Prediction of the interaction strength of an urea-based probe towards ions in water by means of DFT/PCM calculations

Robert Benda, Thomas Vezin, Bérengère Lebental

Prediction of the interaction strength of an urea-based probe towards ions in water by means of DFT/PCM calculations

Robert Benda*
LPICM, CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, Route De Saclay, 91128 Palaiseau, France and CERMICS, Ecole des Ponts and INRIA Paris, 6-8 avenue Blaise Pascal 77455 Marne-la-Vallée, France

Thomas Vezin and Gaël Zucchi
LPICM, CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, Route De Saclay, 91128 Palaiseau, France

Eric Cancès
CERMICS, Ecole des Ponts and INRIA Paris, 6-8 avenue Blaise Pascal 77455 Marne-la-Vallée, France

Bérengère Lebental
LPICM, CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, Route De Saclay, 91128 Palaiseau, France and COSYS-LISIS, Université Gustave Eiffel, IFSTTAR, F-77454 Marne-la-Vallée, France.

We study numerically, by means of DFT calculations complemented with an implicit solvation model, a novel chemical probe bearing urea and aromatic phenyl groups. We probe the interaction in water of the latter with a wide variety of ions relevant to water quality. We perform geometry minimizations using PBE0 functional and aug-cc-pVDZ basis set, and a Polarizable Continuum Model (PCM) to take into account the aqueous solvent. We underline several methodological details concerning the definition of the *binding* or *interaction* energy, and the Basis Set Superposition Error definition in the context of implicit solvation models.
We observe two competing interaction modes for this probe : a urea-enhanced, cation-π interaction (with cations only), and hydrogen bonding occurring between the urea group and anions, the former being more favorable than the latter. A Generalized Kohn-Sham Energy Decomposition Analysis (GKS-EDA) [1] in implicit solvent is performed to analyze the nature of the ions – probe interactions. We unveil two families of hydrogen bonding interactions with urea, through oxygen atoms of polyatomic anions on the one hand, and with halides on the other. Magnesium and sodium ions, and respectively glyphosate and hypochlorite ions, are found as the cations (resp. anions) having the largest binding free energies with the probe.
This is the first time such an exhaustive selectivity study is carried out in the context of DFT/PCM models. Moreover, this methodology can be used as a general way to gain a valuable insight into the sensitivity of organic ligands towards a variety of ions or pesticides in water, without the need of an explicit solvent description. By predicting possible competitive interactions, and understanding their nature, this methodology can thus help to better design functional groups selective to specific targets.

PACS numbers:

I. INTRODUCTION

Chemical sensing in water is a very active field for both experimental and theoretical research [2]. The monitoring of water quality, particularly in real time, is a major issue that is only feasible today for a very limited number of parameters. Species to be detected can be present in the form of neutral or charged species. While anions can originate from both inorganic and organic matter, cations are mainly present as metal ions. The knowledge of the concentration of cations – in particular heavy metal cations – and of anions is of importance to precisely determine the quality

*Electronic address: robert.benda@polytechnique.org

of drinking water [3, 4]. Indeed, both deficiency and excess of these can cause serious disorders to the human body (cramps, hypertension, etc.), diseases (diabetes, cancer, etc.), and can even lead to fatalities [5, 6].

While sensing of cations in water has undergone significant advances in the past decades, *e.g.* thanks to the discovery of crown ethers [7, 8], anion detection in water still remains very challenging. Indeed, most studies addressing anion sensing are performed in anhydrous media only [9], *i.e.* organic solvents [10, 11].
Even more challenging is the selectivity, *i.e.* to develop probes which interact in a preferential way with selected target analytes. Due to frequent chemical and steric similarities, such as size or charge of ions, electronegativity of atoms, different ions may often interact in a very similar way with a given probe. More generally, chemical species bearing similar functional groups may also be difficult to isolate by interaction with a given probe.
A promising approach to develop selective sensors is the design of sensing materials able to recognize and specifically interact with the targeted analytes [8]. A very large number of organic probes selective to some specific ionic species have been reported in the literature. Recently, a colorimetric chemosensor to detect fluoride in pure organic solvent and in water/alcohol mixtures was reported [11]. An aluminium based complex, optically selective to fluoride ion, was evaluated for sensing in water in Ref. [12]. Dorazco-González *et al.* also reported fluorescence-based anion sensing in water [13]. A chemical sensor based on a thio-urea group, selective to the acetate ion, was reported by Choi *et al.* [10]. In the latter, selectivity was observed both optically and electrically, using functionalized single walled carbon nanotubes as electrical transducers. Their resistance changed selectively due to the neighboring ion–probe binding. The thiourea functional group was also recently studied in the context of an optical (fluorescence) chemosensor for fluoride ion [14].

Molecular recognition can be achieved through a proper design of the probe and control of the nature of the interaction in which it will be engaged with the analyte [15]. The possibility to use computational tools is thus of prime importance as it may greatly help predicting the optimal design of sensing molecules. Numerical simulations can be deemed as a preliminary screening to design and select promising probes, among a wide variety of candidates, for a given target species, in a similar fashion as in drug design [16]. In particular, unveiling the detailed geometrical features of the probe – analyte complex, and understanding the nature of the interaction, is only possible thanks to theoretical and numerical studies.
Using molecular dynamics simulations, with an *explicit* description of the solvent, and a classical [17] or polarizable force field as an energy model [18, 19], allows to explore widely the configurational energy landscape of these probe–analyte systems at room temperature and derive relevant indicators of the affinity between analytes and probes, such as free energies. However, such models require careful parametrization of the energy functionals (force fields) for each specific probe or chemical species, implying time consuming training and validation steps.

On the contrary, ab-initio theories, such as Density Functional Theory (DFT) [20, 21], require only few model parameters that are independent of the molecule or system at stake. Performing simple geometry minimizations at zero temperature, with a DFT model and without solvent, already allows to gain a valuable insight into the interaction properties of probe–analyte systems (in vacuo, or surrounded by a few water molecules) and affinities between several molecular or ionic fragments.
For instance, Bhargava *et al.* studied the interaction of the CO_2 molecule with several anions thanks to gas phase DFT calculations [22]. Hay *et al.* studied by similar methods the interactions of one or several urea groups with several anions and discussed the optimal arrangement of the former for ion recognition [23]. Ab-initio molecular dynamics, using DFT as energy model, and performed with *explicit* solvent, allows a more precise and realistic room temperature exploration of the configurational landscape of the system. Yet, this family of methods proves extremely costly in terms of computational ressources, due to the expensive self-consistent computation of the density at every time, and even more importantly to the large number of time steps necessary to correctly sample the solvent degrees of freedom.

While DFT minimizations alone cannot computationally address a large number of solvent molecules, long-range solvent effects may still be integrated with *implicit* solvation models, such as Polarizable Continuum Models (PCM) [24–26]. Modelling the solvent with an implicit

model allows notably to compare the preferential solvation behavior of ions with the interactions occurring with the binding probe, at a relatively low computational cost. This is for instance described in the context of chloride binding in Ref. [27]. When complemented with a PCM model to account for the aqueous environment, DFT may prove very enlightening to investigate ion recognition capabilities, and may possibly suffice to validate the efficiency of a given probe to bind selectively a specific analyte in a given solvent, as was shown recently for thiourea based anion receptors [17, 28], or for glyphosate – heavy metal ions interaction [29].

In this article, we study a molecule denoted F0 and depicted Figure 1. It was first introduced by Zucchi *et al.* [30] as one of the monomers periodically grafted to the backbone of the copolymer for anion sensing. This probe was further studied in Ref. [31]. Its design is inspired from the known anion recognition capabilities of the urea moiety [17, 23, 28, 32–34]. On top of the two possible hydrogen bonds at each urea group, the presence of two urea moieties – one in each of the two branches grafted to the same fluorene monomer – may lead to a synergistic and entropic effect. The two branches could possibly jointly interact with the same anion, leading to a stronger anion complexation as compared with the case of two isolated branches, in particular for anions of large size and high steric hindrance. Such effect was observed for monoatomic anions in the case of thiourea moieties acting as hydrogen donors, in Refs. [17] and [28], with the help of an additional OH hydrogen donor in the former. We denote F1 the derivative of probe F0 consisting only of one branch of F0 (see Figure 1). We report a numerical study of the interaction of this urea-based probe with a wide variety of ions present in drinking water, using DFT calculations with a PCM implicit solvent model. In section II, we detail the methodology used to investigate the ion – probe interaction. An emphasis is put on several methodological aspects. We notably discuss the importance of the basis set size and of the implicit solvent model. In section III, we analyse the binding energies and geometries, as well as the selectivity predicted by these calculations. We also analyse the results of the Generalized Kohn-Sham Energy Decomposition Analysis (GKS-EDA) analysis and of the individual components of the decomposed energy.

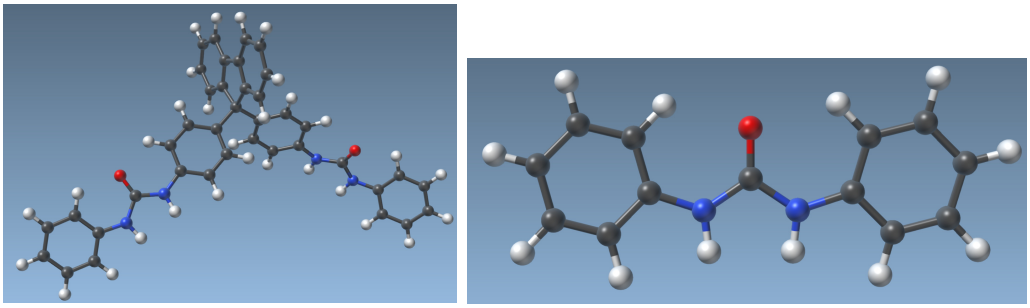

FIG. 1: Left : geometry of F0 probe. The monomer structure contains two sensing moieties, each based on a urea NH-CO-NH group attached to the fluorene-based backbone. Right : F1 probe, derivative of F0 probe consisting only of the sensing moiety part. Both geometries were optimized at the DFT level, using PBE0 exchange-correlation functional and a 6-31G(d) – for F0 probe – or aug-cc-pVDZ – for F1 probe – basis set.

II. METHODOLOGY

A. Investigated systems: ions – probes compounds in water

Definition of the systems:

We perform geometry minimizations of several compounds consisting of a variety of ions initially placed in the vicinity of F0 or F1 probe. Most ions naturally present in drinking water at non-negligible concentration, at pH close to 7, are included in the study [35]. We consider for cations : Na^+, K^+, Ca^{2+}, Mg^{2+} (common minerals, the latter two characterizing water hardness), as well as the common interfering ion NH_4^+. Heavy metal cations, studied *e.g.* in Refs. [29, 36, 37], are not included in this study but are considered for later analysis. Chlorine (Cl^-), hypochlorite (ClO^-), bicarbonate (HCO_3^-), nitrate (NO_3^-), fluorine (F^-), hydrogen phosphate (HPO_4^{2-}), dihydrogen phosphate ($H_2PO_4^-$), sulfate (SO_4^{2-}), formate ion ($HCOO^-$) and bromide

ion (Br^-), are the main anions considered in this study. We also probe the interaction with larger anionic analytes, namely glyphosate (GLY), under its form $C_3H_6NO_5P^{2-}$ at pH $\simeq$ 7 and one of its metabolite, AMPA, under its form $CH_5NO_3P^-$ at pH$\simeq$7. Hydroxide ions (HO^-) and hydronium ions (H_3O^+) are not considered in this study because of their low concentration in water at pH around 7, by comparison to the analytes listed above.

Computing the interaction of these analytes with the full F0 probe proves very costly in terms of computation time. Indeed, F0 is a large molecule containing 75 atoms. We ran several preliminary tests with the large F0 probe and small ions, such as nitrate, with the medium-size basis set 6-31G(d), and did not observe any *simultaneous* interaction with the two branches, for monoatomic and small polyatomic ions (up to 7 atoms). For these ions, later referred to as *small ions*, we therefore only probe the interaction with the smaller probe F1. For the larger ions (glyphosate and AMPA), later referred to as *large ions*, the presence of the two urea moieties in F0 probe happens to play a role and sometimes leads to the expected twofold complexation (by both urea groups) for some specific equilibrium geometries. As a consequence, large ions are studied both with F1 and F0 probes.

Choice of initial configurations:
The initial geometries of the fragments (the probe and the analyte) are optimized at the considered level of theory (*i.e.* DFT exchange-correlation functional and basis set). The initial relative position of these fragments, forming the initial geometry of the complex, is a critical part of the method. Indeed, our choice of method to perform geometry minimizations (see next section) does not allow to explore large deformations or complex pathways in the ion–probe compound energy landscape. Our objective is to probe the interaction of each ion with all possible active sites of the sensing molecule. Thus, for each ion, several initial geometries are tested. For some ions, these different initial geometries lead to different equilibrium geometries (*i.e.* different local minima), whose relative energy can be compared. We then focus on the most stable ones (see section Results below) for further analysis.

In greater details, for small ions, three types of relative initial geometries were tested: the ion placed on the side of the hydrogen atoms of the urea group – approximately in the plane containing F1 – (pos. 1), the ion placed above the urea oxygen atom at about 3 to 4 $\mathring{A}$ (pos. 2), and finally on the side of the oxygen of the urea group, approximately in the plane containing F1 probe (pos. 3). Additional geometries were tested on a case by case basis, but these three initial geometries happened to suffice to describe all local energy minima encountered and all associated equilibrium geometries.

Due to the much higher number of possible initial geometries for the larger probe F0, geometry minimizations of F0 with glyphosate and AMPA ions where first performed with GROMOS classical force field [38], within GROMACS simulation engine [39, 40]. We used the GROMOS parametrization obtained by ATB tool [41] for each compound. The local minima found with GROMACS (at negligible computational cost) were then used as initial geometry for the DFT/PCM calculations. The latter allowed to confirm or infirm the stability of these GROMACS equilibrium geometries, and to compute the associated binding energies at the quantum level. The use of a classical force field can be viewed as a preliminary filter to select some possible interacting geometries of the complex and thus avoid unnecessary initial refinement of the geometry at the (costly) quantum level.

B. Level of theory

DFT model :
The level of theory chosen to perform the geometry minimizations and derive the ion–probe binding energies is DFT, with additional Grimme's D3 semi-empirical London dispersion corrections [42]. The PBE0 functional [43] and the aug-cc-pVDZ Dunning basis set [44], the latter including a significant number of diffuse functions, is used whenever computationally possible. Regarding the computations involving the whole monomer F0, comprising 75 atoms, the smaller 6-31G(d) Pople's type basis set [45–48] is used for geometry minimization, followed by a single point calculation of the energy using the aug-cc-pVDZ basis set.

The general methodology is the following : for small ions initially in the vicinity of the F1 probe, with low configurational disorder and a small number of possible local minima, the geometry is di-

rectly optimized at the PBE0/aug-cc-pVDZ level. For large ions in the vicinity of the F1 probe, the geometry is first optimized using the medium size basis set 6-31G(d), and the resulting equilibrium geometry is used as an input for a second geometry minimization using the larger aug-cc-pVDZ basis set. Finally, for computations involving the whole monomer F0, the first minimization using 6-31G(d) basis set is followed only by a Single Point calculation at PBE0/aug-cc-pVDZ level. Indeed, the geometry optimization at the PBE/aug-cc-pVDZ was too costly computationally-wise.

The aug-cc-pVDZ basis set is larger than the ones generally used to study molecular complexes with large molecules. Glyphosate, for instance, was recently studied at the B3LYP/6-31G(d) or B3LYP/6-31G(d,p) level [29, 49, 50]. Choosing a larger basis allows to capture more precisely the detailed features of non-covalent interaction via the density *tails*. The importance of the latter has been underlined *e.g.* in Ref. [51], in terms of atomic densities. Note that for the F1 probe with a neighboring ion, the Dunning basis set aug-cc-pVDZ is the largest basis that we could use to perform calculations in a reasonable time (a few days per geometry optimization on 8 CPUs).

Finally, to evaluate the impact of this choice of basis set, we asserted that our results at the PBE0/aug-cc-pVDZ level (without Basis Set Superposition Error correction) could be deemed converged with respect to the basis set size. Moreover, the selectivity order between F1 probe and all studied ions at PBE0/aug-cc-pVDZ level was found to be identical at second-order Møller-Plesset theory (MP2) [52] level, using the even larger basis aug-cc-pVTZ. We could thus validate our choice of PBE0/aug-cc-pVDZ level, satisfying both in terms of accuracy and of computation time. See Appendix, Table IV for more details.

PCM model *:*

All the results presented in the following use a conductor-like PCM (C-PCM) implicit solvent model for water [25], unless explicitly mentioned. As a reminder, in a PCM model, the fragments forming the complex are each enclosed in a cavity domain, while the solvent is modelled as an external dielectric medium. The cavity is usually a union of spheres centered on all atoms, of scaled atomic van der Waals radii [53]. When the fragments are far apart, their cavities do not overlap, while they can merge to form one larger cavity, for the whole complex, if they are close enough. Analytical derivatives with respect to atomic coordinates are available for this C-PCM energy [26] and are therefore used for computing the energy gradients needed for the geometry optimizations. The fixed points with variable areas (FIXPVA) tessellation scheme implemented in GAMESS [54] is used to discretize the PCM cavities and solve the C-PCM equations by a boundary element method. The FIXPVA method is well adapted for the computation of the analytic gradients of the PCM contribution to the total energy.

Finally, internal coordinates, generated by the automatic delocalized coordinate algorithm of Ref. [55], were used. Geometry minimizations of non-covalent complexes carried out using such internal coordinates proved to be much more efficient than those using cartesian coordinates [55].

We studied the effect of adding non-electrostatic terms to the PCM electrostatic contribution, and found that the inclusion of cavitation, dispersion and repulsion terms (by Single Point calculations on the optimized complex) did not change qualitatively the selectivity order of the probe. We also studied the sensitivity of the binding energies with respect to the main PCM model parameters, by performing Single Point calculations with the universal solvation model based on solute electron density (SMD) [56]. The latter uses a PCM model with *optimized* parameters, in particular atomic radii (for the definition of the cavity) and atomic surface tensions (for non-electrostatic terms). These parameters are optimized so that the model better reproduces experimental solvation free energies of a wide range of neutral and ionic molecules, on a wide variety of solvents (91 solvents, including water, for the reference database in Ref. [56]). With SMD, we found a qualitatively similar selectivity order for the ion–probe interaction strength, *except* a noticeable difference for calcium ion, which was found to interact with F1 probe as strongly as magnesium within SMD model (uncorrected binding energy of -2.34 eV). The SMD and PCM equilibrium geometries were very close though. To understand the large difference between these models, we performed further calculations using C-PCM model and varying the scaled sphere radius of the calcium ion (used for the definition of the PCM cavity) from 1.92 $\AA$ (corresponding to the default scaled PCM radius of calcium) to 2.72 $\AA$ (corresponding to the calcium radius in SMD model). This resulted in uncorrected binding energies ranging from -0.51 eV to -2.22 eV respectively, *i.e.* a 335 % variation. The former result is associated to the default PCM radius and is reported in this study, while the latter result, using the SMD atomic radius for calcium (and default PCM radii for other atoms) is very close to the SMD uncorrected binding energy (-2.34 eV). Due to this sensitivity of the uncorrected binding energy with respect to the calcium atomic radius used for

the PCM cavity definition, the results associated to Ca^{2+} ion have to be taken with the utmost care. Small ions such as Mg^{2+} are supposed to interact with aromatic rings more strongly than large ions such as Ca^{2+}, which has been studied and validated at high theory levels for the singly charged counterparts Na^+ and K^+ [57]. We therefore chose to keep the default PCM radius for calcium. This interesting model dependency would require further studies which are outside the scope of this article.

Software:

All calculations were performed using the GAMESS package [58] with its 2018 (R3) version, except the Generalized Kohn-Sham Energy Decomposition Analysis (GKS-EDA) [1, 59] calculations which were performed with the latest 2020 GAMESS version (R2) [60].

All graphics of equilibrium geometries below were generated using the SAMSON software [61].

C. Methodological details

Definition of the binding, deformation and interaction energies:

The *binding* energy within a PCM model is defined here as the difference between the PCM total energy of the optimized complex (within its PCM cavity made of the union of both fragment cavities) and the sum of the PCM energies of the optimized, *isolated* fragments A and B, (within their own cavities) :

$$\Delta E_{bind}(AB) := E^{AB}_{AB}(AB)^{PCM} - E^{A}_{A}(A)^{PCM} - E^{B}_{B}(B)^{PCM} \tag{1}$$

where *e.g.* $E^{Y}_{X}(Z)^{PCM}$ denotes the PCM energy of fragment Z, in the optimal geometry of X (indicated as a subscript), using the basis functions of fragment Y only (indicated as a superscript), with $\{X, Y, Z\} \in \{A, B, AB\}$, AB denoting the equilibrium complex. Each of these energies is calculated at the same level of theory.

Note that some authors prefer to study the *interaction* energy [1, 59], because it can be separated into several meaningful physical contributions. The *interaction* energy is defined as the difference between the PCM total energy of the optimized complex and the sum of the PCM energies of the fragments in their geometries *within the optimized complex*, and within their own PCM cavities :

$$\Delta E_{int}(AB) := \left(E^{AB}_{AB}(AB)\right)^{PCM} - \left(E^{A}_{AB}(A)\right)^{PCM} - \left(E^{B}_{AB}(B)\right)^{PCM} \tag{2}$$

Setting aside the Basis Set Superposition Error (BSSE, see section below), the difference between those two energies is called *deformation* energy. It quantifies the energetical cost (or gain) due to the deformation of each fragment, upon interacting with the other and forming the complex. It is defined by the following quantity :

$$\begin{aligned} \Delta E_{def}(AB) &= \Delta E_{def}(A) + \Delta E_{def}(B) \\ &:= \left[\left(E^{A}_{AB}(A)\right)^{PCM} - \left(E^{A}_{A}(A)\right)^{PCM}\right] + \left[\left(E^{B}_{AB}(B)\right)^{PCM} - \left(E^{B}_{B}(B)\right)^{PCM}\right] \end{aligned} \tag{3}$$

Note that there is no consensus in the literature regarding the definition of the interaction strength in the context of an implicit solvent. Some authors chose to use the interaction energy rather than the binding energy, which we deem as the only physically meaningful. The latter accounts for the deformation energy, which can have a significant contribution. The interaction energy may indeed *e.g.* largely overestimate the interaction strength in the case of a high deformation energy (penalty) of one or both of the fragments. Additionally, some authors chose to consider the PCM energy of the fragments within the – larger – complex cavity [62] to define the interaction energy, while we compute in this study the PCM energies of the fragments in their own cavities to define the binding energy.

Definition of the basis set superposition error (BSSE) counterpoise (CP) correction for PCM models:

When comparing the energies of the complex and of its two fragments, a BSSE is made. In a complex, the electronic density of a given fragment can be expanded using the atom-centered basis

functions of the other fragment, which usually leads to an overestimation of the interaction energy. This error can be corrected using the counterpoise (CP) correction method [63]. The rationale behind this correction is to compare only quantities derived exactly at the same level of theory, in particular using the exact same basis set, leading to a compensation of errors. In other words, when minimizing the energy of one fragment, the orbitals centered on the atoms of the other fragment should also be available. This can be done by using *ghost* atoms representing the positions of the other fragment in the complex, carrying their usual atomic basis functions.

When using a PCM model, the appropriate definition of CP correction for BSSE is not clearly defined in literature. To maintain the variational nature of the CP correction, we choose to compute the energies for CP correction *in vacuo* (*i.e.* without PCM), at the equilibrium geometry of the complex :

$$\Delta E^{CP}(AB) := \left[\left(E^{AB}_{AB}(A)\right)^{vac.} - \left(E^{A}_{AB}(A)\right)^{vac.}\right] + \left[\left(E^{AB}_{AB}(B)\right)^{vac.} - \left(E^{B}_{AB}(B)\right)^{vac.}\right] < 0 \quad (4)$$

where $\left(E^{AB}_{AB}(A)\right)^{vac.}$ and $\left(E^{A}_{AB}(A)\right)^{vac.}$ denote the energy *in vacuo* (without PCM model) of fragment A in the geometry of complex AB, respectively with the basis functions of both fragments A and B, or with the basis functions located on the atoms of A only (and similarly for B).

One could propose to calculate the CP correction as $\left(E^{AB}_{AB}(A)\right)^{PCM} - \left(E^{A}_{AB}(A)\right)^{PCM}$ (for the contribution of fragment A), *i.e.* the difference of PCM energy between the fragment A in its geometry in the equilibrium complex AB, with the whole basis of AB, or with its own basis only, respectively. However, computing the PCM contributions to the energy of fragment A or B, in the geometry of the complex AB, using the basis functions on ghost atoms, *and* spheres centered on these ghost atoms to define the PCM cavity, would not make sense from a physical point of view. For instance, if a monoatomic ion is artificially placed into a much larger cavity, defined from the union of its natural cavity *and* of the spheres centered on the (ghost) atoms of the neighboring F0 or F1 probe, the PCM polarization energy is artificially low, due to the (too) large size of the cavity, enabling fewer polarization of the ionic density by the surrounding dielectric medium. Besides, not using spheres centered on ghost atoms would lead to uncontrolled escape charge errors [64].

Due to the variational nature of the energy in vacuo, the CP correction (4) for BSSE is always negative, so that the the *CP corrected binding energy* is always smaller in magnitude than its non-corrected counterpart (if both are negative) :

$$\Delta E^{CP}_{bind}(AB) = \Delta E_{bind}(AB) - \Delta E^{CP}(AB) > \Delta E_{bind}(AB) \quad (5)$$

The *CP corrected binding energy* can be rewritten as :

$$\begin{aligned}\Delta E^{CP}_{bind}(AB) = E^{AB}_{AB}(AB)^{PCM} - E^{A}_{A}(A)^{PCM} - E^{B}_{B}(B)^{PCM} \\ - \left[\left(E^{AB}_{AB}(A)\right)^{vac.} - \left(E^{A}_{AB}(A)\right)^{vac.}\right] - \left[\left(E^{AB}_{AB}(B)\right)^{vac.} - \left(E^{B}_{AB}(B)\right)^{vac.}\right] \quad (6)\end{aligned}$$

There is no further simplification arising from expression (6). Though CP correction, widespread in literature, is generally thought to reduce BSSE [65], some authors point out that it may not be the case in some contexts, specially in the case of post-Hartree Fock calculations [66, 67]. One thus needs to be careful when interpreting results from CP corrected binding energies. Moreover, there is no consensus on how to define CP correction when using an implicit solvent model (see the discussion above). In practice, we show in the case study presented in this article that CP correction does not change significantly the selectivity order of the probe F1. Indeed, it is reported in Table I that CP correction amounts to 5% at most of the (uncorrected) binding energy, except for SO_4^{2-} and HPO_4^{2-} ions.

Entropic corrections :

To get closer to the experimental setup at room temperature, we correct the binding energies thanks to vibrational entropic contributions [68]. The latter allow to quantify the width of the energetic wells associated to every equilibrium geometry. The wider the energetic wells, the longer the system (complex probe–ion) is likely to remain within this favorable region, and the larger (*i.e.* more stabilizing) the entropic contributions. We derive the normal mode frequencies from the computations of the hessian matrices at the complex geometries (at PBE0/aug-cc-pVDZ level,

with PCM) for F1 probe and different ions.

In GAMESS, this contribution for a system A is computed as the sum over all vibrational (normal) modes at the equilibrium geometry (for the considered level of theory) of the mode's vibrational entropy [68, 69] :

$$S_{vib}(A) = \sum_{\omega} R \left[\frac{\hbar\omega}{k_B \left(e^{\frac{\hbar\omega}{k_B T}} - 1 \right)} - \ln \left(1 - e^{-\frac{\hbar\omega}{k_B T}} \right) \right] \tag{7}$$

where the frequency of the normal modes ω runs over all $3M-6$ internal vibrational modes (M being the total number of atoms in A), k_B (resp. $h = 2\pi\hbar$) is Boltzmann (resp. Planck) constant. The resulting formula for the corrected binding *free* energy corrects the (already CP corrected) binding energy thanks to the difference of vibrational entropy between the optimized complex with respect to its isolated fragments :

$$\Delta E_{bind}^{corr}(AB) = \Delta E_{bind}^{CP}(AB) - T\left(S_{vib}(AB) - S_{vib}(A) - S_{vib}(B)\right) \tag{8}$$

which is evaluated at T=298 K in the following.

Selectivity of the probe:
The selectivity of a probe towards a set of analytes is defined as the binding preference order between the probe and the analytes. In the case study presented here, the *corrected binding energy corrected* (see Eq. 8) is chosen as measure of this binding preference order.

Definition of the interaction distance:
The probe and the different analytes can interact in different ways. For anions, the interactions are mainly hydrogen (H) bonds of the type N–H - - X where N–H belongs to one urea of the probe, while X is one atom of the anion. The *interaction distance* is defined as the average length of the H bonds existing in the complex.

Energy Decomposition Analysis (EDA):
In order to rationalize the selectivity order and the nature of the interactions between the probe and the analytes, we perform a Generalized Kohn-Sham Energy Decomposition Analysis analysis (GKS-EDA) in implicit solvent on the optimized complexes [1, 59]. This method is a generalization in the context of DFT of the EDA-PCM method [70], overcoming some shortcomings of the latter regarding Kohn-Sham orbitals [1, 59]. It belongs to a family of methods enabling to decompose the total *interaction* energy (*i.e.* of the complex with respect to its fragments *in their geometries in the complex*) into seven physically meaningful components ; namely an attractive *electrostatic* contribution, a *repulsion* term, weakened by an attractive *exchange* (arising from Pauli principle) term, a *polarization* term (also coined as the induction term in other EDA methods, such as SAPT [71]), a *desolvation* term typical of EDA methods in implicit solvent (quantifying the energy loss due to incomplete solvation shell upon fragment complexation), a *dispersion* term accounting for the contributions of attractive London dispersion interactions, and finally a *correlation* term. As in all EDA methods, there are necessarily arbitrary choices in the decomposition of the total energy, which is the only unambiguous quantity, into individual components [72], and the definitions of the different terms may slightly differ from one conceptual framework to another. Yet, these methods are useful to qualitatively comment on the nature of interactions at stake and in the present case to discuss the relative differences of binding among the studied ions.
Following the choices of Ref. [62] and for the sake of simplicity, we choose to gather the exchange and repulsion terms into a single term (Pauli repulsion energy [62]) and the dispersion and correlation terms into another single term.

Note that the GKS-EDA method is based on the decomposition of the *interaction* energy, which differs from the *binding* energy studied in the present article (see Equation 1), as discussed in Section II C. As mentioned, the interaction energy does not take into account the *deformation* energy associated with the deformation of each fragment in the complex with respect to their standalone equilibrium geometries (see Equation 3). Since the deformation energy is always positive, this method yields energies higher (in magnitude, *i.e.* more favorable) than the corrected binding

energy used in the main analysis of the present article.

Another difference between the interaction energy used in the GKS-EDA scheme and our corrected binding energy lies in the way PCM cavities are defined when computing the energy of the fragments. In the present article, the PCM cavity of the fragment alone (a sphere for a single ion) is used when computing the corrected binding energy of the fragment. On the contrary, in the GKS-EDA scheme, the PCM cavity of the whole complex is considered, for each fragment, in its geometry within the complex [62]. These two differences in definition explain the differences between the CP corrected binding energies reported in Table I (third column) and the total CP corrected GKS-EDA interaction energies in Table V (Supplementary Material), whose decompositions are reported in the main text, section III B. The relation and source of the difference between the GKS-EDA interaction energy and the binding energy are further detailed Appendix IV B and Table VI.

III. RESULTS AND ANALYSIS

A. Interaction of the probes with a single ion

1. F1 probe

The smallest probe (F1), made of a single urea group in-between two phenyl groups, is found to interact with most of the investigated ions. The results of the DFT/PCM energy minimizations at the PBE0/aug-cc-pVDZ level are ranked in Table I from the largest to the smallest corrected (for BSSE, and for entropic terms) binding energy (*i.e.* binding *free* energy at T=298 K), using only the most favorable equilibrium geometries found in this study. The other equilibrium geometries found when using different initial geometries are reported in Table VII (Supplementary Material). By entropic corrections, we mean the (negative) stabilizing energy $-T\left(S_{vib}(AB) - S_{vib}(A) - S_{vib}(B)\right)$ defined section II, at $T = 300K$, reported in the second column of Table I. By CP correction value, we mean for each complex AB, the quantity $-\Delta E^{CP}(AB)$, defined above, which corrects the overestimated binding energy (due to BSSE).

Anions

First, most of the tested anions, whether they are monoatomic or polyatomic, interact more strongly with F1 probe than with water, consistently with the experimental data of anion recognition by urea groups [17, 23, 28, 33, 34]. In both cases, the complexation is due to the formation of two hydrogen bonds of type N–H - - X, where nitrogen and hydrogen atoms belong to urea while X is the targeted ion (or one of its atoms in the polyatomic case). This interaction is mainly of electrostatic nature (see section III B below), with an interaction distance (see Section II C for definition) ranging from 1.65 $\mathring{A}$ (for F^-) to 2.41 $\mathring{A}$ (for Br^-). Glyphosate (GLY) and hypochlorite (ClO^-) are the anions found to interact most favorably with F1 probe at this level of theory, with a binding free energy (corrected binding energy) of -1.114 and -0.982 eV respectively.

The equilibrium geometries obtained for most anions and F1 probe are reported in Figure 2. For polyatomic anions, either one oxygen of the analyte forms two H bonds of the form N–H - - O (one for each N–H group of the urea), or two different oxygens of the analyte form one H bond of the form N–H - - O each. The angles $\widehat{NHX}$ associated with these H bonds (indicated in the last column of Table I) are closer to the most favorable 180° angles [73] in the last case. Finally, we observe that planar ions such as NO_3^- or ClO^- are inclinated perpendicularly to the plane of F1 probe in their equilibrium geometry. Generally speaking one can hope to increase the selectivity of urea-based probes by using *several* urea moieties in specific geometries, as shown in Refs. [23, 74] *e.g.* for NO_3^- ion.

The interaction between F1 probe and large anions is similar to the one described for small anions. Indeed both glyphosate ion (GLY) and AMPA ion interact with F1 through a single one of their functional groups (either PO_3^{2-} or COO^- for GLY, and COO^- for AMPA) bonding to the urea of the F1 probe. As a result, the binding free energy of AMPA (-0.929 eV) is close to the one of $HCOO^-$ alone (-0.873 eV).

Only minor deformations of F1 probe are observed regarding anion complexation discussed above. The only visible deformations are those necessary for F1 to complex simultaneously by two H

bonds the sulfate ion SO_4^{2-} and the dihydrogen phosphate ion (see Figure 2). In particular, no covalent bond formation was found.

Cations

The interaction of F1 probe with cations occurs through two different mechanisms. First, the phenyl rings can enable a cation-pi interaction. Second, the oxygen atom of the urea can form an electrostatic bond with a cation. These two mechanisms result in a hybrid interaction, depicted in Figure 3, associated to a large deformation of the probe F1. Despite being much weaker than in the gas phase – where it can reach up to −5 eV at MP2/aug-cc-pVTZ level [70] –, the cation-π interaction mode [57, 75–78] remains effective in implicit water. This result is consistent with the studies of cation-π interactions in aqueous, biological conditions [77]. We show that this hybrid interaction amounts to a binding free energy of about −1 eV at PBE0/aug-cc-pVDZ level (with PCM), which is consistent with Ref. [70]. Both Mg^{2+} and Na^+ ions exhibit such a large hybrid interaction, respectively -1.555 eV and -1.114 eV. In the case of Ca^{2+}, it is found to be much weaker (-0.530 eV) with the default PCM radius for Calcium (used to build the cavity), although this binding free energy was found to be very sensitive towards this PCM radius parameter (see section II. This is consistent with cation-π theory, which predicts (at least for singly charged ions) a larger interaction for smaller ions [57, 77].

Interestingly, this hydrid interaction is, in the case of Mg^{2+}, more favorable than the maximal anion – F1 interaction found in this study (see Table I). The difference of binding energies between Mg^{2+} and GLY ions is much larger than k_BT=25 meV (at 300 K), so that F1 probe can be expected to be selective towards Mg^{2+} in water at room temperature.

Note that this hybrid interaction could not be estimated at the same PBE0/aug-cc-pVDZ level of theory in the case of the potassium ion, as this family of Dunning basis set has no been parametrized for the K atom to our knowledge.

Ion (charge)	**Corrected binding energy** $\Delta E_{bind}^{corr}(AB)$ (entropic correction)	**CP corrected binding energy** $\Delta E_{bind}^{CP}(AB)$ (CP correction value)	**Binding geometry**	**Distance and** $\widehat{NHX}$ **angle (H–bond)**
Mg^{2+}	-1.555 (-0.018)	-1.537 (0.045)	Urea enhanced cation-π interaction	.
GLY (-2)	-1.114 (-0.376)	-0.738 (0.065)	Complexation by H-bonds of one oxygen atom of the phosphate moiety of GLY	1.73 $\AA$ 155°
Na^{+}	-1.087 (-0.120)	-0.967 (0.037)	Urea enhanced cation-π interaction	.
ClO^{-}	-0.982 (-0.209)	-0.773 (0.040)	H-bond complexation	1.74 $\AA$ 158°
AMPA (-1)	-0.929 (-0.242)	-0.687 (0.059)	Complexation by H-bonds of one oxygen of the carboxylic moiety of AMPA	1.77 $\AA$ 157°
$HCOO^{-}$	-0.873 (-0.194)	-0.679 (0.048)	H-bonds with 2 oxygen atoms of $HCOO^{-}$	1.78 $\AA$ 173°
SO_4^{2-}	-0.861 (-0.225)	-0.636 (0.165)	H-bonds with 2 oxygen atoms of $SO4^{2-}$	1.75 $\AA$ 169°
$H_2PO_4^{-}$	-0.849 (-0.277)	-0.572 (0.074)	H-bonds with 2 oxygen atoms of $H_2PO_4^{-}$	1.85 $\AA$ 171°
NO_3^{-}	-0.751 (-0.359)	-0.392 (0.035)	H-bond with one oxygen atom of NO_3^{-}	1.92 $\AA$ 157°
HPO_4^{2-}	-0.742 (-0.242)	-0.512 (0.260)	H-bonds with 1 oxygen atoms of HPO_4^{2-}	1.72 $\AA$ 156°
F^{-}	-0.704 (-0.094)	-0.610 (0.030)	H-bond complexation	1.65 $\AA$ 157°
HCO_3^{-}	-0.623 (-0.070)	-0.553 (0.042)	H-bond complexation with one oxygen of $HCO3^{-}$ only	1.86 $\AA$ 158°
Cl^{-}	-0.584 (-0.142)	-0.442 (0.023)	H-bond complexation	2.23 $\AA$ 162°
Br^{-}	-0.569 (-0.161)	-0.408 (0.023)	H-bond complexation	2.41 $\AA$ 164°
Ca^{2+}	-0.530 (-0.052)	-0.478 (0.028)	Urea enhanced, very weak cation-π interaction (interaction mainly with oxygen of urea)	.
H_2O	-0.509 (-0.190)	-0.319 (0.035)	H bonds with oxygen atom of H_2O	1.97 $\AA$ 156°
HClO	-0.491 (-0.315)	-0.176 (0.042)	Complexation by H-bonds of oxygen atom of HClO	2.01 $\AA$ 157°
NH_4^{+}	-0.216 (-0.140)	-0.076 (0.034)	H-bond N–H (N of NH_4^{+}) with nitrogen of urea group.	.

TABLE I: Binding energies of different ions with the molecule F1, estimated at PBE0/aug-cc-pVDZ level, with DFT-D3 semi-empirical dispersion corrections [42], with PCM, corrected for BSSE by counterpoise correction (third column), and additionally by entropic corrections (second column). Binding energies are computed with respect to the isolated ions and F1 probe (see section II C). CP correction values naturally do not include DFT-D3 contributions. All energies are in eV.

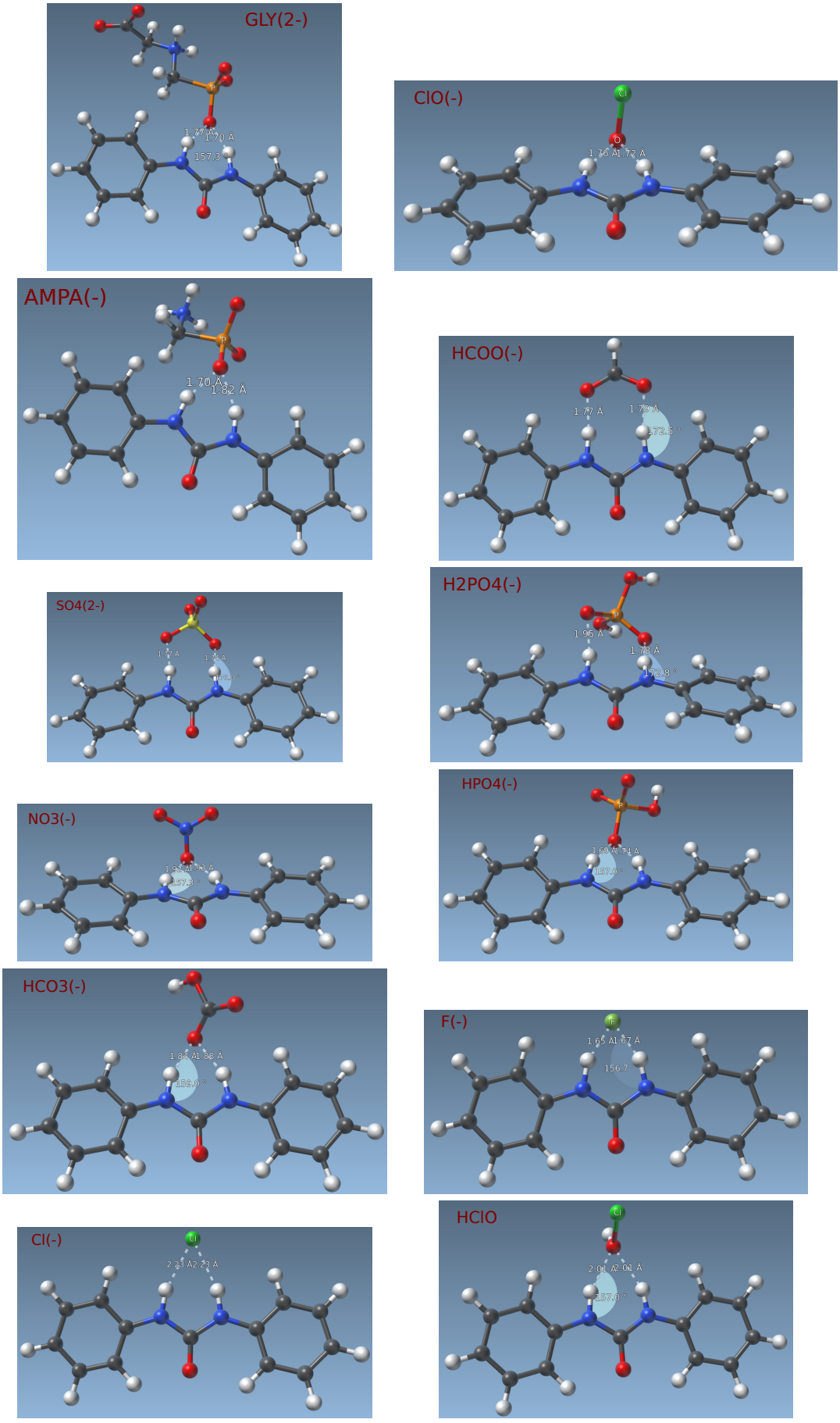

FIG. 2: Most stable equilibrium binding geometries at PBE0/aug-cc-pVDZ level, with PCM, found for F1 probe and *anions*. The ordering corresponds to the largest (upper left) down to the lowest binding free energy (lower right) for anions : GLY (-2), ClO^-, AMPA, $HCOO^-$, SO_4^{2-}, $H_2PO_4^-$, NO_3^-, HPO_4^{2-}, HCO_3^-, F^-, Cl^-, HClO.

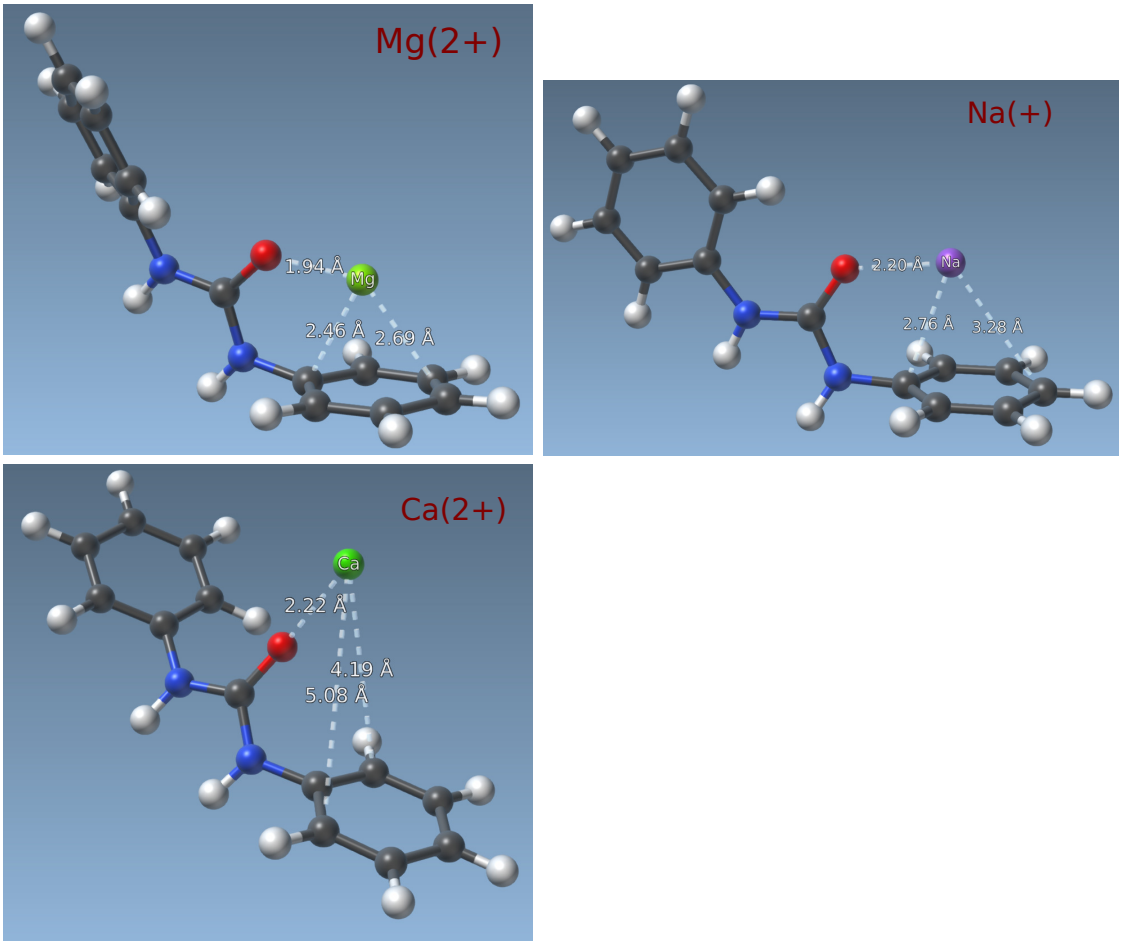

FIG. 3: Optimal *cation*–F1 interacting geometries found at PBE0/aug-cc-pVDZ level, with PCM, for Mg^{2+}, Na^{+} and Ca^{2+} ions. A cation-π interation is observed for all these cations, albeit weaker for Ca^{2+} ion, much farther from the phenyl cycle at equilibrium.

2. F0 probe

The interactions with the F0 probe were studied to investigate whether interactions with both branches of the probe were possible or not. It was found that for small ions, the interaction can only occur with each branch separately. In that case, the binding energy with F0 is the same than with the smaller F1 probe. Therefore, we focused on studying in greater details the interaction of F0 with large ions only. Equilibrium geometries obtained at the PBE0/6-31G(d) level are presented in Figure 4.

An interaction with both branches of F0 was found for both GLY and AMPA. GLY is composed of two negatively charged moieties that may interact with urea: $-PO_3^{2-}$ on one end and $-COO^-$ on the other. As a result, we observed a configuration in which $-PO_3^{2-}$ formed H bonds with the urea of one of the branches of F0, while $-COO^-$ formed H bonds with the urea of the other branch. In this configuration, the two branches of F0 enclose GLY. On the other hand, AMPA is composed of only one moiety that can interact with urea, namely a carboxylate group $-COO^-$. We could still obtain a configuration in which $-COO^-$ interacted with both ureas at the same time, but in this configuration, F0 was greatly deformed as compared with its optimal standalone geometry, with a large associated deformation energy penalty.

Ion	**CP corrected binding energy** (CP correction value)	**Binding geometry**
GLY	-1.098 (0.126)	GLY interacting with both urea moieties of F1 at the same time (Figure 4 left)
AMPA	-0.462 (0.122)	AMPA encaged between the two branches of F0 (Figure 4 right)

TABLE II: CP corrected binding energies (in eV) and binding geometries of GLY and AMPA with F0 probe (whole monomer), estimated at PBE0/aug-cc-pVDZ level, with PCM level, with DFT-D3 semi-empirical dispersion corrections [42].

As can be seen by comparing the results reported in Tables I and II, these *cage* configurations with F0 can be more favorable than the ones found with F1, depending on how much F0 is deformed. For instance, GLY could fit inside the branches of F0 without a too large deformation of the latter. As a result, the CP corrected binding energy between GLY and F0 (-1.098 eV) is larger than the one for GLY and F1 (-0.738 eV). But in the case of AMPA, the probe F0 was much more deformed, resulting in a large deformation energy and a smaller CP corrected binding energy (-0.462 eV) than in the case with F1 (-0.687 eV). Note that for such a large system involving F0 molecule, and the associated very large computational cost of Hessian calculations, entropic contributions to the free energy could not be estimated, so that we only report CP corrected binding energies.

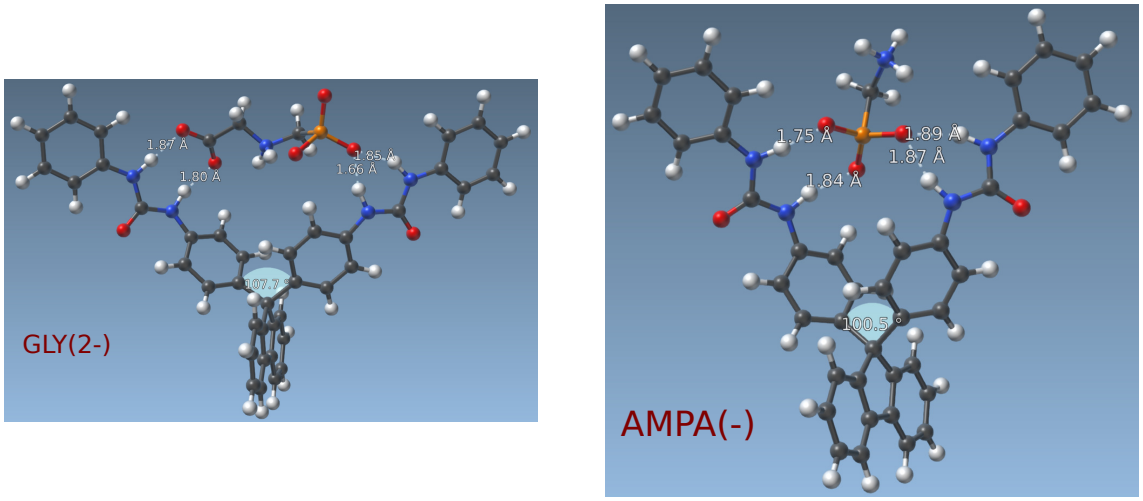

FIG. 4: Interaction between F0 and GLY (left) and AMPA (right). Optimization at PBE0/6-31G(d) level with PCM model.

Note also that for both of these configurations with F0, the CP correction is about twice larger than the one obtained with F1. For instance, the CP correction for AMPA was of -0.059 eV with F1 (see Table I), while it reaches -0.122 eV with F0. The reason lies in the definition of the BSSE error: in a complex, the electronic density of a given fragment can express itself using the atom-centered basis functions of the other fragment, which usually leads to an overestimation of the interaction energy [79]. Compared to the interaction with F1, the interaction between F0 and the analyte involves about twice as many atoms, leading to a larger CP correction.

B. Energy Decomposition Analysis (EDA) for F1 probe – ion compounds

The components of the energy arising from the GKS-EDA decomposition scheme are reported in Table V (Supplementary Material) and in figure 5 below, allowing for a more visual understanding of this partitioning and a comparison of the differences between some representative interaction mechanisms. For all ions (and also for the neutral species H_2O and HClO to a lesser extent), the interaction is mainly of electrostatic nature, as the electrostatic term is always larger than the polarization term, except for NH_4^+ that interacts very weakly. This confirms the non-covalent, mainly ionic character, of the interaction, as the share of the polarization term in the total energy is an indicator of the degree of orbital hybridization and thus of covalent bonding.

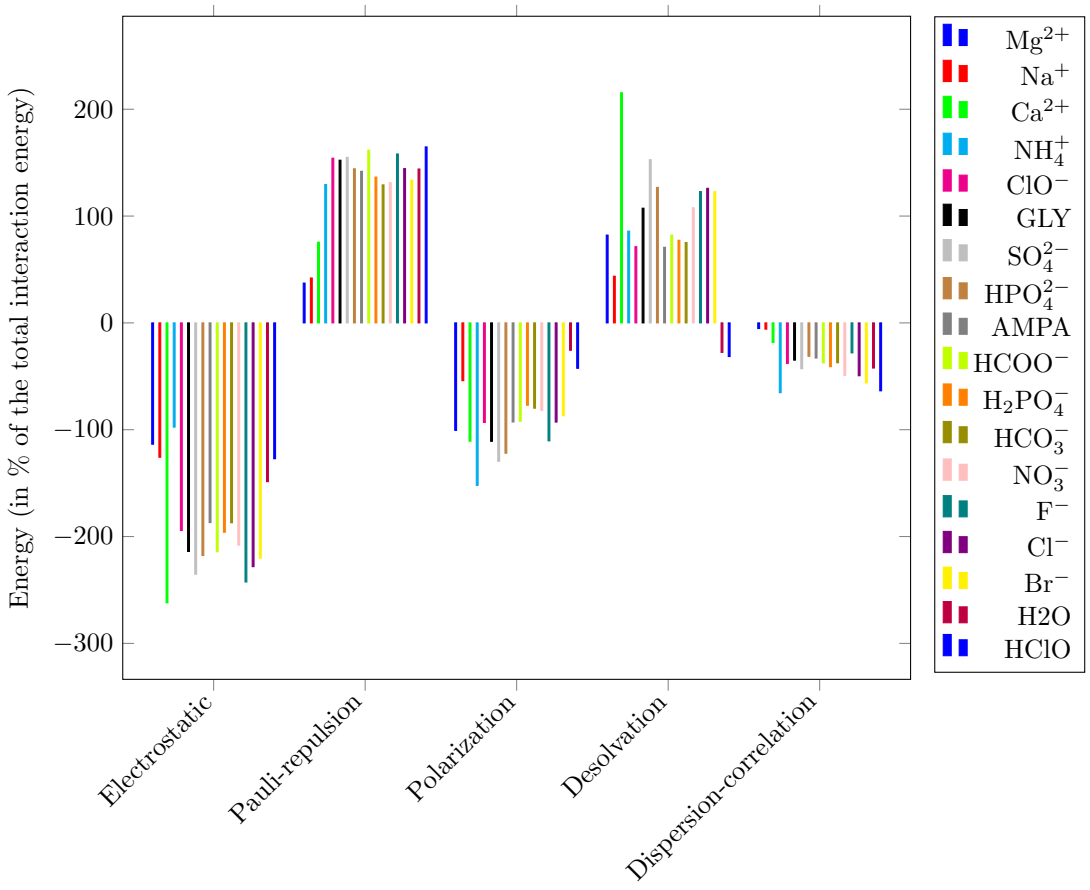

FIG. 5: Histogramm showing the contribution of all terms of the GKS-EDA interaction energy (in % of the total interaction energy, negative terms accounting by convention for attractive contributions, see Table V) of F1 probe with all studied chemical species. We choose to gather respectively cations, anions interacting with F1 probe thanks to an oxygen atom (from ClO^- to NO_3^-), halides (F^-, Cl^-, Br^-), and neutral species (H_2O, HClO). Within these subgroups, ions or molecules are ranked by (uncorrected) binding energies (Table I). As suggested in [62], exchange and repulsion terms were grouped into a Pauli repulsion term, as well as dispersion and correlation terms.

1. *Interaction with cations*

The desolvation energy quantifies the loss of energy of the two interacting fragments, due to the non-complete solvation shell around them, when they lie close one to another in the complex geometry. For cations such as Ca^{2+} (see Table V), and more specifically Mg^{2+}, the desolvation free energy is particularly high (see Table V and histogram 5). This is due to the high solvation free energy of these cations and to the urea enhanced cation-π interaction geometry unveiled above, which forces the ion to loose more than half its complete solvation shell due to the steric hindrance of the phenyl group. This desolvation penalty is compensated (for cations interacting by urea enhanced cation-π interaction) by a strong electrostatic free energy contribution as well as a significant polarization free energy contribution, in particular for Mg^{2+} ion. The polarization contribution, which is particularly high for Mg^{2+}, quantifies the energy gain due to the orbital relaxation of the fragments upon complexation. It thus seems that the orbitals associated to the magnesium ion are the most affected (relative to other ions) by the interaction with F1 probe.

2. *Polyatomic analytes*

In this study, all polyatomic analytes – be them negatively charged or neutral – interact with F1 probe through hydrogen bonds of the type N–H - - O, O belonging to the analyte. As expected from hydrogen bonding, the electrostatic free energy is the major contributor to the interaction energy. The former is about twice as large (or more) as the polarization energy, the second largest attractive contribution to the total interaction energy, as can be seen from Histogram 5.

It is interesting to note that all energetic terms (except desolvation) can be linearly correlated to the length of the H bond (see Figure 6). In particular, considering only polyatomic analyte – F1 complexes, the electrostatic, Pauli repulsion and polarization energies are linearly correlated to the length of the H bond (Figure 6). This linear behavior is consistent with the physical meaning of each term. First, electrostatic free energy originates from attractive Coulomb interactions, whose strength increases when the distance decreases. Second, Pauli repulsion is due to Pauli's exclusion principle, which plays a role when the electronic densities of the fragments overlap. It is thus expected to be positive and increase when the separation between the fragment decreases. Third, the polarization free energy measures the energetic gain associated to the relaxation of the electronic density of the fragments, with respect to their isolated counterparts, upon approaching one to another. Its magnitude is thus also expected to increase when the interaction distance decreases. Finally, dispersion-correlation term is less clearly correlated with the length of the H bond (not shown).

Let us also notice that the desolvation free energy is not correlated with the length of the H bond, but rather with the charge of the ion. Indeed, anions with charge -2 have a desolvation energy around 1.8 eV, while anions with charge -1 have a desolvation energy around 0.9 eV, and the neutral species (H_2O and HClO) have a negligible desolvation energy (about -0.1 eV in both cases). The only exceptions are fluorine (F^-) and glyphosate (GLY).

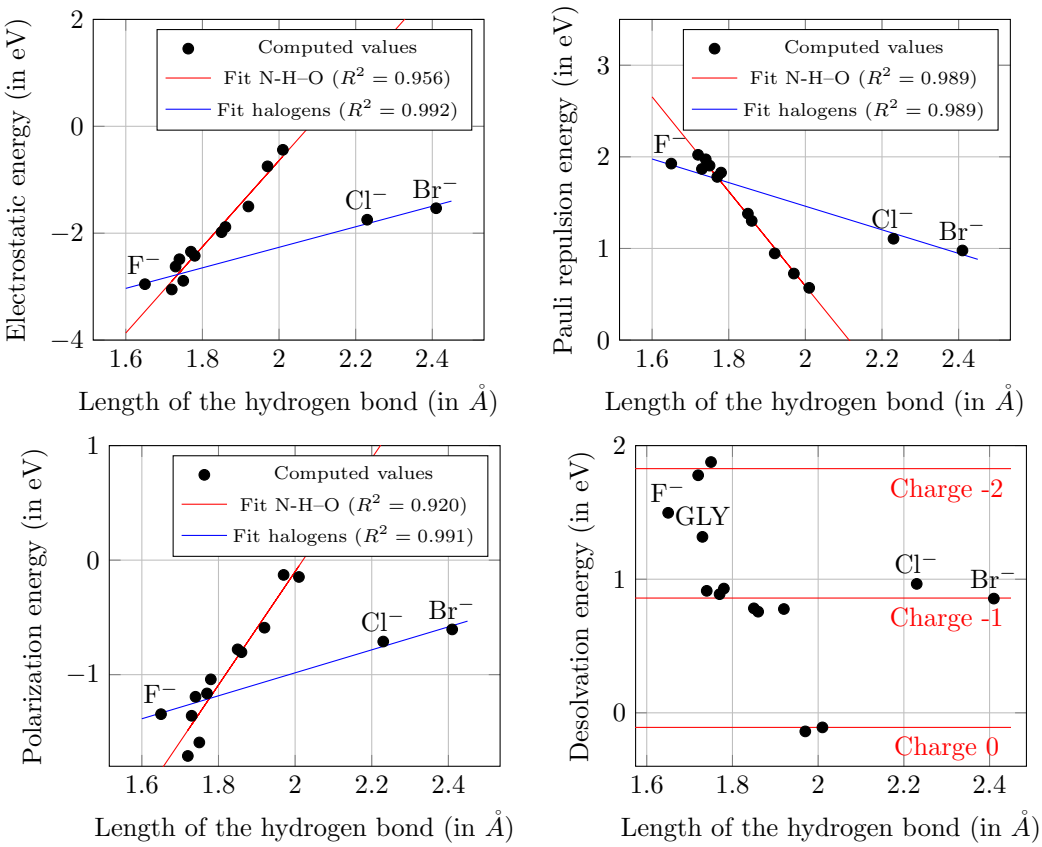

FIG. 6: Dependance of some terms of the GKS-EDA interaction energy (between F1 probe and anions) with the length of the H bond. Two linear fits are performed on these data: one for complexes with N-H - - O bond and one for halides, which interact with F1 thanks to a different kind of H bond.

3. Interaction with halides

As can be seen on Figure 6, the previous analysis for polyatomic ions cannot be extended to the case of halide ions. Indeed, the nature of the bond between halides and F1 is different, as it is of the type N–H - - X, X being the halide ion. The main energy contributions are still following a linear pattern in the case of halides, but its parameters are not the same as in the case of the N-H - - O bond. As expected, the interacting distance between halides and F1 increases with the size of the anion. Therefore, the interacting distance of F^- is smaller than Cl^-, which is smaller than Br^-.

IV. CONCLUSION

In this article, we probed the interaction of a urea-based probe with a wide variety of ions in water by means of DFT calculations complemented with an implicit solvation model (PCM). We derived binding free energies, corrected for the Basis Set Superposition Error and accounting for entropic contributions at room temperature. Two different interaction mechanisms are predicted: one for anions and one for cations. On the one hand, the observed sensitivity of F0 and F1 probes to anions is due to hydrogen bonds with the urea groups, which we show thanks to the GKS-EDA analysis to be of different nature for oxoanions and for halides. On the other hand, the presence of the phenyl groups in the probe leads to competitive interactions with cations – which are always present in solution as counter-ions. The observed cation-π interaction, enhanced by the interaction with the oxygen atom of urea, is predicted to be stronger than the urea – anion interaction, in particular for Mg^{2+} cation.

The thermodynamical predictions of this article are relevant for a range of concentrations allowing possible interaction of the ions with the probes. This family of urea based probes are thus predicted to be *sensitive* to a wide variety of ions in water, the latter having a much larger binding free energy with the probes than water itself (except NH_4^+ ion). Within the usual range of concentration of the studied ions in water, our results let us expect a *selectivity* to Mg^{2+} ion, due to the large binding free energy difference with the second most interacting ion (about 18 times k_BT at room temperature) and thanks to the Boltzmann exponential factor governing the probability of interaction. Strategies to reach selectivity to *anions*, would require the combination of several urea groups as well as impeding the possibility of a cation-π interaction.

Whether these results can be extrapolated into predictive reaction constants (characterizing the probe selectivity) is still an open question, although it has been reported to work well in some contexts [68]. Indeed, this DFT/PCM study, performed by means of geometry minimizations, overlooks dynamical (kinetic) effects, which might be the driving force in some contexts, as well as solvent structure effects [56]. The role of the first solvation shell and the necessity to include some explicit water molecules around the probe [80] is still an open question. To include these effects, explicit solvent molecular dynamics simulations (*e.g.* using polarizable force fields to accurately describe electrostatics and polarization) would be a reliable method. The binding free energies at room temperature could be as well be estimated with such an explicit solvent room temperature dynamics, and compared with the binding free energy derived here (at a lower computational cost) by structural minimizations with DFT/PCM calculations, and additional entropic corrections. Moreover, the DFT/PCM calculations of this study can be used as reference data to parametrize the force field necessary to run dynamical simulations. Such force field parameters optimization and explicit solvent molecular dynamics simulations will be reported in the near future, and compared to the predictions of the present study.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

We deeply thank Peifeng Su for sending us the GAMESS code 2020 version [60] implementing the GKS-EDA method, and for some advice on the best way to use the GKS-EDA method on DFT calculations. We thank Benedetta Mennucci and Filippo Lipparini for useful discussions and advice. Calculations were performed thanks to CERMICS computing cluster. This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement N° 810367).

Appendix

A. Convergence of binding energies with the basis set

1. *PBE0/6-31G(d) level*

As can be seen comparing the results reported in Tables I and III, the selectivity order found at PBE0/6-31G(d) level is significantly different from the selectivity order found at PBE0/aug-cc-pVDZ level (both with PCM). This can be noticed *e.g.* for F^- ion, which displays the largest binding energy at PBE0/6-31G(d) level (Table III), while it only interacts moderately with F1 at PBE0/aug-cc-pVDZ level (Table I), showing that the results of binding energies are not converged for the basis set 6-31G(d).

Ion	**Uncorrected binding energy (eV)**	**Binding geometry**	**Distance H-bond**
F^-	-1.837	H bonds with urea	1.52 Å
Mg^{2+}	-1.810	Cation-π interaction	.
ClO^-	-1.342	H bonds with urea	1.68 Å
Na^+	-1.189	Cation-π interaction	.
SO_4^{2-}	-1.173	H-bonds with 2 oxygen atoms of $SO4^{2-}$	1.72 Å
$HCOO^-$	-1.163	H-bonds with 2 oxygen atoms of $HCOO^-$	1.78 Å
HPO_4^{2-}	-1.145	H-bonds with 1 oxygen atoms of HPO_4^{2-}	1.69 Å
$H_2PO_4^-$	-0.908	H-bonds with 2 oxygen atoms of $H_2PO_4^-$	1.81 Å
HCO_3^-	-0.894	H-bonds with 2 oxygen atoms of HCO_3^-	1.82 Å
AMPA	-0.882	Complexation by H-bonds of the carboxylic moiety on the urea	1.77 Å
K^+	-0.790	Cation-π interaction	.
GLY	-0.722	Complexation by H-bonds of the phosphate moiety on the urea	1.87 Å
Ca^{2+}	-0.716	Very weak cation-π interaction (interaction mainly with oxygen of urea)	.
NO_3^-	-0.756	H-bonds with one O of NO_3^-	1.85 Å
Cl^-	-0.553	H bonds with urea	2.23 Å
H_2O	**-0.548**	H bonds with urea	1.91 Å
HClO	-0.326	H bonds with oxygen atom of HClO	2.01 Å
NH_4^+	-0.138	H-bond N–H (N of NH_4^+) with nitrogen of urea	.

TABLE III: Binding energies (in eV) of different ions with molecule F1 computed at the PBE0/6-31G(d) level, with DFT-D3 corrections [42]) and with a PCM implicit solvent model, using GAMESS code.

2. *PBE0/aug-cc-pVDZ // PBE0/aug-cc-pVTZ and PBE0/aug-cc-pVDZ // MP2/aug-cc-pVTZ levels*

The selectivity order at the level PBE0/aug-cc-pVDZ // PBE0/aug-cc-pVTZ (optimization at PBE0/aug-cc-pVDZ level followed by PBE0/aug-cc-pVTZ single point calculations at PBE0/aug-cc-pVDZ optimal geometry) is the same as at the level PBE0/aug-cc-pVDZ, except a reversal for F^- and $H_2PO_4^-$ (these ions being already very close in terms of uncorrected binding energies at PBE0/aug-cc-pVDZ level, within ±1%), and a reversal for Ca^{2+} and HCO_3^- (see Table IV). Note that the reversal in selectivity order for F^- and $H_2PO_4^-$ also occurs at PBE0/aug-cc-pVDZ level when considering CP corrected binding energies instead of uncorrected binding energies. Let us note that binding energies for *anions* seem to show a tendency to decrease with increasing basis set size (*e.g.* from PBE0/aug-cc-pVDZ level to PBE0/aug-cc-pVDZ // PBE0/aug-cc-pVTZ level), while an opposite behavior is observed here for the cations Mg^{2+} and Ca^{2+} (the binding energy being stabilized for Na^+ ion).

The selectivity order at the level PBE0/aug-cc-pVDZ // MP2/aug-cc-pVTZ (optimization at PBE0/aug-cc-pVDZ level followed by MP2/aug-cc-pVTZ single point calculations at PBE0/aug-cc-pVDZ optimal geometry) is exactly the same as at the level PBE0/aug-cc-pVDZ (see Table IV), which is very satisfactory. Semi-empirical corrections for the London dispersion interactions (DFT-D3) were not included for the MP2 calculations [52], as MP2 accounts well for these

interactions [81] (part of the MP2 energy correction to Hartree-Fock being dispersion energy).
In conclusion, Table IV validates the use of PBE0/aug-cc-pVDZ as level of theory, which turns out being both computationally tractable and in accordance with other higher-level theories. The binding energies are indeed shown to be sufficiently converged with respect to the basis set at the PBE0/aug-cc-pVDZ level, which justifies the choice of aug-cc-pVDZ basis set in conjunction with this DFT exchange-correlation functional.

Ion	Uncorrected binding energy, Opt PBE0/aug-cc-pVDZ with DFT-D3, with PCM	Uncorrected binding energy, SP PBE0/aug-cc-pVTZ with DFT-D3, with PCM	Uncorrected binding energy, SP MP2/aug-cc-pVTZ at PBE0/aug-cc-pVDZ optimum, with PCM
Mg^{2+}	-1.582	-1.646	-1.718
Na^{+}	-1.004	-1.001	-1.380
ClO^{-}	-0.813	-0.762	-0.849
GLY	-0.803	-0.732	-0.842
SO_4^{2-}	-0.800	-0.701	-0.809
HPO_4^{2-}	-0.772	-0.685	-0.758
AMPA	-0.746	-0.677	-0.742
$COOH^{-}$	-0.727	-0.671	-0.730
$H_2PO_4^{2-}$	-0.645	-0.557	-0.658
F^{-}	-0.640	-0.600	-0.601
HCO_3^{-}	-0.595	-0.509	-0.556
Ca^{2+}	-0.506	-0.538	-0.544
Cl^{-}	-0.465	-0.441	-0.526
Br^{-}	-0.431	-0.410	-0.499
NO_3^{-}	-0.427	-0.397	-0.474
H2O	-0.354	-0.300	-0.320
HClO	-0.218	-0.196	-0.260
NH_4^{+}	-0.110	-0.083	-0.127

TABLE IV: Convergence with basis set size of the uncorrected (for BSSE) binding energies (in eV). Single Point calculations at the PBE0/aug-cc-pVTZ or MP2/aug-cc-pVTZ level have been performed at the complex equilibrium geometries found at the PBE0/aug-cc-pVDZ level (right column). The binding energies at the PBE0/aug-cc-pVDZ level (third column) are found to be stable (within ± 15 %) with increasing basis set size (both using DFT and PBE0 functional, or MP2). The selectivity order is preserved at higher levels of theory.

B. Generalized Kohn-Sham Energy Decomposition Analysis : complete data

The total interaction energy found by the GKS-EDA analysis in our optimized complexes is decomposed into electrostatic, exchange, repulsion, polarization, desolvation free energies, dispersion energy and electron correlation energy, as can be seen from Table V. This total *interaction* energy found by the GKS-EDA method can be compared with the *binding* energy defined and computed in our study (last column of the Table V). As explained section III B, these two measures of ion – probe interaction strength can largely differ due to the choices of cavity and to the deformation energy of the probe upon ion complexation. The total GKS-EDA interaction energy may largely overestimate the interaction strength evaluated in terms of the binding energy.

Complex	Electrostatic free energy	Exchange free energy	Repulsion free energy	Polarization free energy	Desolvation free energy	Dispersion energy	Electron correlation	Total *interaction* energy (CP corrected)	CP corrected binding energy (this study)
F1 – Mg^{2+}	-5.494	-0.905	2.708	-4.867	3.977	-0.063	-0.194	-4.839	-1.537
F1 – Na^{+}	-1.909	-0.470	1.108	-0.821	0.663	-0.063	-0.026	-1.518	-0.967
F1 – ClO^{-}	-2.487	-2.477	4.450	-1.193	0.913	-0.118	-0.367	-1.280	-0.773
F1 – GLY	-2.624	-2.362	4.230	-1.359	1.317	-0.158	-0.269	-1.226	-0.738
F1 – SO_4^{2-}	-2.891	-2.468	4.372	-1.592	1.877	-0.169	-0.358	-1.229	-0.636
F1 – HPO_4^{2-}	-3.053	-2.561	4.583	-1.710	1.778	-0.139	-0.299	-1.402	-0.512
F1 – AMPA	-2.346	-2.251	4.032	-1.164	0.888	-0.160	-0.253	-1.255	-0.687
F1 – $HCOO^{-}$	-2.423	-2.368	4.196	-1.041	0.930	-0.126	-0.298	-1.131	-0.679
F1 – $H_2PO_4^{-}$	-1.981	-1.848	3.228	-0.779	0.782	-0.156	-0.258	-1.011	-0.572
F1 – F^{-}	-2.954	-2.394	4.320	-1.345	1.497	-0.047	-0.295	-1.218	-0.610
F1 – HCO_3^{-}	-1.884	-1.725	3.026	-0.804	0.757	-0.127	-0.249	-1.007	-0.553
F1 – Ca^{2+}	-3.991	-0.806	1.957	-1.686	3.281	-0.131	-0.147	-1.523	-0.478
F1 – Cl^{-}	-1.748	-1.427	2.533	-0.711	0.965	-0.074	-0.305	-0.766	-0.442
F1 – Br^{-}	-1.533	-1.272	2.250	-0.605	0.855	-0.087	-0.304	-0.696	-0.431
F1 – NO_3^{-}	-1.500	-1.222	2.168	-0.590	0.777	-0.126	-0.228	-0.721	-0.392
F1 – H_2O	-0.749	-0.933	1.659	-0.129	-0.139	-0.070	-0.143	-0.504	-0.354
F1 – HClO	-0.440	-0.715	1.285	-0.147	-0.109	-0.100	-0.120	-0.346	-0.187
F1 – NH_4^{+}	-0.360	-0.531	1.008	-0.561	0.317	-0.101	-0.140	-0.369	-0.142

TABLE V: GKS-EDA analysis [1, 59] on the PBE0/aug-cc-pVDZ optimized F1 – ions complexes, in the *all basis set* case, *i.e.* BSSE corrected. Energies are in eV.

More specifically, the following relation holds between the total (CP-corrected) *interaction* energy of the GKS-EDA method and our *binding* energy :

$$\Delta E_{bind}(AB) = \Delta E^{CP}_{int,\mathrm{GKS-EDA}}(AB) + \Delta E_{def}(AB) + \left[\Delta E(A)^{PCM}_{AB}\right]^{BSSE}_{\mathrm{cavity}AB} + \left[\Delta E(B)^{PCM}_{AB}\right]^{BSSE}_{\mathrm{cavity}AB} \quad (9)$$

$$+ \left[\Delta E^{A}_{AB}(A)\right]^{PCM}_{\mathrm{solvent-excluded}A/AB} + \left[\Delta E^{B}_{AB}(B)\right]^{PCM}_{\mathrm{solvent-excluded}B/AB} \quad (10)$$

where $\Delta E_{bind}(AB)$ is the (uncorrected) binding energy defined equation 1 in the main text (section II C), $\Delta E^{CP}_{int,\mathrm{GKS-EDA}}(AB)$ is the counterpoise corrected interaction energy of the complex AB computed by the GKS-EDA method (second column of Table VI), $\Delta E_{def}(AB)$ is the deformation energy (third column of Table VI) defined in section II of the article. Finally, the four terms

$$\left[\Delta E(A)^{PCM}_{AB}\right]^{BSSE}_{\mathrm{cavity}AB} + \left[\Delta E(B)^{PCM}_{AB}\right]^{BSSE}_{\mathrm{cavity}AB} + \left[\Delta E^{A}_{AB}(A)\right]^{PCM}_{\mathrm{solvent-excluded}A/AB} + \left[\Delta E^{B}_{AB}(B)\right]^{PCM}_{\mathrm{solvent-excluded}B/AB}$$

are gathered into a so-called solvent excluded and BSSE energy contribution (fourth column of Table VI), useful to decompose the CP corrected binding energy.

$\left[\Delta E(A)^{PCM}_{AB}\right]^{BSSE}_{\mathrm{cavity}AB}$ and $\left[\Delta E(B)^{PCM}_{AB}\right]^{BSSE}_{\mathrm{cavity}AB}$ are Basis Set Superposition Error terms, computed for each fragment A or B, in the geometry of the complex AB, with the supramolecular cavity of the complex AB, as the difference of PCM energy with the whole basis set (of fragments A and B) with respect to the PCM energy using only the fragment (A or B) basis set :

$$\left[\Delta E(A)^{PCM}_{AB}\right]^{BSSE}_{\mathrm{cavity}AB} = E^{AB}_{AB}(A)^{PCM} - \left[E(A)^{\mathrm{Cavity}AB,\mathrm{Basis}A}_{AB}\right]^{PCM} \quad (11)$$

$$\left[\Delta E(B)^{PCM}_{AB}\right]^{BSSE}_{\mathrm{cavity}AB} = E^{AB}_{AB}(B)^{PCM} - \left[E(B)^{\mathrm{Cavity}AB,\mathrm{Basis}B}_{AB}\right]^{PCM} \quad (12)$$

Finally, $\left[\Delta E(A)^{A}\right]^{PCM}_{\mathrm{solvent-excluded}A/AB}$ and $\left[\Delta E(B)^{B}\right]^{PCM}_{\mathrm{solvent-excluded}A/AB}$ are *solvent-excluded* energies, computed for each fragment, in the complex geometry, with its own basis, as the difference of PCM energy taking the whole supramolecular (AB) cavity with respect to the PCM energy

taking the natural fragment cavity of A or B only :

$$\left[\Delta E^{A}_{AB}(A)\right]^{PCM}_{\mathrm{solvent-excluded}A/AB} = \left[E(A)^{\mathrm{Cavity}AB,\mathrm{Basis}A}_{AB}\right]^{PCM} - E^{A}_{AB}(A)^{PCM} \quad (13)$$

$$\left[\Delta E^{B}_{AB}(B)\right]^{PCM}_{\mathrm{solvent-excluded}B/AB} = \left[E(A)^{\mathrm{Cavity}AB,\mathrm{Basis}B}_{AB}\right]^{PCM} - E^{B}_{AB}(B)^{PCM} \quad (14)$$

This decomposition of Table VI highlights the origins of the differences between the GKS-EDA interaction energy and the binding energy defined and used throughout this study. In particular, the deformation energies of F1 probe in the urea enhanced cation-π complex geometries are much larger (in particular for Mg^{2+}) that their counterparts corresponding to hydrogen-bond complexes with anions. However, the solvent excluded and BSSE energy component accounts for the most striking differences between our CP-corrected binding energy and the CP-corrected GKS-EDA total interaction energy, in particular for Mg^{2+}, for which it is responsible of 85 % of the difference. In the latter case, the choice of a supramolecular cavity (cavity of the complex) for the ion alone, in the complex geometry, instead of its natural cavity (sphere), leads to notably high differences.

Complex	**Total *interaction* energy GKS-EDA (CP corrected)**	**Deformation energy**	**Solvent excluded and BSSE energy**	**CP corrected *binding* energy (this study)**
F1 – Mg^{2+}	-4.839	0.500	2.802	-1.537
F1 – Na^{+}	-1.518	0.155	0.396	-0.967
F1 – ClO^{-}	-1.280	-0.056	0.563	-0.773
F1 – GLY	-1.226	-0.001	0.489	-0.738
F1 – SO_4^{2-}	-1.229	0.044	0.549	-0.636
F1 – HPO_4^{2-}	-1.402	0.093	0.797	-0.512
F1 – AMPA	-1.255	0.047	0.521	-0.687
F1 – $HCOO^{-}$	-1.131	0.027	0.425	-0.679
F1 – $H_2PO_4^{-}$	-1.011	0.052	0.387	-0.572
F1 – F^{-}	-1.218	0.087	0.521	-0.610
F1 – HCO_3^{-}	-1.007	0.085	0.369	-0.553
F1 – Ca^{2+}	-1.523	0.189	0.856	-0.478
F1 – Cl^{-}	-0.766	0.023	0.301	-0.442
F1 – Br^{-}	-0.696	0.017	0.248	-0.431
F1 – NO_3^{-}	-0.721	0.022	0.307	-0.392
F1 – H_2O	-0.504	0.012	0.138	-0.354
F1 – HClO	-0.346	0.015	0.144	-0.187
F1 – NH_4^{+}	-0.369	0.067	0.160	-0.142

TABLE VI: Decomposition of the CP-corrected binding energy in terms of the GKS-EDA [1, 59] CP-corrected interaction energy, the deformation energy and the so-called *solvent excluded and BSSE energy*, defined in detail in the text. Energies are in eV.

C. Other local minima for F1–ions compounds found at PBE0/aug-cc-pVDZ level

We have reported in Table I only the most stable local minima unveiled in this study. However, by testing a wide variety of initial relative positions, we found other stable equilibrium configurations, corresponding to lower binding energies (in absolute value), which are reported in Table VII.

Ion (best local minimum in eV)	**Other local minima : Uncorrected binding energy (eV)**	Binding geometry
Mg^{2+} (-1.582)	-1.519	Interaction with the oxygen atom of urea (1.84 $\AA$ distance) and very weak interaction with the two phenyl groups (about 4 $\AA$ distance)
Mg^{2+} (-1.582)	-1.247	Interaction only with the oxygen atom of urea (1.85 $\AA$ distance)
Na^{+} (-1.004)	-0.886	Interaction with the oxygen atom of urea (2.12 $\AA$ distance) and very weak interaction with the two phenyl groups (about 4 $\AA$ distance)
Mg^{2+} (-1.582)	-0.710	Cation-π interaction with one phenyl group only (ion above aromatic cycle)
GLY (-0.803)	-0.637	Complexation by H bond of the carboxylic moiety of GLY with a single N-H of the urea
SO_4^{2-} (-0.800)	-0.627	H – bonds with one oxygen atom of SO_4^{2-} (1.84 $\AA$ distance)
$HCOO^{-}$ (-0.727)	-0.624	Complexation by H-bond of a single O of HCOO on both N-H of the urea
Na^{+} (-1.004)	-0.593	Cation-π interaction with one phenyl cycle only (2.75 $\AA$ average distance of Na^{+} to carbon atoms), F1 probe planar
Ca^{2+} (-0.506)	-0.414	Weak urea-enhanced cation-π interaction (2.25 $\AA$ distance with oxygen atom of urea and about 4 $\AA$ distance from the closer carbon atom of phenyl cycle)
Ca^{2+} (-0.506)	-0.156	Urea-enhanced cation -π interaction (2.28 $\AA$ distance with oxygen atom of urea and about 3.20 $\AA$ average distance from carbon atoms of phenyl cycle)
HClO (-0.218)	-0.190	H–bond between O–H of HClO (located above urea group) and one nitrogen of urea (1.88 $\AA$)

TABLE VII: List of other local minima found for ions (less stable than those of table I), at the PBE0/aug-cc-pVDZ level. The best local minimum energy unveiled in this study (Table I) is recalled in parenthesis after each ion (in terms of uncorrected binding energy). Energies are in eV.

[1] Peifeng Su, Zhen Jiang, Zuochang Chen, and Wei Wu. Energy decomposition scheme based on the generalized kohn–sham scheme. The Journal of Physical Chemistry A, 118(13):2531–2542, mar 2014.

[2] Renato Pereira Orenha, Vanessa Borges da Silva, Giovanni Finoto Caramori, Felipe Silveira de Souza Schneider, Maurício Jeomar Piotrowski, Julia Contreras-Garcia, Carlos Cardenas, Marina Briese Gonçalves, Fernando Mendizabal, and Renato Luis Tame Parreira. On the recognition of chloride, bromide and nitrate anions by anthracene–squaramide conjugated compounds: a computational perspective. New Journal of Chemistry, 44(41):17831–17839, 2020.

[3] Hemant Sharma, Navneet Kaur, Amanpreet Singh, Anil Kuwar, and Narinder Singh. Optical chemosensors for water sample analysis. Journal of Materials Chemistry C, 4(23):5154–5194, 2016.

[4] Dalmieda and Kruse. Metal cation detection in drinking water. Sensors, 19(23):5134, nov 2019.

[5] Kenji Ueshima. Magnesium and ischemic heart disease: a review of epidemiological, experimental, and clinical evidences. Magnesium Research, 18:275–284, December 2005.

[6] H. M. S. Wasana, D. Aluthpatabendi, W. M. T. D. Kularatne, P. Wijekoon, R. Weerasooriya, and J. Bandara. Drinking water quality and chronic kidney disease of unknown etiology (ckdu): synergic effects of fluoride, cadmium and hardness of water. Environmental Geochemistry and Health, 38, February 2016.

[7] Jean-Paul Behr, Jean-Marc Girodeau, Rodney C. Hayward, Jean-Marie Lehn, and Jean-Pierre Sauvage. Molecular receptors. functionalized and chiral macrocyclic polyethers derived from tartaric acid. Helvetica Chimica Acta, 63(7):2096–2111, oct 1980.

[8] C. J. Pedersen. The discovery of crown ethers. Science, 241(4865):536–540, jul 1988.

[9] Philip A. Gale, Ethan N.W. Howe, and Xin Wu. Anion receptor chemistry. Chem, 1(3):351–422, sep 2016.

[10] Seon-Jin Choi, Bora Yoon, Sibo Lin, and Timothy M. Swager. Functional single-walled carbon nanotubes for anion sensing. ACS Applied Materials & Interfaces, 12(25):28375–28382, jun 2020.

[11] Lloyd C. Murfin, Kirstie Chiang, George T. Williams, Catherine L. Lyall, A. Toby A. Jenkins, Jannis Wenk, Tony D. James, and Simon E. Lewis. A colorimetric chemosensor based on a nozoe azulene that detects fluoride in aqueous/alcoholic media. Frontiers in Chemistry, 8, jan 2020.

[12] Barun Kumar Datta, Chirantan Kar, and Gopal Das. Fluorescent naphthalene-based benzene tripod for selective recognition of fluoride in physiological condition. Journal of Chemical Sciences, 127(2):337–342, 2015.

[13] Alejandro Dorazco-González, Marcos Flores Alamo, Carolina Godoy-Alcántar, Herbert Höpfl, and Anatoly K. Yatsimirsky. Fluorescent anion sensing by bisquinolinium pyridine-2,6-dicarboxamide receptors in water. RSC Adv., 4(1):455–466, 2014.

[14] Sha Ding, Aixiang Xu, Mengyao Li, Aokui Sun, Zhe Zhang, Yong Xia, and Yuejun Liu. Theoretical study on the sensing mechanism of an ON1-OFF-ON2 type fluoride fluorescent chemosensor. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 237:118397, aug 2020.

[15] Hosoowi L., Kyeong-Im H., and Woo-Dong J. Design and applications of molecular probes containing porphyrin derivatives. Coordination Chemistry Reviews, 354:46 – 73, 2018. Recent Progress on Fluorescent Probes.

[16] Mariya Popova, Olexandr Isayev, and Alexander Tropsha. Deep reinforcement learning for de novo drug design. Science Advances, 4(7), 2018.

[17] Ming Wah Wong, Huifang Xie, and Soo Tin Kwa. Anion recognition by azophenol thiourea-based chromogenic sensors: a combined DFT and molecular dynamics investigation. Journal of Molecular Modeling, 19(1):205–213, jul 2012.

[18] Nohad Gresh, G Andrés Cisneros, Thomas A Darden, and Jean-Philip Piquemal. Anisotropic, polarizable molecular mechanics studies of inter-and intramolecular interactions and ligand- macromolecule complexes. a bottom-up strategy. Journal of chemical theory and computation, 3(6):1960–1986, 2007.

[19] Nohad Gresh, Krystel El Hage, Elodie Goldwaser, Benoit de Courcy, Robin Chaudret, David Perahia, Christophe Narth, Louis Lagardère, Filippo Lipparini, and Jean-Philip Piquemal. Addressing the issues of non-isotropy and non-additivity in the development of quantum chemistry-grounded polarizable molecular mechanics. In Challenges and Advances in Computational Chemistry and Physics, pages 1–49. Springer International Publishing, 2015.

[20] Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. Physical review, 136(3B):B864, 1964.

[21] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. Physical Review, 140(4A):A1133–A1138, nov 1965.

[22] B.L. Bhargava and S. Balasubramanian. Probing anion–carbon dioxide interactions in room temperature ionic liquids: Gas phase cluster calculations. Chemical Physics Letters, 444(4-6):242–246, aug 2007.

[23] Benjamin P. Hay, Timothy K. Firman, and Bruce A. Moyer. Structural design criteria for anion hosts: strategies for achieving anion shape recognition through the complementary placement of urea donor groups. Journal of the American Chemical Society, 127(6):1810–1819, feb 2005.

[24] E Cances, Benedetta Mennucci, and J Tomasi. A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics. The Journal of chemical physics, 107(8):3032–3041, 1997.

[25] Vincenzo Barone and Maurizio Cossi. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. The Journal of Physical Chemistry A, 102(11):1995–2001, 1998.

[26] Maurizio Cossi, Nadia Rega, Giovanni Scalmani, and Vincenzo Barone. Energies, structures, and electronic properties of molecules in solution with the c-pcm solvation model. Journal of computational chemistry, 24(6):669–681, 2003.

[27] Scott L. Cockroft. Screening solvent effects in anion recognition. Chem, 3(3):383–384, sep 2017.

[28] Ga Young Lee, Katherine L. Bay, and Kendall N. Houk. Evaluation of DFT methods and implicit solvation models for anion-binding host-guest systems. Helvetica Chimica Acta, 102(5):e1900032, apr 2019.

[29] Melissa S. Caetano, Teodorico C. Ramalho, Douglas F. Botrel, Elaine F. F. da Cunha, and Walclee Carvalho de Mello. Understanding the inactivation process of organophosphorus herbicides: A DFT study of glyphosate metallic complexes with Zn2+, Ca2+, Mg2+, Cu2+, Co3+, Fe3+, Cr3+, and Al3+. International Journal of Quantum Chemistry, 112(15):2752–2762, 2012.

[30] Gaël Zucchi, Bérengère Lebental, Loïc Loisel, Sasikumar Ramachandran, Alfredo Flores Guttierez, Xin Yang Wang, Mallesham Godumala, and Laurence Bodelot. Fr 1753131 - capteurs chimiques à base de nanotubes de carbone fonctionnalisés par des polymères conjugués pour l'analyse en milieu aqueux, January 2017.

[31] Robert Benda, Gaël Zucchi, Eric Cancès, and Bérengère Lebental. Insights into the $\pi - \pi$ interaction driven non-covalent functionalization of carbon nanotubes of various diameters by conjugated fluorene and carbazole copolymers. The Journal of Chemical Physics, 152(6):064708, feb 2020.

[32] Massimo Boiocchi, Laura Del Boca, David Esteban Gómez, Luigi Fabbrizzi, Maurizio Licchelli, and Enrico Monzani. Nature of urea-fluoride interaction: Incipient and definitive proton transfer. Journal of the American Chemical Society, 126(50):16507–16514, dec 2004.

[33] Benjamin Schazmann and Dermot Diamond. Improved nitrate sensing using ion selective electrodes based on urea–calixarene ionophores. New J. Chem., 31(4):587–592, 2007.

[34] J. Nagendra Babu, Vandana Bhalla, Manoj Kumar, Rajiv Kumar Puri, and Rakesh K. Mahajan. Chloride ion recognition using thiourea/urea based receptors incorporated into 1,3-disubstituted calix[4]arenes. New Journal of Chemistry, 33(3):675, 2009.

[35] M. Soylak, F. Armagan Aydin, S. Saracoglu, L. Elci, and M. Dogan. Chemical Analysis of Drinking Water Samples from Yozgat, Turkey. Polish Journal of Environmental Studies, 11(2), 2002.

[36] Lubomír Rulíšek and Zdeněk Havlas. Theoretical studies of metal ion selectivity. 1. DFT calculations of interaction energies of amino acid side chains with selected transition metal ions (co^{2+}, ni^{2+}, cu^{2+},

zn^{2+}, cd^{2+}, and hg^{2+}). Journal of the American Chemical Society, 122(42):10428–10439, oct 2000.
[37] Lubomír Rulíšek and Zdeněk Havlas. Theoretical studies of metal ion selectivity.†2. DFT calculations of complexation energies of selected transition metal ions (co^{2+}, ni^{2+}, cu^{2+}, zn^{2+}, cd^{2+}, and hg^{2+}) in metal-binding sites of metalloproteins. The Journal of Physical Chemistry A, 106(15):3855–3866, apr 2002.
[38] Nathan Schmid, Andreas P. Eichenberger, Alexandra Choutko, Sereina Riniker, Moritz Winger, Alan E. Mark, and Wilfred F. van Gunsteren. Definition and testing of the GROMOS force-field versions 54a7 and 54b7. European Biophysics Journal, 40(7):843–856, apr 2011.
[39] Walter R. P. Scott, Philippe H. Hünenberger, Ilario G. Tironi, Alan E. Mark, Salomon R. Billeter, Jens Fennen, Andrew E. Torda, Thomas Huber, Peter Krüger, and Wilfred F. van Gunsteren. The GROMOS biomolecular simulation program package. The Journal of Physical Chemistry A, 103(19):3596–3607, may 1999.
[40] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E Mark, and Herman JC Berendsen. Gromacs: fast, flexible, and free. Journal of computational chemistry, 26(16):1701–1718, 2005.
[41] Alpeshkumar K. Malde, Le Zuo, Matthew Breeze, Martin Stroet, David Poger, Pramod C. Nair, Chris Oostenbrink, and Alan E. Mark. An automated force field topology builder (ATB) and repository: Version 1.0. Journal of Chemical Theory and Computation, 7(12):4026–4037, nov 2011.
[42] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-d) for the 94 elements h-pu. The Journal of Chemical Physics, 132(15):154104, apr 2010.
[43] Carlo Adamo and Vincenzo Barone. Toward reliable density functional methods without adjustable parameters: The PBE0 model. The Journal of Chemical Physics, 110(13):6158–6170, apr 1999.
[44] Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. The Journal of Chemical Physics, 90(2):1007–1023, jan 1989.
[45] P.C. Hariharan and J.A. Pople. Accuracy of AHn equilibrium geometries by single determinant molecular orbital theory. Molecular Physics, 27(1):209–214, jan 1974.
[46] Michelle M. Francl, William J. Pietro, Warren J. Hehre, J. Stephen Binkley, Mark S. Gordon, Douglas J. DeFrees, and John A. Pople. Self-consistent molecular orbital methods. XXIII. a polarization-type basis set for second-row elements. The Journal of Chemical Physics, 77(7):3654–3665, oct 1982.
[47] Jean-Philippe Blaudeau, Mark P. McGrath, Larry A. Curtiss, and Leo Radom. Extension of gaussian-2 (g2) theory to molecules containing third-row atoms k and ca. The Journal of Chemical Physics, 107(13):5016–5021, oct 1997.
[48] Vitaly A. Rassolov, Mark A. Ratner, John A. Pople, Paul C. Redfern, and Larry A. Curtiss. 6-31g* basis set for third-row atoms. Journal of Computational Chemistry, 22(9):976–984, 2001.
[49] Marcia P. M. Costa, Letícia M. Prates, Leonardo Baptista, Maurício T. M. Cruz, and Ivana L. M. Ferreira. Interaction of polyelectrolyte complex between sodium alginate and chitosan dimers with a single glyphosate molecule: A DFT and NBO study. Carbohydrate Polymers, 198:51–60, October 2018.
[50] Mandeep, Archa Gulati, and Rita Kakkar. DFT study of adsorption of glyphosate pesticide on Pt-Cu decorated pyridine-like nitrogen-doped graphene. Journal of Nanoparticle Research, 22(1):17, January 2020.
[51] Alston J. Misquitta, Anthony J. Stone, and Farhang Fazeli. Distributed multipoles from a robust basis-space implementation of the iterated stockholder atoms procedure. Journal of Chemical Theory and Computation, 10(12):5405–5418, nov 2014.
[52] Chr Møller and Milton S Plesset. Note on an approximation treatment for many-electron systems. Physical review, 46(7):618, 1934.
[53] A van Bondi. van der waals volumes and radii. The Journal of physical chemistry, 68(3):441–451, 1964.
[54] Peifeng Su and Hui Li. Continuous and smooth potential energy surface for conductor-like screening solvation model using fixed points with variable areas. The Journal of Chemical Physics, 130(7):074109, feb 2009.
[55] Jon Baker, Alain Kessi, and Bernard Delley. The generation and use of delocalized internal coordinates in geometry optimization. The Journal of Chemical Physics, 105(1):192–212, jul 1996.
[56] Aleksandr V Marenich, Christopher J Cramer, and Donald G Truhlar. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. The Journal of Physical Chemistry B, 113(18):6378–6396, 2009.
[57] Michael S. Marshall, Ryan P. Steele, Kanchana S. Thanthiriwatte, and C. David Sherrill. Potential energy curves for cation-π interactions: Off-axis configurations are also attractive. The Journal of Physical Chemistry A, 113(48):13628–13632, dec 2009.
[58] Michael W. Schmidt, Kim K. Baldridge, Jerry A. Boatz, Steven T. Elbert, Mark S. Gordon, Jan H. Jensen, Shiro Koseki, Nikita Matsunaga, Kiet A. Nguyen, Shujun Su, Theresa L. Windus, Michel Dupuis, and John A. Montgomery. General atomic and molecular electronic structure system. Journal of Computational Chemistry, 14(11):1347–1363, nov 1993.
[59] Peifeng Su, Zhen Tang, and Wei Wu. Generalized kohn-sham energy decomposition analysis and its

applications. WIREs Computational Molecular Science, 10(5), jan 2020.
[60] Giuseppe MJ Barca, Colleen Bertoni, Laura Carrington, Dipayan Datta, Nuwan De Silva, J Emiliano Deustua, Dmitri G Fedorov, Jeffrey R Gour, Anastasia O Gunina, Emilie Guidez, et al. Recent developments in the general atomic and molecular electronic structure system. The Journal of Chemical Physics, 152(15):154102, 2020.
[61] https://www.samson-connect.net/.
[62] Dan Shen, Peifeng Su, and Wei Wu. What kind of neutral halogen bonds can be modulated by solvent effects? Physical Chemistry Chemical Physics, 20(41):26126–26139, 2018.
[63] S.F. Boys and F. Bernardi. The calculation of small molecular interactions by the differences of separate total energies. some procedures with reduced errors. Molecular Physics, 19(4):553–566, oct 1970.
[64] Eric Cancès and Benedetta Mennucci. The escaped charge problem in solvation continuum models. The Journal of Chemical Physics, 115(13):6130–6135, oct 2001.
[65] Marco Lorenz, Bartolomeo Civalleri, Lorenzo Maschio, Mauro Sgroi, and Daniele Pullini. Benchmarking dispersion and geometrical counterpoise corrections for cost-effective large-scale DFT calculations of water adsorption on graphene. Journal of Computational Chemistry, 35(24):1789–1800, 2014.
[66] Ł. M. Mentel and E. J. Baerends. Can the Counterpoise Correction for Basis Set Superposition Effect Be Justified? Journal of Chemical Theory and Computation, 10(1):252–267, January 2014.
[67] Brina Brauer, Manoj K. Kesharwani, and Jan M. L. Martin. Some Observations on Counterpoise Corrections for Explicitly Correlated Calculations on Noncovalent Interactions. Journal of Chemical Theory and Computation, 10(9):3791–3799, September 2014.
[68] Stefan Grimme. Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory. Chemistry - A European Journal, 18(32):9955–9964, August 2012.
[69] Ioan Andricioaei and Martin Karplus. On the calculation of entropy from covariance matrices of the atomic fluctuations. The Journal of Chemical Physics, 115(14):6289–6292, 2001.
[70] Peifeng Su, Hui Liu, and Wei Wu. Free energy decomposition analysis of bonding and nonbonding interactions in solution. The Journal of Chemical Physics, 137(3):034111, jul 2012.
[71] Bogumil Jeziorski, Robert Moszynski, and Krzysztof Szalewicz. Perturbation theory approach to intermolecular potential energy surfaces of van der waals complexes. Chemical Reviews, 94(7):1887–1930, nov 1994.
[72] Juan Andrés, Paul W. Ayers, Roberto A. Boto, Ramon Carbó-Dorca, Henry Chermette, Jerzy Cioslowski, Julia Contreras-García, David L. Cooper, Gernot Frenking, Carlo Gatti, Farnaz Heidar-Zadeh, Laurent Joubert, Ángel Martín Pendás, Eduard Matito, István Mayer, Alston J. Misquitta, Yirong Mo, Julien Pilmé, Paul L. A. Popelier, Martin Rahm, Eloy Ramos-Cordoba, Pedro Salvador, W. H. Eugen Schwarz, Shant Shahbazian, Bernard Silvi, Miquel Solà, Krzysztof Szalewicz, Vincent Tognetti, Frank Weinhold, and Émilie-Laure Zins. Nine questions on energy decomposition analysis. Journal of Computational Chemistry, 40(26):2248–2283, jun 2019.
[73] Abhishek Shahi and Elangannan Arunan. Why are hydrogen bonds directional? Journal of Chemical Sciences, 128(10):1571–1577, 2016.
[74] Blakely W. Tresca, Ryan J. Hansen, Calvin V. Chau, Benjamin P. Hay, Lev N. Zakharov, Michael M. Haley, and Darren W. Johnson. Substituent effects in CH hydrogen bond interactions: Linear free energy relationships and influence of anions. Journal of the American Chemical Society, 137(47):14959–14967, nov 2015.
[75] Jennifer C. Ma and Dennis A. Dougherty. The cation-π interaction. Chemical Reviews, 97(5):1303–1324, aug 1997.
[76] J. P. Gallivan and D. A. Dougherty. Cation-pi interactions in structural biology. Proceedings of the National Academy of Sciences, 96(17):9459–9464, aug 1999.
[77] Dennis A. Dougherty. The cation-π interaction. Accounts of Chemical Research, 46(4):885–893, dec 2012.
[78] Kiran Kumar, Shin M. Woo, Thomas Siu, Wilian A. Cortopassi, Fernanda Duarte, and Robert S. Paton. Cation–π interactions in protein–ligand binding: theory and data-mining reveal different roles for lysine and arginine. Chemical Science, 9(10):2655–2665, 2018.
[79] Klaus R. Liedl. Dangers of counterpoise corrected hypersurfaces. advantages of basis set superposition improvement. The Journal of Chemical Physics, 108(8):3199–3204, 1998.
[80] Kwang-Hwi Cho, Kyoung Tai No, and Harold A Scheraga. Ion pair interactions in aqueous solution: Self-consistent reaction field (scrf) calculations with some explicit water molecules. The Journal of Physical Chemistry A, 104(27):6505–6509, 2000.
[81] Stefan Grimme. Density functional theory with london dispersion corrections. Wiley Interdisciplinary Reviews: Computational Molecular Science, 1(2):211–228, mar 2011.