DISSERTATIONES
CHIMICAE
UNIVERSITATIS
TARTUENSIS
242

KARL MARTI TOOTS

Cheminformatics Approaches for Analysing and Modelling the Gas-Ionic Liquid Distribution of Organic Solutes

DISSERTATIONES CHIMICAE UNIVERSITATIS TARTUENSIS
242

DISSERTATIONES CHIMICAE UNIVERSITATIS TARTUENSIS
242

KARL MARTI TOOTS

Cheminformatics Approaches for Analysing and Modelling the Gas-Ionic Liquid Distribution of Organic Solutes

UNIVERSITY OF TARTU
Press
1632

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

Dissertation has been accepted for the commencement of the degree of *Doctor philosophiae* in Molecular Technology on June 10, 2025, by the Council of Institute of Chemistry, Faculty of Science and Technology, University of Tartu.

Supervisors: Professor Uko Maran, PhD
Institute of Chemistry
University of Tartu, Estonia

Associate professor Sulev Sild, PhD
Institute of Chemistry
University of Tartu, Estonia

Associate professor Jaan Leis, PhD
Institute of Chemistry
University of Tartu, Estonia

Opponent: Dr. Habil. Igor Tetko,
Helmholtz Zentrum München,
Molecular Targets and Therapeutics Center,
Institute of Structural Biology
Germany

Public defense: 5. September 2025 at 10:00 in Ravila 14a, auditorium 1020

Author is grateful for support from the Ministry of Education and Research, Republic of Estonia, through the Estonian Research Council (grant number PRG1509). Author is also grateful to the University of Tartu for the opportunity to complete this doctoral research and to the supervisors, Uko Maran, Sulev Sild, and Jaan Leis, for their invaluable guidance and support. Publication of this thesis is financed by the Institute of Chemistry, University of Tartu.

ISSN 1406-0299 (print) ISSN 2806-2159 (pdf)
ISBN 978-9916-27-968-7 (print) ISBN 978-9916-27-969-4 (pdf)

University of Tartu Press
www.tyk.ee

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

The present thesis consists of four articles with reference by Roman numerals I–IV.

I. Toots, K. M.; Sild, S.; Leis, J.; Acree Jr., W. E.; Maran, U. The Quantitative Structure-Property Relationships for the Gas-Ionic Liquid Partition Coefficient of a Large Variety of Organic Compounds in Three Ionic Liquids. *J. Mol. Liq.* **2021**, *343*, 117573. https://doi.org/10.1016/j.molliq.2021.117573.

II. Toots, K. M.; Sild, S.; Leis, J.; Acree, W. E.; Maran, U. Machine Learning Quantitative Structure-Property Relationships as a Function of Ionic Liquid Cations for the Gas-Ionic Liquid Partition Coefficient of Hydrocarbons. *Int. J. Mol. Sci.* **2022**, *23* (14), 7534. https://doi.org/10.3390/ijms23147534.

III. Toots, K. M.; Sild, S.; Leis, J.; Acree, W. E.; Maran, U. Exploring the Influence of Ionic Liquid Anion Structure on Gas-Ionic Liquid Partition Coefficients of Organic Solutes Using Machine Learning. *Langmuir* **2024**, *40* (45), 23714–23728. https://doi.org/10.1021/acs.langmuir.4c02628.

IV. Toots, K. M.; Sild, S.; Leis, J.; Acree Jr., W. E.; Maran, U. A Multi-Component QSPR Approach to Describe and Predict the Gas-Ionic Liquid Distribution of Organic Solutes using Machine Learning. *J. Mol. Liq*. **2025**, *436*, 128184. https://doi.org/10.1016/j.molliq.2025.128184

LIST OF ABBREVIATIONS

AD	Applicability Domain
AI	Artificial Intelligence
CV	Cross-Validation
FAIR	Findability, Accessibility, Interoperability, and Reusability
GLC	Gas-Liquid Chromatography
GPR	Gaussian Process Regression
HB	Hydrogen Bonding
IL	Ionic Liquid
Log K	Logarithm of the Gas–Ionic Liquid Partition Coefficient
LOOCV	Leave-One-Out Cross-Validation
ML	Machine Learning
MLR	Multiple Linear Regression
OLS	Ordinary Least Squares
OMP	Orthogonal Matching Pursuit
QSAR	Quantitative Structure–Activity Relationship
QSPR	Quantitative Structure–Property Relationship
RBF	Radial Basis Function
RF	Random Forest
RMSE	Root Mean Square Error
R^2	Coefficient of Determination
SVR	Support Vector Regression

INTRODUCTION

Understanding the physical and chemical foundation of the world is driving exponential technological progress in the modern human era. In this technological progress, computers are multiplying human creativity and discovery capabilities at an accelerating pace. Thus, *in silico* approaches, such as qualitative and quantitative structure-property relationships (QSPRs) help to accumulate knowledge by harnessing the processing power of computers. In this regard, the widespread application of artificial intelligence and machine learning methods within the frame of QSPR approach enables modern research in understanding the physicochemical properties of materials and substances. One such class of substances is ionic liquids (ILs) defined as organic salts, i.e., liquids at ambient conditions. Understanding and predicting the distribution properties of solutes in the IL provides a basis for the exploration and development of such application-oriented chemical media.

As potential environmentally safer alternatives to traditional volatile organic solvents, ILs are an important subject of research. Their versatility has led to numerous applications as solvents and extraction media in synthesis, catalysis, electrochemistry, separation and extraction chemistry, and material and biochemical engineering. Despite the unique, and tunable properties of ILs, comparatively few studies have been conducted on their structure-property relationships. Systematic studies have generated experimental data necessary for the development of large-scale generic data-driven models that can be used to provide a better mechanistic understanding of chemical interactions in ionic liquid properties of interest. For example, the large and diverse data set covering the complex interaction system of organic solutes, cations, and anions in ILs requires the development of novel methodological solutions.

The aim of research conducted in this thesis was to investigate the relationships between the gas-ionic liquid partition coefficient (log K) and the structure of organic solutes and the ionic components of ILs using cheminformatics approaches, in order to develop efficient models for predicting log K of chemically multicomponent systems, and to describe the possible interactions occurring in such molecular systems. This involves use of theoretical molecular descriptors and advanced machine learning methods to model and uncover the interaction mechanisms based on the structure of the solute and the IL that govern the partitioning behaviour of the solute depending on the structural features of each component. The hypotheses of this research are (i) that the structural properties of both anions and cations of the ionic liquid significantly affect gas–IL partitioning of organic solutes; (ii) that integrated multicomponent models that include molecular descriptors of solutes, cations, and anions can provide competitive predictive accuracy for log K; (iii) and finally, that both linear and non-linear advanced machine learning approaches through molecular descriptors successfully shed light on the intermolecular forces underlying their distribution behaviour, such as dispersion, Coulomb and dipolar interactions, and hydrogen bonding.

CHAPTER 1. LITERATURE REVIEW

1.1 Ionic Liquids and Applications

Ionic liquids are a unique class of compounds generally defined as organic salts with a low melting point, many of which retain liquid state at room temperature.[1] They are composed entirely of ions, typically organic quaternary ammonium, sulfonium, phosphonium, pyrrolidinium, imidazolium, or pyridinium cations of low charge density, combined with diffusely charged, bulky, water-stable anions, which hinder packing in the crystal lattice because of the mismatched ion sizes and shapes.[2] These structural intricacies, suppressing the strong electrostatic Coulombic attraction between the counterions, are the reason for low melting points and negligible vapor pressures of ILs.[2] In addition, ILs display low viscosity, wide liquid range, high polarity, good electrical conductance, and high thermal stability, setting them apart from molecular solvents and motivating their use in a wide range of applications across various fields.[2,3]

ILs are often sub-categorized by their chemical structure or functionality.[4] Based on structure, they can be classified as protic, chiral, magnetic, polymeric, polyionic, chelating, fluorous, oxidative, monovalent, and dicationic.[5] As discussed by Singh and Savoy (2020)[6], their purity depends to a significant extent on their synthesis strategy, and a low impact on the environment is also important, as in the case of multi-step synthesis methods based on microwave irradiation and ultrasound-assisted reactions. High purity ionic liquids are required for reliable results in scientific research, but depending on the ionic liquid, they can be difficult to purify. The extremely low vapor pressure inherent to ILs complicates the purification through common distillation methods. Nevertheless, most impurities in ionic liquids can still be removed by distillation and the degree of purity can be increased by slow crystallization.[6]

ILs have been widely studied as solvents and electrolytes in catalysis[7–11], electrochemistry[12–17], extraction and separation chemistry[18–21], synthesis[22–27], and numerous other industrial and research applications.[28–32] They are increasingly used as an alternative to volatile organic solvents, mineral acids, bases, solid acids and many other compounds.[6] The main reason here is the very low vapor pressure of ionic liquids. The potential environmental benefit is an important driving force in the research and application of ILs.[33–35] Another remarkable property is their adaptability; the properties of ILs can be fine-tuned by slightly altering their cation or anion components, which has led to the concept of "designer solvents", where ILs can be designed for a specific application.[36] However, many commonly used ILs exhibit some degree of toxicity, as demonstrated by toxicological data collected from evaluating their effects on various organisms.[37] Despite this circumstance, being designer solvents, ILs have opened a window to design safer, biodegradable ILs with properties similar to those of the commonly used solvents enhanced by the low volatility, high thermal and chemical stability, and low flammability of ILs.[37] Beyond these advantages, ILs are good solvents for both organic and inorganic substances, which is why they can be used, for example, to bring different reactants into the same phase.[10,38] Their good dissolving properties

are also used in various syntheses, materials processing, biomass processing, extraction and gas separation.[33] Electrochemical stability in addition to unique solvent properties provides a good basis for the use of ILs for the separation of rare earth metal salts and the electrochemical deposition of pure earth metal.[39] Low melting point and good solubility allow ILs to be used as chemical enhancers in drugs to ensure effective drug delivery.[40] Finally, the ability of ILs to solubilize and stabilize enzymes, proteins, DNA, and RNA is an exceptionally valuable property for biotechnological applications.[33]

As ILs continue to gain increasing popularity in various applications, the need to understand and predict their properties becomes increasingly important. In the last two decades, the number of scientific reports on ILs research has increased exponentially, and the continuous development and synthesis of new ILs necessitate efficient methods for evaluating their suitability for specific purposes.[41] The amount of physicochemical characteristics data of ILs such as density, melting point, thermal stability, conductivity, solubility, viscosity, surface tension, polarity, vapor pressure, and miscibility among others has seen a quick rise.[41]

1.2 Partition Coefficient

Partitioning is the distribution of molecules between two immiscible solvents.[42,43] This is quantitatively characterized by the partition coefficient, i.e., the ratio of compound concentrations in immiscible solvents in an equilibrium.[43] The partition coefficient is useful in estimating the distribution of chemicals in any system.[43] In natural environments, a hydrophobic chemical with a high octanol-water partition coefficient is distributed mainly in hydrophobic areas of the ecosystem.[43] In contrast, a hydrophilic chemical with a low octanol-water partition coefficient is found primarily in wet areas such as marshes, ponds, rivers, and lakes.[43] The partition coefficient is also used in the development of drugs to measure the hydrophobicity of the solute and, based on this, to determine the permeability of the drug through the membrane.[43] If one of the solvents is a gas and the other a liquid, the distribution of the solute between these phases is characterized by the gas-liquid partition coefficient.[43]

Gas-Ionic Liquid Partition Coefficient

The gas–ionic liquid partition coefficient (K) quantifies the distribution of a solute between a gas phase and an IL:[43]

$$K = \frac{c_{IL}}{c_{GAS}},$$

where c_{IL} is the concentration of the solute in the IL phase, and c_{GAS} is its concentration in the gas phase (illustrated in **Figure 1**).[43] The partition coefficient is often expressed in logarithmic form, log K, to facilitate data handling and analysis. K can be found by calculations from the results of isothermal GLC measurements.[44,45] The log K is found as the ratio of the volume of the carrier gas

required to elute the solute to the volume of the stationary liquid phase.[44] Experimental determination of log K can be labor-intensive, time-consuming, and costly, because the methods require significant amounts of pure compounds and extensive laboratory work, making them impractical for screening large numbers of solute–anion–cation combinations. Therefore, it is necessary to understand the effects of molecular (including ionic) structures and develop predictive models that can accurately and efficiently predict IL properties. Computational approaches, such as QSPRs, offer a viable solution by correlating molecular structures with their properties using mathematical models.[46,47]

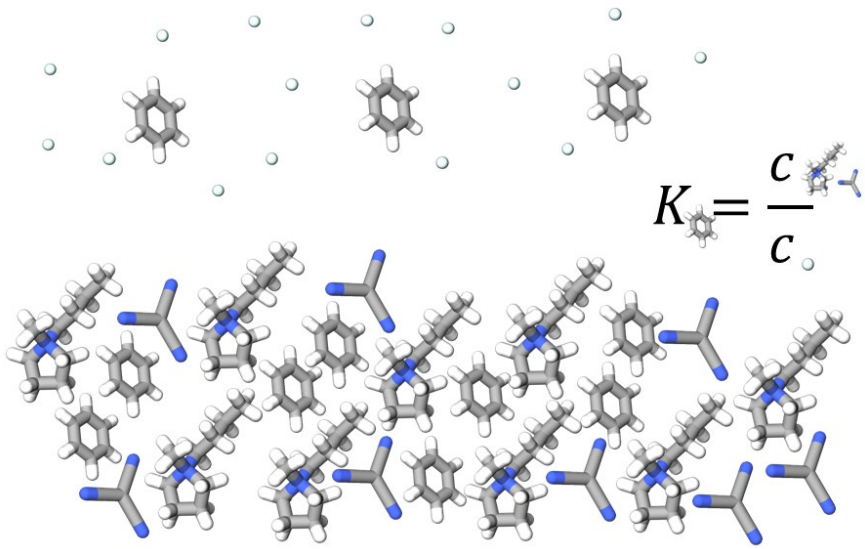

Figure 1. Solute molecules are distributed between a gas and an ionic liquid, for example benzene partitioned between 1-butyl-1-methylpyrrolidinium tricyanomethanide and carrier gas (e.g., helium).

1.3 Quantitative Structure–Property Relationships

QSPRs[48] represent a family of *in silico* modelling approaches used to correlate molecular structure with their physicochemical, biological, or other properties by means of mathematical models. The underlying assumption is that structurally similar compounds tend to exhibit similar behaviours or properties. Consequently, QSPR methods hold both descriptive and predictive power, enabling, for instance, the estimation of a compound's biological effect prior to experimental testing or the identification of structural features that most strongly influence a target property.[48,49]

In a typical QSPR workflow[48], data is gathered regarding the chemical or physical property of interest. The gathered data are then carefully curated[50] to eliminate poor-quality or erroneous measurements in ensuring data quality and consistency. Next, information about each compound's structure is transformed

into numerical parameters known as molecular descriptors, which can be derived experimentally or through computational methods based on molecular structure.[51] As described by Cronin[48], it is widely accepted that at least 5-10 compounds are required per descriptor in a QSPR. But a simpler model with a minimal number of descriptors per compound is preferred for transparency, interpretability, and mechanistic understanding. Before forming descriptor-property relationships, pre-processing is sometimes required to remove uninformative or redundant descriptors, because QSPR modelling may involve very large sets of descriptors. Following this step, the data set is split into a training set used to build or "train" the model and a test set used to validate the model's performance.[48]

A variety of algorithms, ranging from multiple linear regression to random forests, artificial neural networks, or support vector machines, may be employed to build a QSPR model.[49] During model building, "hyperparameters" that control the learning process of the chosen algorithm are also optimized, and the most relevant subset of descriptors is selected.[52] Finally, the trained model's accuracy is measured on the test set (sometimes via *n*-fold cross-validation) to ensure its robustness and avoid overfitting.[48,49] In many QSPR studies, additional external validation criteria are used to further confirm the model's predictive ability on data it has never encountered.[53–57]

1.3.1 Molecular Descriptor Selection

Molecular descriptors are parameters derived experimentally or calculated from chemical structure and serve as input variables in QSPR models.[49,58,59] These descriptors can be broadly classified by their dimensionality—0D (e.g., elemental composition), 1D (e.g., molecular formula), 2D (e.g., topological descriptors), and 3D (e.g., surface areas, volumes, quantum chemical descriptors)—or by the field or method used to derive them (e.g., topological, geometric, quantum-chemical, or thermodynamic).[58] The number of potential descriptors can be very large: commercial software such as DRAGON 6 provides up to 4885 descriptors, whereas open-source libraries (e.g., Mold2, Mordred) offer hundreds or thousands of descriptors.[58,60,61]

Including too many descriptors can harm the model in several ways.[56] Uninformative or constant descriptors add noise without offering predictive insight, while redundant descriptors are often correlated with each other, which can lead to multicollinearity and overfitting.[56] Finally, large descriptor sets can increase computational complexity and reduce interpretability.[49,56] Thus, descriptor selection methods aim to strike a balance between model accuracy and describability; keeping only enough descriptors to capture the essential variation in the data without compromising external predictivity or interpretability.[48,49]

Common selection strategies include forward selection, which sequentially adds the descriptor that provides the greatest improvement; backward elimination, which starts with all descriptors and prunes them iteratively; and embedded methods that perform descriptor selection as part of the modelling process (e.g., regularization-based approaches, genetic algorithms, or methods integrated with

machine learning models).[56] While linear and nonlinear techniques often select different descriptors due to their inherent assumptions about data structure, they share a common goal of maximizing a model's predictive performance while preserving interpretability.[49,56]

Among the various descriptor selection methods, Orthogonal Matching Pursuit (OMP) offers a practical solution for iteratively choosing descriptors that are minimally correlated with each other while strongly correlated with the target property.[62,63] OMP was originally developed in the signal processing domain but has proven useful in QSPR when a linear model is assumed.[62] The core idea behind OMP[63] is to start with an empty set of descriptors and select one descriptor at a time by looking at which candidate has the highest correlation with the current residual i.e., the difference between the experimental property values and the model's predicted values at each iteration. Once the most informative descriptor is found, it is included in a linear regression model, and the predictions are updated. The algorithm then computes a new residual by subtracting these updated predictions from the experimental values. The procedure repeats until a predefined number of descriptors (or another stopping criterion) is reached.[63]

OMP focuses on the largest residual correlation at each step[62] and thereby tends to pick descriptors that are (a) highly relevant to the property being modelled, and (b) less redundant with already chosen descriptors. Compared to forward or backward selection methods, OMP often converges faster and is relatively straightforward to implement. However, it still relies on the assumption of linearity and can miss important descriptors if the relationship with the target property is highly nonlinear. Nonetheless, OMP's computational simplicity and interpretability make it a solid choice for QSPR descriptor selection as a linear approach.[62,63]

1.3.2 Model Validation

Model validation is regarded as a crucial step to verify a model's predictiveness and to define the chemical space in which its predictions are reliable as per Tropsha et al.[53] Furthermore, internal validation strategies used in model training commonly adopt cross-validation (CV) methods, which split the dataset into multiple training and validation folds to check if the model maintains consistent performance when confronted with data not used in its development. For medium- to large-sized datasets, 5-fold or 10-fold CV is often chosen, whereas for smaller datasets typically leave-one-out (LOO) CV is used to include the maximal amount of the limited training data in each split. In external validation, a holdout subset of data is separated and entirely unseen during model construction, thus serving as a test of real-world applicability.[53]

Model performance is reported with various metrics and statistics such as the coefficient of determination r^2, the concordance correlation coefficient CCC[57], and the root mean squared error $RMSE$[53]. The coefficient of determination:[53]

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

quantifies the proportion of variance in the observed data explained by the model, where $y_i$denotes the experimental value, $\hat{y}_i$ the predicted value, and $\bar{y}_i$ the mean of the experimental values.[53] The *CCC* is used to measure both precision and accuracy relative to the line of identity and can be expressed as

$$CCC = \frac{2\sum(\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sum(\hat{y}_i - \bar{\hat{y}})^2 + \sum(y_i - \bar{y})^2 + n(\bar{\hat{y}} - \hat{y})^2},$$

where $\bar{\hat{y}}$ is the mean of predicted values.[57] Meanwhile, the RMSE:

$$RMSE = \sqrt{\frac{1}{n}\sum(y_i - \hat{y}_i)^2}.$$

offers a measure of expected prediction error.[53]

1.3.3 Applicability Domain

The model AD characterizes the space of chemical structures and descriptor values used in building the model where predictions are meaningful.[64] In linear models, leverage analysis derived from the hat matrix H is widely employed for identifying data points that exert a disproportionate influence on the regression parameters.[65] The leverage of each observation is calculated as

$$h_i = x_i^T (X^T X)^{-1} x_i,$$

where x_i is the descriptor vector of the i-th compound, and X is the design matrix of all selected descriptors typically augmented by a constant column for the linear model intercept. Observations with leverage beyond critical thresholds:[66]

$$h_2^* = \frac{2(k+1)}{n},$$
$$h_3^* = \frac{3(k+1)}{n},$$

calculated from the number of model parameters $k + 1$ and sample size n, can have an outsized impact on model coefficients and therefore require closer inspection. Standardized residuals r_i are used to detect data points with unusually large prediction errors:[66]

$$r_i = \frac{\hat{y}_i - y_i}{\hat{\sigma} \cdot \sqrt{1 - h_{ii}}},$$

where $\hat{\sigma}$ is the mean squared error of the linear regression model. Typical thresholds are $r_i > 2$ for medium and $r_i > 3$ to flag large residuals. Proportional to both standardized residuals and leverage values, Cook's distance:[67]

$$D_i = \frac{(y_i - y_{(i)})^2}{p \cdot MSE},$$

further quantifies how removing a given observation i influences the fitted model, with $y_{(i)}$ denoting the predicted value when observation i is omitted, p the number of predictors, and MSE the mean squared error. An influence plot or Cook's plot graphs standardized residuals against leverage values where data point sizes are often scaled proportional to Cook's distances.[67] The Cook's plot provides visualization of influential data points with the most influential outliers typically largest on the plot and positioned furthest from the origin point. Data points with values exceeding 0.5 or 1.0 Cook's distance are considered influential and should be examined in detail to determine whether specific chemical features or measurement issues might have led to their disproportionate impact.[67]

Non-linear models typically require alternative approaches to analyse their AD. One such measure is the distance in the descriptor space, Mahalanobis distance:[68]

$$D_i^2 = (x_i - \mu)^T \Sigma^{-1} (x_i - \mu),$$

where x_i is the descriptor vector of the compound in question, μ is the mean descriptor vector of the training set, and Σ is the empirical covariance matrix. Points that exhibit both a large Mahalanobis distance and a high standardized residual are considered likely to fall outside the reliable prediction region of the model.[68] Other methods for non-linear model AD analysis include k-nearest neighbours and kernel density estimation.[69–71]

1.4 Machine Learning Methods

Machine learning is widely considered a set of modelling methods part of artificial intelligence, but it has grown out into a discipline on its own[72], where computer algorithms take some data and parameters as input, and the algorithm learns from the data in a process called "training" with the goal of generalizing to perform tasks on unseen data.[73] Machine learning methods applied to labelled data is called supervised learning and sub-categorized as classification and regression.[74] In QSPR regression analysis, the goal is to find the relationship between molecular descriptors and the property of interest.[75] The choice of algorithm and molecular descriptor set is often grounded on the quantity and quality of available data, as well as the model's desired predictive accuracy and explainability.[75]

1.4.1 Multiple Linear Regression

MLR is a linear approach to relate log *K* to a linear combination of descriptors:[76]

$$\hat{y} = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$$

In the MLR equation, β_0, β_1, …, β_k represent the regression coefficients, while X_1, X_2, …, X_k denote the selected molecular descriptors.[76]

The OLS method is regularly used to minimize the residual sum of squares between the predictions and target experimental log *K* values:[73]

$$\hat{\beta} = (X^T X)^{-1} X^T y,$$

where X is the matrix of descriptor values and y is the vector of log *K* values. Other methods such as weighted least squares and robust regression offer alternatives to handle specific challenges, such as heteroscedasticity or the presence of outliers, respectively.[77] OLS results are easily interpretable, making it a preferred choice for studies emphasizing explainability.[76]

Regression coefficients calculated from raw descriptor values are scaled by the mean and standard deviation of the descriptor values. Dormann et al.[78] discuss the use of standardized regression coefficients to compare the relative importance of descriptors after scaling by mean and standard deviation. For comparing coefficients in the regression model, the data can be standardized by subtracting the mean and dividing by standard deviation to achieve standardized regression coefficients with 0.0 mean and 1.0 standard deviation. After standardization, the standardized coefficients allow evaluating the relationship of each molecular descriptor to the property based on the sign and size of the coefficient.

Consequently, MLR offers a tool to find relevant descriptors, quantify their importance and relationship to the property, and predict property values of new compounds as per Todeschini and Consonni.[51] They further outline that this method offers good results, especially if the relationship between the descriptors and the property is highly linear. However, the results of MLR analysis depend on the quantity and quality of data, and non-linear relationships require the use of non-linear machine learning methods.

1.4.2 Random Forest Regression

RF[79] regression is a non-linear modelling algorithm, which trains an ensemble of models called regression trees. For new data points, the random forest model prediction is the arithmetic mean of the predicted values of the regression trees. A regression tree (**Figure 2**, top) is a hierarchical model consisting of regression nodes $s_1, s_2, \ldots, s_n$ with a tree-like structure.[80] In the regression tree prediction, starting from the initial node, each subsequent node is selected based on the values of the descriptors, until the "leaf" node is reached. Each leaf node has been assigned a log *K* value in model training, which is the predicted value of the

regression tree. Each regression tree is trained on a randomly sampled batch of the training data, due to which each regression tree structure in the random forest (**Figure 2**, bottom) can have a varying number of differently attributed nodes.[80]

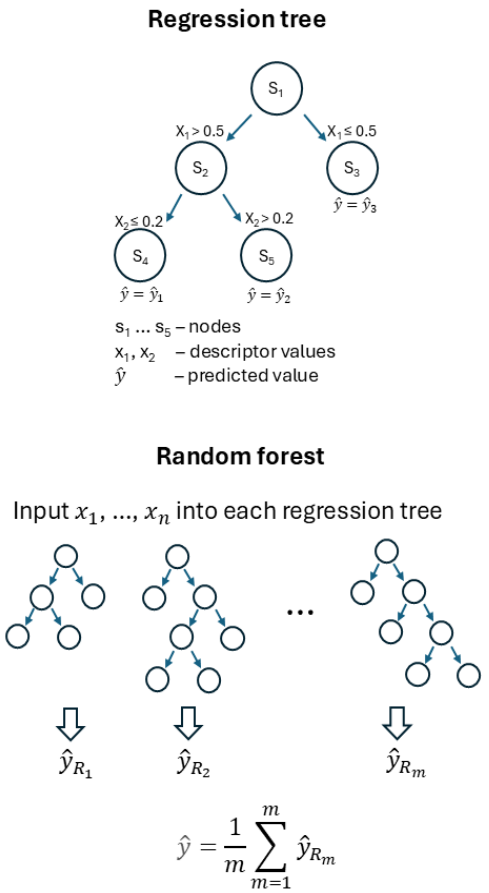

Figure 2. Regression tree (top) and random forest (bottom) model structure and working principle. As an example, starting from hypothetical descriptor values $x_1 = 0.4$ and $x_2 = 0.2$, the regression tree (top) nodes traversed based on descriptor values are $s_1 \rightarrow s_2 \rightarrow s_4$, and the predicted value is $\hat{y}_1 = f(x_1, x_2)$.

The RF algorithm depends on many hyperparameters that affect the resulting individual regression trees and the entire random forest.[80] The most important parameters are the number of regression trees, the maximum height of the tree, i.e., the maximum number of layers of tree nodes, the maximum number of leaf nodes and the minimum number of data points in a node to create child nodes. Another hyperparameter is the choice of whether each regression tree is given all training data or a randomized subsample of training data. The use of the random variant is called bootstrap aggregating or bagging, where some data points occur multiple times and a subset of the data points are left out. If certain data points

are omitted in each tree, it is also possible to use them as an out-of-bag test estimate. Another important parameter is the number of features to choose from when splitting a node into two sub nodes in the regression tree. Common choices are either all descriptors n or a random sample of $\sqrt{n}$ descriptors.[80]

Reasons to use RF for QSPR is mainly due to resilience to overfitting, ease of use, generalizability of ensemble methods, and decent interpretability.[80] RF is easy to use with the default hyperparameters out-of-the-box and generalizes well over large amounts of data and noisy target variables. With a large number of trees and nodes, the relationship between the input and the targets can be gauged by calculating descriptor importance values.[80] It should also be noted that RF is unable to predict a value outside the range of values in the model training data, which sets an applicability restriction on extrapolation. Overall, the RF model structure is straightforward in mechanism and modern software libraries do provide descriptor importance scores, quantifying the proportion that each descriptor contributes in the model to help with interpreting the final model.[79]

1.4.3 Support Vector Regression

The SVR model algorithm solves an optimization problem to find a set of training points called the support vectors, the regression coefficients, and the intercept of the SVR model.[81] Support vectors are training data points at least a tolerance distance ε from the calculated regression line (**Figure 3**). Training points within distance ε form a space called the "ε-insensitive tube" around the regression line. Points outside the tube are penalized in the optimization expression in model training.[81]

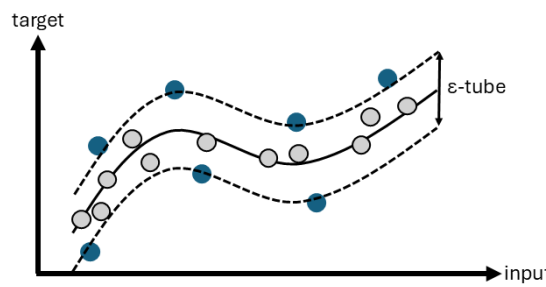

Figure 3. Support vector regression model regression line, the ε-insensitive tube, training points inside the tube (grey) and support vector points (blue).

The SVR model can account for non-linearity with respect to the descriptor inputs due to input transformation through the kernel function.[81] As further described by Schölkopf and Smola[81], the chosen kernel function maps training points into higher dimensional space, characterizing the similarity between two data points. The linear, polynomial, radial basis function (RBF), and sigmoid kernel are some

of the regularly applied kernels. Other than the kernel function hyperparameters, the SVR algorithm also involves hyperparameters C and ε, which influence the optimization expression. The error penalization term C should be configured with attention to the input data, as it influences how much the errors penalize the optimization expression and in turn deviate the regression line.[81] The chosen kernel function may contain additional hyperparameters, for example the RBF function can be adjusted by changing a scale coefficient γ.[73]

The model prediction for a molecule with descriptors x is:

$$\hat{y} = \sum_{i \in SV} \alpha_i \, k(x_i, x) + b,$$

where α_i are the regression coefficients of the support vector training points, x_i are descriptor values for the i-th support vector, k is the kernel function, and b is the intercept.[81]

The SVR method offers flexible non-linear modelling on small to medium-sized data sets with controlled model complexity via hyperparameters and a solid history of application in the space of machine learning methods.[81,82] It should be noted that the method works well on moderate data set sizes, but training time and memory can become demanding for larger modelling tasks due to task time scaling in n^3 with respect to training set size n, although there are approximate methods available.[81] SVR solution depends on only a subset of training points (the support vectors), which can be beneficial for interpretability in some cases.[81] The kernel maps data into high-dimensional spaces capturing complex non-linear relationships, but does not explicitly calculate the intermediate mathematical transforms, making the resulting function less interpretable directly.[81] However, model descriptor interpretation can be aided by other metrics such as calculating the descriptor permutation importance[83], where a descriptor is randomly shuffled and re-scored to determine the effect it has on the model.

1.4.4 Gaussian Process Regression

In the GPR model, a distribution over functions is defined using the training data, a covariance function, and the target log *K* values.[84] The predictions from the fitted GPR model form a full predictive distribution with a mean and standard deviation (**Figure 4**) over the entire input space.[84] The predicted distribution for new data points X_{*i} is calculated from the mean f_* and the covariance matrix $V[f_*]$:[84]

$$f_* = K(X_*, X)[K(X, X)]^{-1} y,$$

$$V[f_*] = K(X_*, X_*) - K(X_*, X)[K(X, X)]^{-1} K(X, X_*),$$

where X is the training descriptor matrix, X_* is the validation descriptor matrix, y is the vector of training log K values, and $K(X, X)$ is the $n\ x\ n$ covariance matrix computed by applying the kernel to all pairs of training data points.

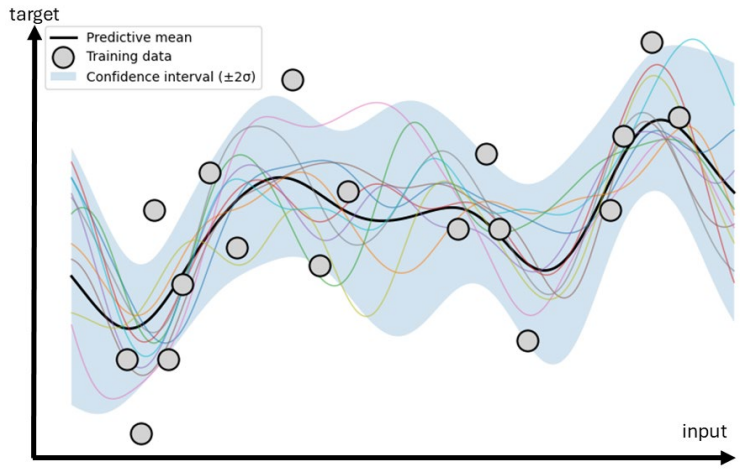

Figure 4. Example gaussian process regressor predicted mean line trained on sample data (grey points). Functions sampled from the distribution are shown in various colors. The kernel used for this example was $constant \cdot RBF$.

Kernels used in constructing the GPR model have a large effect on the derived space of functions. Rasmussen and Williams discuss commonly applied kernels including the constant, white noise, dot product, polynomial, RBF kernel, and combinations.[84] Models benefit from an explicit WhiteKernel to capture small-scale measurement variability not explained by other parts of the kernel. The DotProduct kernel allows modelling direct linear relations between the descriptor and log K efficiently. The RBF kernel can represent a broad class of smooth functions, allowing the model to discover subtle, highly complex relationships based on the similarity of the descriptor inputs.

Reasons to use GPR for QSPR are many, with high accuracy in early-stage research on limited data, uncertainty quantification, and non-linear flexibility based on assumptions about the derived function space, as outlined by Rasmussen and Williams[84]. They add that method works best with small to medium-sized data sets, due to time scaling in n^3 with respect to training set size n, similar to SVR. Calculating the descriptor permutation importance, similar as for SVR, aids in interpreting the relative impact of each descriptor in the final model. Overall, GPR is more complex in mechanism and to interpret, but it provides a powerful tool to model any function.

1.5 Previous Modelling of log *K*

Initial QSPR approaches to model log *K* were based primarily on linear modelling methods, such as MLR.[85–90] A prominent example is the Abraham solvation model, which uses a linear combination of solute descriptors (excess molar refractivity, dipolarity/polarizability, hydrogen-bond acidity and basicity, and solute size) combined with solvent-specific coefficients to predict properties like gas–IL partitioning.[85–89] This approach assigns interaction parameters as coefficients of the IL-specific Abraham model and the parameter values reflect particular solute-solvent interactions that correspond to chemical properties of the IL phase, while the researchers also note that attributing specific interactions to the cationic or anionic part of the IL could allow obtaining maximum application-specific performance.[86,88] Calculating log *K* for a new solute requires determining solute descriptors, which in turn requires sufficient experimental data or the use of commercial software to estimate the descriptor values.[88]

QSPR methodologies based on purely theoretical descriptors calculated from molecular structures allowed research efforts to develop effective QSPR models for partitioning properties of solutes eliminating the need for solute descriptor data collection.[90–95] For example, eight QSPR models were developed for the water-IL partition coefficients of solutes, where each model was based on MLR through 30-60 experimental data points and the molecular descriptors calculated from geometry optimized structures.[90] Calculating the descriptors solely from the structures of molecules makes it possible to apply the models to unavailable or unknown solutes.

Recent research efforts have increasingly incorporated molecular descriptors of the IL ionic components separately, acknowledging the important role of both the cation and anion in influencing solute distribution. Advanced modelling studies, such as those employing ion-specific and group contribution linear solvation energy relationships, have separated the Abraham model coefficients into ionic counterparts and further into contribution coefficients of the structure fragments in the cation and anion of the ionic liquid.[96–108] These improvements emphasize the significant impact of IL ionic composition on solute distribution behaviour.

While conventional linear approaches remain informative, the complexity and non-linearity inherent in IL systems have increasingly driven the adoption of more advanced machine learning techniques, such as SVR applied by Dashtbozorgi et al.[109] and Khooshechin et al.[110], in modelling the log *K* of a series of organic solutes with a constant IL.[109,110] Moreover, various other ML methods such as the artificial neural network, random forest, genetic algorithms, and many others have been successfully used in modelling IL properties.[111] Their success shows that the machine learning domain is promising for gathering further insights about solute partitioning. Application of ML methods to model and describe the effects of each component of the solute-cation-anion multicomponent system and the development of a general model can provide mechanistic understanding and enable the design of more selective and efficient IL media for targeted applications.

CHAPTER 2. DATA AND METHODS

Experimental data sets of gas–ionic liquid partition coefficients (K) at 298.15 K were compiled from various literature sources by Prof. William E. Acree (University of North Texas). The data consisted of log K value series for varying ranges of solutes in diverse ionic liquids. Specifically, the data set includes 6,531 log K values, involving 170 distinct solutes and 138 ionic liquids comprising 79 unique cations and 19 unique anions. log K values span from –1.63 to 11.03, and the structural diversity of both solutes and ionic liquids reflects a broad spectrum of solute-solvent interactions. The original data was organized resulting in a table of solutes in columns and ILs in rows for the purposes of preparation, organizing, visualization, overview, and analysis. To ensure consistency, the entire data set was extensively curated for standardizing the naming format of the solutes and ILs.

2.1 Case 1: log *K* as a function of the solute component

Driven by the overall aim of the study and the first hypothesis, the data matrix was analysed to find the data series, where the structural variation of one of the three components (solute) is the largest. The first set of three data series was for the case where the solute structure is changing and the IL remains constant, the so-called typical case in modelling properties of IL-s (**article I**). The data series included both saturated and unsaturated aliphatic, and aromatic organic compounds, mostly mono- or non-functional common organic solvents. The ILs were selected to have a common anion, $[FAP]^-$, in one pair and a common cation, $[BMPyrr]^+$, in another pair to compare the effect of switching the cation or the anion. The data processing followed a typical workflow depicted in **Figure 5**, explained in detail in the text of **article I**. Starting from the names of the organic solute components and their log K values, the preparation concluded with a data matrix consisting of ~1170 pre-processed molecular descriptors per solute molecule acting as the modelling parameters and corresponding log K values acting as the model output. The data were split into subsets to prepare for model training and validation by 5-fold cross-validation.

The machine learning methods used to derive the relationships between solute structure and log K were MLR and RF. The Sklearn library implementations RandomForestRegressor and MLR[112] were used. In both methods, descriptor selection stopped at four descriptors since the addition of a 5th descriptor did not significantly improve the common training evaluation metric of r^2_{CV5} for 5-fold cross-validation. MLR models descriptor selection followed a custom bottom-up scheme, detailed in **article I**, based on the highest r^2_{CV5} using OMP[112] for which the descriptor values were first also standardized to a mean of 0.0 and a standard deviation of 1.0. In the RF model development, descriptors were selected based on the highest r^2_{CV5} in a custom bottom-up scheme (**article I**). The resulting RF models for each solute series were optimized with a two-step hyperparameter

tuning. The AD of the linear models was analysed using a Cook's influence plot of leverage vs standardized residual values.

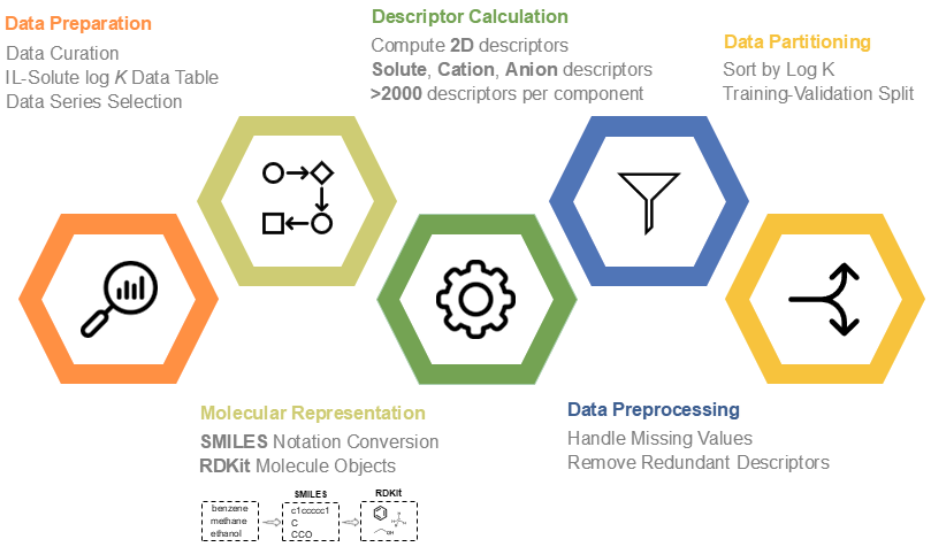

Figure 5. Five-step data processing workflow diagram for modelling of log *K*.

2.2 Case 2: modelling log *K* of solutes based on the IL cation

Due to lack of data in the literature, modelling of log *K* of solutes based on the structure of IL components is less common. Since there are two ionic parts in an IL, this offers the opportunity for non-typical modelling tasks based on the structure of the ion. Therefore, a second set of data series was constructed based on the data matrix where the structure of the cation of the ionic liquid changes, while the anion remains constant. The constant anion was chosen as $[Tf_2N]^-$ because it had the largest number of experimental log *K* values in the full data set. The cationic components of the data sets were diverse in their molecular structure, covering different cation families, functional groups, aliphatic or aromatic rings, branching and length of the aliphatic chain. Three solute data sets were selected for the comparison, hexane, cyclohexane, and benzene, with each solute-cation series having a size of ~60 data points. Data processing followed the same general workflow as depicted in **Figure 5**, resulting in 1180 cation descriptors in each solute data matrix. A 10-fold CV scheme distributed partition coefficient values evenly into each fold. The detailed data and methods are available in the **article II** text.

The linear model was developed using the Scikit-learn MLR[112] class with OrthogonalMatchingPursuit[112] for molecular descriptor selection. A custom bottom-up scheme was applied until diminishing returns measured by improvement to the r^2_{CV10} when adding a descriptor. SVR and GPR methods were used to

develop non-linear models. The implementation applied Scikit-learn library SVR[112] class with the RBF kernel and the GaussianProcessRegressor[112] class with the sum of WhiteKernel, DotProduct, and the RBF kernel from the same library module. This kernel sum was chosen based on the assumption that, the relationships in the log *K* model may be explained by a combination of separate trends or behaviors. In choosing the sum of WhiteKernel, DotProduct, and the RBF kernel combination, the assumed log *K* model considers and captures simple linear trends, allows flexible non-linear corrections and separates out noise. For SVR and GPR models, a custom bottom-up scheme optimized for the highest r^2_{CV10} until diminishing returns. The non-linear model development also included hyperparameter tuning. The AD of the linear models was analyzed based on an influence plot.

2.3 Case 3: modelling log *K* of solutes with the IL anion as a parameter

The third set of data series dealt with the case where only the structure of anion is variable, and cation and solute remain constant. The availability of data allowed the compilation of small but structurally representative data series. The nine data series featured combinations of ethylmethylimidazolium, $[EMIm]^+$, butylmethylimidazolium, $[BMIm]^+$, or hexylmethylimidazolium $[HMIm]^+$ and one of three solutes: benzene, cyclohexane, and methanol. The series varied in size (7–13 data points) and consisted of diverse anions with distinct molecular structures, including ionic counterparts, functional groups, symmetricity, heteroatoms, charge distribution, size, and shape. Following the generic workflow (**Figure 5**), the final data matrices consisted of ~500 standardized descriptors and a corresponding log *K* value per anion. LOO CV was chosen to validate the models derived on the data series because the series contained fewer measurements, so that each data point served as the test sample in turn. More detailed description of data and methods is available in the text of **article III**.

The choice of MLR[112] as the modelling method was guided by the composition and sizes of the data series. The feature selection process was carried out using the OrthogonalMatchingPursuit[112] class from the Scikit-learn library in a bottom-up scheme targeting the highest $r^2_{LOO\ CV}$. The model's AD analysis was carried through using an influence plot and descriptor box-swarm distribution plots.

2.4 Case 4: log *K* of solutes as a function of solute, cation, and anion components

A comprehensive data set with all 6,531 of log *K* values and the entire range of solutes, cations, and anions was compiled to develop generalized multi-component QSPR models in **article IV**. The solutes in the dataset cover diverse chemical structures, from simple hydrocarbons to more diverse halogenated compounds like 2,2,2-trifluoroethanol and 1,2-dichlorobenzene (**article IV,** S1).

On the side of ionic liquids mainly quaternary ammonium and imidazolium cations are represented, which are common in the ionic liquid research, with the cation structures ranging from simple alkyl chains to more complex aromatic moieties (**article IV,** S2). Anions range from simple halides like hexafluorophosphate to more complex species such as bis(trifluoromethylsulfonyl)imide, including most of the typical anions used in ionic liquid applications (**article IV,** S3). The general data workflow (**Figure 5**) was followed, and the final data matrix consisted of 6531 rows consisting of a total of 3245 descriptor columns from the solute, cation, and anion molecular structures and the log *K* column. The data rows were split into the holdout set and training set, and the training set was further split into five sets with even partition coefficient distribution in each for cross-validation.

In model development, MLR[112] class from the Sklearn library was used as the linear implementation accompanied by an iterative bottom-up descriptor selection based on achieving the highest r^2_{CV5}. In non-linear modelling, RF[112] implementation from the Sklearn library was used and descriptors were selected in a similar iterative bottom-up scheme. The AD of optimal linear model was examined via influence plot analysis, employing DBSCAN[113] clustering to group together points with similar leverage-residual profiles. The AD of the optimal RF model was similarly investigated by computing Mahalanobis distance. DBSCAN was again applied to group together and identify points with similar Mahalanobis-residual profiles.

2.5 Classification of Molecular Descriptors According to the Solute-Solvent Interactions

In the case of QSPR, one of the central elements is the interpretation and analysis of the molecular descriptors selected for the model. This is a challenging task because there are many molecular descriptors and their calculation schemes are often complex, combining different elements of the chemical structure that describe their combined effects.

This created the need to find an approach to generalize and classify molecular descriptors and, if possible, link them to the interactions (**Figure 6**) occurring in the gas-IL system. Since the gas-IL system is similar in nature to the gas-liquid system, the framework for classification is provided by the solvation free energy that is related to the solubility at equilibrium conditions and constant temperature. More specifically, the solvation free energy can be divided into components that express non-specific (i.e., indirect) and specific (i.e., direct) interactions in the solute-solvent system.[114] Non-specific are the interactions caused by electrostatic i.e., Coulombic and dipolar interactions, dispersion forces, and the occupation of a space in the solvent, i.e., the creation of a cavity.[115,116] Specific ones are related to hydrogen bonding interactions.[115,116]

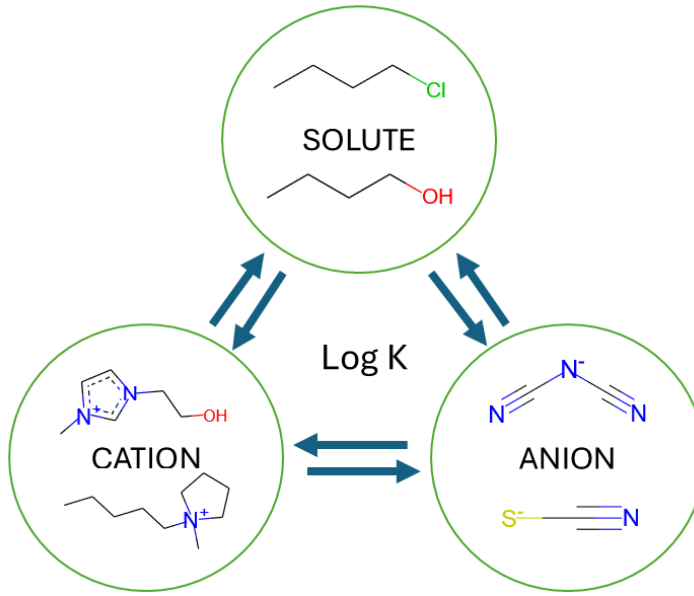

Figure 6. Solute, cation, and anion in interaction with each component of the system.

Analysis of the computational schemes of molecular descriptors allows to link the structural contributions used in the calculations with the corresponding solute-solvent interactions (see for example **article IV, Figure 5**) and thus, through the molecular descriptors, it is possible to provide a substantive view of the mechanistic understanding of the derived relationships. More specifically, providing the space in which the molecule resides is linked (i) to dispersion forces and cavity formation, related to molecule size, shape, surface area, and polarizability; (ii) intermolecular electrostatic interaction, associated with Coulombic and dipolar forces related to polarity and charge distribution, and (iii) the presence of specific functional groups capable of acting as HB donors or acceptors relates to hydrogen bonding interaction capability.[86]

For this purpose, the mathematical formulas of the descriptors were analysed based on the program code[117] and the literature[51] and classified according to the molecular interactions in the solute and solvent systems, which correspond to the concepts of solvation free energy[114–116], and which have also been used elsewhere.[92–95]

2.5 Model Availability and Reporting

For the purpose of ensuring reproducibility and adherence to best practices[118,119], the final models and data were stored in PMML[120] format suitable for QSPR (e.g., QsarDB)[121,122]. Digital object identifiers (DOIs) were assigned to the models, and all descriptor and model development steps were documented following FAIR[123] principles, allowing future researchers to reuse and test the models.

CHAPTER 3. RESULTS AND DISCUSSION

3.1 log *K* models for solute series

The development of three solute log *K* series models resulted in MLR and RF models[124], each consisting of four descriptors (**Article I**). The evaluation metrics demonstrated excellent predictive capability for log *K*, with r^2_{CV5} ranging between 0.88 and 0.94. The RF models offered slight performance gains over the MLR models for two of the three IL datasets including the highest performing [BMPyrr]$^+$[FAP]$^-$-solute series (**Figure 7**), highlighting non-linear effects in describing the partition coefficient. However, the MLR approach offered more straightforward interpretations of how molecular structural factors influence solute partition coefficients. The AD analysis highlighted some solute categories for each model that could improve predictive performance if more experimental data points were introduced.

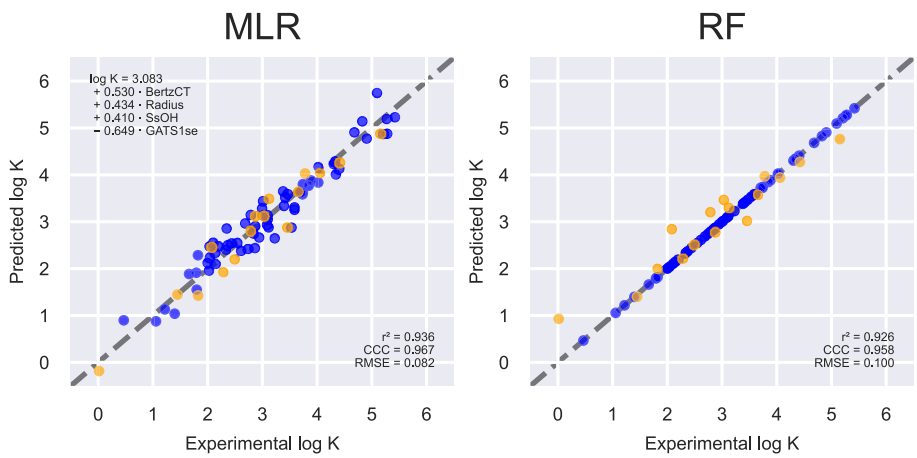

Figure 7. Predicted vs. experimental log *K* scatter plots of the best MLR and RF model (for CV Fold 1), both relationships of form $log\ K_{[BMPyrr]^+[FAP]^-} = f(solute)$ with training set observations in blue and validation set values in orange. Evaluation metrics (r^2, *CCC*, *RMSE*) shown were calculated as the arithmetic mean over five validation folds.

Both RF and MLR model descriptors included hydrogen bonding, dipolar or Coulombic interactions, and dispersion-related properties. MLR standardized regression coefficient analysis revealed that HB, dipolar and Coulombic effects were more prominent for the organic solute, compared to descriptors related to dispersion effects. RF model descriptors were more diverse in interaction information, and therefore, an exact mapping to log *K* was also more complicated. It is noteworthy that the adoption of the RF method was new to modelling of log *K*, and its success proved to be a good starting point for follow-up studies.

Comparisons of the three ILs indicated that the common [FAP]$^-$ anion in the [BMPyrr]$^+$[FAP]$^-$ and [MeoeMPyrr]$^+$[FAP]$^-$ IL-s produced very high correlation,

~0.999 between their respective log *K* data series. By contrast, when the same cation $[BMPyrr]^+$ was paired with two distinct anions in the $[BMPyrr]^+[FAP]^-$ and $[BMPyrr]^+[C(CN)_3]^-$, the correlation, 0.93, was smaller. This difference suggests that the choice of anion exerts a larger overall impact on the solute partitioning behaviour than minor variations in the cation.

3.2 Dependence of log *K* on cation structure

In **article II**, all final models[125] demonstrated robust predictive performance with r^2_{CV10} in the range of 0.71–0.93, indicating that the partition coefficients of relatively simple hydrocarbons can be accurately modeled through descriptors encoding cation structure alone. For hexane and cyclohexane, the r^2_{CV10} values approached 0.90 or higher, reflecting the strong influence of cation structural motifs on non-polar solute solubility. Benzene showed a competitive r^2_{CV10} of around 0.72–0.85, indicating satisfactory predictive accuracy. In each series, both linear and non-linear modelling methods yielded similarly high predictive metrics, but certain data sets showed modest gains in accuracy when non-linear models such as SVR and GPR were used. Models for log *K* of hexane (**Figure 8**) had the highest prediction performance for every applied ML method. Across every solute data series, SVR models (**Figure 9**) had the highest evaluation metrics compared to the other methods. Model diagnostics confirmed that the linear and non-linear models remained robust and generalizable, covering a broad cation structural domain that spans imidazolium, ammonium, pyrrolidinium, and piperidinium cations. Di-cations, morpholinium, and sulfonium cations were among the structures that occasionally produced higher leverage or moderate residuals, reflecting their distinctive molecular features and limited representation in the data sets.

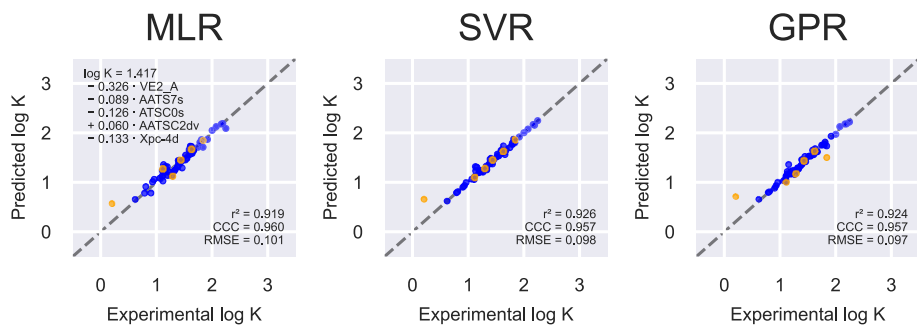

Figure 8. Predicted vs. experimental log *K* scatter plots of the best MLR, SVR, GPR models (for CV Fold 1). All are relationships of the form $log\ K_{hexane\ in\ [cation]^+[Tf_2N]^-} = f(cation)$ with training set observations in blue and validation set values in orange. Evaluation metrics (r^2, *CCC*, *RMSE*) shown were calculated as the arithmetic mean over ten validation folds.

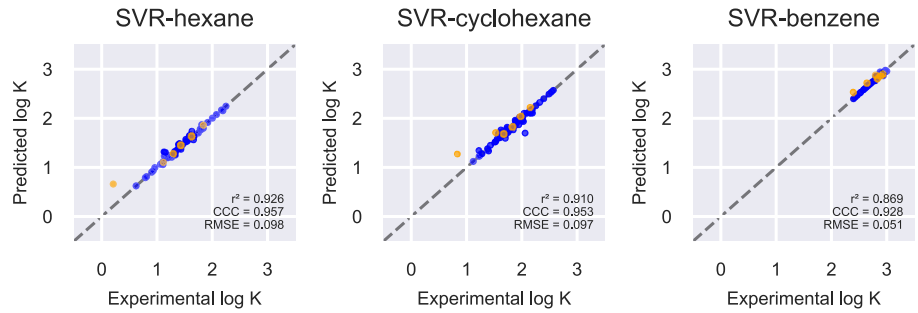

Figure 9. Predicted vs. experimental log *K* scatter plots (for CV Fold 1) of the best ML method (SVR) for the hexane, cyclohexane, and benzene series with training set observations in blue and validation set values in orange. Evaluation metrics (r^2, *CCC*, *RMSE*) shown were calculated as the arithmetic mean over ten validation folds.

Interpretation of descriptors highlighted that more lipophilic non-aromatic cations with larger surface area or extended alkyl chains give higher log *K* values for the studied hydrocarbons and cations with strong electronegative moieties generally lower the partition coefficient. Particularly in the cation series for benzene, cation descriptors associated with dipolar interactions and hydrogen bonding capacity gained more importance along size-related properties. This aligns with the aromatic nature of benzene, which can engage in dipolar interactions depending on the electronic environment of the solute. The solute competes with the anion-cation interaction in solution, and the anion-cation interaction could instead be weaker due to lipophilicity and charge shielding of the cation to produce a higher solubility in the environment for the organic solute.

3.3 Dependence of log *K* on anion structure

Nine derived two-parameter MLR models[126] with varying anion structures demonstrated a correlation coefficient, r^2, in the range 0.58–0.997 (best models showcased in **Figure 10**) and internal validation $r^2_{LOO\ CV}$ up to 0.980 (**Article III**). The CCC for models similarly reached values of 0.88–0.98, underscoring that linear correlations between the descriptors encoding the anion structure and the measured log *K* values were reliable. The results confirm the initial hypothesis and provide valuable information on the role of the structure of the anionic component in determining the partitioning behavior of organic solutes. Validation using LOO CV and diagnostic tools like standardized residuals, leverages, and Cook's distance helped clarify where the MLR models are robust. Despite occasional outliers, none of the data points severely distorted the derived regressions.

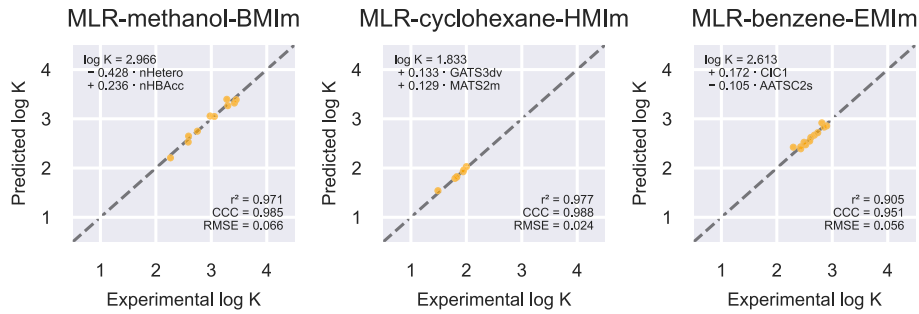

Figure 10. Predicted vs. experimental log K scatter plots of the best MLR models for $\log K = f(anion)$ with LOO CV values in orange. Evaluation metrics (r^2, *CCC*, *RMSE*) shown were calculated as the arithmetic mean over all validation folds (orange points on the figure).

For benzene, the chosen descriptors frequently contained size- and symmetry-related information consistent with greater dispersion interactions. Alongside were descriptors that were measures of electronegativity or valence electron distribution related to increased dipolar or Coulombic interactions in the IL environment. Cyclohexane models tended to show an even stronger emphasis on dispersion-related descriptors. The proportionality of these descriptors with the anion's overall size and branching signified that purely steric and polarizability factors dominated the solvation of cyclohexane in these ILs. In contrast, methanol-based models gave consistently higher importance to descriptors tied to hydrogen bonding and dipolar-Coulomb interactions, such as the counts of heteroatoms or hydrogen bond acceptors within the anion. Anion descriptors related to the dispersion force in the methanol-based models had negative regression coefficients indicating an inversely proportional relationship between log K and dispersion force strength.

Taken together, the descriptors selected by the MLR approach can be categorized similarly to those described in **article I** and **II**: dispersion forces remain relevant for all solutes; dipolar or Coulomb interactions matter particularly for polar functionalities; and hydrogen bonding has strong effects for polar HB-capable solutes like methanol. **Article III** results confirm that anion structure plays an equally vital role in determining log K in many imidazolium-based systems.

3.4 Dependence of log K on the structure of both the solute, cation, and anion

Modelling the entire data set together, including molecular features describing the structure of both the solute and the cationic and anionic parts of the ionic

liquid, gave unexpectedly good results[127] (**Article IV**). The final MLR model incorporated eight solute, one cation, and one anion descriptors, reflecting a dominance of solute-related structural factors in the linear model. Given the size of the dataset, the MLR approach yielded a model (**Figure 11**) with strong predictive performance, with a cross-validated coefficient of determination (r^2_{CV5}) of ~0.80. In addition, the RF model (**Figure 10**) performed excellently, achieving a predictive power, r^2_{CV5}, of ~0.97 and a similarly high external validation r^2 on the holdout set. The improvement observed for RF was due to the ability to capture significant non-linearities associated with IL structural features giving a more balanced representation of cation and anion descriptors, indicating that capturing effects by ionic liquid components requires accounting non-linearity in the modelling. The evaluation of AD for MLR revealed that sparse representation of complex functionalities in the solute, and hydrophobic or sterically hindered ions can challenge a strictly linear approach. The RF model derived using descriptors that describe nonlinear structural effects proved to be more robust, but still showed clusters of outliers, especially for small polar solutes or large alkanes.

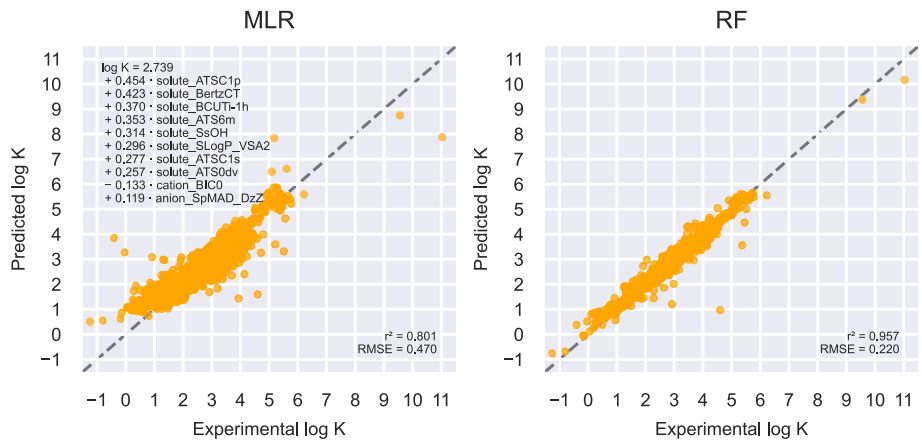

Figure 11. Predicted vs. experimental log *K* scatter plots of the best MLR and RF model for $\log K = f(solute, cation, anion)$ on the external validation set. Evaluation metrics (r^2, *RMSE*) shown were calculated on the external validation set.

Interpretation of the descriptors indicated that solute structure continues to be the primary contributor to variability in gas–ionic liquid distribution, consistent with earlier findings where solute size, polarity, polarizability, and HB were dominant factors. Nevertheless, the cation and anion descriptors showed a more substantial effect in the RF model, underlining that these components must be accounted for. Descriptors representing dispersion interactions, hydrogen bonding, and dipolar forces appeared consistently, highlighting the multifaceted nature of IL-based mixtures.

SUMMARY

The research described in this thesis advances the understanding of log *K* through a comprehensive and methodologically diverse exploration of QSPR modelling approaches. Starting the modelling of the solute–IL system in selected data-series and moving through the analysis of solute, cation, and anion structures, the work demonstrates that machine learning methods, especially RF, SVR, and GPR, often capture solute–IL interactions more effectively than conventional MLR. Nonetheless, the MLR approach consistently provides valuable interpretative insight, particularly regarding HB, Coulomb-dipolar interactions, and dispersion forces. Both linear and non–linear models emphasize the critical impact of cation and anion composition on solute partitioning, reflecting the complexity of the solvating environment created by ILs.

Collectively, the findings suggest that common structural elements can systematically strengthen or weaken partitioning, depending on the balance between dispersion forces, Coulomb–dipolar, and HB interactions. The choice of IL ionic component is repeatedly shown to be crucial in determining log *K*. The work further demonstrates that modelling the entire system by combining descriptors from solute, cation, and anion improves the predictive performance for large and chemically diverse datasets, highlighting the importance of multicomponent approaches. Rigorous curation of experimental data and thorough validation using cross-validation, holdout test data, and influence analyses reinforce the reliability and applicability of each model. By examining the influence of each structural component in the solute–cation–anion system, these studies not only provide mechanistic insights into intermolecular forces but also enable the design of more selective and efficient IL media for targeted industrial, environmental, and research applications.

REFERENCES

(1) An Introduction to Ionic Liquids. In *Fundamentals of ionic liquids: From Chemistry to Applications*; Wiley-VCH, 2017; p 2.

(2) Reichardt, C.; Welton, T. Classification of Solvents. In *Solvents and Solvent Effects in Organic Chemistry*; John Wiley & Sons, Ltd, 2010; pp 65–106. https://doi.org/10.1002/9783527632220.ch3.

(3) Douglas R MacFarlane; Mega Kar; Jennifer M Pringle. Chapters 5, 6, 7, 8 and 9. In *Fundamentals of Ionic Liquids*; John Wiley & Sons, Ltd, 2017; pp 103–244. https://doi.org/10.1002/9783527340033.ch7.

(4) Brahma, S.; Gardas, R. L. History and Development of Ionic Liquids. In *Handbook of Ionic Liquids*; John Wiley & Sons, Ltd, 2024; pp 1–28. https://doi.org/10.1002/9783527839520.ch1.

(5) Javed, F.; Ullah, F.; Zakaria, M. R.; Akil, H. M. An Approach to Classification and Hi-Tech Applications of Room-Temperature Ionic Liquids (RTILs): A Review. *J. Mol. Liq.* **2018**, *271*, 403–420. https://doi.org/10.1016/j.molliq.2018.09.005.

(6) Singh, S. K.; Savoy, A. W. Ionic Liquids Synthesis and Applications: An Overview. *J. Mol. Liq.* **2020**, *297*, 112038. https://doi.org/10.1016/j.molliq.2019.112038.

(7) Welton, T. Ionic Liquids in Catalysis. *Coord. Chem. Rev.* **2004**, *248* (21), 2459–2477. https://doi.org/10.1016/j.ccr.2004.04.015.

(8) Pârvulescu, V. I.; Hardacre, C. Catalysis in Ionic Liquids. *Chem. Rev.* **2007**, *107* (6), 2615–2665. https://doi.org/10.1021/cr050948h.

(9) Zhang, Q.; Zhang, S.; Deng, Y. Recent Advances in Ionic Liquid Catalysis. *Green Chem* **2011**, *13* (10), 2619–2637. https://doi.org/10.1039/C1GC15334J.

(10) Hallett, J. P.; Welton, T. Room-Temperature Ionic Liquids: Solvents for Synthesis and Catalysis. 2. *Chem. Rev.* **2011**, *111* (5), 3508–3576. https://doi.org/10.1021/cr1003248.

(11) Steinrück, H.-P.; Wasserscheid, P. Ionic Liquids in Catalysis. *Catal. Lett.* **2015**, *145* (1), 380–397. https://doi.org/10.1007/s10562-014-1435-x.

(12) Liu, Y.-S.; Pan, G.-B. Ionic Liquids for the Future Electrochemical Applications. In *Ionic Liquids: Applications and Perspectives*; InTech, 2011; pp 627–642.

(13) Faridbod, F.; Ganjali, M. R.; Norouzi, P.; Riahi, S.; Rashedi, H. Application of Room Temperature Ionic Liquids in Electrochemical Sensors and Biosensors. In *Ionic Liquids: Applications and Perspectives*; InTech, 2011; pp 643–658.

(14) Virendra V. Singh; Anil K. Nigam; Anirudh Batra; Mannan Boopathi; Beer Singh; Rajagopalan Vijayaraghavan. Applications of Ionic Liquids in Electrochemical Sensors and Biosensors. *International Journal of Electrochemistry* **2012**, *2012* (Article ID 165683), 19 pages. https://doi.org/10.1155/2012/165683.

(15) Douglas R MacFarlane; Mega Kar; Jennifer M Pringle. Electrochemistry of and in Ionic Liquids. In *Fundamentals of Ionic Liquids*; John Wiley & Sons, Ltd, 2017; pp 177–207. https://doi.org/10.1002/9783527340033.ch7.

(16) Liu, H.; Yu, H. Ionic Liquids for Electrochemical Energy Storage Devices Applications. *J. Mater. Sci. Technol.* **2019**, *35* (4), 674–686. https://doi.org/10.1016/j.jmst.2018.10.007.

(17) Tiago, G. A. O.; Matias, I. A. S.; Ribeiro, A. P. C.; Martins, L. M. D. R. S. Application of Ionic Liquids in Electrochemistry—Recent Advances. *Molecules* **2020**, *25* (24), 5812. https://doi.org/10.3390/molecules25245812.

(18) Huddleston, J. G.; Willauer, H. D.; Swatloski, R. P.; Visser, A. E.; Rogers, R. D. Room Temperature Ionic Liquids as Novel Media for 'clean' Liquid–Liquid Extraction. *Chem Commun* **1998**, No. 16, 1765–1766. https://doi.org/10.1039/a803999b.
(19) Trinh, H. B.; Lee, J.; Lee, J. Chapter 7 – Task-Specific Ionic Liquids for the Separation and Recovery of Rare Earth Elements. In *Ionic Liquid-Based Technologies for Environmental Sustainability*; Jawaid, M., Ahmad, A., Reddy, A. V. B., Eds.; Elsevier, 2022; pp 101–121. https://doi.org/10.1016/B978-0-12-824545-3.00007-6.
(20) Trujillo-Rodríguez, M. J.; Pacheco-Fernández, I.; Pino, V. Chapter FIVE – Magnetic Ionic Liquids in Analytical Sample Separation Techniques. In *Ionic Liquids in Analytical Chemistry*; Carda-Broch, S., Ruiz-Angel, M., Eds.; Elsevier, 2022; pp 141–170. https://doi.org/10.1016/B978-0-12-823334-4.00001-1.
(21) Khan, J. A.; Jabin, S. Chapter 8 – Ionic Liquids as Valuable Assets in Extraction Techniques. In *Advanced Applications of Ionic Liquids*; Siddique, J. A., Ansari, S. P., Khan, A. A. P., Asiri, A. M., Eds.; Elsevier, 2023; pp 199–221. https://doi.org/10.1016/B978-0-323-99921-2.00014-8.
(22) Earle, M. J.; Seddon, K. R. Ionic Liquids. Green Solvents for the Future. *Pure Appl. Chem.* **2000**, *72* (7), 1391–1398. https://doi.org/10.1351/pac200072071391.
(23) Klembt, S.; Dreyer, S.; Eckstein, M.; Kragl, U. Biocatalytic Reactions in Ionic Liquids. In *Ionic Liquids in Synthesis*; John Wiley & Sons, Ltd, 2007; pp 641–661. https://doi.org/10.1002/9783527621194.ch8.
(24) Maase, M. Industrial Applications of Ionic Liquids. In *Ionic Liquids in Synthesis*; John Wiley & Sons, Ltd, 2007; pp 663–687. https://doi.org/10.1002/9783527621194.ch9.
(25) Welton, T.; Endres, F.; Abedin, S. Z. E.; Antonietti, M.; Smarsly, B.; Zhou, Y. Inorganic Synthesis. In *Ionic Liquids in Synthesis*; John Wiley & Sons, Ltd, 2007; pp 569–617. https://doi.org/10.1002/9783527621194.ch6.
(26) Earle, M.; Wasserscheid, P.; Schulz, P.; Olivier-Bourbigou, H.; Favre, F.; Vaultier, M.; Kirschning, A.; Singh, V.; Riisager, A.; Fehrmann, R.; Kuhlmann, S. Organic Synthesis. In *Ionic Liquids in Synthesis*; John Wiley & Sons, Ltd, 2007; pp 265–568. https://doi.org/10.1002/9783527621194.ch5.
(27) Haddleton, D. M.; Welton, T.; Carmichael, A. J. Polymer Synthesis in Ionic Liquids. In *Ionic Liquids in Synthesis*; John Wiley & Sons, Ltd, 2007; pp 619–640. https://doi.org/10.1002/9783527621194.ch7.
(28) Mecerreyes, D. Polymeric Ionic Liquids: Broadening the Properties and Applications of Polyelectrolytes. *Prog. Polym. Sci.* **2011**, *36* (12), 1629–1648. https://doi.org/10.1016/j.progpolymsci.2011.05.007.
(29) Eshetu, G. G.; Armand, M.; Ohno, H.; Scrosati, B.; Passerini, S. Ionic Liquids as Tailored Media for the Synthesis and Processing of Energy Conversion Materials. *Energy Environ. Sci.* **2016**, *9* (1), 49–61. https://doi.org/10.1039/C5EE02284C.
(30) Claus, J.; Sommer, F. O.; Kragl, U. Ionic Liquids in Biotechnology and Beyond. *Solid State Ion.* **2018**, *314*, 119–128. https://doi.org/10.1016/j.ssi.2017.11.012.
(31) Faisal, M.; Saeed, A. 6 – The Role of Ionic Liquid in Medicinal Chemistry. In *Green Approaches in Medicinal Chemistry for Sustainable Drug Design*; Banik, B. K., Ed.; Advances in Green and Sustainable Chemistry; Elsevier, 2020; pp 143–180. https://doi.org/10.1016/B978-0-12-817592-7.00006-X.
(32) Pei, Y.; Zhang, Y.; Ma, J.; Fan, M.; Zhang, S.; Wang, J. Ionic Liquids for Advanced Materials. *Mater. Today Nano* **2022**, *17*, 100159. https://doi.org/10.1016/j.mtnano.2021.100159.

(33) MacFarlane, D. R.; Kar, M.; Pringle, J. M. *Fundamentals of Ionic Liquids*; Wiley-VCH Verlag GmbH & Co. KGaA, 2017. https://doi.org/10.1002/9783527340033.
(34) Ghandi, K. A Review of Ionic Liquids, Their Limits and Applications. *Green Sustain. Chem.* **2014**, *04* (01), 44–53. https://doi.org/10.4236/gsc.2014.41008.
(35) Hospido, A.; Rodríguez, H. Life Cycle Assessment (LCA) of Ionic Liquids. In *Encyclopedia of Ionic Liquids*; Springer Singapore, 2019; pp 1–9. https://doi.org/10.1007/978-981-10-6739-6_54-1.
(36) DESIGNER SOLVENTS. *Chem. Eng. News Arch.* **1998**, *76* (13), 32–37. https://doi.org/10.1021/cen-v076n013.p032.
(37) Mathew, H. T.; Abhisek, K.; Vhatkar, S. S.; Kumar, A.; Oraon, R. Ionic Liquids as Green Solvents: Are Ionic Liquids Nontoxic and Biodegradable? In *Handbook of Ionic Liquids*; John Wiley & Sons, Ltd, 2024; pp 69–96. https://doi.org/10.1002/9783527839520.ch4.
(38) Armarego, W. L. F.; Chai, C. L. L.; D.D.P.; W.L.F.A.; D.R.P.; D.D.P.; W.L.F.A.; D.R.P.; D.D.P.; W.L.F.A.; Armarego, W. L. F.; Armarego, W. L. F.; Chai, C. L. L.; Armarego, W. L. F.; Chai, C. L. L. Preface to the Seventh Edition. In *Purification of Laboratory Chemicals*; Elsevier, 2013; pp 5–10. https://doi.org/10.1016/b978-0-12-382161-4.05001-9.
(39) Bourbos, E.; Giannopoulou, I.; Karantonis, A.; Paspaliaris, I.; Panias, D. Chapter 13 – Electrodeposition of Rare Earth Metals from Ionic Liquids. In *Rare Earths Industry*; Borges De Lima, I., Leal Filho, W., Eds.; Elsevier: Boston, 2016; pp 199–207. https://doi.org/10.1016/B978-0-12-802328-0.00013-9.
(40) Shukla, M. K.; Tiwari, H.; Verma, R.; Dong, W.-L.; Azizov, S.; Kumar, B.; Pandey, S.; Kumar, D. Role and Recent Advancements of Ionic Liquids in Drug Delivery Systems. *Pharmaceutics* **2023**, *15* (2), 702. https://doi.org/10.3390/pharmaceutics15020702.
(41) Mondal, T.; Samanta, P. Study of Physicochemical Properties of Ionic Liquids. In *Handbook of Ionic Liquids*; John Wiley & Sons, Ltd, 2024; pp 51–67. https://doi.org/10.1002/9783527839520.ch3.
(42) Speight, J. G. Introduction Into the Environment. In *Environmental Organic Chemistry for Engineers*; Elsevier, 2017; pp 263–303. https://doi.org/10.1016/b978-0-12-804492-6.00006-x.
(43) Speight, J. G. Molecular Interactions, Partitioning, and Thermodynamics. In *Reaction Mechanisms in Environmental Engineering*; Elsevier, 2018; pp 307–336. https://doi.org/10.1016/b978-0-12-804422-3.00009-2.
(44) Voelkel, A.; Strzemiecka, B.; Adamska, K.; Milczewska, K. Inverse Gas Chromatography as a Source of Physiochemical Data. *J. Chromatogr. A* **2009**, *1216* (10), 1551–1566. https://doi.org/10.1016/j.chroma.2008.10.096.
(45) Marciniak, A. The Solubility Parameters of Ionic Liquids. *Int. J. Mol. Sci.* **2010**, *11* (5), 1973–1990. https://doi.org/10.3390/ijms11051973.
(46) Kar, S.; Sizochenko, N.; Ahmed, L.; Batista, V. S.; Leszczynski, J. Quantitative Structure-Property Relationship Model Leading to Virtual Screening of Fullerene Derivatives: Exploring Structural Attributes Critical for Photoconversion Efficiency of Polymer Solar Cell Acceptors. *Nano Energy* **2016**, *26*, 677–691. https://doi.org/10.1016/j.nanoen.2016.06.011.
(47) Achary, P. G. R. Applications of Quantitative Structure-Activity Relationships (QSAR) Based Virtual Screening in Drug Design: A Review. *Mini Rev. Med. Chem.* **2020**, *20* (14), 1375–1388. https://doi.org/10.2174/1389557520666200429102334.

(48) Cronin, M. T. D. Quantitative Structure–Activity Relationships (QSARs) – Applications and Methodology. In *Challenges and Advances in Computational Chemistry and Physics*; Springer Netherlands, 2009; pp 3–11. https://doi.org/10.1007/978-1-4020-9783-6_1.
(49) Yee, L. C.; Wei, Y. C. Current Modeling Methods Used in QSAR/QSPR. In *Statistical Modelling of Molecular Descriptors in QSAR/QSPR*; Wiley-VCH Verlag GmbH & Co. KGaA, 2012; pp 1–31. https://doi.org/10.1002/9783527645121.ch1.
(50) Ruusmann, V.; Maran, U. From Data Point Timelines to a Well Curated Data Set, Data Mining of Experimental Data and Chemical Structure Data from Scientific Articles, Problems and Possible Solutions. *J. Comput. Aided Mol. Des.* **2013**, *27* (7), 583–603. https://doi.org/10.1007/s10822-013-9664-4.
(51) Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*, 1st ed.; Methods and Principles in Medicinal Chemistry; Wiley, 2000. https://doi.org/10.1002/9783527613106.
(52) Feurer, M.; Hutter, F. Hyperparameter Optimization. In *Automated Machine Learning: Methods, Systems, Challenges*; Hutter, F., Kotthoff, L., Vanschoren, J., Eds.; Springer International Publishing: Cham, 2019; pp 3–33. https://doi.org/10.1007/978-3-030-05318-5_1.
(53) Tropsha, A.; Gramatica, P.; Gombar, V. The Importance of Being Earnest: Validation Is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR Comb. Sci.* **2003**, *22* (1), 69–77. https://doi.org/10.1002/qsar.200390007.
(54) Krein, M.; Huang, T.-W.; Morkowchuk, L.; Agrafiotis, D. K.; Breneman, C. M. Developing Best Practices for Descriptor-Based Property Prediction: Appropriate Matching of Datasets, Descriptors, Methods, and Expectations. In *Statistical Modelling of Molecular Descriptors in QSAR/QSPR*; Wiley-VCH Verlag GmbH & Co. KGaA, 2012; pp 33–64. https://doi.org/10.1002/9783527645121.ch2.
(55) Golbraikh, A.; Tropsha, A. Beware of Q2! *J. Mol. Graph. Model.* **2002**, *20* (4), 269–276. https://doi.org/10.1016/s1093-3263(01)00123-1.
(56) Shahlaei, M. Descriptor Selection Methods in Quantitative Structure–Activity Relationship Studies: A Review Study. *Chem. Rev.* **2013**, *113* (10), 8093–8103. https://doi.org/10.1021/cr3004339.
(57) Chirico, N.; Gramatica, P. Real External Predictivity of QSAR Models: How To Evaluate It? Comparison of Different Validation Criteria and Proposal of Using the Concordance Correlation Coefficient. *J. Chem. Inf. Model.* **2011**, *51* (9), 2320–2335. https://doi.org/10.1021/ci200211n.
(58) Hong, H.; Slavov, S.; Ge, W.; Qian, F.; Su, Z.; Fang, H.; Cheng, Y.; Perkins, R.; Shi, L.; Tong, W. Mold2Molecular Descriptors for QSAR. In *Statistical Modelling of Molecular Descriptors in QSAR/QSPR*; Wiley-VCH Verlag GmbH & Co. KGaA, 2012; pp 65–109. https://doi.org/10.1002/9783527645121.ch3.
(59) Molecular Descriptors from Two-Dimensional Chemical Structure. In *In Silico Toxicology*; The Royal Society of Chemistry, 2010; pp 148–192. https://doi.org/10.1039/bk9781849730044-00148.
(60) Mauri, A.; Consonni, V.; Pavan, M.; Todeschini, R.; Chemometrics, M. DRAGON SOFTWARE: AN EASY APPROACH TO MOLECULAR DESCRIPTOR CALCULATIONS; 2006.
(61) Ivanciuc, O. CODESSA Version 2.13 for Windows. *J. Chem. Inf. Comput. Sci.* **1997**, *37* (2), 405–406. https://doi.org/10.1021/ci950193n.

(62) Cai, T. T.; Wang, L. Orthogonal Matching Pursuit for Sparse Signal Recovery With Noise. *IEEE Trans. Inf. Theory* **2011**, *57* (7), 4680–4688. https://doi.org/10.1109/tit.2011.2146090.
(63) Rubinstein, R.; Zibulevsky, M.; Elad, M. Efficient Implementation of the K-SVD Algorithm Using Batch Orthogonal Matching Pursuit; 2008.
(64) Gramatica, P. Chemometric Methods and Theoretical Molecular Descriptors in Predictive QSAR Modeling of the Environmental Behavior of Organic Pollutants. In *Recent Advances in QSAR Studies: Methods and Applications*; Puzyn, T., Leszczynski, J., Cronin, M. T., Eds.; Springer Netherlands: Dordrecht, 2010; pp 327–366. https://doi.org/10.1007/978-1-4020-9783-6_12.
(65) Hoaglin, D. C.; Welsch, R. E. The Hat Matrix in Regression and ANOVA. *Am. Stat.* **1978**, *32* (1), 17–22. https://doi.org/10.1080/00031305.1978.10479237.
(66) Belsley, D. A.; Kuh, E.; Welsch, R. E. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*; John Wiley & Sons, 2005.
(67) Cook, R. D. Detection of Influential Observation in Linear Regression. *Technometrics* **1977**, *19* (1), 15–18. https://doi.org/10.2307/1268249.
(68) Mclachlan, G. Mahalanobis Distance. *Resonance* **1999**, *4*, 20–26. https://doi.org/10.1007/BF02834632.
(69) Fechner, N.; Hinselmann, G.; Jahn, A.; Zell, A. Kernel-Based Estimation of the Applicability Domain of QSAR Models. *J. Cheminformatics* **2010**, *2* (1), P38. https://doi.org/10.1186/1758-2946-2-S1-P38.
(70) Tian, Y.; Zhang, S.; Yin, H.; Yan, A. Quantitative Structure-Activity Relationship (QSAR) Models and Their Applicability Domain Analysis on HIV-1 Protease Inhibitors by Machine Learning Methods. *Chemom. Intell. Lab. Syst.* **2020**, *196*, 103888. https://doi.org/10.1016/j.chemolab.2019.103888.
(71) Schultz, L. E.; Wang, Y.; Jacobs, R.; Morgan, D. A General Approach for Determining Applicability Domain of Machine Learning Models. *Npj Comput. Mater.* **2025**, *11* (1), 95. https://doi.org/10.1038/s41524-025-01573-x.
(72) Barbierato, E.; Gatti, A.; Incremona, A.; Pozzi, A.; Toti, D. Breaking Away From AI: The Ontological and Ethical Evolution of Machine Learning. *IEEE Access* **2025**, *13*, 55627–55647. https://doi.org/10.1109/ACCESS.2025.3553032.
(73) Flach, P. *Machine Learning: The Art and Science of Algorithms That Make Sense of Data*; Cambridge University Press: USA, 2012.
(74) Mohri, M.; Rostamizadeh, A.; Talwalkar, A. *Foundations of Machine Learning*; The MIT Press, 2012.
(75) Tropsha, A. Best Practices for QSAR Model Development, Validation, and Exploitation. *Mol. Inform.* **2010**, *29* (6-7), 476–488. https://doi.org/10.1002/minf.201000061.
(76) Jobson, J. D. Multiple Linear Regression. In *Springer Texts in Statistics*; Springer New York, 1991; pp 219–398. https://doi.org/10.1007/978-1-4612-0955-3_4.
(77) Kutner, M. H.; Neter, J. *Applied Linear Regression Models*; McGraw-Hill/Irwin, 2004.
(78) Dormann, C. F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; Marquéz, J. R. G.; Gruber, B.; Lafourcade, B.; Leitão, P. J.; Münkemüller, T.; McClean, C.; Osborne, P. E.; Reineking, B.; Schröder, B.; Skidmore, A. K.; Zurell, D.; Lautenbach, S. Collinearity: A Review of Methods to Deal with It and a Simulation Study Evaluating Their Performance. *Ecography* **2013**, *36* (1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x.
(79) Genuer, R.; Poggi, J.-M. Random Forests. In *Use R!*; Springer International Publishing, 2020; pp 33–55. https://doi.org/10.1007/978-3-030-56485-8_3.

(80) Louppe, G. Understanding Random Forests: From Theory to Practice. arXiv June 3, 2015. https://doi.org/10.48550/arXiv.1407.7502.
(81) Schölkopf, B.; Smola, A. J. A Tutorial Introduction. In *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; MIT Press, 2001; pp 1–22.
(82) Rodríguez-Pérez, R.; Bajorath, J. Evolution of Support Vector Machine and Regression Modeling in Chemoinformatics and Drug Discovery. *J. Comput. Aided Mol. Des.* **2022**, *36* (5), 355–362. https://doi.org/10.1007/s10822-022-00442-9.
(83) Altmann, A.; Toloşi, L.; Sander, O.; Lengauer, T. Permutation Importance: A Corrected Feature Importance Measure. *Bioinformatics* **2010**, *26* (10), 1340–1347. https://doi.org/10.1093/bioinformatics/btq134.
(84) Rasmussen, C. E.; Williams, C. K. I. *Gaussian Processes for Machine Learning*; The MIT Press, 2005. https://doi.org/10.7551/mitpress/3206.001.0001.
(85) Abraham, M. H. Scales of Solute Hydrogen-Bonding: Their Construction and Application to Physicochemical and Biochemical Processes. *Chem. Soc. Rev.* **1993**, *22* (2), 73. https://doi.org/10.1039/cs9932200073.
(86) Anderson, J. L.; Ding, J.; Welton, T.; Armstrong, D. W. Characterizing Ionic Liquids On the Basis of Multiple Solvation Interactions. *J. Am. Chem. Soc.* **2002**, *124* (47), 14247–14254. https://doi.org/10.1021/ja028156h.
(87) Anderson, J. L.; Armstrong, D. W. Immobilized Ionic Liquids as High-Selectivity/ High-Temperature/High-Stability Gas Chromatography Stationary Phases. *Anal. Chem.* **2005**, *77* (19), 6453–6462. https://doi.org/10.1021/ac051006f.
(88) Abraham, M. H.; William E. Acree, J. Comparative Analysis of Solvation and Selectivity in Room Temperature Ionic Liquids Using the Abraham Linear Free Energy Relationship. *Green Chem.* **2006**, *8* (10), 906–915. https://doi.org/10.1039/B606279B.
(89) William E Acree, J.; Abraham, M. H. The Analysis of Solvation in Ionic Liquids and Organic Solvents Using the Abraham Linear Free Energy Relationship. *J. Chem. Technol. Biotechnol.* **2006**, *81* (8), 1441–1446. https://doi.org/10.1002/jctb.1589.
(90) Katritzky, A. R.; Kuanar, M.; Stoyanova-Slavova, I. B.; Slavov, S. H.; Dobchev, D. A.; Karelson, M.; Acree, W. E. Quantitative Structure–Property Relationship Studies on Ostwald Solubility and Partition Coefficients of Organic Solutes in Ionic Liquids. *J. Chem. Eng. Data* **2008**, *53* (5), 1085–1092. https://doi.org/10.1021/je700607b.
(91) Katritzky, A. R.; Tatham, D. B.; Maran, U. Correlation of the Solubilities of Gases and Vapors in Methanol and Ethanol with Their Molecular Structures. *J. Chem. Inf. Comput. Sci.* **2001**, *41* (2), 358–363. https://doi.org/10.1021/ci000124v.
(92) Katritzky, A. R.; Oliferenko, A. A.; Oliferenko, P. V.; Petrukhin, R.; Tatham, D. B.; Maran, U.; Lomaka, A.; Acree, W. E. A General Treatment of Solubility. 1. The QSPR Correlation of Solvation Free Energies of Single Solutes in Series of Solvents. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (6), 1794–1805. https://doi.org/10.1021/ci034120c.
(93) Katritzky, A. R.; Oliferenko, A. A.; Oliferenko, P. V.; Petrukhin, R.; Tatham, D. B.; Maran, U.; Lomaka, A.; Acree, W. E. A General Treatment of Solubility. 2. QSPR Prediction of Free Energies of Solvation of Specified Solutes in Ranges of Solvents. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (6), 1806–1814. https://doi.org/10.1021/ci034122x.
(94) Katritzky, A. R.; Tulp, I.; Fara, D. C.; Lauria, A.; Maran, U.; Acree, W. E. A General Treatment of Solubility. 3. Principal Component Analysis (PCA) of the

Solubilities of Diverse Solutes in Diverse Solvents. *J. Chem. Inf. Model.* **2005**, *45* (4), 913–923. https://doi.org/10.1021/ci0496189.

(95) Tulp, I.; Dobchev, D. A.; Katritzky, A. R.; Acree, W.; Maran, U. A General Treatment of Solubility 4. Description and Analysis of a PCA Model for Ostwald Solubility Coefficients. *J. Chem. Inf. Model.* **2010**, *50* (7), 1275–1283. https://doi.org/10.1021/ci1000828.

(96) Sprunger, L.; Clark, M.; Acree, W. E.; Abraham, M. H. Characterization of Room-Temperature Ionic Liquids by the Abraham Model with Cation-Specific and Anion-Specific Equation Coefficients. *J. Chem. Inf. Model.* **2007**, *47* (3), 1123–1129. https://doi.org/10.1021/ci7000428.

(97) Sprunger, L. M.; Proctor, A.; Acree, W. E.; Abraham, M. H. LFER Correlations for Room Temperature Ionic Liquids: Separation of Equation Coefficients into Individual Cation-Specific and Anion-Specific Contributions. *Fluid Phase Equilibria* **2008**, *265* (1–2), 104–111. https://doi.org/10.1016/j.fluid.2008.01.006.

(98) Sprunger, L. M.; Gibbs, J.; Proctor, A.; William E. Acree, Jr.; Abraham, M. H.; Meng, Y.; Yao, C.; Anderson, J. L. Linear Free Energy Relationship Correlations for Room Temperature Ionic Liquids: Revised Cation-Specific and Anion-Specific Equation Coefficients for Predictive Applications Covering a Much Larger Area of Chemical Space. *Ind. Eng. Chem. Res.* **2009**, *48* (8), 4145–4154. https://doi.org/10.1021/ie801898j.

(99) Revelli, A.-L.; Mutelet, F.; Jaubert, J.-N. Prediction of Partition Coefficients of Organic Compounds in Ionic Liquids: Use of a Linear Solvation Energy Relationship with Parameters Calculated through a Group Contribution Method. *Ind. Eng. Chem. Res.* **2010**, *49* (8), 3883–3892. https://doi.org/10.1021/ie901776z.

(100) Jiang, R.; Anderson, J.; Stephens, T.; Acree, W.; Abraham, M. H. Abraham Model Correlations for Predicting Gas-to-Liquid Partition Coefficients and Activity Coefficients of Organic Solutes Dissolved in 1-(2-Methoxyethyl)-1-Methylpyrrolidnium Tris(Pentafluoroethyl)Trifluorophosphate. *Eur Chem Bull* **2013**, *2*, 741–751.

(101) Stephens, T. W.; Chou, V.; Quay, A. N.; Shen, C.; Dabadge, N.; Tian, A.; Loera, M.; Willis, B.; Wilson, A.; Acree, W. E.; Twu, P.; Anderson, J. L.; Abraham, M. H. Thermochemical Investigations of Solute Transfer into Ionic Liquid Solvents: Updated Abraham Model Equation Coefficients for Solute Activity Coefficient and Partition Coefficient Predictions. *Phys. Chem. Liq.* **2014**, *52* (4), 488–518. https://doi.org/10.1080/00319104.2014.880114.

(102) Stephens, T. W.; Hart, E.; Kuprasertkul, N.; Mehta, S.; Wadawadigi, A.; Acree, W. E.; Abraham, M. H. Abraham Model Correlations for Describing Solute Transfer into Ionic Liquid Solvents: Calculation of Ion-Specific Equation Coefficients for the 4,5-Dicyano-2-(Trifluoromethyl)Imidazolide Anion. *Phys. Chem. Liq.* **2014**, *52* (6), 777–791. https://doi.org/10.1080/00319104.2014.929949.

(103) Acree, W. E.; Jiang, B. Abraham Model Correlations for Ionic Liquid Solvents: Computational Methodology for Updating Existing Ion-Specific Equation Coefficients. *Phys. Chem. Liq.* **2016**, *55* (4), 457–462. https://doi.org/10.1080/00319104.2016.1218878.

(104) Jiang, B.; Horton, M. Y.; Acree, W. E.; Abraham, M. H. Ion-Specific Equation Coefficient Version of the Abraham Model for Ionic Liquid Solvents: Determination of Coefficients for Tributylethylphosphonium, 1-Butyl-1-Methylmorpholinium, 1-Allyl-3-Methylimidazolium and Octyltriethylammonium Cations. *Phys. Chem. Liq.* **2016**, *55* (3), 358–385. https://doi.org/10.1080/00319104.2016.1218009.

(105) Lu, A.; Jiang, B.; Cheeran, S.; Acree, W. E.; Abraham, M. H. Abraham Model Ion-Specific Equation Coefficients for the 1-Butyl-2,3-Dimethyimidazolium and 4-Cyano-1-Butylpyridinium Cations Calculated from Measured Gas-to-Liquid Partition Coefficient Data. *Phys. Chem. Liq.* **2017**, *55* (2), 218–237. https://doi.org/10.1080/00319104.2016.1191634.
(106) Yue, D.; Acree, W. E.; Abraham, M. H. Development of Abraham Model IL-Specific Correlations for N-Triethyl(Octyl)Ammonium Bis(Fluorosulfonyl)Imide and 1-Butyl-3-Methylpyrrolidinium Bis(Fluorosulfonyl)Imide. *Phys. Chem. Liq.* **2018**, *57* (6), 733–745. https://doi.org/10.1080/00319104.2018.1519713.
(107) Mutelet, F.; Baker, G. A.; Zhao, H.; Churchill, B.; Acree, W. E. Development of Abraham Model Correlations for Short-Chain Glycol-Grafted Imidazolium and Pyridinium Ionic Liquids from Inverse Gas-Chromatographic Measurements. *J. Mol. Liq.* **2020**, *317*, 113983. https://doi.org/10.1016/j.molliq.2020.113983.
(108) Churchill, B.; Casillas, T.; Acree, W. E.; Abraham, M. H. Abraham Solvation Parameter Model: Calculation of Ion-Specific Equation Coefficients for the N-Ethyl-N-Methylmorpholinium and N-Octyl-N-Methylmorpholinium Cations. *Phys. Chem. Liq.* **2020**, *0* (0), 1–10. https://doi.org/10.1080/00319104.2020.1774879.
(109) Dashtbozorgi, Z.; Golmohammadi, H.; Acree, W. E. Prediction of Gas-to-Ionic Liquid Partition Coefficient of Organic Solutes Dissolved in 1-(2-Methoxyethyl)-1-Methylpyrrolidinium *Tris*(Pentafluoroethyl)Trifluorophosphate Using QSPR Approaches. *J. Mol. Liq.* **2015**, *201*, 21–29. https://doi.org/10.1016/j.molliq.2014.11.025.
(110) Khooshechin, S.; Dashtbozorgi, Z.; Golmohammadi, H.; Acree, W. E. QSPR Prediction of Gas-to-Ionic Liquid Partition Coefficient of Organic Solutes Dissolved in 1-(2-Hydroxyethyl)-1-Methylimidazolium Tris(Pentafluoroethyl)Trifluorophosphate Using the Replacement Method and Support Vector Regression. *J. Mol. Liq.* **2014**, *196*, 43–51. https://doi.org/10.1016/j.molliq.2014.03.012.
(111) Yusuf, F.; Olayiwola, T.; Afagwu, C. Application of Artificial Intelligence-Based Predictive Methods in Ionic Liquid Studies: A Review. *FLUID PHASE EQUILIBRIA* **2021**, *531*, 112898. https://doi.org/10.1016/j.fluid.2020.112898.
(112) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12* (85), 2825–2830.
(113) Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*; KDD'96; AAAI Press: Portland, Oregon, 1996; pp 226–231.
(114) Karelson, M.; Diercksen, G. H. F. Models for Simulating Molecular Properties in Condensed Systems. In *Problem Solving in Computational Molecular Science: Molecules in Different Environments*; Wilson, S., Diercksen, G. H. F., Eds.; Springer Netherlands: Dordrecht, 1997; pp 215–248. https://doi.org/10.1007/978-94-009-0039-4_7.
(115) Karelson, M. Quantum Chemical Treatment of Molecules in Condensed Disordered Media. In *Advances in Quantum Chemistry*; Löwdin, P.-O., Sabin, J. R., Zerner, M. C., Karwowski, J., Karelson, M., Eds.; Academic Press, 1997; Vol. 28, pp 141–157. https://doi.org/10.1016/S0065-3276(08)60212-9.
(116) Karelson, M. Molecular Properties and Spectra in Solution. In *Problem Solving in Computational Molecular Science: Molecules in Different Environments*; Wilson, S., Diercksen, G. H. F., Eds.; Springer Netherlands: Dordrecht, 1997; pp 353–387. https://doi.org/10.1007/978-94-009-0039-4_10.

(117) Moriwaki, H.; Tian, Y.-S.; Kawashita, N.; Takagi, T. Mordred: A Molecular Descriptor Calculator. *J. Cheminformatics* **2018**, *10* (1). https://doi.org/10.1186/s13321-018-0258-y.
(118) Piir, G.; Kahn, I.; García-Sosa, A. T.; Sild, S.; Ahte, P.; Maran, U. Best Practices for QSAR Model Reporting: Physical and Chemical Properties, Ecotoxicity, Environmental Fate, Human Health, and Toxicokinetics Endpoints. *Environ. Health Perspect.* **2018**, *126* (12), 126001. https://doi.org/10.1289/EHP3264.
(119) Sild, S.; Piir, G.; Neagu, D.; Maran, U. CHAPTER 6:Storing and Using Qualitative and Quantitative Structure–Activity Relationships in the Era of Toxicological and Chemical Data Expansion. In *Big Data in Predictive Toxicology*; 2019; pp 185–213. https://doi.org/10.1039/9781782623656-00185.
(120) Guazzelli, A.; Zeller, M.; Lin, W.-C.; Williams, G. PMML: An Open Standard for Sharing Models. *R J.* **2009**, *1* (1), 60. https://doi.org/10.32614/RJ-2009-010.
(121) Ruusmann, V.; Sild, S.; Maran, U. QSAR DataBank – an Approach for the Digital Organization and Archiving of QSAR Model Information. *J. Cheminformatics* **2014**, *6* (1), 25. https://doi.org/10.1186/1758-2946-6-25.
(122) Ruusmann, V.; Sild, S.; Maran, U. QSAR DataBank Repository: Open and Linked Qualitative and Quantitative Structure–Activity Relationship Models. *J. Cheminformatics* **2015**, *7* (1), 32. https://doi.org/10.1186/s13321-015-0082-6.
(123) Wilkinson, M. D.; Dumontier, M.; Aalbersberg, Ij. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L. B.; Bourne, P. E.; Bouwman, J.; Brookes, A. J.; Clark, T.; Crosas, M.; Dillo, I.; Dumon, O.; Edmunds, S.; Evelo, C. T.; Finkers, R.; Gonzalez-Beltran, A.; Gray, A. J. G.; Groth, P.; Goble, C.; Grethe, J. S.; Heringa, J.; 't Hoen, P. A. C.; Hooft, R.; Kuhn, T.; Kok, R.; Kok, J.; Lusher, S. J.; Martone, M. E.; Mons, A.; Packer, A. L.; Persson, B.; Rocca-Serra, P.; Roos, M.; van Schaik, R.; Sansone, S.-A.; Schultes, E.; Sengstag, T.; Slater, T.; Strawn, G.; Swertz, M. A.; Thompson, M.; van der Lei, J.; van Mulligen, E.; Velterop, J.; Waagmeester, A.; Wittenburg, P.; Wolstencroft, K.; Zhao, J.; Mons, B. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3* (1), 160018. https://doi.org/10.1038/sdata.2016.18.
(124) Toots, K. M.; Sild, S.; Leis, J.; Maran, U. Data for: The Quantitative Structure-Property Relationships for the Gas-Ionic Liquid Partition Coefficient of a Large Variety of Organic Compounds in Three Ionic Liquids. QsarDB Repository, QDB., 2021. https://doi.org/10.15152/QDB.241.
(125) Toots, K. M. Data for: Machine Learning Quantitative Structure-Property Relationships as a Function of Ionic Liquid Cations for the Gas-Ionic Liquid Partition Coefficient of Hydrocarbons.. QsarDB Repository, QDB., 2022. https://doi.org/10.15152/QDB.256.
(126) Toots, K. M.; Sild, S.; Leis, J.; Acree, W. E.; Maran, U. Data for: Exploring the Influence of Ionic Liquid Anion Structure on Gas-Ionic Liquid Partition Coefficients of Organic Solutes Using Machine Learning. QsarDB Repository, QDB., 2024. https://doi.org/10.15152/QDB.262.
(127) Toots, K. M.; Sild, S.; Leis, J.; Acree, W. E.; Maran, U. Data for: A Multicomponent QSPR Approach to Describe and Predict Gas-Ionic Liquid Distribution of Organic Solutes Using Machine Learning. QsarDB Repository, QDB., 2025. https://doi.org/10.15152/QDB.266.

SUMMARY IN ESTONIAN

Keemiainformaatika lähenemised orgaaniliste lahustunud ainete gaasi-ioonse vedeliku jaotumise analüüsis ja modelleerimisel

Selles dissertatsioonis kirjeldatud andmepõhised arvutuslikud uuringud aitavad mõista orgaaniliste ainete jaotumist ioonvedelikes, seda kirjeldavat gaas-ioonvedelik jaotus koefitsienti, rakendades põhjalikku ja mitmekesist kvantitatiivsete struktuur-omadus sõltuvuste (QSPR) modelleerimis meetodikat. Töö käsitleb orgaaniliste ühendite gaas-ioonvedelik jaotuskoefitsientide ja nendele vastavate süsteemide modelleerimist valitud andmeseeriates, lähtudes orgaaniliste ühendite, ja ioonvedelike katioonide ja anioonide struktuurist. Ühe tulemusena näidati, et masinõppe meetodid, eriti juhumets (RF), tugivektor regression (SVR) ja gaussian protsessi regression (GPR), kirjeldavad lahustunud aine ja ioonvedeliku interaktsioone enamikel juhtudel tõhusamalt kui sagedasti kasutatud multilineaarne regressioon (MLR). Sellegipoolest annab MLR-lähenemine väärtuslikku tõlgenduslikku teavet, eriti vesiniksideme, kuloniliste-dipolaarsete interaktsioonide ja dispersiooni jõudude osas. Nii lineaarsed kui ka mittelineaarsed mudelid rõhutavad katioonide ja anioonide keemilise struktuuri olulist mõju lahustunud ainete jaotumisele, demonstreerides ioonvedelike loodud lahustava keskkonna keerukust.

Kokkuvõttes näitavad doktoritöö tulemused, et ühised struktuurielemendid võivad süstemaatiliselt tugevdada või nõrgendada jaotuskoefitsienti, sõltuvalt dispersioonijõudude, kuloniliste-dipolaarsete interaktsioonide ja vesiniksideme interaktsioonide tasakaalust. Mitmes publikatsioonis demonstreeriti, et nii ioonvedeliku katioonse kui ka anioonse komponendi valik on log *K* kirjeldamise ja hindamise osas olulised parameetrid. Uurimistulemused näitavad lisaks, et kogu süsteemi modelleerimine, kombineerides suure andmehulga korral keemiliselt mitmekesiste lahustunud aine, katiooni ja aniooni molekulaar-tunnuseid, parandab mudeli ennustusvõimet, mis omakorda rõhutab ka mitmekomponentse lähenemisviisi olulisust. Eksperimentaalsete andmete range kureerimine ja põhjalik valideerimine, kasutades rist-valideerimist, välist valideerimist ja mõjuanalüüsi, tugevdavad iga mudeli usaldusväärsust ja rakendatavust. Uurides iga struktuurse komponendi mõju lahustunud aine ja ioonvedeliku mitmekomponentse süsteemi kontekstis, pakub töö mehhanistlikku ülevaadet molekulidevahelistest vastasmõjudest, ning võimaldab kujundada sobivama ja tõhusama ioonvedelik-keskkonna spetsiifiliste tööstuslike ja uuringuliste rakenduste jaoks.

PUBLICATION

CURRICULUM VITAE

Name: Karl Marti Toots
Date of birth: December 6, 1995
Email: karlmarti65@gmail.com

Education (institution, graduation date, qualification):

2020 materials science M.Sc., University of Tartu
2018 materials science B.Sc., University of Tartu
2015 secondary education, Pärnu Koidula Gymnasium

Career:

2022– Software Developer Tech Lead, Concise Systems
2019–2022 Software Engineer, Playtech
2018–2019 Materials Engineer, Gelatex Technologies
2018–2019 Physics Teacher, Tartu Tamme Gymnasium
2018–2019 Science Workshop Instructor, Spark Makerlab
2017–2017 Lab Intern, Estiko-Plastar

Scientific Activities and Developments

Main areas of research

Science and technology, molecular technology

Patents

Järvekülg, M.; Enn R.; Toots K. M.; Veinla M. Device and method for producing polymer fibres and its uses thereof. EP20190897A, 2020.

Scientific articles in internationally distributed journals

1. Toots, K. M.; Sild, S.; Leis, J.; Acree Jr., W. E.; Maran, U. The Quantitative Structure-Property Relationships for the Gas-Ionic Liquid Partition Coefficient of a Large Variety of Organic Compounds in Three Ionic Liquids. J. Mol. Liq. **2021**, 343, 117573. https://doi.org/10.1016/j.molliq.2021.117573.
2. Toots, K. M.; Sild, S.; Leis, J.; Acree, W. E.; Maran, U. Machine Learning Quantitative Structure-Property Relationships as a Function of Ionic Liquid Cations for the Gas-Ionic Liquid Partition Coefficient of Hydrocarbons. *Int. J. Mol. Sci.* **2022**, *23* (14), 7534. https://doi.org/10.3390/ijms23147534.
3. Toots, K. M.; Sild, S.; Leis, J.; Acree, W. E.; Maran, U. Exploring the Influence of Ionic Liquid Anion Structure on Gas-Ionic Liquid Partition Coefficients of Organic Solutes Using Machine Learning. *Langmuir* **2024**, *40* (45), 23714–23728. https://doi.org/10.1021/acs.langmuir.4c02628.
4. Toots, K. M.; Sild, S.; Leis, J.; Acree Jr., W. E.; Maran, U. A Multicomponent QSPR Approach to Describe and Predict the Gas-Ionic Liquid Distribution of Organic Solutes using Machine Learning. *J. Mol. Liq*. **2025**, *436*, 128184. https://doi.org/10.1016/j.molliq.2025.128184.

Other scientific publications (models and data)

5. Toots, K. M.; Sild, S.; Leis, J.; Acree Jr., W. E.; Maran, U. Data for: The Quantitative Structure-Property Relationships for the gas-ionic liquid partition coefficient of a large variety of organic compounds in three ionic liquids. QsarDB repository, QDB.241. 2021. https://doi.org/10.15152/QDB.241
6. Toots, K. M.; Sild, S.; Leis, J.; Acree Jr., W. E.; Maran, U. Data for: Machine learning Quantitative Structure-Property Relationships as a function of ionic liquid cations for the gas-ionic liquid partition coefficient of hydrocarbons. QsarDB repository, QDB.256. 2022. https://doi.org/10.15152/QDB.256
7. Toots, K. M.; Sild, S.; Leis, J.; Acree, W. E.; Maran, U. Data for: Exploring the influence of ionic liquid anion structure on gas-ionic liquid partition coefficients of organic solutes using machine learning. QsarDB repository, QDB.262. 2024. https://doi.org/10.15152/QDB.262
8. Toots, K. M. Sild, S.; Leis, J.; Acree, W. E.; Maran, U. Data for: A multi-component QSPR approach to describe and predict gas-ionic liquid distribution of organic solutes using machine learning. QsarDB repository, QDB.266. 2025. https://doi.org/10.15152/QDB.266

Conference thesis

9. Karl Marti Toots, Sulev Sild, Jaan Leis, William E. Acree, Uko Maran. Partitioning and solute-solvent interactions of organic compounds in ionic liquids predicted from molecular structure: comparative analysis of data driven methods., eSSENCE-EMMC e-Meeting 2021, Rootsi, Uppsala, 7.–8. juuni 2021
10. Karl Marti Toots, Sulev Sild, Jaan Leis, William E. Acree, Uko Maran. Machine learning Quantitative Structure-Property Relationships as a function of ionic liquid cations for the gas-ionic liquid partition coefficient of hydrocarbons., Graduate School of Functional Materials and Technologies Scientific Conference 2022, Eesti, Tallinn, 17.–18. mai 2022

Research grants and scholarships

Research staff in "Center of Excellence in Molecular Cell Engineering (TK143)" (01.01.2016−01.03.2023); Principal Investigator: Tanel Tenson; University of Tartu, Faculty of Science and Technology, Institute of Technology; Financier: Archimedes Foundation.

Research staff in "Data driven modelling of physical-chemical-biological properties of complex molecular environments and functional materials (PRG1509)" (01.01.2022−31.12.2026); Principal Investigator: Uko Maran; University of Tartu, Faculty of Science and Technology, Institute of Chemistry; Financier: Estonian Research Council.

Other scientific, organizational and professional activities

Teaching in University:
Chemical Principles (LOKT.07.010) in 2016
Chemical Principles (LOKT.07.010) in 2017
Theoretical Chemistry (LOKT.08.001) in 2021
Theoretical Chemistry (LOKT.08.001) in 2022

ELULOOKIRJELDUS

Nimi: Karl Marti Toots
Sünniaeg: 6. detsember 1995
E-post: karlmarti65@gmail.com

Haridus (õppeasutus, lõpetamise aeg, kvalifikatsioon (kraad)):
2020 materjaliteadus M.Sc., Tartu Ülikool
2018 materjaliteadus B.Sc., Tartu Ülikool
2015 keskharidus, Pärnu Koidula Gümnaasium

Teenistuskäik:
2022– Tarkvaraarendusjuht, Concise Systems
2019–2022 Tarkvaraarendaja, Playtech
2018–2019 Materjali insener, Gelatex Technologies
2018–2019 Füüsika õpetaja, Tartu Tamme Gümnaasium
2018–2019 Teadusteemaliste töötubade juhendaja, Spark Makerlab
2017–2017 Laboripraktikant, Estiko-Plastar

Teaduslik ja arendustegevus
Peamised uurimisvaldkonnad
Loodus- ja täppisteaduste valdkond, molekulaartehnoloogia

Patendid
1. Järvekülg, M.; Enn R.; Toots K. M.; Veinla M. Device and method for producing polymer fibers and its uses thereof. EP20190897A, 2020.

Rahvusvahelise levikuga publikatsioonide loetelu
2. Toots, K. M.; Sild, S.; Leis, J.; Acree Jr., W. E.; Maran, U. The Quantitative Structure-Property Relationships for the Gas-Ionic Liquid Partition Coefficient of a Large Variety of Organic Compounds in Three Ionic Liquids. J. Mol. Liq. **2021**, 343, 117573. https://doi.org/10.1016/j.molliq.2021.117573.
3. Toots, K. M.; Sild, S.; Leis, J.; Acree, W. E.; Maran, U. Machine Learning Quantitative Structure-Property Relationships as a Function of Ionic Liquid Cations for the Gas-Ionic Liquid Partition Coefficient of Hydrocarbons. *Int. J. Mol. Sci.* **2022**, *23* (14), 7534. https://doi.org/10.3390/ijms23147534.
4. Toots, K. M.; Sild, S.; Leis, J.; Acree, W. E.; Maran, U. Exploring the Influence of Ionic Liquid Anion Structure on Gas-Ionic Liquid Partition Coefficients of Organic Solutes Using Machine Learning. *Langmuir* **2024**, *40* (45), 23714–23728. https://doi.org/10.1021/acs.langmuir.4c02628.
5. Toots, K. M.; Sild, S.; Leis, J.; Acree Jr., W. E.; Maran, U. A Multicomponent QSPR Approach to Describe and Predict the Gas-Ionic Liquid Distribution of Organic Solutes using Machine Learning. *J. Mol. Liq*. **2025**, *436*, 128184. https://doi.org/10.1016/j.molliq.2025.128184.

Muud publikatsioonid (mudelid ja andmed)

6. Toots, K. M.; Sild, S.; Leis, J.; Acree Jr., W. E.; Maran, U. Data for: The Quantitative Structure-Property Relationships for the gas-ionic liquid partition coefficient of a large variety of organic compounds in three ionic liquids. QsarDB repository, QDB.241. 2021. https://doi.org/10.15152/QDB.241
7. Toots, K. M.; Sild, S.; Leis, J.; Acree Jr., W. E.; Maran, U. Data for: Machine learning Quantitative Structure-Property Relationships as a function of ionic liquid cations for the gas-ionic liquid partition coefficient of hydrocarbons. QsarDB repository, QDB.256. 2022. https://doi.org/10.15152/QDB.256
8. Toots, K. M.; Sild, S.; Leis, J.; Acree, W. E.; Maran, U. Data for: Exploring the influence of ionic liquid anion structure on gas-ionic liquid partition coefficients of organic solutes using machine learning. QsarDB repository, QDB.262. 2024. https://doi.org/10.15152/QDB.262
9. Toots, K. M.; Sild, S.; Leis, J.; Acree, W. E.; Maran, U. Data for: A multi-component QSPR approach to describe and predict gas-ionic liquid distribution of organic solutes using machine learning. QsarDB repository, QDB.266. 2025. https://doi.org/10.15152/QDB.266

Konverentside teesid

10. Karl Marti Toots, Sulev Sild, Jaan Leis, William E. Acree, Uko Maran. Partitioning and solute-solvent interactions of organic compounds in ionic liquids predicted from molecular structure: comparative analysis of data driven methods., eSSENCE-EMMC e-Meeting 2021, Rootsi, Uppsala, 7.–8. juuni 2021
11. Karl Marti Toots, Sulev Sild, Jaan Leis, William E. Acree, Uko Maran. Machine learning Quantitative Structure-Property Relationships as a function of ionic liquid cations for the gas-ionic liquid partition coefficient of hydrocarbons., Graduate School of Functional Materials and Technologies Scientific Conference 2022, Eesti, Tallinn, 17.–18. mai 2022

Uurimistoetused ja stipendiumid

Täitja, "Andmetepõhine keeruliste molekulaarsete keskkondade ja funktsionaalsete materjalide füüsikalis-keemiliste-bioloogiliste omaduste modelleerimine (PRG1509)" (01.01.2022−31.12.2026); Vastutav täitja: Uko Maran; Tartu Ülikool, Loodus- ja täppisteaduste valdkond, keemia instituut; Finantseerija: Sihtasutus Eesti Teadusagentuur.

Täitja, "Molekulaarse Rakutehnoloogia Tippkeskus (TK143)" (01.01.2016−01.03.2023); Vastutav täitja: Tanel Tenson; Tartu Ülikool, Loodus- ja täppisteaduste valdkond, tehnoloogiainstituut; Finantseerija: Sihtasutus Archimedes.

Muu teaduslik organisatsiooniline ja erialane tegevus

Õppetöö läbiviimine

Keemia alused (LOKT.07.010) aines 2016 aastal
Keemia alused (LOKT.07.010) aines 2017 aastal
Teoreetilise keemia (LOKT.08.001) aines 2021 aastal
Teoreetilise keemia (LOKT.08.001) aines 2022 aastal

DISSERTATIONES CHIMICAE UNIVERSITATIS TARTUENSIS

1. **Toomas Tamm.** Quantum-chemical simulation of solvent effects. Tartu, 1993, 110 p.
2. **Peeter Burk.** Theoretical study of gas-phase acid-base equilibria. Tartu, 1994, 96 p.
3. **Victor Lobanov.** Quantitative structure-property relationships in large descriptor spaces. Tartu, 1995, 135 p.
4. **Vahur Mäemets.** The ^{17}O and ^{1}H nuclear magnetic resonance study of H_2O in individual solvents and its charged clusters in aqueous solutions of electrolytes. Tartu, 1997, 140 p.
5. **Andrus Metsala.** Microcanonical rate constant in nonequilibrium distribution of vibrational energy and in restricted intramolecular vibrational energy redistribution on the basis of slater's theory of unimolecular reactions. Tartu, 1997, 150 p.
6. **Uko Maran.** Quantum-mechanical study of potential energy surfaces in different environments. Tartu, 1997, 137 p.
7. **Alar Jänes.** Adsorption of organic compounds on antimony, bismuth and cadmium electrodes. Tartu, 1998, 219 p.
8. **Kaido Tammeveski.** Oxygen electroreduction on thin platinum films and the electrochemical detection of superoxide anion. Tartu, 1998, 139 p.
9. **Ivo Leito.** Studies of Brønsted acid-base equilibria in water and non-aqueous media. Tartu, 1998, 101 p.
10. **Jaan Leis.** Conformational dynamics and equilibria in amides. Tartu, 1998, 131 p.
11. **Toonika Rinken.** The modelling of amperometric biosensors based on oxidoreductases. Tartu, 2000, 108 p.
12. **Dmitri Panov.** Partially solvated Grignard reagents. Tartu, 2000, 64 p.
13. **Kaja Orupõld.** Treatment and analysis of phenolic wastewater with microorganisms. Tartu, 2000, 123 p.
14. **Jüri Ivask.** Ion Chromatographic determination of major anions and cations in polar ice core. Tartu, 2000, 85 p.
15. **Lauri Vares.** Stereoselective Synthesis of Tetrahydrofuran and Tetrahydropyran Derivatives by Use of Asymmetric Horner-Wadsworth-Emmons and Ring Closure Reactions. Tartu, 2000, 184 p.
16. **Martin Lepiku.** Kinetic aspects of dopamine D_2 receptor interactions with specific ligands. Tartu, 2000, 81 p.
17. **Katrin Sak.** Some aspects of ligand specificity of P2Y receptors. Tartu, 2000, 106 p.
18. **Vello Pällin.** The role of solvation in the formation of iotsitch complexes. Tartu, 2001, 95 p.
19. **Katrin Kollist.** Interactions between polycyclic aromatic compounds and humic substances. Tartu, 2001, 93 p.

20. **Ivar Koppel.** Quantum chemical study of acidity of strong and superstrong Brønsted acids. Tartu, 2001, 104 p.
21. **Viljar Pihl.** The study of the substituent and solvent effects on the acidity of OH and CH acids. Tartu, 2001, 132 p.
22. **Natalia Palm.** Specification of the minimum, sufficient and significant set of descriptors for general description of solvent effects. Tartu, 2001, 134 p.
23. **Sulev Sild.** QSPR/QSAR approaches for complex molecular systems. Tartu, 2001, 134 p.
24. **Ruslan Petrukhin.** Industrial applications of the quantitative structure-property relationships. Tartu, 2001, 162 p.
25. **Boris V. Rogovoy.** Synthesis of (benzotriazolyl)carboximidamides and their application in relations with *N*- and *S*-nucleophyles. Tartu, 2002, 84 p.
26. **Koit Herodes.** Solvent effects on UV-vis absorption spectra of some solvatochromic substances in binary solvent mixtures: the preferential solvation model. Tartu, 2002, 102 p.
27. **Anti Perkson.** Synthesis and characterisation of nanostructured carbon. Tartu, 2002, 152 p.
28. **Ivari Kaljurand.** Self-consistent acidity scales of neutral and cationic Brønsted acids in acetonitrile and tetrahydrofuran. Tartu, 2003, 108 p.
29. **Karmen Lust.** Adsorption of anions on bismuth single crystal electrodes. Tartu, 2003, 128 p.
30. **Mare Piirsalu.** Substituent, temperature and solvent effects on the alkaline hydrolysis of substituted phenyl and alkyl esters of benzoic acid. Tartu, 2003, 156 p.
31. **Meeri Sassian.** Reactions of partially solvated Grignard reagents. Tartu, 2003, 78 p.
32. **Tarmo Tamm.** Quantum chemical modelling of polypyrrole. Tartu, 2003. 100 p.
33. **Erik Teinemaa.** The environmental fate of the particulate matter and organic pollutants from an oil shale power plant. Tartu, 2003. 102 p.
34. **Jaana Tammiku-Taul.** Quantum chemical study of the properties of Grignard reagents. Tartu, 2003. 120 p.
35. **Andre Lomaka.** Biomedical applications of predictive computational chemistry. Tartu, 2003. 132 p.
36. **Kostyantyn Kirichenko.** Benzotriazole – Mediated Carbon–Carbon Bond Formation. Tartu, 2003. 132 p.
37. **Gunnar Nurk.** Adsorption kinetics of some organic compounds on bismuth single crystal electrodes. Tartu, 2003, 170 p.
38. **Mati Arulepp.** Electrochemical characteristics of porous carbon materials and electrical double layer capacitors. Tartu, 2003, 196 p.
39. **Dan Cornel Fara.** QSPR modeling of complexation and distribution of organic compounds. Tartu, 2004, 126 p.
40. **Riina Mahlapuu.** Signalling of galanin and amyloid precursor protein through adenylate cyclase. Tartu, 2004, 124 p.

41. **Mihkel Kerikmäe.** Some luminescent materials for dosimetric applications and physical research. Tartu, 2004, 143 p.
42. **Jaanus Kruusma.** Determination of some important trace metal ions in human blood. Tartu, 2004, 115 p.
43. **Urmas Johanson.** Investigations of the electrochemical properties of polypyrrole modified electrodes. Tartu, 2004, 91 p.
44. **Kaido Sillar.** Computational study of the acid sites in zeolite ZSM-5. Tartu, 2004, 80 p.
45. **Aldo Oras.** Kinetic aspects of dATPαS interaction with $P2Y_1$ receptor. Tartu, 2004, 75 p.
46. **Erik Mölder.** Measurement of the oxygen mass transfer through the air-water interface. Tartu, 2005, 73 p.
47. **Thomas Thomberg.** The kinetics of electroreduction of peroxodisulfate anion on cadmium (0001) single crystal electrode. Tartu, 2005, 95 p.
48. **Olavi Loog.** Aspects of condensations of carbonyl compounds and their imine analogues. Tartu, 2005, 83 p.
49. **Siim Salmar.** Effect of ultrasound on ester hydrolysis in aqueous ethanol. Tartu, 2006, 73 p.
50. **Ain Uustare.** Modulation of signal transduction of heptahelical receptors by other receptors and G proteins. Tartu, 2006, 121 p.
51. **Sergei Yurchenko.** Determination of some carcinogenic contaminants in food. Tartu, 2006, 143 p.
52. **Kaido Tämm.** QSPR modeling of some properties of organic compounds. Tartu, 2006, 67 p.
53. **Olga Tšubrik.** New methods in the synthesis of multisubstituted hydrazines. Tartu. 2006, 183 p.
54. **Lilli Sooväli.** Spectrophotometric measurements and their uncertainty in chemical analysis and dissociation constant measurements. Tartu, 2006, 125 p.
55. **Eve Koort.** Uncertainty estimation of potentiometrically measured ph and pK_a values. Tartu, 2006, 139 p.
56. **Sergei Kopanchuk.** Regulation of ligand binding to melanocortin receptor subtypes. Tartu, 2006, 119 p.
57. **Silvar Kallip.** Surface structure of some bismuth and antimony single crystal electrodes. Tartu, 2006, 107 p.
58. **Kristjan Saal.** Surface silanization and its application in biomolecule coupling. Tartu, 2006, 77 p.
59. **Tanel Tätte.** High viscosity $Sn(OBu)_4$ oligomeric concentrates and their applications in technology. Tartu, 2006, 91 p.
60. **Dimitar Atanasov Dobchev**. Robust QSAR methods for the prediction of properties from molecular structure. Tartu, 2006, 118 p.
61. **Hannes Hagu**. Impact of ultrasound on hydrophobic interactions in solutions. Tartu, 2007, 81 p.
62. **Rutha Jäger.** Electroreduction of peroxodisulfate anion on bismuth electrodes. Tartu, 2007, 142 p.

63. **Kaido Viht.** Immobilizable bisubstrate-analogue inhibitors of basophilic protein kinases: development and application in biosensors. Tartu, 2007, 88 p.
64. **Eva-Ingrid Rõõm.** Acid-base equilibria in nonpolar media. Tartu, 2007, 156 p.
65. **Sven Tamp.** DFT study of the cesium cation containing complexes relevant to the cesium cation binding by the humic acids. Tartu, 2007, 102 p.
66. **Jaak Nerut.** Electroreduction of hexacyanoferrate(III) anion on Cadmium (0001) single crystal electrode. Tartu, 2007, 180 p.
67. **Lauri Jalukse.** Measurement uncertainty estimation in amperometric dissolved oxygen concentration measurement. Tartu, 2007, 112 p.
68. **Aime Lust.** Charge state of dopants and ordered clusters formation in CaF_2:Mn and CaF_2:Eu luminophors. Tartu, 2007, 100 p.
69. **Iiris Kahn**. Quantitative Structure-Activity Relationships of environmentally relevant properties. Tartu, 2007, 98 p.
70. **Mari Reinik.** Nitrates, nitrites, N-nitrosamines and polycyclic aromatic hydrocarbons in food: analytical methods, occurrence and dietary intake. Tartu, 2007, 172 p.
71. **Heili Kasuk.** Thermodynamic parameters and adsorption kinetics of organic compounds forming the compact adsorption layer at Bi single crystal electrodes. Tartu, 2007, 212 p.
72. **Erki Enkvist.** Synthesis of adenosine-peptide conjugates for biological applications. Tartu, 2007, 114 p.
73. **Svetoslav Hristov Slavov**. Biomedical applications of the QSAR approach. Tartu, 2007, 146 p.
74. **Eneli Härk.** Electroreduction of complex cations on electrochemically polished Bi(*hkl*) single crystal electrodes. Tartu, 2008, 158 p.
75. **Priit Möller.** Electrochemical characteristics of some cathodes for medium temperature solid oxide fuel cells, synthesized by solid state reaction technique. Tartu, 2008, 90 p.
76. **Signe Viggor.** Impact of biochemical parameters of genetically different pseudomonads at the degradation of phenolic compounds. Tartu, 2008, 122 p.
77. **Ave Sarapuu.** Electrochemical reduction of oxygen on quinone-modified carbon electrodes and on thin films of platinum and gold. Tartu, 2008, 134 p.
78. **Agnes Kütt.** Studies of acid-base equilibria in non-aqueous media. Tartu, 2008, 198 p.
79. **Rouvim Kadis.** Evaluation of measurement uncertainty in analytical chemistry: related concepts and some points of misinterpretation. Tartu, 2008, 118 p.
80. **Valter Reedo.** Elaboration of IVB group metal oxide structures and their possible applications. Tartu, 2008, 98 p.
81. **Aleksei Kuznetsov.** Allosteric effects in reactions catalyzed by the cAMP-dependent protein kinase catalytic subunit. Tartu, 2009, 133 p.

82. **Aleksei Bredihhin.** Use of mono- and polyanions in the synthesis of multisubstituted hydrazine derivatives. Tartu, 2009, 105 p.
83. **Anu Ploom.** Quantitative structure-reactivity analysis in organosilicon chemistry. Tartu, 2009, 99 p.
84. **Argo Vonk.** Determination of adenosine A_{2A}- and dopamine D_1 receptor-specific modulation of adenylate cyclase activity in rat striatum. Tartu, 2009, 129 p.
85. **Indrek Kivi.** Synthesis and electrochemical characterization of porous cathode materials for intermediate temperature solid oxide fuel cells. Tartu, 2009, 177 p.
86. **Jaanus Eskusson.** Synthesis and characterisation of diamond-like carbon thin films prepared by pulsed laser deposition method. Tartu, 2009, 117 p.
87. **Marko Lätt.** Carbide derived microporous carbon and electrical double layer capacitors. Tartu, 2009, 107 p.
88. **Vladimir Stepanov.** Slow conformational changes in dopamine transporter interaction with its ligands. Tartu, 2009, 103 p.
89. **Aleksander Trummal.** Computational Study of Structural and Solvent Effects on Acidities of Some Brønsted Acids. Tartu, 2009, 103 p.
90. **Eerold Vellemäe.** Applications of mischmetal in organic synthesis. Tartu, 2009, 93 p.
91. **Sven Parkel.** Ligand binding to 5-HT_{1A} receptors and its regulation by Mg^{2+} and Mn^{2+}. Tartu, 2010, 99 p.
92. **Signe Vahur.** Expanding the possibilities of ATR-FT-IR spectroscopy in determination of inorganic pigments. Tartu, 2010, 184 p.
93. **Tavo Romann**. Preparation and surface modification of bismuth thin film, porous, and microelectrodes. Tartu, 2010, 155 p.
94. **Nadežda Aleksejeva.** Electrocatalytic reduction of oxygen on carbon nanotube-based nanocomposite materials. Tartu, 2010, 147 p.
95. **Marko Kullapere.** Electrochemical properties of glassy carbon, nickel and gold electrodes modified with aryl groups. Tartu, 2010, 233 p.
96. **Liis Siinor.** Adsorption kinetics of ions at Bi single crystal planes from aqueous electrolyte solutions and room-temperature ionic liquids. Tartu, 2010, 101 p.
97. **Angela Vaasa.** Development of fluorescence-based kinetic and binding assays for characterization of protein kinases and their inhibitors. Tartu 2010, 101 p.
98. **Indrek Tulp.** Multivariate analysis of chemical and biological properties. Tartu 2010, 105 p.
99. **Aare Selberg.** Evaluation of environmental quality in Northern Estonia by the analysis of leachate. Tartu 2010, 117 p.
100. **Darja Lavõgina.** Development of protein kinase inhibitors based on adenosine analogue-oligoarginine conjugates. Tartu 2010, 248 p.
101. **Laura Herm.** Biochemistry of dopamine D_2 receptors and its association with motivated behaviour. Tartu 2010, 156 p.

102. **Terje Raudsepp.** Influence of dopant anions on the electrochemical properties of polypyrrole films. Tartu 2010, 112 p.
103. **Margus Marandi.** Electroformation of Polypyrrole Films: *In-situ* AFM and STM Study. Tartu 2011, 116 p.
104. **Kairi Kivirand.** Diamine oxidase-based biosensors: construction and working principles. Tartu, 2011, 140 p.
105. **Anneli Kruve.** Matrix effects in liquid-chromatography electrospray mass-spectrometry. Tartu, 2011, 156 p.
106. **Gary Urb.** Assessment of environmental impact of oil shale fly ash from PF and CFB combustion. Tartu, 2011, 108 p.
107. **Nikita Oskolkov.** A novel strategy for peptide-mediated cellular delivery and induction of endosomal escape. Tartu, 2011, 106 p.
108. **Dana Martin.** The QSPR/QSAR approach for the prediction of properties of fullerene derivatives. Tartu, 2011, 98 p.
109. **Säde Viirlaid.** Novel glutathione analogues and their antioxidant activity. Tartu, 2011, 106 p.
110. **Ülis Sõukand.** Simultaneous adsorption of Cd^{2+}, Ni^{2+}, and Pb^{2+} on peat. Tartu, 2011, 124 p.
111. **Lauri Lipping.** The acidity of strong and superstrong Brønsted acids, an outreach for the "limits of growth": a quantum chemical study. Tartu, 2011, 124 p.
112. **Heisi Kurig.** Electrical double-layer capacitors based on ionic liquids as electrolytes. Tartu, 2011, 146 p.
113. **Marje Kasari.** Bisubstrate luminescent probes, optical sensors and affinity adsorbents for measurement of active protein kinases in biological samples. Tartu, 2012, 126 p.
114. **Kalev Takkis.** Virtual screening of chemical databases for bioactive molecules. Tartu, 2012, 122 p.
115. **Ksenija Kisseljova**. Synthesis of aza-β^3-amino acid containing peptides and kinetic study of their phosphorylation by protein kinase A. Tartu, 2012, 104 p.
116. **Riin Rebane.** Advanced method development strategy for derivatization LC/ESI/MS. Tartu, 2012, 184 p.
117. **Vladislav Ivaništšev.** Double layer structure and adsorption kinetics of ions at metal electrodes in room temperature ionic liquids. Tartu, 2012, 128 p.
118. **Irja Helm.** High accuracy gravimetric Winkler method for determination of dissolved oxygen. Tartu, 2012, 139 p.
119. **Karin Kipper.** Fluoroalcohols as Components of LC-ESI-MS Eluents: Usage and Applications. Tartu, 2012, 164 p.
120. **Arno Ratas.** Energy storage and transfer in dosimetric luminescent materials. Tartu, 2012, 163 p.
121. **Reet Reinart-Okugbeni**. Assay systems for characterisation of subtype-selective binding and functional activity of ligands on dopamine receptors. Tartu, 2012, 159 p.

122. **Lauri Sikk.** Computational study of the Sonogashira cross-coupling reaction. Tartu, 2012, 81 p.
123. **Karita Raudkivi.** Neurochemical studies on inter-individual differences in affect-related behaviour of the laboratory rat. Tartu, 2012, 161 p.
124. **Indrek Saar.** Design of GalR2 subtype specific ligands: their role in depression-like behavior and feeding regulation. Tartu, 2013, 126 p.
125. **Ann Laheäär.** Electrochemical characterization of alkali metal salt based non-aqueous electrolytes for supercapacitors. Tartu, 2013, 127 p.
126. **Kerli Tõnurist.** Influence of electrospun separator materials properties on electrochemical performance of electrical double-layer capacitors. Tartu, 2013, 147 p.
127. **Kaija Põhako-Esko.** Novel organic and inorganic ionogels: preparation and characterization. Tartu, 2013, 124 p.
128. **Ivar Kruusenberg.** Electroreduction of oxygen on carbon nanomaterial-based catalysts. Tartu, 2013, 191 p.
129. **Sander Piiskop.** Kinetic effects of ultrasound in aqueous acetonitrile solutions. Tartu, 2013, 95 p.
130. **Ilona Faustova**. Regulatory role of L-type pyruvate kinase N-terminal domain. Tartu, 2013, 109 p.
131. **Kadi Tamm.** Synthesis and characterization of the micro-mesoporous anode materials and testing of the medium temperature solid oxide fuel cell single cells. Tartu, 2013, 138 p.
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 Mn^{2+}. 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 signalinga. 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, 137 p.
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 nortropane 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.
178. **Sofja Tšepelevitš.** Experimental studies and modeling of solute-solvent interactions. Tartu, 2018, 109 p.
179. **Märt Lõkov.** Basicity of some nitrogen, phosphorus and carbon bases in acetonitrile. Tartu, 2018, 104 p.
180. **Anton Mastitski**. Preparation of α-aza-amino acid precursors and related compounds by novel methods of reductive one-pot alkylation and direct alkylation. Tartu, 2018, 155 p.
181. **Jürgen Vahter**. Development of bisubstrate inhibitors for protein kinase CK2. Tartu, 2019, 186 p.

182. **Piia Liigand.** Expanding and improving methodology and applications of ionization efficiency measurements. Tartu, 2019, 189 p.
183. **Sigrid Selberg.** Synthesis and properties of lipophilic phosphazene-based indicator molecules. Tartu, 2019, 74 p.
184. **Jaanus Liigand.** Standard substance free quantification for LC/ESI/MS analysis based on the predicted ionization efficiencies. Tartu, 2019, 254 p.
185. **Marek Mooste.** Surface and electrochemical characterisation of aryl film and nanocomposite material modified carbon and metal-based electrodes. Tartu, 2019, 304 p.
186. **Mare Oja.** Experimental investigation and modelling of pH profiles for effective membrane permeability of drug substances. Tartu, 2019, 306 p.
187. **Sajid Hussain.** Electrochemical reduction of oxygen on supported Pt catalysts. Tartu, 2019, 220 p.
188. **Ronald Väli.** Glucose-derived hard carbon electrode materials for sodium-ion batteries. Tartu, 2019, 180 p.
189. **Ester Tee.** Analysis and development of selective synthesis methods of hierarchical micro- and mesoporous carbons. Tartu, 2019, 210 p.
190. **Martin Maide.** Influence of the microstructure and chemical composition of the fuel electrode on the electrochemical performance of reversible solid oxide fuel cell. Tartu, 2020, 144 p.
191. **Edith Viirlaid.** Biosensing Pesticides in Water Samples. Tartu, 2020, 102 p.
192. **Maike Käärik.** Nanoporous carbon: the controlled nanostructure, and structure-property relationships. Tartu, 2020, 162 p.
193. **Artur Gornischeff.** Study of ionization efficiencies for derivatized compounds in LC/ESI/MS and their application for targeted analysis. Tartu, 2020, 124 p.
194. **Reet Link.** Ligand binding, allosteric modulation and constitutive activity of melanocortin-4 receptors. Tartu, 2020, 108 p.
195. **Pilleriin Peets.** Development of instrumental methods for the analysis of textile fibres and dyes. Tartu, 2020, 150 p.
196. **Larisa Ivanova.** Design of active compounds against neurodegenerative diseases. Tartu, 2020, 152 p.
197. **Meelis Härmas.** Impact of activated carbon microstructure and porosity on electrochemical performance of electrical double-layer capacitors. Tartu, 2020, 122 p.
198. **Ruta Hecht.** Novel Eluent Additives for LC-MS Based Bioanalytical Methods. Tartu, 2020, 202 p.
199. **Max Hecht.** Advances in the Development of a Point-of-Care Mass Spectrometer Test. Tartu, 2020, 168 p.
200. **Ida Rahu.** Bromine formation in inorganic bromide/nitrate mixtures and its application for oxidative aromatic bromination. Tartu, 2020, 116 p.
201. **Sander Ratso**. Electrocatalysis of oxygen reduction on non-precious metal catalysts. Tartu, 2020, 371 p.
202. **Astrid Darnell.** Computational design of anion receptors and evaluation of host-guest binding. Tartu, 2021, 150 p.

203. **Ove Korjus.** The development of ceramic fuel electrode for solid oxide cells. Tartu, 2021, 150 p.
204. **Merit Oss.** Ionization efficiency in electrospray ionization source and its relations to compounds' physico-chemical properties. Tartu, 2021, 124 p.
205. **Madis Lüsi**. Electroreduction of oxygen on nanostructured palladium catalysts. Tartu, 2021, 180 p.
206. **Eliise Tammekivi.** Derivatization and quantitative gas-chromatographic analysis of oils. Tartu, 2021, 122 p.
207. **Simona Selberg.** Development of Small-Molecule Regulators of Epi-transcriptomic Processes. Tartu, 2021, 122 p.
208. **Olivier Etebe Nonga.** Inhibitors and photoluminescent probes for in vitro studies on protein kinases PKA and PIM. Tartu, 2021, 189 p.
209. **Riinu Härmas.** The structure and H2 diffusion in porous carbide-derived carbon particles. Tartu, 2022, 123 p.
210. **Maarja Paalo.** Synthesis and characterization of novel carbon electrodes for high power density electrochemical capacitors. Tartu, 2022, 144 p.
211. **Jinfeng Zhao.** Electrochemical characteristics of Bi(hkl) and micro-meso-porous carbon electrodes in ionic liquid based electrolytes. Tartu, 2022, 134 p.
212. **Alar Heinsaar.** Investigation of oxygen electrode materials for high-temperature solid oxide cells in natural conditions. Tartu, 2022, 120 p.
213. **Jaana Lilloja.** Transition metal and nitrogen doped nanocarbon cathode catalysts for anion exchange membrane fuel cells. Tartu, 2022, 202 p.
214. **Maris-Johanna Tahk.** Novel fluorescence-based methods for illuminating transmembrane signal transduction by G-protein coupled receptors. Tartu, 2022, 200 p.
215. **Eerik Jõgi.** Development and Applications of E. coli Immunosensor. Tartu, 2022, 103 p.
216. **Alo Rüütel.** Design principles of synthetic molecular receptors for anion-selective electrodes. Tartu, 2022, 109 p.
217. **Tanel Sõrmus.** Development of stimuli-responsive and covalent bisubstrate inhibitors of protein kinases. Tartu, 2022, 148 p.
218. **Oleg Artemchuk.** Autotrophic nitrogen removal processes for nutrient removal from sidestream and mainstream wastewater. Tartu, 2022, 115 p.
219. **Andre Leesment.** Quantitative studies of Brønsted acidity in biphasic systems and gas-phase. Tartu, 2023, 83 p.
220. **Meeli Arujõe-Sado.** Structural effects in aza-peptide bond formation reaction. Tartu, 2023, 83 p.
221. **Jonas Mart Linge.** Electrochemical reduction of oxygen on silver-based catalysts. Tartu, 2023, 269 p.
222. **Tõnis Laasfeld.** Integrating Image Analysis and Quantitative Modeling for a Holistic View of GPCR Ligand Binding Dynamics. Tartu, 2023, 226 p.
223. **Ernesto de Jesus Zapata Flores.** Derivatization Reagents used in negative mode electrospray LC-MS. Tartu, 2023, 107 p.

224. **Patrick Teppor.** Obtaining platinum-free oxygen reduction catalysts through biomass valorization: a case study of peat. Tartu, 2023, 161 p.
225. **Peeter Valk.** Methanol Oxidation on Platinum-Rare-Earth Metal Oxide Activated Catalysts. Tartu, 2023, 162 p.
226. **Shidong Chen.** Unravelling prehistoric plant exploitation in eastern Baltic: organic residue analysis of plant-based materials by multi-method approach. Tartu, 2023, 245 p.
227. **Yogesh Kumar.** M-N_4 macrocycle-based catalysts for electrocatalysis of oxygen reduction and oxygen evolution. Tartu, 2023, 224 p.
228. **Kerli Martin.** Recognition of carboxylates by synthetic receptors – from structure-affinity studies to solid-contact anion-selective electrode prototyping. Tartu, 2024, 130 p.
229. **Huy Quí Vinh Nguyen.** Development of Carbon Supported Pt–CeO_2 Catalysts for Proton Exchange Membrane Fuel Cells. Tartu, 2024, 198 p.
230. **Heigo Ers**. Adsorption and Structuring Processes at Single Crystal Electrode – Ionic Liquid Interface – Insights from Simulations and *in situ* Studies. Tartu, 2024, 137 p.
231. **Ritums Cepitis.** Modelling Structural and Geometrical Effects in Carbon Dioxide and Oxygen Electrocatalysis. Tartu, 2024, 99 p.
232. **Kaarel Kisand.** Resorcinol-derived carbon-based catalysts for polymer electrolyte fuel cell cathodes. Tartu, 2024, 205 p.
233. **Akmal Kosimov**. Template-assisted Mechanosynthesis (TAMS) for the production of bifunctional transition metal-based catalysts. Tartu, 2024, 123 p.
234. **Larissa Silva Macieli.** Derivatization-targeted LC-MS analysis of compounds containing amino group. Tartu, 2024, 157 p.
235. **Silvester Jürjo.** Separation of rare earth elements from Estonian phosphorite ore using liquid extraction followed by electrochemical reduction. Tartu, 2024, 99 p.
236. **Jan-Michael C. Cayme.** Organic-inorganic interactions in experimental and archaeological ceramics. Tartu, 2025, 156 p.
237. **Miriam Koppel.** The diffusion of H2 adsorbed in carbide-derived carbons: a quasi-elastic neutron scattering study. Tartu, 2025, 138 p.
238. **Kenneth Tuul.** Evaluating lithium-ion pouch cells and hydrogen storage materials under extreme conditions using advanced techniques. Tartu, 2025, 188 p.
239. **Marta-Lisette Pikma.** Exploring the basicity of phosphanes and related compounds. Tartu, 2025, 100 p.
240. **Indrek Saar.** Development of novel on-site chemical analysis tests – from alternative materials and technologies to functional prototypes. Tartu, 2025, 214 p.
241. **Gulnara Yusibova.** TAL MOF-Derived M-N-C Electrocatalysts for Oxygen Reduction and Evolution Reactions. Tartu, 2025, 120 p.