frontiers in Medical Technology

DATA REPORT
published: 15 February 2021
doi: 10.3389/fmedt.2021.644039

The SIRAH-CoV-2 Initiative: A Coarse-Grained Simulations' Dataset of the SARS-CoV-2 Proteome

*Pablo G. Garay, Exequiel E. Barrera, Florencia Klein, Matias R. Machado, Martín Soñora and Sergio Pantano**

Biomolecular Simulations Group, Institut Pasteur de Montevideo, Montevideo, Uruguay

Keywords: coarse-grain, simulation, data sharing, molecular dynamics, virus, COVID-19, microsecond

OPEN ACCESS

Edited by:
Diana Lousa,
New University of Lisbon, Portugal

Reviewed by:
Manuel N. Melo,
New University of Lisbon, Portugal
Donal MacKernan,
University College Dublin, Ireland

****Correspondence:***
Sergio Pantano
spantano@pasteur.edu.uy

Specialty section:
This article was submitted to
Pharmaceutical Innovation,
a section of the journal
Frontiers in Medical Technology

Received: *19 December 2020*
Accepted: *20 January 2021*
Published: *15 February 2021*

Citation:
Garay PG, Barrera EE, Klein F, Machado MR, Soñora M and Pantano S (2021) The SIRAH-CoV-2 Initiative: A Coarse-Grained Simulations' Dataset of the SARS-CoV-2 Proteome. Front. Med. Technol. 3:644039. doi: 10.3389/fmedt.2021.644039

INTRODUCTION

During the last decades, the broad community of computational biophysicist/biochemists has developed computational tools to quickly test molecular hypotheses, support, complement, and even substitute experimental data in a reliable and reproducible fashion. As a by-product of these advances, enormous amounts of data are being generated (1). Unfortunately, good practices about data archiving, documenting, and sharing are not in pace with the formidable capacity to produce information. This often results in suboptimal utilization of efforts and resources, leaving authentic "data treasures" undiscovered. This redounds in a useless replication of work, which often times is only needed as input for further investigation rather than representing an objective themselves (1). It, therefore, becomes increasingly important to make computational biophysics data publicly available, searchable, and downloadable, adhering to the "FAIR" principles (2).

Among many others, the European community has advanced a large and coordinated initiative, the European Open Science Cloud (EOSC), which is aimed at sharing and re-using scientific content increasing transparency and accountability. OpenAIRE is a socio-technical infrastructure for scholarly communication and Open Science (3). It offers data store ensuring long-term preservation of relatively "big size" datasets. Among others, the Zenodo database (4) provides a simple and fast upload system, with the possibility to immediately obtain a DOI identifier for each data set, including the option to update data sets separately.

Early in 2020, the COVID-19 pandemic pervaded virtually all personal and scientific activities with extensive lockdown regimes in most countries across the world. In response to this extraordinary context, the entire scientific community devoted massive efforts to study SARS-CoV-2 at basic and applied levels. The Biocomputing community was not an exception and showed a strong commitment endorsed by hundreds of groups around the globe (5). Many researchers reoriented their priorities, offering a swift response to the emergency, providing fresh structural and dynamical perspectives on viral variability, drug targets, effect of mutations, etc. (5). As a result, only a few months after the beginning of the pandemic, it was possible to find many data-sharing initiatives scattered in different portals and repositories.

In this context, our group undertook the initiative of simulating and sharing the raw data of coarse-grained (CG) simulations of the SARS-CoV-2 proteins reported in the PDB, in the apo state. **Figure 1** shows the representative structures reported in the PDB database until October 30, 2020. We named this "*The SIRAH-CoV-2 Initiative*," which was carried out in collaboration with the Uruguayan National Center for Supercomputing, ClusterUY (https://www.cluster.uy) (8). The raw data for individual CG Molecular Dynamics (MD) simulations is available from the Zenodo database (9–28).

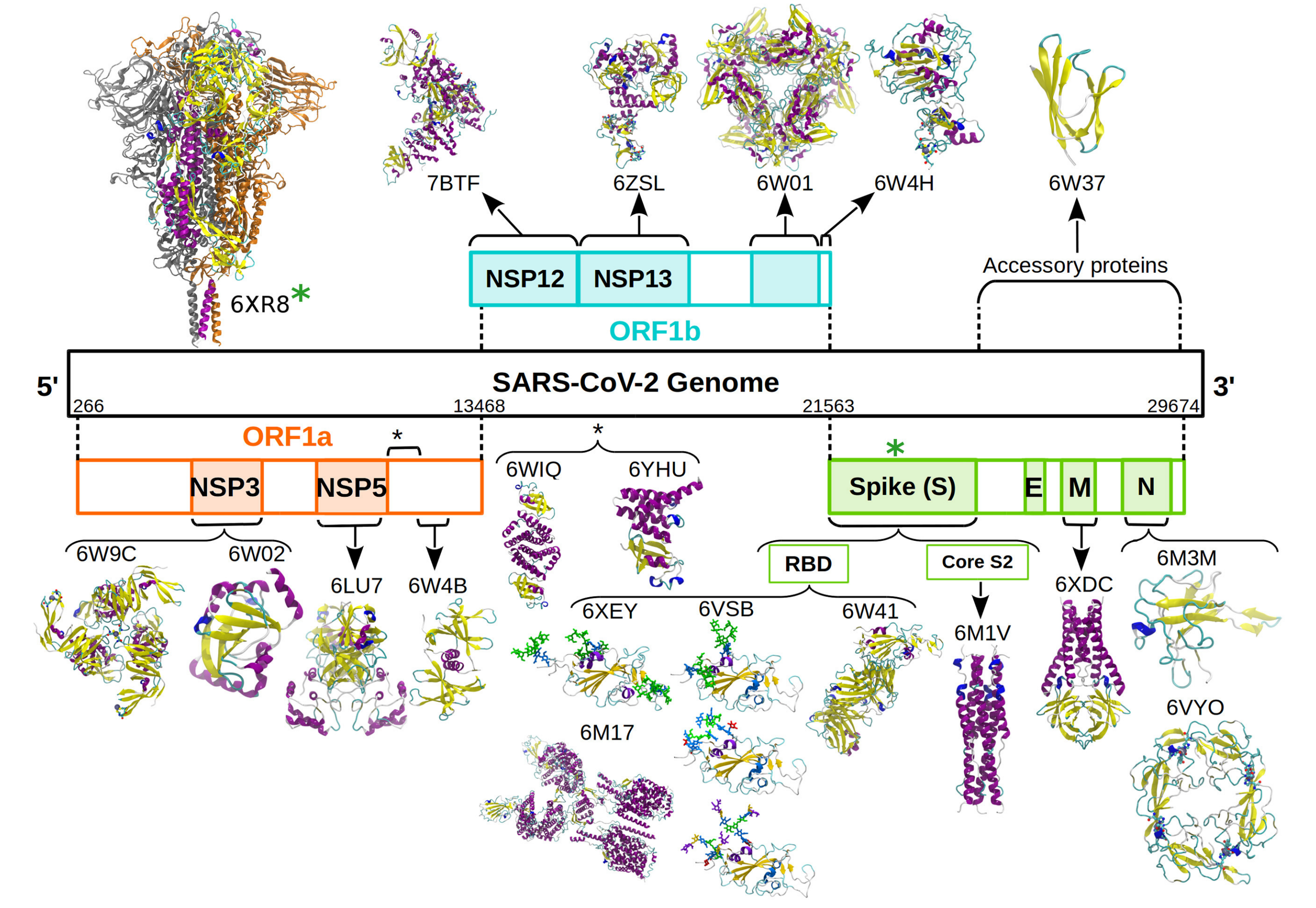

FIGURE 1 | Schematic representation of the SARS-CoV-2 genome and associated proteins. All the proteins simulated are presented as cartoon and colored according to their secondary structure following the standard VMD color code (6). Glycans are presented as sticks and colored according to the SNFG color scheme (7). The D614G mutation was introduced in the soluble domain of the wild type Spike protein (green asterisk).

METHODS

Simulations were performed using the SIRAH force field 2.0 (29) running with the Amber18 suite (http://ambermd.org) at ClusterUY. Interaction parameters for bound divalent cations and glycans were reported by Klein et al. (30) and Garay et al. (31), respectively.

Coordinates were downloaded from the PDB database (PDBs id: 6VYO, 6W01, 6LU7, 6W02, 6W4B, 6M3M, 6W9C, 6W4H, 6W41, 6YHU, 6W37, 6WIQ, 7BTF, 6M17, 6VSB, 6M1V, 6XDC, 6ZSL, 6XEY, 6XR8). Non-protein, non-glycan molecules and ions not coordinated by proteins were removed (e.g., water and molecules present in crystallization buffers). When deemed necessary, missing residues were reconstructed with ModLoop (32). The D614G mutation in the SARS-CoV-2 Spike protein was introduced on the wild type structure (PDB id: 6XR8) by simply deleting the side chain of Asp614 and renaming the residue. Only in this particular case, missing loops were completed using SWISS-MODEL at https://swissmodel.expasy.org. All structures were protonated using PDB2PQR (33) at a pH = 7. The orientation of the protein in PDB id 6XDC was set according to the OPM database (https://opm.phar.umich.edu/proteins/5172), with a pre-equilibrated patch containing POPC, POPE, and POPS phospholipids in a 1:2:1 relation according to the experimental data (34). Interaction parameters for lipids were taken from Barrera et al. (35). The glycosylation trees were added/completed (in PDB ids 6VSB and 6XEY) using the Glycan Modeler & Reader utility from CHARMM-GUI (36).

All parameters are available for download from the SIRAH force field web page (http://www.sirahff.com).

Protonated structures were mapped to CG with SIRAH Tools (37). Solutes were centered in an octahedral box filled with pre-equilibrated SIRAH's CG water molecules named WT4 (38). An ionic strength of 0.15 M was set by randomly replacing WT4 molecules with Na^+ and Cl^- CG ions (39).

Since SIRAH uses a Hamiltonian common to any atomistic MD simulation, the 6–12 terms used to treat Lennard-Jones interactions might lead to large repulsions if initial structures suffer from clashes. Because of this, gentle initialization protocols aimed to resolve steric clashes are strongly recommended.

The simulation protocol consisted of:

1) Solvent and side chains relaxation by 5,000 steps of energy minimization, imposing positional restraints of 2.4 kcal mol^{-1} Å^{-2} on backbone beads corresponding to the nitrogen and carbonylic oxygen (named GN and GO, respectively). When Zinc or glycans are present, these restrains also apply to the beads corresponding to the metal ions and sugar rings (named ZnX, GO2, GNac, GO3, GO4, GC6, GC1, and GO7).
2) Full system relaxation by 5,000 steps of unrestrained energy minimization.
3) Solvent equilibration by 5 ns of MD in the NVT ensemble at 300 K, imposing positional restraints of 2.4 kcal mol^{-1} Å^{-2} on the whole protein and glycans and Zinc ions, when present.
4) Biomolecule relaxation by 25 ns of MD in the NVT ensemble at 300 K, imposing positional restraints of 0.24 kcal mol^{-1} Å^{-2} on the mentioned beads.
5) Same as step 4 with the position restrains of 0.12 kcal mol^{-1} Å^{-2}.
6) Production simulation in the NPT ensemble at 300 K and 1 bar.

We used a time step of 20 fs and a direct cutoff of 1.2 nm for non-bonded interactions and Particle Mesh Ewald (PME) for long-range electrostatics (40, 41). Snapshots were recorded every 200 ps. PME was calculated at every integration step owing to code restrictions, and the neighbor list was updated whenever any atom had moved more than one-half a non-bonded "skin" of 0.2 nm. A Fourier spacing close to 0.1 nm was used. The whole system was coupled to a Langevin thermostat (42) with a collision frequency of 50 ps^{-1} and to a Berendsen barostat (43) with a relaxation time of 1 ps.

The multimicroseconds CG MD trajectories of SARS-CoV-2 proteins include the information required to visualize, analyze, and backmap on VMD (6). Each entry is constituted by three subsets of data associated with the same CG MD simulation. The first set (referred as raw data) contains the system's topology, starting configuration, simulation report, last checkpoint, and trajectory in AMBER format and allows continuing the simulation. The second subset contains a "stripped" version of the MD, not including solvent, while the third contains a "skipped" trajectory with one frame every 10 ns.

Since CG beads in SIRAH are mapped from atoms' position, it is possible to get direct measures from the trajectories using VMD tools. These include Root Mean Square Deviation (RMSD), radius of gyration, etc. Moreover, a tcl script corresponding to SIRAH Tools is present in each tar file of the dataset that performs additional analyses and secondary structure content from the VMD's Tcl/Tk console (6). Typing sirah_help displays all available options. It enables macros for visualizing and coloring residue types, the element corresponding to each CG bead, among others. This tool is totally compatible with all the functions on VMD and used the same color schemes. Provided that Amber Tools (44) is locally installed, it is possible to obtain pseudoatomistic structural models at every point of the trajectory.

UTILIZATION OF THE DATA

We started this initiative to provide our colleagues with a complete and homogeneous set of CG MD simulations that could facilitate the analysis of large-scale dynamics of SARS-CoV-2 proteins.

Aimed to provide the readers with a brief example of the performance of SIRAH in comparison with fully atomistic simulations, we compared the 15 µs long simulation of SARS-CoV-2 Main protease, present in our database, with an 10 µs long all-atoms simulation of the same protein reported by the Taiji's group and deposited in the Mendeley database (45). Reciprocal (or 2D) RMSD comparison showed that both trajectories visit different conformations with a checkered pattern indicative of conformational fluctuations in both trajectories (**Figure 2A**). The RMSD using the experimental structure as a reference showed a higher deviation for the CG trajectories of both chains (**Figure 2B**, bottom), while the gyration radii showed that the CG protein sampled also higher values (**Figure 2B**, middle). Despite these differences the secondary structure elements were well-maintained during the CG trajectory (**Figure 2B**, top). The traces for CG and all-atom simulations were similar, although with a loss of nearly 5% in the content of extended beta conformations in the CG case. A superposition against the experimental structure on both trajectories showed roughly comparable features (**Figures 2C,D**).

A possible use for the dataset is described in the following example. On June 14th, Liu et al. deposited the Cryo-EM structure of the SARS-CoV-2 Spike glycoprotein bound to a human antibody (47). This structure showed that Spike's Receptor Binding Domain (RBD) was glycosylated at Asn331, 343, and 481. Surprisingly, Asn481 neither showed the canonical glycosylation motif nor was previously reported as glycosylated (48). The vicinity of this site to the Angiotensin Converting Enzyme 2 (ACE2) binding zone poses the question of whether glycosylation at Asn481 could modulate the RBD-ACE2 binding (**Figures 2E,F**). To address this question we took the trajectory of the triple glycosylated RBD from the database (27). First, we used the backmapping capabilities of SIRAH to backmap the trajectory. Second we superimposed the backmapped structures from this CG simulation on the X-ray structure of the RBD-ACE2 complex [PDB id 6VW1 (49)]. This generated an ensemble of possible conformations of glycosylated RBD putatively bound to ACE2, which provided rough insights about possible steric clashes. Despite being close to the ACE2 interface, glycosylation at Asn481 would not be expected to create steric clashes with the human receptor (**Figure 2G**). Similarly, glycan-glycan interactions at the RBD-ACE2 interface between Asn481 and Asn90, its closest glycosylation site on ACE2 (asterisk in **Figures 2E,H**) seemed unlikely despite their large flexibility because both moieties remained at opposite sides of the protein-protein interface. Clearly, a thorough analysis would include

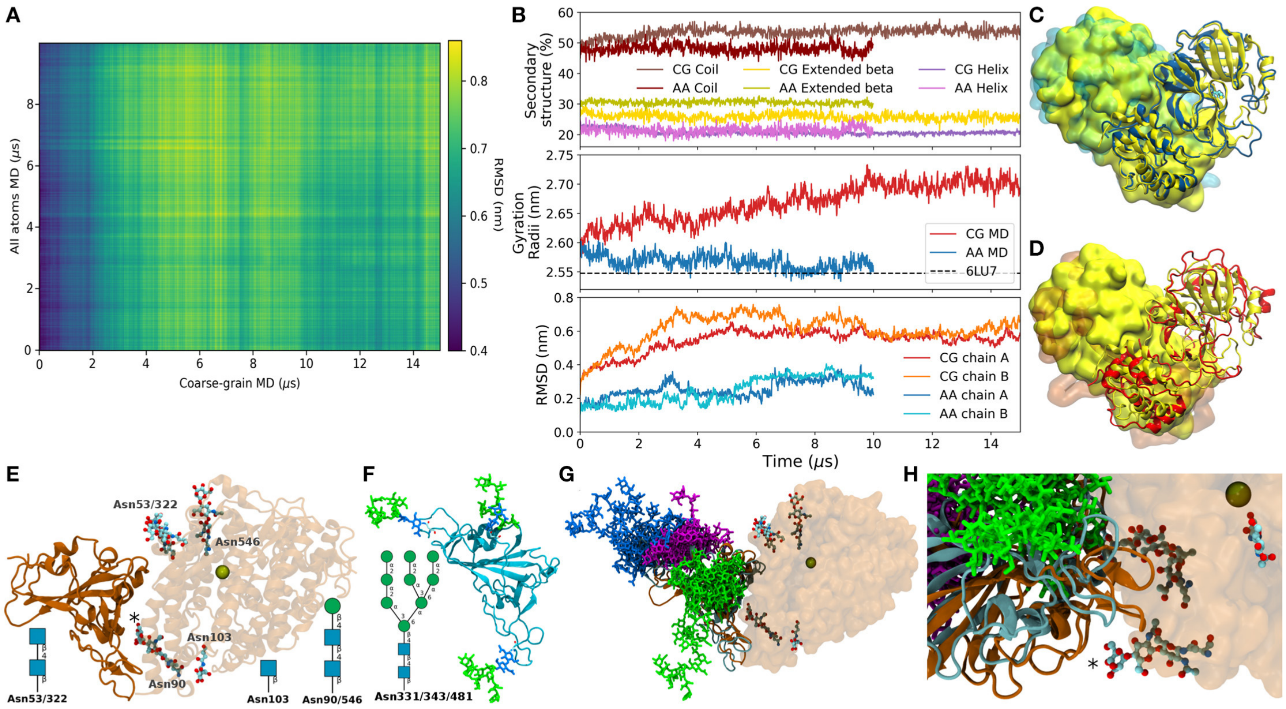

FIGURE 2 | **(A)** Bi-dimensional RMSD of the Cα between the atomic trajectory and the backmapped conformations from the CG trajectory. **(B)** Top: Percentage of secondary structure content. Middle: Gyration Radii of the whole proteins from the atomic (blue), and backmapped trajectory in red. Bottom: RMSD of the Cα between the trajectories and the experimental structure, separated by chains. **(C)** Superposition between the experimental structure (yellow) and the last conformation of the atomic trajectory (blue). Chains A are presented as cartoon and chains B as presented as surface. **(D)** Same as **(C)** with the backmapped structure of the last frame of the SIRAH trajectory in red. **(E)** X-ray structure PDB 6VW1 in cartoon representation, ACE2 is semitransparent, glycosylations are show as sticks colored according to element. The Zinc ion present in the binding site is shown as a space-filling sphere. The asterisk marks the glycosylation at Asn90 on ACE2. The N-glycosylation solved by X-ray on each position are indicated schematically. **(F)** Starting conformer of the RBD glycosylated at Asn331, 343, and 481 colored according to the SNFG color scheme (46). **(G)** Superposition of backmapped structures from the glycosylated RBD and that in the structure 6VW1. Proteins are colored according to panels **(E)** and **(F)**. RBD Glycosylations sites are shown in magenta, blue, and green for Asn331, Asn343, and Asn481, respectively. Only one conformed every 1 μs is shown. **(H)** Close up on the RBD-ACE2 interaction showing the positions of Asn481 (green), and the closest glycans in ACE2.

simulation of complex and all possible glycosylation motifs. Although perfectly possible, this goes beyond the scope of this Data Report.

Nevertheless, besides being useful to foresee large conformational changes and the gross determinants of possible interactions, this kind of information could be useful to decide the expression systems of choice in relation to the length or nature of the glycoforms attainable by prokaryote, insects, or mammalian cells.

Finally, we would like to point out that this is a live initiative and new simulation data will be added. Moreover, we remain open to produce additional simulations upon request.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://zenodo.org.

AUTHOR CONTRIBUTIONS

PGG and SP drafted the manuscript. All authors performed MD simulations and contributed to the writing.

FUNDING

This work was partially funded by FOCEM (MERCOSUR Structural Convergence Fund), COF 03/11, and by the URGENCE COVID-19 fundraising campaign of Institut Pasteur.

ACKNOWLEDGMENTS

PGG, EB, MM, and SP belong to the SNI program of ANII. FK and MS were doctoral fellows of ANII. Some of the graphic cards used in this research were donated by the NVIDIA Corporation. We were extremely grateful to Sergio Nesmachnow, Sebastián Valenzuela, and Santiago Iturriaga from Cluster-UY for their help and willingness to support this project.

REFERENCES

1. Riccardi E, Pantano S, Potestio R. Envisioning data sharing for the biocomputing community. *Interface Focus.* (2019) 9:20190005. doi: 10.1098/rsfs.2019.0005
2. Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data.* (2016) 3:160018. doi: 10.1038/sdata.2016.18
3. Manola N, Rettberg N, Manghi P, Mertens M, Schmidt B, Steiner T, et al. *Achieving Open Science in the European Open Science Cloud: Setting Out OpenAIRE's Vision and Contribution to EOSC.* (2019). doi: 10.5281/zenodo.3610132
4. European Organization for Nuclear Research, OpenAIRE. *Zenodo* (2013). doi: 10.25495/7gxk-rd71
5. Amaro RE, Mulholland AJ. A community letter regarding sharing bimolecular simulation data for COVID-19. *J Chem Inf Model.* (2020) 60:2653–6. doi: 10.1021/acs.jcim.0c00319
6. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph.* (1996) 14:33–8. doi: 10.1016/0263-7855(96)00018-5
7. Neelamegham S, Aoki-Kinoshita K, Bolton E, Frank M, Lisacek F, Lütteke T, et al. Updates to the symbol nomenclature for Glycans guidelines. *Glycobiology.* (2019) 29:620–4. doi: 10.1093/glycob/cwz045
8. Nesmachnow S, Iturriaga S. Cluster-UY: collaborative scientific high performance computing in Uruguay. In: Torres M, Klapp J, editors. *Supercomputing*. Cham: Springer International Publishing (2019). p. 188–202.
9. Klein F, Barrera E, Garay P, Machado M, Soñora M, Pantano S. *SIRAH-CoV2 Initiative: RNA Binding Domain of Nucleocapsid Phosphoprotein (PDB id:6VYO)* (2020). doi: 10.5281/zenodo.3747621
10. Barrera E, Garay P, Klein F, Machado M, Soñora M, Pantano S. *SIRAH-CoV2 Initiative: NSP15 Endonuclease (PDB id:6W01)* (2020). doi: 10.5281/zenodo.3748659
11. Soñora M, Barrera E, Garay P, Klein F, Machado M, Pantano S. *SIRAH-CoV2 Initiative: Main Protease (PDB id:6LU7)* (2020). doi: 10.5281/zenodo.3748665
12. Barrera E, Garay P, Klein F, Machado M, Soñora M, Pantano S. *SIRAH-CoV2 Initiative: Apo ADP Ribose Phosphatase of NSP3 From SARS CoV-2 (PDB id: 6W02)* (2020). doi: 10.5281/zenodo.3751583
13. Soñora M, Barrera E, Garay P, Klein F, Machado M, Pantano S. *SIRAH-CoV2 Initiative: NSP9 RNA Binding Protein (PDBid:6W4B).* (2020). doi: 10.5281/zenodo.3754643
14. Klein F, Barrera E, Garay P, Machado M, Soñora M, Pantano S. *SIRAH-CoV2 Initiative: Nucleocapsid Protein N-Terminal RNA Binding Domain (PDB id:6M3M).* (2020). doi: 10.5281/zenodo.3755835
15. Klein F, Barrera E, Garay P, Machado M, Soñora M, Pantano S. *SIRAH-CoV2 Initiative: Papain-Like Protease (PDB id:6W9C).* (2020). doi: 10.5281/zenodo.3766270
16. Soñora M, Barrera E, Garay P, Klein F, Machado M, Pantano S. *SIRAH-CoV2 Initiative: NSP16–NSP10 Complex (PDB id:6W4H).* (2020). doi: 10.5281/zenodo.3779057
17. Soñora M, Barrera E, Garay P, Klein F, Machado M, Pantano S. *SIRAH-CoV2 Initiative: S1 Receptor Binding Domain in Complex With Human Antibody CR3022 (PDBid: 6W41).* (2020). doi: 10.5281/zenodo.3779002
18. Soñora M, Barrera E, Garay P, Klein F, Machado M, Pantano S. *SIRAH-CoV2 Initiative: nsp7-nsp8 Complex (PDB id:6YHU).* (2020). doi: 10.5281/zenodo.3803620
19. Soñora M, Barrera E, Garay P, Klein F, Machado M, Pantano S. *SIRAH-CoV2 Initiative: ORF7A Enconded Accessory Protein (PDB id:6W37).* (2020). doi: 10.5281/zenodo.3814373
20. Klein F, Barrera E, Garay P, Machado M, Soñora M, Pantano S. *SIRAH-CoV2 Initiative: Co-factor Complex of NSP7 and the C-Terminal Domain of NSP8 From SARS CoV-2 (PDBid:6WIQ).* (2020). doi: 10.5281/zenodo.3820673
21. Soñora M, Barrera E, Garay P, Klein F, Machado M, Pantano S. *SIRAH-CoV2 Initiative: RNA-Dependent RNA Polymerase in Complex With Cofactors Nsp7 and Nsp8 (PDB id:7BTF).* (2020). doi: 10.5281/zenodo.3865479
22. Klein F, Barrera E, Garay P, Machado M, Soñora M, Pantano S. *SIRAH-CoV2 Initiative: Updated Trajectory of SARS-Cov2 Spike's RBD/ACE2-B0AT1 Complex (PDB id:6M17).* (2020). doi: 10.5281/zenodo.3942566
23. Garay P, Barrera E, Klein F, Machado M, Soñora M, Pantano S. *SIRAH-CoV2 Initiative: Glycosylated RBD.* (2020). doi: 10.5281/zenodo.3903522
24. Florencia K, Barrera E, Garay P, Machado M, Soñora M, Pantano S. *SIRAH-CoV2 Initiative: S2 Spike Core Fragment in Postfusion State (PDB id:6M1V).* (2020). doi: 10.5281/zenodo.4019350
25. Barrera E, Garay P, Klein F, Machado M, Soñora M, Pantano S. *SIRAH-CoV2 Initiative: Membrane Embedded SARS-CoV-2 ORF3a (PDB id:6XDC).* (2020). doi: 10.5281/zenodo.4044378
26. Garay P, Barrera E, Klein F, Machado M, Soñora M, Pantano S. *SIRAH-CoV2 Initiative: SARS-CoV-2 helicase (PDB id:6ZSL).* (2020). doi: 10.5281/zenodo.4250123
27. Garay P, Barrera E, Klein F, Machado M, Soñora M, Pantano S. *SIRAH-CoV2 Initiative: RBD Triple Glycosylated at Asn331, 343, and 481 From PDB Structure 6XEY*. (2020). doi: 10.5281/zenodo.4277989
28. Garay P, Barrera E, Klein F, Machado M, Soñora M, Pantano S. *SIRAH-CoV2 Initiative: Spike D614G Mutation (Introduced on PDB id:6XR8).* (2020). doi: 10.5281/zenodo.4311516
29. Machado MR, Barrera EE, Klein F, Sóñora M, Silva S, Pantano S. The SIRAH 2.0 force field: altius, fortius, citius. *J Chem Theory Comput.* (2019) 15:2719–33. doi: 10.1021/acs.jctc.9b00006
30. Klein F, Caceres-Rojas D, Carrasco M, Tapia JC, Caballero J, Alzate-Morales JH, et al. Coarse-Grained parameters for divalent cations within the SIRAH force field. *J Chem Inf Model.* (2020) 60:3935–43. doi: 10.1021/acs.jcim.0c00160
31. Garay PG, Machado MR, Verli H, Pantano S. SIRAH late harvest: coarse-grained models for protein glycosylation. *bioRxiv.* (2020). doi: 10.1101/2020.12.18.423446
32. Fiser A, Sali A. ModLoop: automated modeling of loops in protein structures. *Bioinformatics.* (2003) 19:2500–1. doi: 10.1093/bioinformatics/btg362
33. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* (2004) 32:665–7. doi: 10.1093/nar/gkh381
34. Kern DM, Sorum B, Hoel CM, Sridharan S, Remis JP, Toso DB, et al. Cryo-EM structure of the SARS-CoV-2 3a ion channel in lipid nanodiscs. *bioRxiv.* (2020). doi: 10.1101/2020.06.17.156554
35. Barrera EE, Machado MR, Pantano S. Fat SIRAH: coarse-grained phospholipids to explore membrane–protein dynamics. *J Chem Theory Comput.* (2019) 15:5674–88. doi: 10.1021/acs.jctc.9b00435
36. Park S-J, Lee J, Qi Y, Kern NR, Lee HS, Jo S, et al. CHARMM-GUI Glycan Modeler for modeling and simulation of carbohydrates and glycoconjugates. *Glycobiology.* (2019) 29:320–31. doi: 10.1093/glycob/cwz003
37. Machado MR, Pantano S. SIRAH tools: mapping, backmapping and visualization of coarse-grained models. *Bioinformatics.* (2016) 32:1568–70. doi: 10.1093/bioinformatics/btw020
38. Darré L, Machado MR, Dans PD, Herrera FE, Pantano S. Another coarse grain model for aqueous solvation: WAT FOUR? *J Chem Theory Comput.* (2010) 6:3793–807. doi: 10.1021/ct100379f
39. Machado MR, Pantano S. Split the charge difference in two! A rule of thumb for adding proper amounts of ions in MD simulations. *J Chem Theory Comput.* (2020) 16:1367–72. doi: 10.1021/acs.jctc.9b00953
40. Darden T, York D, Pedersen L. Particle mesh Ewald: an N log(N) method for Ewald sums in large systems. *J Chem Phys.* (1993) 98:10089–92. doi: 10.1063/1.464397
41. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *J Chem Phys.* (1995) 103:8577–93. doi: 10.1063/1.470117
42. Wu X, Brooks BR. Self-guided Langevin dynamics simulation method. *Chem Phys Lett.* (2003) 381:512–8. doi: 10.1016/j.cplett.2003.10.013
43. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys.* (1984) 81:3684–90. doi: 10.1063/1.448118
44. Salomon-Ferrer R, Case DA, Walker RC. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip Rev Comput Mol Sci.* (2013) 3:198–210. doi: 10.1002/wcms.1121

45. Komatsu T, Koyama Y, Okimoto N, Morimoto G, Ohno Y, Taiji M. COVID-19 related trajectory data of 10 microseconds all atom molecular dynamics simulation of SARS-CoV-2 dimeric main protease. *Mendeley Data.* (2020). doi: 10.17632/vpps4vhryg.2
46. Varki A, Cummings RD, Aebi M, Packer NH, Seeberger PH, Esko JD, et al. Symbol nomenclature for graphical representations of Glycans. *Glycobiology.* (2015) 25:1323–4. doi: 10.1093/glycob/cwv091
47. Liu L, Wang P, Nair MS, Yu J, Rapp M, Wang Q, et al. Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature.* (2020) 584:450–6. doi: 10.1038/s41586-020-2571-7
48. Zhao P, Praissman JL, Grant OC, Cai Y, Xiao T, Rosenbalm KE, et al. Virus-receptor interactions of glycosylated SARS-CoV-2 spike and human ACE2 receptor. *Cell Host Microbe.* (2020) 28:586–601.e6. doi: 10.1016/j.chom.2020.08.004
49. Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature.* (2020) 581:221–4. doi: 10.1038/s41586-020-2179-y

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.