Bioinformatics, 38(21), 2022, 4962–4965
https://doi.org/10.1093/bioinformatics/btac633
Advance Access Publication Date: 19 September 2022
Applications Note

Structural bioinformatics

HMI-PRED 2.0: a biologist-oriented web application for prediction of host–microbe protein–protein interaction by interface mimicry

Hansaim Lim [1], Chung-Jung Tsai[1], Ozlem Keskin[2], Ruth Nussinov[1,3,*] and Attila Gursoy [4,*]

[1]Computational Structural Biology Section, Frederick National Laboratory for Cancer Research in the Cancer Innovation Laboratory, NCI-Frederick, Frederick, MD 21702, USA, [2]Department of Chemical and Biological Engineering, Koç University, Istanbul 34450, Turkey, [3]Department of Human Molecular Genetics and Biochemistry, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel and [4]Department of Computer Engineering, Koç University, Istanbul 34450, Turkey

*To whom correspondence should be addressed.
Associate Editor: Alfonso Valencia

Received on March 14, 2022; revised on August 5, 2022; editorial decision on September 12, 2022; accepted on September 15, 2022

Abstract

Summary: HMI-PRED 2.0 is a publicly available web service for the prediction of host–microbe protein–protein interaction by interface mimicry that is intended to be used without extensive computational experience. A microbial protein structure is screened against a database covering the entire available structural space of complexes of known human proteins.

Availability and implementation: HMI-PRED 2.0 provides user-friendly graphic interfaces for predicting, visualizing and analyzing host–microbe interactions. HMI-PRED 2.0 is available at https://hmipred.org/.

Contact: nussinor@mail.nih.gov or agursoy@ku.edu.tr

1 Introduction

Microbes alter host cell signaling and modulate immune responses to maximize their survival. They induce diverse physiological conditions including immune-mediated diseases and cancers (Ruff *et al.*, 2020; Sepich-Poore *et al.*, 2021). Though the importance of microbial effects on human health is now well-accepted, in most cases the detailed mechanisms through which they induce the conditions are yet to be determined (Zhou *et al.*, 2022). Microbial species interact with their hosts by exploiting diverse strategies, including host–microbe protein–protein interactions at multiple stages of the cell's life cycle with many different, albeit partially shared interactors (Gupta *et al.*, 2012; Guven-Maiorov *et al.*, 2017; Schneider and Hoffmann, 2022; Tyl *et al.*, 2022; Walch *et al.*, 2021).

As observed broadly across the eukaryotes, for microbes, mimicry of favored interface motifs is an important and efficient strategy, where a microbial protein can interact with a host protein if the microbial protein has a surface patch similar to the binding surface of a known interactor of the host protein (Franzosa and Xia, 2011; Guven-Maiorov *et al.*, 2019; Lasso *et al.*, 2021). That is, the microbial protein postures as an intrinsic interactor of the host protein and masquerades it, by mimicking the interaction interface, without needing high sequence identity or global structural homology. Interface mimicry studies suggest pathogenic mechanisms of cancers, and neuropsychiatric symptoms associated with the recent coronavirus disease 2019 (COVID-19) pandemic (Guven-Maiorov *et al.*, 2017, 2019; Ovek *et al.*, 2022; Yapici-Eser *et al.*, 2021).

Computational methods have been developed to predict, integrate and analyze general or virus–human protein–protein interactions at a large scale, where the use of non-structural features is incentivized due to the limited structural coverage (Andrighetti *et al.*, 2020; Ding and Kihara, 2018; Dong *et al.*, 2021; Karabulut *et al.*, 2021; Mahajan and Mande, 2017; Wu *et al.*, 2020). On the other hand, it was suggested that the structural space is sufficient for protein complex modeling (Kundrotas *et al.*, 2012). Recent advances in deep learning enabled computational modeling of unsolved protein structures from primary sequences with unprecedented quality, providing ways to reveal atomic details of protein–protein interactions for the structural dark space (Baek *et al.*, 2021; Jumper *et al.*, 2021). Exploiting the protein structural information across the entire available space to accurately and efficiently model protein–protein interactions on a large scale is thus possible and vital.

To date, HMI-PRED (version 1) is the only method that screens the whole structural space to predict host–microbe interactions based on interface mimicry (Guven-Maiorov *et al.*, 2020). That version of HMI-PRED was a more conceptual study that provides the predicted host–microbe interactions with limited microbial protein coverage, search functionalities and outdated structural visualizations; thus, users need to perform filtering, analysis and visualization for their individual proteins of interest. HMI-PRED 2.0 is therefore

a completely rebuilt automated tool with the same core concept of interface mimicry. HMI-PRED 2.0 is equipped with new interface templates as well as computationally modeled microbial proteins to cover broader interactions, enabling more potential host–microbe interactions to be scanned. Visualization of templates and predicted interactions are improved for publication-ready images. More importantly, HMI-PRED 2.0 Library allows users to find host–microbe interactions of interest from the pre-computed interactions in our database with enrichment analysis and network visualization options. We also provide bulk download for predicted interactions as well as the microbial protein structures modeled using AlphaFold 2 (Jumper *et al.*, 2021). Using HMI-PRED 2.0, users can predict, search, analyze visualize and download the predicted host–microbe protein–protein interactions. HMI-PRED 2.0 is freely available at https://hmipred.org/.

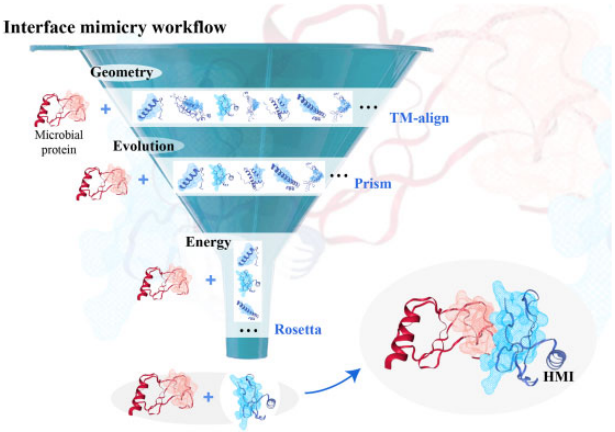

Fig. 1. Interface mimicry workflow, the core prediction mechanism of HMI-PRED 2.0. The input microbial protein (red chain) is scanned through the host protein template database (blue box) to predict host–microbial protein–protein interaction (HMI at the bottom). The prediction process is composed of three main steps. The 3D geometry of the microbial protein is compared against templates, evolutionarily conserved interaction hot spots are checked, and the protein–protein binding energy is estimated. Therefore, HMI-PRED 2.0 predicts host–microbe protein–protein interactions that are geometrically, evolutionarily and energetically favorable, which can be used for downstream analyses (A color version of this figure appears in the online version of this article)

2 Implementations

2.1 Workflow

The core concept of HMI-PRED 2.0 resembles HMI-PRED version 1 and PRISM (Fig. 1) (Baspinar *et al.*, 2014; Guven-Maiorov *et al.*, 2020). First, the extracted surfaces from the microbe protein structures are structurally compared to the interfaces in our template database using TM-Align, with an optional evolutionary hot spot filtering (Tuncbag *et al.*, 2009; Zhang and Skolnick, 2005). The filtered microbe proteins are docked to the complementary proteins (known interactors of mimicked ones) using Rosetta (Gray *et al.*, 2003). Thus, the predicted host–microbe interactions are structurally well-aligned, evolutionarily conserved and energetically favorable. More details about the prediction algorithm, including the estimated false positive rate of 18%, can be found in Guven-Maiorov *et al.* (2020).

2.2 Implementation

HMI-PRED 2.0 is built on top of the Django framework, which is suitable for database-centric web services (Django Software Foundation, 2020). First, our new template database covers 60 868 protein complexes deposited on RCSB PDB (Burley *et al.*, 2020) by January 2021. Users can submit mmCIF-formatted files as input for large structures. Our web service accumulated more than 1.6 million predicted host–microbe interactions for over 20 000 microbial proteins, among which 8520 are modeled using AlphaFold 2 (Jumper *et al.*, 2021). The modeled protein structures as well as predicted interactions can be searched, viewed, analyzed and downloaded via our web service. Proteins are visualized using NGL viewer, which provides publication-quality structure views and interactive options (Rose *et al.*, 2018). We also provide options to visualize the interactions as a network using Cytoscape (Franz *et al.*, 2016) and perform enrichment analysis via String (Szklarczyk *et al.*, 2019) (Fig. 2).

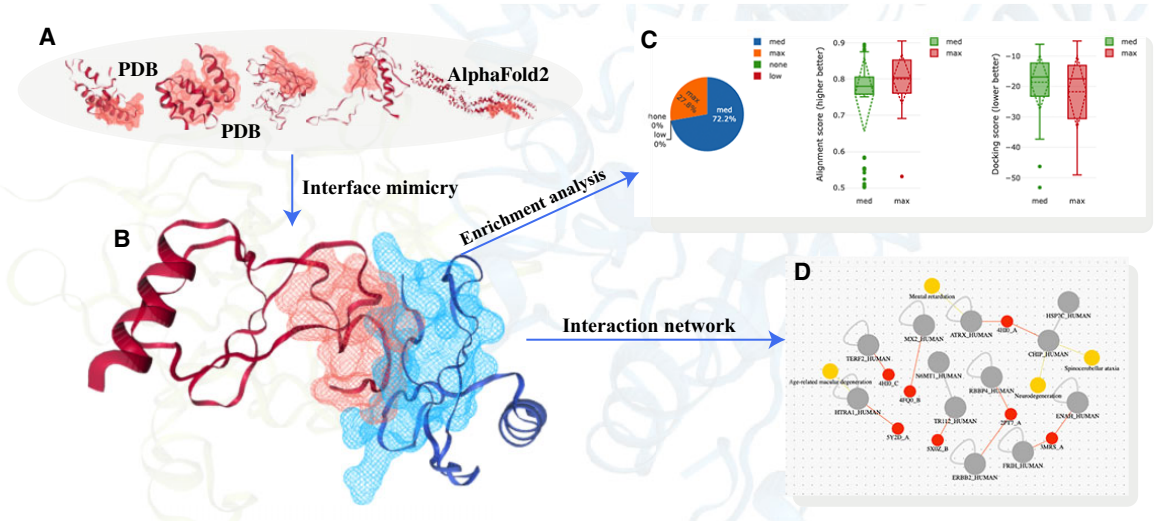

Fig. 2. Illustration of HMI-PRED 2.0 workflow. (**A**) Users can submit input protein structures or search for pre-computed interactions, which include experimentally solved or AlphaFold 2-modeled microbial proteins. (**B**) HMI-PRED 2.0 predicts potential host–microbe protein–protein interactions based on interface mimicry concept. The interface mimicry-based prediction process includes three main stages as shown in Figure 1. At the end of a prediction task, there may be multiple predicted interactions for a microbial input protein. The false positive rate is approximately 18% based on our previous study. (**C** and **D**) HMI-PRED 2.0 web server provides two analysis/visualization functions: human pathway enrichment analysis when multiple human proteins are targeted (C), and interaction network visualization showing diseases associated with targeted human proteins (D). The charts in (C) are conceptual; the actual analysis output is a table with texts

In the section below, we demonstrate the basic usage of our web service. For more detailed usage, users are referred to the tutorials page on our web service.

2.3 Basic usage

Users can either search for pre-computed host–microbe interactions in our database or submit a new prediction task. HMI-PRED 2.0 Library allows users to find host–microbe interactions by several different filter options. One example is shown in the next paragraph.

Users can search predicted host–microbe interactions for all organisms whose name contains 'Epstein–Barr' (case insensitive) and moderate-to-high level of hot spot conservation. Under 'Library—ALL HMI' page, write 'Epstein–Barr' in the 'Microbe organism' box on the filter menu column. Also choose 'med' and 'max' in the 'Hotspot conservation' on the filter menu, and hit 'Apply' (leave the default Max. Rosetta score of −5.0 as is for this example). It returns 90 interactions on the Host–Microbe Interactions table. Click 'Enrichment Analysis' button once the table is shown with results, which will redirect to String with 20 nodes (human proteins predicted to interact with Epstein–Barr viral proteins) and 12 edges. Under 'Analysis' tab, a few important pathways and diseases appear. For example, the associations between the virus and 'MicroRNAs in cancer', 'Ras signaling pathway', 'Kaposi sarcoma-associated herpesvirus infection' and 'non-Hodgkin lymphoma' are supported by literature evidence (Caetano *et al.*, 2021; Fukuda and Longnecker, 2007; Pinzone *et al.*, 2015). Please note that the thorough validation or individual investigation of these associations is out of the scope of this manuscript. Users can apply filters according to their scientific questions and perform analysis accordingly.

In case the microbial protein of interest is not found, users can submit a new prediction task under 'RUN HMI-PRED' page by providing either a valid PDB ID with chain ID or a protein structure file in mmCIF or PDB format. More details can be found on the tutorial pages of our web service.

3 Conclusions

We developed HMI-PRED 2.0, a web service for structure-based host–microbe protein–protein interaction prediction. HMI-PRED 2.0 provides user-friendly interfaces with analysis tools for users without programming experiences. HMI-PRED 2.0 is freely available at https://hmipred.org/.

Acknowledgements

We sincerely thank Mrs Hwayoung Ji for visual design of the web application and manuscript figures. We thank all COSBI group members and especially Fatma Cankara for their help in testing the server. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. This work utilized the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov).

Funding

This work was supported in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health [HHSN261201500003I]. This Research was supported [in part] by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. This research has been partially funded with [Turkish Health Institutions (TUSEB)-4081] and [Scientific and Technological Research Council (TUBITAK)-120C120] projects.

Conflict of Interest: none declared.

References

Andrighetti,T. *et al.* (2020) MicrobioLink: an integrated computational pipeline to infer functional effects of microbiome–host interactions. *Cells*, **9**, 1278.

Baek,M. *et al.* (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, **373**, 871–876.

Baspinar,A. *et al.* (2014) PRISM: a web server and repository for prediction of protein-protein interactions and modeling their 3D complexes. *Nucleic Acids Res*, **42**, W285–9.

Burley,S.K. *et al.* (2020) RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res.*, **49**, D437-D451.

Caetano,B.F.R. *et al.* (2021) Epstein-Barr virus microRNAs in the pathogenesis of human cancers. *Cancer Lett.*, **499**, 14–23.

Ding,Z. and Kihara,D. (2018) Computational methods for predicting protein-protein interactions using various protein features. *Curr. Protoc. Protein Sci.*, **93**, e62.

Dong,T.N. *et al.* (2021) A multitask transfer learning framework for the prediction of virus-human protein–protein interactions. *BMC Bioinformatics*, **22**, 572.

Django Software Foundation. (2020) *Django*. Lawrence, KS.

Franz,M. *et al.* (2016) Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics*, **32**, 309–311.

Franzosa,E.A. and Xia,Y. (2011) Structural principles within the human-virus protein-protein interaction network. *Proc. Natl. Acad. Sci. USA*, **108**, 10538–10543.

Fukuda,M. and Longnecker,R. (2007) Epstein-Barr virus latent membrane protein 2A mediates transformation through constitutive activation of the RAS/PI3-K/AKT pathway. *J. Virol.*, **81**, 9299–9306.

Gray,J.J. *et al.* (2003) Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.*, **331**, 281–299.

Gupta,V. *et al.* (2012) The seroprevalence of Helicobacter pylori and its relationship to malaria in ugandan children. *Trans. R. Soc. Trop. Med. Hyg.*, **106**, 35–42.

Guven-Maiorov,E. *et al.* (2020) HMI-PRED: a web server for structural prediction of host-microbe interactions based on interface mimicry. *J. Mol. Biol.*, **432**, 3395–3403.

Guven-Maiorov,E. *et al.* (2017) Prediction of host-pathogen interactions for *Helicobacter pylori* by interface mimicry and implications to gastric cancer. *J. Mol. Biol.*, **429**, 3925–3941.

Guven-Maiorov,E. *et al.* (2017) Structural host-microbiota interaction networks. *PLoS Comput. Biol.*, **13**, e1005579.

Guven-Maiorov,E. *et al.* (2019) Oncoviruses can drive cancer by rewiring signaling pathways through interface mimicry. *Front. Oncol.*, **9**.

Jumper,J. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.

Karabulut,O.C. *et al.* (2021) ML-AdVInfect: a machine-learning based adenoviral infection predictor. *Front. Mol. Biosci.*, **8**, 647424.

Kundrotas,P.J. *et al.* (2012) Templates are available to model nearly all complexes of structurally characterized proteins. *Proc. Natl. Acad. Sci. USA*, **109**, 9438–9441.

Lasso,G. *et al.* (2021) A sweep of earth's virome reveals host-guided viral protein structural mimicry and points to determinants of human disease. *Cell Syst.*, **12**, 82–91.e3.

Mahajan,G. and Mande,S.C. (2017) Using structural knowledge in the protein data bank to inform the search for potential host-microbe protein interactions in sequence space: application to *Mycobacterium tuberculosis*. *BMC Bioinformatics*, **18**, 201.

Ovek,D. *et al.* (2022) SARS-CoV-2 interactome 3D: a web interface for 3D visualization and analysis of SARS-CoV-2–human mimicry and interactions. *Bioinformatics*, **38**, 1455–1457.

Pinzone,M.R. *et al.* (2015) Epstein-Barr virus- and Kaposi sarcoma-associated herpesvirus-related malignancies in the setting of human immunodeficiency virus infection. *Semin. Oncol.*, **42**, 258–271.

Rose,A.S. *et al.* (2018) NGL viewer: web-based molecular graphics for large complexes. *Bioinformatics*, **34**, 3755–3758.

Ruff,W.E. *et al.* (2020) Host-microbiota interactions in immune-mediated diseases. *Nat. Rev. Microbiol.*, **18**, 521–538.

Schneider,W.M. and Hoffmann,H.-H. (2022) Flavivirus–host interactions: an expanding network of proviral and antiviral factors. *Curr. Opin. Virol.*, **52**, 71–77.

Sepich-Poore,G.D. *et al.* (2021) The microbiome and human cancer. *Science*, **371**, eabc4552.

Szklarczyk,D. *et al.* (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.*, **47**, D607–D613.

Tuncbag,N. *et al.* (2009) Identification of computational hot spots in protein interfaces: combining solvent accessibility and inter-residue potentials improves the accuracy. *Bioinformatics*, **25**, 1513–1520.

Tyl,M.D. *et al.* (2022) Virus–host protein interactions as footprints of human cytomegalovirus replication. *Curr. Opin. Virol.*, **52**, 135–147.

Walch,P. *et al.* (2021) Global mapping of *Salmonella enterica*-host protein-protein interactions during infection. *Cell Host Microbe*, **29**, 1316–1332.e12.

Wu,Z. *et al.* (2020) A comprehensive review and evaluation of computational methods for identifying protein complexes from protein–protein interaction networks. *Brief. Bioinformatics*, **21**, 1531–1548.

Yapici-Eser,H. *et al.* (2021) Neuropsychiatric symptoms of COVID-19 explained by SARS-CoV-2 proteins' mimicry of human protein interactions. *Front. Hum. Neurosci.*, **15**, 656313.

Zhang,Y. and Skolnick,J. (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.*, **33**, 2302–2309.

Zhou,H. *et al.* (2022) Host-microbiome protein-protein interactions capture disease-relevant pathways. *Genome Biol.*, **23**, 72.