



## OPEN ACCESS

EDITED BY  
Jessica Andreani,  
UMR9198 Institut de Biologie Intégrative  
de la Cellule (I2BC), France

REVIEWED BY  
Juan Fernandez-Recio,  
Institute of Vine and Wine Sciences (CSIC),  
Spain  
Arianna Fornili,  
Queen Mary University of London,  
United Kingdom  
Justas Dapkūnas,  
Vilnius University, Lithuania

\*CORRESPONDENCE  
Ezgi Karaca,  
✉ ezgi.karaca@ibg.edu.tr

SPECIALTY SECTION  
This article was submitted to Biological  
Modeling and Simulation,  
a section of the journal  
Frontiers in Molecular Biosciences

RECEIVED 07 October 2022  
ACCEPTED 31 January 2023  
PUBLISHED 01 March 2023

CITATION  
Koşaca M, Yılmazbilek İ and Karaca E  
(2023), PROT-ON: A structure-based  
detection of designer PROTein  
interface MutatiONs.  
*Front. Mol. Biosci.* 10:1063971.  
doi: 10.3389/fmolb.2023.1063971

COPYRIGHT  
© 2023 Koşaca, Yılmazbilek and Karaca.  
This is an open-access article distributed  
under the terms of the [Creative Commons  
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,  
distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# PROT-ON: A structure-based detection of designer PROTein interface MutatiONs

Mehdi Koşaca<sup>1,2</sup>, İrem Yılmazbilek<sup>1,3</sup> and Ezgi Karaca<sup>1,2\*</sup>

<sup>1</sup>Izmir Biomedicine and Genome Center, Dokuz Eylul Health Campus, Izmir, Türkiye, <sup>2</sup>Izmir International Biomedicine and Genome Institute, Dokuz Eylul University, Izmir, Türkiye, <sup>3</sup>Middle East Technical University, Ankara, Türkiye

The mutation-induced changes across protein-protein interfaces have often been observed to lead to severe diseases. Therefore, several computational tools have been developed to predict the impact of such mutations. Among these tools, FoldX and EvoEF1 stand out as fast and accurate alternatives. Expanding on the capabilities of these tools, we have developed the PROT-ON (PROTein-protein interface mutatiONs) framework, which aims at delivering the most critical protein interface mutations that can be used to design new protein binders. To realize this aim, PROT-ON takes the 3D coordinates of a protein dimer as an input. Then, it probes all possible interface mutations on the selected protein partner with EvoEF1 or FoldX. The calculated mutational energy landscape is statistically analyzed to find the most enriching and depleting mutations. Afterward, these extreme mutations are filtered out according to stability and optionally according to evolutionary criteria. The final remaining mutation list is presented to the user as the designer mutation set. Together with this set, PROT-ON provides several residue- and energy-based plots, portraying the synthetic energy landscape of the probed mutations. The stand-alone version of PROT-ON is deposited at <https://github.com/CSB-KaracaLab/prot-on>. The users can also use PROT-ON through our user-friendly web service <http://proton.tools.ibg.edu.tr:8001/> (runs with EvoEF1 only). Considering its speed and the range of analysis provided, we believe that PROT-ON presents a promising means to estimate designer mutations.

## KEYWORDS

interface mutation, protein-protein interaction, interface design, EvoEF1, FoldX

## 1 Introduction

Protein-protein interactions (PPIs) are at the heart of any biological process, such as signaling, molecular transporting, and metabolic activities. Therefore, the disease-causing mutations are often found across protein-protein interfaces (Nishi et al., 2013). Mutation-induced changes across PPIs significantly alter the interface biophysics, thus the binding affinity of the interaction (David et al., 2012; David and Sternberg, 2015; Jemimah and Gromiha, 2020). As an example, Arg191Cys mutation in beta electron transfer protein impairs the protein's interaction with acyl-CoA dehydrogenases (ACDHs) due to a decrease in the enzymatic activity, induced by the Arg191Cys mutation. This decrease leads to the glutaric aciduria IIB metabolic disease. (Yates and Sternberg, 2013; Henriques et al., 2021).

To gauge the binding affinity changes induced by interfacial mutations, several computational tools have been developed (Navío et al., 2019). A recent benchmarking effort has shown that among such tools FoldX (Schymkowitz et al., 2005) and EvoEF1 (Pearce et al., 2019) can predict the structural impact of a mutation with top ranking accuracies in the

fastest manner (Amengual-Rigo et al., 2021). For this, FoldX and EvoEF1 structurally model the mutation under the effect of a force field, followed by the scoring of the new interface with a linear empirical energy formula (composed of van der Waals, electrostatic, hydrogen bonding, and desolvation energies, as well as entropic penalties). Expanding on the capabilities of FoldX and EvoEF1, we developed PROT-ON (PROTein-protein interface mutatiONs), which aims to deliver the most critical PPI mutations, i.e., the designer mutations, that can be used to propose new binders. For this, PROT-ON probes all possible interface mutations with either FoldX or EvoEF1 on a selected protein partner. It then statistically analyzes the scanned energy landscape and filters out the most significant affinity altering mutations according to stability and optionally evolutionary criteria. Together with this list, PROT-ON outputs several residue- and energy-based plots, portraying the synthetic energy landscape of the probed mutations. The stand-alone version of PROT-ON is deposited at <https://github.com/CSB-KaracaLab/prot-on>. Our tool can also be accessed via our user-friendly web interface <http://proton.tools.ibg.edu.tr:8001/> where only EvoEF1-based calculations are performed. The web-based calculations take less than 10 min.

In this manuscript, we firstly explain the architecture of PROT-ON. We then demonstrate its use on MDM2-p53, ACE2-RBD, and MCL1-NOXA protein-protein complexes. With these proof-of-concept PROT-ON cases, we aim to guide our users in finding the most critical PPI positions in a speedy manner, which can translate into possible designer mutations.

## 2 Materials and methods

### 2.1 The PROT-ON pipeline

In the PROT-ON pipeline, the interfacial amino acids of one protein partner are identified by using the classical Euclidian distance formula (Eq. 1). By using this distance calculation, we define the interchain residues that contain at least one atom within a certain cut-off as interfacial amino acids. By default, the distance cut-off is set to 5 Å. Then, by using the relevant FoldX and EvoEF1 functions, the interfacial amino acids of the selected partner are mutated into 19 other amino acid possibilities, followed by calculating the energy score of each mutation. So, if there are 15 interface amino acids defined on one monomer, PROT-ON will calculate  $15 \times 19$  energy scores. Then, all these predicted energy scores will be analyzed with non-parametric box-and-whisker statistics. In box-and-whisker statistics, 50% of the data around a median is defined as an interquartile range (IQR) and represented as a box. Whiskers are extended from the opposite sides of the box by multiple IQRs to cover the data spread. Any data falling outside this whisker expansion is classified as an *outlier* (Supplementary Figure S1). The negative *outliers* are defined as mutations that significantly improve the binding (enhancing mutations,  $\Delta\Delta G \ll 0$ ), while the positive *outliers* are the ones that significantly worsen the binding (depleting mutations,  $\Delta\Delta G \gg 0$ ). We return these mutations as the initial designer mutation set. By default, the IQR threshold is set to 1.5.

$$|AB| = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2}$$

Equations 2, 3. The Euclidian distance formula used to calculate the interatomic distances between two protein partners, i.e., A and B.

The initial designer mutation list is then filtered according to the stability score of each mutation. Here, if a mutation has a FoldX or EvoEF1 stability score  $< 0$ , then the mutation is kept. The stable mutations can optionally be filtered out further according to a user supplied PSSM (position specific scoring matrix) file. PSSM measures the probability of observing a particular mutation at a given position. PROT-ON keeps a depleting mutation, if its PSSM score is  $\leq 0$ . Complementarily, it keeps an enriching mutation prediction if its PSSM score is  $> 0$ . We use these cut-offs as we observed that the most enriching and depleting mutations span such PSSM values. For more, please see Supplementary Figure S2. The mutation set remaining after the consecutive filtering steps are returned to the user as the final designer mutation set.

### 2.2 FoldX and EvoEF1 functions used in PROT-ON

To calculate the impact of PPI mutations, we use FoldX 4.0 and EvoEF1 January-2021 versions. We stick to these versions, as they are the ones that were benchmarked on the SKEMPI 2.0 data set (Jankauskaitė et al., 2019; Amengual-Rigo et al., 2021). For both tools, the input complex is supplied in the PDB format, where the first step is an energy minimization aiming at removing the interatomic clashes (commands: **RepairStructure** for EvoEF1; **RepairPDB** for FoldX). The interfacial mutations are then imposed on the energy-minimized structure on one partner (commands: **BuildMutant** for EvoEF1; **BuildModel** for FoldX). Finally, binding free energy and stability scores are calculated with the empirical energy formulas of each algorithm (commands: **ComputeBinding** and **ComputeStability** for EvoEF1; **AnalyseComplex**, **Stability** for FoldX).

FoldX estimates the free energy of binding ( $\Delta G_{\text{FoldX}}$ ) by using van der Waals forces ( $\Delta G_{\text{vdw}}$ ), electrostatic interactions ( $\Delta G_{\text{el}}$ ), hydrogen bonding ( $\Delta G_{\text{hbond}}$ ,  $\Delta G_{\text{wb}}$ ), and solvation energies of polar ( $\Delta G_{\text{solvP}}$ ), and hydrophobic groups ( $\Delta G_{\text{solvH}}$ ). In addition to these, it also considers the entropy by statistically analyzing the phi-psi distribution for each amino acid ( $\Delta S_{\text{mc}}$ ,  $\Delta S_{\text{sc}}$ ). The empirical formula of FoldX is given in Eq. 2 (Schymkowitz et al., 2005).

$$\Delta G_{\text{FoldX}} = a \cdot \Delta G_{\text{vdw}} + b \cdot \Delta G_{\text{solvH}} + c \cdot \Delta G_{\text{solvP}} + d \cdot \Delta G_{\text{wb}} + e \cdot \Delta G_{\text{hbond}} + f \cdot \Delta G_{\text{el}} + g \cdot \Delta G_{\text{kon}} + h \cdot T\Delta S_{\text{mc}} + k \cdot T\Delta S_{\text{sc}} + l \cdot \Delta G_{\text{clash}}$$

Equation 2. The empirical energy formula of FoldX to calculate the binding free energy change.

EvoEF1 uses similar terms, i.e., van der Waals forces ( $E_{\text{vdw}}$ ), electrostatic interactions ( $E_{\text{elec}}$ ), hydrogen bonding contribution ( $E_{\text{HB}}$ ), desolvation energies of polar and hydrophobic groups ( $E_{\text{solv}}$ ), and reference energies of 20 amino acids to describe atomic interactions ( $E_{\text{ref}}$ ) (Eq. 3) (Pearce et al., 2019).

$$E_{\text{EvoEF1}} = \sum [E_{\text{vdw}}(i, j) + E_{\text{elec}}(i, j) + E_{\text{HB}}(i, j) + E_{\text{solv}}(i, j)] - E_{\text{ref}}$$

Equation 3. The empirical energy formula of EvoEF1 to calculate the binding affinity of a PPI.

For both tools, the impact of the mutation is calculated by subtracting the wild type energy value from the mutant one.

## 2.3 The packages used in the stand-alone and the web service versions of PROT-ON

The standalone PROT-ON uses NumPy (<https://numpy.org/>) (Harris et al., 2020) to create lists and to perform scientific calculations. Pandas (<https://pandas.pydata.org/>) (McKinney, 2011) is used to read data frames and generate output files. Plotly (<https://plotly.com/>) and Kaleido (<https://pypi.org/project/kaleido/>) are used to create heatmap and box-and-whisker plots, as well as to generate static images of creating plots. The latest version of PROT-ON can be cloned from <https://github.com/CSB-KaracaLab/prot-on> for Linux and MacOS systems.

The back end of the PROT-ON web server is developed by Flask (<https://flask.palletsprojects.com/en/2.1.x/>) (Grinberg, 2018) which is a Python microframework. On the other hand, HTML, CSS, JavaScript languages, and Bootstrap (<https://getbootstrap.com/>) web application development toolkit are used to develop the front end of the web server (navigator menu, grid system, table, form, etc.). Fundamental input types like text, slider, email, etc. are written in HTML. CSS, JavaScript, and Bootstrap are used to design input types and create a grid system (container, row, columns). Celery Python package (<https://docs.celeryq.dev/>), RabbitMQ message broker (<https://www.rabbitmq.com/>), and SQLAlchemy toolkit (<https://www.sqlalchemy.org/>) (Brown and Wilson, 2012) are used to process submitting jobs in the background. Celery is a distributed task queue that allows the creation of different tasks running in the background asynchronously. Celery is used to prevent data accumulation in the workstation by running the scheduled task and to process submitted jobs by a worker task. Celery also communicates with a message broker, RabbitMQ (to receive and send job messages and a database) and a database, SQLAlchemy (to store results). The web server is deployed by Nginx (<https://www.nginx.com/>) (Reese, 2008), Unicorn (<https://unicorn.org/>), and the supervisor process control system (<http://supervisord.org/>). Nginx is open-source web-service software that serves as a reverse proxy, and load balancing for the HTTP web server. Unicorn is a WSGI HTTP server written with Python. Flask communicates with Nginx by converting HTTP requests to Python. Flask requires a Supervisor control process system to start automatically if anything goes wrong. Demonstration of processing runs in the background and deployment of the web server is given in [Supplementary Figure S3](#).

PROT-ON service is accessible over <http://proton.tools.ibg.edu.tr:8001>. PROT-ON is hosted on an HP Z2 Tower G4 workstation [64-bit 3.40 GHz Intel (R) Xeon (R) E-2124 G CPU with four cores and 32 GB RAM] connected to <https://ibg.edu.tr>. The system uses Ubuntu OS 22.04 LTS version.

## 3 Results

PROT-ON aims to analyze the mutational landscape of a protein complex to provide with potential (binding enriching or depleting) designer mutations. For this, PROT-ON takes a protein dimer in the format of a PDB file as an input, together with the monomer (chain) of interest. After this, all interfacial amino acids of the selected partner are mutated into other 19 amino acid alternatives. This procedure is

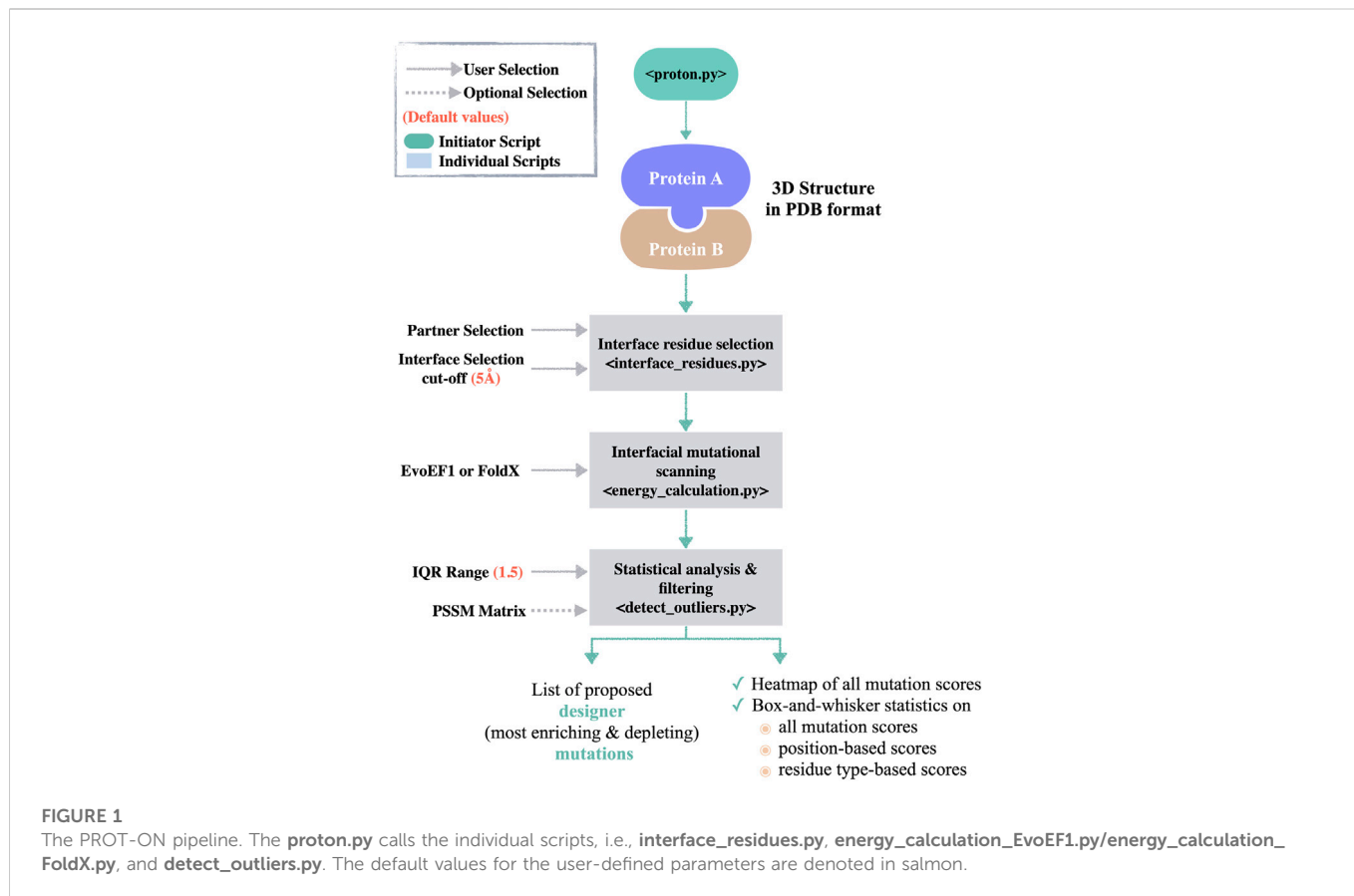
followed upon collecting the energy scores of the scanned mutational landscape. The mutation energy scores are filtered according to statistics-, stability- and optionally evolutionary-based criteria. These filtering steps are designed to ensure proposing biologically feasible mutations, so that they can be utilized in constructing new protein binders. Below, we explain the framework of PROT-ON, as well as our server architecture. We also demonstrate the use of PROT-ON on three different protein-protein complexes.

## 3.1 The PROT-ON framework

The functions of PROT-ON are called by the initiator **proton.py** script ([Figure 1](#)). When the user supplies the protein dimer structure in the PDB format, together with the chain of interest, **interface\_residues.py** script is called. The **interface\_residues.py** calculates the interfacial amino acids within a user provided cut-off (default is 5 Å, see [Methods 2.1](#)). After the selection of mutation calculation engine (either FoldX or EvoEF1), **energy\_calculation\_[selected\_method].py** is called to scan all interfacial mutations on the selected partner. The same script also calculates the estimated binding energy changes between the mutant and wild type states (for the functions used here, please see [Methods 2.2](#)). This step is followed by running the **detect\_outliers.py** script. The **detect\_outliers.py** applies a box-and-whisker statistics on the scanned energy landscape. In box-and-whisker statistics, 50% of the data around a median is defined as an interquartile range (IQR) and represented as a box. Whiskers are extended from the opposite sides of the box by multiple IQRs to cover the data spread. Any data falling outside this whisker expansion are classified as negative and positive *outliers* ([Supplementary Figure S1](#)). Here, our assumption is that the negative *outlier mutations* will significantly improve the binding (enhancing mutations,  $\Delta\Delta G \ll 0$ ), while the positive *outlier mutations* will significantly worsen the binding (depleting mutations,  $\Delta\Delta G \gg 0$ ). Expanding on this assumption, **detect\_outliers.py** script isolates the *outlier* mutations. If needed, the user can play with the IQR range used to define the extent of whisker, thus the extent of *outliers* (1.5 by default). To ensure the selection of biophysically sound mutations, the same script filters out the isolated *outliers* according to their stability scores. Here, we provide the option to supply an external PSSM score file in case the user wants to incorporate evolutionary information into the process (see [Methods 2.1](#)). The **detect\_outliers.py** script further generates a heatmap, as well as residue-based energy analysis to portray the energy landscape scanned by all mutations. The PROT-ON framework is presented in [Figure 1](#). The input/output files generated by the tool are described in <https://github.com/CSB-KaracaLab/prot-on>.

## 3.2 The PROT-ON web service

We present an easy access point to our PROT-ON framework *via* <http://proton.tools.ibg.edu.tr:8001>. Our web service runs only with EvoEF1, as FoldX requires a personal academic license. To submit a run, four mandatory and four optional steps should be followed ([Figure 2](#)). In the 1st step, the run name is specified. In the 2nd step, either the PDB ID of the case is inputted or the PDB coordinates of the complex is uploaded. In the 3rd step, the chain IDs of the protein dimer of interest are selected (from a drop-down menu). In the 4th step, the partner to be analyzed is indicated by switching the relevant toggle. In the



**FIGURE 2**  
 The stepwise annotation of PROT-ON run submission page with 1YCR (MDM2-p53 case) example. Wheat and slate colors represent mandatory and optional fields, respectively.

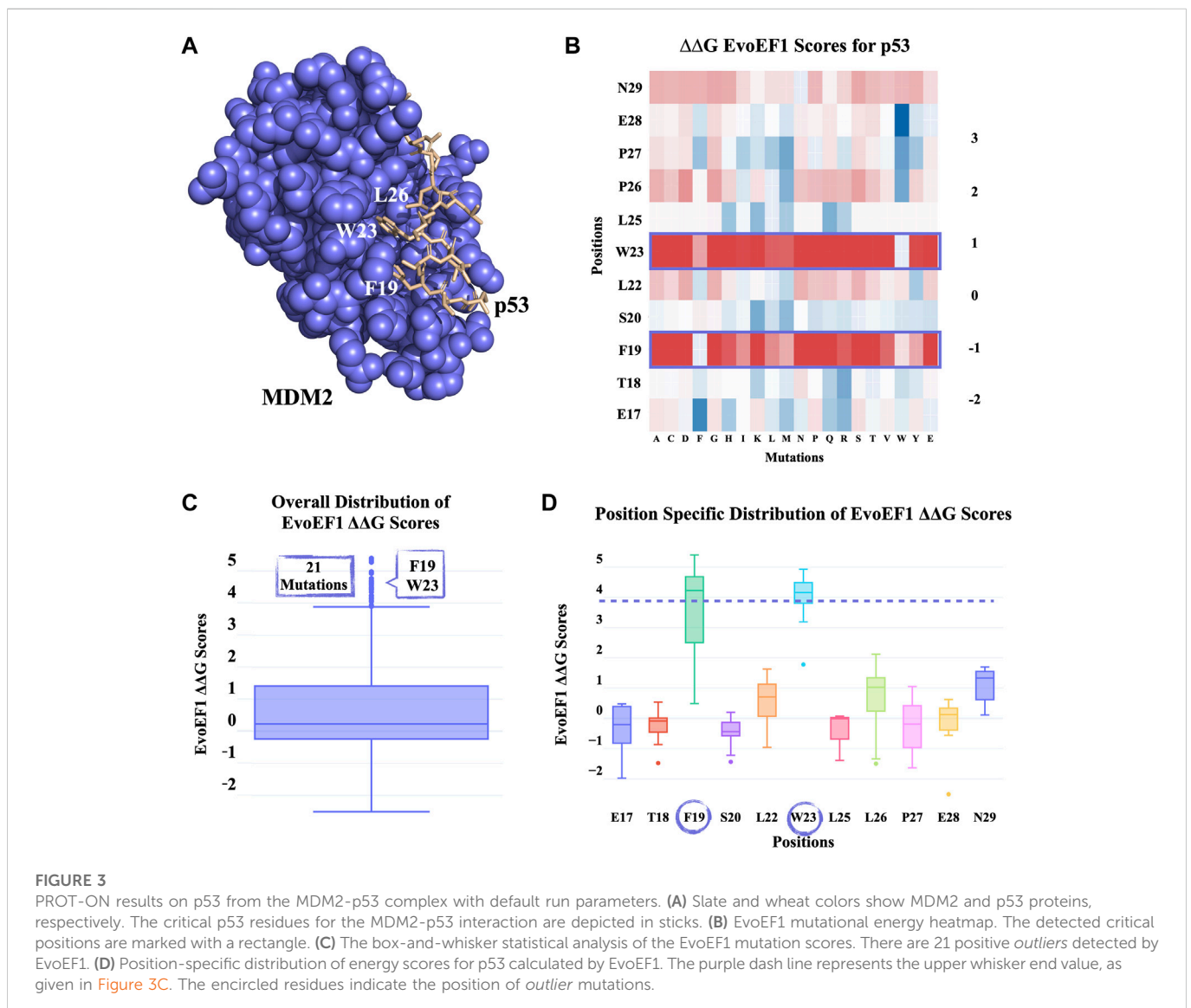
optional 5th step, the cut-off distance for defining the interfacial amino acids is set (choices range from 3.0 Å to 8.0 Å, changing with an increment of 0.1 Å). In the optional 6th step, the user can play with the IQR threshold to define the outliers (choices range from 1.5 to 2.5, changing with an increment of 0.5). In the optional 7th step, a PSSM file in the csv format is supplied. In the 8th and optional step, the e-mail address where the result link will be delivered is entered.

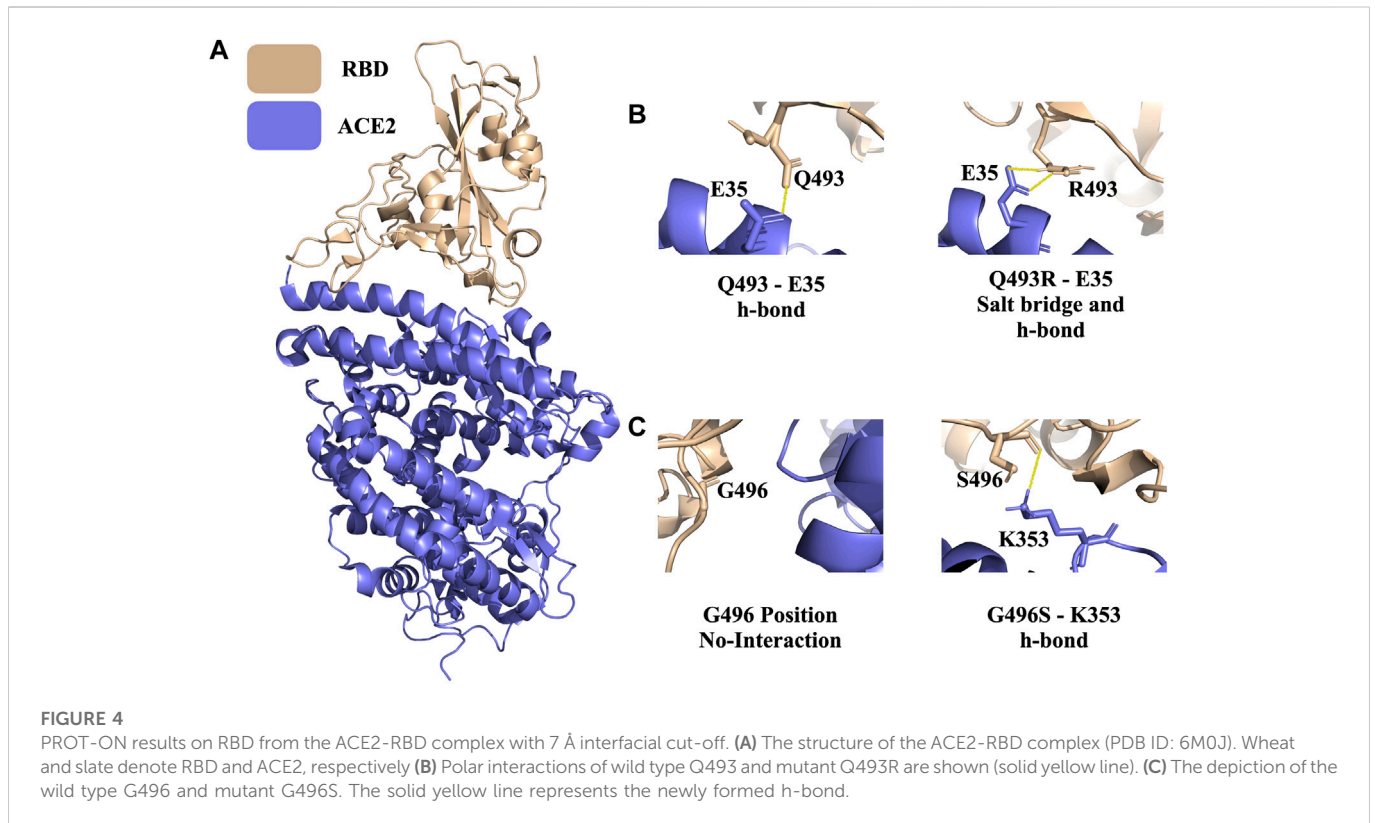
On the result page, a summary of the run, a downloadable run folder (stored for 1 week), three interactive energy plots, and the final designer mutation list is presented (Supplementary Figure S4). The summary section includes the number of interfacial amino acids selected within the defined cut-off distance, the total number of mutations imposed, and the number of final enriching and depleting mutations given the input IQR range. This section is followed by an interactive box-and-whisker plot of the sampled energies and their heatmap, as well as the energy distribution of each interfacial residue. The latter plot can be used to dissect the residues critical for binding. The relevant values on the energy plots can be visualized by hovering the mouse over them. The final enriching and depleting designer mutation list is presented as an interactive table where the user can focus on rows, select, and hide columns.

Below, we demonstrate the use of our web interface on two case examples, i.e., MDM2-p53, ACE2-Spike, and the stand-alone version with FoldX on the MCL1-NOXA complex.

### 3.3 Proof-of-concept server case 1: MDM2-p53 complex

p53 is a tumor suppressor, which decides for the apoptotic state of the cell in case of cellular damage (Moll and Petrenko, 2003). MDM2 is a regulator of p53, specifically controlling the overaccumulation of p53. At the extreme end, in cancer cells, MDM2 is often found to be overexpressed, blocking p53 activation (Nag et al., 2013). Therefore, it has been of utmost importance to cancer research to understand the key interactions between MDM2 and p53. In 1996, the MDM-p53 complex structure was solved with X-ray diffraction at 2.60 Å resolution (PDB ID: 1YCR) (Kussie et al., 1996). This structure showed that p53's key F19, W23, L26 residues penetrate the hydrophobic cleft of MDM2 (Figure 3A). In our first proof-of-





concept case, we run PROT-ON web service on p53 to probe whether we can back-calculate these crucial positions.

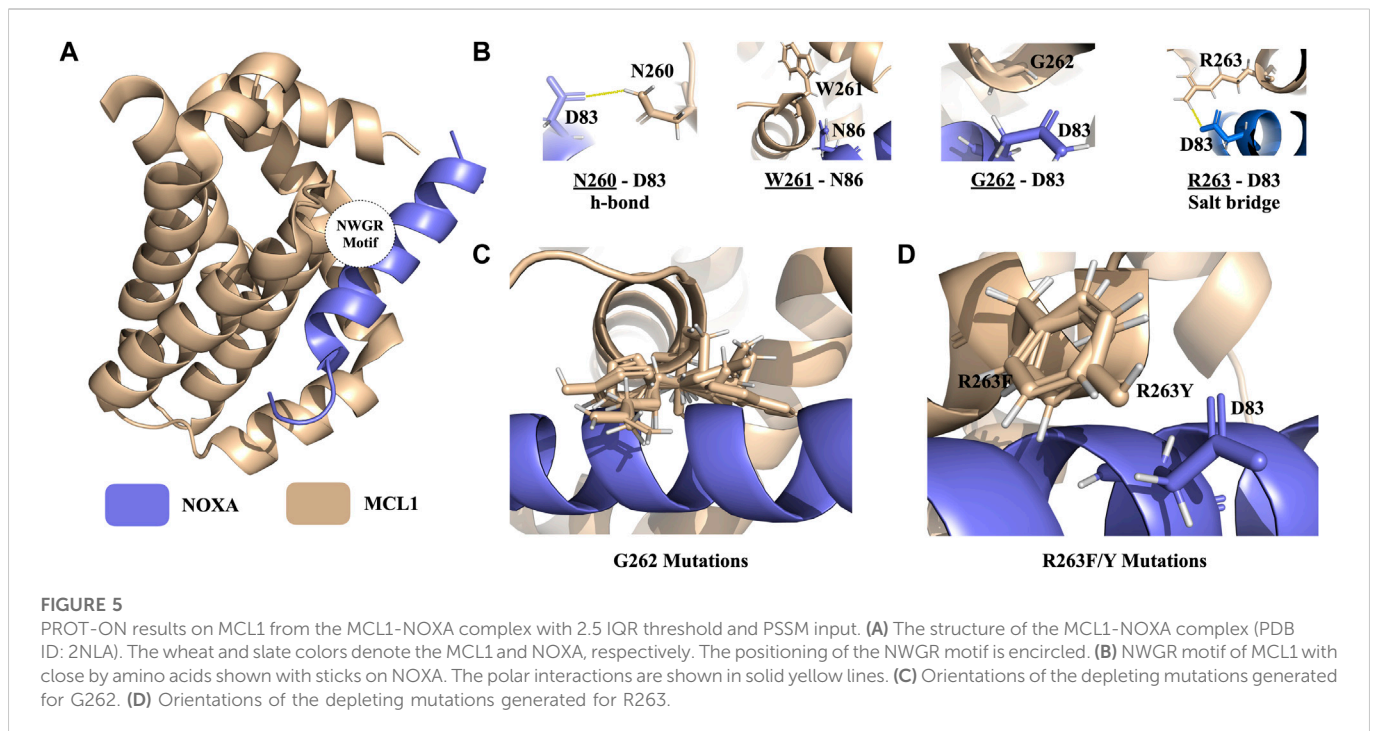
The PROT-ON run with default parameters leads to selection of 11 interfacial p53 residues, which results into the calculation of 209 interfacial mutations. As shown on the all energy heatmaps, EvoEF1 spots two from the three critical amino acids (Figure 3B), which is also reflected on to the box-whisker statistics and residue-based energy analysis (Figures 3C, D; Supplementary Table S1). F19D/E and W23C/K/R/T/V are proposed to be the most depleting designer mutations after imposing the stability filter of EvoEF1. Notice that, F19 and W23 are reported to reduce transactivation when mutated into charged or polar amino acids (Kussie et al., 1996). Even though PROT-ON did not propose any enriching mutations, we suggest that the lowest energy scores of EvoEF1 (E28W) still hold the potential to increase the binding affinity. This might happen by adding more hydrophobicity to the interface towards the two ends of the p53-peptide. This is also endorsed by the fact that the most enriching mutations are converted to an amino acid that has an aromatic side chain (Figure 3D). These run results can be reached at [http://proton.tools.ibg.edu.tr:8001/result/MDM2\\_p53\\_EvoEF1](http://proton.tools.ibg.edu.tr:8001/result/MDM2_p53_EvoEF1).

### 3.4 Proof-of-concept server case 2: ACE2-spike complex

Since late 2019, SARS-CoV-2, a novel SARS virus, has caused a global outbreak, leading to the deadliest pandemic of the 21st century. Since the early days of the outbreak, we know that the

infection cycle starts by having the receptor binding domain (RBD) of the virus' spike protein interacting with the host Angiotensin Converting 2 (ACE2) enzyme (Jubb et al., 2017). The ACE2-RBD crystal structure is resolved in 2020 with X-ray diffraction (2.45 Å resolution, PDB ID 6M0J) (Figure 4A) (Lan et al., 2020). Recently, Tian et al. performed experimental studies to investigate the role of dangerous RBD mutations, like the famous N501Y RBD mutation in Alpha, Beta, and Gamma variants. This mutation led to new  $\pi$ - $\pi$  and  $\pi$ -cation interactions, improving the binding affinity between ACE2 and RBD (Tian et al., 2021). In addition, K417N, G446S, Q493R, G496S, and Q498R are some of the commonly observed interfacial RBD mutations that are seen in the circulating Spike variants (Hodcroft, 2021).

For the ACE2-RBD case, we run PROT-ON with EvoEF1 with a 7 Å interfacial cut-off (to scan a wider range on the interface) by keeping the default IQR threshold. As a result, 39 RBD interfacial amino acids are detected on RBD and a total of 741 interfacial mutations are scanned. In this case, the box-and-whisker statistics defines the previously reported Q493R, G496S, and N501Y variants as binding enhancing mutations (Supplementary Figure S5). After the stability filtering, 36 mutations are selected as the final enriching designer mutations (Supplementary Table S2), with Q493R and G496S being in this final list. Q493R leads to the formation of a new salt bridge with E35 (Figure 4B). There is also a new hydrogen bond forming upon G446S mutation (Figure 4C). PROT-ON also suggests additional mutations on the same critical residues, as observed in the common variants, i.e., K417R, G446R, Q498E/F/I/K/L/M/N/S/T/Y, and N501L (Supplementary Table S2). As a result of the structural



analysis of K417R, we observe an additional salt bridge when compared to existing K417N variant. So, the final PROT-ON designer mutation set should be considered as “possibly dangerous” due to their binding enhancing effect. Our results for this case can be accessed through [http://proton.tools.ibg.edu.tr:8001/result/ACE2\\_RBD\\_EvoEF1](http://proton.tools.ibg.edu.tr:8001/result/ACE2_RBD_EvoEF1) (Supplementary Figure S5).

### 3.5 Proof-of-concept stand-alone case: MCL1-NOXA complex

Protein-protein interaction of the B-cell lymphoma-2 (BCL) family is responsible for the regulation of apoptosis by interacting with their coregulators. This process controls mitochondrial outer membrane permeabilization and causes the release of intermembrane space proteins, followed by the activation of caspases (Xie et al., 2021). MCL1 belongs to the anti-apoptotic class of the BCL family that can inhibit apoptosis and NOXA is its coregulator (Youle and Strasser, 2008). The crystal structure of the MCL1:mNoxaB BH3 complex was identified in 2.80 Å resolution and deposited in the PDB with 2NLA PDB ID (Figure 5A) (Czabotar et al., 2007).

Ivanov et al. (2016) determined a conserved “NWGR” (N260, W261, G262, R263) motif on MCL1 (Figures 5A, B). Xie et al. further explored that mutations on N260 and G262 from the NWGR motif results in the disruption MCL1 and NOXA complex formation. To explore whether PROT-ON can get back the same depleting effects on MCL1, we run the stand-alone version of PROT-ON with FoldX, with the default interface cut-off, and with 2.5 IQR threshold. In this case, we also inputted a PSSM file for MCL1. As an outcome, we obtained 26 interfacial amino acids with 494 interfacial mutations. As shown on the heatmap and

box-and-whisker analysis, N260, G262, and R263 come out as the most detrimental positions irrespective of the mutation type, so agreeing with the suggested importance of the NWGR motif (Supplementary Figure S6B). After stability and PSSM filtering, we are left with a final list of 28 designer-depleting mutations (Supplementary Table S3), where all G262 mutations except for C and T and R263F/Y mutations are proposed as designer-depleting mutations. From the inspection of the G262 mutation models, it is evident that there is no space left for a bulky residue change (Figure 5C). R263F/Y causes the removal of an important salt bridge and thus has an abolishing effect on the binding (Figures 5B–D). Run results for this case can be found at [http://proton.tools.ibg.edu.tr:8001/result/MCL1\\_NOXA\\_FoldX](http://proton.tools.ibg.edu.tr:8001/result/MCL1_NOXA_FoldX) (Supplementary Figure S6).

## 4 Conclusion

Here, we present a new tool, PROT-ON, which aims to find the critical mutations to redesign known protein-protein interactions. The proposed PROT-ON mutations can be especially utilized to speed up the targeted mutation search in protein-based therapeutics. To assist such a search, PROT-ON’s server not only suggest a mutation list, but it also displays several interactive energy plots to portray the scanned mutational landscape. We showcase the capabilities of our tool on three important cases with diverse functions. These proof-of-concept cases validate known critical positions/mutations of the studied complexes. They also offer new mutations to enhance or abolish the known interactions. Our hope is that these and other cases probed by the users will set a milestone in guiding novel experimental research.

## Data availability statement

The datasets presented in this study are provided as online links in the Results section. The stand-alone version of PROT-ON and its use can be reached at: <https://github.com/CSB-KaracaLab/prot-on>.

## Author contributions

All authors contributed equally to the writing of the manuscript. EK conceptualized the study and devised the PROT-ON methodology. MK developed the PROT-ON code and its server and analyzed the MDM2-p53 case. IY analyzed ACE2-RBD and MCL1-NOXA cases.

## Funding

This study is funded by the Turkish Health Institutes (TÜSEB) (Project code: 2019-TA-02, Project Number: 3393). All authors are supported by EMBO IG Grant (Project Number: 4421).

## Acknowledgments

We would like to thank Tülay Karakulak, Eda Şamiloğlu, and Mehmet Ergüven for their contribution to the conceptualization of this study. We

## References

- Amengual-Rigo, P., Fernández-Recio, J., and Guallar, V. (2021). Uep: An open-source and fast classifier for predicting the impact of mutations in protein–protein complexes. *Bioinformatics* 37, 334–341. doi:10.1093/bioinformatics/btaa708
- Brown, A., and Wilson, G. (2012). *The architecture of open source applications, volume ii*. Creative Commons.
- Czabotar, P. E., Lee, E. F., van Delft, M. F., Day, C. L., Smith, B. J., Huang, D. C., et al. (2007). Structural insights into the degradation of Mcl1 induced by BH3 domains. *Proc. Natl. Acad. Sci.* 104, 6217–6222. doi:10.1073/pnas.0701297104
- David, A., Razali, R., Wass, M. N., and Sternberg, M. J. (2012). Protein–protein interaction sites are hot spots for disease-associated nonsynonymous SNPs. *Hum. Mutat.* 33, 359–363. doi:10.1002/humu.21656
- David, A., and Sternberg, M. J. (2015). The contribution of missense mutations in core and rim residues of protein–protein interfaces to human disease. *J. Mol. Biol.* 427, 2886–2898. doi:10.1016/j.jmb.2015.07.004
- Grinberg, M. (2018). *Flask web development: Developing web applications with python*. Sebastopol, CA: O'Reilly Media, Inc.
- Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., et al. (2020). Array programming with NumPy. *Nature* 585, 357–362. doi:10.1038/s41586-020-2649-2
- Henriques, B. J., Olsen, R. K. J., Gomes, C. M., and Bross, P. (2021). Electron transfer flavoprotein and its role in mitochondrial energy metabolism in health and disease. *Gene* 776, 145407. doi:10.1016/j.gene.2021.145407
- Hodcroft, E. B. (2021). *CoVariants: SARS-CoV-2 mutations and variants of interest*.
- Ivanov, D., Dolgui, A., Sokolov, B., Werner, F., and Ivanova, M. (2016). A dynamic model and an algorithm for short-term supply chain scheduling in the smart factory industry 4.0. *Int. J. Prod. Res.* 54, 386–402. doi:10.1080/00207543.2014.999958
- Jankauskaitė, J., Jiménez-García, B., Dapkūnas, J., Fernández-Recio, J., and Moal, I. H. (2019). Skempi 2.0: An updated benchmark of changes in protein–protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics* 35, 462–469. doi:10.1093/bioinformatics/bty635
- Jemimah, S., and Gromiha, M. M. (2020). Insights into changes in binding affinity caused by disease mutations in protein–protein complexes. *Comput. Biol. Med.* 123, 103829. doi:10.1016/j.combiomed.2020.103829
- Jubb, H. C., Higuero, A. P., Ochoa-Montaña, B., Pitt, W. R., Ascher, D. B., and Blundell, T. L. (2017). Arpeggio: A web server for calculating and visualising interatomic interactions in protein structures. *J. Mol. Biol.* 429, 365–371. doi:10.1016/j.jmb.2016.12.004
- Kussie, P. H., Gorina, S., Marechal, V., Elenbaas, B., Moreau, J., Levine, A. J., et al. (1996). Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* 274, 948–953. doi:10.1126/science.274.5289.948
- Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., et al. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581, 215–220. doi:10.1038/s41586-020-2180-5
- McKinney, W. (2011). pandas: a foundational Python library for data analysis and statistics. *Python high Perform. Sci. Comput.* 14, 1–9.
- Moll, U. M., and Petrenko, O. (2003). The MDM2-p53 interaction. *Mol. cancer Res.* 1, 1001–1008.
- Nag, S., Qin, J., Srivenugopal, K. S., Wang, M., and Zhang, R. (2013). The MDM2-p53 pathway revisited. *J. Biomed. Res.* 27, 254–271. doi:10.7555/JBR.27.20130030
- Navio, D., Rosell, M., Aguirre, J., de la Cruz, X., and Fernández-Recio, J. (2019). Structural and computational characterization of disease-related mutations involved in protein–protein interfaces. *Int. J. Mol. Sci.* 20, 1583. doi:10.3390/ijms20071583
- Nishi, H., Tyagi, M., Teng, S., Shoemaker, B. A., Hashimoto, K., Alexov, E., et al. (2013). Cancer missense mutations alter binding properties of proteins and their interaction networks. *PLoS one* 8, e66273. doi:10.1371/journal.pone.0066273
- Pearce, R., Huang, X., Setiawan, D., and Zhang, Y. (2019). EvoDesign: Designing protein–protein binding interactions using evolutionary interface profiles in conjunction with an optimized physical energy function. *J. Mol. Biol.* 431, 2467–2476. doi:10.1016/j.jmb.2019.02.028
- Reese, W. (2008). Nginx: The high-performance web server and reverse proxy. *Linux J.* 2.
- Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L. (2005). The FoldX web server: An online force field. *Nucleic acids Res.* 33, W382–W388. doi:10.1093/nar/gki387
- Tian, F., Tong, B., Sun, L., Shi, S., Zheng, B., Wang, Z., et al. (2021). Mutation N501Y in RBD of spike protein strengthens the interaction between COVID-19 and its receptor ACE2. *BioRxiv*.
- Xie, V. C., Pu, J., Metzger, B. P., Thornton, J. W., and Dickinson, B. C. (2021). Contingency and chance erase necessity in the experimental evolution of ancestral proteins. *Elife* 10, e67336. doi:10.7554/eLife.67336
- Yates, C. M., and Sternberg, M. J. (2013). The effects of non-synonymous single nucleotide polymorphisms (nsSNPs) on protein–protein interactions. *J. Mol. Biol.* 425, 3949–3963. doi:10.1016/j.jmb.2013.07.012
- Youle, R. J., and Strasser, A. (2008). The BCL-2 protein family: Opposing activities that mediate cell death. *Nat. Rev. Mol. Cell Biol.* 9, 47–59. doi:10.1038/nrm2308

also would like to thank Ayşe Berçin Barlas for revising the code of PROT-ON. We would like to thank the HADDOCK web server developers for being us an inspiration for the design our web interface.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2023.1063971/full#supplementary-material>