DISSERTATIONES CHIMICAE UNIVERSITATIS TARTUENSIS
39

QSPR MODELING OF COMPLEXATION AND DISTRIBUTION OF ORGANIC COMPOUNDS

DAN CORNEL FARA

TARTU UNIVERSITY
PRESS

Department of Chemistry, University of Tartu, Estonia

Dissertation is accepted for the commencement of the degree of Doctor of Philosophy in Chemistry on March 23, 2004, by the Doctoral Committee of the Department of Chemistry, University of Tartu.

Opponent: Professor Dr. Ülo Lille, Member of Estonian Academy of Sciences
Tallinn Technical University

Commencement: April 28, 2004 at 2 Jakobi Str., room 430

Tartu Ülikooli Kirjastus
Tiigi 78, Tartu 50410
Tellimus nr. 132

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

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

I. Katritzky, A. R.; Fara, D. C.; Yang, H.; Karelson, M.; Suzuki, T.; Solov'ev, V. P.; Varnek, A. **Quantitative Structure-Property Relationship Modeling of β-Cyclodextrin Complexation Free Energies.** *J. Chem. Inf. Comp. Sci., **ASAP***

II. Katritzky, A. R.; Fara, D. C.; Karelson, M. **QSPR of 3-Aryloxazolidin-2-one Antibacterials.** *Bioorg. Med. Chem., in **press***.

III. Katritzky, A. R.; Kuanar, M.; Fara, D. C.; Karelson, M. **QSPR Treatment of Rat Blood/Air, Saline/Air and Olive Oil/Air Partition Coefficients Using Theoretical Molecular Descriptors**. *Bioorg. Med. Chem., in **press***.

IV. Katritzky, A. R.; Tämm, K.; Kuanar, M.; Fara, D. C.; Oliferenko, A.; Oliferenko, P.; Huddleston, J. G.; Rogers, R. D. **Aqueous Biphasic Systems. Partitioning of Organic Molecules: A QSPR Treatment.** *J. Chem. Inf. Comp. Sci.* **2004**, *44(1)*, 136–142.

V. Andersson, P. L.; Maran, U.; Fara, D.; Karelson, M.; Hermens, J. L. M. **General and Class Specific Models for Prediction of Soil Sorption Using Various Physicochemical Descriptors.** *J. Chem. Inf. Comput. Sci.* **2002**, *42(6)*, 1450–1459.

LIST OF ABBREVIATIONS

CODESSA	COmprehensive DEscriptors for Structural and Statistical Analysis
SMF	Substructural Molecular Fragments
QSPR	Quantitative Structure – Property Relationship(s)
QSAR	Quantitative Structure – Activity Relationship(s)
MLR	Multilinear Regression
BMLR	Best Multilinear Regression
PC	Principal Component
PCA	Principal Component Analysis
PLS	Partial Least Squares
AM1	Austin Model 1
LO	Leave-One-Out cross-validation
LM	Leave-Many-Out cross-validation
MOPAC	Molecular Orbital PACkage
LFER	Linear Free Energy Relationship
NN	Neural Network
GA	Genetic Algorithm
ABS	Aqueous Biphasic System
PEG	Poly-ethylene glycol
TLL	Tie-Line Length

INTRODUCTION

Chemical structure is of paramount importance, not only to chemists, but also to all scientists and to humanity in general. As soon as a chemical structure is written, we have defined all the properties of the compound in question: physical, chemical, biological and technological. Of great importance is that chemical structure is invariant. Chemical structures offer a unique lasting and definitive means of representation of each compound.

Trying to deduce from chemical structure the properties of a compound has been an ongoing effort. Quantum chemists have made enormous advancements, particularly in the last twenty-five years. By using the semiempirical and *ab initio* methods now available, one can determine a great deal of information about a compound including its geometry, charge distribution, the way in which it interacts with UV-VIS and IR radiation and thus its spectral characteristics. Furthermore, one can predict the characteristics of its NMR spectra and many other properties. However, despite these major advantages, there still remain a very large number of properties that presently cannot be satisfactorily predicted using quantum-chemical methods. While this picture will change, it seems likely that for the foreseeable future we will need to approach the correlation and prediction of many properties, especially biological and technological properties, but also many physical and chemical properties, using other methods.

Although there is no hard dividing line, many of the manifestations of molecular structure fall into one of two major classes: (i) the influence of a specific portion of the molecule (as occurs with fatty tails, docking, and similar concepts); (ii) the influence of the whole molecule (as occurs in considerations of solubility, partition coefficients, migration, permeability, bioavailability and similar topics).

The method most frequently applied is that of "Quantitative Structure Property/Activity Relationships" (QSPR/QSAR)[1–11]. The effects of structural variation in a molecule are distinct in the two classes, and their rationalization has been approached from different standpoints. In general, most quantitative structure property relationships (QSPR) fall into class (ii) as manifestations of the whole structure.

This approach attempts to use a set of structures for which a property has been measured and to relate a quantitative value of the property in question with the chemical structures of the compounds that has been measured. Once a reliable equation has been obtained, it is possible to predict that same property for other structures not yet measured or even not yet prepared.

Equations that are set up in this way utilize what are known as "molecule descriptors" of the chemical structures. A descriptor is any parameter that can be determined quantitatively from the structure of a molecule. It is well-known

that the regression equations (QSPR models), whether derived in a purely empirical style from an arbitrary set of molecule descriptors or from a pre-selected set of parameters based on theoretical grounds for a connection with a particular property, can provide valuable insight on which structural characteristics control the physicochemical, biological and environmental behavior of a compound.

In many cases, the QSPR approach has to be accounted as the first step in the development of the whole theoretical modeling scale, by enlightening the major structural influences on the property of interest.

One of the most studied properties in QSAR is the *distribution* of chemicals between two mediums, due to its wide application in medicinal chemistry, drug design and toxicology. The term *distribution coefficient* or *partition coefficient* commonly refers to the equilibrium distribution of a single substance between two solvent phases (pure substances or solutions) separated by a boundary, where at least one of the solvents is a condensed phase. The most used *partition coefficients* are: (i) octanol:water, which provides a thermodynamic measure of the tendency of the substance to prefer a non-aqueous or oily milieu rather than water (i.e. its hydrophilic/lipophilic balance), (ii) human or rat blood:air, tissue:air, blood:brain etc. that allow studying the *distribution* of lipophilic compounds into blood and into adipose tissue for the better understanding of the physiological distribution and toxicology of these substances, (iii) water:soil or soil sorption partition coefficient, which is often one of the key input parameters in models to estimate the mobility and fate of contaminants for environmental risk assessment procedures.

The host-guest *complexation* organic compounds, in particular of cyclodextrins, is another property with large applications in industrial, pharmaceutical, agricultural, and other fields, including the improvement of the solubility and stability of drugs and selectively binding materials that fit into the host cavities.

During the last decades, an increasing interest in the application of the QSPR approach for modeling of the above-mentioned properties has been recorded.

This Thesis presents a compilation of our work on the prediction of several distribution/partition coefficients and complexation of diverse sets of organic compounds. In Chapter 1, an overview of the QSPR modeling algorithm is given. Chapter 2 presents our own results in quantum chemical, molecular mechanical and QSPR modeling tasks that are of potential use in various domains of medicinal, pharmaceutical, and environmental chemistry.

1. LITERATURE OVERVIEW

QSPR – general algorithm

The molecular structure of an organic or inorganic compound determines its properties. However, the direct prediction of compound's properties using *ab initio* theory may be extremely difficult or even impossible. Therefore, the inductive establishment of QSPRs uses an indirect approach in order to tackle this problem (Figure 1). It depends on a set of compounds with known properties or activities that is used for the model building.

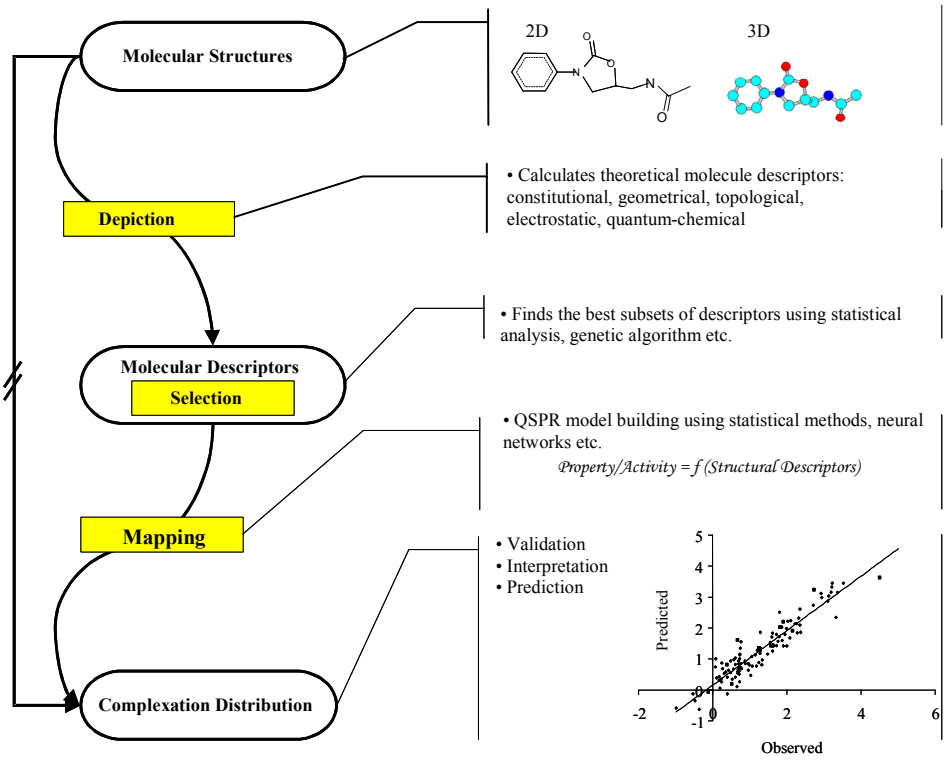

Figure 1. Flow chart of a QSPR model generation.

The QSPR model development and validation involves several major steps, as presented in Figure 1:

i. *decision* (of the property to be modeled);
ii. *construction* (of the data set);
iii. *depiction* (of the molecular structures);
iv. *selection* (of the most informative subset of descriptors);

v. *mapping* (of the QSPR model);
vi. *interpretation* (of the model);
vii. *validation* (of the model);
viii. *prediction* (of the property of interest).

A *Decision* implies choosing of the property to be studied that is highly dependent on the availability of experimental data. A large amount of heterogeneous experimental values lead to more stable and valid QSPR models.

A *Construction* represents the assembling of the data set that consists of 2D and 3D molecular structures obtained by molecular modeling (the three-dimensional rendering of molecules in an environment produced by computer graphics).

The molecular modeling, as part of the computational chemistry, includes the so-called computational methods by which the spatial coordinates of the atoms in molecules and the respective interconnecting bonds are produced. These computational methods can be divided into two major groups: (i) molecular mechanics methods, and (ii) quantum mechanics/quantum chemical methods. Group (ii) is usually subdivided into *ab initio* and semi-empirical methods[12].

In molecular mechanics, the potential energy of a given conformation of molecular systems is calculated as a function of the coordinates of their atomic nuclei, which are treated as Newtonian particles under the influence of a potential energy function or force field. A force field is given by equation (1) together with the data (parameters) required to describe the behavior of different kinds of atoms and bonds.

$$E = E_{\text{stretching}} + E_{\text{bending}} + E_{\text{torsion}} + E_{\text{non-bonded interaction}} \quad (1)$$

where E – potential energy, $E_{\text{stretching}}$ – bond stretching energies, E_{bending} – angle bending energies, E_{torsion} – torsional energies and $E_{\text{non-bonded interaction}}$ – energies of non-bonding interactions. The potential functions generated by molecular mechanics are often useful for comparing different configurations of a molecule.

Most molecular mechanics methods use empirical data to determine individual force constants (for example, bond stretch constants) and equilibrium ("strain-free") values for each geometrical characteristics of a molecule (for example, bond lengths)[13].

The quantum mechanical methods are based on Schrödinger equation[1] in which the electrons are described by the respective molecular wavefunctions. For molecular systems, the Schrödinger equation can only be solved approximately.

Typical *ab initio* electronic structure methods consider the potential energy operator as time-independent and solve only the spatial wavefunction. The a*b initio* calculations can be performed at the Hartree-Fock level of

approximation[12] or using various post-Hartree-Fock theories (configuration-interaction, multiconfiguration self-consistent field etc.).

Semi-empirical methods have been developed as approximations to *ab initio* techniques that were fitted to experimental data (e.g. structures and formation energies of organic molecules, spectroscopic data and ionization energies). In this case, the Hartree-Fock self-consistent field method is used to solve the Schrödinger equation with various approximations. Semi-empirical methods are usually classified according to their treatment of electron-electron interactions[13].

A *Depiction* is based on previous presented computational methods (*ab initio*, semi-empirical) and involves the calculation of the so-called molecular structure descriptors in order to encode the compounds. Those include the constitutional, topological, geometrical, electronic, and quantum chemical classes of descriptors as shown in Figure 2. The constitutional descriptors are fragment additive and reflect mostly the general properties of the compound. The topological descriptors are calculated using the mathematical graph theory applied to the scheme of atoms connections of the structure. The geometrical, electronic and quantum-chemical 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 interactions or hydrogen bonding (donor, acceptor).

A S*election* of descriptors involves finding the most informative subsets of descriptors from those in the descriptor pool due to the fact that models with too many variables lead to bad predictions because of overfitting. Various methods, like the classical forward selection, backward elimination and stepwise regression[15], or, more recently, the genetic algorithms (GA) can be used for the selection of the best combination of descriptors. The genetic algorithms have been shown to be very effective in performing descriptor selection in the case of large descriptor pools[15, 16].

A *Mapping* involves the application of the descriptors using (i) multilinear regression analysis, such as the Heuristic and best multilinear regression (BMLR) methods[17], (ii) multivariate statistical methods, such as principal component analysis (PCA) and partial least squares (PLS)[18–20], or (iii) non-linear methods, such as computational neural networks[21] (NNs) to build mathematical models linking the descriptors directly to the chemical property under investigation.

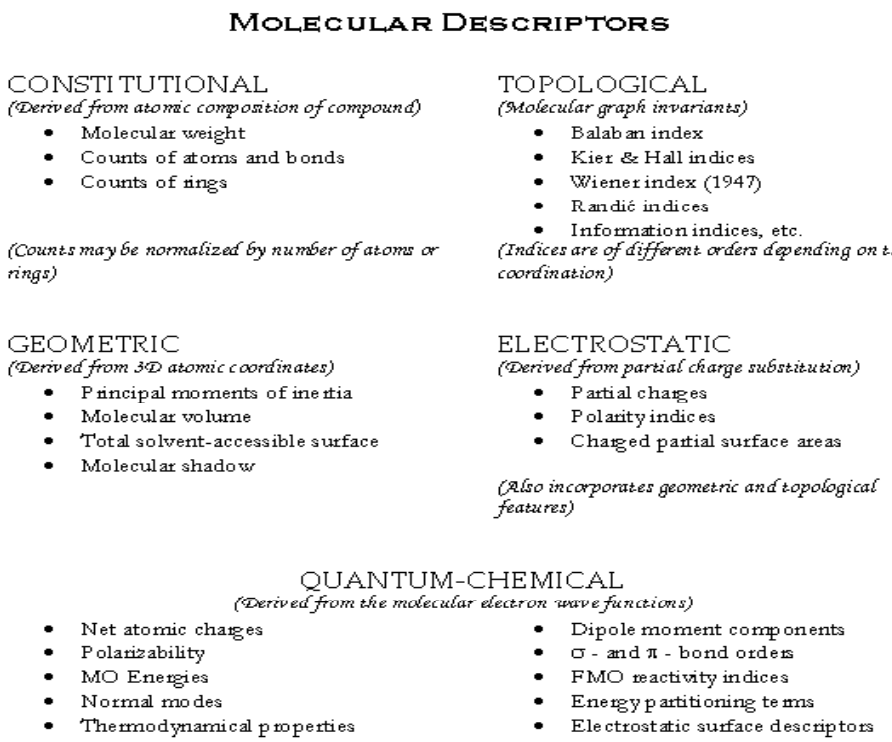

Figure 2. Molecular Descriptors[14].

The multilinear regression methods gave a linear regression equation in which the property, Y, is a linear function of descriptor values, X: $Y = f(X)$. 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.

In the principal component analysis (PCA), the original data matrix is decomposed into new latent variables, and the variation among the objects (e.g. organic compounds), is illustrated in score plots. The corresponding loading plots describe the corresponding variation among the descriptors. In the case of partial least squares (PLS), the latent information in the independent parameters is related to the dependent variable (i.e. property) through a weight vector. In both cases, the number of significant descriptors is determined by a cross-validation procedure[22] and the reliability and stability of the models are estimated by the following statistical measures: the variation explained in the X-matrix (descriptors) (R^2X), the variation explained in Y (property) (R^2Y), the cross-validated explained variance (Q^2), and the root-mean-square error of estimation (RMSEE) and prediction (RMSEP)[23].

A neural network is a non-linear method capable of modeling extremely complex functional relationships. It consists of a process of assimilation in the cognitive system similar to the neurological functions of the brain and is capable to predict new observations from other observations (modeling) after the executing of the so-called "learning process" from existing data. During the last decades, by using diverse "training" algorithms[16] various types of neural network architectures have been developed. Recent investigations involving computational neural networks and genetic algorithms serve as examples of the application of the QSPR methods[24]. Three-layer, feed-forward neural networks trained with a quasi-Newton method have provided excellent results in several QSPR studies[24].

An *Interpretation* of the model implies (i) the explanation of how each of the descriptors from the selected "best" subset contributes to the property of interest, (ii) the assessment of reliable physicochemical meaning to descriptors, especially for those provided by multivariate statistical analysis.

A *Validation* of the developed models is an important aspect of any QSPR study. Once a regression equation is obtained, it is important to determine its reliability and significance. Several procedures are available to assist in this. These can be used to check that the size of the model is appropriate for the data available, as well as to provide some estimate of how well the model can predict property for new molecules.

An Internal validation uses the data set from which the model is derived and checks for the internal consistency. For the internal validation, the parent data set can be divided into several subsets; e.g. the 1st, 4th, 7th, etc. entries go into the first subset (#1), the 2nd, 5th, 8th, etc. into the second subset (#2), and the 3rd, 6th, 9th, etc. into the third subset (#3). Then, three training sets *Set 1, Set 2 and Set 3* are prepared as combination of two subsets (#1 and #2), (#1 and #3), and (#2 and #3), respectively. The corresponding test sets relate to remaining subsets (#3, #2 and #1, respectively). For each training set, the correlation equation is derived with the same descriptors, and the obtained equation is used to predict the values of the property of interest for the compounds from the corresponding test set[25, 26].

An External validation evaluates how well the equation generalizes the data. The original data set is divided into two groups — the training set and the test set. The training set is used to derive a model that further is used to predict the properties of the test set members.

The Cross-validation repeats the regression many times on subsets of the data. Usually, each molecule is left out in turn, and the R^2 (cross-validated correlation coefficient) is computed using the predicted values of the missing molecules (*leave-one-out*, LO). Another possibility is to leave out more than one molecule at a time (i.e. 30%), though not all possible combinations of molecules are used, a concession that makes the computation tractable (*leave-*

many-out, LM). However, if n molecules are removed once from a total set of m, then nXm regressions are performed[27].

In multivariate statistical analysis, cross-validation is often used to determine the number of principal components retained in the model — those that give the highest cross-validated R^2 are to be considered.

The Monte Carlo cross-validation (MCCV)[28] was developed as an asymptotically consistent method in determining the number of components that can avoid an unnecessary large model and therefore decreases the risk of over-fitting in the model.

A *Prediction* involves the use of the developed QSPR models to estimate the property of interest for unknown compounds. If prediction is the main goal of the regression, then a test set data must be reserved strictly for evaluation until the equation is finalized: this reserved data cannot enter the regression process in any form.

This general QSPR methodology can be applied, in principle, to any property that can be related to the molecular structure, and it has been applied for the description and prediction of a wide variety of chemical, physical and environmental properties and activities[24].

2. DIVERSE QSPR MODELS USING CODESSA-PRO METHODOLOGY

In the present work, the CODESSA-PRO approach and software was used to perform QSPR correlations of diverse physicochemical properties with whole theoretical molecule descriptors. The main target of these studies was (i) to explore the applicability of QSPR methodology, (ii) to find the best regression equation for the prediction of the property of interest, and (iii) to explain how molecular interactions characterize the studied property.

2.1 QSPR of β — Cyclodextrin Complexation Free Energies

Article I presents the QSPR modeling of a large data set of experimental free energies of complexation for β-cyclodextrins with 218 organic compounds of different classes. Cyclodextrins (CDs), the cyclic oligomers of α-D-glucose are formed by the action of certain enzymes on starch. The native cyclodextrins include α-, β-, and γ-CD, which are crystalline, homogeneous, nonhygroscopic substances, and form cylindrical or doughnut-shaped molecules with their hydroxyl groups on the outside of the molecule. The characteristic host-guest property of CDs allows them to be used in numerous applications in industrial, pharmaceutical, agricultural, and other fields, including improving the solubility and stability of drugs and selectively binding materials.

Two different QSPR techniques (whole molecule and fragment descriptors) were used. One of them, realized in the program CODESSA-PRO, applies up to 902 different constitutional, geometrical, topological, electrostatic, quantum chemical, and thermodynamic molecular descriptors. Another technique, realized in the TRAIL program, uses several types of various fragment descriptors defined in the Substructural Molecular Fragments method (SMF). Both approaches and their combination led to statistically stable and predictive QSPR models. Indeed, the best models selected at the training stage for the full set of compounds, when applied for three training sets containing 2/3 of all molecules led to correct predictions of free energies for the compounds of corresponding test sets.

The CODESSA-PRO modeling of the binding energies for 1:1 complexation systems between 218 organic guest molecules and β-cyclodextrin resulted in a seven parameter equation with R^2=0.796 and R^2_{cv}=0.779. The fragment based TRAIL calculations gave a better fit with R^2=0.943 and R^2_{cv}=0.848 for 195 data points in the database. However, the analysis of advantages and disadvantages of both approaches lead to a conclusion that a combination of them would be most promising from practical viewpoint.

The seven descriptors involved in the CODESSA-PRO QSPR model can be classified as follows: (i) one as topological (*average complementary information content of 0th order,* $^0\overline{CIC}$), (ii) four as charge-distribution-related (*HACA-2/TMSA,* HACA*; H-acceptors FPSA, version 2,* FPSA; *maximum partial charge — Zefirov — for all atom types,* Q_i*; WNSA-1 weighted PNSA,* WNSA), (iii) one as geometrical (*ZX Shadow / ZX Rectangle,* ${}_SZX^R$), and (iv) one as semi-empirical molecular orbital (*LUMO energy,* ε_{LUMO}). A straightforward interpretation of appearing of these descriptors in the best model was rendered difficult by the complexity of the process involving conformational changes of the host, host-guest interactions and solvation-desolvation effects. However, the indirect links between those descriptors and physical phenomena involved in host-guest complexation were established (cf. Article I). The cross-validated correlation of the model gives R^2_{cv} value 0.779. Also, internal validation showed that the predicted R^2 values are in good agreement with the original QSPR model with the average correlation coefficients of 0.800 and 0.780 for the fitted and predicted sets, respectively (cf. Table 3 of Article I).

Two types of TRAIL calculations were carried out: without the differentiation of similar bonds in molecular fragments in chains and cycles (*Type 1* calculations) and accounting for the difference between them (*Type 2* calculations). Both *Type 1* and *Type 2* calculations performed on the parent set of 218 compounds led to about 40 models with $R^2 > 0.8$. However, only three models corresponded to R^2_{cv} values larger than 0.5. The application of these three models for calculations on three training and test sets led to a single model that gave reasonable statistical criteria for any of these sets. This is a linear model that involves the sequences of atoms and bonds containing from two to four atoms (cf. Table 4 of Article I). The statistical characteristics of these models are $R^2 = 0.943$, $F = 23.9$, $s^2 = 2.72$, $R^2_{cv} = 0.848$ and $s^2_{cv} = 9.42$ for *Type I*, and $R^2 = 0.967$, $F = 29.2$, $s^2 = 1.88$, $R^2_{cv} = 0.877$ and $s^2_{cv} = 13.49$ for *Type 2*.

The validation calculations for TRAIL models showed that average statistical criteria obtained at the training ($R^2 = 0.948$, $s^2 = 2.96$ and $R^2 = 0.971$, $s^2 = 1.92$ for *Type 1* and *Type 2* calculations, respectively) and test (R^2 *(pred.)*= 0.733, s^2 *(pred.)*= 8.80 and R^2 *(pred.)*= 0.783, s^2 *(pred.)*= 6.22 for *Type 1* and *Type 2* calculations, respectively) stages are close to those obtained in CODESSA-PRO calculations (cf. Table 5 of Article I).

For a better comparison between these two different approaches, the same seven theoretical molecular descriptors involved in the CODESSA-PRO model were correlated with the ΔG of complexation of β-cyclodextrins for two reduced data sets having: (i) 195, and (ii) 179 data points, respectively. The same molecules were eliminated from the full data set used as in the case of TRAIL approach (cf. Table 1 of Article I). The statistical characteristics for these two obtained models with CODESSA-PRO are as follows: (i) R^2=0.832, R^2_{cv}=0.817, F=132.0, s^2=4.90, and (ii) R^2=0.838, R^2_{cv}=0.822, F=126.4, s^2=4.78

and show a slight improvement of the correlations; that means that those proposed seven parameters to describe ΔG of complexation of β-cyclodextrins were well selected and they are not significantly dependent on the size and/or composition of the data set.

The free energies of complexation for the full set of compounds calculated with TRAIL correlate well with those obtained by CODESSA-PRO (R^2 = 0.829, s^2 = 4.54 and R^2 = 0.848, s^2 = 4.20 for *Type 1* and *Type 2* calculations, respectively (cf. Figure 5 of Article I)). In principle, this may allow one to apply both models simultaneously to estimate the binding affinity of any other guest molecule similar to those used at the training stage.

2.2 QSPR of 3-Aryloxazolidin-2-one Antibacterials

Article II presents the results of a QSPR modeling of *in vitro* minimum inhibitory concentration (MIC) — required for inhibiting growth of *Staphylococcus aureus* — with 60 3-aryloxazolidin-2-one antibacterials using CODESSA-PRO approach.

The increase during the last decade of bacterial resistance to antibiotics poses a serious concern for medical professionals. Oxazolidinones, a new class of synthetic antibacterials have been detected as promising agents against Gram-positive pathogenic and anaerobic bacteria. By selectively binding to the 50S ribosomal subunit, these compounds inhibit the bacterial translation at the initiation phase of protein synthesis. The observed activity of oxazolidinones, however, depends significantly on the substitution at the phenyl ring (cf. Figure 2 of Article II). Therefore, it is of large practical interest to develop QSPR models describing adequately this dependence.

Two different programs — CODESSA-PRO and CODESSA version 2.0 — were used to calculate various molecular (888) and fragmental (739) descriptors providing a total of 1627 descriptors for the QSPR modeling. These included the constitutional, geometrical, topological, electrostatic, quantum chemical, and thermodynamic descriptors. Based on the conclusions by Gregory *et al.*[29, 30], the fragment descriptors were defined for two distinct parts of the molecules of the given series: (i) the aryl group A, and (ii) the acetamido group B (cf. Figure 1 and Figure 2 of Article II). The multilinear QSPR models were developed using both the molecule and fragment descriptors.

Three QSPR models were obtained using (i) both molecule and fragment descriptors (N = 60, n = 7, R^2 = 0.820, R^2_{CV} = 0.758, F = 33.77, s^2 = 0.082), (ii) only molecule descriptors (N = 60, n = 7, R^2 = 0.795, R^2_{CV} = 0.727, F = 28.79, s^2 = 0.094) and (iii) only fragment descriptors (N = 60, n = 7, R^2 = 0.731, R^2_{CV} = 0.672, F = 20.19, s^2 = 0.123) for minimum inhibitory concentration (MIC).

These results indicate that in characterizing a complex biological property (antibacterial activity) by QSPR, the descriptors related to the whole molecule can be superior to the fragment descriptors. However, the best model obtained utilizes both fragment and molecule descriptors.

An interpretation of how these descriptors characterize the *in vitro* minimum inhibitory concentration (MIC) — required for inhibiting growth of *Staphylococcuc aureus* is given below.

The quantum chemical descriptors related to the H atoms indicate the importance of electrostatic interactions in determining the activity of compounds. Useful information about the most stable and the weakest bond in the molecule by characterizing the molecule as whole, by using σ and π molecular orbital coefficients, is provided. An estimate of the relative reactivity of oxygen atoms in the molecules of antibacterials can be related to the activation energy of the inhibiting growth of *Staphylococcus aureus* — the increase of the relative reactivity of oxygen atoms will cause a substantial decrease of log (1/MIC). It may also be speculated that the biological counterpart in the antibacterial activity is a soft nucleophile (e.g. — SH or double bond) and the conformational changes (rotation, inversion) in the molecule have importance in the biological transport of the molecules.

The best model was verified by using different validation methods that confirmed the correct predictions of the minimum inhibitory concentration (MIC) for the antibacterials. The cross-validated (*leave-one-out*) correlation of the seven-parameter model (eqn. (1)) gives R^2_{CV} value 0.758. Also, the internal validation showed that the predicted R^2 values are in good agreement with our original QSPR model with the average correlation coefficients of 0.825 and 0.700 for the fitted and predicted sets, respectively (cf. Table 4 in Article II).

The poorer correlation using only the fragment descriptors was not unexpected, because in the multilinear QSAR the interfragmental interactions are not accounted for. In the first approximation, the additional cross-terms involving descriptors from different molecular fragments might reflect these interactions. Such a possibility can be examined in future.

2.3 QSPR of Liquid-Air Partition Coefficients

Physiologically based pharmacokinetic (PBPK) models to describe the absorption, distribution, and elimination in animals and humans of volatile organic compounds (VOC) make frequent use of blood:air, saline:air, olive oil:air and tissue/blood partition coefficients to derive metabolic rate constants, and allometric equations. The solubility of a volatile organic chemical in blood, indicated as the blood:air partition coefficient, is one of the most important

physicochemical properties for understanding the pharmacokinetics of organic solvents.

Article III provides the results of a QSPR treatment to link the logarithmic function of rat blood:air, saline:air and olive oil:air partition coefficients (denoted by logK(b:a), logK(s:a), and logK(o:a), respectively) with theoretical molecular descriptors for a training set that consists of 100 diverse organic compounds.

Three QSPR models with squared correlation coefficients of 0.881, 0.926, and 0.922, respectively, were obtained. The verification of the predictive power of these models on a test set of 33 organic chemicals that were not included in the training set gave satisfactory squared correlation coefficients: 0.791 for rat blood:air, 0.794 for saline:air and 0.846 for olive oil:air.

The obtained models are given by eqns. 2, 5, and 6 and the descriptors involved in these equations are listed in Table 2 of Article III.

The analysis of the models indicates that the distribution of various compounds between liquid and air is strongly influenced by:

i. the intermolecular attractive forces in the condensed medium (positive contribution of dispersion and cavity formation effects in liquids and of hydrogen-bonding ability of the compounds);
ii. the compactness of the molecules (preferentially distribution over in solution of the highly compact molecules);
iii. a flexible conformation of the molecules supports the shift of the distribution into the liquid.

The relationship between the rat blood:air, and saline:air and olive oil:air partition coefficients themselves was also investigated, using the last two partition coefficients in order to predict the first (eqns 9–14 in Article III).

Based on these results, two alternative 5-parameters QSPR models were developed to calculate the rat blood:air partition coefficients, using four theoretical molecular descriptors and one external parameter (experimental or predicted value of logK(s:a)) — eq 15 ($R^2 = 0.924$) and eq 16 ($R^2 = 0.894$) in Article III.

Taking into account these promising results, it can be expected that this approach will also be applicable for predicting the partition coefficients of organic compounds in various other biological systems (i.e., tissue:air, tissue-blood etc. for both human and rat).

2.4 QSPR Applied on Aqueous Biphasic Systems

Aqueous biphasic systems (ABS) are formed by the addition of two (or more) water-soluble polymers or a polymer and salt to an aqueous solution above certain critical concentrations or temperature. ABS are unique because each of

the two non-miscible phases has over 80% water on a molal basis and possesses different solvent properties[31, 32]. Due to their highly aqueous and hence mild nature, which is consonant with the maintenance of macromolecular structure, ABS have been employed for the separation of biological macromolecules for over 40 years[32, 33]. ABS systems are also used in industrial biotechnology quality control for the detection of denaturation and degradation of proteins[34]. ABS partition coefficients have been suggested as viable alternatives to logP values in QSAR (quantitative structure-activity relationships) applied to biotechnological products[34, 35].

Article V reports the QSPR investigation of the partitioning of 29 small organic probes in a PEG-2000/$(NH_4)_2SO_4$ biphasic system using CODESSA-PRO approach.

A three-descriptor equation with the following statistical characteristics: N = 29, n = 3, $R^2 = 0.967$, $R^2_{CV} = 0.956$, F = 245.07, $s^2 = 0.017$ was developed for the partition coefficient (logD). All descriptors were derived solely from the chemical structure of the compounds. Using the same descriptors, a three-parameter model was also obtained for logP (octanol/water, $R^2 = 0.878$, $R^2_{CV} = 0.837$, F = 64.68, $s^2 = 0.236$); the predicted logP values were used as an external descriptor for modeling logD ($R^2 = 0.940$, $R^2_{CV} = 0.930$, F = 424.09, $s^2 = 0.029$).

The QSPR models with one, two, and three parameters each contain the gravitational index or the Kier and Hall indices as major descriptors. This indicates the importance of the "shape descriptors" for describing the partition behavior of small organic compounds in aqueous biphasic systems. It has also been shown that despite the best single parameter model includes the Kier and Hall index (order 3), nevertheless, the best two- and three-parameter models utilize the gravitational index for all bonds.

An overall interpretation of the descriptors involved in the models showed:

i. the preferential solubility in the PEG-rich phase of the compounds with complex atomic composition;
ii. a reduced distribution of highly polar solutes into the polymer-rich phase;
iii. a higher basicity of the PEG-rich phase as compared to the salt-rich one.

It is fairly obvious that the models presented in Article IV have some limitations because they have been calculated using logD values in a specific ABS (PEG-2000/$(NH_4)_2SO_4$). The salt and PEG may vary and the current models may not work properly for predicting logD in different ABS. Recently, Rogers *et al.*[36] showed the importance of the TLL (tie line length) for describing the ABS, that can be used for normalizing the logD scale (e.g. logD/TLL).

Taking into account the above-mentioned limitations, it is fair to say that the reported QSPR models help to improve and understand the prediction of partition coefficients in PEG/salt aqueous biphasic systems (ABS) for structurally diverse compounds.

2.5 General and Class Specific Models for Prediction of Soil Sorption

Article V reports the results obtained by testing five different QSPR approaches that are applied in the development of models used for the prediction of soil sorption partition coefficient (K_{OC}, normalized to organic carbon content), which is well-known as an important parameter in environmental risk assessment procedures.

The principal component analysis (PCA) and partial least squares (PLS) methods were applied to develop five general soil sorption models based on a training set of 68 compounds and using the following sets of physicochemical parameters (cf. Table 1 of Article V):

1. the logarithm of the octanol-water partition coefficient (logP) calculated with ClogP (model I)[37];
2. five descriptors calculated using QSPR Properties module implemented in the HyperChem Pro vers. 6.02 package (model II)[38];
3. a diverse set of descriptors that includes topological indices and quantum-chemical parameters (model III)[39];
4. a large set of descriptors derived from the DRAGON software (model IV)[40], and
5. a complex set of molecular descriptors calculated using CODESSA package (model V)[41].

The importance of external validation to determine the reliability and capability of the QSPR models was also studied and carried out using a test set of 274 compounds. The statistical characteristics of the models I — V are given in Table 2 from the Article V. Briefly, the internal validation displayed a variation of the cross-validated explained variance (Q^2) from 0.61 to 0.53, and the external validation gave a root-mean-square error of prediction (RMSEP) from 0.51 to 0.60.

In addition, the full data set was divided into classes of compounds and 14 separate QSPR models were calculated using the three first sets of descriptors. The corresponding models are listed in Table 4 of Article V and show that (i) the average of the variation explained in Y (K_{OC}) range from 0.61 to 0.68, and (ii) the average of the cross-validated explained variance (Q^2) vary from 0.61 to 0.52.

By analyzing all of these uni- and multivariate QSPR models, it has been shown that:

i. a basic univariate logP model is sufficient as a first screen of the compounds soil sorption potential, but it is significantly dependent on the quality of the logP estimate, which directly determines the outcome (for instance, the logP values calculated using ClogP program are more accurate and useful than those from HyperChem Pro vers. 6.02);

ii. for certain types of organic pollutants other descriptors, which reflect the bulk and polar characteristics, are essential, i.e. the connectivity indices that quantify the size and shape of the molecule etc.;
iii. the use of an external validation set is crucial to assess the predictive capacity of the model.

Finally, it can be concluded that the descriptors analyzed in Article V have various advantages and provide essential information in developing QSAR models to predict the soil sorption coefficients of specific classes of organic compounds.

3. CONCLUSIONS

The general algorithm of the QSPR methodology has been applied according to the following eight main steps: *decision* (of the property to be studied), *construction (*of the data set), *depiction* (of the molecular structures), *selection* (of the most informative subset of descriptors), *mapping*, *interpretation* and *validation* (of the QSPR model), and *prediction* (of the property of interest). The methods involved in each step are summarized and briefly defined.

The applicability of this algorithm for modeling the complexation and distribution of organic compounds has been examined and validated by the results presented in five articles: two of them report studies on complexation modeling (Articles I and II), and others the QSPR studies of various partition coefficients (Articles III – V).

I. Two different QSPR approaches using whole molecule (CODESSA-PRO) and fragment (TRAIL) descriptors were used to model the binding energies for 1:1 complexation systems between 218 organic guest molecules and β-cyclodextrin. CODESSA-PRO derived a seven parameter equation with $R^2 = 0.796$ and $R^2_{CV} = 0.779$ using the full data set of compounds. TRAIL calculations gave a better fit with $R^2 = 0.943$ and $R^2_{CV} = 0.848$, but for a limited subset of 195 data points in the database. Both advantages and disadvantages of each approach were discussed. It was revealed that a combination of two approaches is promising from the practical point of view.

II. Whole molecule and fragment descriptors were calculated for 60 3-aryloxazolidin-2-one antibacterials to relate the in vitro minimum inhibitory concentration (MIC) (required to inhibiting growth of *Staphylococcus aureus*). The treatment using CODESSA-PRO descriptors led to a seven-parameter model with $R^2 = 0.820$ and $R^2_{CV} = 0.758$.

III. Theoretical molecular descriptors were calculated for 100 diverse organic compounds with reliable experimental values of rat blood:air, saline:air and olive oil:air partition coefficients. Proceeding from these data, three QSPR models with squared correlation coefficients of 0.881, 0.926, and 0.922, respectively, were obtained and showed a predictive power of $R^2 = 0.791$ for rat blood:air, $R^2 = 0.794$ for saline:air and $R^2 = 0.846$ for olive oil:air by using 33 test organic chemicals not included in the training set.

IV. The distribution of 29 small organic probes in a PEG-2000/$(NH_4)_2SO_4$ biphasic system was investigated using a quantitative structure-property relationship (QSPR) approach. A three-descriptor equation with R^2=0.967 and R^2_{CV}=0.956 for the partition coefficient (logD) was obtained. All descriptors were derived solely from the chemical structure of the compounds. Using the same descriptors, a three-parameter model was also obtained for logP (octanol/water, R^2=0.878,

R^2_{CV}=0.837); predicted logP values were used as an external descriptor for modeling log D.

V. Diverse chemical descriptors divided into five sets were explored for use to screen the soil sorption of organic compounds. The multivariate statistical analysis methods (PCA and PLS) were used to derive uni- and multivariate models. Generally, it was observed that the univariate logP models were capable of capturing most of the variation and gives an indication of the sorption potential. The multivariate models were shown to include essential descriptors for modeling classes of compounds with specific chemical characteristics.

On the basis of the results obtained, it can be concluded that QSPR approach has enabled to aquire new valuable information about complexation and distribution of organic compounds. The reported models have importance for various domains of human activity such as medicine, pharmacy, and environment risk assessment.

REFERENCES

1. Hansch, C.; Leo, A. *Exploring QSAR, Fundamentals and Applications in Chemistry and Biology*; ACS: Washington, 1995.
2. Abraham, M. H. New Solute Descriptors for Linear Free Energy Relationships and Quantitative Structure-Activity Relationships, In *Quantitative Treatments of Solute/Solvent Interactions*; Politzer, P.; Murray, J. S.; Eds.; Elsevier: Amsterdam, 1994; pp 83–133.
3. Abraham, M. H.; Chadha, H. S.; Dixon, J. P.; Rafols, C.; Treiner, C. *J. Chem. Soc. Perkin Trans. 2* **1995**, *5*, 887–894.
4. Balaban, A. T. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 645–650.
5. Hilal, S. H.; Carreira, L. A.; Karickhoff, S. W. Estimation of Chemical Reactivity Parameter and Physical Properties of Organic Molecules Using SPARC. In *Quantitative Treatments of Solute/Solvent Interactions*; Politzer, P.; Murray, J. S.; Eds.; Elsevier: Amsterdam, 1994; pp 291–353.
6. Stuper, A. J.; Brugger, W. E.; Jurs, P. C. *Computer-assisted Studies of Chemical Structure and Biological Function*; John Wiley & Sons: New York, 1979.
7. Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *Chem. Soc. Rev.* **1995**, 279–287.
8. Kier , L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis*; John Wiley & Sons: New York, 1986.
9. Murray, J. S.; Politzer, P. A General Interaction Properties Function (GIPF): An Approach to Understanding and Predicting Molecular Interactions. In *Quantitative Treatments of Solute/Solvent Interactions*; Politzer, P.; Murray, J. S.; Eds.; Elsevier: Amsterdam, 1994; pp 243–289.
10. Randic, M.; Razinger M. On the Characterization of Three-Dimensional Molecular Structure. In *From Chemical Topology to Three-Dimensional Geometry*; Balaban, A. T.; Ed.; Plenum Press: New York, 1996; pp 159–236.
11. Lucic, B.; Trinajstic, N. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 121–132.
12. Lowe, J. P. *Quantum Chemistry*; 2nd ed. Academic Press, San Diego, California, 1993.
13. Burkert, U.; Allinger, N. L. (Eds.), *Molecular mechanics*, Am. Chem. Soc. Monograph, Washington D.C., 1982.
14. Katritzky, A. R.; Fara, D. C. *Annals of West University of Timisoara*, Ser. Chem. **2003**, *12*, 1205–1214.
15. Xu, L.; Zhang, W.-J. *Anal. Chim. Acta* **2001**, *446*, 477–483.
16. Deaven, D. M.; Ho, K. M. *Phys. Rev. Lett.* **1995**, *75*, 288–291.
17. Katritzky, A. R.; Lobanov, V. S.; Karelson, M.; Murugan, R.; Grendze, M. P.; Toomey, J. E. *Rev. Roum. Chim.* **1996**, *41*, 851–867.
18. Gabrielsson, J.; Lindberg, N.-O.; Lundstedt, T. *J. Chemom.* **2002**, *16*, 141–160.
19. Willett, P. *Similarity and Clustering in Chemical Information Systems*; John Wiley: New York, 1987.
20. Jackson, J. E. *A users guide to principal components*; John Wiley & Sons: New York, 1991.
21. Schneider, G.; Wrede, P. *Progr. Biophys. Mol. Biol.* **1998**, *70*, 175–222.
22. Wold, S. *Technometrics* **1978**, *20*, 397–405.
23. *SIMCA-P 8.0*; Umetrics AB, Umeå, Sweden, 2000.

24. Katritzky, A. R.; Fara, D.; Tatham, D.; Petrukhin, R.; Maran, U.; Lomaka, A.; Karelson, M. *Curr. Top. Med. Chem.* **2002**, *2*, 1333–1356
25. Katritzky, A. R.; Wang, Y.; Sild, S.; Tamm, T.; Karelson, M. *J. Chem. Inf. Comput. Sci.* 1998, *38*, 720–725.
26. Maran, U.; Karelson, M.; Katritzky, A. R. *Quant. Struct. Act. Relat.* 1999, *18*, 3–10.
27. Allen, D. M. *Technometrics* **1974**, *16*, 125–127.
28. Xu, Q.-S.; Liang, Y.-Z. *Chemom. Intell. Lab. Syst.*. **2001**, *56*, 1–11.
29. Gregory, W. A.; Brittelli, D. R.; Wang, C.-L. J.; Wuonola, M. A.; McRipley, R. J.; Eustice, D. C.; Eberly, V. S.; Bartholomew, P. T.; Slee, A. M.; Forbes, M. *J. Med. Chem.* **1989**, *32*, 1673–1681.
30. Gregory, W. A.; Brittelli, D. R.; Wang, C.-L. J.; Kezar, H. S.; Carlson, R. K.; Park, C.-H.; Corless, P. F.; Miller, S. J.; Rajagopalan, P.; Wuonola, M. A.; McRipley, R. J.; Eberly, V. S.; Slee, A. M.; Forbes, M. *J. Med. Chem.* **1990**, *33*, 2569–2578.
31. Albertsson, P.-A. *Nature* **1958**, *182*, 709–711.
32. Albertsson, P.-A. *Partition of Cell Particles and Macromolecules*; John Wiley & Sons: New York, 1986.
33. Rogers, R. D.; Eiteman, M. A. *Aqueous Biphasic Separations: Biomolecules to Metal Ions*; Plenum Press: New York, 1995.
34. Zaslavsky, B. Y. *Aqueous Two-Phase Partitioning: Physical Chemistry and Bioanalytical Applications*; Marcel Dekker: New York, 1994.
35. Zaslavsky, B. Y.; Mestechkina, N. M.; Miheeva, L. M.; Rogozhin, S. V.; Bakalkin, G. Ya.; Rjazhsky, G. G.; Chetverina, E. V.; Asmuko, A. A.; Bespalova, J. D.; Korobov, N. V.; Chichenkov, O. N. *Biochem. Pharmacol.* **1982**, *31*, 3757–3762.
36. Willauer, H. D; Huddleston, J. G.; Rogers, R. D. *Ind. Eng. Chem. Res.* **2002**, *41*, 1892–1904.
37. *Biobyte ClogP Manual* http://clogp.pomona.edu/chem/biobyte/manuals/ClogP_Formatted.html
38. *HyperChem Pro 6.02* www.hyper.com
39. Eriksson, L.; Johansson, E.; Müller, M.; Wold, S. *J. Chemometrics* **2000**, *14*, 599–616.
40. *Dragon 1.11* http://www.disat.unimib.it/chm/dragon.htm
41. Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *CODESSA: Reference manual version 2.0*; Gainesville, FL, U.S.A., 1994.

SUMMARY IN ESTONIA

ORGAANILISTE ÜHENDITE KOMPLEKSIMOODUSTUMISE JA JAOTUSE MODELLEERIMINE QSPR MEETODITE ABIL

Kvantitatiivsete struktuur-omadus sõltuvuste (QSPR) metodoloogia üldine algoritm on esitatav kaheksa peamise etapina: uuritava omaduse valik, andmebaasi koostamine, molekulide struktuuride esitusviis, kõige informatiivsemate deskriptorite valimine antud omaduse kirjeldamiseks, QSPR mudeli koostamine, interpreteerimine ja valideerimine, ning omaduse ennustamine saadud QSPR mudeli abil.

Ülaltoodud QSPR algoritmi rakendamise tulemused orgaaniliste ainete kompleksimoodustumise ja faasidevahelise jaotuse modelleerimiseks on esitatud viies artiklis: kaks neist käsitlevad komplekside teket (artiklid I ja II), ülejäänud kolm aga erinevaid jaotuskoefitsiente (artiklid III–V).

I. Modelleeriti kompleksi moodustavate süsteemide 1:1 seostumisenergiaid 218 orgaanilise aine ja β-tsüklodekstriini vahel. Selleks kasutati kahte erinevat QSPR lähenemist, mis lähtusid vastavalt kogu molekuli (CODESSA-PRO) ja molekulaarsete fragmentide (TRAIL) deskriptoritest. CODESSA-PRO abil saadi võrrand korrelatsiooni- ja ristvalideeritud korrelatsioonikoefitsientidega R^2=0.796 ja R^2_{CV}=0.779 kogu andmekogumi jaoks. TRAIL andis statistiliselt mõnevõrra parema tulemuse, vastavalt, R^2=0.943 ja R^2_{CV}=0.848, kuid piiratud andmekogumile (195 ainet). Analüüsiti mõlema meetodi eeliseid ja puudusi. Ilmnes, et nende kahe lähenemise kombinatsioon võib parandada tulemust.
II. Kogu molekuli ja fragmentide deskriptorid arvutati 60 3-akrüüloksasolidiin-2-oon antibakteritsiidse aine jaoks, et leida korrelatsiooni *in vitro* minimaalsete inhibeerimiskontsentratsioonidega (MIC) (vajalik aine kontsentratsioon *Staphylococcus aureus*-e bakterite kasvu inhibeerimiseks). CODESSA-PRO vahendusel saadi QSPR mudel korrelatsioonikoefitsientidega R^2=0.820 ja R^2_{CV}=0.758.
III. Arvutati teoreetilised molekulaardeskriptorid 100 orgaanilise ühendi jaoks erinevatest keemilistest klassidest, millele leidusid kõrgekvaliteedilised eksperimentaalsed jaotuskoefitsiendid: veri:õhk, füsioloogiline lahus:õhk ja oliivõli:õhk. Saadi vastavalt kolm QSPR mudelit, korrelatsioonikoefitsientidega R^2=0.881, R^2=0.926 ja R^2=0.922. Mudelite ennustusvõimet kinnitas jaotuskoefitsientide arvutus 33

täiendava ühendi jaoks (korrelatsioonikoefitsiendid vastavalt R^2=0.791, R^2=0.794 ja R^2=0.846).

IV. QSPR meetodit rakendati 29 madalmolekulaarse aine jaotumise uurimiseks PEG2000/$(NH_4)_2SO_4$ kahefaasilises süsteemis. Jaotuskoefitsiendi (logD) jaoks saadi kolmeparameetriline võrrand korrelatsioonikoefitsientidega R^2=0.967 ja R^2_{CV}=0.956. Kõik deskriptorid olid tuletatud lähtudes vaid ainete keemilisest struktuurist. Kasutades arvutatud deskriptoreid, saadi ka kolmeparameetriline mudel logP (oktanool/vesi jaotuskoefitsient) jaoks, R^2=0.878 ja R^2_{CV}=0.837. Viimase järgi arvutatud logP väärtusi kasutati lisadeskriptorina logD modellerimiseks.

V. Erinevaid keemilisi deskriptoreid, jaotatuna viide rühma, kasutati orgaaniliste ainete jaotuvuse kirjeldamiseks pinnases. Mudelite tuletamiseks kasutati multilineaarse regressiooni ja multivariantse statistilise analüüsi meetodeid (PCA ja PLS). Täheldati, et logP kirjeldab suurema osa variatsioonist terve andmekogumi jaoks ning annab üldise ettekujutuse sorptsiooni potentsiaalist. Aineklasside jaoks välja töötatud multivariantsetes mudelites lisandunud deskriptorid sisaldasid olulist informatsiooni nende spetsiifilise keemilise iseloomu kohta adsorptsiooniprotsessis.

Saadud tulemustest lähtuvalt võib järeldada, et QSPR metodoloogia on edukalt rakendatav orgaaniliste ainete komplekseerumisele ja jaotusele erinevates keskkondades. Esitatud mudelid omavad olulist tähtsust mitmetel inimtegevuse aladel nagu meditsiinis, farmaatsias ja keskkonnahoius.

ACKNOWLEDGMENTS

I would like to thank my doctoral advisor Professor Dr. Mati Karelson for his excellent guidance throughout my research. Being one of his fellows was the most pleasant and valuable experience.

I would like to express my warmest gratitude to Kenan Professor Dr. Alan R. Katritzky for his most valuable support and for giving me the opportunity to learn from his vast experience and knowledge.

I would also like to extend my thanks to my son, Dan, my wife, Dana, and my parents for their love, understanding and moral support during this period.

PUBLICATIONS

Katritzky, A. R.; Fara, D. C.; Yang, H.; Karelson, M.; Suzuki, T.; Solov'ev, V. P.; Varnek, A. **Quantitative Structure-Property Relationship Modeling of β-Cyclodextrin Complexation Free Energies.** *J. Chem. Inf. Comp. Sci., **ASAP***

Katritzky, A. R.; Fara, D. C.; Karelson, M. **QSPR of 3-Aryloxazolidin-2-one Antibacterials.** *Bioorg. Med. Chem., in **press***.

Katritzky, A. R.; Kuanar, M.; Fara, D. C.; Karelson, M. **QSPR Treatment of Rat Blood/Air, Saline/Air and Olive Oil/Air Partition Coefficients Using Theoretical Molecular Descriptors**. *Bioorg. Med. Chem., in **press***.

Katritzky, A. R.; Tämm, K.; Kuanar, M.; Fara, D. C.; Oliferenko, A.; Oliferenko, P.; Huddleston, J. G.; Rogers, R. D. **Aqueous Biphasic Systems. Partitioning of Organic Molecules: A QSPR Treatment.** *J. Chem. Inf. Comp. Sci.* **2004**, *44(1)*, 136–142.

Andersson, P. L.; Maran, U.; Fara, D.; Karelson, M.; Hermens, J. L. M. **General and Class Specific Models for Prediction of Soil Sorption Using Various Physicochemical Descriptors.** *J. Chem. Inf. Comput. Sci.* **2002**, *42(6)*, 1450–1459.

CURRICULUM VITAE

DAN CORNEL FARA

Born:	28 September 1967
Citizenship:	Romanian
Marital Status:	Married, one son (11 years old)
Address:	University of Tartu, Institute of Chemical Physics, Jakobi 2, Tartu, 51014, Estonia Tel: +372 7 375270 E-mail: dfara@theor.chem.ut.ee

Education

1991–1996	Student, Faculty of Chemistry, Biology, Geography, The West University of Timisoara, Romania. *B.Sc.* (chemistry) in 1996.
1996–1997	Graduate student, Faculty of Chemistry, Biology, Geography, The West University of Timisoara, Romania. *M.Sc.* (chemistry) in 1997.
2001–2004	Ph.D. student, Department of Chemistry, University of Tartu, doctoral advisor Prof. Mati Karelson.

Professional experience

1997–1999	Chemistry Teacher, Industrial High School Otelu-Rosu & Caransebes, Romania
1999–2001	Chemist Researcher, Institute of Chemistry, Romanian Academy, Timisoara Branch, Romania
2000–2001	Teaching Assistant, Department of Physical-Chemistry, Faculty of Pharmacy, University of Medicine and Pharmacy Timisoara, Romania
2001–present	Assistant Professor, Department of Physical-Chemistry, Faculty of Pharmacy, University of Medicine and Pharmacy Timisoara, Romania
2001–present	Visiting Scholar, Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, USA

Publications

1. Elenes, F.; Fara, D.; Salló, A.; Seclaman, E.; Kurunczi, L. Multivariate Analysis (PCA, PLS) on a Series of AS-Naphthols, *Annals of West University of Timisoara*, Ser. Chem. **2001**, *10 (1)*, 299–304.
2. Crasmareanu, E.; Salló, A.; Elenes, F.; Seclaman, E.; Fara, D. Estimate Studies of Lipophilicity on a Series of Monoazoic Dyes Derived from the p-aminobenzoic Acid, *Annals of West University of Timisoara*, Ser. Chem. **2001**, *10 (1)*, 305–310.
3. Elenes, F.; Fara, D.; Seclaman, E.; Kurunczi, L.; Simon, Z. Accuracy in DNA Replication and Pairing Energies of Enolic Base Forms. *Rev. Roum. Chim.* **2002**, Volume Date 2001, *46(4)*, 303–307.
4. Katritzky, A. R.; Fara, D.; Tatham, D.; Petrukhin, R.; 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(12)*, 1333–1356.
5. Andersson, P. L.; Maran, U.; Fara, D.; Karelson, M.; Hermens, J. L. M. General and Class Specific Models for Prediction of Soil Sorption Using Various Physicochemical Descriptors. *J. Chem. Inf. Comput. Sci.* **2002**, *42(6)*, 1450–1459.
6. Katritzky, A. R.; Fara, D. C. How Chemical Structure Determines Physical, Chemical and Biological Properties. *Annals of West University of Timisoara*, Ser. Chem. **2003**, *12 (3)*, 1205–1214.
7. Katritzky, A. R.; Tämm, K.; Kuanar, M.; Fara, D. C.; Oliferenko, A.; Oliferenko, P.; Huddleston, J. G.; Rogers, R. D. Aqueous Biphasic Systems. Partitioning of Organic Molecules: A QSPR Treatment. *J. Chem. Inf. Comp. Sci.* **2004**, *44(1)*, 136–142.
8. Katritzky, A. R.; Fara, D. C.; Yang, H.; Tämm, K.; Tamm, T.; Karelson, M. Quantitative Measures of Solvent Polarity. *Chem. Rev.* (Washington, DC, United States) **2004**, *104(1)*, 175–198.
9. Katritzky, A. R.; Fara, D. C.; Yang, H.; Karelson, M.; Suzuki, T.; Solov'ev, V. P.; Varnek, A. Quantitative Structure-Property Relationship Modeling of β-Cyclodextrin Complexation Free Energies. *J. Chem. Inf. Comp. Sci., in press*.
10. Katritzky, A. R.; Fara, D. C.; Karelson, M. QSPR of 3-Aryloxazolidin-2-one Antibacterials. *Bioorg. Med. Chem., in press*.
11. Tämm, K.; Fara, D. C.; Katritzky, A. R.; Burk, P.; Karelson**,** M. A QSPR Study of Lithium Cation Basicities. *J. Phys. Chem. B, in press*.
12. Varnek, A.; Fourches, D.; Solov'ev, V. P.; Baulin, V. E.; Turanov, A. N.; Karandashev, V. K.; Fara, D.; Katritzky, A. R. "In silico" design of new uranyl extractants based on phosphoryl-containing podands: QSPR studies, generation and screening of virtual combinatorial library and experimental tests *J. Chem. Inf. Comp. Sci., in press*.

ELULOOKIRJELDUS

DAN CORNEL FARA

Sündinud:	28. september 1967, Caransebes, Rumeenia
Kodakondsus:	Rumeenia
Perekonnaseis:	abielus, üks poeg (11 a.)
Address:	Tartu Ülikool, keemilise füüsika instituut, Jakobi 2, Tartu, 51014, Eesti
Tel:	+372 7 375270
E-mail:	dfara@theor.chem.ut.ee

Haridus

1991–1996	Timisoara Lääne Ülikooli keemia bioloogia geograafia teaduskonna üliõpilane, Rumeenia. *B.Sc.* (keemia) 1996.
1996–1997	Timisoara Lääne Ülikooli keemia bioloogia geograafia teaduskonna magistrant, Rumeenia. *M.Sc.* (keemia) 1997.
2001–2004	Tartu Ülikooli keemiaosakonna doktorant, juhendaja prof. Mati Karelson.

Erialane kogemus

1997–1999	keemiaõpetaja, Tööstustehnikum, Otelu-Rosu & Caransebes, Rumeenia
1999–2001	teadur, Keemia Instituut, Rumeenia Akadeemia, Timisoara Haru, Rumeenia
2000–present	abiprofessor, keemia osakond, farmaatsia teaduskond, Timisoara Meditsiini ja Farmaatsia Ülikool, Rumeenia
2001–present	külalisteadur, Heterotsükliliste Ainete Keskus, keemia osakond, Florida Ülikool, USA

Publikatsioonid

1. Elenes, F.; Fara, D.; Salló, A.; Seclaman, E.; Kurunczi, L. Multivariate Analysis (PCA, PLS) on a Series of AS-Naphthols, *Annals of West University of Timisoara*, Ser. Chem. **2001**, *10 (1)*, 299–304.
2. Crasmareanu, E.; Salló, A.; Elenes, F.; Seclaman, E.; Fara, D. Estimate Studies of Lipophilicity on a Series of Monoazoic Dyes Derived from the p-

aminobenzoic Acid, *Annals of West University of Timisoara*, Ser. Chem. **2001**, *10 (1)*, 305–310.
3. Elenes, F.; Fara, D.; Seclaman, E.; Kurunczi, L.; Simon, Z. Accuracy in DNA Replication and Pairing Energies of Enolic Base Forms. *Rev. Roum. Chim.* **2002**, Volume Date 2001, *46(4)*, 303–307.
4. Katritzky, A. R.; Fara, D.; Tatham, D.; Petrukhin, R.; 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(12)*, 1333–1356.
5. Andersson, P. L.; Maran, U.; Fara, D.; Karelson, M.; Hermens, J. L. M. General and Class Specific Models for Prediction of Soil Sorption Using Various Physicochemical Descriptors. *J. Chem. Inf. Comput. Sci.* **2002**, *42(6)*, 1450–1459.
6. Katritzky, A. R.; Fara, D. C. How Chemical Structure Determines Physical, Chemical and Biological Properties. *Annals of West University of Timisoara*, Ser. Chem. **2003**, *12 (3)*, 1205–1214.
7. Katritzky, A. R.; Tämm, K.; Kuanar, M.; Fara, D. C.; Oliferenko, A.; Oliferenko, P.; Huddleston, J. G.; Rogers, R. D. Aqueous Biphasic Systems. Partitioning of Organic Molecules: A QSPR Treatment. *J. Chem. Inf. Comp. Sci.* **2004**, *44(1)*, 136–142.
8. Katritzky, A. R.; Fara, D. C.; Yang, H.; Tämm, K.; Tamm, T.; Karelson, M. Quantitative Measures of Solvent Polarity. *Chem. Rev.* (Washington, DC, United States) **2004**, *104(1)*, 175–198.
9. Katritzky, A. R.; Fara, D. C.; Yang, H.; Karelson, M.; Suzuki, T.; Solov'ev, V. P.; Varnek, A. Quantitative Structure-Property Relationship Modeling of β-Cyclodextrin Complexation Free Energies. *J. Chem. Inf. Comp. Sci., in press*.
10. Katritzky, A. R.; Fara, D. C.; Karelson, M. QSPR of 3-Aryloxazolidin-2-one Antibacterials. *Bioorg. Med. Chem., in press*.
11. Tämm, K.; Fara, D. C.; Katritzky, A. R.; Burk, P.; Karelson**,** M. A QSPR Study of Lithium Cation Basicities. *J. Phys. Chem. B, in press*.
12. Varnek, A.; Fourches, D.; Solov'ev, V. P.; Baulin, V. E.; Turanov, A. N.; Karandashev, V. K.; Fara, D.; Katritzky, A. R. "In silico" design of new uranyl extractants based on phosphoryl-containing podands: QSPR studies, generation and screening of virtual combinatorial library and experimental tests *J. Chem. Inf. Comp. Sci., in press*.

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.