J|A|C|S ARTICLES

Published on Web 08/16/2010

De Novo Design of Peptide−Calcite Biomineralization Systems

David L. Masica,[†] Sarah B. Schrier,[‡] Elizabeth A. Specht,[‡] and Jeffrey J. Gray*,[†,‡]

Program in Molecular Biophysics, and Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland 21218

Received January 14, 2010; E-mail: jgray@jhu.edu

Abstract: Many organisms produce complex, hierarchically structured, inorganic materials via protein-influenced crystal growth—a process known as biomineralization. Understanding this process would shed light on hard-tissue formation and guide efforts to develop biomaterials. We created and tested a computational method to design protein-biomineralization systems. The algorithm folds a protein from a fully extended structure and simultaneously optimizes the fold, orientation, and sequence of the protein adsorbed to a crystal surface. We used the algorithm to design peptides (16 residues) to modify calcite ($CaCO_3$) crystallization. We chemically synthesized six peptides that were predicted to bind different states of a calcite growth plane. All six peptides dramatically affected calcite crystal growth (as observed by scanning electron microscopy), and the effects were dependent on the targeted state of the {001} growth plane. Additionally, we synthesized and assayed scrambled variants of all six designed peptides to distinguish cases where sequence composition determines the interactions versus cases where sequence order (and presumably structure) plays a role. Scrambled variants of negatively charged peptides also had dramatic effects on calcite crystallization; in contrast, scrambled variants of positively charged peptides had a variable effect on crystallization, ranging from dramatic to mild. Special emphasis is often placed on acidic protein residues in calcified tissue mineralization; the work presented here suggests an important role for basic residues as well. In particular, this work implicates a potential role for basic residues in sequence-order specificity for peptide−mineral interactions.

Introduction

Biomineralization is a process of crystal nucleation and growth controlled by bioorganic molecules such as proteins.[1−3] Many single-celled organisms mineralize structural and functional materials such as silica, magnetite, and calcite.[1−3] In higher organisms, biomineralization underlies hard-tissue formation.[1−3] These biogenic materials typically confer several advantages relative to synthetic and geologic counterparts. For instance, bioorganic molecules can lower the activation energy of crystal formation,[1,4] circumventing the requirement for extreme temperature, pressure, or pH. Also, biogenic materials often have superior mechanical properties relative to nonbiogenic variants.[5] These advantages make biomineralization a potential alternative to contemporary methods in materials production.[6] Currently, our understanding of protein structure and sequence determinants in biomineralization is not sufficiently detailed to fully realize this potential.

The biomineralization of calcified hard tissues appears to rely on acidic protein residues[7−10] (e.g., aspartate, glutamate, and phosphorylated serine) more than basic residues (e.g., arginine and lysine). This apparent disparity may be surprising considering minerals are composed of both anions and cations. High acidic amino acid content imparts an intrinsic disorder in the predicted binding domains of some mineral-associated proteins.[11−13] This high uniform charge and resulting disorder are at least part of the reason that some biomineralization proteins exhibit binding promiscuity.[14] For instance, osteopontin binds hydroxyapatite,[15] calcite,[16] and calcium oxalate crystals;[17] lithostathine is found associated with different mineral types

[†] Program in Molecular Biophysics.
[‡] Department of Chemical and Biomolecular Engineering.

(1) Sigel, A.; Sigel, H.; Sigel, R. K. O. *Biomineralization. From Nature to Application*; Wiley: New York, 2008; Vol. 4.
(2) Mann, S. *Biomineralization. Principles and Concepts in Bioinorganic Materials Chemistry*; Oxford University Press: Oxford, 2001.
(3) Dove, P. M.; De Yoreo, J. J.; Weiner, S. *Biomineralization*; Mineralogical Society of America, Geochemical Society: Washington, DC, 2003; Vol. 54.
(4) Mann, S. *Nature* **1988**, *332*, 119–124.
(5) Jiang, H.; Liu, X. Y.; Lim, C. T.; Hsu, C. Y. *Appl. Phys. Lett.* **2005**, *86*, 163901–163904.
(6) Vrieling, E. G.; Beelen, T. P. M.; van Santen, R. A.; Gieskes, W. W. C. *J. Biotechnol.* **1999**, *70*, 39–51.
(7) Addadi, L.; Weiner, S. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 4110–4114.
(8) Fu, G.; Valiyaveettil, S.; Wopenka, B.; Morse, D. E. *Biomacromolecules* **2005**, *6*, 1289–1298.
(9) Gotliv, B. A.; Kessler, N.; Sumerel, J. L.; Morse, D. E.; Tuross, N.; Addadi, L.; Weiner, S. *ChemBioChem* **2005**, *6*, 304–314.
(10) Tsukamoto, D.; Sarashina, I.; Endo, K. *Biochem. Biophys. Res. Commun.* **2004**, *320*, 1175–1180.
(11) Kapon, T. M.; Rymarczyk, G.; Nocula-ugowska, M.; Jakob, M.; Kochman, M.; Lisowski, M.; Szewczuk, Z.; Ozyhar, A. *Biomacromolecules* **2008**, *9*, 2118–2125.
(12) Amos, F. F.; Evans, J. S. *Biochemistry* **2009**, *48*, 1332–1339.
(13) Buchko, G. W.; Bekhazi, J.; Cort, J. R.; Valentine, N. B.; Snead, M. L.; Shaw, W. J. *Biomol. NMR Assign.* **2008**, *2*, 89–91.
(14) Delak, K.; Collino, S.; Evans, J. S. *Biochemistry* **2009**, *48*, 499–503.
(15) Addison, W. N.; Masica, D. L.; Gray, J. J.; McKee, M. D. *J. Bone Miner. Res.* **2009**, *0*, 1–44.

 10.1021/ja1001086 © **2010 American Chemical Society**

during litholysis;[18,19] and poly-Asp can modify the crystallization of multiple mineral types and phases.[17,20,21] In vivo, this promiscuity can be advantageous[14] and can be controlled by protein expression, matrix mediation, or boundary organization.[2] Because contemporary in vitro methods in molecular biology cannot achieve the spatiotemporal control utilized by living cells, the tailoring of biomaterials with nanoscale precision in vitro may require the ability to design protein−mineral interactions that have more specificity.

Several groups have developed successful design strategies for biomineralization problems.[22−27] Those strategies relied on modifying naturally evolved proteins or employing directed evolution techniques. Also, one group has predicted novel material-binding peptide sequences in silico.[28] In that work, bioinformatics techniques were used to analyze peptide sequences previously isolated from directed evolution experiments. Ideally, de novo design of biomineralization systems is possible. In the computational de novo design of proteins, a protein sequence is optimized to achieve a desired phenotype, such as binding a mineral surface. De novo design methods do not rely on evolved sequences and may afford added specificity in designing protein−mineral interactions.

Development of de novo protein design algorithms has facilitated the engineering of proteins with novel structural and functional properties.[29−31] In particular, development of the structure-prediction algorithm Rosetta led to unprecedented achievements in many protein design problems (the RosettaDesign algorithm). These achievements include the design of a novel protein fold,[32] β-sandwich proteins,[33] enzyme catalytic sites,[34−36] protein−protein interaction specificity,[37] protein−DNA specificity,[34] protein−peptide specificity,[38] and protein folding pathways.[39] Recently, we developed Rosetta to predict the structure of proteins adsorbed to biominerals by incorporating the energetics of protein−solid surface interactions (RosettaSurface).[40,41]

For the present investigation, we sought to further develop principles for the rational design of biomineralization systems, giving special consideration to peptide charge, composition, and sequence. To accomplish this goal we developed RosettaSurface to simultaneously optimize the fold, orientation, and sequence of a protein adsorbed to a crystal surface (RosettaSurface.Design). We used RosettaSurface.Design to design de novo 10^5 peptide sequences (16 residues) to bind differentially terminated states of a calcite ($CaCO_3$) growth plane. From the predicted libraries, we chemically synthesized six sequences and tested their effect on calcite crystallization. In addition, we synthesized scrambled variants of the six designed peptides and compared the effects of designed and scrambled variant sequences on calcite crystallization. This is, to our knowledge, the first structure-prediction-based algorithm capable of protein design on a solid surface.

Materials and Methods

Algorithm Development and Implementation. The algorithm developed here, RosettaSurface.Design (see flowchart, Figure 1), is based on RosettaSurface.[41] The creation of each designed structure (decoy) begins with the rapid collapse of a fully extended protein molecule, in implicit solvent, using a fragment assembly protocol. After fragment assembly, the backbone is optimized with six "shear" and ∼100 "small" moves[44] coupled with quasi-Newtonian[42] and line minimization, respectively. At all stages thus far the structural model is composed of an all-atom protein backbone and a single pseudoatom to represent each side chain in the starting sequence.

Next, the sequence is optimized by building all-atom side chains (high-resolution decoy) from a backbone-dependent rotamer library.[43] During sequence optimization, sequence space is searched using a simulated-annealing Metropolis Monte Carlo procedure that replaces a side chain at a random position with a side chain from the backbone-dependent rotamer library (∼10^6 substitutions per Monte Carlo iteration).[32] The protein then undergoes $n \leq 5$ cycles of backbone refinement (see Figure 1, right) in implicit solvent (where n is a uniform, random integer between 1 and 5). Each cycle of refinement includes "outer-"and "inner-loop optimization". During outer-loop optimization, a sequence of perturbing small, shear, and crank moves[44] are applied to the protein backbone, each followed by quasi-Newtonian minimization.[42] Next, during inner-loop optimization, the same perturbing moves are applied except each is followed by line minimization along the initial gradient. For each cycle, the outer loop is implemented five times; for each outer loop, the inner loop is implemented five times.

After n refinement cycles an adsorbed-state complex is formed, by introducing the surface in a random orientation and bringing the protein and surface into contact. The adsorbed-state protein

(16) Chien, Y. C.; Hincke, M. T.; Vali, H.; McKee, M. D. *J. Struct. Biol.* **2008**, *163*, 84–99.
(17) Chien, Y. C.; Masica, D. L.; Gray, J. J.; Nguyen, S.; Vali, H.; McKee, M. D. *J. Biol. Chem.* **2009**, *284*, 23491–23501.
(18) Ryall, R. L. *Pediatr. Nephrol.* **1996**, *10*, 656–666.
(19) Bernard, J. P.; Barthet, M.; Gharib, B.; Michel, R.; Lilova, A.; Sahel, J.; Dagorn, J. C.; De Reggi, M. *Br. Med. J.* **1995**, *36*, 630–636.
(20) Politi, Y.; Mahamid, J.; Goldberg, H.; Weiner, S.; Addadi, L. *CrystEngComm* **2007**, *9*, 1171–1177.
(21) Kasugai, S.; Fujisawa, R.; Waki, Y.; Miyamoto, K.; Ohya, K. *J. Bone Miner. Res.* **2000**, *15*, 936–943.
(22) Ajikumar, P. K.; Lakshminarayanan, R.; Ong, B. T.; Valiyaveettil, S.; Kini, R. M. *Biomacromolecules* **2003**, *4*, 1321–1326.
(23) Brown, S. *Nat. Biotechnol.* **1997**, *15*, 269–272.
(24) Capriotti, L. A.; Beebe, T. P. J.; Schneider, J. P. *J. Am. Chem. Soc.* **2007**, *129*, 5281–5287.
(25) DeOliveira, D. B.; Laursen, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 10627–10631.
(26) Elhadj, S.; De Yoreo, J. J.; Hoyer, J. R.; Dove, P. M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19237–19242.
(27) Sarikaya, M.; Tamerler, C.; Alex, K. Y. J.; Schulten, K.; Baneyx, F. *Nat. Mater.* **2003**, *2*, 577–585.
(28) Oren, E. E.; Tamerler, C.; Sahin, D.; Hnilova, M.; Seker, U. O. S.; Sarikaya, M.; Samudrala, R. *Bioinformatics* **2007**, *23*, 2816–2822.
(29) Huang, P. S.; Love, J. J.; Mayo, S. L. *Protein Sci.* **2007**, *16*, 2770–2774.
(30) Bender, G. M.; Lehmann, A.; Zou, H.; Cheng, H.; Fry, H. C.; Engel, D.; Therien, M. J.; Blasie, J. K.; Roder, H.; Saven, J. G.; DeGrado, W. F. *J. Am. Chem. Soc.* **2007**, *129*, 10732–10740.
(31) Lippow, S. M.; Tidor, B. *Curr. Opin. Biotechnol.* **2007**, *18*, 305–311.
(32) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. *Science* **2003**, *302*, 1364–1368.
(33) Hu, X.; Wang, H.; Ke, H.; Kuhlman, B. *Structure* **2008**, *16*, 1799–1805.
(34) Ashworth, J.; Havranek, J. J.; Duarte, C. M.; Sussman, D.; Monnat, R. J.; Stoddard, B. L.; Baker, D. *Nature* **2006**, *441*, 656–659.
(35) Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Rothlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F., III; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D. *Science* **2008**, *319*, 1387–1391.
(36) Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albek, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. *Nature* **2008**, *453*, 190–195.
(37) Kortemme, T.; Joachimiak, L. A.; Bullock, A. N.; Schuler, A. D.; Stoddard, B. L.; Baker, D. *Nat. Struct. Mol. Biol.* **2004**, *11*, 371–379.
(38) Sood, V. D.; Baker, D. *J. Mol. Biol.* **2006**, *357*, 917–927.
(39) Nauli, S.; Kuhlman, B.; Baker, D. *Nat. Struct. Biol.* **2001**, *8*, 602–605.
(40) Makrodimitris, K.; Masica, D. L.; Kim, E. T.; Gray, J. J. *J. Am. Chem. Soc.* **2007**, *129*, 13713–13722.
(41) Masica, D. L.; Gray, J. J. *Biophys. J.* **2009**, *96*, 3082–3091.
(42) Davidon, S.; Fletcher, R.; Powell, M. *The Comput. J.* **1963**, *6*, 163.
(43) Dunbrack, R. L., Jr.; Cohen, F. E. *Protein Sci.* **1997**, *6*, 1661–1681.
(44) Rohl, C. A.; Strauss, C. E.; Misura, K. M.; Baker, D. *Methods Enzymol.* **2004**, *383*, 66–93.

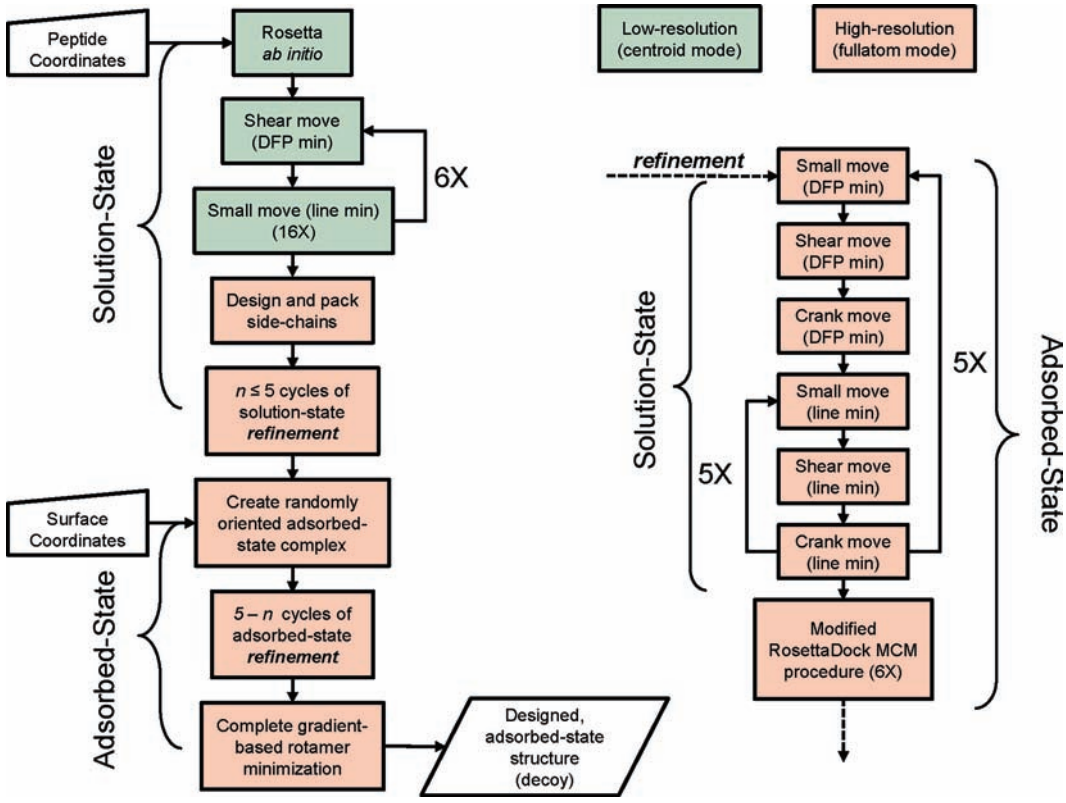

Figure 1. Flowchart for the RosettaSurface. Design algorithm. DFP min = Davidon−Fletcher−Powell minimization;[42] line min = line minimization. small move = perturbations of randomly selected (ϕ,ψ) pairs. shear move = perturbation of randomly selected ϕ angle with compensatory perturbation of equal and opposite magnitude applied to previous ψ angle. crank move = globally nonperturbing fragment insertion followed (ϕ,ψ) perturbation at residues adjacent to, and two residues not adjacent to, the insertion window.

undergoes $5 - n$ cycles of refinement at the interface. Refinement is the same as in the previous paragraph, except each cycle ends with a modified version of RosettaDock's[45] high-resolution procedure. During the modified procedure, small rigid-body perturbations, side-chain repacking, and gradient-based minimization in rigid-body space is repeated six times; side-chain repacking is combinatorial, rather than individual, every third time. Hence, the protein undergoes simultaneous backbone, sequence, and rigid-body optimization on the surface.

The first, third, and fifth refinement cycles include the sequence optimization procedure, regardless of whether or not the protein is adsorbed to the solid surface. After adsorbed-state refinement is complete, RosettaSurface.Design implements one round of coupled rotamer packing and minimization[46] at all side-chain positions and outputs the adsorbed-state coordinates. The generation of each candidate designed structure takes approximately 5 min on a standard desktop processor.

The all-atom energy function is the same as in Masica et al.[41] and includes a nonweighted linear combination of distance-dependent-dielectric electrostatic,[47] Lennard−Jones 6−12 van der Waals,[45] angle-dependent hydrogen bond,[48] and implicit solvent[49] interactions.

The force field parameters for calcite were chosen as follows: The Lennard−Jones well depth, ε, and internuclear separation, σ, were taken from Stockelmann et al.[50] For the hydrogen-bond function, the carbonate oxygen atoms (sp^2 hybridization) are proton acceptors. The Gaussian solvent-exclusion model for solvation free energy was parametrized from calculated water density profiles.[51] Atomic charges were assigned from quantum calculations.[52]

Protein design simulations can include amino acid reference energies to account for the unfolded state and the relative propensities of individual amino acids.[32] During sequence optimization, we used energetic penalties of 2.0, 2.5, 0.5, and 1.0 kcal/mol for the incorporation of Lys, Arg, Asp, and Glu residues, respectively. The penalties prevented the overdesign of charged residues at the charged calcite surface and helped ensure sequence diversity. To optimize the penalties, we performed test design simulations (~100 designed sequences/simulation), varying the penalties from 0 to 5 kcal/mol in increments of 0.5 kcal/mol for each residue. Of the penalties tested, those used in this study provided the best balance between binding energy and sequence diversity.

Starting Materials for Simulation. An extended molecular structure of a 16 residue peptide was constructed using PyMol.[53] In each simulation the initial peptide sequence was 15 alanines (residues 1−15) followed by a tyrosine (residue 16; included for concentration determination in the experimental constructs). The polyalanine starting sequence was chosen because alanine is charge neutral and because polyalanine is able to form (helical) secondary structure, even in the context of small peptides. Residues 1−15 were designed during decoy generation, while the tyrosine at position 16 was fixed.

Sequences for scrambled variants were generated using a random number generator, using as input the sequences of the corresponding

(45) Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D. *J. Mol. Biol.* **2003**, *331*, 281–299.
(46) Wang, C.; Schueler-Furman, O.; Baker, D. *Protein Sci.* **2005**, *14*, 1328–1339.
(47) Warshel, A.; Russell, S. T.; Churg, A. K. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 4785.
(48) Kortemme, T.; Morozov, A. V.; Baker, D. *J. Mol. Biol.* **2003**, *326*, 1239–1259.
(49) Themis, Lazaridis; Karplus, M. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 133–152.
(50) Stockelmann, E.; Hentschke, R. *Langmuir* **1999**, *15*, 5141–5149.
(51) Kerisit, S.; Parker, S. C. *Chem. Commun.* **2004**, *2004*, 52–53.
(52) Catti, M. *Phys. Chem. Miner.* **2001**, *28*, 729–736.
(53) DeLano, W. L. Delano Scientific: Palo Alto, CA, U.S.A., 2002.

designed peptide and conserving the tyrosine residue at position 16 for concentration determination in the experimental construct. This was achieved by removing amino acids at random positions from the designed sequence, without replacement, and placing them sequentially to construct the corresponding scrambled variant. Therefore, designed and scrambled sequences have identical sequence composition but different sequence order.

Coordinates for calcite crystals were generated using CrystalMaker software.[54] Appropriate cuts were made to expose {001} surfaces with either a net negative or net positive charge. Crystals were cut to a thickness of greater than 8 Å to extend beyond the pairwise cutoff in the RosettaSurface energy function.

Peptide Synthesis, Purification, and Verification. All peptides were synthesized using standard fmoc (9-fluorenylmethoxycarbonyl) chemistry on a solid Anaspec fmoc-tyr-wang resin using a Protein Technologies Symphony Quartet automated peptide synthesizer.[55] The synthesis program used double coupling and included a capping step after each amino acid coupling to increase the yield of the desired product. All amino acid reagent concentrations were 200 mM. The peptides were cleaved from the dried resin by exposure to 4 mL of trifluoroacetic acid/anisole/water (95:2.5:2.5) for 2 h at room temperature with constant agitation. After precipitation of the cleavage reaction products with ice-cold diethyl ether and washing with diethyl ether two subsequent times, the crude product was reversed-phase HPLC purified. The samples were injected onto an Agilent ZORBAX C-18 column at 40 °C and eluted using a 0−100% gradient of acetonitrile and water, each with 0.1% trifluoroacetic acid. The UV detector measured absorbance at 254 nm and 280 nm, to detect backbone signal and tryptophan residues, respectively. The collected peaks were then analyzed using a Finnigan LCQ ion trap mass spectrometer with electrospray ionization to determine the peak with the correct product mass.

Crystal Growth Experiments. Calcite crystals were grown from saturated $Ca(HCO_3)_2$ solution. Solution was prepared by bubbling CO_2 gas over 6.7 mM $CaCO_3$ solution for 1 h.[56] Seven microliters of solution was placed on a glass coverslip and silicanized with hexamethyldisilazane (Alfa Aesar), and then the coverslip was inverted over a 24-well plate. Coverslips were sealed with vacuum grease, allowing each drop to grow in its own microenvironment. Crystals were allowed to nucleate at room temperature for 20 min, at which time the seal was broken and either 3 μL of water or 3 μL of peptide solution was added.[25] Final peptide concentration was 0.45 mg/mL. Crystals were grown overnight. Controls (native crystals) were confirmed by light microscopy on an Axiovert 200 M to ensure consistency of experiments. All crystallization experiments were done in triplicate a minimum of three independent times (nine times total per peptide).

Scanning Electron Microscopy. Crystals were air-dried, sputter-coated with platinum, and viewed with a FEI Quanta ESEM 200, operating at 3 kV. All SEMs presented here were chosen to represent the typical biomineralization activity observed for each corresponding peptide.

Results

We sought to alter the growth of calcite crystallization via peptide binding at a high-energy, calcite growth plane. The {001} and {012} calcite surfaces arise from successive deposition of alternating negatively and positively charged ions. Both the {001} and {012} surfaces have been implicated as relevant growth planes during calcite crystallization;[57−59] for this work we restricted computational design to peptides targeting the {001} surface. The cartoon in Figure 2A depicts an almost

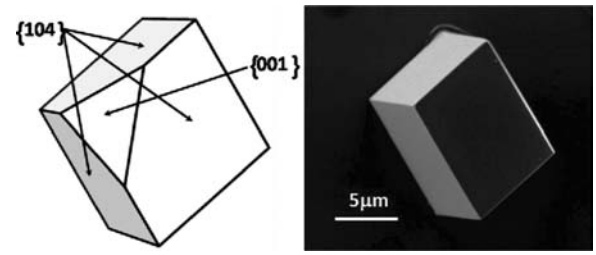

Figure 2. Calcite. (A) Sketch of the targeted {001} growth plane and the stable {104} planes of calcite. (B) SEM of calcite native crystal (grown in the absence of peptides) expressing six stable {104} faces.

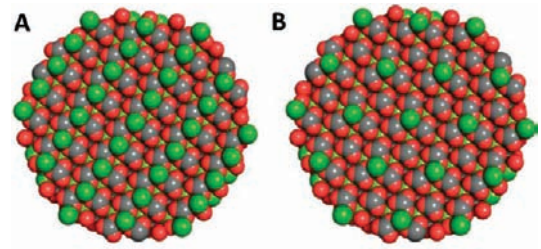

Figure 3. Net positively charged (A) and negatively charged (B) calcite surface terminations used in design simulations. Calcium atoms are green, carbon atoms are gray, and oxygen atoms are red.

mature calcite crystal, expressing six stable {104} planes and one high-energy {001} plane. Figure 2B is a scanning electron microscope (SEM) image of a calcite crystal showing typical rhombohedral morphology.

During experimental crystallization, the surface structure can deviate from that of the bulk, including steps, adsorbed waters, and deviation of atoms from their lattice positions (amorphousity). However, as ions deposit onto a growth plane, significant water must be displaced and a significant number of ions must acquire their lattice positions if the crystalline form of the solid is to develop. For this work, we designed peptides to directly bind surface atoms (not via adsorbed waters), approximately at their lattice positions. We targeted two differentially terminated calcite {001} planes. The terminations were chosen to meet two criteria: (1) have a net charge of either negative or positive and (2) be heterocharged and hence more stable than a homocharged termination. The more stable surface conformers will be more populated during experimental biomineralization. Furthermore, heterocharged surface chemistries select for more heteropolymeric peptide sequences during RosettaSurface.Design implementation. Heteropolymeric sequences scramble in more unique ways, increasing the likelihood that designs and variants will bind differently, i.e., exhibit sequence-order specificity. The two mixed-charge surfaces are shown in Figure 3.

For this study, the design algorithm was supplied a reduced amino acid library containing only Ala, Pro, Gly, Glu, Asp, Arg, and Lys. The reduced amino acid library includes residues that are charged at neutral pH (Asp, Glu, Arg, and Lys) and those that typically play a structural role (Ala, Pro, Gly). While naturally evolved biomineralization proteins comprise all 20 naturally occurring amino acids (plus additional, post-transla-

(54) Palmer, D. CrystalMaker Software: Oxford, U.K., 2001.
(55) Atherton, E.; Sheppard, R. C. *Solid phase peptide synthesis: a practical approach*; IRL: Oxford, 1989.
(56) Mann, S.; Didymus, J. M.; Sanderson, N. P.; Heywood, B. R.; Samper, E. J. A. *J. Chem. Soc., Faraday Trans.* **1990**, *86*, 1873–1880.
(57) Weiner, S.; Addadi, L. *J. Mater. Chem.* **1997**, *7*, 689–702.
(58) Addadi, L. I. A.; Politi, Y.; Nudelman, F.; Weiner, S. *Biomineralization Design Strategies and Mechanisms of Mineral Formation: Operating at the Edge of Instability*; Springer: Netherlands, 2008.
(59) Berman, A.; Ahn, D. J.; Lio, A.; Salmeron, M.; Reichert, A.; Charych, D. *Science* **1995**, *269*, 515–518.

Table 1. Individual Contributions to the Total Interaction Energy (kcal/mol) for the 100 Sequences with Lowest Energy, Designed at the Net Positively (Figure 3A) and Net Negatively Charged (Figure 3B) Calcite {001} Surfaces[a]

	electro	H-bond	attractive	repulsive	solvation
net positively charged surface	-46.6 ± 8.6	-0.2 ± 0.5	-10.2 ± 2.7	0.5 ± 0.6	19.2 ± 4.5
net negatively charged surface	-32.1 ± 6.8	-3.4 ± 1.4	-8.8 ± 2.5	0.6 ± 0.8	20.4 ± 5.2

[a] The distant-dependent dielectric electrostatic (electro), orientation-dependent hydrogen-bond (H-bond), van der Waals attractive (attractive), van der Waals repulsive (repulsive), and Gaussian solvent-exclusion (solvation) energy terms are described in ref 41.

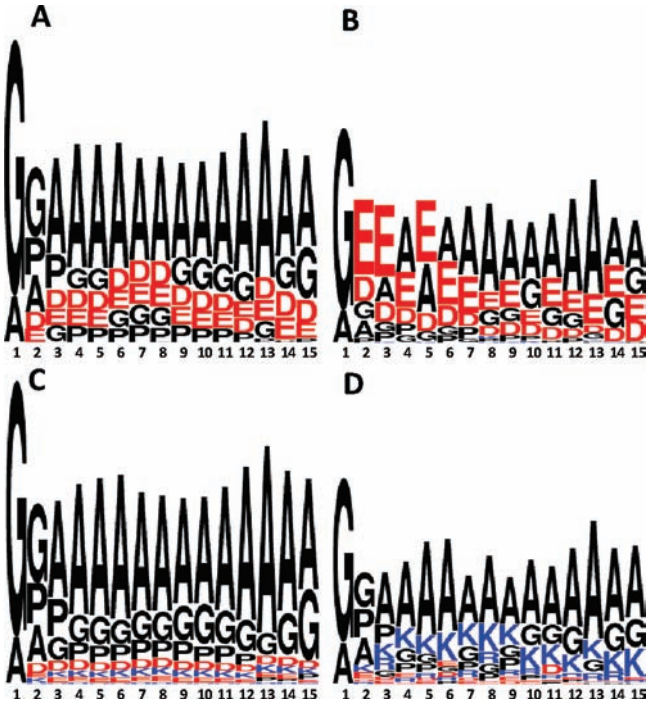

Figure 4. Sequence Logos(71) showing the relative sequence entropy for all 10^5 sequences (A) and the 100 sequences with lowest energy (B) designed at the net positively charged calcite {001} surface (Figure 3A) and all 10^5 sequences (C) and the 100 sequences with lowest energy (D) designed at the net negatively charged calcite {001} surface (Figure 3B).

tionally modified amino acid types), charged amino acids are enriched in the binding domains;[7−10] here we are specifically interested in designing a mineral binding domain. Fewer amino acids makes searching the provided sequence space more computationally tractable and reduces the occurrence of dipeptides with low-coupling efficiency for chemical synthesis. While this approach excludes potential binding sequences, the identification of every potential binding sequence is outside the scope of this study.

We used RosettaSurface.Design to generate 10^5 sequences on both the net positively and net negatively charged surfaces (see Materials and Methods). Because each of the 10^5 optimized sequences was the result of sampling $\sim 10^6$ sequences, $\sim$4.2% of the total sequence space (7^{15} sequences) was sampled. At both of these charged surfaces, the dominant energy term driving sequence optimization is the electrostatic term (Table 1). Hydrogen bonding plays a more significant role at the net negatively charged surface because the calcite surface carbonate oxygens can hydrogen bond and because the negatively charged surface selects for positively charged, hydrogen-bond-donating side chains (Arg and Lys). Solvation accounts for $\sim$98% of the energy terms that disfavor peptide−calcite interaction.

Figure 4 shows analysis of all 10^5 sequences and the 100 sequences with the lowest energy, at each surface. For all four cases, at most positions, Ala is the most populated. Ala enrichment arises because the starting sequence is poly-Ala, and

Table 2. Synthesized Peptide Sequences[a]

Number	Design Sequence	Scrambled Variant	Net Charge
1	GEAEGEEAAAGEGGAY	EGGAAGAEAEEGAEGY	-5
2	GEEAADAAGAEEAGAY	AAAAGAEGEGAAEEDY	-5
3	AKAPKDGRAKEGGAAY	AAGPDAKKARGGEAKY	+2
4	GAAAAARKAEKGAKAY	AARGKGAAAEKAKAAY	+3
5	APPRAKAAKAAAAGKY	AAAKAPAAGKPKARAY	+4
6	GPPPPAKAAKKAAKKY	AKAPKKAPPGAAKKPY	+5

[a] Number defines the naming convention throughout, e.g., design 1, variant 1, etc. Designed and variant sequences are described using one-letter amino acid codes. Net charge is the difference in acidic and basic amino acid content.

therefore the initial conformation favors the conservation of Ala residues. Gly is enriched at position 1 in all four cases. Gly, Pro, and Ala are enriched at position 2, except in the case of the lowest-energy sequences at the net positively charged calcite {001} surface.

Comparing all 10^5 sequences with the 100 sequences of lowest energy at the net positively charged calcite {001} surface (Figure 4A and 4B, respectively), negatively charged residues are enriched in the low-energy sequences. Also, at this surface, aspartate is enriched compared with glutamate in the 10^5 sequences; this results from the energetic bias used for sequence optimization (see Materials and Methods). However, glutamate is the preferred acidic amino acid in the low-energy sequences, showing that the most optimized interactions select for glutamate over aspartate, even in the presence of a higher glutamate reference energy. Positively charged residues are enriched in the low-energy 100 sequences at the negatively charged calcite {001} surface (Figure 4C and 4D). While sequences are diverse in all cases, the electrostatic energy clearly favors incorporation of charged side chains in the 100 sequences with lowest energy, relative to all 10^5 sequences, at each surface.

Ideally, the binding affinity (adsorbed- minus solution-state free energy) could be calculated and used as a discriminating parameter. Predicting the binding affinity would require predicting the solution-state energy for each of the 10^5 candidate sequences in each run and is computationally prohibitive. We chose sequences of both uniform (negative or positive) and mixed charge to investigate the effects of peptide net charge and relative charge composition on calcite biomineralization.

From the 10 sequences with lowest predicted adsorbed-state energy, three sequences were close-to-redundant or had charged residues not interacting with the {001} calcite surface (suggesting potential nonspecific binding in the presence of alternative calcite faces during crystal growth assays) and one had low

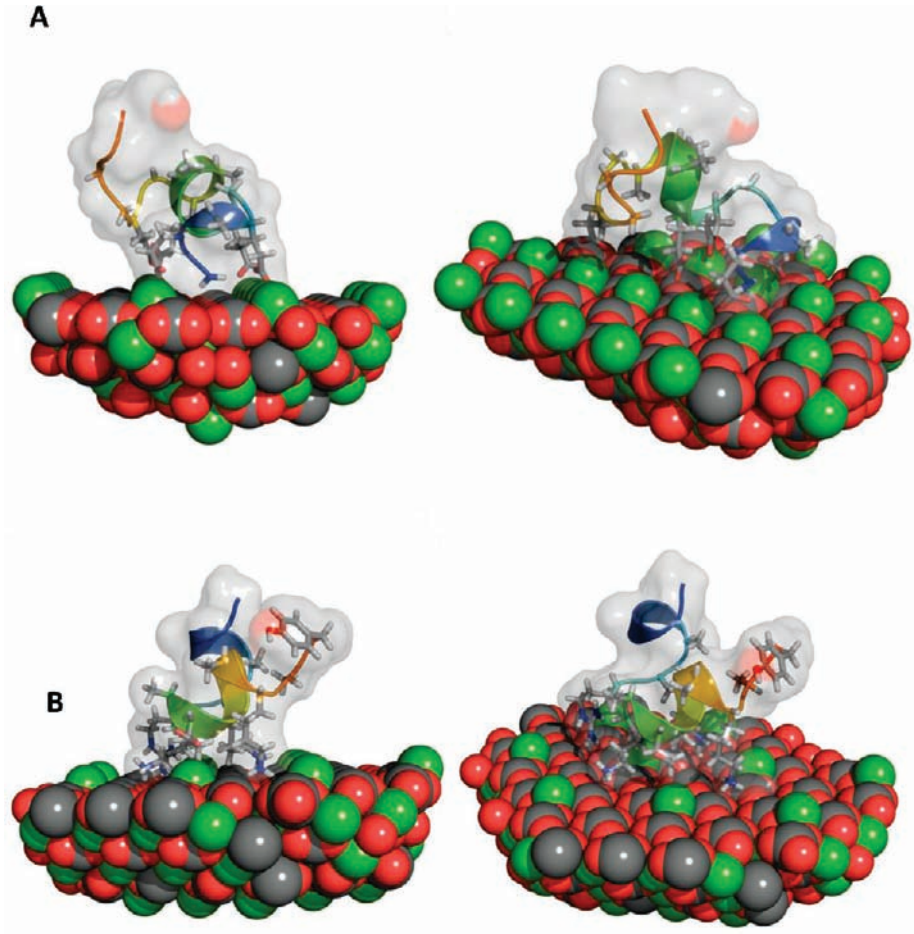

Figure 5. Structures predicted by RosettaSurface.Design for (A) design 1 bound to a net positively charged {001} calcite surface (Figure 3A) and (B) design 4 bound to a net negatively charged {001} calcite surface (Figure 3B).

yield owing to coupling difficulties during chemical synthesis; the remaining six sequences (Table 2) were synthesized and assayed.

Figure 5A and 5B show the predicted structures of two sequences (designs 1 and 4, respectively) evolved by RosettaSurface.Design to bind the calcite {001} growth plane. In Figure 5A, design 1 binds a positively charged calcite surface by aligning all of its aspartic acid side chains with adjacent rows of exposed calcium atoms. In Figure 5B, design 4 binds a negatively charged calcite surface by filling adjacent rows of calcium vacancies with its basic residues. All designed peptides chosen for subsequent chemical synthesis recognize calcite in a similar manner.

We grew calcite crystals in the presence of the synthesized, designed peptides and observed the resulting crystal morphology with a SEM. All six designed peptides affected calcite crystallization significantly (Figure 6A−F). The modification in crystallization resulting from inoculation with negatively charged design 1 (Figure 6A) produced macrostepping and kinking. The only other net negatively charged design, design 2 (Figure 6B), produced many Ca_2CO_3 crystals that were spherical. An average reduction in size and an increased occurrence in crystal twinning were observed with designs 1 and 2.

All of our net positively charged designed peptides (designs 3−6) modified crystallization similar to one another (Figure 6C−F), with design 5 (Figure 6E) being a possible outlier. The distinct feature in these crystals is the heavily stepped corners and edges. These crystals were typically smaller than unmodified crystals and also displayed an increased occurrence in twinning. The biggest difference between crystals grown in the presence of the negatively charged designs and those grown in the presence of positively charged designs is the location of the steps. Crystals grown in the presence of negatively charged designs have considerable stepping on all faces; the resulting crystals lose most of their rhombohedral morphology. In contrast, crystals grown in the presence of positively charged designs have steps located almost exclusively at crystal edges and corners, leaving much of the rhombohedral morphology intact.

Next we synthesized variants of the designed peptides, whose sequences were each randomly scrambled, and grew calcite crystals in the presence of the scrambled variants (Table 2). Scrambled variants provide a test of sequence-order specificity. Because calcite crystal growth modification by charged polymers can be a general (nonspecific) phenomenon, modification via scrambled variants can be used to assess the efficacy of the designed sequences relative to randomly generated sequences. Figure 6G shows crystals grown in the presence of variant 1. Crystals grown while inoculated with design 1 (Figure 6A) and variant 1 (Figure 6G) share a similar morphology, which includes the macrostepping and kinking mentioned above. The other net negatively charged variant, variant 2, produced Ca_2CO_3 crystals bearing almost no gross resemblance to rhombohedral calcite crystals (Figure 6H). While both design 2 and variant 2 affected Ca_2CO_3 crystallization (Figure 6B and 6H, respectively), the effects were generally different.

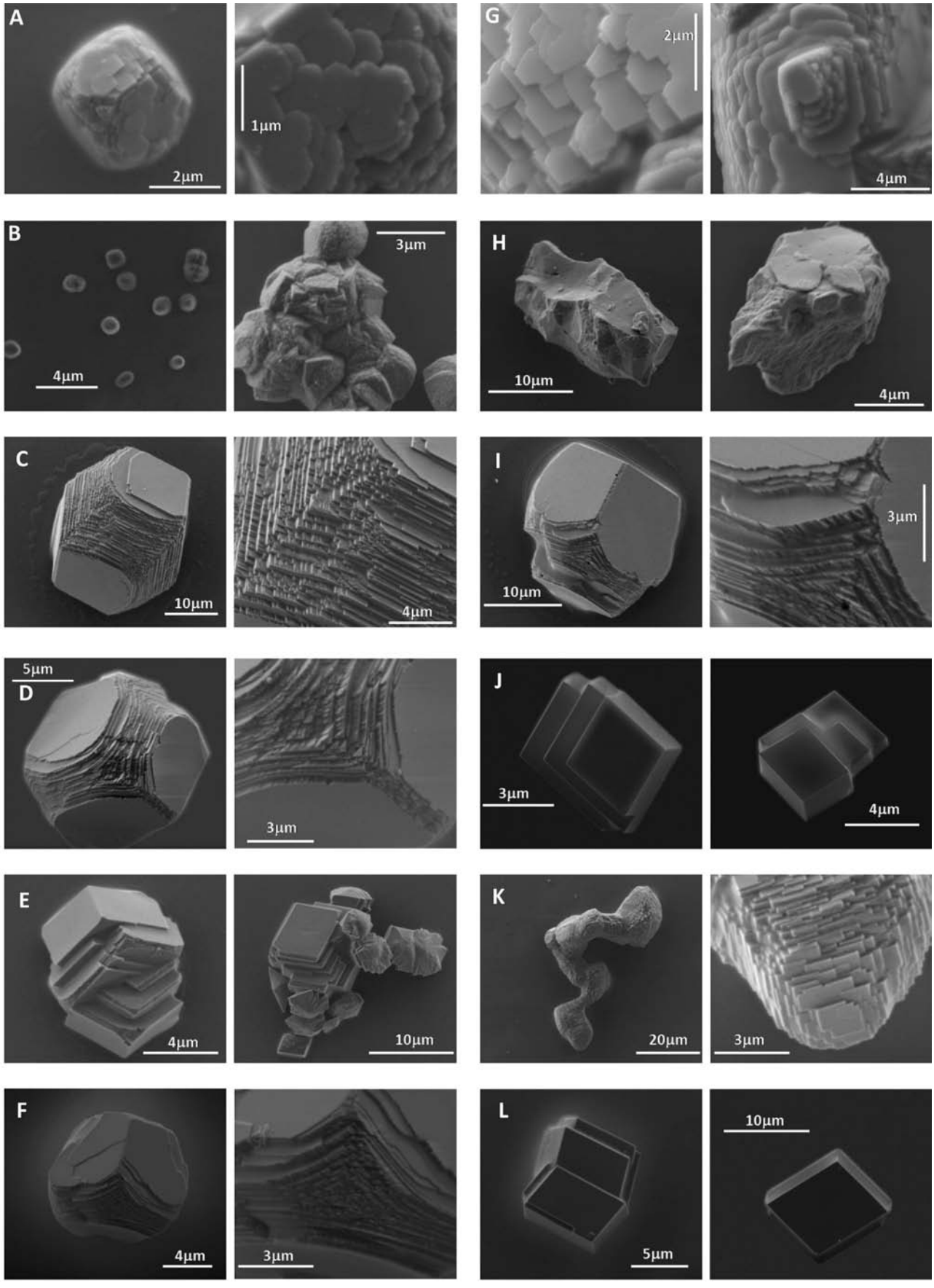

Figure 6. SEM images for calcite crystallized in the presence 0.45 mg/mL peptide. (A) design 1, (B) design 2, (C) design 3, (D) design 4, (E) design 5, (F) design 6, (G) variant 1, (H) variant 2, (I) variant 3, (J) variant 4, (K) variant 5, (L) variant 6.

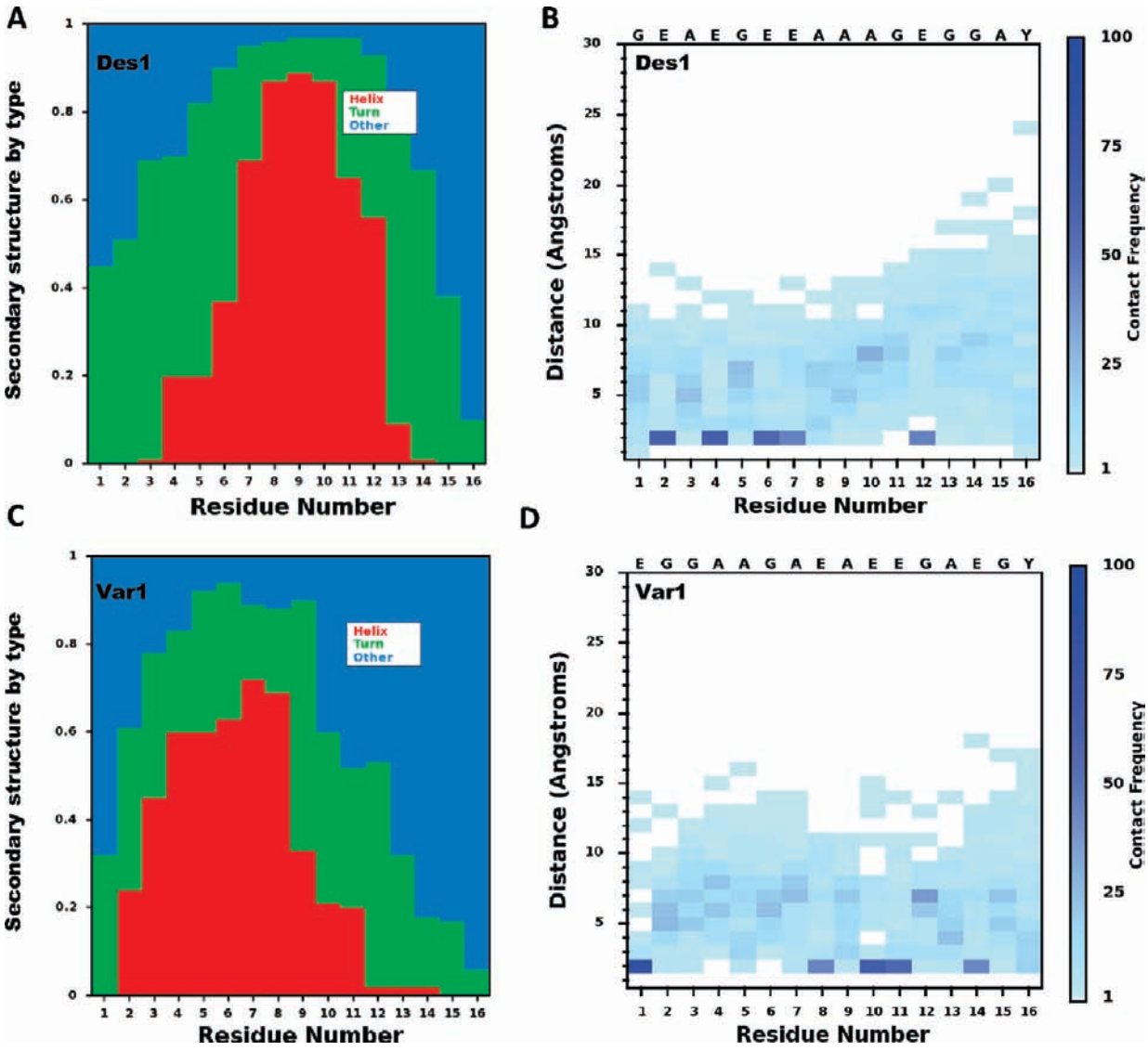

Figure 7. Computed ensemble structure of design 1 and variant 1 adsorbed to net positively charged calcite {001} surface (Figure 3A). (A and C) Distributions of three basic secondary structure motifs for design 1 and variant 1, respectively. The structural designations "Helix" and "Turn" were assigned using the DSSP[70] definitions; classification relied on Rosetta's hydrogen-bond function[48] rather than the generalized hydrogen-bond function implemented by the DSSP package. The structural designation "Other" indicates that hydrogen bonding was either long-range or absent at that residue. (B and D) Pair-wise residue-surface distances for the 100 predicted structures with lowest energy for design 1 and variant 1, respectively. A residue−surface distance reflects the closet atomic contact for that residue and the closest surface atom.

Scrambled variants of the net positively charged designed peptides displayed a range of ability to alter Ca_2CO_3 crystallization. Variant 3, which is net positively charged but contains two negatively charged residues, had an effect similar to that of design 3. Variant 4, which is positively charged and contains only one negatively charged residue, had little effect on crystallization, in contrast to design 4. This difference suggests that sequence order was a partial determinant of activity, not just the sequence composition. Variant 5 has no negatively charged residues and never created steps, but did create twinned crystals. Design 5 also caused crystal twinning, though to a lesser extent than variant 5. Also, design 5 created steps, whereas variant 5 did not. This suggests some sequence-order specificity for design 5. Variant 6, which contains five basic residues and no acidic residues, had little effect on Ca_2CO_3 crystallization. In contrast, design 6 had significant effect on Ca_2CO_3 crystallization, suggesting that sequence order, in addition to composition, played a role.

To compare structure, we generated structural ensembles for each of the six designed peptides and each of the scrambled variants, using RosettaSurface[41] in the absence of sequence optimization. For each of the 12 peptides, we generated 10^5 structures and analyzed the 100 structures with the lowest energy.

In all cases, the short peptide sequences have a moderate amount of flexibility. This flexibility facilitates strong binding at most sequence positions that have charged side chains. Also, in all cases, helix formation is highest in segments with high Ala content; Ala is known to have a high propensity for helix formation. This Ala-induced helix formation can be further stabilized when Ala segments are flanked by charged, surface-binding residues.

Figure 7 compares the secondary structure and peptide-surface contacts of design 1 (Figure 7A and 7B) with variant 1 (Figure 7C and 7D). In design 1, there is a high propensity for helix formation from residues 7 to 11 (Figure 7A), stabilized by high

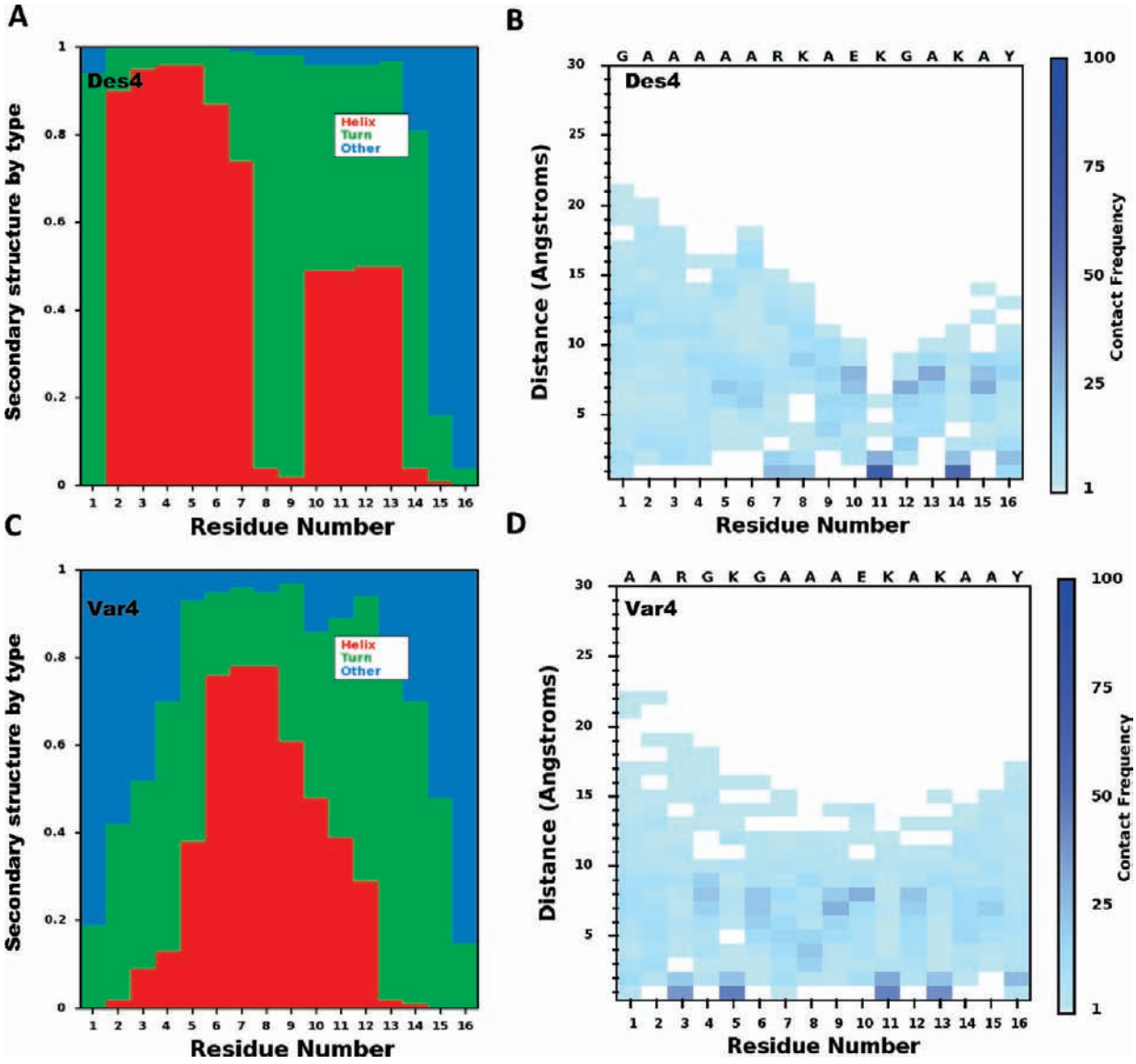

Figure 8. Computed ensemble structure of design 4 and variant 4 adsorbed to the net negatively charged calcite {001} surface (Figure 3B). (A and C) Distributions of three basic secondary structure motifs for design 4 and variant 4, respectively. (B and D) Pairwise residue−surface distances for the 100 predicted structures with lowest energy for design 4 and variant 4, respectively (see Figure 7 for complete description).

Ala content and strong binding of Glu residues flanking that segment (Figure 7B). Variant 1 (Figure 7C) shows moderate helix formation in the N-terminal segment and strong binding in the C-terminal segment. In both design 1 and variant 1, binding is strong for all five acidic residues (Figure 7B and 7D, respectively). Taken together, these observations result from the fact that the designed sequence was optimized for helical backbone hydrogen bonding and calcite binding, whereas the scrambled variant's sequence order was not optimized.

Figure 8 compares the secondary structure and peptide−surface contacts of design 4 (Figure 8A and 8B) with variant 4 (Figure 8C and 8D). In design 4, residues 2 through 7 form a stable poly-Ala helix (Figure 8A) further stabilized by binding of basic residues at positions 7 and 8 (Figure 8B); this motif significantly decreases the occurrence of simultaneous binding of residues 7 and 8. Additionally, there is moderate helix formation in the charged C-terminal binding segment of design 4, which participates in a helix-turn-helix motif. The Ala tripeptide from residues 7 to 9 in variant 4 are part of a short helical segment (Figure 8C) stabilized by charged surface-binding residues (Figure 8D). The greater flexibility of variant 4, compared with design 4, facilitates strong binding via all basic residues. In both design 4 and variant 4, the only acidic amino acid (Glu) is predicted not to locate at the surface owing to competitive binding of a neighboring Lys at the net negatively charged surface.

Discussion

Recently, there has been significant progress in the de novo design of biological macromolecules.[29−39,60] Particularly exciting is the design of proteins with novel properties. For instance, RosettaDesign successfully designed a protein fold not observed in nature[32] and successfully designed enzymes that catalyze nonbiological reactions.[36] Given the diversity of superior functional and structural materials fabricated in biological systems, the potential for novel materials resulting from the de novo design of biomineralization systems is seemingly limitless.

(60) Butterfoss, G. L.; Kuhlman, B. *Annu. Rev.* **2006**, *35*, 49–65.

One advantage of de novo design in biomineralization is the potential to target a specific crystal face. This includes the ability to target faces that are not well represented in the stable, fully developed crystal. For instance, directed evolution techniques are useful for evolving novel material binding peptide sequences. However, these directed evolution techniques suffer from an inability to design for a specific crystal face, and the experiment is typically limited to crystal faces present in the stable substrate material. The ability to design specific interactions in molecular biology is useful for recovering specific phenotypes.

Here we developed the first de novo design approach for biomineralization applications. The RosettaSurface.Design method designed peptides with sequence-order specificity and allowed us to identify basic residues as being a contributor to that specificity. Previously, Elhadj et al.[26] found sequence order had little effect on the kinetics of calcite crystal-growth for a set of peptides containing acidic residues and no basic residues. In our study, net negatively charged sequences (designed and variant) always produced drastic changes in calcite morphology. In contrast, some net positively charged sequences had little or no effect on calcite crystallization. Basic residues can affect calcite crystallization, but it appears that acidic residues can affect crystallization more potently. The results here show a greater dependence on specific sequence order for net positively charged peptides. Therefore, peptide−surface interactions involving basic residues may require greater optimization, in turn providing added specificity.

Some naturally evolved proteins involved in calcified tissue mineralization bind crystal surfaces via basic residues. Basic residues are essential for the interaction of human-salivary statherin with hydroxyapatite (the primary mineral component of tooth enamel).[40,41,61,62] We predicted statherin's basic residues to be important for specific recognition of a phosphate motif on the monoclinic {001} surface of hydroxyapatite (HAp).[40,41] Also, human lysozyme binds hydroxyapatite via basic residues.[63] Previously unrecognized is the fact that human lysozyme and human statherin, two salivary proteins, bind hydroxyapatite with very similar motifs. These motifs comprise four basic residues oriented in an equilateral parallelogram, ∼10 Å on each side; this is the geometry of open phosphate clusters replicated across the HAp {001} surface.[40] In the case of statherin and lysozyme, basic residues are important for binding and possibly specificity—incorporating similar motifs in designed peptides may afford similar binding properties.

Increased specificity may be achieved by requiring greater stability in design constructs. Small peptides are typically flexible and can acquire multiple binding motifs. Increased stability would help overcome some of this intrinsic flexibility and limit the number of accessible binding conformations. Increased peptide chain length and disulfide bonding are strategies that may increase stability. Osteocalcin is a well-studied HAp-biomineralization protein that exhibits these stabilizing factors.[64] The structure of this ∼43 residue protein is stabilized by a single conserved disulfide bond.[65,66] Osteocalcin binds HAp and does so with greater affinity in the presence of reducing agents.[67] Why would evolution conserve the stable, disulfide-bonded structure of osteocalcin, given that its affinity for HAp increases with decreasing structure? One possibility is that the folded conformation of osteocalcin facilitates a specific interaction with HAp. The disulfide-bonded structure of the stable protein can access fewer conformations and should therefore complement fewer HAp surfaces. Considering similar features during de novo design may result in interactions that are more stable and specific. Another strategy for increasing peptide stability is the incorporation of hydrocarbon *staples*.[68,69] Given the success of RosettaDesign in the designing stable, structured, globular proteins, the design of stable protein-adsorbed states may be possible with RosettaSurface.Design.

The de novo design of biomineralization systems presents some unique challenges. For instance, at the nanoscale, many different surface features may be present during the kinetic process of mineral crystallization. In addition, a crystal face may exhibit some disorder as ions encounter the surface and acquire their lattice positions. The similar chemistry and structure of many of these surface conformers may promote promiscuous protein binding in lieu of specific recognition. In the computational design of solution-state proteins, specificity has been achieved using so-called "negative design" to search for sequences that favor one binding partner over others. In the case of biomineralization, the sheer number of alternate surface conformer possibilities makes negative design difficult. The surfaces available to bind are dictated by the relative stabilities of different surface conformers and the kinetic process of crystallization. Thus, future designs will be improved by a greater understanding of the crystallization process, which can help guide the choice of surface conformers used for positive and negative design calculations. Also, increased affinity may result from negative design against the solution state.

Another challenge is relating gross crystal morphology to atomistic phenomena at the protein−surface interface. From SEM images, one cannot ascertain in what crystal plane, termination, or orientation a peptide is bound. Similarly, an algorithm that does not capture crystallization dynamics cannot predict crystal morphologies resulting from predicted interactions. These disparities complicate the interpretation of results. For instance, different peptides could produce different crystal morphologies even if they adsorbed to the same crystal face. Alternatively, different peptides could produce the same morphology even while binding different faces or terminations. The crystal morphology resulting from protein−crystal interactions should have some dependence on the energetics of that interaction. This affinity will dictate binding at different stages of crystal growth and influence crystal growth kinetics and thermodynamics. In summary, peptide-induced morphological changes are more complex than the simple inhibition or enhancement of a single crystal face.

(61) Goobes, R.; Goobes, G.; Shaw, W. J.; Drobny, G. P.; Campbell, C. T.; Stayton, P. S. *Biochemistry* **2007**, *46*, 4725–4733.
(62) Raghunathan, V.; Gibson, J. M.; Goobes, G.; Popham, J. M.; Louie, E. A.; Stayton, P. S.; Drobny, G. P. *J. Phys. Chem. B* **2006**, *110*, 9324–9332.
(63) Aizawa, T.; Koganesawa, N.; Kamakura, A.; Masaki, K.; Matsuura, A.; Nagadome, H.; Terada, Y.; Kawano, K.; Nitta, K. *FEBS Lett.* **1998**, *422*, 175–178.
(64) Hoang, Q. Q.; Sicheri, F.; Howard, A. J.; Yang, D. S. C. *Nature* **2003**, *425*, 977–980.
(65) Frazao, C.; Simes, D. C.; Coelho, R.; Alves, D.; Williamson, M. K.; Price, P. A.; Cancela, M. L.; Carrondo, M. A. *Biochemistry* **2005**, *44*, 1234–1242.
(66) Hauschka, P. V.; Carr, S. A. *Biochemistry* **1982**, *21*, 2538–2547.
(67) Hauschka, P. V.; Wians, F. H., Jr. *Anat. Record* **1989**, *224*, 180–188.
(68) Walensky, L. D.; Kung, A. L.; Escher, I.; Malia, T. J.; Barbuto, S.; Wright, R. D.; Wagner, G.; Verdine, G. L.; Korsmeyer, S. J. *Science* **2004**, *305*, 1466–1470.
(69) Kutchukian, P. S.; Yang, J. S.; Verdine, G. L.; Shakhnovich, E. I. *J. Am. Chem. Soc.* **2009**, *131*, 4622–4627.
(70) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 1239–1259.
(71) Crooks, E. G.; Hon, G.; Chandonia, J. M.; Brenner, S. E. *Genome Res.* **2004**, *14*, 1188–1190.

Deriving a quantitative framework for which crystal growth morphology can be predicted from predicted peptide−surface interactions will require a more exhaustive investigation of peptide sequence and structure determinants in biomineral formation. For instance, the high degree of structural specificity predicted for design 4 (Figure 8A) could be a contributing factor for the specific effect of that sequence on calcite crystal growth (Figure 6D and 6J), relative to variant 4 (Figure 8C). However, there is also a (smaller) difference in structural specificity of design 1 (Figure 7A) relative to variant 1 (Figure 7C), but these two sequences create similar biomineral morphologies (Figure 6A and 6G). A future study could introduce predicted, structurally perturbing mutations into a specific binding peptide, such as design 4, to further investigate the role of structure in specific crystal-growth modification.

One benefit incurred from RosettaSurface.Design is that designed sequences have an increased probability of binding the target face. RosettaSurface.Design is fast and samples ∼35 000 conformations and $\sim 10^6$ sequences for binding the target surface. This approach allows the rapid generation of many peptides that are chemically and geometrically compatible with any given surface. In addition, each round of de novo design can inform subsequent rounds of design. This information can bias the algorithm to capture advantageous phenotypes observed in previous designs and avoid those that were deleterious. In this way the user can apply selective pressure to evolve desired traits. In this study, for instance, the reduced amino acid library and reference energies were chosen based on observations made of naturally evolved systems (see Materials and Methods). And, future RosettaSurface.Design predictions for calcite biomineralization systems could consider the determinants of sequence-order specificity gleaned here. For instance, crystallization modification had a greater dependence on sequence order for net positively charged sequences than net negatively charged sequences. This dependence on sequence order suggests a dependence on molecular structure. This relationship, sequence determines structure, is the important link between sequence-order specificity and crystal-face-binding specificity. Therefore, the requirement to include basic amino acids in subsequent rounds of RosettaSurface.Design may help achieve a desired specific interaction. Additionally, we suggest negative design, expanded amino acid libraries, increased chain length, and increased protein stability as potential future directions for improved design of biomineralization proteins.

Acknowledgment. We thank Marc Ostermeier for providing lab space, Tamara Hendrickson and Megan Ehrenwerth for advice on peptide synthesis, Eleanor Small for her contributions in initial experimental design strategies, Yung-Ching Chien and Marc McKee for their review and input, The Arnold and Mabel Beckman Foundation's Young Investigator Award, and the National Science Foundation's CAREER Award for funding. SEM images were acquired at the Johns Hopkins University Integrated Imaging Center.

JA1001086