A functional ontology-based score for template-based protein docking

Anna Hadarovich[1], Ivan Anishchenko[1,2], Petras J. Kundrotas[2], Ilya A. Vakser[2], Alexander V. Tuzikov[1]

[1]United Institute of Informatics Problems, National Academy of Sciences, 220012, Minsk, Belarus
ahadarovich@gmail.com
[2]Center for Computational Biology and Department of Molecular Biosciences, The University of Kansas, Lawrence, Kansas 66047, USA

Abstract. Comparative docking generates structures of protein-protein complexes using known structures as the templates for modeling. The detection of templates is often based on structural similarity between the modeling target complex and the experimentally determined complexes. The performance of the template-based docking significantly decreases when the templates are only moderately similar to the target. We developed a non-structural score, which is complementary to the structural similarity scoring, based on the target/template similarity of the Gene Ontology annotations. A combined scoring function improves discrimination of wrong templates, enhancing the reliability of the docking.

Keywords: protein-protein interactions, modeling of protein complexes, gene ontology terms

Structural characterization of protein-protein interactions (PPI) is important for understanding life processes at the molecular level. Experimental techniques, due to their inherent limitations, can determine structures only for a fraction of known PPI. Thus most PPI have to be modeled by docking techniques. Template-based docking generates structures of complexes using known structures as the templates for modeling. It is increasingly the method of choice for large-scale structural studies of PPI [1]. The detection of templates is often based on structural similarity between the modeling target complex and the experimentally determined complexes, quantified by TM-score from TM-align algorithm [2]. However, the performance of the template-based docking significantly decreases when the templates are only moderately similar to the target (TM-

score 0.4 – 0.6) [3]. We developed a non-structural score, which is complementary to the structure-based scoring, based on the target/template similarity of the Gene Ontology (GO) annotations [4] (GO-score). A number of algorithms that determine similarity between single GO-terms [5] were applied to protein-protein complexes. Previous studies on similarity of the GO-terms [6,7] utilized only the most populated "molecular function" domain of the GO annotations. We showed that taking into account all three ontology domains (molecular function, biological process, and cellular component) significantly improves detection of the templates. A combined scoring function consisting of the TM-score and the three GO-scores was developed and tested on a non-redundant set of 587 protein-protein complexes, using 4,950 template structures [8] from the DOCKGROUND resource [9]. The results show that the GO-scores improve discrimination of wrong templates, enhancing the reliability of the docking.

References

1. Kundrotas, P., et al.: Templates are available to model nearly all complexes of structurally characterized proteins. Proceedings of the National Academy of Sciences. 109, 9438-9441 (2012).
2. Zhang, Y., et al: TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Research. 33, 2302-2309 (2005).
3. Negroni, J., et al.: Assessing the applicability of template-based protein docking in the twilight zone. Structure. 22, 1356-1362 (2014).
4. Gene Ontology Consortium: Gene Ontology annotations and resources. Nucleic Acids Research. 41, D530-D535 (2012).
5. Couto, F., et al.: Measuring semantic similarity between Gene Ontology terms. Data & Knowledge Engineering. 61, 137-152 (2007).
6. Pesquita, C., et al.: Metrics for GO based protein semantic similarity: a systematic evaluation. BMC Bioinformatics. 9, S4 (2008).
7. Lord, P., et al.: Investigating semantic similarity measures across the Gene Ontology: the relationship between sequence and annotation. Bioinformatics. 19, 1275-1283 (2003).
8. Anishchenko, I., et al.: Structural templates for comparative protein docking. Proteins. 83, 1563-1570 (2015).
9. Douguet, D., et al.: DOCKGROUND resource for studying protein-protein interfaces. Bioinformatics. 22, 2612-2618 (2006).