



## Prediction of interacting proteins from homology-modeled complex structures using sequence and structure scores

<http://www.firstlight.cn> 2007-07-07

Protein-protein interactions support most biological processes, and it is important to find specifically interacting partner proteins among homologous proteins in order to elucidate cellular functions such as signal transduction systems. Various high-throughput experimental methods for identifying these interactions have been invented, and used to generate a huge amount of data. Because these experiments have been applied to only a few organisms, and their accuracy is believed to be limited, it would be valuable to develop computational methods for predicting protein-protein interactions from their amino acid sequences or tertiary structural information. In this study, we describe a prediction method of interacting proteins based on homology-modeled complex structures. We employed the statistical residue-residue contact energy used in a previous study, and two types of new scores, simple electrostatic energy and sequence similarity between target sequences and template structures. The validity of each protein-protein complex model was measured using their single and combined scores. We applied our method to all the protein heterodimers of *Saccharomyces cerevisiae*. To evaluate the prediction performance of our method, we prepared two types of protein-protein interaction dataset: a complete dataset and high confidence dataset. The complete dataset (10,325 protein dimer models) contains all the yeast protein heterodimers whose complex structures can be modeled. Among them, pairs registered in the DIP database are defined as interacting pairs, and those not registered are defined as non-interacting protein pairs. The high confidence dataset (3,219 protein dimer models) is a more reliable subset of the complete dataset extracted using the criteria of the common subcellular localization. Both datasets show that sequence similarity has a much higher discrimination power than the other structure-based scores, but that the inclusion of contact energy results in significant improvement over predictions using sequence similarity alone. These results suggest that the sequence similarity is indispensable for the prediction, whereas structure scores can play supporting roles.

[存档文本](#)