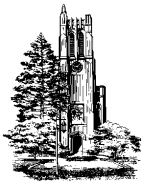
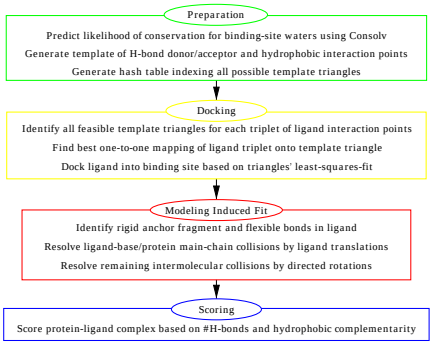
Screening and Docking of Flexible Organic Ligands to Solvated Binding Sites with Induced Complementarity

Volker Schnecke and Leslie A. Kuhn - Michigan State University



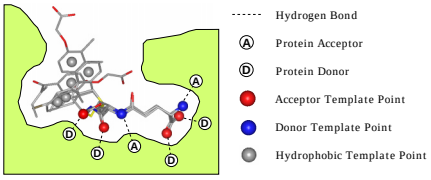
THE SCREENING-TOOL SLIDE

We present the tool SLIDE (for 'Screening for Ligands by Induced-fit Docking'), which is capable of screening datasets of more than 100,000 compounds within a few hours on a regular desktop workstation. The binding site of the protein is represented by a layer of surface residues, water molecules, and a template of favorable hydrogen-bond and hydrophobic interaction points. SLIDE uses multi-level hashing, distance geometry, and mean-field theory based optimization to efficiently dock potential ligands. It usually returns of up to 200 compounds ranked by the number of intermolecular hydrogen bonds and hydrophobic complementarity of the contact surfaces.



TWO TEMPLATE DESIGN STRATEGIES

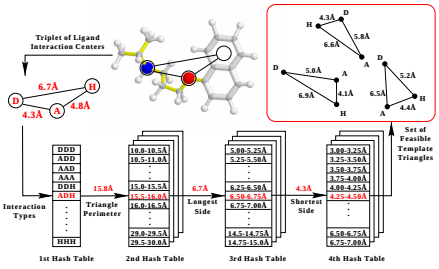
(1) Based on known ligand binding modes:



(2) Automatic template generation:

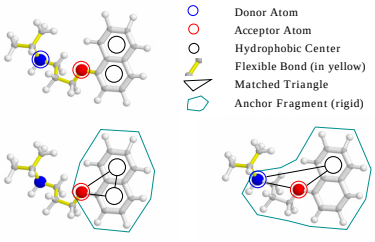
The binding site is filled with randomly placed probe points in feasible interaction distances to the protein surface. Each point is then checked for protein H-bond donors and acceptors or hydrophobic atoms in its proximity. Points of equal types are clustered to yield a template of ~100 points.

DIRECT ACCESS TO FEASIBLE TEMPLATE TRIANGLES BY MULTI-DIMENSIONAL HASHING



LIGAND DOCKING BASED ON TRIANGLE MAPPING

Each admissible one-to-one mapping of a ligand triplet onto a template triangle is used as a basis for docking the ligand into the binding site, and the three ligand points determine the rigid anchor fragment



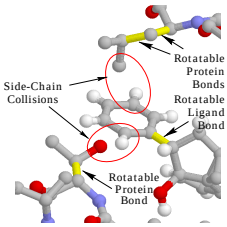
During the search, all possible triplets of ligand interaction centers are checked exhaustively, hence all potential anchor fragments are evaluated and used for docking the ligand, if feasible template triangles exist.

MODELING OF INDUCED FIT

Collisions of the anchor fragment with protein main-chain atoms are resolved by iterative rigid-body translations of the ligand.

For resolving side-chain collisions, an approximately optimal set of both ligand and protein side-chain rotations is determined by the application of mean-field optimization.

The choice of the bonds to be rotated is influenced by the number of atoms that will be displaced, by the rotation angle, and by interdependencies with other bonds in the system.



BINDING-SITE SOLVATION

Consolv, a k-nearest-neighbor/genetic algorithm classifier is used to predict the likelihood of conservation upon ligand binding for each water molecule in the binding site based on physical and chemical features of the local protein environment.

When modeling induced complementarity, waters with large overlaps with ligand atoms are displaced, whereas minor collisions are resolved by directed minimal rotations of ligand side chains.

When scoring the docked ligand, a penalty term based on Consolv's prediction is considered for each water molecule that is displaced by a non-polar atom.

SCORING OF COLLISION-FREE DOCKINGS

$$SCORE(P,L) = A * HPHOB(P,L) + B * HBOND(P,L) - B * DESOLV(P,L)$$

HPHOB(P,L)	hydrophobic complementarity of ligand atoms with their protein neighborhood
HBONDS(P,L)	number of intermolecular H-bonds (incl. water-mediated)
DESOLV(P,L)	number of waters displaced by non-polar ligand atoms, weighted by Consolv's prediction for likelihood of conservation

A and B were empirically tuned to reflect the binding affinities of 89 complexes available in the PDB.

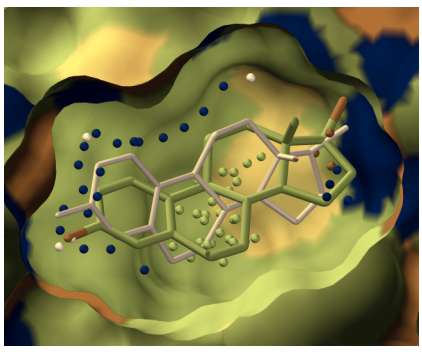
THE SCREENING DATABASES

(1) 67,573 organic compounds from the Cambridge Crystallographic Database System (CSDS), including all compounds of the database with fewer than 100 heavy atoms and at least three interaction points to be matched.

(2) 185,235 five-residue peptide fragments generated from 764 dissimilar chains in the Protein Data Bank (PDB-Select list 8/98).

THE HUMAN ESTROGEN RECEPTOR

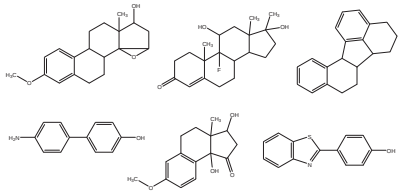
The CSDS subset was screened for ligands to the ligand-binding domain of the estrogen receptor (PDB 1ere). To avoid bias towards estradiol, all side chains in the binding site were energy minimized after removing the ligand. The template was generated automatically and consisted of 64 points. The total screening time was less than 3 hours on a Sun UltraSPARC 1/140.



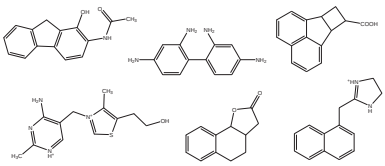
The figure shows the binding site with template points and estradiol in the binding mode from the crystal structure (white) together with an estradiol found in the database (CSDS entry ESTRDP), docked by SLIDE. It was ranked fourth in the set of 123 potential ligands that were identified.

CONTROLLING MOLECULAR DIVERSITY

Depending on size and representation of the template, the search can be biased towards ligands with similar shape and interaction patterns, or to return a large variety of novel potential ligands.



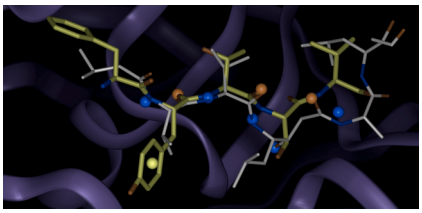
A sample of six potential ligands for the estrogen receptor is shown above. The screening was done with a six-point template based on the positions of the four rings and the two hydroxyl groups in the crystal structure of the bound estradiol in PDB entry 1ere. SLIDE identified and docked 72 ligands out of the 67,500 screened compounds within six minutes, over 80% of these were steroids and steroid-like compounds.



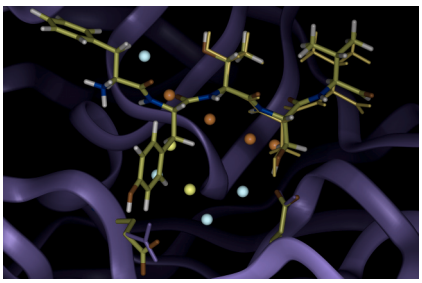
This is a sample of ligands found when screening with the unbiased 64-point template. SLIDE identified 123 potential ligands, and three estradiols were within the top-ten compounds. The total screening time was 160 minutes.

THE ASPARTIC PROTEASE RHIZOPUSPEPSIN

The 185,000 pentapeptides were screened for potential ligands to the aspartic protease rhizopuspepsin using a 9-point template based on average positions of hydrogen-bond donors/acceptors and carbon rings in three known ligands that were superimposed onto the ligand-free structure of the aspartic protease (PDB 2apr). Consolv predicted 10 binding-site waters in this crystal structure as being conserved upon ligand binding. SLIDE found 71 potential ligands, and the total runtime was four hours.



This figure shows the known ligand pepstatin superimposed from PDB 6apr (white), together with a peptide (Phe-Tyr-Thr-Ser-Val) docked by SLIDE. The spheres represent the nine template points.



This figure shows the modeling of induced fit for the same ligand. The conformation of the peptide in the database is shown in yellow, the conformations of protein side-chains in the ligand-free binding site are shown in purple. The final conformation of the complex is shown in green. Conserved binding-site waters are represented by blue spheres, waters displaced by polar ligands are shown in red and waters that were displaced by non-polar ligand atoms are shown in green.

ACKNOWLEDGEMENTS

We gratefully acknowledge funding for this project by the Deutsche Forschungsgemeinschaft (postdoctoral fellowship Schn-576/1-1 to V. S.), and the National Science Foundation (grant BIR 9600831 to L. K.).