Citation
Schuitemaker, A. and Aufort, J. and Koziara, K.B. and Demichelis, R. and Raiteri, P. and Gale, J.D. 2021. Simulating the binding of key organic functional groups to aqueous calcium carbonate species. Physical Chemistry Chemical Physics. 23 (48): pp. 27253-27265. http://doi.org/10.1039/d1cp04226b

Simulating the binding of key organic functional groups to aqueous calcium carbonate species

Alicia Schuitemaker, Julie Aufort, Katarzyna B. Koziara, Raffaella Demichelis, Paolo Raiteri, and Julian D Gale*

Curtin Institute for Computation/The Institute for Geoscience Research (TIGeR), School of Molecular and Life Sciences, Curtin University, PO Box U1987, Perth, WA 6845, Australia

E-mail: J.Gale@curtin.edu.au

Abstract

The interaction of organic molecules with mineral systems is relevant to a wide variety of scientific problems both in the environment and minerals processing. In this study, the coordination of small organics that contain the two most relevant functional groups for biomineralisation of calcium carbonate, namely carboxylate and ammonium, with the corresponding mineral ions are examined in aqueous solution. Specifically, two force fields have been examined based on rigid-ion or polarisable models, with the latter being within the AMOEBA formalism. Here the parameters for the rigid-ion model are determined to target the accurate reproduction of the hydration structure and solvation thermodynamics, while both force fields are designed to be compatible with the corresponding recently published models for aqueous calcium carbonate. The application of these force fields to ion pairing in aqueous solution is studied in order to quantitatively determine the extent of association.

Introduction

Organic-mineral interactions play a key role in a wide range of situations from industrial challenges, such as flotation and minerals processing, through to understanding of environmental geochemistry which can be influenced by naturally-occurring molecules.[1] Organics

are also often used as crystal growth modifiers to manipulate the rate and morphological outcome for mineral crystallisation, including suppressing the formation of undesirable scale. Arguably one of the most complex examples of the influence of organics on mineral formation is that of biomineralisation in which molecules, such as proteins, direct the nucleation, growth, morphology and structure of mineral phases.[2] This can involve regulation of polymorphism, particle size, and how crystallites assemble into a more complex structure, including formation of organic-mineral composites. To try to understand the full complexity of biomineralisation and how nature achieves such exquisite control is a long-standing challenge that involves many facets, from the potential to regulate processes via confinement to direct incorporation.[3] Even the detailed atomic-level knowledge of the interactions between key organic functional groups and components of the mineral-containing aqueous solution is arguably incomplete as a starting point. The objective of this work is therefore to probe such interactions for one of the most common systems in biomineralisation, namely that of aqueous calcium carbonate.

Experimental work has demonstrated that different organic species affect the nucleation and growth of calcium carbonate in numerous ways and at various stages.[4–7] Gebauer *et al.*[4] have classified the influence of organic molecules into nine different categories, showing the wide variety of ways in which organic molecules can change the crystallisation of calcium carbonate. For example, organic molecules can alter which $CaCO_3$ polymorph is formed by stabilising one type of pre-nucleation cluster over another or can change the shape of the crystals by face-specific adsorption.[4]

Although the influence of organic molecules on calcium carbonate has been extensively studied using experimental methods,[8] atomistic knowledge of the interactions of organic species with $CaCO_3$ remains incomplete. Computer simulations offer a way to explore the organic-calcium carbonate interactions to gain atomic level understanding of some of the processes involved in biomineralisation. However, the computational modelling of calcium carbonate and its interactions with organic matter is a complex problem that has a lot of

promise, but also many pitfalls.[9,10] Potentially more accurate, high-level *ab initio* quantum mechanical approaches can only be applied to a small number of atoms for short time scales and are associated with large computational cost, which is why classical molecular dynamics approaches based on force fields are more frequently applied to study the interactions between organic molecules and calcium carbonate.[11]

To date, simulations of organic-mineral interactions have largely been performed with general purpose force fields, or *ad hoc* mixtures of models, that have not been specifically developed to consider mineral-organic interactions.[10] Indeed the interaction parameters are often determined via combination rules due to the paucity of data to fit against.[12] Yet, it is important to consider properties, such as pairing free energies between the organic species and the constituent ions of the mineral phase, for a model to be reliable. To simplify this process, a common approach is to consider the binding of small organics containing the key functional groups of biomolecules with aqueous calcium and carbonate ions, rather than directly tackling association with multiple binding sites simultaneously. For example, the acetate anion is widely used as a proxy for the interaction of calcium with carboxylate groups.

Raiteri *et al.*[13] showed that combining parameters from the CHARMM force field for the organic species with a thermodynamically-optimised force field for calcium carbonate species in water could be effective for carboxylate groups, if the repulsive part of the van der Waals interactions was derived from those between calcium and oxygen of carbonate by scaling based on the ratio of the oxygen charges. This led to pairing free energies for calcium to acetate and calcium to citrate that were slightly too negative, but within thermal energy of some experimental values. Kahlen *et al.*[14] also examined the ion pairing free energy for calcium-acetate as a prototypical example for a carboxylate group. Their study benchmarked popular organic force field variants (GROMOS and OPLS-AA) against a potential of mean force curve obtained from *ab initio* molecular dynamics at the PBE-D3 level of theory. It was found that the variation in the pairing free energies was almost 50 kJ/mol depending on the parameter set, but that the thermodynamics and structures could be improved by

a small adjustment to the Lennard-Jones sigma value. OPLS-AA with a 5.8% increase in the Ca-O(acetate) sigma was proposed as the best combination. A similar modification of Lennard-Jones parameters by increasing sigma by 2.2% was proposed by Church *et al.*[15] for the CHARMM22* force field to improve Ca-carboxylate binding and to correct for systematic over-binding by the default parameters.

A more wide-ranging study of the simulation of organic functional groups interacting with aqueous calcium carbonate solutions was conducted by Saharay and Kirkpatrick.[16] Here they considered the binding with guanidinium, acetic acid, acetate and ethanol as models for the functional groups -NH_2^+, -CO_2H, -CO_2^- and -OH, respectively. Their study was based on the default parameters from CHARMM with TIP3P water for all species. As a result, the hydration free energies of the calcium and carbonate anions were found to be consistently underestimated in an absolute sense and have errors of between 60 and 120 kJ/mol. While the binding free energies were not quantified for individual ions, the potentials of mean force were determined for binding to the $CaCO_3$ ion pair in water. In the case of acetate, the binding free energy was found to be substantially more exergonic than the experimental value for calcium-acetate ion pairing. Although there is no alternative value to compare against, this would appear to be implausible on electrostatic grounds and is almost certainly a consequence of the weaker than expected hydration of the ions.

Recently there has been a detailed study of metal-acetate ion pairing for calcium, magnesium and zinc cations based on both Raman spectroscopy and computer simulation by De Oliveira *et al.*[17] Force field parameters for a rigid-ion model were taken from a number of sources including Amber99 for acetate, SPC/E for water. The use of scaled charges by a factor of 0.8 was also examined, as were the results of using the AMOEBA model for the cases of Ca and Mg. In addition, *ab initio* molecular dynamics for ion pairing with formate was used to provide data for comparison. This study concludes that formally-charged rigid-ion force fields cannot be trusted for ion pairing, while even AMOEBA overestimates the binding free energies by approximately 8 kJ/mol. Overall, the best results were argued to

come from the use of scaled charges. Aside from the thermodynamics, this study concludes that the contact ion pair is the dominant mode of binding with monodentate coordination of carboxylate groups to the metal cations being generally preferred.

Beyond the binding of calcium carbonate species in aqueous solution with organics, it is important to note that there have been a growing number of studies that quote binding thermodynamics for proteins and other biomolecules adsorbed on the surfaces of calcium carbonate, typically in the form of calcite where the basal (104) surface dominates. Many of these values are based on ensemble averages of internal energy, which is of questionable value when entropy is known to be the major driver of association for calcium carbonate binding in water[18] and therefore likely to also make a significant contribution for the binding of other species to these ions in the presence of solvent. Even when free energies of binding are computed, the high dimensionality of the landscape often leads to incomplete sampling and therefore substantial errors in the thermodynamics. For example, if a biomolecule were to undergo a change in conformation at the surface to increase the number of interaction sites then this is unlikely to occur spontaneously on the molecular dynamics timescale, especially when coupled to loss of adsorbed surface water.[19] Given these factors, and the general lack of experimental data against which to compare, the use of such data for benchmarking and calibration is likely to be less fruitful than use of association constants for ion pairs in solution.

This work introduces new force field parameters for small organic molecules interacting with aqueous calcium carbonate species as proxies for the interaction of different organic functional groups. These parameters have been obtained through refinement against several structural and thermodynamic properties of the molecules in an aqueous environment and their interaction with the constituent ions of calcium carbonate. Two different models have been examined - one simpler rigid-ion parameterisation and another more complex set of parameters to include polarisability via the AMOEBA model. Here we examine the performance of both models to capture the thermodynamics of two key functional groups

when associating with aqueous calcium carbonate and the extent to which parameters for the isolated functional groups are additive when combined in a single molecule.

Methodology

Here the objective is to obtain and benchmark force field parameters for two of the most common organic functional groups in biomolecules at close to neutral pH, namely carboxylate ($-CO_2^-$) and amino ($-NH_3^+$), and their interactions with aqueous calcium carbonate systems. Specifically, we consider three different molecules as proxies for these interactions; the first two are the acetate anion and methylammonium cation in order to capture the isolated functional groups. In addition, we also consider the glycine zwitterion to examine the case where both groups are present simultaneously in the same molecule. In order to assess the influence of differences in parameterisation, as well as the incorporation of polarisability in the force field, we examine both rigid-ion and AMOEBA-based models for all three systems. Below we describe how the parameters were obtained and the methods used as part of the benchmarking process. The full list of final force field parameters is given in the Supporting Information (SI).

Force Field Parameterisation

In this work the rigid-ion force field parameters for acetate, methylammonium and glycine zwitterion were generated specifically to be used in conjunction with the model of Raiteri *et al.*[20] for calcium carbonate and the SPC/Fw water model of Wu *et al.*[21] This combination of mineral and water force field has been shown to accurately reproduce the structural and thermodynamic properties of both the solid phases and the constituent ions in solution, at least within the limits of a non-polarisable model. This is in contrast to other parameterisations that have focused on the mechanical properties of the solid[22] at the expense of obtaining the correct solubility. If the objective of the model is to be capable of simulating aqueous inter-

faces and crystallisation processes, then we would argue that the thermodynamic properties of hydration are more pertinent than the elastic response of the solid phase.

The initial GROMOS[23] parameters for the small organic species were obtained from Automated Topology Builder and Repository.[24] The van der Waals parameters for the interaction between the organics and water were iteratively refitted against a range of information relating to the hydration properties, including hydration free energies and data regarding the spatial distribution of water around the solutes. The interactions between carboxylate groups and calcium ions were initialised based on the parameters between this cation and carbonate, but with the Buckingham A coefficient scaled in proportion to the oxygen charges. Carbonate interactions with carboxylate groups were similarly adapted from previous parameters between carbonate and bicarbonate, prior to fitting where data is available.

In addition to fitting a force field for the rigid-ion model, the same data was also used to refine a set of parameters for AMOEBA starting from those derived for organics in proteins[25] combined with a set recently developed for aqueous calcium carbonate systems.[26] The need to revisit the parameterisation of AMOEBA for hydration free energies has been previously highlighted.[27] Charges, dipoles and quadrupoles for all atoms within acetate and methylammonium were taken from the recent AMOEBA+ parameter set.[28] In order to follow the principles of transferability of functional groups, the charges for the carboxylate and ammonium groups were taken from acetate and methylammonium, respectively, and applied to glycine. The multipole moments for the remaining CH_2 in glycine were then set based a separate MP2/aug-cc-pVTZ//MP2/6-31G* quantum mechanical calculation for glycine, using the GDMA software[29] to perform the distributed multipole analysis. The carbon charge for CH_2 was then adjusted slightly to ensure overall charge neutrality. It should be noted that the quantum mechanical data for the glycine zwitterion was obtained using a PCM environment to mimic the presence of water, given that this species is unstable in vacuum. It was found that removing the polarisation groups within the AMOEBA+ parameterisation led to improved results for the present systems and so was adopted here. Most intramolecular

parameters for AMOEBA were taken from the existing 2013 set without modification. For methylammonium, the intramolecular terms were taken from a combination of existing parameters for an -NH_3^+ group, augmented with terms from methyl amine where not available, with the H-C-N-H torsions being modelled on those from ethane. Adjustments were made to the van der Waals parameters for oxygen of carboxylate in order to improve the description of the hydration properties of this functional group, while for the case of methylammonium the corresponding terms for the -NH_3^+ group were also refitted. For the calcium interaction with the oxygens of the carboxylate group, a Buckingham potential was also added, as per this cation interacting with bicarbonate from our previous work.[26] This small extra repulsion partially corrects a slight overbinding observed when comparing the raw AMOEBA results against a gas phase calculation of the interaction of $Ca(H_2O)_5^{2+}$ with the acetate anion. Quantum mechanical reference data was computed at the ωB97X-D3/ma-def2-QZVPP level of theory.

In the subsequent sections, the details of how the reference data for fitting was generated are presented, as well as the methods used to run the simulations to benchmark the results of the force field parameters during the refinement process.

Hydration structure

The first property to be considered during the force field refinement was the local structure of hydration shells when the species are immersed in water. Here a particular focus was on ensuring a reasonable hydrogen bonding structure with the polar functional groups as the dominant interaction sites. This structural information can be characterised in a number of ways. Here we focus on the radial distribution functions (RDFs) of oxygen and hydrogen of water as computed relative to key atoms within the molecules. This was then augmented by the use of 3D isosurfaces in order to capture the angular distribution that is lost when collapsed onto a radial projection. One of the challenges here is to find suitable reference data against which to calibrate force field results, since such data is rarely available from

experiment. To fill this gap we therefore used information from *ab initio* molecular dynamics (AIMD), based on density functional theory (DFT), though this comes with the caveat that at the level of functionals that are computationally tractable for routine use in molecular dynamics this will result in systematic errors. While use of most standard generalized gradient approximations (GGAs) is found to lead to an over-structuring of liquid water, the situation can be improved through the inclusion of dispersion corrections and/or increasing the simulation temperature. In this work we utilize the BLYP functional in combination with the D3 corrections of Grimme and co-workers,[30] as well as performing simulations at 330 K. This combination has been shown by others to yield reasonable RDFs for water in comparison to experiment.[31]

Ab initio molecular dynamics in this study was performed using the Gaussian-augmented planewave approach[32] as implemented in the program CP2K.[33] Here all atoms were represented using GTH pseudopotentials[34] with triple-zeta double polarised (TZV2P) basis sets for the valence electrons. The electron density was expanded in a planewave basis with a cutoff of 400 Ry and a charge-neutralising background was included in the case of acetate and methylammonium. Self-consistency was achieved using the orbital transformation algorithm with full kinetic preconditioning.[35] Calculations were performed with only gamma point sampling of the Brillouin zone as all systems are wide gap insulators with cell dimensions of approximately 14 Å. The simulation cells for acetate, methylammonium and glycine contained 268, 275, and 304 atoms, respectively. Starting configurations were generated through equilibration at constant pressure under cubic constraints based on initial force field parameters. The systems were then run in the NVT ensemble with a timestep of 0.5 fs using a CSVR thermostat with a fast relaxation time of 10 fs as the objective was only to obtain the equilibrated structure rather than the dynamics of the water. For the two charged species, where interactions with water are stronger, the simulations were run for longer than 50 ps, while in the case of the more weakly solvated neutral glycine a shorter run of $\sim$37 ps was used.

The radial distribution functions (RDFs) for the two force field models were obtained from unbiased molecular dynamics simulations (MD). The MD simulations conducted with the polarisable AMOEBA force field were largely performed with the OpenMM code[36] and run on GPUs using mixed precision,[36,37] except for the calculation of the solvation free energy (see later) which used Tinker HP,[38] while the LAMMPS package[39] was employed for the MD simulations with the rigid-ion model. The simulation boxes were approximately 50x50x50 Å^3 in size and contained one of the small organic molecules solvated by ∼4,200 water molecules. A time step of 1 fs was employed and the atomic trajectories were written every 0.1 ps. For both models, the simulations were run at ambient conditions (NPT ensemble, 300 K, 1 atm). The temperature for the rigid-ion force field was controlled by a chain of five Nosé-Hoover thermostats with a relaxation time of 0.1 ps, while for the polarisable model a Langevin thermostat with a relaxation time of 1 ps was employed.

Hydration Free Energy

Having ensured that any set of force field parameters gives a reasonable distribution of water in at least the first hydration shell, the second property used to constrain the organic-water interaction parameters was the hydration free energy. Ensuring reasonable thermodynamics for hydration is essential for the equilibrium between different species, both in aqueous solution and when adsorbing at mineral surfaces, where the organics may act as growth modifiers. A range of experimental and theoretical hydration free energies can be found in the literature for acetate and methylammonium, though values for glycine are more scarce. For consistency across all solutes we have also computed the hydration free energies using three parameterised continuum models, namely SM8[40] at the M06/6-31+G** level of theory, as implemented in the QChem software,[41] the PCM model at the HF/6-31G* level in Gaussian,[42] and the conductor-like polarizable continuum model (CPCM) at the ωB97X-D3/ma-def2-QZVPP level of theory in ORCA.[43] For the SM8 and PCM calculations, the

hydration free energy was determined as the contribution of the model at the optimised geometry in solution, whereas for CPCM the values are the difference between the energies when optimised in solution and gas phase. The exception to this is for glycine, where the species is zwitterionic in solution, but not in the gas phase. Since the force fields used here cannot account for the intramolecular proton transfer, the hydration contribution is taken as the difference between the solution and gas phase energies at the optimised geometry of the former case.

For the force field models, the hydration free energies for acetate, glycine and methylammonium were determined using free energy perturbation[44,45] to immerse the solutes in explicit water. The hydration free energies reported in this work correspond to the average values for the creation and annihilation of the solute molecules. In all cases, Madelung,[46] finite size[47] and standard state corrections[48] were applied as appropriate. The Bennett acceptance ratio (BAR) method,[49] as implemented in TinkerHP,[38] was used for the AMOEBA model and the solute was introduced to/removed from a 25x25x25 Å^3 box of water. This cubic box was first equilibrated isotropically at 300 K and 1 atm using the Andersen thermostat and Berendsen barostat, prior to carrying out the free energy perturbation at constant volume with perturbation of the electrostatic and van der Waals interactions performed separately, with eleven simulations each. Each simulation was run for at least 2.5 ns with a time step of 1 fs. The perturbation parameter, λ, was set to [0.0, 0.2, 0.4, 0.5, 0.6, 0.65, 0.7, 0.75, 0.8, 0.9, 1.0] for the van der Waals interactions and to [0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0] for the electrostatic interactions. A cutoff of 10 Å was used for the van der Waals interactions and the real space part of the electrostatics within an Ewald summation. The criterion for the convergence of the self-consistent induced dipoles was set to 10^{-5}. The reduced computational cost of using the rigid-ion model allowed for the removal of the Coulomb and van der Waals interactions using more steps, namely 20 stages for each, with 5 ns of simulation for each stage (after 500 ps of equilibration). These simulations were performed at ambient conditions (300 K) in the NVT ensemble with a time step of 1

fs. In addition, for both models a vacuum calculation was performed where the Coulomb interactions were reduced as before. This additional gas-phase simulations was necessary to account for the partial screening of the 1-4 interactions, which results in a non-zero internal energy. The temperature and pressure were maintained using a chain of five Nosé-Hoover thermostats and barostats with relaxation times of 0.1 and 1.0 ps, respectively. In addition, the MTK correction term[50] was applied to the equations of motion to account for the fact that volume probability distribution of the original Hoover barostats differs from the true NPT distribution. A cubic simulation box with a side of approximately 50 Å contained one organic solute and approximately 4,200 water molecules for the acetate anion. For the glycine zwitterion and methylammonium, a smaller cubic box ($\sim$ 25 Å in side) with one organic and 515 water molecules was used to reduce the computational cost of the simulations.

Ion Pairing Free Energy

Once the force field parameters had been fitted against the combination of the water structure and hydration free energies, then the models were used to predict the dilute speciation of the organics in aqueous calcium carbonate solutions via considering the stability of ion pairs. The ion pairing free energies were calculated with the PLUMED plug-in[51,52] by using well-tempered,[53] multiple walkers[54] metadynamics.[55,56] For all simulations, a 50x50x50 Å^3 box containing the ion pair and approximately 4,200 water molecules was used. For the AMOEBA force field, the distance between the two ions was biased with a bias factor of 5 and Gaussians were laid every 1 ps with an initial height of k_bT and width of 0.1 Å. The distance between the ions was taken as the separation between the calcium or carbon of the carbonate to the carbon or nitrogen in the functional group of interest. For the rigid-ion simulations, the same bias factor, Gaussian deposition rate and initial Gaussian height was employed. However, for ion pairing free energies with calcium, an additional collective variable was included for the water coordination number of the calcium ion. The width of the Gaussians for the distance and coordination number collective variables were 0.2 Å and

0.1, respectively. The water coordination number is described by Equation 1.

$$c(r) = \sum \frac{1 - \left(\frac{d-d_0}{r_0}\right)^n}{1 - \left(\frac{d-d_0}{r_0}\right)^m} \tag{1}$$

The parameters of the switching function (d_0, r_0, n and m) were determined based on fitting the function to a radial distribution function (Ca-O_w) obtained from unbiased MD. The position and the width of the peak are defined by d_0 and r_0, respectively, whereas n and m control the curvature and are selected so that the function contains all of the first hydration shell given by the first peak in the RDF. The parameters employed during the ion pairing calculations for the small organic species with calcium are d_0 = 2.1 Å, r_0 = 1.0 Å, $n = 4$ and $m = 10$. To limit the collective variable space explored an upper wall was applied to the distance for both models. For the rigid-ion force field, this wall was at a distance of 15 Å using a spring constant of ∼1 eV/Å^2, while for the AMOEBA model, the upper wall was at 14 Å with a spring constant of 10,000 kJ/mol nm^2 (note the spring constants are given in the native units appropriate to the code used). To ensure well-converged sampling was achieved, an aggregate simulation time of at least 200 ns was obtained in all cases. For the rigid-ion model, this was achieved by running the metadynamics simulation using 30 independent walkers, whereas for the AMOEBA model 16 independent walkers were employed.

As in previous works,[20,26,57] after obtaining the free energy profiles from PLUMED, $G(r)$, the ion pairing free energy was obtained from the ions' association constant,

$$K_a = \exp(-\Delta G_a / k_B T) \tag{2}$$

by integrating the potential of mean force,

$$\phi(r) = G(r) + k_B T \ln 4\pi r^2 \tag{3}$$

between the minimum distance at which the ions can be found, R_0, and the limit of the

bound state, R_1,

$$K_a = -4\pi C_0 \int_{R_0}^{R_1} \mathrm{d}r \ \exp(-\phi(r)/k_B T) r^2, \tag{4}$$

after alignment of the long-range tail with the analytical solution for two point charges in a dielectric medium,

$$G(r) = \frac{1}{4\pi\varepsilon_0\varepsilon_r}\frac{q_i q_j}{r} - k_B T \ \ln 4\pi r^2 \tag{5}$$

where q_i and q_j are the charges of the two species, r is their separation and ε_r is the static dielectric constant of the water model used, and C_0=0.0006022Å^{-3} is the standard concentration (1 mol/L) in appropriate units. The value of the dielectric constants for the AMOEBA and SPC/Fw water models were taken from Laury *et al.*[58] and Wu *et al.*,[21] respectively. In the case of interactions with the glycine zwitterion, where there is only a charge-dipole interaction for the electrostatic component of ion pairing, the long-range free energy was aligned based on the analytic entropic solution alone. The alignment between the curves was performed by minimising the difference between the computed free energy and the analytic solutions between 11 and 13 Å. One of the ambiguities in performing the integration of the free energy lies in the selection of the cut-off for the bound state. Because there is no well-defined minimum for the solvent-separated ion pair, the ion pairing free energies were calculated by using only two different cut-offs; specifically the free energy maxima marking the upper bounds for the contact ion pair (CIP) and the solvent-shared ion pair (SSHIP).

Results and Discussion

Solvation Structure and Energetics

The interactions of biomolecules with calcium carbonate minerals occur in an aqueous environment, therefore it is important for an accurate interatomic potential to model solvated organic molecules. This includes a good description of the structure, dynamics and energetics of water associated with each species. The water structure, in the form of radial distribution

functions and 3D atomic density maps, obtained for both the rigid-ion and AMOEBA models are compared to data from *ab initio* MD in Figures 1 and 2. Key quantities that characterise the water structure and dynamics of the different organic species are summarised in Table 1.

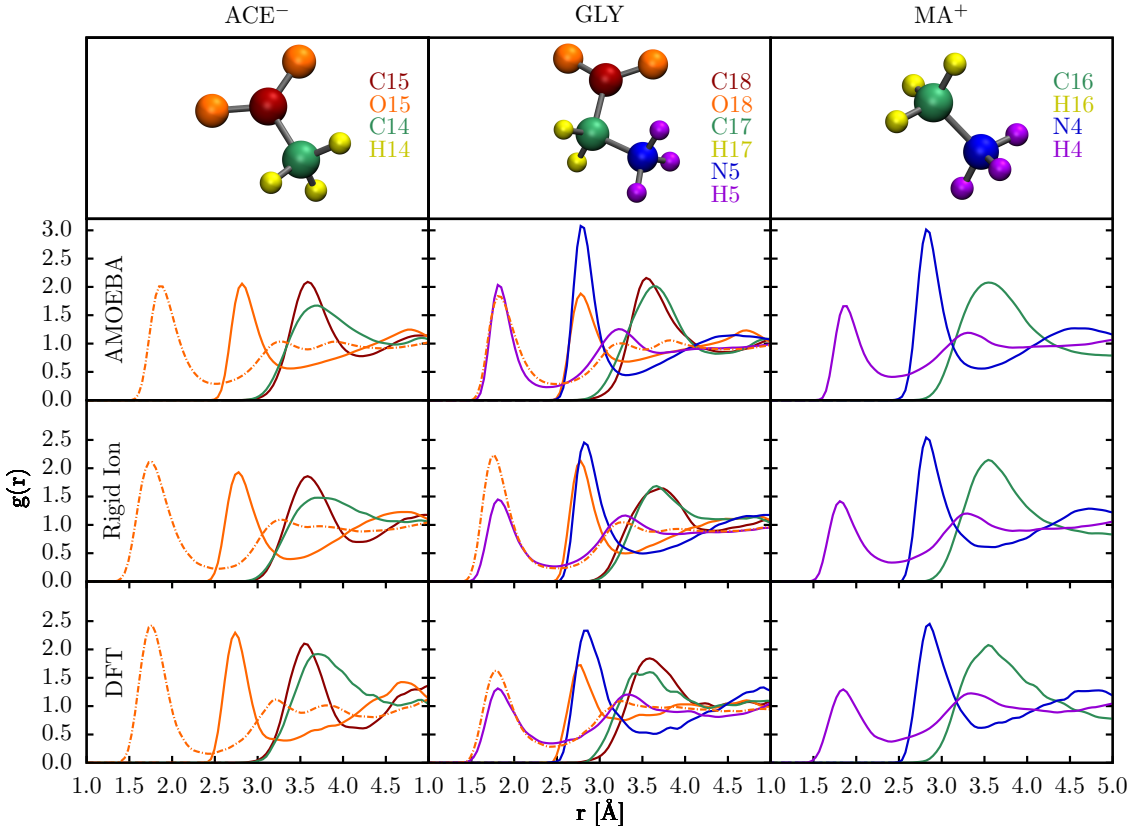

Figure 1: Radial distribution functions for acetate (ACE-), the glycine zwitterion (GLY) and methylammonium (MA+) in water obtained from molecular dynamics simulation. The solid and dashed lines are for O_w and H_w, respectively, with curves colour-coded to reflect the atom of the molecule whose radial distribution function is being shown, as per the scheme given. For clarity, curves are only shown for those atoms where there are significant peaks observed. Corresponding integrals of the RDF curves are provided in Supplementary Information.

The water structure of the two empirical potentials overall compares favourably with the first principles data. The largest scatter occurs for the water coordination number of the methyl group within methylammonium. In comparison to the corresponding functional group in acetate, all approaches exhibit a higher water coordination number, though the extent of this shift is largest for the AIMD. That said, this weak interaction is likely to be

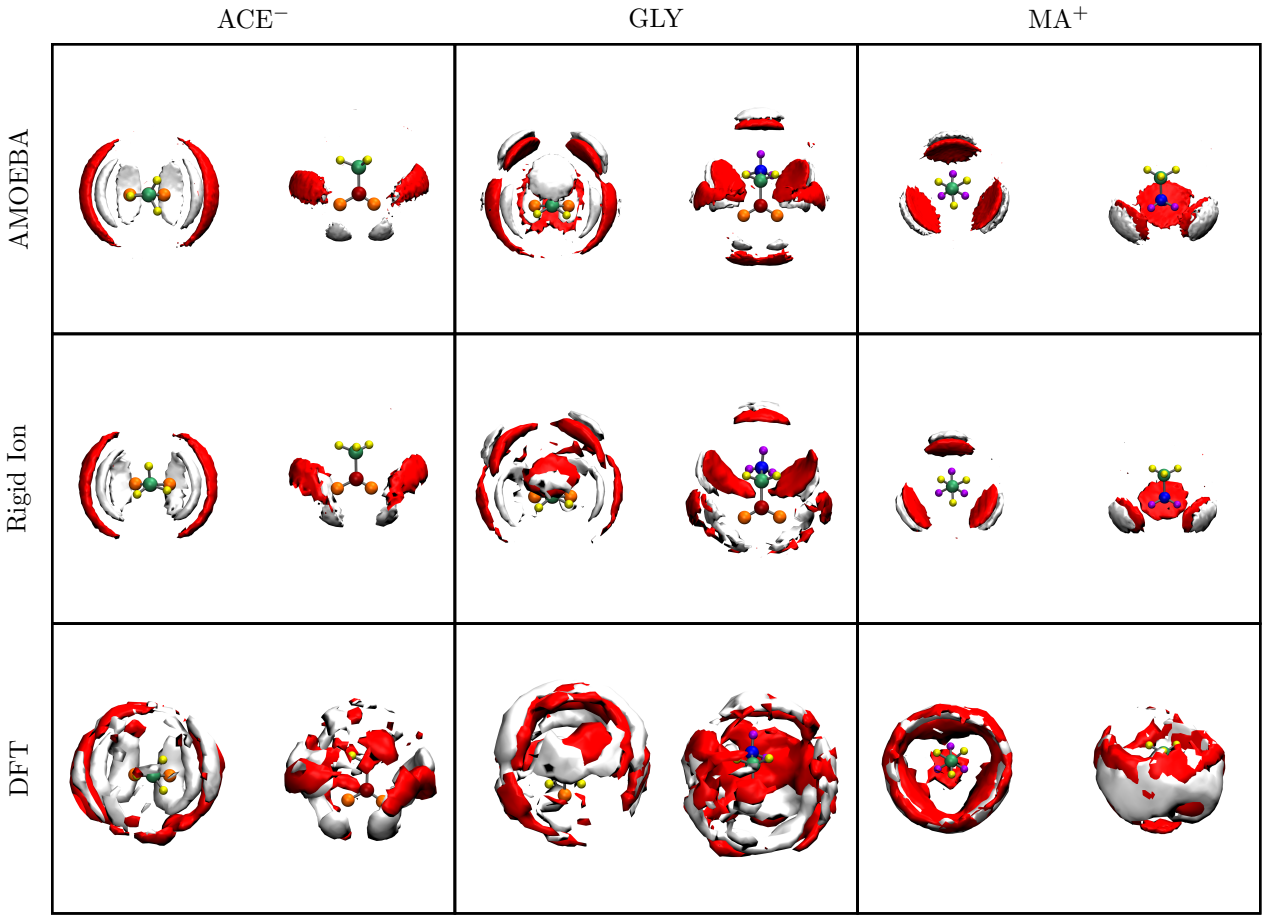

Figure 2: 3D iso-surfaces of the water structure around acetate, glycine and methylammonium obtained from molecular dynamics simulation, including a top and side view for each species. The carbon, nitrogen, oxygen and hydrogen atoms of the organic are depicted in cyan, blue, red and yellow, respectively. The atomic densities for organic molecules are coloured as in Figure 1, while the O_w and H_w atomic densities are shown in red and white, respectively. Iso-surfaces are computed at 0.108 and 0.216 atoms/Å^3 for O and H, respectively.

dominated by entropic effects and thus the dual limitations of the AIMD in regard to box size and timescale will be most apparent for this peak. In addition, only minor differences are observed between the two models, e.g. small shifts in the first solvation peak (Δ ¡ 0.2 Å), the number of coordinating water molecules (Δ ¡ 2 water molecules) and the water residence times (Δ ¡ 0.03 ns) all of which are low indicating that the coordinating waters can be rapidly equilibrated and exchanged on the timescale of a molecular dynamics simulation. Overall,

Table 1: Water structure and dynamics data for the rigid-ion (RI), polarisable AMOEBA model (A) and *ab initio* MD (AIMD) including distances based on the position of the first peak within the radial distribution function (r [Å]), average water coordination numbers (n) and average residence times (τ [ps]) from unbiased MD at 300 K. Note that the water residence times are only given for the force field models due to the limited statistics available for AIMD. For acetate, C14 and C15 represent the carbons of the methyl and carboxylate groups, respectively. For glycine, C17 and C18 represent the carbons of the CH_2 and carboxylate group, again respectively. Atom labels correspond to the force field types as given in the Supplementary Information.

		RI	A	AIMD
$CH_3CO_2^-$	$r_{\text{C15-Ow}}$	3.6	3.6	3.6
	$r_{\text{O15-Ow}}$	2.7	2.8	2.7
	$r_{\text{C14-Ow}}$	3.7	3.7	3.7
	$r_{\text{O15-Hw}}$	1.8	1.9	1.8
	$n_{\text{C15-Ow}}$	7.4	7.7	6.7
	$n_{\text{C14-Ow}}$	13.1	11.7	12.3
	$\tau_{\text{C15-Ow}}$	9	10	-
	$\tau_{\text{C14-Ow}}$	7	5	-
$^+NH_3CH_2CO_2^-$	$r_{\text{C18-Ow}}$	3.7	3.6	3.6
	$r_{\text{O18-Ow}}$	2.7	2.8	2.7
	$r_{\text{C17-Ow}}$	3.6	3.7	3.6
	$r_{\text{N5-Ow}}$	2.8	2.8	2.8
	$r_{\text{H5-Ow}}$	1.8	1.8	1.8
	$r_{\text{O18-Hw}}$	1.8	1.8	1.8
	$n_{\text{C18-Ow}}$	9.2	10.7	9.9
	$n_{\text{N5-Ow}}$	4.0	3.7	4.8
	$\tau_{\text{C18-Ow}}$	10	14	-
	$\tau_{\text{N5-Ow}}$	7	14	-
$CH_3NH_3^+$	$r_{\text{C16-Ow}}$	3.6	3.6	3.6
	$r_{\text{N4-Ow}}$	2.8	2.8	2.8
	$r_{\text{C16-Ow}}$	3.6	3.6	3.6
	$r_{\text{H4-Ow}}$	1.8	1.9	1.8
	$n_{\text{N4-Ow}}$	4.7	4.7	4.6
	$n_{\text{C16-Ow}}$	16.7	16.8	18.0
	$\tau_{\text{C16-Ow}}$	7	8	-
	$\tau_{\text{N4-Ow}}$	5	6	-

for the water structure around the organic molecules, both models perform well and the inclusion of polarisability does not substantially affect the results.

The next step is the consideration of the thermodynamics of both models in water, specifically via the hydration free energies for all three species. The hydration free energies for both models are reported in Table 2, in which they are compared to quantum mechanical

results using a range of continuum solvation models, as computed in this work, as well as other values reported in the literature from both experimental and theoretical approaches. Clearly there is substantial scatter in the theoretical values, both here and elsewhere, which often reflect the approximations and parameters associated within dielectric continuum based models. Although some models are parameterised to try to give good hydration free energies based on the continuum description alone, it has long been recognised that including an explicit first (or even second) hydration shell can improve results. Based on CPCM combined with the ωB97X-D3/ma-def2-QZVPP level of theory (QM3) we have also computed the hydration free energies for acetate and methylammonium by hydrogen bonding two and three molecules of water to the charged functional groups, respectively. Here we employ the monomer method,[59] where the hydration free energy is with respect to separated water molecules rather than a cluster. Based on this, the hydration free energy of acetate is shifted from -298 kJ/mol to -293 kJ/mol, whereas for methylammonium the effect is much larger, taking this quantity from -311 to -290 kJ/mol. These results suggest that the continuum results will represent a lower bound with the real values being less exergonic.

While molecular dynamics considers solvation in a fully atomistic environment, it should be noted that even this approach can be subject to uncertainty, beyond the accuracy of the energy description, due to the alchemical processes involved. For example, for methylammonium with the AMOEBA model, the value quoted is for the Bennett Acceptance Ratio (-264 kJ/mol), but the forward and backward results of free energy perturbation with the current protocol span -262 to -297 kJ/mol and reduction of the hysteresis is difficult to achieve. Overall, both models give plausible hydration free energies to within the uncertainties for acetate. It should be noted that the initial AMOEBA parameters overestimated the hydration free energy of this anion, but it was possible to obtained an improved value while maintaining a good hydration structure. For methylammonium, AMOEBA is closer to the other theoretical results and experiment, though both models appear to underestimate the favourability of hydration. For this species, any further improvement in the hydration free

energies could only be obtained at the expense of a substantial deterioration in the radial distribution functions for water (i.e. allowing water to come too close relative to the distances found from AIMD). For glycine, the rigid-ion model is more consistent with arguably the lowest level of solvated quantum mechanical data, while AMOEBA is closer to the higher level of theory. Given the finding that inclusion of an explicit first hydration shell for acetate and methylammonium shifts the hydration free energy upward by 5 and 21 kJ/mol, respectively, the AMOEBA value is almost certainly the more accurate of the force field values.

Overall, the hydration free energies computed here show good agreement across the force fields and theoretical methods for acetate and glycine, while methylammonium appears to be more challenging. Given that an objective of the current work is to try to parameterise models based on the individual functional groups that can then be combined in other amino acids and beyond, this suggests that achieving transferability is difficult when balancing hydration structure and thermodynamics. Since the van der Waals parameters for acetate and methylammonium are combined in glycine without modification then it is arguably the cancellation of errors that leads to good results for this system (i.e. the values for acetate tend to be at the more exogenic bound, while methylammonium is at the opposite extreme). Across the three species, AMOEBA offers the superior description of hydration properties, as would be expected from the inclusion of polarisation and more sophisticated treatment of electrostatics.

Table 2: Solvation free energies (ΔG [kJ/mol]) at 300 K for the rigid-ion (RI), polarisable AMOEBA model (A) and three quantum mechanical approaches with different continuum solvent models (QM1 = PCM with HF/6-31G(r), QM2 = SM8 with M06/6-31+G**, QM3 = CPCM with ωB97X-D3/ma-def2-QZVPP) compared against previous literature (theoretical and experimental).

	RI	A	QM1	QM2	QM3	Theo	Exp
$CH_3CO_2^-$	-323	-330	-323	-330	-298	-372 to -278[60–66]	-343 to -322[60,63,64,67]
$^+NH_3CH_2CO_2^-$	-199	-181	-199	-215	-188	-247 to -174[68–72]	-
$CH_3NH_3^+$	-218	-264	-298	-334	-311	-364 to -281[61–64,73–76]	-305 to -293[63,64,74]

Ion Pairing Free Energies

Key quantities to consider when examining the ability of a force field to model the interactions of the organic species with a mineral are the pairing free energies between the organic molecule and the constituent ions of the mineral. If the thermodynamics of association of the component ions of the mineral in water with the organic functional groups is reasonable, then this bodes well for the adsorption free energy at the mineral surface also being accurate. The free energy profiles for the individual organic molecules interacting with the constituent ions of calcium carbonate are shown in Figure 3. The ion pairing free energies extracted from these profiles are included in Table 3.

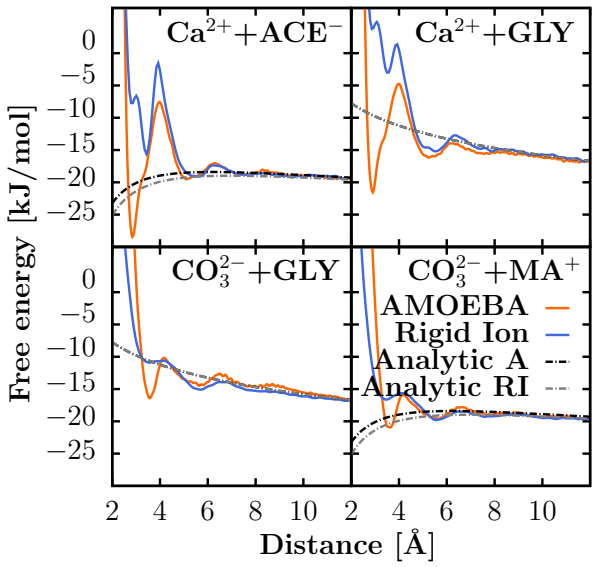

Figure 3: Ion pairing free energy profiles of acetate, glycine and methylammonium, as a function of separation distance, with the constituent ions of calcium carbonate. Here, the distances for molecular entities are taken between the C of carbonate or the carboxylate group, as appropriate, and with N for the interaction of glycine and methylammonium with carbonate. Note that only the interactions between oppositely charged or charged and neutral species are considered, as the interactions between like-charged species are dominated by Coulomb repulsion. Dashed lines indicate the analytic solution for point charges interacting in a continuum with the dielectric constant of the appropriate water model.

Comparison of the free energy profiles shown in Figure 3 shows that both force field mod-

Table 3: Ion pairing free energies [kJ/mol] for the rigid-ion (RI) and polarisable AMOEBA (A) model at 300 K compared against previous theoretical and experimental data. Two different integration limits were employed for each ion pair, namely for the contact ion pair (CIP) and solvent-shared ion pair (SSHIP).

	RI [CIP/SSHIP]	A [CIP/SSHIP]	Theo	Exp
$CH_3CO_2^- + Ca^{2+}$	+3.9/-3.5	-8.0/-8.3	-31.6 to -1.7[13,14]	-7.1 to -3.0[17,77–79]
$^+NH_3CH_2CO_2^- + Ca^{2+}$	+15.8/+0.8	-1.2/-2.5	-85.3, -7.1[80,81]	-1.1[80]
$^+NH_3CH_2CO_2^- + CO_3^{2-}$	+5.5/+0.2	+3.1/-0.3	-	-
$CH_3NH_3^+ + CO_3^{2-}$	+0.1/-4.6	-2.5/-5.1	-	-

els have the correct asymptotic behaviour in the long-range limit for the free energy, which indicates that the configurational space has been adequately sampled, and that both models are in good agreement in regard to the stability of the solvent-separated ion pair. Even for the solvent-shared ion pair, the agreement between the methods is to within thermal energy at ambient conditions, as demonstrated by the binding free energies given in Table 3. However, for the contact ion pair there is a systematic deviation, with the AMOEBA description consistently yielding more stable minima. Indeed, for the interaction between carbonate and glycine, the rigid-ion model contact state is close to being a point of inflection, while the barrier to dissociation is not much greater for methylammonium with carbonate. The greater stability of contact ion pairs in AMOEBA is in line with the expected behaviour due to the extra stability that results from polarisation when the species interact directly, rather than via a screening medium. Within the contact ion pairs of acetate and glycine with calcium there is also the question of how many carboxylate oxygens coordinate to the metal ion? For the rigid-ion model, monodentate coordination is clearly favoured in both cases with only a small activation barrier to leave the bidentate configuration at shorter distance. For AMOEBA, the converse is true in that the bidentate state is strongly preferred, while the monodentate state becomes essentially a point of inflection rather than a distinct minimum.

Overall, both models have favourable binding energies for acetate and methylammonium with the constituent ions of calcium carbonate, whereas the estimated pairing free energies for the glycine zwitterion with calcium and carbonate are not favourable, except for weak as-

sociation when the ions largely maintain their first hydration shell. The binding free energies between acetate and calcium are close to the experimentally expected range for both force field models, if the integration cut-off includes solvent-shared binding modes. These binding free energies for calcium-acetate are also better than many off-the-shelf organic force fields without any re-parameterisation.[14] For example, OPLS-AA without re-parameterisation substantially overestimates the calcium-acetate binding with an ion pairing free energy of approximately -32 kJ/mol as calculated by Kahlen *et al.*[14] The calcium-acetate ion pairing for both models in this work is also an improvement over the previous rigid-ion force field by Raiteri *et al.*,[13] where the ion pairing free energy was determined to be -9.5 kJ/mol. Recent work by de Oliveira *et al.*[17] also explored the binding of calcium-acetate using a free energy perturbation approach with different force fields, including an AMOEBA model. Their AMOEBA force field result of -11.4 kJ/mol for the CIP overestimated binding compared to their experimental reference of -3.2 kJ/mol, whereas the re-parameterised AMOEBA force field from this work gives slightly better agreement with a value of -8.1 kJ/mol. When examining rigid-ion force fields, de Oliveira *et al.* found that a formally-charged model led to overbinding of the calcium-acetate ion pair and advocated for the use of scaled charges to remedy this issue. In the present work we find that provided the hydration free energies are well reproduced, then a formal charge model is able to yield binding free energies that are comparable to the latest experimental value. As a further validation of the calcium-acetate interactions with AMOEBA, we have used gas phase quantum mechanics to compute the binding energy for acetate with $Ca(H_2O)_5^{2+}$ at the ωB97X-D3/ma-def2-QZVPP level of theory. Here the computed internal energy from AMOEBA is -996 kJ/mol in comparison to -958 kJ/mol from the quantum mechanical data, while the Ca-C distances were 2.736 and 2.748 Å, respectively. Although the difference in the binding energies may appear large in gas phase, the absolute values are considerably moderated in water and so the percentage discrepancy (4%) is arguably a better indicator.

The calcium-glycine zwitterion binding free energy for both new models is found to be in

good agreement with experiment.[80] In contrast, our values are at odds with the published value of -85.3 kJ/mol from DFT,[80] which is implausibly strong as a result of equating the enthalpy with the free energy while neglecting the important entropic contribution. The binding free energy of -7.1 kJ/mol obtained using the 3D-RISM solvent model[81] also supports weaker binding, even if it is stronger than found in experiment. One possible issue for this association process is that binding Ca^{2+} to the glycine zwitterion will shift the pKa of the glycine. As recently highlighted for the case of bicarbonate,[82] whose pKa is similar to the second value for glycine, the formation of an ion pair decreases the pKa by 1.7. If a similar effect were to occur for glycine then this would result in a pKa of $\sim$7.9 for the ion pair and implies that partial deprotonation of the ammonium group may occur; something that is not possible within our non-reactive force fields. Within our AIMD simulations of the calcium-glycine ion pair in water we did not observe any tendency for the glycine zwitterion to deprotonate, but this may be a consequence of the small box size where one proton transfer event represents a substantial shift in pH for the aqueous environment.

The overall trend for calcium-carboxylate binding taken across neutral zwitterionic glycine and anionic acetate is, as expected, that the strength of ion pairing increases with the increasing negative charge of the organic species. Previously published data[83] for a rigid-ion simulation of calcium binding to aspartate as a doubly-charged anion gives an ion pairing free energy of approximately -10 to -11 kJ/mol, depending on the specific carboxylate group involved, which is again consistent with this trend. Of course when extrapolated to an extended biomolecule, rather than focusing on small fragments, then the local charge becomes the relevant quantity for determining binding strength, rather than the total net charge. While carbonate binding to alkyl ammonium functional groups also shows a dependence on overall charge, the interaction strength is weaker. Indeed the rigid-ion model only shows binding for the solvent-separated state, which is driven primarily by entropic effects rather than the Coulomb contribution.

Conclusions

This work has developed parameters for small organic molecules to be used as proxies for the interactions of different organic functional groups with calcium carbonate biominerals. The force field refinement considered a variety of properties of the aqueous organic molecules and their association with the constituent ions of calcium carbonate. Two different models were derived, namely a rigid-ion model and a polarisable AMOEBA-like model, and their performance compared against experimental and first principles data for a number of key properties. While the inclusion of polarisation into force fields is obviously desirable, since it should enhance the transferability between different environments, such as from the gas phase to aqueous solution, it is often excluded based on the increased computational cost to solve for the self-consistent dipoles. In the present work, we find that the use of a GPU-based implementation of the AMOEBA model makes it competitive with the parallel CPU-based rigid-ion model in terms of the number of nanoseconds of molecular dynamics that can be achieved per day, at least for the system sizes considered here. Both of the force field models used here are able to reproduce the water structure around the molecules obtained from *ab initio* MD, as well as producing acceptable hydration free energies for each of the species to within the uncertainties associated with such quantities. The one exception is for methylammonium where the hydration free energy is insufficiently exergonic for the rigid-ion model, despite having a solvation environment that is in good agreement with other approaches. This highlights the need for compromise in the absence of the inclusion of polarisation, in that agreement for the hydration free energy could only be achieved if water is allowed to come unrealistically close to the NH_3^+ group.

The ion pairing free energies with the constituent ions of calcium carbonate are also in general agreement between the force field models, except for the contact ion pair state where ion polarisation becomes more significant. Despite this, the overall binding free energies for the species except acetate remain in good agreement because of the tendency to give solvent-shared or solvent-separated states as the most favourable configurations for ion pairing. The

AMOEBA model generated here yields a free energy of binding for Ca-acetate that is closer to the experimental range than in other recent work.[17] The main difference between the two studies is that here we use modified parameters that, in particular, target the experimental hydration free energy of Ca^{2+}. While reparameterisation improves the AMOEBA results for ion pairing in water, a tendency to overbind remains, as found in previous work.

Having calibrated the interactions for the two most common functional groups found in biomolecules that influence the crystallisation of calcium carbonate, the overall conclusion is that each binding interaction in the presence of water contributes of the order of one to three times ambient thermal energy (k_BT). This is important, since it suggests that coordination of biomolecules during crystallisation requires multiple interactions to achieve significant attachment rather than being driven by a single strong binding event. Given that each individual interaction is of the order of thermal energy, binding is reversible, which is consistent with a catalytic role, rather than incorporation (though this is also possible). This is in contrast with some previous computational studies that have reported large binding energies that would lead to irreversible attachment, which stresses the need to consider free energies of binding and to ensure that force fields are carefully benchmarked prior to being applied to such systems.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

RD and JDG thank the Australian Research Council for support though a Discovery Project (DP160100677) and fellowships (FT180100385/FL180100087). AS and RD thank Curtin University for support through a Science and Engineering Faculty Small Grant. The Pawsey Supercomputing Centre and the Australian National Computational Infrastructure are also

acknowledged for the provision of computing time through the NCMAS and Pawsey Partners merit allocation schemes.

References

(1) Kleber, M.; Bourg, I. C.; Coward, E. K.; Hansel, C. M.; Myneni, S. C. B.; Nunan, N. Dynamic interactions at the mineral-organic matter interface. *Nature Reviews Earth and Environment* **2021**, *2*, 402–421.

(2) Meldrum, F.; Coelfen, H. Controlling Mineral Morphologies and Structures in Biological and Synthetic Systems. *Chemical Reviews* **2008**, *108*, 4332–4432.

(3) Weiner, S.; Addadi, L. Crystallization Pathways in Biomineralization. *Annual Review of Materials Research* **2011**, *41*, 21–40.

(4) Gebauer, D.; Cölfen, H.; Verch, A.; Antonietti, M. The Multiple Roles of Additives in $CaCO_3$ Crystallization: A Quantitative Case Study. *Advanced Materials* **2009**, *21*, 435–439.

(5) Verch, A.; Gebauer, D.; Antonietti, M.; Cölfen, H. How to control the scaling of $CaCO_3$: a "fingerprinting technique" to classify additives. *Physical Chemistry Chemical Physics* **2011**, *13*, 16811–16820.

(6) Rao, A.; Berg, J. K.; Kellermeier, M.; Gebauer, D. Sweet on biomineralization: effects of carbohydrates on the early stages of calcium carbonate crystallization. *European Journal of Mineralogy* **2014**, *26*, 537–552.

(7) Lu, H.; Lutz, H.; Roeters, S. J.; Hood, M. A.; Schäfer, A.; Muñoz-Espí, R.; Berger, R.; Bonn, M.; Weidner, T. Calcium-Induced Molecular Rearrangement of Peptide Folds Enables Biomineralization of Vaterite Calcium Carbonate. *Journal of the American Chemical Society* **2018**, *140*, 2793–2796.

(8) Elhadj, S.; De Yoreo, J. J.; Hoyer, J. R.; Dove, P. M. Role of molecular charge and hydrophilicity in regulating the kinetics of crystal growth. *Proceedings of the National Academy of Sciences* **2006**, *103*, 19237–19242.

(9) Harding, J. H.; Duffy, D. M. The challenge of biominerals to simulations. *Journal of Materials Chemistry* **2006**, *16*, 1105–1112.

(10) Demichelis, R.; Schuitemaker, A.; Garcia, N. A.; Koziara, K. B.; De La Pierre, M.; Raiteri, P.; Gale, J. D. Simulation of Crystallization of Biominerals. *Annual Review of Materials Research* **2018**, *48*, 327–352.

(11) Freeman, C. L.; Harding, J. H.; Quigley, D.; Rodger, P. M. Structural control of crystal nuclei by an eggshell protein. *Angewandte Chemie International Edition* **2010**, *49*, 5135–5137.

(12) Freeman, C. L.; Harding, J. H.; Cooke, D. J.; Elliott, J. A.; Lardge, J. S.; Duffy, D. M. New forcefields for modeling biomineralization processes. *Journal of Physical Chemistry C* **2007**, *111*, 11943–11951.

(13) Raiteri, P.; Demichelis, R.; Gale, J. D.; Kellermeier, M.; Gebauer, D.; Quigley, D.; Wright, L. B.; Walsh, T. R. Exploring the influence of organic species on pre- and post-nucleation calcium carbonate. *Faraday Discussions* **2012**, *159*, 61–85.

(14) Kahlen, J.; Salimi, L.; Sulpizi, M.; Peter, C.; Donadio, D. Interaction of Charged Amino-Acid Side Chains with Ions: An Optimization Strategy for Classical Force Fields. *The Journal of Physical Chemistry B* **2014**, *118*, 3960–3972.

(15) Church, A. T.; Hughes, Z. E.; Walsh, T. R. Improving the description of interactions between Ca2+ and protein carboxylate groups, including γ-carboxyglutamic acid: revised CHARMM22* parameters. *RSC Advances* **2015**, *5*, 67820–67828.

(16) Saharay, M.; Kirkpatrick, R. J. Ab initio and metadynamics studies on the role of essential functional groups in biomineralization of calcium carbonate and environmental situations. *Physical Chemistry Chemical Physics* **2014**, *16*, 26843–26854.

(17) Mendes de Oliveira, D.; Zukowski, S. R.; Palivec, V.; Hénin, J.; Martinez-Seara, H.; Ben-Amotz, D.; Jungwirth, P.; Duboué-Dijon, E. Binding of divalent cations to acetate: molecular simulations guided by Raman spectroscopy. *Physical Chemistry Chemical Physics* **2020**, *22*, 24014–24027.

(18) Kellermeier, M.; Raiteri, P.; Berg, J.; Kempter, A.; Gale, J. D.; Gebauer, D. Entropy drives calcium carbonate ion association. *ChemPhysChem* **2016**, *17*, 3535–3541.

(19) De La Pierre, M.; Raiteri, P.; Gale, J. D. Structure and Dynamics of Water at Step Edges on the Calcite $10\bar{1}4$ Surface. *Crystal Growth and Design* **2016**, *16*, 5907–5914.

(20) Raiteri, P.; Demichelis, R.; Gale, J. D. Thermodynamically Consistent Force Field for Molecular Dynamics Simulations of Alkaline-Earth Carbonates and Their Aqueous Speciation. *The Journal of Physical Chemistry C* **2015**, *119*, 24447–24458.

(21) Wu, Y.; Tepper, H. L.; Voth, G. A. Flexible simple point-charge water model with improved liquid-state properties. *The Journal of Chemical Physics* **2006**, *124*, 024503.

(22) Xiao, S.; Edwards, S. A.; Graeter, F. A new transferable forcefield for simulating the mechanics of $CaCO_3$ crystals. *Journal of Physical Chemistry C* **2011**, *115*, 20067–20075.

(23) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry* **2004**, *25*, 1656–1676.

(24) Malde, A. K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P. C.; Oostenbrink, C.; Mark, A. E. An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *Journal of Chemical Theory and Computation* **2011**, *7*, 4026–4037.

(25) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. The Polarizable Atomic Multipole-based AMOEBA Force Field for Proteins. *Journal of Chemical Theory and Computation* **2013**, *9*, 4046–4063.

(26) Raiteri, P.; Schuitemaker, A.; Gale, J. D. Ion Pairing and Multiple Ion Binding in Calcium Carbonate Solutions Based on a Polarizable AMOEBA Force Field and Ab Initio Molecular Dynamics. *The Journal of Physical Chemistry B* **2020**, *124*, 3568–3582.

(27) Bradshaw, R. T.; Essex, J. W. Evaluating parametrization protocols for hydration free energy calculations with the AMOEBA polarizable force field. *Journal of Chemical Theory and Computation* **2016**, *12*, 3871–3883.

(28) Liu, C.; Piquemal, J.-P.; Ren, P. AMOEBA+ classical potential for modeling molecular interactions. *Journal of Chemical Theory and Computation* **2019**, *15*, 4122–4139.

(29) Stone, A. J. Distributed multipole analysis: Stability for large basis sets. *J. Chem. Theory Comput.* **2005**, *1*, 1128–1132.

(30) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **2010**, *132*, 154104.

(31) Bankura, A.; Karmakar, A.; Carnevale, V.; Chandra, A.; Klein, M. L. Structure, dynamics, and spectral diffusion of water from first-principles molecular dynamics. *Journal of Physical Chemistry C* **2014**, *118*, 29401–29411.

(32) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Computer Physics Communications* **2005**, *167*, 103–128.

(33) Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J. cp2k:atomistic simulations of condensed matter systems. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2013**, *4*, 15–25.

(34) Goedecker, S.; Teter, M.; Hutter, J. Separable dual-space Gaussian pseudopotentials. *Physical Review B* **1996**, *54*, 1703–1710.

(35) VandeVondele, J.; Hutter, J. An efficient orbital transformation method for electronic structure calculations. *The Journal of Chemical Physics* **2003**, *118*, 4365–4369.

(36) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Computational Biology* **2017**, *13*, e1005659.

(37) Friedrichs, M. S.; Eastman, P.; Vaidyanathan, V.; Houston, M.; Legrand, S.; Beberg, A. L.; Ensign, D. L.; Bruns, C. M.; Pande, V. S. Accelerating molecular dynamic simulation on graphics processing units. *Journal of Computational Chemistry* **2009**, *30*, 864–872.

(38) Lagardère, L.; Jolly, L.-H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; Gresh, N.; Maday, Y.; Ren, P. Y.; Ponder, J. W.; Piquemal, J.-P. Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chemical Science* **2018**, *9*, 956–972.

(39) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *Journal of Computational Physics* **1995**, *117*, 1–19.

(40) Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Self-consistent reaction field model for aqueous and nonaqueous solutions based on accurate polarized partial charges. *Journal of Chemical Theory and Computation* **2007**, *3*, 2011–2033.

(41) Shao, Y. et al. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Molecular Physics* **2015**, *113*, 184–215.

(42) Frisch, M. J. et al. Gaussian09 Revision E.1. Gaussian Inc. Wallingford CT 2013.

(43) Neese, F. The ORCA program system. *Computational Molecular Science* **2012**, *2*, 73–78.

(44) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *The Journal of Chemical Physics* **1935**, *3*, 300–313.

(45) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *The Journal of Chemical Physics* **1954**, *22*, 1420–1426.

(46) Leslie, M.; Gillan, M. J. The energy and elastic dipole tensor of defects in ionic-crystals calculated by the supercell method. *Journal of Physics C: Solid State Physics* **1985**, *18*, 973–982.

(47) Hummer, G.; Pratt, L. R.; García, A. E. Ion sizes and finite-size corrections for ionic-solvation free energies. *The Journal of Chemical Physics* **1997**, *107*, 9275–9277.

(48) Kastenholz, M. A.; Hünenberger, P. H. Computation of methodology-independent ionic solvation free energies from molecular simulations. II. The hydration free energy of the sodium cation. *The Journal of Chemical Physics* **2006**, *124*, 224501–21.

(49) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics* **1976**, *22*, 245–268.

(50) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics* **1994**, *101*, 4177–4189.

(51) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Computer Physics Communications* **2014**, *185*, 604–613.

(52) Bonomi, M. et al. Promoting Transparency and Reproducibility in Enhanced Molecular Simulations. *Nature Methods* **2019**, *16*, 670–673, Cited by 142.

(53) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Physical Review Letters* **2008**, *100*, 020603.

(54) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics †. *The Journal of Physical Chemistry B* **2006**, *110*, 3533–3539.

(55) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proceedings of the National Academy of Sciences* **2002**, *99*, 12562–12566.

(56) Laio, A.; Gervasio, F. L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics* **2008**, *71*, 126601.

(57) Byrne, E. H.; Raiteri, P.; Gale, J. D. Computational Insight into Calcium–Sulfate Ion Pair Formation. *The Journal of Physical Chemistry C* **2017**, *121*, 25956–25966.

(58) Laury, M. L.; Wang, L.-P.; Pande, V. S.; Head-Gordon, T.; Ponder, J. W. Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model. *The Journal of Physical Chemistry B* **2014**, *119*, 9423–9437.

(59) Bryantsev, V. S.; Diallo, M. S.; Goddard III, W. A. Calculation of solvation free energies of charged solutes using mixed cluster/continuum models. *The Journal of Physical Chemistry B* **2008**, *112*, 9709–9719.

(60) Florián, J.; Warshel, A. Langevin Dipoles Model for ab Initio Calculations of Chemical Processes in Solution: Parametrization and Application to Hydration Free Energies of Neutral and Ionic Solutes and Conformational Analysis in Aqueous Solution. *The Journal of Physical Chemistry B* **1997**, *101*, 5583–5595.

(61) Meng, E. C.; Cieplak, P.; Caldwell, J. W.; Kollman, P. A. Accurate Solvation Free Energies of Acetate and Methylammonium Ions Calculated with a Polarizable Water Model. *Journal of the American Chemical Society* **1994**, *116*, 12061–12062.

(62) Setiadi, J.; Kuyucak, S. A simple, parameter-free method for computing solvation free energies of ions. *The Journal of Chemical Physics* **2019**, *150*, 065101.

(63) Kang, Y. K.; Nemethy, G.; Scheraga, H. A. Free energies of hydration of solute molecules. 3. Application of the hydration shell model to charged organic molecules. *The Journal of Physical Chemistry* **1987**, *91*, 4118–4120.

(64) Cramer, C. J.; Truhlar, D. G. General parameterized SCF model for free energies of solvation in aqueous solution. *Journal of the American Chemical Society* **1991**, *113*, 8305–8311.

(65) Pliego Jr, J. R.; Riveros, J. M. New values for the absolute solvation free energy of univalent ions in aqueous solution. *Chemical Physics Letters* **2000**, *332*, 597–602.

(66) Kamerlin, S. C. L.; Haranczyk, M.; Warshel, A. Progress in ab initio QM/MM free-energy simulations of electrostatic energies in proteins: accelerated QM/MM studies of pKa, redox reactions and solvation free energies. *The Journal of Physical Chemistry B* **2009**, *113*, 1253–1272.

(67) Pearson, R. G. Ionization potentials and electron affinities in aqueous solution. *Journal of the American Chemical Society* **1986**, *108*, 6109–6114.

(68) Dixit, S. B.; Bhasin, R.; Rajasekaran, E.; Jayaram, B. Solvation thermodynamics of amino acids Assessment of the electrostatic contribution and force-field dependence. *Journal of the Chemical Society, Faraday Transactions* **1997**, *93*, 1105–1113.

(69) Chang, J.; Lenhoff, A. M.; Sandler, S. I. Solvation Free Energy of Amino Acids and Side-Chain Analogues. *The Journal of Physical Chemistry B* **2007**, *111*, 2098–2106.

(70) Truong, T. N.; Stefanovich, E. V. Analytical first and second energy derivatives of the generalized conductorlike screening model for free energy of solvation. *The Journal of Chemical Physics* **1995**, *103*, 3709–3717.

(71) Takahashi, H.; Kawashima, Y.; Nitta, T.; Matubayasi, N. A novel quantum mechanical/molecular mechanical approach to the free energy calculation for isomerization of glycine in aqueous solution. *The Journal of Chemical Physics* **2005**, *123*, 124504–9.

(72) Gusarov, S.; Ziegler, T.; Kovalenko, A. Self-consistent combination of the three-dimensional RISM theory of molecular solvation with analytical gradients and the Amsterdam density functional package. *The Journal of Physical Chemistry A* **2006**, *110*, 6083–6090.

(73) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *Journal of the American Chemical Society* **1990**, *112*, 6127–6129.

(74) Lim, C.; Bashford, D.; Karplus, M. Absolute pKa calculations with continuum dielectric methods. *The Journal of Physical Chemistry* **1991**, *95*, 5610–5620.

(75) Fedotova, M. V.; Kruchinin, S. E. Hydration of methylamine and methylammonium ion: structural and thermodynamic properties from the data of the integral equation method in the RISM approximation. *Russian Chemical Bulletin* **2012**, *61*, 240–247.

(76) Cramer, C. J.; Truhlar, D. G. PM3-SM3: A general parameterization for including aqueous solvation effects in the PM3 molecular orbital model. *Journal of Computational Chemistry* **1992**, *13*, 1089–1097.

(77) Shock, E. L.; Koretsky, C. M. Metal-organic complexes in geochemical processes: Estimation of standard partial molal thermodynamic properties of aqueous complexes between metal cations and monovalent organic acid ligands at high pressures and temperatures. *Geochimica et Cosmochimica Acta* **1995**, *59*, 1497–1532.

(78) Daniele, P. G.; Foti, C.; Gianguzza, A.; Prenesti, E.; Sammartano, S. Weak alkali and alkaline earth metal complexes of low molecular weight ligands in aqueous solution. *Coordination Chemistry Reviews* **2008**, *252*, 1093–1107.

(79) Joseph, N. R. The dissociation constants of organic calcium complexes. *Journal of Biological Chemistry* **1946**, *164*, 529–541.

(80) Tang, N.; Skibsted, L. H. Calcium Binding to Amino Acids and Small Glycine Peptides in Aqueous Solution: Toward Peptide Design for Better Calcium Bioavailability. *Journal of Agricultural and Food Chemistry* **2016**, *64*, 4376–4389.

(81) Fedotova, M. V.; Kruchinin, S. E. Ion-binding of glycine zwitterion with inorganic ions in biologically relevant aqueous electrolyte solutions. *Biophysical Chemistry* **2014**, *190-191*, 25–31.

(82) Huang, Y.-C.; Rao, A.; Huang, S.-J.; Chang, C.-Y.; Drechsler, M.; Knaus, J.; Chan, J. C. C.; Raiteri, P.; Gale, J. D.; Gebauer, D. Uncovering the Role of Bicarbonate in Calcium Carbonate Formation at Near-Neutral pH. *Angewandte Chemie International Edition* **2021**, *60*, 16707–16713.

(83) Jiang, W.; Athanasiadou, D.; Zhang, S.; Demichelis, R.; Koziara, K. B.; Raiteri, V., Paolo Nelea; Mi, W.; Ma, J.-A.; Gale, J. D.; McKee, M. D. Homochirality

in biomineral suprastructures induced by assembly of single-enantiomer amino acids from a nonracemic mixture. *Nature Communications* **2019**, *10*, 2318.