

SOFJA TŠEPELEVITŠ

Experimental studies and modeling of
solute-solvent interactions



DISSERTATIONES CHIMICAE UNIVERSITATIS TARTUENSIS

178

SOFJA TŠEPELEVITŠ

Experimental studies and modeling of
solute-solvent interactions



UNIVERSITY OF TARTU
Press

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

Dissertation is accepted for commencement of the degree of Doctor of Philosophy in Chemistry on June 5, 2018 by Council of Institute of Chemistry, University of Tartu

Supervisor: Prof. Ivo Leito, PhD
Institute of Chemistry, University of Tartu, Estonia

Opponent: Prof. Dr.-Ing. Irina Smirnova
Hamburg University of Technology, Germany

Commencement: August 31, 2018 at 10:15
Ravila Street 14a, Tartu (Chemicum)

Publication of this dissertation is granted by University of Tartu, Estonia.

This work has been partially supported by Graduate School of Functional materials and technologies receiving funding from the European Regional Development Fund in University of Tartu, Estonia



European Union
European Regional
Development Fund



Investing
in your future

ISSN 1406-0299

ISBN 978-9949-77-784-6 (print)

ISBN 978-9949-77-785-3 (pdf)

Copyright: Sofja Tšepelevitš, 2018

University of Tartu Press
www.tyk.ee

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS	6
ABBREVIATIONS	7
1. INTRODUCTION	8
2. LITERATURE OVERVIEW	10
2.1. Intermolecular interactions in liquids	10
2.2. Solvation	11
2.3. Computations in fluid phases	11
2.3.1. Principal approaches	12
2.3.2. COSMO-RS and SMD	12
2.3.3. Modeling of hydrogen bonds in solution	14
2.4. Liquid-liquid extraction studies	16
2.4.1. Applications of partition/distribution data	16
2.4.2. Equilibria in two-phase systems	16
2.4.3. Experimental methods	18
2.4.4. Theoretical methods	19
3. EXPERIMENTS AND COMPUTATIONS	20
3.1. Computations	20
3.2. Experimental determination of $\log D$ values	21
3.3. Uncertainty estimation for experimental $\log D$ values	22
4. RESULTS AND DISCUSSION	24
4.1. Modeling of hydrogen bonding in non-aqueous media	24
4.2. Extraction modeling for compounds with unknown structure	29
4.3. Modeling liquid-liquid extraction with COSMO-RS	33
5. SUMMARY	37
REFERENCES	38
SUMMARY IN ESTONIAN	41
ACKNOWLEDGMENTS	42
PUBLICATIONS	43
CURRICULUM VITAE	97
ELULOOKIRJELDUS	99

LIST OF ORIGINAL PUBLICATIONS

- I. Tshepelevitsh, S.; Oss, M.; Pung, A.; Leito, I. Evaluating the COSMO-RS Method for Modeling Hydrogen Bonding in Solution. *ChemPhys Chem* **2013**, *14*, 1909–1919.
- II. Tshepelevitsh, S.; Trummal, A.; Haav, K.; Martin, K.; Leito, I. Hydrogen-Bond Donicity in DMSO and Gas Phase and Its Dependence on Brønsted Acidity. *J. Phys. Chem. A* **2017**, *121*, 357–369.
- III. Tshepelevitsh, S.; Hernits, K.; Jenčo, J.; Hawkins, J. M.; Muteki, K.; Solich, P.; Leito, I. Systematic Optimization of Liquid-Liquid Extraction for Isolation of Unidentified Components. *ACS Omega* **2017**, *2*, 7772–7776.
- IV. Tshepelevitsh, S.; Hernits, K.; Leito, I. Prediction of partition and distribution coefficients in various solvent pairs with COSMO-RS. *J. Comput. Aided Mol. Des.* **2018**, DOI: 10.1007/s10822-018-0125-y.

Author's contribution

- Paper I:** Carried out part of the calculations, contributed to writing the manuscript.
- Paper II:** Carried out about half of the calculations and data treatment, main person responsible for writing the manuscript.
- Paper III:** Planned and supervised the work, carried out part of the experiments, main person responsible for data treatment and writing the manuscript.
- Paper IV:** Planned and supervised the work, carried out part of the experiments and computations, main person responsible for writing the manuscript.

ABBREVIATIONS

ΔG_{HB}	Free energy of hydrogen bond formation
ΔG_{solv}	Free energy of solvation
ΔH_{HB}	Enthalpy of hydrogen bond formation
COSMO-RS	Conductor-like screening model for real solvents
CP	Contact probability approach
DCM	Dichloromethane
DFT	Density functional theory
DMSO	Dimethyl sulfoxide
HB	Hydrogen bond
HBA	Hydrogen bond acceptor
HBD	Hydrogen bond donor
HPLC	High performance liquid chromatography
K_{HB}	Equilibrium constant of hydrogen bond formation reaction
LFER	Linear free energy relationship
$\log D$	Logarithm of distribution ratio
$\log P$	Logarithm of partition coefficient
LOO	Leave-one-out
MD	Molecular dynamics
MeCN	Acetonitrile
MTBE	Methyl- <i>tert</i> -butyl ether
OCE	Outlying charge error
$\text{p}K_{\text{a}}$	Negative logarithm of acid dissociation constant
QM	Quantum mechanics
QSPR	Quantitative structure-property relationship
RMSE	Root mean square error
SM	Supramolecular approach

1. INTRODUCTION

The majority of important chemical processes take place in liquid phase. Numerous biological transformations, greatest part of laboratory chemistry and industrial manufacturing of chemicals occur in solutions. Understanding, quantification and optimization of all those processes require comprehension of the underlying intermolecular interactions. However, liquid phase is the most complicated phase from the computational point of view, since each molecule in solution is influenced by closely located, but unordered and incessantly moving neighboring particles.

The range of compound properties that the scientific and industrial communities would like to accurately predict in liquid phase is very wide. This work addresses prediction of two solute parameters – strength of hydrogen bond in non-aqueous solutions and distribution of organic solutes between immiscible liquids.

Hydrogen bond (HB) plays a crucial role in functioning of biomolecules and greatly influences physical, chemical, catalytic etc properties of compounds. In this work HB is considered from the point of view of synthetic supramolecular chemistry and acid/base studies. HB is the main interaction behind anion binding by most synthetic receptors, functioning of molecular switches and some selective catalysts. On the other hand, processes involving unwanted HB formation (dimerization, homo- and heteroconjugation) are well-known complicating factors in acidity/basicity studies in non-aqueous media. Predicting the strength of HBs at least on relative scale is of much interest to researchers in those fields. However, quantifying HB interactions is not straightforward from computational point of view.

The second part of this thesis reports experimental studies and prediction of distribution of organic solutes between aqueous and organic solutions. Predicting the behavior of the molecule in two-phase liquid systems is of great interest for very different branches of research and technology. In medical, biological and environmental studies one of the most important parameters of a compound is its lipophilicity, which influences biochemical action of the compound and its environmental fate. Lipophilicity is often expressed through partition or distribution coefficient of the solute between aqueous solution and octanol or some other organic solvent. In analytical chemistry and various technological processes knowing the distribution behavior of the compound is needed for setting up effective liquid-liquid extraction procedures. Finally, partitioning of a compound between polymeric material and a liquid (which can also be approximated by partitioning coefficient between two liquids) determines usability of the solute as e.g. a sensor component.

The aim of the present work is to assess the existing methods and/or develop new approaches for prediction of the two abovementioned properties. Preference is to be given to experimental and computational approaches that are easily available and generally affordable, to make the results of this study

conveniently useful for broad audience. The main computational method to be used in this work is COSMO-RS (*Conductor-like Screening Model for Real Solvents*), chosen for its unequalled ability to predict the properties of arbitrary multicomponent liquid mixtures and totally novel solutes and solvents without additional experimental input.

In addition to predicting the properties of the solutes with known structure a possibility of predicting the liquid-liquid extraction outcome for unidentified solutes is addressed. A predictive model using a small number of experimentally determined distribution ratios in selected solvent pairs to predict solute behavior in a larger number of solvent pairs is developed and evaluated. In frequent situations where the identities of all mixture components are not known (pilot synthesis, natural products etc) such model would significantly simplify finding the most effective or selective solvent pair for isolation or purification of components of interest. It will reduce the time, workload and expenses required for extraction process optimization, along with the amount of harmful waste.

2. LITERATURE OVERVIEW

2.1. Intermolecular interactions in liquids

Molecules in liquids reside in close proximity to each other and continually interact in a variety of ways depending on their properties.^[1] Particles possessing permanent di- or multipoles are subject to mutual attraction or repulsion depending on their relative orientation at the given moment. A dipole of one molecule can distort the electron clouds of neighboring particles, inducing so-called instantaneous dipoles in them. Ions, in turn, generate electric field that causes reorientation of dipolar molecules and electron density redistributions in any particles within its effect range.

Intermolecular interactions can be classified into non-specific, or van der Waals forces, and specific forces (hydrogen bonding and electron-pair donor-acceptor interactions).^[1] The term **van der Waals forces** sums up interactions between permanent dipoles (directional forces), between permanent and induced dipoles (induction or Debye forces), and attractive forces resulting from the synchronization of instantaneous dipoles in neighboring molecules (dispersion or London forces). All those forces are short-range – interaction energies are proportional to $1/r^6$, where r is intermolecular distance. Van der Waals forces are rather weak in comparison with covalent interactions, for instance, dispersion forces in alkanes were estimated to be about 4–5 kJ mol⁻¹ per -CH₂-unit^[2] and the net contribution of van der Waals interactions in non-polar liquids to be about 0.3 kJ mol⁻¹ per Å² of molecular contact area.^[3]

Hydrogen bond (HB) is a complex directional interaction between a polarised hydrogen atom and a Lewis base (atom/group with a free electron pair or π -system). HB combines contributions from electrostatic, induction and dispersion interactions, charge-transfer induced covalency, as well as repulsion contribution arising from overlap of electron clouds.^[4] Strong hydrogen bonds have a covalent-like nature, while weak ones are largely electrostatic. The dissociation energy of a medium-strength HB is in the range of 17...63 kcal mol⁻¹^[4] – about an order of magnitude lower than the energy of a covalent bond, yet higher than that of van der Waals forces. Directionality of HB implies that its strength is dependent on the angle between the donor and acceptor atoms/groups, therefore HB is sensitive to steric factors. The strongest HBs result from linear arrangement of the three atoms making up the bond. Hydrogen bonds are involved in proton transfer and can be viewed as “frozen” proton transfer reactions. HB has a crucial role in all kinds of biological processes^[5] and greatly influences physico-chemical (incl. catalytic) properties of substances. It is also one of the cornerstones of supramolecular chemistry and host-guest interactions.

2.2. Solvation

Solvation is defined as a stabilizing interaction of any kind between a solute (or a solid surface) and the surrounding solvent.^[6] The types of forces that are involved depend on the nature of solute and solvent. Effects of solvation on a solute include conformational changes, geometric changes (e.g. lengthening of polar bonds) and redistribution of electron density and surface charges. The energetic effect of solvation is often divided into contributions from cavitation (producing a cavity in the solvent bulk to accommodate a solute), orientation (partial reorientation of the dipolar solvent molecules caused by addition of the solute), isotropic interactions (non-specific interactions like polarisation and dispersion) and anisotropic interactions (hydrogen bonding and other Lewis acid-base processes).^[1] The energetics of solvation are frequently expressed through standard molar Gibbs energy of solvation – the change of Gibbs free energy of the solute upon its transfer from the gas phase to the liquid.

The term **selective** or **preferential solvation** describes the situation where the composition of multicomponent solvent in the close proximity of the solute is different from that of the bulk solvent. The inner solvation shell of the solute will be richer in solvent component(s) that interact with it more favorably, leading to more negative Gibbs free energies of solvation. In some cases even extremely low concentration of some solvent component can lead to non-ideal behavior of the solute due to preferential solvation (e.g. influence of water impurity on the reactions involving ions in low-polarity solvents).

2.3. Computations in fluid phases

Modeling molecular processes in fluids is desirable for numerous practical applications but is a challenging task for computational chemistry.^[7] Contrary to isolated molecules *in vacuo*, which are not influenced by the environment and have properties determined solely by intrinsic structural and electronic features, in fluids the electronic state of every molecule is crucially affected by solvation, i.e. its interactions with rapidly and incessantly changing surroundings. Nowadays the results of *in vacuo* calculations (*ab initio* as well as semi-empirical) are often of comparable quality to the corresponding experimental data,^[8] but the same cannot be said about calculations in liquids. As modeling of liquid environment *ab initio* would be extremely computationally expensive and non-expedient for most practical purposes, the field of fluid modeling is currently dominated by various semi-empirical methods.

The present work focuses on modeling solute properties or bimolecular interactions in liquid environment, as opposed to prediction of bulk properties of liquids (e.g. viscosity, dielectric permittivity etc). The following text addresses the methods used for the description of solute properties and solvent effects.

2.3.1. Principal approaches

There are two principal approaches for describing the effect of the surrounding media on a molecule: explicit and implicit. In the former case the environment is represented by a number of discrete molecules; in the latter case – by a field or continuum with a certain dielectric constant.

Molecular dynamics (MD) simulations^[9,10] take the molecular environment into account explicitly and interactions between particles are modeled in real time. The MD computations yield molecular trajectories, which are used for evaluation of the desired properties. However, all interactions are described through force-fields, which produce rather approximate estimates of interaction energies. Although it is possible to couple MD with high-level calculation of intermolecular interactions,^[10,11] such simulations tend to be extremely computationally expensive.

In **continuum solvation models**^[12] the environment is approximated as a field or structureless medium with a certain relative permittivity. Various interactions of solute with solvent molecules are taken into account by the corresponding parameters in the expression of the total energy. Continuum models are prone to give erroneous results in cases when strong hydrogen bonds form between solute and solvent molecules as such models are intrinsically unable to account for directional interactions with considerable degree of covalency. Also, being parametrized for a certain solvent, classical continuum models are not suitable for multicomponent liquids (including non-dilute solutions where interactions between solute molecules become important) and situations where preferential solvation plays an important role.

Cluster-continuum approach combines implicit and explicit solvation: one or several molecules of the inner solvation shell are modeled explicitly, while more remote solvent particles are represented by continuum.^[13] This approach at least partially solves the problem of accounting for strong directional solute-solvent interactions at moderate computational cost. The main problem is that with larger systems the number of spatial arrangements of the solute molecule and the explicit solvent molecules can become significant, making the computations unpractically resource-demanding.

2.3.2. COSMO-RS and SMD

The two methods employed in this work for modeling solvation – COSMO-RS and SMD – are the examples of computationally affordable and readily available methods that nevertheless produce acceptable to excellent results in their respective application areas. Potentially useful but overly demanding or scarcely available approaches were deliberately avoided.

COSMO-RS^[7,14–17] is a hybrid method, combination of quantum chemical continuum model and statistical thermodynamics developed by Klamt and co-workers. The method is semi-empirical, yet parametrized at atomic, not

molecular, level. COSMO-RS was shown to produce solvation energies of comparable quality to SM8 model^[18] but, unlike the latter, is able to give useful predictions for arbitrary multicomponent liquid mixtures (incl. concentrated solutions) and totally novel solutes and solvents. What is important, this is achieved at very modest computational costs. It has been demonstrated^[19-21] that the obtained absolute values of various properties may be systematically shifted, yet correlation with the respective experimental values is usually good.

COSMO-RS computation consists of two steps.^[7] The first step is a unimolecular QM calculation of every species of interest (including different conformers/tautomers) in the reference state of ideal conductor ($\epsilon = \infty$). It yields molecular geometry, shape of molecular cavity, mapping of charge density on the molecular surface, and the total electronic energy of the molecule in the conductor. In the second step, the charge densities obtained in the QM calculation are used to create the so-called sigma-profile of the molecule – a two-dimensional descriptor containing information on the amounts of molecular surface with certain charge to area ratios (sigma values). The sigma-profiles of the individual mixture components, weighted by the corresponding molar fractions, are combined to produce the sigma-profile of the solution. Thereafter the interaction energy of each compound with its surroundings is evaluated through the energies of pairwise interactions of surface segments belonging to the compound and the entire solution. The thermodynamic parameters (G , H) of each molecule are computed by combining its electronic energy in virtual conductor, interaction energy with the solution (computed as described above), and combinatorial term accounting for shape and size differences of the molecules in the system. This approach is very effective from the position of computational cost, although it neglects some of the steric effects and does not take long-range interactions into account.

The method uses semi-empirical expressions for evaluation of different kinds of intermolecular interactions. Implicit description of hydrogen bonding in COSMO-RS is based on the assumption of bilinear dependence of HB energy on the polarization charge densities on the interacting surface segments. The hydrogen bond expression was initially applied to the pairs of oppositely charged segments with charge densities exceeding certain thresholds. The assumption of bilinear dependence was later verified^[22,23] and elaborated: the latest parametrizations take into account not only segment charge densities but also the number of underlying free electron pairs.

COSMO-RS method makes no distinction between solute and solvent: every compound in the mixture is treated in the same way. It allows taking into account solute-solute interactions and impurities (especially water) that are always present even in high-purity solvents and in some cases have, even in low concentrations, large effect on the solute properties.

SMD^[24] is a continuum solvation model of Marenich, Cramer, and Truhlar that is widely and successfully employed for calculation of solvation energies in various solvents (water, acetonitrile, DMSO, nitromethane etc). Unlike COSMO-RS, SMD is parametrized at solvent level, using such solvent

parameters as refractive index, Abraham's hydrogen bond acidity and basicity, surface tension, and fraction of aromatic carbon and halogen atoms in the molecule. The atomic radii used for construction of the molecular cavity (solute-solvent boundary) have been parametrized using an extensive set of experimental solvation energies.

Hydrogen bonding between solute and solvent is taken into account implicitly through the CDS (cavitation, dispersion, and solvent structural effects) term in the solvation energy equation, so the energetic contribution of HB is not separated from other electrostatic effects.

SMD model can be applied to neutral and charged solutes, but lower accuracy is to be expected in case of ionic species. The mean unsigned errors of calculated solvation free energies reported by method developers are 0.6–1.0 kcal mol⁻¹ for neutrals and ca 4 kcal mol⁻¹ for ions.

Both COSMO-RS and SMD are principally based on unimolecular quantum-chemical calculations, but can be adapted to cluster-continuum or supramolecular approaches. In case of COSMO-RS, cluster-continuum methodology has been found to be less robust than the conventional procedure^[25] (the experience of the author supports this claim^[26]). However, it could be prudent to add explicit solvent molecules to solutes with extreme surface charge hotspots: first, it better approximates the real moiety existing in the solution, as highly charged groups are likely to be strongly solvated; second, the best performance of the method can be expected only with non-extreme surface charges.

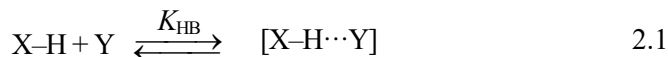
For continuum solvation models like SMD, adding explicit solvent molecules in case of concentrated charges is beneficial due to inherent incapacity of such models to account for strong specific solvation. The positive effect of explicit water molecule on accuracy of aqueous solvation energies of small ions was demonstrated on the example of SM6 model.^[27] In case of SMD, some ions bearing concentrated charge hotspots were represented as ion-water clusters when the model was parametrized for water.^[24]

2.3.3. Modeling of hydrogen bonds in solution

In the present work the HB interaction was studied in the context of molecular design, with the aim to assess how successfully the methods used for designing of novel synthetic receptors, acids/bases etc can describe HB formation. Therefore the work was focused on explicit description of single HBs as opposed to estimating the net interaction of HB-active solute with numerous surrounding molecules. The latter, however, is an important factor determining solute stabilities in different media, and was taken into account implicitly.

Two approaches were used for modeling HB formation reactions – supramolecular (with and without Born-Haber cycle, works I and II) and contact probability approach (work I).

Supramolecular (SM) approach involves quantum chemical modeling of interacting species and the resulting associate:



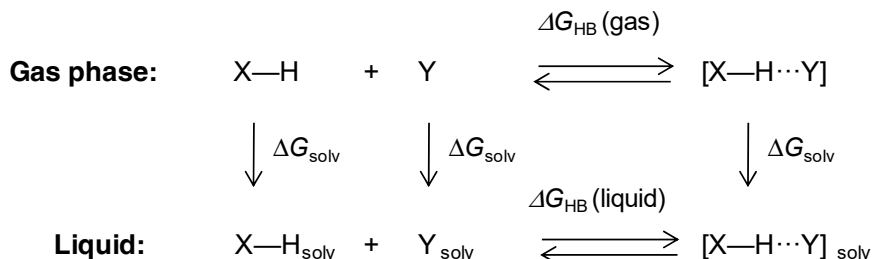
where X-H is a HB donor, Y is a HB acceptor and K_{HB} is the equilibrium constant of HB formation reaction.

The Gibbs free energy change of the HB formation reaction (ΔG_{HB}) is calculated as follows:

$$\Delta G_{\text{HB}} = -RT \ln(K_{\text{HB}}) = G([\text{X-H}\cdots\text{Y}]) - G(\text{X-H}) - G(\text{Y}) \quad 2.2$$

The enthalpy of the reaction (ΔH_{HB}) is calculated in the analogous manner.

The free energies or enthalpies of the involved species can be obtained in different ways, either by computing the energies of all species in the single chosen medium (liquid) or by combining gas-phase energies with solvation energies (ΔG_{solv}) as shown on Scheme 1.



Scheme 1. Thermodynamic (Born-Haber) cycle for calculating the HB energies in solution. Subscript ‘solv’ denotes solvated species.

Dilute donor-acceptor contact probability (CP) method does not require an explicit QM calculation of the hydrogen-bonded complex, thus reducing the computational cost compared to supramolecular approach. The complex formation constant (K_{HB}) is calculated from contact probabilities of HB donor and acceptor in dilute solution according to equation 2.3:

$$K_{\text{HB}} = \frac{\lambda_{\text{d-a}} A_{\text{d}}}{A_{\text{eff}}^{x_{\text{a}}}} = \frac{\lambda_{\text{a-d}} A_{\text{a}}}{A_{\text{eff}}^{x_{\text{d}}}} \quad 2.3$$

where subscripts ‘a’ and ‘d’ denote HB acceptor and HB donor, respectively, λ is the contact probability (obtained in the statistical thermodynamics step of COSMO-RS method), A is the molecular surface area, x is a mole fraction, and A_{eff} is the effective contact area (estimated as 5 \AA^2 for most HB contacts) of the respective species. Reaction enthalpies can be calculated using the van’t Hoff’s plots (ΔH_{HB} is found from the slope of $\ln K_{\text{HB}}$ vs $1/T$ correlation line).

Although computationally inexpensive, the CP method has a number of important limitations. It can only be applied to single, sterically unhindered hydrogen bonds in pure HB-inactive solvents. The solution cannot contain HB donors or acceptors other than the reacting species of interest, which rules out the possibility of taking the effect of water impurity into account. Also, the CP approach cannot account for the energetic and steric effects of proton transfer within the hydrogen-bonded complex. However, due to its attractive simplicity CP method was applied to some HB formation reactions studied in this work to explore its usability.

2.4. Liquid-liquid extraction studies

2.4.1. Applications of partition/distribution data

Partition coefficients are used as a measure of lipophilicity, which is an extremely important property in medicinal chemistry (including drug discovery) and biochemistry. Calculated octanol/water partition coefficient is one of the parameters in Lipinski's „rule of five“^[28] that empirically estimates the usability of a compound as a drug.

Partition coefficients in octanol/water system ($\log P_{o/w}$) are known to correlate with physicochemical properties^[29–32] and ecotoxicological effects^[33–35] of the compounds. Partition coefficients between water and other organic solvents are also used in drug discovery.^[36] Along with the biology- and medicine^[37] related uses, the partition properties of a compound determine its usability as e.g. working compound in membrane-based sensors.^[38] Correct estimates of partition properties are key to efficiency of technological processes involving liquid-liquid extraction,^[39] reverse-phase chromatographic separation and adsorption/absorption phenomena.

2.4.2. Equilibria in two-phase systems

Most practically relevant extraction systems consist of a polar phase (usually aqueous solution) and a less polar phase immiscible with the first one (organic solvent or ionic liquid). The number of equilibria playing role in distribution of the solute between the two phases increases with the complexity of the solute. Partition coefficient P describes the relative abundance of a certain form of the solute S in the two phases:

$$P = \frac{[S]_{\text{org}}}{[S]_{\text{aq}}} \quad 2.4$$

Here and below subscripts ‘org’ and ‘aq’ denote organic and aqueous phase, respectively.

Usually P describes partition of a non-ionized, non-associated form of the solute, but may sometimes refer to ion, hydrate etc.

According to IUPAC definition,^[40] distribution ratio D describes the ratio of net concentrations of all chemical forms of the solute (e.g. neutral, different ionization states, dimer, etc) in two liquid phases. However, when interpreting or modeling extraction process, D is often defined differently depending on the purpose, and less important equilibria may be omitted for the sake of simplicity and clarity. For example, when only single-step solute ionization is considered, the distribution ratio is expressed as follows:

$$D = \frac{[S_{\text{neu}}]_{\text{org}} + [S_{\text{ion}}]_{\text{org}}}{[S_{\text{neu}}]_{\text{aq}} + [S_{\text{ion}}]_{\text{aq}}} \quad 2.5$$

where subscripts ‘neu’ and ‘ion’ denote neutral and ionized forms of the solute. In reality, more equilibria are likely involved in distribution of the solute, e.g. dimerization or ion pairing in the organic phase. As another example, for the solute that ionizes in the aqueous phase and forms dimers in the organic phase, distribution ratio takes the following form:

$$D = \frac{[S_{\text{neu}}]_{\text{org}} + 2 \cdot [S_{\text{dimer}}]_{\text{org}}}{[S_{\text{neu}}]_{\text{aq}} + [S_{\text{ion}}]_{\text{aq}}} \quad 2.6$$

Distribution ratio is relatively easily obtained experimentally, yet is often difficult to interpret theoretically. P and D values are usually presented in logarithmic form ($\log P$ and $\log D$).

In cases when the solute is a monoprotic acid or base (i.e. if the solute can exist as neutral and one ionized form) the relationship between $\log P$ and $\log D$ in organic solvent-water systems is for practical purposes often described by equation 2.7:^[41]

$$\log D = \log P - \log(1 + 10^Z) \quad 2.7$$

where Z is $(\text{pH} - \text{p}K_{\text{a}})$ for acidic and $(\text{p}K_{\text{a}} - \text{pH})$ for basic solutes.

The equation is based on the assumptions that (1) ions are present only in the aqueous phase, i.e., $[S_{\text{ion}}]_{\text{org}} \sim 0$; and (2) mutual saturation of the phases does not change the aqueous acidity/basicity of the solute (its $\text{p}K_{\text{a}}$ value). Although sufficiently accurate for many solute-solvent combinations, the equation produces erroneous results in case of hydrophilic organic solvents in which the ionized forms dissolve relatively well (e.g., butanol, as demonstrated in work IV) and compounds with strongly hydrophobic ionized forms.^[38]

Partitioning of free solute ions into organic phase can be accounted for by using partition coefficients of neutral (P_{N}) and ionized (P_{I}) forms of the solute as follows:^[42]

$$\log D = \log(P_{\text{N}} + P_{\text{I}} \cdot 10^Z) - \log(1 + 10^Z) \quad 2.8$$

However, equation 2.8 is also rather simplified, ignoring the effects of counterions and ion pair formation.

2.4.3. Experimental methods

The most straightforward and reliable method for determination of partition or distribution coefficients is the **shake-flask method**.^[43] It involves establishing equilibrium between the two liquids with dissolved solute, followed by determination of solute concentration in one or both phases. Analysis of only one phase and calculation of the solute concentration in the other by difference requires carefully measuring the initial amount of the solute and an assumption that any solute losses (e.g. adsorption on vessel surfaces) can be neglected. Therefore, if the total amount of the dissolved solute in the two phases can change during equilibration, quantitative analysis of both phases should be carried out to obtain accurate results. Various analytical techniques can be used for determination of solute concentration^[44] – UV-Vis spectrophotometry, NMR,^[45] voltammetry,^[46] potentiometry,^[47] chromatographic separations^[48] with suitable detection method. The downsides of the shake-flask method are relatively narrow applicability range ($\log D$ values approximately from -4 to 4) and generally high time- and resource demand.

Indirect chromatographic method for $\log P/\log D$ determination is based on the correlations between the partition or distribution coefficients of a series of solutes with their respective chromatographic capacity factors. Such correlations exist due to the fact that liquid-liquid extraction and liquid chromatographic separation are ruled largely by the same factors and can be described by analogical LFER (Linear Free Energy Relationship) models.^[49,50] The retention factor of the studied solute is measured, and its $\log P$ is calculated using the previously established correlation. Many types of separation techniques are used, e.g. reverse-phase chromatography,^[51,52] HILIC,^[53] and capillary electrophoresis.^[54] However, a serious downside of this approach is sensitivity to congeneric effects: the compounds included in the correlation should be chosen with care and should bear considerable structural similarity to the solute of interest. Also, any changes in chromatographic conditions (such as mobile phase composition or state of the column) between measurements should be either avoided (which is not always practical) or corrected for,^[55] which adds to the workload.

High-throughput methodologies for $\log P$ and $\log D$ determination that do not rely on chromatographic retention have been developed,^[56-58] but, unfortunately, often require non-conventional instrumentation.

Only procedures involving establishing solute equilibrium between the actual liquid phases (like shake-flask method) can be safely applied to the solutes of uncertain identity. Indirect chromatographic determinations, while being potentially less resource-demanding, require at least partial knowledge of

the solute structure in order to compose a suitable equation for conversion of retention data to $\log P$ values.

2.4.4. Theoretical methods

Various methods that have been developed for calculation of $\log P$ values (usually in octanol/water system) are mainly aimed at estimation of molecular lipophilicity for drug design and toxicology-related applications.^[59] Many of those methods (e.g. UNIFAC^[60] and ACD/LogP^[61]) are substructure-based: the molecule is divided into the set of fragments and its $\log P$ is calculated from the empirically derived constants corresponding to each fragment, incidence of the fragment in the molecule, and various correction factors meant to account for interactions between fragments within the molecule.

Another widely used method is the LFER approach of Abraham,^[62] whereby the thermodynamic properties of a neutral solute are generally expressed as a function of five molecular descriptors as follows:

$$\log P = c + a \cdot \alpha^H + b \cdot \beta^H + s \cdot \pi + v \cdot V + e \cdot E \quad 2.9$$

where α^H is hydrogen bond acidity, β^H – hydrogen bond basicity, π – dipolarity/polarizability, V – molecular volume, and E – excess molar refraction. If the solute undergoes ionization under the employed conditions, some additional descriptors should be included into the equation.^[62] Model coefficients and molecular descriptors can be obtained experimentally or predicted.^[63–65]

COSMO-RS method is also well suitable for extraction modeling. It was previously demonstrated to give good $\log P$ predictions in various solvent pairs: octanol/water,^[66] alkane/water,^[66,67] systems involving ionic liquids^[68] and micelles.^[69] In COSMO-RS partition coefficients are computed from chemical potentials of the solute in two liquid phases (calculated as explained in Section 2.3.2) and molar volumes of the phases (by default estimated using an embedded QSPR equation). The distinct advantage of COSMO-RS over other methods is its ability to provide useful (even when not quite accurate) predictions for arbitrary multicomponent solvents as well as novel solutes and solvents for which no experimental data is available yet. It also allows taking mutual saturation of the solvents into account without any experimental input.^[20,70] equilibrium compositions of the phases of a multi-phased system can be obtained from iterative calculation of chemical potentials of all involved species. In addition, COSMO-RS allows to account for the effects of electrolytes^[71,72] on partition/distribution at least at qualitative level.

3. EXPERIMENTS AND COMPUTATIONS

This section gives an overview of the experimental and computational methodologies used in this work for the studies of hydrogen bonding and partition or distribution equilibria in liquid media. The detailed descriptions of the methods are provided in the respective publications.

3.1. Computations

The **quantum-mechanical (QM) methods** used in this work include DFT BP86/TZVP (with or without COSMO model), DFT M05-2X/6-31G(d) (with or without SMD model), and G4MP2^[73] for high-quality gas-phase energies. Multiple starting geometries were produced for geometry optimization to ensure that all stable conformers of each species in each medium have been identified. In all cases, vibrational analysis has been carried out to verify that the obtained geometries correspond to the true energy minima. As a rule, if imaginary frequencies were found in the spectra the structure was reoptimized and vibrational analysis repeated.

The commercial implementation of COSMO-RS method (*COSMOtherm*) has an intrinsic algorithm for taking different conformers into account in all kinds of calculations, weighted by their relative stability in the given mixture. This function was utilized in modeling of HB formation in work I as well as in $\log P/\log D$ calculations in work IV. In case of the gas-phase ΔG_{HB} computations and calculations involving SMD model, only the energies of the most stable conformers of each involved species were used.

The **free energy of solvation** (ΔG_{solv}) was defined as the difference between the Gibbs free energies of the molecule in the solvent and in the gas phase. In case of COSMO-RS, the chemical potential of the species in solution was obtained using the statistical thermodynamics procedure, and in case of SMD by QM calculation in a continuum emulating DMSO. The corresponding gas-phase energies were obtained using the same DFT functional and basis set as in case of liquid-phase energies. Conformers were chosen and combined so that ΔG_{solv} included the energy of conformational change upon transfer from vacuum to the liquid phase (see work II for details).

LogP values (work IV) were computed using the ‘*logP/logD*’ module of *COSMOthermX* software. The default procedure for $\log P$ computations was used, in which the volume quotient (ratio of molar volumes of the two liquid phases) is used in calculation. Unless experimental data is provided by the user, the molar volumes of multicomponent liquid phases are estimated by an embodied QSPR equation that uses linear extrapolation and ignores excess volumes.

3.2. Experimental determination of $\log D$ values

Detailed description of the experimental procedure is given in work III. The protocol is based on a shake-flask method with reverse-phase liquid chromatographic analysis of both equilibrated liquid phases. The methodology was chosen and elaborated to allow determination of $\log D$ values of the components of multisolute mixtures, including unidentified solutes. It is suitable for real-life situations where one needs to quantitatively describe the behavior of an unknown impurity or reaction by-product in different solvent pairs. The $\log D$ values of the solutes are calculated from the ratios of the corresponding peak areas in the two chromatograms (Figure 1), corrected by the difference of injection volumes, if needed. The methodology is direct, i.e. does not rely on correlations between chromatographic retention time and $\log P/\log D$ value of the solute. Also, this approach does not involve calculations of absolute concentration values and thus does not require calibration, which would not be possible in case of unidentified compounds. The applicability domain (range of $\log D$ values directly measurable with reasonable accuracy) of this methodology is generally -3 to 3. Determination of $\log D$ values outside those limits may be possible, but is usually more labour-intensive and less accurate. Also, the experimental procedure is well adaptable to different sample and solute types, as almost any other type of chromatographic column and detector can be used instead of the ones employed in this work.

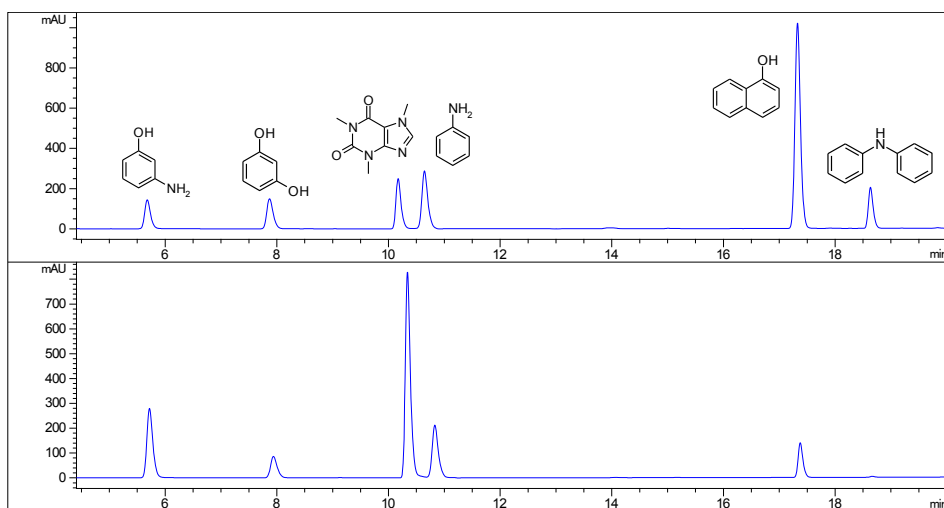


Figure 1. Exemplary chromatograms of equilibrated organic (butanol, upper chromatogram) and aqueous (phosphate buffer pH 7, lower chromatogram) phases after the extraction of a multi-analyte solution.

3.3. Uncertainty estimation for experimental $\log D$ values

A procedure for uncertainty estimation was elaborated for evaluation of the quality of the experimental data in this work.

The experimental $\log D$ values were obtained from the results of at least two replicate measurements, which were conducted on different days with different combinations of solutes in order to vary as many error-inducing factors as possible. The between-day reproducibility of the results was found to depend on the absolute value of the $\log D$, as the extremely low concentration of the solute in one of the phases (in case of very high or very low $\log D$) often causes poor agreement of the replicate measurements. Therefore the target limit for the discrepancy between the replicate measurement results ($\Delta \log D$) was defined as follows:

$$|\Delta \log D| \leq 0.05 \cdot (|\log D| + 1) \quad 3.1$$

The tolerated discrepancy is thus higher for the extreme $\log D$ values. The results of replicate experiments mostly agreed within this limit. In case of higher variability more measurements were carried out, and the values to be used in calculation of the final $\log D$ value were critically chosen based on recorded evidence of experimental problems, as well as chemical intuition and experience. Discrepant results were only discarded if a reasonable explanation for their divergence was found. However, for some values of doubtful accuracy no solid proof of underlying reason was discovered. Such values were not used in calculation of the final $\log D$ values, but were taken into account when calculating the uncertainty of the measurements.

In most cases there were too few replicate measurements to calculate meaningful uncertainty estimates for individual $\log D$ values (values corresponding to a certain solute in a certain solvent pair). Therefore the random components of the uncertainties were found as weighted pooled standard deviations (s_{pooled}) across multiple data points, usually all $\log D$ values of the solute in various solvent pairs. s_{pooled} values were calculated as follows:

$$s_{\text{pooled}} = \sqrt{\frac{\sum_{i=1}^k \left\{ \left(s_i^2 \cdot \frac{n_i - n_{\text{av},i}}{n_i} + s_{\text{av},i}^2 \cdot \frac{n_{\text{av},i}}{n_i} \right) \cdot (n_i - 1) \right\}}{\sum_{i=1}^k (n_i - 1)}} \quad 3.2$$

where k – number of data points (final $\log D$ values) taken into account, n_i – number of replicate measurements corresponding to the i -th data point, s_i – standard deviation of all replicate measurements corresponding to the data point i , $n_{\text{av},i}$ – number of replicate measurements used in calculation of the final $\log D$ value (“good” measurements), $s_{\text{av},i}$ – standard deviation of “good” measurements.

Taking into account only the “good” results of replicate measurements would be incorrect, as the existence of deviating values indicates the possibility

of unidentified experimental problems and the uncertainty estimate should reflect this fact. On the other hand, calculating the uncertainty as a simple pooled standard deviation across all replicate values resulted in counter-intuitive uncertainty estimates. Using weighted contributions according to equation 3.2 provides more sensible uncertainty estimates while being conceptually more correct than omitting deviating values.

The comparison of $\log P$ values from this work with the corresponding literature data showed good agreement and no systematic deviations. Therefore the systematic component of the uncertainty was assumed to be negligible and the uncertainty with estimated 90% coverage probability was found as a double s_{pooled} value.

4. RESULTS AND DISCUSSION

4.1. Modeling of hydrogen bonding in non-aqueous media

Methods for prediction of thermodynamic parameters of hydrogen bonds in non-aqueous media were assessed in works I and II.

Three model reactions were studied, chosen mainly by the availability of the experimental data of adequate quality:

- 1) HB formation between neutral Brønsted acids and bases (work I):
 - substituted acetic acids with a variety of substituted pyridines in chloroform;
 - triphenylphosphine oxide and triphenylphosphine sulfide with a range of HB donors (1-naphthol, 2-naphthol, phenol, carbazole, indole, pyrrole, trifluoro-, trichloro- and tribromoethanol) in carbon tetrachloride.
- 2) Homoconjugation (formation of hydrogen bond between Brønsted acid/base and its conjugated base/acid), on the example of benzoic acids and phenols in various organic solvents (work I).
- 3) HB formation between a Brønsted acid and chloride ion, on the example of different compound classes in DMSO (work II).

Reactions 1 and 2 were modeled using COSMO-RS method with supramolecular approach, and, where possible, calculations by the contact probability method. Model reaction 3 was modeled using thermodynamic cycle combining gas-phase energies with solvation energies by COSMO-RS or SMD methods.

Model reactions 2 and 3 are challenging for several reasons. In both reactions one of the reactants and the resulting complex are anionic, which presents some additional error sources compared to reactions involving only neutral species. One of those is the so-called outlying charge error (OCE).^[74,75] The essence of it is that part of the electron density gets allocated outside of the molecular cavity during the computation, leading to misrepresentation of the net molecular charge and local electron densities. OCE is more pronounced for anions than cations and neutrals.^[74,75] The COSMO model is known to be less sensitive to OCE compared to many other continuum solvation models,^[74] in addition, OCE corrections are applied in the COSMO implementation used in this work.^[76] However, one could still suspect OCE to be an issue in case of anions with highly localized charges. Another consequence of charge localization is strong and possibly specific and/or preferential solvation, which may not be adequately described by a procedure based on unimolecular QM calculations. Finally, the HB energies in solvents, especially the polar ones, are inconveniently narrow-spanned compared to the energies of the same HBs in the gas phase or low-polarity solvents. Such contraction of span is caused by the fact that HB formation in liquid is essentially a displacement reaction: before forming the HB both donor and acceptor must abandon stabilizing contacts with solvent molecules. The stronger capacity of donor and acceptor for HB formation also means their stronger interactions with solvent molecules, leading

to leveling effect of solvent on HB strength. Experimental data available for the model systems 2 and 3 are relatively narrow-spanned, so the differences between individual experimental values are often comparable to the expected computational errors. Although in low-polarity solvents HB energies would be better differentiated, the existence of free (solvated but not associated) ions in them is less likely. Also, impurities such as water would have much greater effect on ion stabilities and, consequently, HB equilibria.

Modeling neutral HB complexes presents fewer difficulties, both from computational standpoint and the quality of experimental data. From the data included in this work, only experimental formation enthalpies of Ph₃PS complexes are critically narrow-spanned.

Figure 2 summarizes the results of HB modeling with COSMO-RS method (without additional empirical corrections or separately calculated gas-phase energies). It is immediately evident that the results are biased, whereas the character (additive and/or multiplicative) and magnitude of the error depends first of all on the compound type. However, for HBs involving only neutral species (Ph₃PO and Ph₃PS with various HBDs, pyridines with substituted acetic acids) the computational results are mostly linearly correlated with the corresponding experimental values and the trends in HB strengths are reproduced relatively well. In case of good correlation a simple empirical correction can be used to obtain sufficiently good estimates of absolute ΔG_{HB} and/or ΔH_{HB} values. It is also apparent that supramolecular approach is superior to the contact probability method in terms of more sensible correlation slopes.

For **pyridine complexes** only ΔH_{HB} values were available from the literature. Those were found to be predicted quite well with the SM approach: despite the bias, calculated and experimental values are linearly correlated, with only slightly different correlation slopes for different carboxylic acids. The CP approach yielded severely contracted range of predicted values, albeit still correlated with the experiment.

For complexes of **triphenylphosphine oxide and sulfide**, both ΔG_{HB} and ΔH_{HB} values were predicted and evaluated against the experimental data. The SM approach was found to systematically overestimate the stabilities of the HBs in terms of Gibbs free energies, while calculated HB enthalpies were mostly positively biased. In contrast, the CP approach produced more realistic ΔG_{HB} values but rather poorly predicted HB formation enthalpies. In case of both CP and SM the predicted ΔG_{HB} values were better correlated with the experiment than ΔH_{HB} values. Also, both Ph₃PO and Ph₃PS complexes fall on the same correlation lines.

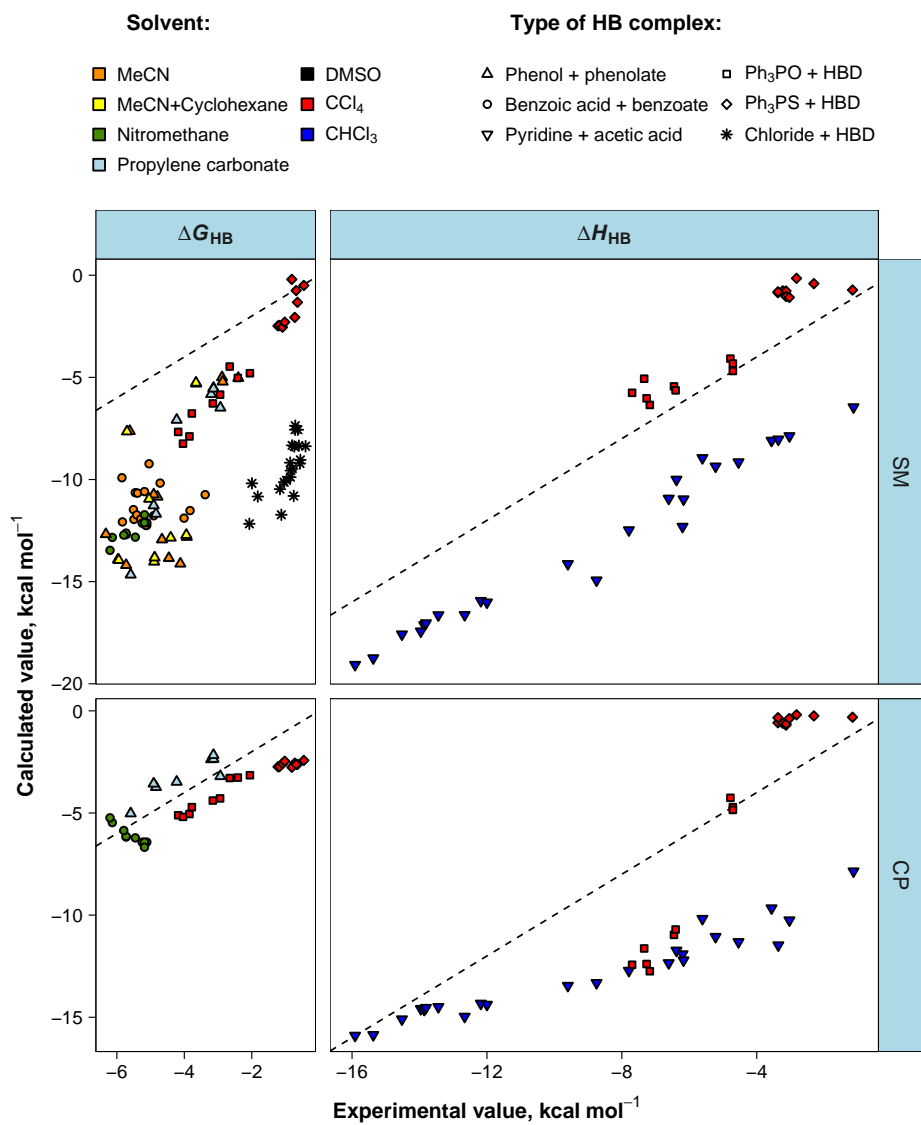


Figure 2. Results of HB modeling with COSMO-RS. Dashed lines mark the diagonals of the plots ($y = x$). **SM:** supramolecular approach, values calculated taking into account water impurity in the solvent as reported in the experiment description, parametrization BP_TZVPD_FINE_C30_1601 for chloride + HBD complexes and BP_TZVP_C30_1201 for other complexes. **CP:** contact probability method, pure solvents, parametrization BP_TZVP_C30_1201.

The ΔG_{HB} values calculated using the SM approach were often more negative than the corresponding ΔH_{HB} values, while the experimental data showed the contrary. This indicates the need for separate empirical corrections for computed ΔG_{HB} and ΔH_{HB} values in case when accuracy on the absolute scale is required. Also, the entropy change of the reaction should be estimated elsewhere than from uncorrected computed Gibbs free energy and enthalpy of the reaction.

Modeling of **homoconjugation** with COSMO-RS gave rather unsatisfactory results, with HB strengths being systematically overestimated and very poorly correlated with the experimental data (Figure 3). The least disappointing results were achieved in case of phenol+phenolate complexes in propylene carbonate: there is a tolerably good correlation between computations and experiment, although in case of SM approach both additive and multiplicative systematic errors are significant.

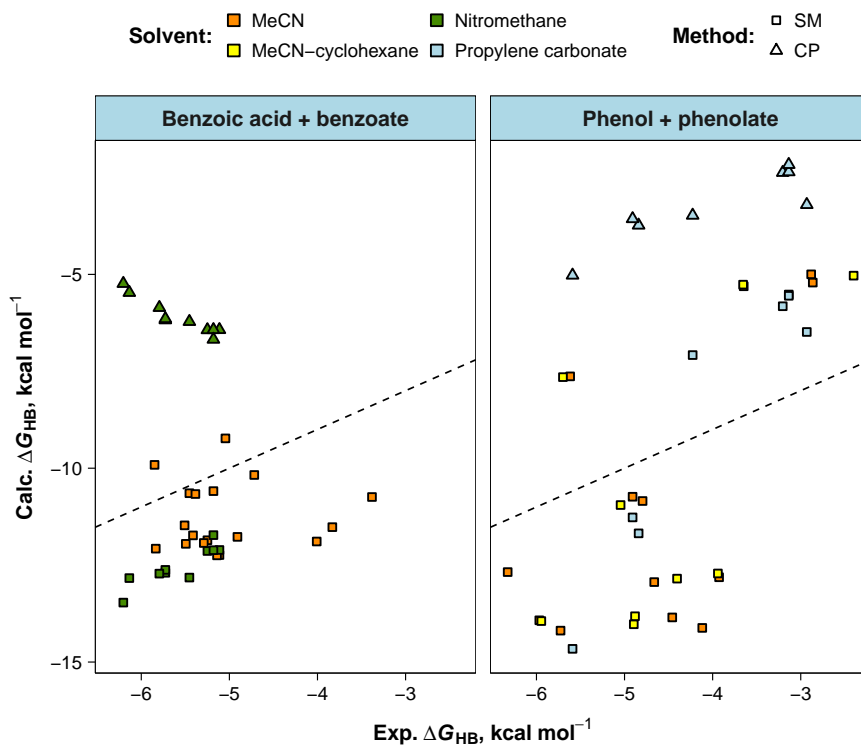


Figure 3. Modeling of homoconjugation reaction with COSMO-RS method (parametrization BP_TZVP_C30_1201, water impurity taken into account in case of SM). Dashed lines are $y = x - 5$.

It was found that taking the solvent's water content into account in case of the SM approach had practically no effect on the overall prediction quality and nature of the systematic errors neither for neutral nor for charged complexes. It is possible that the effect of water impurity on HB equilibria is too small compared to the various error sources of the computational procedure. Using different COSMO-RS parametrizations gave similar results in case of the SM approach, while the CP approach proved to be more sensitive, with nature of the errors notably dependent on parametrization. However, none of the option combinations tried in work I significantly improved prediction accuracy for homoconjugation reaction.

Since modeling of charged HB complexes by the default COSMO-RS procedure was not successful, a more elaborate and methodologically correct approach was used for description of chloride binding with various HB donors. Born-Haber cycle (Scheme 1) combining gas-phase energies and free energies of solvation was employed instead of using only energies computed in ideal conductor. The respective gas-phase energies were computed using G4MP2^[73] method, one of the most accurate methods currently available for medium-size molecules, while solvation energies were calculated using SMD and COSMO-RS methods. Unlike the case of homoconjugation, reference experimental data for chloride binding were highly reliable and self-consistent as they were obtained by an accurate relative binding affinity measurement method.^[77] However, the span of the experimental ΔG_{HB} values in DMSO was still close to the possible magnitude of computational error.

The results are presented in Figure 4 (results calculated by the default COSMO-RS procedure are also given for comparison). Neither SMD nor COSMO-RS provided predictions accurate enough to be used for most practical purposes. In fact, experimental ΔG_{HB} values were better correlated with gas-phase HB energies than computational ΔG_{HB} values taking solvent effects into account. Also, in the case of G4MP2+COSMO-RS there is a systematic shift of ΔG_{HB} values, suspectedly due to the biased prediction of chloride solvation energy. However, the systematic error is not as severe as it is in case of the default COSMO-RS procedure (Figure 4, rightmost plot). On the other hand, COSMO-RS provided somewhat more logical ranking of different compound classes than SMD. G4MP2+SMD protocol erroneously predicted fluorinated alcohols to give the strongest HBs, which can be explained by intrinsic inability of SMD to account for strong solute-solvent HBs. Another peculiarity is the description of amides (urea, thiourea and N,N'-diphenylurea), which all are bidentate HB donors while the rest of the studied donors are monodentate. The experimental ΔG_{HB} values of amide-chloride complexes are not outstanding compared to other compound classes. However, SMD+G4MP2 clearly underestimates the stability of bidentate complexes, while protocols involving COSMO-RS tend to overestimate it.

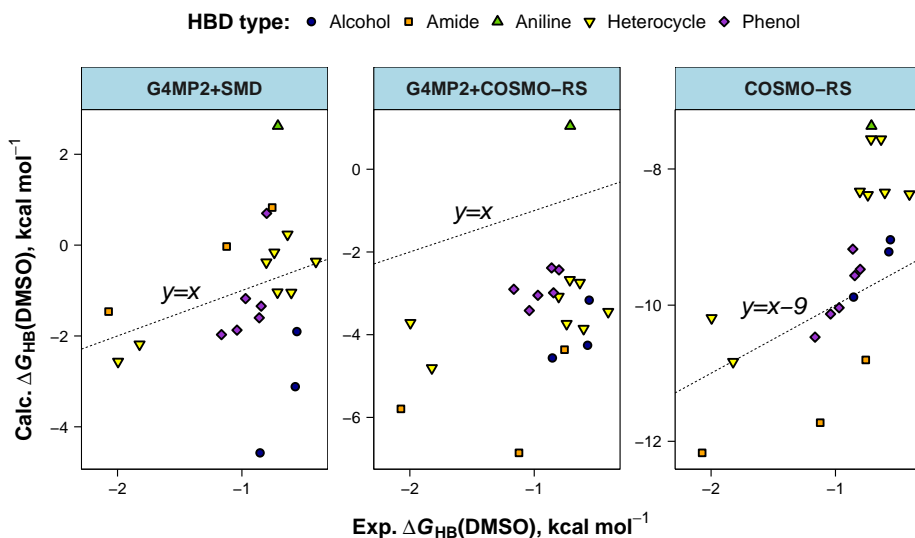


Figure 4. Modeling of chloride association with various HB donors using different methods.

In conclusion, it was found that COSMO-RS method used with the SM approach gives satisfactory relative thermodynamic parameters for HB formation reactions between non-charged species. The CP approach was found to give generally inferior results even for simple model reactions. All tested computational procedures were found inadequate for estimating HB energies in the case of charged complexes with narrowly spanned $\Delta G_{\text{HB}}/\Delta H_{\text{HB}}$ values.

4.2. Extraction modeling for compounds with unknown structure

The possibility to predict the behavior of solutes with unknown structure in different solvent pairs was explored in work III. The conceptual solution to this problem was proposed by researchers from Pfizer Inc. and the work was conducted in cooperation with the company.

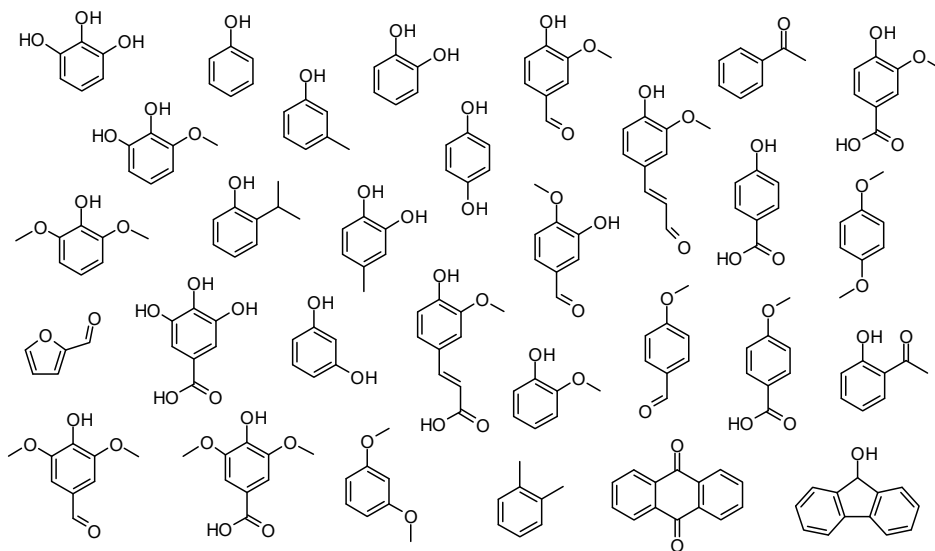
A very common situation in practice is the need to remove unidentified impurities from multicomponent solution, possibly containing other unknown solutes, or to separate the compound of interest from a complex mixture (e.g. plant extract). An ideal extraction solvent would *selectively* and *effectively* separate one or several compounds from the rest of the mixture constituents. With numerous solvents and solvent mixtures on offer, finding the best one experimentally is highly impractical. However, theoretical methods for $\log D$ prediction cannot be used for solvent selection without having structural information on all mixture components, leaving the trial and error approach as the only option. The proposed solution to this problem is a predictive model

relying on a small amount of experimental extraction data that can be obtained without the need to identify any of the involved compounds.

As was described in section 2.4.4, partitioning of a neutral solute between different media can be satisfactorily described by a simple LFER equation based on five solute parameters: the descriptors of HB acidity, HB basicity, dipolarity, polarizability, and molar volume. The modeling approach developed in this work relies on the reverse principle: deducing properties of the solute from its observed distribution patterns in chemically diverse solvent pairs.^[36,78,79] The unknown property – $\log D$ value of the solute in a certain solvent pair – is viewed as a function of the experimental $\log D$ values of this solute in other solvent pairs. Thus, the amount of the experimental work can be reduced to the experiments with just a few solvent pairs (reference solvent pairs).

The experimental methodology for determination of distribution ratios adopted in work III is suitable for the use in practically relevant conditions: with multicomponent, possibly very complex mixtures and without any structural information on the component(s) of interest.

For development of the predictive model a set of 30 neutral or moderately acidic solutes (Scheme 2) were chosen. Previous experience suggested that composing the training set of highly diverse solutes may lead to rather moderate accuracy of predictions, so a set of relatively similar lignin-related compounds (lignin decomposition products or structurally similar molecules) was chosen in this work. The choice of solutes was also influenced by the fact that valorization of lignin as an abundant, renewable natural raw material is attracting considerable interest.^[80,81]



Scheme 2. Structures of lignin-related compounds used in development of the empirical predictive model (work III).

Ten organic solvents were involved, representing various compound classes and HB ability: *n*-hexane, toluene, chlorobenzene, dichloromethane (DCM), methyl *tert*-butyl ether (MTBE), isopropyl acetate, benzyl alcohol, cyclohexanol, *n*-butanol and *n*-octanol. The solvents were combined with two aqueous phases – neutral and acidic – to form 20 solvent pairs. Basic aqueous phase, although potentially useful for separating weakly acidic compounds from non-ionizable ones, proved problematic from the practical point of view and was omitted. Log*P* and log*D* values of 30 solutes in 20 solvent pairs were determined experimentally or, in practically problematic cases, estimated based on the rest of the experimental data.

The developed prediction approach is based on multilinear regression. First, *N* reference solvent pairs are selected in which the unknown solute(s) are studied experimentally. Then, using the training data, multilinear equations are composed to connect log*D* values of the solutes in reference solvent pairs to the log*D* values in the rest of the solvent pairs. The mathematical model takes the following form:

$$\log D(X) = a_X + \sum_{i=1}^N \{\rho_X(\text{RSP}_i) \cdot \log D(\text{RSP}_i)\} \quad 4.1$$

where *X* denotes a solvent pair in which log*D* value of the solute is to be predicted, RSP_{*i*} is the *i*-th reference solvent pair, and *a_X* and *ρ_X* are model parameters (one constant *a* and *N* coefficients *ρ* for each solvent pair *X*).

After log*D* values of the unknown solute in reference solvent pairs have been determined experimentally, equation 4.1 and model parameters calculated from training data are used to obtain the log*D* values in other solvent pairs. Training data can be adjusted according to the circumstances prior to building the model, e.g. unsuitable solvent pairs, if any, and compounds known or suspected to be considerably different from the solute of interest can be excluded.

Internal validation (leave-one-out, LOO) demonstrated good performance of the predictive algorithm under conditions where the properties of the solvent pairs were not influenced by the solutes (i.e. in sufficiently dilute solutions). Minimum number of reference solvent pairs was also optimized in LOO tests, and set to 4 in case of non-ionizable solutes and 5 in case of acidic solutes. The root mean square error (RMSE) of the predictions in LOO test involving full training set and 5 reference solvent pairs was 0.16 log units (for comparison, precision of the experimental data was evaluated at standard deviation level as 0.08 log units). Also, the most suitable solvent pair for separation of a random pair of training set compounds was correctly identified in 60% of cases, and was amongst the predicted 3 best in 90% of cases.

External validation of the predictive model using relatively concentrated solutions of natural products (cedar oil, black cumin oil and colophony) revealed that in cases when the influence of solutes on the properties of the solvent pair is not negligible the predictions can be significantly biased. The largest prediction errors were observed in the case of hexane – the least polar of

the employed organic solvents and thus the most susceptible to the effect of polar additives. As illustrated by Figure 5, predicted $\log D$ values in hexane/water system are considerably more negative than the observed ones, likely because the high concentration of various mixture components in the organic phase significantly increases its polarity and therefore affinity towards polar solutes. Intrinsically polar solvents are much less susceptible to this effect, so prediction errors in their case are smaller. Also, it was confirmed experimentally that strongly polar solutes are more affected by “foreign” components in the solvent systems than the less polar ones.

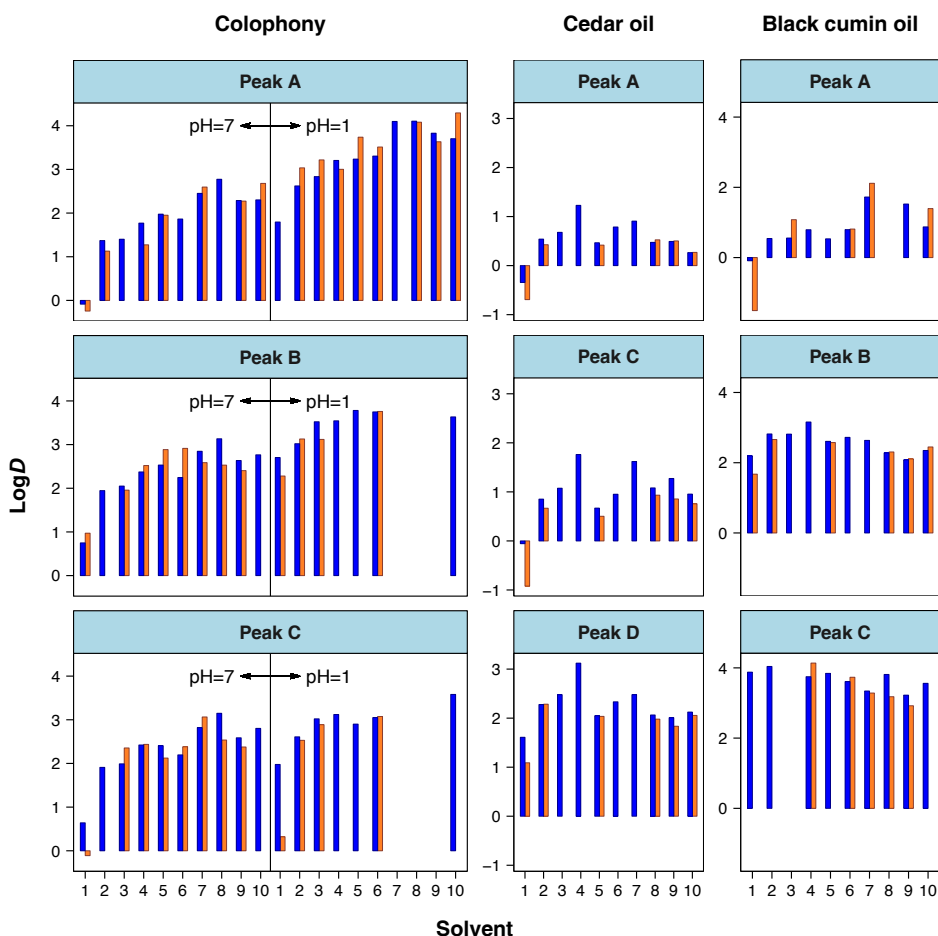
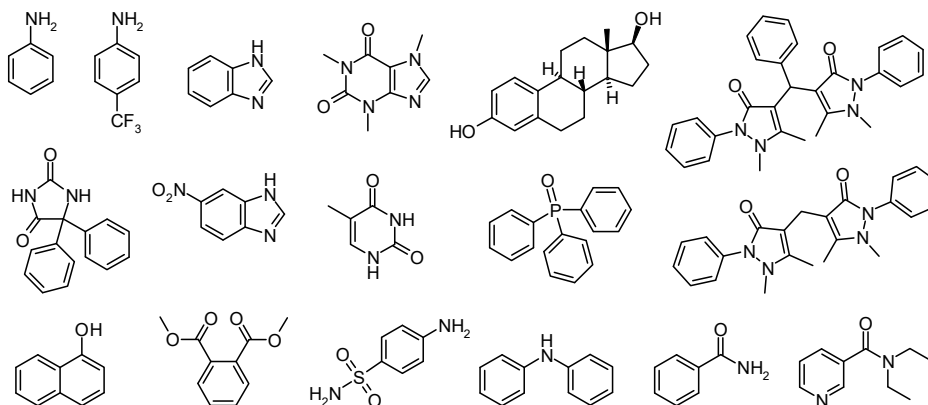


Figure 5. Experimental (blue) and predicted (orange) $\log D$ values of some of the unidentified components of colophony, black cumin oil and cedar oil (data from work III). Experiments showed the chosen components of cumin and cedar oil to be unaffected by pH of the aqueous phase, so the predictions were made using half of the training set (one aqueous phase and 10 organic solvents) and 4 input values. Solvents: 1 – *n*-hexane, 2 – toluene, 3 – chlorobenzene, 4 – DCM, 5 – MTBE, 6 – isopropyl acetate, 7 – benzyl alcohol, 8 – cyclohexanol, 9 – *n*-butanol, 10 – *n*-octanol.

As demonstrated in the course of external validation, the accuracy of the predictive model in real-life situations is expected to be lower (RMSE of 0.3–0.5 log units), yet it still gives useful qualitative or semi-quantitative predictions and allows to reduce the amount of the experimental work required for process optimization.

4.3. Modeling liquid-liquid extraction with COSMO-RS

Modeling of partition and distribution equilibria in different solvent pairs with COSMO-RS was carried out in work IV. Experimental $\log P$ and $\log D$ values of 30 neutral and acidic ligneous compounds (Scheme 2) collected in work III and an additional dataset of 17 neutral and weakly basic drug-related compounds (Scheme 3) were used as reference for evaluation of computational data.



Scheme 3. Structures of drug-like compounds (work IV).

Modeling of extraction outcome was divided into three main stages:

1. Assessment of mutual solubilities of the solvents.
2. Prediction of $\log P$ values of the solutes.
3. Calculation of ionization corrections and $\log D$ values for acidic and basic solutes.

Performance of the method at each stage was assessed separately in addition to the evaluation of the overall prediction quality. Solute-solute interactions were not taken into account in this work: it was assumed that the properties of the extraction system are determined solely by solvents and not affected by solutes. This assumption holds for sufficiently dilute solutions and is probably valid for most experimental data used as reference.

The first stage is prediction of the **equilibrium phase compositions**. Even the solvents that are classified as immiscible and that form two-phase systems when combined are soluble in each other to some extent, and it is important to

take their mutual solubility into account. It was found that even though the predicted solubility values are often biased by 50% and more, the accuracy is sufficient to make the fully *a-priori* $\log P$ predictions close in quality to the prediction with the use of experimental solubility values. The accuracy improvement provided by taking the mutual saturation into account (evaluated against prediction where pure solvents were used as liquid phases) is the most noticeable in case of hydrophilic solvents containing large amount of water when at equilibrium. The improvement in case of highly hydrophobic solvents was less pronounced.

Log P values of various solutes were predicted quite well by COSMO-RS (Figure 6). The solvent-specific systematic errors of prediction (i.e. deviations of the correlation parameters from theoretical values), although statistically significant in some cases, were moderate. The type of solute (ligneous vs drug-like) was found to have little influence on the accuracy of the prediction. The RMSE values for different solvent pairs varied from 0.2 to 0.7, the overall value for fully *a-priori* prediction being 0.6 log units.

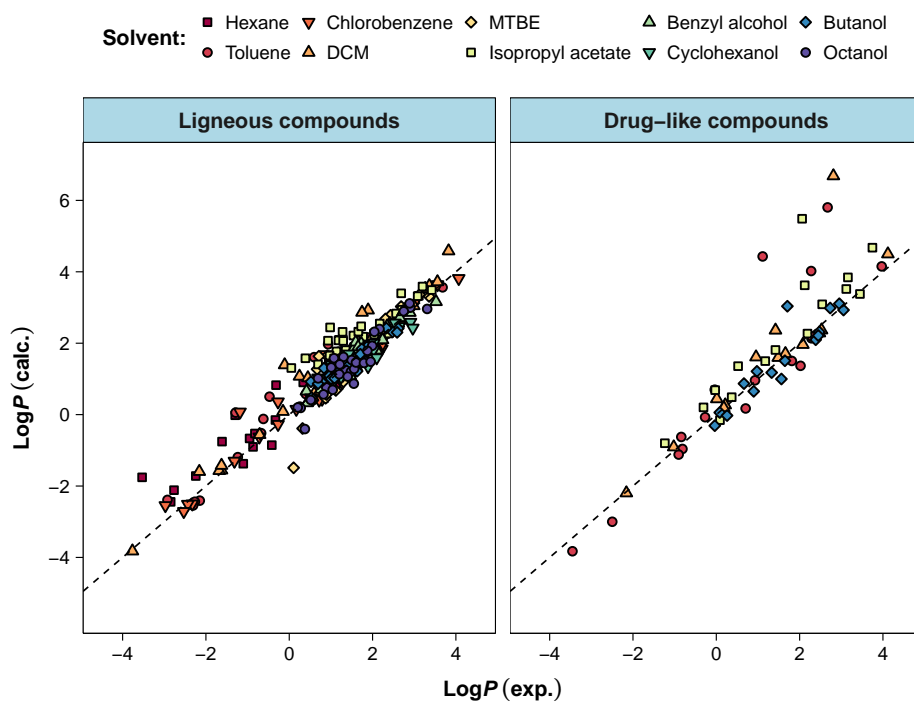


Figure 6. Correlations of experimental and *a-priori* predicted $\log P$ values for the two sets of studied solutes. Dashed lines denote the theoretical regression lines ($y = x$).

Analysis of the data revealed a few solutes that were described consistently poorly in different solvents. The most prominent ones were diantipyrylmethane and diantipyrylphenylmethane – the corresponding $\log P$ values in different solvents were overestimated by over 3 log units. Lesser deviations were observed for compounds in which hydroxy-substituent in the aromatic ring was neighbored by methoxy or hydroxy groups (e.g. syringol). The suspected reason for the deviations is the imperfection of the DFT procedure. It is relatively easily explained in case of syringol-like compounds – the geometry optimization yields only structures with intramolecular hydrogen bonds, while in reality, if the medium contains HB acceptors, intermolecular HB formation might be possible. As the DFT procedure does not produce HBD structures, the compound is likely to appear more hydrophobic in calculations than it is in reality. The deviations of diantipyryl-containing compounds are more difficult to rationalize, but the suspected reason is the failure of the used QM method to correctly describe the polarity of this unusual structural fragment.

The results of **logD prediction** are presented in Figure 7. The default calculation of $\log D$ values in the COSMO $therm$ software proceeds according to equation 2.7, which assumes that ionized forms of the solute do not partition into the organic phase. This approach requires knowing only the aqueous pK_a and $\log P$ values of the solute. COSMO-RS is known to produce adequate pK_a estimates for acids and bases in different media.^[19,25,82] However, predictions may often be biased on absolute scale, requiring empirical corrections to achieve good accuracy. The quality of pK_a predictions varied for different types of ionizable solutes studied in work IV: pK_a values were predicted well (RMSE = 0.35) for weak bases, overestimated by 0.4 ... 0.9 units for carboxylic acids and underestimated by 1.2 ... 1.9 units for weak OH acids. As can be seen from Figure 7, it contributed significantly to the error of *a-priori* $\log D$ prediction.

The other major difficulty in $\log D$ prediction is modeling of the behavior of ions in two-phase systems. While in the aqueous phase solute ions are likely to exist as solvated free species, in the less polar organic phase they can also form stable ion pairs or exist as a part of more complex associates (e.g. ion triplets, quartets, etc, possibly involving hydrogen bonds). Explicit accounting for all involved equilibria requires strenuous calculations of various association constants or using experimental data that are not readily available. Since it is not expedient for most practical purposes, in this work the approaches requiring only unimolecular QM calculations of neutral and ionized solute were used. Specifically, partitioning of ions into the organic phase was accounted for by equation 2.8.

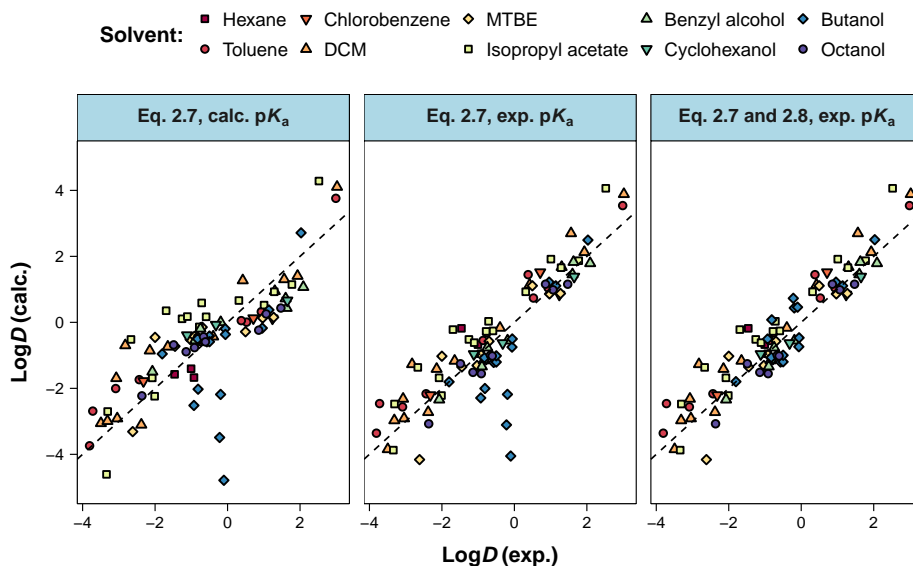


Figure 7. Results of $\log D$ prediction using different equations and pK_a values. Equation 2.8 was only applied to butanol/water system.

The extent of ion partitioning into the organic phase is expected to be the highest in case of polar, HB-active organic solvents and strongly lipophilic ionized forms of the solutes. Indeed, the $\log D$ values in butanol/water solvent pair were the most severely mispredicted. It was found that while using equation 2.8 can markedly reduce $\log D$ prediction errors in case of butanol, it gave strongly biased results for isopropyl acetate and produced practically no change in case of toluene and DCM. Also, for acidic species the equation gave no noticeable improvement compared to the simplified equation 2.7 as all involved anions were predicted to be exceedingly hydrophilic. Therefore it was recommended to only use equation 2.8 in case of hydrophilic HBD solvents and not for HBA-only solvents.

The RMSE of *a-priori* $\log D$ prediction (while using ion partition correction only in case of butanol/water system) was 0.9 for acidic and 0.7 for basic solutes. Figure 7 illustrates the gradual improvement of $\log D$ prediction quality when using experimental pK_a values instead of computational and equation 2.8 instead of the simplified representation.

In conclusion, COSMO-RS was found to yield good-quality predictions of $\log P$ values in diverse solvent pairs without any additional experimental input. $\log D$ predictions were expectedly less accurate, whereas the prediction errors stemmed both from inaccurate pK_a values and too simplified representation of behavior of ionic species.

5. SUMMARY

The aim of this work was to evaluate or develop approaches for modeling intermolecular interactions in solutions and the resulting macroscopic solute properties. Another issue addressed in this work besides conventional structure-based property prediction was the possibility to predict liquid-liquid partitioning of solutes with unknown structure.

First, COSMO-RS method and combined approaches involving COSMO-RS, SMD model and G4MP2 method were evaluated as tools for predicting hydrogen bond energies in non-aqueous solutions. In case of hydrogen bonds between neutral molecules satisfactory results were obtained with the default COSMO-RS procedure. However, the default procedure proved inadequate for modeling narrowly spanned bond energies in negatively charged hydrogen-bonded complexes. The methodology based on thermodynamic cycle combining solvation energies with gas-phase energies also demonstrated insufficient accuracy for the task.

Secondly, empirical prediction model was developed for estimating the outcome of liquid-liquid extraction for compounds with unknown structures. The model based on multilinear regression requires few experimentally obtained $\log D$ values of the solute in different solvent pairs as inputs to predict $\log D$ values in a variety of other solvent pairs. Experimental methodology for determination of $\log D$ values of unidentified components was elaborated, along with necessary data treatment and uncertainty estimation procedures. The experimental method – shake-flask extraction followed by HPLC analysis – is robust and uses conventional laboratory equipment.

The quality of the prediction by the empirical model varied depending on the extent of solute-solute interactions (mainly in the less polar organic phase), yet useful $\log D$ estimates were obtained even when quantitative accuracy was not reached. The root mean square error of prediction varied from <0.2 log units (dilute solutions, high-quality data) to 0.5 log units (concentrated solutions, experimental difficulties).

Finally, experimental partition/distribution data obtained in the course of the work was used to evaluate the quality of COSMO-RS predictions for various solvent pairs and types of solutes. Both overall prediction quality and accuracy of calculated intermediate parameters were assessed. The overall RMSE of *a-priori* $\log P$ prediction was 0.6 log units, varying from 0.2 to 0.7 for different solvent pairs. The accuracy of *a-priori* $\log D$ predictions was expectedly lower due to imperfection of pK_a prediction and complex behavior of ions in two-phase systems. Problematic solutes were identified and the reasons of under-average prediction accuracy in each case were suggested.

REFERENCES

- [1] C. Reichardt, T. Welton, *Solvents and Solvent Effects in Organic Chemistry*, Wiley-VCH, Weinheim/Germany, **2011**.
- [2] P. Góralski, *Fluid Phase Equilib.* **2000**, *167*, 207–221.
- [3] C. A. Hunter, *Chem. Sci.* **2013**, *4*, 834–848.
- [4] E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. J. Dannenberg, P. Hobza, et al., *Pure Appl. Chem.* **2011**, *83*, 1619–1636.
- [5] G. A. Jeffrey, W. Saenger, *Hydrogen Bonding in Biological Structures*, Springer-Verlag, Berlin, **1994**.
- [6] P. Muller, *Pure Appl. Chem.* **1994**, *66*, 1077–1184.
- [7] A. Klamt, F. Eckert, W. Arlt, *Annu. Rev. Chem. Biomol. Eng.* **2010**, *1*, 101–122.
- [8] I. Leito, I. A. Koppel, P. Burk, S. Tamp, M. Kutsar, M. Mishima, J.-L. M. Abboud, J. Z. Davalos, R. Herrero, R. Notario, *J. Phys. Chem. A* **2010**, *114*, 10694–10699.
- [9] P. F. B. Gonçalves, H. Stassen, *J. Comput. Chem.* **2003**, *24*, 1758–1765.
- [10] A. A. Hassanali, J. Cuny, V. Verdolino, M. Parrinello, *Phil. Trans. R. Soc. A* **2014**, *372*, 20120482.
- [11] B. M. Rode, T. S. Hofer, B. R. Randolph, C. F. Schwenk, D. Xenides, V. Vchirawongkwin, *Theor. Chem. Acc.* **2006**, *115*, 77–85.
- [12] J. Tomasi, B. Mennucci, R. Cammi, *Chem. Rev.* **2005**, *105*, 2999–3094.
- [13] J. R. Pliego, J. M. Riveros, *J. Phys. Chem. A* **2001**, *105*, 7241–7247.
- [14] A. Klamt, *J. Phys. Chem.* **1995**, *99*, 2224–2235.
- [15] A. Klamt, V. Jonas, T. Bürger, J. C. W. Lohrenz, *J. Phys. Chem. A* **1998**, *102*, 5074–5085.
- [16] A. Klamt, *COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*, Elsevier Science Ltd, Amsterdam, **2005**.
- [17] A. Klamt, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, e1338.
- [18] A. Klamt, B. Mennucci, J. Tomasi, V. Barone, C. Curutchet, M. Orozco, F. J. Luque, *Acc. Chem. Res.* **2009**, *42*, 489–492.
- [19] F. Eckert, I. Leito, I. Kaljurand, A. Kütt, A. Klamt, M. Diedenhofen, *J. Comput. Chem.* **2009**, *30*, 799–810.
- [20] M. A. R. Martins, C. M. S. S. Neves, K. A. Kurnia, A. Luis, L. M. N. B. F. Santos, M. G. Freire, S. P. Pinho, J. A. P. Coutinho, *Fluid Phase Equilib.* **2014**, *375*, 161–167.
- [21] M. Buggert, L. Mokrushina, I. Smirnova, R. Schomäcker, W. Arlt, *Chem. Eng. Technol.* **2006**, *29*, 567–573.
- [22] A. Klamt, J. Reinisch, F. Eckert, A. Hellweg, M. Diedenhofen, *Phys. Chem. Chem. Phys.* **2012**, *14*, 955–963.
- [23] A. Klamt, J. Reinisch, F. Eckert, J. Graton, J.-Y. L. Questel, *Phys. Chem. Chem. Phys.* **2013**, *15*, 7147–7154.
- [24] A. V. Marenich, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. B* **2009**, *113*, 6378–6396.
- [25] F. Eckert, M. Diedenhofen, A. Klamt, *Mol. Phys.* **2010**, *108*, 229–241.
- [26] M. Lökov, S. Tshepelevitsh, A. Heering, P. G. Plieger, R. Vianello, I. Leito, *Eur. J. Org. Chem.* **2017**, *2017*, 4475–4489.
- [27] C. P. Kelly, C. J. Cramer, D. G. Truhlar, *J. Chem. Theory Comput.* **2005**, *1*, 1133–1152.

- [28] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Delivery Rev.* **1997**, *23*, 3–25.
- [29] T. J. Hou, X. J. Xu, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 2137–2152.
- [30] T. Hou, J. Wang, Y. Li, *J. Chem. Inf. Model.* **2007**, *47*, 2408–2415.
- [31] R. D. Saiakhov, L. R. Stefan, G. Klopman, *Perspect. Drug Discovery Des.* **2000**, *19*, 133–155.
- [32] C. Hansch, A. Leo, S. B. Meikapati, A. Kurup, *Bioorg. Med. Chem.* **2004**, *12*, 3391–3400.
- [33] Y. H. Zhao, X. Yuan, L. M. Su, W. C. Qin, M. H. Abraham, *Chemosphere* **2009**, *75*, 866–871.
- [34] A. M. Voutchkova, J. Kostal, J. B. Steinfeld, J. W. Emerson, B. W. Brooks, P. Anastas, J. B. Zimmerman, *Green Chem.* **2011**, *13*, 2373–2379.
- [35] E. Papa, F. Villa, P. Gramatica, *J. Chem. Inf. Model.* **2005**, *45*, 1256–1266.
- [36] M. Shalaeva, G. Caron, Y. A. Abramov, T. N. O’Connell, M. S. Plummer, G. Yalamanchi, K. A. Farley, G. H. Goetz, L. Philippe, M. J. Shapiro, *J. Med. Chem.* **2013**, *56*, 4870–4879.
- [37] E. Rutkowska, K. Pajak, K. Jóźwiak, *Acta Pol. Pharm.* **2013**, *70*, 3–18.
- [38] S. Selberg, T. Rodima, M. Lõkov, S. Tshepelevitsh, T. Haljasorg, S. Chhabra, S. A. Kadam, L. Toom, S. Vahur, I. Leito, *Tetrahedron Lett.* **2017**, *58*, 2098–2102.
- [39] J. Rydberg, M. Cox, C. Musikas, G. Choppin, Eds., *Solvent Extraction Principles and Practice*, M. Dekker, New York, **2004**.
- [40] N. M. Rice, H. M. N. H. Irving, M. A. Leonard, *Pure Appl. Chem.* **1993**, *65*, 2373–2396.
- [41] A. S. Rustenburg, J. Dancer, B. Lin, J. A. Feng, D. F. Ortwine, D. L. Mobley, J. D. Chodera, *J. Comput.-Aided Mol. Des.* **2016**, *30*, 945–958.
- [42] A. Avdeef, *Absorption and Drug Development: Solubility, Permeability, and Charge State*, John Wiley & Sons, Inc., Hoboken, NJ, USA, **2012**.
- [43] A. Andrés, M. Rosés, C. Ràfols, E. Bosch, S. Espinosa, V. Segarra, J. M. Huerta, *Eur. J. Pharm. Sci.* **2015**, *76*, 181–191.
- [44] A. Berthod, S. Carda-Broch, *J. Chromatogr. A* **2004**, *1037*, 3–14.
- [45] D. Soulsby, J. A. M. Chica, *Magn. Reson. Chem.* **2017**, *55*, 724–729.
- [46] S. M. Ulmeanu, H. Jensen, G. Bouchard, P.-A. Carrupt, H. H. Girault, *Pharm. Res.* **2003**, *20*, 1317–1322.
- [47] C. Ràfols, E. Bosch, R. Ruiz, K. J. Box, M. Reis, C. Ventura, S. Santos, M. E. Araújo, F. Martins, *J. Chem. Eng. Data* **2012**, *57*, 330–338.
- [48] B. Burghoff, J. Schiferli, J. Sousa Marques, A. B. de Haan, *Chem. Eng. Sci.* **2009**, *64*, 2887–2892.
- [49] M. Vitha, P. W. Carr, *J. Chromatogr. A* **2006**, *1126*, 143–194.
- [50] G. Schuster, W. Lindner, *J. Chromatogr. A* **2013**, *1273*, 73–94.
- [51] P. Wiczling, P. Kawczak, A. Nasal, R. Kaliszan, *Anal. Chem.* **2006**, *78*, 239–249.
- [52] K. L. Valkó, *J. Pharm. Biomed. Anal.* **2016**, *130*, 35–54.
- [53] B. Bard, P.-A. Carrupt, S. Martel, *J. Chromatogr. A* **2012**, *1260*, 164–168.
- [54] Z. Jia, *Curr. Pharm. Anal.* **2005**, *1*, 41–56.
- [55] S. Han, J. Qiao, Y. Zhang, L. Yang, H. Lian, X. Ge, H. Chen, *Chemosphere* **2011**, *83*, 131–136.
- [56] C. E. Poulsen, R. C. R. Wootton, A. Wolff, A. J. deMello, K. S. Elvira, *Anal. Chem.* **2015**, *87*, 6265–6270.
- [57] J. J. Bao, X. Liu, Y. Zhang, Y. Li, *J. Chromatogr. B* **2014**, *967*, 183–189.
- [58] M. Abolhasani, C. W. Coley, K. F. Jensen, *Anal. Chem.* **2015**, *87*, 11130–11136.

- [59] R. Mannhold, G. I. Poda, C. Ostermann, I. V. Tetko, *J. Pharm. Sci.* **2009**, *98*, 861–893.
- [60] G. Wienke, J. Gmehling, *Toxicol. Environ. Chem.* **1998**, *65*, 57–86.
- [61] A. A. Petrauskas, E. A. Kolovanov, *Perspect. Drug Discovery Des.* **2000**, *19*, 99–116.
- [62] M. H. Abraham, W. E. Acree, *J. Org. Chem.* **2010**, *75*, 1006–1015.
- [63] J.-C. Bradley, M. H. Abraham, W. E. Acree Jr., A. S. I. D. Lang, *Chem. Cent. J.* **2015**, *9*, 1–10.
- [64] M. H. Abraham, A. Ibrahim, A. M. Zissimos, *J. Chromatogr. A* **2004**, *1037*, 29–47.
- [65] A. M. Zissimos, M. H. Abraham, A. Klamt, F. Eckert, J. Wood, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1320–1331.
- [66] C. Wittekindt, A. Klamt, *QSAR Comb. Sci.* **2009**, *28*, 874–877.
- [67] A. Klamt, F. Eckert, J. Reinisch, K. Wichmann, *J. Comput. Aided Mol. Des.* **2016**, *30*, 959–967.
- [68] T. Jeliński, P. Cysewski, *Chem. Eng. Res. Des.* **2017**, *122*, 176–183.
- [69] L. Mokrushina, M. Buggert, I. Smirnova, W. Arlt, R. Schomäcker, *Ind. Eng. Chem. Res.* **2007**, *46*, 6501–6509.
- [70] A. Klamt, *Fluid Phase Equilib.* **2003**, *206*, 223–235.
- [71] S. Wille, M. Buggert, L. Mokrushina, W. Arlt, I. Smirnova, *Chem. Eng. Technol.* **2010**, *33*, 1075–1082.
- [72] T. Ingram, U. Richter, T. Mehling, I. Smirnova, *Fluid Phase Equilib.* **2011**, *305*, 197–203.
- [73] L. A. Curtiss, P. C. Redfern, K. Raghavachari, *J. Chem. Phys.* **2007**, *127*, 124105.
- [74] A. Klamt, V. Jonas, *J. Chem. Phys.* **1996**, *105*, 9972–9981.
- [75] K. Baldrige, A. Klamt, *J. Chem. Phys.* **1997**, *106*, 6622–6633.
- [76] A. Schäfer, A. Klamt, D. Sattel, J. C. W. Lohrenz, F. Eckert, *Phys. Chem. Chem. Phys.* **2000**, *2*, 2187–2193.
- [77] S. A. Kadam, K. Haav, L. Toom, T. Haljasorg, I. Leito, *J. Org. Chem.* **2014**, *79*, 2501–2513.
- [78] C. Yamagami, A. Hamasaki, Y. Kumagai, T. Moritoki, T. Tanahashi, *Chem. Pharm. Bull.* **2005**, *53*, 398–401.
- [79] A. M. Zissimos, M. H. Abraham, M. C. Barker, K. J. Box, K. Y. Tam, *J. Chem. Soc., Perkin Trans. 2* **2002**, 470–477.
- [80] R. Rinaldi, R. Jastrzebski, M. T. Clough, J. Ralph, M. Kennema, P. C. A. Bruijninx, B. M. Weckhuysen, *Angew. Chem. Int. Ed.* **2016**, *55*, 8164–8215.
- [81] F. G. Calvo-Flores, J. A. Dobado, *ChemSusChem* **2010**, *3*, 1227–1235.
- [82] F. Eckert, A. Klamt, *J. Comput. Chem.* **2006**, *27*, 11–19.

SUMMARY IN ESTONIAN

Soluut-solvent vastasmõjude eksperimentaalne uurimine ja modelleerimine

Käesoleva töö eesmärgiks oli hinnata ning arendada meetodeid intermolekulaarsete vastasmõjude ja nendest tulenevate soluudi makroskoopiliste omaduste modelleerimiseks vedelfaasis. Lisaks molekulaarstruktuuril põhinevatele meetoditele vaadeldi võimalust ennustada tundmatu struktuuriga soluutide jaotust vedelfaaside vahel.

Töö esimese etapina hinnati mitme arvutusmetoodika sobivust vesiniksideme tugevuse ennustamiseks mittevesikeskkonnas. Selleks kasutati COSMO-RS meetodit ning G4MP2 ja COSMO-RS või SMD meetodit ühendavat kombineeritud lähenemist. Neutraalsete molekulide vahelise vesiniksideme puhul andis standardne COSMO-RS arvutus vastuvõetava kvaliteediga tulemusi. Negatiivselt laetud ja lähedaste dissotsiatsioonienergiatega vesiniksidestatud komplekside puhul osutus vaikumisi COSMO-RS protseduuri täpsus ebapiisavaks. Liiga ebatäpseks osutus ka termodünaamilisel tsüklil põhinev metodoloogia, milles kombineeriti soluutide solvatatsioonienergiad gaasifaasienergiatega.

Töö teise etapina töötati välja ennustav mudel vedelik-vedelik ekstraktsiooni tulemuste hindamiseks tundmatu struktuuriga soluutide jaoks. Mudel põhineb multilineaarsel regressioonil ning kasutab sisendina uuritava soluudi jaoks eksperimentaalselt määratud väikest arvu $\log D$ väärtusi erinevates solvendipaarides. Tulemuseks on soluudi $\log D$ väärtused suuremas arvus solvendipaarides. Töö käigus arendati välja eksperimentaalne meetoodika $\log D$ määramiseks ning vajalikud andmetötluse ja määramatuse hindamise protseduurid. Eksperimentaalne $\log D$ määramine koosneb vedelik-vedelik ekstraktsioonist ja mõlema faasi analüüsist HPLC meetodiga. Meetoodika on töökindel ja teostatav tavaliste laboratoorsete vahendite abil.

Ennustuse täpsus sõltus soluutidevaheliste vastasmõjude tugevusest lahustes, eelkõige vähempolaarses orgaanilises faasis. Praktiliselt kasulikke $\log D$ hinnanguid saadi ka siis, kui ennustustäpsus polnud väga kõrge. Ennustuse ruutkeskmine hälve oli vahemikus 0.2 log ühikut (lahjad lahused, täpne määramine) kuni 0.5 log ühikut (kontsentreeritud lahused, eksperimentaalsed raskused).

Töö viimases etapis kasutati töö käigus kogutud eksperimentaalseid andmeid orgaaniliste soluutide jaotuse kohta erinevates solvendipaarides COSMO-RS arvutuste kvaliteedi hindamiseks. Hinnati nii üldist ennustustäpsust kui ka vahepealsete väärtuste täpsust. *A-priori* arvatud $\log P$ väärtuste ruutkeskmine viga oli vahemikus 0.2...0.7 log ühikut sõltuvalt solvendipaarist, keskmisega üle kõigi solvendipaaride 0.6 log ühikut. *A-priori* $\log D$ ennustuste täpsus oli ootuspäraselt madalam. Põhjuseks olid nii ebatäpsed pK_a ennustused kui ka raskused ioonide jaotuse modelleerimisel kahefaasilises süsteemis. Identifitseeriti süsteemaatilistelt halvasti kirjeldatud soluudid ning pakuti välja madala ennustustäpsuse võimalikud põhjused.

ACKNOWLEDGMENTS

First and foremost, I would like express my deep gratitude to my long-time supervisor professor Ivo Leito for constant inspiration, guidance and support throughout my studies. I thank Drs Karin Kipper and Ivar Koppel for their invaluable help in technical matters. I would also like to thank all my colleagues from the Chair of Analytical Chemistry for kind help and great atmosphere. I am thankful to my co-authors for insightful discussions and suggestions.

This work was carried out using the instrumentation of the Estonian Center of Analytical Chemistry (www.akki.ee) and was supported by the EU through the European Regional Development Fund (project TK141 “Advanced materials and high-technology devices for energy recuperation systems” and Graduate School of Functional materials and technologies), as well as by the institutional research grant IUT20-14 from the Estonian Research Council. The computational results presented in this work were in part obtained using the High-Performance Computing Center of the University of Tartu.

PUBLICATIONS

CURRICULUM VITAE

Name: Sofja Tšepelevitš
Date of birth: December 30, 1989
Citizenship: Estonian
Contact: Institute of Chemistry, University of Tartu
14a Ravila Street, 50411, Tartu, Estonia
E-mail: sofja.ts@gmail.com

Education:

2014–... University of Tartu, Institute of Chemistry, PhD student
2012–2014 University of Tartu, Institute of Chemistry, M.Sc (Chemistry)
2009–2012 University of Tartu, Institute of Chemistry, B.Sc (Chemistry)

Professional employment:

2012–... Chemist, University of Tartu, Institute of Chemistry

Scientific publications:

1. Tshepelevitsh, S.; Oss, M.; Pung, A.; Leito, I. Evaluating the COSMO-RS Method for Modeling Hydrogen Bonding in Solution. *ChemPhysChem* **2013**, *14*, 1909–1919.
2. Brioché, J.; Pike, S. J.; Tshepelevitsh, S.; Leito, I.; Morris, G. A.; Webb, S. J.; Clayden, J. Conformational Switching of a Foldamer in a Multi-component System by pH-Filtered Selection between Competing Non-covalent Interactions. *J. Am. Chem. Soc.* **2015**, *137*, 6680–6691.
3. Leito, I.; Koppel, I. A.; Koppel, I.; Kaupmees, K.; Tshepelevitsh, S.; Saame, J. Basicity Limits of Neutral Organic Superbases. *Angew. Chem., Int. Ed.* **2015**, *54*, 9262–9265.
4. Saame, J.; Rodima, T.; Tshepelevitsh, S.; Kütt, A.; Kaljurand, I.; Haljasorg, T.; Koppel, I. A.; Leito, I. Experimental Basicities of Superbasic Phosphonium Ylides and Phosphazenes. *J. Org. Chem.* **2016**, *81*, 7349–7361.
5. Tshepelevitsh, S.; Trummal, A.; Haav, K.; Martin, K.; Leito, I. Hydrogen-Bond Donicity in DMSO and Gas Phase and Its Dependence on Brønsted Acidity. *J. Phys. Chem. A* **2017**, *121*, 357–369.
6. Selberg, S.; Rodima, T.; Lõkov, M.; Tshepelevitsh, S.; Haljasorg, T.; Chhabra, S.; Kadam, S. A.; Toom, L.; Vahur, S.; Leito, I. Synthesis and properties of highly lipophilic phosphazene bases. *Tetrahedron Lett.* **2017**, *58*, 2098–2102.
7. Teearu, A.; Vahur, S.; Rodima, T.; Herodes, K.; Bonrath, W.; Netscher, T.; Tshepelevitsh, S.; Trummal, A.; Lõkov, M.; Leito, I. Method development for the analysis of resinous materials with MALDI-FT-ICR-MS: novel internal standards and a new matrix material for negative ion mode. *J. Mass Spectrom.* **2017**, *52*, 603–617.

8. Lõkov, M.; Tshepelevitsh, S.; Heering, A.; Plieger, P. G.; Vianello, R.; Leito, I. On the Basicity of Conjugated Nitrogen Heterocycles in Different Media. *Eur. J. Org. Chem.* **2017**, *30*, 4475–4489.
9. Tshepelevitsh, S.; Hernits, K.; Jenčo, J.; Hawkins, J. M.; Muteki, K.; Solich, P.; Leito, I. Systematic Optimization of Liquid-Liquid Extraction for Isolation of Unidentified Components. *ACS Omega* **2017**, *2*, 7772–7776.
10. Pezzetta, C.; Tshepelevitsh, S.; Leito, I.; Poli, G. Comment on “Zemplén transesterification: a name reaction that has misled us for 90 years” by B. Ren, M. Wang, J. Liu, J. Ge, X. Zhang and H. Dong, *Green Chemistry*, 2015, *17*, 1390–1394. *Green Chem.* **2018**, *20*, 2392–2394.
11. Tshepelevitsh, S.; Hernits, K.; Leito, I. Prediction of partition and distribution coefficients in various solvent pairs with COSMO-RS. *J. Comput. Aided Mol. Des.* **2018**, DOI: 10.1007/s10822-018-0125-y.

Teaching work:

Supervision

Kertu Hernits, Experimental determination of distribution coefficients of lignin decomposition products and their modeling with COSMO-RS method, Bachelor’s Degree, 2017.

ELULOOKIRJELDUS

Nimi: Sofja Tšepelevitš
Sünniaeg: 30. detsember 1989
Kodakondsus: Eesti
Kontakt: Tartu Ülikool, Keemia instituut
Ravila 14a, 50411, Tartu, Eesti
E-mail: sofja.ts@gmail.com

Haridus:
2014–... Tartu Ülikool, keemia eriala doktorant
2012–2014 Tartu Ülikool, Keemia instituut, magistriõpe (keemia)
2009–2012 Tartu Ülikool, Keemia instituut, bakalaureuseõpe (keemia)

Töökogemus:
2012–... Tartu Ülikool, Keemia instituut, keemik

Teaduspublikatsioonid:

1. Tshepelevitsh, S.; Oss, M.; Pung, A.; Leito, I. Evaluating the COSMO-RS Method for Modeling Hydrogen Bonding in Solution. *ChemPhysChem* **2013**, *14*, 1909–1919.
2. Briocche, J.; Pike, S. J.; Tshepelevitsh, S.; Leito, I.; Morris, G. A.; Webb, S. J.; Clayden, J. Conformational Switching of a Foldamer in a Multi-component System by pH-Filtered Selection between Competing Non-covalent Interactions. *J. Am. Chem. Soc.* **2015**, *137*, 6680–6691.
3. Leito, I.; Koppel, I. A.; Koppel, I.; Kaupmees, K.; Tshepelevitsh, S.; Saame, J. Basicity Limits of Neutral Organic Superbases. *Angew. Chem., Int. Ed.* **2015**, *54*, 9262–9265.
4. Saame, J.; Rodima, T.; Tshepelevitsh, S.; Kütt, A.; Kaljurand, I.; Haljasorg, T.; Koppel, I. A.; Leito, I. Experimental Basicities of Superbasic Phosphonium Ylides and Phosphazenes. *J. Org. Chem.* **2016**, *81*, 7349–7361.
5. Tshepelevitsh, S.; Trummal, A.; Haav, K.; Martin, K.; Leito, I. Hydrogen-Bond Donicity in DMSO and Gas Phase and Its Dependence on Brønsted Acidity. *J. Phys. Chem. A* **2017**, *121*, 357–369.
6. Selberg, S.; Rodima, T.; Lõkov, M.; Tshepelevitsh, S.; Haljasorg, T.; Chhabra, S.; Kadam, S. A.; Toom, L.; Vahur, S.; Leito, I. Synthesis and properties of highly lipophilic phosphazene bases. *Tetrahedron Lett.* **2017**, *58*, 2098–2102.
7. Teearu, A.; Vahur, S.; Rodima, T.; Herodes, K.; Bonrath, W.; Netscher, T.; Tshepelevitsh, S.; Trummal, A.; Lõkov, M.; Leito, I. Method development for the analysis of resinous materials with MALDI-FT-ICR-MS: novel internal standards and a new matrix material for negative ion mode. *J. Mass Spectrom.* **2017**, *52*, 603–617.

8. Lõkov, M.; Tshepelevitsh, S.; Heering, A.; Plieger, P. G.; Vianello, R.; Leito, I. On the Basicity of Conjugated Nitrogen Heterocycles in Different Media. *Eur. J. Org. Chem.* **2017**, *30*, 4475–4489.
9. Tshepelevitsh, S.; Hernits, K.; Jenčo, J.; Hawkins, J. M.; Muteki, K.; Solich, P.; Leito, I. Systematic Optimization of Liquid-Liquid Extraction for Isolation of Unidentified Components. *ACS Omega* **2017**, *2*, 7772–7776.
10. Pezzetta, C.; Tshepelevitsh, S.; Leito, I.; Poli, G. Comment on “Zemplén transesterification: a name reaction that has misled us for 90 years” by B. Ren, M. Wang, J. Liu, J. Ge, X. Zhang and H. Dong, *Green Chemistry*, 2015, *17*, 1390-1394. *Green Chem.* **2018**, *20*, 2392–2394.
11. Tshepelevitsh, S.; Hernits, K.; Leito, I. Prediction of partition and distribution coefficients in various solvent pairs with COSMO-RS. *J. Comput. Aided Mol. Des.* **2018**, DOI: 10.1007/s10822-018-0125-y.

Õppetöö:

Juhendamine

Kertu Hernits, Ligniini laguproduktide jaotuskoeffitsientide eksperimentaalne määramine ja modelleerimine COSMO-RS meetodiga, bakalaureusekraad, kaitstud 2017.

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 ^1H 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 Rinke.** 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 P_2Y receptors. Tartu, 2000, 106 p.
18. **Vello Pällin.** The role of solvation in the formation of iotritch 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*-nucleophiles. 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 Teinmaa.** 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 dATP α S interaction with P2Y₁ 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 Thomborg.** The kinetics of electroreduction of peroxodisulfate anion on cadmium (0001) single crystal electrode. Tartu, 2005, 95 p.
48. **Olavi Loog.** Aspects of condensations of carbonyl compounds and their imine analogues. Tartu, 2005, 83 p.
49. **Siim Salmar.** Effect of ultrasound on ester hydrolysis in aqueous ethanol. Tartu, 2006, 73 p.
50. **Ain Uustare.** Modulation of signal transduction of heptahelical receptors by other receptors and G proteins. Tartu, 2006, 121 p.
51. **Sergei Yurchenko.** Determination of some carcinogenic contaminants in food. Tartu, 2006, 143 p.
52. **Kaido Tämm.** QSPR modeling of some properties of organic compounds. Tartu, 2006, 67 p.
53. **Olga Tšubrik.** New methods in the synthesis of multisubstituted hydrazines. Tartu, 2006, 183 p.
54. **Lilli Sooväli.** Spectrophotometric measurements and their uncertainty in chemical analysis and dissociation constant measurements. Tartu, 2006, 125 p.
55. **Eve Koort.** Uncertainty estimation of potentiometrically measured pH and pK_a values. Tartu, 2006, 139 p.
56. **Sergei Kopanchuk.** Regulation of ligand binding to melanocortin receptor subtypes. Tartu, 2006, 119 p.
57. **Silvar Kallip.** Surface structure of some bismuth and antimony single crystal electrodes. Tartu, 2006, 107 p.
58. **Kristjan Saal.** Surface silanization and its application in biomolecule coupling. Tartu, 2006, 77 p.
59. **Tanel Tätte.** High viscosity Sn(OBu)₄ oligomeric concentrates and their applications in technology. Tartu, 2006, 91 p.
60. **Dimitar Atanasov Dobchev.** Robust QSAR methods for the prediction of properties from molecular structure. Tartu, 2006, 118 p.
61. **Hannes Hagu.** Impact of ultrasound on hydrophobic interactions in solutions. Tartu, 2007, 81 p.
62. **Rutha Jäger.** Electroreduction of peroxodisulfate anion on bismuth electrodes. Tartu, 2007, 142 p.

63. **Kaido Viht.** Immobilizable bisubstrate-analogue inhibitors of basophilic protein kinases: development and application in biosensors. Tartu, 2007, 88 p.
64. **Eva-Ingrid Rõõm.** Acid-base equilibria in nonpolar media. Tartu, 2007, 156 p.
65. **Sven Tamp.** DFT study of the cesium cation containing complexes relevant to the cesium cation binding by the humic acids. Tartu, 2007, 102 p.
66. **Jaak Nerut.** Electroreduction of hexacyanoferrate(III) anion on Cadmium (0001) single crystal electrode. Tartu, 2007, 180 p.
67. **Lauri Jalukse.** Measurement uncertainty estimation in amperometric dissolved oxygen concentration measurement. Tartu, 2007, 112 p.
68. **Aime Lust.** Charge state of dopants and ordered clusters formation in CaF₂:Mn and CaF₂:Eu luminophors. Tartu, 2007, 100 p.
69. **Iiris Kahn.** Quantitative Structure-Activity Relationships of environmentally relevant properties. Tartu, 2007, 98 p.
70. **Mari Reinik.** Nitrates, nitrites, N-nitrosamines and polycyclic aromatic hydrocarbons in food: analytical methods, occurrence and dietary intake. Tartu, 2007, 172 p.
71. **Heili Kasuk.** Thermodynamic parameters and adsorption kinetics of organic compounds forming the compact adsorption layer at Bi single crystal electrodes. Tartu, 2007, 212 p.
72. **Erki Enkvist.** Synthesis of adenosine-peptide conjugates for biological applications. Tartu, 2007, 114 p.
73. **Svetoslav Hristov Slavov.** Biomedical applications of the QSAR approach. Tartu, 2007, 146 p.
74. **Eneli Härk.** Electroreduction of complex cations on electrochemically polished Bi(*hkl*) single crystal electrodes. Tartu, 2008, 158 p.
75. **Priit Möller.** Electrochemical characteristics of some cathodes for medium temperature solid oxide fuel cells, synthesized by solid state reaction technique. Tartu, 2008, 90 p.
76. **Signe Viggor.** Impact of biochemical parameters of genetically different pseudomonads at the degradation of phenolic compounds. Tartu, 2008, 122 p.
77. **Ave Sarapuu.** Electrochemical reduction of oxygen on quinone-modified carbon electrodes and on thin films of platinum and gold. Tartu, 2008, 134 p.
78. **Agnes Kütt.** Studies of acid-base equilibria in non-aqueous media. Tartu, 2008, 198 p.
79. **Rouvim Kadis.** Evaluation of measurement uncertainty in analytical chemistry: related concepts and some points of misinterpretation. Tartu, 2008, 118 p.
80. **Valter Reedo.** Elaboration of IVB group metal oxide structures and their possible applications. Tartu, 2008, 98 p.
81. **Aleksei Kuznetsov.** Allosteric effects in reactions catalyzed by the cAMP-dependent protein kinase catalytic subunit. Tartu, 2009, 133 p.

82. **Aleksei Bredihhin.** Use of mono- and polyanions in the synthesis of multisubstituted hydrazine derivatives. Tartu, 2009, 105 p.
83. **Anu Ploom.** Quantitative structure-reactivity analysis in organosilicon chemistry. Tartu, 2009, 99 p.
84. **Argo Vonk.** Determination of adenosine A_{2A}- and dopamine D₁ receptor-specific modulation of adenylate cyclase activity in rat striatum. Tartu, 2009, 129 p.
85. **Indrek Kivi.** Synthesis and electrochemical characterization of porous cathode materials for intermediate temperature solid oxide fuel cells. Tartu, 2009, 177 p.
86. **Jaanus Eskusson.** Synthesis and characterisation of diamond-like carbon thin films prepared by pulsed laser deposition method. Tartu, 2009, 117 p.
87. **Marko Lätt.** Carbide derived microporous carbon and electrical double layer capacitors. Tartu, 2009, 107 p.
88. **Vladimir Stepanov.** Slow conformational changes in dopamine transporter interaction with its ligands. Tartu, 2009, 103 p.
89. **Aleksander Trummal.** Computational Study of Structural and Solvent Effects on Acidities of Some Brønsted Acids. Tartu, 2009, 103 p.
90. **Eerold Vellemäe.** Applications of mischmetal in organic synthesis. Tartu, 2009, 93 p.
91. **Sven Parkel.** Ligand binding to 5-HT_{1A} receptors and its regulation by Mg²⁺ and Mn²⁺. Tartu, 2010, 99 p.
92. **Signe Vahur.** Expanding the possibilities of ATR-FT-IR spectroscopy in determination of inorganic pigments. Tartu, 2010, 184 p.
93. **Tavo Romann.** Preparation and surface modification of bismuth thin film, porous, and microelectrodes. Tartu, 2010, 155 p.
94. **Nadežda Aleksejeva.** Electrocatalytic reduction of oxygen on carbon nanotube-based nanocomposite materials. Tartu, 2010, 147 p.
95. **Marko Kullapere.** Electrochemical properties of glassy carbon, nickel and gold electrodes modified with aryl groups. Tartu, 2010, 233 p.
96. **Liis Siinor.** Adsorption kinetics of ions at Bi single crystal planes from aqueous electrolyte solutions and room-temperature ionic liquids. Tartu, 2010, 101 p.
97. **Angela Vaasa.** Development of fluorescence-based kinetic and binding assays for characterization of protein kinases and their inhibitors. Tartu 2010, 101 p.
98. **Indrek Tulp.** Multivariate analysis of chemical and biological properties. Tartu 2010, 105 p.
99. **Aare Selberg.** Evaluation of environmental quality in Northern Estonia by the analysis of leachate. Tartu 2010, 117 p.
100. **Darja Lavõgina.** Development of protein kinase inhibitors based on adenosine analogue-oligoarginine conjugates. Tartu 2010, 248 p.
101. **Laura Herm.** Biochemistry of dopamine D₂ receptors and its association with motivated behaviour. Tartu 2010, 156 p.

102. **Terje Raudsepp.** Influence of dopant anions on the electrochemical properties of polypyrrole films. Tartu 2010, 112 p.
103. **Margus Marandi.** Electroformation of Polypyrrole Films: *In-situ* AFM and STM Study. Tartu 2011, 116 p.
104. **Kairi Kivirand.** Diamine oxidase-based biosensors: construction and working principles. Tartu, 2011, 140 p.
105. **Anneli Kruve.** Matrix effects in liquid-chromatography electrospray mass-spectrometry. Tartu, 2011, 156 p.
106. **Gary Urb.** Assessment of environmental impact of oil shale fly ash from PF and CFB combustion. Tartu, 2011, 108 p.
107. **Nikita Oskolkov.** A novel strategy for peptide-mediated cellular delivery and induction of endosomal escape. Tartu, 2011, 106 p.
108. **Dana Martin.** The QSPR/QSAR approach for the prediction of properties of fullerene derivatives. Tartu, 2011, 98 p.
109. **Säde Viirlaid.** Novel glutathione analogues and their antioxidant activity. Tartu, 2011, 106 p.
110. **Ülis Sõukand.** Simultaneous adsorption of Cd²⁺, Ni²⁺, and Pb²⁺ on peat. Tartu, 2011, 124 p.
111. **Lauri Lipping.** The acidity of strong and superstrong Brønsted acids, an outreach for the “limits of growth”: a quantum chemical study. Tartu, 2011, 124 p.
112. **Heisi Kurig.** Electrical double-layer capacitors based on ionic liquids as electrolytes. Tartu, 2011, 146 p.
113. **Marje Kasari.** Bisubstrate luminescent probes, optical sensors and affinity adsorbents for measurement of active protein kinases in biological samples. Tartu, 2012, 126 p.
114. **Kalev Takkis.** Virtual screening of chemical databases for bioactive molecules. Tartu, 2012, 122 p.
115. **Ksenija Kisseljova.** Synthesis of aza-β³-amino acid containing peptides and kinetic study of their phosphorylation by protein kinase A. Tartu, 2012, 104 p.
116. **Riin Rebane.** Advanced method development strategy for derivatization LC/ESI/MS. Tartu, 2012, 184 p.
117. **Vladislav Ivaništšev.** Double layer structure and adsorption kinetics of ions at metal electrodes in room temperature ionic liquids. Tartu, 2012, 128 p.
118. **Irja Helm.** High accuracy gravimetric Winkler method for determination of dissolved oxygen. Tartu, 2012, 139 p.
119. **Karin Kipper.** Fluoroalcohols as Components of LC-ESI-MS Eluents: Usage and Applications. Tartu, 2012, 164 p.
120. **Arno Ratas.** Energy storage and transfer in dosimetric luminescent materials. Tartu, 2012, 163 p.
121. **Reet Reinart-Okugbeni.** Assay systems for characterisation of subtype-selective binding and functional activity of ligands on dopamine receptors. Tartu, 2012, 159 p.

122. **Lauri Sikk.** Computational study of the Sonogashira cross-coupling reaction. Tartu, 2012, 81 p.
123. **Karita Raudkivi.** Neurochemical studies on inter-individual differences in affect-related behaviour of the laboratory rat. Tartu, 2012, 161 p.
124. **Indrek Saar.** Design of GalR2 subtype specific ligands: their role in depression-like behavior and feeding regulation. Tartu, 2013, 126 p.
125. **Ann Laheäär.** Electrochemical characterization of alkali metal salt based non-aqueous electrolytes for supercapacitors. Tartu, 2013, 127 p.
126. **Kerli Tõnurist.** Influence of electrospun separator materials properties on electrochemical performance of electrical double-layer capacitors. Tartu, 2013, 147 p.
127. **Kaija Põhako-Esko.** Novel organic and inorganic ionogels: preparation and characterization. Tartu, 2013, 124 p.
128. **Ivar Kruusenberg.** Electroreduction of oxygen on carbon nanomaterial-based catalysts. Tartu, 2013, 191 p.
129. **Sander Piiskop.** Kinetic effects of ultrasound in aqueous acetonitrile solutions. Tartu, 2013, 95 p.
130. **Ilona Faustova.** Regulatory role of L-type pyruvate kinase N-terminal domain. Tartu, 2013, 109 p.
131. **Kadi Tamm.** Synthesis and characterization of the micro-mesoporous anode materials and testing of the medium temperature solid oxide fuel cell single cells. Tartu, 2013, 138 p.
132. **Iva Bozhidarova Stoyanova-Slavova.** Validation of QSAR/QSPR for regulatory purposes. Tartu, 2013, 109 p.
133. **Vitali Grozovski.** Adsorption of organic molecules at single crystal electrodes studied by *in situ* STM method. Tartu, 2014, 146 p.
134. **Santa Veikšina.** Development of assay systems for characterisation of ligand binding properties to melanocortin 4 receptors. Tartu, 2014, 151 p.
135. **Jüri Liiv.** PVDF (polyvinylidene difluoride) as material for active element of twisting-ball displays. Tartu, 2014, 111 p.
136. **Kersti Vaarmets.** Electrochemical and physical characterization of pristine and activated molybdenum carbide-derived carbon electrodes for the oxygen electroreduction reaction. Tartu, 2014, 131 p.
137. **Lauri Tõntson.** Regulation of G-protein subtypes by receptors, guanine nucleotides and Mn²⁺. Tartu, 2014, 105 p.
138. **Aiko Adamson.** Properties of amine-boranes and phosphorus analogues in the gas phase. Tartu, 2014, 78 p.
139. **Elo Kibena.** Electrochemical grafting of glassy carbon, gold, highly oriented pyrolytic graphite and chemical vapour deposition-grown graphene electrodes by diazonium reduction method. Tartu, 2014, 184 p.
140. **Teemu Näykki.** Novel Tools for Water Quality Monitoring – From Field to Laboratory. Tartu, 2014, 202 p.
141. **Karl Kaupmees.** Acidity and basicity in non-aqueous media: importance of solvent properties and purity. Tartu, 2014, 128 p.

142. **Oleg Lebedev.** Hydrazine polyanions: different strategies in the synthesis of heterocycles. Tartu, 2015, 118 p.
143. **Geven Piir.** Environmental risk assessment of chemicals using QSAR methods. Tartu, 2015, 123 p.
144. **Olga Mazina.** Development and application of the biosensor assay for measurements of cyclic adenosine monophosphate in studies of G protein-coupled receptor signaling. Tartu, 2015, 116 p.
145. **Sandip Ashokrao Kadam.** Anion receptors: synthesis and accurate binding measurements. Tartu, 2015, 116 p.
146. **Indrek Tallo.** Synthesis and characterization of new micro-mesoporous carbide derived carbon materials for high energy and power density electrical double layer capacitors. Tartu, 2015, 148 p.
147. **Heiki Erikson.** Electrochemical reduction of oxygen on nanostructured palladium and gold catalysts. Tartu, 2015, 204 p.
148. **Erik Anderson.** *In situ* Scanning Tunnelling Microscopy studies of the interfacial structure between Bi(111) electrode and a room temperature ionic liquid. Tartu, 2015, 118 p.
149. **Girinath G. Pillai.** Computational Modelling of Diverse Chemical, Biochemical and Biomedical Properties. Tartu, 2015, 140 p.
150. **Piret Pikma.** Interfacial structure and adsorption of organic compounds at Cd(0001) and Sb(111) electrodes from ionic liquid and aqueous electrolytes: an *in situ* STM study. Tartu, 2015, 126 p.
151. **Ganesh babu Manoharan.** Combining chemical and genetic approaches for photoluminescence assays of protein kinases. Tartu, 2016, 126 p.
152. **Carolin Siimenson.** Electrochemical characterization of halide ion adsorption from liquid mixtures at Bi(111) and pyrolytic graphite electrode surface. Tartu, 2016, 110 p.
153. **Asko Laaniste.** Comparison and optimisation of novel mass spectrometry ionisation sources. Tartu, 2016, 156 p.
154. **Hanno Evard.** Estimating limit of detection for mass spectrometric analysis methods. Tartu, 2016, 224 p.
155. **Kadri Ligi.** Characterization and application of protein kinase-responsive organic probes with triplet-singlet energy transfer. Tartu, 2016, 122 p.
156. **Margarita Kagan.** Biosensing penicillins' residues in milk flows. Tartu, 2016, 130 p.
157. **Marie Kriisa.** Development of protein kinase-responsive photoluminescent probes and cellular regulators of protein phosphorylation. Tartu, 2016, 106 p.
158. **Mihkel Vestli.** Ultrasonic spray pyrolysis deposited electrolyte layers for intermediate temperature solid oxide fuel cells. Tartu, 2016, 156 p.
159. **Silver Sepp.** Influence of porosity of the carbide-derived carbon on the properties of the composite electrocatalysts and characteristics of polymer electrolyte fuel cells. Tartu, 2016, 137p.
160. **Kristjan Haav.** Quantitative relative equilibrium constant measurements in supramolecular chemistry. Tartu, 2017, 158 p.

161. **Anu Teearu.** Development of MALDI-FT-ICR-MS methodology for the analysis of resinous materials. Tartu, 2017, 205 p.
162. **Taavi Ivan.** Bifunctional inhibitors and photoluminescent probes for studies on protein complexes. Tartu, 2017, 140 p.
163. **Maarja-Liisa Oldekop.** Characterization of amino acid derivatization reagents for LC-MS analysis. Tartu, 2017, 147 p.
164. **Kristel Jukk.** Electrochemical reduction of oxygen on platinum- and palladium-based nanocatalysts. Tartu, 2017, 250 p.
165. **Siim Kukk.** Kinetic aspects of interaction between dopamine transporter and *N*-substituted nortropine derivatives. Tartu, 2017, 107 p.
166. **Birgit Viira.** Design and modelling in early drug development in targeting HIV-1 reverse transcriptase and Malaria. Tartu, 2017, 172 p.
167. **Rait Kivi.** Allostery in cAMP dependent protein kinase catalytic subunit. Tartu, 2017, 115 p.
168. **Agnes Heering.** Experimental realization and applications of the unified acidity scale. Tartu, 2017, 123 p.
169. **Delia Juronen.** Biosensing system for the rapid multiplex detection of mastitis-causing pathogens in milk. Tartu, 2018, 85 p.
170. **Hedi Rahnel.** ARC-inhibitors: from reliable biochemical assays to regulators of physiology of cells. Tartu, 2018, 176 p.
171. **Anton Ruzanov.** Computational investigation of the electrical double layer at metal–aqueous solution and metal–ionic liquid interfaces. Tartu, 2018, 129 p.
172. **Katrin Kestav.** Crystal Structure-Guided Development of Bisubstrate-Analogue Inhibitors of Mitotic Protein Kinase Haspin. Tartu, 2018, 166 p.
173. **Mihkel Ilisson.** Synthesis of novel heterocyclic hydrazine derivatives and their conjugates. Tartu, 2018, 101 p.
174. **Anni Allikalt.** Development of assay systems for studying ligand binding to dopamine receptors. Tartu, 2018, 160 p.
175. **Ove Oll.** Electrical double layer structure and energy storage characteristics of ionic liquid based capacitors. Tartu, 2018, 187 p.
176. **Rasmus Palm.** Carbon materials for energy storage applications. Tartu, 2018, 114 p.
177. **Jörgen Metsik.** Preparation and stability of poly(3,4-ethylenedioxythiophene) thin films for transparent electrode applications. Tartu, 2018, 111 p.