PRISM: a web server and repository for prediction of protein–protein interactions and modeling their 3D complexes

Alper Baspinar[1], Engin Cukuroglu[1], Ruth Nussinov[2,3], Ozlem Keskin[1,*] and Attila Gursoy[1,*]

[1]Center for Computational Biology and Bioinformatics and College of Engineering, Koc University, 34450 Istanbul, Turkey, [2]National Cancer Institute, Cancer and Inflammation Program, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., National Cancer Institute, Frederick, MD 21702, USA and [3]Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

ABSTRACT

The PRISM web server enables fast and accurate prediction of protein–protein interactions (PPIs). The prediction algorithm is knowledge-based. It combines structural similarity and accounts for evolutionary conservation in the template interfaces. The predicted models are stored in its repository. Given two protein structures, PRISM will provide a structural model of their complex if a matching template interface is available. Users can download the complex structure, retrieve the interface residues and visualize the complex model. The PRISM web server is user friendly, free and open to all users at http://cosbi.ku.edu.tr/prism.

INTRODUCTION

Protein–protein interactions play crucial roles in all biological processes. Recent studies indicate that while the exact number of human PPIs is unknown, estimates range from 130 000 (1) to 650 000 (2). The structures of protein–protein complexes provide interaction details which are crucial for understanding mechanisms of binding, signaling, regulation and effects of mutations on protein function. The number of known PPIs has increased dramatically recently. However, the gap between interactions known to take place and the structures of the complexes continues to widen (3–5). Experimental methods to determine the structures of complexes are time consuming, expensive and challenging to apply on a large scale. Computational approaches are becoming increasingly important as large amounts of sequence and structural data become available. As a computational approach, template-based protein–protein interaction prediction tools (3,6–9) are widely used. These computational techniques also provide valuable insights for protein engineering and drug discovery (10–12). Hence, more efficient and less error prone computational techniques for protein–protein interaction prediction and structural modeling are of paramount importance in the biological sciences.

Here, we present a server version of our template-based protein–protein interaction prediction tool Protein Interactions by Structural Matching (PRISM) Protocol (9,13). The server allows users to carry out protein–protein interaction predictions and model their structural complexes. The users can browse and visualize the results through a web-browser, without any configuration effort on their local machines. The server also stores the prediction results and models for fast future access. We envisage that the repository will grow and become an invaluable resource for the community. This web server is publicly available at http://cosbi.ku.edu.tr/prism. It is free and open to all users, and no login is required.

THE PRISM WEB SERVER

The PRISM rationale reasons that if patches of surfaces of two target proteins are similar to two complementary sides of a template interface and they have similar evolutionary conservation of putative binding residue 'hot spots' (14,15), they may interact to form a complex. Their interface architecture will resemble that of the template. The template protein interfaces are extracted from the Protein Data Bank (PDB) (16). The target proteins are provided by users with PDB and chain IDs; or users can upload their own structures in PDB format. They are processed for prediction of possible interaction sites between them. Our previous studies showed that PRISM generates fast and accurate predictions for protein–protein interactions (10,11). While the first version of PRISM in 2005 (17) only consid-

ered rigid structural similarity, the current PRISM protocol (9) adds flexible refinement and energy minimization. The backbone flexibility is modeled by the first 50 normal modes using Fiberdock (18). For side-chain flexibility, Fiberdock uses a rotamer library and finds optimum combination of rotamers with the lowest total energy. Finally, binding energy scores are found using CHARMM52 force field (19). Flexible refinement resulted in improved predictions as reported in (10). Another change is the definition of the target set. The old server separated the targets into their chains automatically and eliminated the sequentially similar chains. The new server does not perform any elimination, gives more control to the user to define the target structures. The template set used in the new PRISM is the latest version of the interface representatives extracted from PDB (20). In addition to the algorithmic differences, the design of the server is improved significantly to increase the performance. The new server keeps the intermediate results of the rigid dockings between a target and all templates in a database table. If a new query involves a target that has been processed before, the rigid docking results are re-used instead of recomputing. The old server was not intended for online calculations and had a very limited interactive usage. The new one is designed to perform online calculations using cluster computing. However, there are some limitations when using the PRISM server. Firstly, the PRISM method requires the structure of the targets, therefore is limited by the available protein structures. Secondly, if target proteins undergo conformational change upon binding, the PRISM requires structures of corresponding conformations for accurate prediction. Finally, the template set used in predictions does not cover all possible protein–protein modes of interactions. Nevertheless, the size and coverage of the PDB increases exponentially, we expect that the usage of PRISM will increase. Below, we outline the key concepts of the methodology of the protocol and its server.

TEMPLATE INTERFACE DATASET

A protein–protein interface is described as the contact region between two interacting proteins. Two residues are defined as interacting if the distance between any two atoms of the two residues from different proteins is less than the sum of their corresponding van der Waals radii plus 0.5 Å. If the distance between alpha carbon atoms of non-interacting residues and interacting residues in the same protein is smaller than 6 Å they are called nearby residues. The interface scaffold is built by combining both interacting and nearby residues. Evolutionary information is represented by computational hot spots which are predicted and flagged in the template interface (21,22). Some hot spots provide specificity to the protein complex whereas some others contribute to stability (21). The template dataset contains 22 604 structurally non-redundant interface structures. The details of the template dataset is provided in our previous work (20).

TARGET PROTEINS

Target proteins are structures in PDB format provided either as PDB IDs (structures automatically downloaded from PDB site) or input files by the user. The PRISM Server uses the target protein pairs to predict potential interactions. The surfaces of these proteins are structurally compared with the template interfaces to find if they interact directly. It is advisable to use high-resolution (i.e. ≤2.0 Å) structures to obtain reliable results.

PREDICTING AND MODELING PROTEIN–PROTEIN INTERACTIONS

Surface residues of the target proteins are extracted using the relative accessible surface area values (calculated by NACCESS (23)). Each interface in the template interface dataset is split into its constituent chains. The extracted surfaces are structurally aligned with each side of the split interfaces of all template interfaces. PRISM searches whether complementary sides of a template interface are structurally similar to any region on the surface of target structures using the MultiProt structural comparison engine (24). Once structural and hot spot (21,22) similarities are detected, the two target proteins are transformed onto the structurally similar template interface constituting a predicted complex structure. The complex is assessed with FiberDock (18) to resolve steric clashes, especially of side chains, and rank the putative complexes by the global energy binding score. The detailed description of the method is detailed in Tuncbag *et al.* (9).

PRISM WEB SERVER USAGE

The PRISM web server runs the PRISM's prediction algorithm (9,13). Users can access the PRISM functionalities through three pages:

(1) the PRISM main page for predicting protein–protein interactions,

(2) the Predictions page to browse the accumulated results in the database,

(3) the Templates page to browse the protein–template interfaces used for predictions.

PRISM MAIN PAGE

Users can perform two tasks: the first task (two proteins) for predicting interactions between two proteins and the second task (network) for predicting interactions in a network of proteins. To predict interactions between two proteins, users need to supply two PDB IDs. Optionally, users can specify the chains used in the predictions. For example, if the users would like to predict interactions of only the 'A' chain of 2ghu, 2ghuA should be entered as target1 (if the users do not provide chain ID all the chains will be used). Similarly, users need to provide a PDB ID for target2 (i.e. 1cew). Then, users can submit the prediction request (Figure 1a). The PRISM will use the default template set to execute the PRISM algorithm. The request will be put in a job queue to be executed in a cluster environment dedicated to the PRISM server. Users will be given a link to follow the progress of their job status. Additionally, users can provide

Figure 1. Overview of PRISM main page. (**a**) Two proteins interaction prediction. (**b**) Network interaction prediction.
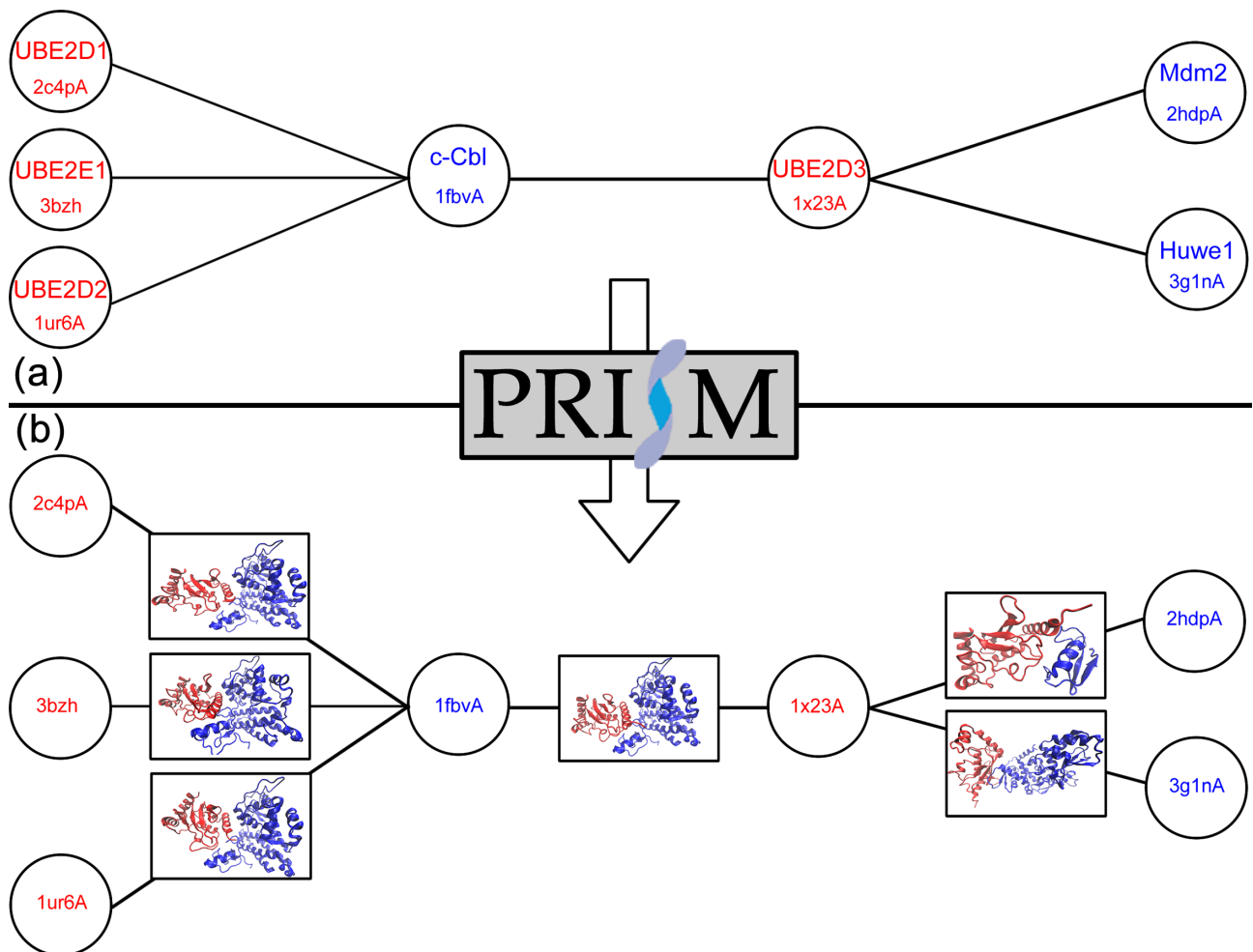

Figure 2. Modeling a small network of protein–protein interactions. (**a**) A node-edge protein–protein interaction network representation of the proteins (UBE2D1, UBE2E1, UBE2D2, UBE2D3, c-Cbl, Mdm2 and Huwe1) is taken from the ubiquitination pathway. (**b**) Six pairs of proteins are given to network prediction task and the results are shown in network representation of proteins as nodes and protein–protein interactions as edges. These proteins and their predicted complexes are shown in boxes.

Figure 3. Overview of the prediction results. Users can search the results by target or interface. The targets' proteins are shown in Target1 and Target2 columns. Interface column shows the template interface name. Energy column shows the FiberDock (18) energy value of the final structure. View button opens a JSmol (25) visualization of the final structure.

their e-mail addresses in the optional e-mail field to be notified when their jobs are submitted and their jobs are completed.

The network prediction is used to predict interactions in a network of proteins. In the pair-list box, the edges of the network are listed as pairs of targets (i.e. target1, target2) where each target is a PDB ID with or without chains as explained above (Figure 1b). The number of edges is limited to 10 pairs due to the heavy computational load. A small network of protein–protein interactions is shown in Figure 2. Six pairs of proteins are given to the network prediction task and the results are shown in a network representation (proteins as nodes and protein–protein interactions as edges) in Figure 2. These proteins (UBE2D1, UBE2E1, UBE2D2, UBE2D3, c-Cbl, Mdm2 and Huwe1) are taken from the ubiquitination pathway and their predicted complexes are shown in boxes.

In the examples above, the targets are protein structures from the PDB. Alternatively, users can upload their own structures in PDB format using the Upload Target buttons. The prediction results will be available to the user for downloading, but the result will not be stored in the PRISM database. PRISM uses the default template set. If users want their own templates, they need to provide PDB ID with two chains (i.e. 1stfEI). PRISM will extract the interface and will use it to perform predictions. The results will not be stored in the PRISM database.

A two-protein prediction might require several hours of computation time. The running time depends on the target proteins' sizes and the number of matched interfaces after structural alignment step. However, if the results or some intermediate results are already in the database from previous requests, the response will be given instantly.

PREDICTIONS PAGE

The Predictions page is used to browse the predictions found and stored in the database so far (Figure 3). The prediction results can be searched by target or interface IDs. For each prediction, PDB and chain IDs of the targets, the interface used in the prediction and the energy score is listed. Additionally, the visualization of the complex structure can be accessed with the view button. The visualization window has a download button as well. The temperature factors of the downloaded PDB file are replaced with interface and non-interface residue information: '1' and '3' for non-interface and interface residues of target1, respectively; '2' and '4' for non-interface and interface residues of target2, respectively. The downloaded PDB file can be easily visualized with other visualization tools using temperature factor coloring methods. In addition to PDB file, the server presents the list of interface residues and their contacts with the 'Contacts of Interface Residues' link in the view window (Figure 3).

TEMPLATES PAGE

The default template set can be browsed and searched on this page. The details of each interface can be accessed by clicking on the interface ID.

CONCLUSION

PRISM has become one of the most widely used computational protein–protein interaction prediction and modeling tools. It generates fast and reliable results using structural complementary of protein binding sites. PRISM results have appeared in the literature in numerous publications. Different from the earlier PRISM, in this work, we describe the web server which is an interactive protein–protein interaction prediction service that runs the PRISM algorithm for

user input structures and provides a searchable repository.

ACCESSION NUMBERS

PDB IDs: 2ghu, 2ghuA, 1cew and 1stfEI.

FUNDING

REFERENCES

1. Venkatesan,K., Rual,J.F., Vazquez,A., Stelzl,U., Lemmens,I., Hirozane-Kishikawa,T., Hao,T., Zenkner,M., Xin,X., Goh,K.I. *et al.* (2009) An empirical framework for binary interactome mapping. *Nat. Methods*, **6**, 83–90.
2. Stumpf,M.P., Thorne,T., de Silva,E., Stewart,R., An,H.J., Lappe,M. and Wiuf,C. (2008) Estimating the size of the human interactome. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 6959–6964.
3. Mosca,R., Ceol,A. and Aloy,P. (2013) Interactome3D: adding structural details to protein networks. *Nat. Methods*, **10**, 47–53.
4. Tyagi,M., Hashimoto,K., Shoemaker,B.A., Wuchty,S. and Panchenko,A.R. (2012) Large-scale mapping of human protein interactome using structural complexes. *EMBO Rep.*, **13**, 266–271.
5. Kuzu,G., Keskin,O., Gursoy,A. and Nussinov,R. (2012) Constructing structural networks of signaling pathways on the proteome scale. *Curr. Opin. Struct. Biol.*, **22**, 367–377.
6. Mukherjee,S. and Zhang,Y. (2011) Protein-protein complex structure predictions by multimeric threading and template recombination. *Structure*, **19**, 955–966.
7. Guerler,A., Govindarajoo,B. and Zhang,Y. (2013) Mapping monomeric threading to protein-protein structure prediction. *J. Chem. Inf. Model.*, **53**, 717–725.
8. Lu,L., Lu,H. and Skolnick,J. (2002) MULTIPROSPECTOR: an algorithm for the prediction of protein-protein interactions by multimeric threading. *Proteins*, **49**, 350–364.
9. Tuncbag,N., Gursoy,A., Nussinov,R. and Keskin,O. (2011) Predicting protein-protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM. *Nat. Protoc.*, **6**, 1341–1354.
10. Tuncbag,N., Keskin,O., Nussinov,R. and Gursoy,A. (2012) Fast and accurate modeling of protein-protein interactions by combining template-interface-based docking with flexible refinement. *Proteins*, **80**, 1239–1249.
11. Kuzu,G., Gursoy,A., Nussinov,R. and Keskin,O. (2013) Exploiting conformational ensembles in modeling protein-protein interactions on the proteome scale. *J. Proteome Res.*, **12**, 2641–2653.
12. Kundrotas,P.J. and Vakser,I.A. (2013) Global and local structural similarity in protein-protein complexes: implications for template-based docking. *Proteins*, **81**, 2137–2142.
13. Aytuna,A.S., Gursoy,A. and Keskin,O. (2005) Prediction of protein-protein interactions by combining structure and sequence conservation in protein interfaces. *Bioinformatics*, **21**, 2850–2855.
14. Bogan,A.A. and Thorn,K.S. (1998) Anatomy of hot spots in protein interfaces. *J. Mol. Biol.*, **280**, 1–9.
15. Clackson,T. and Wells,J.A. (1995) A hot spot of binding energy in a hormone-receptor interface. *Science*, **267**, 383–386.
16. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
17. Ogmen,U., Keskin,O., Aytuna,A.S., Nussinov,R. and Gursoy,A. (2005) PRISM: protein interactions by structural matching. *Nucleic Acids Res.*, **33**, W331–W336.
18. Mashiach,E., Nussinov,R. and Wolfson,H.J. (2010) FiberDock: flexible induced-fit backbone refinement in molecular docking. *Proteins*, **78**, 1503–1519.
19. MacKerell,A.D., Bashford,D., Bellott,M., Dunbrack,R.L., Evanseck,J.D., Field,M.J., Fischer,S., Gao,J., Guo,H., Ha,S. *et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **102**, 3586–3616.
20. Cukuroglu,E., Gursoy,A., Nussinov,R. and Keskin,O. (2014) Non-redundant unique interface structures as templates for modeling protein interactions. *PLoS One*, **9**, e86738.
21. Cukuroglu,E., Gursoy,A. and Keskin,O. (2012) HotRegion: a database of predicted hot spot clusters. *Nucleic Acids Res.*, **40**, D829–D833.
22. Tuncbag,N., Gursoy,A. and Keskin,O. (2009) Identification of computational hot spots in protein interfaces: combining solvent accessibility and inter-residue potentials improves the accuracy. *Bioinformatics*, **25**, 1513–1520.
23. Hubbard,S.J. and Thornton,J.M. (1993) 'NACCESS', Computer Program. Department of Biochemistry and Molecular Biology, University College London5
24. Shatsky,M., Nussinov,R. and Wolfson,H.J. (2004) A method for simultaneous alignment of multiple protein structures. *Proteins*, **56**, 143–156.
25. Hanson,R.M., Prilusky,J., Renjian,Z., Nakane,T. and Sussman,J.L. (2013) JSmol and the next-generation web-based representation of 3D molecular structure as applied to proteopedia. *Isr. J. Chem.*, **53**, 207–216.