UNIVERSITY OF TARTU

FACULTY OF SCIENCE AND TECHNOLOGY

Institute of Chemistry

Master's Thesis (30 EAP)
Materials Science and Technology

Sabina Tatarytskaya

SEARCH AND INVESTIGATION OF PEPTIDE STRUCTURES FOR CARBOXYLATE BINDING

Supervisors:
Research Fellow Astrid Darnell
Research Fellow Sofja Tšepelevitš
Professor Ivo Leito

TARTU 2023

Table of contents

List of abbreviations

AA	Amino acid
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
CCDC	Cambridge Crystallographic Data Centre
COSMO-RS	Conductor-like Screening Model for Realistic Solvation
CSD	Cambridge Structural Database
CSM	Continuum Solvation Model
DFT	Density functional theory
DMSO	Dimethylsulfoxide
GC-MS	Gas chromatography–mass spectrometry
Glu	Glutamic acid
Gly	Glycine
$Glyp^{2-}$	Glyphosate dianion
HB	Hydrogen bond
HBD	Hydrogen bond donor
His	Histidine
Ile	Isoleucine
IMHB	Intramolecular hydrogen bond
LC-MS	Liquid chromatography–mass spectrometry
Leu	Leucine
Lys	Lysine
MC	Macrocyclic receptor
Met	Methionine
Phe	Phenylalanine
Pro	Proline
RM	Reference molecule
Ser	Serine
SMD	Solvation Model based on charge Density
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine

1. INTRODUCTION

Anions, including carboxylates, are essential chemical species that can be found everywhere from food products to active pharmaceutical substances. Therefore, it is crucial to be able to determine them not only accurately but also conveniently. In this regard, electrochemical devices like solid-state ion-selective electrodes (ISE) would be useful.

However, few of them exist for carboxylates, thus provoking interest in their development. Anion sensors require ionophores (anion receptors) – molecules that are able to bind anions – which are usually molecules with numerous hydrogen bond donor (HBD) sites, as working substances in the membrane. Extensive research has been conducted resulting in many promising receptor candidates; however, their selectivity towards any particular anion was generally moderate or low. To increase selectivity markedly, molecules of significantly sophisticated structures have to be designed. This is generally associated with difficulties in synthesis.

For the following reasons, peptides were chosen as molecules worth investigating. First of all, they have numerous HBD sites. Secondly, the geometry of peptide – crucial for suitable orientation of the HBD sites – could be changed via changing its amino acid (AA) sequence. Lastly, the synthesis of even sophisticated peptides is highly standardised and can be done relatively easily.

The main goal of this work is to devise a peptide sequence that can bind carboxylate anions strongly enough. The primary anion of interest is acetate, but peptides can be potentially used for detecting other carboxylates as well.

The specific aims of this master's thesis are the following:

1. Find peptide fragments through a systematic search of the CCDC database according to these criteria:
    a. Suitable number of HBD groups: 3 or 4 NH and/or OH groups is desirable;
    b. Cavity to fit acetate anion;
    c. Similar orientation of HBD groups: facing inwards, towards the structural cavity;
    d. Rather small number of AA (no more than 20);
2. To predict binding affinities of the found receptor candidates in DMSO with 0.5% water by using the COSMO-RS method and compare these values with previously studied anion receptors;
3. To determine and characterize the differences in binding constants between linear and cyclic peptides and the impact of various linkers – organic units that connect parts of a molecule – on binding;
4. Based on these findings, to try to design a peptide-based structure that could be a useful candidate for synthesis and further experimental studies as an ionophore in an ISE for acetate recognition.

2. LITERATURE REVIEW

2.1. Importance of carboxylate anions

Carboxylates are widely distributed both in nature and technology as they occur in plant and animal sources and have numerous roles and uses, such as food preservatives (sodium benzoate), vinegar production (acetic acid), skin care products (salicylic acid), bio- and polymers manufacturing, and anti-inflammatory drugs (aspirin)[1]. Amino acids, both essential and nonessential, are carboxylic acids, and enzymes, which are built up from amino acids, contain carboxyl functional groups[2]. Thus, carboxylic acids are intermediate products in degradation pathways of enzymes, fats, and carbohydrates[3].

Carboxylic acids play an important role in food industry due to the effect on organoleptic properties and stability of products. They are present in foods naturally (citric acid in fruits) or are added artificially (citric acid as acidulant). Additionally, organic acids are effective antimicrobial agents, which makes them useful in dairy and bakery industry. Considering the aforementioned, the content and nature of carboxylic acids is relevant for qualitative and quantitative control, and new analytical methods have to be developed and improved on a regular basis[3].

Pharmaceutical chemistry uses carboxylic acids substantially. While some of them are either precursors or reagents in synthesis, the main areas where carboxylic acids are used can be listed as the following: solubilizers for antihistaminic drug classes; prodrugs that convert to the active substance under specific conditions (antithrombotic drugs); and pharmacophore to interact with an enzyme (blood cholesterol-reducing drugs). Besides, medicines containing carboxylate acids are widely used for pain treatment[3].

Moreover, carboxylic acids take part in several biologically important molecular recognition processes and occur in different metabolic cycles (e.g. the Krebs cycle)[3]. In this regard, organic acid analysis is done to monitor the inborn errors of metabolism; however, here the gas chromatography-mass spectrometry (GC-MS) together with derivatization is the most suitable method[4]. Although GC-MS and LC-MS are widely used and reasonably accurate techniques, they require both time and effort, and in some areas, like industrial fermentation reactors, it would be highly beneficial to constantly monitor carboxylates *in citu* without the need to take samples. Therefore, carboxylate sensors could find an application also in food and pharmaceutical chemistry, where they would be used to control the concentration of carboxylates.

Thus, the search for a molecule capable of binding carboxylate anions to use for analytical purposes (e.g., for using in a sensor) continues to attract scientific interest[1]. The history of anion-binding research and the latest inventions will be covered in a **section 2.3** of this chapter.

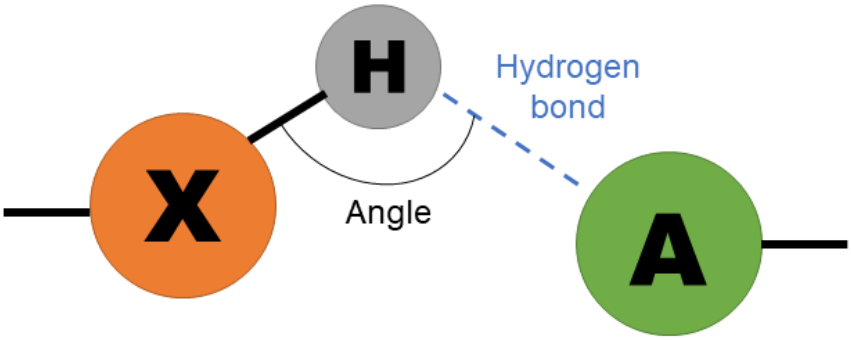
2.2. Hydrogen bond interactions

Hydrogen bond (HB) is a specific type of electrostatic interaction between a hydrogen atom bonded to an electronegative atom (typically O or N) and a lone pair of electrons on an electronegative atom like N, O, and F. The former is also known as an H-bond donor, the latter as an H-bond acceptor. HBs are directional, and their

formation efficiency is highly dependent on the angle of interaction, the environment, and donor and acceptor species[5].

The HB strength is strongly correlated with suitable directionality and the distance between a XH-hydrogen and a HB acceptor atom. HB is considered to be strong when the bond angle falls within a narrow range of 175–180°, and the distances between the hydrogen and the acceptor atom are small (1.2–1.5 Å)[6]. This research kept focus on very strong and strong HBs with angles between 130 and 180° and the distances from XH-hydrogen to A within 2.2–3.0 Å (Scheme 1).

Scheme 1. Hydrogen bond. XH – the donor group, A – acceptor atom.



The surrounding environment has a huge impact on HB formation between anion and receptor of interest. When the medium is rich in water, the chance of interaction between the desired molecules is reduced. The same logic applies for solvents with strong HB donor abilities: they tend to have higher affinities towards anions than potential binding molecules[7].

Hydrogen bonds are crucial in our study because they are the main interactions between a receptor molecule and an anion of interest. To achieve selective binding, the geometry complementarity criterion should be fulfilled. This criterion, in turn, is related to HB formation occurring between the anion and receptor molecule. Moreover, this criterion resembles the “lock-and-key model” in a way that a specific anion could bind with one receptor. On one hand, it could be achieved using the directionality of HBs, because a strong bond will occur only under specific angle and distance, and the ability of anions to form multiple HBs with the binding groups of the receptor simultaneously[8]. On the other hand, achieving selectivity is a challenging task, because the HB interaction with the receptor molecule occurs with the carboxylate anionic centre, which is similar for all carboxylates[9].

In addition, other interactions like electrostatic interactions (ion pair formation, for example) and hydrophobic effect can happen simultaneously and influence the binding strength[8]. However, as this study mostly focused on HBs, they were neglected.

2.3. Previous research on anion binding

The field of anion coordination chemistry received significant attention in the past decades. It is developing moderately fast as the intrinsic properties of anions raise some issues that are challenging to overcome. At the same time, huge number of articles on ligands for both carboxylates and other anions were published[10]. Considering the focus of this work, only studies on carboxylates will be highlighted below.

The performance of an anion receptor is evaluated according to its binding affinity value. It is quantified by the binding (association) constant K_{ass} in the following equilibria, where R is the receptor molecule, A is anion and RA is anion-receptor complex. a designates the activities of different species.

$$R + A^- \rightleftarrows RA^- \quad (1)$$

$$K_{ass} = a_{RA^-}/a_R a_{A^-} \quad (2)$$

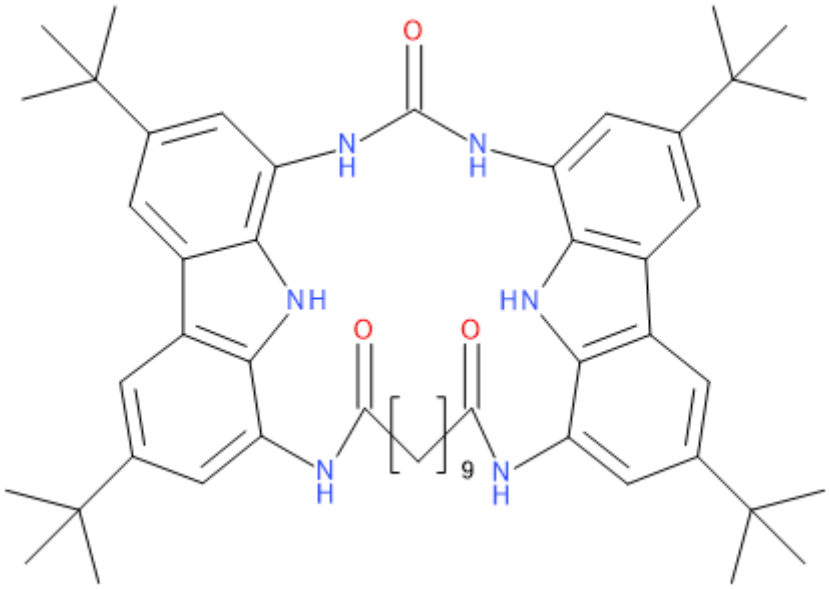
Initially, researchers designed and investigated acyclic receptors containing different units as carboxylate ligands. For instance, Gale and co-workers found that bis-urea based anion receptor with two *ortho*-phenylenediamine groups is capable of binding carboxylates selectively in solution[11]. Later, a series of fluorescent carbazolylurea-based anion receptors were synthesised and reported to have high affinity towards acetate and bicarbonate[12].

More detailed research was then performed by Kadam and co-workers[1]. They studied the binding constants between four carboxylate anions (acetate, trimethylacetate, benzoate and lactate) and thirty-eight different synthetic indole-, urea-, thiourea-, carbazole- and indolocarbazole-based receptors. It was found that the binding is determined by basicity of anion and is the highest for trimethylacetate, although in general, the ability of these receptors to differentiate between small carboxylates is low.

Although the progress with anion receptors was sufficient to help improve their selectivity and sensitivity, there were no comprehensive analysis on the structural properties of carboxylates and their influence on binding with synthetic receptors. Therefore, Martin and co-workers focused their research on it[13]. The quantitative characterization of binding affinity of 11 carboxylates with 22 synthetic multidentate receptors was performed. As a result, more than 200 binding constants were experimentally obtained for all studied receptors. The binding affinity was shown to be strongly influenced by anion basicity, and in case of acetate it is additionally governed by the sterical suitability of the binding site.

Following the finding on importance of the binding cavity, a closer research on the effects of ring size of biscarbazolylurea-based structures was done[14]. It was demonstrated that the ring with 3 to 5 CH_2 groups was too small to accommodate acetate, while the ring with 9 methylene groups (Figure 1) showed the best binding affinities. The experimentally obtained values for binding affinities were in the range of 2.1–5.7 units. Overall, both stronger binding and higher selectivity correlated with the suitability of cavity size.

Figure 1. Macrocyclic receptor with the best observed affinity[14].



A comprehensive data set of binding constants for a series of anion receptors with 12 carboxylates was published recently[9]. This work summarizes the studied affinities and lipophilicities of carbazole-, indolocarbazole-, urea-, and macrocycle-based structures to allow their comparison and ease further search for selective carboxylate binder.

One of the papers[15] describes a study on 12 multidentate anion receptors for binding glyphosate dianion ($Glyp^{2-}$). An interest towards this particular anion lies in its application. It is a widely used herbicide with low concentrations in the environment and high hydrophilicity. The structures of anion receptors were based on indolocarbazole, carbazole, and diphenyl (diaryl) urea moieties. Their design, synthesis and binding affinity measurements in DMSO/water mixtures were performed. Additionally, computational estimations of binding strength were done for comprehensive characterization.

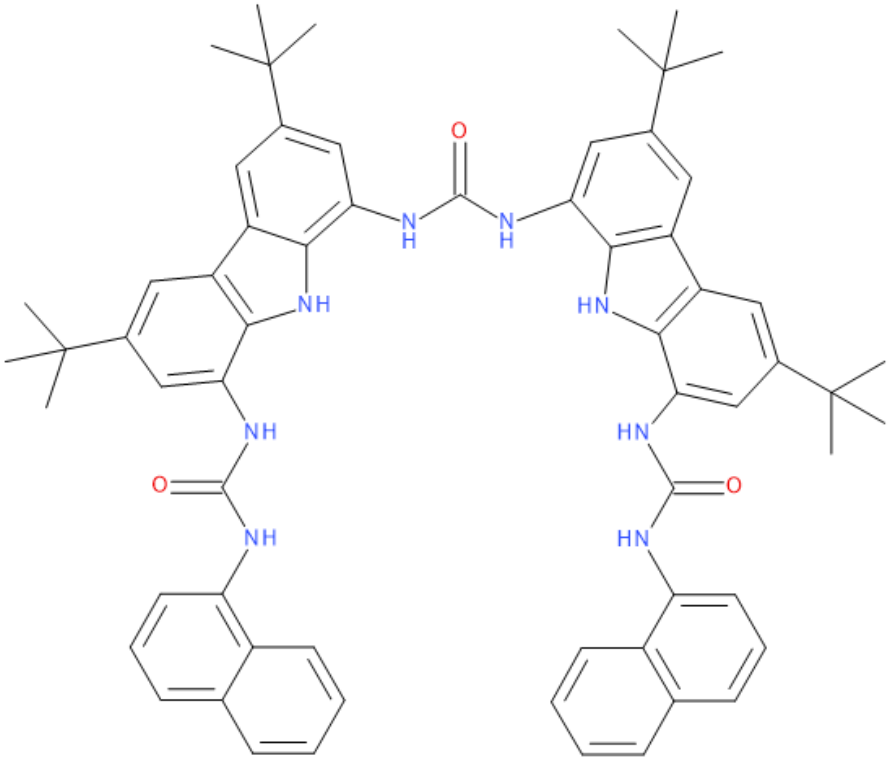
Results of this study show that the properties of moieties, such as flexibility, length and angle, have a strong impact on binding. For instance, the binding affinities with $Glyp^{2-}$ decrease with growth of the linker chain, and the substitution in the aromatic ring of spacer, 1,2-substitution in particular, enables better orientation and leads to stronger binding. Moreover, one of the studied molecules (named “receptor 17”[15], Figure 2) showed outstandingly high binding affinity towards $Glyp^{2-}$ in 5% H_2O/DMSO ($\log K_{ass}$ = 6.5), therefore became the strongest binder.

Binding of dicarboxylates and finding their selective receptors has also been of a huge interest. The review by Butler and Jolliffe[16] summarizes the progress in their detection and recognition. The examples of receptors include simple molecules like diguanidium, reported by Sessler, Thordarson and Gong[17], and more sophisticated molecules like macrocycles or cages and cryptands[18,19].

One interesting and ground-breaking approach, pioneered by Anslyn[20], is the use of sensing arrays. It involves multiple complementary binding agents. Although the individual receptor molecule is not selective towards particular dicarboxylate, their

combination and further analysis of the responses may be used to identify and quantify each analyte separately[16].

Figure 2. Receptor 17 from reference 15.



2.4. Peptides as the molecules of interest

Peptides are short chains of amino acid residues (AA). They typically contain 2 – 50 AA residues. To compare, a protein is a long and continuous peptide string, usually comprising 51+ AA. AAs are linked by peptide bonds, where the N atoms (participating in peptide bonding) of all amino acid residues except proline have one hydrogen atoms and could potentially become a hydrogen-donor bonding site[21].

This work considers peptides as potential carboxylate binders for several reasons. Firstly, as was briefly mentioned in previous paragraph, peptides have multiple HBD fragments. The NH fragment within the amide group connecting AAs can become a binding site for a carboxylate: as the amide groups are located close to each other, the anion could interact with 3 or 4 H atoms at the same time.

Moreover, some of the AA residues (Ser, Asn, Gln, Tyr, Thr, and Trp) have an OH or amide groups in their side chains, which could assist in binding. These groups can either directly interact with the anion or hold the peptide structure in a suitable conformation so that the anion could bind[22].

Secondly, peptide structure design is highly adaptable: the AA sequence can be modified in a way to achieve the best binding. Furthermore, it is possible to synthesize cyclic peptides with a suitable cavity size to allow hydrogen-bonding sites to face in the same direction. Lastly, peptides (e.g. enzymes) naturally interact with anions in our bodies. For example, zinc enzyme carboxy-peptidase A binds with

small inorganic anions like phosphate via hydrogen bonding and anion-metal interaction[2].

Anion binding using cyclic peptides has been previously investigated by other groups. Such receptors are usually cyclic hexapeptides and their modifications (with a linker or solubilizing groups in the structure)[23]. The advantages of these molecular structures are large amount of donor sites pointing into internal cavity, symmetry and constrained conformations[24].

In particular, several publications reported peptide-containing receptors for halides binding. In 2002, Yang *et al.* synthesised and investigated a cyclic hexapeptide derived from D,L-α-aminoxy acids[25]. Due to a preferred highly symmetrical bracelet-like conformation, this peptide demonstrated affinities for halides with the highest value for Cl^- in nonpolar solvents.

Another example is a neutral cyclic hexapeptide with L-proline and 6-aminopicolinic acid subunits described by Kubik et al.[26]. It was reported to interact with both halides and sulphate in a 50% water/methanol mixture, but has a particularly high affinity and selectivity for sulphate. The research on linear hexapeptides was also performed, and Pajewski and co-workers were the first to mention and investigate non-cyclic diglycolylated heptapeptides based on purely natural amino acids[27].

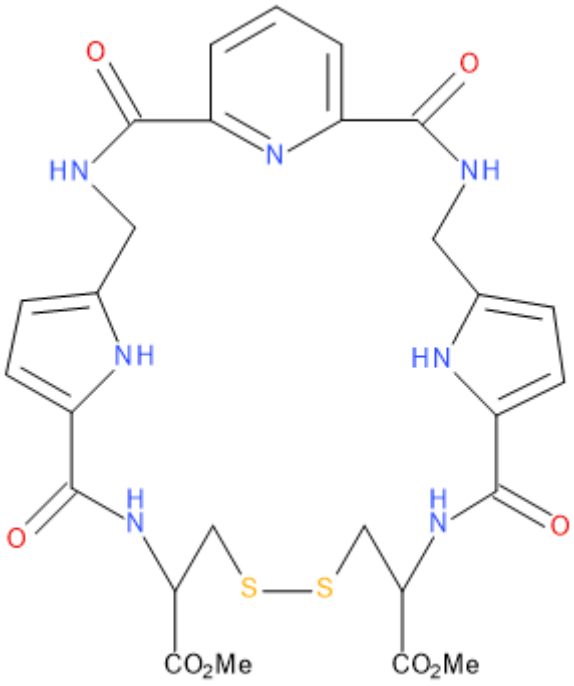
Interestingly, acetate, as well as fluoride, were found to be bound quite well by two pseudo-cyclopeptides[28]. The structure of these receptors contains pyrrole, cystine, and pyridine, which make up a macrocycle. It is very conformationally constrained, thus provides better geometry to host anions of suitable size. The candidate that showed the best affinity and selectivity results is depicted Figure 3.

Later, another study on cyclopeptides was published by Kubik[24]. The anion-complex formation was investigated for two cyclopeptides covalently connected through a linker. It was demonstrated that these receptors differ from single cyclopeptides and from each other in terms of binding thermodynamics. Additionally, they possess considerably different anion selectivity: they fail to bind sulphate and prefer to bind weakly coordinating anions.

Although some of the previously mentioned peptide-based receptors showed promising results on binding, none of them were used and investigated in an actual electrochemical sensor. This is precluded by several challenges[14,29]:

- Aqueous environments. The sensor should work under these conditions, but the majority of studies use different solvent systems because (1) the majority of the investigated molecules are poorly soluble or insoluble in water, and (2) carboxylates are strongly solvated in water, which challenges achieving high-affinity binding because of desolvation penalty.
- Insufficient selectivity. Although this is not the focus of this thesis, the main problem with anion receptors is their low capability to distinguish between different species. This brings us to the point where there is no need to construct a real device when the preliminary results in solution are not satisfactory.
- Lipophilicity. The sensing molecule should be sufficiently hydrophobic to stay in the ISE membrane. However, to interact with anion there must be hydrophilic parts, which can lead to molecule leaking out of the membrane.

Figure 3. Artificial receptor to selectively bind acetate and fluoride[28].



3. TARGET PROPERTIES

The main objective of the research is finding effective anion binders for carboxylate anions. The search for these molecules is initially conducted by targeting the acetate anion, though it could be modified and applied to more complex carboxylates. Peptide-based structures were investigated to try and find suitable receptor candidates that would meet the following **requirements**:

1. The receptor molecule should have a sufficient number of suitably positioned HBD sites to allow receptor-anion complex formation via strong HB interactions. The distance between binding groups and the directional orientation of the X-H hydrogens should be suitable to allow strong interactions with the target anion.
2. The cavity within the receptor structure should have an appropriate size to fit desired anion via additional interactions[13]. An example of such interactions is an interaction between hydrophobic groups of the receptor molecule and the hydrophobic part of the target anion. It should neither be too small, nor too large. In the former situation, the target anion would not fit in, whereas in the latter case significant reduction in selectivity would be expected as multiple anions could be incorporated in the moiety.
3. The conformational change from the free molecule to the molecule bound with acetate should be energetically inexpensive.
4. The molecule should contain the hydrophobic part(s). The hydrophobicity would help with practical use of the receptor molecule in a way that it would fit and stay in the lipophilic sensor membrane (as opposed to being leached out into the aqueous measurement medium) and, thus, make the sensor time-stable.
5. The peptide molecule should not contain AAs with acidic and basic groups due to their charged state in solution. This means that such AAs as Arg, Asp, Cys, Glu, His, Lys, Pyr, and Tyr should be avoided.
6. It should preferably be easily synthesizable to decrease the cost of research and production in the future.

The **strategy** chosen for finding the suitable molecule is performed in 4 steps: (1) search for peptide candidates in the database(s) according to the target properties, (2) choosing the peptide fragment(s) for detailed investigation, (3) low-level of theory computational study of chosen molecules, and (4) their structure optimization and more comprehensive computational study.

The tools and parameters were chosen/adjusted in a way to best suit the aforementioned principles. They are discussed in more details below.

4. METHODS

4.1. CCDC database

The Cambridge Crystallographic Data Centre (CCDC) is a registered non-profit research centre with leading experts in structural chemistry data committed to education and research. One of their main achievements is compilation and distribution of the Cambridge Structural Database (CSD), which contains crystallographic structural information on various chemical compounds[30]. The amount of CSD entities amounts to over 800,000, from which 43% are classified as organic. Those, as well as metal-organic structures, are suitable for pharmaceutical or material development applications. Moreover, numerous peptide structures (up to 24 residues) as well as larger proteins are a substantial part of CSD[31,32]. The CCDC's mission is to empower scientists to reach their research goals by providing high quality data.

Besides the database, CCDC has a software platform that helps database users to find structures that satisfy predefined criteria. The CCDC software was chosen for peptide-based structure screening for several reasons. Firstly, the data within CSD is accurate and consistent because it undergoes extensive validation and cross-checking. Moreover, it contains the structural information on unpublished compounds, which could be useful for finding new appealing structures. Secondly, the software itself is relatively easy to use and provides huge variety of search parameters, which can be optimized to fit the research needs. Lastly, all properties information about the studied molecules, such as bibliographic, chemical and physical data, is also included. It was empirically obtained in a crystalline phase for every investigated substance[31,33,34].

However, it is important to note one serious limitation of the database. It is about crystal structures, i.e. not the structures in solutions. This means that although we consider the molecule suitable (optimal geometry, cavity size, HBD sites orientation, etc.), its behaviour in a liquid environment may change. Furthermore, data is available only for those compounds that can be crystallized, and not all peptides or proteins can.

The CSD software consists of a number of suites and products with specific roles. The CSD-Crossminer was mainly used as it allows to match the molecules in the database with potential target. All the others were used to scan through the database and visualize search results.

The CSD-Crossminer operates in the following way. To start the search for potential anion receptor, the "reference" molecule or ion should be uploaded; in our case it was acetate ion. Then the desired features of a target molecule must be added on the "reference molecule" (it is later explained **in section 5.1.** and shown on Figure 4). These features might be of a different kind: an acceptor, a donor, a particular AA or a hydrophobe. They are functional groups within the database structures, which will show the aforementioned properties to our "reference molecule"[35].

After adding relevant features, the software conducts the search in the databases. There are three databases available at this moment, the largest of which is "csd543_crossminer". However, we used mostly the other database – "pdb_crossminer". It contains data from the Protein Data Bank[36]. These structures

are relatively large proteins, but the software allows screening individual (smaller peptide chain) binding pockets for suitable peptide fragment geometry.

After the search is finished, all the structures with mentioned features are listed. It is possible to view one molecule after another, as well as to see the functional groups which interact with the “reference molecule”. This helps select the promising candidates in the earliest stages.

Considering the advantages of CSD, it proved to be highly useful in our research. Several promising peptide structures were found and investigated, which is described below. However, CSD software was found to have one weakness. The number of results is limited to 10,000 structures, and when the number of found molecules hits this amount, there is no possibility to screen through the rest of the database. This limited the amount of possible candidate molecules. Nevertheless, this disadvantage was overcome by adjusting the search parameters.

4.2. COSMO-RS Method

Nowadays computational chemistry plays an important role in research. Different types of software find their use in various areas from chemical engineering to drug screening and design. Not only do these tools advance research by solving complex chemical problems, but also reduce the cost and time of the experiment, therefore improving the performance[37].

The modelling is especially well-established and easy to conduct for the gas phase. By contrast, the theoretical description of a liquid phase is highly complicated, making the creation of its theoretical models and computations a difficult task to solve. Because the molecules in the solution are constantly interacting with each other, there is a huge number of factors to be taken into consideration. From the thermodynamic and quantum mechanical point of view, compound properties in liquids should be calculated in large ensembles of interacting molecules. This is difficult and resource-intensive, therefore the majority of methods either strongly simplify the situation or require considerable computing power[38].

Our research requires the knowledge of molecular behaviour in the solution, and the computations are required to predict the steric effects and matching between receptor cavities and the anion of interest. To investigate the suitability of the found structures for our application, the Conductor-like Screening Model for Realistic Solvation (COSMO-RS) method[39] was used.

COSMO-RS as a computational method for the liquid phase has several advantages which made us choose it.

- Firstly, it enables calculations in multicomponent mixtures composed of any solvents and at high concentrations of the solute molecules or ions. This is in contrast to other models, like solvation model based on charge density (SMD)[40], which are only capable of computing interactions in pure solvents with known properties. Such methods cannot account for selective solvation, interactions between solute’s molecules or solvation in presence of low water content[41].
- Secondly, COSMO-RS computations are robust, reproducible, less demanding, and the overall accuracy is reasonably high when comparing to the other solvation models[42–44]. For instance, the accuracy of transfer free

energies of neutral compounds improved from ~1.7...1.3 kJ mol^{-1} [38] to 0.55 kJ mol^{-1} on average[45].

- Thirdly, COSMO-RS has shown a satisfactory performance while modelling HB formation within complexes[41].
- Lastly, it is simple to use and powerful in terms of graphics[38].

COSMO-RS was developed as a modification of a dielectric continuum salvation model (CSM). Whereas the dielectric CSMs are based on the approximation of solvent as a dielectric continuum of permittivity ε, COSMO-RS replaces it with a conductor of infinite ε.

Density functional theory (DFT) calculations are used to obtain the required properties of a compound in the conductor[38]. This calculation accounts for molecular structure effects (electric, steric) and intramolecular hydrogen bonds. The DFT calculations are followed by statistical thermodynamics part that takes the van der Waals forces, electrostatic interactions and hydrogen bonds between all the species in the solution into account. The concentrations of different molecular species (the solvent and the solute, for example) are being considered, but no distinction is made between solvent and solute molecules: the method treats them equally.

As a result of COSMO-RS computations, the thermodynamic characteristics, most importantly, molar Gibbs free energy value *G* of every species in the solution is calculated. *G* values of the species enable calculating the Δ*G* of the binding reaction and from Δ*G* the equilibrium constant of the binding reaction can be evaluated[41].

Despite several major advantages, COSMO-RS has its downsides.

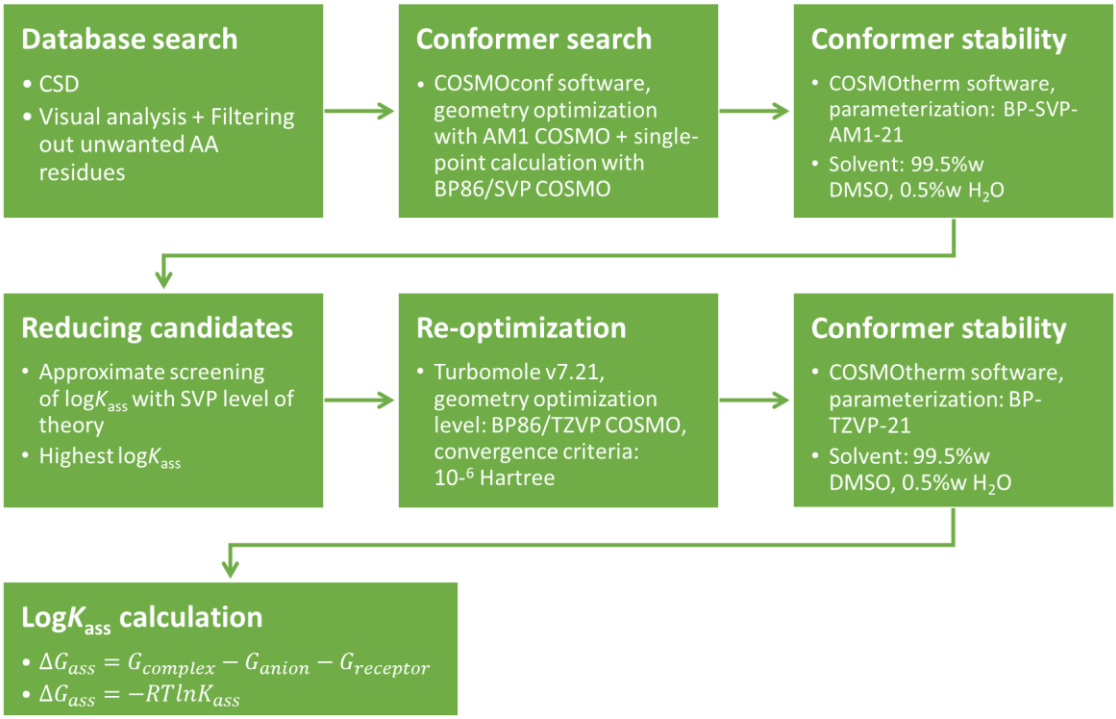
- First of all, the parameterization is extensive, while the consideration of intermolecular interactions is too simplified.
- Secondly, the method does not fully account for some of the long-range interactions between molecules, which may be important for our study.
- Additionally, COSMO-RS tends to strongly overestimate binding affinities in the modelled systems[46].
- If the lowest energy conformers were not found initially by quantum chemistry calculations, the results of COSMO-RS predictions may not be adequate.
- Lastly, while the dominant conformers are generally predicted correctly, the numerical values for conformer abundances may notably differ between experimental data and computational predictions[47].

5. WORKFLOW AND USED PARAMETERS

To meet the purpose of our research, several steps had to be done. As the idea of using peptide as the basis for designing an ionophore for carboxylate sensing is relatively new, the peptides with anion binding capability and agreement with aforementioned target properties have to be found and investigated from the beginning.

Firstly, the search in CSD with specific parameters according to the desired properties was conducted. Next, visual and more detailed analysis of found structures was performed in order to identify the best possible candidates. This was followed by computational estimations of acetate binding affinity in solvent, which give a good base for decision of further research or synthesis of particular molecule.

Scheme 2. General workflow of the work.



5.1. Database search for peptide candidates

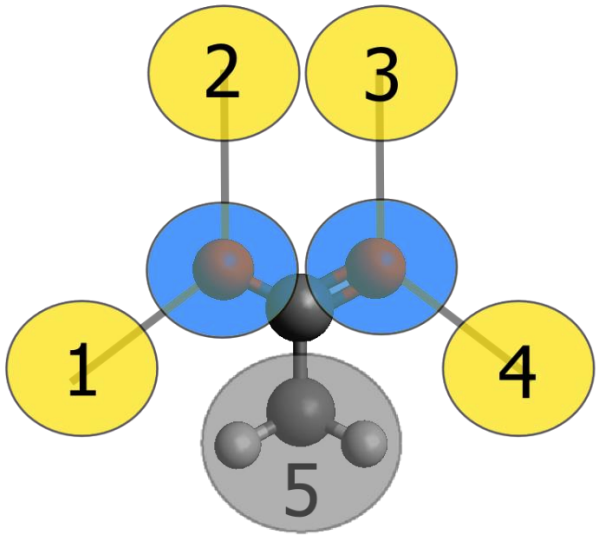
The reference ion to start the search was acetate. The desired feature added was “donor_projected”. It allows to search for HBD functional groups, which are NH or OH groups in our case. All features, including this one, look like a sphere in a software. The “donor_projected” feature (yellow spheres on Figure 4) means that the HBD group in peptide molecule is located in the area of this sphere.

According to earlier studies, in order to bind anion effectively and strongly, there should be at least three hydrogen-providing groups[13]. Therefore, three or four of those spheres were added to the acetate anion as features. The focus of the search was both NH and OH groups, and the software showed all the results containing these functional groups. The radii of the aforementioned spheres varied from query to query because it was observed that the higher the donor projected radius, the bigger the distance between HB groups within the peptide. This leads to widening of the cavity and, therefore, lower possibility of selective peptide-anion interaction.

Consequently, the radii values were chosen to be less or equal than one to yield more suitable results.

For better understanding of the software's operation mode, Figure 4 illustrates the acetate anion with added features. Yellow spheres represent the positions of HB-donor groups of the potential binder, whereas blue ones indicate the location of the HB acceptor groups inside the anion of interest. The numbers indicate the position and characteristics of the features for the Table 1.

Figure 4. Scheme of the database search parameters based on the acetate anion. Yellow spheres (radii between 0.8 and 1.5 Å) – possible positions of potential binding groups (NH and/or OH); blue spheres (radii between 0.8 and 1.2 Å) – position of carboxylic group; grey sphere – position of excluded volume feature. The numbers are used to characterize spheres and shown in Table 1.



Moreover, the software allows to choose between three types of the target molecules: a peptide, a small molecule, or either. This option is applied to the "donor_projected" sphere directly (Positions 1–4 on Figure 4). As the main interest for this search are peptides, the option "peptide" was applied.

To make the search results more suitable for our purposes, the "excluded volume" (Position 5 on Figure 4) feature was added. The idea behind this was to provide the space for anion and, consequently, find a cavity in the peptide of interest. The radius for "excluded volume" remained mainly the same from query to query, as shown in Table 1, because no significant difference in the results appeared when it was modified.

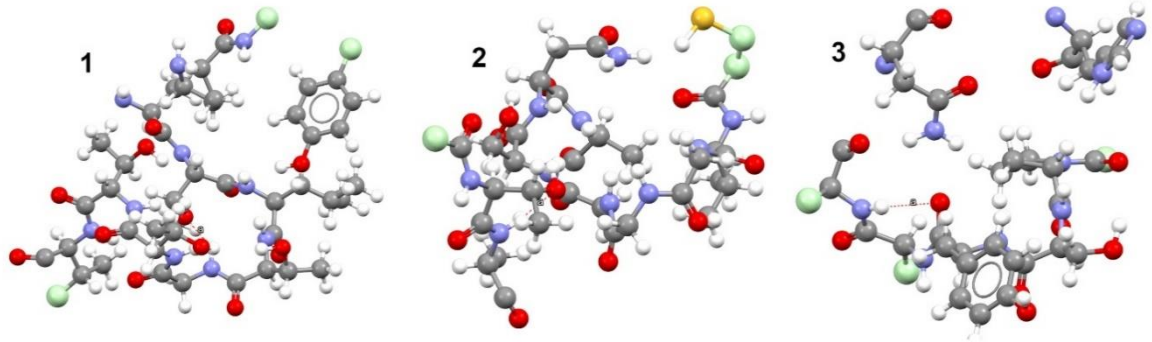
In total there were 17 search queries conducted. The details on some of them are shown in the Table 1. These 4 queries are chosen to indicate the strategy of the search: the location and number of the feature spheres, how their radii were changed, and presence/absence of "excluded" volume.

Table 1. Query parameters for database search.

Query #	Donor-projected groups			Excluded volume	
	Position on Figure 4	R_{donor} (yellow), Å	$R_{acceptor}$ (blue), Å	Position on Figure 4	R, Å
1	1	1	1	-	-
	2	1	1		
	3	1	1		
2	1	1.3	1	5	1.70
	2	1.3	1		
	4	1.3	1		
3	1	0.8	1	5	1.70
	2	0.8	1		
	3	0.8	1		
	4	0.8	1		
4	1	0.9	0.9	5	1.70
	2	0.9	0.9		
	3	0.9	0.9	Covering COO^- group	1.40
	4	0.9	0.9		

Different search queries led to a different number of results. The majority of the molecules found had three or more NH groups, which were close enough to each other to provide a cavity and, therefore, could potentially allow the acetate to bind. Two of the most appealing structures as well as an unsuitable molecule for comparison are illustrated in Figure 5.

Figure 5. Several molecular structures found using CSD-Crossminer. The structures 1 and 2 have a cavity for the anion, the structure 3 lacks it. Colour is used to denote the atoms: red is for oxygen, blue is for nitrogen, yellow is for sulphur, and the green atoms represent the connecting points to the parts of bigger protein.



Apart from peptide-based receptor, the CSD-Crossminer was used later to search for hydrophobic linker. As the hydrophobic interactions can occur with the CH_3 group of acetate, the “hydrophobe” feature size and distance were chosen accordingly. The course of actions was similar to the one described above with some slight modifications: (1) the “hydrophobe” feature was chosen instead to exclude any other type of interactions with the anion, (2) the radii of feature’s spheres were large to allow for broader results (1.2 Å for “hydrophobe” and 2 Å for “excluded volume“), and (3) both types of target molecules were chosen (peptide and small molecule).

Overall, 3 search queries were conducted, and over 5,000 structure fragments were found. In general, those were AAs (up to 3) supposedly from a bigger peptide structure. Screening through obtained results was performed, and the AA sequences found to have better sterical suitability to interact with the acetate anion AA combinations (Tyr-Phe; Gly-Val-Thr-Leu; Ile-Met; and Tyr-Leu) were investigated on their suitability as a linking agent. This is described in **section 6.4** in more details.

5.2. Refining the search results

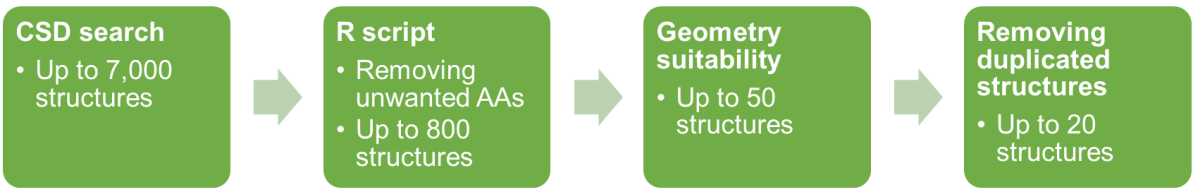
The number of found structures for search queries 1–2, as well as others with similar parameters, was up to 7,000 for each, whereas for queries 3–4 the number reached 3,000. Considering such large number of peptide structures obtained within the search, it was necessary to reduce that number by filtering out less suitable structures before more detailed conformer study. Several selection and filtering criteria were defined:

- Acidic and basic AA residues should be omitted, because their (de)protonation makes the whole peptide more hydrophilic and can cause leaching into water when embedded into a lipophilic receptor membrane.
- The cavity for anion must be visible. Additionally, the number of hydrogen donor groups must be sufficient (3 or 4), and they must be closely located.
- Preferably, AA residues forming the appropriate cavity should be consecutive. It means that AAs with all relevant HB donor sites should form a single peptide chain.

To find peptide molecules that comply with the first requirement, R script was used. It screened through all found peptide structures and removed those with unwanted AA residues (Arg, Asp, Lys, Glu, and His). As a result, the number of structures for queries 1–2 reduced to around 800 and for queries 3–4 to around 20 each. The general flow of reduction of number of structures is represented on Scheme 3.

The occurrence of cavity and suitable orientation of NH and OH groups were checked visually. Some of the structures appeared to be identical, so the duplicates were removed. In addition, the number of both intra- and intermolecular hydrogen bonds and their lengths were approximately estimated. It appeared to be that one particular AA to interact with acetate was similar for each promising peptide structure: Thr.

Scheme 3. Refining the number of peptide candidates.



5.3. Computations in solvent

The software used to implement COSMO-RS method are as follows: *Turbomole*[48] (DFT calculations), *COSMOconf*[49] (conformer search) and *COSMOtherm*[50] (computing energies in liquids and properties dependent on them).

The COSMOconf application was implemented for finding the stable conformations for both free peptide and peptide-acetate complex. One of the pre-programmed conformer search algorithms was used with slight modifications. It included initial conformer generation and step-by-step reduction using geometry optimization with AM1-COSMO method and single-point calculations in ideal conductor with BP86/SVP method. The maximum number of conformers on conformer reduction steps was increased to obtain more results. This algorithm was chosen based on affordable computation time compared with higher-level methods.

The relative stability of conformers in a solvent was computed using COSMOtherm in DMSO with 0.5 mass% water, assuming infinite dilution, with parametrization BP_SVP_AM1_COSMO. This solvent mixture was chosen for comparability of the results with both computational and experimental data available on other receptor candidates. Small amount of water is always present in typical organic solvents, so adding a defined amount of water makes a solvent mixture more reproducible and is a standard technique for studying receptor-anion interactions.

Based on the results of the initial investigation of conformer stability in DMSO/0.5% H_2O, the conformers with predicted 10% or more abundance were taken for the reoptimization at the TZVP level of theory. The used criteria were as follows: the wave function convergence criteria: max difference 10^{-6} Hartree and geometry convergence: max gradient |dE/dxyz| 10^{-3} Hartree/Bohr.

After the geometry optimization at the BP86/TZVP level of theory was done, the stability of conformers was again investigated using COSMOtherm in the same solvent mixture as above. The goal was to obtain the best geometries of peptide molecules to fit acetate, the preferred geometries of free peptide in solution, and compute the estimates of binding affinities from free energies of the found structures in a chosen solvent mixture.

The computed free energies for acetate, free peptide and the host-guest complex in solution were used to estimate the binding (association) affinities of found receptor molecules and compare them. The change in free energy and binding constants can be calculated from Equations (3), (4) and (5). The temperature used is 298.15 K.

$$\Delta G_{ass} = G_{complex} - G_{free\ peptide} - G_{anion} \quad (3)$$

$$\Delta G_{ass} = -RTlnK_{ass} \quad (4)$$

$$logK_{ass} = lnK_{ass}/2.303 \quad (5)$$

The evaluation of binding constants for studied peptide receptor candidates was made in comparison with macrocyclic receptor MC01 from reference 14. The reasons for this were the highest binding affinity logK_{ass} with different carboxylates among all the other receptors and its usage as ionophore in sensor prototype. The structure of this receptor is shown in Figure 1.

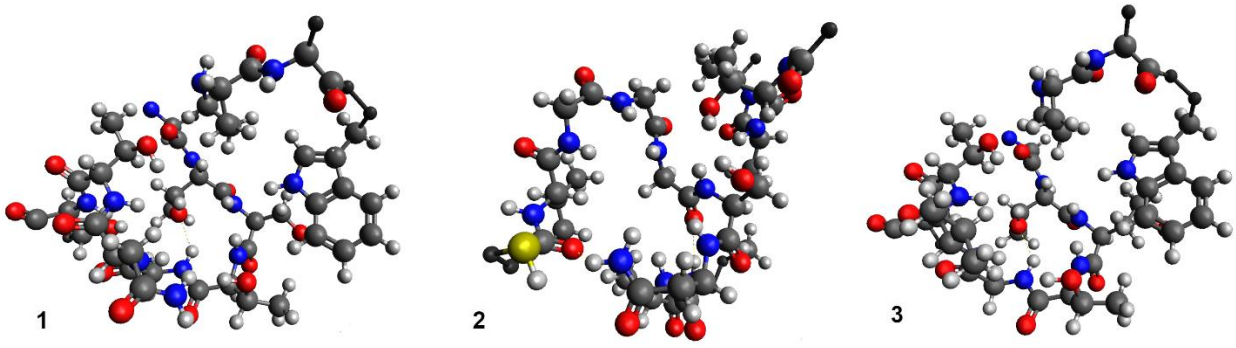
6. RESULTS AND DISCUSSION

6.1. The promising peptide structure

As described in **section 5.1**, 17 database searches were run in total. They differed mainly by the "donor_projected" feature: its radius and location. At the very beginning, more than 5,000 structures were found. Later, when the search criteria were optimized (Queries 3–4 in Table 1), the number of peptide candidates decreased. A more significant reduction (from 800 to 20 structures depending on the number of structures from initial database search) was obtained after application of the R script. It was used for every search query performed.

After visual analysis 15 peptide fragments were identified to be notably promising. Some examples are presented in Figure 6. It was decided to examine one of these structures more thoroughly based on these criteria: (1) it has 4 suitably positioned HBD sites, (2) there is a cavity, and (3) the AA sequence is quite short. The goal was to achieve good binding affinity towards the acetate anion.

Figure 6. Some examples of promising peptide structures.



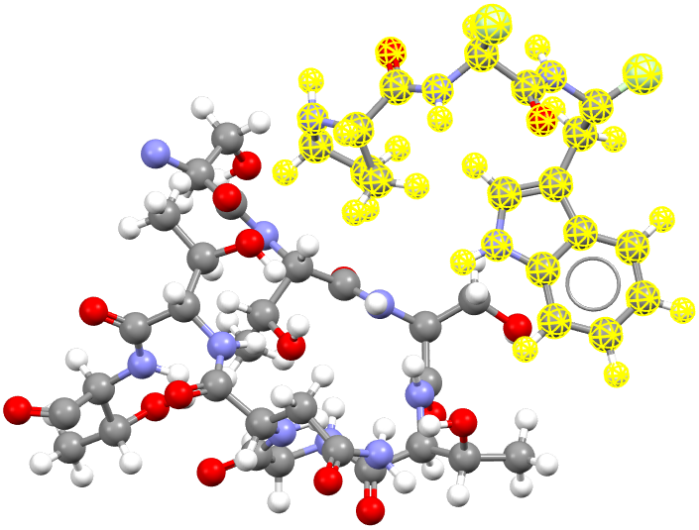
The full protein structure of the CCDC peptide sequence database entry was obtained from RCSB Protein Data Bank (RCSB PDB)[36]. It is an oxidized flavodoxin from a red alga *Chondrus crispus*[51] containing 173 AA in total. The complete AA sequence in the protein is presented in Table 2.

Table 2. The initial protein AA sequence.

KIGIFF**STSTGNTT**EVADFIGKTLGAKADAPIDVDDVTDPQALKDYDLLFLGAPTWNT GADTWNTGADTERSGTSWDEFLYDKLPEVDMKDLPVAIFGLGDAEGYPDNFCDAI EEIHDCFAKQGAKPVGFSNPDDYDYEESKSVRDGKFLGLPLDMVNDQIPMEKRVA GWVEAVVSETGV

The fragment, further referred to as RM1, that was used for the further investigation as an anion receptor was found in CSD. The view of RM1 in the CSD is illustrated in Figure 7. As can be seen, RM1 consists of two amino acid chain fragments. The sequence of longer peptide chain is as follows: Ser-Thr-Ser-Thr-Gly-Asn-Thr-Thr. The other chain, which is highlighted in yellow, consists of Pro, Thr, and Trp. Interestingly, these two chains are not directly connected, but are adjacent chain fragments of a bigger protein.

Figure 7. Image on a database entry that is the basis of the studied candidate RM1 (Ser-Thr-Ser-Thr-Gly-Asn-Thr-Thr). The adjacent chain (Pro-Thr-Trp) in the proten structure denoted by yellow is not directly connected to RM1 and was not used in optimization.



Only the longer chain was used for further investigations because it appears to be more promising due to having several NH groups oriented in the favourable direction to bind with the anion. It is worth noting that Pro-Thr-Trp-chain might play a part in supporting suitable peptide orientation and cavity-like structure. However, it was decided to omit this chain in further investigations because of several reasons.

Firstly, we cannot consider it to be solely responsible for the RM1 geometry. The whole protein structure is essential for guaranteeing the initial conformation of the smaller fragment. In our case it consisted of 60+ AAs, meaning that any part of the chain could influence the geometry of RM1. Secondly, two chains of RM1 are not consecutive, which does not fulfil the requirements: the distance between them is more than 40 AAs. Thus, trying to involve the second chain linked to the first one would make the whole peptide very large and its computational investigation complicated. Finally, the experimental data for RM1 was recorded in solid state, thus its behaviour and chain orientations are likely to be different in a solution.

To ease further reading, the abbreviation RM1 will be still used to signify the peptide fragment found, but Pro-Thr-Trp-chain is excluded from now on.

6.2. Conformational search results

The main aim for the conformational analysis was to determine the most stable conformers in the used solvent – DMSO with 0.5% (m/m) water – and predict the behaviour of RM1 in that solvent. The calculations were performed for both free peptide and peptide complex with anion using the COSMO-RS method and parameters described in **section 5.3**.

First of all, it is important to define the meaning of stable conformer in our research. Stable conformer is defined as a conformer with the energy in conductor no more than 6 kcal/mol higher than the lowest energy conformer of a particular compound[39]. Stable conformers are selected in the last of the three stages of conformer reduction in COSMOconf.

The most abundant conformer is defined in this work as the conformer with the highest statistical weight in the used solvent according to the conformer weight analysis in COSMOtherm. The statistical weight (abundance) of a particular conformer is determined according to its energy in the solvent and is calculated according to the Boltzmann distribution[39,47].

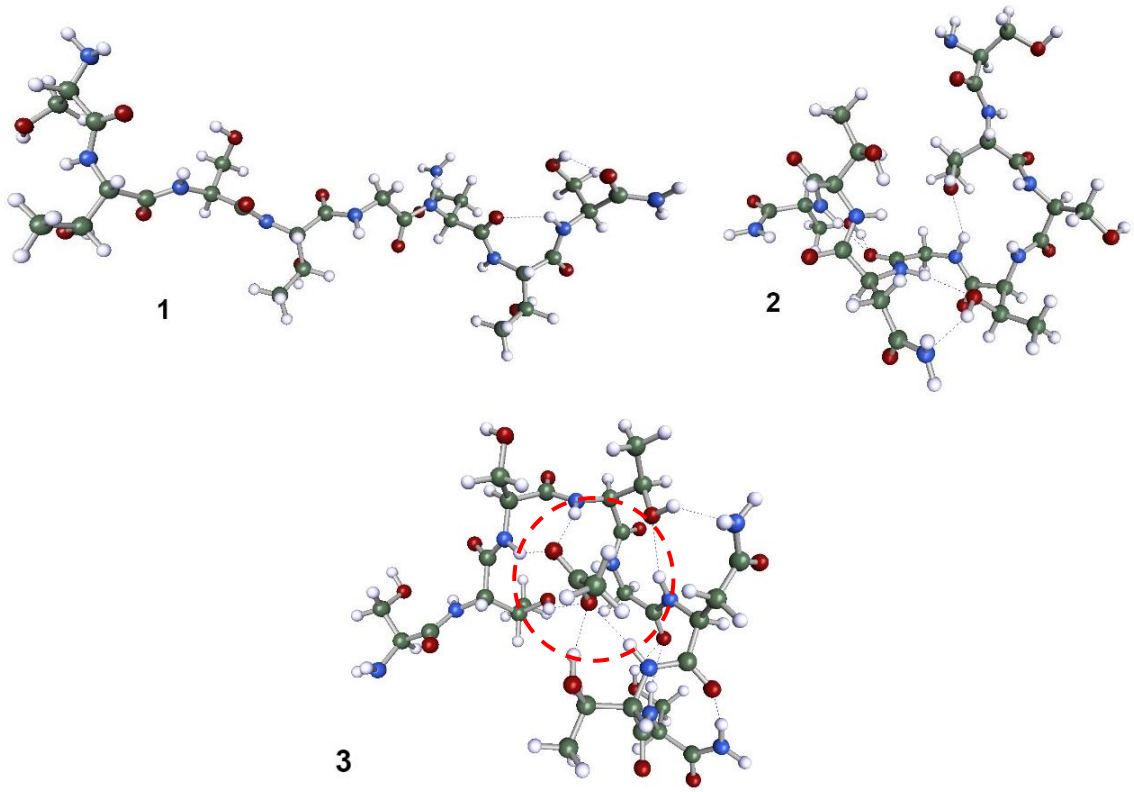
According to the COSMOconf calculations, the free peptide RM1 has 8 stable conformers. However, it tends to unfold in solution when not bound with acetate. That means that in the unbound RM1 there is no cavity and the potential binding groups are not located close to each other. Thus, significant conformational change is expected upon binding. Indeed, for the acetate complex only one stable geometry was found, with peptide chain folded around the anion (Figure 8). It is reasonable as different conformations of free peptide are likely to close up around acetate in a similar manner.

According to these findings, the decision was made to perform further investigations with a cyclic structure of the free peptide. The main reason for this is illustrated on Figure 8. Two geometries of RM1 could be seen there. The first one stands for the most stable free peptide conformer, according to the COSMOconf calculations. The second one represents the desired geometry of RM1. The difference between them is a pre-organized binding pocket that is not present in the most stable conformation.

However, the fact that RM1 is “unfolded” in unbound state does not preclude it as a potential anion binder. In fact, COSMO-RS predicts the logK_{ass} value to be 11.2, which is considered high and promising. For comparison, the computationally predicted logK_{ass} values of acetate with some macrocyclic receptors that have already been used in acetate sensors lie in the range of 14 – 20[14].

RM1 can fold around anion, and this strategy proved to be quite successful for several other receptors (the structure on Figure 2 could be used as an example). It is worth mentioning that the most stable conformer of Receptor 17 from reference 15 does not have a cavity either and, yet, has been more recently successfully used in carboxylate sensing[52]. Nevertheless, such conformational change, in addition to complexation, may require additional energy and is also associated with entropy penalty. Consequently, making RM1 a cyclic structure with suitable size and orientation of the HBD groups would facilitate its binding with anion.

Figure 8. Free geometries of RM1 structure. 1 – most stable conformer 2 – desirable conformation of the structure in complex, 3 – acetate complex (anion is in red dashed circle).



6.3. Peptide structure optimization: alkyl chain linker

To make a cyclic structure out of linear RM1, some kind of linker must be put between the ends of the peptide. The main purpose of the linker is to fix RM1 in a conformation where the HBD sites used for receptor-anion complex formation are facing in the same direction toward the cavity. The chemistry of the linker should be simple enough to synthesize. Additionally, it is desirable that the linker is hydrophobic in order to make the peptide as hydrophobic as possible. The more extensive the hydrophobic parts are in the receptor molecule, the better it will be retained in the sensor's hydrophobic membrane (as opposed to leaching into the measured solution).

Alkyl chain linkers are the simplest option. Their chemical structure is simple, their length can be altered, which makes them widely used[53]. For that reason, their applicability as a linker for RM1 was investigated. It was decided to study alkyl chains of 1 to 7 CH_2 groups and compare the binding affinities of these structures with the linear RM1 and a previously studied macrocycle molecule. The examples of the structures created are shown in Figure 9.

First of all, different methylene linkers were added to the initial RM1 structure. Then, the conformer search in COSMOconf was performed, followed by the COSMOtherm conformer stability computation. The most stable conformers in the solvent mixture were re-optimised with TZVP basis set for higher accuracy. Finally, their energies and, consequently, binding affinities were calculated (Equation (4)).

Detailed structural information and logK_{ass} values are presented in Table 3. To understand better how the number of both AAs in the main peptide chain and alkyl groups in the linker chain influences the binding strength, different modifications of RM1 were created. Either one or both terminal AAs from the initial chain were removed and the alkyl linker was added. The number in the name (RMC) denotes the number of CH_2 groups in the linker chain.

Figure 9. Peptide with alkyl chains as a linker. 1 – RMC1 (methylene group is captured in the red box), 2 – RMC3, 3 – RMC5, 4 – RMC7.

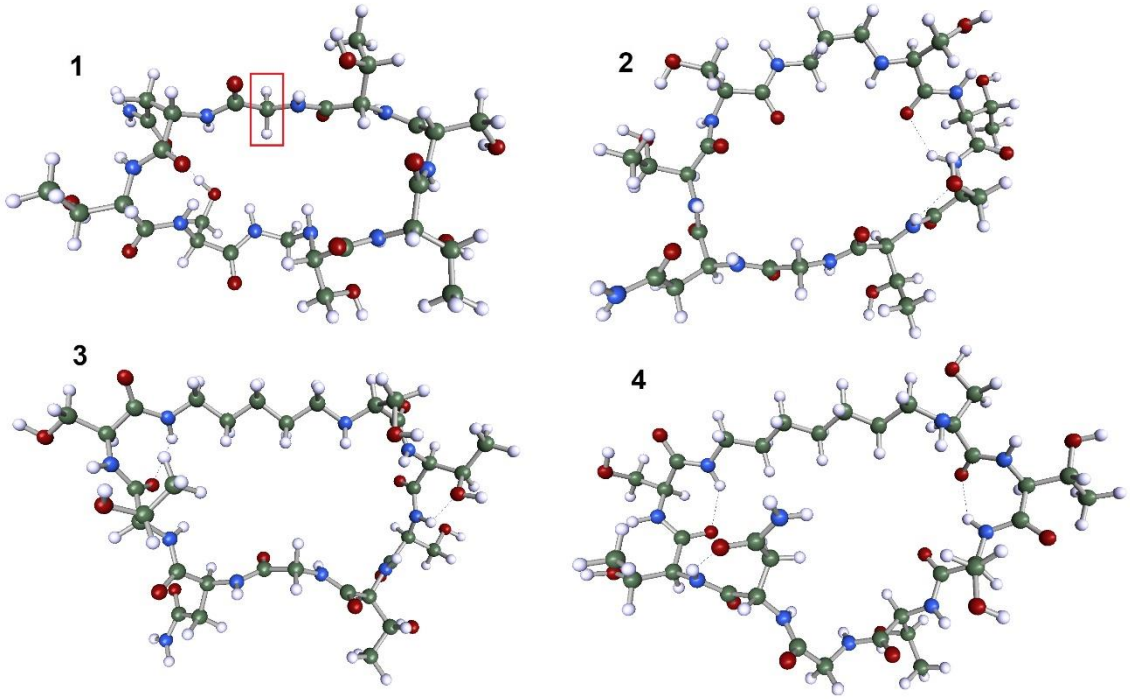
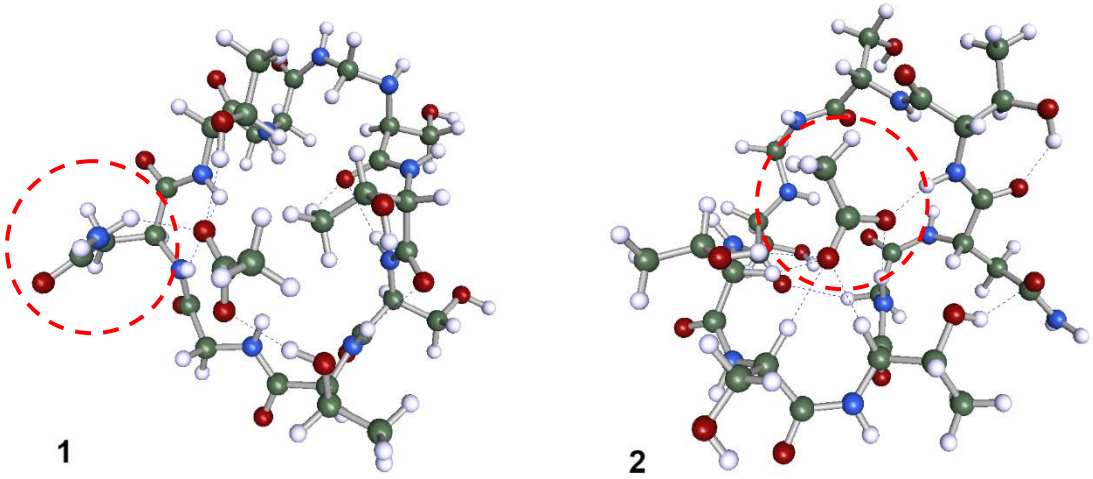


Table 3. Structural information and binding affinity values of cyclic modifications of RM1 with alkyl chain linker. logK_{ass} values computationally predicted through COSMOtherm calculations. Solvent: 99.5% DMSO with 0.5% H_2O, parameterization: BP_TZVP_21.

Name	**logK_{ass}**	**Sequence**								
RM1	11.2	Ser	Thr	Ser	Thr	Gly	Asn	Thr	Thr	-
RMC1	1.6	Ser	Thr	Ser	Thr	Gly	Asn	Thr	Ser	$NHCH_2$
RMC2	7.9	-	Thr	Ser	Thr	Gly	Asn	Thr	Ser	$NH(CH_2)_2$
RMC3	7.1	Ser	Thr	Ser	Thr	Gly	Asn	Thr	$NH(CH_2)_3$	-
RMC4	7.8	-	Thr	Ser	Thr	Gly	Asn	Thr	$NH(CH_2)_4$	-
RMC5	2.9	Ser	Thr	Ser	Thr	Gly	Asn	Thr	Ser	$NH(CH_2)_5$
RMC6	4.6	Ser	Thr	Ser	Thr	Gly	Asn	Thr	Ser	$NH(CH_2)_6$
RMC7	7.1	Ser	Thr	Ser	Thr	Gly	Asn	Thr	Ser	$NH(CH_2)_7$

The initial conformer search via COSMOconf usually yielded 8 – 10 conformers for free peptide fragments, whereas the number of conformers for acetate complex varied from one structure to another. Generally, it was one or two conformers. It is interesting to note that although the conformations where acetate binds to RMC's outside the cavity do exist, their abundance in the solvent is low. For instance, two conformations of RMC1 complex with acetate are shown in Figure 10. According to COSMOtherm calculations, the weights of conformer 1 and 2 in the Figure 10 are 0.22 and 0.53 respectively. This means that the binding of anion will be primarily achieved in the cavity.

Figure 10. Two different conformers of acetate complex of RMC1. 1 – acetate (in red dashed circle) is bound outside of the cavity, 2 – acetate is bound within the cavity.

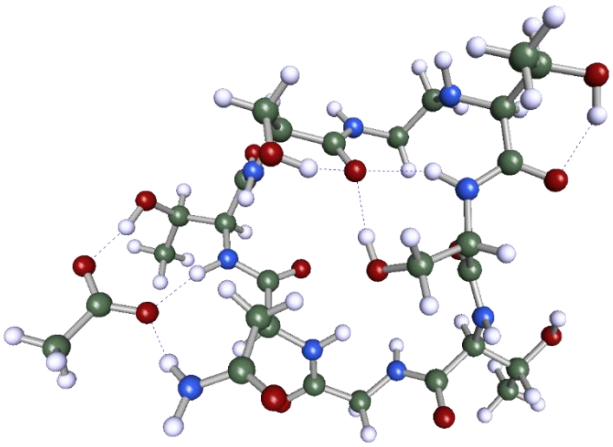


Computationally obtained binding constant values for aforementioned peptide structures are presented in Table 3. It can be seen that all the logK_{ass} values for modified peptide structures are below the logK_{ass} value for the initial RM1. The lowest values are predicted for RMC1 and RMC5, which could be caused, on the one hand, by unsuitable HBD sites orientation or, on the other hand, by initial miscalculations of conformer stability.

Although binding constants are relatively high for RMC2 and RMC7, it does not necessarily mean their suitability as receptors. For example, COSMO-RS calculations predicted acetate to bind with RMC2 outside the cavity (Figure 11), whereas the binding pocket of RMC7 is too big (Figure 9, 4). This finding raises additional doubt on the selectivity of these structures.

Considering the lack of improvement in binding from alkyl linkers, further modifications of RM1 structure were planned. The new peptide structure must remain cyclic, and the linker of our choice in this regard was some AA residue. Detailed description is below.

Figure 11. The complex of RMC2 with acetate. RMC2 is stabilized by the presence multiple IMHBs.



6.4. Peptide structure optimization: amino acid chain linker

The concept behind using an AA chain as a potential linker is similar to the alkyl chain linker concept. The AA linker must be hydrophobic and hold the peptide structure in a favourable configuration for anion binding. Moreover, as AAs are bulkier than methylene groups, they should not cover the binding moiety; on the contrary, it would be preferable if a cavity could be formed due to repulsion between the closely-situated hydrophobic parts (for example, between benzene rings in phenylalanine).

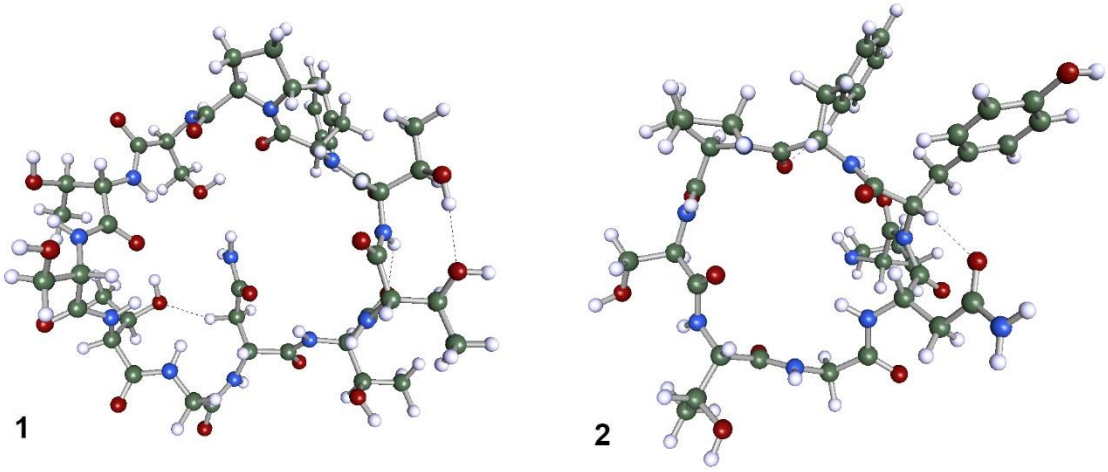
The modification of RM1 structure was done in several stages. Firstly, the linker connected through proline fragments was added to the original peptide. Proline was used as a rigid AA[54] to hold the cavity in place. Then, one AA residue was removed from both ends of RM1, and Pro-connected linker was added. The same procedure was repeated, but two AA residues were removed. In all cases the linker length was two or four AA residues. Investigated linkers were as follows: Tyr-Phe; Gly-Val-Thr-Leu; Ile-Met; and Tyr-Leu. Some examples of final peptides are shown in Table 4. The examples of all modifications of RM1 are presented in the Appendix 1.

Table 4. Optimization of initial RM1 structure with proline-connected linker.

Name	Sequence												Number of HBs in free peptide (up to 3 Å)	Number of HBs in peptide-acetate complex (up to 3 Å)	
														IMHBs	With acetate
RM1	-	-	-	-	Ser	Thr	Ser	Thr	Gly	Asn	Thr	Thr	3	5	6
RMP1	Pro	Tyr	Phe	Pro	Ser	Thr	Ser	Thr	Gly	Asn	Thr	Thr	2	Cavity too large for acetate to be able to approach multiple binding groups for binding	
RMP2	-	Pro	Tyr	Phe	Pro	Thr	Ser	Thr	Gly	Asn	Thr	-	1	4	3
RMP3	Pro	Gly	Val	The	Leu	Pro	Ser	Thr	Gly	Asn	Thr	Thr	1	Cavity too large for acetate to be able to approach multiple binding groups for binding	
RMP4	-	Pro	Ile	Met	Pro	Thr	Ser	Thr	Gly	Asn	Thr	-	2	4	4

The investigation of Pro-connected hydrophobic AA linker revealed that it is not a suitable linking agent for this peptide. All of RMP structures featured one of the following phenomena: the cavity was too large to fit the acetate well (Figure 12, 1), or there were not enough HBD groups facing the same direction for the acetate to approach (Figure 12, 2). For this reason, TZVP geometry optimization and logK_{ass} calculation were omitted.

Figure 12. Geometry of RMP. Examples of receptor cavity size and orientation for some of the candidates. 1 – RMP1, 2 – another peptide receptor candidate (Peptide3 in Appendix).



Next modification was done with the same linkers, but without Pro. The procedure was done similarly, some examples of peptide sequences are presented in Table 5.

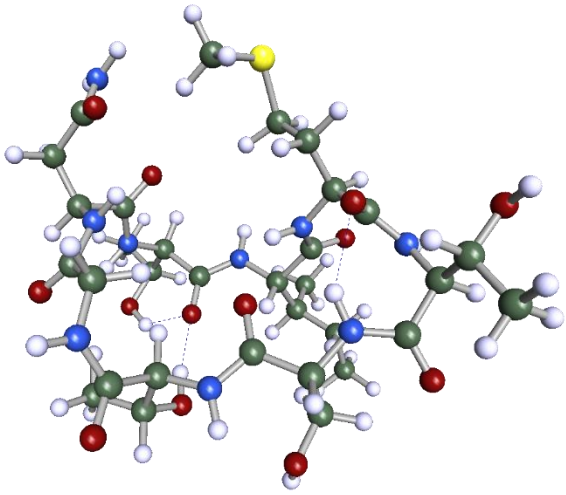
Table 5. Further examples of AA sequences and structural information on hydrogen bonding for receptor candidates based on RM1 structure.

Name	Sequence										Number of HBs in free peptide (up to 3 Å)	Number of HBs in peptide-acetate complex (up to 3 Å)	
												IMHBs	With acetate
RM1	-	-	Ser	Thr	Ser	Thr	Gly	Asn	Thr	Thr	3	5	6
RMA1	-	Tyr	Phe	Thr	Ser	Thr	Gly	Asn	Thr	-	3	6	3
RMA2	Gly	Val	Thr	Leu	Ser	Thr	Gly	Asn	Thr	-	3	4	4
RMA3	-	Ile	Met	Thr	Ser	Thr	Gly	Asn	Thr	-	4	0	5
RMA4	Ile	Met	Ser	Thr	Ser	Thr	Gly	Asn	Thr	-	1	0	4
RMA5	-	Ile	Met	Thr	Ser	Thr	Gly	Asn	-	-	4	1	6

Although some of RMA receptors had up to six HB interactions with acetate, the same issues as with RMP binding agents were observed. On the one hand, when receptor consisted of the initial AA chain and an additional linker chain, the cavity size was too large for acetate to fit close to multiple HBD groups. On the other hand, if the number of AA was similar to RM1 (as in RMA3), the cavity was too small. It was caused by IMHB interactions within free peptide and bulky AA side groups, which prevented access to the moiety (Figure 13).

For other RMA-type candidates, the notable conformation change has to take place in order to bind acetate. Moreover, the breaking of IMHBs is needed to form the receptor-acetate complex, and in some cases the acetate is not located directly in the cavity. All this is associated with large energy penalty and makes these receptors significantly less suitable for anion sensing. This was also supported by the logK_{ass} calculation at SVP level. The TZVP optimization and calculation of binding affinity was done only for a subset of RMA structures (the best of which is presented in Table 7) with the highest logK_{ass} values on SVP level.

Figure 13. Geometry of RMA3.



It is worth noting that logK_{ass} value for RMA5 was high in comparison to other candidates, and acetate complex had optimal geometry (Figure 14). However, the free peptide had 4 IMHBs between NH and OH groups in the structure (two of them were with O atom in the Asn). This is why it was decided to optimize its geometry by replacing Asn with Ile, and Met and Thr with Phe to make it more hydrophobic and reduce the amount of IHMB interactions (RMR5, RMR6 and RMR7).

The last modifications that took place are related to replacement of AAs in the initial RM1 structure. Hydrophobic AA chain as a linker was also added. Details are provided in Table 6.

Figure 14. Geometries of free candidate RMA5 (1) and its complex with acetate (2).

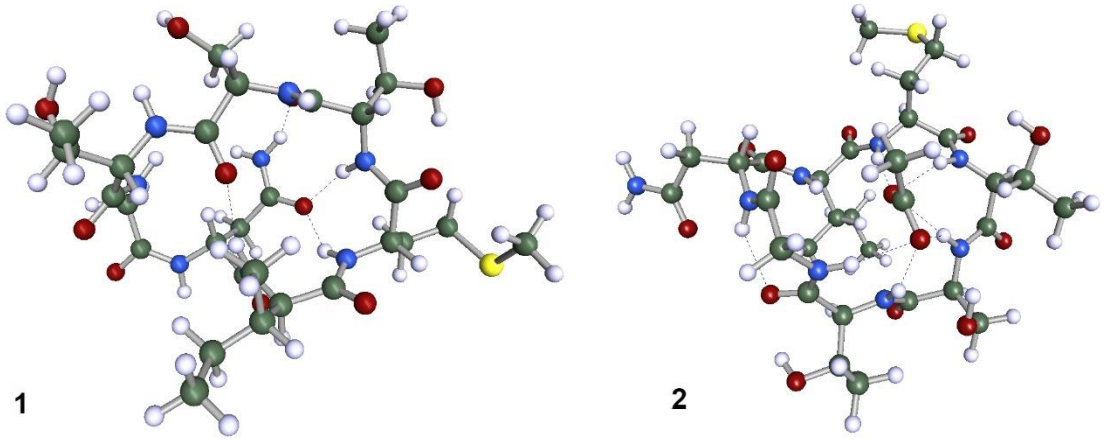


Table 6. Examples of the final optimization of initial RM1 structure.

Name	Sequence											Number of HBs in free peptide (up to 3 Å)	Number of HBs in peptide-acetate complex (up to 3 Å)	
													IMHBs	With acetate
RM1	-	-	-	Ser	Thr	Ser	Thr	Gly	Asn	Thr	Thr	3	5	6
RMR1	-	-	-	Tyr	Phe	Ser	Thr	Gly	Asn	-	-	3	3	4
RMR2	-	-	Tyr	Leu	Thr	Ser	Thr	Gly	Gly	-	-	4	3	5
RMR3	Gly	Val	Thr	Leu	Gly	Ser	Thr	Gly	Asn	Thr	-	2	2	3
RMR4	-	Gly	Val	Thr	Leu	Gly	Gly	Gly	Gly	-	-	4	1	4
RMR5	-	-	Ile	Phe	Thr	Ser	Thr	Gly	Ile	-	-	3	0	4
RMR6	-	-	Ile	Phe	Phe	Ser	Thr	Gly	Ile	-	-	3	1	5
RMR7	-	-	Ile	Phe	Phe	Ser	Phe	Gly	Ile	-	-	0	1	4

The similarity between these structures is sufficient number of HBs with acetate (up to 6). However, the number of IMHBs in free peptides is approximately the same, meaning that they should break upon formation of anion complex. Additionally, some conformation change has to happen in order to achieve binding of acetate.

Binding constants for the optimized structures are shown in Table 7. The values for RM1 and MC01[14] (Figure 1) are also included. The latter value is shown due to experimental data on binding and performance in the membrane of this receptor.

Table 7. Comparison of logK_{ass} values for studied molecules. Reference molecules are highlighted. The experimental value for MC01 is included. RMR5, RMR6 and RMR7 are modifications of RMA5.

Receptor	**logK_{ass}**
RM1	**11.2**
RMP2	0.8
RMP4	7.4
RMA4	8.1
RMR1	11.7
RMR2	12.3
RMR4	15.8
RMA5	15.7
RMR5	15.7
RMR6	6.1
RMR7	11.1
RMR1, linear	12.1
RMA5, linear	6.3
MC01	**20.6**[a]

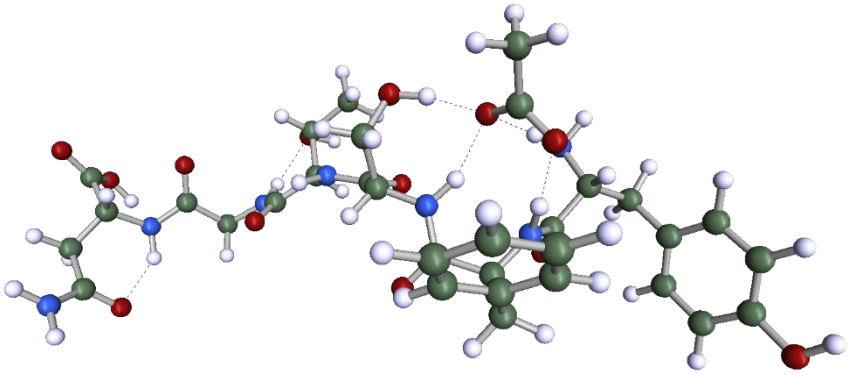
[a] Experimental logK_{ass} value 5.7.[14] COSMO-RS is known to severely overestimate absolute logK_{ass} values while usually satisfactorily predicting trends.[41]

As one can see, binding affinities of RMR1, RMR2, RMR4 and RMA5 are higher than that of the initial peptide RM1, but substantially lower than logK_{ass} of the macrocyclic receptor MC01. It is worth mentioning that COSMO-RS strongly overestimates binding (5.7 in experiment *vs* 20.7 in computations for MC01), and computational logK_{ass} values significantly below 20 (as are the logK_{ass} values of all the investigated peptide receptors) most probably mean poor performance in real-life situation.

Nonetheless, the influence of cyclization was studied by estimating the binding constants for linear analogues of RMR1 and RMA5. The values obtained were 12.1 and 6.3, respectively.

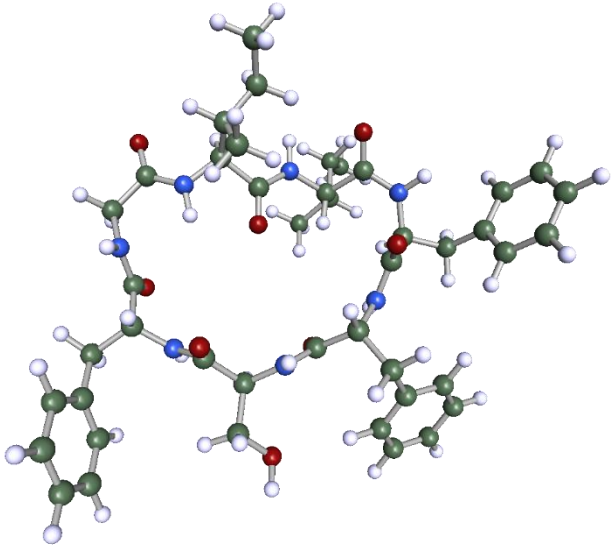
Although the logK_{ass} value for linear RMR1 is high, its selectivity in future is likely to be low as it interacts with carboxylic group via folding around anion (Figure 15). This means that all other anions could bind RMR1 in a similar way. For this reason, further investigation was done with cyclic peptide candidates due to their better geometry for the binding.

Figure 15. Acetate complex of linear analogue of RMR1.



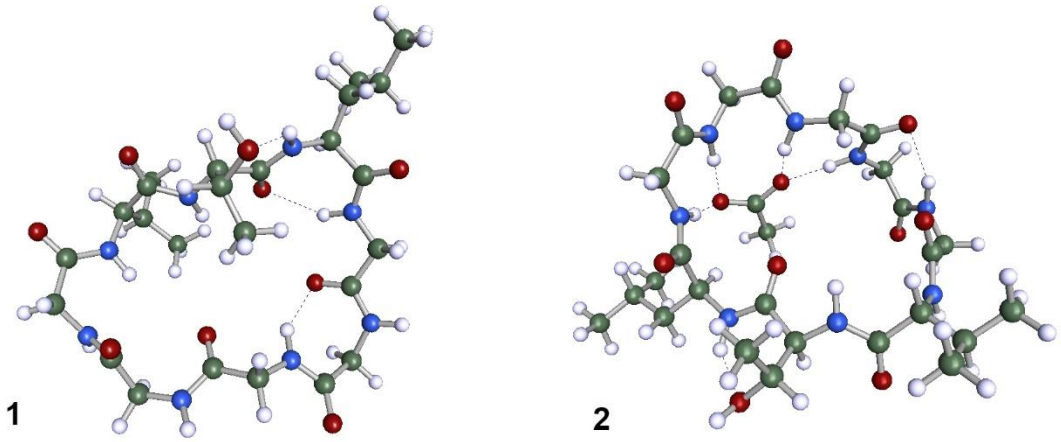
The modifications of RMA5 did not result in a substantial improvement. The binding constant of RMR7 is not significantly lower, but the number of IMHBs is reduced (down to one, Figure 16). However, the HBD sites are not facing the same direction into the cavity, so it decreases the possibility of acetate binding.

Figure 16. The most abundant conformer of RMR7.



The peptide RMR4 showed a remarkably high logK_{ass} value. However, free RMR4 tends to fold inside and has several IMHBs; the HBD groups are not facing inwards the cavity (Figure 17). The number of AAs in its structure is rather small. This does not allow flexibility of the chain to bind anion in close vicinity. Additionally, there are no structural components to hold the geometry in place or distinguish between anions. Therefore, the design of molecules with similar structure might not be suitable for anion sensing.

Figure 17. Most abundant conformations of free RMR4 (1) and its complex with acetate (2).



7. CONCLUSION AND FUTURE OUTLOOK

Finding a suitable peptide-based receptor is a complicated task. It requires extensive research not only because peptide properties vary from one candidate to another, but also because it is a novel and insufficiently studied approach. As a result of this work, one peptide fragment RM1 was thoroughly investigated, and more than 40 of its structural modifications (23 of them are presented in this thesis) were done. Several complications occurred when extensive modifications to initial RM1 structure were performed.

The COSMO-RS method was applied for estimation of conformer stability and calculations of free energy. The geometry optimization was done in Turbomole v.7.21 and the free energy computations were performed using COSMOtherm version 21.0 release 02.11 with BP_SVP_AM1_21 and BP_TZVP_21 parameterization. The binding affinity was calculated from the free energies of peptide-anion complex, free peptide and acetate anion.

The following can be summarized from the study of peptide for acetate binding:

- Linear RM1 has a promising value of binding constant, which can be further improved by structure modifications.
- Cyclization did not always improve binding affinity towards anion. For alkyl chain linkers, the decrease in logK_{ass} value was comparable among all molecules but two. For hydrophobic AA chain linker, not a trend but rather a scatter in logK_{ass} values was observed.
- Substitution of initial AAs in RM1 to more hydrophobic AA does not cause a change or introduces a small decline in binding affinities; it may be caused by IMHB interactions that hold the free peptide together in unfavourable for complex formation geometry.
- As a rule of thumb, bigger cavity means lower probability of interaction with both anionic group and hydrophobic part of target anion. However, COSMO-RS predicts somewhat high values for such cases (RMC7 can be used as an example).
- On the other hand, too small cavity in a receptor molecule can lead to higher number IMHBs and HBD sites facing opposite directions. This means significantly less probable anion complex formation and associated energy difficulties (RMR4 can be used as an example).
- In real-life situations, the binding affinities of RM1 and its structural variations may be several orders of magnitude lower than that obtained with computations. It is caused by the imperfections of COSMO-RS approach, which has been demonstrated earlier[14].

It is complicated to draw a conclusion about potential performance of studied peptide-based receptors in ISE as there is limited experimental data on binding affinities of peptides containing only AAs with carboxylates currently available (including DMSO-water solvent system). One possibility to estimate the binding behaviour of investigated peptide molecules is a comparison between computationally obtained logK_{ass} values of them and previously studied receptors, for which binding constants with carboxylates in real-life solution is known. Nevertheless, this can serve as a direction of improvement: if the binding affinity is higher than that of MC01, it is a good sign to conduct an experiment; in case it is lower, further modelling and computing steps must be taken.

Considering all mentioned above, the studied molecules will most probably serve poorly as ionophores for acetate anion. Thus, their synthesis and imbedding in ISE's membrane at this stage of research is not worth being conducted. However, further investigation and structural improvements of peptides may result in higher binding affinity values and lead to their incorporation into a prototype sensor.

8. ABSTRACT

Carboxylates are widely distributed in various natural and chemically obtained products. Their accurate, easy and fast determination is a compelling task for which an ion-selective electrode (ISE) could be used. ISE works by means of ionophores (anion receptors) – molecules capable of binding the anions. Peptides were chosen to be the basis for the new carboxylate receptor as they possess characteristics necessary for a successful ionophore: many hydrogen bond donor sites for strong binding, flexible and easily tunable structure allowing for tailor-made cavities, easy synthesis.

The aim of this work was to find a peptide sequence that binds carboxylate anions, acetate in particular, investigate its binding affinity ($\log K_{ass}$) and compare it with the values of previously studied macrocyclic receptor. Two methods were used: search in Cambridge Structural Database to find promising structures and COSMO-RS to compute conformation stabilities and $\log K_{ass}$ values in solutions.

Database search yielded 15 potentially suitable peptide fragments, one of which was chosen for further thorough investigation. Firstly, its linear form was computationally studied and revealed a promising $\log K_{ass}$ value. Secondly, to reduce the energy penalty associated with linear receptor-anion complex formation, cyclization and addition of various linkers was explored. The addition of methylene groups as a linking agent did not improve $\log K_{ass}$ value, while hydrophobic amino acid chain linker showed a scatter of results and increased $\log K_{ass}$ value. In particular, one structure of an optimal geometry (RMA5) had the highest binding affinity amongst other candidates, but it was still lower when compared to the macrocyclic receptor.

Future investigations and structural optimizations are needed prior to imbedding these receptor candidates into a prototype sensor.

Keywords: supramolecular chemistry, carboxylate recognition, peptides, COSMO-RS.

CERCS: P300, P410.

Peptiidstruktuuride otsing ja uurimine karboksülaatioonide sidumiseks

Karboksülaatioonid on laialt levinud nii looduslikes kui ka keemiliselt saadud toodetes. Nende täpne, lihtne ja kiire määramine on ahvatlev ülesanne milleks saab kasutada ioonselektiivset elektroodi (ISE). ISE töö põhineb ionofooridel (anioonretseptoritel) – molekulidel mis on võimelised siduma anioone.

Peptiide valiti uue karboksülaatretseptori põhjaks kuna nendel on olemas tõhusa ionofoori jaoks vajalikud omadused: rohkelt vesiniksideme doonoreid tugeva seondumise jaoks, paindlik ja lihtsasti kohandatav struktuur mis võimaldab peenhäälestatud sidumistaskude teket, lihtne süntees.

Käesoleva töö eesmärk oli leida peptiid mis seoks karboksülaatanioone, eeskätt atsetaati, uurida selle seondumisafiinsust (logK_{ass}) ja võrrelda seda varem uuritud makrotsüklilise retseptori väärtusega. Kasutati kahte meetodit: andmebaasiotsingut (Cambridge Structural Database) paljutõotavate struktuuride leidmiseks ja COSMO-RS meetodit konformeeride stabiilsuse ja logK_{ass} väärtuste arvutamiseks lahustes.

Andmebaasiotsing andis 15 potentsiaalselt sobivat peptiidfragmenti, millest üks sai valitud edasiseks põhjalikuks uuringuks. Esiteks, selle lineaarset vormi uuriti arvutuslikult ja leiti lootustäratav logK_{ass} väärtus. Teiseks, et vähendada lineaarse retseptori ja aniooni vahelise kompleksi tekkega seotud energiakulu, prooviti tsükliseerimist ja erinevate vahelülide lisamist. Metüleenrühmade vahelülidena lisamine ei parandanud logK_{ass} väärtust, ent hüdrofoobsetest aminohapetest koosnevad vahefragmendid andsid mitmekesiseid tulemusi ja tõstsid logK_{ass} väärtust. Üks optimaalse geomeetriaga struktuur (RMA5) näitas kõrgeimat seondumisafiinsust teiste kandidaatide hulgas, kuigi see oli ikkagi madalam kui makrotsüklilise retseptori oma. Edasised uuringud ja struktuuri optimeerimine on vajalik enne retseptorkandidaatide kasutamist prototüüpsensoris.

Võtmesõnad: supramolekulaarkeemia, karboksülaatide äratundmine, peptiidid, COSMO-RS

CERCS koodid: P300, P410.

Non-exclusive licence to reproduce the thesis and make the thesis public

I, Sabina Tatarytskaya,

1. grant the University of Tartu a free permit (non-exclusive licence) to:

   reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, my thesis

Search and investigation of peptide structures for carboxylate binding,

supervised by Astrid Darnell, Sofja Tšepelevitš, Ivo Leito,

2. I grant the University of Tartu the permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work from ***05/06/2025*** until the expiry of the term of copyright,
3. I am aware that the author retains the rights specified in points 1 and 2.

4. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Sabina Tatarytskaya
29/05/2023

REFERENCES

(1) Kadam, S. A.; Martin, K.; Haav, K.; Toom, L.; Mayeux, C.; Pung, A.; Gale, P. A.; Hiscock, J. R.; Brooks, S. J.; Kirby, I. L.; Busschaert, N.; Leito, I. Towards the Discrimination of Carboxylates by Hydrogen Bond Donor Anion Receptors. *Chem. Eur. J.* **2015**, 16. https://doi.org/10.1002/chem.201405858.

(2) Dieng, P. S.; Sirlin, C. Recognition of Chiral Carboxylic Anions by Artificial Receptors. *IJMS* **2010**, *11* (9), 3334–3348. https://doi.org/10.3390/ijms11093334.

(3) Badea, G. I.; Radu, G. L. Introductory Chapter: Carboxylic Acids - Key Role in Life Sciences. In *Carboxylic Acid - Key Role in Life Sciences*; Badea, G. I., Radu, G. L., Eds.; InTech, 2018. https://doi.org/10.5772/intechopen.77021.

(4) French, D. Advances in Clinical Mass Spectrometry. In *Advances in Clinical Chemistry*; Elsevier, 2017; Vol. 79, pp 153–198. https://doi.org/10.1016/bs.acc.2016.09.003.

(5) Abelian, A.; Dybek, M.; Wallach, J.; Gaye, B.; Adejare, A. Pharmaceutical Chemistry. In *Remington*; Elsevier, 2021; pp 105–128. https://doi.org/10.1016/B978-0-12-820007-0.00006-4.

(6) Desiraju, G. R.; Steiner, T. *The Weak Hydrogen Bond: In Structural Chemistry and Biology*; IUCr monographs on crystallography; Oxford university press: Oxford, 1999.

(7) Jeffrey, G. A.; Saenger, W. *Hydrogen Bonding in Biological Structures*, Study edition.; Springer Berlin Heidelberg: Berlin, Heidelberg, 1991.

(8) Steed, J. W.; Turner, D. R.; Wallace, K. J. *Core Concepts in Supramolecular Chemistry and Nanochemistry*; John Wiley: Chichester, England ; Hoboken, NJ, 2007.

(9) Rüütel, A.; Tshepelevitsh, S.; Leito, I. One Hundred Carboxylate Receptors. *J. Org. Chem.* **2022**, *87* (21), 14186–14193. https://doi.org/10.1021/acs.joc.2c01725.

(10) Zhao, J.; Yang, D.; Yang, X.-J.; Wu, B. Anion Coordination Chemistry: From Recognition to Supramolecular Assembly. *Coord. Chem. Rev.* **2019**, *378*, 415–444. https://doi.org/10.1016/j.ccr.2018.01.002.

(11) Brooks, S. J.; Gale, P. A.; Light, M. E. Carboxylate Complexation by 1,1′-(1,2-Phenylene)Bis(3-Phenylurea) in Solution and the Solid State. *Chem. Commun.* **2005**, No. 37, 4696. https://doi.org/10.1039/b508144k.

(12) Hiscock, J. R.; Caltagirone, C.; Light, M. E.; Hursthouse, M. B.; Gale, P. A. Fluorescent Carbazolylurea Anion Receptors. *Org. Biomol. Chem.* **2009**, *7* (9), 1781. https://doi.org/10.1039/b900178f.

(13) Martin, K.; Nõges, J.; Haav, K.; Kadam, S. A.; Pung, A.; Leito, I. Exploring Selectivity of 22 Acyclic Urea-, Carbazole- and Indolocarbazole-Based Receptors towards 11 Monocarboxylates: Exploring Selectivity of 22 Acyclic Urea-, Carbazole- and Indolocarbazole-Based Receptors towards 11 Monocarboxylates. *Eur. J. Org. Chem.* **2017**, *2017* (35), 5231–5237. https://doi.org/10.1002/ejoc.201700931.

(14) Rüütel, A.; Yrjänä, V.; Kadam, S. A.; Saar, I.; Ilisson, M.; Darnell, A.; Haav, K.; Haljasorg, T.; Toom, L.; Bobacka, J.; Leito, I. Design, Synthesis and Application of Carbazole Macrocycles in Anion Sensors. *Beilstein J. Org. Chem.* **2020**, *16*, 1901–1914. https://doi.org/10.3762/bjoc.16.157.

(15) Kadam, S. A.; Haav, K.; Toom, L.; Pung, A.; Mayeux, C.; Leito, I. Multidentate Anion Receptors for Binding Glyphosate Dianion: Structure and Affinity:

Multidentate Anion Receptors for Binding Glyphosate Dianion: Structure and Affinity. *Eur. J. Org. Chem.* **2017**, *2017* (11), 1396–1406. https://doi.org/10.1002/ejoc.201601583.
(16) Butler, S. M.; Jolliffe, K. A. Molecular Recognition and Sensing of Dicarboxylates and Dicarboxylic Acids. *Org. Biomol. Chem.* **2020**, *18* (41), 8236–8254. https://doi.org/10.1039/D0OB01761B.
(17) Sun, Y.; Gu, J.; Wang, H.; Sessler, J. L.; Thordarson, P.; Lin, Y.-J.; Gong, H. AAAA–DDDD Quadruple H-Bond-Assisted Ionic Interactions: Robust Bis(Guanidinium)/Dicarboxylate Heteroduplexes in Water. *J. Am. Chem. Soc.* **2019**, *141* (51), 20146–20154. https://doi.org/10.1021/jacs.9b09503.
(18) Kosiorek, S.; Rosa, B.; Boinski, T.; Butkiewicz, H.; Szymański, M. P.; Danylyuk, O.; Szumna, A.; Sashuk, V. Pillar[4]Pyridinium: A Square-Shaped Molecular Box. *Chem. Commun.* **2017**, *53* (100), 13320–13323. https://doi.org/10.1039/C7CC08562A.
(19) Mateus, P.; Delgado, R.; André, V.; Duarte, M. T. Dicarboxylate Recognition Properties of a Dinuclear Copper(II) Cryptate. *Inorg. Chem.* **2015**, *54* (1), 229–240. https://doi.org/10.1021/ic502230q.
(20) Diehl, K. L.; Anslyn, E. V. Array Sensing Using Optical Methods for Detection of Chemical and Biological Hazards. *Chem. Soc. Rev.* **2013**, *42* (22), 8596. https://doi.org/10.1039/c3cs60136f.
(21) Hamley, I. W. *Introduction to Peptide Science*, First edition.; Wiley: Hoboken, NJ, 2020.
(22) Elmes, R. B. P.; Jolliffe, K. A. Anion Recognition by Cyclic Peptides. *Chem. Commun.* **2015**, *51* (24), 4951–4968. https://doi.org/10.1039/C4CC10095F.
(23) Sikora, K.; Jaśkiewicz, M.; Neubauer, D.; Migoń, D.; Kamysz, W. The Role of Counter-Ions in Peptides—An Overview. *Pharmaceuticals* **2020**, *13* (12), 442. https://doi.org/10.3390/ph13120442.
(24) Kubik, S. Anion Recognition in Aqueous Media by Cyclopeptides and Other Synthetic Receptors. *Acc. Chem. Res.* **2017**, *50* (11), 2870–2878. https://doi.org/10.1021/acs.accounts.7b00458.
(25) Yang, D.; Qu, J.; Li, W.; Zhang, Y.-H.; Ren, Y.; Wang, D.-P.; Wu, Y.-D. Cyclic Hexapeptide of D , L -α-Aminoxy Acids as a Selective Receptor for Chloride Ion. *J. Am. Chem. Soc.* **2002**, *124* (42), 12410–12411. https://doi.org/10.1021/ja027073y.
(26) Kubik, S.; Kirchner, R.; Nolting, D.; Seidel, J. A Molecular Oyster: A Neutral Anion Receptor Containing Two Cyclopeptide Subunits with a Remarkable Sulfate Affinity in Aqueous Solution. *J. Am. Chem. Soc.* **2002**, *124* (43), 12752–12760. https://doi.org/10.1021/ja026996q.
(27) Pajewski, R.; Ferdani, R.; Schlesinger, P. H.; Gokel, G. W. Chloride Complexation by Heptapeptides: Influence of C- and N-Terminal Sidechains and Counterion. *Chem. Commun.* **2004**, No. 2, 160. https://doi.org/10.1039/b312008m.
(28) Zhang, Y.; Yin, Z.; He, J.; Cheng, J.-P. Effective Receptors for Fluoride and Acetate Ions: Synthesis and Binding Study of Pyrrole- and Cystine-Based Cyclopeptido-Mimetics. *Tetrahedron Lett.* **2007**, *48* (34), 6039–6043. https://doi.org/10.1016/j.tetlet.2007.06.066.
(29) Wolfbeis, O. S. Editorial: Probes, Sensors, and Labels: Why Is Real Progress Slow? *Angew. Chem. Int. Ed.* **2013**, *52* (38), 9864–9865. https://doi.org/10.1002/anie.201305915.

(30) Nematiaram, T.; Padula, D.; Landi, A.; Troisi, A. On the Largest Possible Mobility of Molecular Semiconductors and How to Achieve It. *Adv. Funct. Mater.* **2020**, *30* (30), 2001906. https://doi.org/10.1002/adfm.202001906.
(31) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2016**, *72* (2), 171–179. https://doi.org/10.1107/S2052520616003954.
(32) Allen, F. H. The Cambridge Structural Database: A Quarter of a Million Crystal Structures and Rising. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58* (3), 380–388. https://doi.org/10.1107/S0108768102003890.
(33) Ward, S. C.; Sadiq, G. Introduction to the Cambridge Structural Database – a Wealth of Knowledge Gained from a Million Structures. *CrystEngComm* **2020**, *22* (43), 7143–7144. https://doi.org/10.1039/D0CE90154G.
(34) Ferrence, G. M.; Tovee, C. A.; Holgate, S. J. W.; Johnson, N. T.; Lightfoot, M. P.; Nowakowska-Orzechowska, K. L.; Ward, S. C. *CSD Communications* of the Cambridge Structural Database. *IUCrJ* **2023**, *10* (1), 6–15. https://doi.org/10.1107/S2052252522010545.
(35) Korb, O.; Kuhn, B.; Hert, J.; Taylor, N.; Cole, J.; Groom, C.; Stahl, M. Interactive and Versatile Navigation of Structural Databases. *J. Med. Chem.* **2016**, *59* (9), 4257–4266. https://doi.org/10.1021/acs.jmedchem.5b01756.
(36) Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242. https://doi.org/10.1093/nar/28.1.235 (www.rcsb.org).
(37) Shakerzadeh, E. Chapter 4 - Theoretical Investigations of Interactions between Boron Nitride Nanotubes and Drugs. In *Boron Nitride Nanotubes in Nanomedicine*; Elsevier, 2016.
(38) Klamt, A.; Eckert, F.; Arlt, W. COSMO-RS: An Alternative to Simulation for Calculating Thermodynamic Properties of Liquid Mixtures. *Annu. Rev. Chem. Biomol. Eng.* **2010**, *1* (1), 101–122. https://doi.org/10.1146/annurev-chembioeng-073009-100903.
(39) Klamt, A. *COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*, 1st ed.; Elsevier: Amsterdam Boston, 2005.
(40) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on the Generalized Born Approximation with Asymmetric Descreening. *J. Chem. Theory Comput.* **2009**, *5* (9), 2447–2464. https://doi.org/10.1021/ct900312z.
(41) Tshepelevitsh, S.; Oss, M.; Pung, A.; Leito, I. Evaluating the COSMO-RS Method for Modeling Hydrogen Bonding in Solution. *ChemPhysChem* **2013**, *14* (9), 1909–1919. https://doi.org/10.1002/cphc.201300186.
(42) Bouillot, B.; Teychené, S.; Biscans, B. An Evaluation of Thermodynamic Models for the Prediction of Drug and Drug-like Molecule Solubility in Organic Solvents. *Fluid Phase Equilib.* **2011**, *309* (1), 36–52. https://doi.org/10.1016/j.fluid.2011.06.032.
(43) Klamt, A. The COSMO and COSMO-RS Solvation Models. *WIREs Comput. Mol. Sci.* **2011**, *1* (5), 699–709. https://doi.org/10.1002/wcms.56.
(44) Xue, Z.; Mu, T.; Gmehling, J. Comparison of the a Priori COSMO-RS Models and Group Contribution Methods: Original UNIFAC, Modified UNIFAC(Do), and Modified UNIFAC(Do) Consortium. *Ind. Eng. Chem. Res.* **2012**, *51* (36), 11809–11817. https://doi.org/10.1021/ie301611w.
(45) Reinisch, J.; Diedenhofen, M.; Wilcken, R.; Udvarhelyi, A.; Glöß, A. Benchmarking Different QM Levels for Usage with COSMO-RS. *J. Chem. Inf. Model.* **2019**, *59* (11), 4806–4813. https://doi.org/10.1021/acs.jcim.9b00659.

(46) Villarroel, E.; Olea, F.; Araya-López, C.; Merlet, G.; Cabezas, R.; Romero, J.; Quijada-Maldonado, E. COSMO-RS Evaluation as a Tool for Prediction of Solvents in Dispersive Liquid-Phase Microextraction: Evaluation of Conventional Solvents and Ionic Liquids as Extractants. *J. Mol. Liq.* **2022**, *354*, 118861. https://doi.org/10.1016/j.molliq.2022.118861.
(47) Pung, A.; Leito, I. Predicting Relative Stability of Conformers in Solution with COSMO-RS. *J. Phys. Chem. A* **2017**, *121* (36), 6823–6829. https://doi.org/10.1021/acs.jpca.7b05197.
(48) TURBOMOLE V7.21 2017, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com.
(49) BIOVIA COSMOconfX. version 21.0. Dassault Systèmes. http://www.3ds.com, 2021.
(50) BIOVIA COSMOtherm, Release 2021; Dassault Systèmes. http://www.3ds.com.
(51) Fukuyama, K.; Matsubara, H.; Rogers, L. J. Crystal Structure of Oxidized Flavodoxin from a Red Alga Chondrus Crispus Refined at 1.8 Å Resolution. *J. Mol. Biol.* **1992**, *225* (3), 775–789. https://doi.org/10.1016/0022-2836(92)90400-E.
(52) Yrjänä, V.; Saar, I.; Ilisson, M.; Kadam, S. A.; Leito, I.; Bobacka, J. Potentiometric Carboxylate Sensors Based on Carbazole-Derived Acyclic and Macrocyclic Ionophores. *Chemosensors* **2020**, *9* (1), 4. https://doi.org/10.3390/chemosensors9010004.
(53) Bancet, A.; Raingeval, C.; Lomberget, T.; Le Borgne, M.; Guichou, J.-F.; Krimm, I. Fragment Linking Strategies for Structure-Based Drug Design. *J. Med. Chem.* **2020**, *63* (20), 11420–11435. https://doi.org/10.1021/acs.jmedchem.0c00242.
(54) Skaf, M. S.; Polikarpov, I.; Stanković, I. M. A Linker of the Proline-Threonine Repeating Motif Sequence Is Bimodal. *J. Mol. Model.* **2020**, *26* (7), 178. https://doi.org/10.1007/s00894-020-04434-0.

Appendix

Appendix 1. The AA sequence of RM1 modifications. Candidates, for which the parallel corresponding linear peptide was investigated, are marked with *.

Name	Sequence													
RM1	-	-	-	-	-	-	Ser	Thr	Ser	Thr	Gly	Asn	Thr	Thr
RMP1	-	-	Pro	Tyr	Phe	Pro	Ser	Thr	Ser	Thr	Gly	Asn	Thr	Thr
RMP2	-	-	-	Pro	Tyr	Phe	Pro	Thr	Ser	Thr	Gly	Asn	Thr	-
Peptide3	-	-	-	-	Pro	Tyr	Phe	Pro	Ser	Thr	Gly	Asn	-	-
Peptide4	-	-	-	-	Tyr	Phe	Ser	Thr	Ser	Thr	Gly	Asn	Thr	Thr
RMA1	-	-	-	-	-	Tyr	Phe	Thr	Ser	Thr	Gly	Asn	Thr	-
RMR1*	-	-	-	-	-	-	Tyr	Phe	Ser	Thr	Gly	Asn	-	-
Peptide7	Pro	Gly	Val	Thr	Leu	Pro	Ser	Thr	Ser	Thr	Gly	Asn	Thr	Thr
Peptide8	-	Pro	Gly	Val	Thr	Leu	Pro	Thr	Ser	Thr	Gly	Asn	Thr	-
RMP3	-	-	Pro	Gly	Val	Thr	Leu	Pro	Ser	Thr	Gly	Asn	-	-
Peptide10	-	-	Gly	Val	Thr	Leu	Ser	Thr	Ser	Thr	Gly	Asn	Thr	Thr
Peptide11	-	-	-	Gly	Val	Thr	Leu	Thr	Ser	Thr	Gly	Asn	Thr	-
RMA2	-	-	-	-	Gly	Val	Thr	Leu	Ser	Thr	Gly	Asn	-	-
Peptide13	-	-	Pro	Ile	Met	Pro	Ser	Thr	Ser	Thr	Gly	Asn	Thr	Thr
RMP4	-	-	-	Pro	Ile	Met	Pro	Thr	Ser	Thr	Gly	Asn	Thr	-
Peptide15	-	-	-	-	Pro	Ile	Met	Pro	Ser	Thr	Gly	Asn	-	-
Peptide16	-	-	-	-	Ile	Met	Ser	Thr	Ser	Thr	Gly	Asn	Thr	Thr

Name	Sequence													
RMA3	-	-	-	-	-	Ile	Met	Thr	Ser	Thr	Gly	Asn	Thr	-
Peptide18	-	-	-	-	-	-	Ile	Met	Ser	Thr	Gly	Asn	-	-
Peptide19	-	-	-	-	-	Tyr	Leu	Thr	Ser	Thr	Gly	Asn	Thr	Thr
Peptide20	-	-	-	-	-	Ile	Met	Thr	Ser	Thr	Gly	Asn	Thr	Thr
Peptide21	-	-	-	-	-	Tyr	Phe	Thr	Ser	Thr	Gly	Asn	Thr	Thr
Peptide22	-	-	-	Gly	Val	Thr	Leu	Thr	Ser	Thr	Gly	Asn	Thr	Thr
Peptide23	-	-	-	-	Tyr	Leu	Ser	Thr	Ser	Thr	Gly	Asn	Thr	-
RMA4	-	-	-	-	Ile	Met	Ser	Thr	Ser	Thr	Gly	Asn	Thr	-
Peptide25	-	-	-	-	Tyr	Phe	Ser	Thr	Ser	Thr	Gly	Asn	Thr	-
Peptide26	-	-	Gly	Val	Thr	Leu	Ser	Thr	Ser	Thr	Gly	Asn	Thr	-
Peptide27	-	-	-	-	-	-	Tyr	Leu	Ser	Thr	Gly	Asn	Thr	-
Peptide28	-	-	-	-	-	-	Ile	Met	Ser	Thr	Gly	Asn	Thr	-
Peptide29	-	-	-	-	-	-	Tyr	Phe	Ser	Thr	Gly	Asn	Thr	-
Peptide30	-	-	-	-	Gly	Val	Thr	Leu	Ser	Thr	Gly	Asn	Thr	-
Peptide31	-	-	-	-	-	Tyr	Leu	Thr	Ser	Thr	Gly	Asn	-	-
RMA5*	-	-	-	-	-	Ile	Met	Thr	Ser	Thr	Gly	Asn	-	-
Peptide33	-	-	-	-	-	Tyr	Phe	Thr	Ser	Thr	Gly	Asn	-	-
Peptide34	-	-	-	Gly	Val	Thr	Leu	Thr	Ser	Thr	Gly	Asn	-	-
Peptide35	-	-	-	-	Ile	Asn	Gly	Gly	Ser	Gly	Gly	-	-	-
Peptide36	-	-	-	-	Ile	Gly	Gly	Gly	Ser	Gly	Gly	-	-	-

Name	Sequence													
Peptide37*	-	-	-	-	-	-	Phe	Phe	Ser	Thr	Gly	Asn	-	-
Peptide38*	-	-	-	-	-	-	Phe	Phe	Ser	Gly	Gly	Gly	-	-
Peptide39	-	-	-	Pro	Tyr	Gly	Pro	Thr	Ser	Thr	Gly	Asn	Thr	-
RMR2	-	-	-	-	-	Tyr	Leu	Thr	Ser	Thr	Gly	Gly	-	-
Peptide41	-	-	-	-	-	Leu	Leu	Thr	Ser	Thr	Gly	Gly	-	-
Peptide42	-	-	-	-	-	Tyr	Gly	Thr	Ser	Thr	Gly	Asn	Thr	-
RMR3	-	-	-	Gly	Val	Thr	Leu	Gly	Ser	Thr	Gly	Asn	Thr	-
Peptide44	-	-	-	-	-	Tyr	Gly	Thr	Ser	Thr	Gly	Asn	Thr	-
Peptide45	-	-	-	Gly	Gly	Val	Thr	Leu	Gly	Thr	Gly	-	-	-
RMR4	-	-	-	-	Gly	Val	Thr	Leu	Gly	Gly	Gly	Gly	-	-
RMR5	-	-	-	-	-	Ile	Phe	Thr	Ser	Thr	Gly	Ile	-	-
RMR6	-	-	-	-	-	Ile	Phe	Phe	Ser	Thr	Gly	Ile	-	-
RMR7	-	-	-	-	-	Ile	Phe	Phe	Ser	Phe	Gly	Ile	-	-