
Subject Section

Web interface for 3D visualization and analysis of SARS-CoV-2 – human mimicry and interactions

Damla Ovek^{1,2}, Ameer Taweel², Zeynep Abali¹, Ece Tezsezen¹, Yunus Emre Koroglu², Chung-Jung Tsai³, Ruth Nussinov^{3,4}, Ozlem Keskin⁵ and Attila Gursoy^{2,*}

¹Graduate School of Science and Engineering, Koc University, Istanbul, 24450, Turkey,

²Department of Computer Engineering, Koc University, Istanbul, 34450, Turkey,

³Computational Structural Biology Section, Frederick National Laboratory for Cancer Research, National Cancer Institute at Frederick, Frederick, MD 21702, U.S.A.,

⁴Department of Human Molecular Genetics and Biochemistry, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel, and

⁵Department of Chemical and Biological Engineering, Koc University, Istanbul, 34450, Turkey.

*To whom correspondence should be addressed.

Associate Editor: XXXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Summary: We present a web-based server for navigating and visualizing possible interactions between SARS-CoV-2 and human host proteins. The interactions are obtained from HML_Pred which relies on the rationale that virus proteins *mimic* host proteins. The structural alignment of the viral protein with one side of the human protein-protein interface determines the mimicry. The mimicked human proteins and predicted interactions, and the binding sites are presented. The user can choose one of the 18 SARS-CoV-2 protein structures and visualize the potential 3D complexes it forms with human proteins. The mimicked interface is also provided. The user can superimpose two interacting human proteins in order to see whether they bind to the same site or different sites on the viral protein. The server also tabulates all available mimicked interactions together with their match scores and number of aligned residues. This is the first server listing and cataloging all interactions between SARS-CoV-2 and human protein structures, enabled by our innovative interface mimicry strategy.

Availability: The server is available at <https://interactome.ku.edu.tr/sars/>.

Contact: hmi-support-group@ku.edu.tr

1 Introduction

Severe acute respiratory syndrome (SARS)-like coronavirus (SARS-CoV-2) is rapidly spreading in humans as COVID19, a respiratory illness with pneumonia and flu-like symptoms (Lai, et al., 2020). The SARS-CoV-2 possesses a total of 29 proteins which consist of 16 nonstructural (Nsp), 4

structural proteins (Sp), and 9 open reading frames (Orf) (Gordon, et al., 2020).

Viruses may hijack host proteins to evade host defense (Dai, et al., 2020; Rawlinson, et al., 2018). Interface based hijacking is a strategy where a pathogen protein competes with host proteins to bind a target protein via interface mimicry. By outcompeting the host counterparts, viral proteins can rewire cellular function, overcoming host defense. Identification of host-pathogen interactions and their structural mechanisms of

action can help design therapeutic strategies. There are web servers that integrate the SARS-CoV-2 genome, protein data, and structures such as Coronavirus3d (<https://coronavirus3d.org>) (Perfetto, et al., 2020; Sedova, et al., 2020) albeit that do not provide a structural protein-protein interaction (PPI) data.

We developed a web-based server that allows users to navigate and visualize the SARS-CoV-2 proteins, the human proteins they mimic, and possible virus protein-human protein interactions due to mimicry. The mimicked proteins and predicted interactions are obtained from HMI_Pred (Guyen-Maiorov, et al., 2020). HMI_Pred uses known protein-protein binding sites (interfaces) from human proteins. The structural alignment of the microbial/viral protein with one side of the human interface determines the mimicry. The interaction prediction relies on the concept that if a viral protein mimics one side of the interface, it can interact with the complementary side of the interface. The concept is similar to PRISM (Baspinar, et al., 2014). HMI_PRED uses TM-align (Zhang and Skolnick, 2005) for structural comparison, and Rosetta (local refinement, Rosetta 2018.09.60072) (Wang, et al., 2007) to predict energetically favorable virus-human complexes. Previous applications showed that HMI_PRED can generate promising host-microbe interaction models which may explain the molecular mechanisms of how microbes hijack host immunity, modify cytoskeletal dynamics, and change cell cycle regulation (Guyen-Maiorov, et al., 2017; Guven-Maiorov, et al., 2019). Conformational changes for a protein are considered by using all possible alternative conformations of that protein extracted from PDB.

The server presents the mimicry and predicted interactions of 18 SARS-CoV-2 proteins whose structures are available in PDB as of Nov 23, 2020 (536 PDB entries including 595 SARS-CoV-2 protein structures). The mimicry is obtained by scanning 17K known non-redundant human interfaces extracted from 53K redundant human interfaces (HMIPRED template set). Different conformations of the 18 proteins are found to mimic 7375 human interactions from 1653 human proteins. One human protein can be mimicked several times.

2 Web Server Usage

The main page displays the structure and mimicry results (Fig. 1). We explain the usage with an example showing the interactors of NSP3. The user can select the individual virus protein from the left top pulldown menu (i.e. NSP3 in **B1** in the figure). The UniProt (i.e. P0DTD1) and PDB (6WX4D) links of the viral protein are provided (**B2**). The PDB ID and the sequence coverage of all available PDB structures of the selected viral protein are listed in the Protege-frame section (the first line in the list corresponds to 6WX4D in **B4**). One can select another PDB structure of the viral protein (green bar) to fetch its structure, mimicked human proteins and interactions. In this example, we continue with the default PDB structure. On the left window, the structure of the selected viral protein (green) and its binding site (light green) with the predicted human protein (**B2**) are visualized in cartoon representation. The viral binding site is the part that is resembling the mimicked human protein's binding site. If the mimicked human interface button is selected (5W8T, Chain A in **B3**), the mimicked human protein (in yellow) is superimposed on the corresponding virus binding site (**B7**). If the viral protein structure comes from a complex, then the "Show known interface residues" button displays the experimental binding site of the viral protein if it is already in a complex in PDB at the same ribbon diagram (red) (**B6**).

The binding partner (dark blue) of the mimicked human protein interacting with the viral protein is also shown with its binding site (light blue) in **B2**. If there are other mimicked human interfaces, they are listed in **B5**. Two human interactors can be displayed simultaneously for a SARS-CoV-

2 protein structure in a superimposed state. Superimposition enables users to compare two human partners at the same time and analyze the virus hijacking of host proteins (**B8**). One needs to click on a second interactor listed in **B5** which will be shown in purple in the 3D visualization window. The second human protein comes from its own mimicked partner. At most two partner structures can be visualized. In order to switch to another mimicked interface, one of the chosen interactors should be unchosen. On the right side, the details of the human interactor (the one that is already ticked in **B5**) are provided if the mouse is over the PDB ID. As one moves the cursor over the possible interactors, information about the chosen human protein can be accessed via the Related links such as UniProt and PDB page links. The prediction scores, interface data and Rosetta scores are also provided. Furthermore, users can go to the gene ontology results via the QuickGo link.

View all data page: On the main page, users can follow the "View All Data" link to download all interactome data as a CSV file or search for a SARS-CoV-2 protein and access its interactions. Viral protein selection can be done via the "Search" section or via the filtering option in the "Microbial Protein" column. After the selection, the columns display human and microbial (viral) partners of the interaction, their gene and protein name, mimicked interface, prediction score (Rosetta score), number of the aligned residues in the interface, interface length, matched ratio (aligned residues/ interface length) and TM score. The user can filter each column according to the choice. Users can download the interface residue information of a prediction and access the 3D visualization of its structure via the "eye" icon.

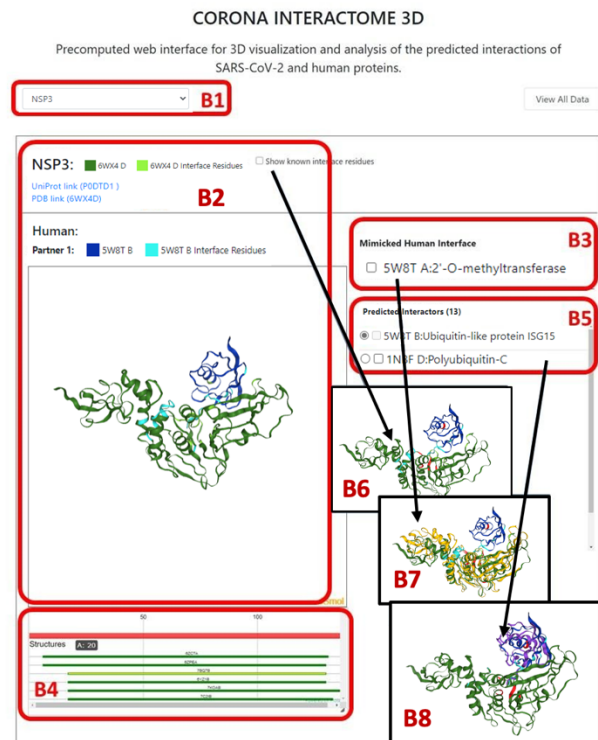


Fig. 1. Screenshot of the CORONA INTERACTOME 3D. The user can select one of the SARS-CoV-2 proteins (B1), NSP3 its structures, mimicked interfaces, and its human partners (B2-B8). SARS-CoV-2 NSP3 preferentially cleaves ISG15 (Shin, et al., 2020). To hijack the host defense, viruses might induce ISG15 deconjugating proteins or prevents the generation of ISGylated proteins and abrogate the antiviral response in the host (Yuan and Krug, 2001).

Web interface for 3D visualization and analysis of SARS-CoV-2 – human mimicry and interactions**Acknowledgements**

All the calculations are performed using the Koc University Advanced Computing Center (KUACC) Facilities. AG and OK are members of Science Academy, Turkey.

Funding

This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract HHSN26120080001E. 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 Research was supported [in part] by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

Conflict of Interest: none declared.

References

- Lai, C.C., et al. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *Int J Antimicrob Agents* 2020;55(3):105924.
- Gordon, D.E., et al. A SARS-CoV-2-Human Protein-Protein Interaction Map Reveals Drug Targets and Potential Drug-Repurposing. *bioRxiv* 2020.
- Rawlinson, S.M., et al. Viral regulation of host cell biology by hijacking of the nucleolar DNA-damage response. *Nat Commun* 2018;9(1):3057.
- Dai, X., et al. Orchestrated efforts on host network hijacking: Processes governing virus replication. *Virulence* 2020;11(1):183-198.
- Perfetto, L., et al. The IMEx Coronavirus interactome: an evolving map of Coronaviridae-Host molecular interactions. *bioRxiv* 2020.
- Sedova, M., et al. Coronavirus3D: 3D structural visualization of COVID-19 genomic divergence. *Bioinformatics* 2020;36(15):4360-4362.
- Guyen-Maiorov, E., et al. HMI-PRED: A Web Server for Structural Prediction of Host-Microbe Interactions Based on Interface Mimicry. *J Mol Biol* 2020;432(11):3395-3403.
- Baspinar, A., et al. PRISM: a web server and repository for prediction of protein-protein interactions and modeling their 3D complexes. *Nucleic Acids Res* 2014;42(Web Server issue):W285-289.
- Zhang, Y. and Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 2005;33(7):2302-2309.
- Wang, C., Bradley, P. and Baker, D. Protein-protein docking with backbone flexibility. *J Mol Biol* 2007;373(2):503-519.
- Guyen-Maiorov, E., et al. Prediction of Host-Pathogen Interactions for *Helicobacter pylori* by Interface Mimicry and Implications to Gastric Cancer. *J Mol Biol* 2017;429(24):3925-3941.
- Guyen-Maiorov, E., et al. Interface-Based Structural Prediction of Novel Host-Pathogen Interactions. *Methods Mol Biol* 2019;1851:317-335.
- Shin, D., et al. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* 2020;587(7835):657-662.
- Yuan, W. and Krug, R.M. Influenza B virus NS1 protein inhibits conjugation of the interferon (IFN)-induced ubiquitin-like ISG15 protein. *EMBO J* 2001;20(3):362-371.