SUMMARY OF POSTER PRESENTED AT THE 12TH PROTEIN SOCIETY SYMPOSIUM IN SAN DIEGO, JULY 1998

Screening for flexible ligands with solvation and conformational change of the protein binding site

Volker Schnecke and **Leslie A. Kuhn**
Protein Structural Analysis and Design Laboratory
Department of Biochemistry
Michigan State University
http://www.bch.msu.edu/labs/kuhn
volker@sol.bch.msu, kuhn@agua.bch.msu

SPECITOPE is a tool for screening large databases for ligand structures matching a given protein binding site. The binding site is modeled by surface atoms, conserved first-shell water molecules, and a template consisting of hydrogen-bonding and hydrophobic interaction centers. During screening, distance geometry and multi-stage hashing techniques are used for mapping polar atoms and hydrophobic centers from ligand candidates onto a subset of these template points. Only ligands that match at least three template points are docked into the binding site and checked for steric fit. Van der Waals collisions are resolved by targeted ligand flexibility and protein side-chain rotations. Solvation of the binding site is taken into account by *Consolv*'s [1] prediction of conserved water molecules likely to mediate ligand binding. Potential ligands are scored on hydrogen-bond and hydrophobic contacts. The original application of SPECITOPE was the screening of peptide structures to find specific epitopes to a protein active site [2]. Now, a set of over 80,000 small organic compounds from the Cambridge Structural Database (CSD) can also be screened for potential ligands. A focus of our research is on targets for antibiotic and cancer therapy: β-lactamase, glutathione transferase, uracil-DNA glycosylase, and thymidylate synthase.

The main difference relative to the original version of SPECITOPE is the use of multi-dimensional hashing to directly access subsets of template points that match a given triangle of ligand points (either carbon-ring centers or H-bond donors/acceptors). This enables the use of a larger template and results in a significant speed increase that allows consideration of full ligand flexibility and solvation. Before screening, all possible triangles consisting of a combination of three template points are generated and stored in a way to be directly accessible via three index tables based on interaction types, the perimeter, and the length of the longest side of the triangle. During screening, all possible triplets of ligand points are tested for matching onto three template points by directly accessing compatible template triangles based on the three hashing features. If a feasible mapping of three ligand points onto a template triangle is identified, the ligand is transformed into the binding site based on a least-squares fit transformation of the matched points.

The choice of the current three ligand points determines the base fragment, which is taken as the rigid part of the ligand. Collisions of the base fragment with the protein main chain are resolved by iterative translations of the ligand as a rigid body. Directed rotations around flexible bonds in the ligand and single bonds in protein side chains are used to resolve all remaining collisions after the base-fragment translation. Protein side-chain rotations are used, if the collisions cannot be resolved by ligand flexibility. Collisions are detected starting from the base fragment or the backbone, respectively, and one single collision is resolved at a time by a minimal rotation around the closest flexible bond. A backtracking approach is used to try different combinations of bonds and directions until all collisions are resolved.

After a collision-free binding mode for a ligand has been detected, the complex is scored based on the number of intermolecular H-bonds, including water-mediated interactions with the conserved binding-site waters [1], and on the hydrophobic complementarity of the protein-ligand interface, as described in [2]. The goal of this scoring is not to predict the binding affinity, since the ligand is not necessarily in its optimized binding mode, but to provide a relative ranking of the potential ligands and to rule out those with poor complementarity. However, since the initial ligand conformations are taken from crystal structures, it can be assumed that their conformations are reasonable.

Consolv [1] is applied to predict conservation of binding-site waters upon ligand binding. *Consolv* is a genetic-algorithm based classifier that can identify favorable solvation sites based on physical and chemical features of the local protein environment. It was developed by our group and is available on our webserver. During screening, these waters are taken as conserved waters, i.e., they cannot be displaced by ligand atoms, and are treated like protein atoms during collision checks. They may be displaced when a protein side-chain is rotated to which they are connected via a H-bond. In the future, we will also consider displacement of bound water sites by polar ligand atoms.

As the first test case, subset of the CSD (Cambridge Crystallographic Database System) consisting of 84,749 compounds, each with at least three interaction points (each can either be a six-atom carbon ring or a polar atom), was screened for potential ligands to glutathione transferase (GST). An eleven-point template for the free protein (PDB 1bay) was generated by superimposing six different ligands into the binding site and taking their conserved hydrophobic centers (in the H-site) and polar ligand atom positions of glutathione as template points. Three binding-site waters were predicted as conserved by *Consolv*. SPECITOPE identified 211 potential ligands within 20 minutes on a Sun SPARC Ultra/140 workstation, including a known ligand that was added to the database for verification, p-bromobenzylglutathione. For most ligands, including the known one, protein side-chain rotations were necessary for docking, reflecting inducible complementarity upon ligand binding, since the ligand-free binding-site conformation was used.

Future plans include the use of *Consolv* not only to predict the conservation of binding-site waters included in the free protein structure, but to dynamically add additional waters during docking and scoring, if a particular position has the environment of a typical solvation site. Ligand and protein side-chain flexibility will be employed to not only resolve collisions, but also to increase the complementarity of the interface, e.g., by maximizing the number of H-bonds. A special type of template point will be added to screen for compounds that are covalently bound to the protein. To establish the covalent bond, this point must be exactly matched by a proper ligand atom, the position of which remains fixed during collision resolving. An automatic template-design approach will be implemented to generate a set of template points for proteins when a variety of ligand-bound structures is not available.

References

[1] M. L. Raymer, P. C. Sanschagrin, W. F. Punch, S. Venkataraman, E. D. Goodman, and L. A. Kuhn, *Predicting Conserved Water-mediated and Polar Ligand Interactions in Proteins Using a K-nearest-neighbors Genetic Algorithm*, Journal of Molecular Biology, Volume 265, Number 4, January 1997, pp. 445–464

[2] V. Schnecke, C. A. Swanson, E. D. Getzoff, J. A. Tainer, and L. A. Kuhn, *Screening a Peptidyl Database for Potential Ligands to Proteins with Side-chain Flexibility*, Proteins: Structure, Function, and Genetics, Volume 33, Number 1, October 1998