Calcium carbonate biomineralization: A theoretical and experimental investigation of biomolecular controls on nucleation and growth

Laura Mae Hamm

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Geochemistry

Patricia M. Dove, Committee Chair
J. Donald Rimstidt
Michael F. Hochella, Jr.
James J. De Yoreo

April 30, 2012
Blacksburg, Virgnia

Keywords: calcium carbonate, biomineralization, molecular dynamics, cation hydration, nucleation rate

Calcium carbonate biomineralization: A theoretical and experimental investigation of biomolecular controls on nucleation and growth

Laura M. Hamm

ABSTRACT

Organisms have evolved a remarkable ability to mineralize complex skeletons and functional biomaterials. These structures are nucleated and grown in close associaiton with macromolecular assemblages of proteins and polysaccharides that are implicated in regulating all stagees of mineralization. Because of this intimate association of organic with inorgaic components, many studies have investigated the effects of particular organic species on mineral morphology, phase, and growth rate. However, the diversity and species-specific nature of the organic assemblages associated with biominerals across a wide variety of taxa, has limited our understanding of how organisms use biomolecules to regulate skeletal formation. It is clear that a mechanistic picture of biomolecular controls on mineralization requires molecular-level investigations of the interplay between organic and inorganic components at all stages of crystallizaiton.

This dissertation presents the findings from theoretical and experimental studies of the physical mechanisms that underlie biomolecule controls on mineral formation. Molecular dynamics simulations probe the effects of acidic molecules on the hydration of alkaline earth cations. After first calculating baseline hydration properties for magnesium, calcium, strontium, and barium, I determine the effects of carboxylate-containing molecules on cation hydration state as well as the kinetics and thermodynamics of water exchange. Experimental work utilizes self-assembled monolayers as proxies for matrix macromolecules in order to understand their effects on $CaCO_3$ nucleation kinetics and thermodynamics. Estimates of nucleation rates and barriers are made from optical microscopy data and correlated with measurements of crystal – substrate rupture force from dynamic force microscopy.

These investigations show that an important function of biomolecules in directing mineralization lies in their ability to modulate cation hydration. Both chemical functionality and molecular conformation are influential in regulating the kinetics and thermodynamics of mineral nucleation, and these effects may be predicted by the strength of interaction between organic and inorganic components. These findings contribute to a mechanistic understanding of how organic matrices act to regulate biomineral formation. They demonstrate a plausible physical basis for how carboxyl-rich biomolecules accelerate the kinetics of biomineral growth and suggest roles for organic species in the nucleation and pre-nucleation stages of mineralization.

Table of Contents

List of Figures

List of Equations

List of Tables

List of Abbreviations

A	kinetic factor for nucleation
ACC	amorphous calcium carbonate
ACP	amorphous calcium phosphate
AFM	atomic force microscopy
Ala	alanine
Arg	arginine
Asn	asparagine
Asp	aspartic acid
Asx	aspartic acid or asparagine
a	species activity
B	slope
b	bond distance
C	normalization constant
$CaCl_2$	calcium chloride
$CaCO_3$	calcium carbonate
CHARMM	chemistry at Harvard macromolecular mechanics
CIP	contact ion pair
Cys	cysteine
DFM	dynamic force microscopy
EXAFS	extended X-ray adsorption fine structure
F	nuclei shape factor
$g(r)$	radial distribution function
ΔG^*	thermodynamic barrier to nucleation
Gln	glutamine
Glu	glutamic acid
Glx	glutamic acid or glutamine
Gly	glycine
HDDT	1,16-hexadecanedithiol
HPLC	high performance liquid chromatography
h	shape-specific constant
Ile	isolucine
J_0	steady-state nucleation rate
K	force constant (Chapter 2) or equilibrium constant (Chapter 3)
K_a	acid dissociation constant
K_{sp}	solubility product
k	rate constant
k_B	Boltzmann's constant
$k(t)$	normalized reactive flux
LAMMPS	large-scale atomic/molecular massively parallel simulator
LJ	Lennard Jones
MD	molecular dynamics

$MgCl_2$	magnesium chloride
MHA	16-mercaptohexadecanoic acid
MUPA	11-mercaptoundecylphosphoric acid
Na_2CO_3	sodium carbonate
NMR	nuclear magnetic resonance
NVT	number of particles, volume, and temperature
$n(r)$	integration of radial distribution function
$P(\ )$	probability distribution
$P^*(\ )$	umbrella sampling biased probability distribution
PMF	potential of mean force
Pro	proline
q	atomic charge
$R(t)$	time correlation function
r	separation distance
r^*	critical radius
SAM	self-assembled monolayer
SEM	scanning electron microscopy
Ser	serine
SSIP	solvent separated ion pair
t	time
T	temperature
Trp	tryptophan
TST	transition state theory
Tyr	tyrosine
U	potential energy
U^*	umbrella sampling biasing potential[*]
UB	Urey-Bradley
UDDT	1,11-undecanedithiol
W(r)	potential of mean force
WHAM	weighted histogram analysis method
XRD	X-Ray diffraction
α	interfacial energy
β	thermodynamic beta ($1/k_B$)
ε	potential well depth
θ	bond angle
$\theta(\ )$	Heaviside function
κ	transmission coefficient
μ	reduced mass
σ	interatomic distance of zero potential (Chapters 2-3) or supersaturation (Chapter 4)
τ	residence time
φ	improper torsion angle
χ	dihedral angle
ω	molecular volume

‡	transition state theory activated complex

Preface

This dissertation is presented in the form of four manuscripts. Chapter 1 is part of a book chapter, *Eukaryotic skeletal formation*, in press in Fundamentals of Geobiology, First Edition. Chapter 2 was publsihed in the *Journal of Physical Chemistry B* in 2010. Chapter 3 was published in part in 2011 in *Proceedings of the Materials Research Society* and the complete study is in review for the *Journal of Physical Chemistry B.* Chapter 4 is in preparation to be submitted for publication.

Chapter 1 The role of matrix proteins in biogenic calcium carbonate formation

1.1 Introduction

The key factor that distinguishes biogenic materials from their synthetic counterparts is the presence of organic matrix macromolecules. All biologically mineralized tissues are associated with and often intimately linked to some group of organic components that have been assembled by specialized cells within the organism. These macromolecules may be instrumental in a myriad of functions related to the formation of the inorganic material - for decades they have been implicated in cell adhesion, ion transport, crystal nucleation, and crystal growth and phase regulation[1]. The fundamental importance of matrix macromolecules was first insinuated with the realization that even though each mineralized tissue possesses a unique organic assemblage, some key chemical aspects of macromolecules are shared across taxa – namely acidic carboxylate moieties and, to a lesser extent, phosphate and sulfate groups[2]. This discovery has led to a thorough investigation of the roles of matrix macromolecules in biomineral formation, although the specific and individual roles of organic components remain elusive.

This section will focus on the functions of matrix macromolecules in the biomineralization of calcium carbonate ($CaCO_3$), an extensively studied and naturally abundant biomineral. Calcifying organisms fabricate elaborate matrices responsible for mineral nucleation and growth, often resulting in surprising crystal phases and morphologies at ambient pressures and temperatures. **Table 1.1** presents a sumary of summary of characteristics of some of the proteins that have been extracted from $CaCO_3$ biominerals. Clarification of the roles played by organic macromolecules will be vital in a complete understanding $CaCO_3$ biomineral formation.

1.2 Soluble and Insoluble Matrix Components

Biogenic $CaCO_3$ tends to be associated with a two-part organic matrix. Quantitatively, the main constituents are usually highly cross-linked, insoluble, and slightly hydrophobic macromolecules; these entities form the structural, three-dimensional framework for mineralization and are ultimately less intimately associated with the resultant biomineral than

Table 1.1. Characteristics of selected matrix proteins extracted from biogenic $CaCO_3$

Organism	Associated Mineral	Protein Name	Solubility	Protein Description	Ref.
Pinctada fucata (pearl oyster)	calcite	Prismalin-14	insoluble	Pro/Ile/Tyr/Arg regions, Gly/Trp rich region, Asp rich region	(3)
Crassostrea virginica (American oyster)	calcite	none	soluble	highly phosphorylated, Asp/Ser/Gly rich	(4)
Pinna nobilis (Mediterranean fan mussel)	calcite	calprismin	soluble	Ala/Asx/Thr/Pro rich, Cys rich terminal sequence	(5)
	calcite	caspartin	soluble	Asx rich	
Haliotis rufescens (Pacific red abalone)	aragonite	AP7	soluble	highly acidic, Asp/Gly rich	(6)
Pyura pachydermatina (New Zealand sea squirt)	ACC	none	soluble	highly acidic, Asx/Glx/Ser/Gly rich	(7)

their soluble counterparts[8]. Besides providing the physical arena for mineralization, these 'framework macromolecules' may also act as a substrate from which other proteins can interact with the mineral phase.

The second set of matrix elements is made up of soluble proteins. They may be unbound, adsorbed onto a matrix substrate, or occluded within a growing mineral. Soluble matrix macromolecules can be grouped roughly into three main categories based on their primary structures and biologic occurances:

1) Aspartic acid (Asp)/asparagine (Asn) rich proteins and glycoproteins – accociated with crystalline $CaCO_3$ phases
2) Glutamic acid (Glu)/glutamine (Gln) and sering rich glycoproteins – associated with amorphous $CaCO_3$ (ACC)
3) Polysaccharide rich macromolecules – not unusually acidc; found in biogenic $CaCO_3$ of many organisms

Categories 1 and 2 are both made up of unusually acidic macromolecules due to the abundance of carboxylate functional groups. The high negative charge associated with these proteins and

glycoproteins makes them good candidates to facilitate interactions at charged crystal faces as well as to accumulate calcium cation at nucleation sites; for these reasons, the acidic macromolecules are presumed to be most influential in $CaCO_3$ nucleation and growth. The soluble matrix macromolecules have been shown to exert remarkable control over crystal nucleation, growth, polymorph, and morphology, even when separated from the rest of the matrix[9-11]. When studying biomineralization processes, however, it is important to consider the matrix in its entirety; several studies have demonstrated the importance of the interplay of framework macromolecules with acidic proteins and mineral phases[12-15].

1.2.1 Characterization of Matrix Macromolecules

A significant challenge in elucidating the individual roles of matrix macromolecules lies in the task of protein characterization. Matrix proteins are typically separated by gel electrophoresis under denaturing conditions or a combination of HPLC and ion exchange after extraction from the demineralized organism[16]. Proteins that are not highly acidic are readily stained with Coomassie blue or silver; sequences for many proteins of this type have been reported[17]. The more interesting acidic proteins, however, have proved difficult to extract and purify while keeping their structure and function intact due to high charge and close association with the mineral phase[4]. Recently, new electrophoretic strategies and the application of cDNA sequencing have allowed the characterization of some new acidic matrix proteins[18, 19]. Further borrowing of molecular genetic techniques may be useful in identification of more matrix elements, an important step in understanding the role of individual proteins and chemical functionalities.

1.3 Overview of $CaCO_3$ Biomineral Phases

Calcium carbonate biominerals commonly occur as one of the two most thermodynamically stable crystalline polymorphs - calcite and aragonite. Aragonite is metastable compared to calcite and differs slightly in crystallographic structure. Vaterite, a third and more metastable crystalline phase, is rare in nature but not unheard of; it has been observed in Ascidian spicules and teleost fish otoliths[20-22]. Synthetic (non-biogenic) examples of the three most common crystalline $CaCO_3$ phases are shown in **Figure 1.1**. Non-crystalline $CaCO_3$, ACC, has recently been recognized as an important crystalline precursor phase in many biomineral

systems[23, 24] but may also be stabilized for long periods of time in organisms including sea squirts and sponges[25, 26]. Organic matrix molecules appear to be instrumental in the nucleation and stabilization of particular $CaCO_3$ polymorphs, but the mechanism by which organisms exercise such strict control over phase is not well understood.

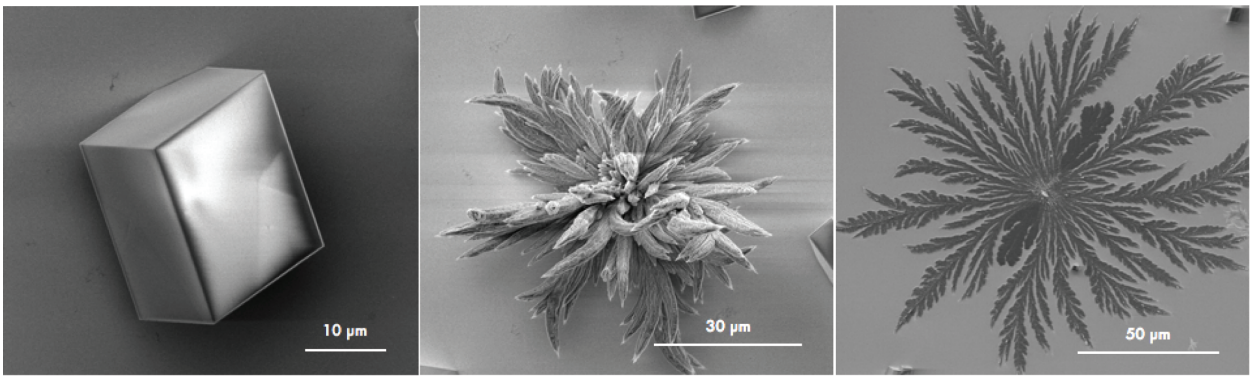


Figure 1.1. Scanning electron micrographs showing synthitic examples of the three most common crystalline $CaCO_3$ phases: calcite (left), aragonite (center), and vaterite (right)

1.4 Modes of regulation

A wide variety of experimental techniques have been applied to examine the role of biomoleclues in $CaCO_3$ biomineralization from the angstrom level to the macroscale. The roles of organic matrices are often inferred from structural observations obtained with scanning and transmission electron microscopy. Calcium carbonate is easily grown from solution or deposited on ordered substrates which mimic the framework macromolecules seen in nature[27, 28]. Countless additives, ranging from synthetic polymers[29-31] to soluble macromolecules extracted from the organism[32-34], have been tested in both solution-based and templated growth. *In situ* experiments based on atomic force microscopy[35, 36] as well X-Ray and neutron scattering methods[37-39] have been key in uncovering the dynamics of macromolecular control over $CaCO_3$ nucleation and growth. The following paragraphs discuss two of the earliest recognized modes by which matrix molecules influence biomineral formation.

1.4.1 Electrostatic Accumulation by Biomolecules

An obvious function of negatively charged biomolecules in $CaCO_3$ mineralization would be to electrostatically attract calcium ions to the site of crystallization. This effect acts to concentrate ions at specific nucleation sites on the matrix, inducing crystal nucleation by

increasing local supersaturation[40, 41]. Matix macromolecules are known to bind Ca^{2+} with stoichiometries exceeding the number of potential anionic binding sites on the matrix; this phenomenon can be explained by local anion binding and secondary calcium binding triggered by the initial binding of calcium to the protein, an effect that has been demonstrated in biomineral systems[42].

The specific chemical interaction of calcium with protein functional groups is an important consideration when thinking about the role of a certain matrix protein[43]. To induce crystal nucleation, a molecule must have a high capacity to attract ions but a relatively low binding affinity; strong binding of calcium would likely inhibit nucleation by rendering the ions unreactive (and some matrix proteins may be assembled specifically for this purpose). In general, calcium binding to carboxylate groups is cooperative and involves at least two or three ligands[44]. It has been suggested that amino acid sequences resembling Asx-X-Asx, where X is a neutral residue, may provide an optimal binding configuration for calcium in order to induce nucleation[45]. These acidic sequences are common in molluscan matrix proteins.

1.4.2 Stereochemical Matching by Biominerals

Biomineral morphology is greatly affected by the both physical constraints of the framework macromolecules and the chemical interactions of soluble proteins with the mineral phase. The stereochemical recognition model provided the first mechanistic explanation of chemical control over crystal shape by proteins. In this model, specific morphologies are stabilized by the binding of proteins to otherwise unstable crystal faces due to stereochemical matching to the crystal lattice, which lowers facial surface energies[46]. Such a model is plausible due to the inherently repeating structures of proteins (polymers of amino acids) and crystal lattices (repeating unit cells). These characteristics allow a possible matching of certain protein side chains with periodic groups exposed at a crystal face[47-50]. Recent work has modified the stereochemical recognition model by demonstrating that peptides and proteins control morphology by altering the free energy of crystal step-edges, the areas where new molecules attach to a growing crystal[35, 51].

1.5 Mollusk as example

The mollusk provides a well studied and intriguing example of an organism that utilizes an ornate organic matrix to exercise strict control over $CaCO_3$ nucleation and growth. Many

species of mollusk are able to deposit both calcite and aragonite through the synthesis of phase-controlling proteins[9, 14], which are secreted by the mantle epithelial cells[52]. A schematic of the mollusk shell and cross-sectional scanning electron micrographs showing the relationship between the two deposited phases is shown in **Figure 1.2.** The following sections review the strategies employed by mollusks to mineralize the nacreous and prismatic shell layers.

1.5.1 Nacreous Layer

Mollusk nacre consists of uniform layers of aragonite single crystal tablets interspaced by layers of organic material[53]. The polysaccharide β-chitin and mildly hydrophobic silk-like proteins rich in Ala and Gly comprise the framework portion of the organic matrix while the soluble fraction is an asseblage of hydrophilic, highly acidic glycoproteins in β-sheet conformation[14, 54, 55]. The highly ordered chitin acts orient the aragonite single crystals and physically limit their size and shape. Histochemical analysis shows that sites of aragonite nucleation are rich in both the carboxylate and sulfate groups of the soluble proteins, suggesting that the two functionalities work cooperatively to direct nucleation[55].

Recent investigations have demonstrated that the silk framework proteins exist in a gel-like state that pre-fills the matrix space to be mineralized[14, 53, 56]. The presence of a gel in which $CaCO_3$ is mineralized forces us to re-think basic concepts in biomineral formation as they are founded on the belief that mineralization occurs from supersaturated solution; more studies on non-solution based crystallization will be necessary to fully understand mollusk nacre formation.

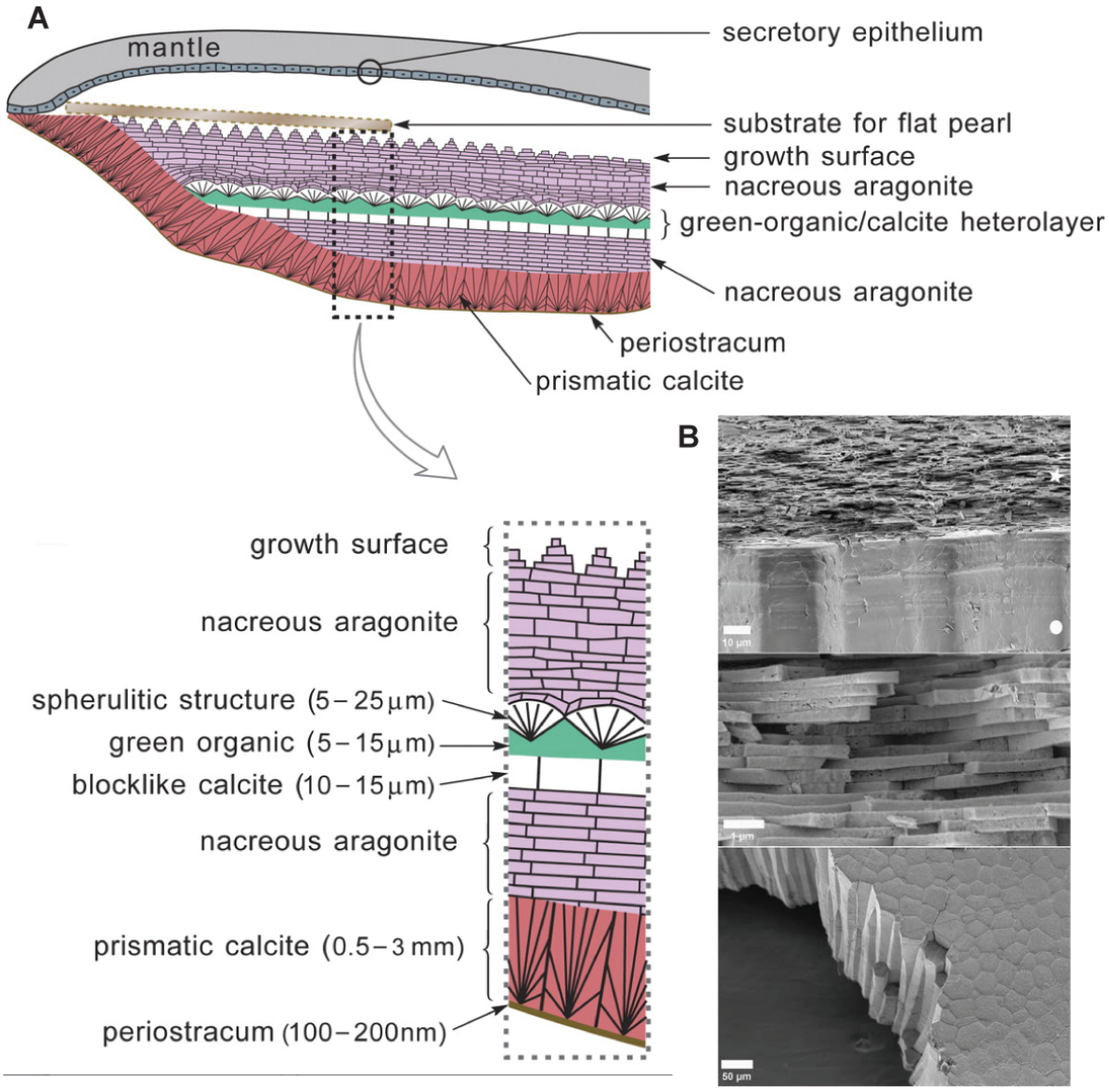


Figure 1.2. A) Schematic of the shell and mantle of the red abalone (*Atrina rigida*) shows the spatial relationship of the two deposited $CaCO_3$ phases. The mineral layers are separated by layers organic species. B) Cross-sectional SEM images of an *Atrina rigida* shell showing the nacreous (white star) and prismatic (white circle) layers (top image), the aragonite tablets of the nacreous layer (middle image), and the calcite prisims of the prismatic layer (bottom image).

1.5.2 Prismatic Layer

The calcite prismatic crystals are each surrounded by a relatively thick layer of organic matrix[57, 58]. β-chitin and Gly-rich proteins are present as framework macromolecules. Although these are similar to those found in nacre, the structures and locations within the mineral differ significantly – prism chitin is highly disordered and no gel-like phase has been observed. The soluble matrix macromolecules are highly acidic (even more so than those in nacre) and associated with both the chitin and the mineral phase; several of these unusually acidic proteins have been isolated and characterized in last ten years[3, 15, 18]. The processes involved in prismatic layer formation have not been studied as thoroughly as in the nacreous layer.

1.6 Frontiers

The importance of organic components in $CaCO_3$ biomineralization has long been recognized. Research in a variety of fields now aims to identify the specific roles of individual matrix macromolecules, understand protein-protein and protein-mineral interactions, and develop of mechanistic understanding of biomineral formation. A recent study on protein-mediated eggshell mineralization recognizes that protein conformational changes initiated by calcium binding may be important in crystallization[39]. Protein conformation switching may be consequential and commonly overlooked; the continued development of *in situ* techniques for monitoring mineralization will help advance our understanding of these processes. Maturity of *in situ* techniques will assist in answering another question that has recently been raised: Are soluble matrix proteins synthesized from scratch or modified *in vivo* to perform new functions as an organism develops?

Investigation of cellular signaling in biomineral development would also benefit the field. A study by Fritz and colleagues showed that severing communication between mollusk epithelial cells and the mineral phase triggers a sequence of protein syntheses resulting in the deposition of calcite prism followed by aragonite nacre[59]. It would be interesting to understand how cells communicate to activate matrix deposition and crystal nucleation.

1.7 References

(1) Addadi, L. W., S.; Geva, M. *Z. Kardiol.* **2001,** *90*, 92-98.
(2) Weiner, S., Troub, W., and Lowenstam, H.A., *Biomineralization and Biological Metal Accumulation*. Reidel Publishing Company: Dordrecht, 1983.
(3) Suzuki, M.; Murayama, E.; Inoue, H.; Ozaki, N.; Tohse, H.; Kogure, T.; Nagasawa, H. *Biochem. J.* **2004,** *382*, 205-213.
(4) Borbas, J. E.; Wheeler, A. P.; Sikes, C. S. *J. Exp. Zool.* **1991,** *258*, 1-13.
(5) Marin, F.; Amons, R.; Guichard, N.; Stigter, M.; Hecker, A.; Luquet, G.; Layrolle, P.; Alcaraz, G.; Riondet, C.; Westbroek, P. *J. Biol. Chem.* **2005,** *280*, 33895-33908.
(6) Kim, I.L.; Morse, D.E.; Evans, J.S. *Langmuir* **2004**, 20(26*)*, 11664-11673.
(7) Aizenberg, J.; Lambert, G.; Addadi, L.; Weiner, S. *Adv. Mater.* **1996**, 8(*3*), 22-226.
(8) Meldrum, F.C. *Int. Mater. Rev.* **2003,** *48*, 187-224.
(9) Belcher, A. M.; Wu, X. H.; Christensen, R. J.; Hansma, P. K.; Stucky, G. D.; Morse, D. E. *Nature* **1996,** *381*, 56-58.
(10) Thompson, J. B.; Paloczi, G. T.; Kindt, J. H.; Michenfelder, M.; Smith, B. L.; Stucky, G.; Morse, D.E.; Hansma, P. K. *Biophys. J.* **2000,** *79*, 3307-3312.
(11) Raz, S., Hamilton, P.C., Wilt, F.H., Weiner, S. and L. Addadi *Adv. Funct. Mater.* **2003,** *13*, 480-486.
(12) Albeck, S., Aizenberg, J., Addadi, L. and S. Weiner *J. Am. Chem. Soc.* **1993,** *115*, 11691-11697.
(13) Falini, G.; Weiner, S.; Addadi, L. *Calcified Tissue Int.* **2003,** *72*, 548-554.
(14) Addadi, L.; Joester, D.; Nudelman, F.; Weiner, S. *Chem. Eur. J.* **2006,** *12*, 981-987.
(15) Nudelman, F.; Chen, H. H.; Goldberg, H. A.; Weiner, S.; Addadi, L. *Farad. Discuss.* **2007,** *136*, 9-25.
(16) Cariolou, M. A. *J. Comp. Physiol.* **1988,** *157*, 717-729.
(17) Sarashina, I.; Endo, K. *Mar Biotechnol.* **2001,** *3*, 362-369.
(18) Tsukamoto, D.; Sarashina, I.; Endo, K. *Biochem. Biophys. Res. Commun.* **2004,** *320*, 1175-1180.
(19) Gotliv, B. A.; Kessler, N.; Sumerel, J. L.; Morse, D. E.; Tuross, N.; Addadi, L.; Weiner, S. *Chembiochem.* **2005,** *6*, 304-314.
(20) Falini, G., Fermani, S., Vanzo, S., Miletic, M. and G. Zaffino *Eur. J. Inorg. Chem.* **2005**, 162-167.
(21) Sollner, C.; Burghammer, M.; Busch-Nentwich, E.; Berger, J.; Schwarz, H.; Riekel, C.; Nicolson, T. *Science* **2003,** *302*, 282-286.
(22) Lowenstam, H. A.; Abbott, D. P. *Science* **1975,** *188*, 363-365.
(23) Olszta, M. J.; Odom, D. J.; Douglas, E. P.; Gower, L. B. *Connect. Tissue Res.* **2003,** *44*, 326-334.
(24) Colfen, H.; Mann, S. *Angew. Chem. Int. Ed. Engl.* **2003,** *42*, 2350-2365.
(25) Weiner, S.; Levi-Kalisman, Y.; Raz, S.; Addadi, L. *Connect. Tissue. Res.* **2003,** *44*, 214-218.
(26) Aizenberg, J.; Lambert, G.; Weiner, S.; Addadi, L. *J. Am. Chem. Soc.* **2002,** *124*, 32-39.
(27) Mann. S., D., J.M., Sanderson, N.P., Heywood, B.R. and E.J. Aso Samper *J. Chem. Soc. Faraday Trans.* **1990,** *86*, 1873-1880.
(28) Falini, G.; Fermani, S.; Gazzano, M.; Ripamonti, A. *J. Chem. Soc., Dalton Trans.* **2000**, 3983-3987.

(29) Wang, T.; Colfen, H.; Antonietti, M. *J. Am. Chem. Soc.* **2005,** *127*, 3246-3247.
(30) Jada, A.; Jradi, K. *Macromol. Symp.* **2006,** *233*, 147-151.
(31) Naka, K. *Top. Curr. Chem.* **2007,** *271*, 119-154.
(32) Wheeler, A. P.; George, J. W.; Evans, C. A. *Science* **1981,** *212*, 1397-1398.
(33) Walters, D. A.; Smith, B. L.; Belcher, A. M.; Paloczi, G. T.; Stucky, G. D.; Morse, D. E.; Hansma, P. K. *Biophys. J.* **1997,** *72*, 1425-1433.
(34) Kim, I. W.; Darragh, M. R.; Orme, C.; Evans, J. S. *Cryst. Growth. Des.* **2006,** *6*, 5-10.
(35) De Yoreo, J. J.; Dove, P. M. *Science* **2004,** *306*, 1301-1302.
(36) Elhadj, S.; De Yoreo, J. J.; Hoyer, J. R.; Dove, P. M. *Proc. Natl. Acad. Sci. USA* **2006,** *103*, 19237-19242.
(37) Pontoni, D. P., Bolze, J., Dingenouts, N., Narayanan, T. and M. Ballauff *J. Phys. Chem. B* **2003,** *107*, 5123-5125.
(38) DiMasi, E.; Kwak, S. Y.; Amos, F. F.; Olszta, M. J.; Lush, D.; Gower, L. B. *Phys. Rev. Lett.* **2006,** *97*, 045503.
(39) Pipich, V.; Balz, M.; Wolf, S. E.; Tremel, W.; Schwahn, D. *J. Am. Chem. Soc.* **2008,** *130*, 6879-6892.
(40) Greenfield, E. M., Wilson, D.C. and M.A. Crenshaw *Amer. Zool.* **1984,** *24*, 925-932.
(41) Mann, S. *Nature* **1988,** *332*, 119-124.
(42) Lee, S. L., Glonek, T. and M.J. Glimcher *J. Calcif. Tissue Int.* **1983,** *35*, 815-818.
(43) Aizenberg, J.; Black, A. J.; Whitesides, G. M. *Nature* **1999,** *398*, 495-498.
(44) Kretsinger, R. H. *Ann. Rev. Biochem.* **1976,** *45*, 239-266.
(45) Weiner, S.; Hood, L. *Science* **1975,** *190*, 987-989.
(46) Addadi, L.; Weiner, S. *Proc. Natl. Acad. Sci. USA* **1985,** *82*, 4110-4114.
(47) Addadi, L.; Moradian, J.; Shay, E.; Maroudas, N. G.; Weiner, S. *Proc. Natl. Acad. Sci. USA* **1987,** *84*, 2732-2736.
(48) DeOliveira, D. B.; Laursen, R.A. *J. Am. Chem. Soc.* **1997**, 10627-10631.
(49) Gilbert, P. U. P. A.; Abrecht, M.; Frazer, B.H. *Rev. Mineral Geochem.* **2005,** *59*, 157-185.
(50) Wolf, S. E.; Loges, N.; Mathiasch, B.; Panthofer, M.; Mey, I.; Janshoff, A.; Tremel, W. *Angew. Chem. Int. Ed. Engl.* **2007,** *46*, 5618-5623.
(51) Orme, C. A.; Noy, A.; Wierzbicki, A.; McBride, M. T.; Grantham, M.; Teng, H. H.; Dove, P. M.; DeYoreo, J. J. *Nature* **2001,** *411*, 775-779.
(52) Marin, F.; Luquet, G. *C. R. Palevol.* **2004,** *3*, 469-492.
(53) Nudelman, F.; Shimoni, E.; Klein, E.; Rousseau, M.; Bourrat, X.; Lopez, E.; Addadi, L.; Weiner, S. *J. Struct. Biol.* **2008,** *162*, 290-300.
(54) Crenshaw, M. A. *Biomineral. Res. Rep.* **1972,** *6*, 6-11.
(55) Nudelman, F.; Gotliv, B. A.; Addadi, L.; Weiner, S. *J. Struct. Bio.* **2006,** *153*, 176-187.
(56) Levi-Kalisman, Y.; Falini, G.; Addadi, L.; Weiner, S. *J. Struct. Biol.* **2001,** *135*, 8-17.
(57) Nakahara, H.; Bevelander, G. *Calcif. Tissue Res.* **1971,** *7*, 31-45.
(58) Gregoire, C. *J. Biophys. Biochem. Cytol.* **1961,** *9*, 395-400.
(59) Fritz, M.; Belcher, A.M.; Radmacher, M.; Walters, D.A.; Hansma, P.K.; Stucky, G.D.; Morse, D.E; Mann, S. *Nature* **1994,** *371*, 49-51.

Chapter 2 Molecular dynamics of ion hydration in the presence of small carboxylated molecules and implications for calcification

Abstract: The aspartate-rich macromolecules found at nucleation sites of calcifying organisms are widely implicated in regulating biomineral formation. Anecdotal evidence suggests their ability to influence the onset and composition of calcified structures may arise from effects on ion hydration. This study investigates the interactions of acidic amino acids and dipeptides with hydrated cations using molecular dynamics (MD). By monitoring the hydration states of Mg^{2+}, Ca^{2+} and Sr^{2+} during their approach to negatively charged molecules, we show that carboxylate moieties of Asp promote dehydration of Ca^{2+} and Sr^{2+}. A contact ion pair (CIP) is not required to disrupt cation hydration and we demonstrate that reductions and rearrangements of first shell water can begin at ion-Asp separation distances as large as ~4.9 Å for Ca^{2+} and ~5.1 Å for Sr^{2+}. CIP formation between Ca^{2+} and Sr^{2+} and carboxylate groups decreases the total first shell coordination number from an average of 8.0 and 8.4 in bulk water to 7.5 and 8.0, respectively. The energy barrier to physically replacing waters about Ca^{2+} with carboxylate oxygen atoms is small (~ 2 kcal/mol) compared to a somewhat larger barrier for Sr^{2+} (~ 4 kcal/mol). This may be explained by differences in the strength of Coulombic interactions between the cations and Asp, resulting in different paths of approach towards Asp for Ca^{2+} and Sr^{2+}. In contrast, the primary solvation shell of Mg^{2+} remains largely unchanged during interactions with Asp until the abrupt physical replacement of water by carboxylate oxygen atoms, which comes at a high energetic cost. These insights support the claim that carboxylated biomolecules increase the growth rate of calcite by lowering the energy barrier to Ca^{2+} dehydration. The findings also suggest a physical basis for the idea that ion-specific behaviors of Ca^{2+} and Mg^{2+} in cellular systems arise from a critical balance between water binding in the ion hydration shells versus their interactions with ligands present in intracellular environments.

2.1 Background

Calcified tissues are comprised of biogenic minerals that nucleate and grow in close association with an organic matrix of proteins and other macromolecules. These molecules are widely believed to have a myriad of regulatory functions in different stages of biomineral formation, including crystal nucleation [1, 2], crystal growth [3-5], and the stabilization of amorphous intermediate phases [6-8]. Matrix biomolecules are also thought to influence the mineral phase [2, 9, 10] and morphology [4, 11-13] that forms. Studies of calcification across many taxa report that the macromolecules are characterized by an abundance of carboxylated amino acids and carbohydrate species[14]. This unique association of biopolymer and mineral has motivated an extensive effort to determine the physical basis by which the compositions and conformations of acidic molecules modulate the onset of different calcified phases and morphologies [2, 15-17]. Although the findings of many *in vitro* and *in vivo* correlations provide important clues, the mechanistic basis for how organic components modulate mineralization is not well understood.

Anecdotal evidence from a number of studies suggests that a critical role of macromolecules in biomineralization may reside in their ability to modify solvent properties in the local environment where mineral nucleation and growth occurs. Biochemists have long-recognized that solvation properties influence the ability of Ca^{2+} and Mg^{2+} to activate a variety of cellular functions[18, 19]. For example, 'ion antagonism', a phenomenon observed in many biological systems, refers to the ability of one inorganic ion to overcome the inhibitory or stimulatory effects of another[20, 21]. To explore the probable role of solvation in modulating biomineral nucleation and growth, a number of *in vitro* studies used simple alcohol-water mixtures to mimic the effects of organics on local water structure and indeed documented that these altered solvation environments influence crystal phase[22, 23], morphology[23, 24] and growth kinetics[25].

A recent experimental study of calcification processes proposed the solvent-modifying effects of biomolecules as the explanation for why amino acids, peptides and proteins accelerate the growth rate of calcite in proportion to their hydrophilicity[26]. A subsequent investigation found significant enhancement of Mg^{2+} uptake during calcite growth in the presence of hydrophilic polypeptides of aspartate[27]. The findings from both of these kinetic studies suggest the physical basis for these effects is rooted in how biomolecules impact the desolvation of Ca^{2+} and Mg^{2+}, but this explanation could not be tested directly. Insights from these experimental

studies, coupled with evidence that electrolyte ions also modulate solution properties during mineral formation[28], indicate that additional approaches are needed to build a mechanistic understanding of how calcification processes may be influenced by macromolecule interactions with hydrated ions. More broadly, many reactions in diverse biologic systems are strongly influenced by ion solvation across a continuum from weaker, solvent-separated interactions to contact ion pair (CIP) formation, where waters of hydration are displaced and direct ligand binding occurs.

Efforts to gain a complete understanding of ion-organic interactions have found a powerful tool in computer simulation. Here we present a fundamental MD investigation that is focused on ion solvation in the presence of small acidic biomolecules. Previous studies using MD have obtained valuable information regarding the structure and dynamics of water at the mineral surfaces of calcite[29-31] and barite[24, 32]. A good example is found in work by Piana and colleagues demonstrating that Asp bound to the (001) face of barite greatly decreases the energy barrier for dehydrating Ba^{2+} and bringing it to the mineral surface[33]. MD methods are well suited for the study of biomineral systems as they allow for the description of large, complex assemblages and readily incorporate force fields developed for the modeling of macromolecules.

Using small organic molecules as simple model compounds for the larger matrix macromolecules found in biomineral systems, this study utilizes MD simulation to investigate the effects of soluble organic molecules on ion hydration in regards to calcification. Our primary objective is to determine the extent to which aqueous cations are dehydrated by carboxylate moieties and to probe the thermodynamics of these processes. We focus on Ca^{2+} because of its importance in calcifying organisms and also examine Mg^{2+} and Sr^{2+}, ions that readily substitute for Ca^{2+} in carbonate biominerals. Our approach is motivated by previous studies which have found that insights from amino acids and small peptides can be scaled up to predict some aspects of macromolecule influences on calcification[26].

Our findings are presented in three sections. First, we show the results of MD simulations that determine the hydration states of Mg^{2+}, Ca^{2+}, and Sr^{2+} in the absence of any organic molecules. Reactive flux water exchange rates for each cation are also reported to demonstrate that the potentials employed produce results that are consistent with experimental trends and previous simulation results. A second section describes the effect of three acidic organic molecules on the hydration of aqueous Ca^{2+}. We chose a single Asp residue and two Asp-

containing dipeptides as the organic components because of their simplicity and proven ability to accelerate calcite growth kinetics. In the third section, we present the potential of mean force (PMF) profiles for the interaction of an Asp molecule with each of the divalent cations in solution. These results quantify the energetic barriers to partial cation dehydration as a function of the Asp – ion separation distance. The findings show differenctial dehydration behavior between Mg^{2+}, Ca^{2+} and Sr^{2+}, suggesting that organic-ion interactions may influence impurity uptake into calcium-containing biominerals.

2.2 Introduction to MD

2.2.1 MD Force Fields

In order to describe atoms and molecules without considering electrons, MD fundamentally relies on the Born-Oppenheimer approximation, where nuclei are viewed as fixed in position relative to electronic motion due to the large difference in masses between electrons and atomic nuclei. Instantaneous forces on the particles can then be calculated by a set of continuous potential energy functions called force fields that describe both bonded and non-bonded interactions. A force field may consist of simple functions such as Hooke's law, or may be more complex as when the functions are a fit of extensive quantum mechanical energy calculations.

Contributions to an MD force field are typically divided into two categories: bonded and nonbonded. The basic interactions between bonded atoms are bond stretching, bond torsion or rotation, and angle bending. Force field parameters assign an equilibrium value to each bond or angle in the system and the potential energy functions describe the changes in energy that result due to deviations from those equilibrium values. Nonbonded interactions are separated into electrostatic contributions and van der Waals forces. Electrostatic considerations are generally calculated by Coulomb's law while the Lennard-Jones (LJ) potential is a commonly used simple model to describe van der Waals interactions. Bonded and nonbonded contributions to the forece field are shown by representations in **Figure 2.1**. The specific MD force field used in our simulations, CHARMM22, is described in the following section.

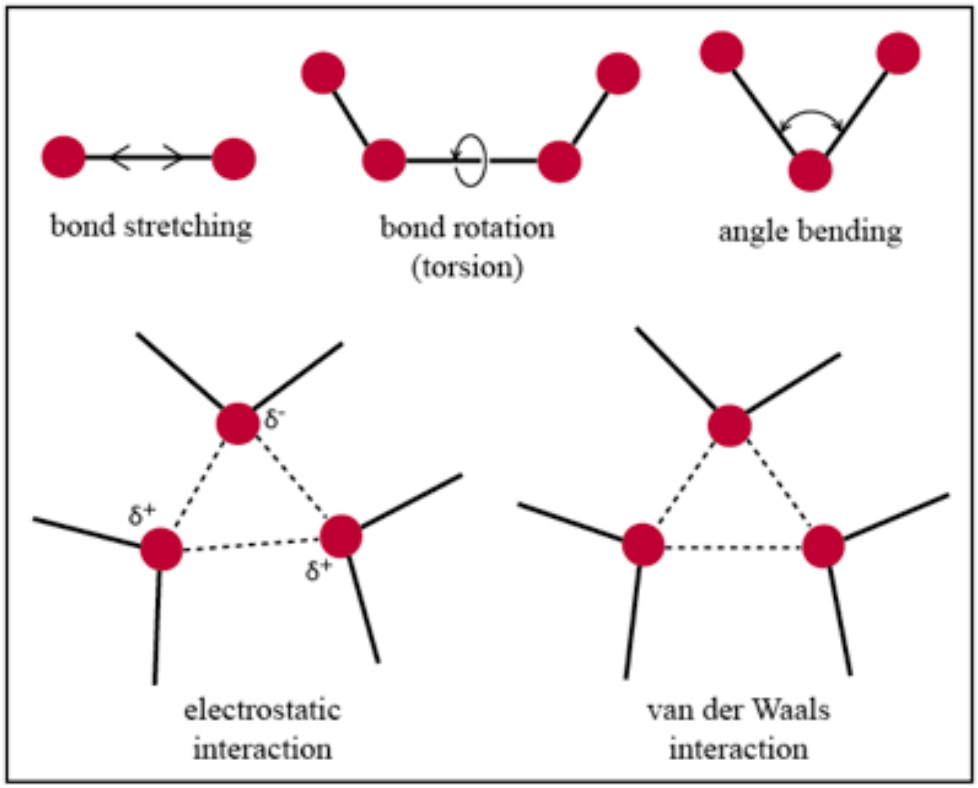


Figure 2.1. Representations of the five fundamental contributions to a molecular dynamics force field.

2.2.2 *The CHARMM22 Force Field*[34]

CHARMM22 is an all-atom empirical potential developed by MacKerell and colleagues to accurately represent protein-protein, protein-solvent, and solvent-solvent interactions. All-atom refers to the treatment of each individual atom individually by the force field, as opposed to the use of extended atom forms (i.e. where a CH_3 group is treated as a single entity). The CHARMM22 force field separates potential energy into six components: bonds (bond stretch), angles (angle bend), Urey-Bradley interactions (2-bond stretch), dihedrals (bond rotation), impropers, and nonbonded (van der Waals and electrostatic interactions). Total potential energy as a function of the vector of the atomic coordinates ($\vec{R}$) is then expressed as a sum over all components:

$$U(\vec{R}) = \sum_{bonds} K_b (b - b_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{UB} K_{UB} (S - S_0)^2 + \sum_{dihedrals} K_\chi [1 + \cos(n\chi - \delta)] + \sum_{impropers} K_{imp} (\varphi - \varphi_0)^2 +$$

$$\sum_{nonbonded} \varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{r_{ij}}$$

Equation 2.1. Poetential energy summation for the CHARMM22 force field.

K_b, K_{UB}, K_θ, K_χ, and K_{imp} = force constants; b, S, θ, χ, and φ = the bond distance, Urey-Bradley 1,3-distance, bond angle, dihedral angle, and improper torsion angle, respectively, at any time in the simulation. The subscript zero indicates an 'equilibrium' or 'natural' distance or angle corresponding to the lowest potential energy state. A harmonic potential form is used to model all the non-bonded interactions except for dihedrals, which are described by a cosine series and where n is the multiplicity or coefficient of symmetry. The potential energy summation includes all possible bonds, angles, and dihedrals, while only select Urey-Bradley and improper dihedral terms are incorporated to optimize the fit of simulations to vibrational spectra.

The force field employs a Lennard-Jones (LJ) 6-12 term and a Coulombic term to describe the nonbonded interactions. Nonbonded terms are applied to all atom pairs separated by three or more covalent bonds. In the LJ contribution, ε_{ij} is the well depth, r_{ij} is the separation distance between atoms i and j, and σ_{ij} is the distance at which the LJ energy is minimized. For pairs of different atoms, LJ parameters are determined from the Lorentz-Berthelodt combination rules: ε_{ij} is the geometric mean of ε_i and ε_j, and R_{ij} is calculated as the arithmetic mean between R_i and R_j. Physically, the '12' term (raised to the twelfth power) represents Pauli repulsion forces resulting from orbital overlap at small atomic separations, while the 6 term describes long range attraction due to dispersion (temporary dipoles). Coulombic contributions are a function of the partial atomic charges on each atom, q_i and q_j as well as r_{ij}. No explicit nonbonded interaction specifically for hydrogen bonding is included in the CHARMM22 force field; LJ and Coulombic contributions are assumed to adequately describe hydrogen bonding.

CHARMM22 bond lengths and angles and dihedral parameters (intramolecular terms) were fit to accurately represent geometries. Ideally, both gas phase and X-ray structure data supplemented by information from *ab initio* calculations were used to determine reasonable geometric parameter ranges for model compounds. Following the establishment of suitable geometries, the force constants in **eq. 2.1** were adjusted by fitting to vibrational data from gas-phase infrared and Raman spectroscopies. Solution and crystal data was used when vibrational information was not available, as often is the case for ionic compounds. As for the geometric

parameters, experimental data was supplemented with *ab initio* results when optimizing the force constants.

Geometric and force constant parameters were finalized by rechecking geometries and adjusting both parameters in an iterative fashion until reasonable agreement with the structures, vibrational spectra, and energy surfaces of the model compounds was reached. Lastly, Urey-Bradley and improper dihedral terms were added in cases where the force field agreement with vibrational spectra was found to be unsatisfactory. The Urey-Bradley contributions are most important when describing in-plane deformations and when differentiating between symmetric and asymmetric bond stretching modes. Improper dihedrals were found necessary to represent out-of-plane modes like the wagging of imidazole hydrogens.

Fitting of the nonbonded (intermolecular) parameters focused on balancing protein-protein, protein-water, and water-water interactions. All parameters were developed specifically for the TIP3P model of water[35], which accurately reproduces the first-shell hydration properties and energetics of liquid water at a low computational cost compared with other models, although it results in an overly high diffusion coefficient. LJ parameters and partial atomic charges were fit to reproduce interaction energies and minimum energy geometries from *ab initio* calculations and experimental data on the model compounds. Thermodynamic properties, molecular volumes, and the magnitude and direction of dipole moments were also considered during the fitting procedure.

The CHARMM22 force field was parameterized based on a set of small model compounds with the assumption that parameters fit to certain chemical functionalities could be transferred to larger molecules such as amino acids. Model compounds were chosen based on the nature of the compound and the availability of data. The same parameters developed for a model compound can then be used for a chemically similar moiety within a larger molecule. Parameters for a simulation are chosen by classifying all atoms in the system as a certain atom type, as defined by the identity and chemical environment of the individual atom.

2.3 Computational Methods

All calculations were performed with the LAMMPS[36] MD software on the Virginia Tech SGI Altix 3700 clusters. Simulations used the CHARMM22 force field[34], and parameters that describe interactions between the metal ions and the peptides were obtained using the Lorentz-

Berthelodt combination rules. Description of the solvent in all simulations was carried out with the TIP3P[37] model of water with the O-H bonds and H-O-H angles kept fixed by the SHAKE[38] algorithm; this model is easily implemented in LAMMPS and well-suited for use with the CHARMM22. To model the hydration of divalent cations, we incorporated supplementary cation-water LJ parameters from Åqvist[39] that were calculated by free energy perturbation to reproduce experimental hydration free energies and radial distribution functions for alkali and alkaline Earth metals. The Åqvist cation-water LJ parameters were modified for use with the TIP3P water model. PMF calculations were duplicated using the SPC model of water as well as alternate ion – oxygen parameters developed by Babu and Lin[40] to determine the dependence of the results on choice of model description. Results from these additional simulations are shown in **Appendix 1 (Figures A1.3 and A1.4)**. Calculations applied the canonical (NVT) ensemble with a Nose/Hoover thermostat set to a temperature of 300 K. All simulation outpurt was analyzed using Python routines written by the author (**Appendix 1**).

2.3.1 Hydration of cations in bulk water

Simulations of solvated Mg^{2+}, Ca^{2+}, and Sr^{2+} consisted of a periodic water box of side length 20 Å containing 261 water molecules (solvent density = 0.975 g/cm^3) and one divalent cation (ion concentration = 0.20 M). Chloride anions were included for charge balance. The MD timestep was set at 0.25 fs and each system was equilibrated for 40 ps before collecting the ion-water oxygen radial distribution information for at least 250 ps. Distribution data was gathered every 10 timesteps (2.5 fs) for the cation-oxygen distance from 3.0-10 Å and was divided into 500 bins. Example input files are provided in **Appendix 1**.

The residence times of water molecules in the first shell of the cations were calculated using the reactive flux method[41]. Within this formalism, the rate constant for water exchange, k, is expressed as a product:

$$k = \kappa k^{TST}$$

Equation 2.2. Expression for the water exchange rate constant determined by reactive flux.

where k^{TST} = the rate constant from transition state theory (TST)[42] and κ = the transmission coefficient or fraction of activated states that result in a successful exchange event. The TST rate constant was related to the PMF along the reaction coordinate r by the relation[41, 43]:

$$k^{TST} = \sqrt{\frac{k_B T}{2\pi\mu}} \frac{r^{\ddagger 2} e^{-\beta W(r^{\ddagger})}}{\int_0^{r^{\ddagger}} r^2 e^{-\beta W(r)} dr}$$

Equation 2.3. Expression for the transition state theory rate constant calculated in the reactive flux method.

The coordinate r in this case refers to the distance between the cation and an exchanging water molecule. μ is the reduced mass of a cation-water pair and $r^{\ddagger}$ = the ion-water distance corresponding to the activated complex where the PMF is at a maximum. Thus, the TST rate constant can be calculated from the PMF along an ion-water distance spanning the region of water exchange, or the region between the first and second hydration shells.

We employed an umbrella sampling MD approach to generate PMF profiles (method described in section ***2.3.3***) for water exchange about Mg^{2+}, Ca^{2+}, and Sr^{2+}. Transmission coefficients were determined as the plateau value of the normalized reactive flux $k(t)$, which was calculated as:

$$k(t) = \frac{\left\langle \dot{r}(0)\theta\left[r(t) - r^{\ddagger}\right]\right\rangle_c}{\left\langle \dot{r}(0)\theta\left[\dot{r}(0)\right]\right\rangle_c}$$

Equation 2.4. Expression for normalized reactive flux, used to determine the transmission coefficients in reactive flux residence time calculations.

$\dot{r}(0)$ = the initial velocity along the reaction coordinate and $\theta[x]$ = the Heaviside function, which is 1 for values of x larger than 0 and 0 otherwise. The superscript c indicates that averages were generated from starting configurations constrained at the transition state, which was defined as the ion-water separation corresponding to the maximum free energy in our PMF profiles ($r^{\ddagger}$). MD systems were constrained to the transition state by a strong harmonic potential (spring constant = 125 – 300 kcal/mol-Å^2). After 2.5 ps of equilibration, atoms were assigned new, random velocities and the trajectory was followed until the system became committed to either

the reactant or product basin. This procedure was repeated up to 10,000 times so that the averages produced a plateau value in $k(t)$.

2.3.2 Biomolecule interactions with Ca^{2+}

The Ca^{2+} – O_{water} radial distribution was monitored as a function of the ion's proximity to three acidic organic molecules: Asp, $(Asp)_2$ and AspLeu. Simulation systems were made up of a periodic water box with side lengths of 24.632 Å that contained 475-485 water molecules, one Ca^{2+} ion and one zwitterionic peptide or dipeptide molecule (**Figure 2.2**). A harmonic constraint with a spring constant from 100-150 kcal/mol- $Å^2$ was introduced to restrict the distance between Ca^{2+} and the center-of-mass of the organic molecule to a specific value between 3 Å and 10 Å. Simulations were equilibrated for 20 ps followed by data collection for 120 ps using a timestep of 0.2 fs. Ion position, peptide or dipeptide center-of-mass position, and Ca^{2+} – O_{water} radial distribution information was collected every timestep and used to determine the average ion - organic separation and Ca^{2+} first shell hydration number for each MD run.

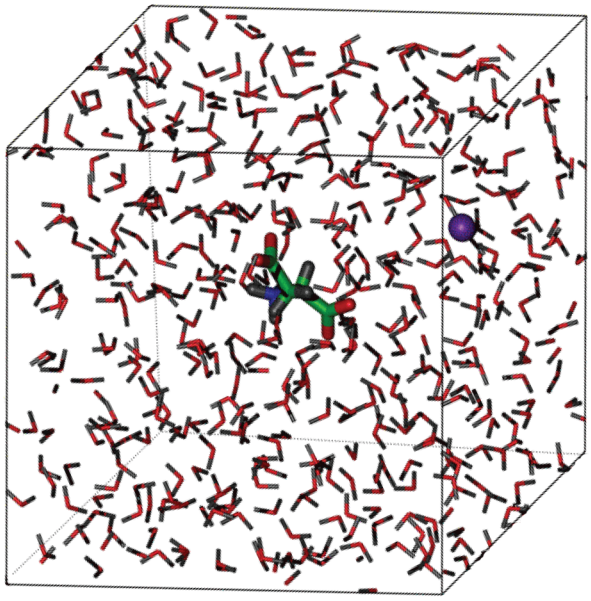


Figure 2.2. Example MD input system consisting of a box of water with Ca^{2+} (purple sphere) and Asp molecule.

2.3.3 Cation-specific interactions with Asp

PMF profiles were generated for Mg^{2+}, Ca^{2+}, and Sr^{2+} over cation—Asp separation distances ranging from 9 Å to 4 Å. Each simulation consisted of a periodic box of side length 24.632 Å containing 475 water molecules (solvent density = 0.970 g/cm^3), one Asp molecule, and one divalent cation. Asp was represented in zwitterionic form with full degrees of freedom. A chloride anion was added to the simulation box to keep the total system charge neutral.

An umbrella sampling approach was applied in order to fully sample the conformational space from large to small cation-Asp separation. This technique modified the MD potential to allow sufficient sampling of the unfavorable states of the system along a defined reaction coordinate[44]. The free energy surface along the reaction coordinate, $W(r)$ or PMF, is related to the probability distribution in that coordinate[45],

$$W(r) = -k_B T \ln P(r) + C$$

Equation 2.5. Expression for potential of mean force calculated with umbrella sampling.

where r = the reaction coordinate of interest, $P(r)$ = the probability distribution in r, k_B = Boltzmann's constant, T = temperature, and C = a normalization constant often chosen so that the most probable distribution corresponds to zero free energy.

The umbrella sampling method divided the simulation into a series of windows designed to collectively sample the entire relevant range of r. Each individual window was centered about a different r and spanned a limited range of r values. A restraining or umbrella potential in harmonic form was added to the force field potential function to restrict the sampling in each window to the desired range. The resulting series of non-Boltzmann probability distributions were unbiased by removing the effects of the restraining potentials[46]. Recombination of the unbiased information then provided a single probability distribution over the full range of r, which was used to produce a PMF profile from **eq. 2.5**.

The procedure for unbiasing and combining data from different sampling windows was established by Torrie and Valleau, who related the biased window frequency distributions to the total unbiased distribution by[44],

$$P(r) = \frac{P*(r)\exp[\beta U^*(r)]}{\left\langle \exp[\beta U^*(r)] \right\rangle_T}$$

Equation 2.6. The relationship between the unbiased probility distribution and the biased distribution calculated with umbrella sampling.

where $P^*(r)$ = the biased probability, $U^*(r)$ = the restraining potential, β = the thermodynamic beta $\frac{1}{k_B T}$, and $\langle\ \rangle_T$ denotes a constant temperature ensemble average over the biased system and represents the work required to maintain the umbrella restraint. For a one-dimensional problem where the restraint potential is dependent solely upon a single parameter, **eqs. 2.5 and 2.6** can be combined into an expression for the free energy surface of the i^{th} umbrella sampling window[47]:

$$W_i(r) = -k_B T \ln\left[P_i^*(r)\right] - U_i^*(r) + C_i$$

Equation 2.7. Potential of mean force as a function of the biased probability distribution calculated from umbrella sampling simulations.

Here, the unbiased probability distribution is no longer needed and the complete free energy surface can be obtained by splicing together $W_i(r)$ for all windows. C_i in **eq. 2.7** is the unknown factor $k_B T \ln\left\langle \exp[\beta U^*(r)] \right\rangle_T$ and was determined for each window based upon the condition that adjacent free energy surfaces have matching slopes.

The coordinate r was defined as the separation distance between the cation and the Asp molecule center-of-mass, which was easily determined at any time from the Cartesian coordinates of the system. Simulations were divided into 14-21 windows, each centered about a different value of r between 9.5 and 3.0 Å. Cation-Asp separation was restrained in each window by a harmonic biasing potential,

$$U^*(r) = \frac{K}{2}(r - r_o)^2$$

Equation 2.8. Expression for the harmonic biasing potential used in umbrella sampling simulations.

where K = a spring constant and r_0 = the equilibrium cation-Asp separation for the window. Values for K ranged from 1.5 to 7.5 kcal/mol-Å^2 depending on the cation and r_0; appropriate K values allow complete sampling of the trajectory and generate sufficient overlap between windows, which is necessary for proper splicing of the data.

The Weighted Histogram Analysis Method (WHAM)[48] was employed to unbias and splice umbrella sampling results. This method was used to locate the point of maximum overlap between adjacent windows, discard overlapping data, calculate $W_i(r)$ from **eq. 2.5**, and compute shifting constants C_i to produce a complete free energy profile. The resultant surface had an arbitrary baseline and is informative to distinguish relative free energies. WHAM was implemented using the program developed by Alan Grossfield[49].

Each simulation was equilibrated for at least 120 ps with a timestep of 0.2 fs before the adding of any restraining potentials. During umbrella sampling, positional coordinates of the cation and Asp molecule were collected every 1.0 fs for 1.0 ns per window. Oxygen nuclear positions were also recorded every 50 fs in order to monitor cation hydration state.

2.4 Results

2.4.1 Hydration of cations in bulk water

Calculated radial distributions functions, $g(r)$, show that ion-oxygen distances for the primary and secondary hydration spheres of Mg^{2+}, Ca^{2+} and Sr^{2+} are consistent with the values expected from the Åqvist parameters[39]. Running integration of the radial functions, labeled $n(r)$ in Figure 1, denotes the number of water molecules found within a given distance of the cation. Primary and secondary hydration numbers determined directly from $n(r)$ and the locations of $g(r)$ peaks are presented in **Table 2.1**.

Table 2.1. Atomic radii, cation-oxygen distances, r_1 and r_2, for the primary and secondary hydration shells and the corresponding average number of waters, n_1 and n_2, found in each shell

	atomic radius[a] (Å)	primary shell		secondary shell	
		r_1 (Å)	n_1	r_2 (Å)	n_2
Mg^{2+}	1.45	2.00	6.0	4.13	12.8
Ca^{2+}	1.94	2.41	8.0	4.58	15.4
Sr^{2+}	2.19	2.57	8.4	4.73	17.7

[a]Ref. 50

A strong $g(r)$ Mg^{2+} – O first shell peak is seen at 2.00 Å (**Figure 2.3a**), slightly shorter than the distances of 2.11 Å observed with X-ray diffraction[51] and 2.08 Å calculated by a combined quantum mechanics/molecular mechanics[52] approach and MD using a polarizable force field[53]. **Table 2.1** shows our calculated coordination number for Mg^{2+} is 6, which is in agreement with both experiment and theory[51-55]. The first shell peak for Ca^{2+} is located at 2.41 Å and has a corresponding $n(r)$ of 8.0 (**Figure 2.3b**). These values are within the reported ranges for first shell location (2.35 – 2.68 Å) and coordination number (6.0 - 9.2) determined both experimentally[56, 57] and computationally[53, 55, 56, 58, 59]. Our simulations determine an Sr^{2+} – O first shell distance of 2.57 Å and an average coordination of 8.4 (**Figure 2.3c**).

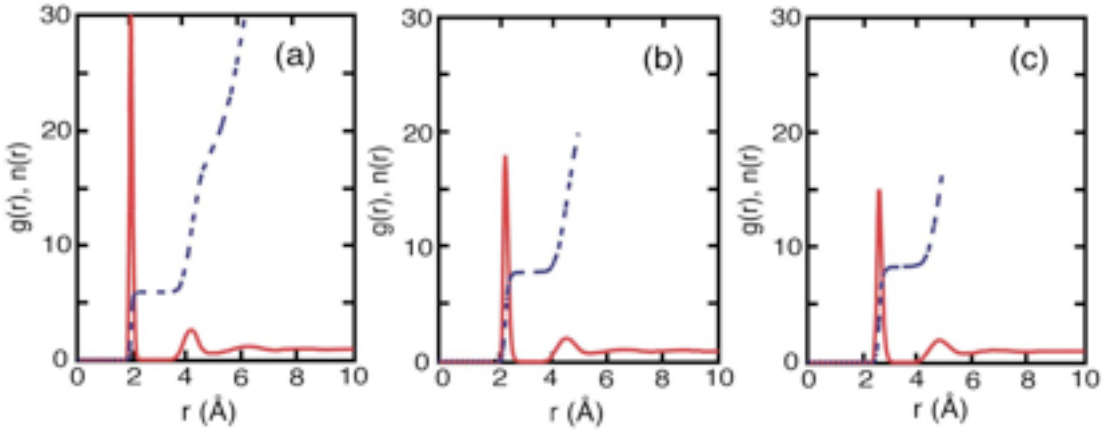


Figure 2.3. Calculated cation – water oxygen radial distribution functions, $g(r)$ (solid red line) and running integration number, $n(r)$ (dashed blue line) for (a) Mg^{2+}, (b) Ca^{2+} and (c) Sr^{2+}.

Fewer studies have been carried out regarding Sr^{2+} hydration, but our results are in reasonable agreement with experimental investigations on strontium-bearing solutions. EXAFS studies have determined Sr^{2+} – O distances and coordination numbers of 2.57 Å and 7.8[60, 61] as well as 2.62 Å and 8.3[61], while XRD indicates a distance of 2.64 Å and a coordination of 8[62].

The smaller Mg^{2+} and Ca^{2+} strongly favor coordination by 6 and 8 waters, respectively, as seen by the relative probabilities of occurrence for each first-shell hydration number, n_1 (**Figure 2.4**). In contrast, the primary hydration sphere of Sr^{2+} is much more dynamic and similar probabilities exist for the 8 – and 9 – coordinated cation, resulting in an average value of 8.4 (**Table 2.1**). As expected, first shell waters about Mg^{2+} are very tightly bound by the ion; while some water molecules enter the region between the first and second hydration shells, no

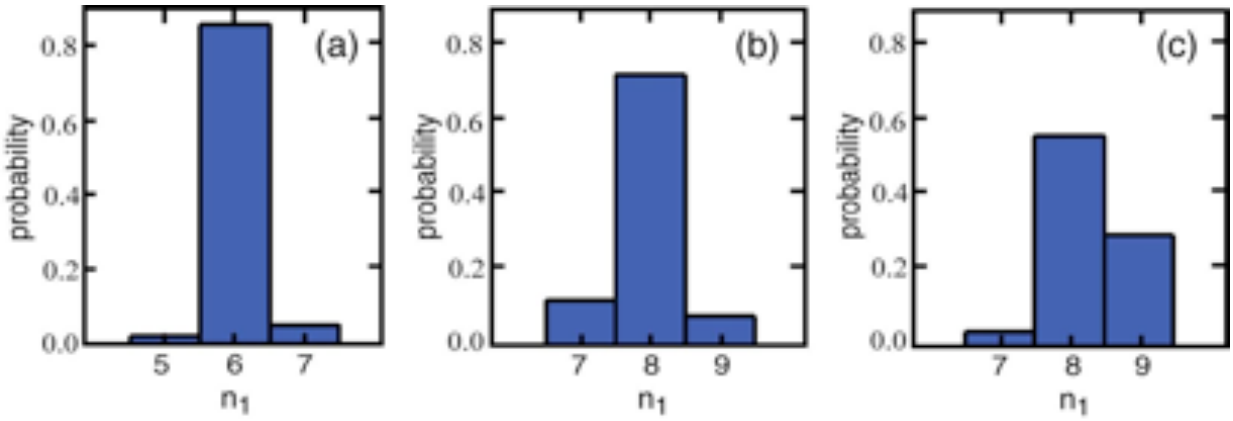


Figure 2.4. Calculated probabilities of finding (a) Mg^{2+}, (b) Ca^{2+}, and (c) Sr^{2+} with different numbers of water molecules in the inner hydration sphere.

water exchange events are observed during the 250 ps simulation. Compared to Mg^{2+}, much greater mobility is observed in the hydration spheres of Ca^{2+} and Sr^{2+} as water exchange events are seen during the simulation time for the two larger cations.

Results for the residence time calculations are compared to experimental and computational values from MD studies using different methods in **Table 2.2**. The calculated PMF profiles (**Figure 2.5**) demonstrate that the barrier for water escape from the first hydration shell is greatest for Mg^{2+} (13.5 kcal/mol) and smallest for Sr^{2+} (3.9 kcal/mol).

Table 2.2. Calculated ($\tau_{calc.}$) and experimental ($\tau_{exp.}$) water residence times for the primary hydration sphere of Mg^{2+}, Ca^{2+} and Sr^{2+}.

cation	residence time (seconds)	
	calculated	experimental
Mg^{2+}	1.3×10^{-8} [a]	6.7×10^{-6} [d]
	$> 1.0 \times 10^{-8}$ [b]	
Ca^{2+}	2.2×10^{-10} [a]	3.2×10^{-9} [e]
	4.9×10^{-11} [b]	
	7.0×10^{-10} [c]	
Sr^{2+}	7.9×10^{-11} [a]	2.5×10^{-9} [e]
	4.3×10^{-11} [b]	

[a]This work (MD reactive flux method); [b]Ref. 63 (MD direct method); [c]Ref. 64 (MD direct method); [d]Ref. 65 (NMR); [e]Ref. 66 (NMR)

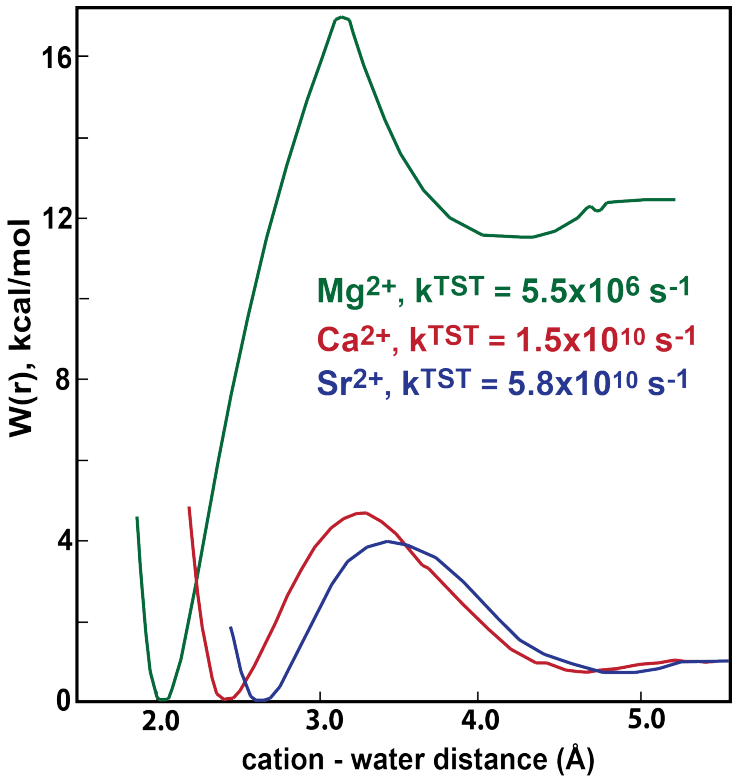


Figure 2.5. Calculated PMFs for a first shell water exchange event around Mg^{2+}, Ca^{2+} and Sr^{2+}, and the transition state theory rate constants calculated by eq. 2.3.

In the group IIa ion series, the water residence time and the barrier to water exchange decrease as ion radius increases and charge density decreases. The fact that this trend is reproduced by our simulations indicate that the methods and potentials employed in this work are suitable for describing ion-solvation dynamics.

2.4.2 Biomolecule-facilitated desolvation of Ca^{2+}

Calculations of the hydration number of Ca^{2+} versus separation distance from Asp, Asp_2 and AspLeu show that the ion retains the expected 8 first-shell waters of hydration at large cation – organic separations (**Figure 2.6**). As shorter ion-organic separations are sampled, the simulations show the onset of significant disruptions of the inner hydration sphere of Ca^{2+} within the interval where solvent-separated Ca^{2+}-biomolecule interactions begin and the CIPs are formed. **Figure 2.7** illustrates the disruption of Ca^{2+} hydration within this interval for a simulation with an average Ca^{2+} – Asp separation of 4.85 Å. At this distance, CIP formation has not occurred and the nearest carboxylate group to the cation is at an average distance of 3.52 Å.

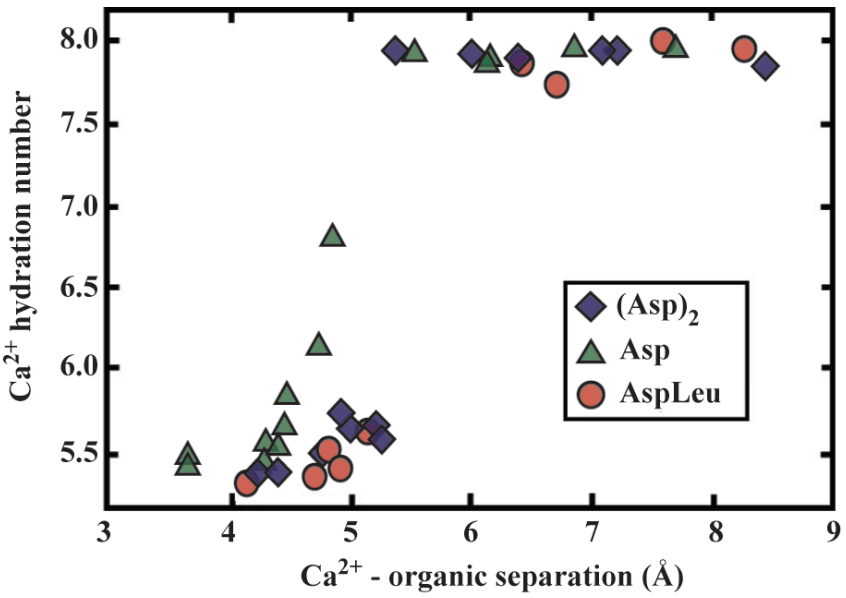


Figure 2.6. The number of water molecules contained in the primary hydration shell of Ca^{2+} as a function of the ion's proximity to three organic molecules.

Asp induces a reduction in the most probable hydration state compared to the probabilities in bulk water (compare **Figure 2.3a and 2.7b**), and the calculated radial distribution function

shows that the average hydration number decreases to 6.92 from 8.0 (compare **Figure 2.7a and 2.3b).**

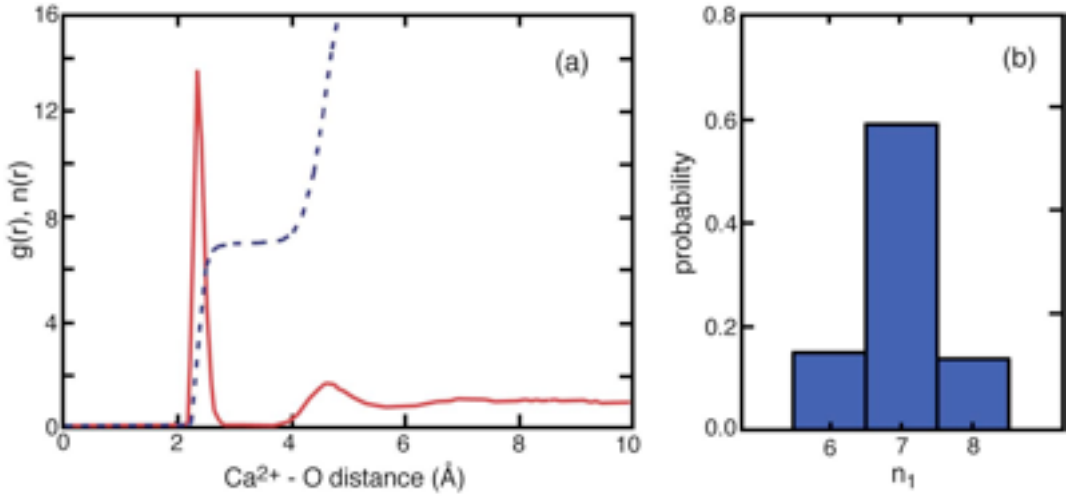


Figure 2.7. (a) Ca^{2+} – O_{water} radial distribution function, $g(r)$ (red solid line) and running integration, $n(r)$ (blue dashed line). (b) Coordination number histogram indicating partial dehydration of Ca^{2+} in proximity to an Asp molecule. Both figures correspond to an average ion – Asp separation of 4.85 Å.

At the smallest Ca^{2+} – organic separation, a CIP is formed with the entry of two carboxylate oxygen atoms into the inner hydration sphere of the ion and the expulsion of an average of 2.5 first shell water molecules. In this scenario, the number of first-shell water molecules decreases to an average of 5.5 and total coordination of Ca^{2+} is reduced to 7.5 (5.5 water oxygen atoms and 2 carboxyl group oxygen atoms).

2.4.3 Cation-specific interactions with Asp

Calculations of the potential of mean force versus the ion – Asp separation distance show cation-specific interactions with the carboxylate moieties of Asp (**Figure 2.8**). For calcium, the PMF plot shows a strong energy minimum representing CIP formation, where two carboxylate oxygen atoms enter into the primary hydration sphere of Ca^{2+}. The energy barrier corresponding to this process is small (≈ 2 kcal/mol). The different hydration states of Ca^{2+} as the ion moves from a solvent-separated ion pair with a carboxyl group of Asp to a CIP are shown in **Figure 2.9**.

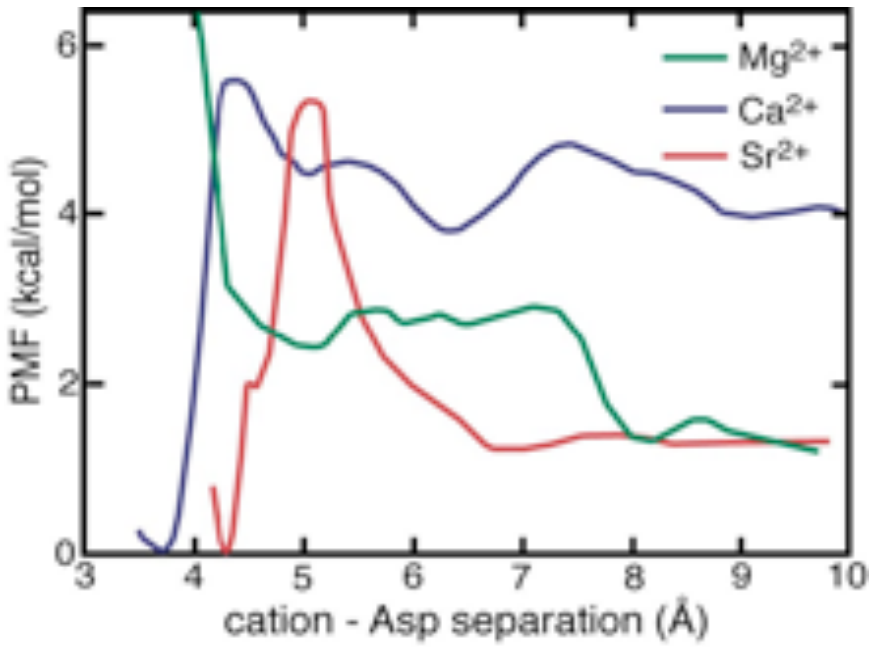


Figure 2.8. Potential of mean force (PMF) diagrams as a function of ion – Asp separation for the interaction of Mg^{2+}, Ca^{2+} and Mg^{2+} with a single Asp molecule.

The PMF for the Sr^{2+} – Asp system shows a weaker energy minimum for the ion – carboxylate CIP formation than for Ca^{2+}, a result that agrees with reported binding constants[67]. The energy barrier for Sr^{2+} is expected to be smaller than that of Ca^{2+} because Sr^{2+} has a lower charge density and a more flexible hydration sphere. However, we observe a larger energetic cost for Sr^{2+} ($\approx$ 4 kcal/mol). This outcome can be explained by differences in ion trajectory between Sr^{2+} and Ca^{2+}. As Ca^{2+} approaches Asp, the ion attracts both organic carboxylates and Asp bends so that both COO^- groups point towards the ion. During the approach of Sr^{2+}, Asp remains extended and only one carboxyl group faces the ion (images in **Appendix A, Figure A1.5**). This difference in trajectories may be explained by differences in the strength of the Coulmbic interactions between the caions and Asp. Since the Coulombic force decays by $1/r^2$ and the Sr^{2+} – O_{water} equilibrium distance is longer than that of Ca^{2+}, the Coulomb interaction between Sr^{2+} and Asp is weaker than for Ca^{2+} when the ions are fully hydrated. As seen for the Ca^{2+} – Asp system, the Sr^{2+} – Asp CIP also has a reduced in first shell coordination number (8.4 versus 8.0).

For Mg^{2+}, high charge density and strong affinity for inner sphere water molecules result in the absence of a free energy minimum corresponding to a Mg^{2+} – Asp ion pair. Experimental measurements of complexation equilibria show magnesium forms strong complexes with

carboxylate groups [67], but crystal structure analyses report a Mg:Asp stoichiometry of 1:2 [68]. Our simulation, however, contains only one Asp molecule and one Mg^{2+} cation, which may not have sufficient charge density to induce partial desolvation of the ion by Asp. To test the alternate 1:2 stoichiometry, we repeated the PMF calculation with a system containing one Mg^{2+} ion and two Asp molecules. During the simulation, one Asp was held at a fixed distance corresponding to a solvent separated Mg^{2+} – Asp pair while the second Asp was moved progressively closer to the cation. The PMF over the separation distance between Mg^{2+} and the second Asp molecule is shown in the **Appendix 1 (Figure A1.6)**. As seen for the simulation with a single Asp, the system with a 1:2 Mg^{2+}:Asp ratio does not predict the favorable partial desolvation of the Mg^{2+}. For both cases, no disturbance in the inner hydration shell of Mg^{2+} is seen prior to the energetically costly physical replacement of a single water molecule with one carboxylate oxygen atom.

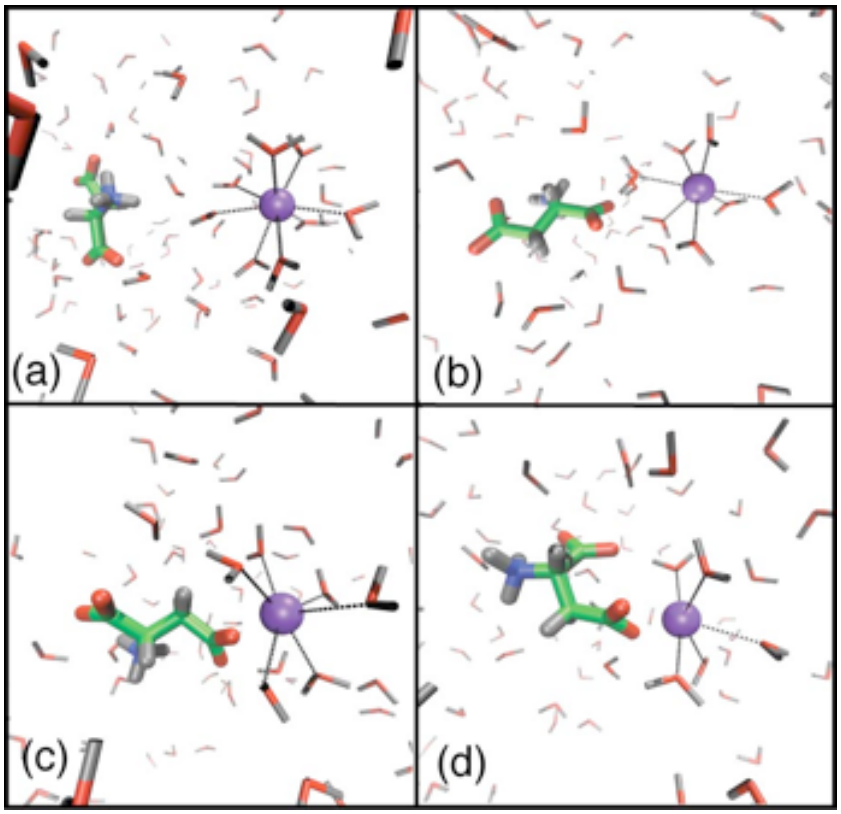


Figure 2.9. MD snapshots depict the partial dehydration of Ca^{2+} as the ion approaches Asp. Dashed lines indicate water molecules in the primary hydration sphere of Ca^{2+}. The number of first shell waters decreases from (a) 8 to (b) 7, (c) 6 and (d) 5. Some water molecules have been removed for clarity. Atom colors are: Ca – purple, N – blue, C – green, O – red, H – grey.

2.5 Discussion

The results demonstrate that simple carboxylated molecules displace solvation water about Ca^{2+} at separations of up to 5.5 Å (**Figures 2.6 and 2.7**) and that the average solvation number is strongly dependent upon separation distance below this threshold. Additionally, the PMF calculations reported here show that carboxylated molecules reduce the energy barrier to the dehydration of Ca^{2+}. Cation dehydration is widely accepted as the rate-limiting step to crystal growth in aqueous solutions[69, 70]. Indeed, a recent investigation by Elhadj et. al. proposed that enhanced $CaCO_3$ growth rates in the presence of carboxyl-rich amino acids and peptides may be explained by the ability of the biomolecules to lower the energy barrier to Ca^{2+} dehydration[26], although this hypothesis could not be tested experimentally. The work presented here provides independent evidence in support of the idea that soluble carboxylated species promote partial cation desolvation, thus contributing to increased crystal growth rates.

Our findings are also consistent with MD simulations reported by Piana and colleagues indicating that surface-bound Asp lowers the activation energy to Ba^{2+} dehydration and acts as a cation 'shuttle' that coordinates it to the (001) face of barite, resulting in increased growth rates[33]. It is plausible that an analogous process occurs on the (104) surface of calcite. Our results, however, suggest that Asp in solution could also promote ion dehydration and mineralization. Reports that certain acidic proteins, which are only weakly attracted to the mineral surface, increase the growth rate of calcium oxalate monohydrate[71] are also in agreement with our physical interpretation - that soluble biomolecules can impact Ca^{2+} solvation in addition to surface-bound organics, producing the measured kinetic enhancements.

The cation-specific effects of simple carboxylated biomolecules on the dehydration of Ca^{2+}, Sr^{2+}, and Mg^{2+} also present an opportunity to test the claim that Asp-rich polypeptides of $(Asp_3Gly)_6$ Asp increase the Mg content in calcite by promoting the desolvation of Mg^{2+} relative to Ca^{2+} [27]. Our simulations for Mg^{2+} - Asp_1 interactions did not find evidence that a simple Asp monomer could disrupt the strong solvation sphere about this charge-dense cation for a Mg:Asp stoichiometry of 1:1 or 1:2. Our test is not definitive, however, because of the many differences between the properties of Asp_1 and $(Asp_3Gly)_6Asp$. A better simulation awaits advances in MD methods that can more accurately model the complexities of larger molecules.

Results of this study may also provide preliminary insights into the role of biomolecules in regulating the composition of amorphous calcium carbonate, a transient precursor phase that is

used by many organisms during calcification of skeletal structures. Recent work suggests that ACC is formed by the aggregation of ion-pairs into prenucleation clusters[72, 73] and that cation dehydration is a crucial step in this process[28]. The low energy barrier to Ca^{2+} interactions with carboxylate groups supports a report[74] that carboxyl-rich matrix macromolecules have systematic and predictable controls on the composition of ACC. Experimental measurements show the Mg content of ACC to be regulated by the binding strength of these biomolecules with Ca^{2+} versus Mg^{2+}. Estimates of the electrostatic potential distribution about these small carboxylated molecules[74] suggest weak sequestering interactions with Ca^{2+}, but not Mg^{2+}, just as reported in this study.

Conway has noted the importance of ion hydration in regulating many types of biophysical interactions[75]. As traditional concepts in ion-biomolecule complexation are revisited, long-standing questions regarding the physical basis for ion selectivity in biological systems may be resolved. For example, ion-specific hydration of Ca^{2+} and Mg^{2+} regulates both the triggering and inhibitory activity of intracellular proteins and enzymes. Matthews suggested the ready desolvation of Ca^{2+} could explain its ability to promote hormonal secretory processes versus the powerful blocking effects of Mg^{2+} [76, 77]. Another study reports that Mg^{2+} binding cannot trigger activity in Ca-proteins while Mg-proteins require a very high charge density to become active[75]. Similarly, Sr^{2+} and Ca^{2+} compete and can combine with the same sites in transmitter release, Sr^{2+} is the less effective ion[78, 79]. Each of these examples from biochemical settings may be accounted for in terms of our findings showing the energetic trends for cation interactions with carboxylated molecules. More work is needed, however, to confidently extend the results of this relatively simple investigation to complex biological systems.

2.6 Conclusions

This theoretical investigation provides mechanistic insights into how biomolecular species impact cation solvation on the molecular level. By monitoring the hydration state of Mg^{2+}, Ca^{2+}, and Sr^{2+} during their approach to negatively charged molecules, we show carboxylate moieties of Asp promote dehydration of Ca^{2+} and Sr^{2+}. Specific binding is not required to disrupt cation hydration and we demonstrate that reductions and rearrangements of first shell water can begin at ion-organic separation distances as large as 4.9 Å for Ca^{2+} and 5.1 Å for Sr^{2+}. At smaller ion-organic separations, the carboxylate-cation CIPs exhibit an overall

decrease in ion coordination number. Weaker Coulombic interactions appear to prevent the larger Sr^{2+} from taking the same path of approach as Ca^{2+}, resulting in a greater energetic cost for dehydration and CIP formation. In contrast to Ca^{2+} and Sr^{2+}, the solvation sphere of Mg^{2+} is largely unaffected by Asp until abrupt physical replacement of water by a carboxylate oxygen, which comes at a high energetic cost.

2.7 Acknowledgment

The research was supported by the US Department of Energy (grant FG02-00ER15112) and the National Science Foundation Graduate Research Fellowship Program. Computational resources were provided by the Virginia Tech Advanced Research Computing Center.

2.8 References

(1) Belcher, A. M.; Wu, X. H.; Christensen, R. J.; Hansma, P. K.; Stucky, G. D.; Morse, D. E. *Nature* **1996,** *381*, 56-58.
(2) Levi, Y.; Albeck, S.; Brack, A.; Weiner, S.; Addadi, L. *Chem. Eur. J.* **1998,** *4*, 389-396.
(3) Addadi, L.; Weiner, S. *Proc. Natl. Acad. Sci USA* **1985,** *82*, 4110-4114.
(4) Volkmer, D.; Fricke, M.; Huber, T.; Sewald, N. *Chem. Commun.* **2004**, 1872-1873.
(5) Fu, G.; Orme, C. A.; Morse, D. E.; DeYoreo, J. J. *Adv. Mater.* **2005,** *17*, 2678.
(6) Aizenberg, J.; Lambert, G.; Addadi, L.; Weiner, S. *Adv. Mat.* **1996,** *8*, 222-226.
(7) Raz, S., Hamilton, P.C., Wilt, F.H., Weiner, S. and L. Addadi *Adv. Funct. Mater.* **2003,** *13*, 480-486.
(8) Politi, Y.; Mahamid, J.; Goldberg, H.; Weiner, S.; Addadi, L. *Cryst. Eng. Comm.* **2007,** *9*, 1171-1177.
(9) Falini, G.; Albeck, S.; Weiner, S.; Addadi, L. *Science* **1996,** *271*, 67-69.
(10) Falini, G., Fermani, S., Vanzo, S., Miletic, M. and G. Zaffino *Eur. J. Inorg. Chem.* **2005**, 162-167.
(11) Wada, N.; Yamashita, K.; Umegaki, T. *J Coll. Int. Sci.* **1999,** *212*, 357-364.
(12) Aizenberg, J.; Hanson, J.; Ilan, M.; Leiserowitz, L.; Koetzle, T. F.; Addadi, L.; Weiner, S. *Faseb Journal* **1995,** *9*, 262-268.
(13) Albeck, S.; Addadi, I.; Weiner, S. *Connect. Tissue Res.* **1996,** *35*, 365-370.
(14) Weiner, S., Troub, W., and Lowenstam, H.A., *Biomineralization and Biological Metal Accumulation*. Reidel Publishing Company: Dordrecht, 1983.
(15) Puverel, S.; Tambutte, E.; Pereira-Mouries, L.; Zoccola, D.; Allemand, D.; Tambutte, S. *Comp. Biochem. Phys. B* **2005,** *141*, 480-487.
(16) Gotliv, B. A.; Kessler, N.; Sumerel, J. L.; Morse, D. E.; Tuross, N.; Addadi, L.; Weiner, S. *Chembiochem.* **2005,** *6*, 304-314.
(17) Levi-Kalisman, Y.; Falini, G.; Addadi, L.; Weiner, S. *J. Struct. Biol.* **2001,** *135*, 8-17.
(18) Cowan, J. A. *Inorg. Chim. Acta* **1998,** *275-276*, 24-27.
(19) Chevalier, F.; Angulo, J.; Lucas, R.; Nieto, P. M.; Martin-Lomas, M. *Eur. J. Org. Chem.* **2002,** *14*, 2367-2376.
(20) Ringer, S. *J. Physiol.* **1882,** *3*, 380-393.

(21) Beani, L.; Bianchi, C.; Santinoceto, L.; Marchetti, P. *Int. J. Neuropharmacol.* **1968,** *7*, 469-481.
(22) Chen, S.; Yu, S.; Jiang, J.; Li, F.; Liu, Y. *Chem. Mater.* **2006,** *18*, 115-122.
(23) Seo, K.-S.; Han, C.; Wee, J.-H.; Park, J.-K.; Ahn, J.-W. *J. Cryst. Growth* **2005,** *276*, 680-687.
(24) Piana, S.; Jones, F.; Gale, J. D. *J. Am. Chem. Soc.* **2006,** *128*, 13568-13574.
(25) Kowacz, M.; Putnis, C. V.; Putnis, A. *Geochim. Cosmochim. Acta* **2007,** *71*, 5168-5179.
(26) Elhadj, S.; De Yoreo, J. J.; Hoyer, J. R.; Dove, P. M. *Proc. Natl. Acad. Sci. USA* **2006,** *103*, 19237-19242.
(27) Stephenson, A. E.; DeYoreo, J. J.; Wu, L.; Wu, K. J.; Hoyer, J.; Dove, P. M. *Science* **2008,** *322*, 724-727.
(28) Kowacz, M.; Putnis, A. *Geochim. Costmochim. Acta* **2008,** *72*, 4476-4487.
(29) Kerisit, S.; Cook, D. J.; Spagnoli, D.; Parker, S. C. *J. Mater. Chem.* **2005,** *15*, 1454-1462.
(30) deLeeuw, N. H.; Parker, S. C. *Phys. Rev. B* **1999,** *60*, 13792-13799.
(31) Perry, T. D.; Cygan, R. T.; Mitchell, R. *Geochim. Cosmochim. Acta* **2007,** *71*, 5876-5887.
(32) Jang, Y. H.; Chang, X. Y.; Blanco, M.; Hwang, S.; Tang, Y.; Shuler, P.; III, W. A. G. *J. Phys. Chem. B* **2002,** *106*, 9951-9966.
(33) Piana, S.; Jones, F.; Gale, J. D. *Cryst. Eng. Comm.* **2007,** *9*, 1187-1191.
(34) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B.* **1998,** *102*, 3586-3616.
(35) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983,** *79*, 926-935.
(36) Plimpton, S. J. *J. Comp. Phys.* **1995,** *117*, 1-19.
(37) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983,** *79*, 926-935.
(38) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977,** *23*, 327-341.
(39) Aqvist, J. *J. Phys. Chem.* **1990,** *94*, 8021-8024.
(40) Babu, C. S.; Lim, C. *J. Phys. Chem. A* **2006,** *110*, 691-699.
(41) Rey, R.; Hynes, J. T. *J. Phys. Chem.* **1996,** *100*, 5611-5615.
(42) Agmon, N. *Chem. Phys.* **1980,** *45*, 249-260.
(43) Ciccotti, G.; Ferrario, M.; Hynes, J. T.; Kapral, R. *J. Chem. Phys.* **1990,** *93*, 7137-7147.
(44) Torrie, G. M.; Valleau, J. P. *J. Comp. Phys.* **1977,** *23*, 187-199.
(45) Leach, A. R., *Molecular modeling: principles and applications - 2nd edition*. Pearson Education Limited: Edinburgh Gate, 2001; p 580-585.
(46) Field, M. J., *A Practical Introduction to the Simulation of Molecular Systems, Second Edition*. 2 ed.; Cambridge University Press: Cambridge, 2007.
(47) Boczko, E. M.; Brooks, C. L. *J. Phys. Chem.* **1993,** *97*, 4509-4513.
(48) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comp. Chem.* **1992,** *13*, 1011-1020.
(49) Grossfield, A. http://membrane.urmc.rochester.edu/Grossfield_Lab/Software.html. http://membrane.urmc.rochester.edu/Grossfield_Lab/Software.html.
(50) Clementi, E.; Raimondi, D. L.; Reinhardt, W. P. *J. Chem. Phys.* **1967,** *47*, 1300-1307.

(51) Caminiti, R.; Licheri, G.; Piccaluga, G.; Pinna, G. *J. Appl. Cryst.* **1972,** *12*, 34-38.
(52) Tongraar, A.; Rode, B. M. *Chem. Phys. Lett.* **2005,** *409*, 304-309.
(53) Piquemal, J. P.; Perera, L.; Cisneros, G. A.; Ren, P.; Pedersen, L. G.; Darden, T. A. *J. Chem. Phys.* **2006,** *125*, 054511.
(54) Lightstone, F. C.; Schwegler, E.; Hood, R. Q.; Gygi, F.; Galli, G. *Chem. Phys. Lett.* **2001,** *343*, 549-555.
(55) Pavlov, M.; Siegbahn, P. E. M.; Sandstrom, M. *J. Phys. Chem. A* **1998,** *102*, 219-228.
(56) Jalilehvand, F.; Spangberg, D.; Lindqvist-Reis, P.; Hermansson, K.; Persson, I.; Sandstrom, M. *J. Am. Chem. Soc.* **2001,** *123*, 431-441.
(57) Fulton, J. L.; Heald, S. M.; Badyal, Y. S.; Simonson, J. M. *J. Phys. Chem. A* **2003,** *107*, 4688-4696.
(58) Koneshan, S.; Rasaiah, J. C.; Lynden-Bell, R. M.; Lee, S. H. *J. Phys. Chem. B* **1998,** *102*, 4193-4204.
(59) Tongraar, A.; Liedl, K. R.; Rode, B. M. *J. Phys. Chem. A* **1997,** *101*, 6299-6309.
(60) Seward, T. M.; Henderson, C. M. B.; Charnock, J. M.; Driesner, T. *Geochim. Cosmochim. Acta* **1999,** *63*, 2409-2418.
(61) Parkman, R. H.; Charnock, J. M.; Livens, F. R.; Vaughan, D. J. *Geochim. Cosmochim. Acta* **1998,** *62*, 1481-1492.
(62) Caminiti, R.; Musinu, A.; Paschina, G.; Pinna, G. *J. Appl. Cryst.* **1982,** *15*, 482-487.
(63) Larentzos, J. P.; Criscenti, L. J. *J. Phys. Chem. B.* **2008,** *112*, 14243-14250.
(64) Koneshan, S.; Rasaiah, J. C.; Lynden-Bell, R. M.; Lee, S. H. *J. Phys. Chem. B* **1998,** *102*, 4193-4204.
(65) Bleuzen, A.; Pittet, P.-A.; Helm, L.; Merbach, A. E. *Magn. Reson. Chem.* **1997,** *35*, 765-773.
(66) Marcus, Y., *Ion Solvation*. John Wiley & Sons: Chichester, 1985.
(67) Lumb, R. F.; Martell, A. E. *J. Phys. Chem.* **1953,** *57*, 690-693.
(68) Schmidbaur, H. *Angew. Chem. Int. Ed.* **1990,** *29*, 1090.
(69) Nielsen, A. E. *J. Cryst. Growth* **1984,** *67*, 289-310.
(70) Sethmann, I.; Putnis, A.; Grassmann, O.; Löbmann, P. *Amer. Miner.* **2005,** *90*, 1213-1217.
(71) Friddle, R. W.; Weaver, M. L.; Qiu, S. R.; Wierzbicki, A.; Casey, W. H.; De Yoreo, J. J. *Proc. Natl. Acad. Sci. USA* **2010,** *107*, 11-15.
(72) Gebauer, D.; Volkel, A.; Colfen, H. *Science* **2008,** *322*, 1819-1822.
(73) Pouget, E. M.; Bomans, P. H.; Goos, J. A.; Frederik, P. M.; de With, G.; Sommerdijk, N. A. *Science* **2009,** *323*, 1455-1458.
(74) Wang, D.; Wallace, A. F.; DeYoreo, J. J.; Dove, P. M. *Proc. Natl. Acad. Sci. USA* **2009,** *106*, 21511-21516.
(75) Conway, B. E., *Ionic hydration in chemistry and biophysics*. Elsevier Scientific: Amsterdam, 1981.
(76) Matthews, E. K., Calcium and Hormone Release. In *A Symposium on Calcium and Cellular Function*, Cuthbert, A. W., Ed. St. Martin's Press: New York, 1970.
(77) Rubin, R. P.; Feinstein, M. B.; Jaanus, S. D.; Paimre, M. *J. Pharmacol. Exp. Ther.* **1967,** *155*, 463-471.
(78) Rahamimoff, R., Role of Calcium Ions in Neuromuscular Transmission. In *A Symposium on Calcium and Cellular Function*, Cuthbert, A. W., Ed. St. Martin's Press: New York, 1970.

(79) Dodge, F. A., Jr.; Miledi, R.; Rahamimoff, R. *J. Physiol.* **1969,** *200*, 267-283.

Chapter 3 Influence of Aspartic Acid on the Kinetics of Water Exchange About Alkaline Earth Cations

Abstract. The mineralization of tissues into calcified skeletal structures is associated with macromolecules that contain a large fraction of acidic residues, including aspartate (Asp). Recent investigations suggest these biomolecules influence the nucleation, growth, and composition of biominerals by regulating the kinetics and thermodynamics of local cation solvation. This study utilizes molecular dynamics (MD) to investigate the impact of Asp on the water exchange rates of group IIA cations. Water residence times of four divalent cations (Mg^{2+}, Ca^{2+}, Sr^{2+} and Ba^{2+}) were calculated using reactive flux and time correlation functions. Residence times were then recalculated for the interaction of hydrated cations with a single Asp molecule as solvent-separated ion pairs (SSIP) or contact ion pairs (CIP). In the presence of Asp, SSIP and CIP formation decreases the primary and secondary sphere water residence times about Ca^{2+}, Sr^{2+}, and Ba^{2+}. The residence times of secondary sphere waters about Mg^{2+} are decreased by Asp, but no measurable effect on primary sphere waters is observed. We estimate an approximately 10% decrease in the primary sphere waters of Mg^{2+} by extrapolating trends in water exchange rates determined for the other IIA cations. The calculations demonstrate the ability of charged biomolecules to enhance hydration shell lability about IIA cations. These findings demonstrate a plausible physical basis for how carboxylated biomolecules increase calcite and barite growth rates and suggest roles in promoting the formation of amorphous precursor phases.

3.1 Introduction

Biominerals are composite structures that consist of crystalline and amorphous phases in intimate association with an organic matrix. Across diverse taxa, a shared feature of the macromolecules comprising this matrix is an abundance of the acidic amino acids[1], Asp and Glu. This well-known association of carboxylated macromolecules with mineralized tissues has led to considerable investigation aimed at deciphering specific roles of biomolecules in mineral nucleation[2, 3] and growth processes[4-6] as well as the accumulation and transformation of amorphous precursors[7-9]. Acidic biomolecules are also implicated in influencing the crystal phase[10-12] and morphology[13-15] of the biominerals that develop.

Previous studies recognized that the physical basis for how biomolecules regulate mineralization may be related to their effects upon cation solvation. Experimental investigations hypothesized that Asp and Asp-containing peptides enhance the growth rates of calcite[16] and barite[17] by contributing to the dehydration of calcium and barium, respectively. A similar explanation is proposed by Stephenson et. al.[18] to explain the ability of carboxylate-rich polypeptides to enhance the incorporation of Mg^{+} into calcite relative to Ca^{2+}. These hypotheses are consistent with the idea that the kinetic barrier to attaching growth units to a crystal surface is related to cation dehydration[19-21]; however, such experimental studies provide only limited insights into the specific effects of biomolecules on cation hydration.

Cation solvation may also play a role in mineralization by non-classical pathways that involve the formation of amorphous precursor phases. The acidic biomolecules commonly found in calcifying organisms are linked to both the formation and stabilization of amorphous carbonate and phosphate phases[22-24]. Furthermore, experimental studies demonstrate a complex interplay between Asp-containing molecules and Mg^{2+} in regulating the formation, stabilization and transformation of these phases[22, 23]. Cation dehydration is also thought to be a crucial step in the earliest stages of mineral formation because the initial aggregation of ions into ion pairs or prenucleation clusters[25, 26] involves significant changes in solvation. Gebauer et al.[25] speculated that prenucleation clusters are favored by the gain in entropy that occurs when water molecules are released from ion hydration shells.

Support for the idea that biomolecules impact cation hydration to influence mineralization processes is found in studies on the thermodynamics of cation solvation. MD simulations by Piana et al.[27] showed that surface-bound Asp decreases the barrier to Ba^{2+}

dehydration and coordination to the (001) surface of barite. In a study focused on calcification, Hamm et al.[28] demonstrated that Asp promotes the partial dehydration of Ca^{2+} and Sr^{2+} at separation distances as large as 4.9 Å and 5.1 Å. Hence, CIP formation was not required to alter the solvation environment. This work also found the energetic barrier to displacing primary sphere waters of Ca^{2+} and Sr^{+} by Asp is significantly lowered compared with the barrier for water exchange in the bulk solution. Both studies, however, are limited because these thermodynamic barriers describe ion association, not dehydration, and they do not directly show a change in the kinetics of water exchange. Little is known about the effects of biomolecules on water exchange rates of the alkaline earth cations, despite the recognized importance of cation water exchange rates in crystal growth[20, 29].

The purpose of this study is to determine the effect of Asp on the kinetics of water exchange about the alkaline earth cations that are significant in biomineralization. We first establish baseline estimates of primary and secondary sphere water exchange rates for the group IIA cations Mg^{2+}, Ca^{2+}, Sr^{2+} and Ba^{2+}. From these "control" simulations, we then determine the effect of Asp on water exchange rates. Asp is used as a simple model compound for the larger caboxylated macromolecules common to biomineralization environments. Previous studies demonstrate that simple model systems using amino acids and peptides can provide insights into the influence of macromolecules on mineralization[16]. We focus on Ca^{2+} due to its integral role in the calcification of diverse biominerals, but also consider Mg^{2+}, Sr^{2+} and Ba^{2+} because these cations occur as minor and trace elements in biominerals. They also have important roles as proxies for the environmental conditions of mineral formation[30]. Additionally, this suite of cations allows us to determine systematic effects of Asp on ion hydration across the alkaline earth elements.

3.2 Methods

All calculations were performed with the LAMMPS[31] MD software on the Virginia Tech SGI Altix 3700 clusters and utilized the canonical (NVT) ensemble with Nosé-Hoover thermostat set to 300 K. Simulations employed the CHARMM22[32] force field because this set of potentials contains parameters for all of the common amino acids and was developed for use with proteins. Energy contributions from bonds, angles, improper dihedrals and 1-3 Urey-Bradley interactions were described by harmonic potentials, while a cosine series accounted for

dihedral interactions. A 6 – 12 Lennard-Jones (LJ) potential, and a Coulombic term comprised the nonbonded terms and long-range Coulombic interactions were calculated using the Ewald summation method. LJ parameters describing the interaction between metal ions and peptides were obtained with the Lorentz-Berthelot combination rules.

Description of the solvent was carried out with the TIP3P[33] model of water while keeping the O – H bond lengths and H – O – H angles fixed by the SHAKE[34] algorithm. This approach is implemented in LAMMPS and well suited for use with the CHARMM22 force field. LJ parameters for the interaction of metal ions with water oxygen atoms were calculated by Åqvist[35] from free energy perturbation in order to reproduce experimental hydration free energies and radial distribution functions for alkali and alkaline earth metals. These parameters were modified for use with TIP3P water using Lorentz-Berthelot rules and converted to LJ potential well depth, ε, and the distance of zero inter-particle potential, σ, using $A = 4\varepsilon\sigma^{12}$ and $B = \varepsilon\sigma^{6}$, where A and B are the parameters of the simplified 6-12 LJ potential. The resulting parameters used in this study are presented in **Table 3.1**. The reliability of this parameter set in reproducing hydration free energies with TIP3P water has been demonstrated[36], and such an approach has previously been applied to study Ca^{2+} and Mg^{2+} hydration[37]. Allouche et al.[38] have shown that the combination of Åqvist parameters, TIP3P water model, and CHARM22 force field is appropriate for describing Ca^{2+} and Mg^{2+} binding to an Asp-containing protein and indicated that hydration properties reported in the original Åqvist study are reliably reproduced.

Table 3.1. Summary of ion 6-12 LJ parameters, A_{ion} and B_{ion}, ion-water oxygen LJ parameters calculated for TIP3P water, A_{io} and b_{io}, and corresponding ε and σ values used in this study.

	A_i[a]	B_i[a]	A_{io}	B_{io}	ε	σ
Mg^{2+}	37	8.32	29352	208.1	0.36879	2.2816
Ca^{2+}	264.1	18.82	209511	470.7	0.26436	2.7632
Sr^{2+}	613.5	20.54	486690	513.7	0.13556	3.1339
Ba^{2+}	1341.5	24.13	1064212	603.5	0.08556	3.4758

[a]Ref. 35

Baseline cation hydration numbers, cation – oxygen distances, and reactive flux primary sphere water residence times in the absence of Asp were determined as reported by Hamm et

al.[39]. Solvation properties and reactive flux residence times calculated with this method and set of paramters, along with experimentally determined values for comparison, are presented in **Tables 3.2 and 3.3**.

Table 3.2. Calculated solvation sphere locations and corresponding hydration numbers for the computational model used in this work. Experimental values from XRD and EXAFS are provided for comparison where available.

	primary sphere				secondary sphere			
	r_1 (Å) calc.	r_1 (Å) exp.	n_1 calc.	n_1 exp.	r_2 (Å) calc.	r_2 (Å) exp	n_2, calc.	n_2, exp.
Mg^{2+}	2.00[a]	2.11[c]	6.0[a]	6[c]	4.13[a]	4.1[h]	12.8[a]	12[h]
Ca^{2+}	2.41[a]	2.46[d]	8.0[a]	8[d]	4.58[a]	4.58[d]	15.4[a]	13[d]
Sr^{2+}	2.57[a]	2.57[e]	8.4[a]	8.3[f]	4.73[a]	4.91[g]	17.7[a]	15[g]
Ba^{2+}	2.76[b]	2.82[i]	8.8[b]	8.1[i]	4.98[b]	-	19.2[b]	-

[a]Ref. 28 [b]This work [c]Ref. 40 [d]Ref. 41 [e]Ref. 42
[f]Ref. 43 [g]Ref. 44 [h]Ref. 45 [i]Ref. 46

Table 3.3. Calculated transition state rate constants, transmission coefficients and resulting water exchange rate constants and water residence times for primary sphere waters determined by the reactive flux method. Experimental estimates of water residence times from NMR are provided for comparison.

	k^{TST} (s^{-1})	κ	k (s^{-1})	τ (s), calc.	τ (s), exp
Mg^{2+}	5.5×10^{6}	0.24	1.32×10^{6}	7.58×10^{-7}	6.7×10^{-6} [a]
Ca^{2+}	1.5×10^{10}	0.30	4.50×10^{9}	2.22×10^{-10}	3.2×10^{-9} [b]
Sr^{2+}	5.8×10^{10}	0.24	1.39×10^{10}	7.18×10^{-11}	2.5×10^{-9} [b]
Ba^{2+}	6.7×10^{10}	0.31	2.08×10^{10}	4.81×10^{-11}	1.3×10^{-9} [b]

[a]Ref. 47 [b]Ref. 48

To determine the effects of Asp on hydration dynamics, water residence times were also calculated using the method of Impey et. al.[49] This allowed for the calculation of both primary and secondary sphere residence times and was more suitable for MD systems including organic species. Because the interaction of Asp with water molecules prevented the calculation of a

single, reliable $W(r)$ describing water exchange, we found that the metal – oxygen distance was not an adequate reaction coordinate to describe the water exchange mechanism in metal – Asp systems.

Using the Impey method, residence times were computed from time correlation functions defined as,

$$R(t) = \frac{1}{N_h} \sum_{i=1}^{N_h} \left[\theta_i(0) \theta_i(t) \right]$$

Equation 3.1. Expression for calculation of the residence correlation function using the Impey method.

where N_h = the number of water molecules in the hydration sphere at time zero. $\theta_i(t)$ = the Heaviside function, which is assigned a value of 1 if the water molecule *i* is in the ion's hydration sphere at time *t* and 0 if it is not. Waters undergoing temporary excursions from the hydration shell, but not involved in an actual exchange event, were counted as present when computing $\theta_i(t)$; temporary excursions were defined as those lasting less than 2 ps.

Input systems for calculating $R(t)$ consisted of 475 water molecules, one Asp molecule, one divalent cation, and a chloride anion for charge balance. Asp was represented in zwitterionic form with full degrees of freedom. The $R(t)$ was calculated for the interaction of Mg^{2+}, Ca^{2+}, Sr^{2+} and Ba^{2+} with Asp for three different states: infinite separation of the cation and Asp, solvent separated ion pair (SSIP), and contact ion pair (CIP). For infinite ion – Asp separation, the correlation functions were also calculated using a smaller water box without Asp to ensure that system size was not affecting water residence times. Simulations that included Asp employed a weak harmonic restraining potential to restrict the cation – Asp distance to the desired separation. The distances corresponding to SSIPs and CIPs for each cation were determined from the locations of minima in cation – Asp potential of mean force profiles[28] and confirmed by monitoring ion, water and Asp positions throughout the duration of each simulation.

Using a timestep of 1.0 fs, systems were equilibrated for 300 ps followed by collection of ion and water positional coordinates every 50 fs. Simulation lengths after equilibration varied from 1 ns for Ba^{2+} to 50 ns for Mg^{2+} because longer water residence times require a longer

period of time to observe full exchange of the hydration sphere. The calculated $R(t)$ were averaged over 190 – 300 simulations until additional simulations did not change the shape of the decay function. The ensemble-averaged correlation functions, $\langle R(t) \rangle$, were fit to exponential functions, which were used to compute water residence time, τ, from:

$$\tau = \int_0^\infty \langle R(t) \rangle dt$$

Equation 3.2. Relationship between residence time and the residence correlation function.

Residence times were calculated by this method for both primary and secondary sphere waters of hydration. Python routines written by the author were used to process simulation output and calculate water residence times (**Appendix 2**).

3.3 Results & Discussion

3.3.1 Baseline Water Residence Times

Results of reactive flux calculations (**Table 3.3**) show that the calculated transition state theory rate constants and water residence times for IIA cations follow the expected trends based on their size and charge density. This trend is reflected in the potential of mean force profiles for a primary sphere water exchange event (**Figure 3.1a**), which demonstrates that the largest energetic barrier for water exchange corresponds to Mg^{2+} while Ba^{2+} has the smallest barrier. The normalized reactive flux, used to determine transmission coefficient, converges to a steady value within 3 ps for all four cations (**Figure 3.1b**). The calculated primary sphere water residence times follow the same trend as experimentally determined values (**Table 3.3**), although residence times are underestimated by approximately an order of magnitude for all four cations[47, 48].

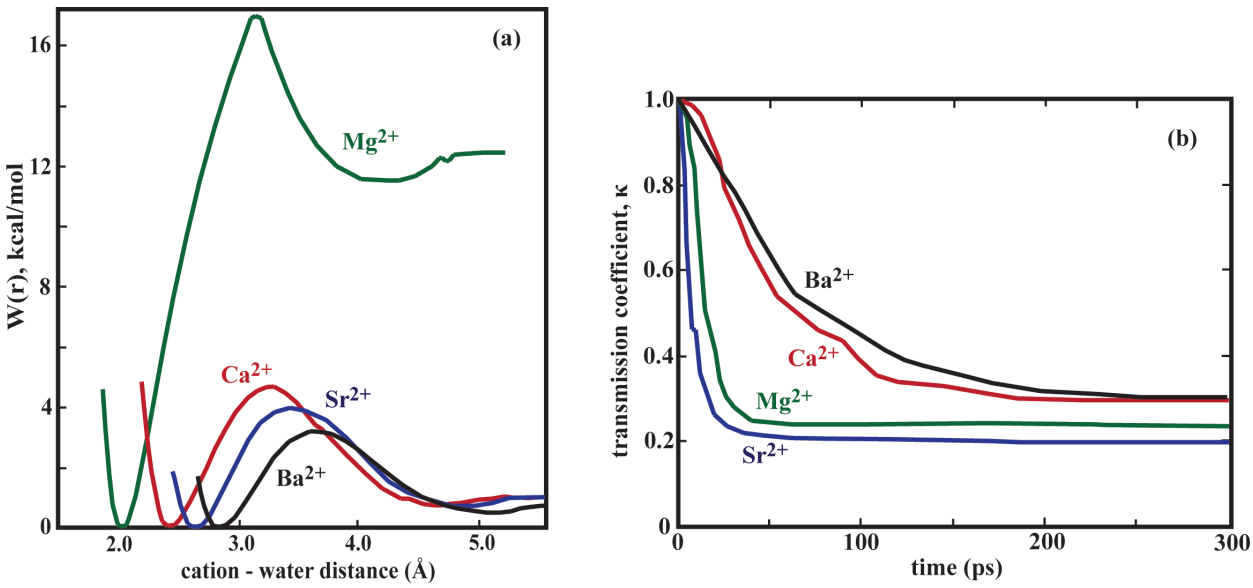


Figure 3.1. (a) Potential of mean force profiles show the energetic barrier to primary sphere water exchange about alkaline earth cations; (b) The normalized reactive flux for each cation, used to calculate transmission coefficients, converges over 3 ps simulations.

The primary and secondary sphere water residence times calculated with the Impey method also follow the expected trends based on cation size (**Table 3.4**). For Ca^{2+}, Sr^{2+} and Ba^{2+}, the averaged correlation functions for primary sphere waters decay over 2 ns (**Figure 3.2a**), indicating a high probability for full exchange of the primary sphere waters of hydration over the course of the simulation. Trajectories of primary sphere waters throughout individual simulations show the water exchange events (**Appendix 2, Figure A2.1**).

Table 3.4. Summary of primary and secondary sphere water residence times calculated from residence correlation functions for baseline conditions (no Asp) and cation – Asp solvent-separated ion pairs (SSIP) and contact ion pairs (CIP).

	primary sphere residence time (s)			secondary sphere residence time (s)		
	no asp	SSIP	CIP	no asp	SSIP	CIP
Mg^{2+}	$>2.5\times10^{-8}$	$>2.5\times10^{-8}$	$>2.5\times10^{-8}$	2.14×10^{-11}	1.74×10^{-11}	1.48×10^{-11}
Ca^{2+}	6.66×10^{-10}	5.12×10^{-10}	2.29×10^{-10}	1.15×10^{-11}	9.3×10^{-12}	8.2×10^{-12}
Sr^{2+}	3.16×10^{-10}	1.38×10^{-10}	7.5×10^{-11}	1.0×10^{-11}	8.0×10^{-12}	6.7×10^{-12}
Ba^{2+}	1.48×10^{-10}	7.3×10^{-11}	5.3×10^{-11}	4.5×10^{-12}	3.5×10^{-12}	2.8×10^{-12}

These plots show full exchange of the primary sphere for Ca^{2+}, Sr^{2+} and Ba^{2+} and suggest faster exchange of waters as cation size increases. Correlation functions for the much shorter-lived secondary sphere waters of hydration are seen to decay over a 100 ps timeframe for all four cations (**Figure 3.2b)**.

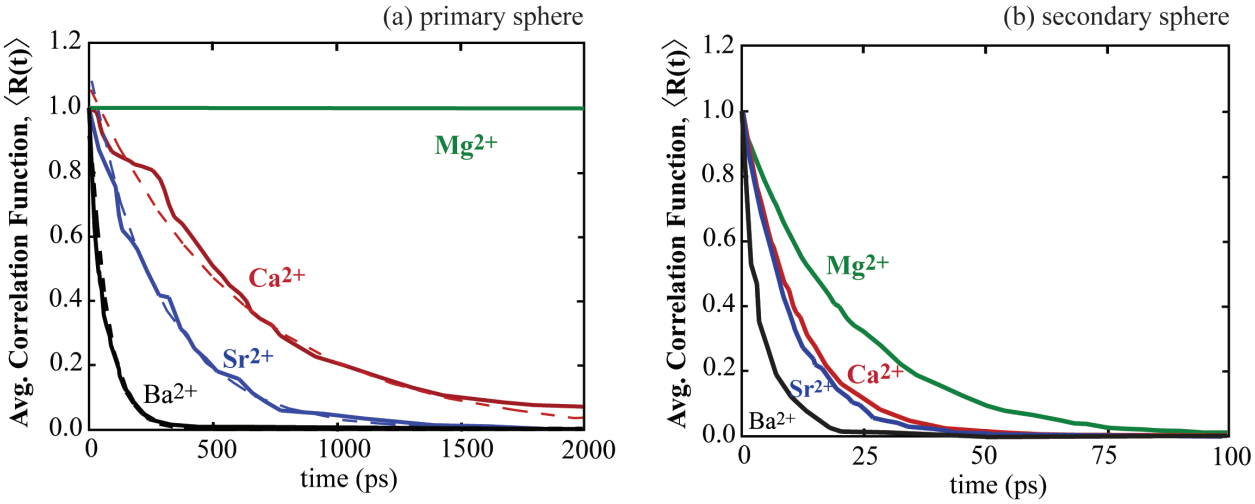


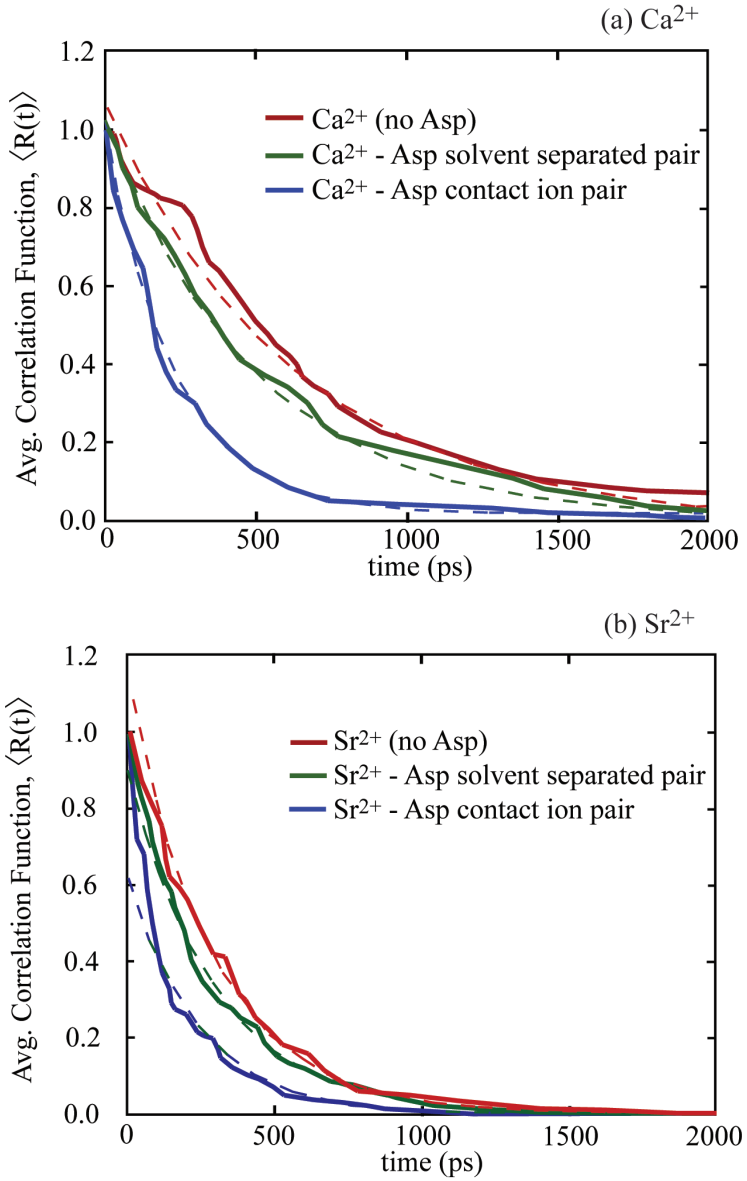
Figure 3.2. Calculated ensemble-averaged time correlation functions for (a) primary sphere and (b) secondary sphere hydration waters show the relative rates of water exchange for group IIA cations. The $\langle R(t) \rangle$ for Mg^{2+} in (a) does not decay, indicating that no primary sphere waters were exchanged during the simulations. Dashed lines in (a) show the fit of exponential decay functions to $\langle R(t) \rangle$.

The primary sphere residence time determined for Mg^{2+} is uncertain because the correlation function for Mg^{2+} does not decay over the 2 ns time frame (**Figure 3.2a**). Rather, the correlation function for Mg^{2+} retains a value of 1.0, indicating that no water exchange occurs. Longer simulations (25 ns – 50 ns) with Mg^{2+} produce the same result, therefore primary sphere residence time for Mg^{2+} for this method is reported as > 25 ns (**Table 3.4**). Analysis of cation and water trajectories confirms that no water molecules exchange in or out of the primary hydration sphere of Mg^{2+} during a 25 ns simulation (**Figure A2.1a**). While it is possible that the combination of the interatomic potentials with the Impey method underestimates the lability of the Mg^{2+} primary hydration sphere, the result concurs with other MD studies[50, 51] employing a different water model and ^{17}O NMR measurements that determine a residence time on the order

of microseconds[47]. Additionally, recalculation using an alternate Mg^{2+} – O interaction potential[52] and water model (SPC) yielded the same result. This outcome suggests that conventional MD approaches are not unable to determine Mg^{2+}water residence times due to the long length of simulation that would be needed to observe water exchange. Such an interpretation is also supported by the long residence time of 0.76 μs determined by reactive flux.

3.3.2 Effect of Asp on Water Residence Times

For Ca^{2+}, Sr^{2+} and Ba^{2+}, interaction with Asp as an SSIP or CIP results in faster decay of the residence correlation functions for primary sphere hydration waters (**Figure 3.3a-c**).



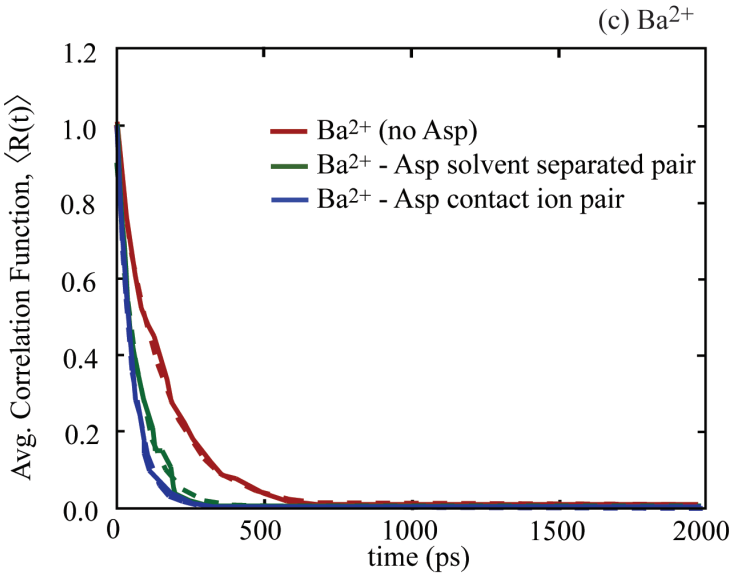
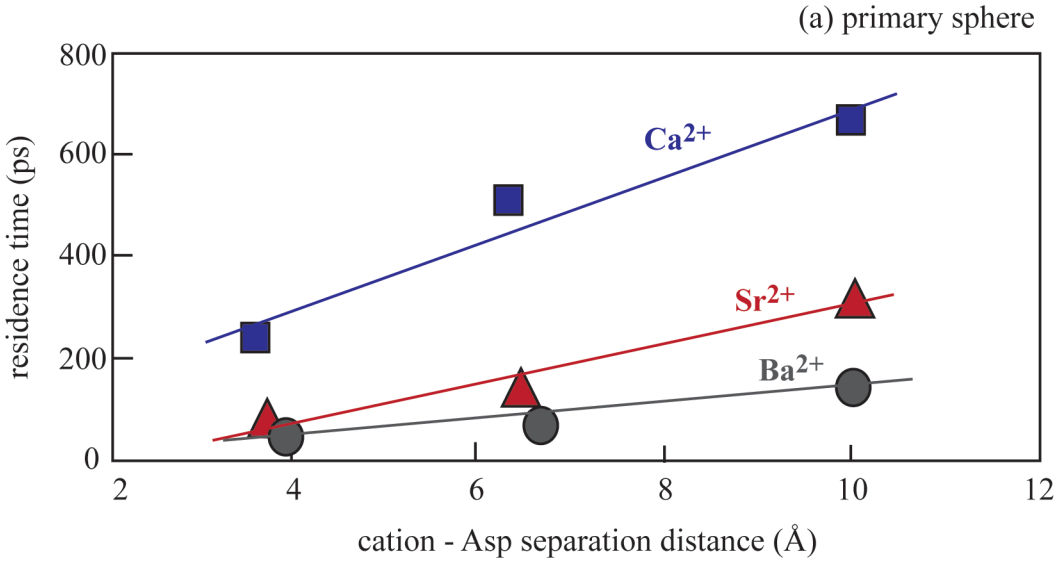


Figure 3.3. Ensemble-averaged time correlation functions for primary sphere hydration waters about (a) Ca^{2+}, (b) Sr^{2+} and (c) Ba^{2+} decay faster when the cations interact with Asp as an SSIP (green lines) or a CIP (blue lines) than when Asp is not present (red lines).

For SSIP formation, this translates into a greater reduction in residence time for Sr^{2+} and Ba^{2+} (56% and 51% decrease in residence time, respectively) compared to the 23% decrease seen for Ca^{2+} (**Table 3.4, Figure 3.4a**). In the case of Asp – cation CIP formation, primary sphere residence times for Ca^{2+}, Sr^{2+} and Ba^{2+} are decreased by 66%, 76% and 64%, respectively (**Table 3.4, Figure 3.4a)**.



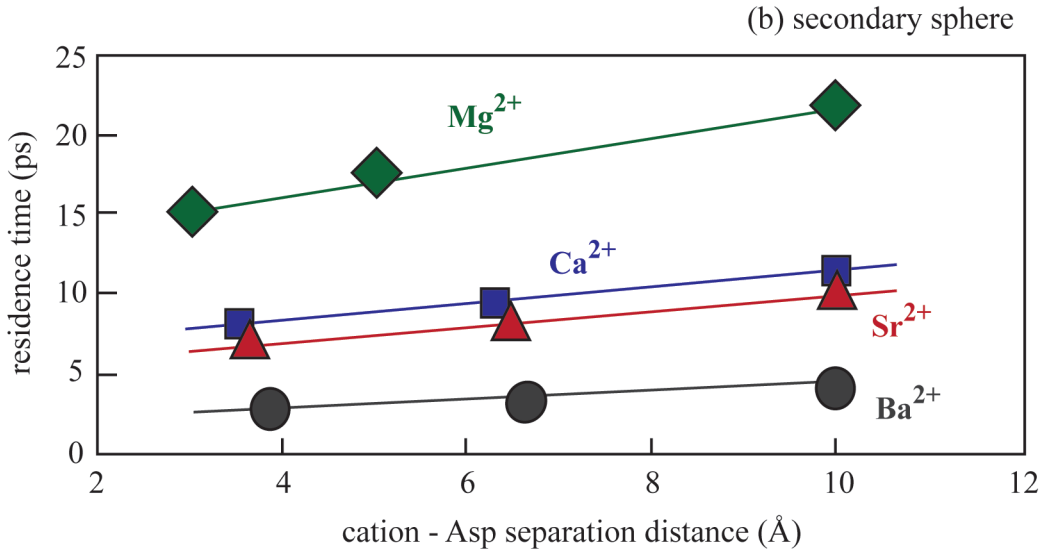


Figure 3.4. Residence times of (a) primary and (b) secondary sphere hydration waters decrease with cation – Asp separation distance. The plotted distances correspond to the average separation between the cation and Asp center of mass for simulations of a M^{2+} – Asp CIP or SSIP, or for simulations without cation – Asp interaction (large separation).

This result for CIPs is in agreement with evidence that the reactivity of metal-oxygen bonds increases when a carboxylate ligand is present in the primary coordination sphere of Ni^{2+} [53]. In contrast to the larger cations, Mg^{2+} interaction with Asp does not result in an observable increase in primary hydration sphere mobility. As in the baseline Mg^{2+} simulations without Asp, no primary sphere exchange events occur and residence time is again reported as $>2.5 \times 10^4$ ps. (**Table 3.4**).

To approximate the residence time effect of SSIP formation on primary sphere waters of Mg^{2+}, we first noted that the relative decrease in residence time of secondary sphere waters is correlated with cation charge density[54] (**Figure 3.5**). Assuming that this trend is indicative of a similar correlation for primary sphere waters, extrapolation to Mg^{2+} predicts an approximately 10% decrease in primary sphere residence time. Based on the reactive flux result for Mg^{2+}, the resulting residence time is around 6.6×10^{-7} s, which is still longer than the time scale accessible with direct residence time calculation. This decrease is only a small perturbation in bulk solution, but the effect may be even more significant at an interface or near a high charge density region on a biomolecule.

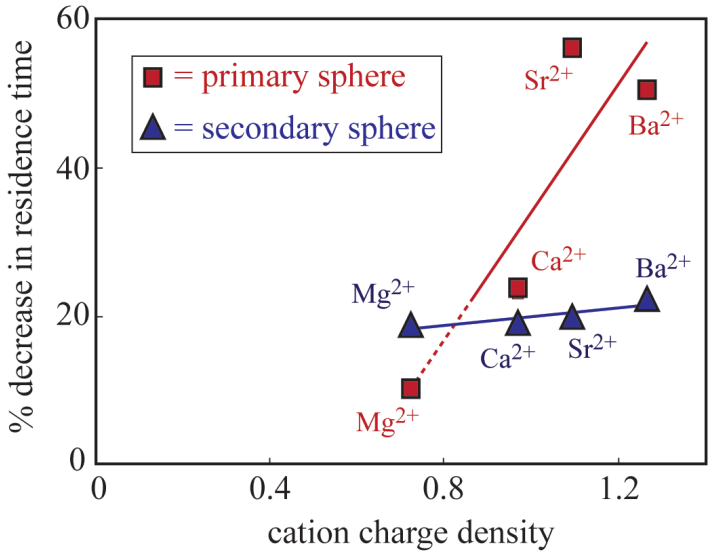


Figure 3.5. Cation – Asp SSIP formation decreases the residence time of primary sphere waters by as much as 56% (red squares) and reduces the lifetime of secondary sphere waters by up to 22% (blue triangles). Extrapolating the trend for primary sphere water molecules to the charge density of Mg^{2+} (dashed red line) predicts a decrease of approximately 10%.

Residence times of the secondary sphere hydration waters about group IIA cations are reduced by Asp (**Figure 3.3b**). The formation of a SSIP with Asp reduces the secondary sphere residence times by 18-20%, while CIP formation decreases the residence times 29-38% (**Table 3.4**). The time correlation functions corresponding to these decreases are shown in **(Appendix 2, Figure A2.2**).

To further test the ability of our model to predict the binding behavior of Asp to divalent cations, we calculated binding constants from Asp – cation free energy profiles[55]. The results demonstrate that our simulations reproduce the trend in the Asp – cation binding affinity predicted by experiment (**Table 3.5**). We determine a binding constant for Ca^{2+} (log K_{calc} = 1.63) that is very close to the experimentally measured value at 25°C (log K_{exp} = 1.60)[56]. The calculated constant for Mg^{2+} (log K_{calc} = 2.83) correctly predicts that Asp has a much stronger affinity for Mg^{2+} than for Ca^{2+} but overestimates the stability of the Mg^{2+}-Asp CIP when compared to the experimental value (log K_{exp} = 2.43)[56]. This overestimate of CIP stability does not impact estimates of water residence time because complexation equilibria are not involved in the Impey residence time calculations.

Table 3.5. Calculated stability constants for M^{2+} - Asp contact ion pairs (K_{calc}) and experimental values (K_{exp}) for comparison.

	log K_{calc}	log K_{exp}[a]
Mg^{2+}	2.83	2.43
Ca^{2+}	1.63	1.60
Sr^{2+}	1.59	1.48
Ba^{2+}	1.32	1.14

[a]Ref. 56

3.4 Implications for mineralization

The simulations show that Asp significantly decreases the water residence times about Ca^{2+}, Sr^{2+}, and Ba^{2+} and demonstrate the ability of a small, charged biomolecule to enhance hydration shell lability. To our knowledge, this is the first estimate of biomolecule effects upon the kinetics of alkaline earth cation solvation. The observed effects of Asp are consistent with experiments showing enhanced calcite and barite growth rates in the presence of Asp because factors that enhance water exchange are expected to catalyze cation dehydration. Findings in this study also demonstrate that soluble organic species, whose functional group chemistry is prevalent in biomineralizing microenvironments, may be active in promoting cation dehydration without the requirement of surface binding. This idea is supported by reports that certain acidic proteins, which are only weakly attracted to mineral surfaces, increase the growth rate of calcium oxalate monohydrate[57].

The findings also suggest that Asp-rich macromolecules may promote the formation of mineral precursors by contributing disorder to the solvent, favoring ion pair and cluster formation. As noted earlier, the calculations show that Asp enhances the mobility of the hydration waters about Ca^{2+}, even without cation binding as a CIP. It is plausible that carboxylated molecules contribute to ACC or cluster formation by enhancing the rate of cation diffusion to existing clusters. This is because trends in cation water exchange rates correlate strongly with the mobility of those ions in water[58] and the growth of amorphous calcium carbonate (ACC) is thought to be diffusion controlled[59].

The cation-specific effects of Asp on water exchange kinetics may also provide insight into the complex interplay between Ca^{2+}, Mg^{2+}, and biomolecules during the formation and transformation of amorphous carbonate and phosphate phases. Mg^{2+} is known to inhibit the rate of crystallization of metastable ACC and amorphous calcium phosphate (ACP)[22, 23]. We find that the residence times of secondary sphere water molecules about Mg^{2+} decrease due to SSIP or CIP formation with Asp. However, no residence time effect on primary sphere waters is observed, most likely because the long residence time of Mg^{2+}, is inaccessible by direct simulation regardless of interaction with Asp. These results suggest that the stabilization of ACC and ACP by Mg^{2+} arises from the large time scales required for the dehydration of magnesium, which must occur during the transition from a hydrated amorphous phase to an anhydrous crystalline phase. Such an interpretation is supported by recent measurements showing that the water content of ACC increases with magnesium incorporation[60].

3.5 Acknowledgment

This research was supported by the U.S. Department of Energy (Grant FG02-00ER15112) and the National Science Foundation Graduate Research Fellowship Program. Computational resources were provided by the Virginia Tech Advanced Research Computing Center.

3.6 References

(1) Weiner, S.; Troub, W.; Lowenstam, H., *Biominerlization and Biological Metal Accumulation*. Reidel Publishing Company: Dordrecht, 1983.
(2) Levi, Y.; Albeck, S.; Brack, A.; Weiner, S.; Addadi, L. *Chem. Eur. J.* **1998,** *4*, 389-396.
(3) Belcher, A. M.; Wu, X. H.; Christensen, R. J.; Hansma, P. K.; Stucky, G. D.; Morse, D. E. *Nature* **1996,** *381*, 56-58.
(4) Puverel, S.; Tambutté, E.; Zoccola, D.; Domart-Coulon, I.; Bouchot, A.; Lotto, S.; Allemand, D.; Tambutté, S. *Coral Reefs* **2005,** *24*, 149-156.
(5) Gotliv, B.-A.; Kessler, N.; Sumerel, J. L.; Morse, D. E.; Tuross, N.; Addadi, L.; Weiner, S. *ChemBioChem.* **2005,** *6*, 304-314.
(6) Levi-Kalisman, Y.; Falini, G.; Addadi, L.; Weiner, S. *J. Struct. Bio.* **2001,** *135*, 8-17.
(7) Aizenberg, J.; Addadi, L.; Weiner, S.; Lambert, G. *Adv. Mater.* **1996,** *8*, 222-226.
(8) Raz, S.; Hamilton, P. C.; Wilt, F. H.; Weiner, S.; Addadi, L. *Adv. Funct. Mater.* **2003,** *13*, 480-486.

(9) Politi, Y.; Mahamid, J.; Goldberg, H.; Weiner, S.; Addadi, L. *CrystEngComm.* **2007,** *9*, 1171-1177.
(10) Falini, G.; Albeck, S.; Weiner, S.; Addadi, L. *Science* **1996,** *271*, 67-69.
(11) Falini, G.; Fermani, S.; Vanzo, S.; Miletic, M.; Zaffino, G. *Eur. J. Inorg. Chem.* **2005**, 162-167.
(12) Suzuki, M.; Saruwatari, K.; Kogure, T.; Yamamoto, Y.; Nishimura, T.; Kato, T.; Nagasawa, H. *Science* **2009,** *325*, 1388-1390.
(13) Volkmer, D.; Fricke, M.; Huber, T.; Sewald, N. *Chem. Commun.* **2004**, 1872-1873.
(14) Wada, N.; Yamashita, K.; Umegaki, T. *J. Coll. Int. Sci.* **1999,** *212*, 357-364.
(15) Albeck, S.; Addadi, L.; Weiner, S. *Connect. Tissue Res.* **1996,** *35*, 365-370.
(16) Elhadj, S.; De Yoreo, J. J.; Hoyer, J. R.; Dove, P. M. *Proc. Natl. Acad. Sci. USA* **2006,** *103*, 19237-19242.
(17) Kowacz, M.; Putnis, C.; Putnis, A. *Geochim. Cosmochim. Acta* **2007,** *71*, 5168-5179.
(18) Stephenson, A. E.; DeYoreo, J. J.; Wu, L.; Wu, K. J.; Hoyer, J.; Dove, P. M. *Science* **2008,** *322*, 724-727.
(19) Nielsen, A. E.; Toft, J. M. *J. Cryst. Growth* **1984,** *67*, 278-288.
(20) Petsev, D. N.; Chen, K.; Gliko, O.; Vekilov, P. G. *Proc. Natl. Acad. Sci. USA* **2003,** *100*, 792-796.
(21) Sethmann, I.; Putnis, A.; Grassmann, O.; Löbmann, P. *Am. Miner.* **2005,** *90*, 1213-1217.
(22) Tao, J. H.; Zhou, D. M.; Zhang, Z. S.; Xu, X. R.; Tang, R. K. *Proc. Natl. Acad. Sci. USA* **2009,** *106*, 22096-22101.
(23) Aizenberg, J.; Lambert, G.; Weiner, S.; Addadi, L. *J. Am. Chem. Soc.* **2002,** *124*, 32-39.
(24) Olszta, M. J.; Odom, D. J.; Douglas, E. P.; Gower, L. B. *Connect Tissue Res.* **2003,** *44*, 326-334.
(25) Gebauer, D.; Völkel, A.; Cölfen, H. *Science* **2008,** *322*, 1819-1822.
(26) Pouget, E. M.; Bomans, P. H. H.; Goos, J. A. C. M.; Frederik, P. M.; With, G. d.; Sommerdijk, N. A. J. M. *Science* **2009,** *323*, 1455-1458.
(27) Piana, S.; Jones, F.; Gale, J. D. *CrystEngComm.* **2007,** *9*, 1187-1191.
(28) Hamm, L. M.; Wallace, A. F.; Dove, P. M. *J. Phys. Chem. B* **2010,** *114*, 10488-10495.
(29) Christoffersen, J.; Dohrup, J.; Christoffersen, M. R. *J. Cryst. Growth* **1998,** *186*, 275-282.
(30) Fischer, G.; Wefer, G., *Use of proxies in paleoceanography: Examples from the South Atlantic*. Springer: Berlin, 1999.
(31) Plimpton, S. J. *J. Comp. Phys.* **1995,** *117*, 1-19.
(32) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998,** *102*, 3586-3616.
(33) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983,** *79*, 926-935.
(34) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977,** *23*, 327-341.
(35) Aqvist, J. *J. Phys. Chem.* **1990,** *94*, 8021-8024.
(36) Aqvist, J. *J. Phys. Chem.* **1994,** *98*, 8253-8255.
(37) Periole, X.; Allouche, D.; Daudey, J. P.; Sanejouand, Y. H. *J. Phys. Chem. B* **1997,** *101*, 5018-5025.
(38) Allouche, D.; Parello, J.; Sanejouand, Y. H. *J. Mol. Biol.* **1999,** *285*, 857-873.

(39) Hamm, L. M.; Wallace, A. F.; Dove, P. M. *J. Phys. Chem. B* **2010,** *114*, 10488-10495.
(40) Caminiti, R.; Licheri, G.; Piccaluga, G.; Pinna, G. *J. Appl. Cryst.* **1972,** *12*, 34-38.
(41) Jalilehvand, F.; Spangberg, D.; Lindqvist-Reis, P.; Hermansson, K.; Persson, I.; Sandstrom, M. *J. Am. Chem. Soc.* **2001,** *123*, 431-441.
(42) Seward, T. M.; Henderson, C. M. B.; Charnock, J. M.; Driesner, T. *Geochim. Cosmochim. Acta* **1999,** *63*, 2409-2418.
(43) Parkman, R. H.; Charnock, J. M.; Livens, F. R.; Vaughan, D. J. *Geochim. Cosmochim. Acta* **1998,** *62*, 1481-1492.
(44) Caminiti, R.; Musinu, A.; Paschina, G.; Pinna, G. *J. Appl. Cryst.* **1982,** *15*, 482-487.
(45) Magini, M.; Licheri, G.; Piccaluga, G.; Paschina, G.; Pinna, G., *XRD of Ions in Aqueous Solutions: Hydration and Complex Formation*. CRC Press: Boca Raton, 1988.
(46) Persson, I.; Sandstrom, M.; Yokoyama, H.; Chaudhry, M. *Z. Natur. A* **1995,** *50*, 21-37.
(47) Bleuzen, A.; Pittet, P.-A.; Helm, L.; Merbach, A. E. *Magn. Reson. Chem.* **1997,** *35*, 765-773.
(48) Marcus, Y., *Ion Solvation*. John Wiley & Sons: Chichester, 1985.
(49) Impey, R. W.; Madden, P. A.; Mcdonald, I. R. *J. Phys. Chem.* **1983,** *87*, 5071-5083.
(50) Larentzos, J. P.; Criscenti, L. J. *J. Phys. Chem. B* **2008,** *112*, 14243-14250.
(51) Bourg, I. C.; Sposito, G. *Geochim. Cosmochim. Acta* **2007,** *71*, 5583-5589.
(52) Babu, C. S.; Lim, C. *J. Phys. Chem. A* **2006,** *110*, 691-699.
(53) Ludwig, C.; Casey, W. H.; Rock, P. A. *Nature* **1995,** *375*, 44-47.
(54) Clementi, E.; Raimondi, D. L.; Reinhardt, W. P. *J. Chem. Phys.* **1967,** *47*, 1300-1307.
(55) Matthews, R. P.; Naidoo, K. J. *J. Phys. Chem. B* **2010,** *114*, 7286-7293.
(56) Smith, R. M.; Martell, A. E., *Critical Stability Constants, Inorganic Complexes*. Plenum, New York, 1976; Vol. 1.
(57) Friddle, R. W.; Weaver, M. L.; Qiu, S. R.; Wierzbicki, A.; Casey, W. H.; De Yoreo, J. J. *Proc. Natl. Acad. Sci. USA* **2010,** *107*, 11-15.
(58) Burgess, J., *Ions in Solution: Basic Principles of Chemical Interactions*. Ellis Horwood Limited: Chichester, 1988.
(59) Tribello, G. A.; Bruneval, F.; Liew, C.; Parrinello, M. *J. Phys. Chem. B.* **2009,** *113*, 11680-11687.
(60) Radha, A. V.; Fernandez-Martinez, A.; Hu, Y.; Jun, Y.-S.; Waychunas, G. A.; Navrotsky, A. *Geochim. Cosmochim. Acta* **2012,** *submitted*.

Chapter 4 Kinetics and thermodynamics of calcite nucleation on self-assembled monolayers

Abstract: The roles of functionalized substrates in directing the kinetics and thermodynamics of $CaCO_3$ formation was studied by measuring calcite nucleation rates on self-assembled monolayers (SAMs). Nucleation rates substrates terminated by COOH, SH, or PO_4 groups were determined with optical microscopy on flow-through systems that constrained the chemistry of the nucleating solution. The experiments spanned a variety of chemical driving forces both above and below the reported solubility of amorphous calcium carbonate (ACC), although ACC formation was not observed in these systems. Using classical nucleation theory to interpret the kinetic data, we calculated calcite – substrate interfacial energies and estimated substrate-specific effects on the thermodynamic barrier to nucleation. Calcite – substrate interaction forces were quantified with independent dynamic force microscopy measurements. A second part of the study investigated the effect of magnesium on calcite nucleation rates and barriers for a carboxylic acid terminated substrate.

The kinetic measurements demonstrate a dependence of calcite nucleation rate and thermodynamics on both functional group chemistry and alkanethiol chain length. For all but one experimental substrate (C11-OH), the energetic barrier to nucleation is significantly reduced compared to homogeneous calcite nucleation. Interfacial free energies for C16 alkanethiols terminated by –COOH and –SH were reduced to 86 and 91 mJ/m^2, respectively, from approximately 110 mJ/m^2 reported for calcite in solution. The kinetic behavior is accurately described with classical models of nucleation, and calcite – substrate interaction force measured by dynamic force microscopy (DFM) is positively correlated with the reduction in interfacial free energy. Rate measurements from solutions containing progressive amounts of magnesium demonstrate reductions in calcite nucleation rate without affecting the thermodynamic barrier to nucleation and are explained by the increase in calcite solubility caused by magnesium incorporation.

4.1 Introduction

The mineralization of tissues to produce functional structures is a complex process that occurs in close association with proteins and other macromolecules. Previous studies suggest these compounds have important roles in regulating the phase, size, and morphology of the biominerals that form[1, 2]. Efforts to understand the interplay between organic and inorganic components during mineralization suggest that both primary and secondary macromolecular structure is associated with regulation of nucleation and growth, and that one function of organic species is to act as substrates that facilitate nucleation[3]. Of particular interest are calcium carbonate ($CaCO_3$) biominerals because of their widespread occurrence across the phylogenetic tree and their diversity of structure and function. The matrix proteins associated with biogenic calcite, the most common $CaCO_3$ skeletal phase, are long recognized to be highly acidic with an abundance of aspartic acid (Asp) and glutamic acid (Glu) residues. This observation has motivated a 20 year effort to understand the effects of carboxylate groups on mineralization. More recently, characterization of proteins associated with calcification have revealed the widespread occurrence of other chemistries, including the thiols of cysteine (Cys) residues and a number of phosphorylated amino acids (**Table 1**). The effects of protein conformation are challenging to assess, but templating studies indicate that spacing and orientation have strong controls over mineralization[4, 5].

Efforts to explore the roles of organic species as substrates for nucleation have made extensive use of Langmuir monolayers[6] and self-assembled monolayers (SAMs)[4] as versatile templates for promoting and controlling $CaCO_3$ formation. By preparing monolayers that present different head groups and regulating monolayer spacing and conformation through chain length, these substrates provide valuable model systems for deciphering factors that control mineralization. Observation-based studies of $CaCO_3$ formation on SAMs have provided significant insights into the role how substrate chemistry and crystal – substrate stereochemical matching may direct crystallographic orientation and morphology[7]. These studies, however usually rely on a macroscopic analysis of the resultant minerals and cannot quantify kinetic and thermodynamic factors underlying substrate – promoted mineralization processes. Furthermore, computational studies suggest that factors including degree of substrate ionization and the interplay between crystal-substrate and crystal-solution interfacial free energies are more important than stereochemical matching for calcite nucleation on organic substrates[8].

Table 4.1. Partial list of characterized proteins extracted from biogenic calcties. Most are rich in carboxylate groups due to Asp and Glu residues, while many are phoosphorylated or contain Cys thiol groups.

Organism	Protein Description	Protein Name	Solubility	Ref.
Pinctada fucata (pearl oyster)	poly(glycine) and Asp/Glu rich regions, Cys	MSI31	insoluble	(9)
	Gly rich	MSI7	insoluble	(10)
	Pro/Ile/Tyr/Arg regions, Gly/Trp rich region, Asp rich region	Prismalin-14	insoluble	(11)
Crassostrea nippona (Iwagaki oyster)	acidic, highly phosphorylated, Asp/Glu/Cys rich	MPP1	soluble	(12)
Crassostrea virginica (American oyster)	highly phosphorylated, Asp/Ser/Gly rich	none	soluble	(13)
Adamussium colbecki (American scallop)	highly phosphorylated, Asp/Ser/Gly rich	none	soluble	(14)
Patinopecten yessoensis (Yesso scallop)	Ser/Gly/Asp rich, basic terminal sequence	MSP1	insoluble	(15)
Pinna nobilis (Mediterranean fan mussel)	Ala/Asp/Thr/Pro rich, Cys rich terminal sequence	calprismin	soluble	(16)
	Asp rich	caspartin	soluble	
Procambarus clarkia (crayfish)	Glu/Ala rich	GAMP	insoluble	(17)
Gallus gallus (chicken)	cysteic acid rich	ovotransferrin	soluble	(18)
Anser anser (goose)	Ala/Glu/Ser/Gly/Trp rich	ansocalcin	soluble	(19)

In this study, we conduct a series of experiments designed to quantify the effect of functional group chemistry and conformation on the kinetics of calcite nucleation. The model substrates are prepared by established methods to produce SAMs with specific head group chemistry (-COOH, PO_4, SH or OH) and chain lengths (C11 or C16). By measuring the rate of nuclei formation onto these substrates in solutions with controlled supersaturations (chemical

driving forces), we estimate substrate – specific rates of nucleation and the interfacial energies associated with heterogeneous calcite nucleation. We find that substrate chemistry and conformation govern nucleation rates through thermodynamic factors with a systematic dependence upon crystal – substrate interaction strength.

A second part of this work quantifies the effect of magnesium ion on calcite nucleation. Owing to the importance of Mg^{2+} in calcification environments and its strong inhibitory properties[20, 21], the chemical basis for how this element regulates calcite formation has received considerable attention. Qualitative studies show Mg^{2+} influences the macroscopic crystal morphology, orientation, and composition of calcites nucleated onto functionalized substrates[10, 11]. By measuring the rate of calcite nucleation onto a single type of substrate, we separate thermodynamic and kinetic contributions to the nucleation process and find that rates become slower through changes in thermodynamics. This behavior is consistent with studies that show Mg^{2+} slows the rate of calcite growth by a thermodynamic effect arising from the higher solubility of Mg-calcite[22].

4.1.1 Background: Substrate – directed nucleation

According to classical nucleation theory, the change in free energy, ΔG, as a function of particle radius, r, is given by[23, 24],

$$\Delta G = \frac{-4Fr^3}{3\omega} k_B T\sigma + 4F\pi r^2 \alpha$$

Equation 4.1. Classical nucleation theory expression for free energy as a function of radius.

where k_B = Boltzmann's constant, T = temperature, α = interfacial energy, ω = molecular volume (6.13×10^{-23} cm^3 per molecule[22] for calcite), σ = the supersaturation of the nucleating solution, and F = a constant that depends on the shape of crystal nuclei. For pure calcite, σ is defined as,

$$\sigma = \ln\left(\frac{a_{Ca^{2+}} a_{CO_3^{2-}}}{K_{sp}}\right)$$

Equation 4.2. Definition of supersaturation, a measure of chemical driving force, with respect to calcium carbonate.

where a_i is the activity of species *i and* K_{sp} is the solubility product of calcite ($K_{sp} = 10^{-8.48}$ at 25° C). The steady-state rate of nucleation J_0, is,

$$J_0 = A\exp\left(\frac{-\Delta G^*}{k_B T}\right)$$

Equation 4.3. Expression for steady-state nucleation rate from classical nucleation theory.

where A = a kinetic factor that includes barriers to diffusion and dehydration and ΔG^* = the maximum in ΔG from **eq. 4.1**. ΔG^* indicates the thermodynamic barrier to forming a nucleus of critical radius (r^*) – the barrier to nucleation – and is defined as:

$$\Delta G^* = \frac{F\omega^2\alpha^3}{\sigma^2 k_B^2 T^2}$$

Equation 4.4. The thermodynamic barrier to forming a nucleus of critical radius (nucleation).

Combining equations **4.3** and **4.4** and transforming to a linear form gives,

$$\ln(J_0) = \ln(A) - B\left(\frac{1}{\sigma^2}\right)$$

Equation 4.5. Linear form of the steady state nucleation expression.

where

$$B = \frac{\Delta G^* \sigma^2}{k_B T} = \frac{F\omega^2\alpha^3}{k_B^3 T^3}$$

Equation 4.6. Definition of the slope of the linearized nucleation rate expression.

By measuring J_0 for a series of σ, the value of B can be determined for each substrate from **eq. 4.5**. This allows the value of α to be estimated from **eq. 4.6**, providing a quantitative assessment

of how the different organic substrates regulate mineralization. This approach has been demonstrated for the nucleation of silica on SAMs[25] as well as apatite nucleation on collagen[26].

4.2 Results and Discussion

4.2.1 Nucleation Kinetics and Thermodynamics

Measurements of crystallite density versus time demonstrate that the appearance of calcite crystallites is linear with time and depends strongly upon σ. **Figure 1** shows typical rate data for the precipitation of calcite onto C16-COOH substrates. Data for all surfaces are provided in **Appendix 3 (Figure A3.1)**. Rates increase in solutions with progressively higher supersaturations and the rates are specific to each type of substrate. Estimates of the steady state

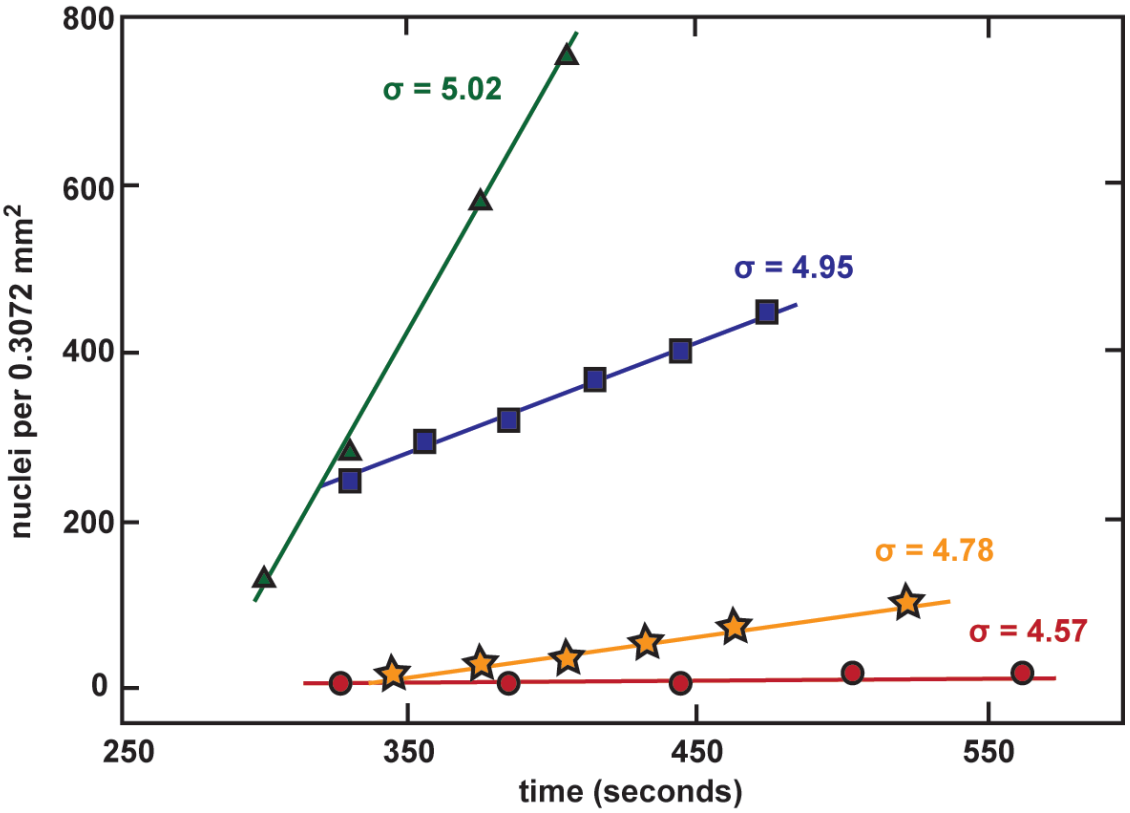


Figure 4.1. The number of calcite nuclei on the surface of C16-COOH (MHA) scales linearly with time during the early stages of each experiment. The slopes of the plotted lines for each supersaturation quantify steady state nucleation rates, J_n. Rates of nucleation increase with the supersaturation, σ, of the mineralizing solution.

nucleation rates were obtained for each experiment from the linear portion of each nuclei versus time plot (see methods) to obtain J_0. As predicted by **eq. 5**, ln (J_0) is linearly dependent on 1/ σ^2 for each substrate (**Figure 2**), indicating that the nucleation behavior is consistent with the predictions of classical nucleation theory. All of the substrates show a strong dependence upon supersaturation except the –OH terminated surfaces (C11-OH). The very slow rates of nucleation measured on these surfaces are consistent with observations reported by previous studies.

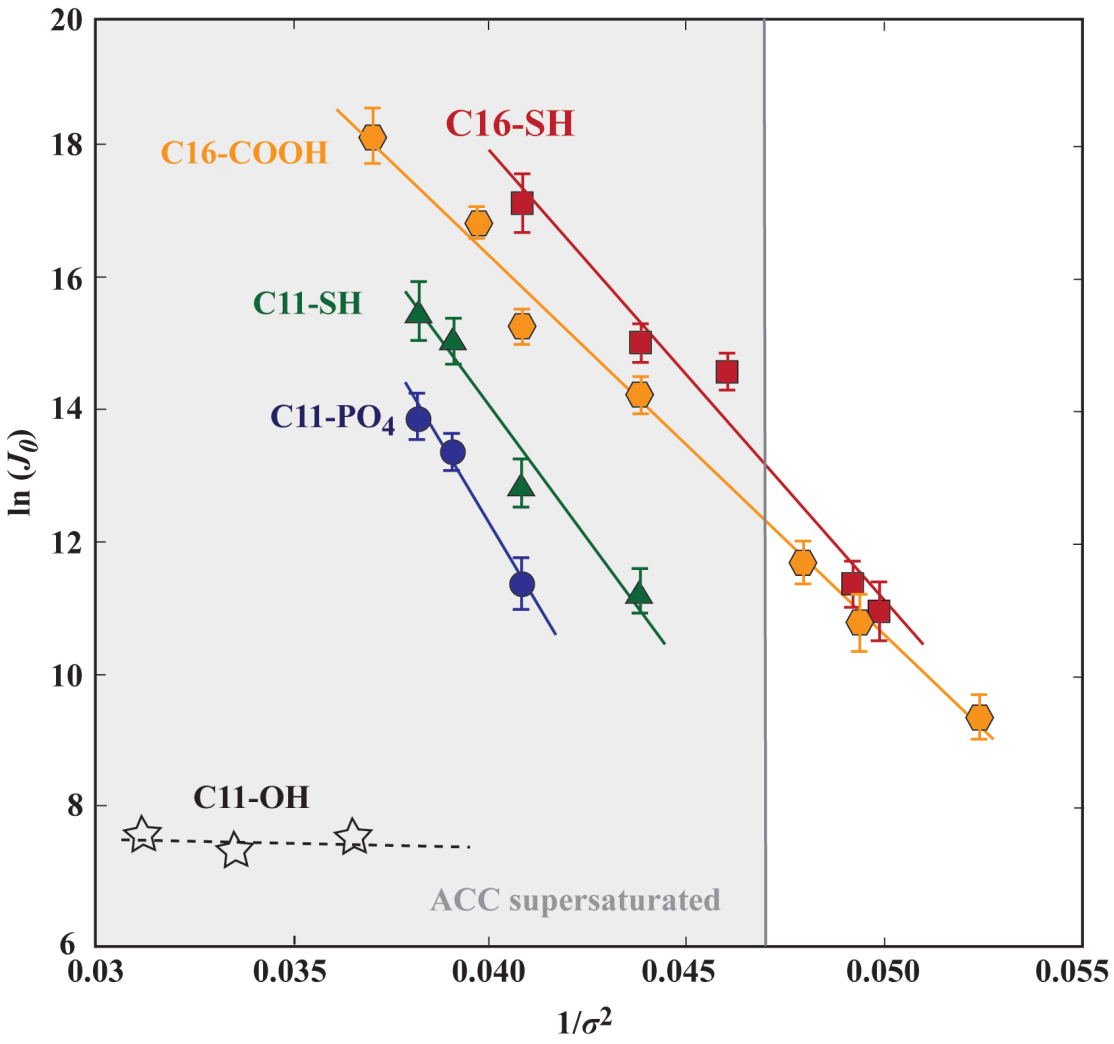


Figure 4.2. Plots of ln (J_0) versus $1/\sigma^2$ for all substrates demonstrate the linear relationship predicted in eq. 4. The slopes of the lines (B) inform substrate specific barriers to nucleation and calcite-substrate interfacial free energies. The shaded region indicates where the nucleating solution is supersaturated with respect to amorphous calcium carbonate (ACC). Error bars are two standard deviations of measurements made at each supersaturation.

Analysis of **Figure 4.2** demonstrates the importance of both chemistry and conformation in regulating nucleation rates and provides quantitative insights into the substrate-specific thermodynamic barrier to nucleation. Using **eq. 4.5** and the slopes of the relationships shown in **Figure 4.2**, we estimate values of B for each of the substrates (**Table 4.2**). Hydroxyl-terminated surfaces (C11-OH) were not included in this analysis because the number of crystallites observed were so few that the resultant nucleation rates are smaller than the experimental errors. Because B is proportional to ΔG^* (**eq. 4.6**), the slopes of the lines plotted in **Figure 4.2** give an indication of how the different substrates influence the thermodynamic barrier to nucleation. This effect is demonstrated by plots of **eq. 4.1** for a single supersaturation that show the dependence of ΔG on radius (**Figure 4.3).** The maximum of each curve indicates ΔG^*, which is reduced for all substrates compared to that predicted for homogeneous calcite nucleation. Both substrate functional group chemistry and alkanethiol chain length have significant effects on the barrier to calcite nucleation. For example, a comparison of substrates with the same chain length but different head groups shows that B and ΔG^* are lower for C16-COOH than C16-SH and C11-SH are lower than C11-PO_4. The importance of chain length is seen by the lower B and ΔG^* values for C16-SH compared to C11-SH.

Table 4.2. Estimates of B, ln(A), and α from measured rates of heterogeneous calcite nucleation. Values of α assume spherical nuclei. The interfacial tension of calcite in solution is approximately[27] 110 mJ/m^2.

substrate	B	ln (A)	α (mJ/m^2)
C16-COOH	-572	39	86
C16-SH	-680	45	91
C11-SH	-806	46	96
C11-PO_4	-983	52	98

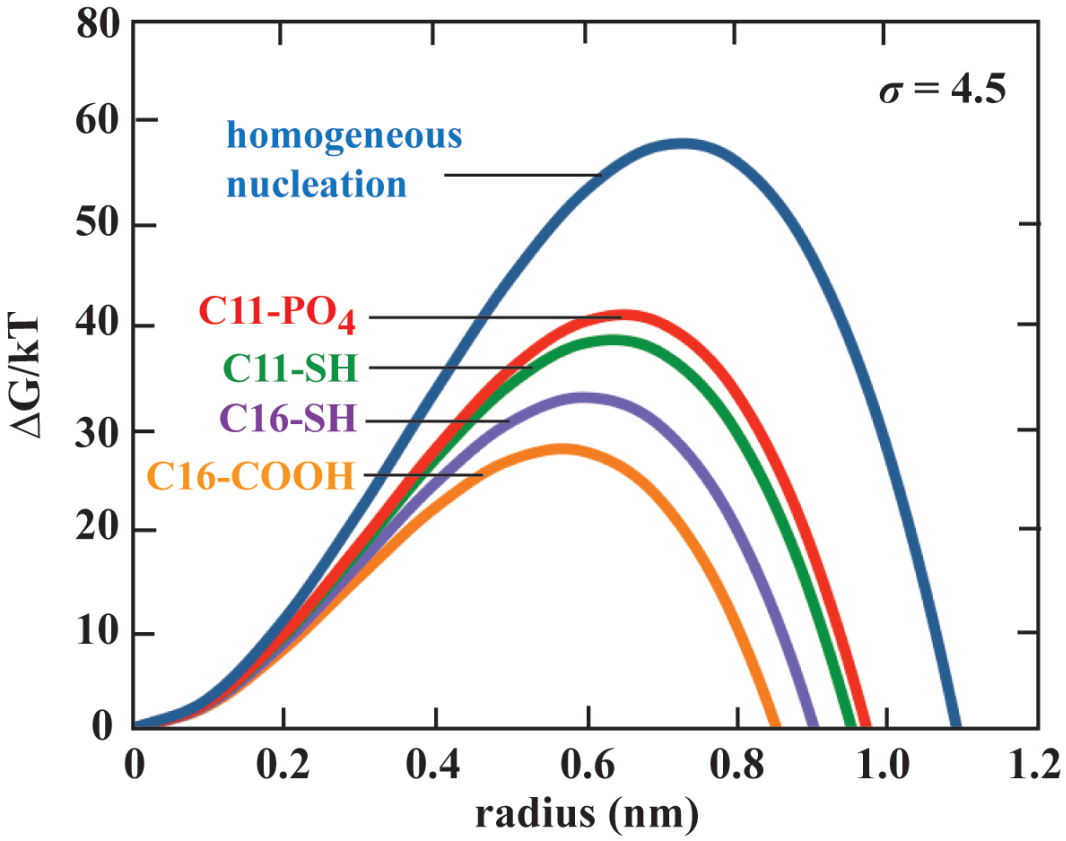


Figure 4.3. The dependence of $\Delta G/kT$ on radius at $\sigma = 4.5$ for heterogeneous calcite nucleation on templated substrates and homogeneous nucleation from solution. The values of α used to predict ΔG (**eq. 1**) are estimated for heterogeneous nucleation assuming spherical nuclei. For heterogeneous nucleation, we assume $\alpha = 110$ mJ/m^2. All experimentally tested substrates reduce the barrier to nucleation (the maximum of each curve) and the corresponding critical radius. Critical radius size is predicted to be on the order of 6 Å for heterogeneous nucleation, slightly smaller than the 7.3 Å radius expected in the homogeneous case.

4.2.2 Interfacial Free Energies

Using **eq. 4.5**, substrate specific interfacial free energies, α, can be estimated from B (**Table 4.2**) by assuming a value for the nuclei shape factor, F. For these calculations, we assume spherical nuclei for all substrates ($F = 16.8$) instead of supposing that nuclei shape resembles the resultant macroscopic calcite crystals. This allows a straightforward estimate of α without making assumptions for the exact ratios of crystal orientations on each substrate (multiple

orientations are seen on the C11 substrates) and generates little error compared to the uncertainty in B. For example, F = 19.7 for calcite nucleation on the (012) plane, the prevalent orientation on C16-COOH, resulting in an α that differs from a spherical nuclei by only 5.3%. Using this approach, we find values of α from 86-98 mJ/m^2. If we assume the interfacial energy for homogeneous nucleation of calcite from solution to be around 110 mJ/m^2 (the average of the range reported by Söhnel and Mullin[27]), all substrates in this study reduce α (**Table 4.2**).

To examine the origins of the substrate-specific differences in interfacial energy, we must evaluate the individual components that make up the α. The α in **eqs. 4.2** and **4.5** incorporate three distinct interfacial energies, crystal-liquid (α_{CL}), crystal-substrate (α_{CS}), and substrate-liquid (α_{SL}) that are related by,

$$\alpha = \alpha_{CL} - h\left(\alpha_{SL} - \alpha_{CS}\right)$$

Equation 4.7. The interfacial energy term calculated from nucleation rate data incorporates three distinct interfacial energy terms.

where h = is a constant that depends on nucleus shape. The α values that we calculate (**Table 4.2**) reflect substrate effects on all three components in **eq. 4.7**, although individual contributions cannot be separated here. Nucleation substrate is expected to effect α_{CS} and α_{SL}, but could also influence α_{CL} by affecting the area of calcite faces available to solution. We assume that α_{CL} will not change significantly between substrates as the same /104/ surfaces are exposed to solution in each case. Furthermore, although the surface pK_a of monolayers are known to be higher than those of the same functionalities in solution[28, 29], all experimental monolayer surfaces should be charged and hydrophilic due to the high pH (10.55) of experimental solutions. Therefore, we expect α_{LS} to be relatively low for all substrates, leaving α_{CS} as the most significant contributor to differences in α between substrates.

4.2.3 Calcite – substrate interaction energies

To test the idea that differences in α_{CS} between substrates primarily drive the substrate-specific α values estimated from nucleation rates, we measured rupture forces between functionalized AFM tips and the (104) face of calcite with dynamic force microscopy (DFM).

These rupture forces quantify the level of interaction between the nucleating phase and the substrate, thereby informing the crystal – substrate interfacial energy, α_{CS}. These forces are anticipated to be strong for heterogeneous nucleation, where the thermodynamic barrier compared to the homogeneous case is significantly reduced[30]. Substrates with a strong interaction with the nucleating crystal should have a relatively small α_{CS}, resulting in a smaller overall α by **eq. 4.7**.

Representative force curves at a single tip retraction velocity for C16-COOH, C16-SH, and gold (**Figure 4.4a**) show that greater force is required to rupture tip – calcite bonds for functionalized tips compared to non-functionalized (gold coated) tips. **Figure 4.4b** demonstrates that the functionalization of tips strongly influences the probe – calcite rupture force and that all experimental alkanethiols increase the rupture force compared to the bare gold tip. The largest forces are associated with the C16-COOH surfaces, while the smallest forces measured with a functionalized tip correspond to the C11-PO_4 surface. The bond-rupture forces measured by DFM depend on experimental loading rates, which are a function of the spring constant of the

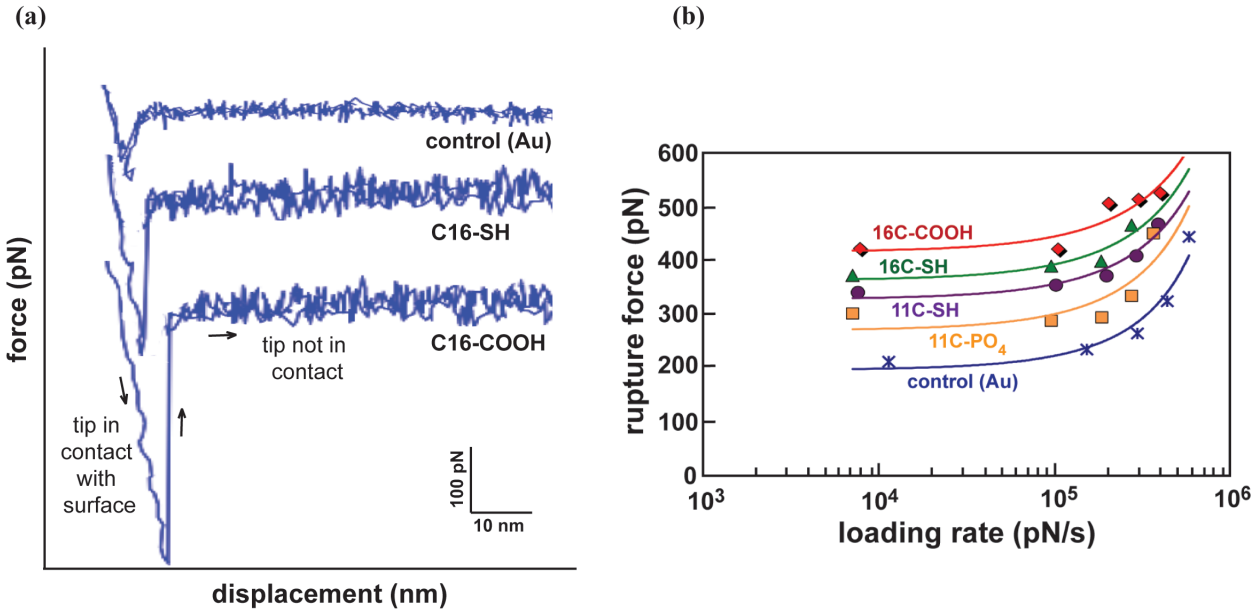


Figure 4.4. (a) Representative force curves for the control (Au-coated AFM tip) and two substrates (C16-COOH and C16-SH) show the differences in interaction strength with the calcite surface at the same tip retraction velocity ($1x10^{-5}$ m/s). (b) Calcite – SAM rupture forces measured by dynamic force microscopy decrease with the experimental loading rate, which is the product of cantilever spring constant and tip retraction velocity. All experimental substrates interact with the calcite (104) surface more strongly than the bare (gold-coated) tip. Exponential fits are drawn in as guides.

AFM cantilever and the velocity at which the cantilever tip is retracted from the crystal surface. Lower loading rates correspond to slower tip retraction, which allows more time for thermally driven solvent collisions to assist in the bond rupture process and results in lower apparent rupture forces[31], as demonstrated in **Figure 4.4b**.

The DFM rupture forces correlate well with the experimentally derived estimates of α (**Figure 5**). Larger substrate – calcite rupture forces are associated with lower α for the experimental substrates. This relationship supports the idea that substrates reduce α through a reduction in the α_{CS} component of **eq. 7**. This is likely explained by binding between the negatively charged SAM terminal groups and calcite surface calcium atoms. Such an interpretation is supported by DFM measurements carried out at lower pH (7.0) showing that rupture forces are reduced most drastically for the alkanethiols terminated by functionalities with the high pK_a. This evidence supports the idea that highly ionized substrates will exhibit strong adhesion to the calcite surface and strongly promote calcite nucleation[8].

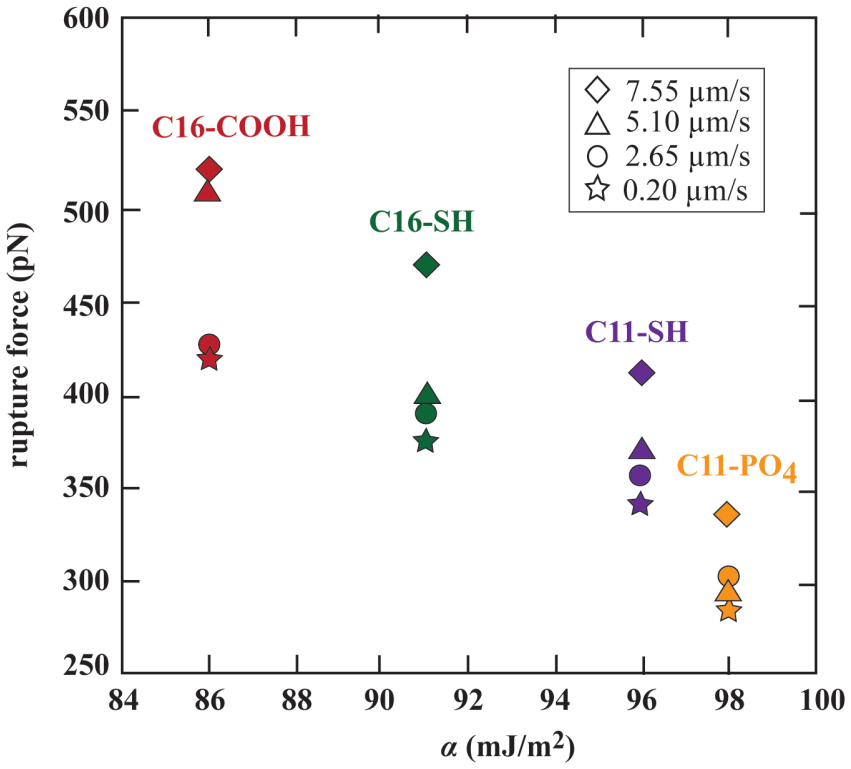


Figure 4.5. Experimenatlly derived estimates of interfacial energies (α) correlate with the rupture forces measured by dynamic force microscopy. Lower interfacial energies correspond to strong calcite – SAM interactions (high rupture force). For each substrate, forces are plotted for four tip retraction velocities (0.20, 2.65, 5.10, and 7.55 µm/s).

It is unlikely that the measured rupture forces describe single-molecule bonds to the calcite surface, but reflect the total interaction between the surface alkanethiols attached to the tip and the calcite surface (see schematic in **Appendix 3**, **Figure A3.2**). Further evidence for this is seen if we convert the measured forces to quantities of energy per mol by the method of Friddle et al[32]. The energies predicted are much larger than expected for a single bond-rupture event, suggesting that the rupture forces we measure are for multiple bonds (**Figure A3.3)**.

4.2.4 Magnesium Effects

Magnesium is recognized as an inhibitor of calcite nucleation and growth from solution[20-22]. Plots of **eq. 4.5** (**Figure 4.6**) demonstrate this effect for calcite nucleation on 16C-COOH from solutions containing up to 2:1 Mg^{2+}:Ca^{2+}. For each experimental supersaturation, the nucleation rate decreases with increasing solution Mg^{2+}. The slopes and intercepts of the $\ln(J_0)$ vs. $1/\sigma^2$ plots, however, are similar for all Mg^{2+} contents, indicating that α and ΔG^* for calcite nucleation on MHA are unchanged by Mg^{2+}. This effect is most likely explained by the fact that Mg^{2+} incorporation into calcite increases its solubility[33], resulting in a decreased effective supersaturation (**Figure A3.4**). This thermodynamic effect has been demonstrated by AFM measurements of calcite growth rate in the presence of magnesium[22]

Although Mg^{2+} is known to encourage aragonite formation[34], we did not observe aragonite nucleation in any of our systems. Solutions with Mg^{2+}/Ca^{2+} of up to 2 nucleate calcite or Mg-calcite with morphologies consistent with reported morphologies of calcite nucleation from solution[35] and directed by carboxylate-terminated SAMs[36] (**Figure A3.5A-D**). For solutions with $Mg^{2+}/Ca^{2+} > 2$ and supersaturations both above and below ACC solubility, we observed the deposition of a vaterite film (**Figure A3.5E-F**). Nucleation rates could not be determined for this phase and further investigation is needed to understand why aragonite is not formed in these conditions. It is possible that aragonite nucleation is discouraged by the particular substrates used in this study or the residence time of the nucleating solution in contact with the substrate is insufficient for aragonite formation.

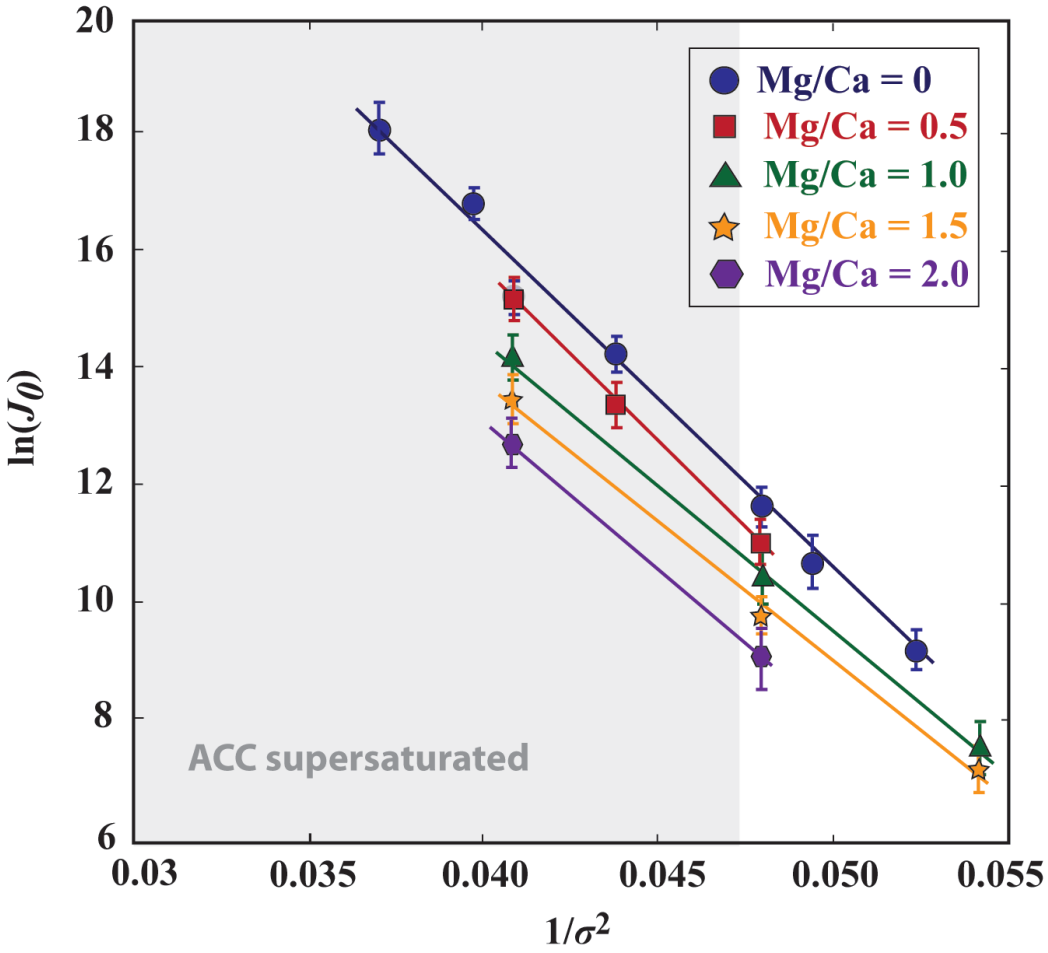


Figure 4.6. Plots of ln (J_0) versus $1/\sigma^2$ for calcite nucleation on C16-COOH from solutions of varying magnesium content show that magnesium reduces nucleation rate while slope, and therefore the thermodynamic barrier to nucleation, remain unchanged.

4.2.5 Absence of ACC

It is notable that even though nucleation rates were measured for solutions supersaturated with respect to amorphous calcium carbonate (ACC, **eq 4.3,** $K_{sp,\,ACC} = 10^{-6.39}$ [37]), SEM micrographs taken of substrates at experimental times less than the induction time for calcite nucleation revealed no evidence of surface-bound ACC. This high-solubility phase is thought to precede the final crystalline product in a wide variety of calcifying organisms[38-40]. In these experiments, the slopes of ln (J_0) vs. $1/\sigma^2$ plots and therefore the thermodynamic barrier to nucleation appears unaffected by the crossover into the ACC solubility field (**Figures 4.2 and 4.6**). This suggests that even if ACC is being formed in solution or undergoing rapid

transformation to calcite on the surface, the process has little effect on the kinetic and thermodynamic parameters that govern the substrate-directed calcite mineralization. These results are consistent with new evidence suggesting that a quiescent environment that has been rapidly supersaturated with carbonate – unlike the environment of nucleation in these experiments – is required to stabilize ACC clusters[41].

4.3 Conclusions

The values of B and α estimated from steady state nucleation rates (**Table 4.2**) demonstrate that four of the experimental substrates (C16-COOH, C16-SH, C11-SH and C11-PO_4) facilitate calcite nucleation by reducing the energetic barrier to nuclei formation compared to the homogeneous case. The findings show that calcite nucleation rates are substrate specific, and that groups other than COOH are effective promoters of calcite mineralization. This suggests that the less-commonly studied components of biomolecules, including thiol and phosphate moieties, may have a more dominant role in $CaCO_3$ formation than previously thought. The substrate-specific rates highlight the likely importance of primary structure and cooperative effects within proteins, which has been previously demonstrated for silica deposition[25]. The differences in the kinetics and thermodynamics of nucleation on substrates with the same head group but different alkanethiol chain length (C16-SH versus C11-SH) imply that SAM structure and packing, which are strongly influenced by chain length[42, 43] are just as influential as chemical functionality in regulating calcite nucleation. This suggests that the secondary structure of biomolecules may be an important consideration in biogenic $CaCO_3$ formation.

The rate data demonstrate that for the systems tested here, substrate directed calcite nucleation obeys classical theory at supersaturations both above and below the ACC solubility level. This observation is consistent with the current thinking that ACC formation is diffusion limited[44], because our experiments were designed to prevent diffusion-limited conditions. Measurements of calcite nucleation from magnesium containing solutions demonstrate that magnesium reduces nucleation rate through a thermodynamic effect. This interpretation is in agreement with insights from calcite growth experiments[22].

4.4 Experimental

4.4.1 Preparation of Functionalized Substrates

Substrates for $CaCO_3$ nucleation were prepared from evaporated gold on mica (Agilent Technologies) using the template stripped gold method[45]. Self-assembled monolayers (SAMs) were deposited on the gold by submerging the substrates in ethanol-based (200 proof anhydrous, Sigma-Aldrich), 1.5 mM solutions of the following five alkanethiols (Sigma-Aldrich®): 16-Mercaptohexadecanoic acid (C16-COOH), 1,16-Hexadecanedithiol (C16-SH), 1,11-Undecanedithiol (C11-SH), 11-Mercaptoundecylphosphoric acid (C11-PO_4), or 11-Mercapto-1-undecanol (C11-OH), for 12-15 hours.

4.4.2 Nucleation Rate Measurements

Solutions were prepared using calcium chloride dihydrate (99+%, Alfa Aesar®), magnesium chloride hexahydrate (99% Sigma-Aldrich®), and sodium bicarbonate (99.997%, Alfa Aesar®) and transferred to two polypropylene 140cc syringes for the $CaCl_2$/$MgCl_2$ and $NaHCO_3$ solutions. After rinsing with ethanol and water, the substrates were placed in a sealed flow chamber connected to the syringes with Tygon tubing (1/16" inner diameter, Cole Parmer). A high precision syringe pump continuously flowed the solutions into the substrate chamber at 30 mL/hour per syringe to maintain steady-state conditions. This flow rate prevented diffusion – limited conditions while also maintaining the reactant solutions in the chamber for sufficiently long residence time to observe nucleation at a wide variety of supersaturations. Solutions were mixed in the tubing 10 cm upstream of the substrate chamber to ensure that nucleation did not occur before the solution came into contact with the substrate (**Figure A3.6**). The experiments were conducted at 25° C and supersaturations of 4.66 – 5.12 (species activities calculated with Geochemist's Workbench®, Urbana, IL) at pH values were 10.55 ± 0.10.

$CaCO_3$ nucleation was monitored by optical microscopy (Olympus BX51, 10x or 20x objective). Because the resolution of the optical microscope was insufficient to directly observe the formation of critical nuclei, we estimated nucleation rates by making the assumption that each optically observed crystal developed from a single nucleus. That is, each nucleus that formed and successfully crossed the energy barrier to reach critical size grew into an optically observable crystallite. This assumption is plausible because the data were collected only within

the time interval where crystallite spacing was very large (separation between observed crystals was much greater than the size at which they became optically visible, **Figure A3.7**). Moreover, the data show their appearance was linearly dependent on time (**Figure 4.1 and Figure A3.1**). We also assumed that all crystallites nucleated heterogeneously on the surface. Parallel experiments that used inverted substrates and an Olympus IX2 inverted microscope validated this assumption with measurements showing similar nucleation rates and observations that the crystallites exhibited similar substrate-specific orientations.

The steady state nucleation rates were calculated directly as the slopes of crystallites versus time plots (in unites of nuclei per m^2 per s) within the time interval where the number of nuclei increased linearly with time, typically within the first few hours of each experiment. Up to three replicate experiments were performed for each substrate/supersaturation condition and the average J_0 from the replicates was used to calculate B and α for each substrate.

4.4.3 Dynamic Force Microscopy (DFM)

Force measurements were made with the MFP-3D AFM (Asylum Research®) using Si_3N_4 cantilevers with triangular tips (Veeco®). Cantilevers were cleaned under UV/ozone for 10 minutes, coated by thermal evaporation with 4 nm Cr followed by 40 nm Au, and re-cleaned under UV/ozone before placement in alkanethiol solutions for SAM deposition. Force measurements between the modified tips and a freshly cleaved calcite (104) surface (Iceland Spar, Wards®) were conducted in a saturated ($\sigma = 0$) $CaCO_3$ solution at pH 10.55. The spring constants of individual cantilevers were calculated using the thermal calibration method[46] before beginning force measurements. Five tip retraction velocities ($2x10^{-7}$ - $1x10^{-5}$ m/s) were sampled while using a constant approach velocity of 200 nm/s and a 1 s dwell time at the surface. Heterogeneity in the calcite surface was factored into the measurements by employing a routine to randomly move the tip to different positions on the surface using a 20 nm steps. For each surface, force curves were measured for each of the five velocities at 100 locations on the surface, resulting in 500 data points per sample.

4.4.4 Scanning Electron Microscopy (SEM)

SEM images were collected on a FEI Quanta 600 FEG Environmental SEM operated at high vacuum and 10-20 kV accelerating voltage. All samples were coated with ~10 nm of Au/Pd before imaging.

4.4.5 Raman Spectroscopy

The $CaCO_3$ polymorph of the crystallites that formed were determined by Raman spectroscopy. Spectra were recorded with a JY Horiba LabRam HR spectrometer using a 632.81 nm He-Ne laser focused through a 100x objective.

4.5 References

(1) Falini, G.; Albeck, S.; Weiner, S.; Addadi, L. *Science* **1996,** *271*, 67-69.
(2) Albeck, S.; Addadi, L.; Weiner, S. *Connect. Tissue Res.* **1996,** *35*, 365-370.
(3) Mann, S. *Nature* **1988,** *332*, 119-124.
(4) Aizenberg, J.; Black, A. J.; Whitesides, G. M. *Nature* **1999,** *398*, 495-498.
(5) Travaille, A. M. *Soft Interactions, Solid Results: Nucleation of Calcite on Organic Monolayers.* Radboud University Nijmegen, Ipskamp, Enschede, Netherlands, 2005.
(6) Rajam, S.; Heywood, B. R.; Walker, J. B. A.; Mann, S.; Davey, R. J.; Birchall, J. D. *J. Am. Chem. Soc. Faraday Trans.* **1991,** *87*, 727-734.
(7) Pokroy, B.; Aizenberg, J. *Crystengcomm.* **2007,** *9*, 1219-1225.
(8) Duffy, D. M.; Harding, J. H. *Langmuir* **2004,** *20*, 7630-7636.
(9) Sudo, S.; Fujikawa, T.; Nagakura, T.; Ohkubo, T.; Sakaguchi, K.; Tanaka, M.; Nakashima, K.; Takahashi, T. *Nature* **1997,** *387*, 563-564.
(10) Zhang, Y.; Xie, L. P.; Meng, Q. X.; Jiang, T. M.; Pu, R. L.; Chen, L.; Zhang, R. Q. *Comp. Biochem. Phys. B* **2003,** *135*, 565-573.
(11) Suzuki, M.; Murayama, E.; Inoue, H.; Ozaki, N.; Tohse, H.; Kogure, T.; Nagasawa, H. *Biochem. J.***2004,** *382*, 205-213.
(12) Samata, T.; Ikeda, D.; Kajikawa, A.; Sato, H.; Nogawa, C.; Yamada, D.; Yamazaki, R.; Akiyama, T. *Febs. J.* **2008,** *275*, 2977-2989.
(13) Borbas, J. E.; Wheeler, A. P.; Sikes, C. S. *J. Exp. Zool.* **1991,** *258*, 1-13.
(14) Halloran, B. A.; Donachy, J. E. *Comp. Biochem. Phys. B* **1995,** *111*, 221-231.
(15) Sarashina, I.; Endo, K. *Mar. Biotechnol.* **2001,** *3*, 362-369.
(16) Marin, F.; Amons, R.; Guichard, N.; Stigter, M.; Hecker, A.; Luquet, G.; Layrolle, P.; Alcaraz, G.; Riondet, C.; Westbroek, P. *J. Biol. Chem.* **2005,** *280*, 33895-33908.
(17) Ishii, K.; Tsutsui, N.; Watanabe, T.; Yanagisawa, T.; Nagasawa, H. *Biosci. Biotech. Bioch.* **1998,** *62*, 291-296.
(18) Gautron, J.; Hincke, M. T.; Panheleux, M.; Garcia-Ruiz, J. M.; Boldicke, T.; Nys, Y. *Connect. Tissue. Res.* **2001,** *42*, 255-267.

(19) Lakshminarayanan, R.; Valiyaveettil, S.; Rao, V. S.; Kini, R. M. *J. Biol. Chem.* **2003,** *278*, 2928-2936.
(20) Bischoff, J. L. *J. Geophys. Res.* **1968,** *73*, 3315-3322.
(21) Berner, R. A. *Geochim. Cosmochim. Acta* **1975,** *39*, 489-504.
(22) Davis, K. J.; Dove, P. M.; De Yoreo, J. J. *Science* **2000,** *290*, 1134-1137.
(23) Turnbull, D.; Fisher, J. C. *J. Chem. Phys.* **1949,** *17*, 71-73.
(24) Markov, I. V., *Crystal Growth for Beginners: Fundamentals of Nucleation, Crystal Growth, and Epitaxy*. World Scientific Publishing: Singapore, 2003.
(25) Wallace, A. F.; DeYoreo, J. J.; Dove, P. M. *J. Am. Chem. Soc.* **2009,** *131*, 5244-5250.
(26) Habraken, W. J. E. M.; Tao, J.; Brylka, L. J.; Friedrich, H.; Bertinetti, L.; Schenk, A. S.; Verch, A.; Dmitrovic, V.; Bomans, P. H. H.; Frederik, P. M.; Laven, J.; Schoot, P. v. d.; Aichmayer, B.; With, G. d.; DeYoreo, J. J.; Sommerdijk, N. A. J. M. **2012**, in review.
(27) Söhnel, O.; Mullin, J. W. *J. Cryst. Growth* **1978,** *44*, 377-382.
(28) Wang, J.; Frostman, L. M.; Ward, M. D. *J. Phys. Chem.* **1992,** *96*, 5224-5228.
(29) Kakiuchi, T.; Iida, M.; Imabayashi, S.; Niki, K. *Langmuir* **2000,** *16*, 5397-5401.
(30) Liu, X. Y.; Lim, S. W. *J. Am. Chem. Soc.* **2003,** *125*, 888-895.
(31) Lo, Y. S.; Zhu, Y. J.; Beebe, T. P. *Langmuir* **2001,** *17*, 3741-3748.
(32) Friddle, R. W.; Battle, K.; Trubetskoy, V.; Tao, J. H.; Salter, E. A.; Moradian-Oldak, J.; De Yoreo, J. J.; Wierzbicki, A. *Angew. Chem. Int. Ed.* **2011,** *50*, 7541-7545.
(33) Bischoff, W. D.; Mackenzie, F. T.; Bishop, F. C. *Geochim. Cosmochim. Acta* **1987,** *51*, 1413-1423.
(34) Folk, R. L. *J. Sed. Res.* **1974,** *44*, 40-53.
(35) Zhang, Y. P.; Dawe, R. A. *Chem. Geol.* **2000,** *163*, 129-138.
(36) Han, Y. J.; Aizenberg, J. *J. Am. Chem. Soc.* **2003,** *125*, 4032-4033.
(37) Brecevic, L.; Nielsen, A. E. *J. Cryst. Growth* **1989,** *98*, 504-510.
(38) Politi, Y.; Metzler, R. A.; Abrecht, M.; Gilbert, B.; Wilt, F. H.; Sagi, I.; Addadi, L.; Weiner, S.; Gilbert, P. *Proc. Natl. Acad. Sci. USA* **2008,** *105*, 17362-17366.
(39) Tao, J. H.; Zhou, D. M.; Zhang, Z. S.; Xu, X. R.; Tang, R. K. *Proc. Natl. Acad. Sci. USA* **2009,** *106*, 22096-22101.
(40) Mahamid, J.; Sharir, A.; Addadi, L.; Weiner, S. *Proc. Natl. Acad. Sci. USA* **2008,** *105*, 12748-12753.
(41) Wang, D.; Echigo, T.; Giuffre, A. F.; Hamm, L. M.; Rimstidt, D. J.; DeYoreo, J. J.; Grotzinger, J.; Dove, P. M. *Farad. Discuss.* **2012,** *159*, in review.
(42) Fenter, P.; Eberhardt, A.; Liang, K. S.; Eisenberger, P. *J. Chem. Phys.* **1997,** *106*, 1600-1608.
(43) Fenter, P.; Eisenberger, P.; Liang, K. S. *Phys. Rev. Lett.* **1993,** *70*, 2447-2450.
(44) Raiteri, P.; Gale, J. D. *J. Am. Chem. Soc.* **2010,** *132*, 17623-17634.
(45) Hegner, M.; Wagner, P.; Semenza, G. *Surf. Sci.* **1993,** *291*, 39-46.
(46) Burnham, N. A.; Chen, X.; Hodges, C. S.; Matei, G. A.; Thoreson, E. J.; Roberts, C. J.; Davies, M. C.; Tendler, S. J. B. *Nanotechnology* **2003,** *14*, 1-6.

Appendix A. Supplementary Information to Chapter 2

A1. Lennard-Jones Interactions

While dispersive and repulsive interactions between atoms can be calculated using quantum mechanics, these calculations require large basis sets to describe electronic structure and are impractical for the simulation of large systems and long time scales. Thus, an empirical function that is rapidly evaluated is needed to accurately model interatomic energy functions. The most frequently used of such functions is the Lennard-Jones (LJ) 6-12 interaction, which describes the van der Waals contribution to potential energy, U_{LJ}(r), as,

$$U_{LJ}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

Equation A1. Form of the 6-12 Lennard-Jones potential describing van der Waals interactions

where ε = the potential well depth and σ = the interatomic distance of zero potential. The term to the 12th power describes the repulsive component while the term to the 6th power represents the attractive component. Some MD software packages simplify the 6-12 LJ potential into the following form,

$$U_{LJ}(r) = \frac{A}{r^{12}} - \frac{B}{r^{6}}$$

Equation A2. Simplified *A*-*B* representation of the 6-12 LJ potential.

where $A = 4\varepsilon\sigma^{12}$ and $B = 4\ \varepsilon\sigma^{6}$.

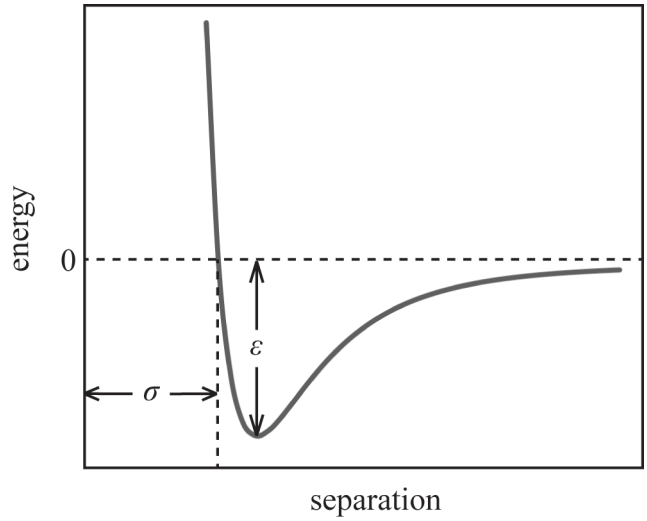


Figure A1. Physical representation of the 6-12 LJ potential

A2. Lorentz-Berthelot Mixing Rules

To describe the non-electrostatic interactions in polyatomic systems, the van der Waals interaction energies between atoms of different types must be quantified. A system containing N atoms requires the determination of $N(N\text{-}1)/2$ parameter sets for cross interactions between unlike atoms. In MD simulations, these parameters are often determined directly from the LJ parameters of like sets of atoms using mixing rules. The Lorentz-Berthelot mixing rules model the LJ well depth between unlike atoms i and j as a geometric mean,

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}}$$

Equation A3. Lorentz-Berthelot summation of well depth for unlike atoms.

while the of distance of zero potential is expressed as an arithmetic mean:

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{ij}}{2}$$

Equation A4. Lorentz-Berthelot summation of zero potential distance for unlike atoms.

A3. Ewald Summation Method

The description of long-range forces in MD simulation can be problematic because these forces often decay over a length scale that is large relative to the size of the simulation box. These potentials decay slower than r^{-n}, where n is the dimensionality of the system, and include electrostatics, dipoles, and gravity. The reaction field method, the cell multiple method, particle-mesh methods, and the Ewald summation method[1] have all been developed to deal with long-range forces. The Ewald summation (used in the work presented in this dissertation) method is a type of Poisson summation where the simulation system is assumed to be infinitely periodic and long-range electrostatic interactions are summed in Fourier space. The real-space simulation system is considered to be a central unit cell surrounded by image cells (**Figure A1.2**). The long-range interaction energy is calculated as a sum between the central cell and the surrounding image cells. All charges are considered to be surrounded by a neutralizing charge distributions with Gaussian forms, which are included in the overall summation.

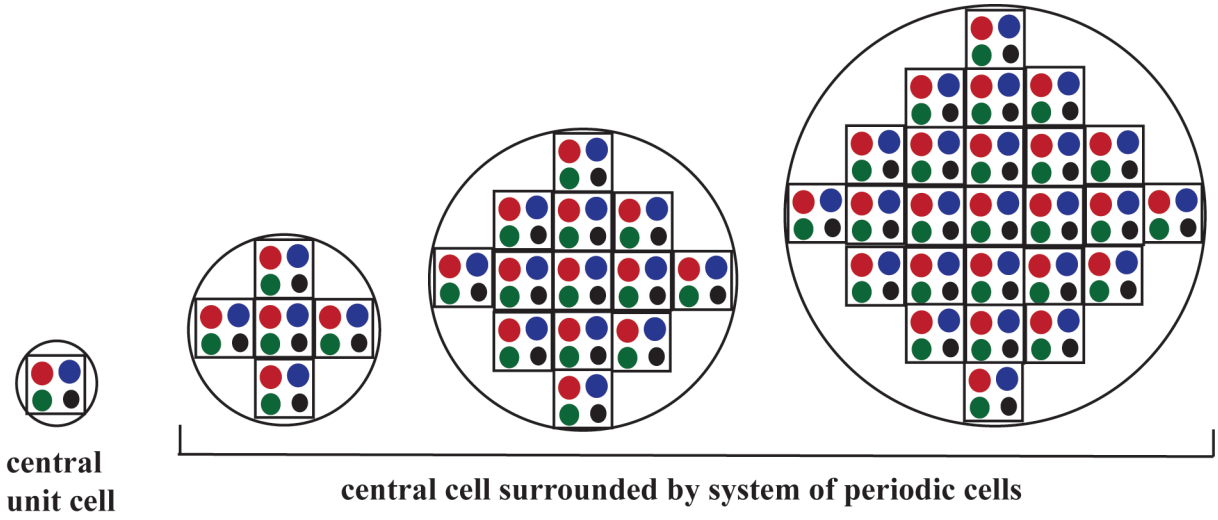


Figure A2. Construction of an infinitely periodic system for Ewald summation of long-range electrostatic forces

A4. Results of Replicate Calculations Using Alternate Lennard-Jones Parameters and Water Models

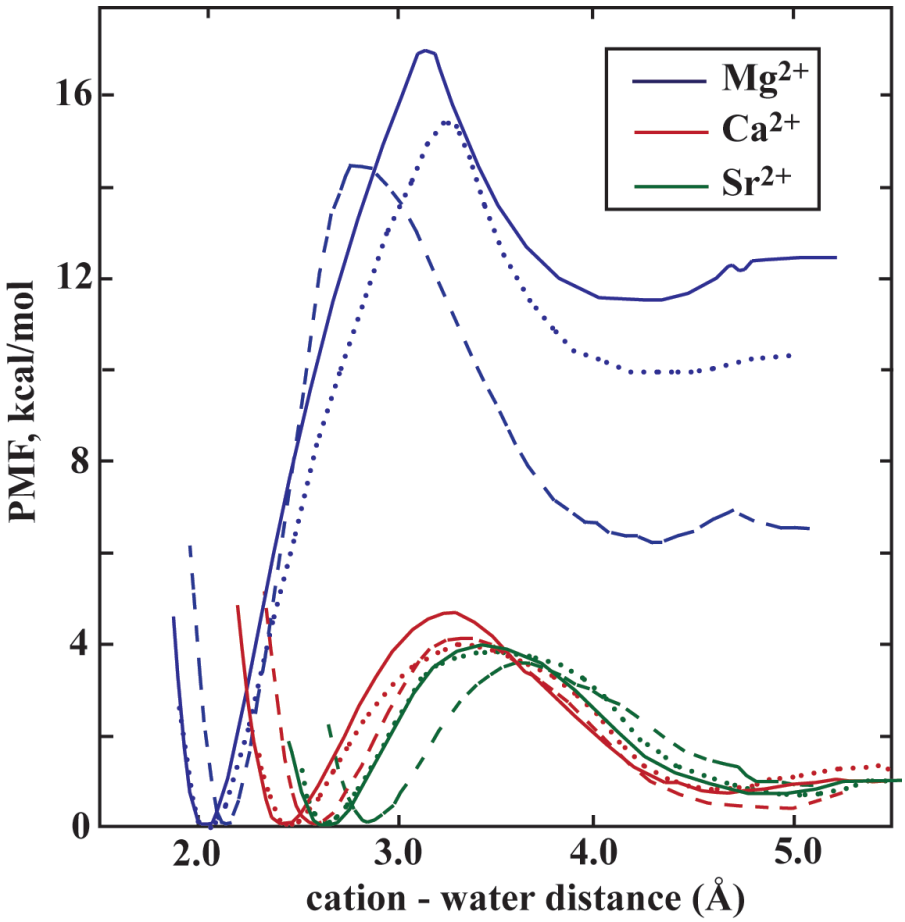
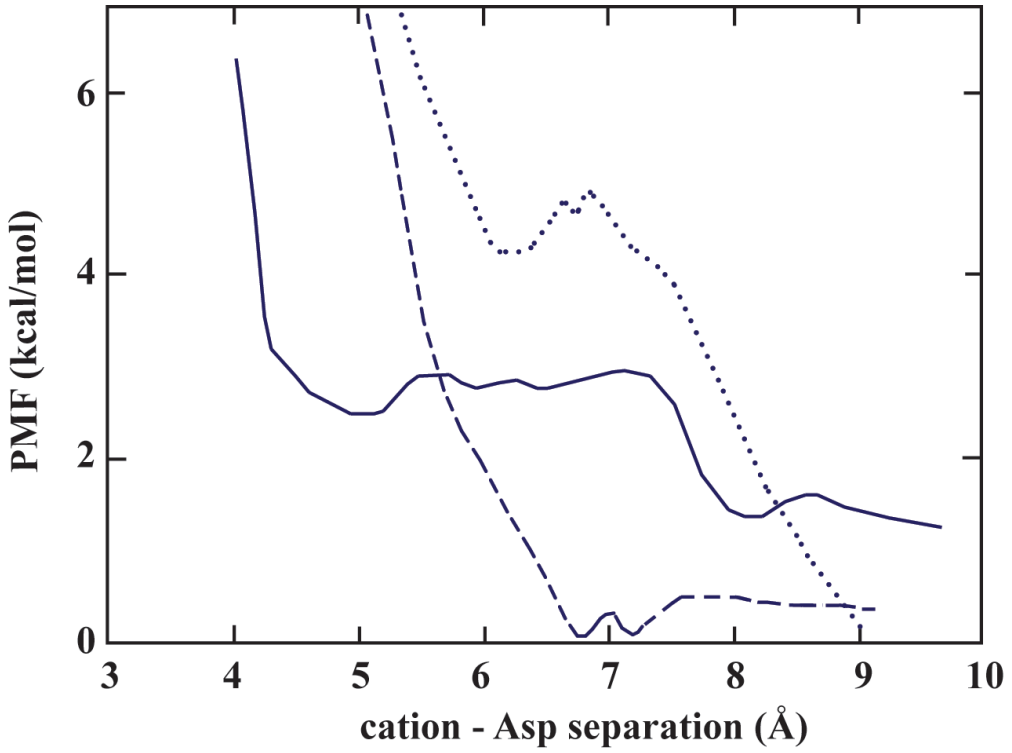
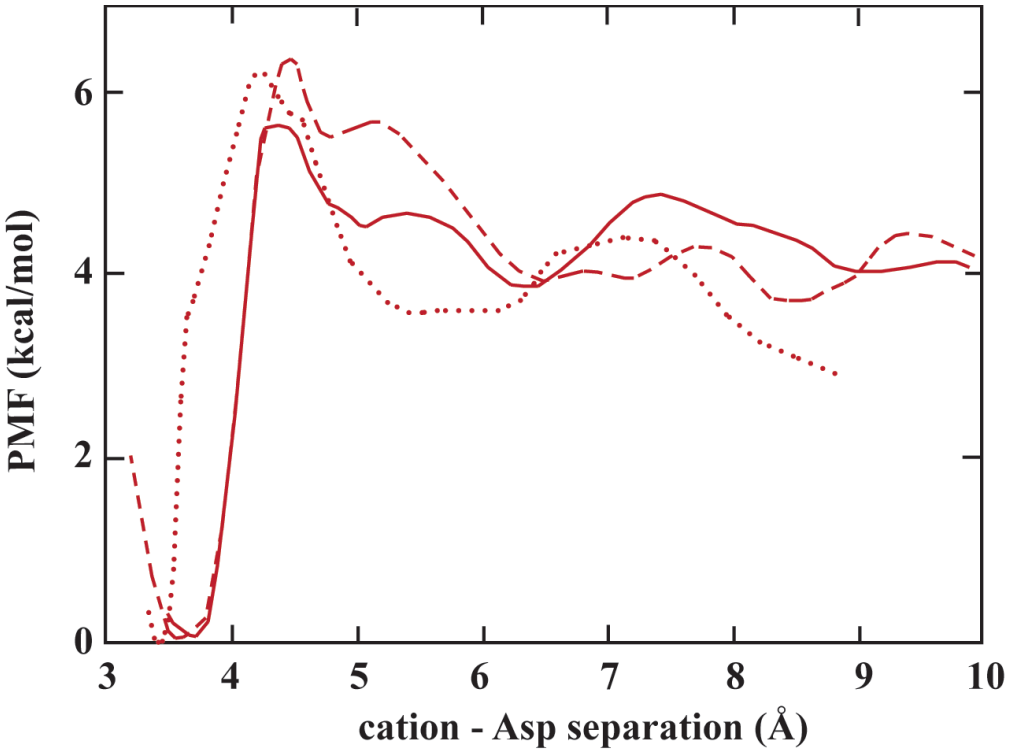


Figure A3. Calculated PMF profiles for a primary sphere water exchange event about Mg^{2+} (blue), Ca^{2+} (red), and Sr^{2+}(green). Solid lines correspond to the simulations presented in Chapter 2, using TIP3P water with the modified Åqvist (1990) ion-oxygen LJ parameters. Dashed lines are results from TIP3P water with paramters from Babu and Lin (2006). Dotted lines correspond to SPC water with the original Åqvist (1990) parameters.

(a) Mg^{2+}


PMF (kcal/mol)
0
2
4
6
3
4
5
6
7
8
9
10
cation - Asp separation (Å)

(b) Ca^{2+}


PMF (kcal/mol)
0
2
4
6
3
4
5
6
7
8
9
10
cation - Asp separation (Å)

(c) Sr^{2+}

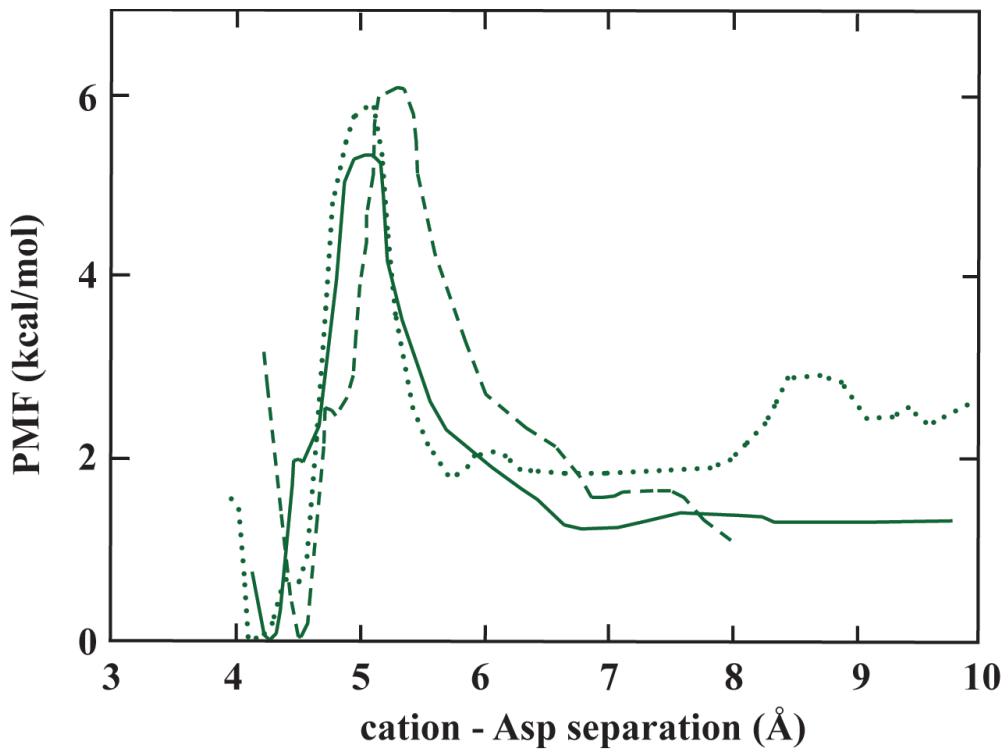


Figure A4. Calculated potential of mean force profiles for the association of (a) Mg^{2+}, (b) Ca^{2+}, and (c) Sr^{2+} with a single Asp molecule for three different water model/ion-oxygen LJ parameter combinations. Solid lines correspond to the simulations presented in Chapter 2, using TIP3P water with the modified Åqvist (1990) ion-oxygen LJ parameters. Dashed lines are results from TIP3P water with paramters from Babu and Lin (2006). Dotted lines correspond to SPC water with the original Åqvist (1990) parameters.

A5. Sample Python analysis routines

A5.1 Script for radial distribution function plotting and integration

```
#!/usr/bin/env python

import os
import math
import pylab
import numpy as np
import matplotlib.pyplot as plt
from pylab import *

#-----------------------------------------------------------------------------------------------
# input LAMMPS rdf file, write timestep, g(r) & n(r) to lists
#-----------------------------------------------------------------------------------------------


Path = '/enter data path/'                      # set data path
ofile = Path + 'shortRDF.txt'
input = open(ofile,'r')
print "------"
print "Generating plots..."

ListA=[]
ListB=[]
ListC=[]

on=0
count=0
for line in input.xreadlines():
        if count>1:
                entries = line.split()          # separate columns of text file
                A = (entries[0])
                columnA = float(A)
                ListA.append(columnA)

                B = (entries[1])
                columnB = float(B)
                ListB.append(columnB)

                C = (entries[2])
                columnC = float(C)
                ListC.append(columnC)
        count=count+1
```

```
input.close()

#-----------------------------------------------------------------------------------------------
# plot g(r) and n(r) vs. timestep for run average data
#-----------------------------------------------------------------------------------------------

ym = (max(ListB)+2)

lineA , =plot(ListA,ListB)
lineA.set_markersize(3)
lineA.set_marker('d')
lineA.set_markerfacecolor('g')
lineA.set_markeredgecolor('r')
lineB , =plot(ListA,ListC)
lineB.set_markersize(3)
lineB.set_marker('d')
lineB.set_markerfacecolor('r')
lineB.set_markeredgecolor('g')
axis([0,10,0,ym])
xlabel('Distance, Angstrom')
ylabel('g(r), n(r)')

savepath = Path + 'RDFplot.png'
pylab.savefig(savepath)
draw()
show()
```

A5.2 Script for reactive flux rate constant calculation

```
import os, sys; import math; import pylab
import numpy; import matplotlib.pyplot as plt
from pylab import *; import scipy
import scipy.optimize; from scipy import polyfit

print "------"

path = '/enter file path/'
ifile = path + 'WHAMout.txt'

#-----------------------------------------------------------------------------------------------
# Define constants
#-----------------------------------------------------------------------------------------------

beta = float(0.0004009168145)  #thermodynamic beta in units of mol/J (1/KbT)
Kb = 8.31427671    #Boltzmann's constant in units of J/mol-K
```

```
#mass = float(0.01242881935575)  #reduced mass for Ca/H2O pair in kg/mol
#mass = float(0.0149323992)   #reduced mass for Sr/H2O pair in kg/mol
#mass = float(0.0100956) #reduced mass for Na/H2O pair in kg/mol
#mass = float(0.01326219) #reduced mass for Fe/H2O pair in kg/mol
mass = float(0.0159266) #reduced mass for Ba/H2O pair in kg/mol
#mass = float(0.01028) #reduced mass for Mg/H2O pair in kg/mol
e = float(2.718281828)  #natural log constant
Rtst = float(3.46)   # ion-water distance corresponding to maximum PMF value

#-----------------------------------------------------------------------------------------
# Plot PMF and unbiased probabilty distribution
#-----------------------------------------------------------------------------------------

input = open(ifile,'r')

coord=[ ]; free=[ ]; gram=[ ]

for line in input.xreadlines():
        if 'Free' not in line and 'inf':
                obama = line.split()
                coordinate = (obama[0]).strip()
                pmf=float(obama[1])
                prob=(obama[3]).strip()
                if 'n' not in str(pmf) and 'n' not in str(prob):
                        coord.append(float(coordinate))
                        free.append(pmf*4184)
                        gram.append((float(prob)))


plot(coord,free,linewidth=2)                                    # plot ion-water PMF
xlabel("Ion-Water Separation, Angstroms")
ylabel("PMF, J/mol")
show()

penguin = raw_input("Press enter for unbiased histogram")

plot(coord,gram,linewidth=2)                                    # plot probability dist.
xlabel("Ion-Water Separation, Angstroms")
ylabel("Unbiased Probability")
show()

penguin = raw_input("Press enter to continue.")

#-----------------------------------------------------------------------------------------
# Calculate exp[-BW(r)] (integrand) for reactive flux calculation
#-----------------------------------------------------------------------------------------
```

```
print "Calculating integrands"
print "------"

input = open(ifile,'r')

coord=[]
integ = []
FE=[]
ALLcoord=[]

for line in input.xreadlines():
        if 'Free' not in line and 'inf':
                obama = line.split()
                coordinate = float((obama[0]).strip())
                pmf=float((obama[1]).strip())
                prob=(obama[3]).strip()
                if 'n' not in str(pmf) and 'n' not in str(prob):
                        FE.append(pmf*4184)
                        ALLcoord.append(coordinate)
                if 'n' not in str(pmf) and 'n' not in str(prob) and coordinate <= Rtst:
                        coord.append(coordinate)
                        raised = coordinate**2*e**(-beta*pmf*4814)
                        integ.append(raised)

#-----------------------------------------------------------------------------------------
# Integrate under curve
#-----------------------------------------------------------------------------------------

TotArea=[]
width = coord[1]-coord[0]

plot(coord,integ,linewidth=2)                           # plot filled curve
xlabel("Ion-Water Separation, Angstroms")               # during integration
ylabel("integrand r^2*exp[-BW(r)]")

i=0
while i < len(coord)-1:
        height = integ[i]
        area=height*width
        print "r =",coord[i],', area =',area
        TotArea.append(area)
        #x=[coord[i],coord[i+1],coord[i],coord[i+1],coord[i]]
        #y=[0,0,integ[i],integ[i],0]
        #plot(x,y,linewidth=0.5)
        #show()
```

```
        i=i+1

AreaSum=sum(TotArea)*1e-30
print "------"
print "Area under curve =",AreaSum,"meters^3"
print "------"

#-------------------------------------------------------------------------
# Plot integrated curve
#-------------------------------------------------------------------------

plot(coord,integ,linewidth=2)
xlabel("Ion-Water Separation, Angstroms")
ylabel("integrand exp[-BW(r)]")
show()

penguin = raw_input("Press enter to continue")

#-------------------------------------------------------------------------------------------------
# Find value of pmf at r=rtst
#-------------------------------------------------------------------------------------------------

print "------"
print "Getting PMF value at Rtst..."
print "      "

i=0
while i<len(ALLcoord):
        if ALLcoord[i]==Rtst or Rtst+0.001>=ALLcoord[i]>=Rtst-0.001:
                Wrtst = FE[i]
                print "r =",ALLcoord[i],"A, W =",Wrtst,"J/mol"
        i=i+1

print "PMF at transition state separation: r =",Rtst,"PMF =",Wrtst,"J/mol"
print "------"

#-------------------------------------------------------------------------------------------------
# Calculate transition state rate constant
#-------------------------------------------------------------------------------------------------

print "Calculating transition state rate constant..."
print "------"
A = ((Kb*300)/(2*pi*mass))**0.5
print "A =",A
top = (Rtst*1e-10)**2*e**(-beta*Wrtst)
```

```
Ktst = A*top/(AreaSum)
print "Ktst =",Ktst,"inverse seconds =",Ktst/1e9,"inverse nanoseconds"
```

A6. Sample LAMMPS input files

A6.1 Input file for calculation of baseline Ca^{2+} hydration properties

```
units           real
atom_style      full

pair_style      lj/charmm/coul/long 8 10
bond_style      harmonic
angle_style     harmonic
kspace_style    ewald 0.0001

read_data       CaCl2.data

pair_coeff      1 1 0.10200 3.1507
pair_coeff      2 2 0.00000 0.0000
pair_coeff      3 3 0.00000 0.0000
pair_coeff      4 4 0.00000 0.0000
pair_coeff      1 2 0.00000 0.0000
pair_coeff      1 3 0.26144 2.75676
pair_coeff      1 4 0.52160 3.7850

neighbor        2.0 bin
neigh_modify    delay 5

thermo_style    multi
thermo          5
group           cal molecule 234
group           oxy type 1
group           big type 1 3

fix             1 all nvt 300.0 300.0 100.0
fix             2 all shake 0.0001 20 10 b 1 a 1
fix             origin tether spring 50.0 0.0 0.0 0.0

group           H2O type <= 702
dump                    2 H2O atom 25 pdb.dump
dump_modify             2 image no scale yes

timestep                0.2
run                     300000
```

```
fix                     3 all rdf 25 Ca_rdf.out 300 3 1

dump_modify             3 scale no

timestep        1.0
run             10000000
```

A6.2 Input file for umbrella sampling calculation of Ca^{2+} - Asp potential of mean force profile

```
units           real
neigh_modify    delay 2 every 1

atom_style      full
bond_style      harmonic
angle_style     charmm
dihedral_style  charmm
improper_style  harmonic

pair_style      lj/charmm/coul/long 8.0 10.0 10.0
pair_modify     mix arithmetic
kspace_style    pppm 0.0001

read_data       Asp.data

pair_coeff      3 10 0.00000 0.0000
pair_coeff      10 11 0.26144 2.75676
pair_coeff      10 12 0.52160 3.7850

group asp molecule 1
group cal molecule 192

special_bonds           charmm
fix                     1 all nvt 300.0 300.0 300.0
fix                     2 all shake 1e-6 500 0 b 9 a 15
fix                     4 asp spring tether 100.0 0.0494 0.467 0.217 0.0
velocity                all create 0.0 12743586 dist gaussian

thermo                  1
thermo_style            multi

dump                    3 all atom 25 pdb.dump
dump_modify             3 image no scale yes

timestep                0.2
run                     100000
```

```
timestep        1.0
run             300000

fix             6 asp com 50 com.out

dump            1 all atom 50 collect.dump
dump_modify     1 scale no


fix             3 asp spring couple cal 10.0 0.0444716 0.402963 0.0808854 9.5

timestep        1.0
run             1000000

fix             3 asp spring couple cal 10.0 0.0444716 0.402963 0.0808854 9.0

timestep        1.0
run             1000000

fix             3 asp spring couple cal 10.0 0.0444716 0.402963 0.0808854 8.50

timestep        1.0
run             1000000


fix             3 asp spring couple cal 10.0 0.0444716 0.402963 0.0808854 8.00

timestep        1.0
run             1000000

fix             3 asp spring couple cal 10.0 0.0444716 0.402963 0.0808854 7.50

timestep        1.0
run             1000000


fix             3 asp spring couple cal 10.0 0.0444716 0.402963 0.0808854 7.00

timestep        1.0
run             1000000

fix             3 asp spring couple cal 10.0 0.0444716 0.402963 0.0808854 6.50

timestep        1.0
run             1000000
```

```
fix             3 asp spring couple cal 10.0 0.0444716 0.402963 0.0808854 6.00

timestep        1.0
run             1000000

fix             3 asp spring couple cal 10.0 0.0444716 0.402963 0.0808854 5.50

timestep        1.0
run             1000000

fix             3 asp spring couple cal 10.0 0.0444716 0.402963 0.0808854 5.00

timestep        1.0
run             1000000

fix             3 asp spring couple cal 10.0 0.0444716 0.402963 0.0808854 4.50

timestep        1.0
run             1000000
fix             3 asp spring couple cal 10.0 0.0444716 0.402963 0.0808854 4.00

timestep        1.0
run             1000000

fix             3 asp spring couple cal 10.0 0.0444716 0.402963 0.0808854 3.50

timestep        1.0
run             1000000

fix             3 asp spring couple cal 10.0 0.0444716 0.402963 0.0808854 3.00

timestep        1.0
run             1000000
```

A7. Supporting Figures for Sr^{2+} and Mg^{2+}

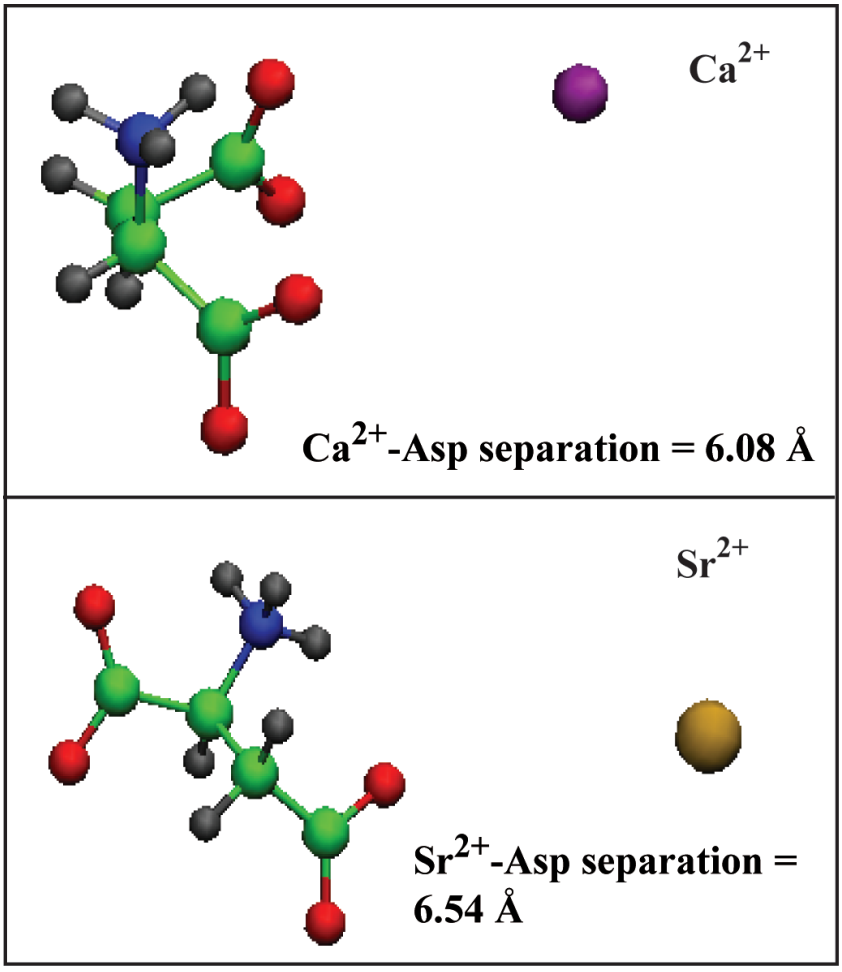


Figure A5. Images taken from simulations used to generate cation – Asp potential of mean force profiles. The cation – Asp separation distances, 6.08 Å for Ca^{2+} (purple sphere) and 6.54 Å for Sr^{2+} (yellow sphere), were chosen to show the different conformations of Asp as the ions approach yet before cation dehydration has occurred. In the case of Ca^{2+}, we observe bending of the Asp molecule so that both carboxylate groups are facing the incoming ion. Alternatively, Sr^{2+}approaches only one of the carboxylate groups and Asp remains extended. Water moledulces have been deleted from these images for clarity.

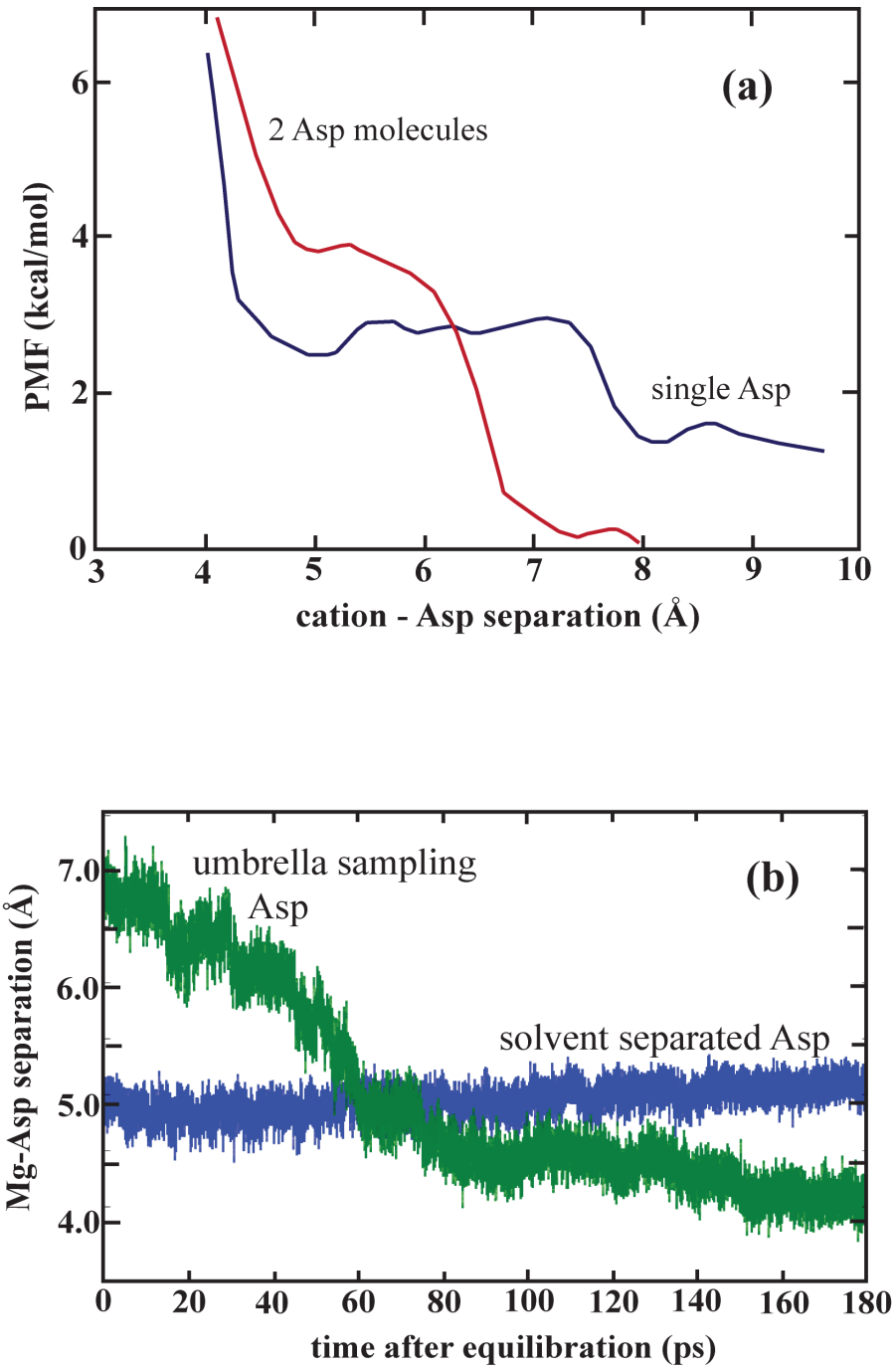


Figure A6. (a) Mg^{2+} - Asp potential of mean force profiles and (b) Mg^{2+} - Asp separation distances for umbrella sampling with one Mg^{2+} cation and two Asp molecules. For this simulation, one Asp was fixed at a distance corresponding to a solvent separated Mg^{2+} - Asp pair (b, blue line) while the second Asp was moved progressively closer to the cation (b, green line). The PMF profile corresponding to the interaction of the second Asp molecule with Mg^{2+} (a, red line) is compared to that from the simulation with a single Asp molecule reported in Chapter 2 (a, blue line).

A8. Reference

(1) Ewald, P. *Ann. Phys.* **1921,** *369*, 253-287.

Appendix B. Supplementary Information to Chapter 3

B1. Python routine for calculation of water residence times with the Impey method

```
import os, sys;          import math
import glob;             import pylab
import numpy;            import matplotlib.pyplot as plt
from pylab import *

#-----------------------------------------------------------------------------------------------
# Input
#-----------------------------------------------------------------------------------------------

Path = '/enter data path/'
ifile = Path + 'ShortDump.dump'

IonNum = int(raw_input("Enter atom type number for ion (from data file): "))
print "------"
Onum = int(raw_input("Enter atom type number for water oxygen: "))
print "------"
step = float(raw_input("Enter length of timestep (in femtoseconds): "))
print "------"
cutoff = float(raw_input("Enter ion-oxygen 1st shell cutoff distance: "))
print "------"

#-----------------------------------------------------------------------------------------------
# Get list of oxygen ID numbers
#-----------------------------------------------------------------------------------------------

input=open(ifile,'r')

firstshell=[]; T=[]
timer=0
for line in input.xreadlines():
        if timer==2:
                break
        if 'TIMESTEP' in line:
                parsed = line.split()
                time = float(parsed[1])
                timer=timer+1
                T.append(time)
        else:
                parsed = line.split()
                type = float(parsed[1])
```

```
                id = float(parsed[0])
                if type == Onum:
                        firstshell.append(id)
input.close()

start = T[0]
print "Starting timestep = ",start

#-----------------------------------------------------------------------------------------------
# Get lists of ion XYZ coordinates
#-----------------------------------------------------------------------------------------------

femto=[]; Pico=[]
input=open(ifile,'r')
time=-1
CAid=[]; CAx=[]; CAy=[]; CAz=[]
for line in input.xreadlines():
        if "TIMESTEP" in line:
                parsed = line.split()
                time = float(parsed[1])
                real_time = (time*step)-(start*step)
                femto.append(real_time)
                pico = float(real_time/1000); Pico.append(pico)
        else:
                parsed=line.split()
                iran=float(parsed[1])
                if iran==IonNum:
                        CAx.append(float(parsed[2]))
                        CAy.append(float(parsed[3]))
                        CAz.append(float(parsed[4]))
input.close()

print "   ",len(CAx),"ion coordinates found"
print "   ",len(femto),"timesteps recorded"

#-----------------------------------------------------------------------------------------------
# Calculate separation distances for each oxygen atom and plot vs. simulation time
#-----------------------------------------------------------------------------------------------

#pngdir = Path + '/AspO_Plots/'
pngdir = Path + '/WaterPlots_Primary/'
os.mkdir(pngdir)

for tag in firstshell:
        time=-1
        input=open(ifile,'r')
```

```
counter=0
seplist=[]

for line in input.xreadlines():
        if "TIMESTEP" in line:
                parsed = line.split()
                time = float(parsed[1])
        else:
                parsed=line.split()
                id = float(parsed[0])
                if id == tag:
                        Ox=(float(parsed[2]))
                        Oy=(float(parsed[3]))
                        Oz=(float(parsed[4]))
                        distsq = (CAx[counter]-Ox)**2+(CAy[counter]-
                        Oy)**2+(CAz[counter]-Oz)**2
                        sepdist = distsq**0.5
                        seplist.append(sepdist)
                        counter=counter+1
input.close()

minx=min(femto)
maxx=max(femto)
miny=1.0
maxy=cutoff+3

lineA , =plot(femto,seplist,linewidth=0.1)
lineA.set_markersize(1)
lineA.set_marker('d')
lineA.set_markerfacecolor('g')
lineA.set_markeredgecolor('r')
axis([minx,maxx,miny,maxy])
xlabel('Time After Equilibration (fs)')
ylabel('Separation Distance (angstrom)')
title(tag)
savepath = pngdir + str(tag) + '.png'
pylab.savefig(savepath)
pylab.clf()


#-----------------------------------------------------------------------------------------
# Calculate ion-oxygen separation distances for each water molecule
# Write to temporary output files
#-----------------------------------------------------------------------------------------
```

```
datadir = Path + '/WaterData_Primary/'
os.mkdir(datadir)

for n in firstshell:
	counter=-1
	newpath2 = datadir + str(n) + '/'
	os.mkdir(newpath2)
	waterout = newpath2 + 'waterdata.txt'
	output = open(waterout,'w')
	input = open(ifile,'r')
	for line in input.xreadlines():
		if "TIMESTEP" in line:
			counter = counter +1
		else:
			parsed = line.split()
			id = float(parsed[0])
			if id == n:
				Ox = float(parsed[2])
				Oy = float(parsed[3])
				Oz = float(parsed[4])
				distsq = (CAx[counter]-Ox)**2+(CAy[counter]-
Oy)**2+(CAz[counter]-Oz)**2
				sepdist = distsq**0.5
				output.write(str(sepdist))
				output.write('\n')

	output.close()
	input.close()

#-----------------------------------------------------------------------------------------------
# Write oxygen residence files
#-----------------------------------------------------------------------------------------------
gap = float(T[1]-T[0])
print "timestep gap =",gap
tstar = float(step*2000/gap) #determine number of steps for t*

for n in firstshell:
	print "====================================="
	print "Water number",n
	io=[]; distlist=[]
	infile = datadir + str(n) + '/waterdata.txt'
	sumfile = datadir + str(n) + '/SumList.txt'
	output = open(sumfile,'w')
	input = open(infile,'r')

	for line in input.xreadlines():
```

```
        dist = float(line.strip())
        distlist.append(dist)
        if dist < cutoff:
                io.append(1)
        if dist >= cutoff:
                io.append(0)

print " ",len(io),"data points"

q=0
iostar=[]; done=0
for k in io:
        leave = 0; left = len(io)-tstar
        if done==1:
                leavings = len(io)-q
                j=1
                while j<=leavings:
                        iostar.append(0)
                        j=j+1
                break
        if float(k)==1:
                iostar.append(1)
        elif float(k)==0 and q<left:
                print "  Water out of first shell at point",q,"checking for re-entry..."
                r=1
                while r<=tstar:
                        check = io[q+r]
                        #print "      timestep",q+r,", separation =",distlist[q+r]
                        if check == 1:
                                print "  Water re-enters shell within 2 ps"
                                leave=1
                                iostar.append(1)
                                break
                        r=r+1
                if leave == 0:
                        print "  Water does not re-enter first shell"
                        iostar.append(0)
                        done = 1
                        #print "Water",n,"leaves first shell at timestep",q*step

        elif float(k)==0 and q>=left:
                print "  Water out of first shell at point",q,"checking for re-entry..."
                leftover=len(io)-q
                r=1
                while r<leftover:
                        check = io[q+r]
```

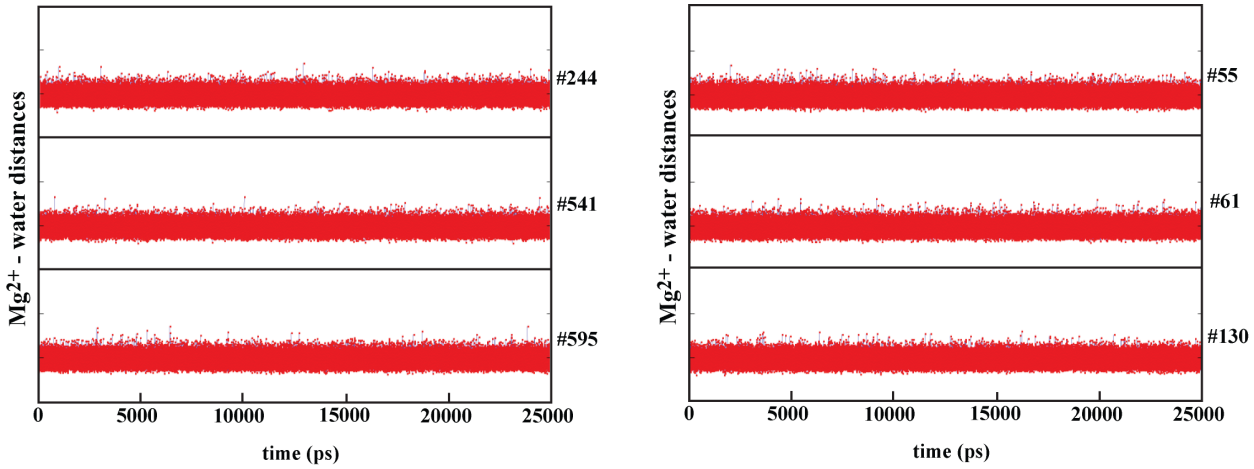
```
                                #print "      timestep",q+r,", separation =",distlist[q+r]
                                if check ==1:
                                        print "  Water re-enters shell within 2 ps"
                                        leave=1
                                        iostar.append(1)
                                        break
                                r=r+1
                        if leave == 0:
                                print "  Water does not re-enter first shell"
                                iostar.append(0)
                                done==1
                                #print "Water",n,"leaves first shell at timestep",q*step
                q=q+1
        for switch in iostar:
                output.write(str(switch))
                output.write('\n')

        #plot(femto,iostar,'b.', markersize=1.5)
        #xlabel('Time After Equilibration (fs)')
        #ylabel('one or zero')
        #title(n)
        #axis([min(femto),max(femto),-1,2])
        #show()
        #q = raw_input("  Hit any key to continue")

        input.close()
        output.close()

#-----------------------------------------------------------------------------------------------
# Write combined output file for R(t) calculation
#-----------------------------------------------------------------------------------------------

bigpath = datadir + 'combined.txt'
output = open(bigpath,'w')
for n in firstshell:
        output.write(str(n).rjust(0))
        sumfile = datadir + str(n) + '/SumList.txt'
        input=open(sumfile,'r')
        y=0
        for line in input.xreadlines():
                switch = line.strip()
                output.write(str(switch).rjust(2))
        output.write('\n')
output.close()
input.close()
```

```
#-----------------------------------------------------------------------------------------
# Compute R(t)
#-----------------------------------------------------------------------------------------

print "========================="
print "    Calculating R(t)..."

columns = float(len(femto))
index = 0; Nh=float(len(firstshell))
Resid=[]

while index < columns:
        sum = float(0)
        input=open(bigpath,'r')
        for line in input.xreadlines():
                parse = line.split()
                summer = float(parse[index])
                sum = sum + summer
        arr = float((1/Nh)*sum)
        Resid.append(arr)
        index = index+1
        input.close()

outpath = Path + 'Rdata.txt'
output = open(outpath, 'w')

i=0
for item in Resid:
        if i>0:
                output.write(str(femto[i])); output.write(' '); output.write(str(item));
output.write('\n')
        i=i+1
output.close()

print "   ",len(Resid),"R values computed"

#-------------------------------------------------------------------------
# Integrate under curve
#-------------------------------------------------------------------------
print "Integrating R(t)..."
int = float(0)
width = Pico[2]-Pico[1]

print "Width =",width
print "------"
```

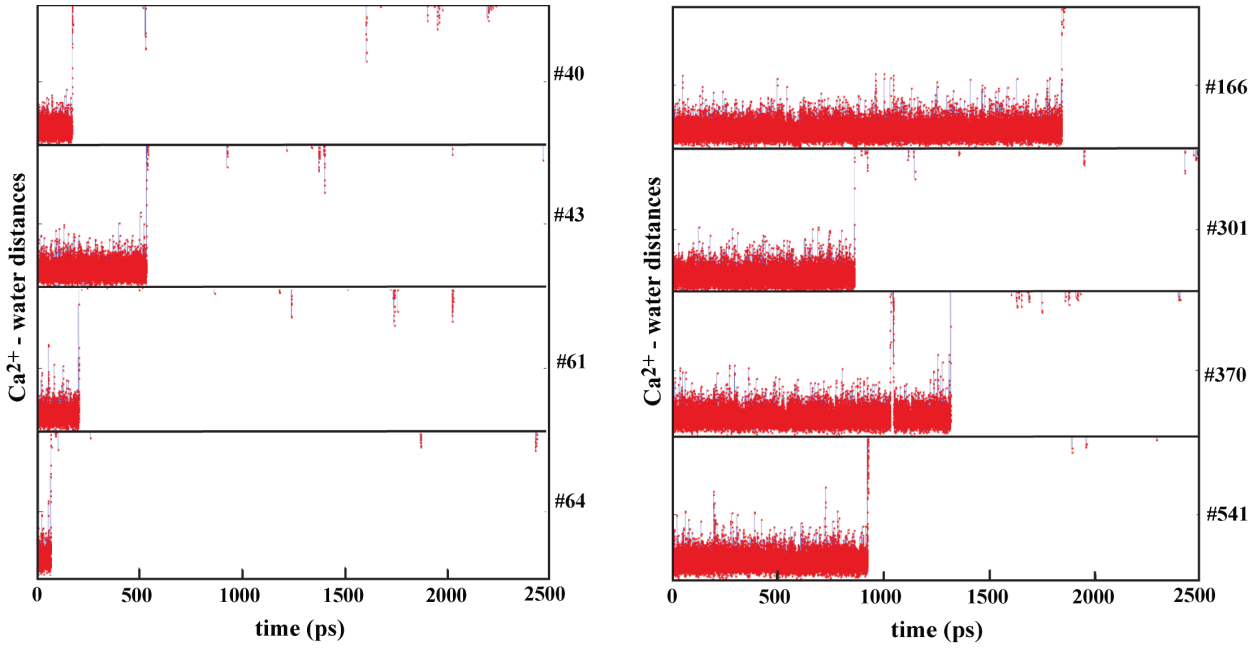
```
#plot(Time,R,linewidth=2)
#xlabel("Time (ps)")
#ylabel("R(t)")

i=0
while i < len(Pico)-1:
        height = Resid[i]
        area=float(height*width)
        int = int+area
        #print "Time =",Time[i],"ps, Int =",int
        #x=[Time[i],Time[i+1],Time[i],Time[i+1],Time[i]]
        #y=[0,0,R[i],R[i],0]
        #plot(x,y,linewidth=0.5)
        i=i+1

#-----------------------------------------------------------------------------------------------
# Plot R(t)
#-----------------------------------------------------------------------------------------------

plot(femto,Resid,'b.', markersize=1.5)
xlabel('Time After Equilibration (fs)')
ylabel('Residence Correlation Function')
title(int)
axis([min(femto),max(femto),0,1])


savepath =  Path + '/DirectR.png'
pylab.savefig(savepath)
show()
q = raw_input("Hit any key to quit")

pylab.clf()
print "------"
print Path
print "Direct integration area =",int,"ps"
```

B2. Hydration Water Trajectories

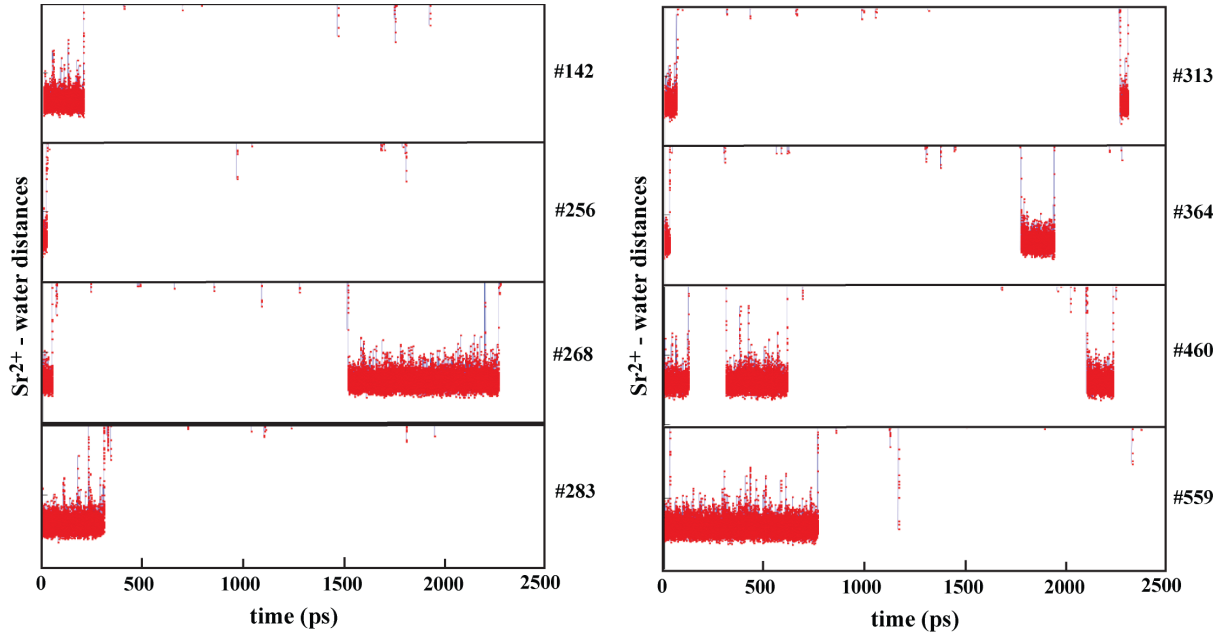
(a) Mg^{2+}



(b) Ca^{2+}



(c) Sr^{2+}



(d) Ba^{2+}

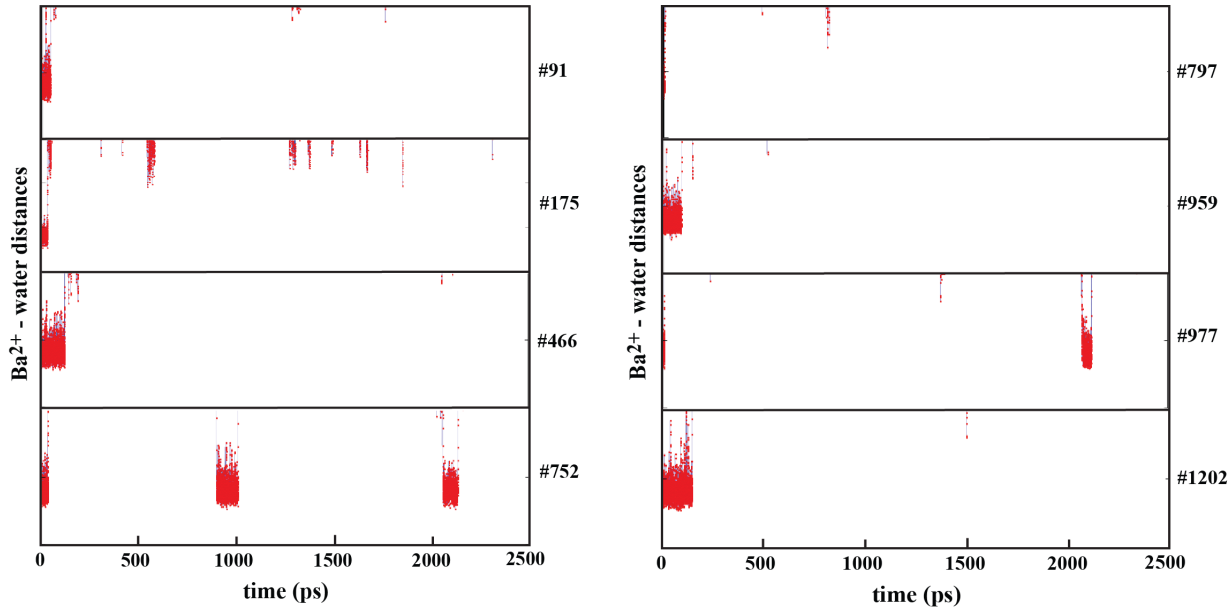
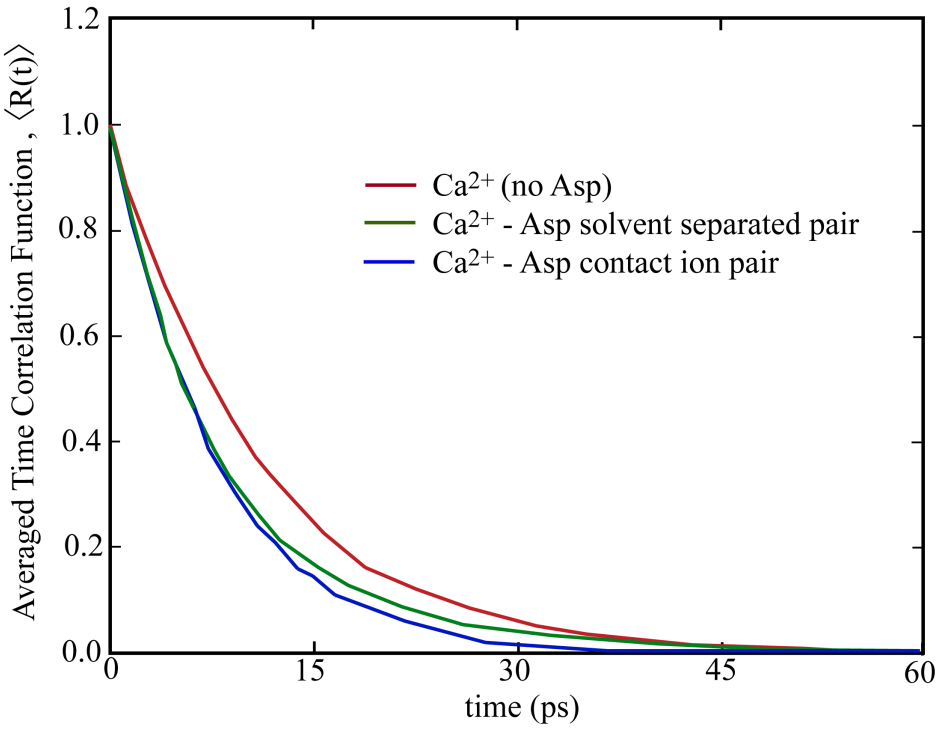


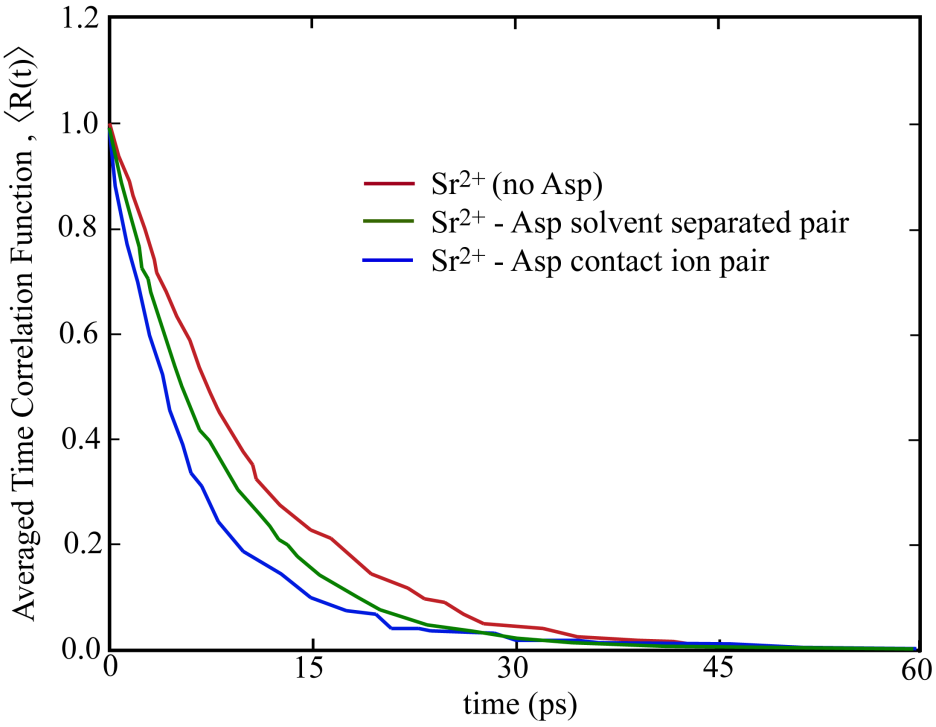
Figure B1. The trajectories of all initially primary sphere hydration waters over the course of a simulation for (a) Mg^{2+}, (b) Ca^{2+}, (c) Sr^{2+} and (d) Ba^{2+}. No water exchange events are observed for Mg^{2+} during a 25 ns simulation; all water molecules remain in the primary sphere. For Ca^{2+}, Sr^{2+} and Ba^{2+}, all primary sphere water molecules are exchanged during a 2500 ps simulation. The average residence times of primary-sphere waters decrease with decreasing cation charge density. The numbers appearing at right are id numbers given to each water molecule for bookkeeping purposes.

B3. Secondary sphere water residence correlation functions for caiton – Asp interaction

(a) Ca^{2+}



(b) Sr^{2+}



(c) Ba^{2+}

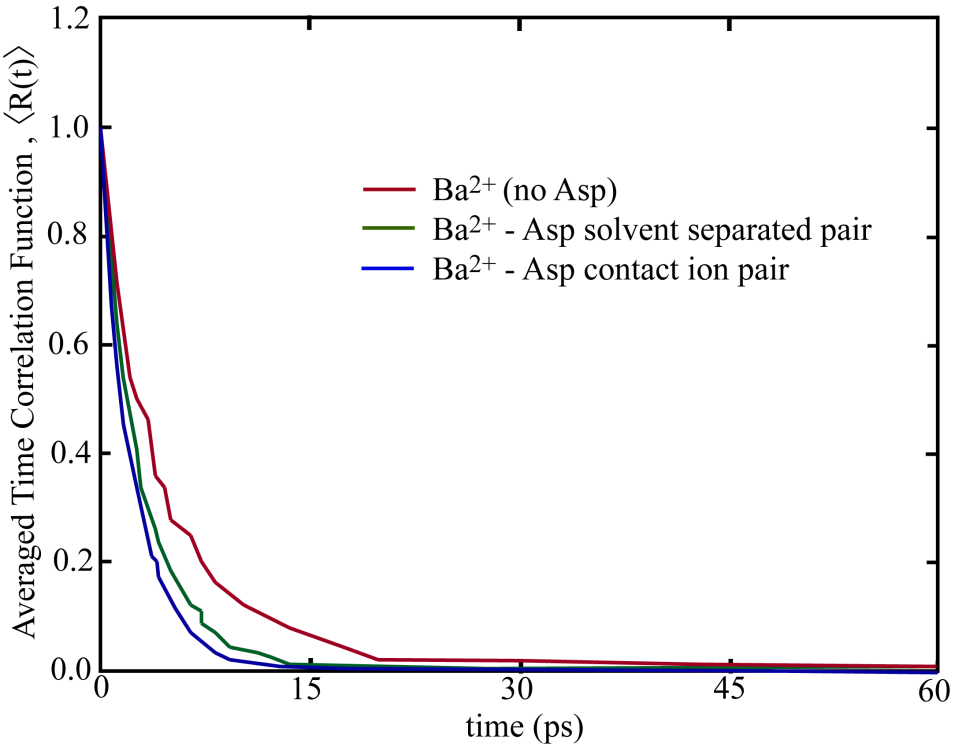
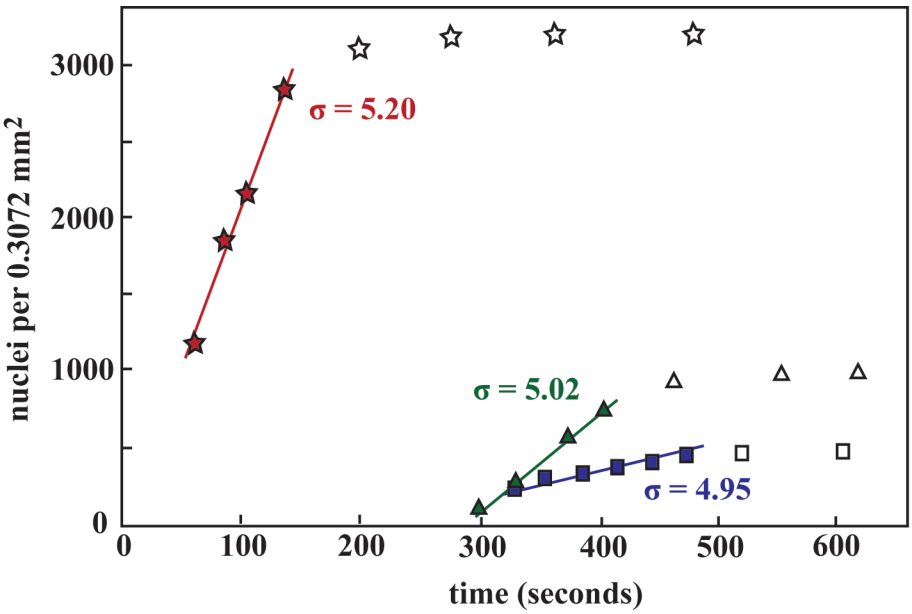


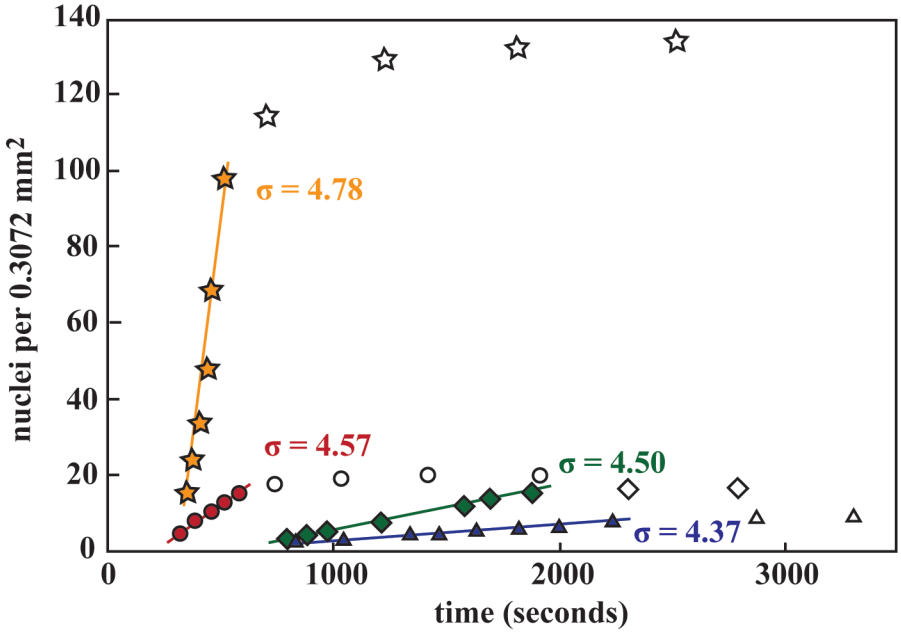
Figure B2. Ensemble averaged time correlation functions for secondary sphere hydration waters of (a) Ca^{2+}, (b) Sr^{2+} and (c) Ba^{2+} interacting with a single Asp molecule.

Appendix C. Supplementary Information to Chapter 4

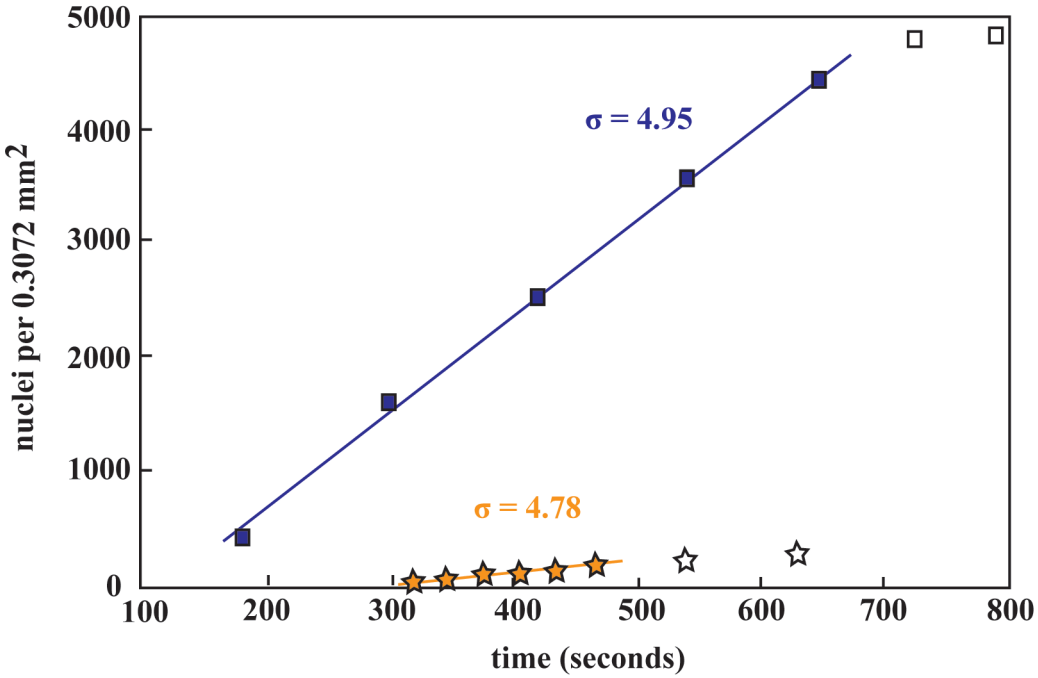
C1. Raw nucleation data

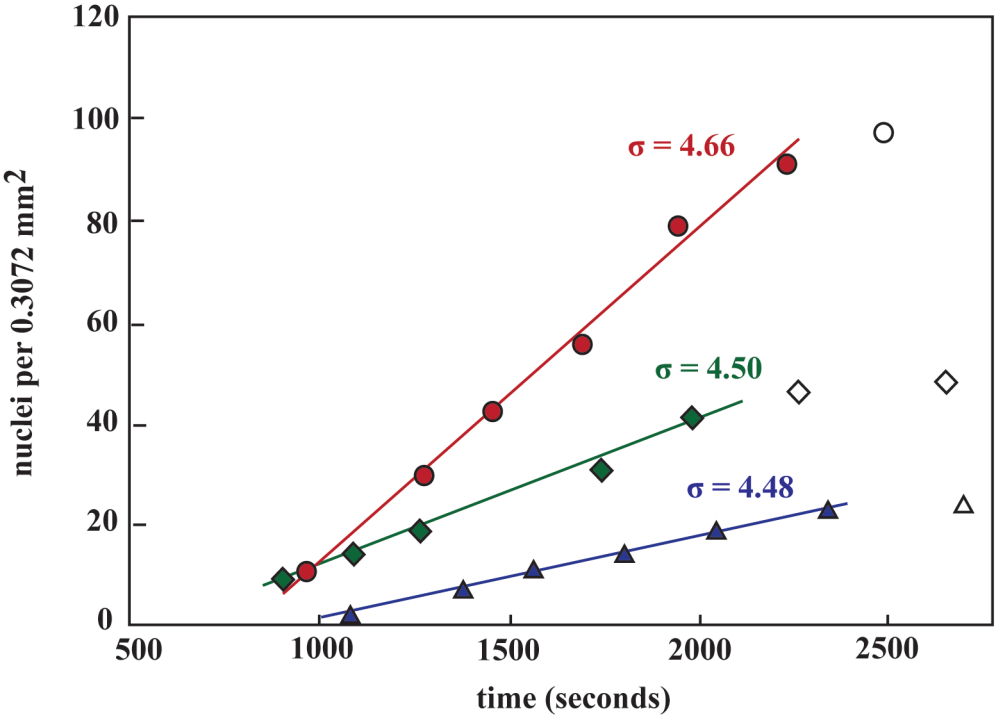
(a) C16-COOH



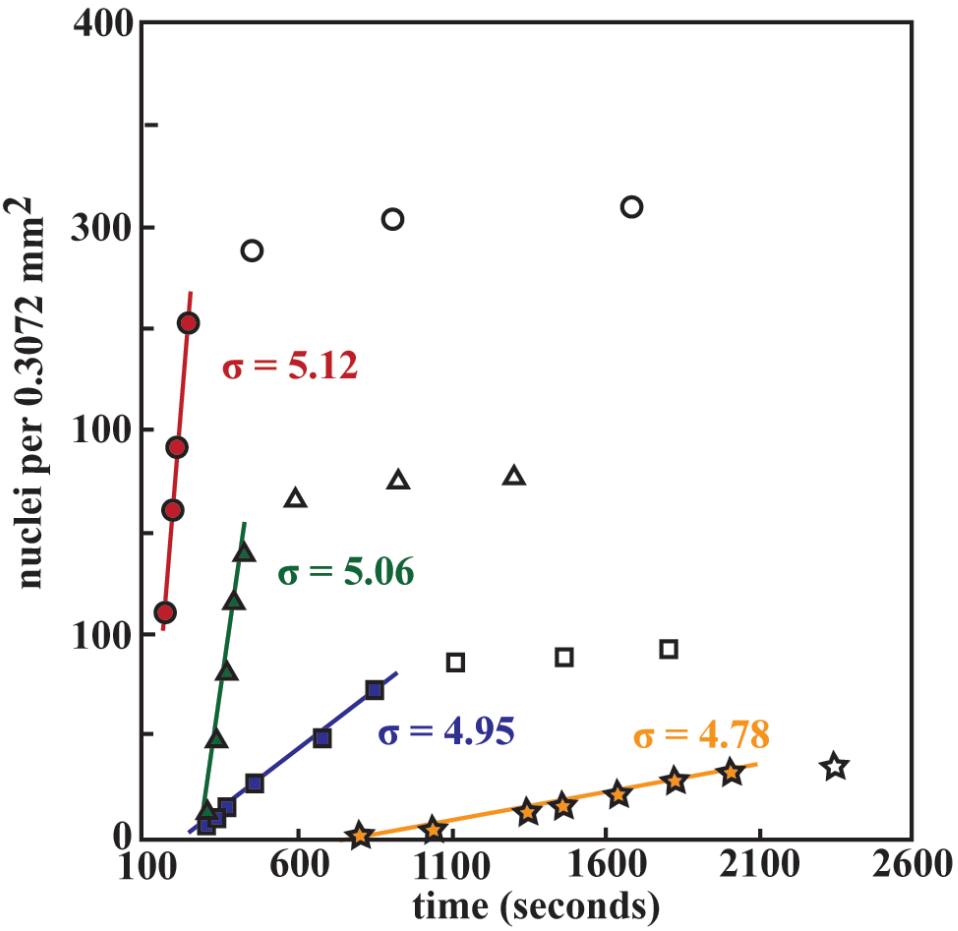


(b) C16-SH


5000
4000
3000
2000
1000
0
nuclei per 0.3072 mm2
σ = 4.95
σ = 4.78
100
200
300
400
500
600
700
800
time (seconds)


120
100
80
60
40
20
0
nuclei per 0.3072 mm2
σ = 4.66
σ = 4.50
σ = 4.48
500
1000
1500
2000
2500
time (seconds)

(c) C11-SH


400
300
100
100
0
nuclei per 0.3072 mm2
σ = 5.12
σ = 5.06
σ = 4.95
σ = 4.78
100
600
1100
1600
2100
2600
time (seconds)

(d) C11-PO_4

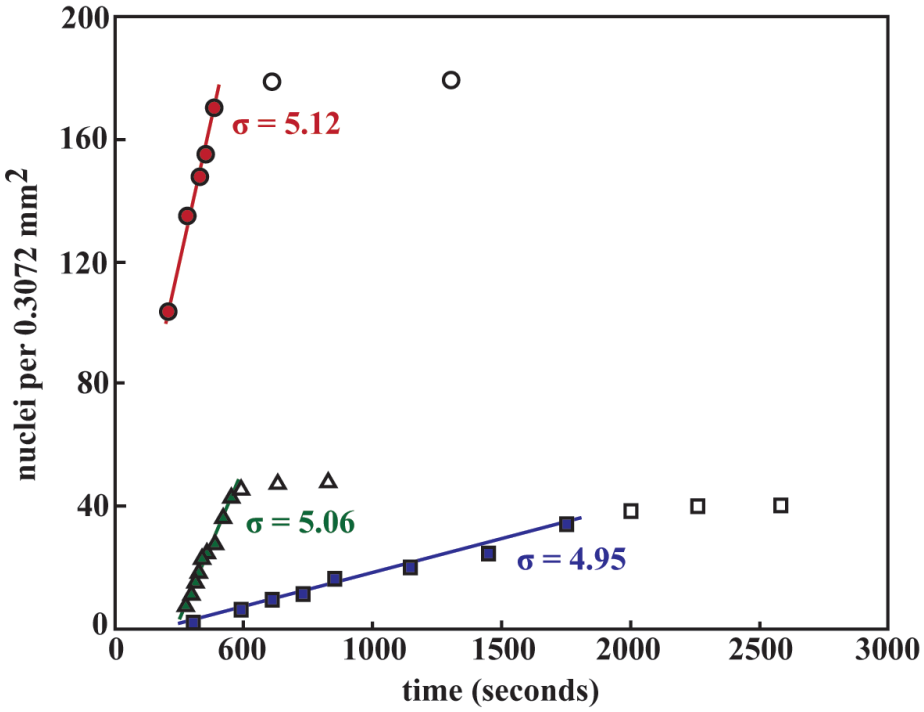


Figure C1. Raw optical microscopy data showing the appearance of calcite crystallites over time on substrates functionalized with (a) C16-COOH, (b) C16-SH, (c) C11-SH and (d) C11-PO_4. Data for C16-COOH and C16-SH are presented on two different sets of axes because the range of experimental supersaturations results in vastly different nucleation rates. Steady state rates were determined from the regions where crystallite appearance was linear with time (filled in data points with linear regressions). At greater times, the appearance of new crystallites begins to level out until no new surface-directed nucleation is observed (unfilled data points). These data were omitted in the determination of nucleation rates.

C2. Dynamic force microscopy (DFM)

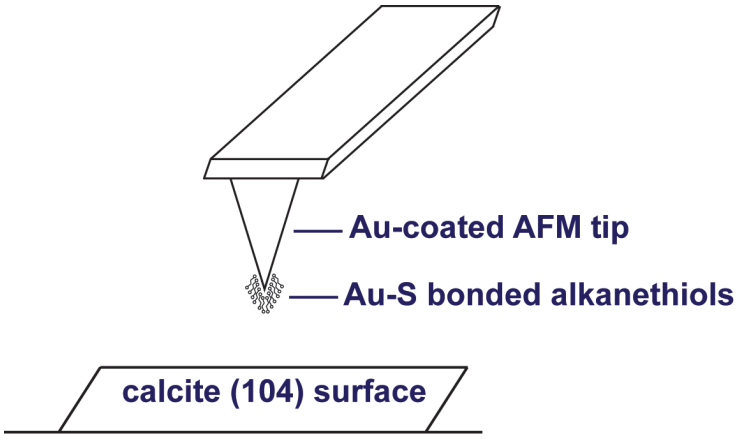


Figure C2. Schematic of the DFM setup used to measure substrate – calcite rupture forces. Alkanethiol molecules are tethered to the gold-coated catilever tip by Au-S bonds. Rupture forces are determined by moving the modified tip into contact with a hydrated calcite surface and then measuring the force required to retract the tip back to a position of zero deflection.

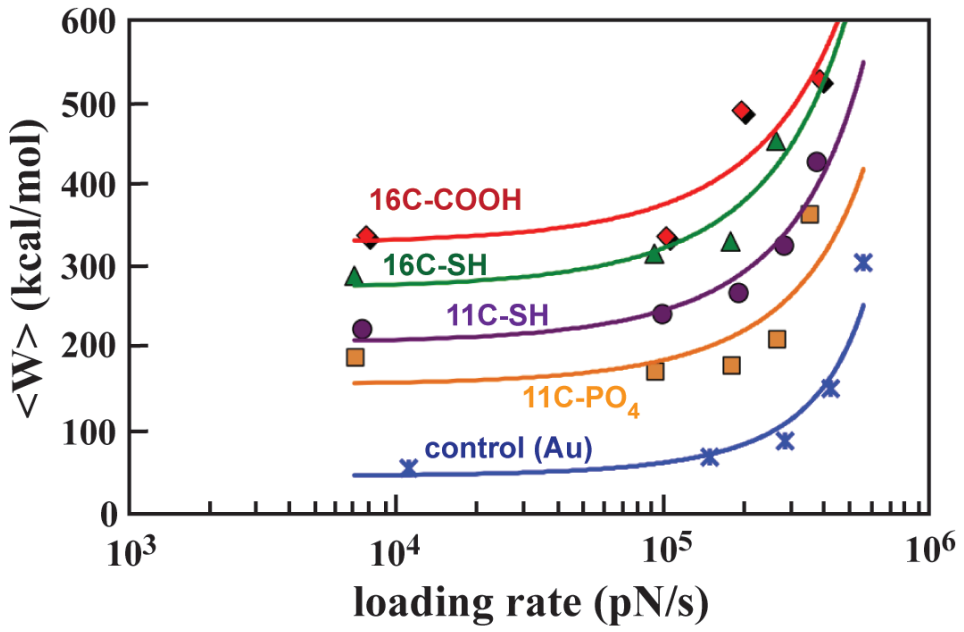


Figure C3. Mean work of dissociating a functionalized tip from the calcite (104) surface as a function of DFM loading rate. <*W*> is calculated from measured rupture forces assuming that the force is associated with a single molecule bond rupture. Because the <*W*> projected for zero loading rate is much higher than expected for a single bond rupture (above 300 kcal/mol for 16C-COOH compared to an expected single-molecule value on the order of 70 kcal/mol), we conclude that the measured rupture forces are for multiple bonds and not a single event.

C3. Nucleation from magnesium-containing solutions

The effective supersaturation, σ_{eff}, accounts for the increase in calcite solubility that occurs with magnesium incorporation and was determined by the following relations:

$$xMg^{2+} + yCa + (x+y)CO_3^{2-} \leftrightarrow Mg_xCa_y(CO_3)_{x+y}$$

Equation C1. Reaction for the formation of magnesian calcite.

$$\sigma_{eff} = \ln\left(\frac{a^{x}_{Mg^{2+}} a^{y}_{Ca^{2+}} a^{(x+y)}_{CO_3^{2-}}}{K'_{sp}}\right)$$

Equation C2. Expression for effective supersaturation of magnesium-containing calcites.

The stoichiometric constants, x and y, and solubility constants for Mg-calcite, K'_{sp}, were estimated from Mucci and Morse[1] based on solution compositions.

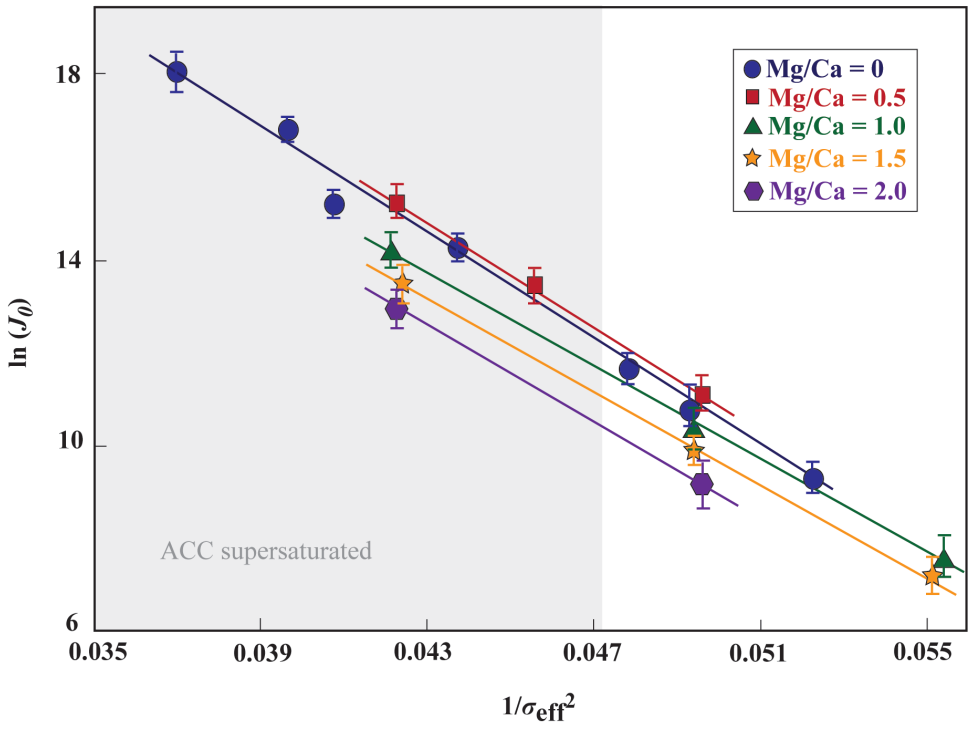


Figure C4. Plots of ln (J_0) versus $1/\sigma_{eff}^2$ for calcite nucleation on C16-COOH from solutions of varying magnesium content. The shaded region indicates supersaturation with respect to amorphous calcium carbonate (ACC).

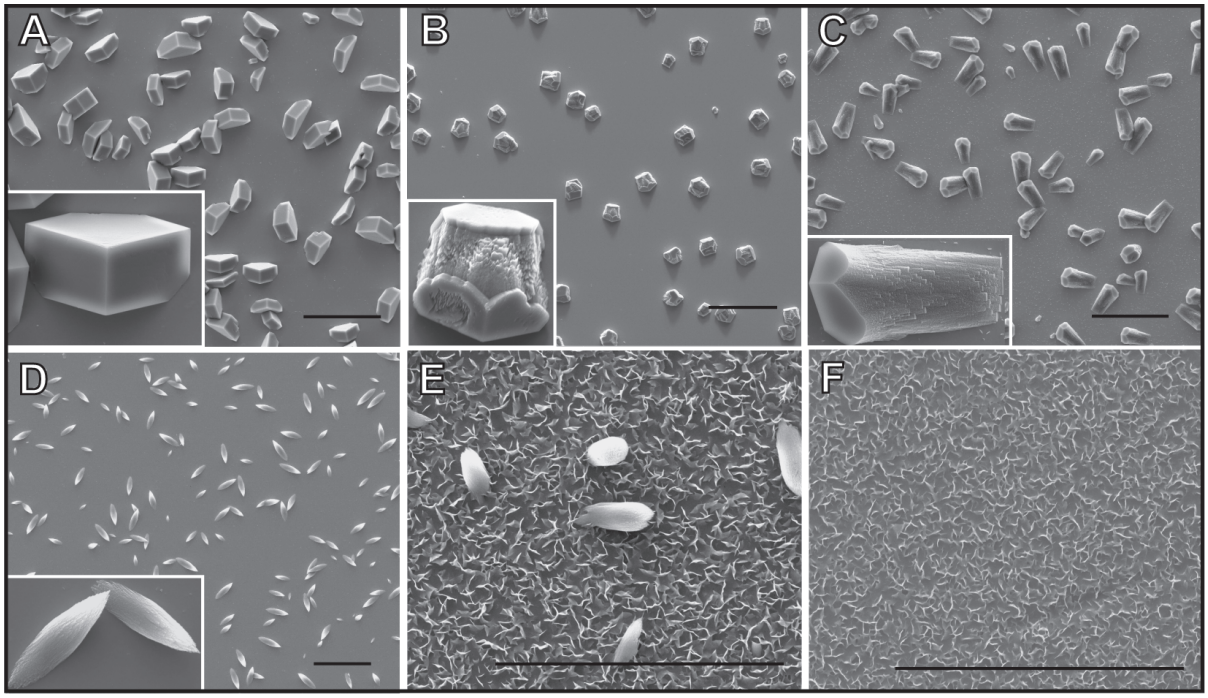


Figure C5. SEM images show the effect of Mg on $CaCO_3$ nucleation on 16C-COOH from solutions with $\sigma_{calcite}$ = 4.73, pH = 10.55, and Mg^{2+}/Ca^{2+} = (A) 0; (B) 0.5; (C) 1.0; (D) 1.5; (E) 2.5; (F) 3.0. Scale bars = 50 μm. In the absence of Mg^{2+}, precipitates are rhombohedral calcites preferentially oriented with the (012) plane parallel to the substrate. The addition of Mg^{2+} to the nucleating solution results in calcite crystallites with roughened faces, rounded edges, and c-axis elongation. At $Mg^{2+}/Ca^{2+} > 2$, the Mg-calcite is accompanied by a vaterite film, and at $Mg^{2+}/Ca^{2+} > 2.5$, only vaterite is observed. The substrates in (E) and (F) were removed from the flow-through nucleation chamber after 3 hours. In replicate experiments, no transformation of the vaterite was observed after 12 hours of exposure to supersaturated solutions.

C4. Optical microscopy method for measuring nucleation rates

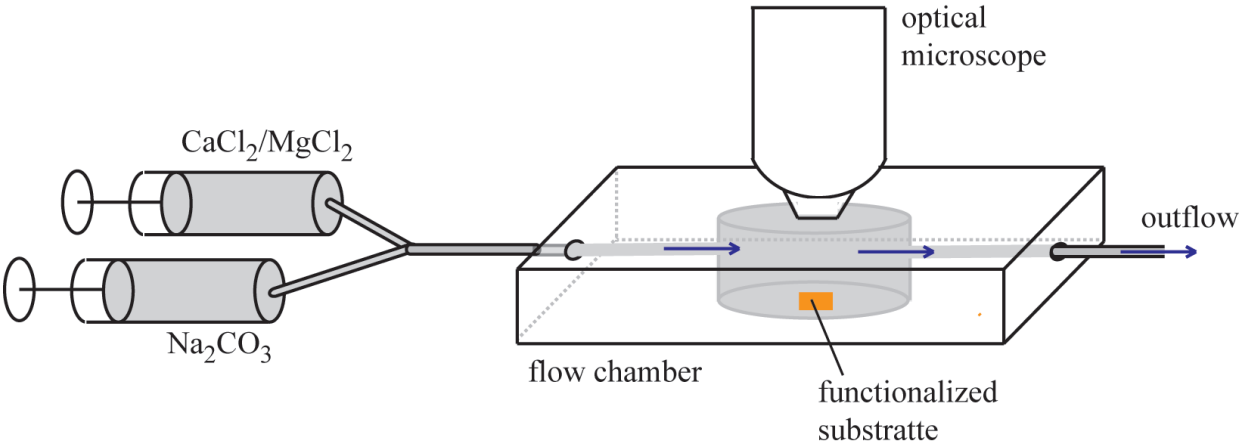


Figure C6. Schematic of the experimental setup used to measure calcite nucleation rates on functionalized substrates. A high precision syringe pump continuously flows the two solutions into the flow chamber containing the substrate. The crystals that nucleate on the substrate are monitored form above using optical microscopy.

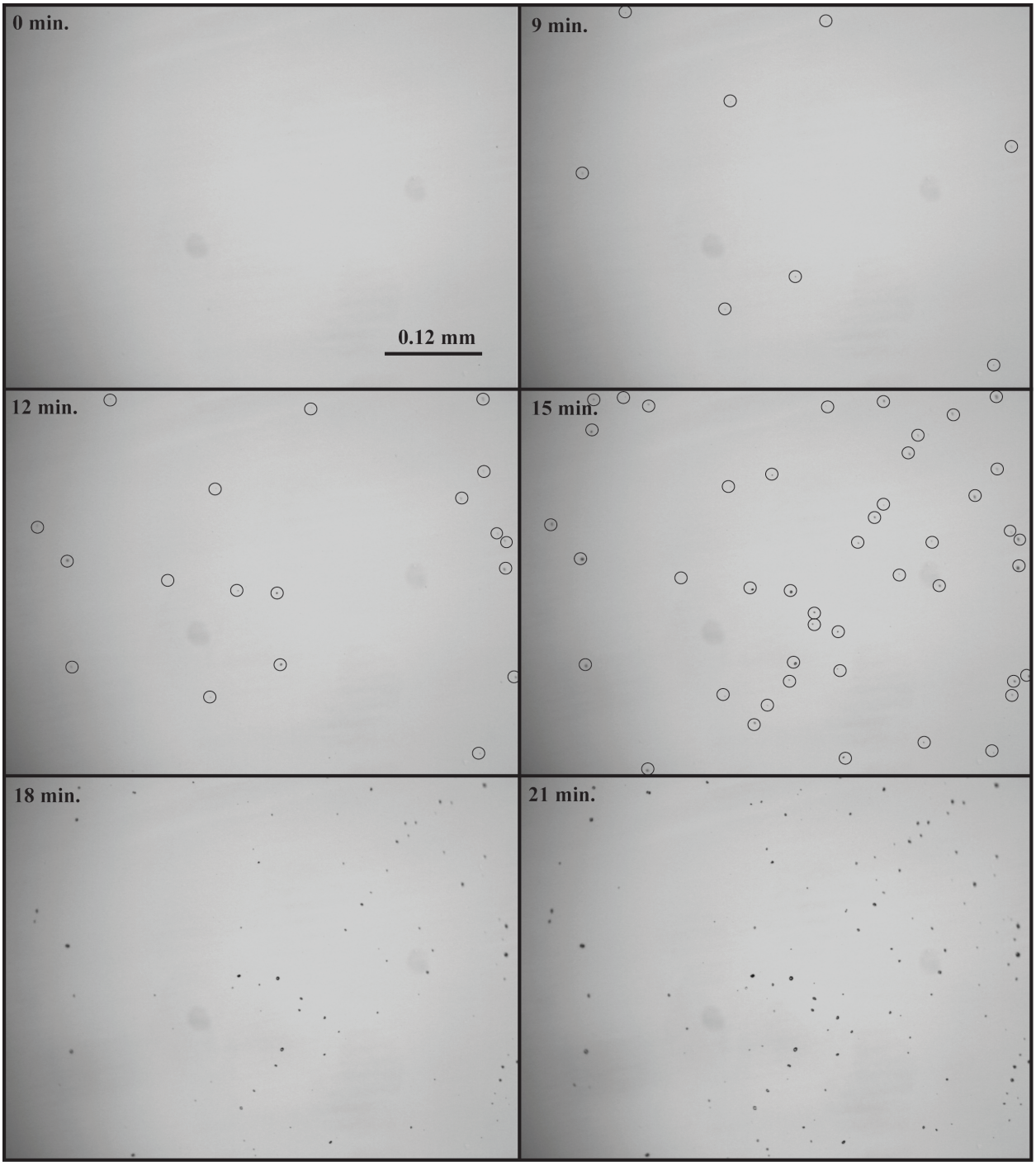


Figure C7. Optical microsopy images showing the appearance of visible crystals over time on 16C-COOH (MHA) for $\sigma = 4.78$. Crystals are indicated with drawn-in circles on the images taken at 9, 12, and 15 minutes. The size at which the crystals become visible is much smaller than the distance separating neighboring crystals. This, along with the linear relationship between crystal density and time, validates our assumption that the observed crystals each develop from a single nucleus.

C5. Python routine for analysis of DFM data

```
#!/usr/bin/env python

import struct
from time import clock
import numpy
import glob
import os
import pylab
import matplotlib
import matplotlib.pyplot as plt
from pylab import *

#-----------------------------------------------------------------------------------------------
# Input datapath and calibrated cantilever spring constant
#-----------------------------------------------------------------------------------------------

path = "/datapath/"
spring = float(raw_input("Enter calibrated cantilever spring constant in n/m: "))

#-----------------------------------------------------------------------------------------------
# Create output folders
#-----------------------------------------------------------------------------------------------

CurvePath = path + 'ForceCurves/'; os.makedirs(CurvePath)
FilePath = path + 'AnalysisFiles/'; os.makedirs(FilePath)
TextPath = path + 'Text/'

#-----------------------------------------------------------------------------------------------
# Create output folders & files
#-----------------------------------------------------------------------------------------------

n=1
AvgForces = []              #list of average rupture forces, 1 per velocity
Velocities = []             #list of velocities
AllDev = []                 #lis of stardard deviation in force for each velocity

while n<=5:
        Forces=[]           #list of rupture forces for each velocity

        velocity = 2e-7+0.00000245*(n-1); Velocities.append(velocity)
        VelPath = TextPath + 'Vel' + str(n) + '/'
        VelCurvePath = CurvePath + 'Vel' + str(n) + '/'

        os.makedirs(VelCurvePath)  #make force curve image directory for single velocity
```

```
        RuptureFile=FilePath+'RuptureForces'+str(n)+'.txt'  #make rupture force file
        FilterFile=FilePath+'FilterFiles' + str(n)+'.txt'   #make filtered rupture force file

        output=open(RuptureFile,'w'); output2=open(FilterFile,'w')
        output.write('Velocity in m/s: '); output.write(str(velocity)); output.write('\n')
        output2.write('Files numbers used for analysis at velocity = '); output2.write(str(velocity))
        output2.write(' m/s'); output2.write('\n')

        print VelPath ; print 'Velocity = ',velocity,' m/s'
        print "------------"

        FileList = glob.glob(VelPath+'*.txt*')      #get list of text files in velocity folder
        number = int((len(FileList)-2)/3)           #calc. num. of force measurements in folder
        print number, 'force curves recorded at velocity #', n
        print "------------"

        CurveNum = 1
        while CurveNum <= number:

                x=[]; y=[]; z=[]; Trunc=[]; FrontTrunc=[]           #reset lists
                print "Analyzing velocity",n,', force curve #',CurveNum

                DeflFile = VelPath + 'Defl' + str(CurveNum) + '.txt'    #make deflection and
                LVDTFile = VelPath + 'LVDT' + str(CurveNum) + '.txt'    #LVDT files

#-------------------------------------------------------------------------------------------
# Analyze LVDT (x) values, store to list for plotting
#-------------------------------------------------------------------------------------------

                xinput = open(LVDTFile,'r')             #open LVDT file
                for line in xinput.xreadlines():
                        datax = float(line.strip())     #input LVDT values
                        nanometer = datax*-1*1e9        #convert LVDT to nanometers
                        x.append(nanometer)
                xinput.close()

#-------------------------------------------------------------------------------------------
# Analyze Deflection (y) values, store to list for plotting
#-------------------------------------------------------------------------------------------

                yinput = open(DeflFile,'r')             #open deflection file
                for line in yinput.xreadlines():
                        datay = float(line.strip())     #import deflection values
                        picoN = datay * spring * 1e12   #convert deflection to pN
                         y.append(picoN); yinput.close()
#-------------------------------------------------------------------------------------------
```

```
# Plot Deflection vs. LVDT, determine deflection baseline, & calculate rupture force
#-----------------------------------------------------------------------------------------------

        if len(x) != len(y) and CurveNum not in exclude:
                print 'Bad files: Unequal # of points for Deflection & LVDT. Skipping
                measurement.'
                print '------------'

        else:
                MIN=[]; BL=[]; BLavg=[]; filter=[]; Fmax=[]; Fmin=[]
                TruncNum = int(len(y)*0.5); Trunc=y[0:TruncNum]
                #print TruncNum
                #print len(Trunc),'values in truncated force list'
                FrontTrunc = x[0:TruncNum]
                forcemin = min(Trunc)

                print "Minimum of force:", forcemin, 'pN'

                r=0                                     #find location of min. defl. value
                while r<len(Trunc):
                        Tvalue = Trunc[r]
                        if Tvalue == forcemin:
                                break
                        r=r+1


                p=n+100                                 #calculate deflection baseline
                while p<n+130:
                        BLavg.append(Trunc[p])
                        p=p+1
                BaseForce=sum(BLavg)/len(BLavg)

                s=n+300
                while s<n+700:                          #calculate noise level of baseline
                        if s+300>=len(Trunc):
                                break
                        filter.append(Trunc[s])
                        s=s+1
                threshold = max(filter)-min(filter)
                if threshold < 0:
                        threshold = threshold*(-1)
                print "Noise level =",threshold,"pN"

                q=0
                while q<len(Trunc):
                        MIN.append(forcemin); BL.append(BaseForce)
```

```
            Fmax.append(max(filter)); Fmin.append(min(filter))
            q=q+1

        DeltaForce = BaseForce-forcemin          #calculate rupture force
        print 'Rupture force =',DeltaForce,'pN'


        lineA , =plot(FrontTrunc,Trunc)          #plot force vs. LVDT
        lineA.set_markersize(1)
        lineA.set_marker('d')
        lineA.set_markerfacecolor('g')
        lineA.set_markeredgecolor('r')

        lineB , =plot(FrontTrunc,MIN)            #plot force minimum line
        #lineB.set_markersize(1)
        #lineB.set_marker('d')
        #lineB.set_markerfacecolor('g')
        #lineB.set_markeredgecolor('r')

        lineC , =plot(FrontTrunc,BL)             #plot force baseline line

        lineD , =plot(FrontTrunc,Fmax)           #plot filter maximum line
        lineE , =plot(FrontTrunc,Fmin)           #plot filter minimum line


        axis([min(FrontTrunc),max(FrontTrunc),min(Trunc)-
        50,(forcemin+DeltaForce+200)])
        axis([min(FrontTrunc),max(FrontTrunc),min(Trunc)-50,max(Trunc)])
        xlabel('LVDT, nm')
        ylabel('Force, pN')
        titleline=str((str(CurveNum)+', spring constant= '+str(spring))+' N/m')
        title(titleline)

        savepath = VelCurvePath + str(CurveNum) + '.png'     #save curve image
        pylab.savefig(savepath)

        draw()
        show()

#-------------------------------------------------------------------------------------------
# Filter force curves
#-------------------------------------------------------------------------------------------

        filter = raw_input("Keep curve? Inuput 'y' to analyze, hit any other key to
        discard: ")                              #visually check curves
        if filter=='y':
```

```
                        if DeltaForce > threshold:              #filter noisy baselines
                            Forces.append(DeltaForce)
                            output.write(str(DeltaForce)); output.write('\n')
                            output2.write(str(CurveNum)); output2.write('\n')
                    pylab.clf()

                CurveNum = CurveNum + 1

        VelForceAvg = float(sum(Forces)/len(Forces))        #calculate average rupture force
        AvgForces.append(VelForceAvg)                       # for each retraction velocity


        dev=[]
        for force in Forces:
            sdv = (force-VelForceAvg)**2; dev.append(sdv)     #calculate standard deviation
        StDev = ((1/float(len(Forces))*sum(dev))**0.5)
        AllDev.append(StDev)


        print "Average force for velocity #",n,"=",VelForceAvg,'+/-',StDev,'pN'
        print discard_count, "measurements discarded due to threshold"
        print "------------"

#-------------------------------------------------------------------------------------------------
# Plot and save a histogram of rupture forces for each retraction velocity
#-------------------------------------------------------------------------------------------------

        hist(Forces,bins=20,normed=True,align='center')
        xlabel('rupture force, pN')
        ylabel('P')
        savepath = path + 'Histogram' + str(n) + '.png'
        savefig(savepath)

        pylab.clf()

        n=n+1
        output.close(); output2.close()

#-------------------------------------------------------------------------------------------------
# Plot average rupture force vs. tip retraction velocity
#-------------------------------------------------------------------------------------------------

lineA , =plot(Velocities,AvgForces)
lineA.set_markersize(1)
lineA.set_marker('d')
lineA.set_markerfacecolor('g')
```

```
lineA.set_markeredgecolor('r')
xlabel('velocity, m/s')
ylabel('force, pN')

savepath = path + 'ForcesCurve.png'  #save png of force vs. velocity
pylab.savefig(savepath)

#draw()
#show()

#-----------------------------------------------------------------------------------------------
# Output average rupture forces and retraction velocities to a text file
#-----------------------------------------------------------------------------------------------

outputfile = path + 'AvgForces.txt'
output = open(outputfile,'w')
n=0
while n < len(Velocities):
        output.write(str(Velocities[n])); output.write('          ')
        output.write(str(AvgForces[n])); output.write('          ')
        output.write(str(AllDev[n])); output.write('\n')
        n=n+1

output.close()
pylab.clf()
```

C6. Reference

(1) Mucci, A.; Morse, J.W. Geochim. Cosmochim. Acta **1984**, *48*, 815-822.