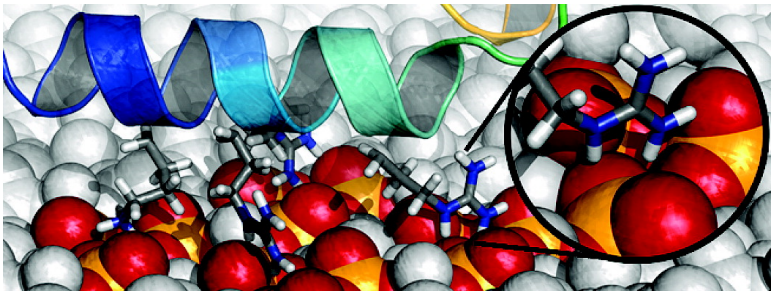
Article

Structure Prediction of Protein–Solid Surface Interactions Reveals a Molecular Recognition Motif of Statherin for Hydroxyapatite

Kosta Makrodimitris, David L. Masica, Eric T. Kim, and Jeffrey J. Gray

J. Am. Chem. Soc., **2007**, 129 (44), 13713-13722• DOI: 10.1021/ja074602v • Publication Date (Web): 12 October 2007



J|A|C|S
ARTICLES
Published on Web 10/12/2007

Structure Prediction of Protein−Solid Surface Interactions Reveals a Molecular Recognition Motif of Statherin for Hydroxyapatite

Kosta Makrodimitris,[†] David L. Masica,[‡] Eric T. Kim,[§] and Jeffrey J. Gray*[,†,‡,||]

Contribution from the Department of Chemical and Biomolecular Engineering, Program in Molecular and Computational Biophysics, Departments of Biomedical Engineering and Computer Science, and Institute for NanoBioTechnology, Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218

Received July 9, 2007; E-mail: jgray@jhu.edu

Abstract: A molecular description of protein−surface interactions could open new avenues in bionanotechnology and provide a deeper understanding of *in vivo* phase boundary biophysics. However, current experimental techniques can provide only inferential or incomplete information about the protein−surface interface. We present a novel computational method for modeling the interactions of proteins with solid surfaces using comprehensive sampling and an atomistic description. The approach relies on an all-atom Monte Carlo plus-minimization search algorithm that rapidly and simultaneously optimizes rigid-body and side-chain conformations. We apply the method to the statherin−hydroxyapatite system, an evolved protein−surface interaction that is likely to have one or a few specific structural solutions. The algorithm converges on a set of low energy, entropically favorable structures that are consistent with previous experimental results, namely protein−surface intermolecular distances acquired by solid-state NMR. The simulations isolate particular residues as being primary contributors to the adsorption free energy (hydrogen bonding, van der Waals, and electrostatic energies), in agreement with previous mutagenesis, deletion, and single amino acid experiments. We also report the discovery of a molecular recognition motif where the N-terminal α-helix of statherin places all four of its basic residues to match the periodicity of open phosphate triad clusters across the [001] monoclinic face of the hydroxyapatite surface. Results suggest new experiments that could further elucidate the structural features of this important biological system.

Introduction

Natural systems have evolved the capacity to control the crystal growth and morphology of biological hard tissues such as silica, calcite, and hydroxyapatite (HAp) through the use of proteins.[1,2] Similarly, directed evolution techniques have been successfully implemented to biopan for peptides that modulate mineralization of nonbiological materials such as gold and silver.[3] Rational design of proteins for specific interactions with such solid-phase materials could allow us to engineer metals, semiconductors, and tissues with nanoscale precision. These applications would most likely require knowledge of the protein−surface interface with atomic resolution and an understanding of the principles of molecular recognition.

Despite its importance, there are currently no protein−surface complexes whose atomic coordinates have been experimentally determined. Experimental measurements of the interface are very challenging and describe either bulk phenomena (stability of adsorbed protein on the surface, solvent effects such as pH or ionic strength, protein concentration, etc.) or atomic information limited to a few regions.[4] Computational approaches could help integrate experimental data and create plausible structural models of proteins on solid surfaces.[4] Several research groups have simulated proteins and peptides on surfaces using detailed atomistic approaches such as molecular dynamics (MD),[5−8] Monte Carlo methods,[9−11] a combination of MD with docking protocols,[12,13] and minimization techniques.[14−19] These high-

[†] Department of Chemical and Biomolecular Engineering.
[‡] Program in Molecular and Computational Biophysics.
[§] Departments of Biomedical Engineering and Computer Science.
[||] Institute for NanoBioTechnology.

(1) Weiner, S.; Addadi, L. *J. Mater. Chem.* **1997**, *7*, 689−702.
(2) Mann, S. *Nature* **1993**, *365*, 499−505.
(3) Sarikaya, M.; Tamerler, C.; Schwartz, D. T.; Baneyx, F. O. *Annu. Rev. Mater. Res.* **2004**, *34*, 373−408.
(4) Gray, J. J. *Curr. Opin. Struct. Biol.* **2004**, *14*, 110−115.
(5) Agashe, M.; Raut, V.; Stuart, S. J.; Latour, R. A. *Langmuir* **2005**, *21*, 1103−1117.
(6) Braun, R.; Sarikaya, M.; Schulten, K. *J. Biomater. Sci., Polym. Ed.* **2002**, *13*, 747−757.
(7) Dalal, P.; Knickelbein, J.; Haymet, A. D. J.; Sonnichsen, F. D.; Madura, J. D. *Phys. Chem. Commun.* **2001**, *4*, 32−36.
(8) Raut, V. P.; Agashe, M. A.; Stuart, S. J.; Latour, R. A. *Langmuir* **2005**, *21*, 1629−1639.
(9) Jorov, A.; Zhorov, B. S.; Yang, D. S. C. *Protein Sci.* **2004**, *13*, 1524−1537.
(10) Mungikar, A.; Forciniti, D. *Biomacromolecules* **2006**, *7*, 239−251.
(11) Noinville, V.; Vidalmadjar, C.; Sebille, B. *J. Phys. Chem.* **1995**, *99*, 1516−1522.
(12) Huq, N. L.; Cross, K. J.; Reynolds, E. C. *J. Mol. Model.* **2000**, *6*, 35−47.
(13) Yarovsky, I.; Hearn, M. T. W.; Aguilar, M. I. *J. Phys. Chem. B* **1997**, *101*, 10962−10970.
(14) Cormack, A. N.; Lewis, R. J.; Goldstein, A. H. *J. Phys. Chem. B* **2004**, *108*, 20408−20418.
(15) Madura, J. D.; Wierzbicki, A.; Harrington, J. P.; Maughon, R. H.; Raymond, J. A.; Sikes, C. S. *J. Am. Chem. Soc.* **1994**, *116*, 417−418.
(16) Oren, E. E.; Tamerler, C.; Sarikaya, M. *Nano Lett.* **2005**, *5*, 415−419.

resolution techniques demand large CPU times and are limited to small proteins and short time and length scales. On the other hand, coarse-grained simulations have been employed by either modeling the protein structure in low resolution[20−23] or describing the surface mesoscopically with parametrization strategies[24−30] to access longer time and length scales and compare with experimental data under different conditions within resolution limits.[26,31]

In this work, we describe a new method for studying protein−surface interactions based on protein structure prediction approaches. The protocol, RosettaSurface, exploits the Rosetta energy function (dominated by van der Waals, solvation, and hydrogen-bonding energies) that has been applied successfully in biological problems[32] such as protein folding,[33,34] design,[35] and docking.[36−38] Rosetta employs rapid structure prediction techniques which sample conformation space broadly using discrete fragment and side-chain rotamer libraries, a semiempirical energy function based on physical molecular potentials and statistics from high-resolution protein structures, and continuous minimization methods. This unique combination of strategies samples extensively while maintaining an all-atom description of the entire system.

We apply our new method to human salivary statherin on HAp. HAp (Figure 1a) is a calcium phosphate mineral and is the primary inorganic component of both bone and tooth enamel. Statherin, a 43-residue doubly phosphorylated (Sep2, Sep3) human salivary protein (Figure 1b), binds HAp *in vivo* and plays an important role in crystal growth inhibition. This system is an ideal test for RosettaSurface for two reasons. First, since statherin evolved to bind HAp and the adsorption is known to be reversible,[39] one or a few configurations of the statherin−HAp complex are likely to dominate the molecular recognition event. Therefore, we seek the global free-energy minimum conformation of the surface-bound statherin. Second, the structure of statherin on HAp has been probed experimentally, so we can assess the accuracy of the predictions and extend existing models. Experimental measurements that give insight into the adsorbed microstructure include fragmentation and substitution experiments to localize the interactions,[40] intramolecular NMR measurements while adsorbed,[41,42] dynamic NMR studies which demonstrate broad regions which interact with HAp,[42] solid-state NMR measurements which provide distance constraints between side chains and the surface,[43−45] and

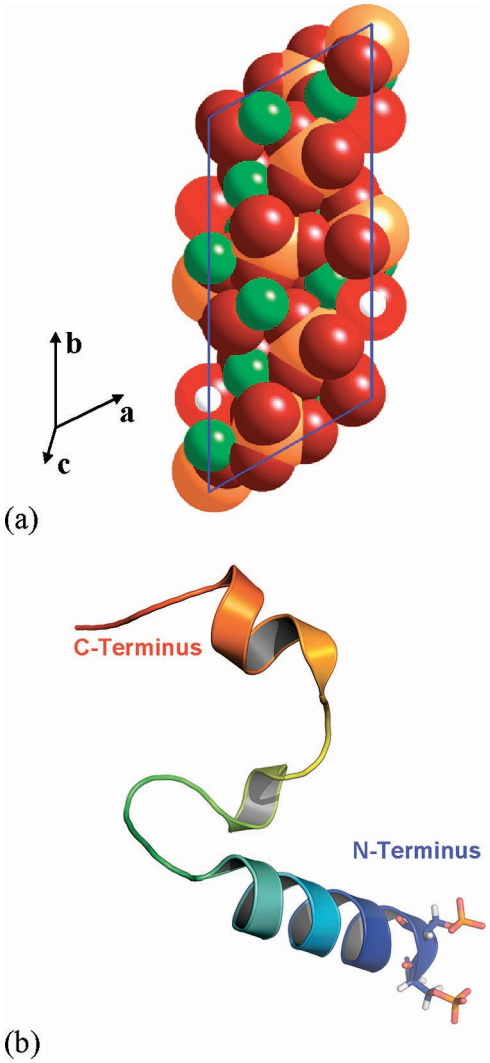


Figure 1. (a) Crystal structure of hydroxyapatite, $Ca_5(PO_4)_3(OH)$, [001] plane with lattice parameters $a = 9.4$ Å, $b = 18.8$ Å, $c = 6.9$ Å.[47] Coloring (for all HAp figures): Ca, green; P, orange; H, white; O, dark red for phosphate group O atoms which function as hydrogen bond proton acceptors and red for O atoms which can function as either proton donors or acceptors. (b) Statherin structure selected from the set of Goobes et al.[48] and refined using intramolecular experimental distance constraints (see methods). Sticks depict the two phosphoserine side chains Sep2 and Sep3.

(17) Raffaini, G.; Ganazzoli, F. *Langmuir* **2003**, *19*, 3403−3412.
(18) Raffaini, G.; Ganazzoli, F. *Phys. Chem. Chem. Phys.* **2006**, *8*, 2765−2772.
(19) Wierzbicki, A.; Madura, J. D.; Salmon, C.; Sonnichsen, F. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1006−1010.
(20) Knotts, T. A.; Rathore, N.; de Pablo, J. J. *Proteins* **2005**, *61*, 385−397.
(21) Skepo, M.; Linse, P.; Arnebrant, T. *J. Phys. Chem. B* **2006**, *110*, 12141−12148.
(22) Sun, Y.; Welsh, W. J.; Latour, R. A. *Langmuir* **2005**, *21*, 5616−5626.
(23) Zhou, J.; Chen, S. F.; Jiang, S. Y. *Langmuir* **2003**, *19*, 3472−3478.
(24) Grossfield, A.; Sachs, J.; Woolf, T. B. *Proteins* **2000**, *41*, 211−223.
(25) Makrodimitris, K.; Fernandez, E. J.; Woolf, T. B.; O'Connell, J. P. *Mol. Simul.* **2005**, *31*, 623−636.
(26) Makrodimitris, K.; Fernandez, E. J.; Woolf, T. B.; O'Connell, J. P. *Anal. Chem.* **2005**, *77*, 1243−1252.
(27) Ravichandran, S.; Madura, J. D.; Talbot, J. *J. Phys. Chem. B* **2001**, *105*, 3610−3613.
(28) Roth, C. M.; Lenhoff, A. M. *Langmuir* **1995**, *11*, 3500−3509.
(29) Ladiwala, A.; Xia, F.; Luo, Q. O.; Breneman, C. M.; Cramer, S. M. *Biotechnol. Bioeng.* **2006**, *93*, 836−850.
(30) Sethuraman, A.; Vedantham, G.; Imoto, T.; Przybycien, T.; Belfort, G. *Proteins* **2004**, *56*, 669−678.
(31) Makrodimitris, K.; Fernandez, E. J.; Woolf, T. B.; O'Connell, J. P. *Biotechnol. Prog.* **2005**, *21*, 893−896.
(32) Schueler-Furman, O.; Wang, C.; Bradley, P.; Misura, K.; Baker, D. *Science* **2005**, *310*, 638−642.
(33) Bradley, P.; Malmstrom, L.; Qian, B.; Schonbrun, J.; Chivian, D.; Kim, D. E.; Meiler, K.; Misura, K. M. S.; Baker, D. *Proteins* **2005**, *61*, 128−134.
(34) Bradley, P.; Misura, K. M. S.; Baker, D. *Science* **2005**, *309*, 1868−1871.
(35) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. *Science* **2003**, *302*, 1364−1368.
(36) Gray, J. J.; Moughon, S. E.; Kortemme, T.; Schueler-Furman, O.; Misura, K. M.; Morozov, A. V.; Baker, D. *Proteins* **2003**, *52*, 118−22.
(37) Daily, M. D.; Masica, D.; Sivasubramanian, A.; Somarouthu, S.; Gray, J. J. *Proteins* **2005**, *60*, 181−186.
(38) Chaudhury, S.; Sircar, A.; Sivasubramanian, A.; Berrondo, M.; Gray, J. J. *Proteins* **2007**, *in press*.
(39) Goobes, R.; Goobes, G.; Campbell, C. T.; Stayton, P. S. *Biochemistry* **2006**, *45*, 5576−5586.
(40) Raj, P. A.; Johnsson, M.; Levine, M. J.; Nancollas, G. H. *J. Biol. Chem.* **1992**, *267*, 5968−5976.
(41) Long, J. R.; Dindot, J. L.; Zebroski, H.; Kiihne, S.; Clark, R. H.; Campbell, A. A.; Stayton, P. S.; Drobny, G. P. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12083−7.
(42) Long, J. R.; Shaw, W. J.; Stayton, P. S.; Drobny, G. P. *Biochemistry* **2001**, *40*, 15451−5.
(43) Gibson, J. M.; Popham, J. M.; Raghunathan, V.; Stayton, P. S.; Drobny, G. P. *J. Am. Chem. Soc.* **2006**, *128*, 5364−5370.
(44) Gibson, J. M.; Raghunathan, V.; Popham, J. M.; Stayton, P. S.; Drobny, G. P. *J. Am. Chem. Soc.* **2005**, *127*, 9350−1.
(45) 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.

thermodynamic measurements on statherin mutants.[39,46] Through these studies, statherin is known to have an α-helix near the N-terminus, and several cartoon models have been postulated.[40,42,43,45,46]

Our goal is to develop methods for modeling proteins on solid surfaces. To our knowledge, this is the first time a comprehensive suite of tools from the protein structure prediction field has been applied to the protein−surface interaction problem. In this study, we aim to predict the structure and orientation of statherin on an HAp surface, to identify contributions to the binding energy, and to provide a framework for additional directed investigations by solid-state NMR and mutagenesis experiments. The predicted models give a clearer picture of the statherin−HAp complex and the binding mechanism that validates and extends current experimental hypotheses.

Computational Methods

Statherin Model. For the RosettaSurface simulations presented in this paper, the statherin backbone was held fixed. This assumption allows for more sampling of rigid-body space and tests the extent of agreement with experiment that can be obtained from a fixed backbone simulation. Typically, Rosetta structure prediction applications, such as RosettaDock, RosettaDesign, and RosettaAbInitio, are parametrized and tested using large data sets of high-resolution protein crystal structures from the Protein Data Bank (PDB).[49] Unfortunately, no such benchmark exists for protein−surface simulations, and for this study validation depends on solid-state NMR distance measurements and binding assays.

Candidate low-resolution structures of statherin were predicted previously using Rosetta.[44,48] To prepare an input structure for docking, we chose the best-scoring structure from the largest cluster of similar structural models (31 "decoys"). This model agrees with intramolecular atomic distances determined by solid-state NMR[48] and circular dichroism studies[50] that indicate that some or most of the 15 N-terminal residues are α-helical. The agreement with experimental data obtained from surface-adsorbed statherin and the fact that Rosetta selects for compact folded structures suggests that the model represents the adsorbed structure of statherin rather than the solution-phase structure which is suspected to be extended. Furthermore, prior to docking we refined the model with Rosetta[51,52] to obtain a high-resolution, all-atom statherin structure at a local energy minimum using the intramolecular distance constraints measured by solid-state NMR: Pro33 C_α to Tyr34 C_α, 3.0−3.25 Å; Pro33 C_α to Tyr38 N, 5.0−5.8 Å; Tyr34 C_α to Tyr38 N, 3.5−4.5 Å; and Tyr34 C_α to Pro23 C_α, 10−11 Å.[48] Rosetta refinement optimizes the structure's energy via combinatorial rotamer packing and gradient-based minimization of backbone and side-chain torsion angles.[51,52]

Hydroxyapatite Model. Biological HAp is amorphous on a 100 nm length scale but is approximately crystalline over the much smaller length-scale of the protein (<2 nm).[45] Stoichiometric synthetic HAp crystallizes in the monoclinic space group at room temperature and is closely related to the hexagonal (high temperature) HAp crystal structure except for the loss of inversion symmetry.[45,53] HAp's fastest growth rate occurs in the direction of the **c** axis (Figure 1a) by deposition of ions in the [001] plane.[54−57] The [001] plane of HAp is the most stable face under dehydrated and hydrated conditions and is the major experimental cleavage plane.[58−60] We generated coordinates for HAp ($P2_1/b$ monoclinic) with the [001] face exposed using CrystalMaker software.[61] While solid-state NMR can distinguish neither the face for statherin adsorption nor whether statherin binds to a flat surface face, step edge, or defect, both experimental[42,62,63] and simulation studies[12,64,65] have assumed [001] to be the binding plane and the predominant surface at the enamel−saliva interface.[56,58,65,66] The lattice parameters of the unit cell[47] are shown in Figure 1a.

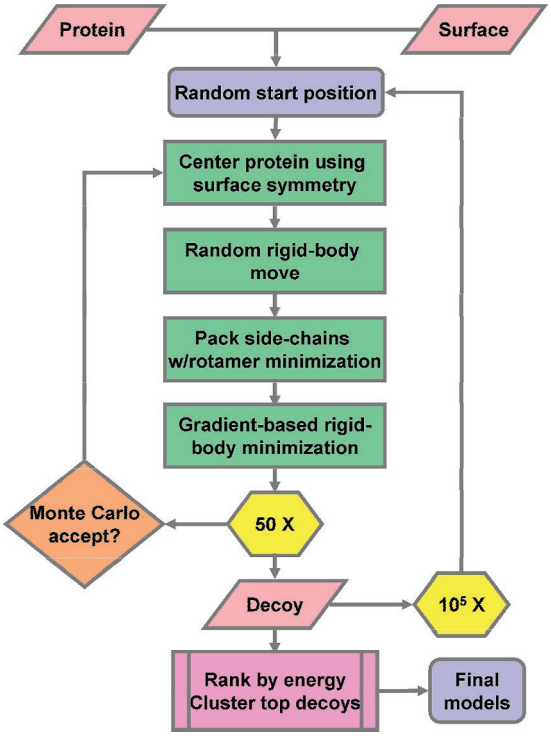


Figure 2. RosettaSurface flowchart.

Simulation Algorithm. As with other protein structure prediction problems, the primary scientific challenges for predicting the structure of proteins on solid surfaces are (1) sampling the immense conformational space of the protein on the surface and (2) accurately calculating a potential function to identify the lowest free-energy structure. The algorithm we developed is designed to address these challenges.

Figure 2 shows the flowchart of the new RosettaSurface protocol. First, the protein is randomly oriented and positioned away from the

(46) Goobes, R.; Goobes, G.; Shaw, W. J.; Drobny, G. P.; Campbell, C. T.; Stayton, P. S. *Biochemistry* **2007**, *46*, 4725−33.
(47) Elliott, J. C.; Mackie, P. E.; Young, R. A. *Science* **1973**, *180*, 1055−1057.
(48) Goobes, G.; Goobes, R.; Schueler-Furman, O.; Baker, D. B.; Stayton, P. S.; Drobny, G. P. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16083−8.
(49) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(50) Naganagowda, G. A.; Gururaja, T. L.; Levine, M. J. *J. Biomol. Struct. Dyn.* **1998**, *16*, 91−107.
(51) Tsai, J.; Bonneau, R.; Morozov, A. V.; Kuhlman, B.; Rohl, C. A.; Baker, D. *Proteins* **2003**, *53*, 76−87.
(52) Rohl, C. A.; Strauss, C. E.; Misura, K. M.; Baker, D. *Methods Enzymol.* **2004**, *383*, 66−93.
(53) Hochrein, O.; Kniep, R.; Zahn, D. *Chem. Mater.* **2005**, *17*, 1978−1981.
(54) Barnett, B. L.; Strickland, L. C. *Acta Crystallogr., Sect. B* **1979**, *35*, 1212−1214.
(55) Margolis, H. C.; Beniash, E.; Fowler, C. E. *J. Dent. Res.* **2006**, *85*, 775−793.
(56) Simmer, J. P.; Fincham, A. G. *Crit. Rev. Oral Biol. Med.* **1995**, *6*, 84−108.
(57) Zhan, J. H.; Tseng, Y. H.; Chan, J. C. C.; Mou, C. Y. *Adv. Funct. Mater.* **2005**, *15*, 2005−2010.
(58) Mkhonto, D.; de Leeuw, N. H. *J. Mater. Chem.* **2002**, *12*, 2633−2642.
(59) Palache, C.; Berman, H.; Frondel, C. The system of mineralogy of James Dwight Dana and Edward Salisbury Dana. 7th ed.; John Wiley and Sons, Inc.: New York 1951, Vol. 2, p 880−881.
(60) Simon, P.; Carrillo-Cabrera, W.; Formanek, P.; Gobel, C.; Geiger, D.; Ramlau, R.; Tlatlik, H.; Buder, J.; Kniep, R. *J. Mater. Chem.* **2004**, *14*, 2218−2224.
(61) Palmer, R. CrystalMaker Software Ltd: Yarnton, Oxfordshire, U.K., 1995.
(62) Hauschka, P. V.; Carr, S. A. *Biochemistry* **1982**, *21*, 2538−2547.
(63) Neves, M.; Gano, L.; Pereira, N.; Costa, M. C.; Costa, M. R.; Chandia, M.; Rosado, M.; Fausto, R. *Nucl. Med. Biol.* **2002**, *29*, 329−338.
(64) Wright, J. E. I.; Zhao, L. Y.; Choi, P.; Uludag, H. *Biomaterials: From Molecules to Engineered Tissues* **2004**, *553*, 139−148.
(65) Zahn, D.; Hochrein, O. *Phys. Chem. Chem. Phys.* **2003**, *5*, 4004−4007.
(66) Radlanski, R. J.; Seidl, W.; Steding, G.; Jager, A. *Anat. Anseiger* **1989**, *168*, 405−412.

surface where there is no intermolecular interaction. The protein is then translated (1) parallel to the surface an integer combination of HAp unit cell vectors **a** and **b** and (2) normal to the surface to position the protein in glancing contact over the central unit cell (Figure S2 in the Supporting Information). The simulation algorithm is a modified version of the all-atom refinement stage of the Monte Carlo plus-minimization (MCM)-based[67] RosettaDock algorithm.[68] Each cycle of the MCM algorithm begins with a Gaussian distributed random translation of mean 0.1 Å in each Cartesian direction and a Gaussian distributed random rotation of mean 0.05 radians around each Cartesian axis with the rotation origin at the centroid of the interface residues. Next, the interfacial side chains are optimized with a rotamer packing algorithm using an expanded rotamer set[69,70] and off-rotamer minimization in torsion space as described by Wang et al.[71] During this step, only the residues with side-chain centroid positions less than 7−10 Å (depending on the amino acid size) from the interface are repacked.[68,71] Finally, the rigid-body translation and rotation are optimized using a gradient-based, iterative Davidson−Fletcher−Powell minimization.[72] The minimized structure is compared with the previously accepted structure using the standard Metropolis criterion to determine whether to accept or reject the new position. The MCM procedure, which optimizes both side-chain and rigid-body positions in each iteration, is repeated for 50 cycles, resulting in a final "decoy" structure.

The best-scoring 200 decoys are retained and compared based on the root-mean-square deviation (rmsd) of the statherin C_α atoms after superposition of the HAp and optimal translation across equivalent crystal surface unit cells (Figure S2 in the Supporting Information). Sets of decoys within 2.5 Å rmsd of each other are designated as a cluster using a hierarchical clustering algorithm.[73] Good clustering indicates convergence of the simulation and favorable conformational entropy.

Protein Energy Function. The Rosetta energy function has been optimized using several distinct structure prediction applications including folding, docking, and design.[52] In brief, the all-atom protein energy function includes van der Waals interactions using a modified Lennard−Jones potential;[68] solvation using a pairwise Gaussian solvent-exclusion model;[74] hydrogen-bonding energies using an orientation-dependent function derived from high-resolution protein structures and quantum calculations;[75,76] side-chain internal energies using a rotamer probability term;[69] residue−residue pair interactions derived statistically from a database of protein structures;[70] and an electrostatic term with a distance-dependent dielectric constant and a cutoff of 5.5 Å.[77] The energy function includes the solvent entropy explicitly but neglects the conformational entropy of the protein. The energy terms are weighted as in protein−small-molecule docking studies.[78]

Since statherin contains two phosphoserines (Sep) crucial for the interaction with HAp (Figure 1a), we added and parametrized a phosphorylated serine residue for the Rosetta package. Lennard−Jones parameters and partial charges for electrostatics are obtained from CHARMM27 nucleic acid parameters.[79,80] Sep phosphate oxygen atoms assume an sp^3 hybridization state and function as proton acceptors in the hydrogen bond calculation. The rotamer database for the side-chain packing algorithm requires appropriate sets of side-chain torsion angles.[69] Therefore, to sample the rotamer conformations with the appropriate probability, we analyzed all of the 118 protein structures with Sep residues in the Protein Data Bank.[49] In the Supporting Information, Figure S1 shows histograms of the observed Sep χ angle frequencies, and Table S1 lists the average χ angles and standard deviations used for the side-chain packing algorithm. Due to the small number of structures, χ_1, χ_2, and χ_3 are assumed to be independent.

Protein−Surface Energy Function. The force field parameters for HAp are chosen as follows. The Lennard−Jones well depth, ϵ, and the equilibrium internuclear separation, σ, are taken from the CHARMM22 all-hydrogen parameter file for proteins[81,82] and CHARMM27 for nucleic acids.[79,80] For purposes of weighting the score function contributions, the Lennard−Jones interactions are partitioned into those interactions which are net attractive ($r > 0.89\ \sigma_{ij}$) and those which are net repulsive and cause clashes in the structure ($r < 0.89\ \sigma_{ij}$). As in previous Rosetta simulations,[68] the repulsive Lennard−Jones term is linearly extrapolated at small radii ($r < 0.6\ \sigma_{ij}$) to remove the singularity and improve minimization performance. For the hydrogen-bonding calculation, the HAp phosphate oxygen atoms (sp^3 hybridization) are designated as proton acceptors, and the hydroxyl oxygen atoms function as both proton acceptors and donors.[65] The Gaussian solvent-exclusion model for the solvation free energy is extended to the phosphate group atoms with empirical parameters from micelle measurements.[83] The partial atomic charges have been derived from quantum-chemical calculations and dynamical simulated annealing optimization.[84] A distance-dependent dielectric constant[77] approximates the varying dielectric constant in the system. Electrostatic energies are smoothly truncated at 8.0 Å and capped at a maximum value matching that at the sum of the van der Waals radii of the interacting atoms. The full-atom energy of the protein-surface interaction is a linear combination of attractive and repulsive Lennard−Jones interactions (E_{att} and E_{rep}), solvation (E_{sol}), hydrogen bonding (E_{hb}), and electrostatics (E_{coul}):

$$E_{p-s} = w_{att}\,E_{att} + w_{rep}\,E_{rep} + w_{sol}\,E_{sol} + w_{hb}\,E_{hb} + w_{coul}\,E_{coul}$$

Weights for combining the energies have been fitted to experimental data of multiple ligand molecules (updated set since the original paper,[78] personal communication, Jens Meiler) $w_{att} = 0.8$, $w_{rep} = 0.4$, $w_{hb} = 2.0$, $w_{sol} = 0.6$, $w_{coul} = 0.25$.

Structure Evaluation. Candidate decoy structures are inspected using a combination of external postprocessing tools to ensure structures display "protein-like" features. Of particular interest are interfacial shape complementarity, interfacial voids that are inaccessible to solvent, and a completely satisfied hydrogen bond network.

Shape complementarity was assessed using the Fast Atomic Density Evaluator (FADE).[85,86] FADE creates a lattice of points at the molecular interface about which each molecule is proximally determined to present a crevice, protrusion, or flat surface by calculating a density exponent d characterizing the rate of increase in atomic density, ρ, as a function of radius, r, where $\rho \approx r^d$. Then, $d < 2.8$ is defined as a protrusion, d

(67) Li, Z. Q.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6611−6615.
(68) Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D. *J. Mol. Biol.* **2003**, *331*, 281−299.
(69) Dunbrack, R. L., Jr.; Cohen, F. E. *Protein Sci.* **1997**, *6*, 1661−81.
(70) Kuhlman, B.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10383−10388.
(71) Wang, C.; Schueler-Furman, O.; Baker, D. *Protein Sci.* **2005**, *14*, 1328−1339.
(72) *Numerical Recipes in C++: The Art of Scientific Computing*, 2nd ed.; Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P., Eds.; Cambridge University Press: Cambridge, U.K., 2002.
(73) Ihaka, R.; Gentleman, R. *J. Comput. Graph. Stat.* **1996**, *5*, 299−314.
(74) Lazaridis, T.; Karplus, M. *Proteins* **1999**, *35*, 133−152.
(75) Kortemme, T.; Morozov, A. V.; Baker, D. *J. Mol. Biol.* **2003**, *326*, 1239−59.
(76) (a) Morozov, A. V.; Kortemme, T.; Tsemekhman, K.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6946−6951. (b) Morozov, A. V.; Misura, K. M. S.; Tsemekhman, K.; Baker, D. *J. Phys. Chem. B* **2004**, *108*, 8489−8496.
(77) Warshel, A.; russell, S. T.; Chrug, A. K. *Proc. Natl. Acad. Sci.-Biol.* **1984**, *81*, 4785−4789.
(78) Meiler, J.; Baker, D. *Proteins* **2006**, *65*, 538−548.
(79) Foloppe, N.; MacKerell, A. D. *J. Comput. Chem.* **2000**, *21*, 86−104.
(80) MacKerell, A. D.; Banavali, N. K. *J. Comput. Chem.* **2000**, *21*, 105−120.
(81) MacKerell, A. D., et al. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(82) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(83) Lazaridis, T.; Mallik, B.; Chen, Y. *J. Phys. Chem. B* **2005**, *109*, 15098−15106.
(84) Hauptmann, S.; Dufner, H.; Brickmann, J.; Kast, S. M.; Berry, R. S. *Phys. Chem. Chem. Phys.* **2003**, *5*, 635−639.
(85) Kuhn, L. A.; Siani, M. A.; Pique, M. E.; Fisher, C. L.; Getzoff, E. D.; Tainer, J. A. *J. Mol. Biol.* **1992**, *228*, 13−22.
(86) Mitchell, J. C.; Kerr, R.; Ten Eyck, L. F. *J. Mol. Graph. Model.* **2001**, *19*, 324−329.

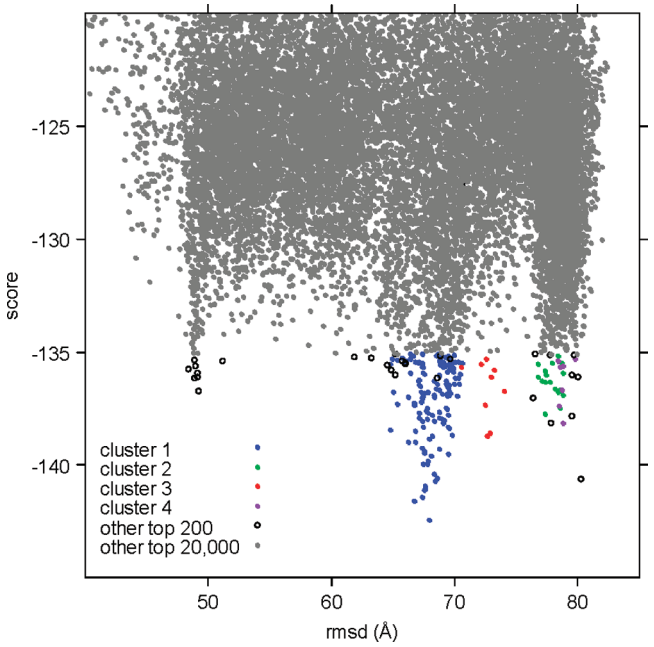


Figure 3. Statherin−HAp energy landscape. Scores of the top quartile of the 80 000 decoys are plotted as a function of the distance (rmsd) from an arbitrary starting position. The top-scoring 200 decoys are grouped into clusters indicated by color.

> 2.8, a crevice, and $d \approx 2.8$, as flat. Density exponents are converted to a score using the following equation $s = (d_1 - d_0)(d_2 - d_0)$, where d_1 and d_2 are the density exponents of molecules 1 and 2, and d_0 is the density exponent for a volume of perfectly packed spheres (3.0 in theory, in practice set to 2.8 to reflect average efficiency of protein packing). A negative value for s represents good shape complementarity and is obtained only when an intermolecular crevice and protrusion meet at an interface.

Solvent accessibility was calculated using NACCESS version 2.1.1.[87] Since Rosetta models solvent implicitly, NACESS helps determine the probability of finding a water molecule at an interfacial cavity and satisfying a hydrogen bond. NACCESS uses the "rolling ball" method[88] with a probe radius of 1.4 Å. Protein and surface atomic radii were set to RosettaSurface van der Waals radii, and all hydrogen atoms were included in the solvent accessibility calculations.

Contour plots of shape complementarity and solvent accessibility are superimposed on the molecular coordinates and viewed using PyMOL[89] to reveal regions of the interface where poor solvent accessibility and poor shape complementarity are coincident. Regions displaying poor shape complementarity should be solvent-accessible to solvent to avoid vacuum at an interface. Similarly, solvent accessibility plots, hydrogen bond networks, and molecular coordinates are superimposed to ensure that all hydrogen bond donors and acceptors that are not involved in inter- or intramolecular hydrogen bonds can be satisfied by solvent.

Results

The RosettaSurface simulation generated 80 000 decoy structures in approximately 150 CPU-days on a cluster of 2.4 GHz Linux processors. Considering that each decoy completed 50 MCM cycles, 4×10^6 minima were sampled in the rigid-body conformation space. Furthermore, $\sim 10^3$ conformations were sampled at each rigid-body position during the side-chain packing steps.

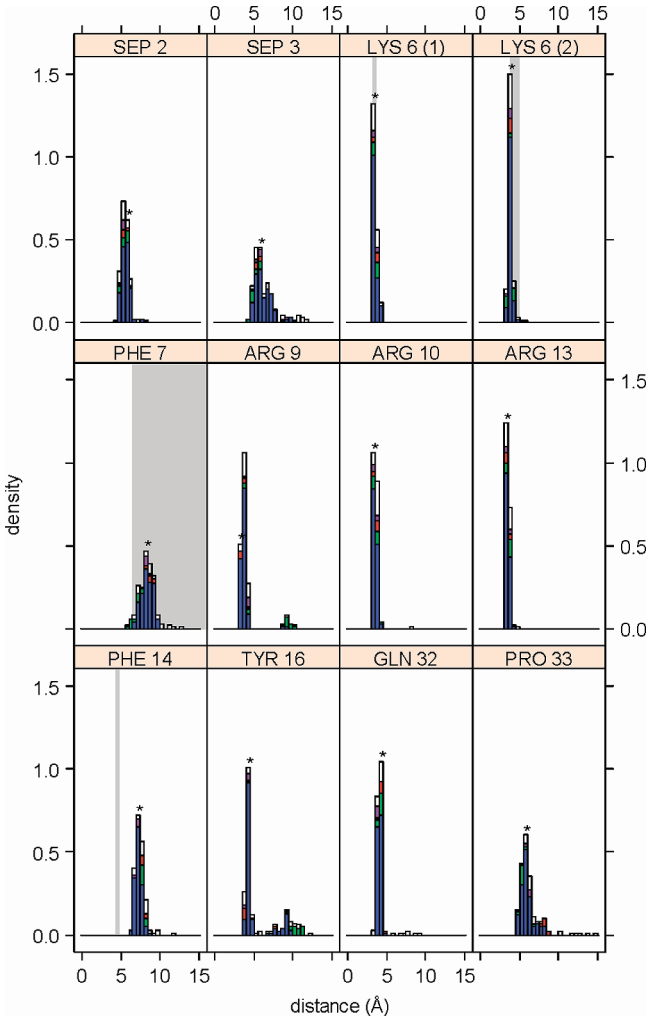


Figure 4. Distributions of statherin−HAp distances among the 200 top-scoring structures. Distances are measured from the nearest P atom of HAp to the P in Sep, the N_ζ in Lys, the center of the aromatic ring in Phe, the nearest N_η in Arg, the hydroxyl O in Tyr, the N_ϵ in Gln, and the N in Pro. For Lys6, the distance to the nearest and next-nearest HAp P is shown. Grey regions indicate distances measured experimentally by solid-state NMR,[43−45] and bar colors denote decoy clusters. Stars indicate bins containing model 1.

Figure 3 shows scores of the top quartile of the 80 000 decoys as a function of the distance from an arbitrary starting position. The sampling is extensive, and several local minima emerge. Of the 200 top-scoring decoy structures, 171 comprise four clusters using a cluster radius of 2.5 Å rmsd. Clustering statistics gathered periodically during the simulations indicate that the main clusters had not changed significantly during the last 40% of the run. The four largest clusters contain 138, 17, 9, and 7 members, and members of the four largest clusters are indicated with colors in Figure 3. The largest cluster includes the best scoring structure. The clustering indicates that the simulation converges on several related structures, predicting specific protein−surface orientations with favorable conformational entropies about the global energy minima.

Nine residues (Sep2, Sep3, Lys6, Arg9, Arg10, Arg13, Tyr16, Gln32, and Pro33) appear repeatedly at the interface in the top-scoring structures. Figure 4 shows the distribution of several statherin atom-HAp atom distances among the 200 top-scoring structures for these nine residues plus residues for which distances have been measured experimentally (Lys6, Phe7,

(87) Hubbard, S. J.; Thornton, J. M. Department of Biochemistry and Molecular Biology, University College London: London, U.K., 1993.
(88) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379−400.
(89) DeLano, W. L. DeLano Scientific: Palo Alto, CA, U.S.A., 2002.

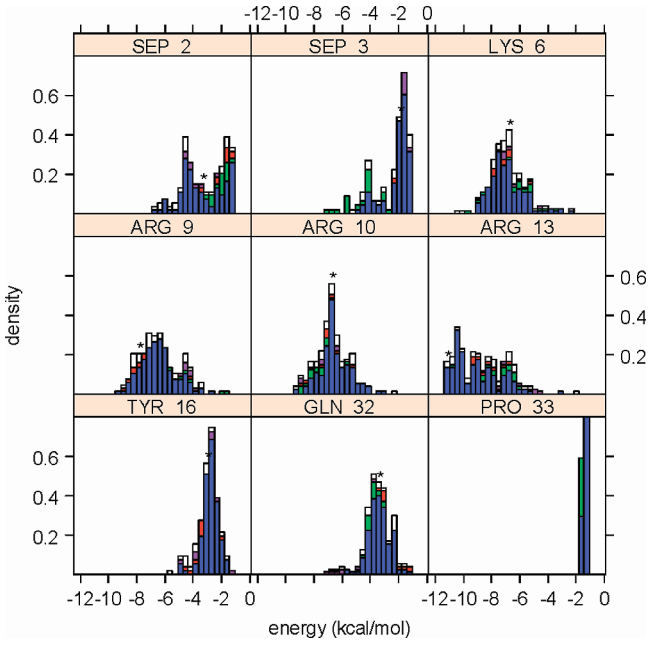


Figure 5. Statherin−HAp interaction energy distributions among the 200 top-scoring structures for selected residues. Bar colors denote decoy clusters, and stars indicate bins containing model 1.

Phe14). These histograms reveal a broad similarity among structures within clusters as well as between clusters. Distances to Lys6, Arg10, Arg13, and Gln32 are all under 5 Å and narrowly distributed. Arg9 is similar, except that structures in cluster 2 show larger distances. The Sep residues and Pro33 have broader distributions from 4 to 8 Å and a few decoys with larger distances.

For Lys6, the nearest and next-nearest N_ζ−P distance are both plotted to allow comparison with models proposed from solid-state NMR measurements (gray regions in Figure 4). Both distances are within the experimental range for most decoys; however the decoys show more equal nearest and next-nearest distances compared to that proposed by Gibson et al.[44] The Phe residues have also been experimentally probed. The simulation decoys show broad distributions, with most of the Phe7 aromatic rings being over 6.5 Å from the surface in agreement with the NMR measurements.[43] However, the structures have Phe14 distances a few Ångstroms above the experimental range in all models.

Figure 5 shows distributions of the protein−surface intramolecular energies over the 200 top-scoring models for each of the residues near the interface. The basic residues (particularly Arg13) dominate the energetics, although there is considerable variation in individual energies between models. The Sep residues both have bimodal distributions with some models having large energies and some having almost negligible energies. For example, the cluster 2 models have excellent Sep3 energies of 6−8 kcal/mol which perhaps are achieved at the expense of Sep2 energies which are limited to 1−3 kcal/mol.

In a typical Rosetta simulation, a single structure with both the lowest score (global free energy minimum) and best clustering (most favorable conformational entropy) suggests accurate results.[68] In this case, there is such a structure that scores better than any other and clusters with 138 of the 200 top-scoring decoys. While many of the models are plausible by energy, the size difference of the top clusters coupled with

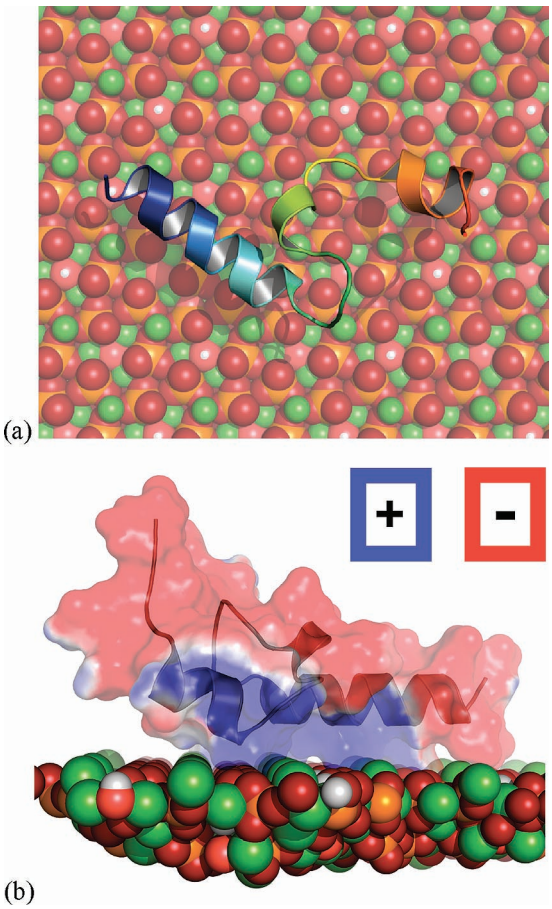


Figure 6. Orientation of model 1. (a) Top view. (b) Side view showing the shape complementarity and charge distribution of statherin (calculated in isolation by PyMOL's vacuum electrostatics module).

extensive sampling carried out suggests the relevance of this single structure. With limited experimental constraints available, it is impossible to determine which model(s) is (are) likely to be experimentally relevant. To simplify the analysis we focus on that single, dominant structure and refer to it as model 1; later we return to the other models to identify similar and distinguishing characteristics.

Figure 6 shows top and side views of the global orientation of model 1. The system buries 849 Å^2 of statherin's 4343 Å^2 of solvent-exposed surface area, and the N-terminal helix is pointed from N-to-C in a direction about 30° to the right of the negative **a** axis (Figure 1a). In Figure 6b, the semitransparent surface surrounding the cartoon (backbone) representation of statherin represents the van der Waals surface of the protein and reveals the interfacial shape complementarity. To highlight the global charge distribution, Figure 6b also shows the electrostatic potential of isolated statherin superimposed on model 1 as calculated by PyMOL's vacuum electrostatics module.[89] The HAp surface bears a net negative charge and complements the positively charged binding surface of model 1.

In Figures 4 and 5, stars mark the bins containing model 1 measurements. Table 1 further details the dominant energy terms and percent buried surface area for selected residues at the statherin−HAp interface in model 1. Hydrogen-bonding energy is the largest favorable term followed by van der Waals and then electrostatic energies, offsetting the unfavorable desolvation

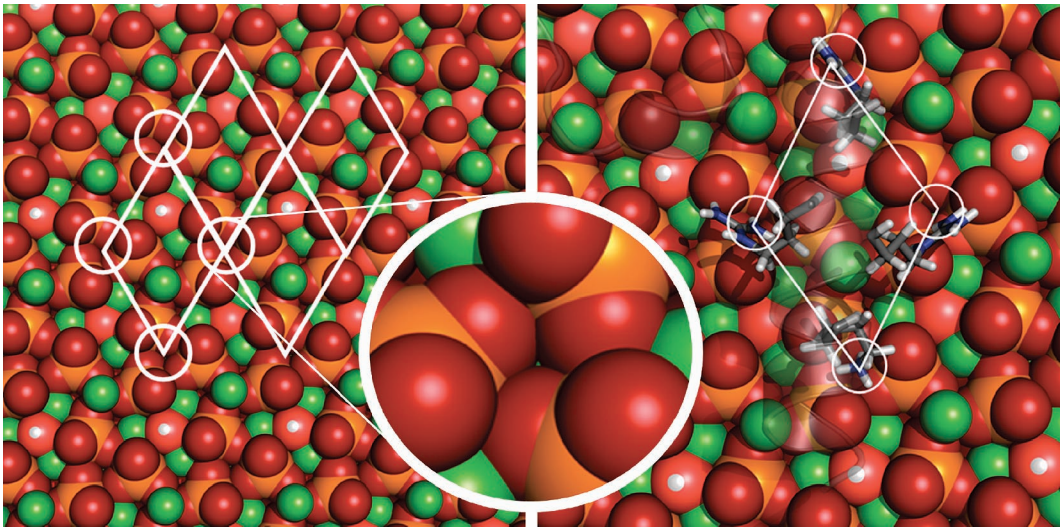


Figure 7. Left: Repetition of the IPOT pattern across the HAP surface (white parallelograms). Inset: Close-up view of an IPOT. Right: Four basic amino acids which comprise the IPOT recognition motif.

Table 1. Residue−Surface Interaction Energies (kcal/mol) and Percentages of the Residues' Solvent-Accessible Surface Areas Which Are Buried by Adsorption for the Best Statherin−HAp Structure (Model 1)

amino acid	van der Waals	solvation	hydrogen bonding	coulomb	total	% desolvated
Sep2	−1.9	1.2	−2.5	−0.4	**−3.5**	66
Sep3	−0.9	0.4	0.0	−1.3	**−1.8**	71
Lys6	−3.1	4.8	−5.7	−3.0	**−6.9**	93
Arg9	−2.8	3.5	−7.1	−1.4	**−7.8**	43
Arg10	−3.9	4.6	−5.5	−1.9	**−6.6**	93
Arg13	−4.4	5.7	−10.5	−2.1	**−11.2**	91
Tyr16	−1.8	1.0	−2.1	−0.1	**−2.2**	57
Gln32	−2.2	1.6	−2.7	−0.1	**−3.4**	85
Pro33	−0.8	0.3	0.0	−0.1	**−0.7**	86
total	**−21.7**	**23.2**	**−36.1**	**−10.5**	**−45.0**	

penalty. The individual residue energies which underlie the molecular recognition phenomenon are noted in detail below.

The RosettaSurface simulations reveal a promising molecular recognition motif for statherin binding on HAp. In the left pane of Figure 7, white circles encapsulate an atom cluster we denote as the Interstice of the Phosphate-Oxygen Triad (IPOT). IPOTs occur wherever three phosphate oxygens meet on the HAp surface, forming a negative charge center with a trivalent hydrogen bond capacity. The pocket has a depth of ∼1.6 Å and a diameter of ∼2.1 Å (inset). In Figure 7, the white parallelograms (left) show the periodicity of the IPOT motif replicated across the HAp surface; the dimensions measure approximately 9.4 Å by 9.4 Å in a regular parallelogram with an acute angle of 60°.

Remarkably, in model 1 this geometry matches perfectly that displayed by the four basic amino acids of statherin (Figure 7, right). All four basic amino acids, Lys6, Arg9, Arg10, and Arg13, fall directly into an IPOT. These residues provide the largest contribution to every favorable energy term (van der Waals, electrostatics, and hydrogen bonding) and cumulatively the largest contribution to the total binding energy (Table 1). Lys6 is positioned such that all polar side-chain hydrogens are donated to HAp phosphate oxygens (Figure 8a). Arg9 and Arg10 both form hydrogen bonds to HAp phosphate oxygens with two of their three side-chain nitrogen atoms (Figure 8a,b), similar to the "arginine fork" observed in the recognition of nucleic acid backbone phosphates.[90] Arg13 exhibits the strongest hydrogen bond energy and the strongest van der Waals interaction, combining to create the strongest overall interaction (−11.2 kcal/mol). All hydrogen bond donors and acceptors in the interface are fulfilled explicitly or accessible to solvent, except for one donor from the N_η of Arg13. Lys6, Arg10, and Arg13 are over 90% desolvated; Arg9 retains about half of its solvent accessibility (Table 1). Figure 9 shows the shape complementarity as measured using FADE; the basic amino acids fit well into the IPOT pockets and score very highly.

Previous studies[40] have suggested a special HAp recognition role for statherin's N-terminal Sep residues. Considering that the adsorption of Sep residue phosphates to the growing HAp surface would closely mimic the deposition of phosphate ions from solution, Sep residues could be partially responsible for statherin's crystal growth inhibitory capabilities. In model 1, both of the phosphoserines (Sep2, Sep3) interact strongly with the HAp surface (Table 1, Figure 5). Figure 8c shows the direct interaction of both phosphoserine residues with the HAp surface. Sep3 is striking in that its tetrahedral phosphate coordinates a HAp Ca^{2+} atom in a fashion similar to that of phosphate groups in the bulk of the HAp crystal. The favorable portion of this interaction is dominated almost entirely by electrostatics and to a lesser extent by van der Waals forces (Table 1). The interaction of Sep2 is mediated in a different and rather intricate manner. Sep2 is the only residue in which an intermolecular hydrogen bond pair has statherin as the proton acceptor and HAp as the donor. One of the Sep2 phosphate oxygens accepts a hydrogen bond from a HAp hydroxyl atom in a moderately favorable interaction. The Sep2−HAp interaction is further reinforced by a moderate attractive Lennard−Jones component and a weak electrostatic interaction.

Tyr16 at the end of the N-terminal helix and Gln32 in the C-terminus (Figure 8a,b) are moderately strong donors and interact with moderate attractive van der Waals forces (Table 1). Pro33's only favorable energy term is a relatively weak

(90) Calnan, B. J.; Tidor, B.; Biancalana, S.; Hudson, D.; Frankel, A. D. *Science* **1991**, *252*, 1167−1171.

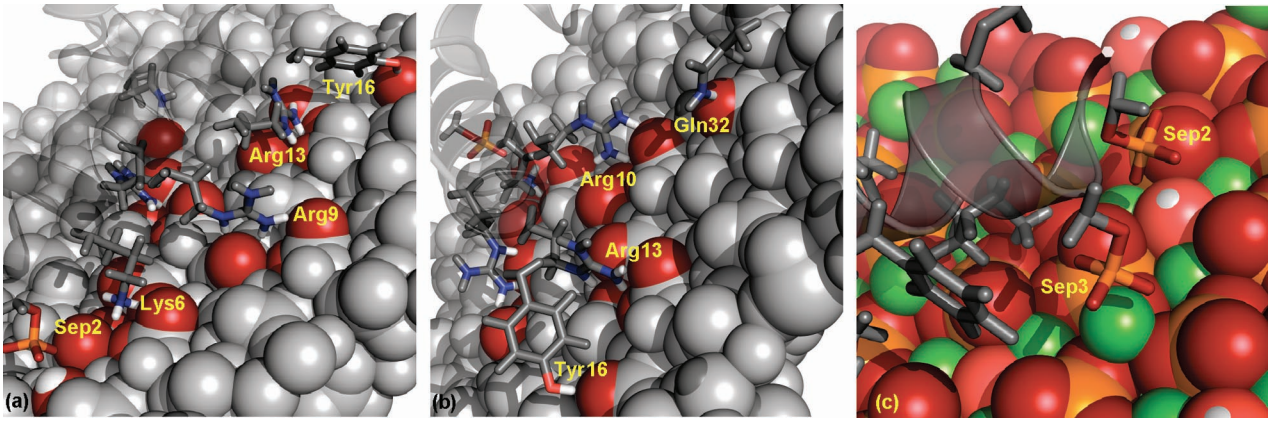


Figure 8. Selected structural details of model 1. (a) N-terminal view and (b) C-terminal view of basic, aromatic, and phosphoserine residues with hydrogen-bonding HAp atoms in color. (c) Sep2 makes hydrogen bonds and Sep3 chelates a Ca^{+2} ion (green).

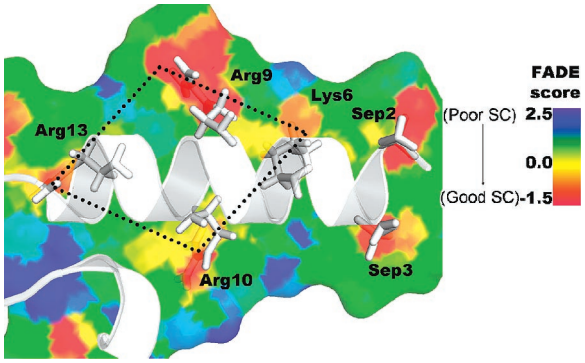


Figure 9. Map of interface FADE scores.[86] The four basic residues of the IPOT recognition motif (outlined in dots) and the two phosphoserines show shape complementarity typical of protein−protein interfaces.

attractive van der Waals interaction. This prediction agrees with experimental observations that the C-terminus interacts with HAp more strongly than the central region but weaker than the N-terminus of statherin.[40,42] Gln32 interacts with an IPOT, and its binding might help constrain the global orientation of the statherin. For all three residues, Tyr16, Gln32, and Pro33, the electrostatic contribution to the total binding energy is negligible compared to hydrogen bond contribution.

The difference in interface quality between the N-terminal region and the rest of the protein is also shown in the FADE scores, which measure shape complementarity. For reference, the FADE scores in the set of protein−protein interfaces from the docking benchmark set 2.0[91] average -0.103 ± 0.038 per lattice point. In model 1 of statherin on hydroxyapatite, the average FADE score for the entire interface of model 1 is poor, at -0.014 per lattice point. However, the score for the N-terminal region is -0.062 per lattice point, almost within the standard deviation range of protein−protein interfaces.

Many of the structures in the set of 200 top-scoring decoys share recognition features with model 1. As seen in the histograms of distances and energies, most models contain the IPOT-recognition motif with each of the four basic residues binding an IPOT pocket. Exceptions are the cluster 2 structures, in which Arg9 does not make contact, and the cluster 4 structures, in which Arg13 and Arg9 share a single IPOT in a less-optimal fashion. Figure S4 in the Supporting Information shows the orientational distribution of the top 200 decoys. Cluster 2 and 3 structures are rotated in the HAp [001] plane by 60° and 90°, respectively, relative to cluster 1. In these different orientations, it can be difficult to make energetically favorable interactions with all of the surface-binding groups, and there is a distribution of better and less-optimal interactions (Figure 5). A top-view comparison of the four model structures (Figure S5) and coordinates of the best-scoring structures from each of the four largest clusters (models 1−4) are available in the Supporting Information.

Discussion

Many studies have shown diverse face and phase recognition of an inorganic solid by proteins (e.g., refs 92−95). Molecular recognition of a protein−solid surface system may differ from that of a protein−protein or protein−small molecule system. Proteins draw from the chemical complexity of 20 unique amino acids that can be arranged in numerous combinations in primary and tertiary space, and protein topography is geometrically complex, usually not flat or repetitive. Therefore, a protein−protein recognition motif is rarely a repetitive pattern, but rather a complex arrangement of residues of side-chain lengths, hydrophobicities, pK_a's, and ionic strengths exquisitely complemented by the evolved binding partners. In contrast, solid crystal surfaces typically present a pattern whose chemistry and geometry is repetitive on the scale of a unit cell. Therefore, one may expect a surface-binding protein to possess a complementary pattern in the tertiary arrangement of its side chains.[96] Furthermore, biomaterials such as silica, calcite, and HAp typically terminate ionically and yield a net surface charge. If a protein is to complement the charge at the interface, it must bear an array of proximal like-charges in its binding domain. Such arrangements have been observed in antifreeze proteins[97] and osteocalcin.[98]

Statherin's array of proximal basic residues of one lysine and four arginines in residues 6−13 (KFLRRIGR) with no interven-

(91) Mintseris, J.; Wiehe, K.; Pierce, B.; Anderson, R.; Chen, R.; Janin, J.; Weng, Z. *Proteins* **2005**, *60*, 214−6.
(92) Jia, Z.; DeLuca, C. I.; Chao, H.; Davies, P. L. *Nature* **1996**, *384*, 285−8.
(93) Fujisawa, R.; Kuboki, Y. *Biochim. Biophys. Acta* **1991**, *1075*, 56−60.
(94) DeOliveira, D. B.; Laursen, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 10627−10631.
(95) Belcher, A. M.; Wu, X. H.; Christensen, R. J.; Hansma, P. K.; Stucky, G. D.; Morse, D. E. *Nature* **1996**, *381*, 56−58.
(96) Mann, S. *Nature* **1988**, *332*, 119−124.
(97) Liou, Y. C.; Tocilj, A.; Davies, P. L.; Jia, Z. *Nature* **2000**, *406*, 322−4.
(98) Flade, K.; Lau, C.; Mertig, M.; Pompe, W. *Chem. Mater.* **2001**, *13*, 3596−3602.

ing acidic residues would be expected to destabilize the secondary and tertiary structure. However, residues 2−12 are measured to be helical by NMR,[42] and Rosetta predicted residues 3−15 as helical.[48] When the above sequence is threaded on an α-helical backbone, the basic residues are positioned perfectly to mimic the periodicity of acidic phosphate clusters on the HAp surface. In the RosettaSurface models, this lattice-matching provides a significant contribution to the free energy of adsorption. While the thermodynamic contribution of the basic residues has been measured,[46] to our knowledge this is the first structural identification of this molecular recognition motif. Most previous studies have focused on the Sep residues, although it has been previously noted that the Sep residues can be substituted by similarly charged Asp residues and still retain function.[40]

In previous Rosetta studies, alternate force fields (choice of energy functions, parametrizations, and weightings) have been tested using known protein structures from the Protein Data Bank.[49] After using unvalidated selections of force field parameters from the literature, it is encouraging that the results are broadly consistent with known experimental data. Early experiments on statherin fragments[40] and more recent dynamic NMR measurements[42] identify the 15 N-terminal residues as the primary contributors to the HAp adsorption free energy; in fact, the 15-mer is sufficient for α-helix formation, HAp binding, and inhibition of HAp growth.[40] Similarly, in the final models, the N-terminal residues dominate the interface and contribute the most to the binding energy while the intervening and C-terminal regions contribute weakly. Single amino-acid experiments have shown that Sep, Arg, Lys, Asp, Glu, Tyr, and Gln interact strongly with HAp;[99−104] these residue types comprise the majority of the 15-residue N-terminal fragment. Furthermore, Lys was shown to inhibit HAp crystal growth,[100] which is thought to be the primary function of statherin *in vivo*, further implicating Lys' role in HAp binding. Finally, calorimetric data of mutated statherin fragments verify the importance of the basic amino acids (Lys6, Arg9, Arg10, Arg13) in HAp binding.[46] RosettaSurface both samples and selects for decoys that are in agreement with these experimental results.

The current experimental picture of the statherin−HAp system is one of low resolution, and many of the predicted structures broadly match this picture despite differences up to a 20 Å rmsd. Despite the fact that hydroxyapatite's [001] surface does not exhibit any formal rotational symmetry, the existence of different clusters of low-energy structures rotated 60° and 90° reveals similar features on the surface across different rotations. New measurements of the residues with broader distance distributions (*e.g.*, Sep2, Sep3, Tyr16, and Pro33) would help to discriminate among the current models, but it will remain difficult to distinguish between the set of top structures due to the pseudo-rotational symmetry of the surface.

The simple model predicted here captures an encouraging amount of the previous experimental data, but there are many aspects of the system still not addressed fully. The statherin backbone conformation was determined by *de novo* predictions and held fixed during docking. While the structure agrees with solid-state NMR measurements of backbone distances in the adsorbed state, there are several angular measurements that are not consistent with the predicted backbone.[48] Statherin is known to undergo a structural transition upon adsorption, so it is reasonable that accurate backbone structures are only achievable in the presence of the hydroxyapatite. Simultaneous backbone and rigid-body optimization of statherin was beyond the scope of this study but would be an important goal for future studies. Backbone flexibility might resolve the suboptimal interactions of some side-chain groups (*e.g.*, Sep3 in model 1) while retaining the optimal interactions of other groups. One of the currently dominant clusters might be better at achieving a higher degree of electrostatic and shape complementarity, especially beyond the N-terminal region, resolving the remaining uncertainties in the protein orientation. Also, some unraveling of the end of the first helix might allow the structure to meet the experimentally measured Phe14−HAp distance and allow solvent access to the one unsatisfied hydrogen-bonding group in Arg13 while maintaining the IPOT recognition motif.

Additional remaining questions about the statherin−hydroxyapatite system concern the HAp surface. In this study, we assumed that statherin binds to a clean [001] monoclinic face of HAp terminated in a balance of Ca^{2+} and phosphate groups, as expected in many *in vitro* experiments. However, the crystallographic binding face of biological HAp for statherin has not been unambiguously determined, and crystallographic faces ([010], [100], etc.), lattice types (hexagonal rather than monoclinic), terminations, and step edges or defects may play a role in statherin binding. For example, the deposition of one additional Ca^{2+} atom at the appropriate location would provide better electrostatic and van der Waals interactions for one of the Sep residues in model 1. Future simulations on this system would include some of these competing scenarios to address these open questions.

The RosettaSurface method developed here has several advantages over previous approaches. The multistart MCM algorithm is an efficient and parallel method to sample many local minima in the free energy landscape of the protein−surface system. The energy function, while needing validation for protein−surface interactions, has been broadly successful in a variety of protein structure prediction problems. The description is atomically detailed allowing direct examination of structural features and a detailed decomposition of energetic components. The reversible, evolved statherin−hydroxyapatite system revealed a recognition motif, but this is not expected in the general case: prediction of many protein−surface systems could be confounded by the nonequilibrium nature of the adsorption process and the existence of multiple conformational states on the surface. Furthermore, lateral protein−protein interactions are likely to play a role, perhaps even for the statherin−HAp system,[39] requiring additional representations and sampling. Therefore, development of a general tool for protein−surface interactions remains a challenging problem.

Conclusion

RosettaSurface simulations have captured some important features of the statherin−hydroxyapatite system. Importantly,

(99) Moreno, E. C.; Kresak, M.; Hay, D. I. *Calcif. Tissue Int.* **1984**, *36*, 48−59.
(100) Koutsopoulos, S.; Dalas, E. *J. Colloid Interface Sci.* **2000**, *231*, 207−212.
(101) Koutsopoulos, S.; Dalas, E. *J. Cryst. Growth* **2000**, *217*, 410−415.
(102) Koutsopoulos, S.; Dalas, E. *J. Cryst. Growth* **2000**, *216*, 443−449.
(103) Kresak, M.; Moreno, E. C.; Zahradnik, R. T.; Hay, D. I. *J. Colloid Interface Sci.* **1977**, *59*, 283−292.
(104) Tanaka, H.; Miyajima, K.; Nakagaki, M.; Shimabayashi, S. *Chem. Pharm. Bull.* **1989**, *37*, 2897−2901.

it supported previous experimental measurements, assigned residue-specific energy contributions to the known HAp interaction domain of statherin, and identified the IPOT recognition motif. Knowledge of natural biomaterial recognition will help to rationally design novel protein−surface interactions and harness the medical and engineering potential of such systems. Protein−surface interactions present a unique challenge to the structural biology community in that no current experimental method affords the ability to solve the molecular structures. Molecular modeling currently provides the only means to investigate the protein−surface interface with atomic resolution. Combinations of experimental and computational approaches are promising routes to reveal the molecular structure of proteins at solid surfaces.

Acknowledgment. We thank Gil Goobes, Gary Drobny, and Pat Stayton of the University of Washington for useful discussions and interpretation of the solid-state NMR data and the statherin−HAp system. Ora Schueler-Furman provided *ab initio* predictions of the statherin structure in advance of publication. Protein structure figures were generated with PyMOL.[89] We thank Arvind Sivasubramanian, Maria Makrodimitri, and Tom Woolf for reviewing the manuscript. K.M. was partially supported by the ACS Petroleum Research Fund (41233-G5), and E.K. was supported by a Provost's Undergraduate Research Award from Johns Hopkins University. This work was funded by the Arnold and Mabel Beckman Foundation through a young investigator grant.

Supporting Information Available: Rotamer statistics (Figure S1) and packing parameters (Table S1) for phosphorylated serines, symmetry-related protein movement (Figure S2), orientational distribution of the top-scoring structures (Figure S3), top view of the four model structures (Figure S4), complete ref 81, and coordinates of representative predicted structures of statherin on HAp (files model1.pdb, model2.pdb, model3.pdb, and model4.pdb). This material is available free of charge via the Internet at http://pubs.acs.org.

JA074602V