SUMMARY OF POSTER PRESENTED AT THE UCSF-MDI MEETING "MOLECULAR RECOGNITION IN DRUG DESIGN: DOCKING AND SCORING", SAN FRANCISCO, FEBRUARY 1998

Screening for Potential Ligands to Proteins with Inducible Side-chain Complementarity

Volker Schnecke and **Leslie A. Kuhn**
Department of Biochemistry
Michigan State University

SPECITOPE is a tool for screening large structural databases for potential ligands to a given protein binding site. When developing a tool to screen more than 100,000 compounds to identify about 100 potential ligands, care must be taken to reduce the rate of false positives and to prevent false negatives, e. g. to keep the known ligands included in the database. SPECITOPE uses efficient techniques to eliminate infeasible candidates early in the search, before docking them into the binding site. The screening is based on a template consisting of four or five (in general, N) points above the protein surface, which are to be matched by hydrogen-bonding donor or acceptor atoms of the ligand. One way to design such a template is to superimpose several complexes to identify conserved ligand hydrogen-bonding positions.

During the search, distance-geometry techniques are employed to rule out those candidates that cannot form the required hydrogen bonds based on distance constraints. If the largest intra-template distance exceeds the longest distance between polar atoms by 2.0 Å, this molecule can be ruled out, since no N-element subset of its polar atoms can match the template. Otherwise, all possible combinations of N atoms out of all polar atoms in this potential ligand are tested against the template. First, the number of hydrogen-bond donors and acceptors in a set is compared against the donor/acceptor requirements of the template points. If they are compatible, the interatomic distances for the N polar atoms are compared to the intra-template distances. For comparing the distances, the root-mean-square deviation (RMSD_{list}) between the sorted lists of the distances in both sets is computed. The RMSD_{list} provides a lower bound for the distance matrix error (DME) of any matching (one-to-one correspondence) of the N atoms to the template points. If the RMSD_{list} is above 0.7 Å, the current set of atoms is discarded, since none of its possible matchings can yield a DME below 0.7Å. If the set passes the RMSD_{list} check, the matching with minimal DME is found by computing all permutations of the N polar ligand atoms, and they are transformed by a least squares-fit transformation onto the corresponding template points. For our test cases (see below), on average only 30% of the screened molecules pass to this step. If the root-mean-square deviation between template points and ligand atoms is below 1.0 Å, then the whole molecule is transformed into the binding site; otherwise, this atom set is discarded. About 70% of the tested molecules exceed this RMSD threshold, so that on average only 10% of the screened molecules are finally docked into the binding site.

Once a molecule has been docked, it is checked for steric fit, and different approaches are used to resolve collisions. In the current version of SPECITOPE, peptides are screened; their main chains are assumed to be rigid, and the single bonds in the side chains are rotatable. A fixed number of overlaps of main-chain atoms of both protein and ligand are resolved by iterative transformations of the ligand as a rigid body. For this, the optimal directions for resolving each single overlap are computed, and the average translation direction is determined by adding the single displacement vectors. The whole ligand is then transformed in this direction by the minimal amount necessary to resolve all main-chain collisions. If new overlaps originate from this transformation, the approach is iterated up to 100 times. After resolving all main-chain collisions, the side chains of the peptidyl ligands are checked for overlaps

with any protein atom. The minimal rotation required to resolve all collisions in a side chain is computed for the closest single bond to the colliding atoms. If no rotation around this bond results in an overlap-free conformation, the next single bond closer to the peptide backbone is tested. If it is not possible to resolve the overlaps by rotating a ligand side chain, the same approach is used for the protein side chain whose atoms collide with the ligand. If an overlap-free conformation is found, the complex is evaluated by an empirically derived scoring function, which considers the dominant hydrophobic and hydrogen-bond terms, and was tuned on 30 PDB structures of complexes with small peptidyl ligands. Since a ligand is docked in a collision-free conformation, not necessarily in its optimal binding mode, the goal of the scoring function is not to model the binding affinity, but to eliminate molecules that lack chemical complementarity to the target and to provide a ranking for the potential ligands. However, since the screened molecules are already in a favorable conformation and only small conformational changes are made, it can be assumed that the final conformation is reasonable.

SPECITOPE has been tested by screening a database of 140,000 potential peptidyl ligands taken as overlapping fragments from the structures of 553 dissimilar ($<$ 25% sequence-identical) chains in the PDB. For four distinct targets, a serine protease (subtilisin), a DNA repair enzyme (human uracil-DNA glycosylase), an aspartic protease (rhisopuspepsin), and a glycosyltransferase (cyclodextrin glycosyltransferase), the database was screened in less than three hours on a desktop workstation (Sun SPARC Ultra/140) for ligands to the free target structures. For those cases where known peptidyl ligands existed and were included in the database, they were identified by SPECITOPE and ranked within the top five of the potential ligands found. For our four protein applications, active-site side-chain conformational change was required for docking the known ligand to the free protein structure, and SPECITOPE modeled this essential inducible complementarity correctly.

Currently, SPECITOPE is being extended for screening the Cambridge Crystallographic Database System (CSDS) of small organic compounds. The approach to deal with flexibility will be adapted to organic molecules by resolving overlaps with large rigid parts (ring structures) by protein side-chain rotation and ligand transformation, and for the remaining atoms by directed rotation of flexible bonds. During the conformational search, there will be an explicit optimization of hydrogen-bonding in addition to resolving collisions. The scoring function will be tuned for complexes with small organic compounds, and we will include terms for the hydrophilicity of the solvent exposed parts of the bound ligand and for rewarding good van der Waals packing. In the current version of SPECITOPE, the search is somewhat biased by the explicit choice of the four or five template points. Although this is desirable for proteins where much is known about ligand binding, for other cases we will stochastically sample over more template points and consider hydrophobic interactions in the template.

Reference:

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*, in review for Proteins: Structure, Function, and Genetics.

Contact:

Dr. Volker Schnecke
Protein Structural Analysis and Design Laboratory
Department of Biochemistry
Michigan State University
East Lansing, MI 48824–1319

WWW: http://www.bch.msu.edu/labs/kuhn/
Email: volker@sol.bch.msu.edu
Phone: (517) 355 3455
Fax: (517) 353 9334