

in which **f** is a measure of the biological activity and **C** represents relevant structural features. Their work was published in the middle of the XIX century and at this time they were only able to associate C with the nature of the quaternising group[1].

At about the same time Richardson[2] showed that the toxicity effect of ethers and alcohols could be directly related to their water solubility if reciprocal function was used. Richet[3] found a simple relationship between the narcotic effect of alcohols and their molecular weight and Meyer[4] and Overton[5] independently reported QSARs where the narcotic action of a series of diverse organic compounds was related to their oil/water partition coefficients. In the 1930's chemists began the exploration of the relationship between the chemical structure and the equilibrium or rate constants of chemical reactions[6]. Among all published in this field studies one gained a lot of popularity. This was the work of Hammett who invented a scale of electronic effects of substituents using equation

$$\rho * \sigma_x = \log K_x - \log K_H \tag{2}$$

where $\mathbf{K_X}$ and $\mathbf{K_H}$ are the equilibrium constants for a reaction involving respectively an X substituted compound and its unsubstituted parent, $\boldsymbol{\rho}$ is a reaction constant which is characteristic of a particular reaction, and $\boldsymbol{\sigma_x}$ is the substituent constant for the substituent X. The Hammett equation has been used to create $\boldsymbol{\sigma}$ scales useful for characterization of different types of electronic effects[7].

In the 1960's Corwin Hansch[8] made a significant contribution by proposing a hydrophobicity model system based on octanol/water partition coefficients (logP), simply defining them as a ratio between the concentrations of a

compound *Y* in octanol and water. The choice of the octanol as a reference phase was predetermined by the sensation that it might simulate the lipid components of the cellular membranes, while water represented the aqueous phase of biological systems. The original proposal involved a substituent constant $\boldsymbol{\pi}$, defined using equation:

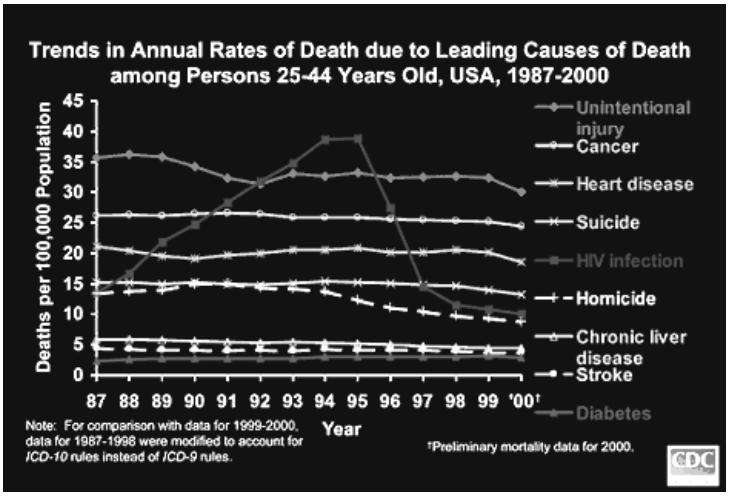
$$\pi_x = \log P_x - \log P_H \qquad (3)$$

One of the most significant contributions made by Hansch was his supposition that the use of only one variable (such as logP) might not be enough for the "explanation" of the biological potency. The generalised form of the so called "Hansch equation"[9], is shown in equation (4):

$$\text{Log } 1/C = a\pi + b\pi^2 + c\sigma + dE_s + \text{constant} \qquad (4)$$

The terms in equation 4 are the hydrophobic and electronic substituent constants $\boldsymbol{\pi}$ and $\boldsymbol{\sigma}$ and the steric substituent constant $\mathbf{E_S}$, due to Taft[10]. This equation consists of a linear combination of terms (even though $\boldsymbol{\pi}^2$ is a square) and thus this is the well known multiple linear regression equation.

With the increased computational power of the early 80's the computer aided molecular design (CAMD) was adopted in pharmaceutical and agrochemical industries. The earliest success stories involved a method known as Quantitative Structure-Activity Relationships (QSAR) in which mathematical models are constructed between a biological response and a quantitative description of chemical structure. The success of CAMD is illustrated in this chart from the Centre for Disease Control, which shows that AIDS deaths dropped sharply in 1995, when HIV protease inhibitors were developed by using the methods of QSAR.



Currently, drug design is one of the most extensive and important fields of the bioinformatics. The perspective of discovering new drugs based on the knowledge of their structure, target site and mechanism of action is a tempting goal for many scientists working in the field. With recent estimates of drug development costs in the order of $800 million and increased pressure to reduce consumer drug costs, it is not surprising that the pharmaceutical industry is highly interested in reducing the overall expense associated with drug development[11]. It is also important to notice that the process starting with drug development and ending with bringing the final product to market could take from 5 to 10 years and this period might be significantly reduced by applying the methods of QSAR.

As understood today, the QSAR approach is a rational approach for the optimization of leading structure using a set of purely theoretically calculated descriptors which are then related to the available experimental property values. The underlying premise of QSAR is that there is an existing relationship connecting the biological/pharmacological activity of a compound to its physico-chemical properties. This approach allows important for the activity structural features to be identified and hence narrows the search for perspective candidates. Once, the final form of the QSAR functional dependence is known then it is relatively easy to determine how each structural change influences the activity. The QSAR equations also allow prediction of activity of new compounds' based only on structural data and hence saves time synthesizing molecules unlikely to be active.

1. LITERATURE OVERVIEW

1.1. QSAR general methodology

The most important role in drug design as understood today is played by the binding site – the target location for the drug molecules. For depiction of these so-called "ligand-receptor" interactions, historically, two main approaches were developed: one now called classical or Hansch approach and its alternative – the 3D-QSAR.

Classical QSAR relates the biological activity to physicochemical properties or indicator variables which encode certain structural features[12–16]. In addition to lipophilicity, polarizability, and the electronic properties, steric parameters are also frequently used to describe the different size of substituents. In some cases, indicator variables have been attributed to differentiate racemates and active enantiomers[13,14].

Instead of dealing with scalar quantities (2D-QSAR descriptors), the 3D-QSAR method deals with fields, which are usually only two – steric and electrostatic. This method assumes that "looking" at the ligand (drug molecule) binding site does not see atoms and bonds: i) from a far distance, it would "feel" the electrostatic potential of the molecule and ii) at a closer distance, the relatively hard body of the molecule with its charge distribution pattern at the solvent-accessible surface.

In 1979, Cramer and Milne made the first attempt to compare molecules by aligning them in space and by mapping their molecular fields to a 3D grid[17]. Several important facts made the application of this approach broader:

1) in 1986, Svante Wold proposed the use of partial least squares (PLS) analysis, instead of principal component analysis (PCA), to correlate the field values with the biological activities;
2) in 1988, a key publication appeared in the *Journal of the American Chemical Society*[18] and the method was called comparative molecular field analysis (CoMFA);
3) appropriate software became commercially available[19].

Despite their differences, these two main QSAR approaches involve several common major steps:

A) Selection of a dataset

Over the years, a large number of QSAR studies on drugs and potential drug candidates have been published[20]. With few exceptions[21–25] the reported models are based on small datasets, since it is not easy for the academic groups to access large QSAR databases and publish studies based on them. Good QSAR models for large data sets developed in the pharmaceutical companies, on the

other hand, were often not published in order to protect intellectual property. Larger data sets are typically found in QSPR (Quantitative Structure-Property Relationships) studies[26], but the reported in such studies results have only a limited value for drug design QSARs. However, there are two publications[27,28] that try to estimate the influence of the quantitative parameters of the datasets (size, ratio training:test datasets and experimental error) on the quality of the QSAR modeling results.

Some essential practical rules concerning the criteria for dataset selection were extracted from the references cited above and are summarized below:

1) Size of the dataset – at least 25–30 molecules characterized with similar mechanism of action should be selected. Usually these are congener compounds participating in specific interactions.
2) Range of the property values – the property values should cover a range of at least 1 logarithmic unit.
3) Experimental error – reliable QSAR models can be obtained if the relative experimental error is less than 15%.
4) The optimal ratio between the training and the test datasets lies within the boundaries 2:1…4:1.

B) Property transformations

1) Data distribution – the property values should possess a normal (Gaussian) distribution[29]. If they are not normally distributed a transformation of the original property values is necessary. Some of the most frequently used transformation functions are: LogBA, Log(1/BA), 1/BA and $1/BA^2$. Transformations implying sinus, tangent or hyperbolic functions are "statistical unicorns – beasts that exist on paper but not in reality"[13].
2) All mass concentration based property values should be transformed to such which are molar concentration based.

C) Model building

This is one of the most critical steps in the general QSAR methodology. It includes:

1) Molecular structure optimization

For the purposes of drug design and especially when performing 3D-QSAR analysis one should ensure that the energy of the optimized structure is close to that of the global minima. Two alternative approaches for exploration of the conformational space are currently in use – systematic conformational analysis and the method of random searching.

The systematic search method covers whole conformational space only when all rotatable bonds are defined as optimization parameters. If many single bonds are present and/or the chosen rotational angle is small, the number of the

conformers generated can be exceedingly large. Random conformational searching differs from the systematic search in that it is not guaranteed to cover the whole conformational space.

Thus, in complex cases, it is desirable to have a random sampling of the conformational space. As a result of the method application a set of different conformers will be generated. The completeness of the set of conformations produced can be increased by augmenting the random searching cycles on the molecule[30].

At the next stage full geometry optimization of the studied compounds is necessary. Usually, the bioactive molecules are complex structures and the wide-spread *ab initio* quantum chemical methods as highly time consuming are inappropriate. However, a variety of sufficiently accurate semiempirical quantum chemical methods AM1[31], PM3[32], MNDO[33] can be used.

2) Descriptors generation

The molecular descriptors are mathematical values describing the physical and chemical properties of molecules. Some of them are empirical indices (such as substituent constants and various electronegativity-related parameters), most actively used in the early QSAR studies. Nowadays, electronegativity remains very popular and broadly employed descriptor[34–40], while the substituent constants are often disregarded by the modern QSAR. With the increase of the computational power the quantum chemical, electronic, geometrical, constitutional and topological descriptors become preferable. The quantum-chemical, electronic and geometrical descriptors are usually derived from the results of empirical schemes or molecular orbital calculations and they encode the molecule's ability to participate in polar or hydrogen bonding (donor, acceptor) interactions. The constitutional descriptors are fragment additive and reflect mostly the general properties of compounds related to their structures. The topological descriptors are calculated using mathematical graph theory applied to the scheme of atom connections of the structure. Available software programs that can calculate these descriptors are CODESSA PRO[41], Oasis[42], Dragon[43], Chem-X[44], TSAR[45] and others.

As discussed above the 3D-QSAR methods usually use two types of fields, which corresponding descriptors are simply electrostatic and steric interaction energies between a set of aligned molecules and "fictitious" atom calculated at the 3D-grid intersections.

3) QSAR model generation

The model generation procedure involves identification of the most informative subsets of descriptors, which explain a sufficient amount of the variance of experimental data. Several different statistical methods for calculating QSAR equations are in use:

- Genetic function approximation (GFA).

4

- G/PLS.
- Stepwise multiple linear regression.
- Partial least squares (PLS).
- Principal components regression.

The genetic function approximation (GFA) algorithm can be used as an alternative to standard regression analysis for constructing QSAR equations. The application of this method leads to multiple models generated by evolving random initial models using a genetic algorithm. Each cycle performs a crossover operation to recombine terms of better scoring models, thus improving the parameters of the model. The method is good for generating QSAR equations when dealing with a large number of descriptors.

Genetic partial least squares (G/PLS) is a variation of GFA which is derived from two methods: genetic function approximation (GFA) and partial least squares (PLS). Both GFA and PLS are intended to help in situations where the datasets have more descriptors than samples.

The multiple linear regression method (MLR) and its BMLR modification (B stands for Best) generates QSAR equations by performing standard regression calculations using multiple variables in a single equation. The use of multiple linear regression, assumes that all variables are independent (not correlated).

Partial least squares (PLS) regression is based on linear transformation of a large number of original descriptors to a new variable space consisting only of a small number of orthogonal latent variables. In other words, the generated latent variables are orthogonal linear combinations of original descriptors. The PLS method is typically applied when the independent variables are correlated, or when the independent variables exceeds the observations. Under these conditions, it produces a more robust QSAR models than MLR. This is the standard statistical method used in 3D-QSARs.

The principal components analysis (PCA) method does not generate a QSAR model but searches for relationships among the independent (X) variables. It then creates new variables called principal components, which represent most of the information contained in the independent variables.

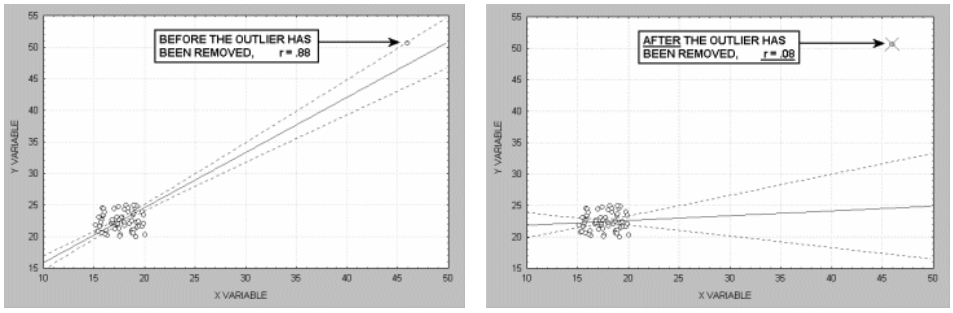
The principal components regression (PCR) method uses linear regression to create a model using the generated by PCA principal components as independent variables.

D) Selection of an equation

1) Identification and removal of outliers.

An outlier is an atypical value not belonging to the distribution of the rest of the values in the data set. Datapoints having deviation two times greater than standard deviation of the data are usually considered as outliers. This definition is correct only when the distribution is unimodal and symmetrical (most cases).

For skewed data, median is a better indicator of central location than mean. In such cases, an observation can be considered as an outlier if it is more than 1.5 inter-quartile distance away from the closest quartile. The extreme outliers are those which lie in more than 3 inter-quartile distances from the closest quartile. Among all possible equations generated those characterized with less, but explainable outliers should be selected for further treatment.



2) 5:1 thumb rule

Giving enough parameters any data set can be fitted to a regression line. The consequence of this is that regression analysis generally requires significantly more compounds than parameters. A useful rule of thumb is that the ratio between the objects and the variables should be at least five to one for the MLR analysis – otherwise there is a corresponding large risk of by chance correlation[46].

3) Principle of parsimony (Occam's Razor)

The principle of parsimony was postulated by William of Occam and states that among a set of equally good explanations for a given phenomenon, the correct explanation is the simplest explanation. It is called Occam's razor because he was "trimming down" his explanations down to the bare minimum. In QSAR modeling, the principle of parsimony means that:

- models should have as few parameters as possible;
- models should be pared down until they are minimally adequate;
- the simple explanations should be preferred as a better alternative to the complex ones.

The process of model simplification is an integral part of hypothesis testing. In general, a variable is retained in the model only if its removal causes a significant decrease of the statistical parameters compared to those of the current model.

Of course by simplifying the model we should be careful not to lose the essential parts. Einstein made a clever addition to Occam's razor, saying that "A model should be as simple as possible. But no simpler"[47].

E) Validation and Interpretation of the model

1) Domain of applicability

Even the most exhaustive, confident and validated QSAR model cannot be expected to reliably predict the modeled property for the entire possible set of chemicals. Its specific domain of applicability must be defined, and the predictions for only those chemicals that fall into this domain can be considered reliable. The chemical domain of applicability is a region in the functional space defined by the modeled response and the descriptors of the model for which a given QSAR should be able to generate reliable predictions. This region is defined by the chemicals included in the training set, and can be characterized in various ways, by using a William's plot or leverage plot for example.

2) Internal validation

Several statistical methods now known as Leave-One Out (LOO), Leave-Many Out (LMO), bootstrapping, scrambling and etc[48] were developed with the main aim to find robust criteria for verification of "internal predictivity". Most methods are based on iterations where, on each cycle, different numbers of compounds from the training set are omitted. The QSAR is then developed on the basis of the data for the training chemicals, followed by the use of the omitted chemicals to make predictions. This procedure could be repeated a number of times (thousands for instance), so that statistics can be derived from the comparison of predicted with known data. Cross-validation techniques allow the assessment of internal predictivity, in addition to the robustness of the model (stability of QSAR model parameters). These types of internal validation procedures provide only a reasonable estimation of the internal predictive power of the QSAR model; it has been demonstrated that the LOO and LMO Q^2 does not indicate a significant external predictive power (known as Kubinyi paradox).

3) External validation

In most cases, finding new compounds for external validation purposes is generally difficult. The new data should be available in a statistically significant amount (in fact, results on few data points might give too optimistic or pessimistic information regarding the model predictivity) and belonging to the same chemical domain as the compounds used for model development. This is the best way of external validation, performed after the model development.

When additional data (in useful quantity and quality) is unavailable, an adequate splitting of the initial data set into training (used to build up a model) and validating (used to validate the model's predictability) sets can be done.

The external validation relies on a procedure where chemical structures the selected for inclusion in the validation set are different from those included in the training set, but are representative of the same chemical domain. The QSAR model developed using only training set chemicals is then applied to the

"unknown" validation set to verify, more reliably, the predictive ability of the model.

4) Interpretation

One of the most important aspects of drug design targeted QSAR is the extraction of structure-activity relationship information that is encoded in the model. To perceive one model as credible it is crucial to explain its biochemical significance. Unless the equation is very simple, a routine examination of a model generated using conventional QSAR techniques would be insufficient for the following reasons: i) the coefficients of the equation are not always directly interpretable because they usually represent a complex picture of two or even more structural trends typically buried in the model; ii) the structural descriptors involved may include those derived by using a purely mathematical construct that may not have a direct physico-chemical interpretation – e.g. molecular connectivity indices[49], and iii) a descriptor selected for inclusion in the model by using conventional regression algorithms, may be acting as a surrogate measure for structural features not characterized accurately enough by another descriptor with a more intuitive physico-chemical interpretation. For example, the topological changes in the molecules can be related to changes in their geometry and thus could act as a measure of the shape these molecules present to the other interacting molecules. However, the changes in branching can affect the distribution of the electronic charge, which can cause changes in the reactivity or polarity. Similarly, a measurement of the distribution of electronic charge can, in some cases, act as a better measure of branching, and therefore of shape, than do the typical topological descriptors[50]. Thus, simply knowing the type of features measured by a certain descriptor could be insufficient in itself.

The development of a multilinear model involves selecting a set of one or more descriptors that provide a statistical correlation with the experimental property values. A preconceived notion of the physico-chemical interpretation of the descriptors involved could result in a misunderstanding of the underlying structure-activity relationship. There are two important pieces of information one needs in order to generate a clear QSAR interpretation: i) the knowledge of what features of the structure are measured by a given descriptor, and ii) the knowledge of how structural changes influence the experimentally observed property. In the case of whole-molecule descriptors used, e.g. molecular connectivity indices etc., the structural features changes measured by a certain descriptor may be occurring in several places. However, the important structural changes which affect the observed property may be localized at a particular position in the molecule. Thus to make a clear QSAR interpretation one must know both the essential structural features and their relation to the descriptors involved.

D) Prediction

This is the final stage of the QSAR modeling involving the use of all successful models (2D- and 3D – QSARs) to estimate the bioactivity of known or unknown structures or to predict new potentially active drug candidates. Usually the 3D-QSAR models are more accurate in their predictions since the exact location of the positively and negatively contributing substituents can be located. Some of the 2D-QSAR models could achieve similar levels of accuracy when fragment descriptors are used. The predictions then should be analyzed in order to validate their reliability and if inaccurate predictions were made the reasons need to be explored more thoroughly.

1.2. Hansch type analysis – multilinear modeling

The general purpose of multilinear regression is to learn more about the relationship between several independent or predictor variables and a dependent or criterion variable[51]. The method uses the following assumptions:

- The random errors ε_i have expected value 0;
- The random errors ε_i are uncorrelated;
- The random errors ε_i have the same variance.

A linear regression with p parameters (including the regression intercept β_1) and n data points (sample size), with $n \geq (p+1)$ allows construction of the following vectors and matrix with associated standard errors:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{21} & x_{31} & \dots & x_{p1} \\ 1 & x_{22} & x_{32} & \dots & x_{p2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{2n} & x_{3n} & \dots & x_{pn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{5}$$

or, from vector-matrix notation above,

$$y = X_* \beta + \varepsilon \tag{6}$$

Each data point can be given as $(\vec{x}_i, y_i)$, i=1,2,….,n. For $n = p$, standard errors of the parameter estimates could not be calculated. For n less than p, parameters could not be calculated. The estimated values of the parameters are given by:

$$\hat{\beta} = (X^T X)^{-1} X^T \vec{y} \tag{7}$$

Using the assumptions provided by the Gauss-Markov Theorem, it is possible to analyse the results and determine whether or not the model determined using least-squares is valid. The number of degrees of freedom is given by $n - p$.

The residuals, representing "observed" minus "calculated" quantities, are useful to analyse the regression. They are determined from:

$$\hat{\vec{\varepsilon}} = \vec{y} - X\vec{\beta} \tag{8}$$

The standard deviation, $\vec{\sigma}$ for the model is determined from

$$\vec{\sigma} = \sqrt{\frac{\hat{\vec{\varepsilon}}^T \hat{\vec{\varepsilon}}}{n-p}} = \sqrt{\frac{\vec{y}^T \vec{y} - \hat{\vec{\beta}}^T X^T \vec{y}}{n-p}} \tag{9}$$

The variance in the errors can be described using the Chi-square distribution:

$$\hat{\sigma}^2 \sim \frac{\chi^2_{n-p}\sigma^2}{n-p} \tag{10}$$

The 100(1 − α)% confidence interval for the parameter, β_i, is computed as follows:

$$\vec{\beta}_i \pm t_{\frac{\alpha}{2},n-p}\vec{\sigma}\sqrt{(X^T X)^{-1}_{ii}} \tag{11}$$

where t follows the Student's t-distribution with $n - p$ degrees of freedom and $(X^T X)^{-1}_{ii}$ denotes the value located in the i^{th} row and column of the matrix.

The regression sum of squares *SSR* is given by:

$$SSR = \sum(\vec{y}_i - \bar{y})^2 = \hat{\beta}^T X^T \vec{y} - \frac{1}{n}(\vec{y}^T \vec{u}\vec{u}^T \vec{y}) \tag{12}$$

where $\vec{u}$ is an n by 1 unit vector.

The error sum of squares *ESS* is given by:

$$ESS = \sum(y_i - \hat{y}_i)^2 = \vec{y}^T \vec{y} - \hat{\beta}^T X^T \vec{y} \tag{13}$$

The total sum of squares *TSS'* is given by

$$TSS = \sum(y_i - \bar{y})^2 = \vec{y}^T \vec{y} - \frac{1}{n}(\vec{y}^T \vec{u}\vec{u}^T \vec{y}) = SSR + ESS \tag{14}$$

Pearson's regression coefficient, R^2 is then given as

$$R^2 = \frac{SSR}{TSS} = 1 - \frac{ESS}{TSS} \tag{15}$$

Several statistical characteristics give information about the "goodness" of the model: R^2 – squared correlation coefficient, R^2_{CV} – squared cross-validated correlation coefficient, F – Fisher criterion value, s^2 – squared standard error[52].

Usually the QSAR study deals with a large number of molecular descriptors. The search of the best multiple linear regression model among such a large descriptor space is not a trivial task. Various regression techniques were proposed for the selection of the "best set" of regression predictors as backward elimination, forward selection, ridge regression, heuristic etc[53,54]. In the current thesis we have used an alternative algorithm called Best Multilinear Regression (BMLR)[52,55] which combines MLR and descriptor selection procedure. This method is encoded in CODESSA PRO software[41] and it is fully automatised. It is a modification of the forward stepwise algorithm[54] and the main steps regarding equation building and variable selection are:

i) in a given descriptor space BMLR searches and selects all orthogonal within some limits descriptor pairs;
ii) in the created descriptor subspace in step ii), it builds all two-parameter regression equations and selects these descriptors which show the highest correlation coefficient (R) in the respective two-parameter regressions;
iii) by using the best two-parameter equations in ii), it creates three-parameter regressions by adding an additional noncollinear descriptor. Then Fisher criterion F and the crossvalidated coefficient (R_{cv}) are assessed and if there is no improvement over the best two-parameter regressions in ii), the procedure is halted. Otherwise, BMLR continues with the selection of the best three and higher parameter regression equations according to R.

The BMLR method is able to find the "best" regression for a short computational time in a descriptor space of hundreds of variables. There is full control over the algorithm settings so that certain descriptor space can be explored more precisely for best correlations.

1.3. 3D-QSAR analysis – brief overview

Since its introduction about 20 years ago CoMFA (Comparative Molecular Field Analysis) has become one of the most powerful tools for QSAR and drug design. In fact, CoMFA has pioneered a new paradigm of three-dimensional

QSAR[56] where the electrostatic and steric molecular fields are related to specific structural features (substituents, etc.) and their spatial interaction.

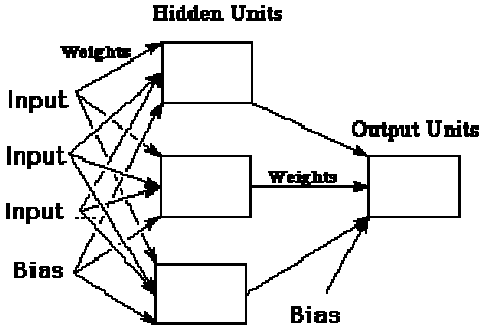
CoMFA describes 3D structure-activity relationships in a quantitative manner. As a most important precondition, all molecules forming the dataset have to interact with the same kind of receptor (or enzyme, ion channel, transporter) targeting identical binding sites in the same relative geometry. In the next step, a certain subgroup of molecules is selected to form the training set, which is used to derive the CoMFA model. The remaining molecules are considered to be a test set used for validation of the derived model. Atomic partial charges are calculated and several low energy conformations are generated. Once the pharmacophore hypothesis is derived all molecules from the training set could be superimposed in a similar manner, in order to produce a rational and consistent alignment[57]. In this case the common substructure for all molecules should have the same conformation, and the unmatching parts should be superimposed as much as possible by adjusting the internal torsional angles[58]. A sufficiently large box with a defined grid constant (usually 2Å) is then positioned around the set of aligned molecules. For each molecule at each grid point, the steric (Lennard-Jones potential) and electrostatic interaction energies are measured by a probe positively or negatively charged atom, a hydrogen bond donor or acceptor, or a lipophilic probe. The generated "field" values are then stored in tables, usually including several thousands of columns, which at the next step are correlated to the binding affinities or with other biological property values. The PLS analysis is usually used as a standard statistical tool for CoMFA data processing. Normally, cross-validation is used to check the internal predictivity of the derived model.

The result of the PLS analysis corresponds to a regression equation with thousands of coefficients. For easier interpretation most often the results are presented as a set of contour maps. These contour maps show all favorable and unfavorable regions for the steric and electrostatic interactions around the molecules concerning different types of substituents in certain positions. Predictions for the test set and/or for other compounds can be made, either by a visual inspection of these contour maps or, in a quantitative manner, by calculating the fields of these molecules and by inserting the grid values into the PLS model.

1.4. ANN approach, general scheme

Neural networks are typically used when a large number of observations are available and a nonlinear relationship is expected, or when the problem is not understood well enough to apply some of the other available methods. The "architecture" of the artificial neural network consists of a number of "neurons" or "hidden units" that receive data from the outside, process the data using transformation functions, and produce a signal. The "neurons" actually act as

non-linear transformation functions. When more than one of these neurons are used, non-linear models can be fitted. These networks have been applied to the modeling of numerous problems, including QSAR[59–68]. Neural networks are known for their ability to model a wide set of functions without knowing the model a priori. The back propagation network receives inputs signals which are then multiplied by each neuron's weight (Figure 1). For each neuron these products are summed and a non-linear transfer function is applied. The role of the bias is to shift the transfer function to the left or right. The resultant sums from the previous step are then multiplied by the output weights, transformed, and interpreted. Since a back-propagation network is a supervised method, the desired output must be known for each input vector. The difference between the desired output and the network's predicted output define the ANN model error, which needs to be minimized. This error is then propagated backward through the network, adjusting the weights so that on the next cycle, the generated predictions will come closer to the desired output.



There are four important factors which must be considered when using neural networks:

i) The design of the network is critical with respect to the number of hidden units involved – if too many hidden units are used the network will overfit or memorize the data; conversely, if too few hidden units are used the network will fail to generalize and will become unstable.
ii) The length of the training time must always be considered – if excessive training periods are used the network might become overtrained, and thus destabilized.
iii) Appropriate test and training sets must be defined. The training set should adequately represent the entire dataset and be sufficiently large in order to properly train the neural network.
iv) The results obtained from neural networks can be difficult to interpret and apply to drug design.

2. APPLICATIONS OF 2D-, 3D- AND ANN QSARs TO VARIOUS DRUG DESIGN PROBLEMS

2.1. QSAR analysis of the anticancer activity of 2,5-disubstituted 9-aza-anthrapyrazoles

Article I proposes 2D- and 3D-QSAR models for a series of thirty five 2,5-disubstituted indazolo[4,3-gh]isoquinolin- 6(2H)-ones with well-expressed cytotoxicity against tumor cell line LoVo. A logarithmic transformation of data was used, which resulted in a distribution close to normal.

The 2D-QSAR analysis carried out followed the scheme of the general Hansch approach. All descriptors used were derived solely from molecular structure without reference to experimental data. The best 3 parameter QSAR model obtained is given below:

$LogIC_{50} = 3.25*MACB + 2470.40*MPCNA + 0.10*HDSA2 – 587.81$
$n = 35; R^2=0.66; R^2_{cv} = 0.56; F=20.25; s^2 = 0.28$

The abbreviations used are as follows:
MACB – Minimum Electron-Nuclear Attraction Energy for a C-C bond
MPCNA – Maximum Partial Charge for a Nitrogen Atom
HDSA2 – Area-Weighted Surface Charge of Hydrogen-Bonding Donor Atoms

The MACB descriptor could be related to the quite stable structure of condensed rings forming the backbone of the 9-aza-APs.

The MPCNA descriptor depicts the charge distribution for nitrogen atoms within the molecule. These atoms are expected to play a crucial role in the binding of 9-aza-APs to the DNA.

The HDSA2 descriptor describes the solvent-accessible surface area of the H-bonding donor atoms. Hydrogen bonding can be realized if: i) the nucleophilic nitrogen atoms of the 9-aza-APs act as proton acceptors and ii) the NH or OH act as hydrogen donors.

The results obtained from 3D-QSAR are in agreement with the hypothesis of intercalation of the antitumor agents into the DNA molecules. The intercalation can be greatly facilitated by hydrogen bonding formation.

As most active 2-[2-(dimethylamino)ethyl]-5-[[2-(methylamino)-ethyl]amino] indazolo[4,3-g,h]isoquinolin-6(2H)-one dimaleate was chosen as a template for aligning.

By using the implemented in Chem-X WLS method, 3D-maps were separately generated for the steric and electrostatic fields. It was found that R^2 of the best produced 3D-QSAR model describing the steric interactions is about two times higher than Q^2. This result indicates that the steric interactions do not

appear to have a major role in explaining the biological activity of the compounds.

The visual inspection of the WLS map for the electrostatic interactions shows that the presence of aliphatic substituents containing an amino group in second position, hydrocarbon residues containing strongly electronegative atoms in the fifth position as well as a keto group in sixth position of the heterocyclic system all lead to increased antitumor activity. It was also found that the presence of a nitrogen atom in the ninth position is also an essential factor for the activity.

In conclusion, the results produced by both 2D- and 3D-QSARs are fully compatible with each-other and lead to similar conclusions.

2.2. 2D and 3D-QSARs of a series of indole amide hydroxamic acids

Article II presents 2D- and 3D-QSAR models explaining the inhibitor activity of a series of 36 indole amide hydroxamic acids (suberoylanilide hydroxyamic acid derivatives) to histone deacetylases (HDACs). In addition a comparison between these two alternative approaches was carried out. On the basis of the obtained results, new potentially active inhibitors were predicted and reported.

The histone deacetylase inhibitors broadly used as cancer therapeutics are small molecules regulating gene transcription. It has been shown that their activity is related to the insertion of the aliphatic chain of SAHA into the hydrophobic cleft of the enzyme active site, whereas the polar head group serves as a bidentate chelator of the catalytic Zn^{2+} ion.

CODESSA PRO and Chem-X software packages were used to build respectively 2D- and 3D-QSAR models. The BMLR (best multilinear regression) method was used to generate the following multiparameter model:

$$\mathrm{LogIC_{50}} = -25.31(\pm 3.20)\ \mathrm{ABOC} - 9.40*10^{-3}(\pm 9.98*10^{-4})\ \mathrm{TMSA} + 34.59(\pm 3.67)$$

$$N = 36;\ R^2 = 0.778;\ R^2_{cvOO} = 0.721;\ R^2_{cvMO} = 0.727;\ F = 57.71;\ s = 0.328$$

In this model **ABOC** designates the average bond order for atom C and **TMSA** represents the total molecular surface area (Zefirov PC). As a van der Waals radii based descriptor, TMSA describes mainly steric interactions. The ABOC descriptor defines the degree of unsaturation/aromaticity of the structure and thus its flexibility, which could be directly related to the transport properties and the surface recognition profile of the molecule.

For the purposes of 3D-QSAR the most active compound from the series (N-[7-(hydroxyamino)-7-oxoheptyl]-4,6-dimethoxy-1H-indole-2-carboxamide) was

subjected to a systematic conformational search analysis followed by a full geometry optimization in water. Its optimized structure was then used as a template for the alignment procedure. A random separation of the initial dataset into training and test sets was done. The PLS analysis was used to extract the principal components. WLS analysis was then applied to visualize the steric and electrostatic fields maps.

As a result of the performed 3D-QSAR analysis it was found that:

- The presence of bulky substituents at 4^{th} or 6^{th} positions in the aromatic system leads to higher activity. The optimal groups for position six, which are expected to increase the inhibitor effect are OCH_3 or OC_2H_5.
- Substituents with WDV radii between 3 and 4 Å (CH_3, NH_2 and Cl) at the 3^{rd} position will contribute positively to the biological effect.
- The presence of electronegative residues in the 7^{th} position of the aromatic system and hydroxamic acid residue lead to increased biological activity.
- The presence of hydrocarbon or electron-donating residues at all other positions will increases the biological effect.

In addition a docking approach was used to find the size and shape of the ligand molecule that matches the receptor site best. Thus, we proceeded from all 36 structures collected in the database and the PDB structure of human enzyme HDAC8. The active site of the enzyme was selected and used to generate 3-center pharmacophore key as a complimentary part to the active site. The docking results enabled the identification of the following substructures characterizing every HDAC inhibitor: (i) surface recognition, (ii) linker and (iii) metal chelator. The metal chelator is responsible for the interaction between Zn^{2+} ion and the ligand. The surface recognition substructure binds the molecule to the right place in the enzyme structure and the linker substructure connects these two parts.

On the basis of the obtained 2D- and 3D- QSAR models new potentially active candidates were predicted: 7-chloro-6-ethoxy-N-(7-(hydroxyamino)-7-oxoheptyl)-4-methoxy-1H-indole-2-carboxamide, 6-ethoxy-N-(7-(hydroxyamino)-7-oxoheptyl)-4-methyl-1H-indole-2-carboxamide and 6-ethoxy-4-ethyl-N-(7-hydroxy-7-(hydroxyamino)heptyl)-1H-indole-2-carboxamide.

Both 2D- and 3D-QSAR approaches led to the conclusion that steric interactions play the most important role in the explanation of the revealed inhibitor effect.

2.3. QSAR study of antiplatelet agents

Article III proposes a QSAR model for the antiplatelet activity of 60 benzoxazinone derivatives against factor Xa. The platelets are small components in the blood a-nucleate disc shaped cells which play a central role in normal haemostasis and are key participants in pathologic thrombosis. Due to the significant limitations of the existing antiplatelet drugs, the search for new antiplatelet aggregation agents is now a challenging task for scientists working in the field.

Both linear and nonlinear modeling procedures were applied to reveal the relationship between the IC_{50} values and the structural characteristics of the antiplatelet agents. A logarithmic transformation was applied to the IC_{50} values in order to obtain distribution closer to normal.

The application of the breaking rule suggested a QSAR model involving the following 5 descriptors: *Maximum e-n attraction for atom N*, *Minimum exchange energy for bond C-O*, *HOMO-1 energy*, *Average nucleophilic reactivity index for atom N* and *Min (>0.1) bond order for atom H*. The quality of the model is determined by its higher statistical parameters: $R^2 = 0.821$, $R^2_{cv} = 0.773$, $F = 49.55$, $s^2 = 0.621$. Inspection of the differences between the predicted and observed $LogIC_{50}$ values led to the registration of two outliers. However, their removal was considered to be unnecessary and they were retained in the model.

Concurrently to the standard multilinear QSAR approach, an ANN methodology was also applied. The method used implements the feed-forward backpropagation algorithm. A sensitivity analysis was performed over the most important descriptors extracted from CODESSA PRO. This was done by building 1-1-1 ANN models and the first 5 descriptors that showed lowest error at the output were selected. For the purpose of ANN modeling the initial number of compounds was split into training (40 compounds) and test (20 compounds) datasets.

To ensure that the compounds from the test set are representative of the general population (the training set) the following procedure was applied:

i. All compounds were ordered in ascending order of their $logIC_{50}$ values.
ii. Every third compound was selected to be part of the test set.
iii. All the remaining compounds formed the training set.

The generated ANN model displayed the following statistical characteristics: the squares of the coefficients of determination for the training and test datasets were 0.825 and 0.909 respectively and their corresponding root-mean-squared (RMS) errors were 0.360 and 0.211.

The ability of the ANN model, to classify correctly the $logIC_{50}$ values was further analyzed by separation of the antiplatelets into 5 subcategories with low, moderate, middle, good and high activity. Confusion matrices representing the

ability of the ANN to predict exactly certain class of antiplatelet activity (calculated as N/Ns where the N the number of the antiplatelet agents in a certain class and Ns is the number of the successive 100% predictions) were also built. The inspection of the confusion matrices showed that the compounds with higher biological activity are better predicted than those displaying lower activity.

As a result of our study we concluded that the interaction of benzoxazinones with Factor Xa involvs two corresponding N-H and C-O bond pairs.

2.4. Correlation of Blood:Brain Penetration Using Structural Descriptors

Article IV presents QSAR models relating the blood-brain partition coefficients (logBB) of a diverse set of 113 compounds to a set of whole-molecule and fragment descriptors using CODESSA PRO and ISIDA programs. The ability of the drugs to penetrate the blood-brain barrier (BBB) is of fundamental importance for drug design. The blood-brain partition coefficient (logBB) is a determining factor for the effectiveness of the drugs targeting the central nervous system.

The Best Multi Linear Regression (BMLR) algorithm as implemented in CODESSA PRO was used to generate the best possible model from a set of 800+ collinear descriptors. Concurrently the ISIDA (In Silico Design and Data Analysis) program was employed to generate up to 1300 structure-property models involving one linear and two non-linear fitting equations using a large number of fragment descriptors.

The application of the "breaking point" rule indicated that the optimal number of parameters for the CODESSA PRO model is 5.

$$LogBB = -0.148*NDB + 0.240*CLogP - 0.161*HDCPSA - 0.138*KFI + 6.771*MPCAT + 0.179$$

$n= 127$, $k=5$, $R^2=0.631$, $R^2_{CV} = 0.586$, $F= 41.4$, $s^2= 0.249$

To improve the model all outliers characterised with deviations two times greater than the standard deviation of the data were removed. The final QSAR model is:

$$LogBB = 0.224*CLogP - 0.176*KFI - 0.131*NDB - 0.163*HDCPSA + 4.744*MPCAT + 0.378$$

$n = 113$, $k = 5$, $R^2 = 0.781$, $R^2_{CV} = 0.752$, $F= 76.1$, $s^2 = 0.123$

The notations used are as follows: NDB – number of double bonds, HDCPSA – H-donors CPSA, KFI – Kier flexibility index and MPCAT – maximum partial charge for all atom types.

To validate the model an additional set of 19 compounds was used as an external test set. The predictive correlation coefficient R^2_{ext} and the standard deviation s^2 were respectively found to be 0.766 and 0.032.

The alternative ISIDA approach generated more than 2800 fragment descriptors including atoms, bonds and atoms/bonds sequences from 2 to 8 atoms and augmented atoms. Compound "116" was identified as an outlier and it was excluded from the initial dataset of 113 compounds. A total number of 48 models with $R^2_{cv} \geq 0.6$ resulted in an "average model" with $R^2 = 0.872$ and $s^2 = 0.047$. The use of the same set of 19 compounds for external validation (as in case of CODESSA PRO modeling) resulted in $R^2_{ext} = 0.827$ and $s^2 = 0.0025$.

Despite the lower statistical parameters reported, CODESSA PRO was found to be superior to ISIDA because of:

i) The smaller difference between R^2 and R^2_{ext} values (higher consistency of the model);
ii) fewer descriptors in the initial descriptor pool (lower probability for "by chance correlations");
iii) easier interpretation of the descriptors in the model and avoidance of "averaged model" characteristics.

However, both of the proposed QSAR models may become reliable tools for analog design to improve CNS penetration and for de novo modeling of $CNS^{+}/^{-}$ libraries.

2.5. QSAR modeling of the antifungal activity against Candida albicans

Article V investigates the antifungal activity of a series of 83 diverse organic compounds from the following chemical classes: cyanoboranes, fluconazoles, carbonylaminobenzoxazoles and imidazolylmethylindoles (salts are excluded). All of them are characterized by their well expressed activity against dimorphic fungus *Candida albicans* which causes a variety of superficial and deep-seated mycoses.

The initial dataset was divided into training (72 compounds) and test (11 compounds) subsets. The minimum inhibitory concentration (MIC) values were studied. Since, different literature sources were used all MIC values were transformed into μmol/L. A logarithmic function was used to approach Gaussian distribution of the data.

A modified QSAR procedure involving the following steps was applied:

1. All compounds were arranged in ascending order according to their logMIC values.
2. The initial set was separated into three subsets (conditionally denoted as A, B and C) by selection of every third point from the original dataset.
3. Three new datasets were constructed using the following binary sums: A+B, A+C and B+C.
4. The BMLR method was applied to the sets obtained in step 3.
5. The complimentary parts to each of these three subsets (C, B and A respectively) were used as test sets.
6. All descriptors appeared in the generated in step 4 models were tested to obtain a general model including all existing compounds.
7. The general model was again validated using a classical internal crossvalidation procedure.

"Classical" QSAR models with up to 8 descriptors were generated using the BLMR procedure. Among all those generated, the QSAR model involving six theoretical descriptors was preferred:

$$LogMIC = -9.029*RNCA + 0.06771*HDCSA - 112.3*AVHA + 7.501*RNC + 0.01565*LogP^2 - 146.0*AERICA + 114.5$$

$R^2 = 0.788$, $F = 47.140$, $s^2 = 0.130$.

The interpretation of the descriptors involved is as follows:

RNCA – "Relative number of C atoms" – represents the ratio between the carbon atoms and all other atoms in the molecule. The descriptor value depends in a complex way on the number of double and triple bonds and the presence of heteroatoms in the molecule. Hence, more double or triple bonds and less heteroatoms will enhance the property.

HDCSA – "H-donors charged surface area" – The larger overlapping between the hydrogen donor atoms and the H atoms decreases the probability for creation of stronger donor-acceptor bonds and thus causes a negative impact on the antifungal effect.

AVHA – "Average valence of a H atom" – is a measure for the stability of bonds formed between the hydrogen atoms and the molecular backbone. The minus sign of the regression coefficient suggests that the more stable compounds should be characterized by a higher antifungal effect.

RNC – "Relative negative charge" – The positive regression coefficient implies that the increase of binding affinity associated with the heteroatoms will lower the antifungal effect.

$LogP^2$ – a parabolic relationship usually appears when the drugs are structurally non-specific and it could be a reasonable representation of the pharmacokinetic phase.

AERICA – "Average electrophilic reactivity index for a C atom" measures the energy stabilization when the system acquires an additional electronic charge ΔN, thus, the higher the descriptor value (the molecule will act as a stronger electrophile) the higher its antifungal effect.

3. CONCLUSIONS

The most frequently used QSAR approaches such as 2D-, 3D and ANN were sucessfully applied to various drug design problems. It was concluded that the results obtained using 2D- and 3D-QSARs are fully compatible and usually lead to a similar interpretation. The application of ANN leads to substantially less informative results and could only be used for estimation of the activity of unknown compounds. However, we are convinced that the simultaneous application of all these methods would help scientists to avoid their disadvantages in the process of searching for more accurate predictions.

To summarize, the combined application of the QSAR methods reported above could significantly enhance our knowledge of the mechanism of drug action, as well as to provide scientists with reliable tools for accurate structural predictions.

REFERENCES

1. Brown, C.; Frazer, T. *Trans. Roy. Soc. Edinburgh*, **1868**, *25*, 151.
2. Richardson, J. *Medical Times and Gazette*, **1868**, *2*, 703.
3. C. Richet, *C. R. Seances Soc. Biol*., **1893**, *9*, 775.
4. Meyer, H. *Arch. Experim. Pathol. und Pharmakol*., **1899**, *42*, 109.
5. Overton, E. *Z Physik. Chem*., **1897**, *22*, 189.
6. Bell, R. P. *in* "Correlation Analysis in Chemistry: Recent Advances" eds N.B. Chapman and J. Shorter, Plenum Press, New York, **1978**, pp 55–84.
7. Charton, M., *in* "Advances in Quantitative Structure-Property Relationships" ed. M. Charton, JAI Press, Inc., Greenwich, **1996**, pp 171–219.
8. Hansch, C.; Maloney, P. P.; Fujita, T. and Muir, R.M. *Nature*, **1962**, *194*, 178.
9. Hansch, C.; Muir, R. M.; Fujita, T.; Maloney, P. P.; Geiger F. and Streich, M. *J. Am. Chem. Soc.*, **1963**, *85*, 2817.
10. R.W. Taft and I.C. Lewis, *J. Am. Chem. Soc*., **1959**, *81*, 5343.
11. Delisle, R. K.; Lowrie, J. F.; Hobbs, D. W. and Diller, D. J. *Curr. Comp.-Aided Drug Des.* **2005**, *4*, 325.
12. Ramsden, C. A. (ed.), "Quantitative Drug Design", Vol. 4, of "Comprehensive Medicinal Chemistry", eds. Hansch, C.; Sammes, P. G. and Taylor, J. B. Pergamon, Oxford, **1990**.
13. Kubinyi, H. "QSAR: Hansch Analysis and Related Approaches", VCH, Weinheim, **1993**.
14. Kubinyi, H. *in* "Burger's Medicinal Chemistry", Vol. I, ed. M. E. Wolff, Wiley, New York, 5th edn., **1995**, pp. 497–571.
15. Hansch C. and Leo, A. "Exploring QSAR. Fundamentals and Applications in chemistry and Biology", American Chemical Society, Washington, DC, **1995**.
16. van de Waterbeemd, H. (ed.), "Structure Property Correlations in Drug Research", Academic Press, Austin, TX, **1996**.
17. Cramer, R. D. and Milne, M. *Abstracts of Papers of the Am. Chem. Soc. M.*, April **1979**, Computer Chemistry Section, no. 44.
18. Cramer, R. D.; Patterson, D. E. and J. D. Bunce, *J. Am. Chem. Soc.*, **1988**, *110*, 5959.
19. SYBYL/QSAR, Molecular Modelling Software, Tripos Inc., 1699 S. Hanley Road, St. Louis, MO 63944, USA.
20. Livingstone, D. J. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 195.
21. Mattioni, B. E.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 94.
22. Kauffman, G. W.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1553.
23. Xiao, Z.; Xiao, Y.-D.; Feng, J.; Golbraikh, A.; Tropsha, A.; Lee, K.- H. *J. Med. Chem.* **2002**, *45*, 2294.
24. Burden, F. R.; Winkler, D. A. *J. Med. Chem.* **1999**, *42*, 3183.
25. Sheridan, R. P.; Feuston, B. P.; Meiorov, V. N.; Kearsley, S. K. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1912.
26. Katritzky, A. R.; Fara, D. C.; Petrukhin, R. O.; Tatham, D. B.; Maran, U.; Lomaka, A.; Karelson, M. *Curr. Top. Med. Chem.* **2002**, *2*, 1333–1356.
27. Gedeck, P.; Rohde, B. and Bartels, C. *J. Chem. Inf. Model.* **2006,** *46,* 1924.

28. Lao, Y.; Leong, H. W. PRICAI 2002: Trends in Artificial Intelligence : 7th Pacific Rim International Conference on Artificial Intelligence, Tokyo, Japan, August 18–22, 2002. Proceedings, p 345
29. http://www.rightwingnation.com/asmotext/stats/correlation.html
30. Saunders, M. *J. Am. Chem. Soc.,* **1987**, *109*, 3150
31. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart. AM1: A New General Purpose Quantum Mechanical Model. *J. Am. Chem. Soc.*, **1985**, *107*, 3902.
32. J. J. P. Stewart, Optimization of Parameters for Semiempirical Methods IV: Extension of MNDO, AM1, and PM3 to more Main Group Elements, *J. Mol. Model*. **2004**, *10*, 155–164.
33. M. J. S. Dewar and W. Thiel. Ground States of Molecules, 38. The MNDO Method. Approximations and Parameters. *J. Am. Chem. Soc.*, **1977,** 99:4899–4907.
34. Yang, Z.Z.; Wang C.S.; *J. Theor. Comp. Chem.*, **2003**, *2*, 273.
35. Arulmozhiraja, S.; Morita, M. *Chem. Res. Toxil.,* **2004**, *17*, 348.
36. Liu, S.S.; Cui, S.H.; Yin, D.Q.; Shi, Y.Y.; Wang, L.S. *Chin. J. Chem.*, **2003**, *21*, 1510.
37. Chiu, T.L.; So, S.S. *J. Chem. Inf. Comp. Sci.*, **2004**, *44*, 147.
38. Chin, T.L.; So, S.S. *J. Chem. Inf. Comp. Sci.*, **2004**, *44*, 154.
39. Lin, Z.H.; Wu, Y.Z.; Quan, X.J.; Zhou, Y.G.; Ni, B.; Wan, Y. *Lett. Pept. Sci.*, **2002**, *9,* 273.
40. Agrawal, V.K.; Mishra, K.; Khadikar, P.V. *Oxid Commun*, **2003**, *26*, 14.
41. CODESSA PRO Software, University of Florida, 2002 – www.codessa-pro.com/
42. Mekenyan, O. G.; Karabunarliev, S. H.; Ivanov, J. M.; Dimitrov, D. N. *Comput. Chem.* **1994**, *18*, 173.
43. http://www.talete.mi.it/download.htm
44. Chem-X, Chem-X was developed at Chemical Design Ltd. (now part of Accelrys), 1994
45. http://www.accelrys.com/products/tsar/
46. Crown, H., Ed. *Comprehensive Drug Design;* Pergamon Press: New York 1990, p 19.
47. Crawley M. J. "Statistics: An Introduction using R", Wiley, New York, **2005** pp. 7–8
48. Eriksson, L.; Johansson, E.; Muller, M.; Wold, S. *J. Chemometrics* **2000**, *14*, 599.
49. Stanton, D. *J. Chem. Inf. Comput. Sci.* **2003,** *43,* 1423.
50. Stanton, D. T.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 109.
51. Darlington, R. B. "Regression and linear models". New York: McGraw-Hill, 1990.
52. Karelson, M. *Molecular descriptors in QSAR/QSPR*. Wiley-Interscience: New York, 2000.
53. Hoskuldsson, A. *Chemom. Intel. Lab. Sys.* **2001**, *55*, 23–38.
54. Drapper, N. R.; Smith, H. "Applied Regression Analysis". Wiley, New York, 1981.
55. Katritzky, A. R.; Mu, L.; Lobanov, V. S.; Karelson, M. *J. Phys. Chem*. **1996**, *100*, 10400.
56. Kubyini, H., ed. "3D QSAR in Drug Design, Theory, Methods, and Applications" ESCOM Science Publishers, B.V., Leiden, 1993.
57. DePriest, S.A., et al., *J. Am. Chem. Soc.*, **1993**, *115***,** 5372.
58. Calder, J.A., et al., *J. Comput.-Aided Mol. Des.*, **1993**. *7***,** 45.
59. McCulloch, W. S.; Pitts, W. *Bull. of Math. Bio.*, 1943, **5**, 115.

9

60. Hoskins, J. C.; Himmelbau, D. M. *Comput. Chem. Eng.*, **1988**, *12*, 881.
61. Qian, N.; Sejnowski, t. J. *J. Mol. Biol.*, **1988**, *202*, 865.
62. Sigman, M. E.; Rives, S. S. *J. Chem. Inf. Comput. Sci.*, **1994**, *34*, 617.
63. Rumelhart, D. B. "Parallel Distributed Processing", Feldman, J. A.; Hayes, P. J.; Rumelhart, D. B., Ed.; The MIT Press, London, **1982**, *1*, 318.
64. Aoyama, T.; Suzuki, Y.; Ichikawa, H. *J. Med. Chem.*, **1990**, *33*, 905.
65. Andrea, T. A.; Kalayeh, H. *J. Med. Chem.*, **1991**, *34*, 2824.
66. So, S.; Richards, W. G. *J. Med. Chem.*, **1992**, *35*, 3201.
67. Manallack, D. T..; Ellis, D. D.; Livingstone, D. J. *J. Med. Chem.*, **1994**, *34*, 3758.
68. Song, X.; Chen, Z.; Yu, R. *Computers Chem.*, **1993**, *17*, 303.

SUMMARY IN ESTONIAN

QSAR meetodi töövõtteid (2D, 3D, ANN) kasutati edukalt lahendamaks erinevaid ravimidisaini ja arendusega seotud probleemide uurimiseks. Leiti, et 2D ja 3D meetodil saadud QSAR mudelid on võrreldavad ja sarnaselt interpreteeritavad. ANN meetodi tulemused olid vähem informatiivsed ja võimaldasid ligikaudselt hinnata aktiivsust ainult mitteteadaolevatele ühenditele. Sellest hoolimata oleme veendunud, et mainitud kolme meetodi üheaegne kasutamine aitab teadlasi edasistel uuringutel saada täpsemaid tulemusi ainete omaduste ennustamisel.

Ülalmainitud meetodite kombineeritud kasutamine QSAR analüüsi meetodi juures aitab paremini mõista ravimite toime mehhanismi ning pakkuda teadlastele usaldusväärsemat ja täpsemat töövahendit struktuuranalüüside läbiviimiseks.

ACKNOWLEDGMENTS

I would like to thank my doctoral advisor Professor Mati Karelson for his excellent guidance throughout my research.
I would like to express my sincere gratitude to Kenan Professor Alan R. Katritzky for his support and guidance.

PUBLICATIONS

I

Slavov, Svetoslav; Atanassova, Mariyana; Galabov, Boris.
QSAR analysis of the anticancer activity of 2,5-disubstituted 9-aza-anthrapyrazoles. QSAR & Combinatorial Science (2007), 26(2), 173–181.

II

Katritzky, Alan R.; Slavov, Svetoslav H.; Dobchev, Dimitar A.; Karelson, Mati.
Comparison between 2D and 3D-QSAR approaches to correlate inhibitor activity for a series of indole amide hydroxamic acids.
QSAR & Combinatorial Science (2007), 26(3), 333–345.

III

Katritzky, Alan R.; Pacureanu, Liliana M.; Slavov, Svetoslav; Dobchev, Dimitar A.; Karelson, Mati. **QSAR study of antiplatelet agents.** Bioorganic & Medicinal Chemistry (2006), 14(22), 7490–7500.

IV

Katritzky, Alan R.; Kuanar, Minati; Slavov, Svetoslav; Dobchev, Dimitar A.; Fara, Dan C.; Karelson, Mati; Acree, William E.; Solov'ev, Vitaly P.; Varnek, Alexandre. **Correlation of blood-brain penetration using structural descriptors.** Bioorganic & Medicinal Chemistry (2006), 14(14), 4888–4917.

V

Katritzky, Alan R.; Slavov, Svetoslav; Dobchev, Dimitar A.; Karelson, Mati. **QSAR modeling of the antifungal activity against Candida albicans for a diverse set of organic compounds.** Submitted for publication to *Journal of Computer-Aided Molecular Design.*

CURRICULUM VITAE

SVETOSLAV HRISTOV SLAVOV

Born:	10/29/1974, Sliven, Bulgaria
Citizenship:	Bulgarian
Marital Status:	Married with two children
Address:	University of Florida Department of Chemistry Center for Heterocyclic Compounds CRB 229, Gainesville 32611, FL, USA
Tel:	+1 352 392 0554
E-mail:	svetoslavslavov2000@yahoo.co.uk

Education

1992–1996	Student, Faculty of Chemistry, University of Sofia, Bulgaria, B.Sc. in 1996.
1996–1997	Faculty of Chemistry, University of Sofia, Bulgaria, M.Sc. in 1997.
2004–2007	Ph.D. student, Department of Chemistry, University of Tartu, doctoral advisor Prof. Mati Karelson

Professional Experience

2001–2005	Assistant professor – Sofia University "St. Kliment Ohridski", Bulgaria
2003–2003	Young Researcher, IMAGETOX, 5th European Framework Program, Milano, Italy
2003–2003	Young Researcher, IMAGETOX, 5th European Framework Program, Tartu, Estonia
2005–present	Visiting Scholar, Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, USA

Publications

1. Slavov, Svetoslav; Djunlieva, Maria; Ilieva, Sonia; Galabov, Boris. **Quantitative structure-activity relationship analysis of the substituent effects on the binding affinity of derivatives of trimetazidine.** Arzneimittel Forschung (2004), 54(1), 9–14.
2. Slavov, Svetoslav; Atanassova, Mariyana; Galabov, Boris. **QSAR analysis of the anticancer activity of 2,5-disubstituted 9-aza-anthrapyrazoles.** QSAR & Combinatorial Science (2007), 26(2), 173–181.
3. Katritzky, Alan R.; Slavov, Svetoslav H.; Dobchev, Dimitar A.; Karelson, Mati. **Comparison between 2D and 3D-QSAR approaches to correlate inhibitor activity for a series of indole amide hydroxamic acids.** QSAR & Combinatorial Science (2007), 26(3), 333–345.
4. Katritzky, Alan R.; Slavov, Svetoslav H.; Dobchev, Dimitar A.; Karelson, Mati. **Rapid QSPR model development technique for prediction of vapor pressure of organic compounds.** Computers & Chemical Engineering (2007), 31(9), 1123–1130.
5. Katritzky, Alan R.; Slavov, Svetoslav H.; Dobchev, Dimitar A.; Karelson, Mati. **QSPR modeling of UV absorption intensities.** Journal of Computer-Aided Molecular Design (2007), 21(7), 371–377.
6. Katritzky, Alan R.; Pacureanu, Liliana M.; Slavov, Svetoslav; Dobchev, Dimitar A.; Karelson, Mati. **QSAR study of antiplatelet agents.** Bioorganic & Medicinal Chemistry (2006), 14(22), 7490–7500.
7. Katritzky, Alan R.; Kuanar, Minati; Slavov, Svetoslav; Dobchev, Dimitar A.; Fara, Dan C.; Karelson, Mati; Acree, William E.; Solov'ev, Vitaly P.; Varnek, Alexandre. **Correlation of blood-brain penetration using structural descriptors.** Bioorganic & Medicinal Chemistry (2006), 14(14), 4888–4917.
8. Katritzky, Alan R.; Slavov, Svetoslav; Dobchev, Dimitar A.; Karelson, Mati. **QSAR modeling of the antifungal activity against Candida albicans for a diverse set of organic compounds.** Submitted for publication to *Journal of Computer-Aided Molecular Design.*

ELULOOKIRJELDUS

SVETOSLAV HRISTOV SLAVOV

Sündinud:	10/29/1974, Sliven, Bulgaaria
Kodakondsus:	Bulgaaria
Perekonnaseis:	abielus, 2 lapse isa
Isikukood:	7410295821
Aadress:	University of Florida Department of Chemistry Center for Heterocyclic Compounds CRB 229, Gainesville 32611, FL, USA
Tel:	+1 352 392 0554
E-mail:	svetoslavslavov2000@yahoo.co.uk

Haridus

1992–1996	Keemia teaduskond, Sofia Ülikool, Bulgaaria, B.Sc. 1996.
1996–1997	Keemia teaduskond, Sofia Ülikool, Bulgaaria, M.Sc. 1997.
2004–2007	Tartu Ülikool, Füüsika-Keemiateaduskond, doktoriõpe, juhendaja prof. Mati Karelson

Teenistuskäik

2001–2005	Assistent – Sofia Ülikool “St. Kliment Ohridski”, Bulgaaria
2003–2003	Nooremteadur, IMAGETOX, 5 Euroopa Raamprogramm, Milaano, Itaalia
2003–2003	Nooremteadur, IMAGETOX, 5 Euroopa Raamprogramm, Tartu, Eesti
2005–praeguseni	Külalisteadur, Heterotsükliliste Ühendite Keskus, Keemia teaduskond, Florida Ülikool, USA

36

Publikatsioonid

1. Slavov, Svetoslav; Djunlieva, Maria; Ilieva, Sonia; Galabov, Boris. **Quantitative structure-activity relationship analysis of the substituent effects on the binding affinity of derivatives of trimetazidine.** Arzneimittel Forschung (2004), 54(1), 9–14.
2. Slavov, Svetoslav; Atanassova, Mariyana; Galabov, Boris. **QSAR analysis of the anticancer activity of 2,5-disubstituted 9-aza-anthrapyrazoles.** QSAR & Combinatorial Science (2007), 26(2), 173–181.
3. Katritzky, Alan R.; Slavov, Svetoslav H.; Dobchev, Dimitar A.; Karelson, Mati. **Comparison between 2D and 3D-QSAR approaches to correlate inhibitor activity for a series of indole amide hydroxamic acids.** QSAR & Combinatorial Science (2007), 26(3), 333–345.
4. Katritzky, Alan R.; Slavov, Svetoslav H.; Dobchev, Dimitar A.; Karelson, Mati. **Rapid QSPR model development technique for prediction of vapor pressure of organic compounds.** Computers & Chemical Engineering (2007), 31(9), 1123–1130.
5. Katritzky, Alan R.; Slavov, Svetoslav H.; Dobchev, Dimitar A.; Karelson, Mati. **QSPR modeling of UV absorption intensities.** Journal of Computer-Aided Molecular Design (2007), 21(7), 371–377.
6. Katritzky, Alan R.; Pacureanu, Liliana M.; Slavov, Svetoslav; Dobchev, Dimitar A.; Karelson, Mati. **QSAR study of antiplatelet agents.** Bioorganic & Medicinal Chemistry (2006), 14(22), 7490–7500.
7. Katritzky, Alan R.; Kuanar, Minati; Slavov, Svetoslav; Dobchev, Dimitar A.; Fara, Dan C.; Karelson, Mati; Acree, William E.; Solov'ev, Vitaly P.; Varnek, Alexandre. **Correlation of blood-brain penetration using structural descriptors.** Bioorganic & Medicinal Chemistry (2006), 14(14), 4888–4917.
8. Katritzky, Alan R.; Slavov, Svetoslav; Dobchev, Dimitar A.; Karelson, Mati. **QSAR modeling of the antifungal activity against Candida albicans for a diverse set of organic compounds.** Submitted for publication to *Journal of Computer-Aided Molecular Design.*

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.

37

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.