Ligand-Protein Docking using Genetic Algorithms and Support Vector Machines

Najmanovich, R., Sobolev, V. and Edelman, M.
Department of Plant Sciences, Weizmann Institute of Science

Goals. Development of a ligand-protein docking algorithm based on genetic algorithms. Prediction of side chain flexibility upon ligand binding using support vector machines.

Background. Ligand-protein docking predictions aim at determining the structure of the ligand-protein complex given the protein atomic coordinates. In certain cases, one may consider both protein and ligand as rigid bodies, in other cases, it might be necessary to consider either the ligand, the protein or both molecules as completely or partially flexible. Genetic algorithms are useful in optimization tasks involving large number of variables and rough landscapes and may be suitable for docking simulations including ligand and side-chain flexibility. A previous work (Najmanovich et. al., 2000) determined that few side-chains undergo conformational changes upon ligand binding. To determine which side chains need to be set flexible, we utilize support vector machines in order to create a classifier system able to predict which side-chains are likely to be flexible during the docking simulation.

Results. We developed a new reproduction technique in our genetic algorithm implementation (Population Boom) that improves convergence. The search algorithm succeeds in finding optimal solutions in both global simulations (searching the whole protein surface) and local simulations (an approximate position for the binding site is known). Our preliminary results using support vector machines to predict side chain flexibility show a classification accuracy of 74.0%±0.9%. Moreover, when used to classify all side-chains on a protein, those side chains predicted to be flexible are present in the surface, in regions of high flexibility such as loops or chain termini.