Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Advances in template-based protein docking by utilizing interfaces towards completing structural interactome

Serena Muratcioglu[1,3,4], Emine Guven-Maiorov[1,3,4], Ozlem Keskin[1,3] and Attila Gursoy[2,3]

The increase in the number of structurally determined protein complexes strengthens template-based docking (TBD) methods for modelling protein–protein interactions (PPIs). These methods utilize the known structures of protein complexes as templates to predict the quaternary structure of the target proteins. The templates may be partial or complete structures. Interface based (partial) methods have recently gained interest due in part to the observation that the interface regions are reusable. We describe how available template interfaces can be used to obtain the structural models of protein interactions. Despite the agreement that a majority of the protein complexes can be modelled using the available Protein Data Bank (PDB) structures, a handful of studies argue that we need more template proteins to increase the structural coverage of PPIs. We also discuss the performance of the interface TBD methods at large scale, and the significance of capturing multiple conformations for improving accuracy.

Addresses
[1] Department of Chemical and Biological Engineering, Koc University, 34450 Istanbul, Turkey
[2] Department of Computer Engineering, Koc University, 34450 Istanbul, Turkey
[3] Center for Computational Biology and Bioinformatics, Koc University, 34450 Istanbul, Turkey

Corresponding author: Gursoy, Attila (agursoy@ku.edu.tr)
[4] S.M., E.G.M. contributed equally to this work.

Current Opinion in Structural Biology 2015, **35**:87–92

This review comes from a themed issue on **Protein–protein nteractions**

Edited by **Ozlem Keskin** and **Alexandre Bonvin**

http://dx.doi.org/10.1016/j.sbi.2015.10.001

Introduction

Proteins, their diverse 3D structures and interactions lead to a large repertoire of cellular functions. Proteins may form homo-oligomers, hetero-oligomers, transient or permanent complexes to perform their functions. A better understanding of the molecular basis of the protein functions necessitates characterization of the protein–protein interactions (PPIs) at the atomic level. PPIs can be determined by high-throughput experiments, such as yeast-two-hybrid, protein arrays, pull-down experiments, and mass spectrometry. However, these techniques do not provide 3D structural details on how proteins interact. Construction of the whole structural interactome is one of the major objectives of structural biology [1,2,3••,4]. Structures of PPIs depict how signals are relayed to downstream effectors, how regulation is achieved, and whether parallel pathways can be activated simultaneously [5,6]. They also help in figuring out mutation mechanisms [7] and in drug discovery [8].

Structures can be obtained by experimental techniques, such as X-ray crystallography, NMR, and electron microscopy (EM). X-ray and NMR provide high-resolution structures of macromolecules, whereas EM offers low-resolution of larger macromolecular assemblies. Although the number of resolved protein structures increases rapidly with advances in the experimental techniques, the structural space of the whole proteome is still far from complete [9]. The number of experimentally identified interactions is much greater than the number of resolved PPI structures [2]. Therefore, computational approaches can fill in the gap that experiments cannot resolve [10,11•,12,13]. Computational modelling of residue-level details of the PPIs is a promising way to elucidate the comprehensive interactome and overall dynamic picture of the cells [3••,4,9,10,11•,14].

Various computational approaches have been developed to forecast the structures of PPIs, through co-evolution, co-expression, sequence similarity, and structural similarity [2,13,15,16]. These methods can be grouped into two major classes: *ab initio* docking and template-based docking [17,18••,19]. *Ab initio* docking does not require any prior knowledge and depends on shape and electrochemical complementarity [1]. However, template-based docking (TBD) employs experimentally determined structures (templates) to model the structures of target proteins (targets) [11•,15,20,21]. Detailed assessment studies demonstrated that TBD strategies generate more reliable results [11•,19] and they are more successful at predicting the structures of PPIs that undergo conformational changes [18••,22,23]. In addition, it is more convenient to use TBD in large-scale proteome-wide applications, since they lower the computational cost and have less false-positive rates [19,23].

There are different evaluation criteria relating targets to templates in TBD: overall sequence similarity, sequence similarity of the predicted interface, sequence and structural motifs derived from known interactions, co-evolving residues in the interface, global structural similarity, local structural similarity, and the combination of both sequence and structural similarities [11•,12,18••,21,24,25]. Methods based on global similarity have been successfully used [24] in cases where more than 30% sequence similarity exists [26••]. Many of the proteins, however, do not have sufficient sequence identity, thus, such methods are not applicable at large scale. Using targets from previous CAPRI (Critical Assessment of Prediction of Interaction) experiments, Rodrigues *et al.*, however, showed that poor models built by using templates having as low as 20% sequence identity, can still yield acceptable predictions (within 3 Å interface RMSD — iRMSD), provided that the interface information is reliable [27]. Local similarity based methods, on the other hand, can be applied to cases of low sequence identity to obtain high-quality predictions. These methods have become more popular for large-scale predictions particularly after several studies reporting that protein interfaces are sufficient to cover most PPIs [28,29•,30].

Predicting and modelling PPIs using template interfaces

An interface structure consists of the binding regions of two interacting proteins. Known interface structures can be used as templates. Given two target proteins, the first step in an interface TBD method is to search for a template such that one of the target proteins aligns to one side of the template, and the other target protein aligns to the other side of the same template. The next step is to superimpose the target proteins onto the matched template to construct the model. Potentially there might be many models from different templates, thus, in the last step; these models are ranked using scoring functions. Figure 1a shows the basic flow of the TBD methods that exploit interfaces at large scale.

PRotein Interactions by Structural Matching (PRISM) is the first interface based TBD algorithm (to our knowledge) to predict the structures of PPI complexes by employing only interface structure similarity (not necessarily fold similarity) without sequence or global structural similarity [15]. Later on, PRISM was extended with flexible refinement [20]. The inputs are the structures of unbound proteins (targets) and the template interface dataset. The target input is not limited to monomers, that is, multimeric proteins can also serve as the input. The output is the structural models of the PPI complexes. After the structural alignment phase carried out by Multiprot [31], the resulting models are filtered using a combination of structural and evolutionary constraints, such as RMSD, the number of contacting residues, and the number of matching hotspots (critical residues at interfaces). The output structures are flexibly refined to relieve the steric clashes and then finally scored using FiberDock [32]. The current template library is comprised of 22 604 non-redundant template interface structures constructed from the co-crystallized complex structures in PDB [33]. The success of PRISM has been assessed in [23] using Docking Benchmark [34] and in [19] with other approaches. Overall, it finds acceptable models for most cases with low conformational changes upon binding. Some proteins undergo large conformational changes and computational prediction/modelling tools perform poorly in such cases. The presence of multiple conformations of a protein in PDB, such as bound, unbound, and post-translationally modified conformations increases the success of TBD methods. Kuzu *et al.* [22] showed that even the difficult cases with large conformation changes in the docking benchmark could be accurately predicted if the structures of multiple conformations are utilized.

Table 1 lists the PPI prediction methods that employ interface knowledge [13,16,17,35–38]. Some of them, like PrePPI, make use of machine learning approaches in which both structural and non-structural clues, such as sequence, co-expression, and co-localization data are combined [13]. The integration of such diverse and independent data can increase the accuracy. HADDOCK [39] is also a method that benefits from data integration.

Another advantage of interface TBD methods over template-free methods is the lower computational time, an important factor for genome-wide predictions. This major advantage stems from the fact that the number of unique interface templates (K) is smaller than the size of the proteome (N) (Figure 1b) [33,40]. The computational time decreases from $O(N^2)$ docking operations of template-free approaches to O(NK) structural alignment operations of interface based TBD approaches (Figure 1c) [23].

Rationale on using template interfaces

Similar protein–protein interfaces are observed between different proteins [28,40,41]. Reuse of similar interfaces provides an explanation to why interface-based template docking is successful for modelling PPIs. A study conducted by Keskin and Nussinov [41] reveals that similar binding structures occur even in the absence of global structural similarity. The interfaces derived from the multi-chain PDB entries are classified into three types based on the similarity of global structures of corresponding complexes: (a) similar interfaces derived from similar global structures, (b) similar interfaces derived from different global structures, and (c) interfaces whose members have only one side similar derived from dissimilar global structures. Detailed analysis of these interface clusters indicates that despite having globally different structures, proteins can interact with different partners using common interface motifs. These proteins prefera-