

UNIVERSITY OF TARTU
FACULTY OF SCIENCE AND TECHNOLOGY
Institute of Chemistry

Astrid Pung

Computational estimation of receptor- anion binding in solution

Master`s thesis

Supervisor: Ivo Leito, PhD

Tartu 2014

Table of contents

Abbreviations	3
1. Introduction.....	4
2. Overview of literature	5
2.1. Definition and general working principle of receptor molecules	5
2.2. Intermolecular forces and their role in determining the binding strength	6
2.3. Hydrogen bonds.....	6
2.4. Binding constants	7
2.5. Selectivity	8
2.6. Receptor molecule arrays	8
3. Computational methods	9
3.1. DFT calculations and the COSMO-RS method	9
3.2. General information on calculation parameters.....	10
3.3. Choosing starting geometries	11
4. Choosing receptor molecules, solvents and anions for the study	11
4.1. Small receptor molecules.....	11
4.2. Large receptor molecules.....	14
4.3. Studied solvents.....	17
4.4. Studied anions.....	18
5. Results and discussion	22
5.1. General information.....	22
5.2. Indolocarbazole and urea based small receptor complexes with the acetate anion...	23
5.3. Binding of the glyphosate dianion by receptors 1a to 1j	26
5.4. Binding of the glyphosate, succinate, malonate, glutarate and adipate dianions by receptors 2a to 2f	28
5.5. Binding of the glyphosate dianion by receptors 3a to 3c	30
5.6. Binding of the glyphosate dianion by receptors 4a – 4d	31
5.7. Effects of the choice of parametrization.....	33
5.8. Effects of the water content	34
6. Summary	35
7. Kokkuvõte.....	37
8. References.....	39
9. Appendix. Lowest energy conformers for receptors 1a - 1j.	43

Abbreviations

AN - acceptor numbers

Adj - adjusted (reduced) computational association constant (binding constant) values

Calc - computationally determined association constant (binding constant) values

BP - Becke-Perdew, a class of DFT functionals

COSMO-RS - COnductor- like Screening MOdel for Real Solvation

CSM - continuum solvation model

DFT - density functional theory

DN - donor numbers

Exp - experimentally determined association constant (binding constant) values

HB - hydrogen bond

HBA - hydrogen bond acceptor

HBD - hydrogen bond donor

K_{ass} - the association constant (binding constant) of a receptor-anion complex

MeCN - acetonitrile

SEP - strain estimation parameter

TZVP - triple zeta valence plus polarization

UV - ultraviolet spectral range

Vis - visual spectral range

1. Introduction

Design and preparation of synthetic receptors for different anions has become an area of intense research. [1] The obvious goal is finding receptors exhibiting as high as possible sensitivity and selectivity towards specific analytes, which then could be exploited for constructing sensors or analytical separation systems. Achieving this is not easy, however, as the binding strength of a receptor towards a specific anion depends on the complex interplay of several factors. [1,2]

It is quite work-intensive to experimentally determine how efficient a given receptor structure will be in binding a given anion. Therefore, a computational methodology for prediction of binding would be very useful. In this study the suitability of the COSMO-RS [3] computational method is evaluated for description of anion binding to synthetic receptors and the COSMO-RS method is applied to estimate the differences between binding affinity of synthetic receptors towards different mono- and dianions.

Acetate anion is used for assessing the suitability of the COSMO-RS methodology for carboxylate anion binding as there are quite abundant experimental binding data available. Dianions are of interest, because many of them (succinate, oxalate, etc.) are important in biochemistry. Finally, the glyphosate dianion is of specific interest as it is probably the most widely used pesticide in Europe [4] and at neutral pH is a dianion.

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

1. To study on the example of acetate anion the possibility of using the COSMO-RS method for qualitatively or semiquantitatively predicting trends in the binding strength of receptor-anion complexes in a solution.
2. To use the COSMO-RS method to predict the trends in receptor-dianion complex formation originating from different linkers, binding fragments and steric effects.
3. From these predictions, to try to find promising receptor molecule structures that could be used as parent structures for developing a sensitive molecular receptor for the glyphosate dianion.

2. Overview of literature

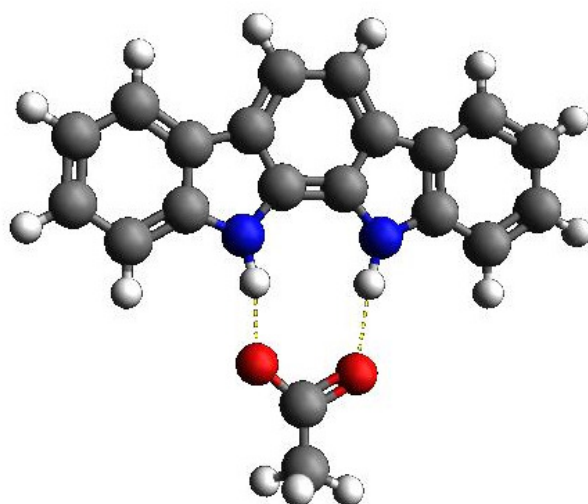
2.1. Definition and general working principle of receptor molecules

Receptor molecules are a type of molecule that can be used to detect an ion or another molecule from a solution. They can be selective to anions, cations or to neutral molecules. [1]

The methods of interaction between the analyte and the receptor molecule are varied, but can include hydrogen bond formation, solvophobic effects, electrostatic interactions, ion pair formation etc. Several types of interactions can occur at the same time, which improves the binding strength of the formed complex. [1,5]

The formed receptor-anion complexes can be detected in multiple ways. When the complex formation changes the absorption spectrum of the solution, UV/Vis spectroscopy can be used. When the change occurs in the visual wavelength range, the change of color can sometimes be observed with the naked eye. In some cases, complex formation causes fluorescence quenching or excitation, in which case fluorescence spectroscopy can be used for the method of detection. The receptor-analyte complex formation can sometimes cause a change in the electrical properties of the solution, in which case a voltammetric detection method could be used. [5-7]

Scheme 1. Example of receptor-anion complex formation. Indolocarbazole receptor complex with the acetate anion.



2.2. Intermolecular forces and their role in determining the binding strength

The binding strength of a receptor towards a specific anion depends on the complex interplay of several factors. [1,8,9] Firstly, the suitability of spatial arrangement of the binding sites of the receptor for the anion of interest (complementarity) [1] is of crucial importance. Connected to this is the issue of flexibility/rigidity (preorganization) [10] of the receptor backbone. When the preferred conformation of the free receptor in solution is very similar to its conformation in the receptor-anion complex, no energetically unfavorable conformation change is needed upon complexation with the anionic guest. [1] Secondly, the binding strength – usually governed by the hydrogen bond [11,12] (HB) donicity of the binding sites of the receptor – should be high. Thirdly, a receptor molecule with more than one binding site is able to form stronger complexes due to more interactions with the guest. [1] Binding strength is also affected by the existence of co-operativity effects [1] (similar to the chelate effect in the binding of macrocycles and metal cations) [13] between the binding sites. If the binding sites of a receptor molecule produce combined interactions, then binding is stronger than in the case of independent binding sites. [1,2]

These principles are not as straightforward to apply as it seems. Firstly, the more flexible is the receptor's backbone the better it can sterically fit around the anion. At the same time, this flexibility immediately introduces an entropy penalty, because the free receptor has more degrees of freedom than the receptor in the complex. Secondly, the stronger is the HBD ability of the receptor, the stronger will also be its interactions with solvent molecules, especially water (the most relevant solvent for any ion sensing application), as well as other anions. [1]

2.3. Hydrogen bonds

A hydrogen bond is a molecular interaction between a hydrogen bond donor (HBD) and a hydrogen bond acceptor (HBA) group. It is directional, and on its properties is somewhere between electrostatic interactions and covalent bonds. [11,12,14]

Hydrogen bond donor fragments in the receptor molecules are hydrogens linked to an electronegative atom (N,O), the HBA fragment is an electronegative atom with a free ion pair or an electron rich part of the molecule. The more positive is the partial charge of the HBD fragment hydrogen and the more negative is the partial charge of the HBA fragment, the stronger is the interaction. For that reason, ions form tend to stronger hydrogen bonds than neutrals. [11,12,14]

Hydrogen bonds are directional but the rigidity and force constants of HBs are a lot weaker than those of covalent bonds. The efficiency of HB formation and the strength of the created bond depends strongly on the angle between the atoms forming the HB. [12] The range of possible hydrogen bond formation angles is narrow for strong hydrogen bonds (175-180°) and the weaker hydrogen bonds are more flexible and allow for angles of 90-150°. [2] Directionality is less important for weak HBs because the strength of the interaction is low enough to be similar to electrostatic interaction. [12]

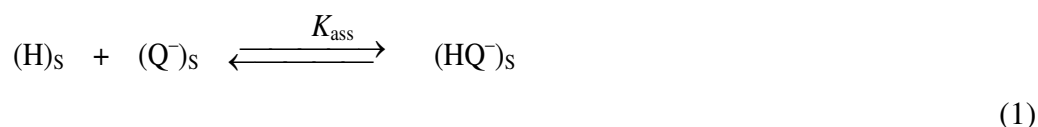
HB formation is strongly affected by the surrounding chemical environment. The water content and HBD abilities of the solvent have a very important role in determining the formation of hydrogen bonds in the solvent environment. When a solvent environment contains some amounts of water or the solvent itself has good HBD abilities, both the chance of other HBs occurring and their strength are reduced. [12]

In a receptor-anion complex, it is possible for the ion to be involved in several hydrogen bond interactions with the receptor molecule. Using directional bonds helps to change the binding more selective to the anion geometry as stronger hydrogen bond interactions need the receptor and anion to have complementary geometries. [1,2]

Other interactions between molecules than HB formation can also happen at the same time and have an additional influence on the binding strength. [1]

2.4. Binding constants

The binding of a guest species is an equilibrium process that can be characterized by K_{ass} , the association (binding) constant. In the case of a host molecule H binding an anionic guest Q- with a stoichiometric ratio of 1:1 and forming a complex HQ-, according to Equation 1: [2,15]



The lowercase index S in Equation 1 specifies solvation by the solvent molecules for all particles.

The association constant K_{ass} and the change in free energy ΔG can be calculated from Equation 2 and 3:

$$K_{\text{ass}} = \frac{a(HQ^-)}{a(H) \cdot a(Q^-)} \quad (2)$$

$$\Delta G_{\text{ass}} = G(HQ^-) - G(H) - G(Q^-) \quad (3)$$

K_{ass} ja ΔG_{ass} are linked through Equation 4:

$$\Delta G_{\text{ass}} = -RT \ln K_{\text{ass}} \quad (4)$$

When experimental methods are used for determining the binding constant value, it is calculated by using the concentrations of the species present at the equilibrium. [15]

When computational methods are used in this study for estimating the binding constants, Equation 4 is used. The free energy G values in a chosen solvent system are obtained for the host (receptor), guest (anion of interest) and the host-guest (anion-receptor) complex through COSMO-RS calculations. The change in free energy ΔG is calculated for the reaction and the association constant found using Equation 4.

It is possible that the binding does not happen in a stoichiometric 1:1 ratio, in which case the calculations would include additional association constants and become more complex. All calculations carried out in this study have been made assuming 1:1 stoichiometry for the binding.

2.5. Selectivity

Generally, the receptor molecules are not selective to only one analyte. An anion-selective receptor molecule, for example, can bind one anion stronger, but usually still can potentially also form weaker complexes with other anions. Selectivity can be improved in several ways. The complementarity, co-operativity and preorganization effects should be taken into account when designing a receptor. [1] For another method of improving selectivity, arrays of several receptors can be used to create a more selective response to the analyte of interest. [14] Eliminating the interfering particles from the analysis solution as a stage of sample preparation is another option. [16]

Receptor molecules could also be used at the preparative stage of analysis for the purpose of preconcentration. In that case, selectivity is not vital as the receptor molecules are not used to directly determine the analyte but to bind it and convert it into a more concentrated form.

2.6. Receptor molecule arrays

An array of receptors can be used instead of a single receptor for improving selectivity of the analysis. When a group of receptors is used together, different receptors will have different responses to different anions. By analyzing these responses, a receptor array will be able to more selectively determine the presence of anions in the cases where a single receptor might not be very useful due to low selectivity. [17]

3. Computational methods

3.1. DFT calculations and the COSMO-RS method

Quantum chemical calculations are based on finding approximate solutions to the Schrodinger equation. Through the calculations the initial geometry and respective wave functions are converged to the geometry of the minimum energy, the respective wave function and the total energy of the system. [18]

The main sequence of operations of quantum chemical calculations is the following. First, the starting geometry of the studied molecule or ion is defined. An approximate wave function is created for the starting geometry and the wave function is iteratively changed towards a lower energy. Every new wave function is compared with the convergence criteria. When the wave function meets the criteria, the changes of energy and geometry gradient during the calculations are compared. When the result meets the optimum for a given system and geometry, the calculation is finished, otherwise the geometry is modified towards the gradient and used a new wave function is created for that geometry. The calculation process continues until an optimal geometry is found. [18]

In the DFT method, an approximation is made to simplify the calculations. A system of multiple electrons with a large number of coordinates is replaced with a functional of electron density that only has 3 coordinates. Using this approximation, the calculation speed increases by a large margin while the accuracy still remains good. For this reason, the DFT method can be used to carry out calculations of large molecules with suitable accuracy. [18]

Historically, the quantum chemistry calculations have been carried out mainly in the gas phase. At the same time, for practical uses it would be important to be able to compute the properties of compounds in solutions. Compared to the gas phase, the solvent environment also has interactions between molecules that must be taken into account by the calculations.

At first, methods of molecular dynamics were used for the study of solvents. These methods work for some applications but they cannot be used for situations where accurate description of molecules is needed. Another method for the study of the behavior of compounds in solvents is using correlation equations that have been developed by using a large number of experimental data. The problem of this method is that it does not help predict the behavior of molecules when they are different from the molecules used for creating the correlation equation. Also, that method does not take into account that the interactions of different functional groups can largely depend on the properties of the used solvent. This means the method cannot make predictions for compounds that have not been previously synthesized or that have been studied in different conditions. [3]

Currently in the field of quantum chemical calculations, the solvents are usually simulated by continuum solvation models (CSMs [19]). These models are based on the simplified approximation of the real solvent by a dielectric continuum of permittivity ϵ . CSMs nowadays are parameterized on the solvation energies of organic compounds. Each solvent for the

method needs separate parametrization and CSMs provide no concept for mixtures or for the description of temperature effects. [20]

COSMO is a technical modification of the dielectric CSMs. It replaces the dielectric boundary conditions with simpler scaled-conductor boundary condition. COSMO-RS (Conductor-Like Screening Model for Realistic Solvation) as a calculation method combines the COSMO model with statistical thermodynamics of interacting molecular surface fragments in order to better take into account specific interactions between molecules. [21]

The COSMO-RS method uses the DFT method on the state of molecules embedded in a conductor with a dielectric constant of $\epsilon=\infty$ as a reference starting point for the simulation. The COSMO calculation yields the energy, electron density, polarization charge densities and geometry of the molecule in a virtual conductor. It takes into account the electronic and steric effects of the molecular structure, including intramolecular hydrogen bonds. [20]

The RS part of the calculation is based on statistical thermodynamics of interacting charged molecular surface segments. Starting from the reference of molecules being separate in a conductor, it approximates a closely packed liquid system by iterative introduction of molecular contacts. From the thermodynamic point of view, COSMO-RS is based on Gibbs free energy concept of non-compressible fluids. The statistical thermodynamics part treats all molecules in the solution (solvent, solvent impurities, solute molecules) similarly. It accounts for the van der Waals and electrostatic interactions as well as the hydrogen bonds between any species in the solution. [3]

The strengths of the COSMO-RS method are the ability to carry out computations in mixed solvents, at high concentrations of solute molecules and to be able to take into account the effects of temperature. [20] The downsides of the method are quite extensive parametrization and handling the intermolecular interactions in a simplified way through pairs of interacting surface segments. This method takes steric effects only partially into account and does not take into account the possible long-range interactions between molecules. Another limitation of the COSMO-RS method is its rather poor performance when small anions with highly localized charge are involved (e.g F⁻). This can be due to the high polarization charge densities that bring the interaction terms to the limits and possibly due to the outlying charge error (OCE [21]) [22].

The main reason of choosing COSMO-RS as the computational method for this work is that it is not parameterized for any specific solvent and it has the ability to handle solvent mixtures.

3.2. General information on calculation parameters

The geometry optimization and application of the COSMO model were carried out using the program Turbomole v.6.4 [23] using the following criteria: Becke-Perdew functional [24,25], TZVP basis set, wave function convergence criteria: max difference 10^{-6} Hartree, geometry convergence: max gradient $|dE/dxyz| 10^{-3}$ Hartree Bohr⁻¹.

Calculations for finding the free energies of studied particles were carried out using the program COSMOthermX14 with the parameterization BP_TZVP_C30_1401 [26]. The

environment temperature chosen for the calculations was 25 °C and the solvent used was MeCN with 0.5% water.

For the calculations, a set of conformers with the lowest energies was chosen to be used for every molecule in order to achieve comparability of the results with real situation in solution. The method of conformer treatment of the COSMO-RS program takes into account an ensemble of molecular conformations, including ones that may be of low free energy in polar or nonpolar solvents. The method calculates relative statistical weights (abundances) for each conformation used according to the Boltzmann statistics and evaluates the thermodynamic properties of a multi-conformational compound as corresponding averages. [20]

The goal was to find the optimized geometries of certain host-guest complex conformations and to calculate the estimates of association constants from free energies of these molecules in a chosen solvent mixture found using the COSMOtherm program. Two steps were taken to analyze the calculation results:

Firstly, the optimized geometries were used to analyze the steric effects of binding.

Secondly, the estimates of association constant values were compared to estimate differences in binding strengths for different molecules. This enabled predicting which receptor structure would have the highest binding affinity towards a given anion. The binding constants were calculated according to Equation 4.

3.3. Choosing starting geometries

For the calculations, the conformers used were chosen according to the following principles:

- Enough conformers – i.e. the most stable ones – should be taken into account for each molecule in order to get comparable results.
- In the anion-receptor complexes, the anion is often in a conformation close to its conformation as a free anion and the receptor is often in a conformation close to its conformation as a free receptor.
- When a receptor molecule has a rigid structure, a smaller number of conformers exist for the molecule. When the molecule includes some flexible fragments, the possible main conformations occurring from that must be taken into account.
- When the binding groups in a free receptor can face the same direction or the opposite direction, both types of such conformers must be taken into account.
- Approximate numbers of conformers chosen for the studied particles were the following: 5-10 for the dianions, 2-3 for the dianion complexes, 5-10 for the receptors.

4. Choosing receptor molecules, solvents and anions for the study

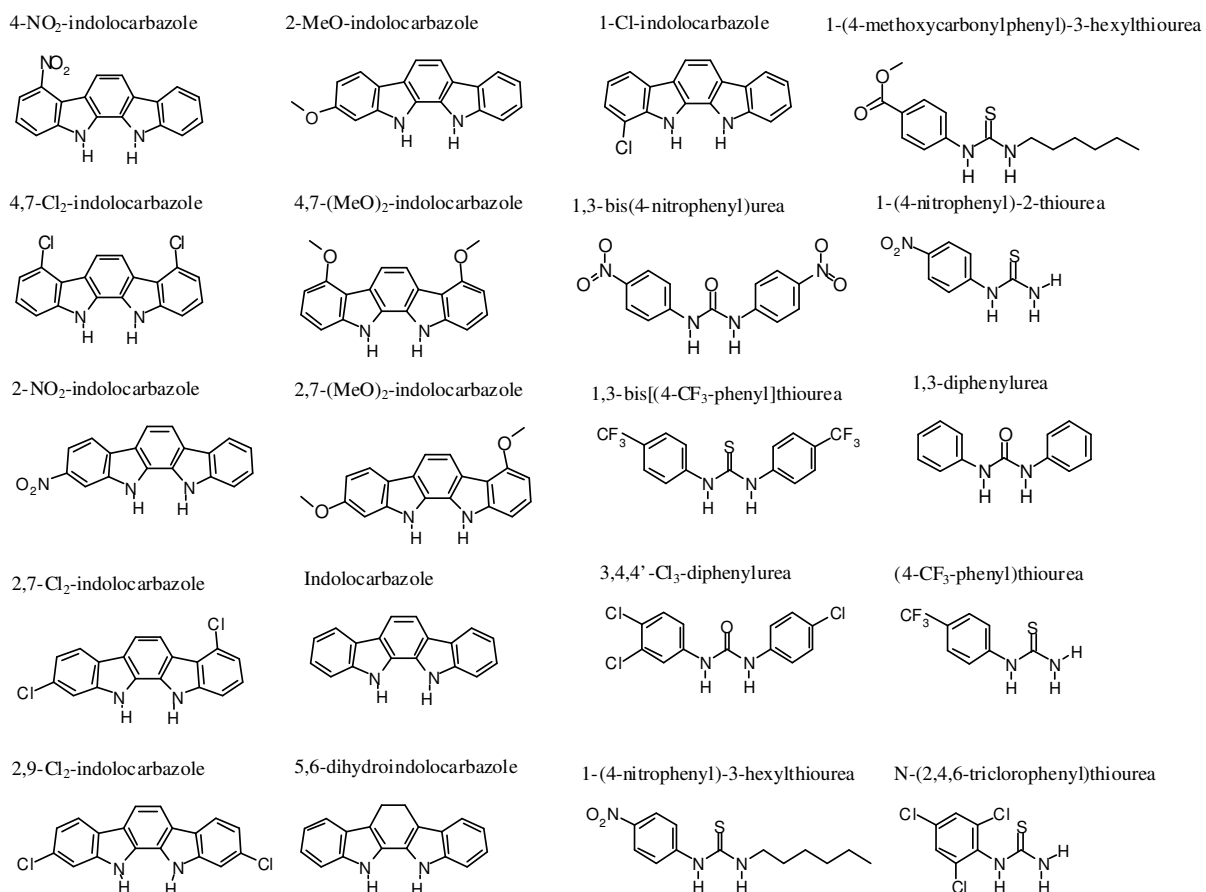
4.1. Small receptor molecules

In my previous work, miscellaneous small receptor molecules belonging to the classes of indolocarbazoles, ureas and pyrrolotetrazoles were studied in order to evaluate the suitability of the COSMO-RS method for predicting the strength of receptor-anion binding. For this, a

number of receptor molecules were found from articles [27-32] with experimental results of measured binding constants available. The same binding constant values were found computationally through the use of the COSMO-RS method and compared with the experimental values. The results of the small receptor calculations and the comparison of calculated and experimental values have been published in an article [22]. It was found that while the method cannot be used for predicting the absolute values of the binding constants, the trends between binding strengths of different anions or receptors can generally still be predicted.

In this work, the binding of a set of small indolocarbazole and urea based receptors with the acetate anion was studied. The receptors were chosen based on the measurements which have been carried out in our group using one method and the same solvent with a known water content. [33,34] The structures of the receptors are presented in Scheme 2.

Scheme 2. Studied small receptor molecules.



The following criteria explains why this set of experimental data was good to study with the COSMO-RS method:

- **The receptor molecules are small and structurally quite rigid.** This would reduce the overall number of conformers for the molecules and make the calculations faster.
- **The number of experimentally determined binding constant values is suitably large.** In order to be able to compare the results of the calculation and experiment better, a large number of experimental binding constants found with the same method in the same research group makes the comparison easier and more reliable.
- **The used receptor molecules are bidentate and bind using 1:1 stoichiometry.** When anion-receptor complexes form with a different stoichiometry than 1:1, the reaction formula and the dimension of the binding constant will change. For that reason, the binding constants of complexes with a different binding stoichiometry will not be comparable. It is important for the receptor to be bidentate as it enables anions with multiple anionic centers to bind better. One of the goals for the study was studying the binding of the glyphosate anion specifically, which can be bound from both sides bidentately.
- **The studied anion has a charge of -1.** Ionic charge is one of the main parameters that determines the strength of binding. For that reason, comparing ions of a different charge is not reasonable.
- **Water content of the solvents used in the experiment is known (0.5%).** Because even small trace level water content affects the properties of the solvent and the effectivity of the binding process by a large amount, it would be very helpful to know it and to be able to take it into account.

4.2. Large receptor molecules

The study of larger receptor molecules was carried out specifically for the dianions with the goal to find a suitable receptor molecule for the glyphosate dianion. Four different groups of molecules were studied in these calculations. The goal of the calculations was determining the suitable linker length, rigidity and making some specific suggestions on receptor molecules that could be used for binding the dianionic form of glyphosate.:

1. For investigating the effect of linker length on the receptor binding affinity, a number of bis-indolocarbazole receptors with linear alkyl chains with 1 to 10 carbon atoms (**1a – 1j**, Scheme 3). was studied

Indolocarbazole fragments were chosen as binding groups because they do not have HB acceptor sites and thus are unable to form intramolecular HBs (differently from, e.g., urea fragments). Formation of intramolecular HBs would complicate the interpretation of the results, because on receptor-anion complex formation the intramolecular HB would have to break.

The linkers have been connected to the binding groups in 2,2' positions. This position was chosen to try to ensure that there is a suitable distance between the binding groups of the receptor for the glyphosate anion to fit into. Other linker positions would create a steric hindrance to binding or make the glyphosate dianion unable to bind to both of the binding centers at the same time.

Alkyl linkers of length of 1 to 10 were studied to try and find the most suitable linker length. This group was used as a model system for determining the most suitable linker length for glyphosate complex formation with the bis-indolocarbazole receptors. Alkyl linkers are not easy to synthesize, have no special properties that could be useful to anion binding and have a rigidity that is rather limited by the preferred conformers of the linker. For that reason, some alternative linkers were studied in group **2**.

2. In group **2**, a set of molecules with varying linkers (**2a-2f**, Scheme 3) were analyzed as model systems to test the effects of varying the linker rigidities. Similarly to the receptors in group **1**, indolocarbazole fragments were used as the binding groups this set of receptors as well. The linkers were chosen for the ease of synthesis compared to simple alkyl linkers and to make it possible to look for differences in the binding strength depending on the dianion lengths and linker rigidities. The binding strengths of some of the molecules with these linkers could also be compared with experimental results in the future. The binding of 5 dianions (malonate, succinate, glutarate, glyphosate and adipate) was studied with the receptors of group **2** to test the effects of anion length to the binding strength of the complex as well.

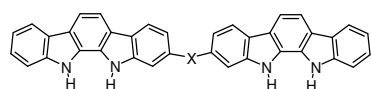
3. In group **3**, the possibility of using a different binding group for binding the phosphonate end of glyphosate was investigated. 1,1'-(1,2-Phenylene)bis[3-(2-aminophenyl)urea] was chosen as an alternative binding group for the following reasons. Firstly, it has more binding centers than the indolocarbazole fragment, which enables stronger interactions with the phosphonate end of the glyphosate dianion. The binding centers in the fragment are sterically

situated in such a way that they sterically fit around the phosphonate side of the dianion and can give up to 5-7 hydrogen bonds. While a large number of hydrogen bonds will individually be weaker due to less perfect distances and angles, they overall strength of the binding interaction will still increase.

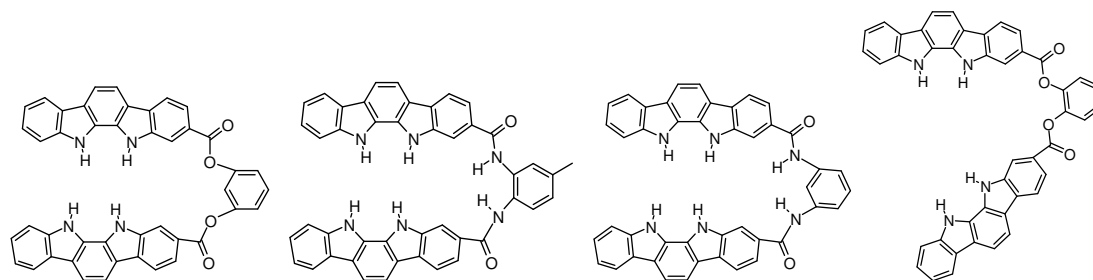
Simple alkyl chains were used as linker groups in this group of molecules in order to try to determine a suitable linker length for this type of receptor (**3a-3c**, Scheme 3).

4. In group **4**, a set of molecules (group **4**, Scheme 3) was chosen for research from an article [35]. Two more molecules of similar design were added to this group. These molecules were chosen for the study in hopes of finding a large molecule capable of binding glyphosate from both ends without the need to add a linker to the center of the molecule. The molecules vary by the length and type of the molecule fragments on the sides of the receptor. These fragments were varied in order to add additional binding groups (**4c**, **4d**) and to test the steric effects of the side functional groups to the binding strength of the glyphosate dianion.

Scheme 3. Studied large receptor molecules.



X= -CH₂- (**1a**), -(CH₂)₂- (**1b**), -(CH₂)₃- (**1c**), -(CH₂)₄- (**1d**), -(CH₂)₅- (**1e**), -(CH₂)₆- (**1f**), -(CH₂)₇- (**1g**), -(CH₂)₈- (**1h**), -(CH₂)₉- (**1i**), -(CH₂)₁₀- (**1j**)

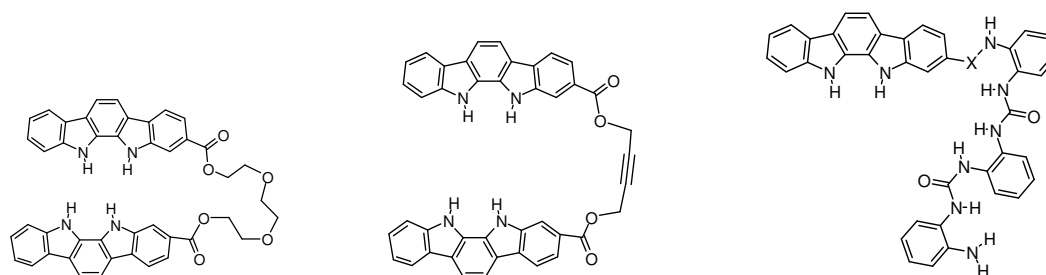


2a

2b

2c

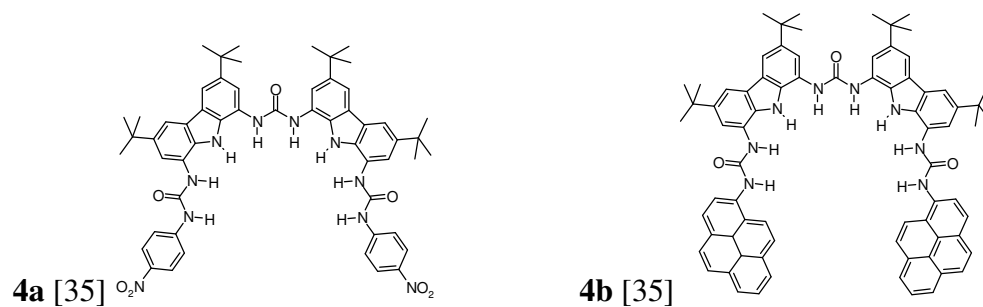
2d



2e

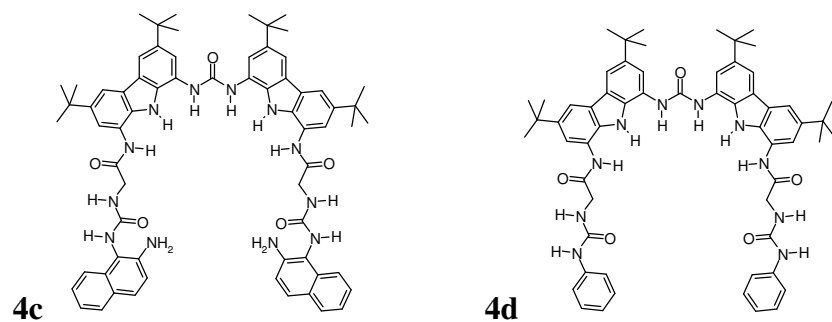
2f

3 X= -(CH₂)₃- (**3a**), -(CH₂)₄- (**3b**), -(CH₂)₅- (**3c**)



4a [35]

4b [35]



4c

4d

4.3. Studied solvents

Different parameters are used to characterize the properties of solvents, some of the most important ones being polarity, polarizability, acidity, basicity and the hydrogen bond donor and acceptor properties of a solvent. [36,37]

The polarity of a solvent characterizes the extent of separation of charges in the solvent molecules. It is characterized by the dielectric constant ϵ , the dipole moment of a solvent molecule μ and the Hildebrand parameter δ . The dielectric constant shows how much weaker the interaction between two charges is in the environment compared to vacuum. It characterizes the ability of a solvent to make electrolytical dissociation happen easier. The dipole moment characterizes the separation of opposite charges in the molecule. The Hildebrand parameter characterizes the work needed to overcome intramolecular forces during solvation. [36]

The polarizability of a solvent characterizes the deformation ability of its electron cloud of the solvent molecules. The refractive index n_D can be used to characterize the extent of polarizability of a molecule. The larger is the value of the parameter, the larger is the polarizability of a solvent. [36]

The basicity of a solvent characterizes the Brønsted and Lewis basicity and the hydrogen-bond acceptor properties of a solvent. The Lewis basicity can be characterized by the electron pair donor numbers (DN). [36]

The acidity of a solvent characterizes the Brønsted and Lewis acidities and the hydrogen bond donor properties of a solvent. The parameters that can be used to characterize the acidity of a solvent are the Kamlet-Taft α parameter and the electron pair acceptor numbers (AN). The electron pair acceptor numbers characterize the electron-pair acceptor capacity (the Lewis acidity) and the α parameter characterizes the hydrogen bond donicity of the solvent molecules. [36,37]

Additionally, the acidity and basicity of the environment are characterized by the autoprotolysis constant pK_{auto} . Autoprotolysis characterizes the ability of the solvent to dissociate into a proton and an anion of the solvent:



When the $pK_{\text{auto}} > 20$, the solvent is considered aprotic. [36]

Table 1. Some parameters of acetonitrile and water. [36,37,38] ^a

	ϵ	μ	δ	n_D	DN	α	AN	pK_{auto}
MeCN	35.94	3.95	24.3	1.344	14.1	0.19	18.9	33.3
H₂O	80.1	1.87	47.9	1.333	18	1.17	54.8	14

^a ϵ - dielectric constant, μ - dipole moment, δ - Hildebrand parameter, n_D - refractive index, DN - donor number, α - Kamlet-Taft α parameter, AN - acceptor number, pK_{auto} - autoprotolysis constant.

From the parameters shown in Table 1, the following can be said about the characteristics of acetonitrile and water as solvents:

- From the comparison of the dielectric constant and Hildebrand parameter values of water and MeCN, it can be seen that the polarity of water is much higher than the polarity of acetonitrile.
- Neither of the viewed solvents have a very high polarizability.
- It can be seen from the AN and DN numbers and the Kamlet-Taft α parameter that water has both higher acidity and higher basicity than MeCN
- As the pK_{auto} value is 33.3 for acetonitrile, the solvent can be considered aprotic. The pK_{auto} value for water is under 20 and the solvent can dissociate into a proton and a hydroxide anion.

From this comparison, it can be seen why even a small water content in the acetonitrile solvent can hinder the receptor-anion complex formation. Water molecules are small and polar. They form a hydrate layer around anions, which can sterically prevent the anions from binding with the receptor. Water molecules themselves can also be both hydrogen bond donors and acceptors and can thus prevent the hydrogen bond interaction between the receptor and anion by interacting with the anion or receptor on their own.

Organic solvents always contain a small amount of water. This amount can vary, leading to irreproducible solvent properties. Adding a small well-defined amount of water to an organic solvent will make the solvent composition better reproducible and is the standard approach used in receptor-anion binding studies.

The solvent used in this study was acetonitrile containing 0.5% of water. This is on one hand based on the above considerations. On the other hand, there are ample experimental data available with this solvent composition, which is useful for checking the suitability of the computational approach.

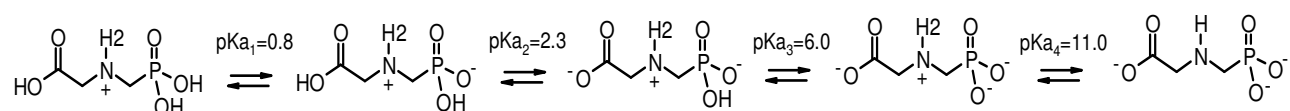
4.4. Studied anions

For the evaluation of suitability of the computational approach for receptor-anion binding studies, binding of acetate anion was studied and compared with experimental data.

For the receptors-dianion binding studies, the dianions of succinate, malonate, glutarate, glyphosate and adipate were studied.

Glyphosate (N-phosphonomethylglycine) is an amino acid that is widely used as an active ingredient in herbicides around the world (for example, under the commercial name Roundup) [4]. It dissociates in aqueous solution in three parts and can have the following forms at different pH values [39] as is seen on the following Scheme.

Scheme 4. Glyphosate forms and pK_a values in aqueous solution. [39]

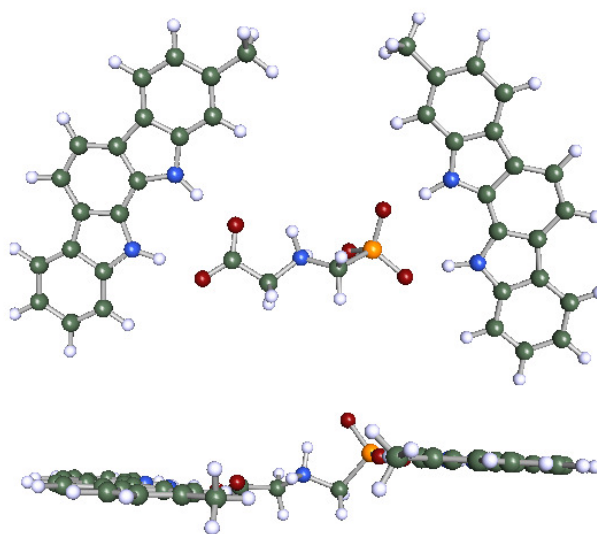


In a slightly basic or neutral environment, the majority of glyphosate in a solution can be found in a dianionic form with an overall charge of -2.

A host molecule that could bind glyphosate selectively would need to have the following characteristics.

To improve binding strength, glyphosate should be able to form a complex through several hydrogen bonds. The host molecule should have binding sites for both ends of the anion, preferably two or more hydrogen bond donor sites for both ends of glyphosate. Steric effects can largely improve or hinder binding. For that reason, the host should be able to achieve conformations that enable glyphosate to stay in its preferred conformation state. Similarly, to avoid a negative energetic effect in the case of binding, the free host molecule should have a preferred conformation similar to that in the host-guest complex.

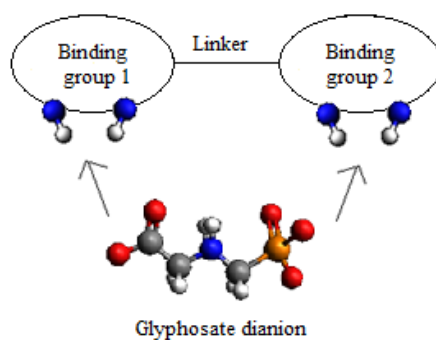
Scheme 5. Views of the preferred conformation of glyphosate in a receptor complex from two different angles.



The preferred conformation of the glyphosate dianion is shown on Scheme 5. The following points of interest can be noted about it:

- The anionic centers are not in the same plane with each other. This can be noted in the preferred conformations of the glyphosate-receptor complexes as well.
- The NH_2^+ hydrogens in the anion give intramolecular hydrogen bond interactions with the oxygen atoms of the anionic centers of the anion.
- Glyphosate in its preferred conformer is in a rather compact state, when the anion has to change its conformation in order to form a complex it will have a large negative effect on the binding strength.
- The anionic centers of glyphosate are similar to acetate and phosphate anions and both can be bound by a bidentate binding group.
- Indolocarbazole fragment suits well for the binding of the carboxylate end of glyphosate due to the binding group shape and rigidity.
- Alternatively, the phosphonate end of glyphosate could be bound by a larger binding group as it can give a larger number of hydrogen bond interactions than the carboxylate end.

Scheme 6. Design scheme for dianion receptors.



The goal of using the receptor design shown on Scheme 6 is to bind the glyphosate dianion by both of its anionic centers. Indolocarbazole would make a good bidentate binding group for the following reasons: it is small, structurally rigid, binds strongly with many bidentate anions and could be derivatized with functional groups to enhance the binding strength.

Linker structure will be important in the design of this kind of receptors. Too short or too rigid linker could mean that the receptor or the glyphosate dianion has a preferred conformation very different from one in the complex and that the change of conformation is also made harder for the receptor molecule. Too long and flexible linker could lead to a significant entropy penalty on complex formation and therefore the lowering of the binding strength of the complex.

5. Results and discussion

5.1. General information

In the following tables, the calculated binding constants are displayed. For the study of the small receptors, the computational results are also compared to the experimental [33,34] results. In order to better compare the results, the first table has two rows for calculated values. In the first row, there are the results obtained directly using the COSMO-RS method. In the second row, there are the adjusted values that have been calculated using Equation 6:

$$\log K_{\text{ass}}^{\text{adj}} = \log K_{\text{ass}}^{\text{calc}} - (\log K_{\text{ass}}^{\text{calc}} (\text{parent compound}) - \log K_{\text{ass}}^{\text{exp}} (\text{parent compound})) \quad (6)$$

where $\log K_{\text{ass}}^{\text{adj}}$ is the adjusted value of $\log K_{\text{ass}}^{\text{calc}}$ used to improve the comparability of experimental and calculated results, $\log K_{\text{ass}}^{\text{calc}}$ is the calculated value of the $\log K_{\text{ass}}$ and $\log K_{\text{ass}}^{\text{exp}}$ is an experimentally determined value of $\log K_{\text{ass}}$ taken from the published experimental data [33,34]. The reduced values have been found through the use of a parent compound of a receptor family. In this case, indolocarbazole was chosen as the parent compound. The reduced values are calculated on the basis of the difference of the calculated and the experimental values of the base compound. The reduced values are also not used for the larger receptors as there are no experimental values to carry out a comparison with.

A parameter was developed and calculated for expressing the steric strain involved with complex formation. The strain estimation parameter (SEP) was found from the energy differences of the optimized geometry of the glyphosate anion from a complex with a chosen base compound (receptor binding groups without a linker) and the optimized geometry of the glyphosate anion from a complex with a studied receptor. The difference was calculated using Equation 7:

$$\text{SEP} = E(\text{gly, complex}) - E(\text{gly, base}) \quad (7)$$

The strain estimation parameter characterizes how much the glyphosate anion is bent out of its preferred conformation while bound in a complex with a receptor. With a smaller strain estimation parameter, the negative energetic effects of glyphosate having to bind out of its preferred conformation are also smaller and the overall binding is stronger. It would be best to find a receptor where the strain estimation parameter is very low.

5.2. Indolocarbazole and urea based small receptor complexes with the acetate anion

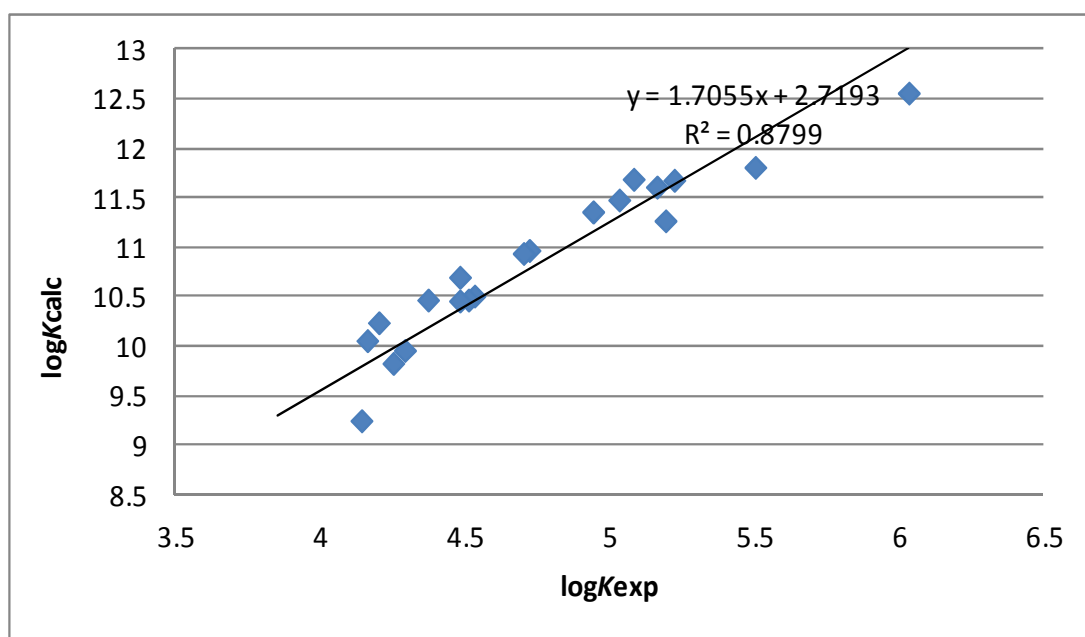
In order to evaluate the ability of the COSMO-RS method to predict the binding strength of small receptor complexes with anions the binding constants of a number of indolocarbazole- and urea-based receptors with the acetate anion were calculated. The calculated $\log K_{\text{ass}}$ values were compared with the experimental $\log K_{\text{ass}}$ values.

The results are presented in Table 2, The results are organized from largest to smallest, using the calculated data. It can be seen from Scheme 7 - Scheme 9 that while the calculated estimates of binding constants are significantly larger than the experimentally determined $\log K_{\text{ass}}$ values, trends are generally predicted correctly by the calculations. The regression analysis data for Scheme 7 - Scheme 9 is shown in Table 3.

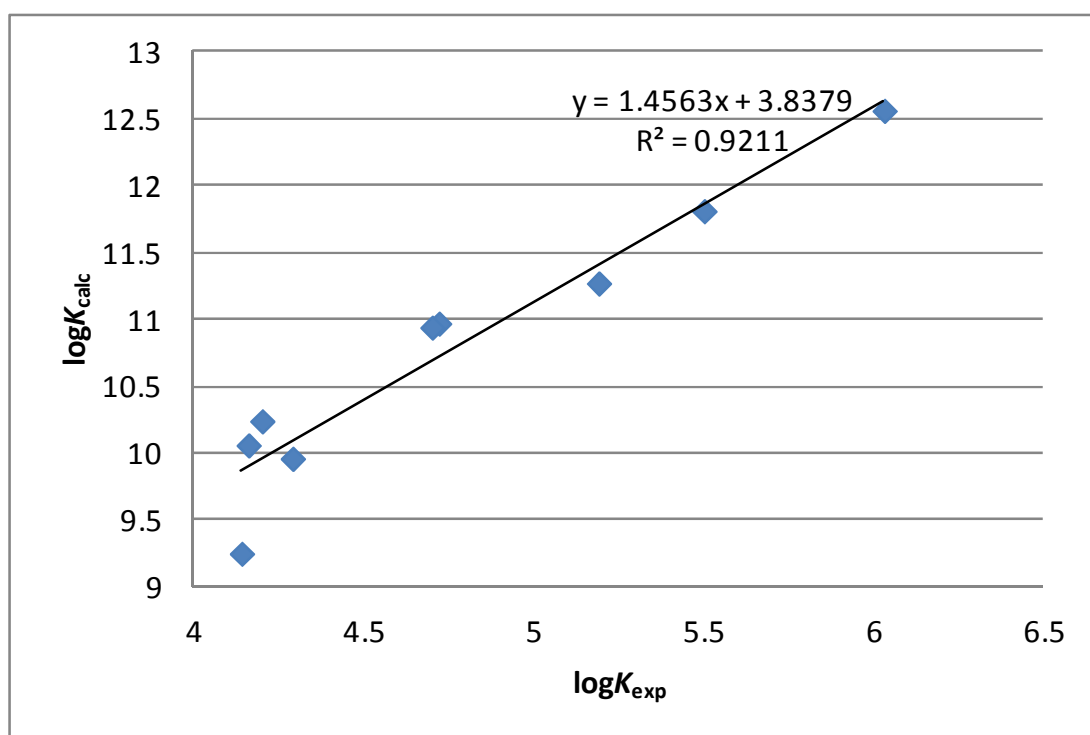
Table 2. Comparison of computational and experimental [33,34] binding constant values of indolocarbazole and urea-based receptor complexes with the acetate anion. Solvent: MeCN with 0.5% water.

Molecule	$\log K_{\text{ass}}^{\text{calc}}$	$\log K_{\text{ass}}^{\text{adj}}$	$\log K_{\text{ass}}^{\text{exp}}$
1,3-bis(4-nitrophenyl)urea	12.56	6.34	6.03
1,3-bis[(4-CF ₃ -phenyl)thiourea	11.81	5.59	5.5
2-NO ₂ -indolocarbazole	11.69	5.47	5.08
4-NO ₂ -indolocarbazole	11.68	5.46	5.22
4,7-Cl ₂ -indolocarbazole	11.61	5.39	5.16
2,7-Cl ₂ -indolocarbazole	11.48	5.26	5.03
2,9-Cl ₂ -indolocarbazole	11.36	5.14	4.94
3,4,4'-Cl ₃ -diphenylurea	11.27	5.05	5.19
1-(4-nitrophenyl)-3-hexylthiourea	10.97	4.75	4.72
1-(4-nitrophenyl)-2-thiourea	10.94	4.72	4.7
Indolocarbazole	10.7	4.48	4.48
2-MeO-indolocarbazole	10.51	4.29	4.53
4,7-(MeO) ₂ -indolocarbazole	10.47	4.25	4.51
5,6-dihydroindolocarbazole	10.47	4.25	4.37
2,7-(MeO) ₂ -indolocarbazole	10.46	4.24	4.48
(4-CF ₃ -phenyl)thiourea	10.24	4.02	4.2
1-(4-methoxycarbonylphenyl)-3-hexylthiourea	10.06	3.84	4.16
1,3-diphenylurea	9.96	3.74	4.29
1-Cl-indolocarbazole	9.83	3.61	4.25
N-(2,4,6-trichlorophenyl)thiourea	9.25	3.03	4.14
1,10-Cl ₂ -indolocarbazole	8.34	2.12	3.85

Scheme 7. Comparison of experimental and calculated results: All small receptors.



Scheme 8. Comparison of experimental and calculated results: Urea-based receptors.



Scheme 9. Comparison of experimental and calculated results: Indolocarbazole-based receptors.

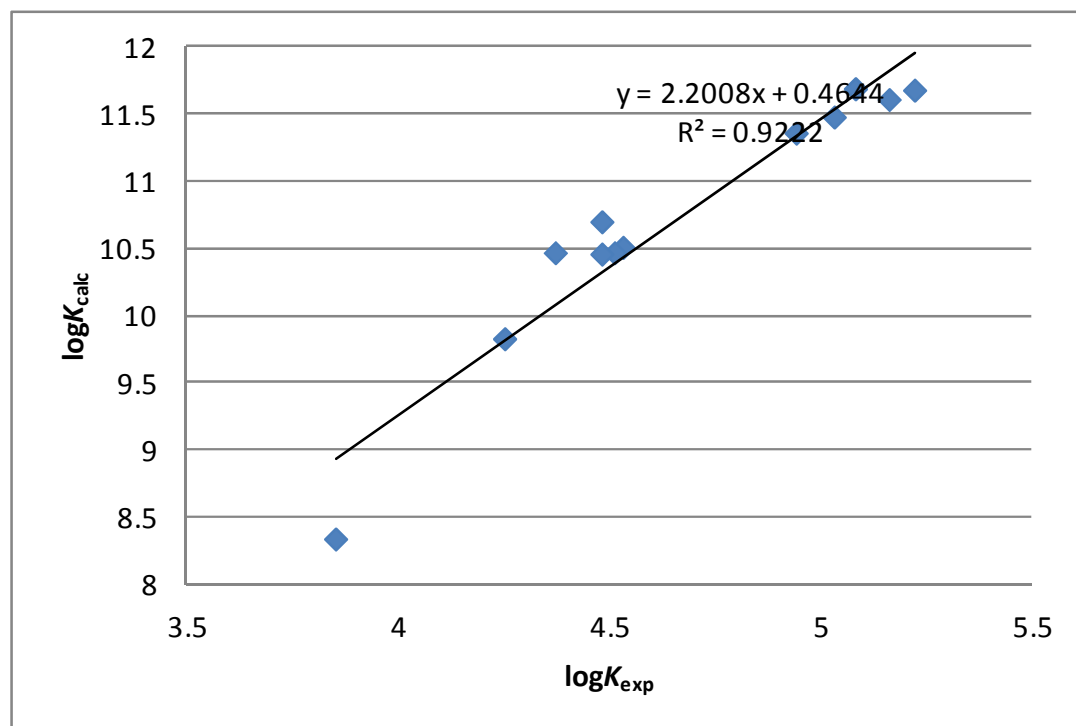


Table 3. Regression analysis of data of studied small receptors. ^a

	a	s(a)	b	s(b)	R²	s	N
All receptors	1.71	0.14	2.72	0.68	0.88	0.34	21
Urea-based receptors	1.46	0.16	3.84	0.77	0.92	0.31	9
Indolocarbazole-based receptors	2.20	0.20	0.46	0.95	0.92	0.28	12

^a a - slope; b - intercept; s(a) - standard deviation of the slope; s(b) - standard deviation of the intercept; R² - the correlation coefficient; s - the standard deviation of the linear regression; N - the number of data points.

From the regression analysis data, the following can be seen. The correlation coefficient for the data of all receptors is 0.88, which shows that a correlation between the experimental and calculated results exists. Compared with the correlation coefficients for receptor groups, it can be seen that the COSMO-RS method has a better correlation with the experiment when trends are compared between the molecules of the same class. This is also visible from the standard deviations of the regressions as the s values are smaller in the cases where the data of different receptor groups is analyzed separately. These results confirm that the COSMO-RS method is suitable for predicting trends in receptor-anion binding.

5.3. Binding of the glyphosate dianion by receptors 1a to 1j

It is expected that if the linker length is unsuitable for the size of the anion, then both the receptor molecule and the anion will have suboptimal conformations in the receptor-anion complex, invoking steric strain and decreasing the interaction intensity. These effects are often strong enough to notably affect the K_{ass} values.

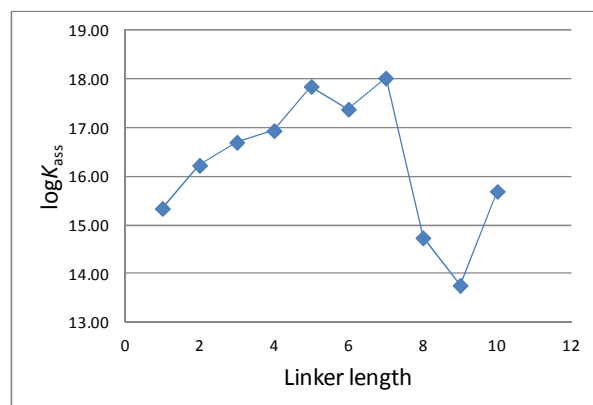
Table 4. Computational binding constants of receptors **1a** to **1j** with glyphosate dianion. Solvent: MeCN with 0.5% water.

	1a	1b	1c	1d	1e	1f	1g	1h	1i	1j
linker length	1	2	3	4	5	6	7	8	9	10
$\log K_{\text{ass}}^{\text{calc}}$	15.34	16.23	16.70	16.94	17.85	17.38	18.02	14.74	13.76	15.69
SEP	5.29	2.09	5.97	0.93	-0.34	0.28	-0.33	5.68	0.56	1.70

In Table 4, indolocarbazole-based receptors with alkyl linkers have been studied using MeCN solvent with 0.5% water content.

As is expected, variance in the length of the linker group in the receptor molecule has a strong effect on the glyphosate dianion binding strength but the relationship between binding strength and linker length is not simple.

Scheme 10. Effects of linker length to the binding strength of bis-indolocarbazole receptors and glyphosate dianion.



As can be seen from Scheme 10, with short linkers the binding constants are low and increase with the increase of the linker strength. The reason is that the shortest linkers do not enable the binding groups of the receptor to fit around the anionic centers of the glyphosate anion. The linker lengths C₄ – C₇ lead to the strongest binding and here an interesting “alternating” pattern is observed – the receptor molecules having linkers with odd number of carbon atoms

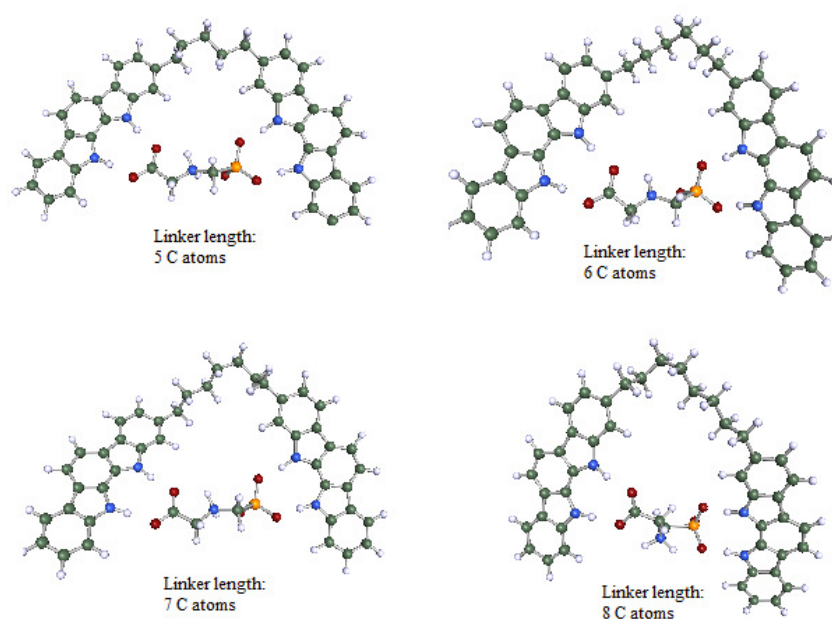
bind glyphosate stronger. The obvious reason is some extent of “pre-orientation” of the indolocarbazole fragments in the case of odd number of carbon atoms: when the linker length is an even number of carbon atoms, the alkyl chain can stay in its preferred conformation even during complex formation. When the linker has an odd number of carbon atoms, it will be in a bent conformation in the receptor-glyphosate complex. Due to the conformation of the linker, the binding groups will also be able to be in a conformation more suitable for the complex formation. Due to this, it will be easier for glyphosate to fit between the binding groups of the receptor. For this reason, the best linker lengths for the glyphosate anion are 7 and 5 and 6 has a lower binding strength.

The binding affinity drops sharply starting from the C₈ linker. Examining the glyphosate complex geometry reveals that this linker is unsuitable because the size difference of the glyphosate dianion and the linker forces either the anion or the linker to bend in several locations, which carried a negative energetic effect. Similar situation is observed with linker C₉ and C₁₀.

The results thus demonstrate that binding affinity is strongly dependent on linker length but the nature of the dependence is not simple.

Glyphosate conformation in the complex plays a very important role in determining the binding strength of the complex formation. When the linker is too short, the strain estimation parameter increases as the glyphosate anion cannot fit well between the binding groups. When the linker gets too long, its preferred conformation will push the binding groups too far apart for glyphosate to be able to bind as efficiently. The strain estimation parameters are the lowest on the case of C₅ and C₇ linkers and these receptors also have the highest binding constant values.

Scheme 11. Lowest energy conformers for receptors of linker length of 5,6,7 and 8 C atoms. All images for the lowest energy conformers of this receptor group are displayed in the Appendix.



5.4. Binding of the glyphosate, succinate, malonate, glutarate and adipate dianions by receptors **2a** to **2f**

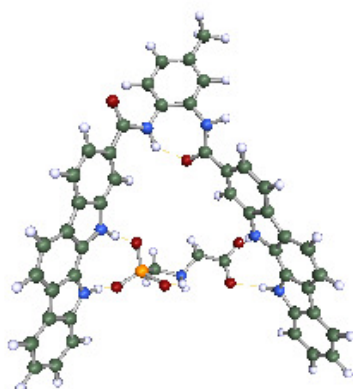
For the receptors of group 2, calculations were carried out with various dianions in order to see the changes in binding strength of anions of a similar type depending on the length of the anion. In Table 5, the logK_{ass} values of glyphosate complexes with group 2 receptor molecules are shown. In Table 6, the logK_{ass} values of all dianion complexes with group 2 receptors are included.

Table 5. Computational binding constants of receptors **2a** to **2f** with glyphosate dianion. Solvent: MeCN with 0.5% water.

	2a	2b	2c	2d	2e	2f
logK _{ass} ^{calc}	18.28	11.83	17.08	16.09	17.77	19.47

It can be seen that the range of binding affinities is almost eight orders of magnitude. Situation with **2b** the receptor with lowest binding affinity demonstrates the potential negative effect of intramolecular hydrogen bond formation in the receptor molecule. When a receptor molecule has an intramolecular hydrogen bond, it is likely that this hydrogen-bonded conformation is its preferred conformation. When this conformation is unsuitable for glyphosate binding, then the binding strength of the complex lowers considerably as there is a large effect of either breaking the hydrogen bond or of the glyphosate dianion having to bend far from its preferred conformation. From Scheme 12, the molecule **2b** complex with glyphosate can be seen. In the preferred conformation, the intermolecular hydrogen bond does not break but causes the receptor to stay in a conformation where the glyphosate anion cannot efficiently fit between the binding groups of the receptor.

Scheme 12. Receptor **2b** complex with the glyphosate dianion.



The binding affinities of the remaining receptors of the series **2** range over 3 orders of magnitude and neither too short (**a**, **c**, **d**) nor too long linkers (**e**) do well. The strongest binding is displayed by **2f**, which apparently has a linker making optimal spacing between the indolocarbazole fragments.

Table 6. Computational binding constants of bis-indolocarbazole receptors **2a** to **2f** with malonate, succinate, glutarate, glyphosate and adipate dianions. Solvent: MeCN with 0.5% water.

	2a	2b	2c	2d	2e	2f
Malonate	18.99	17.99	15.41	21.72	17.52	22.43
Succinate	17.91	16.81	15.43	19.93	18.37	20.90
Glutarate	11.11	12.79	15.87	11.73	19.63	20.06
Glyphosate	18.28	11.83	17.08	16.09	17.77	19.47
Adipate	12.57	18.84	17.83	19.48	18.98	19.49

As can be seen from Table 6, the binding constants vary noticeably depending on the anion size (alkyl chain length). Receptor **2e** has a long and flexible linker and it could be expected that it can bind the larger anions better. Receptor **2b** has a very short effective linker due to the intermolecular hydrogen bond and it is expected to bind smaller anions better.

Table 6 reveals, however, that these expectations are not fully met by calculation results. On one hand, **2b** binds adipate best, on the other hand **2e** bind glutarate best. It turns out that the adipate anion is flexible enough to bend and can often bind efficiently with the smaller receptors. Receptor **2e** has a preferred conformation in a complex where the distance between the linking groups is too short for adipate and glyphosate to bind efficiently. Receptor **2d** is similar to **2b** in the linker shape, however it cannot give an intermolecular hydrogen bond with itself. For this receptor, the smallest anion complexes have the highest binding strength. Receptors **2a** and **2c** are similar in linker construction, but **2a** has the strongest affinity towards malonate, while **2c** binds adipate strongest. Receptor **2f** has a less flexible linker and it is more suitable for the smaller anions. However, compared to other receptors, the longer anions have higher complex binding strengths with **2f** as well. Receptors **2c** and **2f** are the only ones where a clear-cut relationship is observed between anion size and binding affinity: **2c** binds larger anions better, **2f** binds smaller anions better.

5.5. Binding of the glyphosate dianion by receptors **3a** to **3c**

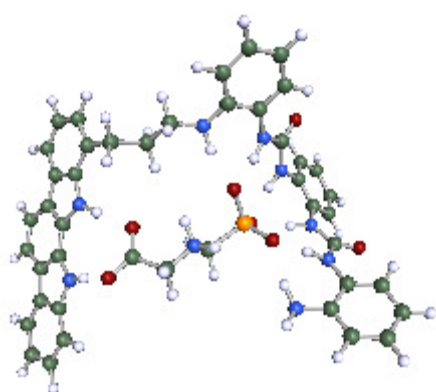
A receptor molecule for glyphosate could use alternative binding groups to indolocarbazole fragments as well. The urea fragment used in this group of receptors was specifically chosen to try to bind the phosphonate end of the glyphosate dianion stronger by creating more hydrogen bond interactions between the anion and the binding group of the receptor.

Table 7. Computational binding constants of receptors **3a** to **3c** with glyphosate dianion. Solvent: MeCN with 0.5% water.

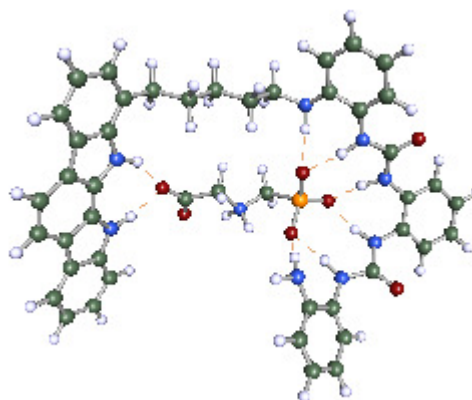
	3a	3b	3c
$\log K_{\text{ass}}^{\text{calc}}$	23.10	22.66	21.52

Changing one of the indolocarbazole groups of the receptor to a group more suitable for the phosphonate end of glyphosate had a positive effect on the binding strength. The 1,1'-(1,2-Phenylene)bis[3-(2-aminophenyl)urea] binding group sterically fits with the phosphonate group and can also give more hydrogen bond interactions with it than the indolocarbazole fragment could (see Scheme 13). The problem with these molecules however was finding a suitable linker length and rigidity again to avoid negative energetic effects due to steric blocking. From the calculations, the most suitable linker length for this type of receptor would be 3 carbon atoms. From the preferred conformer of the complex between receptor **3a** and the glyphosate dianion (shown on Scheme 13), it can be seen that the indolocarbazole fragment in the complex is planar, the alkyl chain is in its preferred conformation, the glyphosate anion does not need to bend itself for complex formation and the urea group is also in a sterically favourable conformation. Receptor **3c** (shown on Scheme 14) on the other hand can only have interactions with one of the oxygen atoms of the carboxylate fragment of the glyphosate anion, which lowers the overall binding strength considerably.

Scheme 13. Receptor **3a** complex with glyphosate.



Scheme 14. Receptor **3c** complex with glyphosate.



5.6. Binding of the glyphosate dianion by receptors 4a – 4d

Some larger receptor structures were selected to be studied for glyphosate binding strength as well. Two of these structures are taken from the literature [35], the other two are designed with a similar structure.

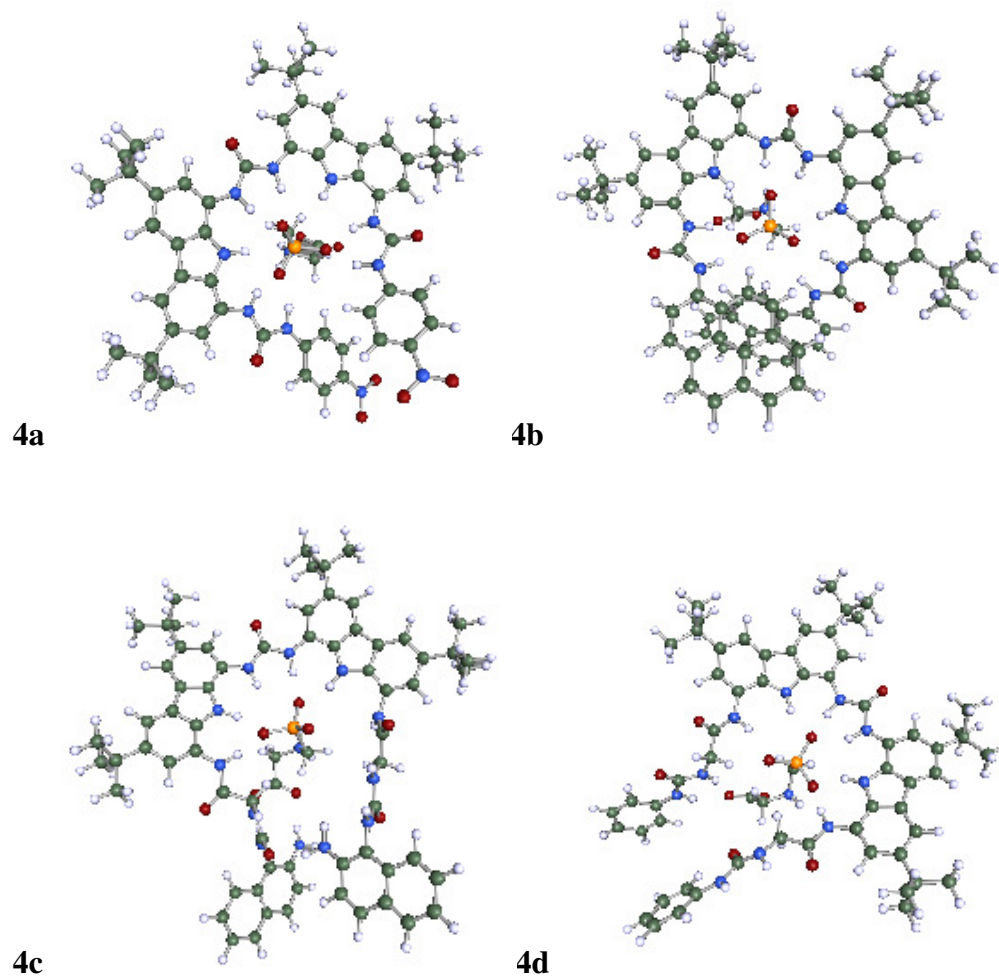
Table 8. Computational binding constants of receptors from reference [35] with glyphosate dianion. Solvent: MeCN with 0.5% water.

	4a	4b	4c	4d
$\log K_{\text{ass}}^{\text{calc}}$	23.12	23.69	16.07	21.57

The receptors in this group are all large and can give a number of hydrogen bonds with the bound anion (see Scheme 15). Receptors **4a** and **4b** have a preferred conformation where the glyphosate dianion is bound by the phosphonate end only. The first two receptors also have the highest binding constant values. There are two causes of this: (1) a larger number of potential hydrogen bond interactions are formed between the phosphonate group and the receptor binding groups and (2) The distance between the binding groups of the receptor is more suitable for the anionic centers of the phosphonate end than for the entire glyphosate dianion.

When glyphosate is bound by the phosphonate end, 7-8 hydrogen bond interactions can be formed, one interaction less can form in the case of glyphosate being bound from both of its anionic centers. Only binding with the phosphonate end also allows the glyphosate anion to stay in a sterically preferred conformation while binding with both sides causes changes in the preferred conformation and carries a negative energetic effect that will lower the binding strength.

Scheme 15. Lowest energy conformers of complexes between glyphosate dianion and receptors of group 4.

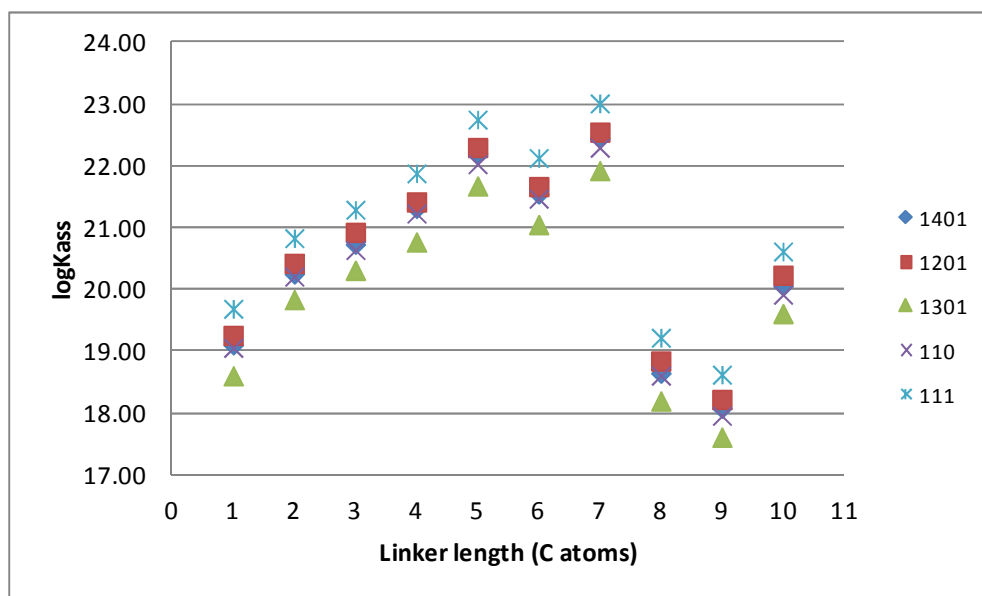


5.7. Effects of the choice of parametrization

The results of the estimates of binding constants using the COSMOtherm program depend on the used parametrization. Five different parametrizations were studied: BP_TZVP_C21_110 [40], BP_TZVP_C21_0111 [41], BP_TZVP_C30_1201 [42], BP_TZVP_C30_1301 [43] and BP_TZVP_C30_1401 [26]. The comparison was made by calculating the binding constants of the complexes of bis-indolocarbazole receptors with alkyl linkers and the glyphosate dianion.

There have been a number of changes relevant to binding constant calculations made between the different parametrization versions of the COSMOtherm software. The most important implemented changes that concern the binding constant calculations are the improvement of the sigma profile parameter and the changes in the calculation of conformer weights through the updates. [44-48]

Scheme 16. The effects of the choice of parametrization on the calculation results. Calculation example: Receptor **1a-1j** complexes with glyphosate dianion. Solvent: MeCN with 0.5% water.

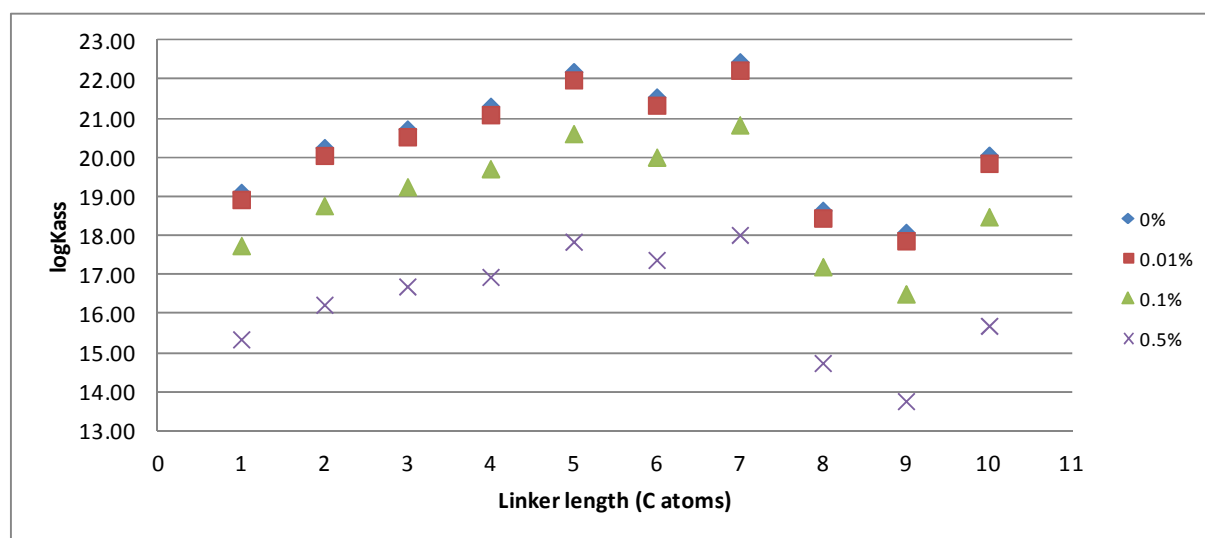


Scheme 16 displays the effects of chosen parameterization on the calculation results of receptors **1a** to **1j**. The binding constant values calculated using different parameterizations are shown in different colors. Comparing the results obtained with 5 different parameterizations, it can be seen that while the parameterization does have an effect on the absolute calculated binding constant values, it does not have an effect on the trends between binding constants. Unfortunately there is no parameterization that would lower the $\log K_{\text{ass}}$ values notably closer to the experimental values. In this study, the newest parameterization, BP_TZVP_C30_1401 was used for the calculations.

5.8. Effects of the water content

As explained above, molecular interactions in aprotic solvent environments are largely affected by even a small water content. Water molecules solvate anions intensely, first of all by strong hydrogen bonds. This hinders complex formation, because in order to bind with the receptor, the anion has to be desolvated. For these reasons, even a small water content in a solution will lower the binding constants of receptor-anion complex formation. [39, 41] The comparison of the effects of water content to the binding constants is carried out with the binding constants of the complexes of receptors **1a** to **1j** and the glyphosate dianion.

Scheme 17. The effect of water content on binding constant values. Receptor **1a** to **1j** complexes with the glyphosate dianion. Solvent: MeCN with 0%; 0.01%; 0.1% and 0.5% water.



The comparison of the effects of water content of the binding constant values can be seen from Scheme 17. Even a small water content has an effect on the binding strengths. Water content in a solution lowers the binding strengths of the complexes. Changing water content in the solvent from 0 to 0.5% leads to decrease of the binding constant by up to an order of magnitude. However, it is visible from the graph that the trends between binding strengths do not change noticeably in the used group of receptor-anion binding constants. In this study, a water content of 0.5% was chosen as that is an amount that will take into account the effects of water content of a solution and a solution with a water content of 0.5% could also be used in experiments.

6. Summary

Astrid Pung

Computational estimation of receptor-anion binding in solution

The purpose of this study was to investigate the possibilities of predicting the values of association constants between small receptor molecules and anions using the COSMO-RS computational method and also to try to find promising receptor molecule structures that could be used as parent structures for developing a sensitive molecular receptor for the glyphosate dianion.

For the calculations, geometry optimization was carried out using Turbomole v 6.4 and the calculation of free energies of studied particles in a chosen solvent environment was done using the program COSMOthermX version C3.0 revision 14.01 parametrization BP_TZVP_C30_1401. Association constants were calculated from the free energies of the particles using the equation $\Delta G_{\text{ass}} = -RT \ln K_{\text{ass}}$.

For evaluating the suitability of COSMO-RS for this work, calculations were carried out in order to reproduce the association constant data available from the literature. Receptor molecules from the classes of indolocarbazoles and ureas were investigated.

For glyphosate, a number of potential receptor molecules consisting of indolocarbazole and urea fragments were constructed and some structures were obtained from literature. Calculation of the association constants was carried out in a similar manner as with the smaller molecules.

From the study of small receptors and comparison of experimental and computational data, the following can be concluded:

1. The COSMO-RS method cannot predict the absolute values of binding constants as the predicted values are systematically by a number of log units higher than the experimental values.
2. For comparing the trends between binding constants of different receptors, the method, on the example of acetate anion, works well.
3. Computations are a useful tool for comparison of the optimized geometries and using them to analyze the steric effects on binding strengths.

For the design of a receptor for the glyphosate dianion, the following can be concluded:

1. For bis-indolocarbazole receptors with alkyl linkers, the optimal linker length is 7 C atoms, followed by the receptor with 5 C atoms.

2. The linker of a receptor must be of a suitable length and rigidity for the glyphosate anion to be able to fit between the binding groups of the receptor.
3. The preferred conformations of the free receptor molecule should be as close as possible to the conformation of the receptor in the complex. When the free receptor has a conformation very different to that of the complex, negative energetic effects will be caused by complex formation and the binding strength will decrease. The preferred conformations of the receptor depend on multiple factors such as the preferred conformations of the linker and binding groups, the rigidity of the linker and binding groups and the ability of the receptor molecule to form intermolecular hydrogen bonds.
4. For designing a receptor molecule for the glyphosate anion, using different binding groups for the carboxylate and phosphonate end of glyphosate can also improve the binding strength. Using a larger binding group for the phosphonate end of glyphosate and choosing a receptor that can give a larger number of hydrogen bond interactions with the phosphonate group can greatly improve the binding strength of the complex.

Using larger molecules for glyphosate binding is also a possibility. Due to a large number of potential hydrogen bond interactions, two of the studied large receptors gave the highest estimated binding constants of the study. However, computational study of large receptor molecules is not easy.

7. Kokkuvõte

Astrid Pung

Arvutuslik retseptor-anioon seondumise uurimine lahusekeskkonnas

Käesoleva uurimistöö eesmärgiks oli uurida COSMO-RS arvutusmeetodi suutlikkust väikeste retseptormolekulide ja anioonide vaheliste sidumiskonstantide väärtuste ennustamisel ning proovida leida selliseid retseptormolekulide struktuure, mida oleks võimalik kasutada baasstruktuuridena glüfosaadi dianiooni suhtes tundliku retseptormolekuli loomiseks.

Arvutustes viidi geomeetria optimeerimine läbi programmiga Turbomole v 6.4. Uuritud osakeste vabaenergiate väärtused valitud lahusti segu keskkonnas leiti kasutades arvutusprogrammi COSMOthermX versiooni C3.0 revisjoni 14.01 parametrisatsiooni BP_TZVP_C30_1401. Sidumiskonstantide väärtused leiti vabaenergiatest võrrandi $\Delta G_{\text{ass}} = -RT \ln K_{\text{ass}}$ kaudu.

Hindamaks COSMO-RS meetodi sobivust selleks tööks, üritati arvutustega reprodutseerida kirjandusest pärinevaid eksperimentaalseid sidumiskonstantide väärtuseid. Töös uuriti retseptormolekule indolokarbasoolide ning ureate aineklassidest.

Glüfosaadi jaoks disainiti indolokarbasoolide ning ureate fragmentidest mitmeid retseptormolekule, samuti kasutati mõningaid retseptormolekule kirjandusest. Sidumiskonstantide väärtusi prooviti arvutada sama meetodiga, mida kasutati väiksemate molekulide korral.

Väikeste molekulide arvutuslike- ja eksperimentaalsete tulemuste võrdlusest saab välja tuua järgmist:

1. COSMO-RS meetod ei suuda ennustada sidumiskonstantide absoluutseid väärtusi - ennustatud väärtused on süstemaatiliselt mitmete suurusjärkude võrra eksperimentaalselt leitud väärtustest kõrgemad.
2. Erinevate retseptorite seondumiskonstantide vaheliste trendide võrdlemiseks töötab meetod atsetaadi aniooni näite korral hästi.
3. Arvutused on kasulikud optimeeritud geomeetria võrdlemiseks ning neid saab kasutada selleks, et uurida steeriliste efektide mõju seondumise tugevusele.

Glüfosaadi dianiooni jaoks sobiva retseptormolekuli disainimisel on vajalik arvestada järgmist:

1. Alküllinkeriga indolokarbasooli fragmentidel põhineva retseptormolekuli jaoks on optimaalseks linkeri ahela pikkuseks 7 C aatomit, järgmisena 5 C aatomit.

2. Retseptori linkerrühma pikkus ja jäikus peab olema sobiv glüfosaadi aniooni mahtumiseks retseptori siduvate rühmade vahele.
3. Vaba retseptormolekuli eelistatud konformatsioonid peaksid olema võimalikult sarnased retseptori konformatsioonile kompleksis. Kui vaba retseptori konformatsioon on kompleksi omast väga erinev, põhjustab kompleksi moodustumine negatiivseid energeetilisi efekte ning seondumise tugevus väheneb. Retseptori eelistatud konformatsioonid sõltuvad mitmest faktorist nagu linkerrühma ja siduvate rühmade eelistatud konformatsioonid, nende jäikus ja suurus ning retseptormolekuli võime moodustada molekulisisesid vesiniksidemeid.
4. Glüfosaadi aniooni jaoks sobiva retseptormolekuli disainimisel võib erinevate siduvate rühmade kasutamine glüfosaadi dianiooni karboksülaat- ja fosfonaatoste jaoks samuti suurendada seondumise tugevust. Fosfonaatrühma sidumisel suurema siduva rühma kasutamine ning fosfonaatrühmaga suuremat hulka vesiniksideid moodustada saava retseptori valimine võib märgatavalt suurendada kompleksi seondumise tugevust.

Samuti on võimalik kasutada glüfosaadi sidumiseks suuremaid molekule. Suure hulga vesiniksidemete moodustamise võime tõttu on kahel töös uuritud suurel retseptormolekulil selles töös uuritud molekulidest kõige kõrgemad sidumiskonstantide väärtused. Samas on suurte retseptormolekulide puhul keeruline nende arvutuslik uurimine.

8. References

1. Steed, J. W., Turner D. R., Wallace K. J. *Core concepts in supramolecular chemistry and nanochemistry*, 1st ed.; Wiley: Chichester, England, 2007.
2. *Analytical methods in supramolecular chemistry*; Schalley, C. A., 2nd ed.; Wiley: Germany, 2012; Vol. 1, pp.2-12; 27-28.
3. Klamt, A. *COSMO-RS : from quantum chemistry to fluid phase thermodynamics and drug design*, Elsevier: Amsterdam, 2005.
4. Franz, J. E.; Mao, M. K.; Sikorski J. A. *Glyphosate: a unique global herbicide*, ACS: Washington DC, USA, 1997.
5. Anslyn, E. V. Supramolecular Analytical Chemistry. *J. Org. Chem.* **2007**, 72 (3), 687-699.
6. Beer, P. D.; Gale, P. A. Anion Recognition and Sensing: The State of the Art and Future Perspectives. *Angew. Chem. Int. Ed.* **2001**, 40 (3), 486-516.
7. Gale, P. A. Anion and ion-pair receptor chemistry: highlights from 2000 and 2001. *Coord. Chem. Rev.* **2003**, 240, 191–221.
8. Schneider, H.-J. Binding Mechanisms in Supramolecular Complexes. *Angew. Chem. Int. Ed.* **2009**, 48 (22), 3924 – 3977.
9. Safont-Sempere, M. M.; Fernández, G.; Würthner, F. Self-Sorting Phenomena in Complex Supramolecular Systems *Chem. Rev.* **2011**, 111 (9), 5784–5814.
10. Cram, D. J. Preorganization – from solvents to spherands. *Angew. Chem. Int. Ed. Engl.* **1986**, 25, 1039–1134.
11. Grabowski, S. J. *Hydrogen bonding - new insights*. Springer: Dordrecht, 2006.
12. Jeffrey, G. A.; Saenger, W. *Hydrogen bonding in biological structures*. Springer-Verlag: Berlin, 1994.
13. Hancock, R. D. Chelate ring size and metal ion selection. The basis of selectivity for metal ions in open-chain ligands and macrocycles. *J. Chem. Edu.* **1992**, 69 (8), 615–621.
14. *Encyclopedia of supramolecular chemistry*; Atwood J. L.; Steed J. W., Ed.; Marcel Dekker: New York, USA, 2004; Vol.1, pp.658-665.
15. Anslyn, E. V.; Dougherty, D. A. *Modern Physical Organic Chemistry*. University Science Books: USA, 2006.
16. Skoog, D. A.; West, D. M.; Holler, J.F. *Fundamentals of analytical chemistry*. 6th ed.; Saunders college publishing : USA, 1992.

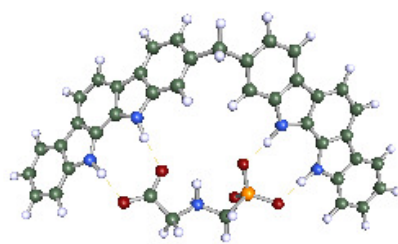
17. Zyryanov, G. V.; Palacios, M. A.; Anzenbacher, P. Jr. Rational Design of a Fluorescence-Turn-On Sensor Array for Phosphates in Blood Serum. *Angew. Chem. Int.* **2007**, 46, 7849–7852.
18. P. W. Atkins, *Quanta: A Handbook of Concepts*, 2nd ed.; Oxford University Press: Oxford, 1991.
19. Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation model. *Chem. Rev.* **2005**, 105 (8), 2999-3093.
20. 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, 101-122.
21. Klamt, A.; Schüürmann, G. COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perkin Trans.* **1993**, 2, 799-805.
22. Tshepelevitsch, S.; Oss, M.; Pung, A.; Leito; I. Evaluating the COSMO-RS Method for Modelling Hydrogen Bonding in Solution. *ChemPhysChem.* **2013**, 14 (9), 1909 – 1919.
23. Ahlrichs, R.; Armbruster, M. K.; Bachorz, R.A.; Bar, M.; Baron, H.-P.; Bauernschmitt, R.; Bischoff, F. A.; Bocker, S.; Crawford, N.; Deglmann, P.; Ehrig, M.; Eichkorn, K.; Elliott, S.; Furche, F.; Glob, A.; Haase, F.; Haser, M.; Horn, H.; Hattig, C.; Hellweg, A.; Hofener, S.; Huber, C.; Huniar, U.; Kattannek, M.; Klopper, W.; Kohn, A.; Kolmel, C.; Kollwitz, M.; May, K.; Nava, P.; Ochsenfeld, C.; Ohm, H.; Patzelt, H.; Rappoport, D.; Rubner, O.; Schafer, A.; Schneider, U.; Sierka, M.; Tew, D. P.; Treutler, O.; Unterreiner, B.; Von Arnim, M.; Weigend, F.; Weis, P.; Weiss, H.; Winter, N. Turbomole Version 6.4, 2012.
24. A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, **1988**, 38, 3098-3100.
25. J. P. Perdew. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B*, **1986**, 33, 8822-8824.
26. Eckert, F.; Klamt, A. COSMOtherm, Version C3.0, Release 14.01, COSMOlogic GmbH&CoKG, Leverkusen, Germany, 2014; <http://www.cosmologic.de/>.
27. Curiel, D.; Cowley, A.; Beer, P. D. Indolocarbazoles: a new family of anion sensors. *Chem. Commun.* **2005**, 236–238.

28. Wang, T.; Wang, H.-F.; Yan, X.-P. Fabrication of anion complexes from 5,6-dihydrodiindolo[3,2-a:2',3'-c]phenazine as a building block. *CrystEngComm*. **2010**, *12*, 3177–3182.
29. Courtemanche, R. J. M.; Pinter, T.; Hof, F. Just add tetrazole: 5-(2-Pyrrolo)tetrazoles are simple, highly potent anion recognition elements. *Chem. Commun.* **2011**, *47*, 12688–12690.
30. Boiocchi, M.; Del Boca, L.; Gomez, D. E.; Fabbrizzi, L.; Licchelli, M.; Monzani, E. Nature of Urea-Fluoride Interaction: Incipient and Definitive Proton Transfer. *J. Am. Chem. Soc.* **2004**, *126* (50), 16507-16514.
31. Boiocchi, M.; Del Boca, L.; Esteban-Gomez, D.; Fabbrizzi, L.; Licchelli, M.; Monzani, E. Anion-Induced Urea Deprotonation. *Chem. Eur. J.* **2005**, *11* (10), 3097 – 3104.
32. Caltagirone, C.; Gale, P. A.; Hiscock, J. R.; Brooks, S. J.; Hursthouse, M. P.; Light, M. E. 1,3-Diindolylureas: high affinity dihydrogen phosphate receptors. *Chem. Commun.* **2008**, *26*, 3007–3009.
33. Haav, K.; Kadam, S. A.; Toom, L.; Gale, P. A.; Busschaert, N.; Wenzel, M.; Hiscock, J. R.; Kirby, I. L.; Haljasorg, T.; Lõkov, M.; Leito, I. Accurate Method To Quantify Binding in Supramolecular Chemistry. *J. Org. Chem.* **2013**, *78* (16), 7796–7808.
34. Kadam, S. A.; Haav, K.; Toom, L.; Haljasorg, T.; Leito, I. NMR Method for Simultaneous Host-Guest Binding Constant Measurement. *J. Org. Chem.* **2014**, *79* (6), 2501–2513.
35. Sanchez, G.; Espinosa, A.; Curiel, D.; Tarraga, A.; Molina, P. Bis(carbazolyl)ureas as Selective Receptors for the Recognition of Hydrogenpyrophosphate in Aqueous Media. *J. Org. Chem.* **2013**, *78* (19), 9725–9737.
36. Reichardt, C. *Solvents and Solvent effects in Organic Chemistry*, 3rd. ed.; VCH: Weinheim, Germany, 2003.
37. Abboud, J.-L. M.; Notario, R. Critical Compilation of Scales of Solvent Parameters. Part I. Pure, Non-Hydrogen Bond Donor Solvents. *Pure Appl. Chem.* **1999**, *71* (4), 645–718.
38. Stenutz, R. Tables for chemistry; <http://www.stenutz.eu/chem/> last updated on 15.03.2013.
39. Chen, Z.; He, W.; Beer, M.; Megharaj, M.; Naidu, R. Speciation of glyphosate, phosphate and aminomethylphosphonic acid in soil extracts by ion chromatography

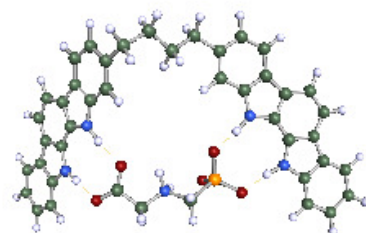
with inductively coupled plasma mass spectrometry with an octopole reaction system. *Talanta*. **2009**, 78, (3), 852–856.

40. Eckert, F.; Klamt, A. COSMOtherm, Version C2.1, Revision 01.10; COSMOlogic GmbH&CoKG: Leverkusen, Germany, 2006; <http://www.cosmologic.de/>.
41. Eckert, F.; Klamt, A. COSMOtherm, Version C2.1, Revision 01.11; COSMOlogic GmbH&CoKG: Leverkusen, Germany, 2011; <http://www.cosmologic.de/>.
42. Eckert, F.; Klamt, A. COSMOtherm, Version C3.0, Revision 12.01; COSMOlogic GmbH&CoKG: Leverkusen, Germany, 2012; <http://www.cosmologic.de/>.
43. Eckert, F.; Klamt, A. COSMOtherm, Version C3.0, Revision 13.01; COSMOlogic GmbH&CoKG: Leverkusen, Germany, 2013; <http://www.cosmologic.de/>.
44. Klamt, A. COSMOtherm C21_0110 Release Notes. Nov. 2009.
45. Klamt, A. COSMOtherm C21_0111 Release Notes. Dec. 2010.
46. Klamt, A. COSMOtherm C30_1201 Release Notes. Dec. 2011.
47. Klamt, A. COSMOtherm C30_1301 Release Notes. Dec. 2012.
48. Klamt, A. COSMOtherm C30_1401 Release Notes. Dec. 2013.

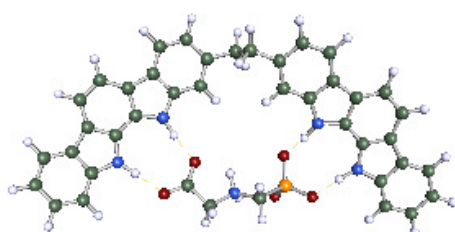
9. Appendix. Lowest energy conformers for receptors 1a - 1j.



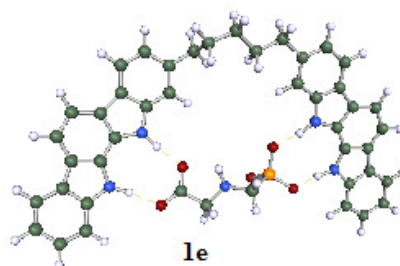
1a



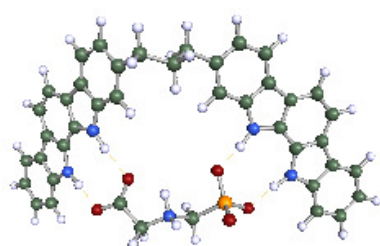
1d



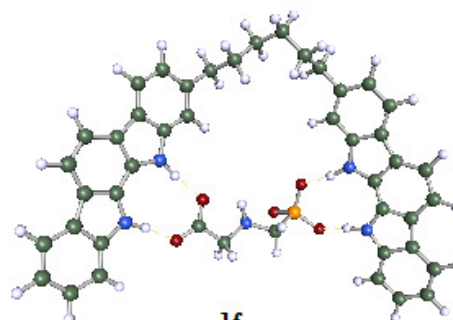
1b



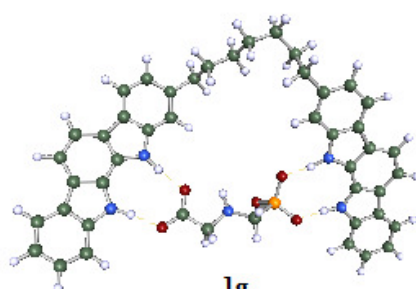
1e



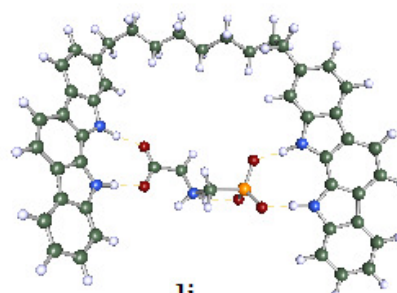
1c



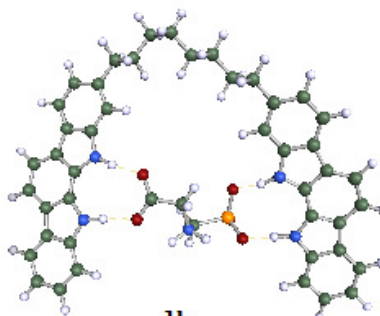
1f



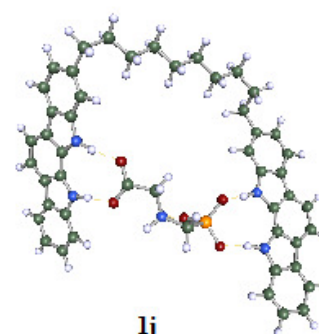
1g



1i



1h



1j

Lihlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina,

Astrid Pung,

(autori nimi)

1. annan Tartu Ülikoolile tasuta loa (lihlitsentsi) enda loodud teose

Computational estimation of receptor-anion binding in solution,

(lõputöö pealkiri)

mille juhendaja on

Ivo Leito,

(juhendaja nimi)

1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2.üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu alates **09.06.2017** kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, **02.06.2014**