Published in final edited form as:

Methods Enzymol. 2013 ; 532: 343–366. doi:10.1016/B978-0-12-416617-2.00016-3.

Using the RosettaSurface Algorithm to Predict Protein Structure at Mineral Surfaces

Michael S. Pacella^{*}, Da Chen Emily Koo[†], Robin A. Thottungal[‡], and Jeffrey J. Gray^{§,1}

^{*}Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, USA

[†]T. C. Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, Maryland, USA

[‡]Program in Molecular Biophysics, Johns Hopkins University, Baltimore, Maryland, USA

[§]Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland, USA

Abstract

Determination of protein structure on mineral surfaces is necessary to understand biomineralization processes toward better treatment of biomineralization diseases and design of novel protein-synthesized materials. To date, limited atomic-resolution data have hindered experimental structure determination for proteins on mineral surfaces. Molecular simulation represents a complementary approach. In this chapter, we review RosettaSurface, a computational structure prediction-based algorithm designed to broadly sample conformational space to identify low-energy structures. We summarize the computational approaches, the published applications, and the new releases of the code in the Rosetta 3 framework. In addition, we provide a protocol capture to demonstrate the practical steps to employ RosettaSurface. As an example, we provide input files and output data analysis for a previously unstudied mineralization protein, osteocalcin. Finally, we summarize ongoing challenges in energy function optimization and conformational searching and suggest that the fusion between experiment and calculation is the best route forward.

1. INTRODUCTION

The interaction of proteins with solid surfaces plays a pivotal role in diverse areas including biomaterials, bioprocessing, nanotechnology, solar energy, materials synthesis, and mineralogy. Treatments for diseases resulting from unregulated biomineralization, such as atherosclerosis, dental calculus, and kidney stone formation, could also be better developed with structural guidance. Protein–surface interactions govern the immune response and biocompatibility of implanted medical devices, the growth and morphology of crystals during biomineralization, and the self-assembly of nanoscale systems (Gray, 2004; Latour, 2008). A better understanding of these interactions would give engineers the ability to

design proteins to grow custom materials: a revolutionary, environmentally friendly approach to nanoengineering.

Despite the integral role of protein–surface interactions in these fields, a fundamental understanding of their nature has remained elusive due to limited experimental methods capable of resolving the atomic structure of proteins at solid interfaces (Cohavi et al., 2010, Collier, Vellore, Yancey, Stuart, & Latour, 2012; Goobes, Stayton, & Drobny, 2007). Computational methods offer an alternative approach to explore these interactions at an atomic scale (Latour, 2008).

A wide range of computational methods is currently employed to study protein–surface interactions. In general, these approaches can be broken down into three categories: quantum mechanical (QM)-, molecular dynamics (MD)-, and Monte Carlo (MC)-based docking. QM methods, which treat electronic degrees of freedom (DOFs) explicitly, can be accurate for determining both interaction energies and equilibrium structures of molecules adsorbed on surfaces. However, such methods are limited by poor scaling with the number of atoms in the system and performing geometry optimizations on small peptides using this approach is just beginning to become feasible (Rimola, Aschi, Orlando, & Ugliengo, 2012). Still, QM calculations on smaller systems can guide parameterization of molecular mechanics force fields or score functions (Iori, Di Felice, Molinari, & Corni, 2009; Latour, 2008). Recent QM studies have parameterized the interaction of individual amino acids with hydroxyapatite (HAp) (Corno, Rimola, Bolis, & Ugliengo, 2010) and gold (Iori et al., 2009).

MD methods integrate Newton's laws of motion for each atom in the system and follow the trajectory through time. Theoretically, if MD is run long enough, the time-average of the simulation should approach the thermodynamic ensemble average, allowing for the calculation of thermodynamic properties such as the free energy change of adsorption (Freeman, Harding, Quigley, & Rodger, 2011; Wang, Stuart, & Latour, 2008). Determining the sufficient length of time, however, is difficult and variable from system to system. Whether MD simulations have been run long enough to draw conclusions about the equilibrium behavior of proteins on solid surfaces remains an open question (Latour, 2008), particularly in determining the role of water-mediated protein–surface interactions. The advantage of MD lies in the ability to calculate kinetic properties such as adsorption rate constants from the resulting trajectories. MD also captures correlated motions in a natural way, making it ideal for systems with many correlated DOFs (such as systems containing explicit solvent).

MC-based docking algorithms employ a move set of random perturbations that update explicit DOFs in a molecular system; each move is either accepted or rejected based on the Metropolis criterion. Similarly to MD, if a sufficient number of moves are attempted in MC, the simulation average should converge to the thermodynamic ensemble average. By incorporating minimization into MC-based docking, the ability to sample from a thermodynamic ensemble is lost, but the likelihood of locating a global minimum on the potential energy surface increases dramatically (Li & Scheraga, 1987). Although it lacks the accuracy of QM methods and the kinetic information provided by MD, MC-plus-minimization (MCM) docking is perhaps the best method for locating global minima and

determining structural complementarity rapidly. MCM can also perform protein design calculations by expanding the move set to include altering the chemical identity of amino acid side chains, thus searching for a sequence that minimizes the potential energy for a given structure (Kuhlman & Baker, 2000; Kuhlman et al., 2003).

In the field of biomaterials and bioprocessing, the conventional wisdom has been that proteins adsorb to solid surfaces nonspecifically and become denatured upon their adsorption (Gray, 2004; Michaels & Matson, 1985; Ratner & Bryant, 2004). However, in the field of biomineralization, organisms demonstrate exquisite control over crystal growth and morphology using proteins (Belcher et al., 1996; Berman et al., 1993; Weiner & Dove, 2003). With this apparent contradiction in mind, we designed RosettaSurface, a computational tool that simultaneously optimizes the fold, side-chain conformations, and rigid-body orientation of a protein on a solid surface, with the goal of broadly testing the following hypothesis: Does structural recognition analogous to that found in protein–protein complexes exist in the interaction of proteins with crystal surfaces? While QM calculations give accurate energies and equilibrium structures and MD methods provide kinetic data, RosettaSurface employs MCM docking as the proposed route to test potential structural complementarity. Additionally, because of Rosetta’s success in predicting protein folding (Raman et al., 2009; Simons, Bonneau, Ruczinski, & Baker, 1999) and protein–protein interfaces (Gray et al., 2003; Sircar, Chaudhury, Kilambi, Berrondo, & Gray, 2010), the algorithms and score functions contained within that framework are logical starting points. RosettaSurface can perform *ab initio* or constraint-based calculations. Including constraints allows users to create a feedback loop between experimental constraint determination and computational refinement based on these constraints.

In this chapter, we present the current version of the RosettaSurface algorithm and the methods it employs. The core RosettaSurface algorithm is available both as a C++ executable and as a modifiable PyRosetta script (Chaudhury, Lyskov, & Gray, 2010). For usage on high-performance computing clusters, we recommend the C++ executable. For users interested in modifying or expanding the existing algorithm, or crafting their own algorithm, PyRosetta scripting is recommended. For both versions, command lines for preprocessing, execution, and postprocessing/analysis are provided.

2. ALGORITHM EVOLUTION AND PRIOR RESULTS

RosettaSurface was developed in stages to address progressively more complex challenges. In the first iteration, the algorithm focused exclusively on sampling rigid-body and side-chain conformational DOFs between a previously refined solution-state structure of the protein of interest and the surface. Restriction to these DOFs precluded the prediction of protein conformational changes upon adsorption but allowed a thorough coverage of the rigid-body and side-chain conformational space. This algorithm was first applied to the 43-residue salivary protein statherin docking to the [001] plane of HAp (Makrodimitris, Masica, Kim, & Gray, 2007). The algorithm converged on a set of low-energy structures (Fig. 16.1A and B) consistent with previously published solid-state NMR (ssNMR) data and mutagenesis experiments, and it also suggested a helical lysine and arginine molecular recognition motif for the phosphate clusters in HAp.

The next iteration of the algorithm extended sampling to include protein backbone torsional DOFs, not only allowing us to model conformational changes upon adsorption but also increasing the size of conformational space. Again, results were generated for the model statherin/HAp system and benchmarked against experimental ssNMR distance and angle measurements (Masica & Gray, 2009). This algorithm captured an experimentally observed adsorption-induced folding event in statherin's helical binding domain and yielded top structures that matched 15 published ssNMR distance and angle measurements. The algorithm was adapted to incorporate user-provided structural constraints to bias sampling toward conformations that are consistent with existing experimental information. This study was the first to demonstrate that a combination of experimental constraint determination and structure prediction could effectively determine high-resolution protein structures at biomineral interfaces.

The method of establishing a feedback loop between experimental constraint determination and structure prediction was further expanded upon in the next study on statherin/HAp (Masica, Ash, Ndao, Drobny, & Gray, 2010). Areas of structural ambiguity in computational results were targeted for isotopic labeling and ssNMR constraint measurement. Several iterated rounds of experiment and computation yielded a final adsorbed-state structure with minimal ambiguity, shown in Fig. 16.1C. In addition, the ability of predicted structures to comply with the experimental constraints on different HAp faces suggested which faces were most likely to be the binding face.

Since the original algorithm development, RosettaSurface has also been used to study the binding of tissue-nonspecific alkaline phosphatase to HAp (McKee et al., 2011), phosphorylated and nonphosphorylated leucine-rich amelogenin to HAp (Masica, Gray, & Shaw, 2011; Tarasevich et al., 2013), an acidic serine- and aspartate-rich motif of osteopontin to HAp (Addison, Masica, Gray, & McKee, 2010), and other apatite-binding peptides (Addison, Miller, et al., 2010).

3. COMPUTATIONAL METHODS

3.1. Overview

Recently, we incorporated RosettaSurface into the Rosetta 3 software suite (Leaver-Fay et al., 2011), giving object-oriented access to new features such as noncanonical amino acids (Renfrew, Choi, Bonneau, & Kuhlman, 2012), a generalized framework for handling symmetry (DiMaio, Leaver-Fay, Bradley, Baker, & Andre, 2011), variable pH (Kilambi & Gray, 2012), and the scripting interfaces RosettaScripts (Fleishman et al., 2011) and PyRosetta (Chaudhury et al., 2010). Rosetta 3's object-oriented design and Python interface make it possible for users to craft their own MCM docking or folding algorithms and tailor score functions specifically to their needs.

The RosettaSurface algorithm addresses the two principal challenges in structure prediction: (1) sampling of the conformational space available to both the protein and solid surface and (2) scoring of candidate structures (decoys) to identify the global free energy minimum.

Conformational sampling is carried out in two stages: (1) a solution-state stage, where the internal DOFs of the protein are optimized in implicit solvent, independent of the surface, and (2) an adsorbed-state stage, where the internal DOFs of the protein and the rigid-body orientation of the protein relative to the surface are simultaneously optimized on the surface. This process is repeated many times to generate a large ensemble of decoys for both the solution and adsorbed states. Starting from a fresh random fragment assembly process to generate each new decoy avoids oversampling kinetically trapped structures in false minima on the energy landscape and allows for straightforward parallelization.

3.2. Search algorithm

Figure 16.2 presents a flowchart of the RosettaSurface algorithm. The algorithm employs a multiscale MCM protocol based on Rosetta docking (Gray et al., 2003) and folding (Simons et al., 1999) strategies. A key feature in our algorithm is variation in the number of cycles before transitioning from the “solution” to the “adsorbed” stages. In the solution stage, the protein is able to sample both extended and compact structures. Once adsorbed onto the surface, large conformational changes often produce steric clashes between protein and surface atoms, preventing acceptance in a traditional MC algorithm (a computational limitation). In the algorithm, the protein begins the solution stage as a fully extended peptide chain and undergoes several optimization cycles, each resulting in a more compact and “protein-like” structure, containing features that occur frequently in known crystal structures. Varying the number of optimization cycles in solution before transitioning to the adsorbed state ensures that both extended (few optimization cycles in solution) and compact (many optimization cycles in solution) structures are sampled in the adsorbed state. Physically, this feature allows the calculation to capture protein conformational changes upon adsorption (such as unfolding or structural transitions). A detailed description of the algorithm is in the Supporting Information.

An assumption in the search algorithm is that the relative positions of the atoms within the solid surface remain static throughout the simulation and that the surface atoms form a flat plane (no step edges or local defects). These complexities have yet to be incorporated into RosettaSurface.

3.3. Score function

The standard Rosetta score function contains both statistical and physical terms that are optimized to recover known protein crystal structures during folding and docking. However, benchmarking and parameterization are crucial when extending an existing score function or force field into the context of protein–surface interactions (Collier et al., 2012). Due to a lack of experimental benchmarks, very few parameter sets have been developed to specifically describe protein–surface interactions. Additionally, few existing parameter sets designed to simulate protein behavior in solution have been verified to be transferrable to the context of protein–surface interactions (Collier et al., 2012; Vellore, Yancey, Collier, Latour, & Stuart, 2010).

Rosetta’s default all-atom score function for protein folding (score12) serves as the starting point for the RosettaSurface score function. To account for electrostatic interactions between

protein side chains and charged surface ions, a Coulomb interaction term with a distance-dependent dielectric is added (Kilambi et al., 2012). The final score function is a weighted linear combination dominated by van der Waals (attractive and repulsive), electrostatic (distance-dependent dielectric), hydrogen-bond, and implicit solvent interactions (Table 16.1):

$$E_{total} = W_{vdW} E_{vdW} + W_{elec} E_{elec} + W_{Hbond} E_{Hbond} + W_{solv} E_{solv}$$

where the weights W_i and the energy functions E_i are defined in Rohl, Strauss, Misura, and Baker, (2004). The statistical score function terms derived from protein structures are only included for intraprotein interactions. Rosetta's accuracy in modeling the intraprotein component of protein-surface interactions is the main advantage of using score12 as a starting point (Leaver-Fay et al., 2013).

When RosettaSurface is run in a constraint-based fashion, the score function is modified to include additional terms that penalize structures deviating from experimental measurements. This penalty ensures that conformational sampling is restricted to structures in agreement with experimental data.

Our studies to date are limited to surfaces of ionic crystals. In this context, electronic delocalization effects are minimal, and assignment of partial charges for electrostatic calculations is a fair approximation. In addition to ignoring electronic effects, we have also employed an implicit solvation model (Lazaridis & Karplus, 1999). Although the reduced DOFs offered by implicit solvation save computation time, hydrogen-bond interactions mediated by explicit water molecules at the solid-liquid interface are lost (Freeman et al., 2011; Latour, 2008; Sun & Latour, 2006). The precise role of explicit waters in protein-surface interactions remains an area of active research.

4. PROTOCOL CAPTURE

In this section, we provide a step-by-step protocol to generate and postprocess candidate adsorbed protein structures. As an example, we dock osteocalcin to the [100] surface of HAp. Osteocalcin is known to play a role in bone mineralization, but the mechanism of its interaction remains speculative (Ducy et al., 1996). Osteocalcin's solution-state crystal structure is known (PDB ID=1Q8H) and a model binding structure has been proposed assuming no unfolding upon binding (Hoang, Sicheri, Howard, & Yang, 2003). Since osteocalcin contains three γ -carboxyglutamic acids, it has only become possible to model this protein in the new Rosetta 3 framework with the noncanonical amino acid parameterization (Renfrew et al., 2012).

Input files containing the crystal surface coordinates, the protein sequence, and experimentally determined constraints (if any) are needed to run RosettaSurface. Descriptions of each of these file types are included in the Supporting Information, and the input files for the osteocalcin/HAp example are provided in the protocol_capture directory included with Rosetta 3.5 (www.rosettacommons.org).

4.1. Running RosettaSurface

4.1.1 C++ executable—After obtaining and compiling the latest version of the Rosetta source code (see <http://www.rosettacommons.org>), the RosettaSurface executable can be found in the directory `rosetta_source/bin`. The command line for execution of RosettaSurface is `rosetta_source/bin/surface_docking.gccrelease@flags` where `gccrelease` may be replaced if other compilers (`icc` and `macos`) were used to build Rosetta and `flags` is a plain text file containing a list of all arguments the user wishes to pass to RosettaSurface (Table 16.2). For scientific benchmarking, $O(10^5)$ structures are needed; thus, a high-performance computing cluster is employed, typically using the Condor job distribution system (Thain, Tannenbaum, & Livny, 2005).

4.1.2 PyRosetta script—In addition to the C++ RosettaSurface script, we have written a Python script employing the PyRosetta libraries to allow a user to easily change and adjust parameters of the protocol to suit specific needs. For example, the number of cycles of refinements may be decreased if sufficient constraint data are available, allowing fewer cycles to obtain a meaningful structure in less time. Examples of parameters that can be changed are listed in Table 16.3.

The PyRosetta scripts can be found in the `apps` directory within PyRosetta and can be run in a similar way to the C++ executable: `PyRosetta/apps/surface_docking/surface_docking.py@flags`

The PyRosetta options differ slightly (Table 16.4). Decoy generation requires 1–15 min, depending on the length of the protein and the distribution of the solution-state and adsorbed-state refinement cycles.

4.2. Postprocessing and interpretation of results

For both the C++ and PyRosetta implementations, the raw data output from RosettaSurface consists of decoys and their corresponding scores. Decoys are output in PDB format in the working directory and are appended with a numerical identification tag. Total scores and individual score terms are appended in each decoy's PDB file and summarized in a score file (`.fasc`) in the working directory. A breakdown of the hydrogen bonds and the constraint energies is also appended to the end of each PDB file.

Typically, the top-scoring 1% of decoys are considered for structural analysis. The visualization tool PyMOL (Schrodinger, 2010), which can be interfaced with PyRosetta (Baugh, Lyskov, Weitzner, & Gray, 2011), allows the user to identify interactions that contribute most strongly to the adsorption energy. The script `GetTop.sh` extracts the top decoys into a separate folder, and the script `PostProcessRS.sh` generates the secondary structure, contact map, and adsorbed-state surface contact map plots inside that folder. An additional script, `extract_scores_and_constraints.py`, is used to output the total score and constraint score of each decoy in the directory. A summary of these scripts and their usage is in Table 16.5.

The sample files included with the installation can be found in the directory `rosetta_demos/protocol_capture/2013/surface_docking`, which includes a README file with additional

instructions regarding the environment and paths. A sample flags file is provided to generate ten osteocalcin decoys with a disulfide constraint to bond residues 23 and 29. The commands, run from the sample directory, are

1. `surface_docking.py@flags` (Usecommand in §4.1.1 for C++ executable) **a.** Wait until all 10 decoys have been generated (no `.in_progress` files left)
2. `GetTop.sh Ads 4`
3. `cd TOP4.Ads/`
4. `PostProcessRS.sh`

In order for this example to run quickly, these commands generate 10 decoys (step 1 in the flags file) and postprocess the four top-scoring decoys. For production runs, and for the plots described in sections 4.2.1–4.2.4, we generated 100,000 structures and analyzed the top-scoring 100.

4.2.1 Secondary structure—A plot is generated to illustrate the distribution of secondary structures at each residue in the set of top-scoring structures. Secondary structure can assess convergence between top-scoring decoys, and comparison of the secondary structure profiles between adsorbed- and solution-state structures can identify adsorption-induced conformational changes.

Output for the case of osteocalcin on HAp is in Fig. 16.3A. The secondary structure distribution of the top adsorbed-state decoys shows a mix of secondary structures at the N-terminus and three major helical structures, matching those observed in the crystal structure (bar above the plot). Residues 1–12 are disordered in the crystal structure (Hoang et al., 2003), so they are excluded, not comparable. The protocol suggests that the solution-state secondary structures are largely retained after adsorption.

4.2.2 Protein–protein contact map—While the secondary structure plot provides local structural information, the contact map provides information about the tertiary structure of the protein. Pairs of residues with at least one atom–atom contact within 8 Å are aggregated over the 100 top-scoring structures and the resulting pairwise frequencies are illustrated in a nonmirrored plot. The *x*- and *y*-axes represent the residue numbers and the gradient from white to blue represents minimum to maximum frequencies.

An example of this two-dimensional matrix for the osteocalcin adsorbed-state is shown in Fig. 16.3B. Proximal contacts, such as those found in helices, appear as 4-residue wide stripes along the diagonal, while distal interactions are represented by patches off the diagonal.

Figure 16.3C is a difference plot between the known solution-state structure and the top adsorbed-state decoys, which shows how the tertiary structure changes upon adsorption. (To generate this plot, include a native.pdb file containing the solution-state structure and run the script `PlotContactMap.sh`.) Contacts that dominate in the adsorbed-state and solution-state decoys are colored blue and red, respectively, with white representing equivalent frequencies. In this example, the first 12 residues are not comparable and are omitted. The

map shows that the prominent interhelical interactions in the solution-state structure are broken during adsorption.

4.2.3 Protein–surface contact map—Similar to the protein–protein contact map, the protein–surface contact map gives information about contacts in the top adsorbed-state decoys. The y-axis is the perpendicular distance of each residue to the surface (in angstroms), and the width of the distance distribution can be used to assess convergence in the models. Residues with ill-defined positions are the most useful protein–surface atom pairs to inform with ssNMR experiments.

Figure 16.3D illustrates the protein–surface contact map for osteocalcin. The three helices exhibit alternating dark spots close to the surface, showing frequent adsorption among the top-scoring structures. Some of the most dominant protein–surface interactions include the three γ -carboxyglutamic acids, two aspartic acids, and two arginine residues and minor contacts with some unmodified glutamic acids. The large spread of distances at both termini shows that a range of conformations are represented in the top-scoring decoys. ssNMR measurements, therefore, would be helpful near the termini to narrow down the candidate structures.

4.2.4 Interpretation of osteocalcin results—Together, the RosettaSurface results in Fig. 16.3 suggest that the adsorbed osteocalcin maintains three helices, adopts an unfolded structure, and interacts with the HAp mineral primarily through three γ -carboxyglutamic acid residues (Gla17, Gla21, and Gla24) of the first helix, two aspartic acid residues (Asp30 and Asp34) of the second helix, and two arginine residues (Arg43 and Arg44) of the third helix. The interacting residues on the first two helices are similar to those proposed by Hoang et al. in their rigid, crystal structure-based, model (Hoang et al., 2003) and both models show symmetrically equivalent calcium atoms coordinated by the three γ -carboxyglutamic acid residues. In fact, the residue contacts proposed by Hoang et al. are all local in sequence, so that many alternate or unfolded conformations can reproduce the same strong contacts with the calcium ions in HAp.

To test whether the unfolded conformations are truly lower in energy, or whether RosettaSurface may have been unable to sample folded binding structures like Hoang et al.'s model, we modified the PyRosetta implementation of RosettaSurface to hold the osteocalcin backbone fixed for rigid docking to the surface. While structures like Hoang et al.'s model were generated during rigid docking (Fig. 16.4B), their energies were higher than those of the lowest-scoring unfolded, adsorbed osteocalcin structures (Fig. 16.4C). It is possible that the unfolded RosettaSurface models are closer to the true structure of adsorbed osteocalcin. However, it is also possible that the balance of interface energy and folding energy is inaccurate; thus, further experimental data are needed to resolve the true structure. The conformations generated by RosettaSurface can be used to suggest sites for ssNMR labeling and measurement. For example, a distance constraint between a side-chain nitrogen atom of Arg43 and a HAp phosphorous atom could provide information on whether or not osteocalcin's third helix interacts with HAp.

5. FUTURE CHALLENGES

The RosettaSurface algorithm provides a flexible platform to explore adsorbed conformations of proteins on mineral surfaces. Because of the paucity of experimental data, it is difficult to validate an energy function (Latour, 2008), and thus, despite the success of the Rosetta energy function for folding, docking, and design, the ranking of adsorbed structures is subject to error. Work is needed to test alternate formulations and parameterizations of the energy function, ideally with additional atomic-resolution structural data on a variety of experimental systems. A second limitation is in RosettaSurface's sampling. Exploring relevant areas of conformational space is particularly difficult for long proteins, since transitional pathways may be hindered by the surface (Masica et al., 2011).

For progress in the study of biomineralization, a combination of computational and experimental approaches is ideal. The weaknesses in energy functions and conformational sampling can be compensated with even sparse experimental data to constrain the search. The constraint-based docking framework employed by RosettaSurface establishes a feedback loop between experimental constraint determination and computational refinement, making it a valuable tool for the study of biomineralization (Masica, Ash, et al., 2010). As the fundamental components of the algorithm improve, design of biomineralization proteins will also improve (Masica, Schrier, Specht, & Gray, 2010; Schrier, Sayeg, & Gray, 2011).

Acknowledgments

Moon-Young Liza Lee developed early versions of RosettaSurface in PyRosetta. Funding was provided by NSF CAREER grant 0846324 and NIH training grant T32 GM008403 (for RAT).

A. APPENDIX: A.1. Detailed algorithm description

Figure 16.2 in the main text shows the algorithm flowchart. The protein starting structure is an extended chain in implicit solvent, independent from the surface. The protein is initially represented in “centroid mode” where all backbone atoms are explicitly modeled but side chains are represented by a single “pseudoatom” that captures the low-resolution features of the side-chain chemistry.

The extended chain is then rapidly collapsed using a fragment assembly protocol. The fragment assembly protocol follows Simons, Kooperberg, Huang, and Baker (1997), except the radius of gyration score term is removed to allow a greater sampling of extended structures in both the solution and adsorbed stages to capture conformational changes upon adsorption. After each fragment assembly step, the protein backbone is optimized with “shear” and “small” moves coupled with conjugate gradient and line minimization, respectively.

At this point, the protein side chains are converted from “centroid” to “full-atom” mode. In “full-atom” mode, all-atom (including hydrogen) side chains are built from a backbone-dependent rotamer library (Dunbrack & Cohen, 1997) using a simulated annealing protocol (Kuhlman & Baker, 2000). A rotamer library, derived from PDB statistics, represents local minima on the QM energy landscape (Renfrew, Butterfoss, & Kuhlman, 2008).

The protein now undergoes a total of 5 “refinement cycles” with adsorption occurring randomly between cycles 1 and 5. This is done by generating a random number n between 1 and 5 and performing n cycles of solution stage refinement and $5-n$ cycles of adsorbed stage refinement. n is varied for each individual decoy, ensuring an approximately equal amount of sampling in both the solution and adsorbed states and yielding a range of compact and extended final structures.

Solution stage refinement includes “outer-“ and “inner-loop optimization.” During outer-loop optimization, a sequence of backbone perturbation moves are applied, each followed by conjugate-gradient minimization. During inner-loop optimization, the same perturbing moves are applied except each is followed by line minimization along the initial gradient. Separating the refinement stage in this fashion ensures that the least computationally expensive (and least globally perturbing) line minimization calculation is performed in the innermost loop. The outer loop is implemented five times; for each outer loop, the inner loop is repeated five times.

After n cycles of solution stage refinement, an adsorbed-state complex is formed by introducing the solution-refined protein to the surface in a random orientation and bringing the two into contact. The adsorbed protein then undergoes $5-n$ outer cycles of refinement on the surface. Adsorbed stage refinement is identical to solution stage refinement with the addition of a modified version of RosettaDock’s (Gray et al., 2003) high-resolution docking protocol at the end of each cycle. This modified docking protocol consists of a sequence of small rigid-body perturbations, side-chain repacking, and gradient-based minimization in rigid-body space repeated six times; side-chain repacking is combinatorial (full repack), rather than sequential (rotamer trials), every third time. Combinatorial repacking accounts for the possibility of correlated side-chain motions but is more computationally expensive than sequential repacking, which optimizes each side-chain rotamer independently. Hence, the protein undergoes simultaneous backbone, side-chain, and rigid-body optimization on the surface. Finally, an additional 2 cycles of adsorbed stage refinement are performed to arrive at a low-energy structure. With user-input “surface vectors,” the protein can be translated to a symmetrically equivalent position at the center of the slab after each round of docking. Because rigid-body docking involves translation in the plane of the surface, an additional “orientation” move is necessary to ensure that the protein is not translated off of the surface slab used for simulation yet remains in a symmetrically equivalent position on the surface.

A.2. Surface parameters

To maintain the residue-centric organization of Rosetta 3, RosettaSurface treats the surface as a sequence of “residues” whose positions remain static throughout the simulation. For ionic surfaces, this implementation is straightforward: each monatomic/polyatomic ion is treated as a distinct residue. The chemical properties of a residue are encoded by its corresponding params file in the Rosetta database. This file includes information such as bond connectivity, equilibrium geometry, atom types, and partial charges. Please see the Rosetta 3.4 manual for a description of each field in a params file.

A.3. Formatted PDB and surface vector files

RosettaSurface reads starting coordinates from a modified PDB format file (www.rcsb.org). The protein should initially be positioned above the surface face the user wishes to dock, with the protein coordinates appearing *after* the coordinates of the surface atoms in the file. The surface atoms should be designated as heteroatoms (HETATM record), belonging to a separate chain from the protein, in the input file with atom and residue designations that match those defined in the params file (see Section A.2).

Coordinates for the surface slab may be produced with CrystalMaker[®] (CrystalMaker Software Ltd, Oxford, England www.crystallmaker.com), a tool that enables users to prepare surfaces of predefined dimensions for specific crystal faces. The desired length/width of the surface slab will depend on protein size and must accommodate extended conformations without introducing edge effects. A slab thickness of ~10 Å is sufficient, as it exceeds the pairwise interaction cutoff of the Rosetta score function.

Three surface atoms, which define the edges of the surface unit cell, must be specified in a separate surface vectors file. These atoms define a pair of surface unit-cell vectors: translations by any integer linear combination of these vectors are symmetrically equivalent. These vectors can be used to translate the protein to a symmetrically equivalent surface patch at any point during the docking simulation. An example of the required syntax can be seen in the surface vectors file provided in the protocol_capture directory.

A.4. Fragment files

The fragment assembly protocol in the low-resolution stage requires pre-generated files that specify candidate backbone conformations, or fragments, for a given 9-mer or 3-mer subsequence. Fragments are most easily obtained using the Robetta server (Kim, Chivian, & Baker, 2004).

A.5. Constraint files

If RosettaSurface is run in a constraint-based fashion, the user must specify the type and parametric form of the constraint in a .cst file (Bowers, Strauss, & Baker, 2000). Three constraint types are currently allowed: atom pair, angle, and dihedral. Ambiguous constraints allow the user to provide a list of several constraints, with the score function only including the contribution of the lowest scoring constraint to the total energy. This type of constraint is needed to incorporate ssNMR distance measurements between protein and surface atoms because there are many symmetrically equivalent copies of the same surface atom in the simulation slab. Common parametric forms include harmonic, circular harmonic, bounded, Gaussian, constant, and linear scaling. An example constraint file for the statherin/HAp system is provided in the protocol_capture directory. For a description of constraint file formats, please see the Rosetta 3.4 manual.

In addition to distance and angle constraint files, a file containing residues with disulfide bonds can also be specified to ensure that these bonds are formed. The disulfide file contains

residue number pairs, delimited by a space, in separate lines. An example disulfide file is provided in the protocol_capture directory.

Multiple ambiguous constraints must be placed in separate files with filenames as follows:

PDB file = [name of input file].pdb

Constraints files

1. [name of input file]_ads01.cst, [name of input file]_ads02.cst etc.
2. [name of input file]_sol01.cst, [name of input file]_sol02.cst etc.

Constraints for the solution and adsorbed states are separated, since protein–surface constraints do not apply to the solution-state refinements.

REFERENCES

- Addison WN, Masica DL, Gray JJ, McKee MD. Phosphorylation-dependent inhibition of mineralization by osteopontin ASARM peptides is regulated by PHEX cleavage. *Journal of Bone and Mineral Research*. 2010; 25:695–705. [PubMed: 19775205]
- Addison WN, Miller SJ, Ramaswamy J, Mansouri A, Kohn DH, McKee MD. Phosphorylation-dependent mineral-type specificity for apatite-binding peptide sequences. *Biomaterials*. 2010; 31:9422–9430. [PubMed: 20943264]
- Baugh EH, Lyskov S, Weitzner BD, Gray JJ. Real-time PyMOL visualization for Rosetta and PyRosetta. *PLoS One*. 2011; 6:e21931. [PubMed: 21857909]
- Belcher AM, Wu XH, Christensen RJ, Hansma PK, Stucky GD, Morse DE. Control of crystal phase switching and orientation by soluble mollusc-shell proteins. *Nature*. 1996; 381:56–58.
- Berman A, Hanson J, Leiserowitz L, Koetzle TF, Weiner S, Addadi L. Biological-control of crystal texture—A widespread strategy for adapting crystal properties to function. *Science*. 1993; 259:776–779. [PubMed: 17809339]
- Bowers PM, Strauss CEM, Baker D. De novo protein structure determination using sparse NMR data. *Journal of Biomolecular NMR*. 2000; 18:311–318. [PubMed: 11200525]
- Bradley P, Misura KMS, Baker D. Toward high-resolution de novo structure prediction for small proteins. *Science*. 2005; 309:1868–1871. [PubMed: 16166519]
- Chaudhury S, Lyskov S, Gray JJ. PyRosetta: A script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics*. 2010; 26:689–691. [PubMed: 20061306]
- Cohavi O, Corni S, De Rienzo F, Di Felice R, Gottschalk KE, Hoefling M, et al. Protein-surface interactions: Challenging experiments and computations. *Journal of Molecular Recognition*. 2010; 23:259–262. [PubMed: 19950121]
- Collier G, Vellore NA, Yancey JA, Stuart SJ, Latour RA. Comparison between empirical protein force fields for the simulation of the adsorption behavior of structured LK peptides on functionalized surfaces. *Biointerphases*. 2012; 7:24. [PubMed: 22589067]
- Corno M, Rimola A, Bolis V, Ugliengo P. Hydroxyapatite as a key biomaterial: Quantum-mechanical simulation of its surfaces in interaction with biomolecules. *Physical Chemistry Chemical Physics*. 2010; 12:6309–6329. [PubMed: 20485772]
- DiMaio F, Leaver-Fay A, Bradley P, Baker D, Andre I. Modeling symmetric macromolecular structures in Rosetta3. *PLoS One*. 2011; 6:e20450. [PubMed: 21731614]
- Ducy P, Desbois C, Boyce B, Pinero G, Story B, Dunstan C, et al. Increased bone formation in osteocalcin-deficient mice. *Nature*. 1996; 382:448–452. [PubMed: 8684484]
- Dunbrack RL, Cohen FE. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Science*. 1997; 6:1661–1681. [PubMed: 9260279]

- Fleishman SJ, Leaver-Fay A, Corn JE, Strauch EM, Khare SD, Koga N, et al. RosettaScripts: A scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One*. 2011; 6:e20161. [PubMed: 21731610]
- Freeman CL, Harding JH, Quigley D, Rodger PM. Simulations of ovocleidin-17 binding to calcite surfaces and its implications for eggshell formation. *Journal of Physical Chemistry C*. 2011; 115:8175–8183.
- Goobes G, Stayton PS, Drobny GP. Solid state NMR studies of molecular recognition at protein-mineral interfaces. *Progress in Nuclear Magnetic Resonance Spectroscopy*. 2007; 50:71–85. [PubMed: 19768124]
- Gray JJ. The interaction of proteins with solid surfaces. *Current Opinion in Structural Biology*. 2004; 14:110–115. [PubMed: 15102457]
- Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, et al. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *Journal of Molecular Biology*. 2003; 331:281–299. [PubMed: 12875852]
- Hoang QQ, Sicheri F, Howard AJ, Yang DSC. Bone recognition mechanism of porcine osteocalcin from crystal structure. *Nature*. 2003; 425:977–980. [PubMed: 14586470]
- Iori F, Di Felice R, Molinari E, Corni S. GoLP: An atomistic force-field to describe the interaction of proteins with Au(111) surfaces in water. *Journal of Computational Chemistry*. 2009; 30:1465–1476. [PubMed: 19037859]
- Kilambi KP, Gray JJ. Rapid calculation of protein pKa values using Rosetta. *Biophysical Journal*. 2012; 103:587–595. [PubMed: 22947875]
- Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research*. 2004; 32:W526–W531. [PubMed: 15215442]
- Kortemme T, Morozov AV, Baker D. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *Journal of Molecular Biology*. 2003; 326:1239–1259. [PubMed: 12589766]
- Kuhlman B, Baker D. Native protein sequences are close to optimal for their structures. *Proceedings of the National Academy of Sciences of the United States of America*. 2000; 97:10383–10388. [PubMed: 10984534]
- Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. *Science*. 2003; 302:1364–1368. [PubMed: 14631033]
- Latour RA. Molecular simulation of protein-surface interactions: Benefits, problems, solutions, and future directions. *Biointerphases*. 2008; 3:FC2–FC12. [PubMed: 19809597]
- Lazaridis T, Karplus M. Effective energy function for proteins in solution. *Proteins*. 1999; 35:133–152. [PubMed: 10223287]
- Leaver-Fay A, O'Meara MJ, Tyka M, Jacak R, Song Y, Kellogg EH, et al. Scientific benchmarks for guiding macromolecular energy function improvement. *Methods in Enzymology*. 2013; 523:109–143. [PubMed: 23422428]
- Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, et al. Rosetta3: An object-oriented software suite for the simulation and design of macromolecules. *Methods in Enzymology*. 2011; 487:545–574. *Computer Methods, Pt C*. [PubMed: 21187238]
- Li ZQ, Scheraga HA. Monte-Carlo-minimization approach to the multiple-minima problem in protein folding. *Proceedings of the National Academy of Sciences of the United States of America*. 1987; 84:6611–6615. [PubMed: 3477791]
- Makrodimitris K, Masica DL, Kim ET, Gray JJ. Structure prediction of protein-solid surface interactions reveals a molecular recognition motif of statherin for hydroxyapatite. *Journal of the American Chemical Society*. 2007; 129:13713–13722. [PubMed: 17929924]
- Masica DL, Ash JT, Ndao M, Drobny GP, Gray JJ. Toward a structure determination method for biomineral-associated protein using combined solid-state NMR and computational structure prediction. *Structure*. 2010; 18:1678–1687. [PubMed: 21134646]
- Masica DL, Gray JJ. Solution- and adsorbed-state structural ensembles predicted for the statherin-hydroxyapatite system. *Biophysical Journal*. 2009; 96:3082–3091. [PubMed: 19383454]

- Masica DL, Gray JJ, Shaw WJ. Partial high-resolution structure of phosphorylated and non-phosphorylated leucine-rich amelogenin protein adsorbed to hydroxyapatite. *The Journal of Physical Chemistry C, Nanomaterials and Interfaces*. 2011; 115:13775–13785.
- Masica DL, Schrier SB, Specht EA, Gray JJ. Denovodesign of peptide-calcite biomineralization systems. *Journal of the American Chemical Society*. 2010; 132:12252–12262. [PubMed: 20712308]
- McKee MD, Nakano Y, Masica DL, Gray JJ, Lemire I, Heft R, et al. Enzyme replacement therapy prevents dental defects in a model of hypophosphatasia. *Journal of Dental Research*. 2011; 90:470–476. [PubMed: 21212313]
- Michaels AS, Matson SL. Membranes in biotechnology—State of the art. *Desalination*. 1985; 53:231–258.
- Raman S, Vernon R, Thompson J, Tyka M, Sadreyev R, Pei JM, et al. Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins*. 2009; 77:89–99. [PubMed: 19701941]
- Ratner BD, Bryant SJ. Biomaterials: Where we have been and where we are going. *Annual Review of Biomedical Engineering*. 2004; 6:41–75.
- Renfrew PD, Butterfoss GL, Kuhlman B. Using quantum mechanics to improve estimates of amino acid side chain rotamer energies. *Proteins*. 2008; 71:1637–1646. [PubMed: 18076032]
- Renfrew PD, Choi EJ, Bonneau R, Kuhlman B. Incorporation of non-canonical amino acids into Rosetta and use in computational protein-peptide interface design. *PLoS One*. 2012; 7:e32637. [PubMed: 22431978]
- Rimola A, Aschi M, Orlando R, Ugliengo P. Does adsorption at hydroxyapatite surfaces induce peptide folding? Insights from large-scale B3LYP calculations. *Journal of the American Chemical Society*. 2012; 134:10899–10910. [PubMed: 22680347]
- Rohl CA, Strauss CE, Misura KM, Baker D. Protein structure prediction using Rosetta. *Methods in Enzymology*. 2004; 383:66–93. [PubMed: 15063647]
- Schrier SB, Sayeg MK, Gray JJ. Prediction of calcite morphology from computational and experimental studies of mutations of a de novo-designed peptide. *Langmuir: The ACS Journal of Surfaces and Colloids*. 2011; 27:11520–11527. [PubMed: 21797243]
- Schrodinger, LLC. The PyMOL molecular graphics system. version 1.3r12010.
- Simons KT, Bonneau R, Ruczinski I, Baker D. Ab initio protein structure prediction of CASP III targets using ROSETTA. *Proteins*. 1999; 37(Suppl. 3):171–176. [PubMed: 10526365]
- Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *Journal of Molecular Biology*. 1997; 268:209–225. [PubMed: 9149153]
- Sircar A, Chaudhury S, Kilambi KP, Berrondo M, Gray JJ. A generalized approach to sampling backbone conformations with RosettaDock for CAPRI rounds 13-19. *Proteins*. 2010; 78:3115–3123. [PubMed: 20535822]
- Sun Y, Latour RA. Comparison of implicit solvent models for the simulation of protein-surface interactions. *Journal of Computational Chemistry*. 2006; 27:1908–1922. [PubMed: 17019723]
- Tarasevich BJ, Perez-Salas U, Masica DL, Philo J, Kienzle P, Krueger S, et al. Neutron reflectometry studies of the adsorbed structure of the amelogenin, LRAP. *Journal of Physical Chemistry B*. 2013; 117:3098–3109.
- Thain D, Tannenbaum T, Livny M. Distributed computing in practice: The Condor experience. *Concurrency and Computation: Practice & Experience*. 2005; 17:323–356.
- Vellore NA, Yancey JA, Collier G, Latour RA, Stuart SJ. Assessment of the transferability of a protein force field for the simulation of peptide-surface interactions. *Langmuir*. 2010; 26:7396–7404. [PubMed: 20222735]
- Wang F, Stuart SJ, Latour RA. Calculation of adsorption free energy for solute-surface interactions using biased replica-exchange molecular dynamics. *Biointerphases*. 2008; 3:9–18. [PubMed: 19768127]
- Warshel A, Russell ST. Calculations of electrostatic interactions in biological systems and in solutions. *Quarterly Reviews of Biophysics*. 1984; 17:283–422. [PubMed: 6098916]
- Weiner S, Dove PM. An overview of biomineralization processes and the problem of the vital effect. *Biomineralization*. 2003; 54:1–29.

Figure 16.1.

Predicted structures from RosettaSurface. Top (A) and side (B) views of the lowest-energy structure of statherin docked onto the [001] face of HAp using the initial version of the algorithm. (C) A representative structure of statherin bound to HAp, colored by the local root-mean-square deviation of 100 structures that best satisfied experimental constraints. *Reprinted from Makrodimitris et al. (2007) and Masica, Ash, et al. (2010) with permission.*

Figure 16.2.

Flowchart of the complete RosettaSurface algorithm. Small/shear moves perturb the protein backbone and are described in Bradley, Misura, and Baker (2005); linmin and dfpmin_armijo are energy minimization routines described in the Rosetta user's manual (<https://www.rosettacommons.org/manuals>). N_{res} , protein length; W_{rep} , repulsive van der Waals weight.

Figure 16.3.

Analysis plots generated by RosettaSurface. (A) Secondary structure profile for osteocalcin adsorbed to HAp [100]. (B) Intraprotein contact map for osteocalcin adsorbed to HAp [100]. (C) A difference plot between crystal structure and adsorbed-state decoys. (D) Protein-surface contact map for osteocalcin adsorbed to HAp [100]. γ represents gamma-carboxyglutamic acid. Line above each plot represents the secondary structure of the crystal structure and is colored as in (A). Boxes highlight residues predicted to interact with HAp in Hoang's model.

Figure 16.4.

Predicted structures for the osteocalcin/HAp system. (A) Predicted solution-state structure of osteocalcin (blue) superimposed with its crystal structure (red). Lowest-energy decoys from rigid (B) and flexible (C) backbone docking. Similar contacts are made between the γ -carboxyglutamic acids and the surface in both models, indicating ambiguity in whether or not osteocalcin adopts extended conformations on HAp.

Table 16.1

Major score terms used in RosettaSurface

Score term	Weight	Description
fa_atr	0.800	Attractive range of Lennard-Jones 6–12 potential
fa_rep	0.440	Repulsive range of Lennard-Jones 6–12 potential
hack_elec	1.000	Coulomb interactions with distance-dependent dielectric ^a
hbond_lr_bb	1.170	Long-range orientation-dependent backbone–backbone hydrogen bonding ^b
hbond_sr_bb	0.585	Short-range orientation-dependent backbone–backbone hydrogen bonding ^b
hbond_bb_sc	1.170	Orientation-dependent backbone–side-chain hydrogen bonding ^b
hbond_sc	1.100	Orientation-dependent side-chain–side-chain hydrogen bonding ^b
fa_sol	0.650	Implicit solvation model based on EEF1 from Lazaridis and Karplus ^c

^aWarshel and Russell (1984).^bKortemme, Morozov, and Baker (2003).^cLazaridis and Karplus (1999).

Table 16.2

Command-line arguments for the RosettaSurface C++ executable

Arguments	Description
-s filename	PDB file containing protein and surface
-in:file:frag3 filename	Name of 3-mer fragment file
-in:file:frag9 filename	Name of 9-mer fragment file
-in:file:surface_vectors filename	Name of surface vectors file
-database filepath	Path to rosetta_database (default \$ROSETTA3_DB)
-nstruct integer	Number of structures to generate
-constraints:cst:file filename	Name of constraints file
-multiple_processes_writing_to_one_directory	<i>Optional.</i> Enables parallelization across a computing cluster
-show_simulation_in_pymol integer	<i>Optional.</i> Number of seconds between PyMOL updates (enables live viewing of simulation in PyMOL)

Table 16.3

Examples of modifiable PyRosetta algorithm parameters

Parameter	Variable/function name	Default value
Solution-state (full-atom)		
Refinement cycles	sol_cycles	$n - 5$ (random)
Outer cycles	sol_outer_cycles	5
Inner cycles	sol_inner_cycles	5
Adsorbed-state		
Refinement cycles	ads_cycles	$5 - n$
Outer cycles	ads_outer_cycles	5
Inner cycles	ads_inner_cycles	5
Other parameters		
Number of small/shear moves before minimization	cen_relax.set_nmoves(n) fa_relax.set_nmoves(n)	6
Maximum angle of perturbation for small/shear moves	cen_relax.set_max_angle(n) fa_relax.set_max_angle(n)	30/outer_cycle

Table 16.4

Command-line arguments for RosettaSurface Python script

Arguments	Description
-h, -help	Show help message with these descriptions and exit
-s, -start filename	Specify start PDB file containing protein and surface
-database path	Specify Rosetta database path in PyRosetta (default \$PYROSETTA_DATABASE)
-n, -nstruct integer	Specify number of structures to generate
-f3, -frag3 filename	<i>Optional.</i> Specify 3-mer fragment file. If not specified, <i>ab initio</i> and centroid relax will be skipped, which is useful if solution-state structure is known
-f9, -frag9 filename	<i>Optional.</i> Specify 9-mer fragment file. If not specified, <i>ab initio</i> and centroid relax will be skipped, which is useful if solution-state structure is known
-c, -constraints	<i>Optional.</i> Specifies that ssNMR constraints are to be loaded from [startpdbname]_ads##.cst (adsorbed-state) and [startpdbname]_sol##.cst, (solution-state) where ## are integers
-nosmallshear	<i>Optional.</i> If included, small/shear refinements are not used. Use if rigid-body docking is desired. If not included, refinements are used (default)
-d, -disulf filename	<i>Optional.</i> Specify disulfide constraints file

Table 16.5

Scripts for postprocessing analysis

Script name	Working directory	Usage	Output
GetTop.sh	Main directory with all decoys	GetTop.sh <Ads/Sol> <no. of decoys>	TOP<no. of decoys>. <Ads/Sol> folder containing decoys
PostProcessRS.sh	TOP folders for Ads and Sol	PostProcessRS.sh	SecStruct.png ContactMap.png SurfaceContactMap.png
extract_scores_and_constraints.py	TOP folders for Ads and Sol	extract_scores_and_constraints.py <ads/sol/both>	table_<ads/sol>_scores