Biophysical Journal, Volume 96

Supporting Material

Solution- and adsorbed-state structural ensembles predicted for the statherin-hydroxyapatite system

David L Masica and Jeffrey J Gray

Supplementary material for:

"Solution- and adsorbed-state structural ensembles predicted for the statherin-hydroxyapatite system"

David L Masica and Jeffrey J Gray

Materials and Methods

Statherin. We built an extended molecular structure of human salivary statherin using PyMol (1). The atomic parameter set is the same as in Makrodimitris *et al* (2), except in this study the γ oxygen atom of phosphoserine (Sep) is modeled as an esterified oxygen (3).

Hydroxyapatite. We built monoclinic hydroxyapatite crystals using CrystalMaker (4). Appropriate cuts were made to expose flat (001), (010), and (100) surfaces with approximately neutral (mixed charge) terminations (see figure 6). The atomic parameter set is the same as in Makrodimitris *et al* (2), except in this study all HAp surface atoms are included when calculating the solvent exclusion term (5).

The Algorithm. The algorithm developed for this study (see flow chart in figure S1) is based on the multiscale Metropolis Monte Carlo-plus-minimization (MCM) implementation of the Rosetta docking and folding strategies (6, 7, 8). The protocol implements an initial, fast, low-resolution step to generate diverse protein folds. Subsequent high-resolution steps were optimized to create mostly clash-free protein-like structures in a short amount of time. The protocol energy minimizes the solution-state protein structure before adsorbing the protein to the solid surface, thus generating large ensembles of candidate structures (decoys) in both the solution and adsorbed states. There are a total of 5 "refinement cycles" (see below) implemented to generate each solution- and adsorbed-state decoy; for each decoy, adsorption occurs at random between refinement cycles one and five. This choice allows for approximately equal sampling at all stages of refinement in both the solution and adsorbed states when a large number of decoys are generated.

The creation of each decoy begins with the rapid collapse of a fully extended protein molecule using a fragment assembly (FA) protocol. The FA protocol is the same as in Simons *et al.* (8), except in this study we removed the radius of gyration term to allow sampling of extended structures. After FA, the backbone is optimized with six "shear" and ~250 "small" moves (9) coupled with conjugate-gradient (10) and line minimization respectively. At all stages thus far

the structural model is comprised of an all-atom protein backbone and a single pseudo atom to represent each side chain (8).

Next, all-atom (including hydrogen) side chains are built from a backbone-dependent rotamer library (11) using a simulated annealing protocol (12), and the protein undergoes $n \leq 5$ cycles of backbone refinement (see figure S1, right). Each cycle of refinement includes "outer-"and "inner-loop optimization". During outer-loop optimization, a sequence of perturbing small-, shear-, and crank-moves (9) are applied to the protein backbone, each followed by conjugate-gradient minimization (10). 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 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 (7) high-resolution sequence. In this study, a sequence of 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, side-chain, and rigid-body optimization on the surface.

Solution-state coordinates are output immediately prior to the formation of the adsorbed state. After adsorbed-state refinement is complete, RosettaSurface implements one round of coupled rotamer packing and minimization (13) at all side-chain positions and outputs the adsorbed-state coordinates.

The all-atom energy function is similar to that used previously (2) with the following modifications: (i) To account for the affect of the highly charged binding domain on folding, protein self electrostatics are included using a simple distance-dependent-dielectric model (14). (ii) Because statistically derived energy terms and coefficients (weights) had little effect on final decoy discrimination, both adsorbed- and solution-state decoys were scored using a nonweighted physical potential comprising van der Waals (7), electrostatic (14), hydrogen bond (15), and implicit solvent (5) interactions. (iii) For ssNMR constrained simulations, a harmonic potential with a spring constant of 10 Rosetta Energy Units (REU)/Å (or °) was used to bias predictions toward satisfying all 15 published ssNMR distance and angle measurements (16-21).

Structural analysis. The structural designations "Helix" and "Turn" were assigned using the DSSP (22) definitions; classification relied on Rosetta's hydrogen-bond function (15) 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. Side-chain solvent-accessible surface area was calculated using Naccess (23) with van der Waals radii adjusted to match those in the Rosetta force field. Residues were considered "bound" if they had a nonzero contribution to total adsorption energy.

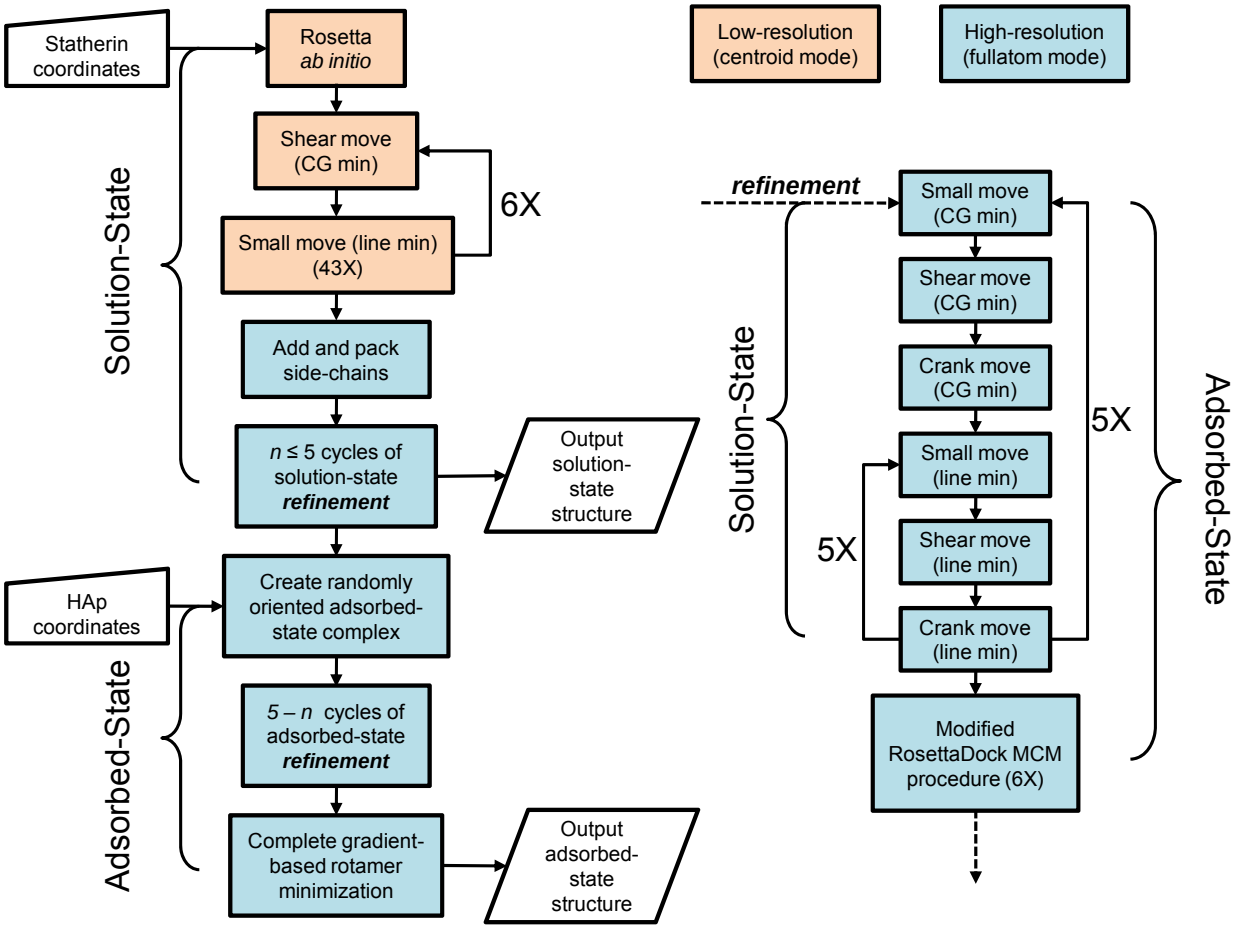

FIGURE S1 Flowchart for the RosettaSurface algorithm. *CG min* = conjugate gradient minimization (9); *line min* = line minimization. *small move* = perturbations of randomly selected (φ,ψ) pairs. *shear* move = perturbation of randomly selected φ angle with compensatory perturbation of equal and opposite 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.

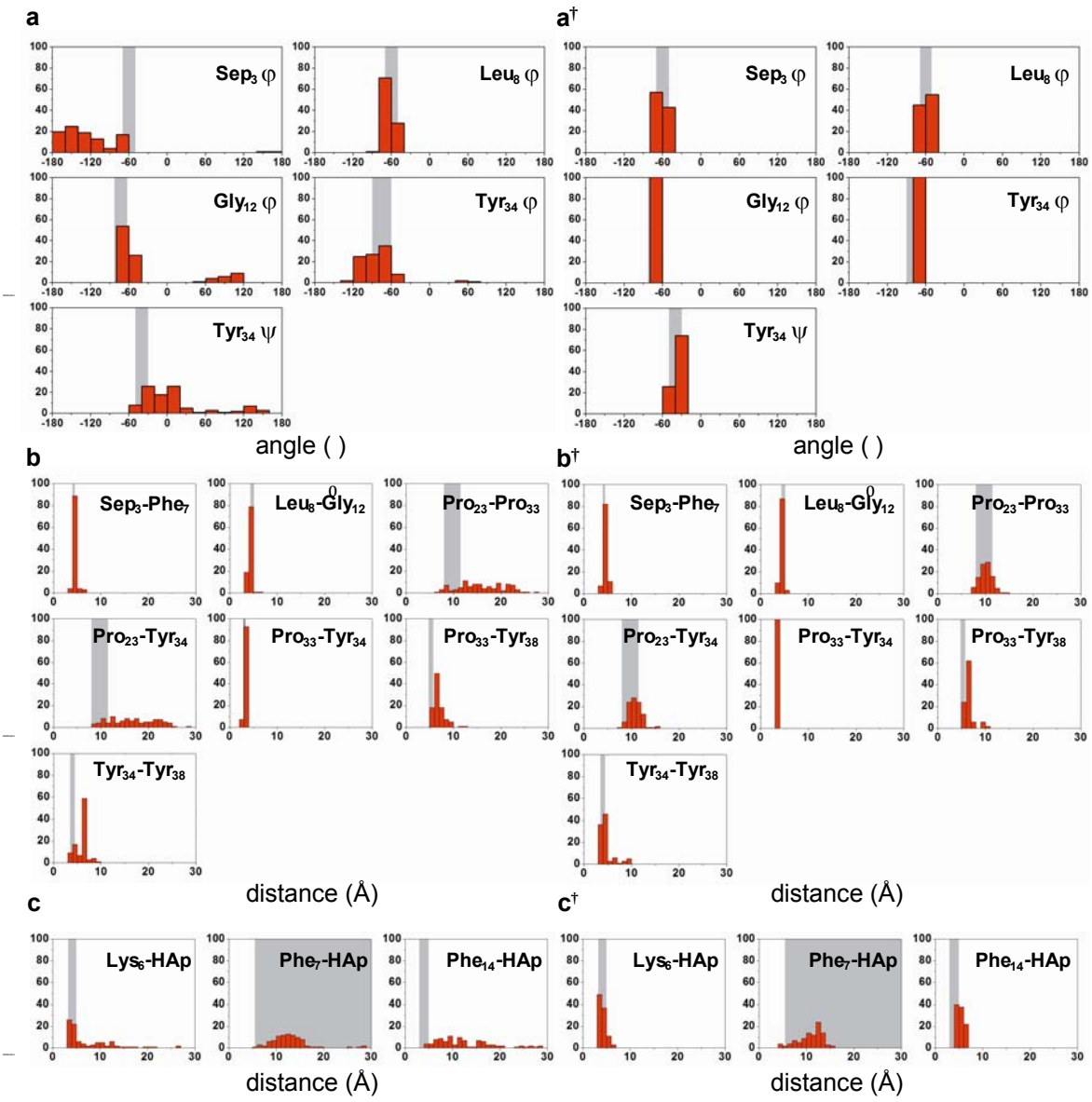

FIGURE S2 Comparison of RosettaSurface predictions and all previously published solid-state NMR measurements. Histograms from unbiased (*a, b, and c*) and NMR-biased ($a^{\dagger}$, $b^{\dagger}$, and $c^{\dagger}$) simulations for statherin adsorbed to the 001 monoclinic face of HAp. (a and $a^{\dagger}$) Statherin torsion angles, (*b* and $b^{\dagger}$) statherin intramolecular distances, and (*c* and $c^{\dagger}$) statherin-HAp intermolecular distances. Red histograms show the distributions for the 100 top-scoring decoys and gray bars show the NMR measurement (width depicts experimental error).

RosettaSurface predictions agreed with measured distances and angles indicative of helical structure in the N-terminal binding domain such as the φ angles of Leu8 and Gly12 and distance between Sep3 & Phe7 and Leu8 & Gly12. The intermolecular distance from the Lys6 side-chain

nitrogen to the closest HAp phosphorus is in close agreement with experiment as is the short-range Pro33-Tyr34 distance measurement. Moderate agreement was observed at the φ and ψ angle of Tyr34; with that ψ angle showing some dispersion. RosettaSurface predictions were close to the measured Tyr34-Tyr38 and Pro33-Tyr38 distance measurements. However, the following measurements disagreed: the long-range intramolecular distances Pro23-Pro33 and Pro23-Tyr34, and the intermolecular distance between Phe14 and the nearest HAp phosphorus. These distances vary between top-scoring models in our simulation and no single measurement is predicted.

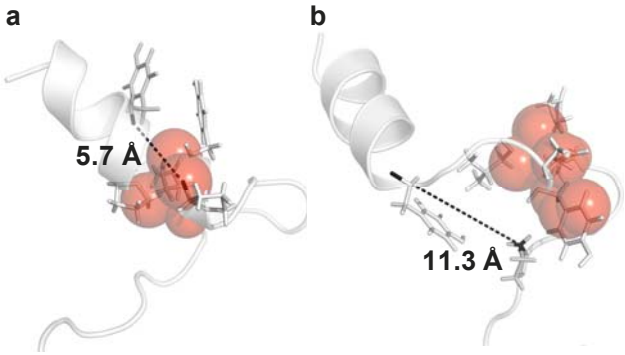

FIGURE S3 Clashes created in statherin's C-terminal segment when NMR constraints are met between (*a*) Pro33-Tyr38 and (*b*) Pro23-Tyr34. Residues that contain clashing atoms (red spheres) and isotopically labeled residues are shown in stick representation. Labels indicate distance (Å).

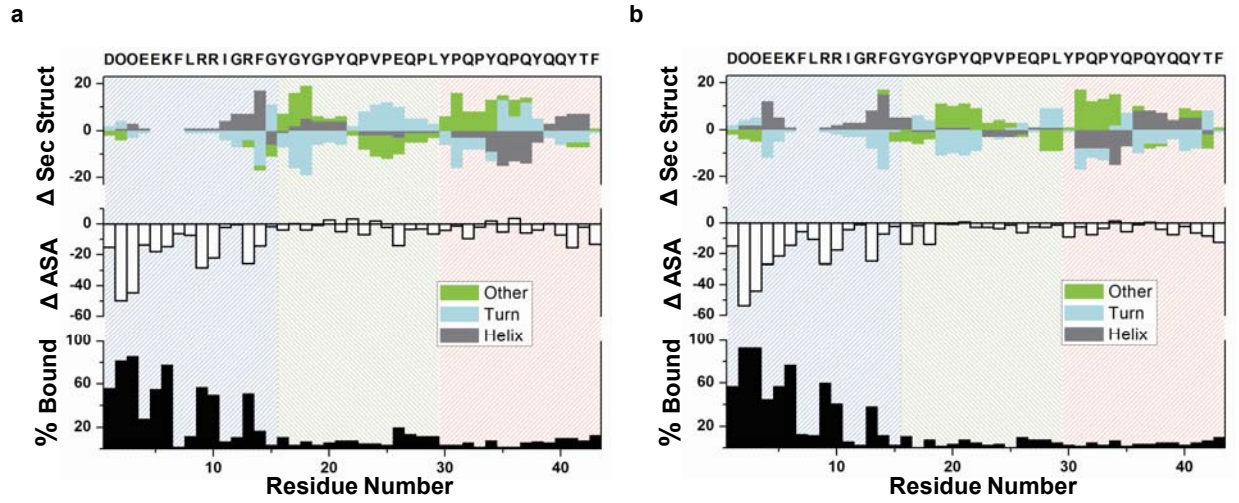

FIGURE S4 Three adsorption phenomena plotted against residue number for the 100 top-scoring decoys at the 010(a) and 100(b) surfaces. See figure 2a caption for a complete description of these three phenomena.

There is a moderate difference in binding with respect to the three surfaces, most notably between the (001) and (100) surfaces (lower panels, figures 2a and S3b respectively). Residues Sep2, Sep3, Lys6, Arg9, Arg10, and Arg13 bind with approximately equal frequency, between 48% and 69%, at the (001) surface. Conversely, at the (100) surface, Arg13 only adsorbed 38%

of the time (58% at the (001) surface), and Sep2 & Sep3 each adsorbed 93% of the time (~68% at the (001) surface). Furthermore, Asp1 bound approximately as or more frequently than Arg9, Arg10, or Arg13 at both the (010) and (100) surfaces (lower panel, figures S3a and S3b respectively). And in the case of the (100) surface, Glu4 binds more frequently than Arg10 or Arg13.

References

1. DeLano, W. L. 2002. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA.
2. Makrodimitris, K., D. L. Masica, E. T. Kim, and J. J. Gray. 2007. Structure prediction of protein-solid surface interactions reveals a molecular recognition motif of statherin for hydroxyapatite. J Am Chem Soc 129:13713-13722.
3. Foloppe, N., A. D. MacKerell., Jr. 2000. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. Journal of computational chemistry 21:86-104.
4. Palmer, D. 2005. CrystalMaker. Version 7.1. CrystalMaker.
5. Lazaridis, T., M. Karplus. 1999. Effective energy function for proteins in solution. Proteins: Structure, Function, and Genetics 35:133-152
6. Bradley, P., K. M. S. Misura, and D. Baker. 2005. Toward High-Resolution de Novo Structure Prediction for Small Proteins. American Association for the Advancement of Science. 1868-1871.
7. Gray, J. J., S. Moughon, C. Wang, O. Schueler-Furman, B. Kuhlman, C. A. Rohl, and D. Baker. 2003. Protein–Protein Docking with Simultaneous Optimization of Rigid-body Displacement and Side-chain Conformations. Journal of Molecular Biology 331:281-299.
8. Simons, K. T., C. Kooperberg, E. Huang, and D. Baker. 1997. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions. Journal of Molecular Biology 268:209-225.
9. Rohl, C. A., C. E. Strauss, K. M. Misura, and D. Baker. 2004. Protein structure prediction using Rosetta. Methods Enzymol 383:66-93.
10. Fletcher, R. 1963. Powell, MJD: A Rapidly Convergent Descent Method for Minimization. Computer Journal 6:163-168.
11. Dunbrack Jr, R. L., and F. E. Cohen. 1997. Bayesian statistical analysis of protein side-chain rotamer preferences. Protein Science: A Publication of the Protein Society 6:1661.
12. Kuhlman, B., and D. Baker. 2000. Native protein sequences are close to optimal for their structures. National Acad Sciences. 10383-10388.
13. Wang, C., O. Schueler-Furman, and D. Baker. 2005. Improved side-chain modeling for protein-protein docking. Protein Science 14:1328-1339.
14. Warshel, A., S. T. Russell, and A. K. Churg. 1984. Macroscopic Models for Studies of Electrostatic Interactions in Proteins: Limitations and Applicability. Proceedings of the National Academy of Sciences 81:4785-4789.
15. Kortemme, T., A. V. Morozov, and D. Baker. 2003. An Orientation-dependent Hydrogen Bonding Potential Improves Prediction of Specificity and Structure for Proteins and Protein–Protein Complexes. Journal of Molecular Biology 326:1239-1259.

16. Raghunathan, V., J. M. Gibson, G. Goobes, J. M. Popham, E. A. Louie, P. S. Stayton, and G. P. Drobny. 2006. Homonuclear and heteronuclear NMR studies of a statherin fragment bound to hydroxyapatite crystals. J. Phys. Chem. B 110:9324-9332.
17. Goobes, G., R. Goobes, O. Schueler-Furman, D. Baker, P. S. Stayton, and G. P. Drobny. 2006. Folding of the C-terminal bacterial binding domain in statherin upon adsorption onto hydroxyapatite crystals. Proceedings of the National Academy of Sciences 103:16083-16088.
18. Gibson, J. M., J. M. Popham, V. Raghunathan, P. S. Stayton, and G. P. Drobny. 2006. A solid-state NMR study of the dynamics and interactions of phenylalanine rings in a statherin fragment bound to hydroxyapatite crystals. Journal of the American Chemical Society 128:5364-5370.
19. Gibson, J. M., V. Raghunathan, J. M. Popham, P. S. Stayton, and G. P. Drobny. 2005. A REDOR NMR study of a phosphorylated statherin fragment bound to hydroxyapatite crystals. Journal of the American Chemical Society 127:9350-9351.
20. Long, J. R., W. J. Shaw, P. S. Stayton, and G. P. Drobny. 2001. Structure and dynamics of hydrated statherin on hydroxyapatite as determined by solid-state NMR. Biochemistry 40:15451-15455.
21. Shaw, W. J., J. R. Long, J. L. Dindot, A. A. Campbell, P. S. Stayton, and G. P. Drobny. 2000. Determination of statherin N-terminal peptide conformation on hydroxyapatite crystals. J. Am. Chem. Soc 122:1709-1716.
22. Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577-2637.
23. Hubbard, S. J., and J. M. Thornton. 1993. NACCESS Computer Program. Department of Biochemistry and Molecular Biology, University College London 2.