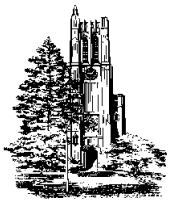
Realistic Modeling of Polar Interactions in Ligand Screening

Maria I. Zavodszky*, Paul C. Sanschagrin*, Leslie A. Kuhn*§

***Protein Structural Analysis and Design Laboratory, Department of Biochemistry and Molecular Biology, and §Center for Biological Modeling, Michigan State University, East Lansing, MI 48824, USA**

zavodszk@sol.bch.msu.edu



ABSTRACT

To adequately represent the binding site of a protein, SLIDE uses a set of hydrogen bonding and hydrophobic template points. The chemistry of the binding site is reflected by the labels attached to these points (acceptor, donor, donor/acceptor and hydrophobic). Their positions are chosen to reproduce sites of favorable interactions of ligand atoms with the protein while also reflecting the shape of the binding pocket.

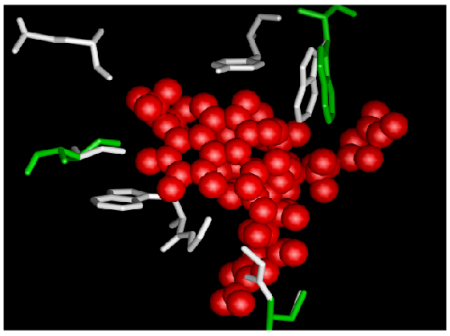
Currently it is computationally too expensive to include a very large number of template points (>150) for screening. Therefore, instead of randomly or uniformly sampling and then labeling the points above the protein surface, we are now applying a knowledge-based approach for identifying optimal positions for hydrogen bonding (presented here) and for representing significant hydrophobic surface patches (presented by Paul Sanschagrin). Hydrogen-bonding template points are now positioned according to the statistically favored stereochemistry of hydrogen bonds observed in high resolution crystal structures from the Protein Data Bank. Our results show that concentrating on the optimal sites for hydrogen bonding is superior over the random or uniform sampling of the binding site followed by the selective retention of points appropriate for positioning ligand hydrogen bond donors or acceptors. The new template design allows us to dock more known ligands into the binding site of the ligand-free thrombin structure than using the random sampling method. Also, for most of these ligands, the RMSD between the best-docked orientation and the position of that ligand from the crystal structure is lower. When combined with the new hydrophobic template design, further improvements in docking can be attained.

KEY FEATURES OF SLIDE

Our screening tool, SLIDE (Screening for Ligands by Induced-fit Docking Efficiently), is capable of screening through a database of 80 000 compounds within a day on a regular desktop workstation. It does this by using multi level hashing and distance geometry to rule out the infeasible ligand candidates as early as possible from the screening process and spend the most time-consuming final docking step on promising candidates.

Solvation of the binding site can also be taken into consideration. Consolv, a genetic-algorithm-based classifier developed in our laboratory, is applied to predict conservation of binding site water molecules upon ligand binding. Conserved waters are treated as part of the protein but can be displaced at a later step if the docked ligand collides with them.

Another novel feature of SLIDE is that it models flexibility by allowing both protein side chain rotations and full ligand flexibility. Our hypothesis (which applies well to thrombin and is under investigation for other proteins) is that both the protein and the ligand change their conformation as little as possible upon binding to each other. SLIDE uses mean-field theory based optimization to select the minimal set of bonds to be rotated in order to resolve interatomic collisions between the docked ligand candidate and the protein.



Side chains from the binding site of thrombin rotated by SLIDE (green) upon docking a known ligand (red) to it. The original position of the protein side chains are shown in white.

MAIN STEPS IN SCREENING WITH SLIDE

Preparation

Generate template with hydrogen bonding and hydrophobic interaction points for the binding site of the protein of interest.
Determine interaction points of the prospective ligand candidates.
Create hash tables indexing all possible template triangles.

↓

Docking

Identify feasible template triangles for each triplet of ligand interaction points.
Dock ligand into the binding site using triangles' least-squares-fit.

↓

Modeling induced fit

Identify rigid anchor fragment (determined by matched interaction point triangle) and flexible bonds in ligand.
Resolve anchor fragment - main chain collisions by ligand translation.
Resolve side chain collisions by directed rotations.

↓

Scoring

Score protein-ligand complex based on number of H-bonds and hydrophobic complementarity.

TEMPLATE DESIGN STRATEGIES

The binding site of the protein is described by a template of favorable interaction points onto to which ligand atoms are matched during the screening process. The chemistry of the binding site is represented by the labels attached to these points (acceptor, donor, donor/acceptor, hydrophobic) while their position also reflects the shape of the binding pocket. Because it is computationally too expensive to use a very large number of template points (>150) for screening, it is important to place those points where optimal interaction can occur between the protein and the ligand. Therefore, we focused on how to model favorable subsites for hydrogen bonding (described here) and hydrophobic interactions (presented by Paul Sanschagrin) using a knowledge based approach.

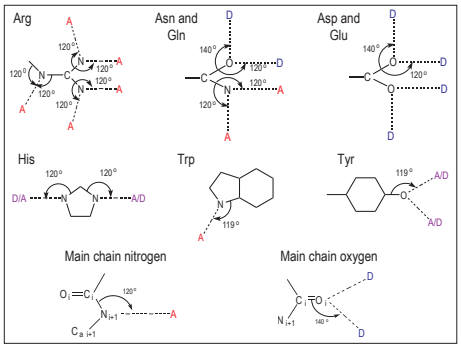
I. Previously used grid template generation:

- The binding site is filled with a large number of points (generally 10,000 to 40,000) placed on a fine grid (0.5 Å grid spacing).
- A layer of points, 2.5 to 5.0 Å away from any protein atom is selected.
- Each point is checked to determine if it could serve as a hydrogen bond donor, acceptor, or be a hydrophobic interaction point with the protein and is labeled as such. Those points that cannot be classified as any of the above types are eliminated.
- Points of the same class are clustered using complete linkage clustering to yield a feasible number of template points, usually around 150.

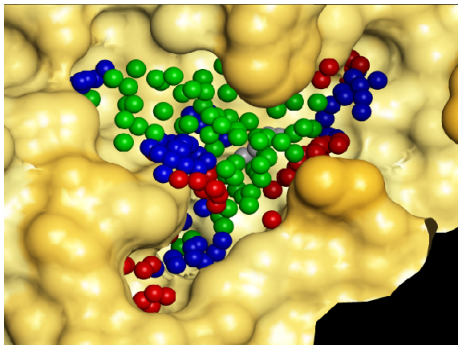
II. New hydrogen-bonding template generation:

- All atoms capable of hydrogen bonding are identified in the binding site.
- A certain number of template points (usually 10-15) are placed at and around the optimal hydrogen bonding position from each of these atoms.
- The template points are approximately at 1 Å distance from each other and labeled as donors, acceptors or donors/acceptors, depending on the atom type they belong to.
- Those template points that are closer then 2.5 Å from any protein atom are discarded.

After they are generated separately, the H-bonding and hydrophobic template points are merged into one template. If the total number of template points obtained this way is larger than 150, complete linkage clustering can be used to cluster points of the same class.



Optimal hydrogen bonding positions around protein atoms



The binding site of thrombin filled with template points colored according to type: donor - blue, acceptor - red, donor/acceptor - white, hydrophobic - green.

TESTING THE NEW METHOD

Hydrogen bonding templates containing approximately equal number of points were generated for the active site of the ligand-free thrombin (PDB code 1vr1) using the grid- and the knowledge-based methods. The same hydrophobic template points were merged with both hydrogen bonding templates.

We evaluated the new method of hydrogen bonding template generation by comparing the following results for grid-based and knowledge-based templates:
- numbers and scores of known ligands docked successfully
- differences between the docked orientations and the crystal structure positions of the known ligands (RMSD)
- percentage of known ligands among the top scoring dockings returned after screening trough the combined database of known ligands and CSD-compounds.

Screening databases:

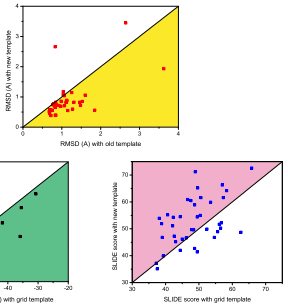
(1) 56 known thrombin ligands from crystal structures deposited into the Protein Data Bank (the list of PDB codes and names presented by Paul Sanschagrin).

(2) a subset of 14691 randomly selected compounds from the Cambridge Crystallographic Database (CSD).

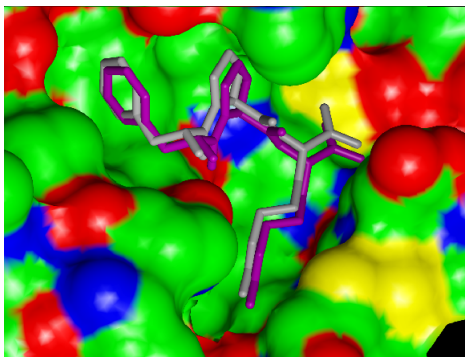
RESULTS

	New method	Grid-method
# of known ligands docked successfully	44/56	39/56
Combined database		
# known ligands among the top 10	6	3
# known ligands among the top 50	19	14
# known ligands among the top 100	29	24
Known ligands docked with both templates		
# higher scores as judged by SLIDE	25/39	10/39
# higher scores as judged by DrugScore	28/39	7/39
# ligands with lower RMSD values	34/39	5/39

Note: 4 out of the 39 known ligands docked to both templates received equal scores from both scoring functions.



These figures show the best RMSD values and the best scores given to the known ligands docked successfully by SLIDE using both templates. The points in the shaded areas highlight those dockings that received higher scores or were closer to the crystal structure position using the new template.



The active site of thrombin with the known ligand PPACK . The crystal structure position of the ligand is colored purple and the best orienation produced by SLIDE with the new template is grey.

CONCLUSIONS

The new template allows SLIDE to identify more known ligands and to dock them into the binding site of the protein in positions that are closer to the orientations observed in the crystal structures of the protein-ligand complexes.

The new, knowledge-based template generation method, which identifies optimal positions for hydrogen bonding of the ligand with the protein, proved to be superior over the previous method based on uniform sampling of the binding site.

REFERENCES

Gohlke, H., Hendlich, M., and Klebe, G. Knowledge-based scoring function to predict Protein-ligand interactions. J. Mol. Biol. 295, 337-356, 2000.

Ippolito, J. A., Alexander, R. S., Christianson, D. W. Hydrogen bond stereochemistry in protein structure and function. J. Mol. Biol., 215, 457- 1990.

McDonald, I., Thornton, J. M. Atlas of Side-Chain and Main-Chain Hydrogen Bonding. http://www.biochem.ucl.ac.uk/~mcdonald/atlas/.

Schnecke, V., Kuhn, L. A. Virtual screening with solvation and ligand-induced complementarity. Perspectives in Drug Discovery and Design, 20: 171-190, 2000.

Schnecke, V., Kuhn, L. A. Database screening for HIV protease ligands: the influence of binding-site conformation and representation on ligand selectivity. Intell. Syst. Mol. Biol., 242-251. AAAI Press, 1999.

Schnecke, V., Swanson, C.A., Getzoff, E.D., Tainer, J.A., Kuhn, L. A. Screening a peptidyl database for potential ligands to proteins with side-chain flexibility. Proteins, 33:74-87, 1998.

Raymer, M.L., Sanschagrin, P.C., Punch, W.F., Venkataraman, S., Goodman, E.D., Kuhn, L. A. (1997) Predicting conserved water-mediated and polar ligand interactions in proteins using a k-nearest-neighbors genetic algorithm. J. Mol. Biol. 265: 445-464, 1997.