Determining an Optimal Scoring Function for Protein Complex Predictions

Brian G Pierce1  Zhiping Weng2
bpierce@bu.edu  zhiping@bu.edu

1 Bioinformatics Program, Boston University, 44 Cummington Street, Boston, MA 02215, USA
2 Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, MA 02215, USA

Keywords: protein docking, minimization, simplex, scoring function

1. Introduction

Predicting the structures of protein complexes computationally (also referred to as protein docking) has become a valuable tool in bioinformatics research. An important aspect of this research is docking refinement, where a set of predictions is rescored and the structures are minimized in order to improve the structures of the predictions and the rankings of hits. We have been involved in two stages of the refinement:

1. Determining energy functions that will optimally rescore and rerank the predictions.
2. Using these energy functions to search 6 degrees of rotational/translational freedom to minimize structure predictions.

2. Methods

In order to distinguish between structure predictions, it is necessary to have an effective scoring function. Our scoring function has seven terms that are summed linearly: van der Waals attractive and repulsive, electrostatics short range attractive and repulsive, electrostatics long range attractive and repulsive, and desolvation. The first six terms are based generally on the scoring terms used in RosettaDock [3], while the desolvation term is based on the Atomic Contact Energy (ACE) of Zhang et al. [5]. These terms are chosen as they provide a reasonably comprehensive representation of the energies involved in protein complex formation, i.e. shape complementarity (van der Waals), electrostatics, and hydrophobicity (ACE).

With these terms as the basis for our energy function, we then determined how these terms should be weighted relative to one another to best improve the structures of our predictions. Our input dataset was produced using the crystal structures of complexes from our Benchmark 1.0 [2], perturbing these structures slightly in 6 degrees of freedom (3 rotational and 3 translational). For each perturbation, each energy term was output separately, along with the root mean square deviation (RMSD) of the atoms between the perturbation and the crystal structure. To produce a useful dataset it was necessary to determine proper spatial and angular domains for the perturbations. As the structures to be minimized are the output from our initial stage docking program ZDOCK [1], we used ZDOCK predictions and find how much they must be moved to line up with the crystal complexes. We found that 3.0 Å and 0.3 radians was an effective range for the perturbations.

The weighting for the energy terms was produced by minimization in the seven-dimensional (7-D) space of the energy term weights (ETWs). The value to be minimized, based on the RMSDs, was a “slope” score that effectively determined the sharpness of the energy funnel. These slope scores were then
summed over the entire benchmark for a given set of ETWs. To search the possible energy weights we used a 7-D downhill simplex as shown in Numerical Recipes [4] (see Figure 1). To avoid becoming trapped in local minima, it was found to be effective to perform several restarts both upon completion of the simplex and by reinitializing the simplex with a completely new starting position.

In addition, we used the 7-D simplex to determine weights for reranking ZDOCK predictions. In this case we used the number of hits in the top N predictions after reranking to score a particular set of ETWs.

3. Results

Here is a typical set ETWs for structure minimization as output by the simplex:

- van der Waals attractive = 1.0
- van der Waals repulsive = 0.01
- Electrostatics short range attractive = 0.29
- Electrostatics short range repulsive = 0.39
- Electrostatics long range attractive = 0.18
- Electrostatics long range repulsive = 0.49
- Desolvation = 0.61

Two of the weights shown above are lower than other weights in the same category. The repulsive van der Waals is lower most likely due to the necessity for docking to be flexible: that is, even though there is some degree of atomic overlap (clash) in the perturbation, the perturbation may have a good RMSD. The long range electrostatics attractive energy is low possibly due to charge shielding at such distances (4 to 12 Å); the repulsive force at long distances is apparently more important to enforce proper binding orientations.

It is clear that when using these weights to determine the score, there is an improvement in the shape of the energy funnel (see Figure 2).

4. Discussion

Using minimization techniques with our docking benchmark, we have determined optimal sets of weights to use in docking refinement. These weights shed some light upon the energy terms that are important in docking and protein interactions. In combination with several search algorithms, the weight sets have been applied to Benchmark 1.0 and are being used to improve the predictions for Benchmark 2.0.

References