
SCoV2-MD Documentation

Release 1.0

GPCR drug discovery group

Oct 28, 2021

CONTENTS:

1	Simulation search	1
2	Workbench	3
3	Virus evolution	9
4	Data download	11
5	Simulation submission	13
6	Citing SCoV2-MD	15

SIMULATION SEARCH

1.1 Quick search

To quickly search for simulations including a specific SARS-CoV-2 protein, the user can just input the protein name (e.g. Spike, NSP3, ...) into the search bar on the [Home](#) page. Alternatively, the user can select a protein of interest from the interactive graphical representation of the SARS-CoV-2 virus, also found on the [Home](#) page.

In both cases, the user will be redirected to a table with all the simulations that include the selected protein. Click on “Open” to visualize and analyze the simulation with our [Workbench](#) page. There, you will also find the details of the simulation, as well as links to download the simulation data.

1.2 Browser

Our [Browse](#) page provides an easy and quick tool to search for any simulation in our platform. Use the *Search* input to filter out the simulations shown in the table. It is possible to search by any of the terms included in the table (Included proteins, Uniprot entry name, Uniprot ID, etc.).

The table rows contain a summary of the most important features of the simulations. This includes links to further details of the receptor and molecules present in the simulated system. Click on “Open” to visualize and analyze the simulation with our [Workbench](#) page. There, you will also find the details of the simulation, as well as links to download the simulation data.

WORKBENCH

Contents

- *Workbench*
 - *Simulation report and files*
 - *Viewer*
 - *Toolkit*

The Workbench includes a set of online tools for the interactive visualization (*Viewer*) and analysis (*Toolkit*) of individual simulations. It also links the time-resolved protein dynamics with information on existing SARS-CoV-2 variants, their phylogeny, and the corresponding individual isolates. This allows evaluating the structural impact of variants taking into account protein dynamics.

2.1 Simulation report and files

By clicking on *Simulation report and files*, you will access the details of the system setup and simulation protocol, as well as links to download the simulation data.

2.2 Viewer

The viewer section of the Workbench page builds on the web-based structure viewer [NGL 2.00](#) and [MDsrv 0.3.5](#).

2.2.1 Mouse controls

- **Left button hold and move:** rotate camera around the center.
- **Middle button hold and move:** zoom camera in and out.
- **Middle button click:** center camera on an atom.
- **Right button hold and move:** translate camera in screen plane.
- **Left button click:** pick atom or distance.
 - When an atom is clicked, a label with information about it appears. Click at the background to deselect it, the label will disappear. To maintain a label, double-click on an atom. Double-click again on the atom to remove the label.

- To draw a distance line between two atoms just single-click one atom after the other. Distances can be removed by double-clicking on one of the atoms at the edges.
- It is also possible to remove all the atom labels and distances at once, with the **Clear labels/dists. button**.

2.2.2 Selection tools

Quick selection

Quick-selection buttons allow to rapidly display or hide the molecules present at the dynamics. Hover the buttons with your mouse to see the abbreviated name of these molecules, if any, which can be used to create your own selections.

Relevant domains

This section includes a list of relevant domains and annotations present in the protein, extracted from [Uniprot](#). Select or un-select a domain to represent it on the structure. More information on each domain can be found by clicking on the info button:

- **Seq. position:** Position of the domain in the Uniprot sequence.
- **Seq. (WT):** Wild type sequence of the domain.
- **Selection:** Selection of the domain based on the [NGL selection language](#).
- **Source:** Source Uniprot entry.

Custom selection

Use the text input field to specify your personalized representations. You can choose a representation type (licorice, cartoon, etc.) and a coloring scheme (color by element, by chain, etc.). Selections must be expressed using the [NGL selection language](#).

2.3 Toolkit

2.3.1 General

Interaction network

Interaction network, based on [Flare Plots](#), provides an interactive circular representation of contact networks. Noncovalent residue-residue interactions formed in the simulation are represented as lines connecting residue pairs. Hover or click a residue to highlight the lines representing the interactions in which it participates.

There are several options available

- **Interaction type: Select the type of interaction to display on the plot.**
 - **Hydrogen bonds:**
 $|AD| < 3.5\text{\AA}$
 $AHD < 70^\circ$
Where A (acceptor) and D (donor) are any atom except hydrogen, carbon or sulphur.
 - **Salt bridges:**
 $|AC| < 4.0\text{\AA}$

Where:

A (anion): ASP/OD1+OD2, GLU/OE1+OE2

C (cation): LYS/NZ, ARG/NH1+NH2, HIS/ND1+NE2

Based on [GetContacts](#).

– **Pi-cation:**

$|AC| < 6.0\text{\AA}$

$\text{CAn} < 60^\circ$

Where:

A (aromatic): center(PHE/CG+CE1+CE2), center(TRP/CD2+CZ2+CZ3),
center(TYR/CG+CE1+CE2), center(HIS/CG+CD2+CE1)

C (cation): LYS/NZ, ARG/NH1+NH2, HIS/ND1+NE2

Based on [GetContacts](#).

– **Pi-stacking:**

$|A1A2| < 7.0\text{\AA}$

$(n1, n2) < 30^\circ$

$(n1, A1A2) < 45^\circ$

$(n2, A1A2) < 45^\circ$

Where:

A1, A2 (aromatic rings): center(PHE/CG+CE1+CE2), center(TRP/CD2+CZ2+CZ3),
center(TYR/CG+CE1+CE2), center(HIS/CG+CD2+CE1)

Based on [GetContacts](#).

– **T-stacking:**

$|A1A2| < 5.0\text{\AA}$

$60^\circ < (n1, n2) < 90^\circ$

$(n1, A1A2) < 45^\circ$

$(n2, A1A2) < 45^\circ$

Where:

A1, A2 (aromatic rings): center(PHE/CG+CE1+CE2), center(TRP/CD2+CZ2+CZ3),
center(TYR/CG+CE1+CE2), center(HIS/CG+CD2+CE1)

Based on [GetContacts](#).

– **Van der Waals:**

$|AB| < \text{Rvdw}(A) + \text{Rvdw}(B) + 0.5$

Where A and B are any non-hydrogen atoms.

Based on [GetContacts](#).

– **Water bridges:** Two different residues forming a Hydrogen bond with the same water molecule. Based on [GetContacts](#).

– **Extended water bridges:** Two different residues forming a Hydrogen bond with two different water molecules which also form a hydrogen bond between them. Based on [GetContacts](#).

– **Hydrophobic:**

$|AB| < \text{Rvdw}(A) + \text{Rvdw}(B) + 0.5$

Where:

A, B: ALA+CYS+PHE+GLY+ILE+LEU+MET+PRO+VAL+TRP and element C or S

Based on [GetContacts](#).

• **Display:**

– **Interacting pairs:** Allows to filter out interactions formed by residues that are 5 or less residues apart.

- **Simulation:** It is possible to summarize the interactions formed through all the trajectory frames. The frequency of each interaction is represented by the thickness of the lines connecting residues.
- **Show in structure:** Click to display structural representations of the residues selected (clicked) at the flare plot. Unclick to hide them. If there are no residues selected at the flare plot, nothing will happen.
- **Clear plot:** Click to delete all selections made on the plot.
- **Download data:** Click to download the plot data.

Root mean square deviation (RMSD)

This tool computes the RMSD of all the conformations in a target trajectory to a reference conformation using the `_rmsd` module of MDtraj. If more than one trajectory is available, the user can select the trajectory to be used. Also, it is possible to specify the frames to be considered. The first frame of the trajectory is used as a reference structure by default, but this can be specified by the user. It is also possible to choose which atoms are going to be considered in the calculation: only alpha carbons, non-hydrogen protein atoms, protein C-alpha, etc. Results are shown in a plot (RMSD by time or by frame). It is possible to download the plot as an image or all the obtained data as a CSV file.

Root mean square fluctuation (RMSF)

The RMSF is computed for all the alpha carbons of the protein using the `rmsf` module of MDtraj. It is calculated based on the average structure of the simulation, obtained by averaging the coordinates of each atom during the trajectory. The trajectory is aligned to the obtained average structure before the RMSD is calculated.

2.3.2 Variant impact

This tool interactively references the MD simulations with information on existing SARS-CoV-2 variants obtained from [GISAID](#). Each of the variants is annotated with static (mutation-dependent) and time-dependent (computed on the basis of the simulation dynamics) descriptors of their impact on multiple aspects of the protein's structure, dynamics, and function. By assigning user- or pre-defined weights to these descriptors, the descriptors are combined in an impact score.

Sequence selector

This tool displays the wild-type sequence of the protein or proteins present in the simulation, extracted from Uniprot (residue name and sequence position). Place your mouse on a residue to see the structure residue ID corresponding to that position. Have in mind that the residue ID may not correspond to the Uniprot sequence position.

Positions that have a known variant are colored in the sequence. It is possible to filter the variants that are displayed on the sequence to show only those present in a particular isolate of interest. This can be done with the **“Isolate”** dropdown.

Click on a position with known variants to highlight it in the structure. Please note that the model may incorporate some mutations with respect to the wild-type sequence, so the residue name of the sequence (wild type) can, in some cases, not correspond to the residue name in the structure.

When a position with known variants is clicked, a pop-up with the list of variants appears. Click on one of the variants to obtain more information about its descriptors and impact score.

The **“Display in viewer”** checkbox control whether or not the variant information is displayed at the Viewer. If it is unselected, clicked positions will not be highlighted on the structure, and it will not be possible to color the structure based on the impact score.

Descriptors of variant impact

Our descriptors include:

- **Mutation effect descriptors, representing the impact of the amino acid change.** Unless indicated otherwise, they were obtained from the [MutaPro](#) database.
 - BLOSUM90 score
 - Charge difference, based on the charge of the wild type and mutated amino acid
 - Epstein’s coefficient of difference
 - Experimental exchangeability
 - Grantham’s distance
 - Miyata’s distance
 - Sneath’s index
 - Several scales of change in hydrophobicity: Kyte-Doolittle, Eisenberg-Weiss, Engelman, Hessa, Hopp-Woods, Janin, Moon-Fleming, Wimley-White, Zhao-London
 - Variant effect predictions extracted from the database [Mutfunc SARS-CoV-2](#), including conservation, structural consequences and [experimental antibody escape data](#).
- **Time-dependent predictors, consisting of parameters extracted from the wild type simulation for the residue affected by the mutation.**
 - RMSD of the residue with relation to the first trajectory frame, obtained using the [rmsd](#) module of MDtraj.
 - RMSF of the residue atoms, obtained with the [rmsf](#) module of MDtraj.
 - Solvent-accessible surface (SASA) of the residue, based on the Shrake and Rupley algorithm, calculated with the [shrake_rupley](#) algorithm of MDtraj.
 - Chi1 angle of the residue, which corresponds to the first side chain torsion angle formed between the 4 atoms over the CA-CB axis, if available. The Chi1 angle is calculated using the [compute_chi1](#) module of MDtraj.
 - Number of contacts that the residue makes with other protein residues, obtained with [GetContacts](#). This includes hydrogen bonds, salt bridges, hydrophobic contacts, pi-cation contacts, pi-stacking contacts, t-stacking contacts, Van der Waals, water bridges, extended water bridges and total contacts.

Click on the + sign beside each time-dependent predictor to see the value of the predictor by simulation time. By clicking on the plot, the user can set the Viewer to the corresponding trajectory frame.

- **User-provided descriptors**, which can be provided by the user as a CSV-formatted file, based on the template provided at the top of the section (example [here](#)). Upload the CSV, with comma separated values, using the “Browse” and “Upload” buttons to generate the corresponding sliders.
- Of note, custom descriptors enable the inclusion of arbitrary external data, as well as non-linearities and interactions, in the prediction models.

Impact score

The descriptors are combined in an impact score, defined as the sum of each weighted descriptor. Users may assign predefined or custom weight combinations to the descriptors to reflect various aspects of the structural impact of the variant.

Predefined weights can be enabled with buttons in the *Impact score* section that load the corresponding sets of coefficients in the “weight” sliders inside the *Mutation effect* and *Time-dependent* boxes. They include:

- **Model-based weights: obtained based on regularized regression models (LASSO) trained on 23 per-variant covariates, used to estimate the impact of each variant on the following:**
 - **Fitted to antibody escape:** trained to estimate, for each variant, its impact in terms of the potential for [antibody evasion](#).
 - **Fitted to binding:** trained to estimate, for each variant, its impact in terms of the change in [binding affinity](#) between the RBD and the ACE2 receptor.
 - **Fitted to expression:** trained to estimate, for each variant, its impact in terms of the change in [expression](#) of the RBD on yeast cells.
- **Other weights:**
 - **RMSF and contacts:** weights set to consider, equally, RMSF and all contact types.

To set a customized set of weights, the user can manually set the “weight” sliders inside the *Mutation effect* and *Time-dependent* boxes.

If a variant is selected in the *Sequence selector*, the obtained score is presented together with a *q value*, showing its rank in the distribution of all variants ($q=0$, 0.5, and 1 respectively mean that the selected amino acid variant achieves the minimum, median and maximum effect score with respect to the other ones observed in the available sequences). Also, the distribution of all variants is shown as a histogram, on which the impact score of the selected variant is highlighted. A significant shift of the variant score to the right or left from the distribution peak reflects a variant with a high propensity to disturb protein function based on the selected descriptors.

Finally, a color scale is generated based on the impact score of the variants in the protein sequence. Both the *Sequence selector* and the protein structure in the Viewer can be automatically colored based on this scale. In residues with more than one variant, the color corresponds to the average score of all the residue variants.

VIRUS EVOLUTION

In the *Virus evolution* section of the [Home](#) page, the user will find a visualization tool for viral phylogenetic data, extracted from [GISAID](#). This tool allows tracking SARS-CoV-2 isolates by region, gender, age, etc. as well as follow the emergence of mutations as a function of time. Each circle in the tree reflects one entry for a viral genome (isolate). By clicking on a viral genome, the user will find more information about it. Also, if the genome includes novel mutations that imply an amino acid change, the user can directly investigate the variant location on the 3D structure of implicated viral proteins as well as stream time-resolved dynamics on the fly.

DATA DOWNLOAD

All files related to the simulations stored in this platform are available for download at the [Workbench](#) page of each simulation, by clicking on *Simulation report and files*.

The [Workbench](#) page can be accessed from our [Browser](#).

SIMULATION SUBMISSION

In order to ensure data quality, we currently restrict data submission to MD data which will be published (peer-reviewed process).

You can upload your data prior to publication for the peer-review process (restricted access only for user and reviewer). Upon paper acceptance, your MD data will be published on this online resource.

Do you want to submit a simulation? Please contact us at info@gpcrmd.org with subject 'SCoV2-MD submission' and we will provide you a registration link.

CITING SCoV2-MD

If you use SCoV2-MD in your work, please cite:

- Torrens-Fontanals, M., Peralta-García, A., Talarico, C., Guixà-González, R., Giorgino, T., Selent, J. SCoV2-MD: a database for the dynamics of the SARS-CoV-2 proteome and variant impact predictions. *Nucleic Acids Research* (2021). doi:[10.1093/nar/gkab977](https://doi.org/10.1093/nar/gkab977)