

# Supporting Information for "Developing force fields when experimental data is sparse: AMBER/GAFF-compatible parameters for inorganic and alkyl oxoanions"

Sadra Kashefolgheta,<sup>‡</sup> and Ana Vila Verde,<sup>\*‡</sup>

## 1 Instruction

### 1.1 AMBER users

The parameters are provided in the \*.top files, which is a gromacs format file. These files contain the original as well as the optimized values in both gromacs and amber formats. For amber users, the amber atom types are given as comments in the \*.top files (";" character is used to comment). You should ignore our suggested names for atom types, since they were chosen to be compatible with gromacs naming conventions. Keep in mind that in many cases, the amber users need to define a new LJtype instead, as described below.

All the parameters for the bonds, angles and dihedrals are retained from original amber force field. The charges and masses are given in the \*.itp files of the corresponding ions, under the [atoms] section.

Amber users are first supposed to prepare their own \*.prmtop and \*.inpcrd files using LEaP or similar programs, and then load the \*.prmtop file into the topology file editor, ParmEd. You can use ParmEd to modify the self-interactions, called *SinglType*, as well as the pairwise interactions, called *PairType*.

\*\*\*Warning!!!\*\*\* This study is intended to provide parameters for specific anions and cations. If you have atoms in neutral groups or atoms in ionic groups other than those optimized in this study, and they share the same atom types (i.e. O, O2, OH, OS, N3, H) with the anions and cations that we parameterized, you first need to create new van der Waals atom types, using "addLJType" command in the LEaP, to avoid unnecessary complications. As an example, to add new LJtype for atoms 3, 5, 6 and 9, the command line should read:

---

<sup>‡</sup> Department of Theory & Bio-systems, Max Planck Institute for Colloids and Interfaces, Science Park, Potsdam 14476 Germany. Fax: +49 (0)331 567 9602; Tel: +49 (0)331 567 9608; E-mail: ana.vilaverde@mpikg.mpg.de

```
addLJType @3,5,6,9
```

where @ is a keyword to select an atom number, in this case atoms 3,5,6 and 9. The command will assign r and epsilon values of the first selected atom, here atom 3, to the rest of atoms, and give them a new LJtype.

To change the self interaction parameters, you can use the "changeLJSingleType" command. This will change the interaction of a <mask> (atom selection) with itself as well as with all other atomtypes. Note that, all the atoms in the <mask> should have the same atom types, otherwise it will raise an error. As an example, to change the self-interaction of atoms 3, 5, 6 and 9, to which we just assigned a new atom type, the command line should look like:

```
changeLJSingleType @3 r(Å) epsilon(kcal/mol)
```

where r and epsilon are the radius and Lennard-Jones well depth, respectively, and are given in the \*.top files provided by us, under the [ *atomtypes* ] section. It is important to note that, we only need to select atom 3, to apply the change to the rest of the atoms with the same LJtype. Therefore, here applying the change to atom 3 would effectively apply the change to the atoms 5, 6 and 9.

To change the pairwise interactions (i.e. anion-cation interactions in this study), one can use "changeLJPair" command. As an example, to change the pairwise interaction between atoms @3,5,6,9 – all of which have the same LJtype – and all sodium ions, the command line should read:

```
changeLJPair @3 :Na+ Rij(Å) epsilonij(kcal/mol)
```

where ":Na+" is to select all Na+ residues, R<sub>ij</sub> is the minimum distance between i and j, and epsilon<sub>ij</sub> is the LJ well depth between atoms i and j. The R<sub>ij</sub> and epsilon<sub>ij</sub> values are given in the \*.top files provided by us, under the [ *nonbond\_params* ] section.

Detailed instructions on how to use ParmEd can be found in the amber manual.

## 1.2 Gromacs users

The original and optimized parameters are provided in \*.top files along with the corresponding \*.itp files. Since this study is intended to provide parameters for specific anions and cations, we should avoid unintended changes to the atoms in neutral groups or atoms in ionic groups other than those optimized in this study, that share the same atom types (i.e. O, O2, OS, OH, N3, H) with the anions and cations that we parameterized. Therefore, for each optimized parameter a new atom type is introduced (e.g. O is replaced with OSUL for  $\text{SO}_4^{2-}$ ). We also changed the atom types N3 and H of  $\text{NH}_4^+$  and  $\text{CH}_3\text{NH}_3^+$  to (NAMO, HAMO) and (NAMM, HAMM) respectively, to distinguish between the two cations. We recommend that you change the atom type of amine head groups in your simulation to (NAMM, HAMM), since often there are some other atoms with type N3 or H, which should not be changed (e.g. H type in guanidinium).

## 2 Methods

### 2.1 Correction terms to the solvation free energies

The free energy contributions arising from Coulombic interactions calculated from simulations ( $\Delta G_{chg}$ ) are known to vary markedly depending on the method by which electrostatic interactions are calculated<sup>1-5</sup>. To avoid this method dependence it is necessary to apply multiple correction terms to the calculated values so that they can be compared with experimental free energies of solvation. The largest of these terms is the Wigner correction term for the ion in the gas phase; this term is already included in the free energy values calculated using Gromacs.<sup>6,7</sup>

Another important correction term eliminates the error incurred on by using particle-based summation schemes when calculating the long-range electrostatic interactions as Gromacs<sup>6,7</sup> does, instead of using the more adequate molecule-based summation schemes. This term is not applied here, however, because it varies linearly with the charge of the ion and, because we are concerned only with *differences* in hydration free energy (see equation M5; “M” indicates an object that can be found in the main text of the article), this term cancels out.

The free energies calculated using our chosen setup are intrinsic: they do not contain a contribution of crossing the macroscopic air-water interface<sup>8</sup>. However, often compilations of hydration free energies from experiment include this contribution – they are referred to as real free energies of hydration of single ions<sup>8</sup>. Because the surface potential term is linear in the charge of the ion and thus cancels out in equation M5, it can be ignored in our approach, independently of

which experimental data set is used for parameterization.

The free energy values calculated from simulations correspond to the transfer of an ion in a given volume in the gas to the same volume of liquid. In contrast, the experimental values correspond to a process where the ion is transferred from an ideal gas at 1 atm to an ideal aqueous solution at 1 M concentration, i.e., the experimental values of the solvation free energy contain an entropic term that is absent from simulation. This entropic term,  $\Delta G_{press}$ , has the form:

$$\Delta G_{press} = -T\Delta S_{press} = -k_B T \ln \left( \frac{V_l}{V_g} \right) \quad (1)$$

where  $V_l = 1$  l/mol is the molar volume of an ideal aqueous solution of concentration 1 M and  $V_g = 24.465$  l/mol is that of an ideal gas at 1 atm. This term will not cancel out in the difference shown in equation M5, so  $\Delta G_{press} = 1.888$  kcal/mol must be added to the solvation free energy values calculated using simulations.

Transferring a molecule from the gas phase to the solution phase changes its electronic density. A polarization correction is also included in calculating solvation free energies to take this change into consideration. This correction was calculated using ab initio methods as follows. The polarization energies were calculated in Gaussian 03. The structures are first optimized using the IEF-PCM<sup>9-13</sup> model with the B3LYP<sup>14-16</sup>/aug-cc-pVTZ level of theory, followed by a single point energy calculation in vacuum, using the same level of theory. This energy corresponds to a solvated structure (more precisely, electron density,  $\rho_{solv}$ ), measured in gas phase,  $E^{gas}[\rho_{solv}]$ . Thereafter, we re-optimize the molecules in the gas phase to obtain the minimum energies in the vacuum,  $E^{gas}[\rho_{gas}]$ . Subtracting the former from the latter energy will give us good estimates of the polarization energies:<sup>17</sup>

$$\Delta G_{pol} = E^{gas}[\rho_{solv}] - E^{gas}[\rho_{gas}] \quad (2)$$

The default values in Gaussian 03 for the dielectric constant, fast response dielectric constant ( $\epsilon_{inf}$ ) and probe radius ( $R_{solv}$ ) are 78.355, 1.776 and 1.385 Å, respectively for this model.

All the correction terms that were applied in this work to the calculated  $\Delta G_{solv}$  are summarized in equation 3.

$$\Delta G_{solv}^{Comp} = \Delta G_{cav} + \Delta G_{chg} + \Delta G_{pol} + \Delta G_{press} \quad (3)$$

## 2.2 Activity Derivative Calculations Using Ion-Pair Potentials

The molar activity derivative is defined as,

$$a_{cc} = \left. \frac{\partial \ln a_c}{\partial \ln \rho_c} \right|_{P,T} \quad (4)$$

where  $a_c$  is the solution molar activity and  $\rho_c$  is the number density of the solution, defined as  $\rho_c = n_c/V = (n_+ + n_-)/V$ ;  $n_+$  and  $n_-$  are the number of cations and anions, respectively, in the solution volume  $V$ . Using simulations, this derivative can be easily calculated from the radial distribution functions (rdfs) in dilute solutions, as formulated in the Kirkwood-Buff (KB) theory and described in detail elsewhere.<sup>18-23</sup> The original KB theory is developed for the grand canonical ensemble,  $\mu VT$ ; in this ensemble, the KB integrals are defined as:

$$G_{ij} = 4\pi \int_0^\infty r^2 [g_{ij}^{\mu VT}(r) - 1] dr \quad (5)$$

where  $g_{ij}(r)$  are the rdfs of the two species  $i$  and  $j$  and  $r$  is the distance between them. To avoid the complexity of using a grand canonical ensemble in simulations, we use an alternative formulation of the KB integrals, defined in terms of rdfs,  $g_{ij}^{NVT}(r)$ , obtained in the canonical ensemble<sup>24,25</sup>:

$$G_{ij} = \int_0^{2R} [f_{ij} g_{ij}^{\mu VT}(r) - 1] 4\pi r^2 (1 - 3x/2 + x^3/2) dr \quad (6)$$

where  $x = r/(2R)$  and  $f_{ij}$  is a concentration-dependent correction factor to the radial distribution function.<sup>26-28</sup> The interactions in the simulation box can be divided into two subsets: water(w)-co-solvent(c) interactions and co-solvent(c)-co-solvent(c) interactions. The corresponding KB integral terms are given by:

$$G_{cc} = \frac{1}{4}(2G_{+-} + G_{++} + G_{--}) \quad (7)$$

and

$$G_{cw} = G_{w+} + G_{w-} \quad (8)$$

Using these expressions for KB integrals the molar activity derivative of the solution is given by:

$$a_{cc} = \frac{1}{1 + \rho_c(G_{cc} - G_{cw})} \quad (9)$$

The activity derivative obtained this way can be compared to the molar activity derivative obtained from experiment, calculated using equation 4. Experimental activity derivatives are typically reported in the molal scale, so a correction factor must be applied to the molar activity derivatives to allow for comparisons with the equivalent molal quantity. At a

0.5 m salt concentration, however, the correction factor is very small and can safely be ignored.

### 2.3 Potential energy scan – counterpoise correction

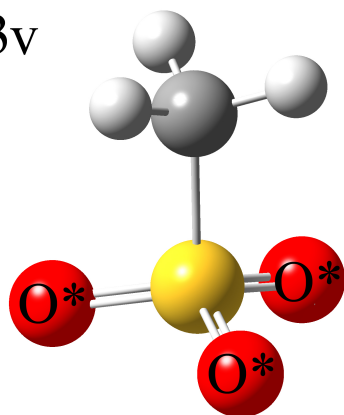
When studying ion-ion interactions with a finite basis set, the basis set superposition error (BSSE)<sup>29</sup> could become significantly large. The BSSE occur when two species A and B are approaching each other and form a complex A-B. The resulting A-B complex is additionally stabilized due to the fact that each of the A and B are using extra basis functions from one another.<sup>30</sup> However, the overlap integral decreases as the distance between A and B increases, and becomes zero when the two species beyond a few angstroms apart; therefore, the total basis functions used vary as a function of the intermolecular distance between A and B. Under these circumstances, the binding energy is calculated as  $\Delta E_{Binding} = E^{AB}(AB) - E^A(A) - E^B(B)$ , where the superscripts are the basis functions used for the energy calculations. Using varying basis functions during 1D potential energy scans thus leads to artifacts in the binding energy and the minimum distance between the two ions. To resolve this problem we use the Boys and Bernardi<sup>31</sup> counterpoise correction (CP), which is shown to correct for the basis set superposition error by using the AB basis to calculate the energies for each of A and B. The expression for the binding energy used in this work, which includes the CP correction, is thus:

$$\Delta E_{Binding}^{CP} = E^{AB}(AB) - E^{AB}(A) - E^{AB}(B) \quad (10)$$

### 2.4 Symmetry and QM Potential Scan

Depending on the symmetry of the anions, different anion-oxygens could experience different chemical environments. Figure 1 illustrates molecules with different symmetry point groups. When performing potential energy scans for the interaction of these ions with each of water, sodium,  $\text{NH}_4^+$  and  $\text{CH}_3\text{NH}_3^+$ , the three oxygens of a  $C_{3v}$  molecule, Figure 1a, would yield the same result. However, for the molecule with  $C_s$  symmetry, Figure 1b, two of the oxygens are similar and the third oxygen interacts differently with water and the counter ions. However, in classical force fields, all terminal oxygens are treated as equal and are assigned identical LJ parameters. To resolve this discrepancy between quantum and classical descriptions of these anions, we performed the potential energy scans for every distinct oxygen in the anions. The final  $\Delta E_{QM \rightarrow cl}(R_m)$  and  $\Delta R(\text{min})_{QM \rightarrow cl}$  values are obtained as an average of different scans.

a)  $C_{3v}$



b)  $C_s$

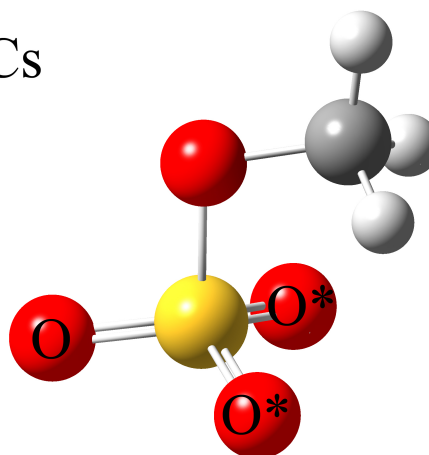


Fig. 1 The three oxygens in a) experience similar chemical environments due to  $C_{3v}$  symmetry. The molecules with geometries similar to b) have only two similar oxygens and the third one is distinguishable, due to lower symmetry. Oxygens labeled with \* experience similar chemical environments within that molecule.

### 3 Results

#### 3.1 Effect of Orientation on Water-Ion Interactions

We examined whether our proposed approach would yield consistent results if potential energy scans in other orientations were used. The comparison between the two different orientations, listed in Table 1, shows that water-ion interactions could be optimized regardless of the orientation of water around the ions, using our approach. The values of  $\Delta E_{QM \rightarrow cl}$  for the reference species ( $\text{HSO}_4^-$ ) are essentially independent of the orientation (I or II; see Figure 2), and the same is observed for the other species ( $\text{H}_2\text{PO}_4^-$  and  $(\text{CH}_3)_2\text{PO}_4^-$ ) parameterized based on the reference species. It is important to note that the absolute minimum energies are different in different orientations. However, since the method is developed to capture the differences in the energies  $\Delta E$ , the parameterization could be done regardless of the orientation.

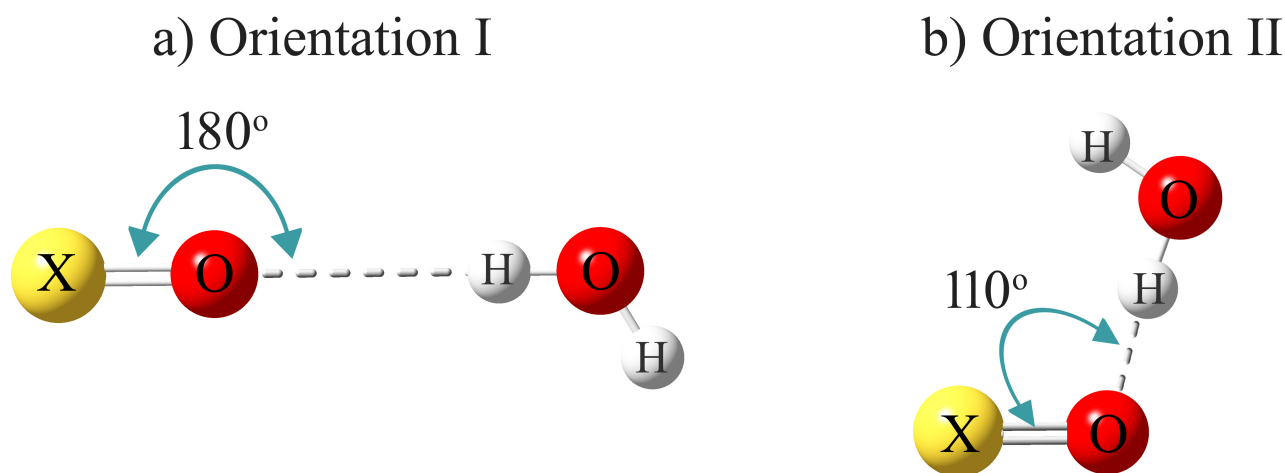


Fig. 2 The two orientations of water around anions. X=S,P

Table 1 Differences between the classical and quantum potential energy scans for the mentioned anion-water dimers, in two different orientations. R denotes the distance between the two closest nuclei of the two species.

Orientation	Anion	Optimization <sup>a</sup>	Parameters <sup>b</sup>	$E_{QM}(R_m)$ kcal/mol	$E_{Classic}(R_m)$ kcal/mol	$\Delta E_{QM \rightarrow cl}^c(R_m)$ kcal/mol	$\Delta R_{m, QM \rightarrow cl}^d$ Å
I	$\text{HSO}_4^-$	Exp	.. <sup>e</sup>	-10.4	-11.7 <sup>f</sup>	-1.3 <sup>f</sup>	-0.14 <sup>f</sup>
	$\text{H}_2\text{PO}_4^-$	Comp	.. <sup>e</sup>	-13.2	-14.6 <sup>f</sup>	-1.4 <sup>f</sup>	-0.12 <sup>f</sup>
	$(\text{CH}_3)_2\text{PO}_4^-$	Comp	.. <sup>e</sup>	-13.4	-14.5 <sup>f</sup>	-1.1 <sup>f</sup>	-0.11 <sup>f</sup>
II	$\text{HSO}_4^-$	Exp	.. <sup>e</sup>	-12.4	-13.6 <sup>f</sup>	-1.2 <sup>f</sup>	-0.14 <sup>f</sup>
	$\text{H}_2\text{PO}_4^-$	Comp	.. <sup>e</sup>	-14.1	-15.2 <sup>f</sup>	-1.1 <sup>f</sup>	-0.14 <sup>f</sup>
	$(\text{CH}_3)_2\text{PO}_4^-$	Comp	.. <sup>e</sup>	-14.6	-15.7 <sup>f</sup>	-1.1 <sup>f</sup>	-0.14 <sup>f</sup>

<sup>a</sup> Parameters optimized based on hydration free energies (Exp), repeated here to facilitate comparisons, or based on the ab initio approach (Comp); <sup>b</sup> Expressed as a scaling factor relative to the original GAFF parameters (given in the Supporting Information); <sup>c</sup>  $\Delta E_{QM \rightarrow cl}(R_m)$  is defined in Equation M2; <sup>d</sup>  $\Delta R_{m, QM \rightarrow cl}$ : the difference between the position of the dimer energy minimum obtained using ab initio calculations and using classical parameters. <sup>e</sup> No scaling; <sup>f</sup> GAFF parameters; \* X=S,P



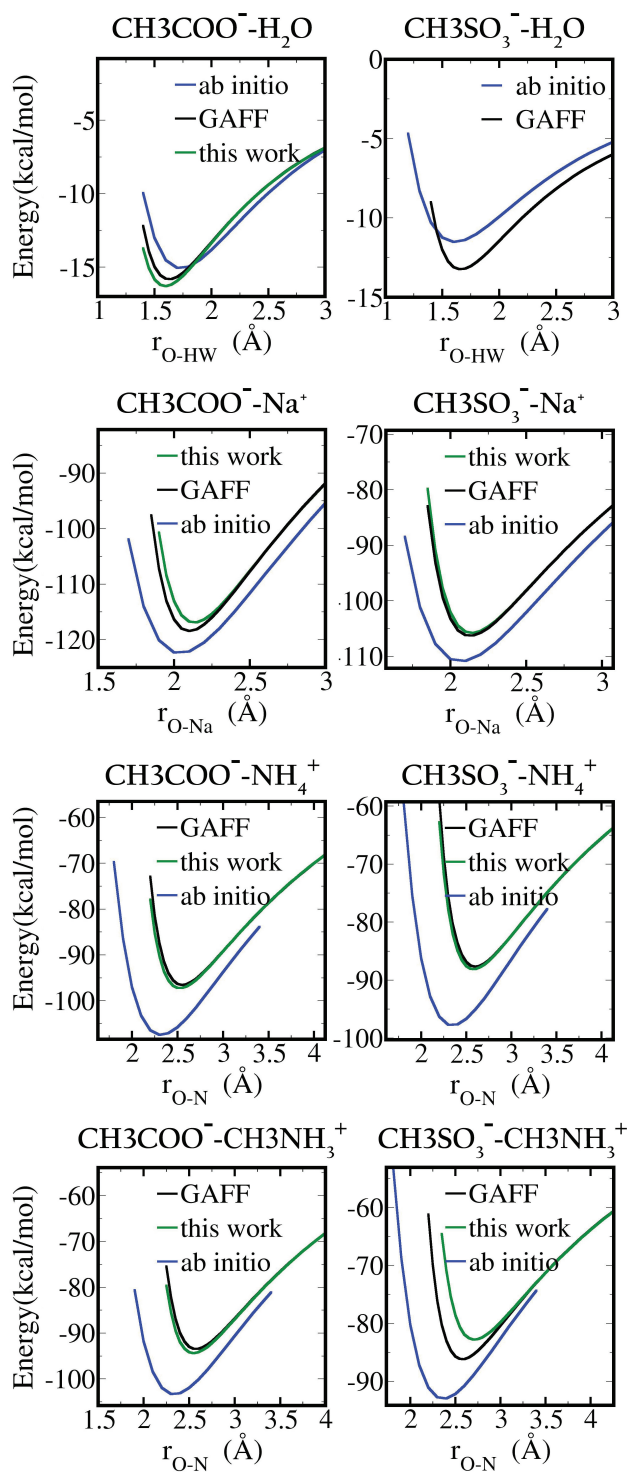


Fig. 3 Classical and ab initio potential energy scans for the interaction of  $\text{CH}_3\text{COO}^-$  and  $\text{CH}_3\text{SO}_3^-$  with each of  $\text{H}_2\text{O}$ ,  $\text{Na}^+$ ,  $\text{NH}_4^+$  and  $\text{CH}_3\text{NH}_3^+$ . Both GAFF/AMBER (black) parameters and optimized parameters from this study (green), are plotted.

**Table 2** Differences between the classical and quantum potential energy scans for all anion-NH<sub>4</sub><sup>+</sup> and anion-CH<sub>3</sub>NH<sub>3</sub><sup>+</sup> dimers, and values of the optimized anion-cation parameters. For anions in which the oxygens are not symmetrically identical, as discussed in section 2.4, separate scans were performed for oxygens (O) and (O\*); see Figure 4. R denotes the distance between the two closest nuclei of the two species; refer to Figure M2 and accompanying text for the orientations.

Anion	Counter Species	$E_{QM}(R_m)$ kcal/mol	$E_{Classic}^{GAFF}(R_m)$ kcal/mol	$E_{Classic}^{OPT}(R_m)$ kcal/mol	$R_{m,QM}$ Å	$R_{m,Classic}^{GAFF}$ Å	$R_{m,Classic}^{OPT}$ Å
CH <sub>3</sub> SO <sub>3</sub> <sup>-</sup>	H <sub>2</sub> O	-11.5	-13.2	.. <sup>a</sup>	1.81	1.66	.. <sup>a</sup>
	Na <sup>+</sup>	-110.9	-106.3	-105.9	2.08	2.13	2.14
	NH <sub>4</sub> <sup>+</sup>	-97.8	-88.1	-87.7	2.35	2.58	2.60
	CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup>	-93.0	-86.2	-82.8	2.37	2.58	2.71
CH <sub>3</sub> SO <sub>4</sub> <sup>-</sup> (O)	H <sub>2</sub> O	-10.3	-11.8	.. <sup>a</sup>	1.85	1.70	.. <sup>a</sup>
	Na <sup>+</sup>	-108.7	-105.0	-104.6	2.09	2.14	2.14
	NH <sub>4</sub> <sup>+</sup>	-96.4	-88.1	-87.3	2.35	2.59	2.62
	CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup>	-92.7	-85.9	-82.7	2.37	2.58	2.72
CH <sub>3</sub> SO <sub>4</sub> <sup>-</sup> (O*)	H <sub>2</sub> O	-10.6	-11.9	.. <sup>a</sup>	1.84	1.70	.. <sup>a</sup>
	Na <sup>+</sup>	-106.5	-101.3	-100.9	2.09	2.14	2.15
	NH <sub>4</sub> <sup>+</sup>	-94.1	-84.7	-83.9	2.35	2.59	2.63
	CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup>	-90.5	-82.7	-79.5	2.37	2.59	2.72
H <sub>2</sub> PO <sub>4</sub> <sup>-</sup> (O)	H <sub>2</sub> O	-13.0	-14.8	.. <sup>a</sup>	1.77	1.64	.. <sup>a</sup>
	Na <sup>+</sup>	-118.1	-118.7	-116.4	2.06	2.10	2.15
	NH <sub>4</sub> <sup>+</sup>	-103.8	-97.3	-95.7	2.35	2.55	2.61
	CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup>	-99.8	-95.1	-92.0	2.36	2.55	2.65
H <sub>2</sub> PO <sub>4</sub> <sup>-</sup> (O*)	H <sub>2</sub> O	-13.4	-14.5	.. <sup>a</sup>	1.76	1.64	.. <sup>a</sup>
	Na <sup>+</sup>	-110.6	-104.0	-101.9	2.07	2.12	2.17
	NH <sub>4</sub> <sup>+</sup>	-96.0	-84.6	-83.3	2.37	2.58	2.63
	CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup>	-92.3	-82.9	-80.2	2.38	2.58	2.68
(CH <sub>3</sub> ) <sub>2</sub> PO <sub>4</sub> <sup>-</sup> (O)	H <sub>2</sub> O	-13.2	-14.4	.. <sup>a</sup>	1.76	1.65	.. <sup>a</sup>
	Na <sup>+</sup>	-117.5	-114.4	-113.3	2.06	2.11	2.13
	NH <sub>4</sub> <sup>+</sup>	-103.1	-94.0	-93.3	2.35	2.56	2.58
	CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup>	-99.1	-91.9	-88.9	2.36	2.56	2.66
(CH <sub>3</sub> ) <sub>2</sub> PO <sub>4</sub> <sup>-</sup> (O*)	H <sub>2</sub> O	-13.5	-14.6	.. <sup>a</sup>	1.75	1.64	.. <sup>a</sup>
	Na <sup>+</sup>	-113.9	-109.4	-108.3	2.07	2.11	2.14
	NH <sub>4</sub> <sup>+</sup>	-99.6	-89.8	-89.2	2.35	2.56	2.58
	CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup>	-95.7	-87.9	-85.0	2.37	2.56	2.67
CH <sub>3</sub> COO <sup>-</sup>	H <sub>2</sub> O	-15.0	-15.8	-16.3	1.73	1.63	1.59
	Na <sup>+</sup>	-122.4	-118.5	-117.0	2.04	2.10	2.13
	NH <sub>4</sub> <sup>+</sup>	-107.4	-96.5	-97.2 <sup>b</sup>	2.32	2.55	2.53 <sup>b</sup>
	CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup>	-103.3	-94.4	-93.4	2.34	2.55	2.57
SO <sub>4</sub> <sup>2-</sup>	H <sub>2</sub> O	-20.7	-19.6	-15.9	1.67	1.61	1.88
	Na <sup>+</sup>	-196.9	-191.8	-174.5	1.98	2.05	2.33
	NH <sub>4</sub> <sup>+</sup>	-180.6	-163.6	-150.2	2.22	2.49	2.82
	CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup>	-175.4	-159.6	-145.4	2.23	2.49	2.85
CH <sub>3</sub> PO <sub>4</sub> <sup>2-</sup> (O)	H <sub>2</sub> O	-22.4	-21.2	-17.7	1.65	1.59	1.81
	Na <sup>+</sup>	-200.3	-197.1	-178.2	1.99	2.05	2.34
	NH <sub>4</sub> <sup>+</sup>	-182.9	-166.9	-152.8	2.23	2.48	2.81
	CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup>	-177.7	-162.9	-146.9	2.24	2.48	2.86
CH <sub>3</sub> PO <sub>4</sub> <sup>2-</sup> (O*)	H <sub>2</sub> O	-22.7	-21.3	-17.8	1.64	1.59	1.80
	Na <sup>+</sup>	-199.5	-196.5	-177.6	1.99	2.05	2.34
	NH <sub>4</sub> <sup>+</sup>	-182.1	-166.3	-152.2	2.23	2.48	2.81
	CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup>	-177.0	-162.3	-146.3	2.24	2.48	2.86

<sup>a</sup>see Figure 4 to identify the oxygen atoms; <sup>d</sup> GAFF parameters; Optimization was not required; <sup>b</sup> No scaling of anion-cation interactions.

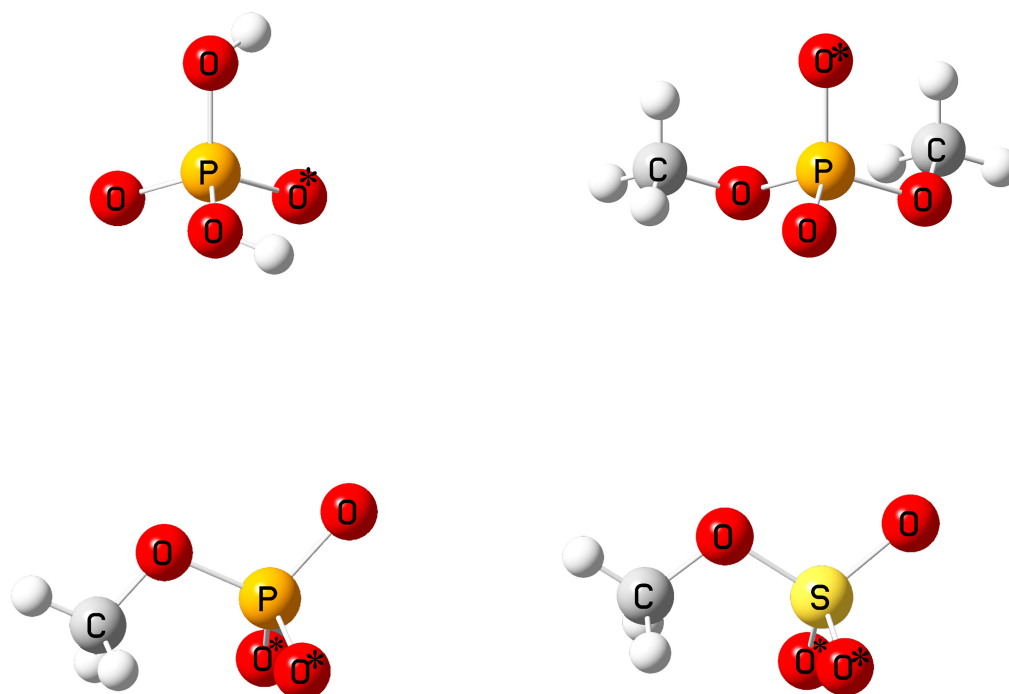


Fig. 4 The oxygen atoms marked by stars, experience more crowded environments than the other oxygens and therefore separated scan where performed for each oxygen type.

Table 3 Statistical error estimation in calculating the solvation free energies,  $\Delta G_{solv}^{Comp,OPT}$ , of anions by running 3 different trials of the same simulation.

Anion	Trial	$\Delta G_{solv}^{Comp,OPT}$ kcal/mol	Standard Deviation kcal/mol
CH <sub>3</sub> SO <sub>4</sub> <sup>-</sup>	1	-88.4 <sup>a</sup>	0.1
	2	-88.5 <sup>a</sup>	
	3	-88.3 <sup>a</sup>	
SO <sub>4</sub> <sup>2-</sup>	1	-272.0 <sup>a</sup>	0.1
	2	-272.1 <sup>a</sup>	
	3	-272.1 <sup>a</sup>	

<sup>a</sup> sum of electrostatic and cavity formation energies, no correction terms included.

**Table 4** Calculated activity derivative  $a_{cc}^{Calc,OPT}$  in different time intervals, and comparison to the overall values.

Salt	Time Interval ns	$a_{cc}^{Calc,OPT}$	Standard Deviation
CH <sub>3</sub> SO <sub>3</sub> <sup>-</sup> -Na <sup>+</sup>	0-50	0.94	
	50-100	0.99	
	100-150	0.93	
	0-150	0.95	0.03
SO <sub>4</sub> <sup>2-</sup> -NH <sub>4</sub> <sup>+</sup>	0-50	0.61	
	50-100	0.59	
	100-150	0.59	
	0-150	0.60	0.01

### 3.2 Potential of mean force calculations

We performed potential of mean force calculations to obtain the free energy as a function of distance between Na<sup>+</sup> and CH<sub>3</sub>COO<sup>-</sup>, for both GAFF and optimized parameters, so we could compare them with recent ab initio molecular dynamics calculations<sup>32</sup>.

We set up all simulations by placing a single anion-cation pair inside a cubic box of size  $\sim 3.4 \times 3.4 \times 3.4$  nm<sup>3</sup> and solvate the molecules using TIP3P water. The systems are then equilibrated for 200 ps in the NPT ensemble at 298.15 K. The temperature is kept constant using the Nose-Hoover scheme<sup>33,34</sup> with a time constant of 0.1 ps. Simulations are done with a pressure of 1 bar, using the Parrinello-Rahman<sup>35</sup> approach with a time constant of 0.5 ps. The remaining simulation details are identical to those used to calculate free energies of solvation (section M2.1.2).

Running umbrella sampling requires generating several configurations along a reaction coordinate. We take the advantage of steered molecular dynamics simulations (SMD) and the implemented pull code in Gromacs to pull ion-pairs apart, along the cation-anion direction. We set the pull rate to 0.001 nm/ps and pull force constant to 5000 kJ mol<sup>-1</sup> nm<sup>-2</sup>. The initial distance between cation and anion is set to be zero using  $pull\_init1 = 0$ , to force the two molecules to approach as close as possible. We run the pulling simulations for 1.5 ns, which corresponds to a maximum anion-cation separation of 15 Å, sufficient for the purpose of our study. We select multiple configurations with anion-cation separations between the minimum and the maximum separation spanned during the SMD simulation; each selected configuration is the starting point for an umbrella sampling window. For each window we run an independent NPT simulation for 10 ns. We use the WHAM<sup>36</sup> analysis program implemented in Gromacs to compute the potential of mean force.

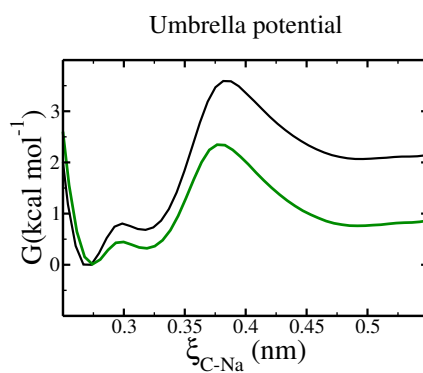


Fig. 5 Potential of mean force between  $\text{Na}^+$  and the carbon atom of the carboxylate group in  $\text{CH}_3\text{COO}^-$ , calculated using Umbrella Sampling. The curves are shifted so that the absolute minima are at 0. Black=GAFF, green=optimized parameters.

### 3.3 Density of chosen salt solutions

Table 5 Density of chosen salts at 0.5 m. The experimental densities are from ref. 37 and are obtained by linear interpolation between the two nearest values. The numbers in parenthesis are the standard deviations.

salt	exp g/cm <sup>3</sup>	GAFF g/cm <sup>3</sup>	this work g/cm <sup>3</sup>
$\text{NaCH}_3\text{COOH}$	1.0184	1.01246 (0.00004)	1.01444 (0.00002)
$(\text{NH}_4)_2\text{SO}_4$	1.035	1.0282 (0.001)	1.03675 (0.00005)
$\text{Na}_2\text{SO}_4$	1.058	1.0451 (0.002)	1.0574 (0.00008)

## References

- 1 M. A. Kastholz and P. H. Hünenberger, *J. Chem. Phys.*, 2006, **124**, 224501.
- 2 M. A. Kastholz and P. H. Hünenberger, *J. Chem. Phys.*, 2006, **124**, 124106.
- 3 P. Hunenberger and M. Reif, *Single-ion solvation - experimental and theoretical approaches to elusive thermodynamic quantities*, RSC, 2011.
- 4 M. M. Reif and P. H. Hünenberger, *J. Chem. Phys.*, 2011, **134**, 144103.
- 5 M. M. Reif and P. H. Hünenberger, *J. Chem. Phys.*, 2011, **134**, 144104.
- 6 H. J. C. Berendsen, D. van der Spoel and R. van Drunen, *Comput. Phys. Commun.*, 1995, **91**, 43–56.
- 7 D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701–18.
- 8 P. Hunenberger, in *Advanced Computer Simulation Approaches for Soft Matter Sciences I*, ed. C. Holm and K. Kremer, Springer, 2005, vol. 173 of *Advances in Polymer Science*, pp. 105–147.
- 9 E. Cancès, B. Mennucci and J. Tomasi, *J. Chem. Phys.*, 1997, **107**, 3032.
- 10 B. Mennucci, E. Cancès and J. Tomasi, *J. Phys. Chem. B*, 1997, **101**, 10506–10517.
- 11 J. Tomasi, B. Mennucci and E. Cancès, *J. Mol. Struct.:THEOCHEM*, 1999, **464**, 211–226.
- 12 S. Miertuš, E. Scrocco and J. Tomasi, *Chem. Phys.*, 1981, **55**, 117–129.
- 13 S. Miertuš and J. Tomasi, *Chem. Phys.*, 1982, **65**, 239–245.
- 14 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
- 15 A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100.
- 16 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.
- 17 T. Steinbrecher, J. Latzer and D. A. Case, *J. Chem. Theory Comput.*, 2012, **8**, 4405–4412.
- 18 J. G. Kirkwood and F. P. Buff, *J. Chem. Phys.*, 1951, **19**, 774–7.
- 19 P. G. Kusalik and G. N. Patey, *J. Chem. Phys.*, 1987, **86**, 5110–5116.
- 20 M. Fyta and R. R. Netz, *J. Chem. Phys.*, 2012, **136**, 124103.
- 21 E. H. Rubensson, *J. Chem. Theory Comput.*, 2011, **7**, 1233–1236.
- 22 M. B. Gee, N. R. Cox, Y. Jiao, N. Benteinitis, S. Weerasinghe and P. E. Smith, *J. Chem. Theory Comput.*, 2011, **7**, 1369–1380.

- 23 B. Klasczyk and V. Knecht, *J. Chem. Phys.*, 2010, **132**, 024109.
- 24 P. Krüger, S. K. Schnell, D. Bedeaux, S. Kjelstrup, T. J. H. Vlugt and J. M. Simon, *J. Phys. Chem. Lett.*, 2013, **4**, 235–238.
- 25 A. Vila Verde, M. Santer and R. Lipowsky, *Phys. Chem. Chem. Phys.*, 2016, **18**, 1918–1930.
- 26 J. L. Lebowitz and J. K. Percus, *Phys. Rev.*, 1961, **122**, 1675–1691.
- 27 A. P. Lyubartsev and S. Marčelja, *Phys. Rev. E - Statistical, Nonlinear, and Soft Matter Physics*, 2002, **65**, 1–6.
- 28 J. D. Ramshaw, *Mol. Phys.*, 1980, **41**, 219–227.
- 29 B. Liu and A. D. Mclean, *J. Chem. Phys.*, 1973, **59**, 4557–4558.
- 30 C. D. Sherrill, *Counterpoise Correction and Basis Set Superposition Error*, <http://vergil.chemistry.gatech.edu/notes/cp.pdf>, 2010.
- 31 S. Boys and F. Bernardi, *Mol. Phys.*, 1970, **19**, 553–566.
- 32 M. D. Daily, M. D. Baer and C. J. Mundy, *J. Phys. Chem. B*, 2016, **120**, 2198–2208.
- 33 S. Nosé, *Molecular Physics*, 1984, **52**, 255–268.
- 34 W. G. Hoover, *Physical Review A*, 1985, **31**, 1695–1697.
- 35 M. Parrinello, *Journal of Applied Physics*, 1981, **52**, 7182.
- 36 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *Journal of Computational Chemistry*, 1992, **13**, 1011–1021.
- 37 *CRC handbook of Chemistry and Physics*, ed. W. M. Haynes and D. R. Lide, CRC press, Boca Raton, FL, USA, 2011.