


 Cite this: *RSC Adv.*, 2019, 9, 1653

# Molecular dynamics simulation of protein-mediated biomineralization of amorphous calcium carbonate

 R. Sandya Rani  and Moumita Saharay \*

The protein-mediated biomineralization of calcium carbonate ( $\text{CaCO}_3$ ) in living organisms is primarily governed by critical interactions between the charged amino acids of the protein, solvent, calcium ( $\text{Ca}^{2+}$ ) and carbonate ( $\text{CO}_3^{2-}$ ) ions. The present article investigates the molecular mechanism of lysozyme-mediated nucleation of amorphous calcium carbonate (ACC) using molecular dynamics and metadynamics simulations. The results reveal that, by acting as nucleation sites, the positively charged side chains of surface-exposed arginine residues form hydrogen bonds with carbonates and promote aggregation of ions around them leading to the formation and growth of ACC on the protein surface. The newly formed ACC patches were found to be less hydrated due to ion aggregation-induced expulsion of water from the nucleation sites. Despite favorable electrostatic interactions of the negatively charged side chains of aspartate and glutamate with calcium ions, these residues contribute minimally to the growth of ACC on protein surface. The activation barrier for the growth of partially hydrated ACC patches on lysozymes was determined from the free energy profiles obtained from metadynamics simulations.

 Received 12th October 2018  
Accepted 21st December 2018

DOI: 10.1039/c8ra08459a

[rsc.li/rsc-advances](http://rsc.li/rsc-advances)

## Introduction

Calcium carbonate ( $\text{CaCO}_3$ ) is a commonly found biomineral in the environment. The diversity in the morphology, origin, and composition exhibited by naturally occurring  $\text{CaCO}_3$  makes it an intriguing system to investigate the mechanism, energetics and thermodynamics of biomineralization. Apart from the three anhydrous (calcite, vaterite, and aragonite) and two hydrous (monohydrocalcite,  $\text{CaCO}_3 \cdot \text{H}_2\text{O}$ ; and ikaite,  $\text{CaCO}_3 \cdot 6\text{H}_2\text{O}$ ) crystalline polymorphs, calcium carbonate can also exist in metastable amorphous phases.<sup>1</sup> Amorphous calcium carbonate (ACC) serves as a precursor in the biomineralization of almost all types of biogenic  $\text{CaCO}_3$ .<sup>2–8</sup>

In living organisms, the process of mineral formation is primarily governed by the interactions between the charged biopolymers,<sup>9–12</sup> ionic constituents ( $\text{Ca}^{2+}$  and  $\text{CO}_3^{2-}$ ), and water. For instance, the negatively charged proteins in corals and sea urchin spicules are shown to stabilize the amorphous phase of  $\text{CaCO}_3$  even in *in vitro* experiments in the absence of any biological cells.<sup>13–15</sup> A recent study on the formation of coral skeleton demonstrated that the ACC phase is indeed the metastable precursor to the crystalline  $\text{CaCO}_3$ , and that the growth rate of aragonite crystals through the non-classical crystallization process<sup>11,12</sup> involving aggregation of ACC particles is  $\sim 100$  times faster than the classical route of ion-by-ion formation of

crystals.<sup>16</sup> Earlier studies have also shown that the binding affinity of Ovocleidin-17 (OC-17), a hen eggshell protein, for ACC nanoparticles is much higher than that for the calcite surface<sup>17,18</sup> suggesting that favorable binding of the protein to ACC is critical for biomineralization and eggshell formation. The lower calcite-protein binding affinity is leveraged for the detachment of the protein from the crystal after the ACC to crystal transition.<sup>18</sup> Thus, OC-17 or a class of functionally similar proteins/biopolymers can be regarded as catalysts to facilitate the crystallization of  $\text{CaCO}_3$ .

Although previous computer simulation studies have investigated the mechanism of the attachment of the protein<sup>19,20</sup> to the ACC nanoparticles and  $\text{CaCO}_3$  crystal surfaces, little is known about the role of protein in promoting the formation of ACC from the ionic solution. Another key feature that is yet to be understood is the atomistic details of the mechanism and energetics of protein-ACC association/dissociation process. Some proteins/biopolymers hinder the crystallization of  $\text{CaCO}_3$  by stabilizing<sup>3,21,22</sup> the amorphous phase, and the biomineralization is activated as and when the living organism requires it. The nucleation of  $\text{CaCO}_3$  in the presence of negatively charged biopolymers<sup>10</sup> is thought to occur in two steps: (i) the binding of  $\text{Ca}^{2+}$  with the biopolymer to form  $\text{Ca}^{2+}$ -bound globules, (ii) the diffusion and association of  $\text{CO}_3^{2-}$  ions with the globules to form polymer-mediated ACC phase that then grows in size. These polymers, referred to as 'ion sponge', critically determine where and how the  $\text{CaCO}_3$  nucleation will take

Department of Physics, Osmania University, Hyderabad, India. E-mail: [saharaymou@gmail.com](mailto:saharaymou@gmail.com)



place and proceed further. However, the precise molecular details associated with these steps are still missing.

*In vitro* experiments have shown that the positively charged lysozyme increases the nucleation rate of  $\text{CaCO}_3$  crystals.<sup>23,24</sup> The large negative  $\zeta$ -potential ( $-14$  mV) around this protein favors the formation of negatively charged double layers of  $\text{CO}_3^{2-}$  around its surface.<sup>23</sup> To understand the microscopic details of the onset of protein-mediated aggregation of  $\text{Ca}^{2+}$  and  $\text{CO}_3^{2-}$  and to identify the vital molecular attributes and the key residues that govern the mineralization process,<sup>25,26</sup> we have performed a series of atomistic molecular dynamics simulations of lysozyme, a hen eggshell protein. Our results reveal that the positively charged side chains of solvent-exposed arginine residues act as nucleation sites for the formation and growth of ACC on protein surface, while the negatively charged side chains of glutamate and aspartate residues contribute minimally to biomineralization.

## Simulation details

We have investigated the following four model systems: (i) lysozyme in water (hereafter referred to as LW), (ii) hydrated ACC (hACC) having 1 : 0.67 composition of  $\text{CaCO}_3$  :  $\text{H}_2\text{O}$ , (iii) lysozyme in  $\text{CaCO}_3$  solution with two different  $\text{CaCO}_3$  :  $\text{H}_2\text{O}$  concentrations ( $\text{S1}_{\text{LC}}$  and  $\text{S2}_{\text{LC}}$ ). The dimensions and the number of constituents of the simulated model systems are provided in Table 1.

The LW model allows us to characterize the lysozyme–water interactions in the absence of  $\text{Ca}^{2+}$  and  $\text{CO}_3^{2-}$  ions, whereas  $\text{S1}_{\text{LC}}$ ,  $\text{S2}_{\text{LC}}$  models are used to probe the influence of these ions on lysozyme. The initial coordinates of the hen egg-white lysozyme (PDB Id: 2cds) was taken from the X-ray crystal structure. To neutralize the net positive charge on lysozyme, 8  $\text{Cl}^-$  were added in LW system, while 4 extra  $\text{CO}_3^{2-}$  were added in  $\text{S1}_{\text{LC}}$ ,  $\text{S2}_{\text{LC}}$  models. The initial configurations of  $\text{S1}_{\text{LC}}$ ,  $\text{S2}_{\text{LC}}$  were generated by randomly placing desired numbers of  $\text{Ca}^{2+}$  and  $\text{CO}_3^{2-}$  ions in cubic water boxes of sizes specified in Table 1. The lysozyme was then solvated in this ionic solution by removing the ions and water molecules that were in hard contact with the protein using CHARMM-GUI<sup>27–29</sup> package. CHARMM36 (ref. 30) force field (FF) parameters were used for the protein and ions ( $\text{Ca}^{2+}$ ,  $\text{CO}_3^{2-}$ ) together with the standard TIP3P water model.

For hACC,  $\text{H}_2\text{O}$  :  $\text{CaCO}_3$  ratio was maintained at 0.67 to be consistent with our earlier investigation.<sup>31</sup> The MD simulations were performed using NAMD<sup>32</sup> package. The systems were equilibrated for 1 ns in NPT ensemble followed by 15 ns of MD simulations in NVT ensemble with the time step of 2 fs. The

temperature and pressure were maintained at 303 K and 1 atm, respectively. The temperature was controlled by the Langevin thermostat using a damping coefficient of  $1 \text{ ps}^{-1}$ . The electrostatic interactions were evaluated using particle mesh Ewald method with a grid spacing of 1 Å. Periodic boundary conditions were used and a distance cutoff of 12 Å was chosen to calculate the non-bonded interaction energies between different species.

To understand the flexibility of the near neighbor environment around the positively charged side chains, we studied the evolution of hydrogen bonds formed by the side chains with the nearest  $\text{CO}_3^{2-}$  and water molecules. An arginine side chain was considered to be hydrogen bonded to  $\text{CO}_3^{2-}$  if the CZ–O<sub>c</sub> (refer Fig. 1 for atom labels) distance was less than 4.5 Å or to a water molecule if the CZ–O<sub>w</sub> distance was less than 4.0 Å.<sup>33–35</sup> For each type of h-bond, the time correlation function,  $C_{\text{hb}}(t)$ , which is defined as follows,

$$C(t) = \langle h(0)h(t) \rangle / \langle h \rangle$$

was calculated with a time resolution of 2 ps. Here,  $h(t) = 1$  when two groups are h-bonded, and  $h(t) = 0$  otherwise. The ' $\langle \rangle$ ' represents the ensemble average.

We performed well-tempered metadynamics<sup>36–38</sup> simulations, using NAMD package, to estimate the free energy required for the association of ions and dehydration of hydrated  $\text{Ca}^{2+}$ – $\text{CO}_3^{2-}$  clusters around selected side chains of lysozyme. The separation ( $r_{\text{CV}}$ ) between the tagged ions and the residue of interest was chosen as the collective variable (CV). The Gaussian hills were added at a regular interval of 100 fs, and the width of each Gaussian hill was 1 Å. The total run length for the metadynamics simulation was 100 ns.

## Results and discussion

### Ionic structure around lysozyme

Fig. 2 compares the arrangement of ions around lysozyme in the initial configuration of  $\text{S1}_{\text{LC}}$  with that in a snapshot taken after 15 ns of MD simulation. The ions are distributed randomly around lysozyme in the initial structure, while dense ionic patches are observed around specific surface residues of lysozyme in the equilibrated system. The aggregation of ions occur primarily around arginine residues, but to some extent also around glutamate and aspartate. The clustering of  $\text{Ca}^{2+}$  and  $\text{CO}_3^{2-}$  at specific surface sites of the protein underscores the key role of protein-mediated aggregation of ions in the formation of biominerals. The average mass density of the dense ionic

Table 1 System specification

System	$\text{CaCO}_3$ pairs	Water molecules	Orthorhombic box dimension (Å <sup>3</sup> )
LW	—	30 284	70.0 × 70.0 × 70.0
$\text{S1}_{\text{LC}}$	235	14 289	76.2 × 76.2 × 76.2
$\text{S2}_{\text{LC}}$	474	14 238	76.5 × 76.5 × 76.5
hACC	1620	1094	50.0 × 50.0 × 50.0



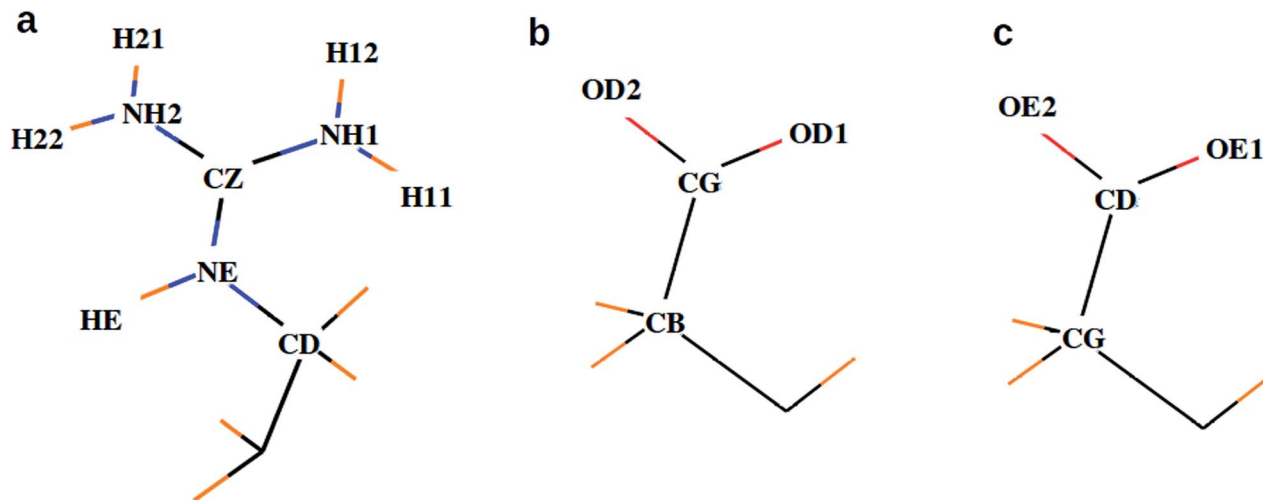


Fig. 1 Atom names used for the descriptions of side chains in (a) arginine, (b) aspartate, and (c) glutamate.

patches around arginine is  $\sim 1.3 \text{ g cm}^{-3}$ , which is comparable with that of hydrated ACC.<sup>26</sup> However, the local environment around glutamate ( $0.7 \text{ g cm}^{-3}$ ) and aspartate ( $0.71 \text{ g cm}^{-3}$ ) are relatively less dense than the microenvironment of arginine residues.

The clusters of ions are observed to grow in size during the course of MD simulation. The largest ion cluster is formed around ARG14, ARG125, and ARG128. Due to their close proximity, these three surface residues form an 'attractive triangle', which favors the rapid growth of amorphous calcium carbonate on lysozyme. The positively charged guanidinium ( $\text{Gdm}^+$ ) functional groups of arginine residues play a crucial role in the nucleation and growth<sup>25</sup> of crystalline  $\text{CaCO}_3$  by favorably interacting with the negatively charged  $\text{CO}_3^{2-}$  ions. The  $\text{Gdm}^+ - \text{CO}_3^{2-}$  and  $\text{Ca}^{2+} - \text{CO}_3^{2-}$  interactions facilitate the ion association and further growth of amorphous  $\text{CaCO}_3$ -like domains around these residues.

### Radial distribution function

To further investigate the local environment around the key residues of the protein, we examined the distribution of ions and water around them by calculating the radial distribution functions (RDFs) that describes how the density of neighbors vary around a reference atom/particle as a function of radial separation between the two.

**Local environment around ions.** Fig. 3 compares the RDFs involving  $\text{Ca}^{2+}$ ,  $\text{CO}_3^{2-}$ , and water oxygen ( $\text{O}_w$ ) of hACC with that of  $\text{S2}_{\text{LC}}$ . The water structure around  $\text{Ca}^{2+}$  and  $\text{CO}_3^{2-}$  are mostly similar in both the systems as evident from the peak positions and the peak heights of  $\text{Ca}-\text{O}_w$  and  $\text{C}_c-\text{O}_w$  distributions. However, the population of water molecules (Table 2) around  $\text{Ca}^{2+}$  and  $\text{CO}_3^{2-}$  differs significantly between these two systems due to the difference in  $\text{CaCO}_3 : \text{H}_2\text{O}$  ratio (1620 : 1094 in hACC and 474 : 14 238 in  $\text{S2}_{\text{LC}}$ ).

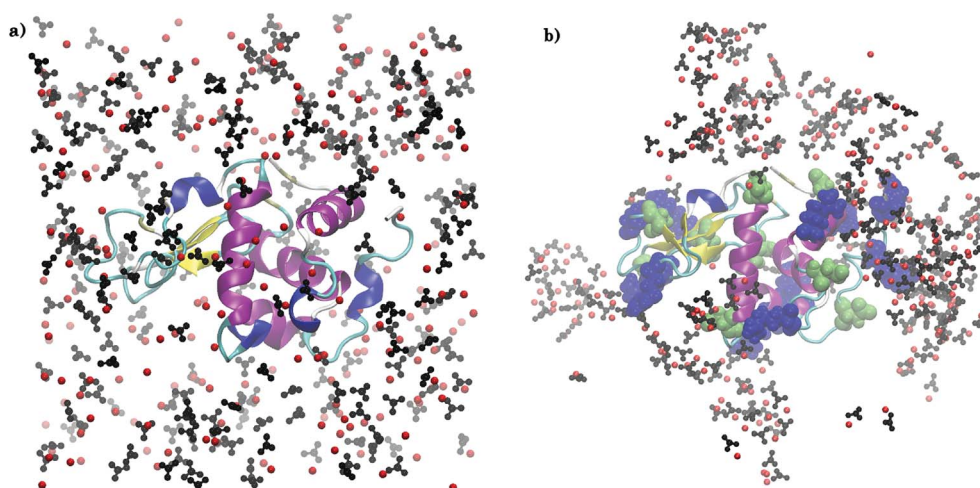


Fig. 2 Arrangement of  $\text{Ca}^{2+}$  (red spheres) and  $\text{CO}_3^{2-}$  (black) ions around arginine (blue), aspartate (green), and glutamate (green) residues in  $\text{S1}_{\text{LC}}$  for (a) initial configuration and (b) after 15 ns of MD simulation. Water molecules are not shown for clarity. VMD<sup>39</sup> software package was used for visualization.



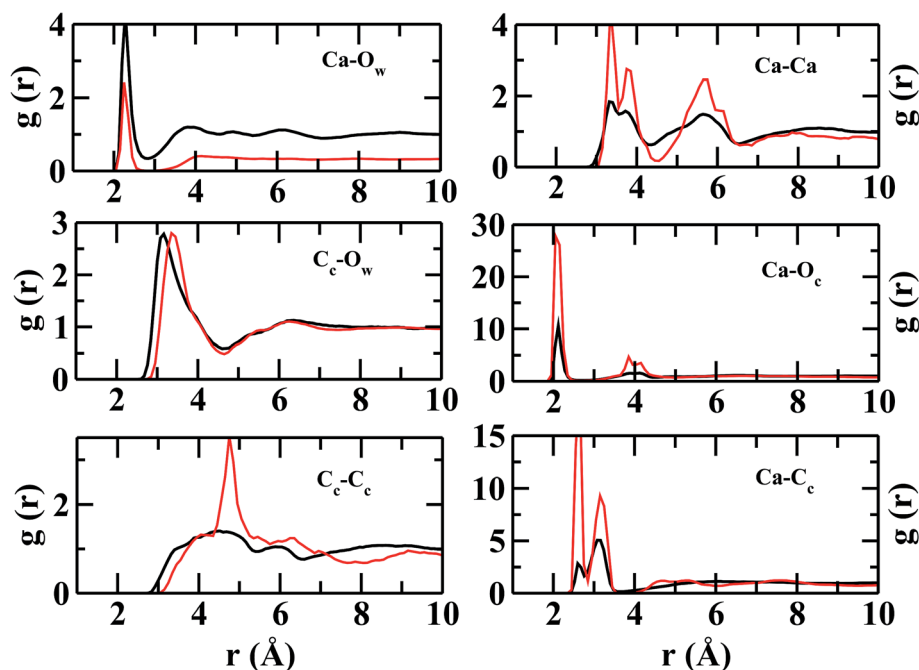


Fig. 3 Intermolecular radial distribution functions for hACC (black) and S2<sub>LC</sub> (red). All partial RDFs in S2<sub>LC</sub>, except for C<sub>c</sub>–O<sub>w</sub>, are compressed by a factor of 3 for visual comparison with the respective RDFs in hACC.

Table 2 The peak positions and coordination numbers ( $N(r)$ ) in the ion–ion and ion–water radial distribution functions obtained from hACC and S2<sub>LC</sub>

Atom pair	S2 <sub>LC</sub>		hACC	
	Peak position $r$ (Å)	$N(r)$	Peak position $r$ (Å)	$N(r)$
C <sub>c</sub> –C <sub>c</sub>	4.8	3.0	4.6	9.2
Ca–C <sub>c</sub>	2.6, 3.2	2.3	2.6, 3.2	4.6
Ca–O <sub>w</sub>	2.2	3.0	2.2	0.8
Ca–Ca	3.3, 3.8	1.3	3.3, 3.7	4.4
Ca–O <sub>c</sub>	2.1	3.2	2.07	5.4
C <sub>c</sub> –O <sub>w</sub>	3.3	12.8	3.1	4.04

Although the peak intensities in ion–ion RDFs are relatively higher in S2<sub>LC</sub> than in hACC, peak positions remain mostly unchanged. The ion–ion RDFs indicate that the ions are held together strongly to form a semi-rigid and well-structured coordination shells in S2<sub>LC</sub>. The present structural analysis reveals that local environment around lysozyme favors the aggregation of ions and the formation of amorphous calcium carbonate, although, the concentration of ions is much lower than the hydrated ACC (hACC) studied here.

**Distribution of ions around lysozyme.** Here we investigated the near neighbor structures around arginine, glutamate, and aspartate side chain residues. The high intensity peaks in the protein–ion RDFs indicate well resolved first and second coordination shells of ions around these key residues, preferably for arginine (Fig. 4), in both S1<sub>LC</sub> and S2<sub>LC</sub>.

In Fig. 4, the RDFs between N<sub>Gdm</sub> (guanidinium-nitrogen) and C<sub>c</sub> (carbonate carbon) atom pair shows three maxima at

3.34, 5.18, and 7 Å. The coordination number at the first minimum is 1.24. The first and second maxima are separated by a plateau of width  $\sim 0.7$  Å with negligible intensity, showing a crystal-like ordering of C<sub>c</sub> atoms in this region. The number of O<sub>c</sub> (carbonate oxygen) around N<sub>Gdm</sub> is  $\sim 1.4$  at the first minimum (2.9 Å) of  $g_{N_{Gdm}-O_c}(r)$  (Fig. 4). The underlying geometry could possibly indicate a configuration where one of the H atoms of a N<sub>Gdm</sub> forms single h-bond with O<sub>c</sub>. This preferential arrangement compares well with the geometry optimized isolated cluster<sup>25</sup> of Gdm<sup>+</sup>–CaCO<sub>3</sub>–2H<sub>2</sub>O, where the distance between N<sub>Gdm</sub> and O<sub>c</sub> is  $\sim 2.65$  Å. The second and third maxima in  $g_{N_{Gdm}-O_c}(r)$  appear at 3.4 and 4.6 Å, respectively. The coordination numbers at the corresponding minima in S1<sub>LC</sub> are 2.3 (2.0 in S2<sub>LC</sub>) and 4.7 (4.5 in S2<sub>LC</sub>), respectively.

The first maximum in  $g_{N_{Gdm}-Ca}(r)$  (Fig. 4) shows two sub peaks at 4.2 and 5.5 Å in both the systems. This characteristic arrangement of Ca<sup>2+</sup> around N<sub>Gdm</sub> is somewhat similar to that of Ca<sup>2+</sup>–environment around C<sub>c</sub> in hACC (Fig. 3), where the first peak of  $g_{Ca-C}(r)$  is also divided into two sub peaks. This indicates two preferred arrangements: for longer distance, Ca<sup>2+</sup> is coordinated with only one N<sub>Gdm</sub> or O<sub>c</sub> making a ‘monodentate’ geometry whereas for shorter distance Ca<sup>2+</sup> is equidistant from two N<sub>Gdm</sub> or O<sub>c</sub> making a ‘bidentate’ geometry. With increased ionic concentration in S2<sub>LC</sub>, the intensity of these peaks decreases which implies a less-structured coordination shell with same number of neighboring Ca<sup>2+</sup>.

The C<sub>c</sub> coordination number around CZ (Fig. 4) is 2 within the first coordination shell for  $r_{CZ-C_c} < 5.4$  Å, which indicates that each pair of N<sub>Gdm</sub> atoms (NH1, NH2; and NH2, NE) of Gdm<sup>+</sup> interacts with a single CO<sub>3</sub><sup>2-</sup>. The arrangement and density of anions up to the second coordination shell is mostly





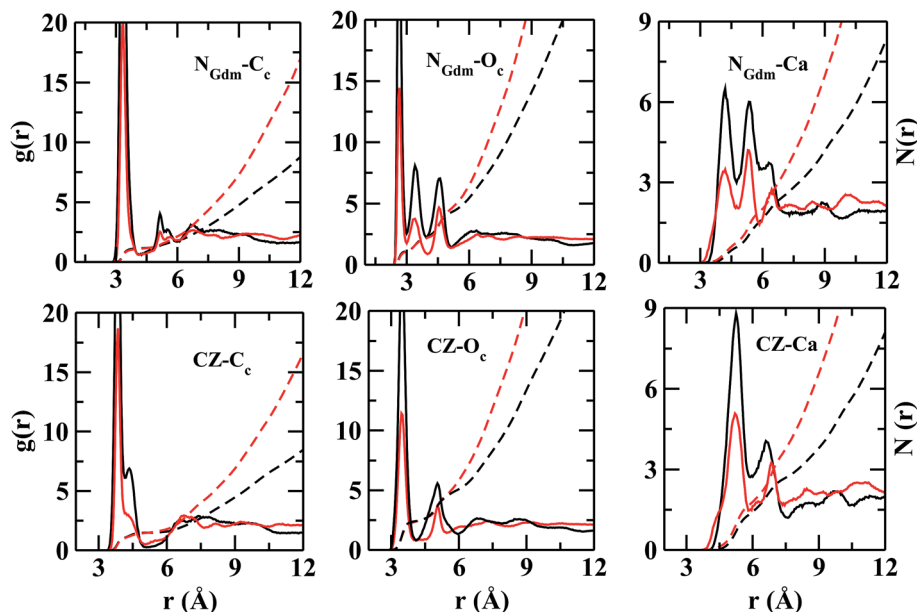


Fig. 4 Partial radial distribution functions ( $g(r)$ , solid line) and running coordination numbers ( $N(r)$ , dashed line) between the head groups of arginine with ions in S1<sub>LC</sub> (black) and S2<sub>LC</sub> (red). Atom labels: guanidinium-nitrogen ( $N_{Gdm}$ ), guanidinium-carbon (CZ), carbon of  $CO_3^{2-}$  ( $C_c$ ), oxygen of  $CO_3^{2-}$  ( $O_c$ ).

similar in both the systems while the ion density increases significantly for S2<sub>LC</sub> beyond second coordination shell.

However, the near neighbor environment around the negatively charged functional groups (Fig. 5) of ASP and GLU residues are mostly populated by  $Ca^{2+}$ . The RDF between carbonyl-oxygen (OD) of ASP and  $Ca^{2+}$  shows two sharp maxima at 2.2 and 4.2 Å, separated by a zero-intensity plateau of 1 Å width which signifies a crystal-like near neighbor ordered structure up to ~5 Å from ASP. In addition, the  $g_{CG-Ca}(r)$  (Fig. 5) also shows two high-intensity peaks at ~2.2 and 3.2 Å signifying two preferred geometries of the cations with respect to the  $COO^-$  group of ASP/GLU, bidentate and monodentate respectively. The radial distributions of  $C_c$  and  $O_c$  (Fig. 5) around  $COO^-$  group of ASP/GLU are poorly structured with coordination numbers less than 1 up to the first minima at ~5.0 Å for both the distributions.

### Density map

**Distribution of  $Ca^{2+}$  and  $CO_3^{2-}$  ions around arginine.** Among the three  $sp^2$ -hybridized nitrogen atoms (NH1, NH2, and NE) of the  $Gdm^+$  group of arginine side chain, NH1 and NH2 can form two h-bonds and NE can establish a single h-bond with the  $CO_3^{2-}$  ions. MD trajectory-based visual inspection of the local environment of arginine residues reveals that each arginine is coordinated to two near neighbor  $CO_3^{2-}$  ions to form the ‘twin N-twin O’ (TNTO) h-bond geometry (Fig. 6a). In this geometry, NH2 forms two h-bonds with these  $CO_3^{2-}$  ions, while NH1 and NE form single h-bonds with them. The  $CO_3^{2-}$  is mostly in the bidentate geometry with the nitrogen atom pairs forming in-plane h-bonds with the  $Gdm^+$  plane. The average length of arginine- $CO_3^{2-}$  h-bonds is ~1.65 Å.

The calculated atomic density distribution (Fig. 6), *i.e.* density map, of  $CO_3^{2-}$  around  $Gdm^+$  shows two density peaks of

$C_c$  atoms within the CZ- $C_c$  distance cutoff of 4.3 Å, the first minimum in  $g_{CZ-C_c}(r)$  (Fig. 4). In contrast,  $O_c$  density map within the first minimum of  $g_{CZ-O_c}(r)$  at 3.5 Å shows four high-density regions along the N-H bond directions.

In general, the non-polar compounds approach the arginine side chain from the directions normal to  $Gdm^+$  plane, while the polar compounds, *e.g.* water, carboxylates/carbonyls, form in-plane h-bonds<sup>40,41</sup> with  $Gdm^+$ . Our results reveal that in the lysozyme- $CaCO_3$  systems, the  $CO_3^{2-}$  ions replace the water molecules from the first coordination shell of arginine to form the TNTO coordination geometry. The second coordination shell is primarily formed by the  $Ca^{2+}$  ions with approximately three near-neighbor (NN) cations for each  $CO_3^{2-}$  (Fig. 6). The observed distribution of NN  $Ca^{2+}$  and  $CO_3^{2-}$  around arginine facilitates the formation of further layers of ions leading to their aggregation around the surface-exposed arginine side chains of lysozyme. In S1,2<sub>LC</sub>, the coordination number of water oxygen ( $O_w$ ) around arginine is less than that of in LW. The aggregation of ions around  $Gdm^+$  in lysozyme- $CaCO_3$  systems by expelling water molecules indicates that arginine residues act as nucleation sites for the formation of biominerals. The observed water expulsion at the protein-ACC interface is in-line with Freeman *et al.*<sup>18</sup>'s work that showed that the water population at the ACC-protein interface is much less than that at the calcite-protein interface.

**Distribution of  $Ca^{2+}$  and  $CO_3^{2-}$  ions around aspartate/glutamate.** The distribution of ions and water molecules around the negatively charged carbonyl groups of aspartate or glutamate side chains is entirely different from that of arginine side chain. In this case (Fig. 7), the  $Ca^{2+}$  ions in the first coordination shell are located mostly along the C=O bond and along the bisector vector of the angle OD1-CG-OD2. The density distribution along the bisector is similar to the most stable geometry



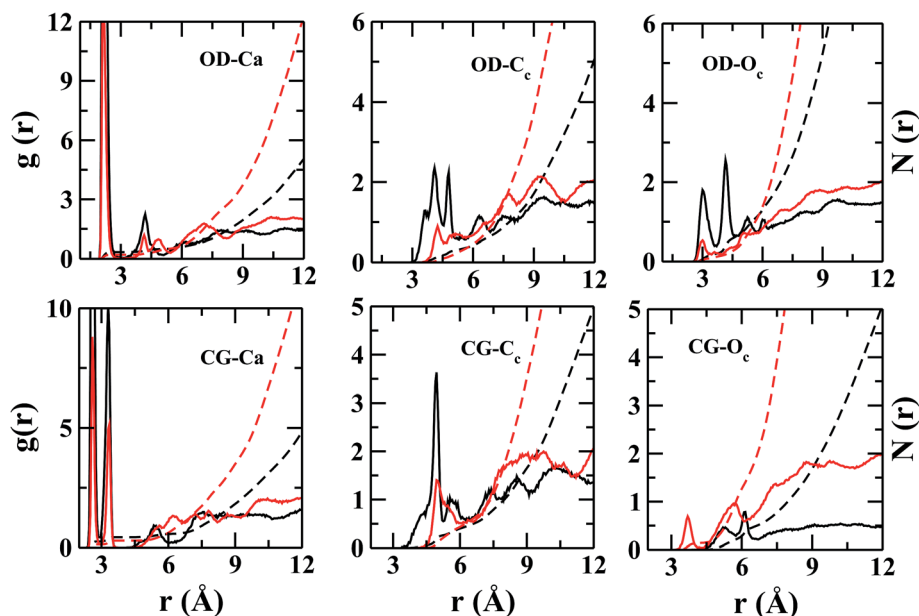


Fig. 5 Partial radial distribution functions ( $g(r)$ , solid line) and running coordination numbers ( $N(r)$ , dashed line) between the head groups of aspartate with ions in S1<sub>LC</sub> (black) and S2<sub>LC</sub> (red). Atom labels: carbonyl-oxygen (OD), carbonyl-carbon (CG), carbon of  $\text{CO}_3^{2-}$  ( $\text{C}_c$ ), oxygen of  $\text{CO}_3^{2-}$  ( $\text{O}_c$ ).

of isolated cluster<sup>25</sup> containing acetate ( $\text{AcO}^-$ ),  $\text{Ca}^{2+}$ , and  $\text{CO}_3^{2-}$ , where  $\text{Ca}^{2+}$  is coplanar to the  $\text{COO}^-$  plane and equidistant from the oxygen atoms. However, the cation population along the lone pairs of the oxygen was not observed in the  $\text{H}_2\text{O}$  containing clusters of acetate.<sup>25</sup> The ensemble averaged coordination number of  $\text{Ca}^{2+}$  around CG is  $\sim 0.9$  up to CG- $\text{Ca}^{2+}$  distance of 7.0

Å. Unlike arginine side chain, the first layer of ions around  $\text{COO}^-$  is surrounded by water molecules (Fig. 7).

The binding energies of  $\text{Ca}^{2+}\text{-CO}_3^{2-}$  and  $\text{Ca}^{2+}\text{-H}_2\text{O}$  complexes are  $\sim 520.8 \text{ kcal mol}^{-1}$  (ref. 25) and  $56.9 \text{ kcal mol}^{-1}$ ,<sup>42</sup> respectively. Among these two attractive interactions, the hydrophilic nature of calcium dominates which results in the

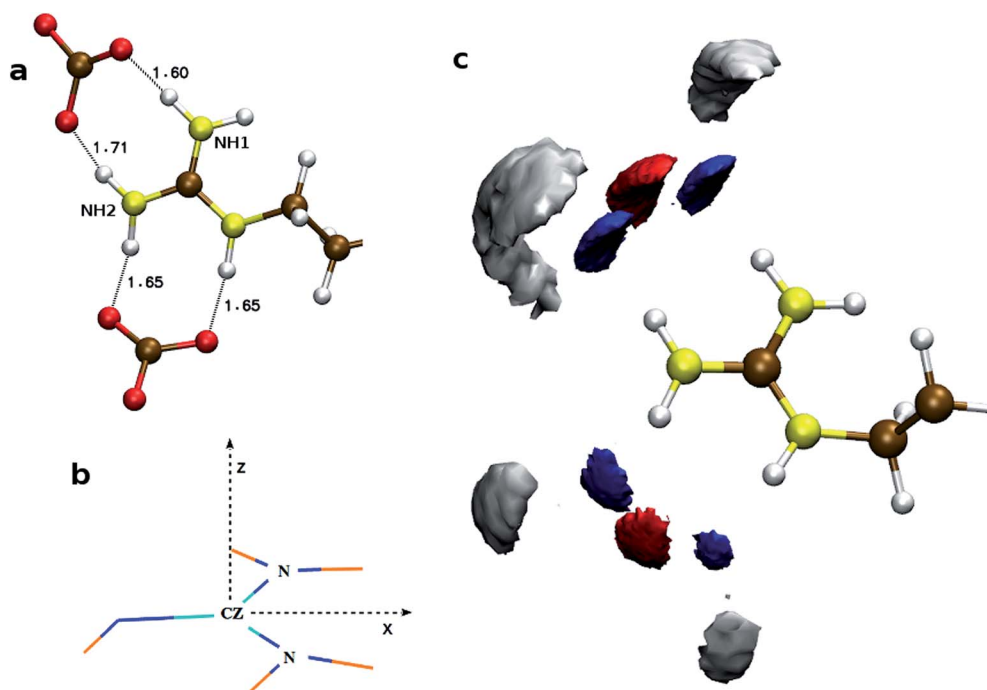


Fig. 6 Distributions of anions and cations around the side chain of arginine. (a) Snapshot of TNTO h-bonding with a pair of  $\text{CO}_3^{2-}$  ions. (b) For density map calculations, the sidechain is translated to bring CZ atom at the origin, z-axis is normal to the plane of  $\text{Gdm}^+$  and the bisector vector of angle  $\text{NH1-CZ-NH2}$  is along x-axis. (c) Probability density map of  $\text{C}_c$  (red),  $\text{O}_c$  (blue), and  $\text{Ca}^{2+}$  (gray) around arginine residue.



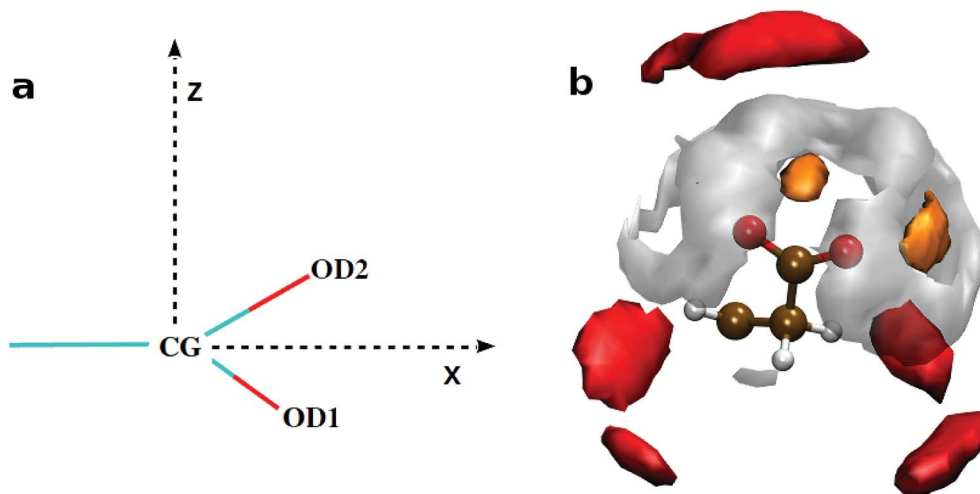


Fig. 7 Distributions of ions and water molecules around the side chains of aspartate or glutamate. (a) For density map calculations, the side chain is translated to bring CG (for aspartate) atom at the origin, z-axis is normal to the plane of OD1–CG–OD2 and the bisector vector of angle OD1–CG–OD2 is along x-axis. (b) Probability density map of  $\text{Ca}^{2+}$  (orange),  $\text{C}_c$  (red) and  $\text{O}_w$  (gray) around the aspartate residue. The distance cut off from CG to  $\text{Ca}^{2+}$ ,  $\text{C}_c$  and  $\text{O}_w$  are 3.5, 5.6 and 4.2 Å, respectively. Color scheme: carbon (Ochre), oxygen (red) and hydrogen (white).

separation of the  $\text{CO}_3^{2-}$  from its  $\text{Ca}^{2+}$  counterpart *via* a water layer (Fig. 7) in between. These water molecules are ‘tightly bound’ by both types of ions. As the life time of  $\text{Ca}$ – $\text{H}_2\text{O}$  h-bond is large, the possibility of breaking this h-bond within the simulation time limit is slim. Thus, the removal of water from the nearest hydration layer and aggregation of ions around aspartate/glutamate seems unlikely.

A convoluted picture of the distribution of ions and the ion density with reference to a given side chain residue are presented here by charge density plots (Fig. 8). The average charge density ( $\rho$ ) around arginine demonstrates sharp oscillations around the neutral axis,  $\rho = 0$ , up to  $\sim 7$  Å, indicating the formation of four alternating layers of cations and anions. However, the charge density around the  $\text{COO}^-$  bearing groups (aspartate/glutamate) shows only one prominent positive peak with a much weaker next layer of anions before the distribution merges to the neutral axis. Thus, we can conclude that arginine may play the major role in the nucleation and growth of  $\text{CaCO}_3$  crystal in lysozyme, whereas the other two residues may facilitate the protein to bind to the positively charged surface of the  $\text{CaCO}_3$  crystal.

### Hydrogen bond dynamics

Fig. 9 shows the calculated hydrogen bond time correlation functions,  $C_{\text{hb}}(t)$ , for  $\text{S1}_{\text{LC}}$  and  $\text{S2}_{\text{LC}}$  systems. The time correlation for water–arginine h-bonds ( $C_{\text{hb}}^w(t)$ ) decays much faster than that for  $\text{CO}_3^{2-}$ –arginine h-bonds ( $C_{\text{hb}}^c(t)$ ) in both the systems. The slow decay of the  $C_{\text{hb}}^c(t)$  implies that the  $\text{CO}_3^{2-}$  ions remain h-bonded for an extended period of time, whereas the relatively faster decay of  $C_{\text{hb}}^w(t)$  signifies frequent breaking and formation of h-bonds with water. The higher concentration of ions in  $\text{S2}_{\text{LC}}$  results in further slowing down of the decay of  $C_{\text{hb}}^c(t)$  suggesting that the ions are held tightly to the protein surface. These tightly bound  $\text{CO}_3^{2-}$  ions hinder water molecules

to establish strong h-bonds with arginines and thereby promote faster decay of  $C_{\text{hb}}^w(t)$  in  $\text{S2}_{\text{LC}}$ .

Each time correlation function was fitted with a multi-exponential function to obtain average time constant or relaxation time ( $\tau_{\text{hb}}$ , Table 3). The values of  $\tau_{\text{hb}}$  for  $\text{CO}_3^{2-}$  are orders of magnitude higher than those for water, which is indicative of higher residence times of  $\text{CO}_3^{2-}$  ions in the vicinity of the nucleation sites. The shorter relaxation times of ‘loosely bound’ water molecules may be necessary to facilitate the dehydration of the ACC patches on the protein surface.

### Free energies for the association of ions near arginine side chain

Free energy profiles were calculated to understand the energetics of ion association, growth of ion density, and the expulsion of water at the protein surface. As we have already studied the structure and energetics of ions in the first coordination shell of various side chain residues using radial distribution functions, density maps, and cluster analysis,<sup>25</sup> our focus here is on the energetics of higher coordination shells. Since the binding affinity between arginine and  $\text{CO}_3^{2-}$  depends on the thickness of the ionic-layer present on the protein surface, it is likely that the role of the ion–ion interactions would dominate over arginine– $\text{CO}_3^{2-}$  interactions in controlling the growth beyond some threshold thickness of ACC.

The calculated free energy profile estimates the energy cost to bring an anion or a  $\text{Ca}^{2+}$ – $\text{CO}_3^{2-}$  ion pair that is located away from the protein surface to the vicinity of arginine side chain. The separation ( $r_{\text{CV}}$ ) between the  $\text{C}_c$  atom of a distant  $\text{CO}_3^{2-}$  and the CZ atom of ARG61 was chosen as a collective variable (CV) to describe arginine– $\text{Ca}^{2+}$ – $\text{CO}_3^{2-}$  association. Fig. 10 shows the calculated free energy profile as a function of  $r_{\text{CV}}$  in the range from 6.0 Å to 20.0 Å.

In the initial configuration, the chosen  $\text{CO}_3^{2-}$  ion pairs with a neighboring  $\text{Ca}^{2+}$  to form an isolated  $\text{CaCO}_3$  unit surrounded by water. This particular  $\text{Ca}^{2+}$  ion was observed to be bound to



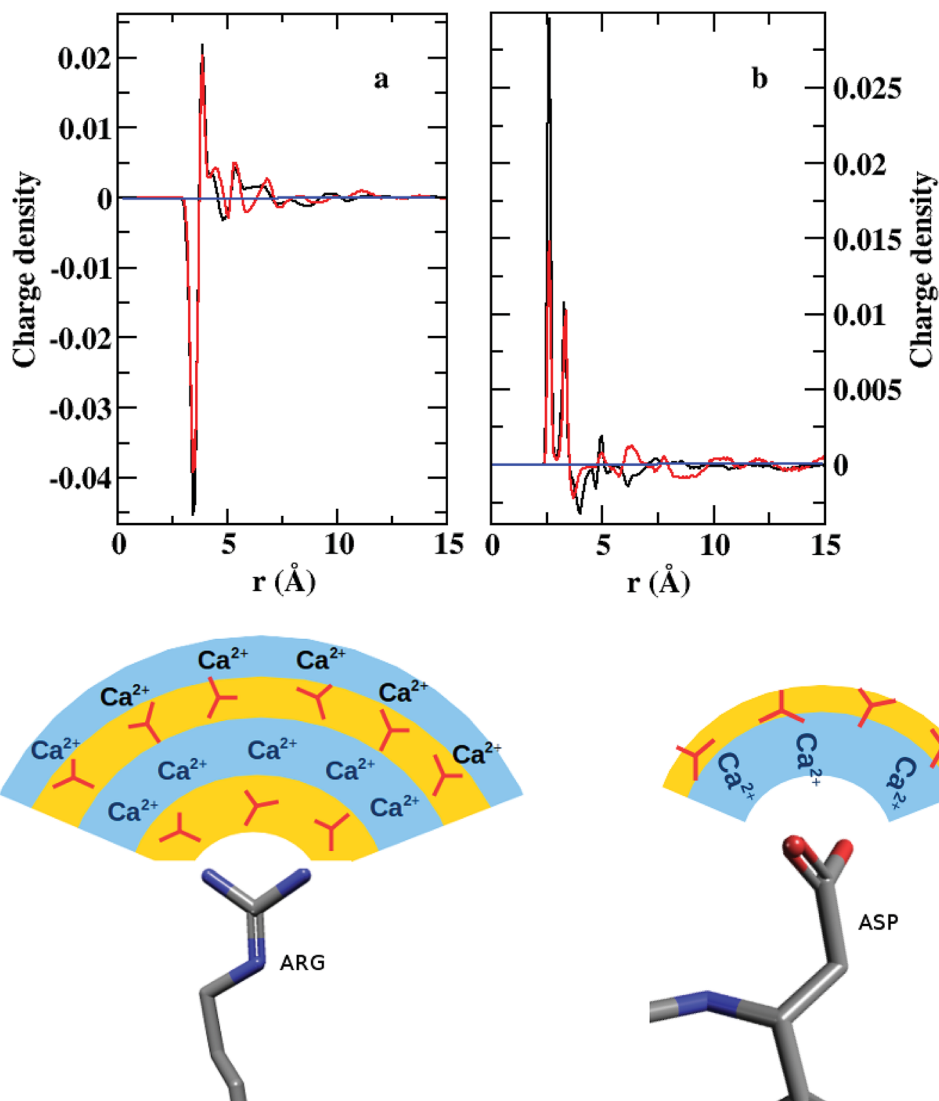


Fig. 8 (Up) Charge density ( $\rho$ ) around (a) CZ of arginine and (b) CG of aspartate in S1<sub>LC</sub> (black) and S2<sub>LC</sub> (red). (Down) Schematic representation of the ionic layers around arginine and aspartate. Color scheme for ball and stick representation: nitrogen (blue), carbon (gray), oxygen (red)

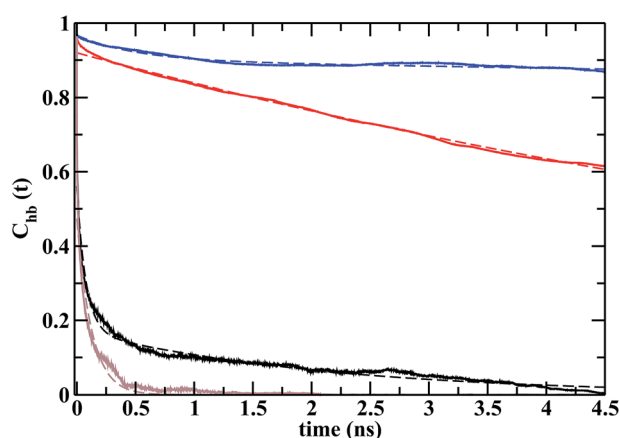


Fig. 9 Time correlation functions of the hydrogen bonds between the polar head groups of arginine with CO<sub>3</sub><sup>2-</sup> in S1<sub>LC</sub> (red) and S2<sub>LC</sub> (blue) and with water in S1<sub>LC</sub> (black) and S2<sub>LC</sub> (gray). Each plot was fitted with multiexponential function (dashed lines) to calculate h-bond relaxation time.

the selected CO<sub>3</sub><sup>2-</sup> throughout the range of  $r_{CV}$ . The free energy profile shows a shallow and broad minimum at  $r_{CV} = 11.4$  Å (Fig. 10a). The binding energy ( $\Delta F$ ) between the dissociated state at  $r_{CV} = 20$  Å and the global minimum is  $\sim 8.4$  kcal mol<sup>-1</sup>. This value compares well with the binding energy ( $\sim 10$  kcal mol<sup>-1</sup>) between OC-17 (a hen eggshell protein) and calcite surface, where arginine residues serve as the anchors to the crystal surface.<sup>18</sup> However, this barrier height ( $\Delta F$ ) is 4 kcal mol<sup>-1</sup> higher than that estimated for a single CO<sub>3</sub><sup>2-</sup>

Table 3 Hydrogen bond relaxation time ( $\tau_{hb}$ )

System	Function	$\tau_{hb}$ (ns)
S1 <sub>LC</sub>	$C_{hb}^c(t)$	10.10
	$C_{hb}^w(t)$	0.69
S2 <sub>LC</sub>	$C_{hb}^c(t)$	154.53
	$C_{hb}^w(t)$	0.12





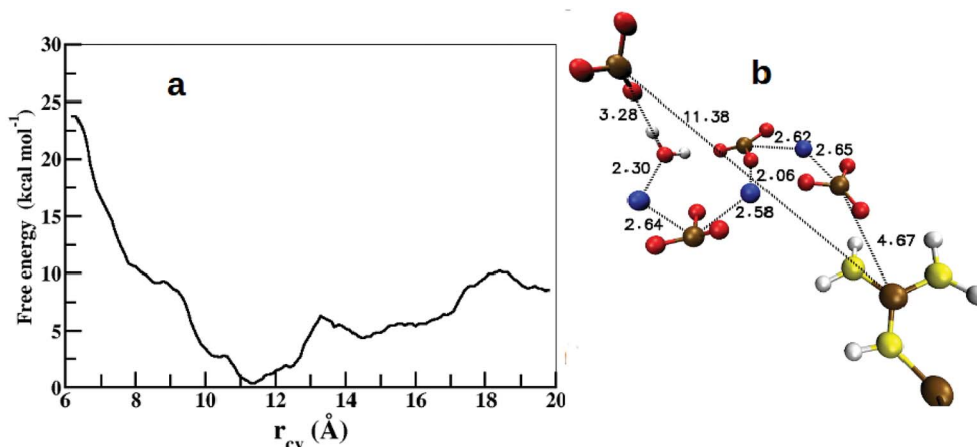


Fig. 10 (a) Free energy profile for the association of  $\text{CO}_3^{2-}$  around the arginine side chain along the reaction coordinate ( $r_{\text{cv}}$ ). (b) Snapshot from the global minimum of the free energy profile showing the arrangement of ions and water. Color scheme: same as in Fig. 6.

around a Gdm<sup>+</sup> functional group<sup>25</sup> and that between an isolated arginine residue and ACC,<sup>26</sup> which can be attributed to the inhomogeneous environment of the carbonate ion in the present study. The interactions between an anchoring agent (here, arginine) and the ions ( $\text{Ca}^{2+}$  and  $\text{CO}_3^{2-}$ ) play a crucial role in the initial stages of the ion aggregation, but further growth of ACC is generally governed mainly by the interactions between ions and water molecules. Thus, the barrier height reported here is an approximate estimate on the energetics for the growth of existing ion aggregates irrespective of the anchoring agent.

Further analysis of the metadynamics trajectory shows that the addition of this single  $\text{CO}_3^{2-}$  in the arginine environment increases the ion-cluster size by at least three more  $\text{CaCO}_3$  pairs *via* near-neighbor electrostatic interactions (Fig. 10b).

## Conclusions

In this work, we report protein-mediated formation of amorphous calcium carbonate (ACC) from the free ions using molecular dynamics (MD) simulations of a hen eggshell protein, lysozyme, in ionic solutions of calcium and carbonate. Our investigation provides comprehensive understanding on the role of selected side chain residues of lysozyme that facilitate the ion association to develop a dense amorphous phase of ions driven primarily by the strong electrostatic interactions between the functional group and its respective counter ions ( $\text{Ca}^{2+}$  or  $\text{CO}_3^{2-}$ ). The radial distribution functions (RDFs) among the ions and water in the ionic clusters around lysozyme are in good agreement with that in the hydrated ACC resembling a similar local environment in both the systems. Three key residues, arginine, glutamate, and aspartate, serve as potential nucleation sites for the crystallization of  $\text{CaCO}_3$ . However, the structure and energetics results indicate that the influence of arginine dominates over the rest, which could be a general scenario in all positively charged proteins, *e.g.* lysozyme, responsible for biomineralization.

Two model systems of lysozyme were generated with different ion concentrations, and both show essentially similar

results, apart from a denser ACC phase in the solution with higher concentration of ions. The three dimensional density profile of ions around arginine shows that, on the average, each arginine is coordinated by at least two  $\text{CO}_3^{2-}$  ions forming four h-bonds with the Gdm- $\text{NH}_2$ -groups making a twin-nitrogen-twin-oxygen (TNTO) geometry. This strong association with the nearest neighbors initializes the foundation for the successive alternating layers of positive and negative charges replacing the solvent water molecules. Moreover, the charge density plots also confirm this layering of charges suggesting that arginine can play a pivotal role in the nucleation of  $\text{CaCO}_3$  crystals. To elucidate the growth of ionic layers around a given arginine residue, we performed metadynamics calculations to estimate the required free energy for ion association in the higher coordination shell. The computed free energy profile shows a barrier height of  $\sim 8 \text{ kcal mol}^{-1}$  between the bound and unbound states of a  $\text{CO}_3^{2-}$  ion. We also observed that the addition of a single  $\text{CO}_3^{2-}$  may fetch few more  $\text{Ca}^{2+}$  and  $\text{CO}_3^{2-}$  pairs *via* near-neighbor electrostatic interactions and thus increases the size of the ion cluster around arginine. In contrast, the local environment around glutamate and aspartate retain the strongly bound solvent water through hydrophilic interactions with  $\text{Ca}^{2+}$  that inhibit the formation of next layer of  $\text{CO}_3^{2-}$  ions.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

This work was supported by the University Grants Commission-Basic Scientific Research (UGC-BSR) through Startup Grant. We thank Dr M. Krishnan for helpful discussions.

## References

- 1 Y. U. Gong, C. E. Killian, I. C. Olson, N. P. Appathurai, A. L. Amasino, M. C. Martin, L. J. Holt, F. H. Wilt and



- P. Gilbert, Phase transitions in biogenic amorphous calcium carbonate, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 6088–6093.
- 2 J. Aizenberg, G. Lambert, S. Weiner and L. Addadi, Factors involved in the formation of amorphous and crystalline calcium carbonate: a study of an ascidian skeleton, *J. Am. Chem. Soc.*, 2002, **124**, 32–39.
- 3 L. Addadi, S. Raz and S. Weiner, Taking advantage of disorder: amorphous calcium carbonate and its roles in biomineralization, *Adv. Mater.*, 2003, **15**, 959–970.
- 4 Y. Politi, T. Arad, E. Klein, S. Weiner and L. Addadi, Sea urchin spine calcite forms *via* a transient amorphous calcium carbonate phase, *Science*, 2004, **306**, 1161–1164.
- 5 R. Lakshminarayanan, J. S. Joseph, R. M. Kini and S. Valiyaveetil, Structure-Function Relationship of Avian Eggshell Matrix Proteins: A Comparative Study of Two Major Eggshell Matrix Proteins, Ansocalcin and OC-17, *Biomacromolecules*, 2005, **6**, 741–751.
- 6 R. Lakshminarayanan, X. J. Loh, S. Gayathri, S. Sindhu, Y. Banerjee, R. M. Kini and S. Valiyaveetil, Formation of transient amorphous calcium carbonate precursor in quail eggshell mineralization: an *in vitro* study, *Biomacromolecules*, 2006, **7**, 3202–3209.
- 7 A. Radha, T. Z. Forbes, C. E. Killian, P. Gilbert and A. Navrotsky, Transformation and crystallization energetics of synthetic and biogenic amorphous calcium carbonate, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 16438–16443.
- 8 J. H. Cartwright, A. G. Checa, J. D. Gale, D. Gebauer and C. I. Sainz-Díaz, Calcium carbonate polymorphism and its role in biomineralization: how many amorphous calcium carbonates are there?, *Angew. Chem., Int. Ed.*, 2012, **51**, 11960–11970.
- 9 G. Falini, S. Albeck, S. Weiner and L. Addadi, Control of aragonite or calcite polymorphism by mollusk shell macromolecules, *Science*, 1996, **271**, 67–69.
- 10 P. J. Smeets, K. R. Cho, R. G. Kempen, N. A. Sommerdijk and J. J. De Yoreo, Calcium carbonate nucleation driven by ion binding in a biomimetic matrix revealed by *in situ* electron microscopy, *Nat. Mater.*, 2015, **14**, 394–399.
- 11 P. J. M. Smeets, *Towards understanding pathway complexity in calcium carbonate mineralization*, PhD thesis, Technische Universiteit Eindhoven, 2016.
- 12 P. J. Smeets, A. R. Finney, W. J. Habraken, F. Nudelman, H. Friedrich, J. Laven, J. J. De Yoreo, P. M. Rodger and N. A. Sommerdijk, A classical view on nonclassical nucleation, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **5**, E7882E7890.
- 13 T. Mass, J. L. Drake, E. C. Peters, W. Jiang and P. G. Falkowski, Immunolocalization of skeletal matrix proteins in tissue and mineral of the coral *Stylophora pistillata*, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 12728–12733.
- 14 M. Reggi, S. Fermani, V. Landi, F. Sparla, E. Caroselli, F. Gizzi, Z. Dubinsky, O. Levy, J.-P. Cuif, Y. Dauphin, S. Goffredo and G. Falini, Biomineralization in Mediterranean corals: the role of the intraskeletal organic matrix, *Cryst. Growth Des.*, 2014, **14**, 4310–4320.
- 15 G. Falini, S. Fermani and S. Goffredo, Coral biomineralization: a focus on intra-skeletal organic matrix and calcification, *Semin. Cell Dev. Biol.*, 2015, 17–26.
- 16 T. Mass, A. J. Giuffre, C.-Y. Sun, C. A. Stifler, M. J. Frazier, M. Neder, N. Tamura, C. V. Stan, M. A. Marcus and P. U. Gilbert, Amorphous calcium carbonate particles form coral skeletons, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, E7670–E7678.
- 17 C. L. Freeman, J. H. Harding, D. Quigley and P. M. Rodger, How does an amorphous surface influence molecular binding?—ovocleidin-17 and amorphous calcium carbonate, *Phys. Chem. Chem. Phys.*, 2015, **17**, 17494–17500.
- 18 C. L. Freeman, J. H. Harding, D. Quigley and P. M. Rodger, Structural control of crystal nuclei by an eggshell protein, *Angew. Chem., Int. Ed.*, 2010, **49**, 5135–5137.
- 19 A. Dey, P. H. Bomans, F. A. Müller, J. Will, P. M. Frederik, G. de With and N. A. Sommerdijk, The role of prenucleation clusters in surface-induced calcium phosphate crystallization, *Nat. Mater.*, 2010, **9**, 1010.
- 20 D. Gebauer and H. Cölfen, Prenucleation clusters and non-classical nucleation, *Nano Today*, 2011, **6**, 564–584.
- 21 S. Raz, P. C. Hamilton, F. H. Wilt, S. Weiner and L. Addadi, The transient phase of amorphous calcium carbonate in sea urchin larval spicules: the involvement of proteins and magnesium ions in its formation and stabilization, *Adv. Funct. Mater.*, 2003, **13**, 480–486.
- 22 F. Tewes, O. L. Gobbo, C. Ehrhardt and A. M. Healy, Amorphous calcium carbonate based microparticles for peptide pulmonary delivery, *ACS Appl. Mater. Interfaces*, 2016, **8**, 1164–1175.
- 23 L. N. Hassani, F. Hindré, T. Beuvier, B. Calvignac, N. Lautram, A. Gibaud and F. Boury, Lysozyme encapsulation into nanostructured CaCO<sub>3</sub> microparticles using a supercritical CO<sub>2</sub> process and comparison with the normal route, *J. Mater. Chem. B*, 2013, **1**, 4011–4019.
- 24 A. Sergeeva, R. Sergeev, E. Lengert, A. Zakharevich, B. Parakhonskiy, D. Gorin, S. Sergeev and D. Volodkin, Composite magnetite and protein containing CaCO<sub>3</sub> crystals. External manipulation and vaterite calcite recrystallization-mediated release performance, *ACS Appl. Mater. Interfaces*, 2015, **7**, 21315–21325.
- 25 M. Saharay and R. J. Kirkpatrick, *Ab initio* and metadynamics studies on the role of essential functional groups in biomineralization of calcium carbonate and environmental situations, *Phys. Chem. Chem. Phys.*, 2014, **16**, 26843–26854.
- 26 R. Innocenti Malini, A. Finney, S. Hall, C. Freeman and J. Harding, The Water– Amorphous Calcium Carbonate Interface and Its Interactions with Amino Acids, *Cryst. Growth Des.*, 2017, **17**, 5811–5822.
- 27 S. Jo, T. Kim, V. G. Iyer and W. Im, CHARMM-GUI: a web-based graphical user interface for CHARMM, *J. Comput. Chem.*, 2008, **29**, 1859–1865.
- 28 B. Brooks, *et al.*, CHARMM: the biomolecular simulation program, *J. Comput. Chem.*, 2009, **30**, 1545–1614.
- 29 J. Lee, *et al.*, CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM



- simulations using the CHARMM36 additive force field, *J. Chem. Theory Comput.*, 2015, **12**, 405–413.
- 30 R. B. Best, X. Zhu, J. Shim, P. E. Lopes, J. Mittal, M. Feig and A. D. MacKerell Jr, Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone  $\phi$ ,  $\psi$  and side-chain  $\chi_1$  and  $\chi_2$  dihedral angles, *J. Chem. Theory Comput.*, 2012, **8**, 3257–3273.
- 31 M. Saharay, A. O. Yazaydin and R. J. Kirkpatrick, Dehydration-induced amorphous phases of calcium carbonate, *J. Phys. Chem. B*, 2013, **117**, 3328–3336.
- 32 J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, Scalable molecular dynamics with NAMD, *J. Comput. Chem.*, 2005, **26**, 1781–1802.
- 33 F. H. Stillinger, *Theory and molecular models for water. Advances in Chemical Physics: Non-Simple Liquids*, 1975, pp. 1–101.
- 34 A. Chandra, Effects of ion atmosphere on hydrogen-bond dynamics in aqueous electrolyte solutions, *Phys. Rev. Lett.*, 2000, **85**, 768.
- 35 S. Balasubramanian, S. Pal and B. Bagchi, Hydrogen-bond dynamics near a micellar surface: origin of the universal slow relaxation at complex aqueous interfaces, *Phys. Rev. Lett.*, 2002, **89**, 115505.
- 36 P. Raiteri, A. Laio, F. L. Gervasio, C. Micheletti and M. Parrinello, Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics, *J. Phys. Chem. B*, 2006, **110**, 3533–3539.
- 37 A. Barducci, G. Bussi and M. Parrinello, Well-tempered metadynamics: a smoothly converging and tunable free-energy method, *Phys. Rev. Lett.*, 2008, **100**, 020603.
- 38 A. Barducci, M. Bonomi and M. Parrinello, Metadynamics, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2011, **1**, 826–843.
- 39 W. Humphrey, A. Dalke and K. Schulten, VMD – Visual Molecular Dynamics, *J. Mol. Graphics*, 1996, **14**, 33–38.
- 40 C. T. Armstrong, P. E. Mason, J. R. Anderson and C. E. Dempsey, Arginine side chain interactions and the role of arginine as a gating charge carrier in voltage sensitive ion channels, *Sci. Rep.*, 2016, **6**, 21759.
- 41 T. Oroguchi and M. Nakasako, Influences of lone-pair electrons on directionality of hydrogen bonds formed by hydrophilic amino acid side chains in molecular dynamics simulation, *Sci. Rep.*, 2017, **7**, 15859.
- 42 M. Pavlov, P. E. Siegbahn and M. Sandström, Hydration of beryllium, magnesium, calcium, and zinc ions using density functional theory, *J. Phys. Chem. A*, 1998, **102**, 219–228.

