Grid Generation and Matching for Small Molecule Docking

Peter Freddolino, John Stone, Klaus Schulten
Theoretical and Computational Biophysics, Beckman, UIUC

Introduction

Computational drug design requires the matching of libraries of thousands of drug molecules into the environment of a known target; this task is known as "docking". Computational docking approaches generally involve the generation of a large number of trial conformations for each drug molecule, followed by ranking of these conformations according to some scoring function. Both the conformational generation and scoring steps are quite time consuming, and methods to allow faster conformational searches would allow docking studies to be both more thorough (by testing more conformations) and to test more drug molecules.

Description

One of the most commonly used approaches to computational docking, used by UCSF DOCK, is to generate drug molecule conformations by matching the distances between points in the drug to distances between accessible points in the target. Scoring is then performed using a precalculated grid of energies, and the best scoring conformations are refined using a steepest-descent minimization algorithm.

Generation of the grids used in scoring involves the calculation of Coulombic (electrostatic) and Lennard-Jones (dispersive) energies due to the target (usually a protein) at each point on a grid surrounding the target. In addition, each grid point is flagged if it is too close to an atom in the target (known as a "bump") and if it is close enough to touch the target (known as a "contact"). The bump, contact, and energy grids are used during docking to allow rapid evaluation of how good a conformation is.

Once these grids are created, a drug molecule can be docked by generating a large variety of orientations of the molecule and scoring each conformation based on the interactions of its atoms with the scoring grid. In rigid docking, the drug is assumed to be fixed in shape, and conformations are generated by matching the distances between atoms in the drug with a set of generated points in space near the surface of the target. Conformations with more than a specified number of bumps (atoms on grid points where the bump flag is set) are eliminated, and the remaining conformations are ranked by energy as calculated from the energy grid.

Objective

The objective of the project is to implement the grid generation and rigid docking steps of DOCK as described above. Implementation of a carefully tuned form of grid calculation should be fairly easy to accomplish; an implementation of rigid docking will be somewhat challenging, but should also be possible to finish in the time allowed for the project. The grid generation implementation will be benchmarked using a set of target proteins ranging in size from 100 to 2000 atoms, and rigid docking of ligands will be benchmarked on a set of ligands ranging from 5 to 20 atoms, docked into a pregenerated set of 76 target points. Analysis of the limitations of the hardware for this problem is required, including a comparison of the different bottlenecks encountered for the grid generation and conformational sampling steps.

Background

No specialized knowledge is necessary to work on the application, since the algorithm to be implemented is straightforward. Background in basic physics and chemistry is useful to better appreciate the problem and the rationale for the steps being taken. Programming expertise is assumed.

Resources

Algorithm references, papers, and sample code will be made available online:
http://www.ks.uiuc.edu/Research/vmd/projects/ece498/

Contact Information

Peter Freddolino: petefred@ks.uiuc.edu
John Stone: johns@ks.uiuc.edu