RSC Advances

Journal Name

ARTICLE TYPE

Improving the description of interactions between Ca^{2+} and protein carboxylate groups, including γ-carboxyglutamic acid: Revised CHARMM22* parameters[†]

Andrew T. Church[a], Zak E. Hughes[a] and Tiffany R. Walsh[*a]

A reliable description of ion pair interactions for biological systems, particularly those involving polyatomic ions such as carboxylate and divalent ions such as Ca^{2+}, using biomolecular force-fields is essential for making useful predictions for a range of protein functions. In particular, the interaction of divalent ions with the double carboxylate group present in γ-carboxyglutamic acid (Gla), relevant to the function of many proteins, is relatively understudied using biomolecular force-fields. Using force-field based metadynamics simulations to predict the free energy of binding between Ca^{2+} and the carboxylate group in liquid water, we show that a widely-used biomolecular force-field, CHARMM22*, substantially over-estimates the binding strength between Ca^{2+} and the side-chains of both glutamic acid (Glu) and Gla, compared with experimental data obtained for the analogous systems of aqueous calcium–acetate and calcium–malonate. To correct for this, we propose and test a range of modifications to the σ value of the heteroatomic Lennard-Jones interaction between Ca^{2+} and the oxygen of the carboxylate group. Our revised parameter set can recover the same three association modes of this aqueous ion pair as the standard parameter set, and yields free energies of binding for the carboxylate–Ca^{2+} interaction in good agreement with experimental data. The revised parameter set recovers other structural properties of the ion pair in agreement with the standard CHARMM22* parameter set.

The interactions between ions and biomolecules are central to the behaviour of many important biological systems including, but not limited to: ion channels,[1–4] conotoxins,[5–7] and bio-mineralisation processes.[8–10] Since many of the peptides/proteins in these systems have complex structure/property relationships, clear elucidation of their corresponding conformational ensembles will enable advances in the design of new *de novo* biomolecules for a variety of applications.[10–13] Experimental approaches have been invaluable in providing structural information about these systems.[1,2,7,8,14,15] However, resolving definitive structural information at the atomistic level from experimental observations alone can be challenging.[4,7,11,16–18] Molecular simulation-based approaches can provide complementary insights into the role played by ionic species in the structure/function relationship of these proteins/peptides.[4,19–23] While molecular simulation can be a powerful tool, for it to provide *meaningful* physical insights into the behaviour of these systems it is vital that the description of the potential energy landscape (PEL) of the system under study is physically reasonable.

Quantum mechanics (QM) based techniques have the potential to offer a detailed and reliable description of the PEL. First-principles molecular dynamics (FPMD) simulations have provided information about the behaviour of biologically relevant ionic species.[24–29] However, the high computational expense of such methods limits the time- and length-scales that can be accessed, making them impractical for the simulation of large biomolecules. While not providing as reliable a description of the PEL as QM techniques, force-field (FF) based simulations offer an acceptable compromise between physical veracity and computational expense. For this reason they have been widely utilised for the simulation of complex biomolecules. Some of the most widely used biomolecular FFs include the CHARMM,[30,31] AMBER[32,33] and GROMOS[34] families of force fields.

Despite their maturity, and as prompted by dual advances in experimental techniques and conformational sampling approaches,

a Institute for Frontier Materials, Deakin University, Geelong, Australia. Fax: +61 (0)3 5227 1103; Tel: +61 (0)3 5227 3116; E-mail: tiffany.walsh@deakin.edu.au
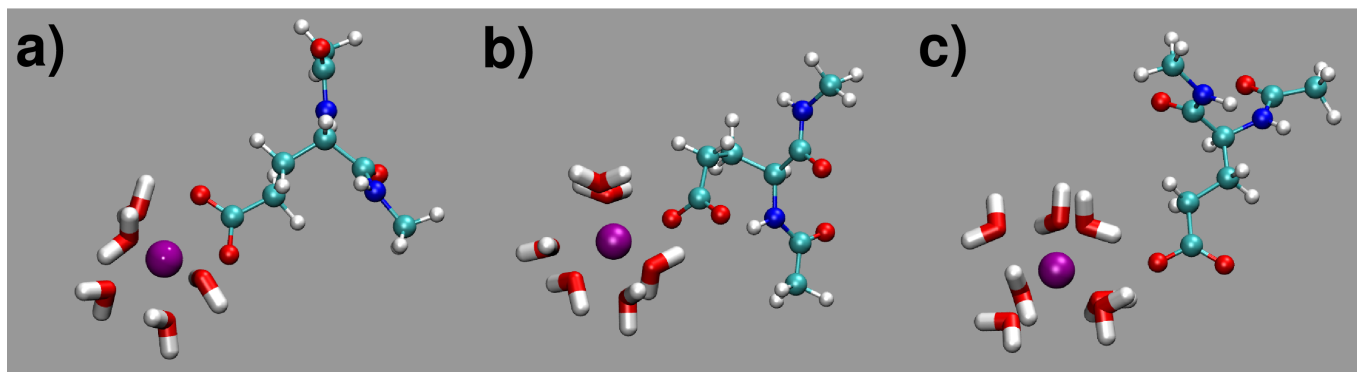
Fig. 1 Illustration of the three classes of co-ordination mode between Ca^{2+} and a carboxylate group: **a)** biCIP, **b)** monoCIP, and **c)** SIP. Images are snapshots taken from simulations for the aqueous Glu–Ca^{2+} system.

the parameters of biomolecular FFs are periodically refined to improve their performance.[31,35–37] To describe LJ interactions that have not been explicitly parametrised, FFs such as CHARMM and AMBER use the Lorentz-Berthelot combining rules to generate heteroatomic LJ parameters.[38,39] While the parameters derived from combining rules are able to reasonably describe many of the inter-atomic interactions within biomolecular systems, for some interactions, *e.g.* the interactions between different ions or the interactions between species at the hard-soft matter interface, they may break down,[40–46] leading to an unphysical description of the system.

The interplay of interactions between ions, biomolecules and water is complex; several previous studies have sought ion parameters that can capture the balance between the ion-water, ion-biomolecule and ion-ion interactions appropriately.[37,41,44,47–50] Properties such as the interaction strength of ions with the side-chains of charged residues, the structuring of solvation shells, and ion coordination are all thought to be important factors that contribute to the ability of a FF to recover this interplay. This interplay can sometimes be highly challenging to recover if the interactions between species are described through the use of default FF parameters and combining rules alone.[27,29,37,42,45,47]

One particular interaction that has been previously demonstrated to require careful attention is that between Ca^{2+} ions and carboxylate groups,[29,42,45] such as those that appear in the acidic residues of proteins and peptides. Previous studies have shown that a number of common biomolecular FFs do not correctly capture the strength of this interaction.[29,42,45] Such a deficiency in the FF will have an influence on the saturation rate of ion pairings in biomolecules and may prevent the FF from accurately modeling the functionality of the system.

The ability of the FF to capture structural details, particularly the different coordination modes (see Figure 1) of divalent ions with carboxylates, is also critical. The contact ion pair (CIP) refers to the case where there is direct (*i.e.* non-solvent mediated) contact between the carboxylate oxygen atoms and the divalent ion, and the solvent-shared ion pair (SIP) refers to a solvent-separated mode of interaction. There are two distinct CIP states, the first referred to here as the biCIP state (see Figure 1a) represents a bidentate coordination, while the monoCIP state (Figure 1b)

represents a monodentate coordination. The term SIP refers to the solvent-separated ion-pair (Figure 1c). The ability to distinguish and correctly coordinate ions to the specific binding locations on acidic residue side-chains is thought to be crucial in systems including signal transduction pathways and catalytic processes.[51–54] In the hydroxyapatite-binding sequence BMP2, previous studies have suggested that the orientation of a Glu side chain, and hence the orientation of the carboxylate group relative to the surface, plays a role in the strength of the surface-residue interaction.[55] The effects that different binding modes can exert on hydration shells of Ca^{2+} when interacting with carboxylate groups have also been reported.[56] Such properties highlight the potential importance that coordination modes may play in protein/ion binding.

A further factor to consider regarding the Ca^{2+}-carboxylate interaction is that many hydroxyapatite-binding bio-mineralisation sequences contain the post-transitionally modified γ-carboxyglutamic acid (Gla) residue. The Gla residue contains two carboxylate groups and is thought to be critical in triggering folding and binding of these sequences.[11,17,57–59] In Figure 2 we show the difference in structure between the Glu and Gla residue side-chains. Both are negatively-charged; the Glu side-chain carries a $-1e$ charge, while the charge on Gla is $-2e$. The Gla residue is also found in many other natural systems,[60–64] such as conantokins, *e.g.* conantokin-Pr1, which adopts a helical conformation in the presence of Ca^{2+} or Mg^{2+}, but is unstructured in the absence of these ions.[5] Though Gla is widely found throughout nature and is central to many well-studied mineralisation sequences, the Gla residue has not been well explored in molecular simulation studies. This may in part be due to the lack of well-defined parametrisation of Gla in many common biomolecular force-fields.

Therefore, there is a clear need to test, and if necessary modify, FF parameters related to the association of carboxylate groups (such as in Glu and Gla) with divalent ions such as Ca^{2+}. The modification of a FF such that it more accurately describes the Ca^{2+}–carboxylate interaction should be done with care. While we wish to reproduce the features (binding strength, coordination mode, interaction with the Gla residue) described above, we also need to ensure that any changes we make do not inadver-
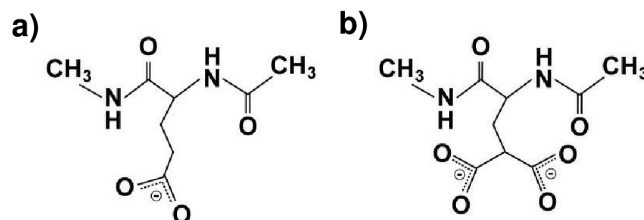
Fig. 2 Chemical structure of **a)** glutamic acid (Glu), and **b)** γ-carboxyglutamic acid (Gla).

tently disrupt the behaviour of other parts of the system. The best way of meeting these criteria is specifically targeting the modification of the non-bonded interactions between Ca^{2+} and carboxylate side-chains. [29,42]

Recently, Kahlen *et al.* [29] reported the optimisation of Ca^{2+}–O_{carb} Lennard-Jones (LJ) parameters for a range of force-fields. To test these modifications, these authors calculated potential of mean force (PMF) profiles for the aqueous calcium acetate complex, and used these to derive the Ca^{2+}–acetate association constant, enabling direct comparison with experimental data. The FFs were optimised via an increase of the heteroatomic σ LJ parameter associated with the Ca^{2+}–O_{carb} pair. This optimisation procedure was reported for both the GROMOS [34] and OPLS [65] FFs; however, the CHARMM22* force-field was neither evaluated nor modified. These modifications were taken on the basis of an integrated property (the association constant), which may not admit a unique PEL as a solution; in other words, many variations of the PEL as a function of the calcium–acetate separation may also produce reasonable association constants. In particular, the delicate balance in free energy between the biCIP, monoCIP and SIP states cannot be resolved via comparison with integrated properties (such as the association constant) alone. In principle, FPMD simulations may offer a viable route to determining this free energy balance, but in practice, at present the uncertainties associated with the resulting free energy profiles computed using this approach [29] prevent definitive resolution of this question.

Here, we have investigated and quantified the strength of the Ca^{2+}–carboxylate interaction, and also the resulting local hydration structure and coordination modes. To determine the relevant PMF profiles for these interactions, we carried out metadynamics [66,67] simulations of the capped Glu and Gla amino acids interacting with a calcium ion in solution using the CHARMM22* FF, [30,31] and compared the resulting association constants with those experimentally obtained for both calcium acetate and calcium malonate. Our results show that the standard CHARMM22* FF parameters for the calcium–carboxylate interaction yield a substantial over-estimate of the binding free energy compared with experimental data. This over-estimation was found to be particularly pronounced for the Gla residue. We then optimised the CHARMM22* [30,31] Ca^{2+}–O_{carb} heteroatomic LJ parameters for the Glu residue, as well as the γ-carboxyglutamic acid (Gla) residue, to match our predicted association constants with experimentally-determined values obtained for both calcium acetate and calcium malonate. The structural data generated using this new LJ parameter resulted in close agreement with previously reported findings.

Methods

We considered two separate systems in our molecular dynamics simulations. The first comprised a single capped Glu amino acid, hydrated with 2163 water molecules, one Ca^{2+} ion, and a Cl^- counterion to ensure overall charge neutrality of the simulation cell. In the second system, we considered a single capped Gla amino acid, with the same number of water molecules, and one Ca^{2+} ion. In each instance, the amino acid N- and C-termini were capped with acetyl and N-methyl groups, respectively.

Molecular Dynamics Simulations

We used the modified TIPS3P force-field [68,69] to describe the interactions between liquid water and its environment. For the biomolecules and the Ca^{2+} ion, we used the CHARMM22* force-field. [30,31] Force-field details for the Gla residue are provided in the Electronic Supplementary Information (ESI†) in Section 'Gla Force-field Summary'. The procedure for how we modified the CHARMM22* force-field to recover the experimental binding constants, for calcium acetate and calcium malonate, are provided below. We used the GROMACS 4.6.1 software package [70] for all of the simulations reported herein. In all simulations, Newton's equations of motion were solved using the leapfrog algorithm [71] with a time-step of 1 fs. The LJ interactions were modeled using the switch functional with a decay starting at 1 nm, going to zero by 1.2 nm. Long-range electrostatic interactions were handled using the Particle Mesh Ewald (PME) [72] approach with a 1.2 nm cutoff.

Once both systems were constructed, they were each placed in a 4 nm cubic cell and were then equilibrated. First, each system was energy minimised and subsequently subjected to a short (200 ps) simulation in the Canonical (NVT) ensemble at our target temperature of 310 K, maintained using the Nosé-Hoover thermostat. [73,74] Next, we ran a short (0.5 ns) simulation in the isothermal-isobaric (NPT) ensemble [75] at 1 atm pressure and 310 K, with a pressure-coupling constant of 0.4 ps. All of our production simulations were carried out at 310 K. This temperature is of interest because it corresponds to the most likely physiological scenario for probing interactions between mammalian (human) Gla-containing proteins and Ca^{2+} under aqueous conditions. The final frame from these simulations was used as the initial starting configuration of our metadynamics simulations (see below).

To modify the force-field in our work, we have chosen to utilise the approach outlined by Project *et al.* [42] and utilised by Kahlen *et al.* [29], where slight modifications to the heteroatomic LJ Ca^{2+}–O_{carb} σ parameter were implemented. [29,42] Modifications to the
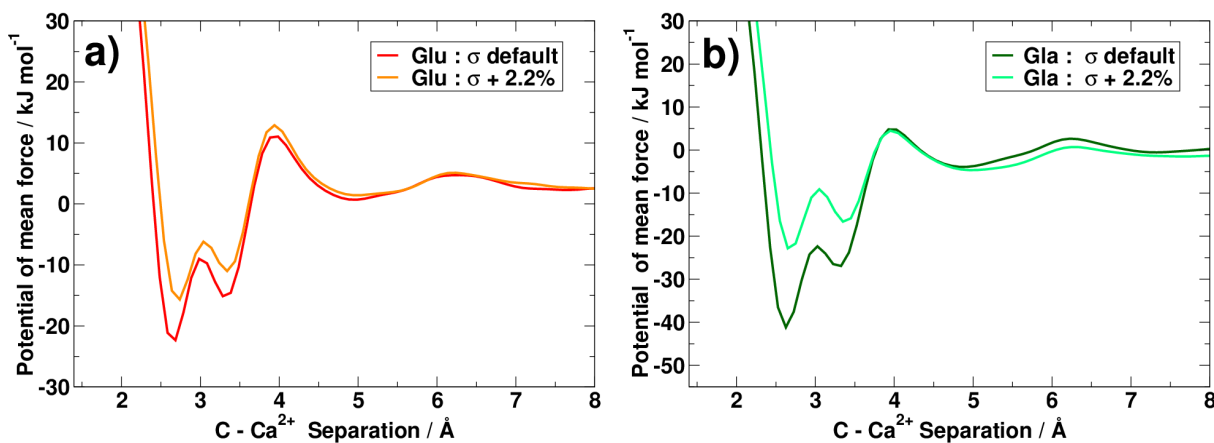
Fig. 3 Predicted potential of mean force profiles for the interaction between Ca^{2+} and the carboxylate side chain of **a)** glutamic acid (Glu), and **b)** carboxyglutamic acid (Gla), for both the unmodified CHARMM22* force-field, and for the $\sigma + 2.2\%$ force-field modification (see text for details).

bespoke heteroatomic LJ parameters can be readily implemented in GROMACS to over-rule the standard Lorentz-Berthelot mixing rules, such that only the interaction between the target pair of atoms is affected. Following Kahlen *et al.*[29], we systematically increased the σ parameter of the heteroatomic LJ Ca^{2+}–O_{carb}, where each increase is expressed as a percentage of the unmodified value, in increments of 0.2%, scanning the range of ($\sigma + 2.0\%$)–($\sigma + 3.0\%$), resulting in a total of seven different parameter values (including the unmodified value). Table S4 in the ESI† provides the explicit σ values used for each case.

We carried out seven well-tempered metadynamics[76] simulations for the aqueous Glu–calcium system, one for each value of the LJ Ca^{2+}–O_{carb} σ parameter. For the Gla-calcium system, we used our metadynamics results obtained for Glu to narrow down the range of sigma values. In this instance we deemed it sufficient to consider the range ($\sigma + 2.0$)–($\sigma + 2.4\%$), in addition to the unmodified LJ parameters, giving a total of four metadynamics simulations for Gla.

We used the PLUMED[77] plugin along with GROMACS to perform these calculations. The collective variable (CV) for the Glu/calcium simulations was defined as the distance between the central carbon of the carboxylate group and the calcium ion. Because the Gla residue contains two carboxylate groups, the collective variable for the aqueous Gla/calcium system was defined as the distance between the calcium ion and the central carbon of the carboxylate group that was closest to the calcium ion at each time frame. Gaussian hills were deposited every 1 ps with a hill height of $0.1\,\mathrm{kJ\,mol^{-1}}$ and a width of 0.025 nm. A bias factor of 10 was used throughout. We ran these simulations in the NVT ensemble, with the temperature maintained at $310\,\mathrm{K}$ using the Nosé-Hoover thermostat. The full range of the CV explored in our simulations was $\sim 32\,\text{Å}$. We ran our metadynamics simulations until the resulting potential of mean force (PMF) profile was showing only minimal evolution with time. This yielded a simulation time of 100-120 ns for each metadynamics simulation. All resulting PMF profiles obtained for a given system (either Glu or Gla) were set to zero at a CV separation of $13\,\text{Å}$.

Following our metadynamics simulations, we carried out some

additional standard MD simulations for the purposes of generating specific structural data. These simulations were also carried out in the NVT ensemble at the same target temperature and for the same simulation settings as specified above.

Analysis

Experimental data for the strength of interactions typically takes the form of association constants (K_a). To compare the strength of the interaction predicted by our MD simulations against the association constants obtained from experimental data, in each case the PMF obtained from each of the metadynamics simulations was processed to allow the calculation of a predicted association constant, following Chialvo *et al.*[80], as was reported by Kahlen *et al.*[29]. To accomplish this, each PMF was first converted to a radial distribution function of the ion-ligand distance (gx_{AC}^{id}) based on equation 1.

$$g^{id}(r) = exp\left(-\frac{V^{PMF}(r)}{k_b T}\right) \tag{1}$$

Next, $g^{id}(r)$ was integrated following Equation 2, therefore yielding a value for the association constant for the Ca^{2+}–carboxylate interaction.

$$\lim_{P_0 \to 0} K_a = 4\pi \int_0^{R_{cut}} g_{AC}^{id}(r) r^2 dr \tag{2}$$

The upper limit on the integration was taken as the distance corresponding to the separation between the ion and the ligand in the disassociated state, which is reached at the point where the ion environment is that of bulk aqueous solution. We define 'dissociated' as the ion pair separations that exceed those corresponding to the first local maximum beyond the local SIP minimum. The association constants can then be converted into Gibbs free energy values based on equation 3.

$$\Delta G = -RT \ln(K_a) \tag{3}$$

Once obtained, our predicted values for the Gibbs free energy of ion ligand adsorption can be compared against experimentally-derived values taken from the literature.
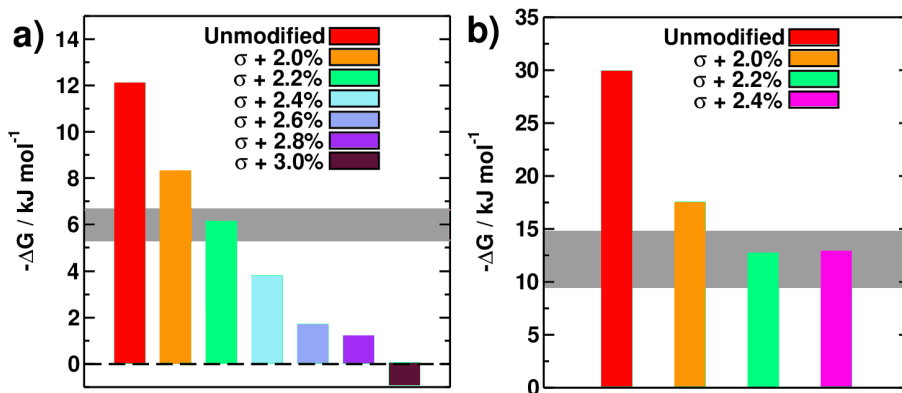
Fig. 4 Predicted free energy of binding for the interaction between Ca^{2+} and the carboxylate side chain of **a)** glutamic acid (Glu), and **b)** carboxyglutamic acid (Gla), for both the unmodified CHARMM22* force-field, and for different force-field modifications (see text for details). The gray-colored horizontal bar indicates the range of experimentally-determined values. [78,79]

While there is a lack of published experimental studies regarding the strength of the interaction between Glu/Gla and Ca^{2+} under aqueous conditions, there are comparable data reported for the aqueous Ca^{2+}-acetate and Ca^{2+}-malonate systems. [78] These molecules are sufficiently similar in structure to the side-chains of Glu and Gla to permit reasonable comparisons between our predictions and experimental data. We therefore tuned our modification of the heteroatomic σ LJ parameter associated with the Ca^{2+}–O_{carb} pair to ensure we could recover the relevant experimental binding free energies.

To investigate the structure of water around the interaction pair, we calculated the radial distribution function between the oxygen atoms of the carboxylate group and the oxygen atom in water, for structures in the biCIP configuration. We calculated the hydrogen-bond angle distribution for the same set of configurations, as well as the coordination number of water molecules around Ca^{2+}. For the hydrogen-bond analysis, our identification of hydrogen bonds was based on the standard geometric definition of Jedlovszky et al. [81] where the interaction must satisfy an O···O distance of less than 3.5Å.

Results and Discussion

The experimentally-observed $\log(K)$ of the binding constants for aqueous calcium acetate in the standard state are reported to be in the range of 0.90–1.12. [78] Here, we reasonably assume that K remains approximately constant in the temperature range of 298–310 K, in order to calculate the free energy of binding at 310 K using Equation 3. This assumption is justifiable considering that $\ln(K)$ varies slowly with varying K. These $\log(K)$ values would thus correspond to a free energy of binding at 310 K in the range of -5.3 to -6.6 kJ mol^{-1}. In the first instance, we generated the PMF profile for Glu at 310 K in the case of the unmodified CHARMM22* FF parameters, as shown in Figure 3a). Three distinct minima were observed in the profile, the first at a calcium–carbon separation of 2.6-3.0Å, the second between 3.2-3.6Å with the third minimum located at a separation of $\sim$ 5Å. Visual inspection of the configurations associated with each of these minima suggests that the first minimum corresponds with the biCIP configuration, the second minimum with the monoCIP

configuration and the third minimum with SIP coordination (see Figure 1). This assignment of structures compares favorably with the previous work reported by Kahlen et al. [29] for the aqueous calcium-acetate system.

The resulting calcium–carboxylate binding free energy was calculated to be -12.1 kJ mol^{-1} at 310 K using the unmodified CHARMM22* FF, approximately twice the experimental value. Following this, we calculated the PMF profiles for this interaction for a total of six modified Ca^{2+}–O_{carb} LJ σ values, increasing this value of σ incrementally in steps of 0.2%. The absolute values of sigma are provided in Table S4 of the ESI†; we also provide all of the final PMF profiles of each metadynamics simulation in Figure S3 of the ESI†. The resulting free energies of binding are summarised in Figure 4a) (see Table S5 in the ESI† for numerical values), revealing that the modification of $\sigma + 2.2\%$ yielded the best comparison with the experimental binding data. The resulting potential of mean force profile is shown in Figure 3a), corresponding to a binding free energy of -6.1 kJ mol^{-1} calculated at 310 K.

In all cases, the dominant minimum corresponded to the biCIP structure. In general, our modification to the σ value tended to reduce the difference in well depth between the biCIP and monoCIP configurations. The free energy difference between the biCIP and monoCIP minima was $\sim$7.5 kJ mol^{-1} with the default CHARMM22* parameters, while for the optimal $\sigma + 2.2\%$ modification this difference reduced to $\sim$4.5 kJ mol^{-1}. As mentioned in the Introduction, our comparison with an integrated value such as the binding free energy does not allow us to reach a definitive conclusion about the relative balance in free energy between the three coordination modes.

As a function of the CV value in the free energy profile, each minimum corresponds with the ensemble average of all configurations that feature the particular C···Ca^{2+} separation corresponding to the position of the minimum. Visual inspection of the configurations associated with each of the three minima suggests that the first minimum is dominated (calculated to be $\sim$100%) by configurations with the biCIP co-ordination mode. Likewise, the second and third minima correspond to the monoCIP and SIP co-ordination modes, respectively (see Figure 1). This assignment
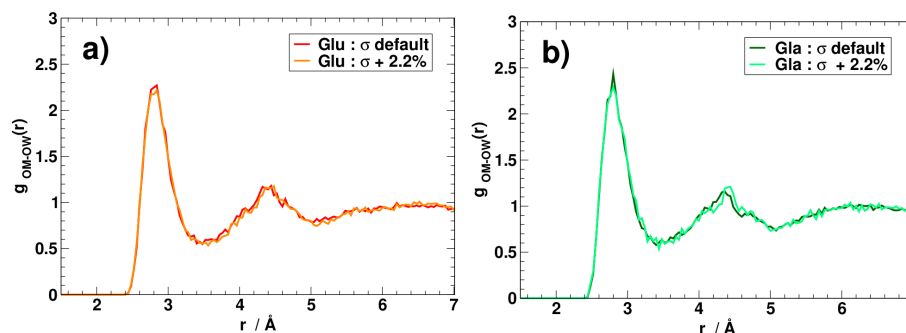
Fig. 5 Radial distribution function between carboxylate oxygen atom and water oxygen atom, calculated for configurations taken from the biCIP state, for **a)** glutamic acid (Glu), and **b)** γ-carboxyglutamic acid (Gla).

of structures compares favorably with the previous work reported by Kahlen *et al.* [29] for the aqueous calcium-acetate system.

Previously published results reported by Project *et al.* [42] utilised regular MD simulations with the CHARMM22 force-field to investigate the aqueous Ca^{2+}–carboxylate interaction. While these authors did not report a binding constant for this particular force-field, their findings hinted that the CHARMM22 force field may actually under-predict this interaction strength. However, the CHARMM22* force-field contains explicit differences in the partial charges of the acidic groups compared with the older CHARMM22 force-field, which may in part explain this discrepancy with our findings. This would seem to be a more likely source of discrepancy compared with the differences in sampling between our work and that of Project *et al.* [42]. Nonetheless, techniques such as metadynamics allow us to improve our degree of sampling compared with regular MD, and in turn allow us to investigate these interactions in more detail.

For calcium malonate, the experimentally-observed range of $\log(K)$ values is reported to be 1.61–2.50 in the standard state.[78,79] Again, we assume that K is approximately constant in the temperature range of 298–310 K, allowing us to infer the binding free energies at 310 K. On this basis, these $\log(K)$ values correspond to free energies in the range of -9.5 to -14.8 kJ mol^{-1} at 310 K. Based on our findings for Glu, we narrowed our investigation of the range of modifications to the Ca^{2+}–O_{carb} LJ σ value to values adjacent to $\sigma + 2.2\%$. In Figure 3b), we show the free energy profile for the association of calcium and the Gla side-chain for the unmodified CHARMM22* FF. As seen for Glu, the free energy profile in all cases supports three distinct minima, where the coordination mode corresponding to each minimum is the same as that described for Glu. There was no evidence to support the existence of a fourth minimum corresponding to a coordination between the two carboxylate groups in Gla. The resulting calcium–carboxylate binding free energy was calculated to be -29.9 kJ mol^{-1} at 310 K for the unmodified CHARMM22* FF, again approximately twice the experimental value.

We considered three modifications, $\sigma + 2.0\%$, $\sigma + 2.2\%$ and $\sigma + 2.4\%$; the resulting free energies of association are summarised in Figure 4b). While the free energy of association calculated from the profiles generated using both the $\sigma + 2.2\%$ and $\sigma + 2.4\%$ modifications fall within the range of experimental values, we suggest that the $\sigma + 2.2\%$ modification provides an acceptable

compromise between performance and transferability. The resulting PMF profile for $\sigma + 2.2\%$ is shown in Figure 3b). The free energy of association calculated using the $\sigma + 2.2\%$ was -12.5 kJ mol^{-1} at 310 K.

We investigated the details of the solvent structuring around the ion pair in the biCIP coordination mode. We calculated the radial distribution function (RDF) between the carboxylate oxygen atom (denoted herein as OM) and the water oxygen atom (denoted herein as OW), and compared our findings for both the unmodified CHARMM22* FF and our $\sigma + 2.2\%$ modification, as shown in Figure 5. Two peaks in the RDF are apparent, the first and most predominant at a OM–OW separation of ~2.8 Å, corresponding to the first solvation shell, and a second peak at ~4.4 Å indicating the second solvation shell. These data show that our modification has imparted no appreciable difference to these RDFs, for both Glu and Gla. These RDFs compare favorably with those previously reported by Kahlen *et al.* [29] for aqueous calcium acetate. In Figure 6 we show the distribution of the hydrogen-bond angle between the carboxylate oxygen atom (OM) and water in the first solvation shell around the carboxylate in the biCIP configuration, calculated for both the unmodified CHARMM22* FF and our $\sigma + 2.2\%$ modification. Again, very little difference is apparent between the unmodified and modified profiles. These profiles broadly agree with those previously published for aqueous calcium acetate,[29] particularly in terms of the positions of the peaks in these profiles. In summary, our data indicate that the proposed FF modification did not adversely impact on the local solvation structure when compared with results from the unmodified CHARMM22* FF.

Finally, we also investigated the coordination number of water around Ca^{2+}. The coordination of waters around Ca^{2+} in the biCIP configuration for both Glu and Gla were found to be unaffected by the FF modification, as indicated in Figure 7a). The number of water molecules in the first solvation shell compared very well with FPMD predictions reported by Kahlen *et al.* [29]. We also show the coordination of waters around Ca^{2+} in the bulk configuration for the aqueous calcium–Glu system (data for Gla are indistinguishable by eye from those of Glu and are not shown). Again, the number of water molecules in the first solvation layer compare favorably with the FPMD values reported previously.[29]

Overall, the unmodified CHARMM22* FF appears to overbind the aqueous calcium–carboxylate ion pair compared with anal-
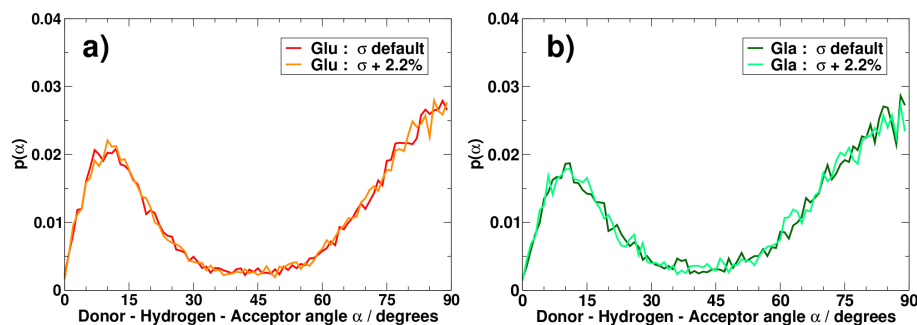
Fig. 6 Distribution function of the hydrogen bond angle between the water hydrogen atom, water oxygen atom and the carboxylate oxygen atom, calculated for configurations taken from the biCIP state, for **a)** glutamic acid (Glu), and **b)** γ-carboxyglutamic acid (Gla).
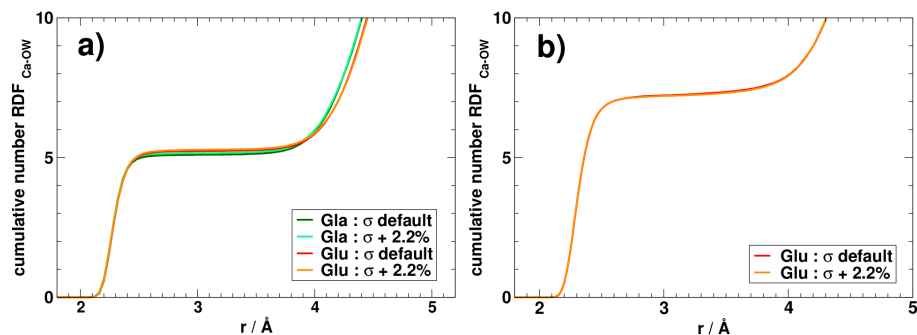
Fig. 7 Number of water molecules around Ca^{2+} calculated for configurations taken from the biCIP state, for **a)** the biCIP state of Glu and Gla, and **b)** the bulk solvation state of Glu.

ogous experimental data. Our findings indicate that a targeted correction to the heteroatomic LJ σ parameter of $+2.2\%$ for the Ca^{2+}–O_{carb} pair, which can be readily implemented in GROMACS, can substantially improve the description of the binding strength of the calcium–carboxylate ion pair in solution, without adversely impacting on the structuring of the solvent around this ion pair.

Some discussion of the CV used in these metadynamics simulations is warranted, because appropriate choice of this parameter is a key determinant in the utility of the metadynamics approach.[82] It should be noted that the calculation of the free-energies of Glu/Gla with Ca^{2+} could be performed using an alternative CV, such as that defined by the distance between Ca^{2+} and oxygen. However, while the choice of a different CV might affect the PMF profiles obtained, any change to the predicted association free energy is likely to be negligible. The free energy differences were calculated from the integration over two broad sets of configurations, the associated and unassociated states. Minor changes to the CV are unlikely to significantly change the ratio of the populations of these two states. To underscore this point, we re-calculated the Glu PMF using a different CV, namely the distance between Ca^{2+} and one of the carboxylate oxygen atoms, using the default CHARMM parameters. As expected, the PMF for this CV appears completely different (see Figure S4 in the ESI[†]). However, upon integration of this PMF, the association constant, and consequently the association free energy, based on this PMF yield very similar results compared with our original simulations. Specifically, ΔG_{ads} was found to be $-12.3\,\mathrm{kJ\,mol^{-1}}$, whereas our original CV yielded a value of $-12.1\,\mathrm{kJ\,mol^{-1}}$. Moreover, by defining the distance between Ca^{2+} and the closest car-

boxylate carbon atom as the CV, we are consistent with previous recent studies.[29,45] Our data also indicate that this CV is appropriate. A poorly-chosen CV can show hysteresis in the resulting free energy profile, due to relatively poor sampling of the energy landscape in directions orthogonal to the CV.[82] In Figure S5 of the ESI[†] we show the evolution of the CV with simulation time; these data do not indicate such problems with our simulations.

The CHARMM22* FF favors the biCIP coordination mode over the monoCIP mode in terms of stability, in both unmodified and modified forms of the FF. While there is some experimental evidence from ^{13}C nuclear magnetic resonance (NMR) observations that the monoCIP mode may be preferred,[83] the FPMD free energy profiles reported by Kahlen *et al.*[29] suggest that the biCIP and monoCIP minima could be roughly equivalent in terms of free energy. However, the large error bars inherent to the free energy profiles generated using FPMD approaches means that this finding is not conclusive. We note from our profiles that our FF correction decreased the energy difference between the biCIP and monoCIP minima; for Glu this was reduced from $\sim 7.5\,\mathrm{kJ\,mol^{-1}}$ with the default CHARMM22* parameters, to $\sim 4.5\,\mathrm{kJ\,mol^{-1}}$ for the optimal $\sigma + 2.2\%$ FF modification. Similarly, for Gla this difference reduced from $\sim 14.5\,\mathrm{kJ\,mol^{-1}}$ in the unmodified case to $\sim 6\,\mathrm{kJ\,mol^{-1}}$ for the modified FF. However, the precise energetic ordering of the biCIP and monoCIP minima remains to be to fully resolved in future.

Conclusions

In summary, via metadynamics simulations, we have shown that the binding free energy between Ca^{2+} and the carboxylate side-

chain of glutamic acid, Glu, and γ-carboxyglutamic acid, Gla, under aqueous conditions is substantially overestimated using the CHARMM22* force-field, compared with analogous experimental data. We have identified and tested a targeted correction to the σ value of the heteroatomic Lennard-Jones interaction between Ca^{2+} and the oxygen of the carboxylate group that can be readily implemented in the GROMACS simulation software package. This modification satisfactorily recovered the binding strength of the aqueous Ca^{2+} and carboxylate ion pair compared with experimental values. Structural properties of the solvent structuring around the ion pair were also found to be unperturbed compared with the unmodified force-field.

Acknowledgments

References

1 D. A. Doyle, J. M. Cabral, R. A. Pfuetzner, A. L. Kuo, J. M. Gulbis, S. L. Cohen, B. T. Chait and R. MacKinnon, *Science*, 1998, **280**, 69–77.

2 Y. Jiang, A. Lee, J. Chen, M. Cadene, B. Chait and R. MacKinnon, *Nature*, 2002, **417**, 515–522.

3 J. J. Kasianowicz, *Chem. Rev.*, 2012, **112**, 6215–6217.

4 C. Maffeo, S. Bhattacharya, J. Yoo, D. Wells and A. Aksimentiev, *Chem. Rev.*, 2012, **112**, 6250–6284.

5 R. W. Teichert, E. C. Jimenez, V. Twede, M. Watkins, M. Hollmann, G. Bulaj and B. M. Olivera, *J. Biol. Chem.*, 2007, **282**, 36905–36913.

6 K. H. Gowd, V. Twede, M. Watkins, K. S. Krishnan, R. W. Teichert, G. Bulaj and B. M. Olivera, *Toxicon*, 2008, **52**, 203–213.

7 K. B. Akondi, M. Muttenthaler, S. Dutertre, Q. Kaas, D. J. Craik, R. J. Lewis and P. F. Alewood, *Chem. Rev.*, 2014, **114**, 5815–5847.

8 Q. Q. Hoang, F. Sicheri, A. J. Howard and D. S. C. Yang, *Nature*, 2003, **425**, 977–980.

9 L. A. Estroff, *Chem. Rev.*, 2008, **108**, 4329–4331.

10 M. B. Dickerson, K. H. Sandhage and R. R. Naik, *Chem. Rev.*, 2008, **108**, 4935–4978.

11 L. A. Capriotti, T. P. Beebe, Jr. and J. P. Schneider, *J. Am. Chem. Soc.*, 2007, **129**, 5281–5287.

12 J. Aizenberg, *MRS Bull.*, 2010, **35**, 323–330.

13 S. Kim and C. B. Park, *Adv. Funct. Mater.*, 2012, **23**, 10–25.

14 S. Collino and J. S. Evans, *Biomacromolecules*, 2008, **9**, 1909–1918.

15 J. W. F. Robertson, J. J. Kasianowicz and S. Banerjee, *Chem. Rev.*, 2012, **112**, 6227–6249.

16 C. B. Ponce and J. S. Evans, *Cryst. Growth Des.*, 2011, **11**, 4690–4696.

17 R. A. Atkinson, J. S. Evans, P. V. Hauschka, B. A. Levine, R. Meats, J. T. Triffitt, A. S. Virdi and R. J. P. Williams, *FEBS J.*, 1995, **232**, 515–521.

18 A. H. Brown, P. M. Rodger, J. S. Evans and T. R. Walsh, *Biomacromolecules*, 2014, **15**, 4467–4479.

19 J. Dzubiella, *J. Am. Chem. Soc.*, 2008, **130**, 14000–14007.

20 M. V. Fedorov, J. M. Goodman and S. Schumm, *J. Am. Chem. Soc.*, 2009, **131**, 10854–10856.

21 C. L. Freeman, J. H. Harding, D. Quigley and P. M. Rodger, *Phys. Chem. Chem. Phys.*, 2012, **14**, 7287.

22 B. Corry and M. Thomas, *J. Am. Chem. Soc.*, 2012, **134**, 1840–1846.

23 C. Boiteux, I. Vorobyov and T. W. Allen, *Proc. Natl. Acad. Sci. U.S.A.*, 2014, **111**, 3454–3459.

24 E. F. Aziz, N. Ottosson, S. Eisebitt, W. Eberhardt, B. Jagoda-Cwiklik, R. Vácha, P. Jungwirth and B. Winter, *J. Phys. Chem. B*, 2008, **112**, 12567–12570.

25 T. Todorova, P. H. Hünenberger and J. Hutter, *J. Chem. Theory Comput.*, 2008, **4**, 779–789.

26 S. Hug, G. K. Hunter, H. Goldberg and M. Karttunen, *Physics Procedia*, 2010, **4**, 51–60.

27 J. Timko, A. De Castro and S. Kuyucak, *J. Chem. Phys.*, 2011, **134**, 204510.

28 C. N. Rowley and B. Roux, *J. Chem. Theory Comput.*, 2012, **8**, 3526–3535.

29 J. Kahlen, L. Salimi, M. Sulpizi, C. Peter and D. Donadio, *J. Phys. Chem. B*, 2014, **118**, 3960–3972.

30 A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586.

31 S. Piana, K. Lindorff-Larsen and D. E. Shaw, *Biophys. J.*, 2011, **100**, L47–L49.

32 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.

33 J. M. Wang, P. Cieplak and P. A. Kollman, *J. Comp. Chem.*, 2000, **21**, 1049–1074.

34 J. Hermans, H. J. C. Berendsen, W. F. Vangunsteren and J. P. M. Postma, *Biopolymers*, 1984, **23**, 1513–1518.

35 J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell, Jr. and R. W. Pastor, *J. Phys. Chem. B*, 2010, **114**, 7830–7843.

36 N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark and W. F. Gunsteren, *Eur. Biophys. J.*, 2011, **40**, 843–856.

37 M. M. Reif, M. Winger and C. Oostenbrink, *J. Chem. Theory Comput.*, 2013, **9**, 1247–1264.

38 H. A. Lorentz, *Ann. Phys.*, 1881, **12**, 127–136.

39 D. Berthelot, *Compt. Rend. Acad. Sci. Paris*, 1898, **126**, 1703–

1855.

40 D. M. Duffy and J. H. Harding, *Surface Science*, 2005, **595**, 151–156.

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

42 E. Project, E. Nachliel and M. Gutman, *J. Comput. Chem.*, 2008, **29**, 1163–1169.

43 R. Di Felice and S. Corni, *J. Phys. Chem. Lett.*, 2011, **2**, 1510–1519.

44 M. Fyta and R. R. Netz, *J. Chem. Phys*, 2012, **136**, year.

45 P. Raiteri, R. Demichelis, J. D. Gale, M. Kellermeier, D. Gebauer, D. Quigley, L. B. Wright and T. R. Walsh, *Farad. Discuss.*, 2012, **159**, 61–85.

46 R. A. Latour, *Colloids and Surfaces B: Biointerfaces*, 2014, **124**, 25–37.

47 B. Hess and N. F. A. van der Vegt, *Proc. Natl. Acad. Sci. U.S.A.*, 2009, **106**, 13296–13300.

48 I. S. Joung and T. E. Cheatham, III, *J. Phys. Chem. B*, 2009, **113**, 13279–13290.

49 A. Savelyev and A. D. MacKerell, Jr., *J. Phys. Chem. B*, 2014, **118**, 6742–6757.

50 A. Savelyev and A. D. MacKerell, Jr., *J. Phys. Chem. B*, 2015, **119**, 4428–4440.

51 T. Dudev and C. Lim, *J. Phys. Chem. B*, 2004, **108**, 4546–4557.

52 S. Gagne, M. Li and B. Sykes, *Biochemistry*, 1997, **36**, 4386–4392.

53 J. J. Falke, S. K. Drake, A. L. Hazard and O. B. Peersen, *Q. Rev. Biol*, 1994, **27**, 219–290.

54 C. Babu, T. Dudev, R. Casareno, J. Cowan and C. Lim, *J. Am. Chem. Soc.*, 2003, **125**, 9318–9328.

55 X. Dong, Q. Wang, T. Wu and H. Pan, *Biophys. J.*, 2007, **93**, 750–759.

56 L. M. Hamm, A. F. Wallace and P. M. Dove, *J. Phys. Chem. B*, 2010, **114**, 10488–10495.

57 J. O'Young, Y. Liao, Y. Xiao, J. Jalkanen, G. Lajoie, M. Karttunen, H. A. Goldberg and G. K. Hunter, *J. Am. Chem. Soc.*, 2011, **133**, 18406–18412.

58 A. L. Boskey and P. G. Robey, *P. G. Robey*, Academic Press, Fourth Edition edn, 2013, ch. The Regulatory Role of Matrix Proteins in Mineralization of Bone, pp. 235–255.

59 M. Goiko, J. Dierolf, J. S. Gleberzon, Y. Liao, B. Grohe, H. A. Goldberg, J. R. de Bruyn and G. K. Hunter, *PLoS ONE*, 2013, **8**, e80344.

60 M. Huang, A. C. Rigby, X. Morelli, M. A. Grant, G. Huang, B. Furie, B. Seaton and B. C. Furie, *Nat. Struct. Mol. Biol.*, 2003, **10**, 751–756.

61 Q. Dai, M. Prorok and F. J. Castellino, *J. Mol. Biol.*, 2004, **336**, 731–744.

62 L. Yang, S. Prasad, E. Di Cera and A. R. Rezaie, *J. Biol. Chem.*, 2004, **279**, 38519–38524.

63 S. E. Cnudde, M. Prorok, Q. Dai, F. J. Castellino and J. H. Geiger, *J. Am. Chem. Soc.*, 2007, **129**, 1586–1593.

64 Y. Z. Ohkubo and E. Tajkhorshid, *Structure*, 2008, **16**, 72–81.

65 G. Kaminski, E. M. Duffy, T. Matsui and W. L. Jorgensen, *J. Phys. Chem.*, 1994, **98**, 13077–13082.

66 A. Barducci, G. Bussi and M. Parrinello, *Phys. Rev. Lett.*, 2008, **100**, 020603.

67 A. Barducci, M. Bonomi and M. Parrinello, *WIREs Comput. Mol. Sci.*, 2011, **1**, 826–843.

68 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys*, 1983, **79**, 926–935.

69 E. Neria, S. Fischer and M. Karplus, *J. Chem. Phys.*, 1996, **105**, 1902–1921.

70 B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *Journal of Chemical Theory and Computation*, 2008, **4**, 435–447.

71 R. W. Hockney, S. P. Goel and J. Eastwood, *J. Comp. Phys.*, 1974, **14**, 148.

72 T. A. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089.

73 S. Nosé, *Mol. Phys.*, 1984, **52**, 255.

74 W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695.

75 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.

76 G. B. A. Barducci and M. Parrinello, *Phys. Rev. Lett.*, 2008, **100**, 020603.

77 M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia and M. Parrinello, *Comput. Phys. Commun.*, 2009, **180**, 1961–1972.

78 P. G. Daniele, C. Foti, A. Gianguzza, E. Prenesti and S. Sammartano, *Coord. Chem. Rev.*, 2008, **252**, 1093–1107.

79 M. Prorok and F. Castellino, *J. Biol. Chem.*, 1998, **273**, 19573–19578.

80 A. A. Chialvo, P. T. Cummings, H. D. Cochran, J. M. Simonson and R. E. Mesmer, *J. Chem. Phys*, 1995, **103**, 9379–9387.

81 P. Jedlovszky, J. P. Brodholt, F. Bruni, M. A. Ricci, E. U. Soper and R. Vallauri, *J. Chem. Phys.*, 1998, **108**, 8528–8540.

82 A. Barducci, M. Bonomi and M. Parrinello, *WIREs Comput. Mol. Sci.*, 2011, **1**, 826–843.

83 A. Kondoh and T. Oi, *Z. Naturforsch. A Phys. Sci.*, 1998, **53**, 77–91.

RSC Advances Accepted Manuscript