PCCP

Check for updates

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

Sadra Kashefolgheta (iD) and Ana Vila Verde (iD) *

We present a set of Lennard-Jones parameters for classical, all-atom models of acetate and various alkylated and non-alkylated forms of sulfate, sulfonate and phosphate ions, optimized to reproduce their interactions with water and with the physiologically relevant sodium, ammonium and methylammonium cations. The parameters are internally consistent and are fully compatible with the Generalized Amber Force Field (GAFF), the AMBER force field for proteins, the accompanying TIP3P water model and the sodium model of Joung and Cheatham. The parameters were developed primarily relying on experimental information – hydration free energies and solution activity derivatives at 0.5 m concentration – with *ab initio*, gas phase calculations being used for the cases where experimental information is missing. The *ab initio* parameterization scheme presented here is distinct from other approaches because it explicitly connects gas phase binding energies to intermolecular interactions in solution. We demonstrate that the original GAFF/AMBER parameters often overestimate anion–cation interactions, leading to an excessive number of contact ion pairs in solutions of carboxylate ions, and to aggregation in solutions of divalent ions. GAFF/AMBER parameters lead to excessive numbers of salt bridges in proteins and of contact ion pairs between sodium and acidic protein groups, issues that are resolved by using the optimized parameters presented here.

1 Introduction

Monoatomic and small polyatomic ions are key players in many biological processes, such as DNA folding,[1] blood coagulation and anti-coagulation,[2,3] protein stability[4] and protein crystallization.[5] The molecular mechanisms by which ions act are currently incompletely understood, in part because inferring molecular scale details from the signals detected in experiment cannot be unambiguously done. In contrast, molecular scale details are directly accessible from simulations and atomistic models, making the development of such models of critical importance.

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

Because proteins and other biomolecules are typically large (50–3000 kDa), only simulations using classical, non-polarizable force fields give access to the long length and time scales relevant for many processes taking place in bio-systems. To be useful, force fields must strike the correct balance between the various intermolecular interactions: biomolecule–water, biomolecule–biomolecule, water–water, water–ion, ion–counterion and ion–biomolecule interactions; in this last case, the main contribution would come from the interaction between the ions in solution and the charged sites of the biomolecules. Classical, fixed-charge force fields for biomolecules in water have been under evolution for decades and are now quite successful at predicting binding affinities[6] and the folded structure of small proteins,[7–9] the self-assembly of phospholipid bilayers,[10] the importance of electrostatics in DNA, and DNA–protein interactions.[11,12] These force fields were parameterized against experimental observables that reflect the balance between ion–water, biomolecule–water and water–water interactions (*e.g.*, free energies of hydration).[13,14] More recently, force fields for aqueous solutions of alkali, alkali-earth and halide ions were developed that also show reasonable anion–cation interactions. They are thus able to reproduce experimental osmotic pressures or solution activities at salt concentrations below 1 M,[15–19] and in a few cases up to much higher concentrations.[20] Simulations based

This journal is © the Owner Societies 2017

Phys. Chem. Chem. Phys., 2017, **19**, 20593–20607 | 20593

on these force fields have already offered unprecedented insight into the molecular scale origin of ion-specific effects in biological systems.[21]

In contrast to the advanced state of development of classical molecular models for alkali and halide ions for explicit water simulations, there are currently only a few models of small, polyatomic anions that have been parameterized to reflect the correct balance between their interactions with water, with cationic amino acids or with commonly used counterions.[22–24] Such models are of interest because polyatomic anions such as sulfates, phosphates or acetates are present in the physiological context in inorganic and/or alkylated form, and the molecular scale details regarding their interaction with biomolecules are not yet understood.[25–27] Simulations of biomolecular systems performed so far with these ions have simply used general parameters based on commonly used biomolecular force fields such as CHARMM or AMBER:[13] the partial charges are obtained from low level quantum mechanical calculations and RESP fitting,[28] and the Lennard-Jones parameters for each atom type are the same as those used in the chosen biomolecular force field. Parameters obtained this way are typically used for simulations without further tests – either comparison with experimental data or with *ab initio* calculations – under the assumption that the ion parameters must be reliable because they were obtained with the established protocol used to develop the biomolecular force field. However, these protocols have only been proven reasonable for neutral biomolecules[13,14] and they are known to fail for charged species, *e.g.*, they yield hydration free energies substantially more negative than those experimentally measured.[13]

In this work, we aim to overcome the above limitations of existing classical, fixed-charge models of polyatomic ions for simulations of biomolecular systems in explicit water. We focus specifically on the polyatomic anions methyl-sulfonate ($CH_3SO_3^-$), methyl-sulfate ($CH_3SO_4^-$), acetate (CH_3COO^-), dimethyl-phosphate ($(CH_3)_2PO_4^-$), methyl-hydrogenphosphate ($CH_3HPO_4^-$) and methyl-phosphate ($CH_3PO_4^{2-}$); these anions were selected because they are good analogues of anionic polymers with longer alkyl chains. We also develop parameters for the non-methylated versions of the sulfate and phosphate ions (SO_4^{2-}, HSO_4^-, HPO_4^{2-}, $H_2PO_4^-$). The parameterization approach we present ensures that the ion parameters reflect the correct balance of the ion interactions with water, with the commonly used counterion Na^+ and with analogues (methylammonium, $CH_3NH_3^+$, and ammonium, NH_4^+) of the amino acid lysine, while ensuring that the parameters for all the ions remain internally consistent. The parameters developed here are compatible with widely used force fields for molecular simulations: the AMBER[13] force field for proteins, the accompanying TIP3P[29] water model, and the sodium model of Joung and Cheatham which is often used as counterion.[15]

Intramolecular parameters for the anions already exist;[13,30–32] the main issue is to parameterize their intermolecular interactions. In the context of the AMBER[13] force field for proteins, intermolecular interactions depend on the Lennard-Jones parameters and on the partial charges of the atoms composing the interacting species. We obtain the partial charges of the ions

following the standard procedure used in AMBER[13] – low level *ab initio* calculations followed by RESP fitting – so only the Lennard-Jones parameters of the ions are optimized.‡

Difficulties optimizing Lennard-Jones parameters for polyatomic anions

To parameterize intermolecular potentials, it would be desirable to reproduce high quality thermodynamic data that strongly depends on intermolecular interactions: *e.g.*, free energies of hydration, which reflect ion–water interactions, or solution activities, which reflect anion–cation interactions. This approach was followed by several groups to develop better models of alkali halides.[15,16,19] However, experimental data exists only for a few of the cases of interest here. For example, hydration free energies exist for CH_3COO^-, but not for $CH_3SO_3^-$, $CH_3SO_4^-$, $CH_3PO_4^{2-}$, $CH_3HPO_4^-$ or $(CH_3)_2PO_4^-$.[33] Osmotic pressure or solution activity data for solutions containing the cationic amino acids arginine and lysine, or suitable small-molecule analogues, and polyatomic anions are also not available. For solutions involving polyatomic anions and Na^+, data exists only for $CH_3SO_3^-$ and CH_3COO^-.

Given that there is no complete set of experimental information that we can use to obtain an internally consistent set of parameters for all the anions mentioned above, an alternative would be to develop classical models solely based on *ab initio* calculations. One possibility would be to develop parameters to match observables (*e.g.*, the potential of mean force between two ions) calculated from *ab initio* molecular dynamics (MD) in explicit solvent. The main shortcoming of this approach is that, currently, *ab initio* MD can only access timescales of tens of picoseconds, so the calculated free energy profiles have high statistical uncertainty.[34,35] A second possible shortcoming could be that a low level of theory must necessarily be used in these calculations, which could introduce systematic errors. Recent results, however, indicate that DFT with appropriate dispersion corrections yields reasonable results: structural properties of iodate in water calculated with MD-DFT simulations agree well with results from X-ray absorption fine structure spectroscopy (XAFS).[36] Another possibility would be to develop parameters based solely on single-point *ab initio* calculations, for example water–ion dimers, dry ion pairs or even small hydrated systems (*e.g.*, an ion-pair or a single ion surrounded by a first hydration layer); dry ion pairs or water–ion dimers are here referred to as systems in the gas-phase. Multiple studies, however,[14,37] show that these approaches fail to produce parameters that are adequate for simulations of aqueous systems: the gas-phase intermolecular interaction energies calculated using *ab initio* methods differ significantly from those calculated using classical models optimized to reproduce properties of aqueous systems. How, then, can we obtain an internally consistent set of parameters for the ions in question, given the existing limitations in experimental data and in *ab initio* methods?

---

‡ In principle, the opposite approach – keeping GAFF Lennard-Jones parameters and tuning the charge distribution of the ions to reproduce the desired target properties – would also be possible. Doing so, however, would require establishing a new and well-defined procedure to obtain partial charges for polyatomic ions, which would be more time-consuming than the present approach and does not present obvious advantages.

20594 | *Phys. Chem. Chem. Phys.*, 2017, **19**, 20593–20607

This journal is © the Owner Societies 2017

Proposed approach

We propose an approach to develop non-bonded parameters for ionic species in aqueous systems, that combines experimental data and *ab initio* simulations. This combination is done in a manner that yields an internally consistent set of parameters for the anion–cation and ion–water interactions.

Here we briefly described the main points behind the approach; in the Methods section we present a detailed algorithm that may be followed by interested users. Let us consider species A, B and B′. B and B′ are similar, *i.e.*, they have identical charge and similar composition (*e.g.*, $CH_3SO_4^-$ and CH_3COO^-), yet different enough that they may not necessarily share identical parameters (*e.g.*, the oxygen Lennard-Jones parameters in $CH_3SO_4^-$ and CH_3COO^- are not necessarily identical). AB and AB′ denote dimers of these species interacting non-covalently. Lennard-Jones parameters for the intermolecular interactions in AB in water have been developed based on experimental data (Fig. 1, step 1), but analogous experimental data for the AB′ system does not exist. Using high level *ab initio* calculations, we calculate the intermolecular energies in the AB and AB′ systems in the gas phase, as a function of the intermolecular distance (Fig. 1, step 2). We also perform the analogous potential energy scan for the AB system using the classical, optimized parameters. As illustrated in cartoon form in Fig. 1 (step 2), the interaction energy for

Fig. 1 The parameterization approach in this study.

system AB from *ab initio* calculations differs from that calculated using the classical model. This difference reflects the fact that the classical parameters were optimized for the intermolecular interactions of AB in water, and that the polarization state of molecules or ions in a dimer in the gas phase is very different from their polarization state in water.[14,38] Therefore, the AB parameters offer a poor description of the intermolecular interactions of this dimer in the gas phase. We then assume that the polarization state of similar species should be perturbed to a similar extent when they are immersed in water. If this assumption is true, then the difference in the minimum interaction energy between the classical and *ab initio* methods for AB, $E_{cl,AB}(R_{min}) - E_{QM,AB}(R_{min})$, should be very similar to the same difference for the AB′ system, $E_{cl,AB'}(R_{min}) - E_{QM,AB'}(R_{min})$. In this article we show that this assumption indeed holds. This finding implies that classical parameters for the intermolecular interactions of the AB′ system can be determined by requiring that

$$\Delta E_{AB',QM \to cl}(R_{min}) = \Delta E_{AB,QM \to cl}(R_{min}) \tag{1}$$

as illustrated in Fig. 1 (step 3). In this expression,

$$\Delta E_{AB,QM \to cl}(R_{min}) = E_{cl,AB}(R_{min}) - E_{QM,AB}(R_{min}) \tag{2}$$

is the difference in the energy minimum of the indicated dimer calculated using *ab initio* and that calculated using the optimized classical parameters, as illustrated in Fig. 1 (step 2). This parameterization approaches uses the energy difference, $\Delta E_{AB,QM \to cl}(R_{min})$ at the energy minimum only, because this parameter is expected to dominate system properties. A similar approach has proven reasonable to develop parameters for neutral compounds like alcohols[14,37] but until now has not been applied to charged species.

We use this approach to develop a set of classical parameters for polyatomic anions which is internally consistent and which is fully compatible with the AMBER[13] force field for proteins, the TIP3P[29] water model and the sodium ion model of Joung and Cheatham.[15] The experimental properties we use for parameterization are the solution activity derivatives and the hydration free energies. Solution activities exist only for $NaCH_3SO_3$, $NH_4CH_3SO_3$, $NaCH_3COO$, Na_2SO_4 and $(NH_4)_2SO_4$; reliable hydration free energies exist only for HSO_4^-, CH_3COO^- and SO_4^{2-}. We optimize the parameters for the ions based on this experimental information, and then use *ab initio* calculations and the approach described above to obtain parameters for the remaining cases.

2 Methods

2.1 Molecular dynamics simulations

2.1.1 Force field and parameterization details. The TIP3P water model[29] is used in all simulations. The Lennard-Jones (LJ) parameters of the TIP3P oxygen are $\varepsilon_{OO} = 0.6364$ kJ mol^{-1} and $\sigma_{OO} = 0.315061$ nm, with the partial charge of $q_O = -0.8340e$.[29] The partial charge for each of the two hydrogens is $q_H = +0.4170e$; the LJ parameters assigned to the two hydrogens in TIP3P water equal zero.[29]

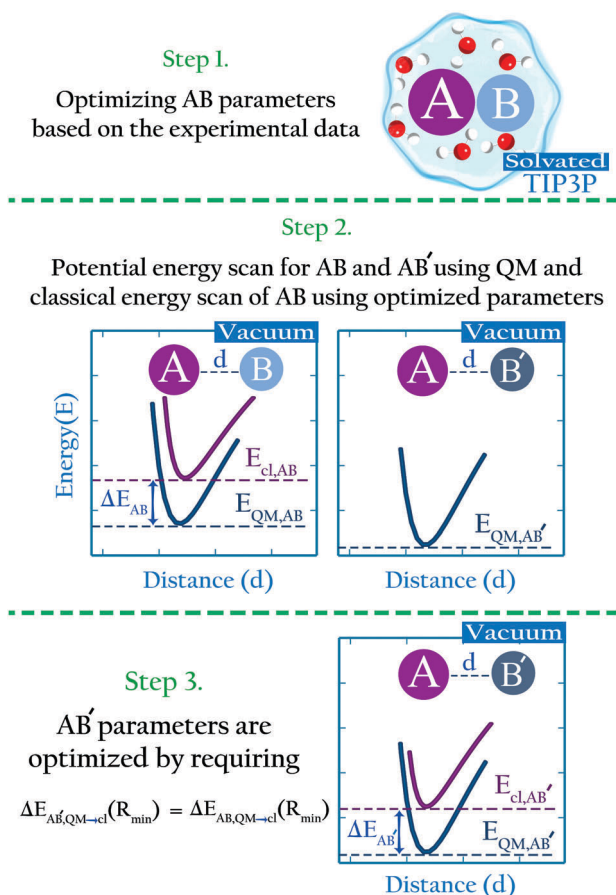This journal is © the Owner Societies 2017

Phys. Chem. Chem. Phys., 2017, **19**, 20593–20607 | 20595

For sodium (Na$^+$) and chloride (Cl$^-$) we use Joung and Cheatham parameters,[15] which are shown to reproduce reasonably well solvation free energies and ion–water binding energies when used in combination with TIP3P water.[29]

We introduce a general notation for the anions, A^{n-}, here investigated: (RO)$_k$XO$_m^{n-}$, where X = {S,P,C} and R = {H,CH$_3$}. The anions have up to three different oxygen types, represented as follows: atom type O2 represents terminal oxygens, atom type OH represents acidic oxygens and atom type OS represents X–O–X oxygens (*e.g.*, C–O–P in (CH$_3$)$_2$PO$_4^-$). We use $\mathcal{O}$ to represent any of the three oxygen types, $\mathcal{O} = \{O2, OH, OS\}$. The hydrogen atoms connected to the nitrogen in NH$_4^+$ and CH$_3$NH$_3^+$ are indicated as H$_N$.

Intramolecular bond, angle and dihedral parameters for the anions, NH$_4^+$ and CH$_3$NH$_3^+$ are taken from GAFF;[13,30–32] for a few angles and dihedrals of the anions, it was necessary to change the strength of the potential, as indicated in the parameter files given as ESI,† to avoid unphysical intramolecular clashes. The partial charges for the atoms composing the anions, NH$_4^+$ and CH$_3$NH$_3^+$ are obtained using the standard RESP fitting protocol used in the AMBER[13] force field, as described in Section 2.2.1.

The ε and σ parameters determine the Lennard-Jones interaction energy of two species i and j as

$$U_{ij} = 4\varepsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right) \qquad (3)$$

Unless otherwise stated, the Lorentz–Berthelot combination rules are used to obtain ε_{ij} and σ_{ij} from the self-interaction parameters, denoted by the subscripts ii and jj, as follows:

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}}, \quad \sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2} \qquad (4)$$

The Lennard-Jones parameters of the GAFF[13,30–32] force field are used in most cases; the only exceptions we propose occur in interactions involving the anion oxygens. For several of the anions, the self-interaction parameters of the oxygens are optimized to better reproduce their interactions with water; these optimized self-interaction parameters, given in the Results section, should be used with eqn (4) to obtain all intermolecular interaction parameters except in the few cases indicated here, for which better estimates are found. We test whether Lorentz–Berthelot combination rules hold for interactions between the anions and Na$^+$, NH$_4^+$ or CH$_3$NH$_3^+$, as detailed below and in the Results section; for most of these cases we find that they do not, so we propose optimized parameters for interactions between the anion-oxygens and Na$^+$ ($\mathcal{O}\cdots$Na$^+$), and between the anion-oxygens and nitrogen and H$_N$ atoms in NH$_4^+$ or CH$_3$NH$_3^+$ ($\mathcal{O}\cdots$H$_N$ and $\mathcal{O}\cdots$N), that should be used instead of eqn (4).

The optimization of the anion-oxygen Lennard-Jones parameters proceeded in multiple stages. These are now described in the form of a road map that may be followed when applying our approach to parameterize intermolecular interactions for any species in water:

(1) Select a starting parameter set for the species to be parameterized. Intramolecular parameters are indispensable because the approach we propose applies only to non-bonded

interactions. GAFF provided good initial parameters for the oxoanions.

(2) Select which interactions should be optimized; select which experimental target properties exist for parameterization. In our case, the difference in the hydration free energies between the anions and a reference ion (Cl$^-$) is selected to optimize ion–water interactions:

$$\Delta\Delta G_{\text{solv}} = \Delta G_{\text{solv}}(A^{n-}) - n \times \Delta G_{\text{solv}}(Cl^-) \qquad (5)$$

We choose to optimize parameters to reproduce the experimental value of $\Delta\Delta G_{\text{solv}}$ instead of ΔG_{solv} because of the lower uncertainty[39] associated with experimental values of $\Delta\Delta G_{\text{solv}}$ relative to ΔG_{solv}, and because we aim to create a force field compatible with the Na$^+$ and Cl$^-$ ions developed by Joung and Cheatham.[15] Details of the free energy calculations are given in Section 2.1.2 and in the ESI† (SI 2.1).

The solution activity derivative at 0.5 m is selected to optimize anion–cation cation interactions. Details of these simulations are given in Sections 2.1.2 and in the ESI† (SI 2.2).

Ion–water interactions should be optimized before anion–cation interactions; the remainder of this road map reflects this order.

(3) Evaluate the quality of the initial parameter set with respect to the primary experimental target property, and select which ion parameters will be optimized. We calculated the hydration free energies of the anions for which that information was available. We found that GAFF parameters failed in some cases, leading to $\Delta\Delta G_{\text{solv}}$ that differed from the target more than 2 kcal mol^{-1}, *i.e.*, more than the uncertainty associated with the experimental target values.[33,39] For the ions for which GAFF parameters proved insufficient, we optimize ion–water interactions.

Because we only have one target experimental property – the hydration free energy – for parameterization of ion–water interactions, we opt to optimize only the self-interaction Lennard-Jones parameters associated with the anion oxygens. The oxygens form the outer part of the anions, and for that reason dominate the Lennard-Jones interactions of the anions with any nearby species. If multiple target properties exist, then parameters for multiple atoms may be optimized.

(4) Optimize self-interaction parameters based on hydration free energies, for the species for which these energies exist. The self-interaction $\varepsilon_{\mathcal{O}\mathcal{O}}$ or $\sigma_{\mathcal{O}\mathcal{O}}$ parameters of the oxygens in CH$_3$COO$^-$ and SO$_4^{2-}$ are optimized to reproduce the experimentally determined difference in solvation free energy between those anions and Cl$^-$, $\Delta\Delta G_{\text{solv}}$.

Given that only one target property was selected for optimization, an infinite number of combinations of $\varepsilon_{\mathcal{O}\mathcal{O}}$ and $\sigma_{\mathcal{O}\mathcal{O}}$ values could yield equally good agreement with the target, possibly at the expense of the performance of the model regarding other observables. Because the original GAFF parameters are physically-based – they are derived to reproduce the density and heats of vaporization of organic liquids[13,30–32] – we retain them for either $\sigma_{\mathcal{O}\mathcal{O}}$ or $\varepsilon_{\mathcal{O}\mathcal{O}}$, and optimize the remaining parameter only. We optimize $\sigma_{\mathcal{O}\mathcal{O}}$ for the cases where the original GAFF[13,30–32] parameters led to deviations greater than 10 kcal mol^{-1} from the target $\Delta\Delta G_{\text{solv}}$, but we optimize $\varepsilon_{\mathcal{O}\mathcal{O}}$ for the cases where the deviation was lower than 10 kcal mol^{-1}.

20596 | *Phys. Chem. Chem. Phys.*, 2017, **19**, 20593–20607

This journal is © the Owner Societies 2017

This threshold energy was chosen because a change of $\Delta\Delta G_{\text{solv}} > 10$ kcal mol^{-1} is approximately equal to half the $\Delta\Delta G_{\text{solv}}$ between Na$^+$ and K$^+$, two ions which differ substantially in radius.§ [33]

To find the optimal parameters, we calculate the hydration free energies for a range of the scaling factor, f, applied to $\varepsilon_{\mathcal{OO}}$ or to $\sigma_{\mathcal{OO}}$ parameters; 4–6 different values of f ranging from 0.7 to 1.3 are used. We fit a linear function to the results. The optimal parameters are found by solving the linear function for the experimental SolvFE difference, $\Delta\Delta G_{\text{solv}}^{\text{Exp}}$.

(5) Optimize the self-interaction parameters of the oxygen atoms of the remaining anions (for which reliable hydration free energies are not available) following the *ab initio* approach described in the introduction.

(a) For each species still to be parameterized (B′), identify a similar species that was already parameterized (B). Similar species must have the same charge and must have similar general composition, primarily evaluated in terms of the characteristics (*e.g.*, identity, hybridization, polarization) of the atoms being parameterized. Here we considered oxoanions of the same charge as similar species.

(b) For the water-B and water-B′ dimers, select an initial relative orientation that is expected to yield similar and strong intermolecular interactions. For example, relative orientations where one of the water hydroxyl groups is pointing to the O2 oxygens in both B and B′ are similar; relative orientations where the water hydroxyl groups point to an O2 oxygen in one of the species, but to an OH or OS oxygen in the other species are not similar. Strong intermolecular interactions refer to orientations that are expected to have particularly favourable interaction energies.

For the chosen initial structures, perform gas phase *ab initio* potential energy scans as described in Section 2.2.2. If water-B and water-B′ differ in hyperconjugation effects (this situation did not happen in our systems), repeat the *ab initio* calculations with different relative orientations (identical for the two dimers) until one is found that yields similar hyperconjugation effects for both dimers. The importance of hyperconjugation effects in our parameterization approach is described in the context of anion–cation gas phase scans in Section 3.2.2.1.

(c) For the water-B dimer, perform a gas phase potential energy scan with the classical model and the optimized parameters, and using the same structures used in the *ab initio* scan. Calculate the difference, $\Delta E_{\text{water-B,QM}\rightarrow\text{cl}}(R_{\text{min}})$, between the position of the energy minimum in the *ab initio* and the classical curves. This calculation is done by first fitting each set of data to a function of the form $A\exp(B \times R) + C(R^{-6}) + (D/R)$,

where A, B, C and D are fitting parameters, corresponding to the Buckingham potential form with an added electrostatic ($1/R$) term, and then identifying the minimum of the fitted functions.

(d) For the water-B′ dimer, perform multiple gas phase potential energy scan with the classical model and multiple values of $\varepsilon_{\mathcal{OO}}$ or $\sigma_{\mathcal{OO}}$ parameters, until a parameter value is found for which $\Delta E_{\text{water-B,QM}\rightarrow\text{cl}}(R_{\text{min}}) = \Delta E_{\text{water-B}',\text{QM}\rightarrow\text{cl}}(R_{\text{min}})$ (eqn (1)) to within 0.5 kcal mol^{-1}. We tested values of $\varepsilon_{\mathcal{OO}}$ or $\sigma_{\mathcal{OO}}$ ranging from 0.6 to 1.3 times the original values, in intervals of 0.01 for $\varepsilon_{\mathcal{OO}}$ and 0.001 for $\sigma_{\mathcal{OO}}$. For similar reasons as those indicated in point 4, we optimize $\varepsilon_{\mathcal{OO}}$ when the original $\Delta E_{\text{water-B}',\text{QM}\rightarrow\text{cl}}(R_{\text{min}})$ differs less than 1 kcal mol^{-1} from the target, and optimize $\sigma_{\mathcal{OO}}$ otherwise.

(6) Evaluate whether the Lorentz–Berthelot combination rules (eqn (4)) together with the current parameter set – already including the optimized self-interaction parameters – allows the reproduction of experimental properties that strongly depend on anion–cation interactions; select which ion parameters will be optimized. We tested whether the current parameter set yields reasonable solution activity derivative at 0.5 m, for the salts for which experimental solution activities are available. The combination rules proved insufficient for all test cases, as described in the Results section, leading to deviations larger than 0.03 – the statistical uncertainty associated with our calculations – from the target values.

As for the case of the self-interaction parameters, we only have one target property against which to optimize anion–cation interactions – the solution activity derivative. For this reason, we opt to tune the Lennard-Jones ε or σ parameters corresponding to the interaction between the anion-oxygens and Na$^+$, and between the anion-oxygens and N and H$_{\text{N}}$ in NH$_4^+$. Both parameters for N and for H$_{\text{N}}$ are optimized because in this case the buried atom, N, is expected to contribute significantly to the intermolecular interactions; hydrogen atoms always have weak or zero Lennard-Jones potentials.

(7) Optimize anion–cation interaction parameters based on experimental properties, for the salts for which they exist. The Lennard-Jones ε or σ parameters corresponding to the interaction between the anion-oxygens and Na$^+$ ($\mathcal{O}\cdots$Na$^+$), and between the anion-oxygens and N and H$_{\text{N}}$ in NH$_4^+$ ($\mathcal{O}\cdots$N and $\mathcal{O}\cdots$H$_{\text{N}}$) are parameterized to reproduce the activity derivative of 0.5 m solutions of salts for which that information was available.

For similar reasons as those indicated in point 4, we optimize either ε or σ, and retain the parameters obtained *via* the combination rules (eqn (4)) for the remaining parameter. We optimize σ when the initial solution activity derivative deviates from the target more than 0.1. and optimize ε otherwise. The determination of the optimal parameters followed an analogous procedure as that described in point 4.

(8) Optimize the intermolecular interaction parameters for the remaining salts (for which experimental target properties are not available) following the *ab initio* approach described in the introduction. We optimized the Lennard-Jones parameters determining the $\mathcal{O}\cdots$Na$^+$, $\mathcal{O}\cdots$N and $\mathcal{O}\cdots$H$_{\text{N}}$ interactions.

(a) For each anion–cation pair to be parameterized (AB′), identify a similar pair for which intermolecular interactions are

---

§ We note that the Lennard-Jones potential depends linearly on ε but depends on σ to the power of 12, so similar fractional changes in σ and ε result in much larger changes in the potential for the case of σ; optimizing $\sigma_{\mathcal{OO}}$ for the cases where the initial parameters deviate more than our threshold from the target $\Delta\Delta G_{\text{solv}}$ and optimizing $\varepsilon_{\mathcal{OO}}$ otherwise thus results in optimized parameters that remain in the same magnitude as the original GAFF parameters. If multiple reliable experimental observables are available, then both parameters can be optimized thus avoiding this somewhat arbitrary decision. Because of this arbitrariness, our optimized parameter set is not unique; however, because it is grounded on experimental data, it is nevertheless a marked improvement over the original GAFF parameters.

already parameterized (AB). The pairs must share a common ion (A), and B and B′ must be similar species; see point 5a for the conditions defining similarity between B and B′.

(b) For the AB and AB′ dimers, select an initial relative orientation that is expected to yield similar and strong intermediate interactions. Here we consider only orientations where the O2 oxygens face the cations. Perform gas phase, *ab initio* potential energy scans as described in point 5b, evaluate hyperconjugation effects, and find new orientations if necessary. A detailed discussion of the impact of hyperconjugation effects on anion–cation interactions is given in Section 3.2.2.1.

(c) For the AB dimer, perform a gas phase potential energy scan with the classical model and the optimized parameters, and with the same configurations used in the *ab initio* scan. Calculate $\Delta E_{\text{AB,QM}\rightarrow\text{cl}}(R_{\min})$ (see eqn (1)).

(d) For the AB′ dimer, perform multiple gas phase potential energy scans with the classical model and multiple values of ε or σ for the $\mathcal{O}\cdots\text{Na}^+$, $\mathcal{O}\cdots\text{N}$ and $\mathcal{O}\cdots\text{H}_\text{N}$ interactions until a parameter value is found for which $\Delta E_{\text{AB,QM}\rightarrow\text{cl}}(R_{\min}) = \Delta E_{\text{AB',QM}\rightarrow\text{cl}}(R_{\min})$ to within 0.5 kcal mol^{-1}. Further details analogous to those in point 5d.

In the Results section we discuss further aspects concerning the application of this approach to our systems. The final parameters, expressed as a scaling factor relative to the original GAFF[13,30–32] parameters to facilitate comparisons, are shown in the Tables presented in the Results section, as well as in the ESI† in Gromacs-readable files.

2.1.2 General simulation details. We perform two types of molecular dynamics (MD) simulations using the GROMACS simulation package[40,41] and the TIP3P water model:[29] free energy perturbation simulations to calculate hydration free energies, and molecular dynamics simulations to calculate solution activity derivatives. All simulation boxes are cubic with periodic boundary conditions applied in the *XYZ* directions. We set a non-bonded potential cutoff at 1.2 nm distance. Beyond this cutoff distance, the electrostatic interactions are calculated with the particle mesh Ewald (PME) method.[42] Lennard-Jones interactions are smoothly shifted to zero between 1.0 and 1.2 nm. Long range dispersion corrections were applied to the energy and pressure. A leap-frog stochastic dynamics (SD) integrator[43] is used to integrate the equations of motion in all simulations with the temperature fixed at 298 K. For *NPT* simulations we maintained the pressure at 1 bar using the Berendsen barostat[44] with a coupling time of 1.0 ps. All bonds with H-atoms were restrained using the LINCS algorithm.[45] We used a 2 fs time step.

The solution activity derivatives are calculated according to the Kirkwood–Buff theory,[46] with expressions for the integrals formulated specifically for closed systems.[47] A summary of the most relevant aspects of this theory and how it was applied here is given in the ESI.† The simulation box for the calculation of the activity derivative was prepared by putting 72 anions and 72 cations (144 when using divalent anions) in a cubic box with dimensions 4.0 × 4.0 × 4.0 nm^3 and solvating the system by adding ≈8000 TIP3P[29] water molecules using editconf and solvate tools from Gromacs 5.0.[40,41] The salt concentration simulated is thus 0.5 m. The system is equilibrated for 10 ns

in the *NPT* ensemble. We select three distinct configurations from the equilibration run, with a box volume similar to the average box volume obtained during the equilibration. Each of these configurations is then used as the initial state for a 50 ns simulation in the *NVT* ensemble. The total simulation time for each production run is thus 150 ns; this large simulation time is necessary to provide enough sampling to calculate activity derivatives. The statistical uncertainty associated with the calculated activity derivatives is 0.03, as shown in Table S4 of the ESI.†

The starting configuration for the solvation free energy calculations is prepared by placing a single ion in a simulation box of dimensions ∼4.0 × 4.0 × 4.0 nm^3 and solvating it in TIP3P water.[29] The free energy is obtained along the path coordinates of λ_{LJ} and λ_{C}, which are the scaling factor to the Lennard-Jones and the electrostatic potentials, respectively. These scaling factors range from 0 to 1 (1 for the decoupled state and 0 for the fully coupled state). To ensure adequate ensemble overlap, we use 21 steps for each of λ_{LJ} and λ_{C}; the steps are evenly spaced for λ_{C}; for the λ_{LJ} spacing we use 21 unevenly spaced points, where $\lambda_{\text{LJ}} \in$ {0.00 0.06 0.12 0.18 0.24 0.30 0.36 0.42 0.46 0.50 0.52 0.54 0.56 0.58 0.60 0.64 0.68 0.72 0.76 0.80 1.00}. Soft core potentials are used during the LJ coupling. For each λ we perform an equilibration of 100 ps followed by a production run of 1 ns in the *NPT* ensemble. The total solvation free energies are calculated from the partial results along the path coordinates using Bennett's acceptance ratio (BAR),[48] as implemented in Gromacs 5.0.[40,41] The estimated uncertainty associated with the calculated solvation free energies is 0.1 kcal mol^{-1}, as shown in Table S3 of the ESI.† This value is estimated by performing different simulations for the same systems, using 0.5 ns simulations for each 21 lambda values. For divalent anions 41 lambda spacing was used as well for comparision, and the result did not change from the one obtained from 21 lambdas. The solvation free energies calculated directly from simulation must be corrected before they can be compared with experiment.[39] In the ESI† (SI 2.1) we describe in detail which correction terms are included in the values here reported.

2.2 *Ab initio* calculations

2.2.1 Charge fitting. The classical charges for the new species are obtained consistently with the AMBER[13] force field for proteins. The molecules are first optimized using MP2/6-31G* level of theory to find the global minimum geometries. Thereafter we use HF/6-31G* level to reoptimize the structures and to calculate the Merz–Kollman[49] atomic charges fitted to the electrostatic potential. All calculations are performed using the Gaussian 03 software package.[50] The Gaussian output is later plugged into the Antechamber module of the AMBER[13] software suit to generate the classical atomic point charges using RESP[28] fitting.

2.2.2 Potential energy scan. To optimize the Lennard-Jones parameters for the interaction of the anions with water and with cations for the cases where experimental data is unavailable, we require rigid-body potential energy scans of anion–water or anion–cation systems in the gas phase, as described in the Introduction and illustrated schematically in Fig. 1. The binding
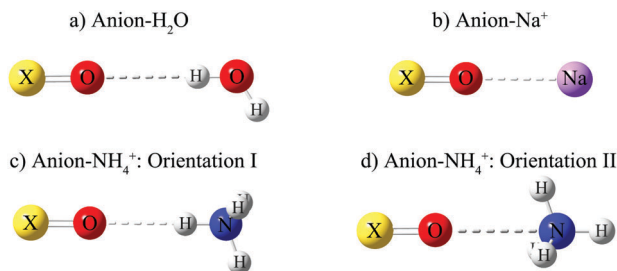
20598 | *Phys. Chem. Chem. Phys.*, 2017, **19**, 20593–20607

This journal is © the Owner Societies 2017

Fig. 2 Example configurations for the 1-D potential energy scans of anions: (a) with H_2O, (b) with Na^+, (c) with NH_4^+ in orientation I, (d) with NH_4^+ in orientation II. Orientation I results in a stronger hydrogen bond than orientation II.[53]

Fig. 3 Difference in solvation free energy, $\Delta\Delta G_{solv}^{Comp} = \Delta G_{solv}^{Comp}(SO_4^{2-}) - 2 \times \Delta G_{solv}^{Comp}(Cl^-)$, as a function of the scaling factor $f_{\sigma_{O2-O2}}$ for the SO_4^{2-} anion. The red line shows the target experimental value, $\Delta\Delta G_{solv}^{Exp}$.

potential energy curves as a function of the intermolecular distance, given in terms of the distance between the two closest nuclei of the two species, are obtained for each dimer using the MP2[51] level of theory with augmented-cc-pvtz[52] basis set in the Gaussian 03 package.[50] The structure of each isolated species is first optimized with the same level of theory. After optimization, we perform the gas phase 1-D potential energy while holding the structure of the molecules fixed during the scan and using counterpoise correction; details regarding this correction are given in the ESI† (SI 2.3). Typical configurations for the anion–cation and anion–water potential energy scans performed are shown in Fig. 2. We do not relax the structures during the potential scans because doing so would result in structures that change as a function of the intermolecular distance, and that also differ between the *ab initio* and the classical calculations for the same intermolecular distance. Given that our aim is to compare the minimum intermolecular energy between *ab initio* and classical calculations, this comparison should be done for identical structures.

3 Results and discussion

The optimized parameters are presented in the ESI,† in.top/.itp format for gromacs users; the original values are shown as comments, to facilitate comparisons. Amber-format parameters are given in the same files in the form of comments.

3.1 Determining self-interaction parameters for the oxygens of the anions

3.1.1 Parameters developed based on solvation free energies.
The computational solvation free energies (SolvFEs) are calculated as described in Section 2.1.2. We attempt to reproduce the experimental target data ($\Delta\Delta G_{solv}^{Exp}$; see eqn (5)) by scanning a range of ε_{O2-O2} or σ_{O2-O2}. For HSO_4^-, scanning of the ε_{OH-OH} or σ_{OH-OH} parameters would in principle also be necessary but in practice was not done because the original GAFF[13,30–32] parameters proved adequate. The procedure followed at this step (point 4 in the road map in the Methods section) is illustrated in Fig. 3 for SO_4^{2-}; the strongly linear dependence of $\Delta\Delta G_{solv}^{Comp}$ on σ_{O2-O2} is apparent.

The optimized parameters for HSO_4^-, SO_4^{2-} and CH_3COO^-, as well as their experimental and computational SolvFEs, are
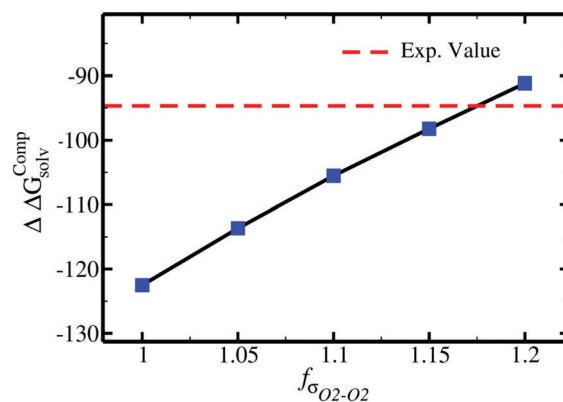
given in Table 1. The GAFF[13,30–32] parameters lead to a -0.6 kcal mol^{-1} deviation from the target value for HSO_4^-. Because the uncertainty associated with experimentally determined solvation free energies of single ions[39] is at least of the order of this deviation, we do not further parameterize HSO_4^- against the SolvFE and we use the GAFF[13,30–32] Lennard-Jones parameters for this anion.

For acetate (CH_3COO^-), the original ε_{O2-O2} needs to be scaled by a factor of 0.77 to retrieve the experimental SolvFE for this ion. The optimized parameter thus reproduces the difference in hydration free energies between Cl^- and CH_3COO^- better than the GAFF[13,30–32] parameters. We note, however, that the deviation from experiment using the GAFF[13,30–32] parameters is only $+2.3$ kcal mol^{-1}, so using GAFF parameters to investigate the interactions of this anion with water is acceptable. In contrast, GAFF[13,30–32] parameters yield a far too low hydration free energy for SO_4^{2-}: the deviation from experiment is -27.9 kcal mol^{-1}. A scaling factor of 1.17 to the original σ_{O2-O2} parameter is required to reproduce the target $\Delta\Delta G_{solv}^{Exp}$ of the sulfate. GAFF[13,30–32] parameters dramatically fail to reproduce the SolvFE of multivalent ions, emphasizing the importance of parameter optimization with respect to SolvFE for those anions.

3.1.2 Parameters developed using *ab initio* calculations.
To optimize the oxygen parameters for $CH_3SO_4^-$, $CH_3SO_3^-$, HPO_4^{2-} and $CH_3PO_4^{2-}$, for which reliable hydration free energies measured experimentally could not be found in the literature, we use the *ab initio* approach described in the Introduction.

For each of the $CH_3SO_4^-$, $CH_3SO_3^-$, HPO_4^{2-} and $CH_3PO_4^{2-}$ anions, we perform an *ab initio* gas phase potential energy scan of the anion–water dimer, as described in Section 2.2.2. For parameter optimization, we require the same type of potential energy scan to be performed for a reference species for which classical parameters already exist: we use SO_4^{2-} as the reference compound for $CH_3PO_4^{2-}$ and HPO_4^{2-}, and HSO_4^- as the reference compound for $CH_3SO_4^-$, $CH_3SO_3^-$, $H_2PO_4^-$ and $(CH_3)_2PO_4^-$ because of their similarities in total charge and structure. Example scans, for $CH_3SO_3^-$–water and CH_3COO^-–water, are shown in Fig. S3 of the ESI;† the energy and interparticle separation at the energy minimum are shown in

Table 1 Self-interaction parameters for O2 derived against experimental solvation free energies. The parameters are expressed in terms of a scaling factor relative to the GAFF values (shown in the ESI)

Anion	Parameters	$n \times \Delta G_{solv}^{Comp}(Cl^-)$ (kcal mol^{-1})	$\Delta G_{solv}^{Comp,GAFF}(A^{n-})$ (kcal mol^{-1})	$\Delta G_{solv}^{Comp,OPT}(A^{n-})$ (kcal mol^{-1})	$\Delta\Delta G_{solv}^{Comp,GAFF}$ (kcal mol^{-1})	$\Delta\Delta G_{solv}^{Comp,OPT}$ (kcal mol^{-1})	$\Delta\Delta G_{solv}^{Exp}$ (kcal mol^{-1})	Deviations^c (kcal mol^{-1})	
								GAFF	OPT
HSO$_4^-$	—^{a}	$1 \times (-87.3)$	-85.1	—^{a}	2.2	—^{a}	2.8^b	-0.6	—^{a}
CH$_3$COO$^-$	$0.77 \times \varepsilon_{O2-O2}$	$1 \times (-87.3)$	-92.0	-93.0	-4.7	-5.7	-6.2^b	$+1.5$	$+0.5$
SO$_4^{2-}$	$1.17 \times \sigma_{O2-O2}$	$2 \times (-87.3)$	-297.1	-270.0	-122.5	-95.4	-94.6^b	-27.9	-0.8

^{a} No scaling. ^{b} Ref. 33. ^{c} $\Delta\Delta G_{solv}^{Comp} - \Delta\Delta G_{solv}^{Exp}$; $\Delta\Delta G_{solv}$ is defined in eqn (5).

Table 2 Differences between the classical and quantum potential energy scans for all anion–water dimers, and values of the optimized anion–water parameters

Anion	Opt.^{b}	Parameters^c	$\Delta E_{QM\to cl}^{GAFF}(R_m)^d$ (kcal mol^{-1})	$\Delta E_{QM\to cl}^{OPT}(R_m)^d$ (kcal mol^{-1})	$\Delta R_{m,QM\to cl}^{GAFF}$ ^{e} (Å)	$\Delta R_{m,QM\to cl}^{OPT}$ ^{e} (Å)
HSO$_4^-$	Exp	—^{a}	-1.3	—^{a}	-0.14	—^{a}
CH$_3$SO$_4^-$	Comp	—^{a}	-1.4	—^{a}	-0.15	—^{a}
CH$_3$SO$_3^-$	Comp	—^{a}	-1.7	—^{a}	-0.15	—^{a}
H$_2$PO$_4^-$	Comp	—^{a}	-1.4	—^{a}	-0.12	—^{a}
CH$_3$HPO$_4^-$	f	—^{a}	NC	—^{a}	NC	—^{a}
(CH$_3$)$_2$PO$_4^-$	Comp	—^{a}	-1.1	—^{a}	-0.11	—^{a}
CH$_3$COO$^-$	Exp	$0.77 \times \varepsilon_{OO}$	-0.8	-1.2	-0.10	-0.14
SO$_4^{2-}$	Exp	$1.17 \times \sigma_{OO}$	1.0	4.8	-0.06	0.22
HPO$_4^{2-}$	Comp	$1.135 \times \sigma_{OO}$	1.1	4.8	-0.05	0.16
CH$_3$PO$_4^{2-}$	Comp	$1.135 \times \sigma_{OO}$	1.3	4.8	-0.05	0.16

^{a} No scaling. ^{b} Parameters optimized based on hydration free energies (Exp), repeated here to facilitate comparisons, or based on the *ab initio* approach (Comp). ^{c} Expressed as a scaling factor relative to the original GAFF parameters (given in the ESI). ^{d} $\Delta E_{QM\to cl}(R_m)$ is defined in eqn (2). ^{e} $\Delta R_{m,QM\to cl}$: the difference between the position of the dimer energy minimum obtained using *ab initio* calculations and using classical parameters. ^{f} See text for explanation; NC: not calculated.

Table S2 of the ESI† for all cases. We then proceed to optimize the self-interaction parameters as described in point 5 of the road map given in the Methods section.

The optimized self-interaction parameters for the oxygens in all the anions are shown in Table 2. The same table also shows $\Delta R_{QM\to cl}(R_{min})$, the difference in the position of the potential energy minimum calculated using *ab initio* and using classical parameters. For the divalent anions, the optimized parameters result in larger anion–water separation at the energy minimum of the dimer than the original GAFF[13,30–32] parameters. This increase has been observed in other parameterizations of multivalent oxoanions based on experimental data,[54] and illustrates the inadequacy of applying GAFF[13,30–32] parameters, derived for neutral molecules, to heavily charged species.

The original GAFF parameters for CH$_3$SO$_4^-$ yield $\Delta E_{QM\to cl}(R_{min}) = -1.4$ kcal mol^{-1}; for CH$_3$SO$_3^-$, H$_2$PO$_4^-$ and (CH$_3$)$_2$PO$_4^-$ they yield $\Delta E_{QM\to cl}(R_{min}) = (-1.7, -1.4, -1.1)$ kcal mol^{-1} (see Table 2), respectively. Tuning the oxygen parameters of these anions to reproduce the target $\Delta E_{QM\to cl}(R_{min}) = -1.3$ kcal mol^{-1} obtained for HSO$_4^-$ leads to changes in hydration free energy lower than 2 kcal mol^{-1}. Because this difference is slight, we opt to retain the original GAFF[13,30–32] parameters for CH$_3$SO$_4^-$, H$_2$PO$_4^-$, (CH$_3$)$_2$PO$_4^-$ and CH$_3$SO$_3^-$. Given that our results show that the same Lennard-Jones parameters can be used for H$_2$PO$_4^-$ and for (CH$_3$)$_2$PO$_4^-$, we propose that the same parameters are also used for CH$_3$HPO$_4^-$.

The original GAFF[13,30–32] parameters lead to a $\Delta E_{QM\to cl}(R_{min}) = +1.1$ kcal mol^{-1} for HPO$_4^{2-}$, and to $\Delta E_{QM\to cl}(R_{min}) = +1.3$ kcal mol^{-1}

for CH$_3$PO$_4^{2-}$. Because these values differ markedly from the target $\Delta E_{QM\to cl}(R_{min}) = +4.8$ kcal mol^{-1} obtained for SO$_4^{2-}$, the oxygens of the two divalent ions are optimized to reproduce the target $\Delta E_{QM\to cl}(R_{min})$. The value of $\Delta E_{QM\to cl}(R_{min})$ depends on the parameters of the O2 oxygens and is essentially independent of the parameters of the other oxygens; however, for consistency, the σ parameter of the OH oxygens in HPO$_4^{2-}$ is scaled by the same prefactor used to scale σ_{O2}; following the same logic, the σ parameter of the OS oxygens in CH$_3$PO$_4^{2-}$ is also scaled by the same scaling factor applied to σ_{O2} of that species. Applying these scaling factors to OH or OS is recommended, despite the fact that they should lead to relatively small changes in the overall interaction between these species and water; *e.g.*, scaling or not the OH oxygen in HPO$_4^{2-}$ leads to changes of ~ 2 kcal mol^{-1} in SolvFE.

The calculated uncorrected ($\Delta G_{cav} + \Delta G_{chg}$) SolvFE of HPO$_4^{2-}$ are -311.8 kcal mol^{-1} and -285.8 kcal mol^{-1}, for GAFF[13,30–32] and our parameters, respectively. This large change in solvation free energy illustrates the critical importance of optimizing ion–water interactions for divalent anions.¶

3.2 Developing specific anion–cation parameters

3.2.1 Parameters developed based on solution activity derivatives. After finding optimal self-interaction parameters

¶ We do not compare these results to the only available experimental value[55] for SolvFE of HPO$_4^{2-}$ because that value (not shown) is highly unreliable. That value was not directly measured, but instead was estimated based on lattice energies.

20600 | *Phys. Chem. Chem. Phys.*, 2017, **19**, 20593–20607

This journal is © the Owner Societies 2017

for the anions based on ion–water interactions, we evaluate whether those parameters can be used with the combination rules shown in eqn (4) to obtain reasonable interactions between the anions and two cations of interest: the commonly used counterion Na$^+$, and NH$_4^+$, which is often considered to be analogue to the terminal side chain group of the amino acid lysine. We assess the quality of those parameters by calculating, as described in Section 2, the activity derivatives of 0.5 m solutions of salts for which experimental solution activities exist. These results are given in Table 3, under column $a_{cc}^{Calc,b}$ and are referred to in the text as arising from using unoptimized anion–cation parameters. For comparison, we also calculate activity derivatives using exclusively the original set of GAFF[13,30–32] parameters for the anions (denoted as $a_{cc}^{Calc,c}$ in Table 3).

The unoptimized anion–cation parameters fail, often dramatically, to adequately describe anion–cation interactions in all cases; the original set of GAFF[13,30–32] parameters also fails. These failures point to the need of specifically optimizing the anion–cation interactions. Modifications to the anion–cation parameters (see Table 3) are indispensable to attain agreement with experiment for solutions of Na$_2$SO$_4$ or (NH$_4$)$_2$SO$_4$: for both salts, using either unoptimized anion–cation parameters or using the original GAFF[13,30–32] parameters led to dramatic aggregation, as illustrated in Fig. 4a for Na$_2$SO$_4$. Such large aggregation is clearly unphysical, because our simulations are done at 0.5 m, far below the solubility limit of Na$_2$SO$_4$ (2.0 m at 25 °C),[57] and of (NH$_4$)$_2$SO$_4$ (5.8 m at 25 °C).[58,59] As shown in Fig. 6, the optimized anion–cation parameters lead to very few contact ion-pairs formed per ion, with both ions forming primarily solvent shared ion-pairs, i.e., configurations where the ions share their first hydration layer.

For NaCH$_3$COO, using unoptimized anion–cation parameters leads to a solution activity derivative, $a_{cc}^{Calc} = 0.35$, which is far lower than the target value of $a_{cc}^{Exp} = 1.00$; the original GAFF[13,30–32] parameters perform even worse, yielding $a_{cc}^{Calc} = 0.22$. To achieve
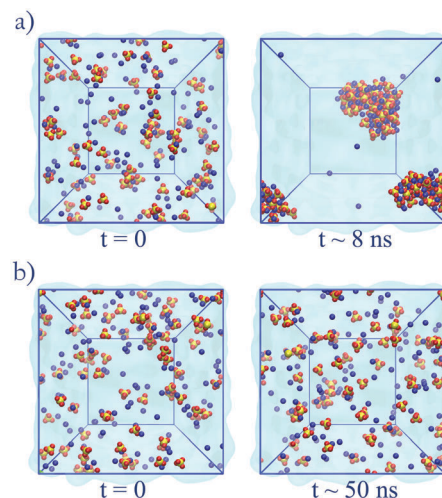

Fig. 4 Simulations of Na$_2$SO$_4$. (a) Original GAFF[13,30–32] parameters: aggregation becomes visible at 500 ps and large aggregates are seen at 8 ns. (b) Optimized anion–cation parameters: aggregation is not observed within the time scope of the simulation.

good agreement with experiment, it is necessary to modify σ_{ONa}. The optimized parameters lead to only 1/5 of the contact ion pairs obtained using the original GAFF[13,30–32] parameters, as shown in Fig. 5a. Our results thus indicate that the interactions between Na$^+$ and acidic amino acids are markedly overestimated when using the sodium model of Joung and Cheatham[15] in the AMBER[13] force field with TIP3P water.

Our optimized potentials for Na$^+\cdots$CH$_3$COO$^-$ agree much better with recent ab initio molecular dynamics results based on DFT methods with dispersion corrections[35] than the original GAFF potentials for TIP3P water: our potential of mean force calculations (ESI,† Fig. S5) for Na$^+\cdots$CH$_3$COO$^-$ show that the difference in free energy between the minimum near 0.33 nm and that near 0.5 nm using our optimized parameters (≈ 0.7 kcal mol^{-1}) is comparable to the same free energy difference (≈ 1 kcal mol^{-1}) obtained using ab initio MD.[35] In contrast, for the original GAFF parameters in TIP3P water this difference is 1.5 kcal mol^{-1}. We note, however, that our classical

Table 3 Solution activity derivatives from experiment (a_{cc}^{Exp}) and from simulation (a_{cc}^{Calc}). The optimal anion–cation parameters are expressed in terms of the values calculated from eqn (4) using the original GAFF values, to facilitate comparisons

Anion	Cation	Parameters	$a_{cc}^{Calc\ b}$	$a_{cc}^{Calc\ c}$	$a_{cc}^{Calc\ d}$	$a_{cc}^{Exp\ e}$
CH$_3$SO$_3^-$	Na$^+$	$1.1 \times \varepsilon_{ONa}$	0.95	0.89	0.89	0.93
	NH$_4^+$	$1.1 \times \begin{cases} \varepsilon_{ON} \\ \varepsilon_{OH_N} \end{cases}$	0.89	0.85	0.85	0.90
CH$_3$COO$^-$	Na$^+$	$1.022 \times \sigma_{ONa}$	0.99	0.35	0.22	1.00
SO$_4^{2-}$	Na$^+$	$1.13 \times \sigma_{ONa}$	0.64	a	a	0.62 ± 0.02
	NH$_4^+$	$1.13 \times \begin{cases} \sigma_{ON} \\ \sigma_{OH_N} \end{cases}$	0.60	a	a	0.62 ± 0.02

a Could not be calculated because of aggregation. b Optimized σ_{OO}, ε_{OO} and anion–cation parameters. c Optimized σ_{OO} and ε_{OO}, with anion–cation parameters derived from combination rules (for some salts, this set of parameters coincides with the original GAFF parameters). d Original GAFF parameters. e Ref. 56.
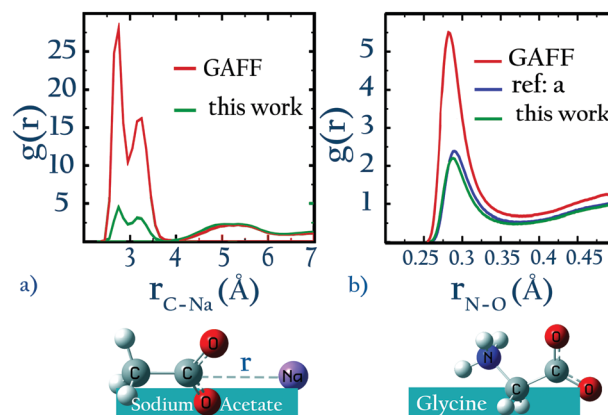

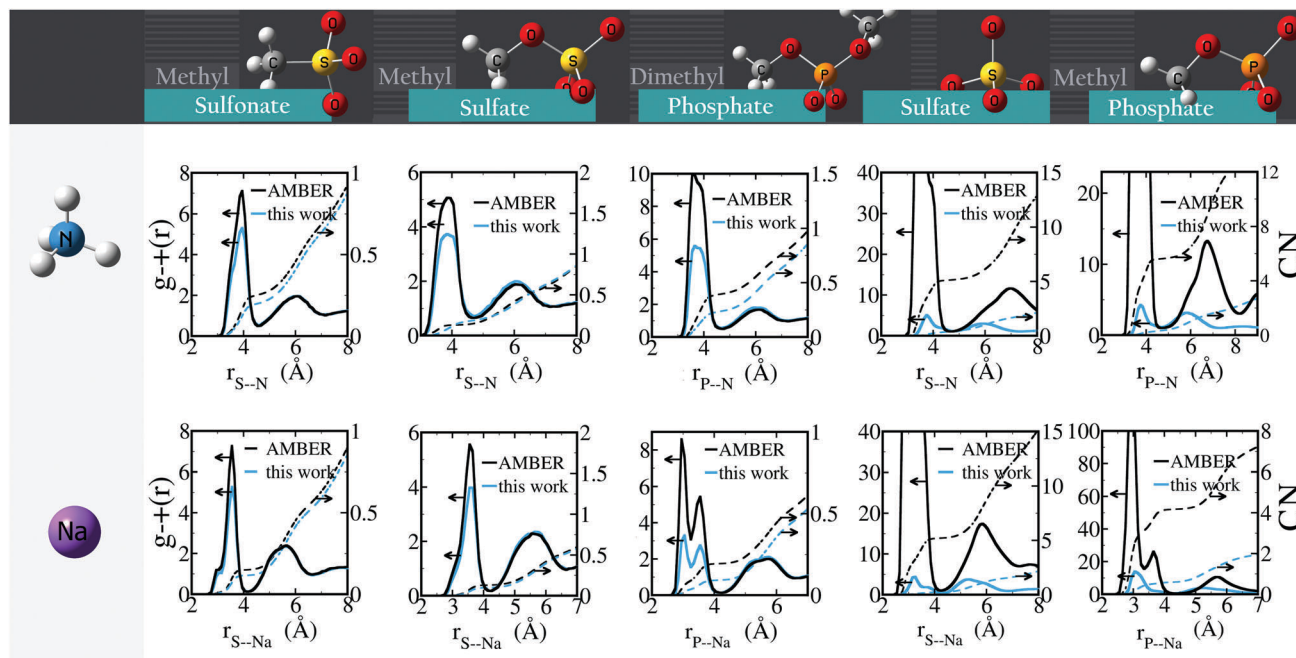Fig. 5 (a) 0.5 m solution of sodium acetate, (b) 3 m solution of glycine (a = ref. 22).

This journal is © the Owner Societies 2017

Phys. Chem. Chem. Phys., 2017, 19, 20593–20607 | 20601

Fig. 6 Radial distribution functions of anion–cation pairs, $g{-}{+}(r)$ (solid lines, left y-axes) and corresponding number integral (CN; dashed lines, right y-axes) using AMBER/GAFF parameters and using the optimized parameters.

simulation results still show marked differences from those obtained using *ab initio* MD: our parameters yield a free energy minimum at 0.27 nm which is not seen in *ab initio* MD. In the absence of reliable experimental information reporting on the (non)existence of the ion pair at 0.27 nm separation, it remains unclear which of these descriptions is better.

For NaCH$_3$SO$_3$ and NH$_4$CH$_3$SO$_3$, the activity derivative calculated without optimizing anion–cation interactions differs less than 5% from experiment. Despite this seemingly small difference, we provide optimized anion–cation parameters for these salts (see Table 3) because they lead to quantitative differences in the number of ion-pairs: Fig. 6 show that the optimized anion–cation parameters lead to 25% fewer contact ion-pairs – defined as ions in direct contact with each other – than the non-optimized ones.

3.2.2 Parameters developed using *ab initio* calculations. For several of the salts of interest here, the solution activities are not available in the literature or have high associated uncertainty. High uncertainty is often found for salts of amphiprotic species like phosphate, because solutions of those salts simultaneously contain H$_2$PO$_4^-$, HPO$_4^{2-}$ or PO$_4^{3-}$, making it difficult to deconvolute the contribution of each species to the solution activity.[56] For these salts, we optimize anion–cation interactions again following the approach described in the introduction, described in detail in point 8 of the road map given in the Methods section.

3.2.2.1 Determining optimal anion–cation orientations to be used in gas-phase potential energy scans. In *ab initio* calculations, contact ion-pairs in the gas phase benefit from extra delocalization stability due to the overlap of cation and anion orbitals; this extra stabilization is sometimes referred to as intermolecular

hyperconjugation energy.[60–63] However, quantum effects such as electron delocalization and charge transfer between two ions are not expected to be dominant in aqueous solution; in this case the charge transfer preferentially occurs between the ions and the water molecules in the first hydration shell.[64–66]

The fact that quantum effects stabilize ion–ion interactions in the gas phase but not in solution represents a problem when using our approach to develop classical parameters for ions in aqueous solution. Given that the magnitude of the hyperconjugation energy varies amongst contact ion-pairs – depending on the electronic structure and the relative orientation of the two ions – our *ab initio* approach might introduce artifacts if the two ion-pairs being compared benefit from different hyper-conjugation stabilities. To resolve this issue, we must ensure that the anions for which we want to make the *ab initio* comparison have similar interactions with the same cation. We assess the similarity of interactions by performing natural bond orbital (NBO)[60] analysis. This analysis proved to well-capture the electron density transfers,[61–63] on the geometries shown in Fig. 2a–c and at distances corresponding to the minimum energy on the 1-D potential curves for each ion-pair. To perform this analysis we use the geometries from MP2 calculations and perform B3LYP[67–69]/aug-cc-pVTZ calculations to obtain the electron densities. Using the B3LYP method is necessary to obtain the well-defined one-electron densities required by NBO calculations.[61]

The results of the NBO analysis are shown in Table 4. In ion-pairs with Na$^+$, the electron density transfers from the lone pair (LP) of the oxygen in the anion to the unfilled valence-shell (LP*) in Na$^+$. The hyperconjugation energies vary in a very small range, between 6.5–10.0 kcal mol^{-1} for monovalent and between

20602 | *Phys. Chem. Chem. Phys.*, 2017, **19**, 20593–20607

This journal is © the Owner Societies 2017

Table 4 Second order perturbation theory analysis of Fock matrix in NBO basis for anion–cation pairs

Ion-pair	Donor (i)	Acceptor (j)	$E(2)(R_m)^a$ (kcal mol^{-1})	Sum
Anion–Na$^+$				
CH$_3$SO$_3^-\cdots$Na$^+$	LP(1–3)O	LP*(1–3)Na	7.5	
CH$_3$SO$_4^-\cdots$Na$^+$	LP(1–3)O	LP*(1–3)Na	6.5	
CH$_3$COO$^-\cdots$Na$^+$	LP(1–3)O	LP*(1–3)Na	10.0	
H$_2$PO$_4^-\cdots$Na$^+$	LP(1–3)O	LP*(1–3)Na	7.0	
(CH$_3$)$_2$PO$_4^-\cdots$Na$^+$	LP(1–3)O	LP*(1–3)Na	7.0	
SO$_4^{2-}\cdots$Na$^+$	LP(1–3)O	LP*(1–3)Na	11.5	
CH$_3$PO$_4^{2-}\cdots$Na$^+$	LP(1–3)O	LP*(1–3)Na	11.0	
Anion–NH$_4^+$: orientation I				
CH$_3$SO$_3^-\cdots$NH$_4^+$	LP(2)O	BD*(1)H–N	48.0	88.0
	BD*(2)S–O	BD*(1)H–N	40.0	
CH$_3$SO$_4^-\cdots$NH$_4^+$	LP(2)O	BD*(1)H–N	45.0	94.0
	BD*(2)S–O	BD*(1)H–N	49.0	
CH$_3$COO$^-\cdots$NH$_4^+$	LP(2)O	BD*(1)H–N	51.0	
SO$_4^{2-}\cdots$NH$_4^+$	LP(2)O	BD*(1)H–N	77.0	130.0
	BD*(2)S–O	BD*(1)H–N	53.0	
CH$_3$PO$_4^{2-}\cdots$NH$_4^+$	LP(2)O	BD*(1)H–N	85.5	233.5
	BD*(2)S–O	BD*(1)H–N	148.0	
Anion–NH$_4^+$: orientation II				
CH$_3$SO$_3^-\cdots$NH$_4^+$	LP(1)O	BD*(1)N–H(1–3)	4.5	
CH$_3$SO$_4^-\cdots$NH$_4^+$	LP(2)O	BD*(1)N–H(1–3)	4.5	
CH$_3$COO$^-\cdots$NH$_4^+$	LP(2)O	BD*(1)N–H(1–3)	6.0	
H$_2$PO$_4^-\cdots$NH$_4^+$	LP(1)O	BD*(1)N–H(1–3)	5.0	
(CH$_3$)$_2$PO$_4^-\cdots$NH$_4^+$	LP(1)O	BD*(1)N–H(1–3)	5.0	
SO$_4^{2-}\cdots$NH$_4^+$	LP(2)O	BD*(1)N–H(1–3)	8.0	
CH$_3$PO$_4^{2-}\cdots$NH$_4^+$	LP(2)O	BD*(1)N–H(1–3)	8.5	

^{a} Stabilization energy $E(2)$ estimated as $E(2) = \Delta E_{ij} = q_i \dfrac{F(i,j)^2}{\varepsilon_j - \varepsilon_i}$, where q_i is the donor orbital occupancy, ε_i, ε_j are diagonal elements (orbital energies) and $F(i,j)$ is the off-diagonal elements of NBO Fock matrix.[70] LP = lone pair. LP* = unfilled valence-shell. BD* = valence anti-bonding.

11.0–11.5 kcal mol^{-1} for divalent anions, suggesting that the donor–acceptor interactions are similar for all ion-pairs involving Na$^+$. Therefore, the configuration shown in Fig. 2b can be used for the *ab initio* calculations needed to derive optimized anion–cation parameters. In contrast to the ion-pairs involving Na$^+$, those involving NH$_4^+$ in configuration I (see Fig. 2c) exhibit two different electron density transfers, one from the LP of the oxygen to the valence antibonding (BD*) orbital of (H–N) and the other from BD* of (S–O) to BD* of (H–N). The total hyperconjugation energies lie between 51.0–94.0 kcal mol^{-1} for monovalent and 130.0–283.5 kcal mol^{-1} for divalent ions. This wider range of hyperconjugation energies arises from different donor–acceptor interactions in these ion-pairs, making impossible the type of comparison necessary for our parameterization approach. To resolve this problem we searched for different relative orientations of the ions involved in NH$_4^+$ ion pairs, to find an orientation in which hyperconjugation energies are both minimized and are similar for the various ion pairs. For orientation II (see Fig. 2d), the hyper-conjugation energies are in the narrow range of 4.5–6.0 kcal mol^{-1} for monovalent anions and 8.0–8.5 kcal mol^{-1} for divalent anions. We use this configuration to find the parameters for the ion-pairs containing ammonium and methyl ammonium.

3.2.2.2 Parameter optimization. For parameter optimization, we require potential energy scans for reference species for which

classical parameters already exist: we use NaCH$_3$SO$_3$ as the reference for NaCH$_3$SO$_4$, NaH$_2$PO$_4$ and Na(CH$_3$)$_2$PO$_4$; Na$_2$SO$_4$ as the reference for Na$_2$CH$_3$PO$_4$; NH$_4$CH$_3$SO$_3$ as the reference for NH$_4^+$ or CH$_3$NH$_3^+$ salts of all monovalent anions; (NH$_4$)$_2$SO$_4$ as the reference for NH$_4^+$ or CH$_3$NH$_3^+$ salts of all divalent anions. Parameter optimization then follows point 8 of the road map given in the Methods section. Example scans, for CH$_3$COO$^-$ and CH$_3$SO$_3^-$, are shown in Fig. S3 of the ESI.† In Table S2 of the ESI† we show the energy and interparticle separation at the energy minimum for all cases.

The optimized anion–cation interaction parameters are shown in Tables 5 and 6. The anion–cation distances for monovalent anions are slightly larger using the optimized parameters than using GAFF.[13,30–32] The same trend is observed for the divalent anions, although for those systems the increase in anion–cation distance is substantially larger than for the monovalent anions. Such an increase is not surprising: as mentioned above and as observed also by others,[54] it reflects the fact that Lennard-Jones parameters optimized for neutral species are not appropriate for charged species.

Our results show that optimum parameters for the NH$_4^+$– anion interactions differ markedly from those for CH$_3$NH$_3^+$– anion interactions. This result indicates that transferring parameters determined for NH$_4^+$ to other amines, which yields reasonable results for the interactions of these cations with water,[71] should be avoided for anion–cation interactions. The need for different NH$_4^+$– anion and CH$_3$NH$_3^+$–anion classical parameters that was detected with our approach illustrates the power of that approach to optimize intermolecular interactions: these two cations can be considered qualitatively similar (as defined in points 5a and 8a in the road map given in the Methods) for the purpose of our *ab initio* approach but, simultaneously, this approach is sensitive enough to detect how the interactions of the nitrogen and H$_N$ with other species quantitatively differ between NH$_4^+$ and CH$_3$NH$_3^+$, because of the impact of the methyl group on the electronic distribution around those atoms. Similarly, the parameters governing the interactions between the $\mathcal{O}$ in H$_2$PO$_4^-$ and the various cations are not identical to those for CH$_3$SO$_4^-$, again because the different substituents alter how the $\mathcal{O}$s interact with nearby species.

Our results also show that (CH$_3$)$_2$PO$_4^-$–cation interaction parameters are almost identical to those of H$_2$PO$_4^-$–cation. We thus propose that either parameter set is suitable for CH$_3$HPO$_4^-$–cation systems.

Similarly to the systems for which parameters are optimized based on experimental data, we find that the optimized parameters reduce the net anion–cation attraction relative to GAFF[13,30–32] para-meters (Fig. 6). These results suggest that GAFF[13,30–32] parameters in general overestimate cation–anion interactions and may lead to aggregation, particularly in solutions with higher salt concentration. The most dramatic changes are seen in solutions of divalent ions: GAFF[13,30–32] parameters lead to very rapid (<1 ns) aggregation of the ions and crystal formation, whereas our parameters yield solubilized systems. For all solutions containing monovalent anions, the optimized parameters decrease the number of contact ion pairs while leaving the number of solvent-shared ion-pairs unchanged. The overestimation of monovalent anion–cation interactions by GAFF is particularly marked for Na$^+\cdots$CH$_3$COO$^-$, NH$_4^+\cdots$(CH$_3$)$_2$PO$_4^-$ and Na$^+\cdots$(CH$_3$)$_2$PO$_4^-$.

This journal is © the Owner Societies 2017

Phys. Chem. Chem. Phys., 2017, **19**, 20593–20607 | **20603**

Table 5 Differences between the classical and quantum potential energy scans for all anion−Na⁺ dimers, and values of the optimized anion−Na⁺ parameters

Anion	Opt.[a]	Parameters[b]	$\Delta E_{QM \to cl}^{GAFF}(R_m)$[c] (kcal mol⁻¹)	$\Delta E_{QM \to cl}^{OPT}(R_m)$[c] (kcal mol⁻¹)	$\Delta R_{m,QM \to cl}^{GAFF}$[d] (Å)	$\Delta R_{m,QM \to cl}^{OPT}$[d] (Å)
CH_3COO^-	Exp	$1.022 \times \sigma_{\mathcal{O}Na}$	3.9	5.3	0.06	0.08
$CH_3SO_3^-$	Exp	$1.1 \times \varepsilon_{\mathcal{O}Na}$	4.6	5.1	0.05	0.06
$CH_3SO_4^-$	Comp	$1.1 \times \varepsilon_{\mathcal{O}Na}$	4.7	5.1	0.05	0.06
$H_2PO_4^{-}$ [e]	Comp	$1.02 \times \sigma_{\mathcal{O}Na}$	3.0	5.2	0.05	0.09
$(CH_3)_2PO_4^{-}$ [e]	Comp	$1.01 \times \sigma_{\mathcal{O}Na}$	3.8	4.9	0.05	0.07
SO_4^{2-}	Exp	$1.13 \times \sigma_{\mathcal{O}Na}$	5.1	22.4	0.07	0.35
$CH_3PO_4^{2-}$	Comp	$1.132 \times \sigma_{\mathcal{O}Na}$	3.1	22.0	0.06	0.35

[a] Parameters optimized based on activity derivatives (Exp), repeated here to facilitate comparisons, or based on the *ab initio* approach (Comp). [b] Expressed as a scaling factor relative to the original GAFF parameters, calculated using eqn (4) from the GAFF self-interaction parameters given in the ESI. [c] $\Delta E_{QM \to cl}(R_m)$ is defined in eqn (2). [d] $\Delta R_{m,QM \to cl}$: the difference between the position of the dimer energy minimum obtained using *ab initio* calculations and using classical parameters. [e] The parameters for $(CH_3)_2PO_4^-$ and $H_2PO_4^-$ are very similar; either set can be used for $CH_3HPO_4^-$.

Table 6 Differences between the classical and quantum potential energy scans for all anion−NH₄⁺ and anion−CH₃NH₃⁺ dimers, and values of the optimized anion−cation parameters

Anion	Cation	Opt.[a]	Parameters[b]	$\Delta E_{QM \to cl}^{GAFF}(R_m)$[c] (kcal mol⁻¹)	$\Delta E_{QM \to cl}^{OPT}(R_m)$[c] (kcal mol⁻¹)	$\Delta R_{m,QM \to cl}^{GAFF}$[d] (Å)	$\Delta R_{m,QM \to cl}^{OPT}$[d] (Å)
$CH_3SO_3^-$	NH_4^+	Exp	$1.10 \times \begin{cases} \varepsilon_{\mathcal{O}N} \\ \varepsilon_{\mathcal{O}H_N} \end{cases}$	9.7	10.1	0.23	0.25
	$CH_3NH_3^+$	Comp	$1.05 \times \begin{cases} \sigma_{\mathcal{O}N} \\ \sigma_{\mathcal{O}H_N} \end{cases}$	6.8	10.2	0.21	0.34
$CH_3SO_4^-$	NH_4^+	Comp	$1.20 \times \begin{cases} \varepsilon_{\mathcal{O}N} \\ \varepsilon_{\mathcal{O}H_N} \end{cases}$	9.1	9.9	0.24	0.27
	$CH_3NH_3^+$	Comp	$1.05 \times \begin{cases} \sigma_{\mathcal{O}N} \\ \sigma_{\mathcal{O}H_N} \end{cases}$	7.4	10.6	0.21	0.35
$H_2PO_4^{-}$ [f]	NH_4^+	Comp	$1.02 \times \begin{cases} \sigma_{\mathcal{O}N} \\ \sigma_{\mathcal{O}H_N} \end{cases}$	8.9	10.4	0.20	0.26
	$CH_3NH_3^+$	Comp	$1.04 \times \begin{cases} \sigma_{\mathcal{O}N} \\ \sigma_{\mathcal{O}H_N} \end{cases}$	7.0	10.0	0.19	0.30
$(CH_3)_2PO_4^{-}$ [f]	NH_4^+	Comp	$1.01 \times \begin{cases} \sigma_{\mathcal{O}N} \\ \sigma_{\mathcal{O}H_N} \end{cases}$	9.5	10.1	0.21	0.22
	$CH_3NH_3^+$	Comp	$1.04 \times \begin{cases} \sigma_{\mathcal{O}N} \\ \sigma_{\mathcal{O}H_N} \end{cases}$	7.5	10.4	0.19	0.30
CH_3COO^-	NH_4^+	Comp	—[e]	10.9	10.2	0.23	0.20
	$CH_3NH_3^+$	Comp	$1.02 \times \begin{cases} \sigma_{\mathcal{O}N} \\ \sigma_{\mathcal{O}H_N} \end{cases}$	8.9	9.9	0.21	0.24
SO_4^{2-}	NH_4^+	Exp	$1.13 \times \begin{cases} \sigma_{\mathcal{O}N} \\ \sigma_{\mathcal{O}H_N} \end{cases}$	16.9	30.3	0.26	0.59
	$CH_3NH_3^+$	Comp	$1.14 \times \begin{cases} \sigma_{\mathcal{O}N} \\ \sigma_{\mathcal{O}H_N} \end{cases}$	15.8	30.0	0.25	0.61
$CH_3PO_4^{2-}$	NH_4^+	Comp	$1.13 \times \begin{cases} \sigma_{\mathcal{O}N} \\ \sigma_{\mathcal{O}H_N} \end{cases}$	16.0	30.0	0.25	0.58
	$CH_3NH_3^+$	Comp	$1.15 \times \begin{cases} \sigma_{\mathcal{O}N} \\ \sigma_{\mathcal{O}H_N} \end{cases}$	14.7	30.7	0.23	0.62

[a] Parameters optimized based on activity derivatives (Exp) or based on the *ab initio* approach (Comp). [b] Expressed as a scaling factor relative to the original GAFF parameters, calculated using eqn (4) from the GAFF self-interaction parameters given in the ESI. [c] $\Delta E_{QM \to cl}(R_m)$ is defined in eqn (2). [d] $\Delta R_{m,QM \to cl}$: the difference between the position of the dimer energy minimum obtained using *ab initio* calculations and using classical parameters. [e] No scaling of anion−cation interactions. [f] The parameters for $(CH_3)_2PO_4^-$ and $H_2PO_4^-$ are very similar; either set can be used for $CH_3HPO_4^-$.

GAFF[13,30–32] also particularly overestimates the interactions between $CH_3NH_3^+$ and CH_3COO^-. This effect is illustrated in Fig. 5b, by comparing the radial distribution of a 3 m solution of glycine zwitterion using GAFF[72] and using optimized parameters. Overestimation of interactions between anionic and cationic side-chains, *i.e.*, salt-bridge formation, has also been observed by others.[73] This artifact can be prevented by using our optimized anion–cation interaction parameters.

3.3 Parameters obtained using the *ab initio* approach yield correct solution properties

As described in the introduction, our approach to optimize parameters based on *ab initio* calculations relies on an assumption:

20604 | *Phys. Chem. Chem. Phys.*, 2017, **19**, 20593–20607

This journal is © the Owner Societies 2017

that, for similar anions, the difference, $\Delta E_{QM \to cl}(R_{min})$, between the minimum binding energy of dimers from *ab initio* calculations and that calculated using optimized classical parameters should be similar for similar dimers (see Fig. 1). To evaluate whether this assumption holds, we compare the classical and *ab initio* gas phase potential energy scans for the $CH_3COO^- \cdots H_2O$ and the $HSO_4^- \cdots H_2O$ systems, and also for $CH_3SO_3^- \cdots Na^+$ and the $CH_3COO^- \cdots Na^+$ systems. These are the only systems for which Lennard-Jones parameters were optimized based on experimental data and which contain.

For the anion–water case the scans are performed in the system configuration shown in Fig. 2a; these results are shown in Fig. 7. We find (see Table 2) that $\Delta E_{QM \to cl}(R_{min}) = -1.3$ kcal mol^{-1} for HSO_4^-, and $\Delta E_{QM \to cl}(R_{min}) = -1.2$ kcal mol^{-1} for CH_3COO^-; the similarity in the $\Delta E_{AB,QM \to cl}(R_{min})$ values for these two anions confirms that the *ab initio* approach proposed here holds for anion–water dimers. The scans also confirm that developing parameters by requiring that the classical potential energy curve overlaps with the *ab initio* is unsuitable. Parameters developed that way would greatly overestimate anion–water interactions.

For the anion–Na$^+$ systems the potential energy scans are performed in the configuration shown in Fig. 2b. We find, as shown in Table 5, that $\Delta E_{QM \to cl}(R_{min}) = +5.1$ kcal mol^{-1} for $CH_3SO_3^-$ and $\Delta E_{QM \to cl}(R_{min}) = +5.3$ kcal mol^{-1} for CH_3COO^-. The two values are very similar, confirming that the approach holds also for anion–cation systems.

Our parameters for the $(CH_3)_2PO_4^- \cdots CH_3NH_3^+$ interaction are very close to recently proposed parameters based that reproduce osmotic pressure measurements made on DNA arrays: we propose that $\sigma_{\mathcal{O}N} = 1.04 \times \sigma_{\mathcal{O}N,GAFF}$, which corresponds to a +0.124 Å increase in the $\sigma_{\mathcal{O}N}$ value. This value is very close to the 0.14 Å increase proposed by Aksimentiev and Yoo.[22]

As a further check of our *ab initio* approach, in Fig. 5b we also compare the anion–cation rdf for glycine obtained using our optimized parameters for $CH_3COO^- \cdots CH_3NH_3^+$ interactions, with that obtained using parameters optimized from recent osmotic pressure measurements.[22] The similarity between our parameter values and those of Aksimentiev (see ref. 22, Table 2), and also of the radial distribution function further confirms the success of our approach.

Finally, for a few of the salts for which the experimental solution activities are available, we calculate the densities of
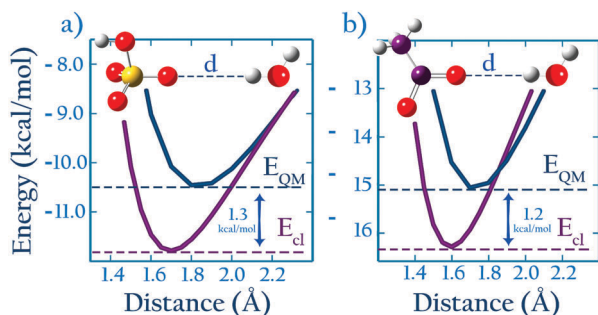
Fig. 8 Densities of three salts calculated using the optimized and the original GAFF parameters experimental densities.[59] The standard deviation of the simulation values is of order the height of the symbols. See ESI,† Table S6, for the numerical values.

0.5 m solutions and compare it against experimental data. The results, shown in Fig. 8, indicate that our parameters improve agreement with experiment, especially in the case of the divalent salts for which the densities obtained using the original GAFF parameters deviated substantially from the reference values.

4 Concluding remarks

We present an approach to develop classical Lennard-Jones parameters for ions, that can be used when experimental data for parameterization is sparse. The approach is based on MP2 level *ab initio* calculations but simultaneously integrates experimental data, to create a parameter set for multiple ions that is internally consistent and grounded on experiment. We chose the MP2 level of theory for the gas phase *ab initio* scans because MP2 reproduces relative binding energies in systems with hydrogen bonding or dispersive interactions well.[74] DFT methods that correct for the dispersion terms, such as DFT-D[75,76] and M06-2X,[77–79] however, might also yield reasonable results at a lower computational cost; this possibility remains to be tested.

Our tests show that our simple approach is very powerful, and is suitable to obtain both ion–water and anion–cation interaction parameters. The approach may be used to rapidly optimize parameters for one or multiple atom types in any species – mono- and polyatomic; mono- and polyvalent. When optimizing multiple atom types, however, it is desirable that multiple target experimental properties exist.

We use this parameterization approach to obtain a set of Lennard-Jones ion–water and anion–cation parameters for salts of various sulfate, sulfonate, phosphate and carboxylate ions and the cations Na$^+$, NH$_4^+$ and CH$_3$NH$_3^+$, in TIP3P water.[29] The entire parameter set is compatible with the GAFF[13,30–32] force field and the sodium model of Joung and Cheatham.[15] GAFF[30] uses the same atom types as the AMBER[13] force field for proteins, so the parameters presented are also compatible with AMBER. The large range of anions tested and parameterized here and the compatibility with GAFF/AMBER should prove useful to investigate ion-specific effects in biology that arise from the presence of multivalent ions.

Fig. 7 (a) HSO$_4^-$–H$_2$O (b) CH$_3$COO$^-$–H$_2$O, 1D energy scan with both QM and classical methods. The curves in (a) are averaged over scans of different symmetries, as described in the ESI.†
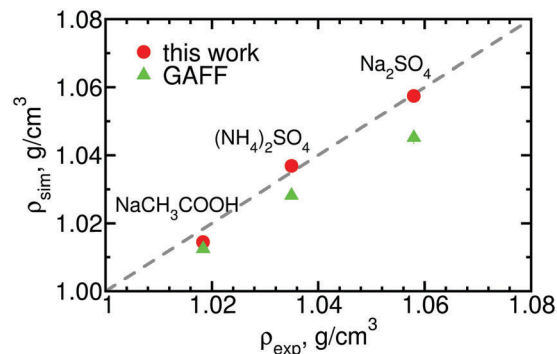
This journal is © the Owner Societies 2017

Phys. Chem. Chem. Phys., 2017, **19**, 20593–20607 | **20605**

Our results show that the GAFF[13,30–32] Lennard-Jones parameters, which were developed based on data for neutral species, often yield incorrect ion–TIP3P water and anion–cation interactions; the results suggest that the same may occur for other similar, general, force fields. GAFF[13,30–32] divalent anions consistently interact far too favorably with TIP3P water; in contrast, ion–TIP3P water interactions for monovalent anions are typically well-represented in GAFF.[13,30–32] GAFF[13,30–32] parameters typically overestimate anion–cation attraction. Excessive anion–cation attraction is particularly marked for divalent anions, for which GAFF[13,30–32] parameters lead to complete aggregation of 0.5 m solutions within timescales of nanoseconds. GAFF[13,30–32] also markedly overestimates the attraction between acetate and the Na^+ ion of Joung and Cheatham,[15] as well as attraction between acetate and $CH_3NH_3^+$; these findings demonstrate that strong Na^+–protein or salt bridges between cationic and anionic amino acids in proteins in simulations using AMBER[13] and the Na^+ ion from Joung and Cheatham[15] are unphysical.

Lennard-Jones parameters developed for NH_4^+–anion interactions are often used for primary amines – and thus of the amino acid lysine.[22] Our results indicate, however, that that usage is flawed because the parameters governing the interaction of NH_4^+ with the anions tested here are substantially different from those of $CH_3NH_3^+$. Applying parameters derived for NH_4^+–anion interactions to simulate $CH_3NH_3^+$–anion systems will typically overestimate the attraction between the amine group and the anion.

Parameters developed for oxygen atoms in sulfate and phosphate are often considered interchangeable.[22] We find this approximation holds for the divalent sulfate or phosphate ions we investigated here (SO_4^{2-}, $CH_3PO_4^{2-}$), but not for monovalent variants of those anions ($CH_3SO_4^-$, $CH_3SO_3^-$, $H_2PO_4^-$, $(CH_3)_2PO_4^-$). These differences between mono and divalent ions likely result from the fact that the net anion–cation attraction has a larger contribution from the electrostatic component for the divalent cations.

Acknowledgements

References

1 H. G. Garg, L. Yu, C. A. Hales, T. Toida, T. Islam and R. J. Linhardt, *Biochim. Biophys. Acta, Mol. Basis Dis.*, 2003, **1639**, 225–231.

2 S. A. Smith, N. J. Mutch, D. Baskar, P. Rohloff, R. Docampo and J. H. Morrissey, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 903–908.

3 J. T. Gallagher, *Handb. Exp. Pharmacol.*, 2012, **207**, 347–360.

4 R. Baldwin, *Biophys. J.*, 1996, **71**, 2056–2063.

5 K. D. Collins, *Methods*, 2004, **34**, 300–311.

6 W. L. Ash, M. R. Zlomislic, E. O. Oloo and D. P. Tieleman, *Biochim. Biophys. Acta, Biomembr.*, 2004, **1666**, 158–189.

7 D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan and W. Wriggers, *Science*, 2010, **330**, 341–346.

8 K. Lindorff-Larsen, N. Trbovic, P. Maragakis, S. Piana and D. E. Shaw, *J. Am. Chem. Soc.*, 2012, **134**, 3787–3791.

9 C. Zhang and J. Ma, *J. Chem. Phys.*, 2010, **132**, 244101.

10 A. A. Skjevik, B. D. Madej, C. J. Dickson, C. Lin, K. Teigen, R. C. Walker and I. R. Gould, *Phys. Chem. Chem. Phys.*, 2016, **18**, 10573–10584.

11 A. Savelyev and G. A. Papoian, *J. Am. Chem. Soc.*, 2007, **129**, 6060–6061.

12 A. van der Vaart, *Biochim. Biophys. Acta, Gen. Subj.*, 2015, **1850**, 1091–1098.

13 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.

14 A. D. Mackerell and M. Karplus, *J. Phys. Chem.*, 1991, **2**, 10559–10560.

15 I. S. Joung and T. E. Cheatham, *J. Phys. Chem. B*, 2008, **112**, 9020–9041.

16 B. Klasczyk and V. Knecht, *J. Chem. Phys.*, 2010, **132**, 024109.

17 D. Horinek, S. I. Mamatkulov and R. R. Netz, *J. Chem. Phys.*, 2009, **130**, 124507.

18 M. Fyta, I. Kalcher, J. Dzubiella, L. Vrbka and R. R. Netz, *J. Chem. Phys.*, 2010, **132**, 024911.

19 M. Fyta and R. R. Netz, *J. Chem. Phys.*, 2012, **136**, 124103.

20 V. Satarifard, S. Kashefolgheta, A. Vila Verde and A. Grafmueller, *J. Chem. Theory Comput.*, 2017, **13**, 2112–2122.

21 M. Kanduc, A. Schlaich, E. Schneck and R. R. Netz, *Langmuir*, 2016, **32**, 8767–8782.

22 J. Yoo and A. Aksimentiev, *J. Chem. Theory Comput.*, 2016, **12**, 430–443.

23 A. Vila Verde, M. Santer and R. Lipowsky, *Phys. Chem. Chem. Phys.*, 2016, **18**, 1918–1930.

24 A. Vila Verde and R. Lipowsky, *J. Phys. Chem. B*, 2013, **117**, 10556–10566.

25 M. Weinhart, D. Gröger, S. Enders, J. Dernedde and R. Haag, *Biomacromolecules*, 2011, **12**, 2502–2511.

26 K. D. Collins, *Biophys. Chem.*, 2012, **167**, 33–49.

27 Y. J. Zhang and P. S. Cremer, *Curr. Opin. Chem. Biol.*, 2006, **10**, 658–663.

28 C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.

29 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926.

30 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.

31 J. Wang, P. Cieplak and P. A. Kollman, *J. Comput. Chem.*, 2000, **21**, 1049–1074.

32 W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1988, **110**, 1657–1666.

20606 | *Phys. Chem. Chem. Phys.*, 2017, **19**, 20593–20607

This journal is © the Owner Societies 2017

33 Y. Marcus, *Ion properties*, Marcel Dekker, Inc., New York, USA, 1997.

34 E. Pluharova, O. Marsalek, B. Schmidt and P. Jungwirth, *J. Phys. Chem. Lett.*, 2013, **4**, 4177–4181.

35 M. D. Daily, M. D. Baer and C. J. Mundy, *J. Phys. Chem. B*, 2016, **120**, 2198–2208.

36 M. D. Baer, V.-T. Pham, J. L. Fulton, G. K. Schenter, M. Balasubramanian and C. J. Mundy, *J. Phys. Chem. Lett.*, 2011, **2**, 2650–2654.

37 W. L. Jorgensen, *J. Phys. Chem.*, 1986, **90**, 1276–1284.

38 I. Leontyev and A. Stuchebrukhov, *Phys. Chem. Chem. Phys.*, 2011, **13**, 2613–2626.

39 P. Hunenberger and M. Reif, *Single-ion solvation – experimental and theoretical approaches to elusive thermodynamic quantities*, RSC, 2011.

40 H. J. C. Berendsen, D. van der Spoel and R. van Drunen, *Comput. Phys. Commun.*, 1995, **91**, 43–56.

41 D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701–1718.

42 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.

43 W. F. van Gunsteren and J. C. Berendsen, *Mol. Simul.*, 1988, **1**, 173–185.

44 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

45 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.

46 J. G. Kirkwood and F. P. Buff, *J. Chem. Phys.*, 1951, **19**, 774–777.

47 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.

48 C. H. Bennett, *J. Comput. Phys.*, 1976, **22**, 245–268.

49 U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, 1984, **5**, 129–145.

50 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *Gaussian 03, Revision E.01*, Gaussian, Inc., Wallingford, CT, 2004.

51 C. Møller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 618–622.

52 T. H. Dunning Jr, *J. Chem. Phys.*, 1989, **90**, 1007.

53 T. Steiner, *Angew. Chem., Int. Ed.*, 2002, **41**, 48–76.

54 T. Steinbrecher, J. Latzer and D. A. Case, *J. Chem. Theory Comput.*, 2012, **8**, 4405–4412.

55 P. George, R. J. Witonsky, M. Trachtman, C. Wu, W. Dorwart, L. Richman, W. Richman, F. Shurayh and B. Lentz, *Biochim. Biophys. Acta, Bioenerg.*, 1970, **223**, 1–15.

56 *Activity coefficients in electrolyte solutions*, ed. K. S. Pitzer, CRC Press, Boca Raton, Fl., USA, 2nd edn, 1991, pp. 75–153.

57 R. W. Potter II and M. A. Clynne, *J. Res. U.S. Geol. Surv.*, 1978, **6**, 701–705.

58 J. Barthel, *Ber. Bunsen-Ges.*, 1985, **89**, 722–723.

59 *CRC handbook of Chemistry and Physics*, ed. W. M. Haynes and D. R. Lide, CRC Press, Boca Raton, Fl, USA, 2011.

60 A. E. Reed, L. a. Curtiss and F. Weinhold, *Chem. Rev.*, 1988, **88**, 899–926.

61 E. Mrázková and P. Hobza, *J. Phys. Chem. A*, 2003, **107**, 1032–1039.

62 P. Hobza and Z. Havlas, *Theor. Chem. Acc.*, 2002, **108**, 325–334.

63 B. J. Van der Veken, W. A. Herrebout, R. Szostak, D. N. Shchepkin, Z. Havlas and P. Hobza, *J. Am. Chem. Soc.*, 2001, **123**, 12290–12293.

64 C. P. Petersen and M. S. Gordon, *J. Phys. Chem. A*, 1999, **103**, 4162–4166.

65 M. Soniat, PhD thesis, University of New Orleans, 2014.

66 M. Soniat and S. W. Rick, *J. Chem. Phys.*, 2012, **137**, 44511.

67 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.

68 A. D. Becke, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1988, **38**, 3098–3100.

69 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1988, **37**, 785–789.

70 E. D. Glendening, J. K. Badenhoop, A. E. Reed, J. E. Carpenter and A. Bohmann, C. M. Morales, C. R. Landis and F. Weinhold, *J. NBO 6. 0.*, Theoretical and Chemistry Institute, University of Wisconsin, Madison, 2013.

71 W. L. Jorgensen and J. Gao, *J. Phys. Chem.*, 1986, **90**, 2174–2182.

72 A. H. C. Horn, *J. Mol. Model.*, 2014, **20**, 2478.

73 S. Sukenik, Y. Boyarski and D. Harries, *Chem. Commun.*, 2014, **50**, 8193–8196.

74 K. E. Riley, M. Pitoňak, J. Černý and P. Hobza, *J. Chem. Theory Comput.*, 2010, **6**, 66–80.

75 J.-i. Fujisawa, N. Tajima, K. Tamaki, M. Shimomura and T. Ishihara, *J. Phys. Chem. C*, 2007, **111**, 1146–1149.

76 P. Jurečka, J. Černý, P. Hobza and D. R. Salahub, *J. Comput. Chem.*, 2007, **28**, 555–569.

77 Y. Zhao and D. G. Truhlar, *J. Chem. Theory Comput.*, 2007, **3**, 289–300.

78 Y. Zhao and D. G. Truhlar, *J. Chem. Phys.*, 2006, **125**, 194101.

79 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241.

This journal is © the Owner Societies 2017

Phys. Chem. Chem. Phys., 2017, **19**, 20593–20607 | 20607