Susan Tang
CS379: Case Study

ClusPro: an automated docking and discrimination method for the prediction of protein complexes

Docking is a very important method in the field of drug design. Most drugs are either small molecules or proteins, and by using a computational approach like docking researchers can rapidly screen vast databases of potential drug candidates against protein targets. The accuracy and efficiency of any docking methodology is dependent on two modules: the search algorithm and the scoring function. In the search algorithm, a select set of energetically reasonable poses of the ligand vs. target are explored to generate a large set of docked conformations. Then, a scoring function takes a pose as input and assigns a value to the pose, reflecting the probability that the pose depicts a binding interaction with high affinity. However, the problem with most existing docking programs is that the scoring functions cannot properly account for all factors attributing to interaction energetics, and thus yield a significant number of false positives. Therefore, it would be valuable to establish a method that can take the entire set of docked conformations and rank them using a more accurate approach than current scoring functions. This is precisely what the authors of the paper have performed.

Camacho et al. developed an algorithm that first filters the immense set docked conformations by an existing docking program using electrostatics and desolvation potential as criteria, and then applies hierarchical clustering of pairwise RMSD values to discriminate the filtered set of docked conformations. The output is a set of top ranked docked conformations that is likely to include 1 or more near-native conformation. In fact, when they tested the accuracy of the discrimination step, they found that when using the 2000 docked conformations of 48 protein-protein pairs, 31 pairs yielded top 10 predictions that included at least 1 near-native conformation with an average RMSD of 5 angstroms from the native structure. Their automated web server, ClusPro, ranked as high as #3 in the CAPRI (Critical Assessment of Predictions of Interactions) contest for the successful prediction of complex Nidogen-G3 and 1KLO, which had an RMSD of 6.5 angstroms from native.

In the filtering portion of their method, the authors used criteria that have strong contributions to the free energy of interaction. However, because protein-protein interactions can be composed of either strong charge complementarity or weak charge complementarity, the contribution of electrostatic potential and desolvation energy is weighted differently depending on the interaction. If the two proteins are of opposite charge, the contribution of electrostatics is much greater while if the two proteins are of like charge, the contribution of desolvation is stronger. Therefore, the authors kept these two factors separate when performing filtering, such that 500 structures with lowest desolvation energy is selected and 1500 structures with lowest electrostatic energy are selected. The desolvation potential is evaluated using atomic contact potential (ACP) and the electrostatic potential is evaluated using Coulombic potential w/ distance dependent dielectric of 4r. The greater number of electrostatic energy conformations chosen reflects the somewhat noisy, fluctuating nature of electrostatic potential.

Once the filtered set of docked conformations is obtained, the authors proceed with hierarchical clustering of the set to rank the conformations. For each conformation, they fix the receptor and select the residues in the ligand that are part of the binding interface. These residues are defined to be within 10 angstroms from any atom in the receptor. Using a particular conformation's set of ligand residues, they calculate pairswise RMSD values for this conformation with each of the other 1999 docked conformations. Therefore, a 2000 x 2000 matrix of pairwise RMSD's is generated and used for clustering, where the maximum distance from the center of the cluster is defined to be 9 angstroms RMSD. The cluster with the most members (conformations) is selected

first and its center (a set of coordinates) becomes the highest ranked prediction. Then, the next largest cluster is selected and its center becomes the second highest prediction, and so forth. The rationale behind this approach is that within the free energy landscape of partially solvated receptor-ligand complexes, the true binding site is usually characterized by a very broad local minima. This means that the most probable, near-native conformation will be one that has the greatest number of similar (neighboring) structures. The idea of a cluster with a defined radius captures this concept well, and allows the authors to distinguish which conformations have the most nearby structures.

ClusPro, an automated web server that integrates existing docking methods and the novel filtering / discrimination algorithm outlined in this paper, has important implications in protein drug discovery. Drugs work via different mechanisms, but two prevalent mechanisms are: 1) binding of receptors to stimulate or inhibit a cell signaling pathway 2) binding competitively to a protein to block an inappropriate or excessive protein interaction that is characteristic of the disease. By effectively discriminating among a large set of docked conformations and producing reliable predictions of near-native binding interactions, ClusPro allows researchers to search databases of protein drug candidates and select for those with the most favorable binding affinity based on the predicted docked conformation. Subsequently, researchers can pursue this refined set of drug candidates and perform experiments (ie: BiaCore analysis) that better forecast how strong the pair of proteins will bind in-vivo by measuring kinetic rate constants of the interaction in solution.

The specificity of a protein drug candidate can also be tested by applying ClusPro to a drug candidate and various other known receptors in the body. An ideal drug needs to have high binding specificity to its target and not bind degeneratively to other molecules. Binding non-specifically to other molecules may produce unwanted, fatal effects and render the candidate drug unsuitable as a lead. Overall, using computational methods to generate higher-confidence experimental candidates reduces time, cost, and labor.