

SAINT PETERSBURG STATE UNIVERSITY

Manuscript copyright

Vladimirova Nadezhda Igorevna

**QSPR approaches in analytical chemistry: *in silico* prediction of properties of
ion-selective electrodes and deep eutectic solvents**

Scientific speciality 1.4.2. Analytical chemistry

Dissertation for the degree of candidate of chemical sciences

Translation from Russian

Supervisor:

Doctor of Chemical Sciences,

Kirsanov Dmitry Olegovich

Saint Petersburg

2024

Contents

INTRODUCTION	4
Chapter 1. LITERATURE REVIEW	12
1.1. Quantitative Structure-Property Relationships: theoretical foundations, methodology and applications in chemistry.....	12
1.2. Descriptors	16
1.2.1. Structural descriptors	17
1.2.2. Geometric (3D) descriptors	21
1.2.3. Quantum chemical descriptors.....	22
1.2.4. Electrostatic descriptors	24
1.3. Mathematical methods for data processing	25
1.3.1. Metrics and validation.....	27
1.3.2. Multiple Linear Regression	29
1.3.3. Partial Least Squares	30
1.3.4. Support Vector Regression	33
1.3.5. Neural Networks	34
1.4. Application of QSPR in analytical chemistry.....	36
1.5. Aim and tasks of the work	40
Chapter 2. APPLICATION OF QSPR MODELING FOR PREDICTING SELECTIVITY OF CARBONATE IONOSELECTIVE ELECTRODES	44
2.1. Experimental section.....	45
2.1.1. Selection of descriptors.....	45
2.1.2. Ion-selective sensors with new ligands in composition	46
2.1.3. QSPR modeling.....	49
2.2. Discussion of results	51
Conclusion to Chapter 2	58

Chapter 3. APPLICATION OF QSPR MODELING FOR PREDICTING THE SENSITIVITY OF IONOSELECTIVE ELECTRODES TO HARD METAL CATHIONS (Cu ²⁺ , Cd ²⁺ AND Pb ²⁺)	60
3.1. Experimental section.....	60
3.1.1. Selection of descriptors.....	60
3.1.2. QSPR modeling.....	61
3.2. Discussion of results.....	63
Conclusion to Chapter 3.....	72
Chapter 4. APPLICATION OF QSPR MODELING TO PREDICT PHYSICOCHEMICAL PROPERTIES OF DEEP EUTECTIC SOLVENTS BASED ON CHOLINE CHLORIDE.....	73
4.1. Experimental section.....	73
4.1.1. Selection of descriptors.....	73
4.1.2. QSPR modeling.....	76
4.2. Discussion of results.....	78
4.2.1. Density.....	78
4.2.2. Conductivity.....	82
4.2.3. Viscosity.....	84
4.2.4. Refractive index.....	86
4.2.5. Statistical validation of models.....	88
4.2.6. Evaluation of predictive performance of models.....	90
Conclusion to Chapter 4.....	91
CONCLUSIONS.....	93
ABBREVIATIONS LIST.....	94
LIST OF REFERENCES.....	95
APPENDIX.....	104

INTRODUCTION

In modern science, there has been a rapid development of computational approaches that are increasingly being used to solve complex chemical problems. One of such approaches is Quantitative Structure-Property Relationships (QSPR), which allows predicting the properties of new compounds and materials based on their molecular structure. QSPR is based on the use of mathematical models that link molecular descriptors - characteristics that describe the structure of a molecule numerically - to the properties of interest. This approach greatly accelerates and cheapens the process of developing new substances and materials with specified characteristics, reducing the need for costly and time-consuming experiments. QSPR allows researchers to obtain valuable information about the properties of compounds, bypassing the stage of their synthesis and experimental testing, which helps in the development of new drugs, solvents, all kinds of materials. This method has a wide range of applications in various fields including pharmaceuticals, environmental science, materials science, and chemical engineering.

Classical QSPR approaches require large amounts of data to build reliable models, usually thousands of compounds whose structure and properties are used to create a mathematical model. This significantly hampers the application of QSPR in analytical chemistry, where the ability to select large representative homogeneous datasets is very limited. For example, in order to predict the analytical performance of potentiometric sensors with polymer plasticized membranes based on the structural formula of the ionophore, a set of experimental data on sensors made on the basis of structurally similar ionophores is required. In this case, the composition of membranes (polymer, solvent-plasticizer, ion exchange additive) should be identical in this training sample. The conditions of potentiometric measurements (concentration ranges for determining sensitivities, set of interfering ions for determining selectivity) and experimental protocols for determining sensor characteristics should also coincide. Obviously,

different scientific groups working in this field conduct experiments in slightly different ways and direct comparison of the results obtained by different researchers is not always possible. Under such conditions, the creation of a training sample for QSPR containing at least a few tens of compounds characterized under similar conditions seems to be a serious problem. Similar limitations exist in many other areas of analytical chemistry, such as the development of extractants with desired properties. Overcoming such limitations and investigating the potential of QSPR for the development of new materials in analytical chemistry is an important and urgent task.

In the present work, two popular current research directions are chosen to explore the potential of QSPRs in solving analytical chemistry problems: the development of potentiometric sensors and the development of deep eutectic solvents (DES) with desired properties. As specific tasks in these directions, the work is focused on investigating the possibility of *in silico* prediction of sensitivity and selectivity of potentiometric sensors with polymer plasticized membranes based on ionophores and on the possibility of predicting the density, conductivity and viscosity of DESs. Sensors based on phosphoryl acetamides with pronounced sensitivity to heavy metal ions, carbonate-selective sensors and DESs based on choline chloride were chosen as model systems for research.

Relevance of the topic

The development of new approaches to QSPR modeling, providing the possibility of building predictive models based on a limited set of training data, would be an extremely useful tool for solving urgent problems in analytical chemistry. The choice of specific directions of this study is dictated by the following considerations. The development of new ion-selective electrodes (ISEs) for the determination of heavy metals and hydrophilic anions is an important task in the field of creating tools for environmental monitoring, food quality control, and medical diagnostics. QSPR modeling can significantly accelerate and optimize this process, allowing prediction of

sensor properties based on the structure of ionophores and other membrane components, thus reducing the time and cost of developing new ISEs.

Predicting the physicochemical properties of deep eutectic solvents is also a highly sought-after capability. DESs represent a new class of green solvents with unique properties that make them promising for applications in various fields such as green chemistry, pharmaceuticals and energy. QSPR modeling will make it possible to predict the properties of DESs based on their composition, which will greatly simplify the search for optimal formulations for specific applications, contributing to the development of sustainable technologies and reducing the negative impact on the environment.

Aims and tasks of the study

The **aim** of this study is to develop and validate QSPR models, adapted to deal with limited data sets, for predicting the properties of ion-selective electrodes and deep eutectic solvents.

To achieve this goal, the following **tasks** are set:

1. Development of QSPR models to predict the selectivity of potentiometric sensors to hydrophilic carbonate anions based on ionophores with acceptor substituents near the carbonyl group.
2. Development of QSPR models for predicting the sensitivity of potentiometric sensors to heavy metal cations (Ca^{2+} , Cd^{2+} , Pb^{2+}) based on diphenylphosphoryl acetamide ionophores.
3. Development of QSPR models for predicting the density, conductivity and viscosity of choline chloride-based DESs with organic acids as hydrogen bond donors.

Scientific novelty:

QSPR models for predicting the selectivity of ISE to hydrophilic anions (carbonate ions) based on limited training datasets were proposed and validated for the first time.

New approaches to modeling the sensitivity of ISEs to heavy metal cations (Ca^{2+} , Cd^{2+} , Pb^{2+}) based on substructure molecular fragments using limited training data sets were proposed; the models were validated on independent test sets of compounds.

QSPR models for predicting the physicochemical properties of choline chloride-based DESs with organic acids as hydrogen bond donors were proposed and validated for the first time. The models were validated on independent test sets of compounds.

For all QSPR models the key molecular descriptors that have a significant influence on the properties of compounds have been identified, which contributes to the development of approaches to obtaining materials with desired properties.

The efficiency of QSPR approaches when working with small training samples is shown and methods of validation of QSPR models under conditions of limited availability of compounds with experimentally determined target properties are proposed.

Practical significance of the study:

The development of QSPR models for the directed design of ionophores providing the given characteristics (selectivity to carbonate ions and sensitivity to heavy metal cations) of ion-selective electrodes essentially facilitates the search for new ligands for ISE.

The development of QSPR-model for prediction of physicochemical properties of deep eutectic solvents based on choline chloride with organic acids allows to select DESs with necessary properties for a specific task, bypassing labor-intensive stages of experimental study of their properties.

Demonstration of the possibility of successful application of the QSPR approach for prediction of properties of analytical chemistry objects characterized by limited data sets.

Thesis outline:

1. QSPR models built on the basis of substructural molecular fragments describing ionophores for ISE using limited training data sets provide prediction of the selectivity of potentiometric sensors to carbonate anion with an error not exceeding 1.5 $\log K^{ce\pi}(HCO_3^-/Cl^-)$.

2. QSPR models based on substructural molecular fragments describing phosphoryl acetamide ionophores provide prediction of the sensitivity of ion-selective sensors to heavy metal cations (Ca^{2+} , Cd^{2+} , Pb^{2+}) with an error not exceeding 8 mV/dec.

3. QSPR models based on substructural molecular fragments and semi-empirical descriptors calculated by the PM3 quantum-chemical method using limited training datasets provide predictions of density, viscosity, and conductivity of choline chloride-based DESs with organic acids as hydrogen bond donors. The estimation errors do not exceed 0.065 g/cm for density, 4.6 mS/cm for conductivity, 3.6 mPa·s for viscosity.

Main results of the work.

The results were presented at the following scientific events: The 1st International Electronic Conference on Chemical Sensors and Analytical Chemistry; XII International Conference of Young Scientists "Mendeleev 2021", St. Petersburg; 13th International Symposium WSC-13, Moscow; The 2nd International Electronic Conference on Chemical Sensors and Analytical Chemistry; as well as three publications in peer-reviewed international journals indexed by Scopus and Web Of Science [1–3]:

1. Vladimirova, N., Polukeev, V., Ashina J., Babain V., Legin, A., Kirsanov, D. Prediction of Carbonate Selectivity of PVC-Plasticized Sensor Membranes with Newly Synthesized Ionophores through QSPR Modeling (2022) *Chemosensors*, 10(2), 43. DOI: 10.3390/chemosensors10020043

2. Vladimirova N., Puchkova E., Dar'in D., Turanov A., Babain V., Kirsanov D. Predicting the Potentiometric Sensitivity of Membrane Sensors Based on Modified

Diphenylphosphoryl Acetamide Ionophores with QSPR Modeling (2022) *Membranes*, 12(10), 953. DOI: 10.3390/membranes12100953

3. Vladimirova N., Bochko T., Shishov A., Dmitry Kirsanov. Predicting the properties of deep eutectic solvents based on choline chloride and carboxylic acids and their mixtures with water using QSPR approach (2024) *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 692. DOI: 10.1016/j.colsurfa.2024.133961

Main scientific results:

1. Vladimirova N., Puchkova E., Dar'in D., Turanov A., Babain V., Kirsanov D.. Predicting the Potentiometric Sensitivity of Membrane Sensors Based on Modified Diphenylphosphoryl Acetamide Ionophores with QSPR Modeling (2022) *Membranes*, 12(10), 953. DOI: 10.3390/membranes12100953

This study investigated the applicability of the QSPR method for predicting the potentiometric sensitivity of plasticized polymer membrane sensors based on the chemical structures of ionophores described by substructural molecular descriptors as model inputs. The QSPR model was trained on literature data on heavy metal sensitivities of previously studied structurally similar ionophores and showed high predictive ability for sensitivities to Cu²⁺, Cd²⁺ and Pb²⁺. The model predictions for four new diphenylphosphorylacetamide-based ionophores were compared with potentiometric experimental data for these ionophores, and satisfactory agreement was observed.

The co-researcher contributed significantly to this study by collecting and analyzing the literature data, formalizing the ionophore structures, constructing the QSPR model, analyzing the obtained results and interpreting the contribution of different descriptors to the sensitivity of the sensors. The applicant was also actively involved in the preparation of the scientific publication.

2. N. Vladimirova, T. Bochko, A. Shishov, D. Kirsanov. Predicting the properties of deep eutectic solvents based on choline chloride and carboxylic acids and their mixtures

with water using QSPR approach (2024) *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 692. DOI: 10.1016/j.colsurfa.2024.133961

In this work, a QSPR approach for predicting the properties of HERs is investigated. Regression models were constructed based on the organic acid structures of the HERs. The models related solvent properties and molecular descriptors of the corresponding carboxylic acids represented by substructural fragments and quantum chemical descriptors. Statistically significant correlations were established between the acid descriptors and the density, viscosity and electrical conductivity of HERs. The effect of different descriptors on these properties was analyzed based on regression coefficients. Given the rather limited data set, the statistical validity of the regression models was assessed using nested cross validation.

The co-researcher was very actively involved in this study, performed data collection and analysis, formalization of carboxylic acid structures using substructural molecular fragments and quantum chemical descriptors, construction and validation of QSPR models, and interpretation of the results. The co-researcher also prepared the initial draft of the scientific publication.

3. Vladimirova, N., Polukeev, V., Ashina J., Babain V., Legin, A., Kirsanov, D.. Prediction of Carbonate Selectivity of PVC-Plasticized Sensor Membranes with Newly Synthesized Ionophores through QSPR Modeling (2022) *Chemosensors*, 10(2), 43. DOI: 10.3390/chemosensors10020043

In this work, the potential of the QSPR approach for predicting the potentiometric selectivity of plasticized polymer membrane sensors based on novel ligands is investigated. Sensors with selectivity to carbonate were considered as the object of study. Using selectivity data for 40 ionophores available in the literature and their substructural molecular fragments as descriptors, a QSPR model was constructed that demonstrated sufficient accuracy in predicting selectivity for novel ligands having similar molecular fragments to the ligands used to train the model.

The co-researcher contributed significantly to this work by collecting and analyzing literature data, formalizing ionophore structures, constructing the QSPR model, analyzing model results, including establishing the contribution of substructural molecular descriptors to the model, and experimentally determining the selectivity of sensors with novel ionophores in the composition. The applicant was also actively involved in the preparation of the scientific publication.

Chapter 1. LITERATURE REVIEW

1.1. Quantitative Structure-Property Relationships: theoretical foundations, methodology and applications in chemistry

For modern chemical research, the search for compounds with predetermined, desired properties is an urgent task. The traditional experimental approach, in which a hypothesis is proposed and then a series of experiments are performed, although it has a long history and is the basis for understanding chemical phenomena, often faces limitations when it comes to the efficiency of selecting structures with predetermined properties. First, the cycle of research - formulating hypotheses, conducting experiments, observing results, and refining procedures often takes a long time to achieve the desired result. Each step, based on previous knowledge, theoretical foundations and new assumptions put forward, requires strict adherence to the experimental methodology, analyzing and then adjusting to the latest results. In addition, each series of experiments requires material costs and labor of experimenters. Finally, the performed synthesis does not guarantee that the obtained compound or material will have the desired properties.

With the development of computational techniques, computational approaches have become an increasingly important complement to experimental work. These computational methods, including quantum calculations, molecular dynamics modeling, and the search for quantitative structure-property relationships, allow us to solve both theoretical problems, such as the study of interaction mechanisms, and practical problems, such as predicting the properties of previously unexplored compounds. Such calculations provide an opportunity for preliminary evaluation of new compounds and materials, which is carried out faster and with much lower material costs than experiments performed in the laboratory.

QSPR is a computational approach that allows to establish a relationship between the chemical structure of a molecule and its properties. The general idea of this approach is to calculate, using various computational methods, a mathematical relationship (model) between the structure of a compound and its properties, and then use this model to calculate the properties of new molecules not used in the model. This calculation is carried out on the basis of parameters representing information about the properties of compounds in numerical form - descriptors. The general scheme of the approach is presented in Figure 1, which shows that the calculation of the connection between the structure of a compound and its properties is carried out through the stages of descriptor collection and their further statistical analysis.

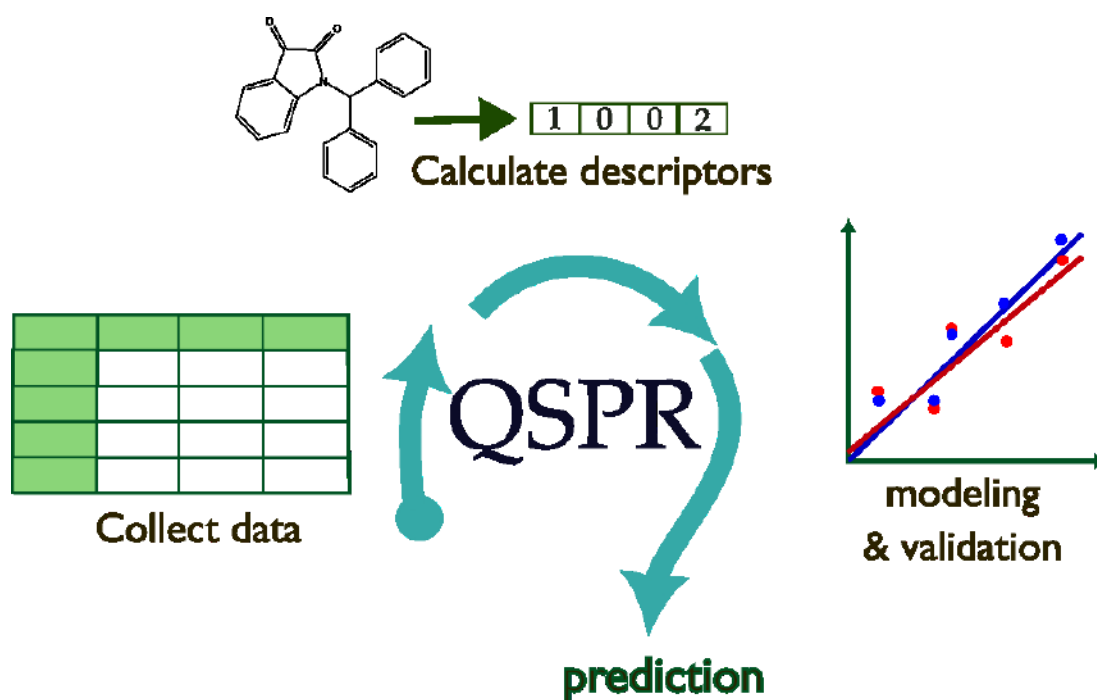


Figure 1: Schematic of the methodology used in papers applying the QSPR approach.

The ideas underlying QSPR seem intuitive, and they were discussed long before the term was officially coined. The first paper in which these ideas are most explicitly articulated is a paper by Alexander Kram Brown and Thomas R. Fraser [4], which argues

that there must be a relationship between the chemical structure and the physiological action of a substance. In such a case, if the structure changes by ΔC and the properties at the same time change by $\Delta \Phi$, then it becomes possible to find the dependence between C and Φ .

The term "QSPR" was formally proposed by Corwin Hansch in the 1960s [5]. [5]. The works of Hansch and his group belong to the field of biochemistry[6,7], where the QSPR approach is often referred to as QSAR -- Quantitative Structure-Activity Relationships.

From the time of Hansch's work to the present day, it is the calculation of the activity of compounds that represents the widest area of application of QSPR, in which such a direction as the search for new drugs should be singled out. For example, in one of the recent works [8] the structure of a quinoxaline derivative, a new drug against leishmaniasis, was proposed. The new substance in this work was not only proposed using descriptor analysis, OSIRIS Property Explorer program and Weber and Lipinski rules, but its properties were also predicted by QSPR model. After the synthesis of the compound with the proposed structure and its investigation, it turned out that the predicted property - the concentration index of half-maximal inhibition is close to that obtained experimentally (the percentage deviation of the predicted value from that obtained experimentally was less than 4%).

The structures of novel coumarin-isoxazole-sulfonamide hybrid compounds as antibiotics were evaluated using the QSAR approach[9], and 1H-pyrazole derivatives were studied by a new variation of QSAR, 5D-QSAR, to identify the best inhibitor of epidermal growth factor receptor, such inhibitors are used in cancer treatment [10].

Why is drug discovery such a significant area of application for QSPR? It's all about the specifics of the approach: first, the most important advantage of QSPR is the significant reduction of costs associated with the development of new materials and compounds. Traditional experimental approaches to the development of new materials are often resource-intensive and require significant financial, labor, and time

investment. However, models calculated with QSPR allow predicting the properties of compounds prior to experimentation. This is especially important when measurements are time-consuming or technically challenging.

In addition, QSPR plays an important role in shortening the development time of new materials. Calculating the properties of compounds prior to experimentation speeds up the development, synthesis and testing cycle by identifying promising candidates and eliminating unsuitable ones. This is particularly useful in fast-moving industries where speed to market is critical.

Finally, in the area of product development requiring animal testing, QSPR's ability to provide predictive data reduces the number of tests, which is ethically important.

Because of all these advantages, QSPR has been successfully applied in a multitude of areas of chemistry, the previously mentioned drug discovery being just one example of an application.

For example, in the field of environmental chemistry (ecological chemistry), QSPR is used to predict the behavior of chemicals from an environmental perspective, including biodegradability [11,12], toxicity to aquatic organisms [13,14], bioaccumulation potential [15,16] and contaminant persistence assessment [17]. Consideration of these factors contributes to the development of safer chemicals and regulatory compliance. Similar goals are pursued by studies applying QSPR in the field of agrochemistry, where it predicts the toxicity of pesticides [18,19], herbicides [20,21] and fungicides [22,23], helping to develop compounds with improved efficacy and reduced environmental impact.

The QSPR approach has been widely used to predict the properties of new materials including polymers [24,25], nanomaterials [26,27] and composites [28], these materials are then applied in areas such as electronics, energy storage and photovoltaics. Much research has been devoted to the development of new catalysts as the resulting QSPR models help to calculate their activity [29,30].

In food chemistry, QSPR models can predict flavor aroma [31,32], aroma [33] and stability of food additives and preservatives [34], facilitating the development of new food products and improving their safety.

In the cosmetic industry, QSPR models predict the absorption of chemicals through the skin [35], comedogenicity [36] and irritation risk [37] from new ingredients, providing an opportunity to develop safer and more effective cosmetic products.

In chemical engineering, QSPR helps in the design and optimization of chemical processes by predicting the thermodynamic properties and reactivity of starting substances and products [38,39], a significant amount of work is devoted to predicting the properties of ionic liquids [40,41].

In the petroleum and fuel industry, the QSPR approach is used to predict the properties of fuels and lubricants [42–44], such as viscosity, flash point and combustion efficiency, which helps in developing more efficient and environmentally friendly fuel formulations.

1.2. Descriptors

In the QSPR approach, the concept of descriptors plays a key role and links the structure of a molecule and its physical, chemical or biological properties. In a general sense, descriptors are numerical values that quantitatively describe some aspect of a molecule, most commonly the molecular structure, and serve as the basis for building predictive models. The variety and choice of these descriptors are critical to the accuracy and applicability of QSPR models, as they directly affect the predictive power of the model. Regardless of the nature of the descriptors, they have the following mandatory requirements [45]: independence from labeling and numbering of atoms, independence from molecular rotation, definition by an unambiguous algorithm, and well-defined

applicability to molecular structures. In addition, it is desirable that descriptors be simple and can be applied to a wide range of molecules.

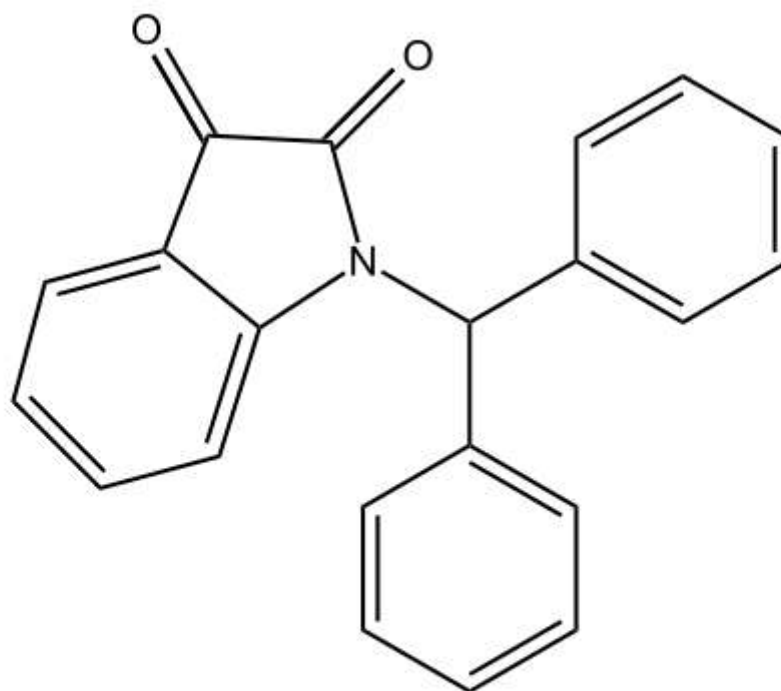
Most descriptors can be categorized as structural, geometric (3D), electrostatic, and quantum.

1.2.1. Structural descriptors

Structural descriptors are the simplest descriptors that quantify the structural characteristics of molecules without delving into their electronic or spatial characteristics. Structural descriptors include the number of atoms of a particular type, the number of bonds, and the presence of certain functional groups. They are easy to calculate and provide basic but important information about the size and composition of a molecule. Various methods and tools exist for deriving structural descriptors, each with unique advantages and applications.

Prominent among the methods are:

- SMILES (Simplified Molecular-Input Line-Entry System): SMILES is a notation, represents a chemical structure as a line of text. It describes the structure based on the sequence of atoms and the types of bonds between them. SMILES strings can be easily converted into numerical descriptors using a variety of software tools, including the widely used ChemDraw, making them very convenient for studies using the QSPR approach. Figure 2 shows the structure of the 1-benzhydrylindoline-2,3-dione molecule and its record in SMILES.



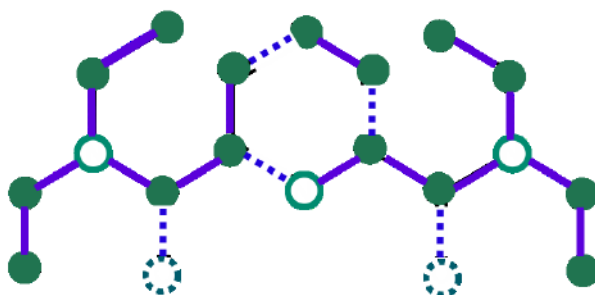
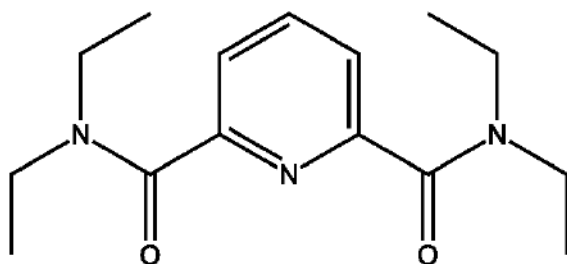
1-benzhydrylindoline-2,3-dione




O=C(C(C1=CC=CC=C1)2)=O)N2C(C3=CC=CC=C3)C4=CC=CC=C4

Figure 2. An example of recording the structure of a molecule using SMILES

- Molecular graphs are an approach that represents a molecule as a graph, where each vertex is labeled with the corresponding chemical element (e.g., C = carbon, H = hydrogen) and each edge is labeled with the type of covalent bond (single -, double =, triple, aromatic). Hydrogen atoms are usually omitted in this representation of the molecule, since the valence of each atom is known. Using this approach, various graph-based descriptors such as adjacency matrices or topological indices can be derived. Figure 3 shows the molecule N2,N2,N6,N6,N6-tetraethylpyridine-2,6-dicarboxamide as a graph.

N2,N2,N6,N6-tetraethylpyridine
-2,6-dicarboxamide



Vertex	Symbol	Valency
	O	2
	N	3
	C	4



Edge	Chemical bond
	single
	double

Figure 3. Representation of the N2,N2,N6,N6-tetraethylpyridine-2,6-dicarboxamide molecule as a graph. Different types of nodes according to atomic elements and different types of edges depending on the chemical bonding in the molecule

Various free and commercial programs are used to find descriptors based on graphs. Among the most popular programs are the proprietary (with a paid

license) DRAGON and the openly available PaDEL and ISIDA. The latter was developed by Vitaly Petrovich Solovyov, Doctor of Chemical Sciences, leading researcher of the Laboratory of New Physicochemical Problems of the Institute of Physical Chemistry of the Russian Academy of Sciences.

- Molecular fingerprints are binary or numerical vectors that reflect the presence or absence (and sometimes the number) of certain substructures or patterns in a molecule. There are several types of fingerprints, including pathway-based fingerprints, Extended-Connectivity Fingerprints (ECFP) [46], and Molecular ACCess System keys (MACCS). Figure 4 shows the structure recordings using MACCS and ECFP. One cell of the fingerprint is called a bit. A MACCS bit is 1 if the molecule contains a substructure corresponding to the bit; otherwise it is 0. ECFP searches for the relative position of each atom in the molecule using the 1 bond layer and records the structural information in a bit array.

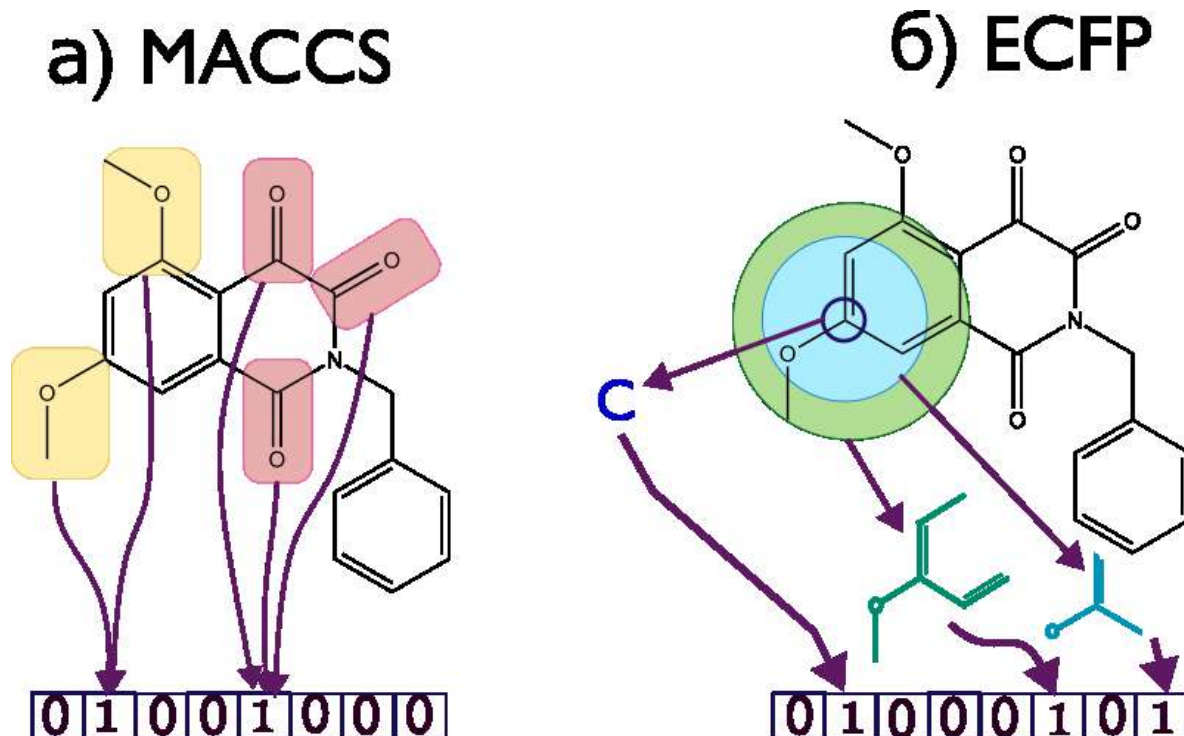


Figure 4. Representation of molecular structures using (a) MACCS and (b) ECFP

1.2.2. Geometric (3D) descriptors

Geometric (3D) descriptors are descriptors that include three-dimensional information about a molecule, such as shape, volume, and surface area. Geometric descriptors are important for understanding properties that depend on the spatial arrangement of atoms and are often applied to objects of study that are large molecules such as proteins. However, models trained on datasets composed solely of 3D descriptors are often inferior in predictive power to models composed solely of structural descriptors, so geometric descriptors are often used as a complement to structural descriptors [47]. Geometric descriptors can be divided into several groups, each of which reflects different aspects of molecular geometry and its influence on the behavior of molecules.

1. Molecular shape descriptors. These descriptors quantitatively characterize the overall shape of a molecule, which can affect its ability to interact with biological targets or to incorporate into receptor sites. For example, eccentricity, asphericity, and shape indices that describe how elongated, flat, or spherical the molecule is.
2. Volume and surface area descriptors. This group includes descriptors that measure the surface area and volume of a molecule, such as solvent-accessible surface area and molecular volume. These descriptors are important for predicting solubility, permeability, and other properties related to how a molecule interacts with its environment.
3. Descriptors of steric effects. These descriptors quantitatively determine the spatial arrangement of atoms in a molecule and the resulting steric effects that can affect the reactivity of molecules and their interactions. Separately, we can single out the parameter Sterimol [48], which takes into account the size of atoms in the molecule and allows us to estimate the steric accessibility.
4. Moment of inertia descriptors. These descriptors are related to the distribution of mass within the molecule and its rotational inertia around different axes. They

provide information about the balance of the molecule and the mass distribution in three-dimensional space.

5. 3D indices. The 3D Wiener index (W3D) is the best known topological descriptor used to quantify the branching of molecules and is defined as the sum of the lengths of the shortest paths between all pairs of vertices (atoms) in a molecule [49]. W3D takes into account the geometric distances between pairs of atoms, not only their topological distances. This modification of the Wiener number allows W3D to more accurately characterize the spatial configuration of a molecule.

The 3D Balaban Index (J3D) is another topological descriptor that measures the degree of branching and cyclicity within a molecule. Like the Wiener index, J3D is calculated based on geometric rather than topological distances between atoms [50].

6. 3D autocorrelation descriptors. Unlike the previous descriptors, 3D indices, 3D autocorrelation descriptors quantify not only the distance between pairs of atoms, but in addition the distribution of certain molecular properties (such as charge, mass, or hydrophobicity) over the structure of a molecule [51].
7. Weighted Holistic Invariant Molecular (WHIM) descriptors. WHIM descriptors reflect information about the shape, size, symmetry and distribution of atoms in a molecule and are calculated from the (x, y, z)-coordinates of the atoms. WHIM descriptors are designed to be constant regardless of the position or rotation of the molecule [52].

1.2.3. Quantum chemical descriptors

Quantum chemical descriptors play an important role in computational chemistry, allowing us to understand the properties and behavior of molecules using quantum mechanics-based calculations. Various computational methods including ab initio,

density functional theory (DFT) and semiempirical approaches are used to derive these descriptors.

Ab initio methods are aimed at solving the Schrödinger equation for molecular systems. Calculations by these methods are performed without reference to empirical parameters or fitting to experimental data and the only input data for ab initio calculations are physical constants. This makes these methods highly accurate, but extremely long and computationally demanding.

Density functional theory offers a different approach, based on the electron density of a system rather than its wave function. DFT assumes that all properties of a system can be determined from its electron density, and calculations using this method are less computationally intensive.

Semi-empirical methods simplify the computational process by incorporating empirical data and simplifications into quantum mechanical calculations. This approach significantly reduces the required computational resources. Semiempirical methods, such as PM3, strike a balance between computational efficiency and the accuracy required to effectively study the properties of molecules.

Quantum-chemical descriptors include such parameters as:

- total energy of the molecule, obtained as a result of quantum mechanical calculations, includes all kinetic and potential energies of electrons and nuclei.
- heat of formation, which characterizes the change in energy associated with the formation of a molecule from its constituent atoms in their standard states.
- repulsion and attraction energies, quantifying the electrostatic repulsion between electrons and the attraction between electrons and nuclei, respectively.
- protonation energy, which characterizes the energy change associated with the attachment of a proton to a molecule.

- electron density, showing the distribution of electron density in a molecule, highlighting regions with more and fewer electrons.

1.2.4. Electrostatic descriptors

Electrostatic descriptors quantify the distribution and influence of electrical charges within a molecule. The characteristics described by these descriptors affect the properties of the molecule such as solubility, binding affinity, and reactivity. The same programs used to calculate electrostatic descriptors of molecules are used to calculate quantum descriptors: Gaussian, HyperChem, and Open3DGRID.

Among electrostatic descriptors the following can be distinguished: empirical atomic partial charges of Gasteiger-Marsili -- semi-empirical charges calculated using the method of electronegativity equalization and giving an approximate idea of the charge distribution between atoms of a molecule. Empirical Zefirov atomic partial charges -- charges derived from an empirical method that takes into account the electronic structure of the molecule.

Mulliken atomic partial charges - charges derived from quantum mechanical calculations, reflecting the distribution of electron density around atoms. [53]

Minimum and maximum atomic partial charges represent the most negative and most positive charges in a molecule, indicating regions of high and low electron affinity, respectively.

Polarity parameters – descriptors that quantitatively characterize the overall polarity of a molecule and reflect its dipole character and potential for polar interactions. Dipole moment – a vector quantity measuring the separation of positive and negative charges in a molecule and indicating its overall polarity. Molecular polarizability – a descriptor describing how easily a molecule's electron cloud can be distorted by an external electric field, affecting its interaction with light and other molecules. Molecular hyperpolarizability – higher-order polarizability that contributes

to nonlinear optical properties. Local polarity of a molecule characterizes the change in polarity in different regions of the molecule.

Average ionization energy reflects the energy required to remove an electron from a molecule, related to its ionization potential and stability. The minimum and maximum electrostatic potential at the surface of a molecule indicate the regions with the lowest and highest electrostatic potential, which is important for understanding how molecules interact with charged particles. The total variation in surface electrostatic potential is a quantitative estimate of the variability of the electrostatic potential across the surface of a molecule. The electrostatic balance parameter describes the balance between positive and negative sites within a molecule, which affects its overall electrostatic behavior.

In addition, Charged Partial Surface Area (CPSA) descriptors can be separately identified. CPSA descriptors are designed to describe polar intermolecular interactions. CPSA descriptors proved to be applicable, for example, in the study of toxicity of aqueous solutions, acting as an alternative to LUMO energy level measures to describe global and local electrophilicity in the case of noncovalent molecular interactions [54].

1.3. Mathematical methods for data processing

To establish quantitative relationships between the molecular structure of compounds and their properties, the QSPR approach uses various mathematical methods to reveal complex relationships. The main purpose of these methods, in particular those based on linear regression, is to model quantitative properties such as selectivity, density or biological activity.

Among the regression methods used in QSPR, multiple linear regression (MLR), partial least squares (PLS) are predominant. Each method aims at obtaining a mathematical regression equation linking molecular descriptors to properties. In the

general case of linear regression, this equation is as follows:

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n + \varepsilon, \quad (1)$$

Where y is the dependent variable representing the property being modeled,

$a_0, a_1 \dots a_n$ - regression coefficients determining the change in Y as a result of changing each variable X by one unit, provided that all other variables X remain unchanged,

$x_1, x_2, \dots x_n$ - independent variables, which are represented by molecular descriptors when applying the QSPR approach,

ε is an error that takes into account the deviation of the observed values from the dependence defined by the model

Thus, the main goal of regression methods is to find the optimal coefficients that best describe the relationship between descriptors and properties. For this purpose, the coefficients need to be calculated in such a way as to minimize the prediction error. The final model is characterized by various metrics to assess its quality.

However, the task of model building is not limited to the calculation of coefficients. Given that data sets in chemical research are often small - typically consisting of dozens of data points - the risk of model overfitting is very high. Overfitting occurs when the model describes random noise rather than the trend that makes up the regression. Validation techniques such as cross-validation are used to reduce this contribution to the model. Cross-validation involves dividing the data into several subsets. In this case, some of them are used to train the model and others are used to test it. This approach ensures that the model generalizes well to new, unknown data, rather than simply repeating the dependencies of the training set.

1.3.1. Metrics and validation

Assessing the quality of models is important to ensure their predictive ability. Typically, this assessment is done using various statistical metrics that quantify how well models predict properties. Before moving directly to a discussion of these metrics, it is worth noting the concept of residuals as a fundamental idea for understanding these metrics.

Residuals are the difference between the observed values of the target variable and the predictions made by the model:

$$\text{Residual} = \text{actual value} - \text{predicted value}$$

Figure 5 illustrates a linear regression built on one variable (x) and predicting the dependent variable (y). The red dots represent the actual observations, the purple line corresponds to the linear regression model, and the green lines highlight the residuals for each observation, visualizing the errors in the prediction.

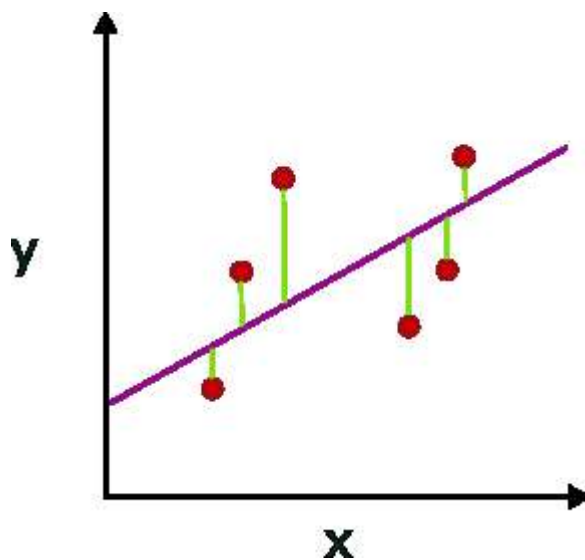


Figure 5. Example of residuals for a linear model with one feature.

This visual representation is not just a method to see how well the model is performing on its prediction task, but also a way to help identify patterns of error that may indicate certain model problems.

Based on the concept of residuals, Mean Squared Error (MSE) determines the average of these squared differences, providing a comprehensive measure of the model's prediction error:

$$MSE = \frac{1}{n} \sum_{i=1}^n residual^2, \quad (2)$$

The Root Mean Squared Error (RMSE) provides a measure of error in the same units as the predicted property. $RMSE = \sqrt{MSE}$

Coefficient of Determination, or R^2 -- another important metric derived from residuals. It measures the proportion of variance of the dependent variable and gives an indication of the strength of the relationship between the model inputs and outputs:

$$R^2 = 1 - \frac{\sum residual^2}{\sum (y_i - \bar{y})^2}, \quad (3)$$

It follows from the equation that the maximum value of $R^2 = 1$, then the larger the value of R^2 , the better the model is, the more accurately the model reproduces the observed values of the dependent variable.

When calculating the model, it is important to evaluate the quality of prediction of unknown data by the model. The simplest method of such evaluation is to divide the initial data set into two parts: training and test samples. The model is trained on the training set and then evaluated using the metrics described above on the test set, which acts as new data unknown to the model. This method is as simple as possible, but excludes the test set from the training data, which can be a significant limitation when working with small datasets.

To maximize the use of available data, especially if it is limited, the data partitioning method has evolved into cross-validation (cross-validation). Cross-validation involves repeatedly partitioning a data set into training and test sets, allowing

each subset of data to be used for both training and testing. The most common form of this technique is K-fold cross-validation.

In K-fold cross-validation, the data set is divided into K equal parts. The model is trained on K-1 parts, and the remaining part is used as a test set. This process is repeated K times, with each of the K parts being used as a test set exactly once. The final model performance is usually the arithmetic mean of the model performance estimates on these K test sets. This method not only improves the robustness of the model estimates, but also utilizes the data more efficiently by ensuring that each data point contributes to both training and validation.

An extreme variant of K-fold cross-validation is full cross-validation (leave-one-out cross-validation, LOO CV), where K is equal to the number of samples in the dataset. This means that each part contains exactly one data point. The model is trained on all but one sample of the data and tested on that excluded sample. The LOO CV method is particularly useful for very small datasets, but can be computationally expensive when dealing with large datasets, making it less practical in such scenarios.

1.3.2. Multiple Linear Regression

Multiple Linear Regression (MLR) is one of the fundamental and still the most frequently used statistical methods in QSPR. Its popularity stems from its simplicity, interpretability and an easy-to-understand way to model the relationship between a dependent variable (the property of interest) and several independent variables (molecular descriptors). As one of the earliest methods used in QSPR research, MLR is a precursor to more sophisticated computational methods in the field. The fundamental idea behind this method is the assumption that there is a linear relationship between the property of interest and the molecular descriptors.

From a formal point of view, the MLR model is described by the general linear regression equation given earlier. And the goal of MLR -- to find the set of coefficients (a_0, a_1, \dots, a_n) that best fit the observed data -- is consistent with the general goal of linear

regression. Typically, finding the coefficients is implemented using the least squares method.

Due to its simplicity, the MLR method works efficiently only on very simple research objects because of the following significant limitations: the number of descriptors should not exceed the number of samples and there should not be a linear relationship between the descriptors. In addition, it should be taken into account that the method assumes the existence of a linear relationship between each descriptor and the dependent variable.

1.3.3. Partial Least Squares

Partial Least Squares (PLS) regression, also known as Projection on Latent Structures, is a key statistical method in the field of QSPR, especially valued for its efficiency in dealing with complex, multivariate data. The PLS method was developed by Svante Wold specifically for chemometrics applications [55] and is indispensable in QSPR studies where available experimental data may be scarce. Unlike the previously described MLR method, the PLS method allows modeling dependence using latent variables (LV) in variable sets where the number of descriptors (independent variables) greatly exceeds the number of samples and where the variables may be linearly dependent. These features of the method are indispensable for finding dependence in cases where descriptors represent frequencies of the spectrum [56].

PLS regression aims at finding a relationship between two matrices, X (independent variables represented in QSPR by molecular descriptors) and Y (responses, e.g., physicochemical properties), by projecting the independent variables onto new latent variables. The original data is a space with dimensionality corresponding to the number of descriptors, but to find a linear dependence we reduce the original dimensionality by projecting the available data onto new axes - latent variables.

The process of finding these variables is as follows: the first stage is to normalize and center the descriptor data. This ensures that each descriptor is transformed to account for values so that variables with large numerical ranges do not carry more weight in the analysis. The analysis will then focus on the distribution and covariance of the descriptors rather than their absolute values.

The directions of maximum covariance are then determined using an iterative process: the algorithm starts by finding the direction in the descriptor space that has the highest covariance with the dependent variable. Next, a latent variable is constructed as a linear combination of descriptors weighted by their contribution to this direction of maximum covariance.

The process continues by successively finding latent variables, each of which represents a direction in descriptor space that reflects the next highest level of covariance with the dependent variable, orthogonal to the directions already considered. This orthogonality ensures that each new latent variable reflects unique information that has not yet been accounted for by previous latent variables. Figure 6 shows the result of finding the latent variables for the sample set.

The final PLS model is built using these latent variables as inputs. The number of latent variables to be included in the model is usually determined through validation to balance the transfer of information to the model and the risk of overfitting.

Thus, in addition to the matrix of independent variables X and the matrix of properties Y , the matrices of accounts and loadings are also used to find the dependence (Figure 7): the matrix of accounts (T) represents the projection of the raw data onto the latent variables, while the matrix of loadings (P): Loadings show how much each descriptor contributes to each latent variable.

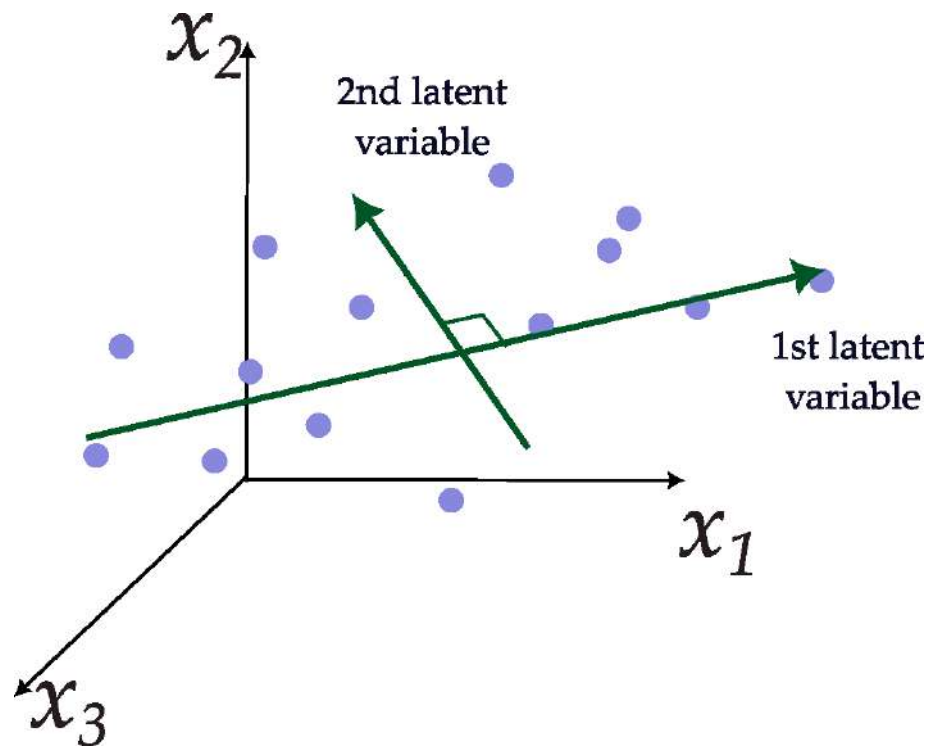


Figure 6: Illustration of the iterative process of constructing latent variables in PLS regression. The initial axes represent normalized and centered descriptors. The point cloud represents the distribution of objects in the descriptor space. Latent variables are found as directions of maximum covariance with the dependent variable, orthogonal to each other

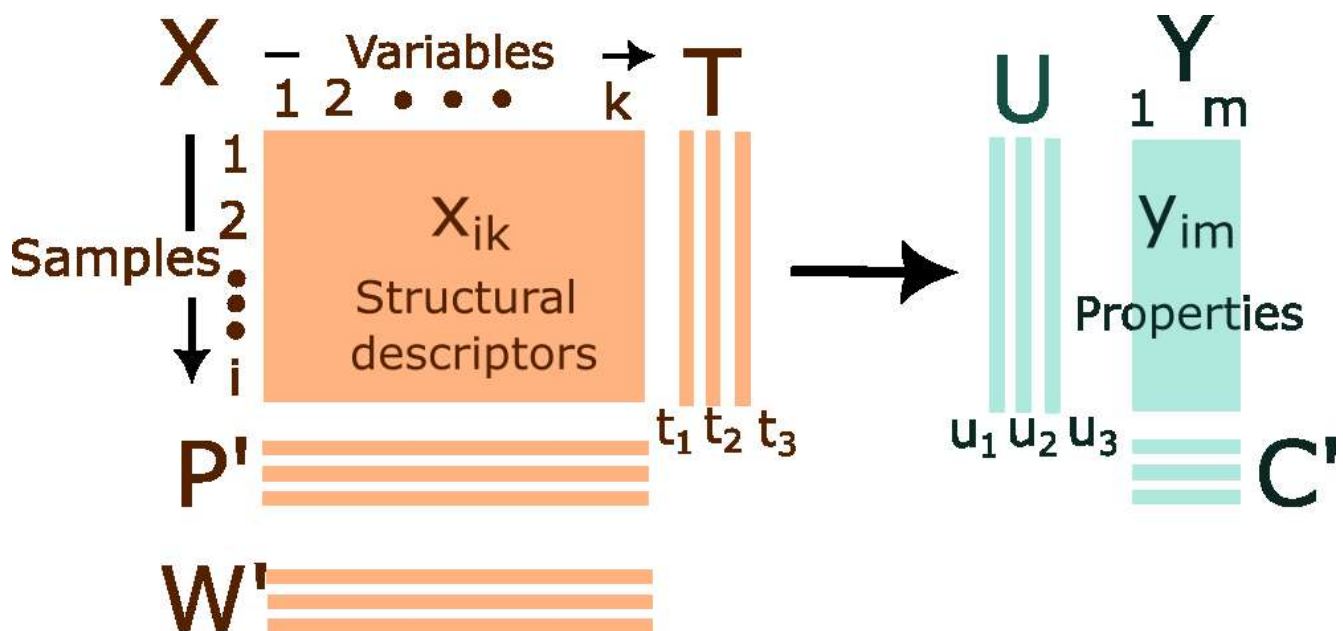


Figure 7. PLS data can be presented as two tables, matrices, X and Y

The basic idea of PLS can be described by the following equations:

$$X = TP^T + E, \quad (4)$$

$$Y = UQ^T + F, \quad (5)$$

where T and U are the count matrices for X and Y, respectively, representing the projections of X and Y onto the latent variables.

P and Q are loadings matrices showing how each original variable in X and Y contributes to the components.

E and F are residual matrices containing information not explained by the model.

Thus, the equation for finding the PLS model:

$$Y = XB + F, \quad (6)$$

where B is the matrix of regression coefficients calculated to maximize the covariance between X and Y.

Since the model allows not only predicting properties but also understanding the relationship between structure and properties through the interpretation of counts and loadings, PLS regression is a widely applicable method in QSPR, allowing researchers to uncover relationships between molecular descriptors and desired properties even under challenging data acquisition scenarios. [57,58]

1.3.4. Support Vector Regression

Support vector regression (SVR) is an approach known for its effectiveness in dealing with complex and nonlinear relationships. Derived from the principles of support vector machine (SVM) used for classification, SVR extends the application of the method to regression problems. Unlike traditional regression methods that minimize the error between predicted and actual values, SVR focuses on fitting the error within a given

threshold, which makes it useful when dealing with data sets with noise [59].

The basic idea of SVR is to find a hyperplane in a multi-dimensional descriptor space that best fits the data. The concept of hyperplane is central to understanding the SVR method and its predecessor, the SVM method. A hyperplane is an n-1 dimensional subplane in n-dimensional space (a line in two-dimensional space, a plane in three-dimensional space) that is used to separate data points in space in order to find a classification or regression.

The goal of SVR is to build a model that predicts an output value y for an input vector X with a variance of no more than ϵ for each data point.

The hyperplane equation in SVR is as follows:

$$f(x) = w^T X + b, \quad (7)$$

where w is the vector of descriptor weights, X is the descriptor matrix, and b is the displacement of the hyperplane with respect to the origin.

1.3.5.

Neural

Networks

Neural Networks or Artificial Neural Networks are the most complex class of machine learning models under consideration. The main attraction of neural networks is their ability to model complex, non-linear relationships that are often too complex for traditional statistical methods. The construction of such relationships is achieved by mimicking the structure and function of the human brain, using layers of interconnected nodes or "neurons" to process data through a series of transformations and connections.

Neurons take signals from one or more inputs and convert them into an output signal. Each neuron can be described as a collection of two components (Figure 8) - an adder (Σ) and an activation function (f). The adder receives input signals and returns their weighted sum, which is then converted by a nonlinear activation function into an output signal. Various functions such as sigmoid, hyperbolic tangent (\tanh), ReLU (Rectified Linear Unit) and others can be used as activation function. The choice of a

particular function depends on the task and the architecture of the neural network. From the mathematical point of view, this scheme can be described as follows:

$$y = f(w \cdot x + b), \quad (8)$$

Where y is the output signal, x is the input vector, w is the weight vector, b is the bias, and f is the activation function.

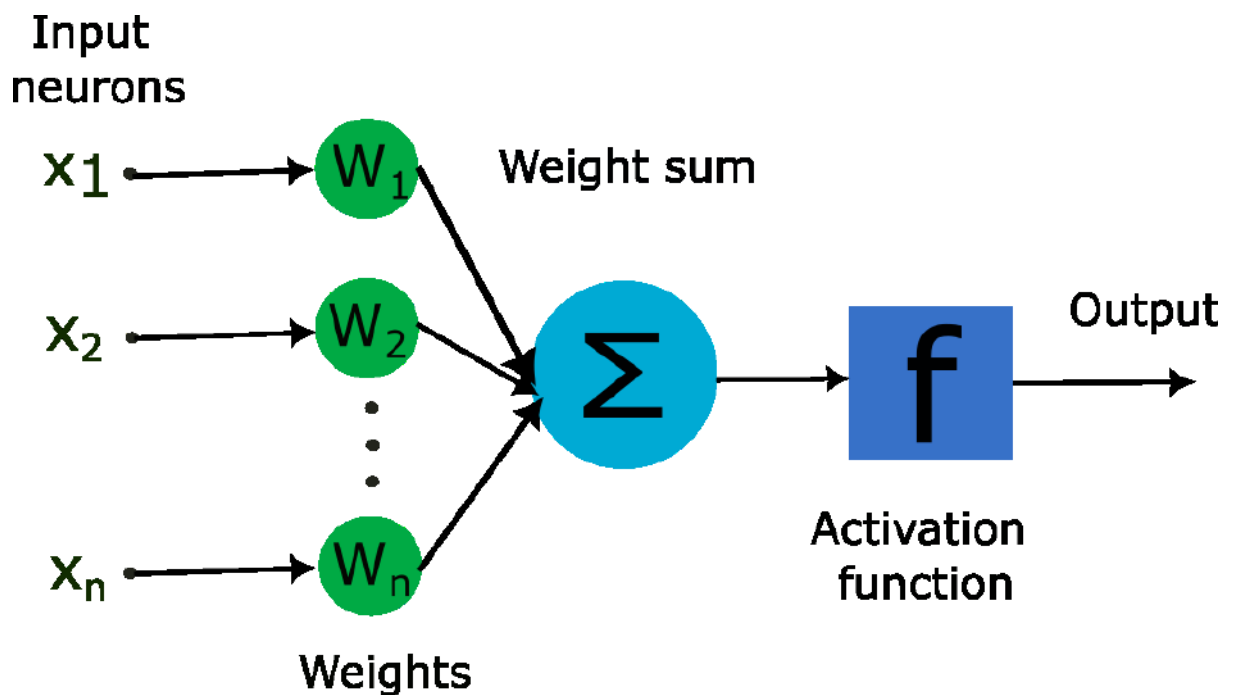


Figure 8. Scheme of a neuron of the artificial neural network.

The fundamental idea of NN involves the arrangement of neurons in layers: an input layer, one or more hidden layers, and an output layer (Fig. 9). Each neuron in one layer connects to neurons in the next layer. Then the process of training the neural network, during which the difference between the predicted and actual results is minimized, is a change in the bias and weighted sums in the adders of neurons.

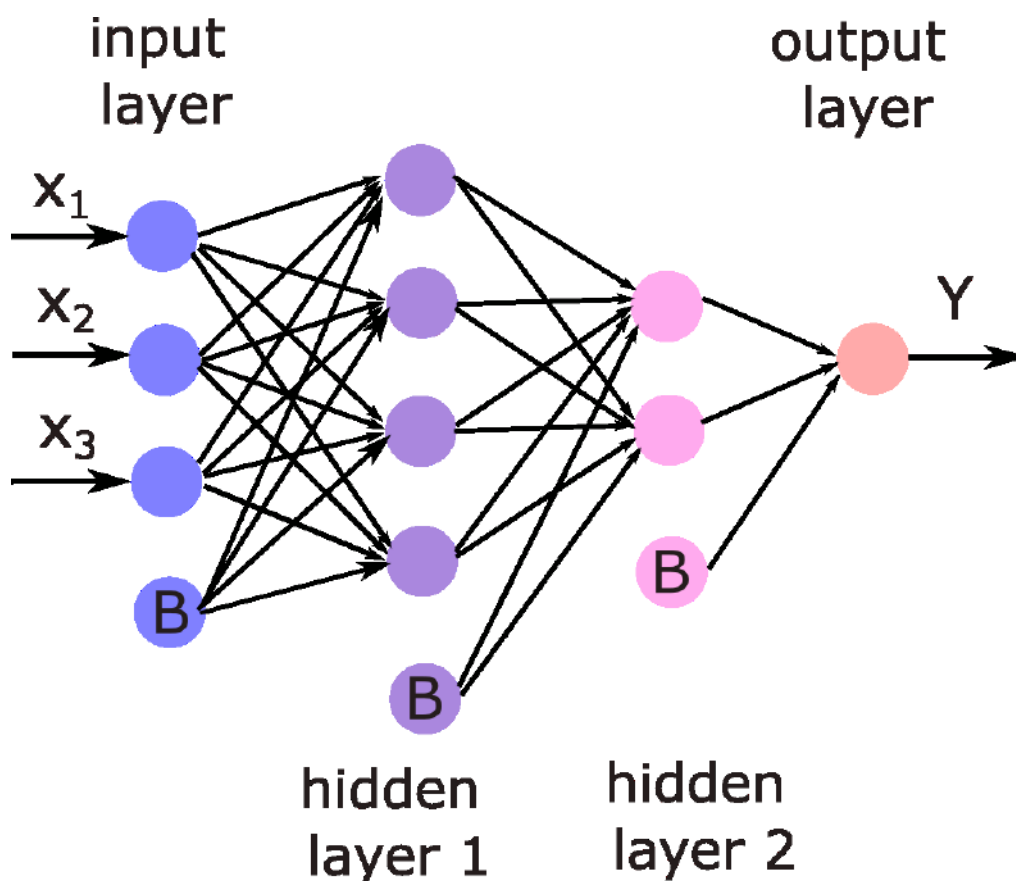


Figure 9. Multilayer neural network with three inputs ($X_i = 1, 2, 3$), two hidden layers, and one output (Y). B - bias neurons.

Neural networks, due to their complex structure, allow us to establish dependencies in datasets where descriptors are nonlinearly related to a property. However, NNs have disadvantages - they are easily overtrained, especially when the number of descriptors is large, and are difficult to interpret, which makes it difficult to understand the influence of descriptors on properties.

1.4. Application of QSPR in analytical chemistry

Despite significant progress in QSPR, the development of accurate and reliable models for property prediction continues to be hampered by the limited amount and variety of experimental data. The variability of experimental conditions and the complexity of chemical systems, especially in the case of multicomponent mixtures such as deep

eutectic solvents, make it difficult to collect the homogeneous data sets needed to build generalizable models.

Nevertheless, QSPR modeling has been successfully applied in a number of areas of analytical chemistry where it has been possible to collect sufficient data from experiments performed under similar conditions. Such areas include the development of ion-selective sensors, optimization of extraction of target components from natural products using DES, and the development of new chromatographic phases.

One of the most important applications of QSPR is the development of ion-selective sensors. In a study by Solovyov et al. (2019) [60], applied QSPR using structural fragments of chemical compounds as descriptors to predict the potentiometric sensitivity of polymeric membrane sensors to heavy metal ions such as Cu^{2+} , Zn^{2+} , Cd^{2+} , and Pb^{2+} . This work demonstrates that despite a limited data set, QSPR models can achieve significant predictive power, as evidenced by the low RMSE value (about 5 mV/dec) and high coefficient of determination (R^2 about 0.8) in cross-validation. This suggests that QSPR can effectively complement traditional experimental approaches by predicting how different ligands in PVC-plasticized membranes will respond to specific ions.

Subsequently, Martynko et al. (2020) studied the selectivity of $\text{Mg}^{2+}/\text{Ca}^{2+}$ sensor membranes using different amide ligands [61]. This study demonstrates how QSPR models can facilitate the early stages of sensor development by identifying promising candidates prior to synthesis and measurements.

QSPR modeling can not only predict sensor sensitivity, but also enable the development of more selective and efficient sensor systems. For example, a study by Paurbashir (2023) uses a combination of genetic algorithms and support vector machines to predict selectivity to carbonates for sensors with PVC-plasticized membrane [62].

Similarly, Kiani-anbui and colleagues applied QSPR to model electrode selectivity coefficients to lanthanum [63]. Their approach involves using density functional theory

to calculate descriptors, which allows for more detailed and accurate prediction of sensor performance.

A novel application of QSPR is illustrated in Phan Ti Diem-Tran et al. (2023), where QSPR models were developed to predict the stability and potentiometric sensitivity of complexes formed by heavy metal ions with organic fluorescent compounds [64]. This study focused on coumarin-like ligands, which are known for their fluorescent properties. QSPR models showed excellent performance, with quadratic correlation coefficients (R^2) as high as 1.00, demonstrating the accuracy of the method in predicting ligand interactions with metals such as Cu^{2+} , Cd^{2+} , and Pb^{2+} . By analyzing a virtual database of coumarin-like structures, the researchers identified several new ligands with promising sensitivity and stability constants. Subsequent quantum chemical calculations and toxicological profiling helped to evaluate their suitability as fluorescent chemosensors for heavy metals in industrial wastewater.

Deep eutectic solvents are a new type of solvent introduced by Abbott and colleagues about two decades ago [65]. DESs typically consist of a hydrogen bond donor (HBD) and a hydrogen bond acceptor (HBA), which together form a eutectic mixture with a melting point significantly lower than that of each component individually. This feature allows DESs to remain liquid at room temperature, making them ideal for a variety of chemical applications due to their increased solubility, reduced volatility and adjustable viscosity.

Since DESs are multicomponent systems, it is common to model any one of the components when applying the QSPR approach to them.

Lemaoui et al. first applied QSPR to predict the properties of deep eutectic solvents using QSPR models and COSMO-RS sigma profiles as molecular descriptors [66,67]. The predicted DES conductivity values obtained from the developed models were in good agreement with experimental data.

Salakhshouri et al. extended the application of QSPR for DES by developing a hybrid model combining QSPR and Gaussian process regression to predict CO_2 solubility

in different DES compositions [68]. This study demonstrated the potential of QSPR models as a tool to accurately predict interactions between DES components and dissolved gases, which can be used to design and optimize solvent systems for carbon capture.

In addition, QSPR models have found applications in optimizing the extraction efficiency of target components from natural products using DESs. For example, Hu et al. used QSPR models in combination with COSMO-SAC models to investigate the effect of DES composition on the extraction efficiency of oleocanthal from olive oil [69]. Thanks to the models built, it was possible to accurately predict the extraction results and select the most efficient DES compositions for the process.

The development of chromatographic phases is a necessary process for the development of chromatography, an extremely widely used analytical technique. The QSPR approach can be applied to develop new phases designed for specific analytical purposes, thereby improving the efficiency and specificity of chromatographic separation of complex mixtures. The first studies in the field of retention time prediction were carried out by Fedorova et al. in [70], where they studied retention time prediction using deep learning models in high performance liquid chromatography (HPLC). Different architectures of deep learning models were trained on molecular fingerprints and SMILES representations. The best performance was demonstrated by a one-dimensional convolutional neural network (1D CNN), which used SMILES as input.

Based on the study by Fedorova, Obradovich et al. further improved the predictive capabilities of QSPR models in their comparative study of descriptor calculation tools for modeling the retention mechanism of isomeric compounds [71]. This study extended previous work by applying a broader set of computational tools and machine learning techniques, including SVM, MLR and XGBoost, as well as feature selection methods such as genetic algorithm and stepwise regression.

Despite its successes, QSPR modeling in analytical chemistry faces a number of unresolved challenges that require further research.

First, the potential of QSPR to predict the properties of deep eutectic solvents remains poorly understood. Existing approaches based on modeling of individual components of DESs do not always allow taking into account complex interactions between components, which may limit the accuracy of predicting the properties of the mixture as a whole. The development of new QSPR models capable of predicting the properties of DESs based on their composition is an urgent task, the solution of which will expand the field of application of these solvents.

Secondly, there is an urgent need to develop new approaches to QSPR modeling of sensor selectivity, especially for the determination of hydrophilic anions such as carbonate ions. Due to the limited number of available ionophores and the difficulty in achieving high selectivity, the development of efficient sensors for carbonate ions is a significant challenge. Given the importance of carbonate detection in environmental studies, food industry and other fields, the development of new QSPR models for predicting and optimizing the selectivity of sensors for carbonate ions is a highly relevant task.

1.5. Aim and tasks of the work

Analysis of the applicability of the QSPR approach to various objects of analytical chemistry, combined with a review of existing descriptors and modeling methods, has shown that, despite the promising nature of this approach, the amount of research in this area is limited. This is due to the difficulties in obtaining large homogeneous datasets required for the construction of classical QSPR models.

Nevertheless, solving this problem seems to be crucial for the development of analytical chemistry. Therefore, in this work, we have chosen two research areas with high potential for practical application and allowing us to collect sufficient experimental data to build reliable QSPR models: ion-selective electrodes and deep eutectic solvents.

Accordingly, the main objective of this study is to extend the scope of QSPR applications in analytical chemistry by developing models to predict the properties of ion-selective electrodes and deep eutectic solvents. To achieve this goal, the following three objectives were set:

1. Development and validation of QSPR models to predict the selectivity of ion-selective sensors to carbonate-ions (CO_3^{2-}).

Determination of anions such as carbonate ion (CO_3^{2-}) using ion selective electrodes (ISEs) is a challenging task in analytical chemistry. The development of selective ligands for anion binding is generally difficult compared to cations because inorganic anions are characterized by a lower charge-to-radius ratio, a greater variety of geometries, and their forms of existence in solution are strongly pH dependent.

In addition, the development of highly selective sensors for carbonate ions is difficult due to their hydrophilicity, which prevents effective binding to ionophores in the hydrophobic sensor membrane.

Currently, the most successful commercial ionophores for the determination of carbonate ions are based on trifluoroacetophenone and its derivatives. The key structural element of these compounds is a carbonyl group with reduced electron density on oxygen due to the strong acceptor influence of the $-\text{CF}_3$ group. This suggests that other compounds containing acceptor substituents near the carbonyl group may also exhibit high selectivity to carbonate ions.

However, experimental validation of each potential ionophore is time and material intensive. The development of QSPR models capable of predicting the selectivity of ISEs to carbonate ions based on the structural characteristics of ionophores will significantly accelerate and optimize the process of searching for new effective ligands. Such models use structural fragments of ionophore molecules as descriptors, as well as membrane plasticizer characteristics encoded using specialized software.

2. Development and validation of QSPR models to predict the sensitivity of ion-selective sensors to heavy metal cations (Ca^{2+} , Cd^{2+} , Pb^{2+}).

The determination of heavy metals (Ca^{2+} , Cd^{2+} , Pb^{2+}) in the environment, food and biological samples is an important task for environmental monitoring, quality control and medical diagnostics. Ion-selective electrodes represent a convenient and affordable tool for the determination of metal ion concentrations, but their application in real-world applications is often limited by insufficient sensitivity.

One of the promising directions for improving the characteristics of ISE is the development of new ionophores - substances that provide selective binding of metal ions in the sensor membrane. Diphenylphosphorylacetylides represent a new class of ligands with high extraction ability towards metal ions, which makes them promising candidates for the creation of new ionophores. [72]

However, experimental validation of each potential ionophore is a costly and time-consuming process. The development of QSPR models to predict the sensitivity of ISEs based on the structural characteristics of ionophores will greatly accelerate and simplify the process of searching for new effective ligands. Structural fragments of ionophore molecules encoded by specialized software are used as descriptors in such models.

3. Development and validation of QSPR models for predicting physicochemical properties of DESs with organic acids as hydrogen bond donors.

Deep eutectic solvents are a new class of solvents with unique properties such as low volatility, high solubility and tunable viscosity. These properties make DESs promising for applications in various fields including extraction, electrochemistry, catalysis, and new materials development. However, finding optimal HER compositions for specific applications is difficult due to the large number of possible combinations of components and the difficulty in experimentally determining their properties.

This objective is to develop QSPR models that allow predicting the physicochemical properties of DESs, such as density, conductivity, viscosity and

refractive index, based on the structural characteristics of organic acids acting as hydrogen bond donors in the HER composition.

Not only structural fragments of organic acid molecules are used as descriptors, but also semi-empirical descriptors such as HOMO and LUMO energies, dipole moment and heat capacity calculated using the PM3 quantum chemical method. This will allow to take into account the electronic properties of the molecules and improve the prediction accuracy.

The choice of DES with organic acids as hydrogen bond donors is due to their wide use and structural diversity, which allows the properties of DES to be varied over a wide range. Density, conductivity, viscosity and refractive index are key physicochemical properties of DES that determine their applicability in various fields, namely: density and viscosity affect the efficiency of extraction and separation of mixture components, conductivity is a critical parameter for electrochemical processes such as electrolysis and electrodeposition, physicochemical properties of DES can affect the activity and selectivity of catalysts.

The development of such models will make it possible to optimize the composition of DES for specific applications, accelerate the process of development of new DES with improved characteristics.

Chapter 2. APPLICATION OF QSPR MODELING FOR PREDICTING SELECTIVITY OF CARBONATE IONSELECTIVE ELECTRODES

This chapter investigates the possibility of using QSPR to predict the potentiometric selectivity of plasticized polymer membrane sensors based on new ligands. Special attention is paid to the development of sensors with selectivity to hydrophilic anions, which is an important task in environmental monitoring, water and food quality control, and medical diagnostics.

Despite significant progress in the development of ion-selective electrodes, the development of sensors with high selectivity to carbonate ions remains a challenge. This is due to the hydrophilicity of the carbonate ion and the limited number of available ionophores that can effectively bind it in the hydrophobic sensor membrane.

QSPR modeling represents a promising approach to solve this problem, allowing the prediction of sensor selectivity based on the structural characteristics of ionophores. In this chapter, the application of QSPR modeling to predict the selectivity of carbonate ISEs based on a dataset of 40 ionophores described in the literature is discussed. Structural fragments of ionophore molecules were used as descriptors. In addition, data on the selectivity of chloride ionophores and membrane plasticizer characteristics were added to expand the dataset and to account for the influence of the sensor matrix on its response.

Analysis of the obtained QSPR models allows us to evaluate their predictive ability and identify key structural features of ionophores responsible for selectivity to carbonate ions, which can be used for the directed search for new ligands with improved characteristics and the development of highly selective carbonate sensors.

2.1. Experimental section

2.1.1. Selection of descriptors

Following the methodology outlined in Chapter 1.1, the first stage of the experiment is to collect data for further processing and model building. "The dataset for QSPR modeling included 40 structures with known $\log K^{sel}(HCO_3^-/Cl^-)$ described in the literature (the structures are listed in Table 1 in the Appendix A) [73–79]. Since the number of papers reporting the successful development of carbonate potentiometric sensors is small, in order to expand the data set with additional samples, several ionophores showing low carbonate selectivity were also added since these substances were chloride ionophores. In these cases, the reported $\log K^{sel}(Cl^-/HCO_3^-)$ were converted to $\log K^{ce\pi}(HCO_3^-/Cl^-)$ according to the Nikolsky-Eisenman equation:

$$E = E_I^0 + \frac{RT}{z_I F} \ln \left(a_I + \sum K_{IJ} a_J^{z_I/z_J} \right), \quad (9)$$

where a_I , a_J , z_I , z_J – are the activities and charges of the main and interfering ions, respectively, and K_{IJ} is the selectivity coefficient. The value of the selectivity coefficient can be found according to the expression:

$$K_{IJ} = a_I / a_J^{z_I/z_J}, \quad (10)$$

then we can calculate the logarithm of the selectivity coefficient K_{IJ} under the condition that the charges of ions I and J are equal. Thus, in our case

$$\log K^{ce\pi}(Cl^-/HCO_3^-) = -\log K^{ce\pi}(HCO_3^-/Cl^-)$$

Uniformity of experimental conditions is very important to obtain a quality data set for QSPR modeling. In collecting data from the literature, we limited ourselves to those papers that used polyvinyl chloride (PVC) as the polymer for the sensor membrane matrix. The limited number of studies on carbonate sensor selectivity makes it difficult to compile a suitable dataset for modeling that uses the same membrane polymers and plasticizers. Therefore, we further considered those works that used o-nitrophenyl octyl

ether (NPOE), dioctyl adipate (DOA) or bis(2-ethylhexyl) sebacinate (DOS) as plasticizers. In this case, the dielectric constant of these plasticizers was added to the data set as a descriptor to account for differences in sensor properties due to different plasticizers in the membrane compositions. Tridodecylmethylammonium (TDMA) was used as an anion exchanger in all literature sources.

In addition, we considered the conditions under which sensor selectivity measurements were made and only considered literature data obtained in the pH range of 7.0-8.6. This was done to ensure that HCO_3^- is the dominant ionic form in solutions.

The method of measuring selectivity was also evaluated - all $\log K^{ce\pi}(HCO_3^- / Cl^-)$ in the selected literature were obtained by the split-solution method, with the exception of eight substances from [77], where the reduced potential method was used. However, it was decided to keep these eight samples in the data set as the two methods usually give comparable results. The selectivity of $\log K^{ce\pi}(HCO_3^- / Cl^-)$ in the literature dataset ranged from -5.8 to 6.2. The mean value was -1.4 and the median value was -2.6.

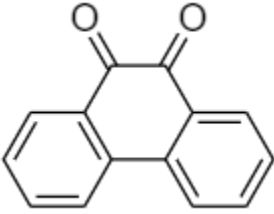
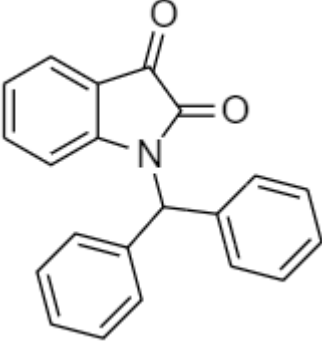
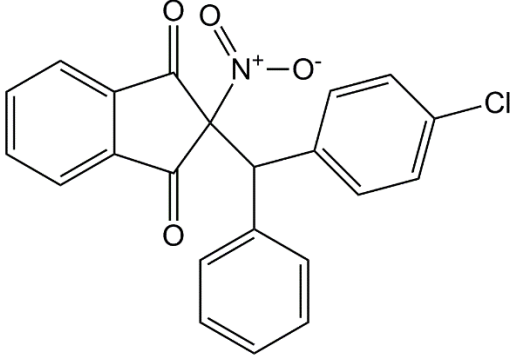
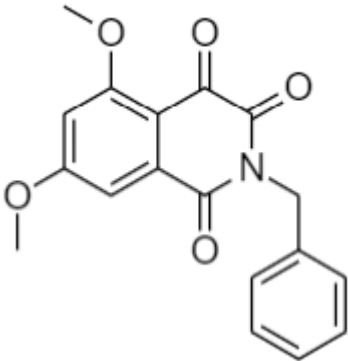
2.1.2. Ion-selective sensors with new ligands in composition

To study the possibility of predicting the selectivity of carbonate sensors using the obtained QSPR model based on literature data, sensors based on new ligands were designed and investigated.

The structures of the ligands are summarized in Table 1. The choice of these compounds for synthesis was due to the fact that most of the known carbonate ionophores are based on trifluoroacetophenone and its derivatives. The carbonyl group with reduced electron density on oxygen due to the strong acceptor influence of the -CF₃ group is the key moiety responsible for binding the carbonate ion. Substance 1, commercially available, was purchased from Merck (Darmstadt, Germany) and was used without further purification. Ligands 2-4 (Table 1) were obtained from an original

synthesis. All four substances contain different acceptor substituents near the carbonyl group, suggesting their potential ability to efficiently bind carbonates.

Table 1. Structures of the new ligands investigated in the experiment

#	Structure of the ligand	IUPAC name
1		Phenanthrenequinone S1
2		1-(diphenylmethyl)-1H-indol- 2,3-dione S2
3		2-((4- chlorophenyl)(phenyl)methyl)- 2-nitro-1H-indan-1,3-dione S3
4		5,7-Dimethoxy-2-benzyl- 1,3,4(2H)-isoquinolintrione S4

Four new ligands were used to prepare PVC-plasticized sensor membranes. The compositions of the membranes are summarized in Table 2. Each membrane contained 50 mmol/kg ionophore and 10 mmol/kg TDMA-NO₃ as an anion exchanger. The ratio between PVC and solvent-plasticizer was 1:2, and the total weight of the membrane was 300 mg. NPOE was used as the solvent-plasticizer.

Table 2. Compositions of sensor membranes, wt. %.

Sensor	PVC	TDMA NO ₃	Plasticizer	Ligand
S1	32.79	0.60	65.57	1.04
S2	32.61	0.60	65.22	1.57
S3	32.48	0.60	64.96	1.96
S4	32.59	0.60	65.18	1.63

Polymer membranes for sensors were prepared according to a standard protocol. Suspensions of all membrane components were dissolved in 3 ml of freshly distilled tetrahydrofuran (THF) in a glass beaker using a magnetic stirrer. The resulting solutions were poured into 20 mm diameter Teflon beakers and left to dry for 48 h. Discs of 8 mm diameter were cut from the initial membranes and attached to the PVC sensor bodies using a mixture of PVC and cyclohexanone. The thickness of the prepared membranes was 0.4 mm. After the adhesive dried, the inner parts of the obtained electrodes were filled with a mixture of 0.01 M NaHCO₃ and 0.001 M NaCl. Chloride anion is necessary for the functioning of the Ag/AgCl electrode, and bicarbonate is necessary to ensure its constant content in the membrane phase required by the Nikolsky formalism. Water for the preparation of all aqueous solutions was obtained from a GFL 2102 distiller (GFL Burgwedel, Germany). The conductivity of the monodistillate is 2.2 s/cm at 25 C. Finally, internal Ag/AgCl electrodes were incorporated into the sensors. Three sensors were fabricated for each membrane composition.

Potentiometric measurements were performed using a multi-channel digital mV-meter KHAN-32 (Sensor Systems LLC, St. Petersburg, Russia) connected to a PC for data acquisition via a USB port. The measurements were performed in relation to the standard reference Ag/AgCl electrode ESr-10101 (Izmeritel'naya Tekhnika, Moscow, Russia). To control pH values in sample solutions, a glass pH-sensor PY-41 (GOnDO Electronic Co., Ltd., Taipei, Taiwan) was used.

Sensor sensitivity was studied in aqueous solutions of inorganic salts (sodium sulfate, sodium chloride, sodium nitrate, sodium bicarbonate, sodium monosodium phosphate) in the concentration range from 10^{-6} to 10^{-2} M. Sensitivity values were calculated as the slopes of the linear portions of the sensor response curves (10^{-4} - 10^{-2} M). The sensors were washed with several portions of distilled water before, after and between measurements until constant potential values were achieved.

Selectivity coefficients were obtained by the split-solution method, which is also known as the bi-ionic potential method [80].

E_I and E_J were calculated for solutions of the major and interfering ions with a concentration of 10^{-3} M, respectively.

2.1.3. QSPR modeling

Structural descriptors in the form of molecular graphs represented by substructural molecular fragments (SMF) were used as descriptors to describe the molecular structures of ligands.

The program "MolFrag", which is a part of the software package "ISIDA QSPR", was used to obtain SMFs [81]. SMF in ISIDA can be described in two ways: either as sequences of atoms and/or bonds (topological pathway) or as a selected ("augmented") atom (atom-centered fragments) with its environment, which can be atoms, bonds, or both. Both ways are illustrated in Figure 10.

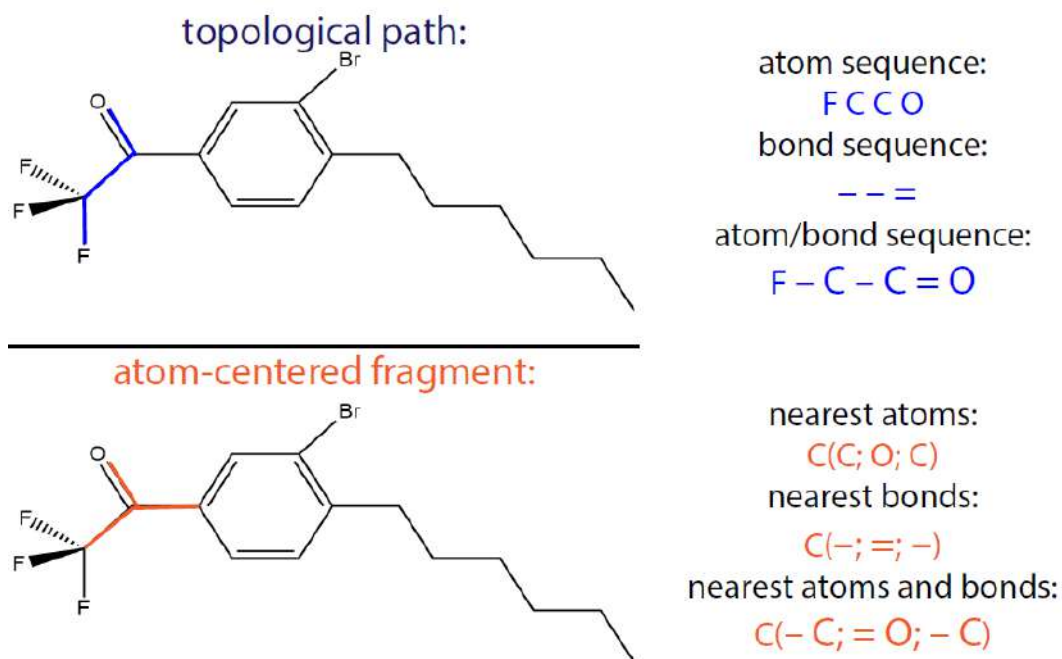


Figure 10. Two approaches to obtain substructure molecular fragments

In this study, sequences of atoms and bonds that satisfied the following conditions were chosen as structural descriptors to describe molecules: the descriptor represents the shortest path from one atom to another; its length ranges from 4 to 10 atoms. The latter restriction is due to the fact that short fragments do not carry significant information about the structure of the molecule, are represented in a larger number relative to longer descriptors, and are present in most compounds. In contrast, descriptors with lengths longer than 10 atoms are represented in an extremely small number of compounds in the dataset, often only one.

The descriptor generation resulted in a 40 * 1103 matrix corresponding to the number of literature compounds * calculated structural descriptors, which was supplemented with another column containing the dielectric constants of the plasticizers used in the preparation of the sensor membrane. The resulting matrix was used to construct a regression model relating these descriptors to the selectivity values $\log K^{sel}(HCO_3^- / Cl^-)$ of the corresponding potentiometric sensors.

The PLS algorithm was used to construct multivariate regressions. PLS models were calculated in The Unscrambler 9.7 software package (CAMO, Norway). To optimize the number of latent variables and the number of descriptors in the model, full cross-validation was applied using the mean-square error of cross-validation as a criterion.

Before modeling, the descriptor values in the X matrix were auto-scaled (the column mean was subtracted from each item and the result was divided by the column standard deviation). The resulting model was used to predict $\log K^{sel}(HCO_3^-/Cl^-)$ for the new ligands.

2.2. Discussion of results

The molecular descriptor matrix obtained for the 40 ligands was related to their $\log K^{sel}(HCO_3^-/Cl^-)$ values using PLS regression. The number of variables in the original model was optimized by analyzing the regression coefficient values of the model. All variables having regression coefficients in the range $[-5 \cdot 10^{-3}; 5 \cdot 10^{-3}]$ were excluded from consideration, reducing the number of variables to 585 from the original 1103. The resulting QSPR model is presented in Figure 11. Based on the statistical performance of the model (RMSECV and R^2 values), it can be concluded that the obtained dependence is suitable for semi-quantitative prediction of the potentiometric selectivity of ionophores. The values of $\log K^{sel}(HCO_3^-/Cl^-)$ range from -6 to +6 in the modeled data set, and the RMSECV value is 1.5. Thus, the model is able to distinguish between weakly, moderately and highly selective carbonate ligands. This seems to be a promising result, given the wide chemical diversity of the ligands, the relatively small (in the scale of the QSPR approach) data set, the use of literature data obtained under dissimilar conditions, and the overall simplicity of the approach.

Analysis of the PLS model regression coefficients allows us to judge the importance of certain descriptors and their contribution to the selectivity values. The greatest contribution is made by the fragments with the highest absolute values of regression

coefficients. The calculated fragments (model variables) were filtered by the presence of distinguishing ligands in the chemical structures. Structural descriptors present in at least five ligands were selected for analysis. The remaining fragments were then analyzed for repeatability in other fragments, and the smallest fragments were selected from a number of equivalent fragments. Thus, the C-C-C-C-O-C=C fragment contains a shorter C-O-C=C fragment, and this shorter fragment was retained for analysis. Finally, the fragments were sorted by the value of their respective regression coefficients, and only fragments with absolute values of $b > 0.01$ were considered. The resulting variable importance diagram is presented in Figure 12.

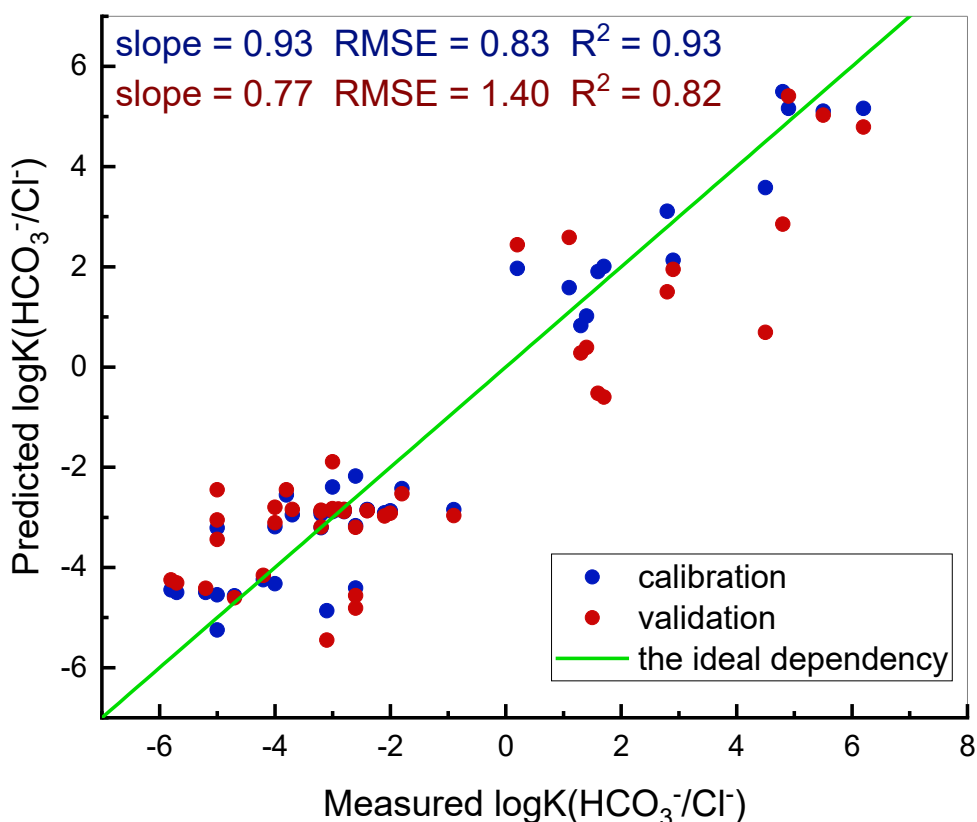


Figure 11. "Measured and predicted values" plot of the cross-validated QSPR model for predicting membrane sensor selectivity. The green line corresponds to a perfect match between measured and predicted values

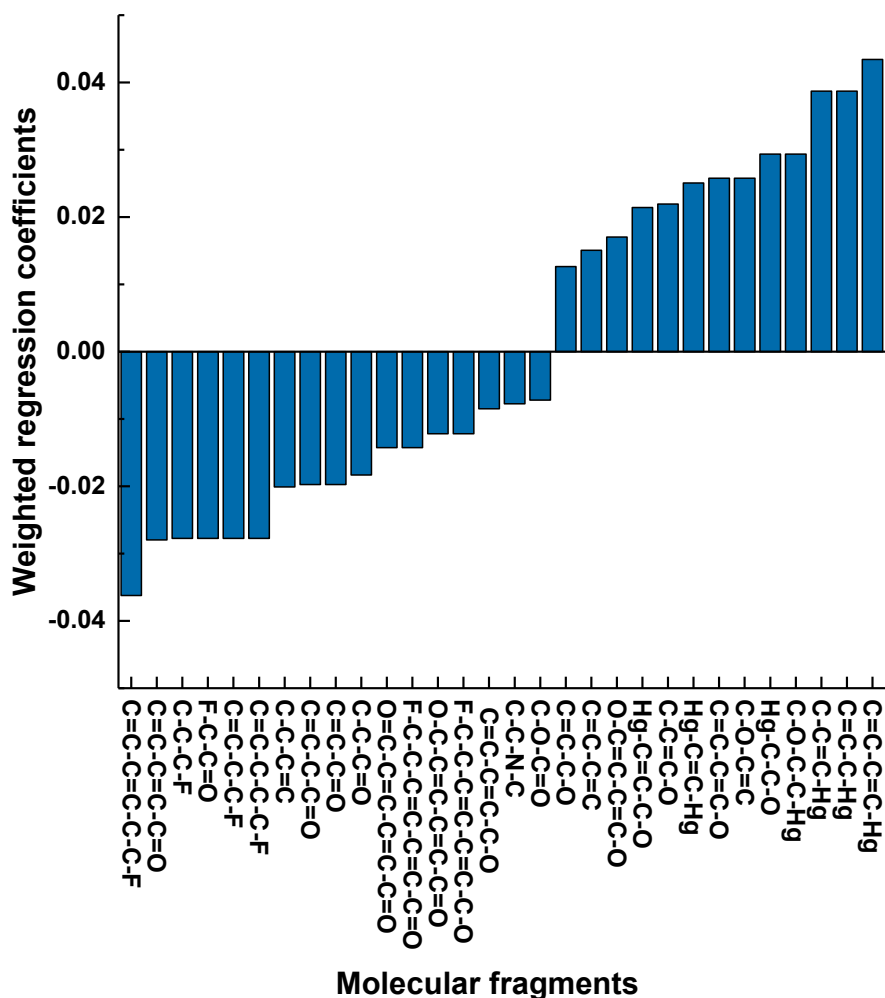


Figure 12. Fragments with the largest contribution to $\log K^{sel}(HCO_3^-/Cl^-)$, where "-" is a single bond and "=" is a double bond

The diagram shows that the largest negative contribution corresponds to the fragment C=CC=C-C-C-C-F. Since the modeled value is $\log K^{sel}(HCO_3^-/Cl^-)$, and the more negative it is, the higher the selectivity to HCO_3^-/Cl^- , implying that such C=C-C-C=C-C-C-C-F fragments contribute more selectivity to the hydrogen carbonate anion. It should be noted that among the 17 fragments with negative contribution, there are only 2 fragments that do not contain fluorine (F-) or C=O groups. These are the fragments C=C-C-C=C-C-O and C-C-N-C, which also have a significant negative contribution among the selected fragments (rank 15 and 16 out of 17).

The ten fragments with the highest negative contribution come from the trifluoroacetophenone fragment. Trifluoroacetophenone is incorporated into a significant number of existing carbonate ionophores. An excellent example of an ionophore with this group and high selectivity $\log K^{sel}(HCO_3^-/Cl^-) = -3.1$ is N, N-Dioctyl-3, 12-bis(4-trifluoroacetylbenzoyloxy)-5-cholan-24-amide (carbonate ionophore VII from the Merck catalog). The structure of this compound is shown in Figure 13. It is known that the TFA group promotes the binding of the carbonate anion through the formation of hydrogen bonds [82].

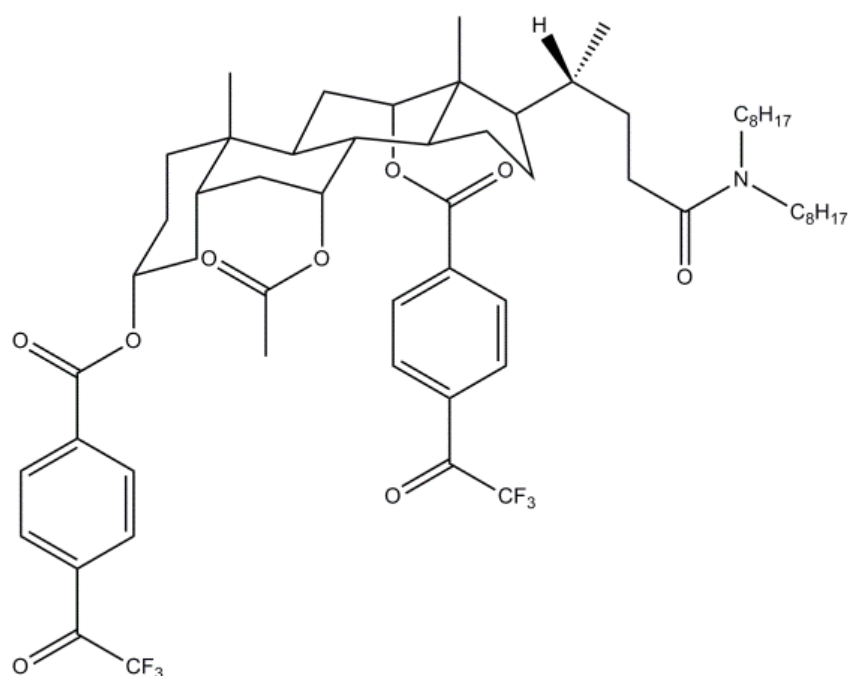


Figure 13. Chemical structure of N, N-Dioctyl-3, 12-bis(4-trifluoroacetylbenzoyloxy)-5-cholan-24-amide

The fragment with the largest positive contribution is C=C-C=C-Hg. Only 6 fragments out of 13 fragments with positive contribution do not contain mercury (Hg-) in their composition. In descending order of contribution values, these are fragments C-O-C=C, C=C-C=C-O, C-C=C-O, O-C=C-C-C=C-O, C=C-C=C and C=C-C-O. These fragments belong to the ionophores proposed for chloride determination and therefore have low selectivity towards carbonate [79]. An example of such an ionophore is $\{\mu\text{-}[4,5\text{-dimethyl-}$

3,6-bis(octyloxy)-1,2-phenylene}}bis(trifluoroacetato-0) dimercurium, whose structure contains all fragments with positive contributions, resulting in $\log K^{sel}(HCO_3^-/Cl^-) = 5.5$. The structure of this ligand is given in Figure 14.

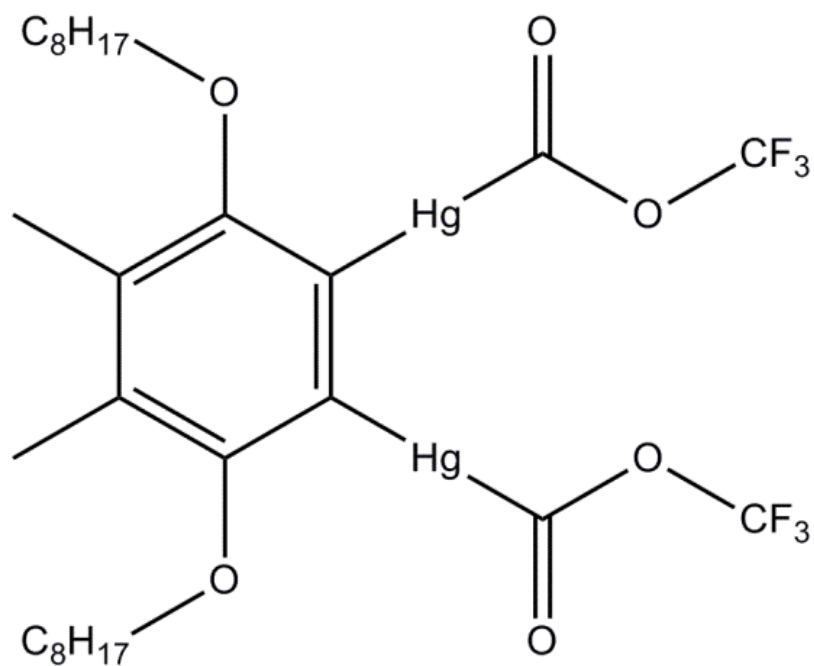


Figure 14. Chemical structure of $\{\mu\text{-}[4,5\text{-dimethyl-}3,6\text{-bis(octyloxy)-}1,2\text{-phenylene}]\}\text{bis(trifluoroacetato-}0\text{) dimercurium}$

Thus, it can be stated that the constructed QSPR model is in good agreement with chemical considerations about the structure of the ligands, and the significance of the different fragments is consistent with chemical intuition.

Polymeric sensor membranes prepared using four new ligands were studied for their sensitivity to inorganic anions and HCO_3^-/Cl^- selectivity. Typical response curves of the new sensors are shown in Figure 15 and the calculated sensitivity values are shown in Table 3. It should be noted that sensors S2 and S3 did not demonstrate sufficient sensitivity to carbonate.

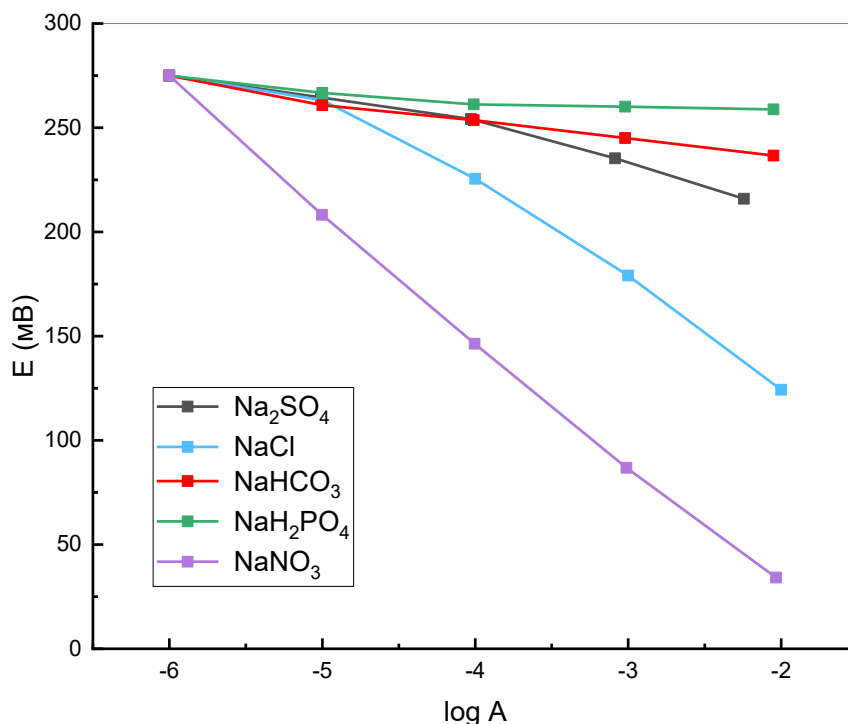


Figure 15. Typical potentiometric response curves of the S2 sensor. The EMF readings are shifted to zero for clarity

Table 3. Sensitivity values of sensors in the solution of tested anions, mV/dec

	SO ₄ ²⁻	Cl ⁻	HCO ₃ ⁻	H ₂ PO ₄ ⁻	NO ₃ ⁻
S1	1.7 ± 0.3	-6.4 ± 1.0	-43.4 ± 3.8	29.6 ± 1.1	-27.5 ± 0.9
S2	-16.6 ± 1.1	-41.0 ± 1.2	-7.2 ± 3.4	-1.5 ± 1.2	-57.8 ± 0.8
S3	-24.0 ± 1.6	-42.5 ± 3.8	-16.4 ± 4.2	-3.5 ± 1.4	-64.9 ± 5.9
S4	-1.2 ± 1.2	-23.8 ± 1.0	-28.7 ± 1.6	12.4 ± 1.4	-47.1 ± 0.7

Using a QSPR model derived from literature data, the selectivity of these four novel membrane sensor compositions was predicted. For this purpose, each of the ligands was described using the same set of SMFs and the resulting data were fit to a filtered set of 585 variables. The ligand structure descriptions thus obtained were used in the QSPR model to calculate the predicted $\log K^{sel}(HCO_3^-/Cl^-)$. The results of this prediction

are summarized in Table 4 together with the selectivity data obtained in a conventional potentiometric experiment using the split-solution method.

Table 4. Comparison of predicted and experimental values of $\log K^{sel}(HCO_3^- / Cl^-)$ for the new ionophores

Sensor	Experimental value $\log K^{sel}(HCO_3^- / Cl^-)$	Predicted value $\log K^{sel}(HCO_3^- / Cl^-)$
S1	-3.6	-2.5
S2	-0.7	-3.4
S3	-1	-2.9
S4	-2.5	-2.7

Considering the RMSECV value of the QSPR model equal to 1.4 $\log K^{sel}(HCO_3^- / Cl^-)$, the agreement between the predicted and experimental selectivities is found to be satisfactory in two out of four cases (S1, S4). Compound S1 has C=C-C-C=C-C=O, C-C-C=C, C=C-C-C=O, C=C-C=O, C=C-C=O, C-C-C=O fragments with negative contribution and C=C-C=C fragment with positive contribution. The model predicted the logarithm of the selectivity coefficient with a value of -2.5, while the experimentally found selectivity coefficient is -3.6. Compound S2 contains a C-C-N-C fragment and the remaining fragments are the same fragments as in the first compound. All these fragments except C=C-C-C=C contribute negatively to the logarithm of the selectivity coefficient. Thus, there is a marked deviation between the predicted $\log K^{sel}(HCO_3^- / Cl^-)$ of -3.4 and the experimental value of -0.7. There are only a few fragments in compound S3 that have a significant effect on the model. These are C-C-C-C=C fragments with negative contribution and C=C-C-C=C with positive contribution. Since the model has only two structural fragments with significant influence for this compound, the difference between the model prediction (-3.0) and the experimental data (-1.0) is also high. The rather large discrepancy between prediction and experiment for S2 and S3 can also be

explained by the fact that these sensors actually did not show sufficient response to carbonate (Table 3).

Compound S4 includes fragments such as C=C-C-C=C-C=O, C=C-C=O, C=C-C=O, C-C-C=O, C-C-C=C, O-C-C-C=C-C=C-C=C-C=O and C-C-N-C with negative contribution and C-O-C=C, C=C-C=C-O, C-C=C-O and C=C-C=C with positive contribution. Since the number of important fragments is quite high, the agreement between experiment and prediction is also quite high (-2.7 vs. -2.5, respectively).

Thus, in order to correctly predict the properties of new ligands using the QSPR model, it is necessary that the same important fragments are present in their structure as in the ligands used to build the model. This conclusion follows from chemical considerations and mathematical principles, since the diversity of the training set determines the scope of applicability of the model.

Conclusion to Chapter 2

A QSPR model linking the structure of carbonate ionophores to the selectivity $\log K^{sel}(HCO_3^-/Cl^-)$ of the corresponding potentiometric sensors with plasticized polymer membrane was developed based on literature data on 40 ionophores. Analysis of the regression coefficients in the PLS model allowed the identification of important structural fragments affecting selectivity.

Four new ligands with different acceptor substituents at the carbonyl group synthesized for this work were used to test the predictive ability of the model. Two of these ligands showed potentiometric sensitivity to hydrocarbonate ions. Good agreement between predicted and experimental selectivity values was observed for these ligands, especially in the case of ligands sharing important molecular moieties with the ionophores used for model construction.

These results indicate the potential of the QSPR approach for the development of new anion-selective sensors, allowing the prediction of selectivity based on the structural characteristics of the ionophores.

Chapter 3. APPLICATION OF QSPR MODELING FOR PREDICTING THE SENSITIVITY OF IONOSELECTIVE ELECTRODS TO HARD METAL CATHIONS (Cu²⁺, Cd²⁺ AND Pb²⁺)

This chapter investigates the possibility of applying the QSPR method to predict the potentiometric sensitivity of plasticized polymer membrane sensors based on the chemical structure of the ionophore. The QSPR model was developed based on literature data on the sensitivity of previously studied, structurally similar ionophores. The resulting model was shown to be highly effective in establishing the relationship between ionophore structures and their sensitivity to Cu²⁺, Cd²⁺, and Pb²⁺. The model predictions for four new diphenylphosphorylacetamide-based ionophores were compared with experimental data obtained for these ionophores and showed satisfactory agreement, confirming the validity of the proposed approach.

Despite the successful application of QSPR modeling for predicting the sensory properties of ionophore-based potentiometric sensors, including sensitivity and selectivity parameters, to date this approach has not been used to predict the sensitivity of new membrane sensors based on yet unexplored ionophores. This work aims to fill this gap by developing a QSPR model based on literature data on sensitivity to Cu²⁺, Cd²⁺, and Pb²⁺ for structurally similar ionophores. The choice of these particular metals was dictated by their toxicity and hence the expressed interest in these elements in environmental studies.

3.1. Experimental section

3.1.1. Selection of descriptors

To build QSPR models for predicting potentiometric sensitivity, literature data on the responses of sensors with PVC-plasticized membranes to metal ions Cu²⁺, Cd²⁺ and Pb²⁺ were collected. The choice of these particular metals was dictated by their toxicity and hence the expressed interest in these elements in environmental studies. Most of

the data set (35 structures and data on their potentiometric responses) was taken from [60]. To improve the training set on phosphorus-containing ionophores, six substances from [83] were added (#36-41, Table 1, Appendix)

The structures of ionophores, their IUPAC names and literature references are summarized in Table 1 in the Appendix.

The structures of the ionophores were described using molecular descriptors. In this work, similar to the choice of descriptors described in Chapter 2.1.1, substructural molecular fragments obtained using the ISIDA software package were used for this purpose. The main idea behind the derivation of such molecular descriptors is that a substructural fragment is a sequence of atoms of a chemical structure connected by chemical bonds. Each fragment corresponds to the number of times that fragment occurs in the molecule. These numbers are then assembled into a matrix where each row corresponds to a particular ionophore and each column indicates the number of times that a particular fragment occurs in the ionophore. In this study, fragments with lengths ranging from 2 to 9 atoms were considered. The resulting descriptor matrix has dimensionality 41×1095 , where 41 is the number of ionophores in the training set and 1095 is the number of different molecular descriptors found describing the ionophore structures that make up the training set.

3.1.2.

QSPR

modeling

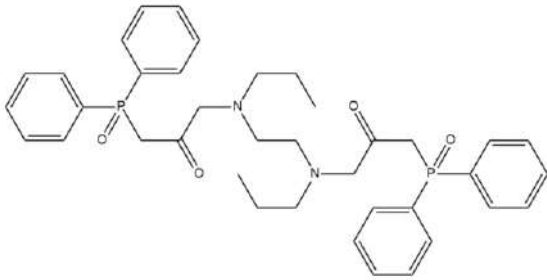
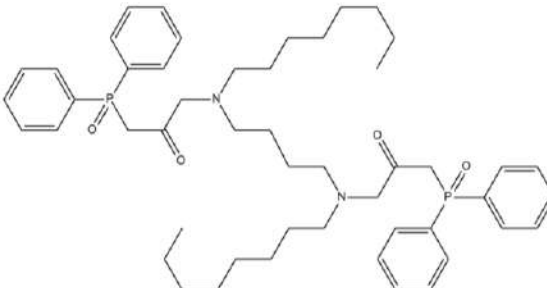
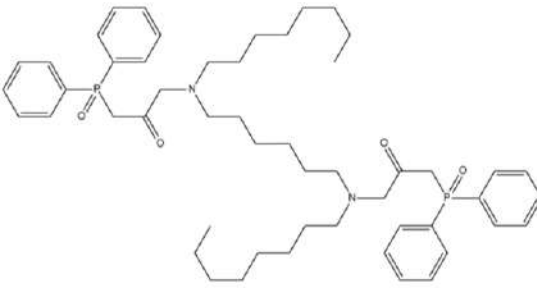
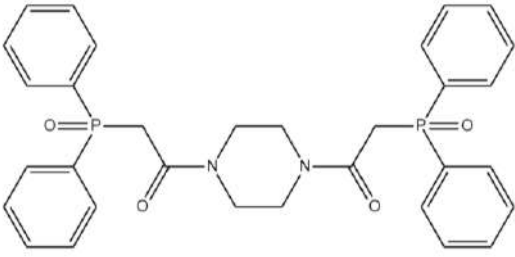
PLS regression was applied to find a mathematical relationship linking the chemical structures of ionophores represented by molecular descriptors and the potentiometric sensitivity of these ionophores to copper, cadmium and lead ions.

PLS models were subjected to full cross validation to find the optimal number of latent variables, this type of cross validation was chosen because of the relatively small size of the training set.

The calculated QSPR models were used to predict the potentiometric sensitivities to copper, cadmium and lead for the four newly synthesized ionophores. The structures

of the new substances and their names according to IUPAC nomenclature are summarized in Table 5.

Table 5. Structures of new ionophores and their names according to IUPAC nomenclature

	Chemical structure	IUPAC name
1		3,3'-(ethane-1,2-diy)lbis(propylazandiyl))bis(1-(diphenylphosphoryl)propan-2-one)
2		3,3'-(butane-1,4-diy)lbis(octylazandiyl))bis(1-(diphenylphosphoryl)propan-2-one)
3		3,3'-(hexane-1,6-diy)lbis(octylazandiyl))bis(1-(diphenylphosphoryl)propan-2-one)
4		1,1'-(piperazine-1,4-diy)lbis(2-(diphenylphosphoryl)ethanone)

3.2. Discussion of results

At the first stage of the experiment, three separate PLS models were constructed linking molecular descriptors of ionophores from the training set to their potentiometric sensitivities to copper, cadmium, and lead. Figure 16 shows a typical result of such modeling: a plot of the dependence of the measured sensitivity on the predicted sensitivity for cadmium. It can be seen that a reasonably good correlation is observed, and the cross-validation error was about 4 mV/dec. These results suggest that the formalized description of the ionophore using substructural molecular fragments can be used to predict the sensitivity of the corresponding plasticized polymer sensor membranes.

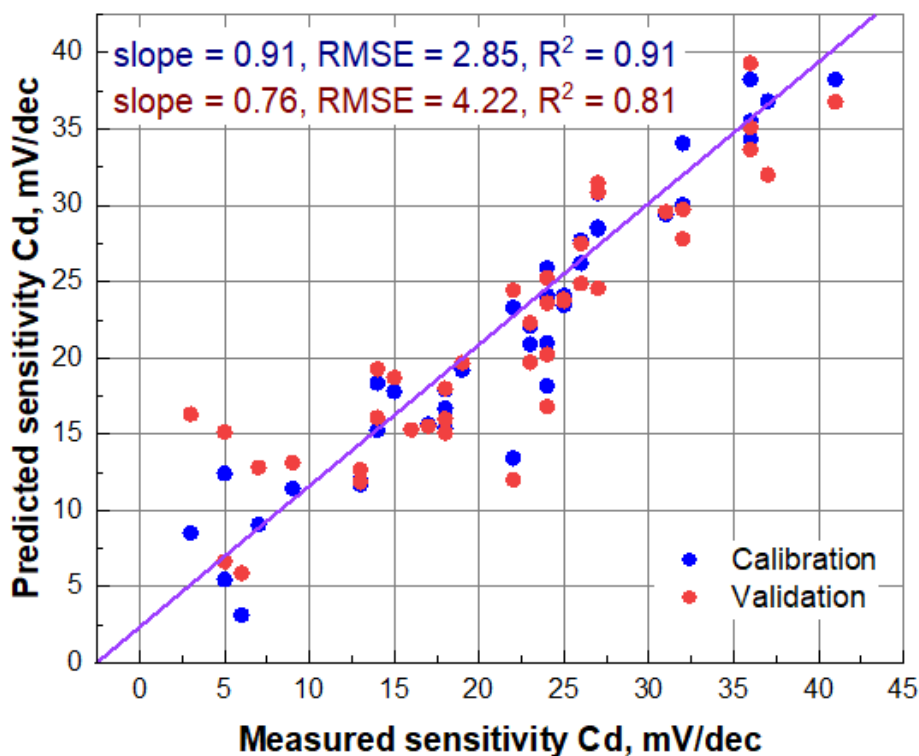


Figure 16. Graph of the entered-found model of cadmium sensitivity (four latent variables) with a marked line of perfect dependence.

The values of regression coefficients in PLS modeling can be used to judge the significance of molecular fragments for a particular target property. A graph of regression coefficients is a convenient way to visualize these values. Weighted regression coefficients are calculated from the weighted data matrix with descriptors during PLS regression; a brief explanation can be found in the Supplementary Materials. Figure 17 shows the plot for the PLS model of cadmium sensitivity: the X-axis shows the descriptor number and the Y-axis shows the numerical value of the specific regression coefficient for each descriptor. The higher the absolute value of the regression coefficient, the more significant was the contribution of the corresponding descriptor. It can be seen that some descriptors did not contribute significantly to the model, and thus could only carry noise. In order to optimize the model, we excluded all variables (descriptors) within the marked area in Figure 17 that had small regression coefficient values between -0.05 and +0.05. Getting rid of irrelevant variables resulted in improved models. A similar procedure was carried out for the PLS models of copper and lead and the resulting statistics are summarized in Table 6. Measured-predicted plots for copper and lead are presented in Figures 18 and 19.

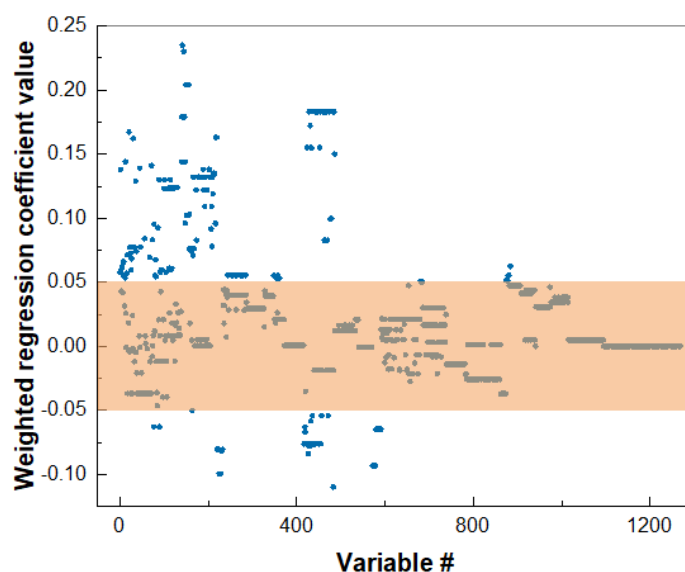


Figure 17. Regression coefficient plot for the PLS model of cadmium sensitivity with a marked region of fragments with small contributions to the model.

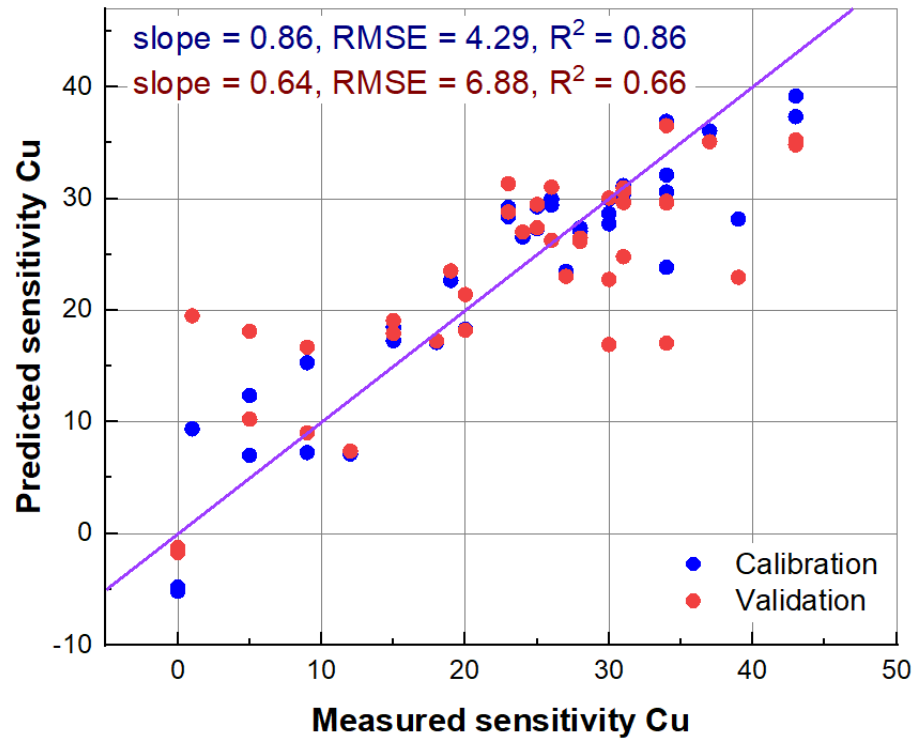


Figure 18. Graph of the entered-found model of copper sensitivity (two latent variables) with the line of perfect dependence marked

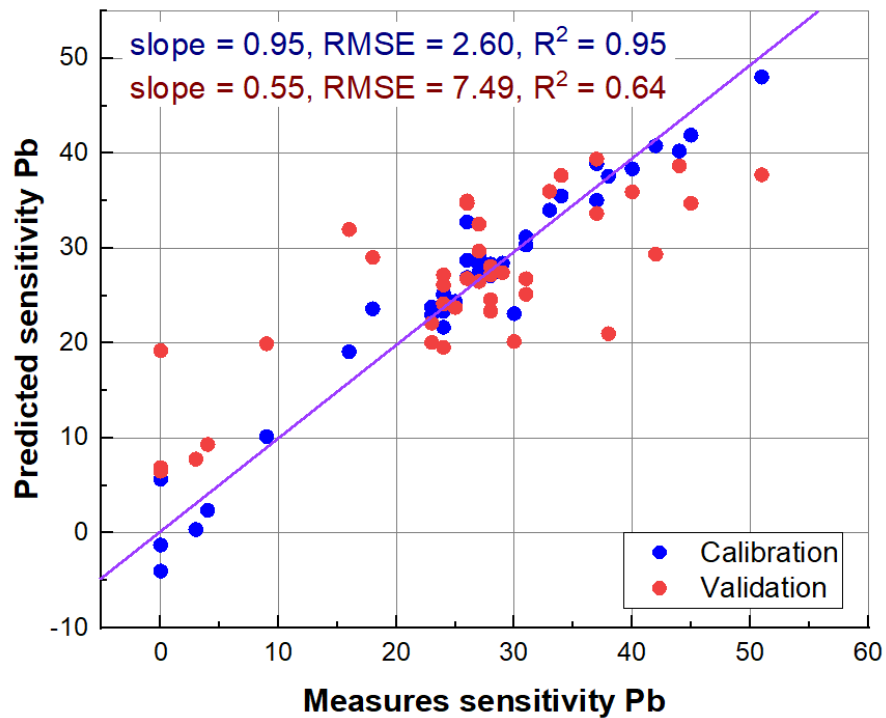


Figure 19. Graph of the entered-found model of lead sensitivity (six latent variables) with the line of perfect correlation marked

Table 6. Model characteristics for cadmium, copper, and lead sensitivities.

		slope	RMSE	R²
Cd	calibration	0.91	2.85	0.91
	validation	0.76	4.22	0.81
Cu	calibration	0.86	4.29	0.86
	validation	0.64	6.88	0.66
Pb	calibration	0.95	2.60	0.95
	validation	0.55	7.49	0.64

It can be seen that the worst performance of the model was observed for lead, where the RMSE at cross-validation was more than 7 mV/dec. Obviously, such a model is more suitable for qualitative evaluation of ionophores in terms of their suitability for sensor development for lead than for numerical prediction. Somewhat better results were obtained in the case of copper, but they were still semi-quantitative in terms of prediction accuracy.

An important indicator of the validity of the QSPR model was the consistency of important molecular fragments with established patterns of chemistry. Figure 20 shows the fragments that were important for characterizing cadmium sensitivity. They were selected on the condition that they occur in at least five structures to avoid random correlations. In plotting the diagrams, we omitted fragments with hydrocarbon chains at the ends if the nested fragments had similar contributions. As an example of such fragments, we can consider the pairs C-C-C-P-C=C-C and C-C-P-C=C. Figures 21 and 22 show these plots for copper and lead, respectively.



Figure 20. Fragments with high contribution to the cadmium sensitivity value



Figure 21. Fragments with high contribution to the copper sensitivity value

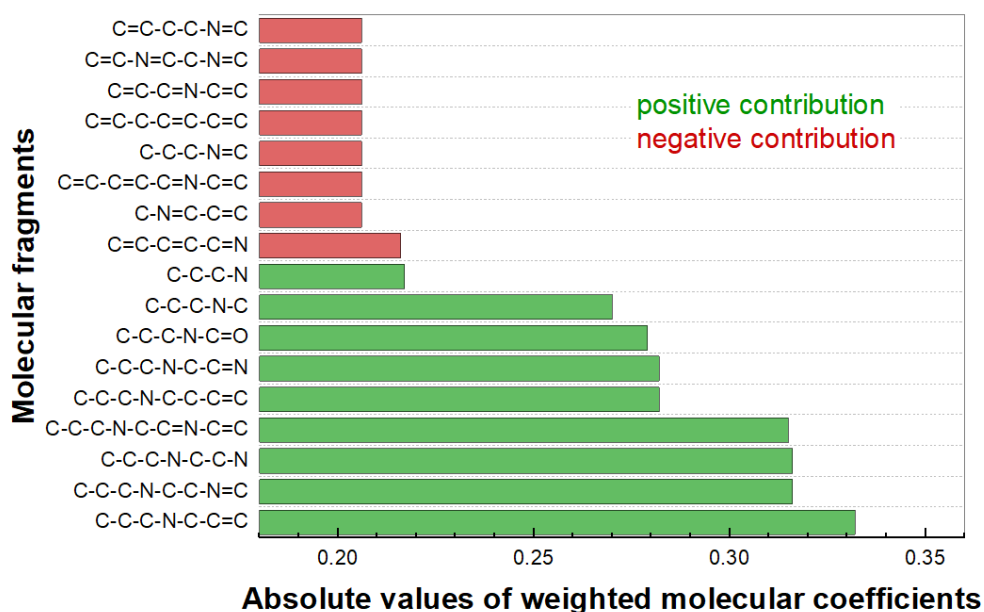


Figure 22. Fragments with high contribution to lead sensitivity value

Analysis of the graph in Figure 20 shows that the fragments with the highest contribution to cadmium sensitivity include a phosphorus atom or an N,N-dimethylpicolinamide fragment (Figure 23a). In contrast to copper sensitivity (Figure 21), the structure fragments (C=C-C=C-C=C-C=N, C-N=C) have mostly negative contributions.

4,40-dibromo-N6,N60-diethyl-N6,N60-bis(4-hexylphenyl)-[2,20-bipyridine]-6,60-dicarboxamide had the highest sensitivity to cadmium in the training data set (41 mV/dec). The chemical structure of this compound is shown in Figure 23b.

The compound contained fragments with both positive (highlighted in green) and negative (highlighted in red) contributions. In general, the identified important structural fragments followed the chemical logic of complexation: ionophores with a picolinamide fragment (Figure 23a) showed a pronounced sensitivity to cadmium when used in plasticized polymer membranes for potentiometric sensors (see also Table 6). The same was observed for the green fragment in Figure 23b, which represented a portion of picolinamide, although the absence of an oxygen atom in the green fragment seems questionable. Another discrepancy with experimental observations is the red

fragment in Figure 23b - in general, the presence of a pyridine structural unit was observed for ionophores with both high and medium sensitivity to Cd^{2+} . We believe that these discrepancies can be explained by the rather small size of the data set used (only 41 ionophores), whereas traditional QSPR studies typically involve hundreds or thousands of entries. It should be noted that in the case of potentiometric experiments, the creation of such a large dataset is hardly possible due to differences in the experimental protocols used by researchers to study potentiometric sensors (membrane compositions, sensitivity calculations, sample pH).

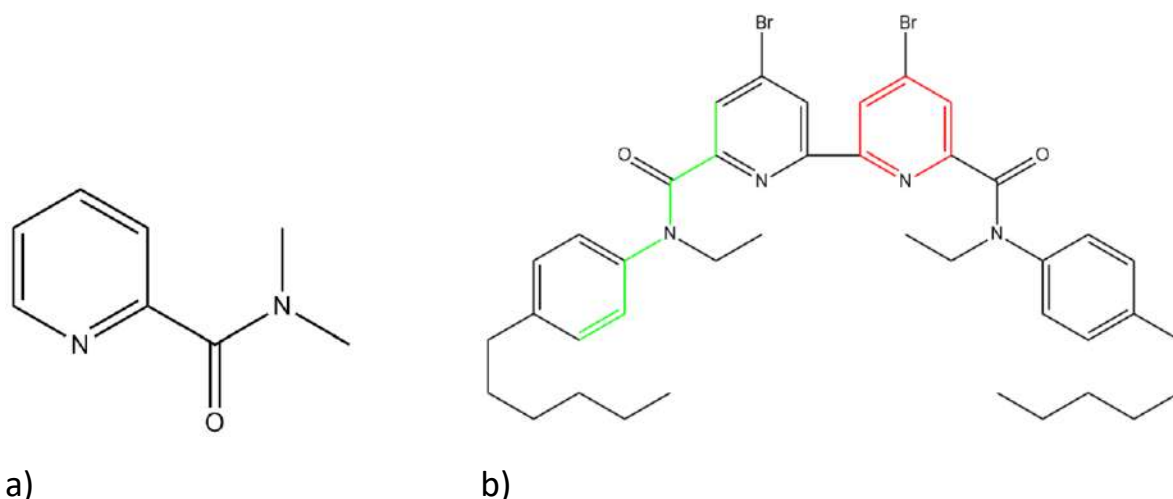


Figure 23. *N,N*-dimethylpicolinamide fragment (a), 4,4'-dibromo-*N*6,*N*6'-diethyl-*N*6,*N*6'-bis(4-hexylphenyl)-[2,2'-bipyridine]-6,6'-dicarboxamide with separated C-C-C-C-N-C-C=C (green) and C=C-C=C-C=C-C=N (red) fragments (b).

The obtained PLS models were applied to predict the potentiometric sensitivity of four new ionophores, whose structures are summarized in Table 5. The chemical structures of these compounds were analyzed using a previously optimized set of molecular fragments. These four sets of molecular descriptors then served as input data for three PLS models to predict the sensitivity to each of the metal cations. The results are summarized in Table 7.

The agreement between model predictions and experimental data was highest for sensitivity to cadmium. The predicted values of sensitivity to copper and lead showed some inconsistency with the experimental data. In particular, for compounds 2 and 3, the predicted values of sensitivity to copper differed from the experimental ones by 2 and 6 mV/dec, respectively. This is explained by the absence in the structure of these compounds of the fragments that contribute most to the sensitivity to copper (C-N-C-C-C-N=C, C-N-C-C-C-N and others). For lead, the largest deviation of the prediction from the experiment was observed for compound 4, which is probably due to the presence of phosphorus in its composition. Despite the presence of compounds with phosphorus in the training set, there are no phosphorus-containing fragments among the fragments with the highest sensitivity to lead.

Table 7. *Experimental and predicted sensitivity. The standard deviation values for the experimental data obtained on three identical sensors in three repeated measurements did not exceed 1.5 mV/dec*

	Cd²⁺		Cu²⁺		Pb²⁺	
	Experiment	Prediction	Experiment	Prediction	Experiment	Prediction
1	24.6	24.7	20.3	20.9	32.9	29.7
2	22.1	22.5	18.3	16.4	31.7	29.2
3	24.7	22.5	21.1	15.3	34.6	29.1
4	23.0	23.1	23.4	19.4	34.9	27.2

This study investigated the sensitivity of sensors based on new ionophores to nickel, cobalt, zinc, and some other metals. However, only data for copper, cadmium and lead were included in the QSPR modeling, since only for these ions the corresponding literature data were available. To construct the initial descriptor matrix, it was necessary to characterize the sensitivity of each substance from the database to

each of the ions considered. Since some of the ligands in our database have not been studied in sensors for nickel, cobalt and zinc, we had to limit the list of metals. In agreement with previous studies, substances with pronounced extractivity towards lanthanides showed a significant potentiometric response towards d-elements in aqueous solutions. The complete table with all obtained sensitivities for all investigated ionophores is given in Table 8, where a certain structural dependence of sensitivity values for cobalt and nickel can be traced.

Table 8. Potentiometric sensitivity of sensors to metal cations (± 1 mV/dec)

Sensor	Co ²⁺	Ni ²⁺	Cu ²⁺	Zn ²⁺	Cd ²⁺	Pb ²⁺
1	18	13	21	18	25	33
2	20	16	18	15	22	32
3	22	18	21	19	25	34
4	18	16	23	20	23	35

An increase in the length of the alkyl chain between nitrogen atoms in the ligand (substances 1, 2 and 3) correlated with an increase in the sensitivity of the sensors to cobalt and nickel. At the same time, this trend was not observed for the other metals. Substance 4 with a piperazine bridge showed a different behavior from substances with alkyl bonds. It can be assumed that the mechanism of complexation between the ligand and the target ion, based on the interaction between the oxygen of the phosphine oxide and the keto group with some influence of nitrogen atoms, is common to all the studied substances. However, further studies are needed to understand the interaction mechanism in more detail.

The comparison between the predictions of the QSPR model and the results of the potentiometric experiment, presented in Table 7, shows a high degree of agreement in most cases. A particularly exact match was observed for cadmium, while for lead the predicted values were systematically lower than the experimental values. Nevertheless,

the relative order of increase in sensitivity from copper to lead agreed well with the model predictions.

Conclusion to Chapter 3

This chapter demonstrates the possibility of successfully applying a QSPR model trained on literature data on the potentiometric sensitivity of plasticized polymer membrane sensors to predict *in silico* the potentiometric behavior of new ionophores without the need for their synthesis and experimental characterization. The validity of the model predictions for sensitivity to Cu^{2+} , Cd^{2+} , and Pb^{2+} was confirmed experimentally for four new diphenylphosphorylacetamide ionophores. The high degree of agreement between the model predictions and the experimental sensitivity values indicates the significant potential of the QSPR approach for the development of new plasticized membrane sensors with specified characteristics.

Chapter 4. APPLICATION OF QSPR MODELING TO PREDICT PHYSICOCHEMICAL PROPERTIES OF DEEP EUTECTIC SOLVENTS BASED ON CHOLINE CHLORIDE

This chapter investigates the possibility of applying the QSPR method to predict the physicochemical properties of deep eutectic solvents based on choline chloride with organic acids as hydrogen bond donors. The aim of this study is to develop efficient QSPR models capable of predicting the properties of DESs based on their molecular structure and composition, which will optimize the choice of components and ratios to obtain DESs with desired properties. This is especially relevant due to the growing interest in DESs as environmentally safe and efficient solvents for various applications.

A dataset of 36 samples of six organic acids with different water contents was used to construct QSPR models. This approach allows to consider the influence of both the nature of the acid and the water content on the properties of DESs. The substructural molecular fragments characterizing the chemical structure of the acids, semi-empirical parameters obtained by PM3 quantum-chemical calculation reflecting the electronic properties of the molecules, and the percentage of water content in each solution were chosen as independent variables. The dependent variables were viscosity, refractive index, electrical conductivity and density determined experimentally. These properties were chosen due to their importance for the practical application of DES in various fields.

The predictive ability of the developed models was further validated on two DESs that were not included in the training set. This allows us to evaluate the effectiveness of the models for predicting the properties of new, previously unstudied DESs.

4.1. Experimental section

4.1.1. Selection of descriptors

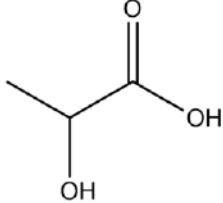
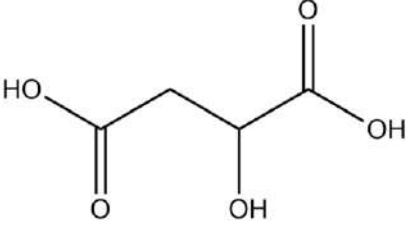
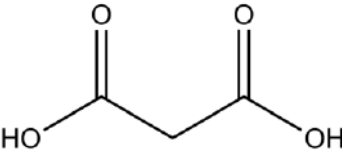
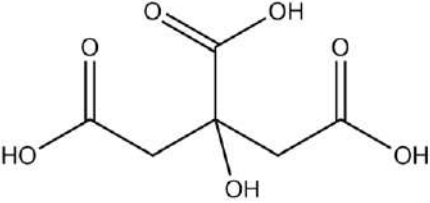
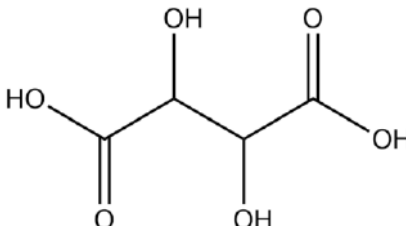
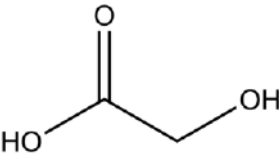
In developing QSPR models for carboxylic acid-based DES, the aim was to establish the relationship between the descriptors and the dependent variables represented by

viscosity, density, conductivity and refractive index. A dataset containing 36 samples of six organic acids with different water contents was collected for modeling using the QSPR approach. Of these, two samples were excluded from the calculations because they are crystals in concentrated form (0% water content). Unlike previous studies described in Chapters 2 and 3, the set of independent variables included not only substructure molecular fragments, but also semi-empirical parameters derived from the PM3 quantum chemical calculation, as well as the percentage of water content in each solution. The dependent variables, viscosity, refractive index, conductivity, and density, were obtained experimentally and provided by colleagues from the Department of Analytical Chemistry, Tatiana Bochko and Andrey Shishov. Table 2 in the Appendix A summarizes the measured DES characteristics, and Table 9 lists the trivial and IUPAC names and structures of the investigated organic acids.

The structural descriptors were calculated using the ISIDA software package similar to the experiments described in Chapters 2 and 3. Since the chemical compounds in the training set of this experiment are much smaller than in the previously described experiments, the size of the substructural molecular fragments was chosen from two to four atoms.

Quantum chemical descriptors were determined with HyperChem software from Hypercube, Inc. (<http://www.hypercubeusa.com/>) using the parametric PM3 method. PM3 enables quantum calculations of the electronic structure of molecules in computational chemistry. Several descriptors such as the energies of the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO), dipole moment and heat capacity (calculated for 300K) were determined using HyperChem. These calculations added additional data to the training set and allowed us to analyze in more detail the nature of the dependencies obtained.

Table 9. Organic acids comprising the training set, their structures and names according to IUPAC nomenclature

№	Trivial name	Structure	IUPAC name
1	lactic		2-hydroxypropanoic acid
2	malic		2-hydroxybutanedioic acid
3	malonic		propanedioic acid
4	citric		2-hydroxypropane-1,2,3-tricarboxylic acid
5	tartaric		2,3-dihydroxybutanedioic acid
6	glycolic		2-hydroxyethanoic acid

The integrated use of structural descriptors together with semi-empirical parameters provides a robust representation of the independent variables within the QSPR modeling framework, enhancing the predictive power of the models.

To establish a mathematical relationship between a set of descriptors for the studied DES and their viscosity, refractive index, density and conductivity, we used PLS regression combined with cross-validation to determine the optimal number of latent variables that provide the greatest predictive power. Descriptors were also selected based on their contribution to the models and to reduce noise, those descriptors that had a negligible effect on the model were excluded from the dataset.

Due to the limited size of the training dataset, it was necessary to perform a thorough check on the quality of the model and its ability to generalize to the data. Nested cross validation, also known as double cross validation, was applied to test the statistical significance of the models. This type of cross-validation is computationally costlier than K-fold or even full cross-validation, but avoids overfitting the model. Nested cross-validation involves two stages, shown in Figure 24 for clarity: an inner stage for model selection or hyperparameter tuning, primarily the number of latent variables, and an outer stage, the familiar full cross-validation, for evaluating model performance. The nested cross-validation has two iterations of splitting the data into several training and validation sets, thus providing a more reliable estimate of efficiency and reducing the risk of model overfitting.

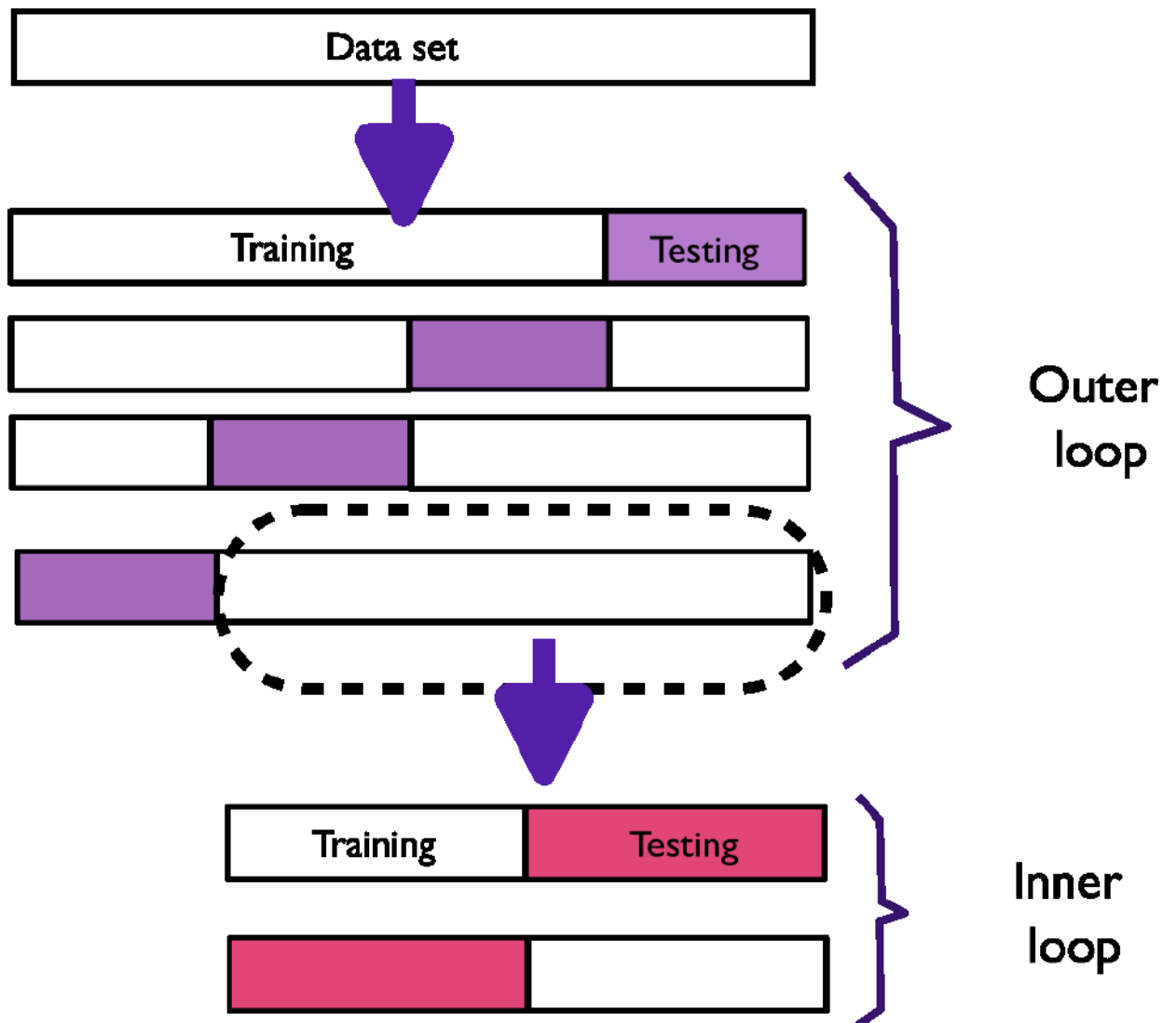


Figure 24. Scheme of nested cross-validation.

As an additional stage of model validation, permutation testing was performed. In permutation tests, a number of new sets of values of the dependent variable are created from the values of the dependent variable. For this purpose, the values of the original sample are swapped, mixed, and for each resulting set the model performance indicators are calculated; this process is illustrated in Figure 25. Thus, it is assumed that all models except the one where the data are not shuffled have no predictive power. So, by comparing the performance of the real model with this distribution, we can assess the significance of the model's predictive power and ensure that it is not the result of random correlation.

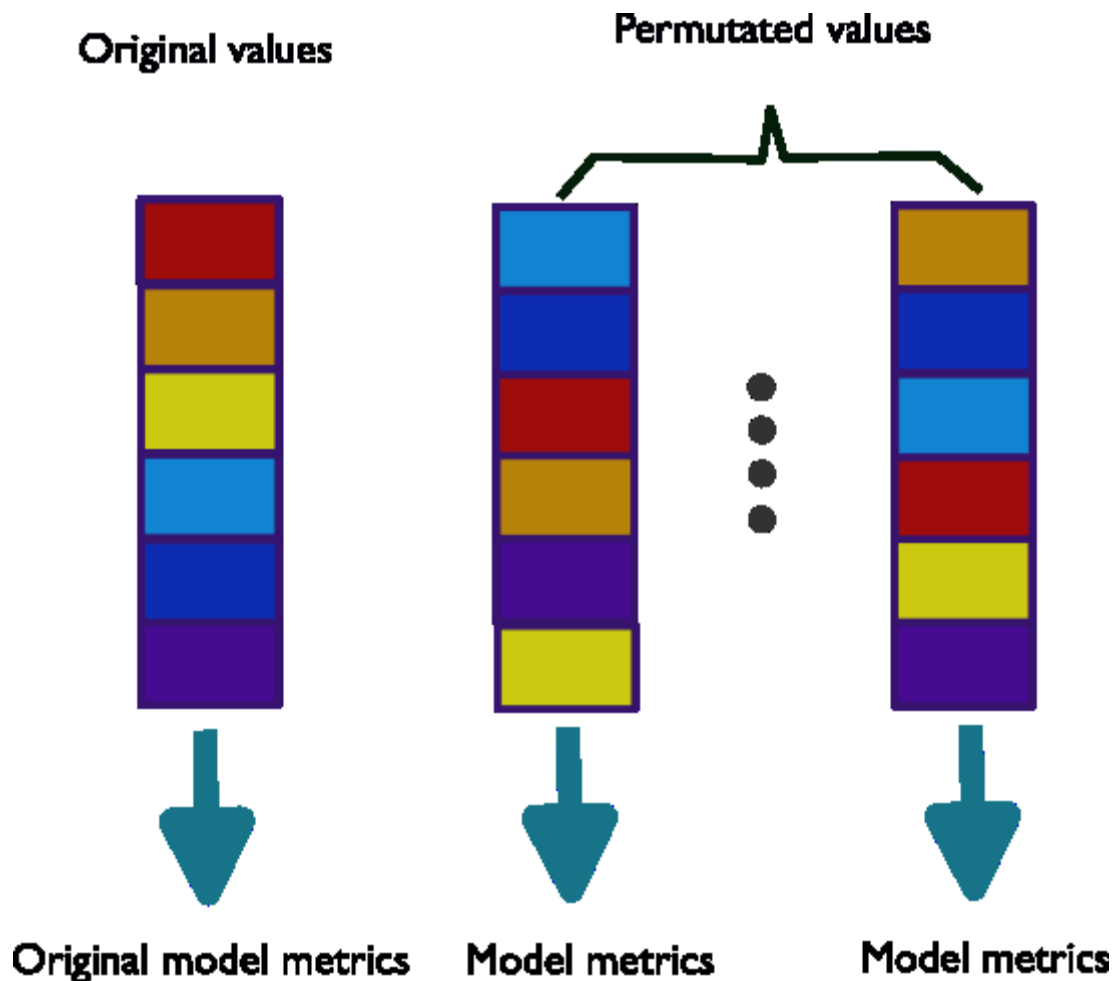


Figure 25. Schematic of permutation testing.

Regression modeling was performed in the R programming environment (R version 4.1.3) using the caret package for PLS and the nestedcv package for nested cross-validation.

4.2. Discussion of results

4.2.1. Density

As a first step in analyzing the data, we plotted the relationship between density and refractive index and water content as shown in Figures 26-27. These plots revealed a well-defined linear relationship between these variables and percent water content. Which is not surprising considering that the water content varies from 0 to 50 wt% in

the samples. PLS modeling was applied to these data in order to gain a deeper understanding of which parameters the values of the variables depend on and how these dependencies can be explained.

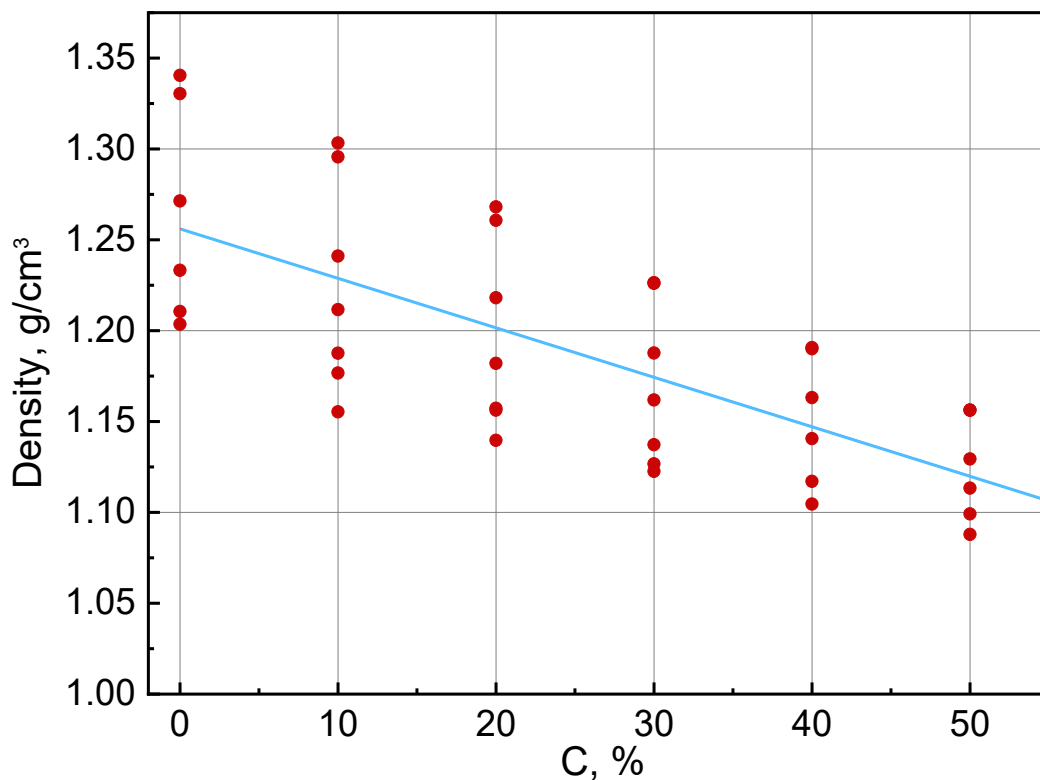


Figure 26. Scatter plot showing the relationship between density and water content. The blue line shows the relationship obtained by simple linear regression

To minimize the influence of water on the DES characteristics, a data set limited to samples with a fixed water content (30 wt%) was used in the simulations. This value of water content was chosen because of its central position in the range studied. Similar studies carried out at other water contents gave similar results, which confirms the validity of the chosen approach.

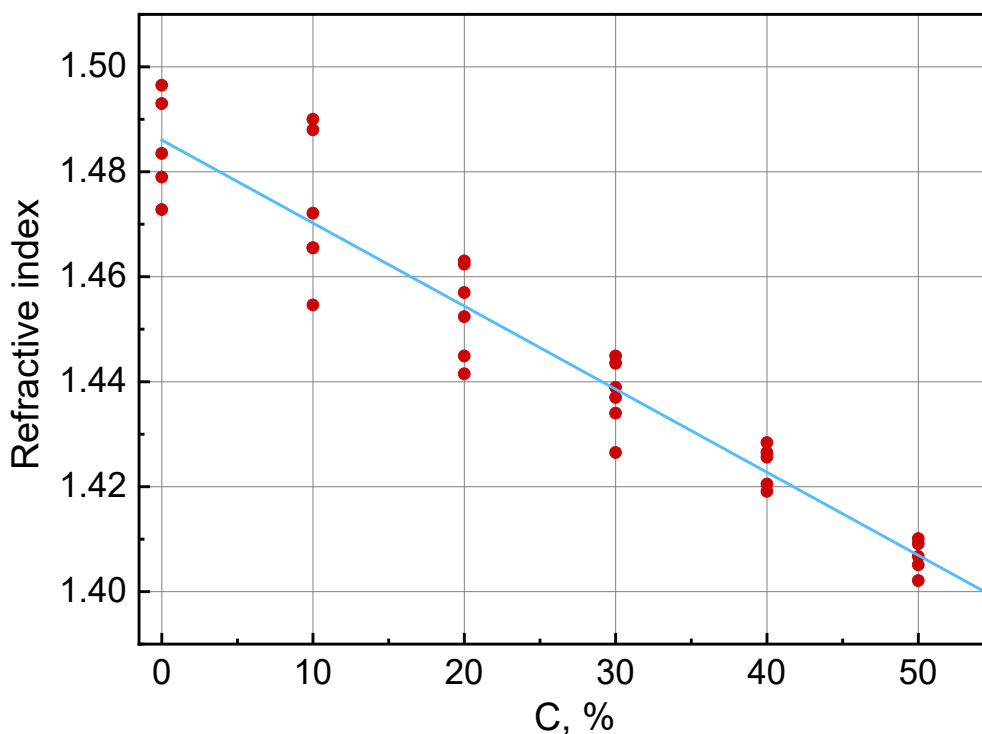


Figure 27. Scatter plot showing the relationship between refractive index and water content. The blue line shows the dependence obtained by simple linear regression

A PLS model based on the molecular descriptors of the organic acids comprising the DES was constructed to predict the density. Figure 28 shows the entered-found plot for this model, indicating that there is a statistically significant correlation between descriptors and density.

This correlation can be further explored using the PLS regression coefficients. The absolute value of each coefficient is proportional to the contribution of the corresponding variable to the final model, with the sign of the coefficient indicating its negative or positive contribution. The corresponding plot of regression coefficients for the PLS model describing DEC with the same water content is shown in Figure 29. Analysis of this plot suggests that the main influence on density is hydrogen bonding, since molecular fragments containing COH and C=O moieties show a positive correlation due to their propensity to hydrogen bond. An increase in the number of hydrogen bonds leads to a denser arrangement of molecules, which leads to an increase in the overall

density [83]. These interactions are also reflected in the hydration energy: a greater number of bonds corresponds to a lower hydration energy and hence a higher density. In addition, the strengthening of hydrogen bonds generally reduces the lipophilicity of acids. This is due to the polar properties of hydrogen bonds, which increase solubility in water, resulting in acids becoming less soluble in nonpolar solvents, as reflected in the negative correlation with $\log p$ (log of the water-octanol distribution coefficient).

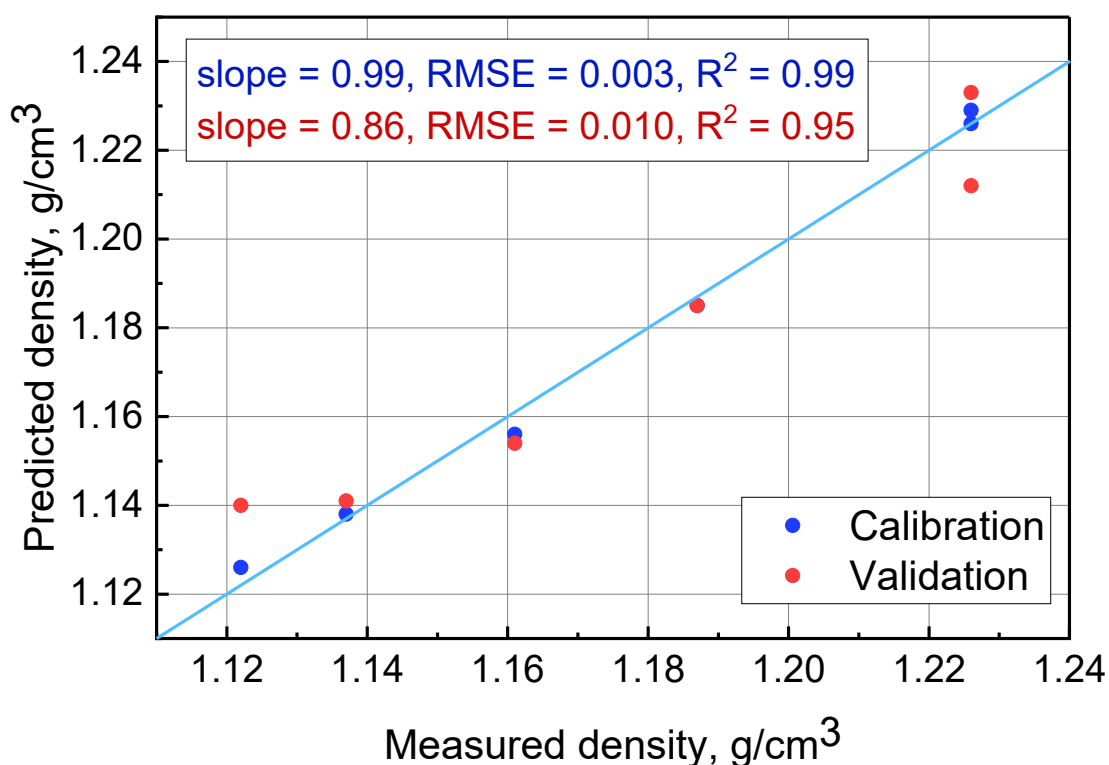


Figure 28. Measured-predicted plot for PLS density model based on samples with water content = 30% (2 latent variables, full cross-validation)

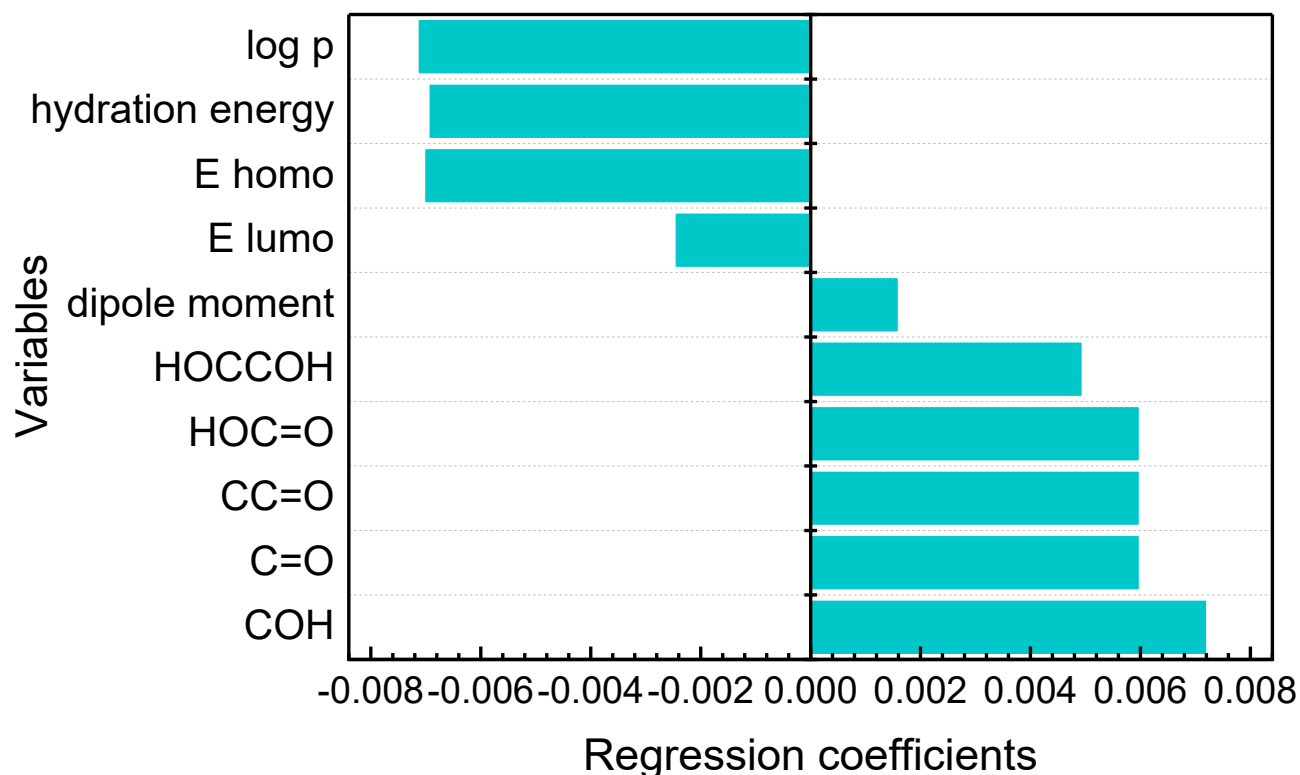


Figure 29. Regression coefficients of variables with the largest contributions to the PLS density model

4.2.2. Conductivity

Using the same methodology, PLS modeling for predicting the conductivity of DES was performed - a subset of samples with a constant water content of 30 wt.% was investigated, and only variables with high absolute values of regression coefficients were included in the model. A plot of the relationship between measured and predicted values for the corresponding PLS model is shown in Figure 30. The model demonstrated a sufficiently high predictive ability.

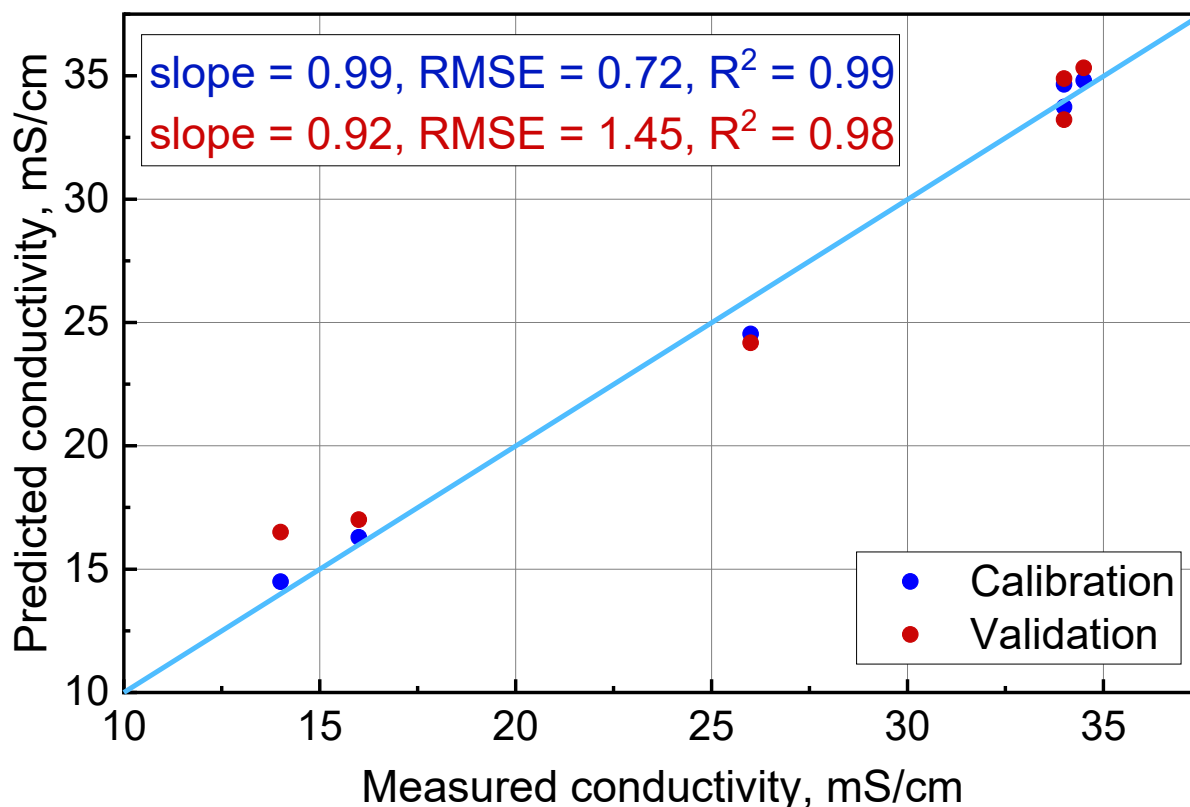


Figure 30. Measured-predicted plot for the conductivity model based on 6 samples with water content = 30% (2 latent variables)

Analysis of regression coefficients (Fig. 31) shows that conductivity is significantly affected by several factors: positively by log P (lipophilicity), total and free energy, and negatively by polar COH groups and entropy. Increasing log P increases conductivity, probably due to the effect on ion mobility [85]. In addition, the presence of long carbon chains can inhibit the mobility of charge carriers, thereby reducing the conductivity.

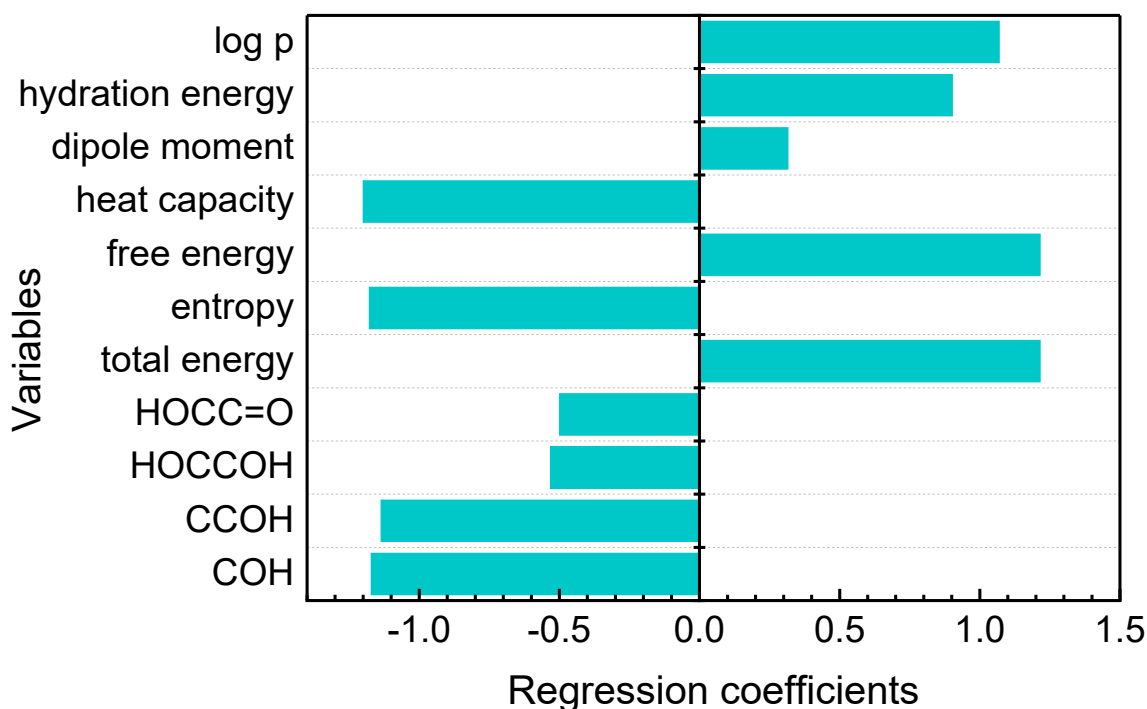


Figure 31. Regression coefficients for PLS model of conductivity based on samples with water content = 30%.

4.2.3. Viscosity

The first experiments on predicting DES viscosity using PLS modeling showed that the models have a clearly nonlinear structure due to the very wide range of variation in experimentally determined viscosity values. To circumvent this problem, we performed regression modeling with respect to the logarithm of the viscosity values. Fig. 32 shows the corresponding plot of the dependence of measured values on predicted values after excluding descriptors with insignificant regression coefficients.

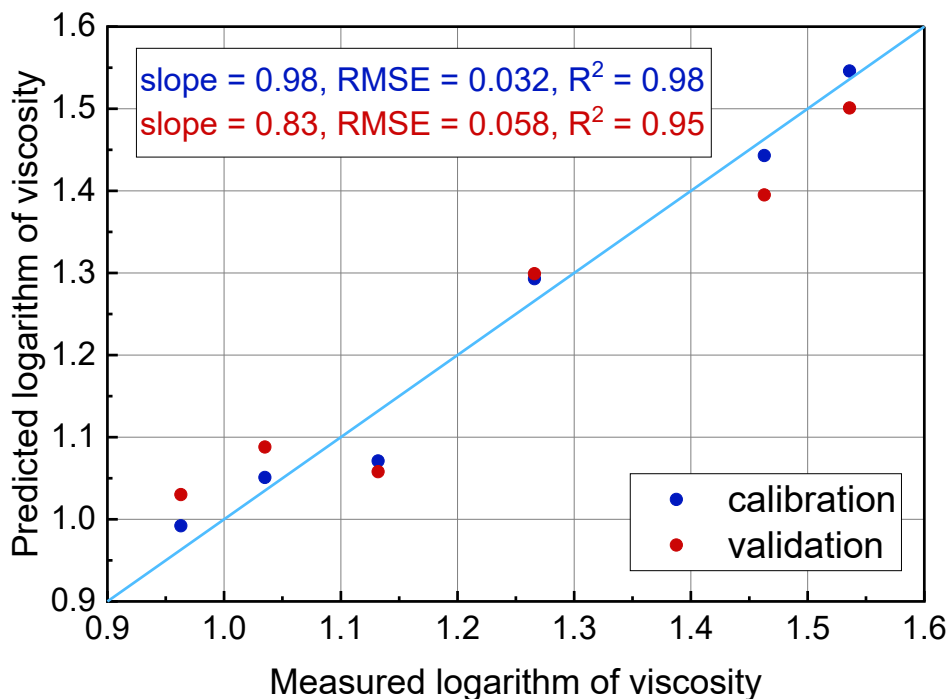


Figure 32. Measured-predicted plot for the viscosity logarithm model based on 6 samples with water content = 30% (1 latent variable)

The regression coefficients for the remaining variables are presented in Figure 33.

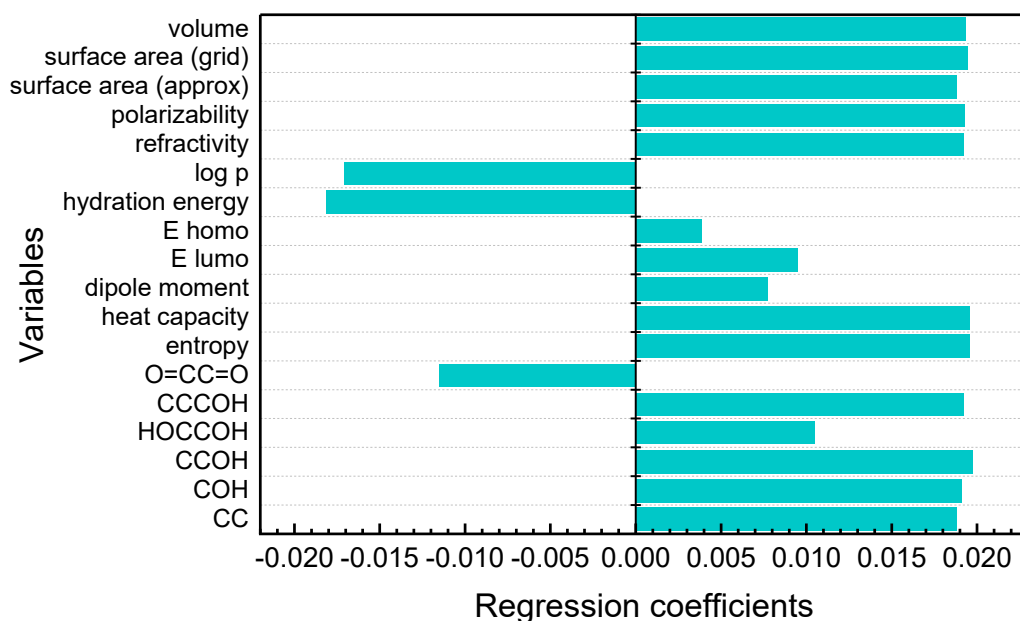


Figure 33. Regression coefficients for the PLS viscosity model based on samples with water content = 30% for the optimized set of variables

More than half of the presented variables have a significant effect on the logarithm of viscosity. In particular, log P and hydration energy have a negative effect on the logarithm of viscosity, that is, the higher the log P or hydration energy of an organic acid, the lower its viscosity. On the contrary, ten variables including CC, COH, CCOH, CCCOH, enthalpy, heat capacity, polarizability, refractivity, surface area and volume of the acid molecules have a positive influence. The observed correlation between lipophilicity and viscosity can probably be explained by the interaction of DES components with water; higher lipophilicity suggests less interaction with water, resulting in lower viscosity. The effects of surface area and volume suggest that larger molecules contribute to higher viscosity, which is consistent with general physical considerations and other experiments [86]. The authors of [86] showed that the viscosity in a series of 1-alkanol solutions correlates with the number of C atoms in the molecule (which is similar to the CC descriptor in our study, correlating with the total number of C atoms). The presence of hydroxyl (-OH) groups in the molecular fragments indicates the possibility of hydrogen bond formation, which possibly increases the resistance to flow. In addition, polarizability, heat capacity, and refractivity may have a significant effect on viscosity as a result of enhanced intermolecular interactions or forces that contribute to the overall viscosity.

4.2.4. Refractive index

The next property studied was the refractive index of the DES. Following the similar reasoning in Section 4.2.1 with density analysis, simulations were performed for a set of samples with the same water content (30 wt.%). A measured-predicted plot for the corresponding PLS model is shown in Fig. 34.

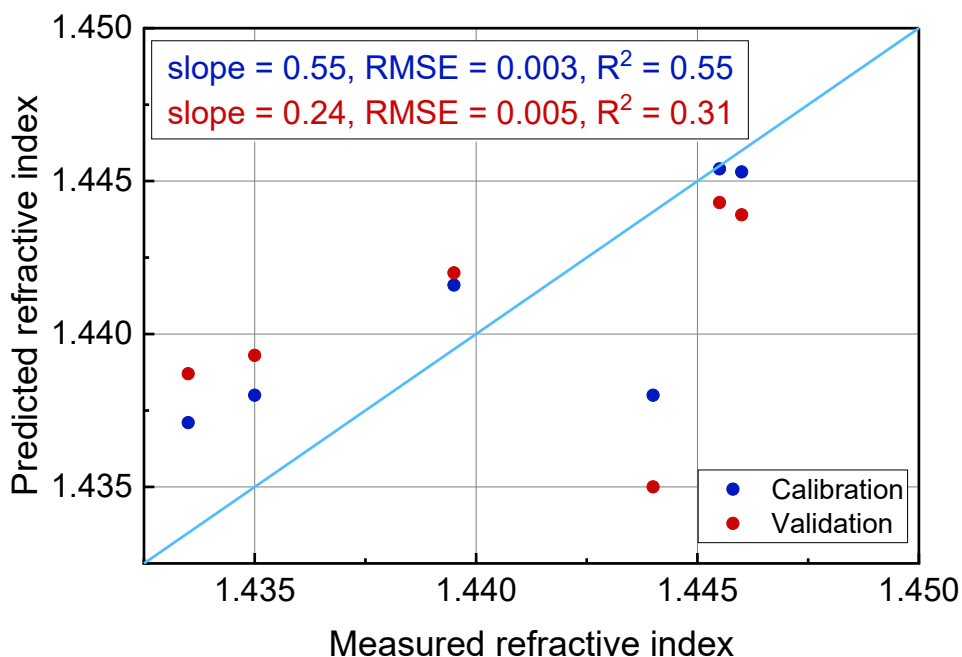


Figure 34. Measured-predicted plot for the refractive index model for samples with water content = 30% (1 latent variable)

It can be seen that the correlation between the studied set of descriptors and refractive index was rather weak (R^2 at calibration 0.55, R^2 at validation 0.31), with the glycolic acid sample being a clear outlier. These results mean that the descriptors chosen are insufficient to predict the optical properties of DES, the set obtained is not sufficient to explain the structure-refractive index dependence. In this case, further extended study will be required to obtain the dependence.

One of the promising approaches to solve this problem seems to be the inclusion of descriptors reflecting the electrodynamic behavior of molecules into the model. Since the nature of the refractive index is related to the interaction of electromagnetic radiation with matter, it is affected by the ability of the molecular structure to polarize in response to the electromagnetic field. This polarization, in turn, depends on the distribution and dynamics of electronic charges and molecular momentum. Examples of such descriptors are topological charge indices [87].

4.2.5. Statistical validation of models

The ultimate goal of QSPR modeling is to evaluate the applicability of these models for predicting the properties of new DESs. Given the limited number of samples, the models were thoroughly validated using nested cross-validation and permutation testing as described in Section 4.1.2. This approach allows for a preliminary assessment of the predictive ability of the regression models under limited sample conditions. In the next step, the predictive ability of the model was evaluated for two acid-based *DESs* that were not used in training (Section 3.6).

Plots of the ratio of measured to predicted values for the nested cross-validation PLS models are shown in Figure 35. The RMSE and R^2 values indicate that all properties except refractive index can be predicted with reasonably high accuracy. For example, the prediction of density in the range of 1.12 - 1.23 g/cm³ is achieved with an RMSE of 0.012. In the case of refractive index, the model did not perform satisfactorily with nested cross-validation, indicating the need for additional descriptors or alternative modeling approaches.

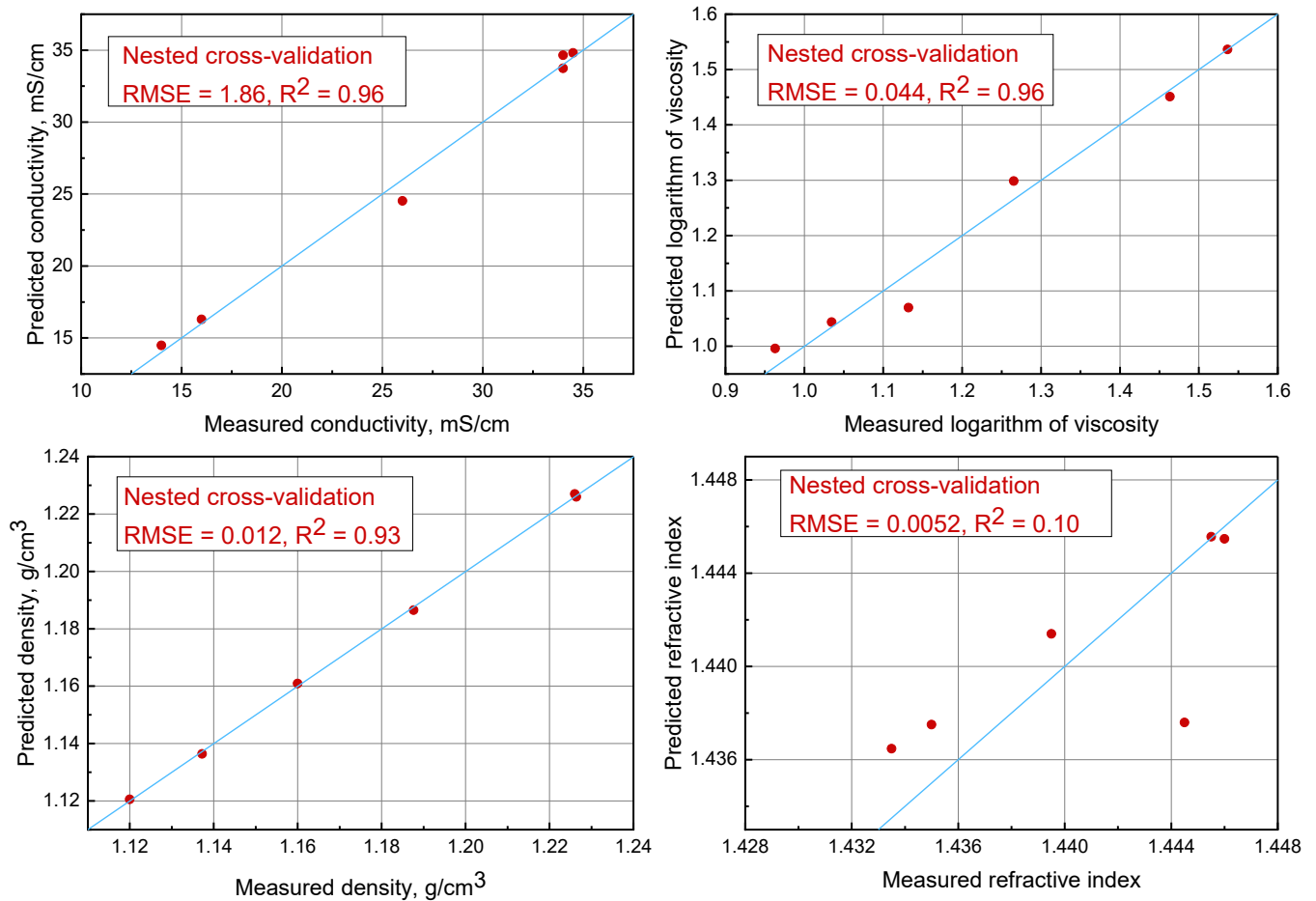


Figure 35. Measured-predicted plots for PLS models with nested cross-validation.

The blue line shows the hypothetical perfect dependence

Similar results were obtained in permutation testing (Fig. 36). The graphs show that the set of descriptors used does not allow building a reliable model for refractive index prediction, while density, conductivity and viscosity correlate well with the optimized descriptors.

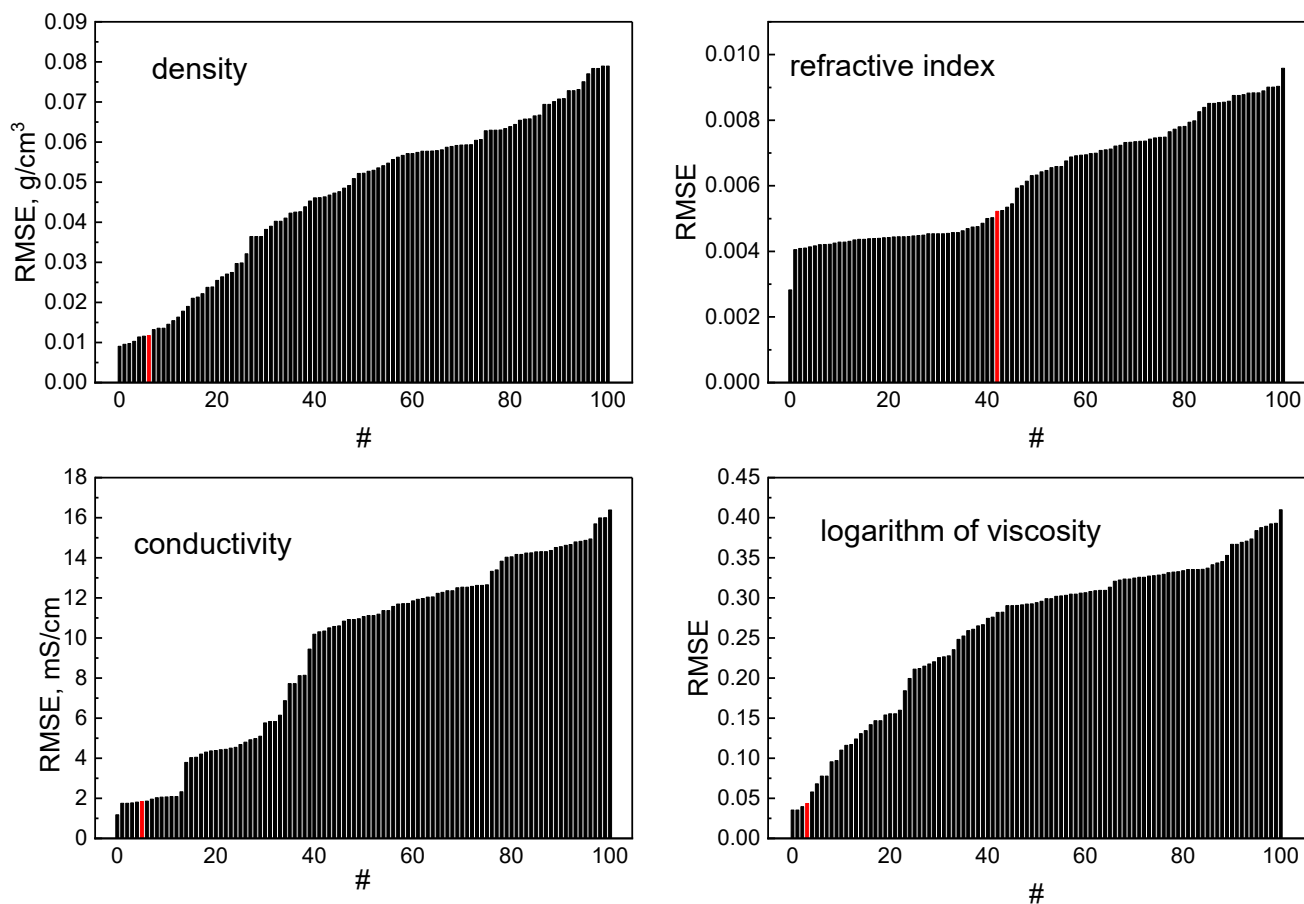
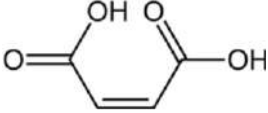
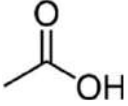


Figure 36. Results of the permutation test for modeling DES properties. The RMSE of the original model without permutations is indicated by the red line

4.2.6. Evaluation of predictive performance of models

To evaluate the predictive ability of the developed QSPR models, descriptors for maleic and acetic acids, which were not included in the training set, were calculated. Based on these descriptors, the density, viscosity, conductivity and refractive index of two corresponding DESs containing 30 wt.% water were predicted. The predictions obtained were compared with the experimental data obtained for the synthesized DESs. The comparison results are presented in Table 10.

Table 10. Properties of DES based on maleic and acetic acids with 30 wt.% water obtained in experiment and predicted by QSPR models. Experimental values are given in the columns "Exp.", predicted values are given in the columns "Pred."

Chemical structure of an organic acid	Refractive index		Density, g/cm ³		Viscosity, mPa·s		Conductivity, mS/cm	
	Pred.	Exp.	Pred.	Exp.	Pred.	Exp.	Pred.	Exp.
	1.438	1.444	1.210	1.145	17.4	13.8	29.0	28.5
	1.435	1.429	1.102	1.085	7.1	7.4	42.5	37.9

As can be seen from the results obtained, there is satisfactory agreement between the predicted and measured values of these new DESs. The largest discrepancy is observed for the viscosity in the case of maleic acid - this is probably due to the presence of a C=C double bond in this molecule, a feature that was not present in the training set molecules.

Some deviation of the predicted value from the experimental one is also observed for the conductivity of HER with acetic acid in its composition. This is probably due to the peculiarities of the acetic acid structure (e.g., the smallest molecule size among the training and test data sets), which are not fully accounted for in the model.

Despite some inconsistencies, the results of the study confirm the promising use of the QSRP method for modeling various physical properties of eutectic solvents.

Conclusion to Chapter 4

It has been shown that mathematical modeling using the QSPR approach can be used to predict the properties of DESs based on molecular descriptors characterized by structural fragments of organic acids. The structural fragments of the molecules and the

results of semi-empirical quantum calculations are sufficient to quantitatively predict the density, conductivity and viscosity values of DESs. The demonstrated feasibility of QSPR in this research area indicates the relevance of further dedicated studies to extend the range of modeled properties and described deep eutectic solvents. Such modeling can significantly facilitate the development of DES systems for specific practical applications

CONCLUSIONS

1. For a sample of 40 ligands described in the literature for anion-selective electrodes based on substructural molecular fragments and plasticizer conductivity, QSPR models were created that allow predicting the potentiometric selectivity of polymer plasticized membrane sensors based on such ligands to carbonate anion. The model allows to obtain selectivity prediction for new compounds with an average error of no more than $1.5 \log K^{sel}(HCO_3^-/Cl^-)$.

2. On the basis of substructural molecular fragments for 41 compounds containing amide and phosphoryl functional groups, QSPR models have been developed to predict the potentiometric sensitivity of polymer plasticized membrane sensors based on such ligands to heavy metal cations (Cu^{2+} , Cd^{2+} , and Pb^{2+}). The model allows us to obtain sensitivity predictions for new compounds with an error not exceeding 8 mV/dec.

3. QSPR models have been developed based on a sample of deep eutectic solvents based on choline chloride with 6 different organic acids as hydrogen bond donors, for which molecular descriptors in the form of substructure molecular fragments and quantum chemical descriptors have been obtained, which allow predicting density with an error not exceeding 0.065 g/cm, conductivity with an error not exceeding 4.6 mS/cm and viscosity with an error not exceeding 3.6 mPa·s.

4. Approaches such as nested cross-validation and permutation testing allow the validation and reliability of QSPR models based on limited samples of training data.

5. Regression coefficient analysis for the PLS regression in QSPR models reveals the key structural descriptors that determine the desired properties of the new materials.

ABBREVIATIONS LIST

CPSA - Charged partial surface area

DES - deep eutectic solvent

DFT - density functional theory

ECFP - Extended-Connectivity Fingerprints

HPLC - high performance liquid chromatography

ISE - ion selective electrode

J3D - 3D Balaban index

LV - latent variable

LOO CV - leave-one-out cross-validation

MACCS - Molecular ACCess System

MLR - Multiple Linear Regression

PLS - Partial Least Squares

SMF - substructure molecular fragment

SMILES - Simplified Molecular-Input Line-Entry System

SVM - support vector machine

SVR - support vector regression

QSAR - Quantitative Structure-Activity Relationships

QSPR - Quantitative Structure-Property Relationships

W3D - 3D Wiener Index

LIST OF REFERENCES

1. Vladimirova N. et al. Prediction of Carbonate Selectivity of PVC-Plasticized Sensor Membranes with Newly Synthesized Ionophores through QSPR Modeling // Chemosensors. 2022. Vol. 10, № 2. P. 43.
2. Vladimirova N. et al. Predicting the Potentiometric Sensitivity of Membrane Sensors Based on Modified Diphenylphosphoryl Acetamide Ionophores with QSPR Modeling // Membranes (Basel). 2022. Vol. 12, № 10. P. 953.
3. Vladimirova N. et al. Predicting the properties of deep eutectic solvents based on choline chloride and carboxylic acids and their mixtures with water using QSPR approach // Colloids Surf A Physicochem Eng Asp. 2024. Vol. 692. P. 133961.
4. Brown A.C., Fraser T.R. On the Connection between Chemical Constitution and Physiological Action; with special reference to the Physiological Action of the Salts of the Ammonium Bases derived from Strychnia, Brucia, Thebaia, Codeia, Morphia, and Nicotia. // J Anat Physiol. 1868. Vol. 2, № 2. P. 224–242.
5. Hansch Corwin., Anderson S.M. The Structure-Activity Relationship in Barbiturates and Its Similarity to That in Other Narcotics // J Med Chem. 1967. Vol. 10, № 5. P. 745–753.
6. Hansch Corwin., Fujita Toshio. ρ - σ - π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure // J Am Chem Soc. 1964. Vol. 86, № 8. P. 1616–1626.
7. HANSCH C. et al. Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients // Nature. 1962. Vol. 194, № 4824. P. 178–180.
8. Silva de Jesus Passaes A.C. et al. Quinoxalines against *Leishmania amazonensis*: SAR study, proposition of a new derivative, QSAR prediction, synthesis, and biological evaluation // Sci Rep. 2023. Vol. 13, № 1. P. 18136.

9. Esfahani S.N. et al. Synthesis of some novel coumarin isoxazol sulfonamide hybrid compounds, 3D-QSAR studies, and antibacterial evaluation // *Sci Rep.* 2021. Vol. 11, № 1. P. 20088.
10. Qin D. et al. 5D-QSAR studies of 1H-pyrazole derivatives as EGFR inhibitors // *J Mol Model.* 2022. Vol. 28, № 12. P. 379.
11. Acharya K. et al. A quantitative structure-biodegradation relationship (QSBR) approach to predict biodegradation rates of aromatic chemicals // *Water Res.* 2019. Vol. 157. P. 181–190.
12. Najafi M. et al. QSPR model for estimation of photodegradation average rate of the porphyrin-TiO₂ complexes and prediction of their biodegradation activity and toxicity: Engineering of two annihilators for water/waste contaminants // *J Mol Struct.* 2022. Vol. 1249. P. 131463.
13. Ai H. et al. QSAR modelling study of the bioconcentration factor and toxicity of organic compounds to aquatic organisms using machine learning and ensemble methods // *Ecotoxicol Environ Saf.* 2019. Vol. 179. P. 71–78.
14. Wang C. et al. Assessment of bromide-based ionic liquid toxicity toward aquatic organisms and QSAR analysis // *Ecotoxicol Environ Saf.* 2015. Vol. 115. P. 112–118.
15. Cai J. et al. Mechanistic studies of congener-specific adsorption and bioaccumulation of polycyclic aromatic hydrocarbons and phthalates in soil by novel QSARs // *Environ Res.* 2019. Vol. 179. P. 108838.
16. Arnot J.A., Gobas F.A.P.C. A Generic QSAR for Assessing the Bioaccumulation Potential of Organic Chemicals in Aquatic Food Webs // *QSAR Comb Sci.* 2003. Vol. 22, № 3. P. 337–345.
17. Puzyn T. et al. QSPR-based estimation of the atmospheric persistence for chloronaphthalene congeners // *Atmos Environ.* 2008. Vol. 42, № 27. P. 6627–6636.

18. Villaverde J.J. et al. QSAR/QSPR models based on quantum chemistry for risk assessment of pesticides according to current European legislation // SAR QSAR Environ Res. 2020. Vol. 31, № 1. P. 49–72.
19. Yang L. et al. QSAR modeling the toxicity of pesticides against *Americamysis bahia* // Chemosphere. 2020. Vol. 258. P. 127217.
20. Rodrigues N.E. et al. QSAR-Guided Proposition of N-(4-methanesulfonyl)Benzoyl-N'-(Pyrimidin-2-yl)Thioureas as Effective and Safer Herbicides // Bull Environ Contam Toxicol. 2022. Vol. 108, № 6. P. 1019–1025.
21. Tukur S., Shallangwa G.A., Ibrahim A. Theoretical QSAR modelling and molecular docking studies of some 4-hydroxyphenylpyruvate dioxygenase (HPPD) enzyme inhibitors potentially used as herbicides // Heliyon. 2019. Vol. 5, № 11. P. e02859.
22. Isyaku Y., Uzairu A., Uba S. QSAR STUDY OF 2-SUBSTITUTED PHENYL-2-OXO-, 2-HYDROXYL- AND 2-ACYLLOXYETHYLSULFONAMIDES AS FUNGICIDES // The Journal of Engineering and Exact Sciences. 2019. Vol. 5, № 3. P. 0283–0290.
23. Kaushik P., Shakil N.A., Rana V.S. Synthesis, Biological Evaluation, and QSAR Studies of 3-Iodochromone Derivatives as Potential Fungicides // Front Chem. 2021. Vol. 9.
24. Knizhnik A.A. et al. New Polymers In Silico Generation and Properties Prediction // Nanomanufacturing. 2023. Vol. 4, № 1. P. 1–26.
25. Cravero F. et al. Computational modelling of mechanical properties for new polymeric materials with high molecular weight // Chemometrics and Intelligent Laboratory Systems. 2019. Vol. 193. P. 103851.
26. Buglak A.A., Zherdev A. V., Dzantiev B.B. Nano-(Q)SAR for Cytotoxicity Prediction of Engineered Nanomaterials // Molecules. 2019. Vol. 24, № 24. P. 4537.
27. Toropov A.A., Kjeldsen F., Toropova A.P. Use of quasi-SMILES to build models based on quantitative results from experiments with nanomaterials // Chemosphere. 2022. Vol. 303. P. 135086.

28. Kar S. et al. Power Conversion Efficiency of Arylamine Organic Dyes for Dye-Sensitized Solar Cells (DSSCs) Explicit to Cobalt Electrolyte: Understanding the Structural Attributes Using a Direct QSPR Approach // *Computation*. 2016. Vol. 5, № 4. P. 2.
29. Ratanasak M., Hasegawa J., Parasuk V. Design and prediction of high potent *ansa*-zirconocene catalyst for olefin polymerizations: combined DFT calculations and QSPR approach // *New Journal of Chemistry*. 2021. Vol. 45, № 18. P. 8248–8257.
30. Nkulikiyinka P. et al. Prediction of Combined Sorbent and Catalyst Materials for SE-SMR, Using QSPR and Multitask Learning // *Ind Eng Chem Res*. 2022. Vol. 61, № 26. P. 9218–9233.
31. Ojha P.K., Roy K. Development of a robust and validated 2D-QSPR model for sweetness potency of diverse functional organic molecules // *Food and Chemical Toxicology*. 2018. Vol. 112. P. 551–562.
32. Chen K. et al. Multiple quantitative structure–pungency correlations of capsaicinoids // *Food Chem*. 2019. Vol. 283. P. 611–620.
33. Belhassan A. et al. QSPR Study of the Retention/Release Property of Odorant Molecules in Water, Dairy and Pectin gels // *Mater Today Proc*. 2019. Vol. 13. P. 621–629.
34. Buckley H.L. et al. Design and Testing of Safer, More Effective Preservatives for Consumer Products // *ACS Sustain Chem Eng*. 2017. Vol. 5, № 5. P. 4320–4331.
35. Rezaei S. et al. Exploring 3D-QSPR models of human skin permeability for a diverse dataset of chemical compounds // *Journal of Receptors and Signal Transduction*. 2019. Vol. 39, № 5–6. P. 442–450.
36. Oztan Akturk S., Tugcu G., Sipahi H. Development of a QSAR model to predict comedogenic potential of some cosmetic ingredients // *Computational Toxicology*. 2022. Vol. 21. P. 100207.
37. Kim J.Y., Kim K.-B., Lee B.-M. Validation of Quantitative Structure-Activity Relationship (QSAR) and Quantitative Structure-Property Relationship (QSPR)

- approaches as alternatives to skin sensitization risk assessment // *J Toxicol Environ Health A*. 2021. Vol. 84, № 23. P. 945–959.
38. Liu R. et al. QSPR models for sublimation enthalpy of energetic compounds // *Chemical Engineering Journal*. 2023. Vol. 474. P. 145725.
 39. Wang B. et al. Prediction of Minimum Ignition Energy from Molecular Structure Using Quantitative Structure–Property Relationship (QSPR) Models // *Ind Eng Chem Res*. 2017. Vol. 56, № 1. P. 47–51.
 40. Zhang S. et al. Evaluating the properties of ionic liquid at variable temperatures and pressures by quantitative structure–property relationship (QSPR) // *Chem Eng Sci*. 2021. Vol. 231. P. 116326.
 41. Yan F. et al. QSPR models for the properties of ionic liquids at variable temperatures based on norm descriptors // *Chem Eng Sci*. 2020. Vol. 217. P. 115540.
 42. Roubehie Fissa M. et al. Development of QSPR-ANN models for the estimation of critical properties of pure hydrocarbons // *J Mol Graph Model*. 2023. Vol. 121. P. 108450.
 43. Li R. et al. Machine learning-quantitative structure property relationship (ML-QSPR) method for fuel physicochemical properties prediction of multiple fuel types // *Fuel*. 2021. Vol. 304. P. 121437.
 44. Jiao L. et al. QSPR Studies on the Octane Number of Toluene Primary Reference Fuel Based on the Electrotopological State Index // *ACS Omega*. 2020. Vol. 5, № 8. P. 3878–3888.
 45. Randić M. Molecular bonding profiles // *J Math Chem*. 1996. Vol. 19, № 3. P. 375–392.
 46. Rogers D., Hahn M. Extended-Connectivity Fingerprints // *J Chem Inf Model*. 2010. Vol. 50, № 5. P. 742–754.
 47. Bahia M.S. et al. A comparison between 2D and 3D descriptors in QSAR modeling based on bio-active conformations // *Mol Inform*. 2023. Vol. 42, № 4.

48. Verloop A., Hoogenstraaten W., Tipker J. Development and Application of New Steric Substituent Parameters in Drug Design // Drug Design. Elsevier, 1976. P. 165–207.
49. Bogdanov B., Nikolić S., Trinajstić N. On the three-dimensional wiener number // J Math Chem. 1989. Vol. 3, № 3. P. 299–309.
50. Mihalic Z., Nikolic S., Trinajstic N. Comparative study of molecular descriptors derived from the distance matrix // J Chem Inf Comput Sci. 1992. Vol. 32, № 1. P. 28–37.
51. Hollas B. THE AUTOCORRELATION OF A TOPOLOGICAL STRUCTURE: A NEW MOLECULAR DESCRIPTOR // J Math Chem. 2003. Vol. 33, № 2. P. 91–101.
52. Gramatica P. WHIM Descriptors of Shape // QSAR Comb Sci. 2006. Vol. 25, № 4. P. 327–332.
53. Mulliken R.S. Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I // J Chem Phys. 1955. Vol. 23, № 10. P. 1833–1840.
54. Stanton D.T. et al. Charged partial surface area (CPSA) descriptors QSAR applications // SAR QSAR Environ Res. 2002. Vol. 13, № 2. P. 341–351.
55. Wold S., Sjöström M., Eriksson L. PLS-regression: a basic tool of chemometrics // Chemometrics and Intelligent Laboratory Systems. 2001. Vol. 58, № 2. P. 109–130.
56. Athanasios Kondylis. Smooth PLS Regression for Spectral Data // REVSTAT-Statistical Journal. 2022. Vol. 20, № 4. P. 463–479.
57. Gopala Krishna J., Roy K. QSPR modeling of absorption maxima of dyes used in dye sensitized solar cells (DSSCs) // Spectrochim Acta A Mol Biomol Spectrosc. 2022. Vol. 265. P. 120387.
58. Eichenlaub J., Rakowska P.W., Kloskowski A. User-assisted methodology targeted for building structure interpretable QSPR models for boosting CO₂ capture with ionic liquids // J Mol Liq. 2022. Vol. 350. P. 118511.
59. Xu H., Caramanis C., Mannor S. Robustness and Regularization of Support Vector Machines. 2008.

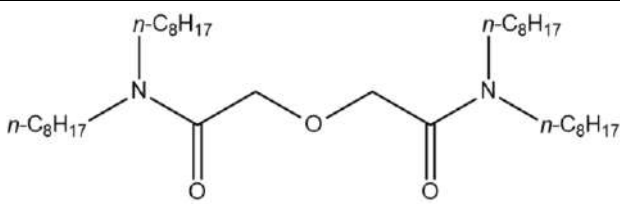
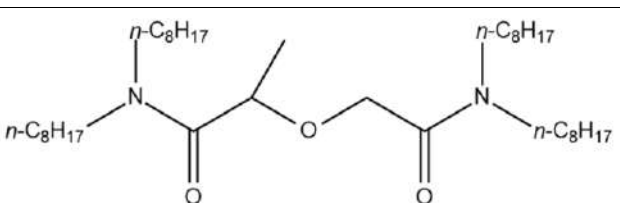
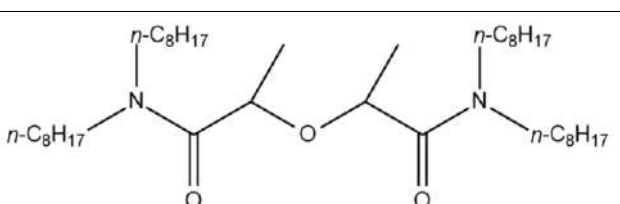
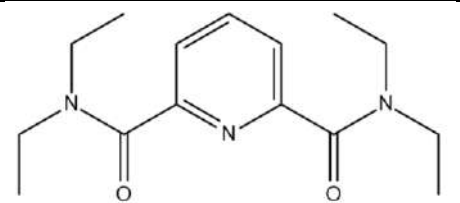
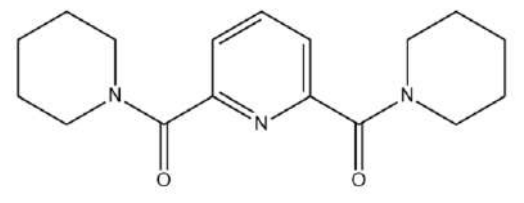
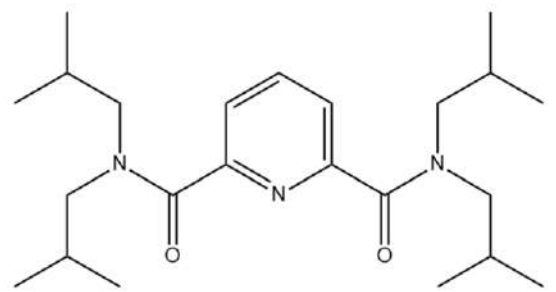
60. Soloviev V. et al. QSPR modeling of potentiometric sensitivity towards heavy metal ions for polymeric membrane sensors // *Sens Actuators B Chem.* 2019. Vol. 301. P. 126941.
61. Martynko E. et al. QSPR Modeling of Potentiometric Mg^{2+}/Ca^{2+} Selectivity for PVC-plasticized Sensor Membranes // *Electroanalysis.* 2020. Vol. 32, № 4. P. 792–798.
62. Eslam Pourbasheer. High Accurate Prediction of Carbonate Selectivity of PVC-Plasticized Membranes Sensors by Genetic Algorithm-Support Vector Machine // *Analytical and Bioanalytical Electrochemistry.* 2023. Vol. 15, № 2. P. 150–165.
63. Roya Kiani-anboui, Zeinab Mozafari. A Novel QSPR Approach in Modeling Selectivity Coefficients of the Lanthanum-Selective Electrode // *Analytical and Bioanalytical Electrochemistry.* 2023. Vol. 15, № 12. P. 1031–1045.
64. Diem-Tran P.T. et al. Stability Constant and Potentiometric Sensitivity of Heavy Metal–Organic Fluorescent Compound Complexes: QSPR Models for Prediction and Design of Novel Coumarin-like Ligands // *Toxics.* 2023. Vol. 11, № 7. P. 595.
65. Abbott A.P. et al. Preparation of novel, moisture-stable, Lewis-acidic ionic liquids containing quaternary ammonium salts with functional side chains // *Chemical Communications.* 2001. № 19. P. 2010–2011.
66. Lemaoui T. et al. Quantitative structure properties relationship for deep eutectic solvents using σ -profile as molecular descriptors // *J Mol Liq.* 2020. Vol. 309. P. 113165.
67. Boubliia A. et al. Molecular-based artificial neural network for predicting the electrical conductivity of deep eutectic solvents // *J Mol Liq.* 2022. Vol. 366. P. 120225.
68. Salahshoori I., Baghban A., Yazdanbakhsh A. Novel hybrid QSPR-GPR approach for modeling of carbon dioxide capture using deep eutectic solvents // *RSC Adv.* 2023. Vol. 13, № 43. P. 30071–30085.

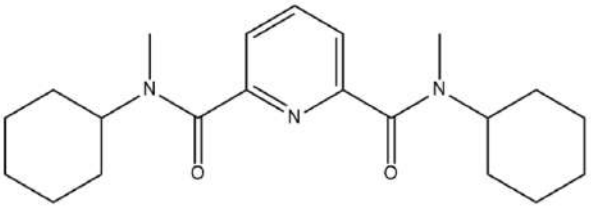
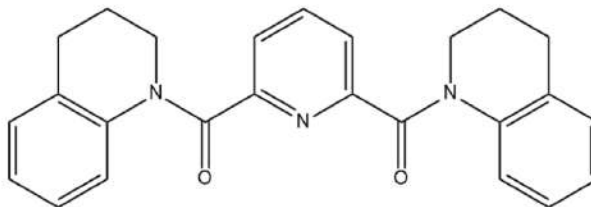
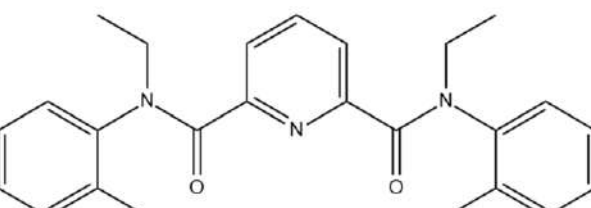
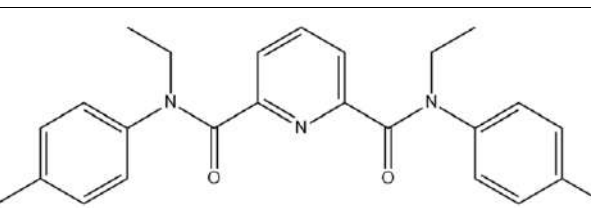
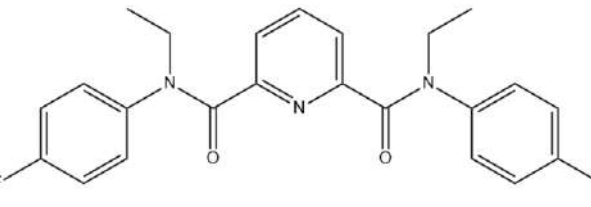
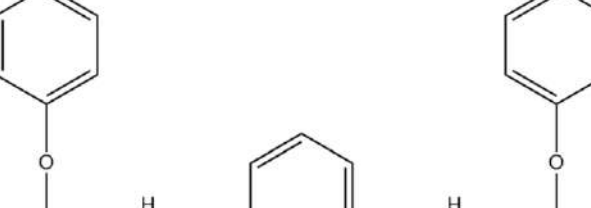
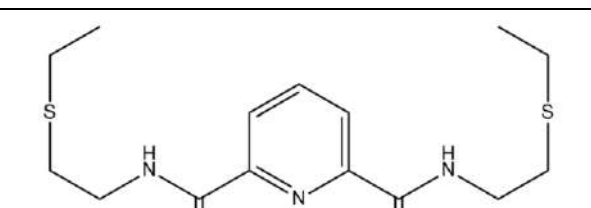
69. Hu J. et al. COSMO-SAC and QSPR combined models: A flexible and reliable strategy for screening the extraction efficiency of deep eutectic solvents // *Sep Purif Technol.* 2023. Vol. 315. P. 123699.
70. Fedorova E.S. et al. Deep learning for retention time prediction in reversed-phase liquid chromatography // *J Chromatogr A.* 2022. Vol. 1664. P. 462792.
71. Obradović D. et al. A comparative study of the predictive performance of different descriptor calculation tools: Molecular-based elution order modeling and interpretation of retention mechanism for isomeric compounds from METLIN database // *J Chromatogr A.* 2024. Vol. 1719. P. 464731.
72. Turanov A.N., Karandashev V.K., Yarkevich A.N. Extraction of Rare Earth Elements(III) from Perchlorate Solutions with Modified Diphenylphosphorylacetamides // *Russian Journal of Inorganic Chemistry.* 2021. Vol. 66, № 4. P. 572–577.
73. Sokalski T. et al. Observations on the behaviour of some trifluoroacetophenone derivatives as neutral carriers for carbonate ion-selective electrodes // *Analyst.* 1996. Vol. 121, № 2. P. 133–138.
74. Behringer C. et al. Anion selectivities of trifluoroacetophenone derivatives as neutral ionophores in solvent-polymeric membranes // *Anal Chim Acta.* 1990. Vol. 233. P. 41–47.
75. Maj-Zurawska M. Carbonate ion selective electrodes with trifluoroacetophenone derivatives in potentiometric clinical analyser // *Talanta.* 1997. Vol. 44, № 9. P. 1641–1647.
76. Hong Y.K. et al. Effect of varying quaternary ammonium salt concentration on the potentiometric properties of some trifluoroacetophenone derivative-based solvent-polymeric membranes // *Electroanalysis.* 1997. Vol. 9, № 11. P. 865–868.
77. Shim J. Ion-selective electrodes based on molecular tweezer-type neutral carriers // *Talanta.* 2004. Vol. 63, № 1. P. 61–71.

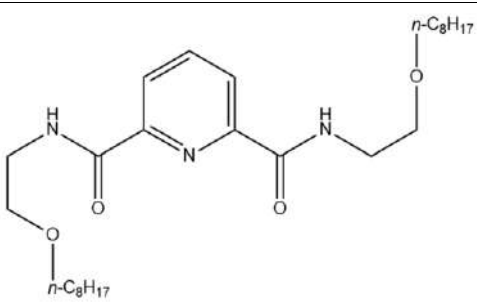
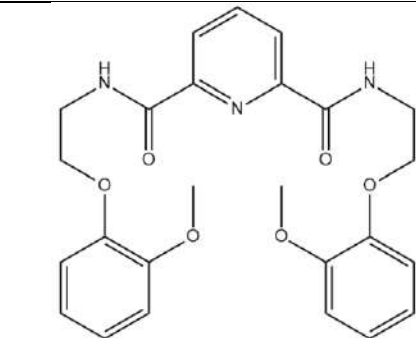
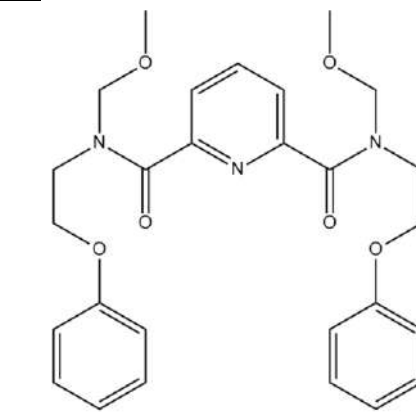
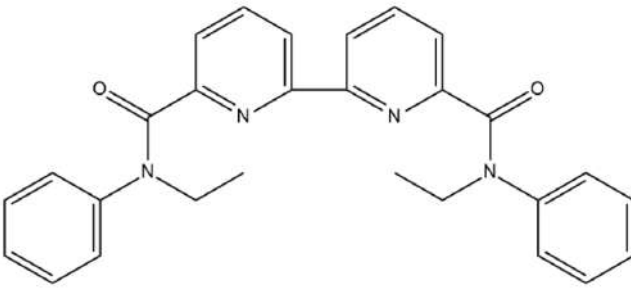
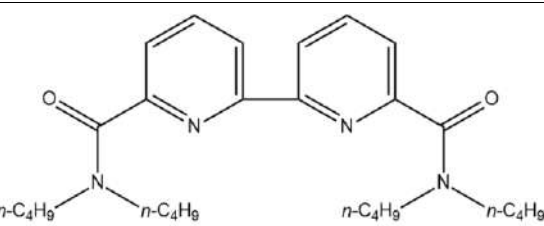
78. KAMATA S., NOMURA S., OHASHI K. Anion-selective membrane electrode based on bis(diphenylphosphino)alkane-copper(II) complexes. // BUNSEKI KAGAKU. 1990. Vol. 39, № 11. P. 677–681.
79. Rothmaier M. et al. Response mechanism of anion-selective electrodes based on mercury organic compounds as ionophores // Anal Chim Acta. 1996. Vol. 327, № 1. P. 17–28.
80. Umezawa Y. et al. Potentiometric Selectivity Coefficients of Ion-Selective Electrodes. Part I. Inorganic Cations (Technical Report) // Pure and Applied Chemistry. 2000. Vol. 72, № 10. P. 1851–2082.
81. Solov'ev Vitaly, Varnek Alexandre. Qspr Models on Fragment Descriptors [Electronic resource] // <http://vpsolovev.ru/wp-content/uploads/sites/9/2017/05/isida-qspr-help-2017.pdf>.
82. Meyerhoff M.E. et al. Role of trifluoroacetophenone solvents and quaternary ammonium salts in carbonate-selective liquid membrane electrodes // Anal Chem. 1987. Vol. 59, № 1. P. 144–150.
83. Kirsanov D. et al. Potentiometric Sensor Array for Analysis of Complex Rare Earth Mixtures // Electroanalysis. 2012. Vol. 24, № 1. P. 121–130.
84. Zhang J. et al. One Step Closer to an Ideal Insensitive Energetic Molecule: 3,5-Diamino-6-hydroxy-2-oxide-4-nitropyrimidone and its Derivatives // J Am Chem Soc. 2021. Vol. 143, № 32. P. 12665–12674.
85. Every H. et al. Ion diffusion in molten salt mixtures // Electrochim Acta. 2000. Vol. 45, № 8–9. P. 1279–1284.
86. Saleh M.A. et al. Density and Viscosity of 1-Alkanols // Phys Chem Liquids. 2004. Vol. 42, № 6. P. 615–623.
87. Galvez J. et al. Charge Indexes. New Topological Descriptors // J Chem Inf Comput Sci. 1994. Vol. 34, № 3. P. 520–525.

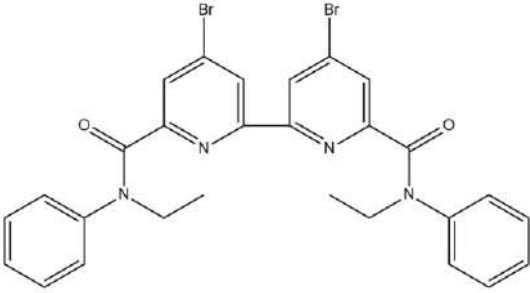
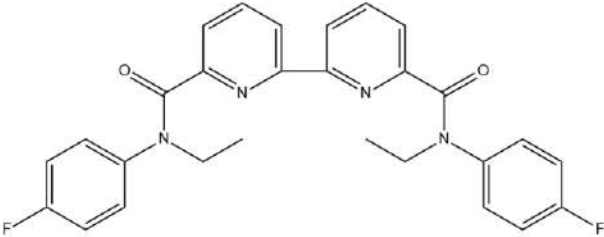
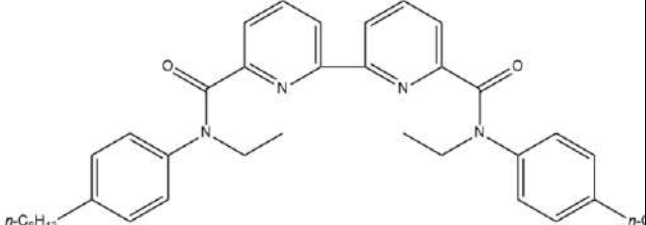
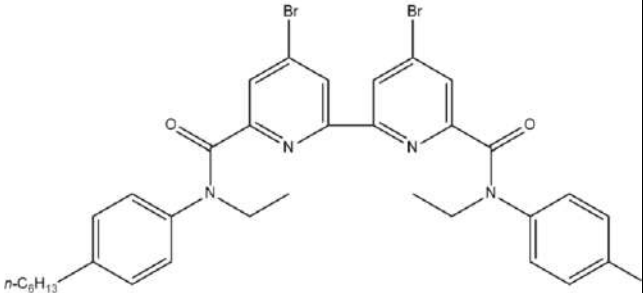
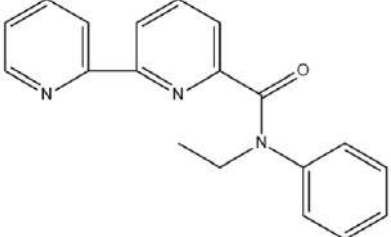
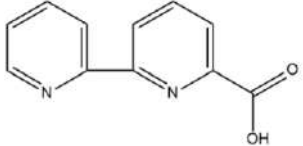
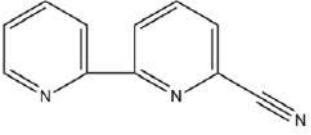
APPENDIX A

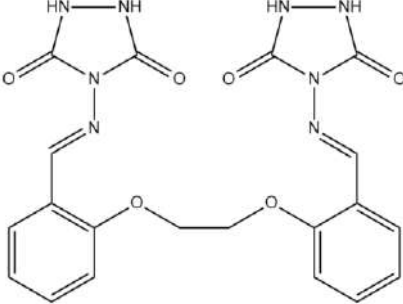
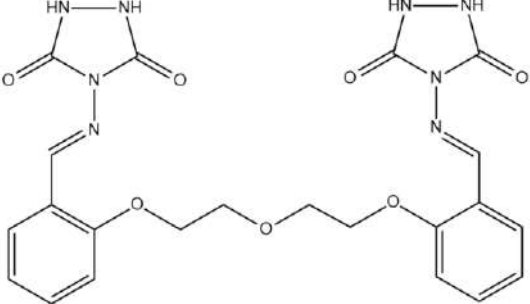
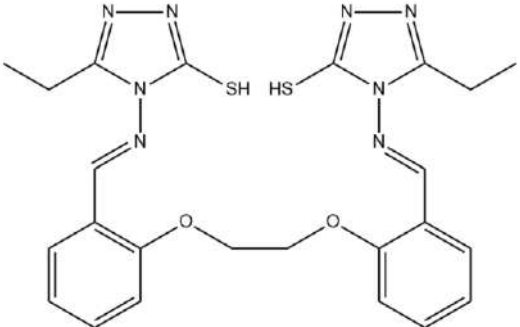
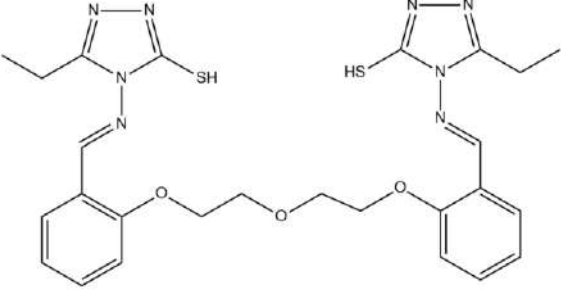
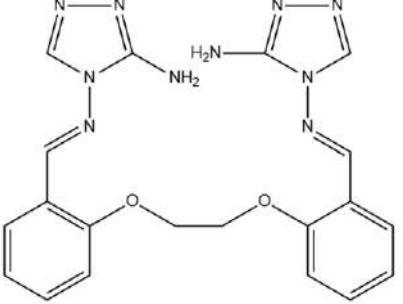
Table 1. Structures and heavy metal sensitivities of ionophores

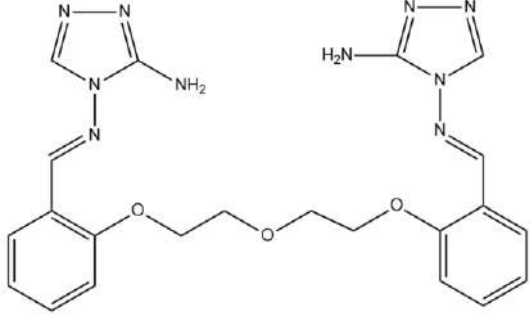
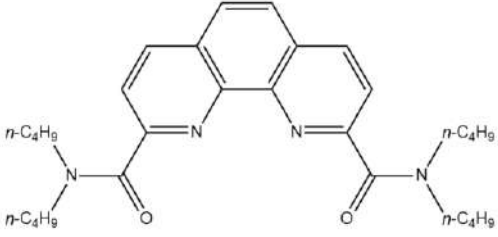
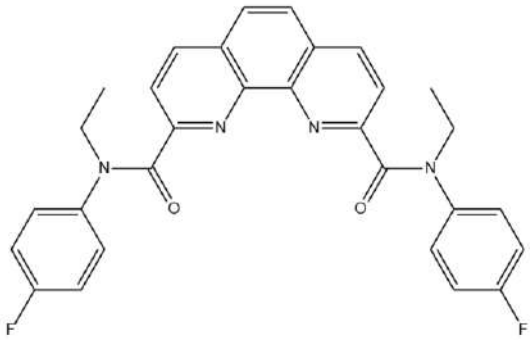
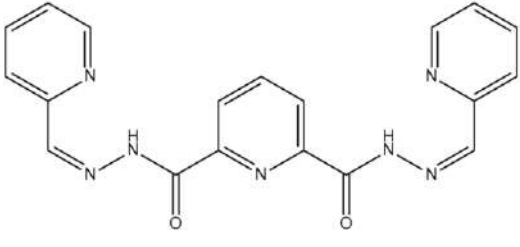
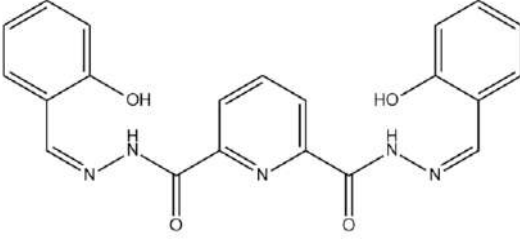
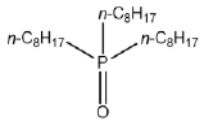
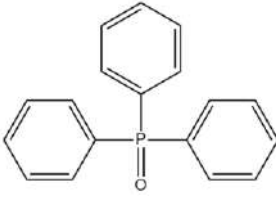
#	Ionophore structure	Sensitivity, mV/dec		
		Cd ²⁺	Cu ²⁺	Pb ²⁺
1		9	5	30
2		13	12	24
3		13	9	18
4		14	25	27
5		14	23	26
6		27	34	51

7		21	31	37
8		26	34	34
9		32	43	44
10		32	43	45
11		19	37	37
12		23	28	28
13		22	27	28

14		24	31	31
15		25	28	29
16		22	25	42
17		36	31	24
18		37	39	26

19		36	26	28
20		31	30	23
21		36	34	23
22		41	31	27
23		3	-12	0
24		6	1	0
25		5	-10	3

26		16	15	24
27		15	15	24
28		17	19	25
29		24	34	24
30		18	28	27

31		18	20	28
32		26	24	31
33		27	23	26
34		7	0	0
35		5	0	4
36		18	9	9
37		-10	5	-20

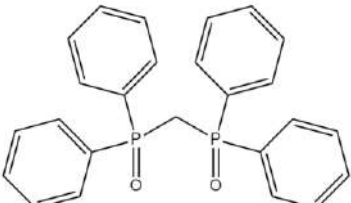
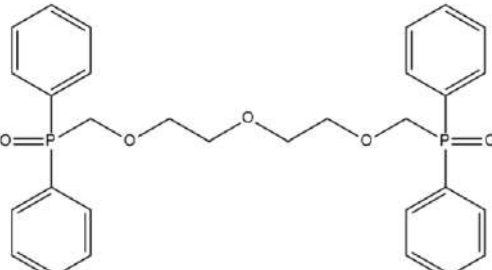
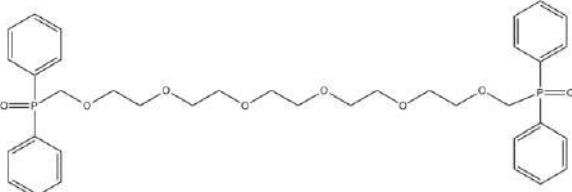
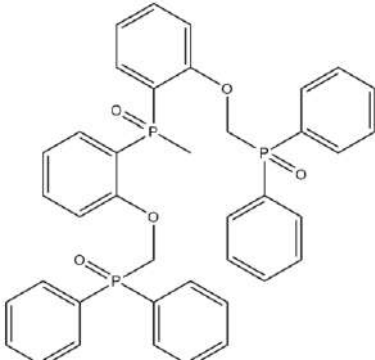
38		23	30	38
39		25	30	33
40		27	26	40
41		24	20	16

Table 2. Properties of DESs based on acids and their mixtures with water. Asterisk (*) denotes calculated values

Acid in DES composition	Water content, wt. %					
	0	10	20	30	40	50
	Refractive index					
Lactic	1.4811 (*)	1.4660	1.4505	1.4350	1.4170	1.4035
Tartaric	1.4910	1.4910	1.4635	1.4460	1.4290	1.4115

Malic	1.4875	1.4870	1.4575	1.4395	1.4245	1.4060
Citric	1.4955	1.4960	1.4645	1.4455	1.4295	1.4130
Malonic	1.4830	1.4665	1.4500	1.4335	1.4200	1.4020
Glycolic	1.4856	1.4675	1.4565	1.4445	1.4360	1.4290
Density, g/cm ³						
Lactic	1.172 (*)	1.155	1.139	1.122	1.104	1.087
Tartaric	1.330	1.297	1.260	1.226	1.190	1.156
Malic	1.271	1.241	1.218	1.187	1.163	1.129
Citric	1.340	1.306	1.268	1.226	1.189	1.156
Malonic	1.232	1.211	1.181	1.161	1.140	1.113
Glycolic	1.203	1.176	1.157	1.137	1.117	1.099
Viscosity, mPa·S						
Lactic	-	153.6	36.32	13.55	6.752	4.297
Tartaric	4504994	2435.7	157.5	29.07	10.09	5.099
Malic	19098.6	400.8	69.2	18.43	8.499	4.345
Citric	10099000	8327	342.7	34.39	11.27	5.598
Malonic	2208.41	147	21.8	10.83	5.944	3.454
Glycolic	797.6	73.1	20.08	9.185	4.996	3.373
Conductivity, mS/cm						
Lactic	-	5	16	34	55	75

Tartaric	0.01	0.015	5	16	34	53
Malic	0.077	0.13	9	26	43	63
Citric	-	0.002	3	14	29	51
Malonic	0.38	4.47	16.15	34.5	53	77
Glycolic	1	8	18	34	56	79