Bioinformatics, 37(22), 2021, 4258–4260
doi: 10.1093/bioinformatics/btab397
Advance Access Publication Date: 20 May 2021
Applications Note

Structural bioinformatics

3DBionotes COVID-19 edition

Jose Ramon Macias [1,*], Ruben Sanchez-Garcia [1], Pablo Conesa[1], Erney Ramirez-Aportela[1], Marta Martinez Gonzalez[1], Carlos Wert-Carvajal [1], Alberto M. Parra-Perez[1], Joan Segura Mora[2], Sam Horrell[3], Andrea Thorn[4], Carlos O.S. Sorzano [1] and Jose Maria Carazo[1,*]

[1]Spanish National Bioinformatics Institute (INB ELIXIR-ES), Biocomputing Unit, National Centre of Biotechnology (CNB-CSIC), Instruct Image Processing Centre, Campus de Cantoblanco, 28049 Madrid, Spain, [2]Research Collaboratory for Structural Bioinformatics Protein Data Bank, San Diego Supercomputer Center, University of California, San Diego, La Jolla, CA, USA, [3]Diamond Light Source Ltd. (DLS), Oxfordshire, UK and [4]Institute for Nanostructure and Solid State Physics, HARBOR, Universität Hamburg, Hamburg, Germany

*To whom correspondence should be addressed.

Associate Editor: Yann Ponty
Received on November 23, 2020; revised on March 26, 2021; editorial decision on May 11, 2021; accepted on May 19, 2021

Abstract

Summary: The web platform 3DBionotes-WS integrates multiple web services and an interactive web viewer to provide a unified environment in which biological annotations can be analyzed in their structural context. Since the COVID-19 outbreak, new structural data from many viral proteins have been provided at a very fast pace. This effort includes many cryogenic electron microscopy (cryo-EM) studies, together with more traditional ones (X-rays, NMR), using several modeling approaches and complemented with structural predictions. At the same time, a plethora of new genomics and interactomics information (including fragment screening and structure-based virtual screening efforts) have been made available from different servers. In this context, we have developed 3DBionotes-COVID-19 as an answer to: (i) the need to explore multiomics data in a unified context with a special focus on structural information and (ii) the drive to incorporate quality measurements, especially in the form of advanced validation metrics for cryo-EM.

Availability and implementation: https://3dbionotes.cnb.csic.es/ws/covid19.

Contact: jr.macias@cnb.csic.es

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

The 3DBionotes-WS web platform has been operational for several years as part of the online offer of the Spanish Institute of Bioinformatics, the Spanish Node of the Research Infrastructure (RI) ELIXIR and Instruct-ERIC RI (Segura *et al.*, 2019). It is, in fact, one of ELIXIR Recommended Interoperability Resources (https://elixir-europe.org/platforms/interoperability/rirs). A major goal is to provide an interactive graphical environment over the web where structural and multiomics data can be intuitively explored, complemented by a powerful API.

The COVID-19 outbreak has changed science: worldwide, scientists came together across disciplines and national borders in order to fight the pandemic. Structural biologists are no exception and the role of cryogenic electron microscopy (cryo-EM) in elucidating key viral structures is paramount [for a brief outline, see (Kearns, 2020)], complementing more traditional approaches, such as X-ray crystallography, NMR and fold predictions. However, SARS-CoV-2 maps did not achieve very high resolution (in most cases close to 3 Å, particularly for cryo-EM), which made it difficult to build atomic models suitable for drug development. Additionally, the pressure to publish these structures as fast as possible has never been so high. Early in the pandemic, specific resources have been created to address structural needs (https://github.com/thorn-lab/coronavirus_structural_task_force), acknowledging the requirement to pay special attention not only to data quantity, but also to data quality. In this way, validation information on cryo-EM maps provided by the Coronavirus Structural Task Force has been integrated into 3DBionotes, which have evolved to supply quality measurements, keeping its orientation toward the web (and its API) and focusing on integrative analysis. Indeed, its COVID-19 edition described here, combines in the same analysis framework key viral genomics, interactomics and structural information, including drug screening approaches [both experimental fragment-based screening (Douangamath *et al.*, 2020) and virtual screening]. In the following, we describe 3DBionotes-COVID-19 edition, illustrating its use and value for users with some case studies in Supplementary Material.

2 Results

The design of 3DBionotes COVID-19 adds an additional layer of interactive information over the classical design of 3DBionotes. The new edition has a specific landing page to manage multiple sources of structural information that, once selected, launches a new version of 3DBionotes accessing COVID-19 specific information (Fig. 1). We describe this design in the following:

2.1 Landing page

This page acts as a structural information organizer, collecting data from SARS-CoV-2 and related coronavirus, as well as their interactions with host proteins. We automatically harvest structures deposited in PDB and EMDB together with predicted models from SWISSModel (Bienert *et al.*, 2017), AlphaFold (Jumper *et al.*, 2020) and BSM-Arc (Hijikata *et al.*, 2020). When validation and quality information are available from PDB-REDO (Joosten *et al.*, 2014) and the Coronavirus Structural Task Force (Croll *et al.*, 2021) special tags are incorporated for every entry, pointing to the rerefined models. Entries are organized into five categories: PDB, EMDB, Interactions with other proteins (PPI) and Ligands, Related (to SARS-CoV or others) and Computational Models. In 'Ligands', we initially incorporate experimental information from fragment-based screening (Douangamath *et al.*, 2020) as well as our own structure-based repurposing virtual screening (https://covid19drugrepurposing.cnb.csic.es).

Every entry is displayed with its reference and a static view of the model, when available, that serves the user as a preliminary visual hint of the structure (Fig. 1A). A pop-up panel is displayed with a brief description of the entry and a set of external links when the pointer is placed on the image for a few moments. Upon clicking on the entry, data are transferred to a new instance of 3DBionotes that pays special attention to multiomics and cryo-EM quality information, as detailed in the next section, opening the 3D viewer (Fig. 1B) and starting the annotation collection process. Users can go back to the landing page at any time in their analysis.

2.2 New annotations

3DBionotes-WS, in general, collects and organizes a wide range of annotations of the selected macromolecule. The COVID-19 release includes access to a collection of specifically developed servers adding new functionality, as it can be appreciated in the case studies detailed in Supplemental Material. Among them, we highlight:

- Cryo-EM quality information at the amino acid level coloring the map being displayed in the 3D viewer, including:
- Local resolution information on cryo-EM maps, calculated using a deep learning approach which does not require half maps (Ramírez-Aportela *et al.*, 2019).
- Quantitative validation metrics, such as Q-scores (Pintilie *et al.*, 2020) and FSC-Q-scores (Ramírez-Aportela *et al.*, 2021).
- SARS-CoV-2 main protease fragment screening by PanDDA analysis (Pearce *et al.*, 2017).
- Genomics variants, with source data from CNCB (https://bigd.big.ac.cn/ncov/variation).
- Functional mapping of Protein–Protein Interactions (PPI), with source data from Korbin's lab (http://draco.cs.wpi.edu/wuhan) (Srinivasan *et al.*, 2020).

2.3 Selected case studies

In Supplementary Material, we present how 3DBionotes can be used in four different use cases, namely:

1. SARS-CoV-2 spike protein cryo-EM map validation analysis based on local resolution metrics.
2. Use of improved structural models using new refinement methods collected from the Coronavirus Structural Task Force.
3. Analysis of SARS-CoV-2 spike variant D614G.
4. Study of drug screening on SARS-CoV-2 main protease (NSP5).

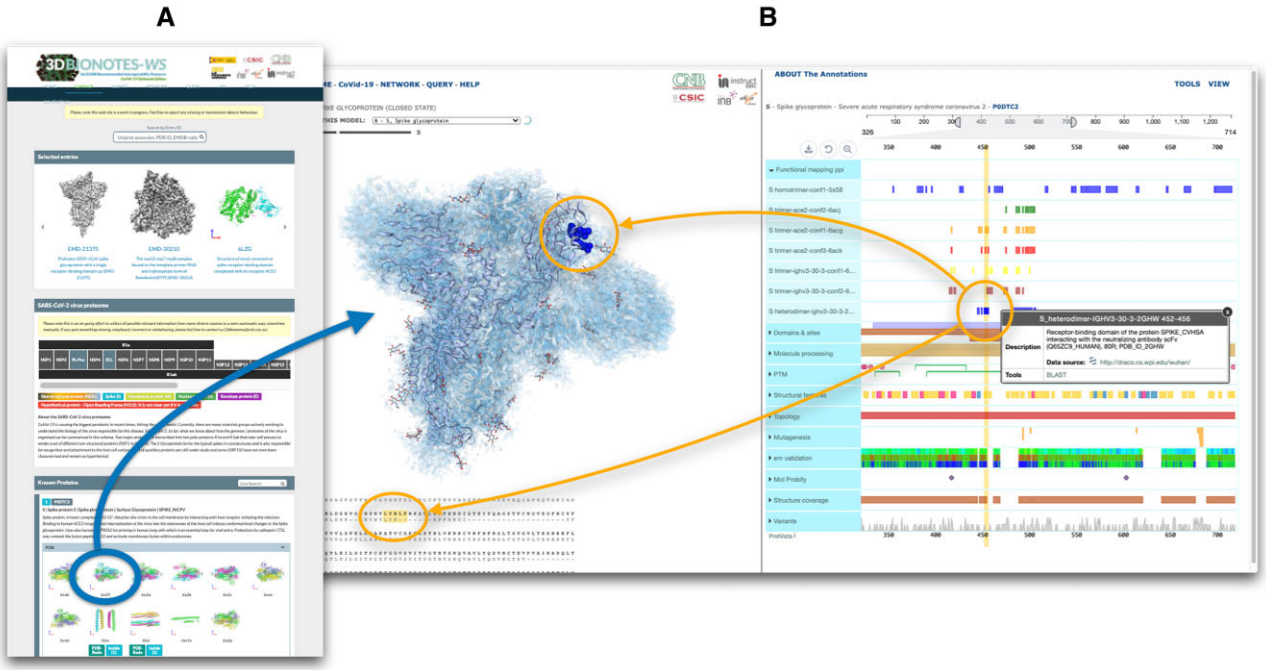


Fig. 1. 3DBionotes COVID-19 application screenshots. (**A**) Landing page, showing some of the main sections: representative examples, a simplified schema of the virus proteome that serves as index with links to the corresponding subsection for every protein, followed by various panels with the structures. (**B**) 3D viewer and annotations, showing the example of EMD-21452, corresponding to SARS-CoV-2 spike glycoprotein (closed state). By clicking in any of the symbols representing an annotation, all the residues associated with it will be highlighted in the protein sequence alignment as well as in the atomic structure. At the same time, those residues will also be highlighted with vertical yellow bars so it is easier to locate in relation with other annotations types. Additionally, a panel will pop-up with more detailed information about the annotation, including links to the origin of the data source

3 Conclusions

3DBionotes-COVID-19 is fully accessible at https://3dbionotes.cnb.csic.es/ws/covid19, providing a unique analysis environment tailored to COVID-19 information. It has all the advantages of 3DBionotes in terms of complex interactive analysis over the web and API access, offering both the possibility to work with structural data already deposited in public databases and with new user data, plus a series of new services geared toward quality (cryo-EM validation and curated structural models) and information integration. We demonstrate the usefulness of this interactive resource on four selected cases.

Funding

We acknowledge financial support from CSIC [PIE/COVID-19 number 202020E079], the Comunidad de Madrid through grant CAM [S2017/BMD-3817], the Spanish Ministry of Science and Innovation through projects [SEV 2017-0712, FPU-2015/264, PID2019-104757RB-I00/AEI/10.13039/501100011033], the Instituto de Salud Carlos III: PT17/0009/0010 [ISCIII-SGEFI/ERDF] and the European Union and Horizon 2020 through grant: CORBEL [INFRADEV-01-2014-1, Proposal 654248] and EOSC Life [INFRAEOSC-04-2018, Proposal: 824087]. This work was supported by Instruct-ULTRA [Grant 731005], an EU H2020 project to further develop the services of Instruct-ERIC. Contributions from the Coronavirus Structural Task Force were supported by the German Federal Ministry of Education and Research [Grant 05K19WWA] and Deutsche Forschungsgemeinschaft [project TH2135/2-1]. The authors acknowledge the support and the use of resources of Instruct, a Landmark ESFRI project.

Conflict of Interest: none declared.

References

Bienert,S. *et al.* (2017) The SWISS-MODEL repository-new features and functionality. *Nucleic Acids Res*., **45**, D313–D319.

Croll,T.I. *et al.* (2021) Making the invisible enemy visible. Nature Structural & Molecular Biology, **28**, 404–408. doi:10.1038/s41594-021-00593-7

Douangamath,A. *et al.* (2020) Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat. Commun*., **11**, 5047.

Hijikata,A. *et al.* (2020) Knowledge-based structural models of SARS-CoV-2 proteins and their complexes with potential drugs. *FEBS Lett*., **594**, 1960–1973.

Joosten,R.P. *et al.* (2014) The PDB_REDO server for macromolecular structure model optimization. *IUCrJ*, **1**, 213–220.

Jumper,J. *et al.* (2020) Computational predictions of protein structures associated with COVID-19. Version 3, DeepMind website. https://deepmind.com/research/open-source/computational-predictions-of-protein-structures-associated-with-COVID-19.

Kearns,S. (2020) Structure of the pandemic. *Structure*, **28**, 874–878.

Pearce,N.M. *et al.* (2017) A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nat. Commun*., **8**, 15123.

Pintilie,G. *et al.* (2020) Measurement of atom resolvability in cryo-EM maps with Q-scores. *Nat. Methods*, **17**, 328–334.

Ramírez-Aportela,E. *et al.* (2019) DeepRes: a new deep-learning- and aspect-based local resolution method for electron-microscopy maps. *IUCrJ*, **6**, 1054–1063.

Ramírez-Aportela,E. *et al.* (2021) *FSC-Q: A CryoEM Map-to-Atomic Model Quality Validation Based on the Local Fourier Shell Correlation*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Segura,J. *et al.* (2019) 3DBIONOTES v3.0: crossing molecular and structural biology data with genomic variations. *Bioinformatics*, **35**, 3512–3513.

Srinivasan,S. *et al.* (2020) Structural genomics of SARS-CoV-2 indicates evolutionary conserved functional regions of viral proteins. *Viruses*, **12**, 360.

Title: 3DBionotes COVID-19 Edition (Supplementary Material)

Authors:

Jose Ramon Macias (1), Ruben Sanchez-Garcia (1), Pablo Conesa (1), Erney Ramírez-Aportela (1), Marta Martinez Gonzalez (1), Carlos Wert-Carvajal (1), Alberto M. Parra-Perez (1), Joan Segura Mora (2), Sam Horrell (3), Andrea Thorn (4), Carlos O.S. Sorzano (1), Jose Maria Carazo (1)

1 - Spanish National Bioinformatics Institute (INB ELIXIR-ES). Biocomputing Unit, National Centre of Biotechnology (CNB-CSIC). Instruct Image Processing Centre

2 - Research Collaboratory for Structural Bioinformatics Protein Data Bank. San Diego Supercomputer Center, University of California, San Diego, La Jolla

3 - Diamond Light Source Ltd. (DLS), Oxfordshire, UK

4 - Institute for Nanostructure and Solid State Physics, HARBOR, Universität Hamburg, Germany

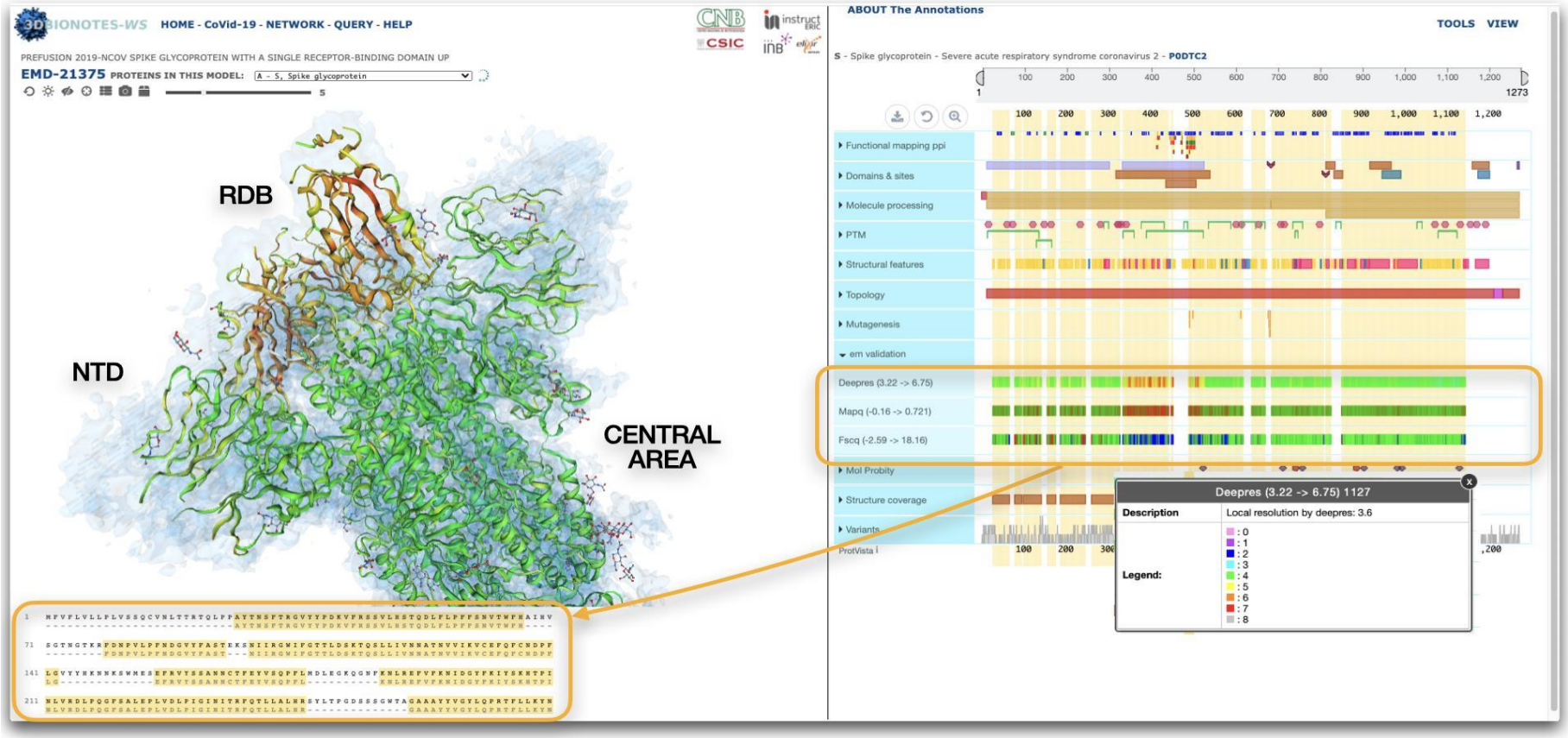
Use Cases:

(1) SARS-CoV-2 spike protein cryo-EM map validation analysis based on local resolution metrics.
(2) Use of improved structural models using new refinement methods collected from the Coronavirus Structural Task Force.
(3) Analysis of SARS-CoV-2 spike variant D614G.
(4) Study of drug screening on SARS-CoV-2 main protease (NSP5).

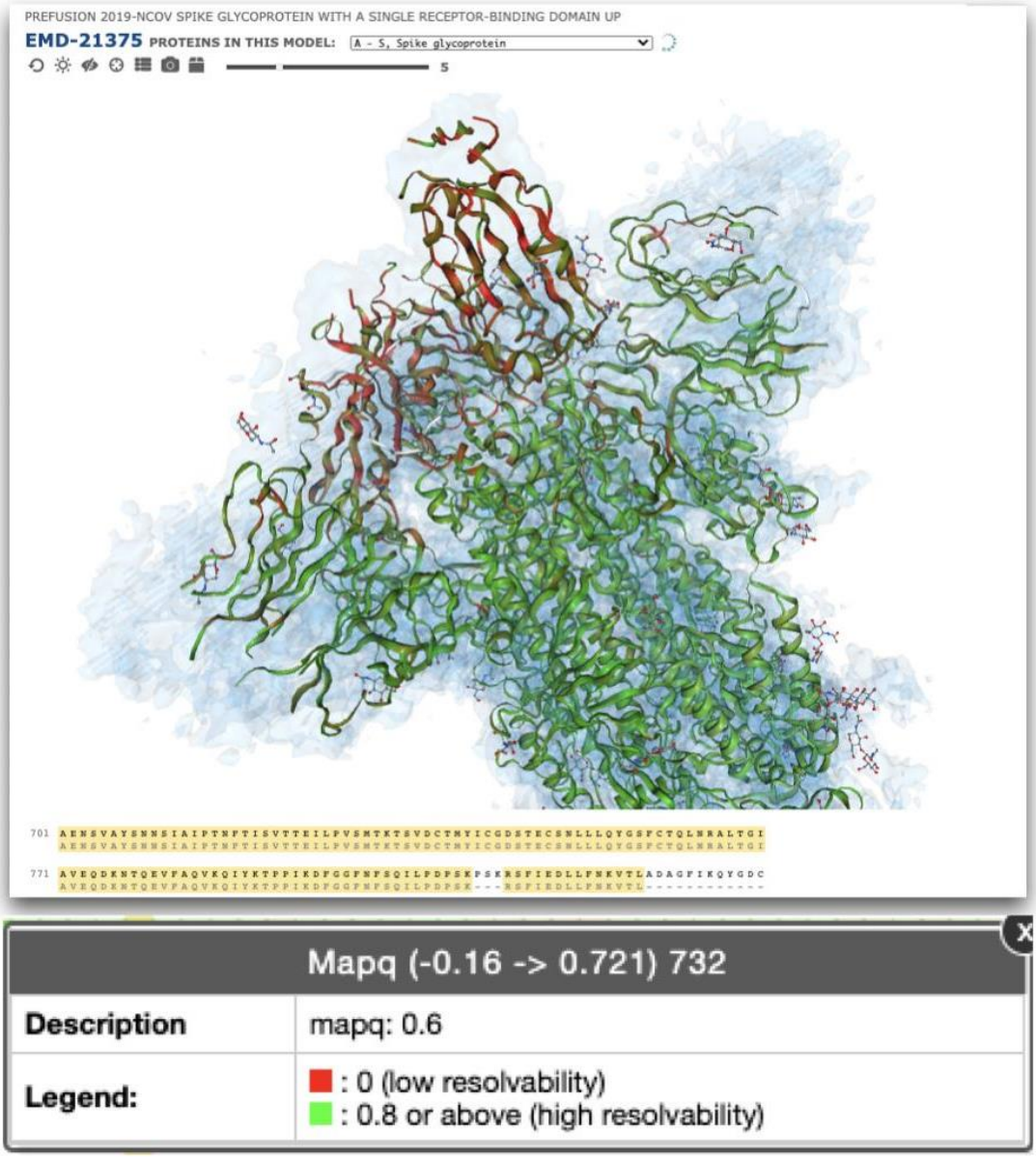
1) SARS-CoV-2 spike protein cryo-EM map validation analysis based on local resolution metrics

The current edition of 3DBionotes-WS incorporates new sets of residue annotations to inform about the quality of cryo-EM maps and the fit of atomic models. Assessment values of map quality were obtained using deepRes (Ramírez-Aportela et al., 2019), a method that computes map local resolution, while the fit and resolvability of the built atomic model were studied with FSC-Q (Ramírez-Aportela et al., 2021) and Q-score (Pintilie et al., 2020). Scores provided by these tools advise users, from a local point of view, about how well the structural model is supported by the experimental data.

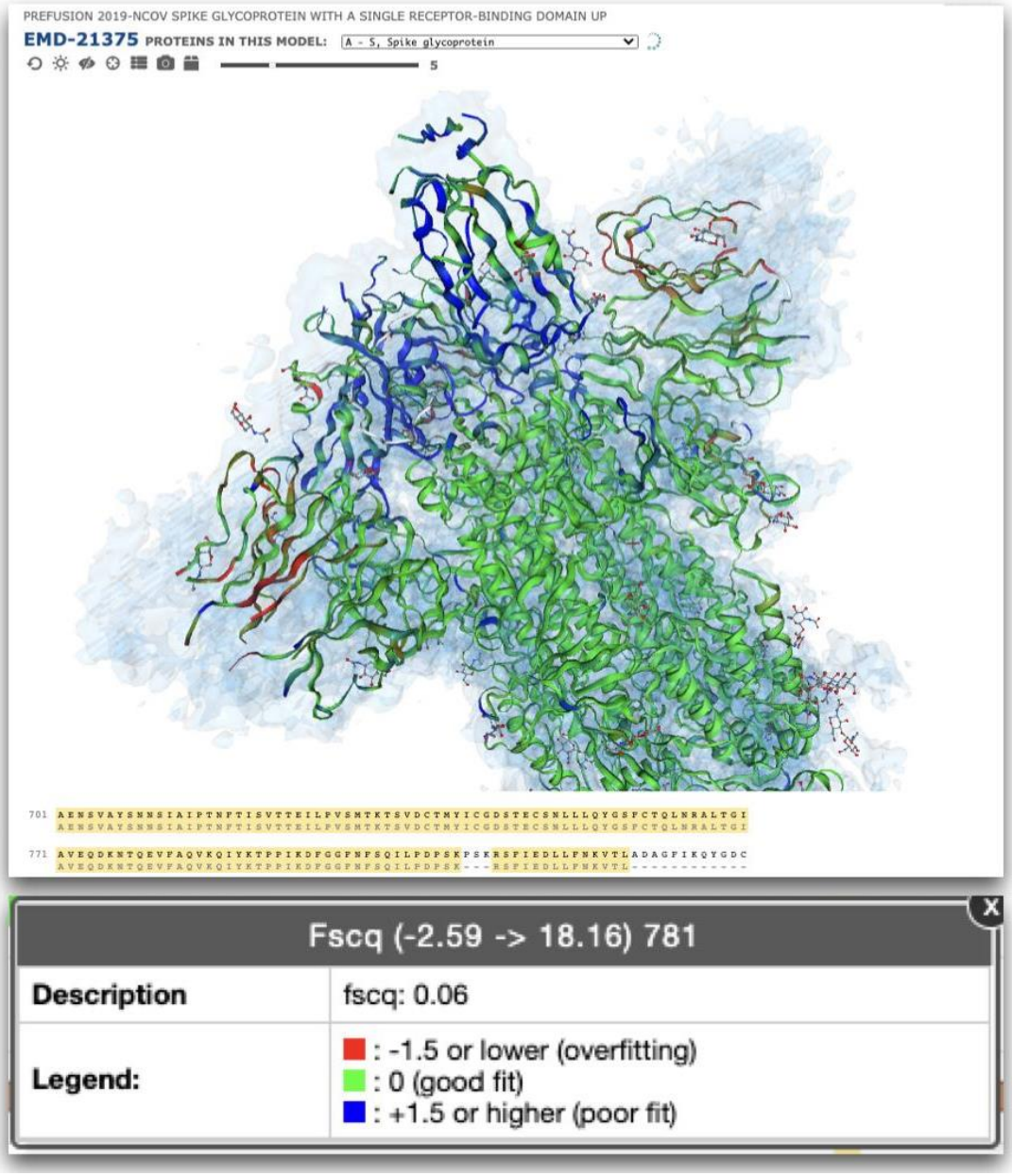
Local resolution of the SARS-CoV-2 prefusion spike glycoprotein map (EMD-21375) appears detailed per modeled residue of the atomic structure (PDB ID: 6VSB) in Supp. Fig. 1A. This figure shows that the central area of the spike has better resolution (~ 3.5 - 4.0 Å) than the receptor-binding domain (RBD) (~ 5.0 - 7.0 Å) and the N-terminal domain (NTD) (~ 4.5 - 5.0 Å). This difference is probably due to the better image alignment achieved in the central area during the map reconstruction process, compared to RBD and NTD regions. Since the two latter regions are flexible (for a detailed analysis see (Melero *et al.*, 2020)), their variable positions in the cryo-EM image cause blurring in the resulting map, thus hindering the tracing of several amino acids of the atomic model. Supp. Fig. 1B shows the low resolvability of these areas, while Supp. Fig. 1C details a worse model to map fitting in those continuous flexible domains compared to the central core of the spike.



Supplementary Figure 1A: Cryo-EM map validation analysis. Local resolution calculated for the spike glycoprotein map (EMD-21375) projected on the atoms of the atomic model (PDB ID: 6VSB).



Supplementary Figure 1B: Cryo-EM map validation analysis. Q-score values are represented. Red indicates atoms with low resolvability.



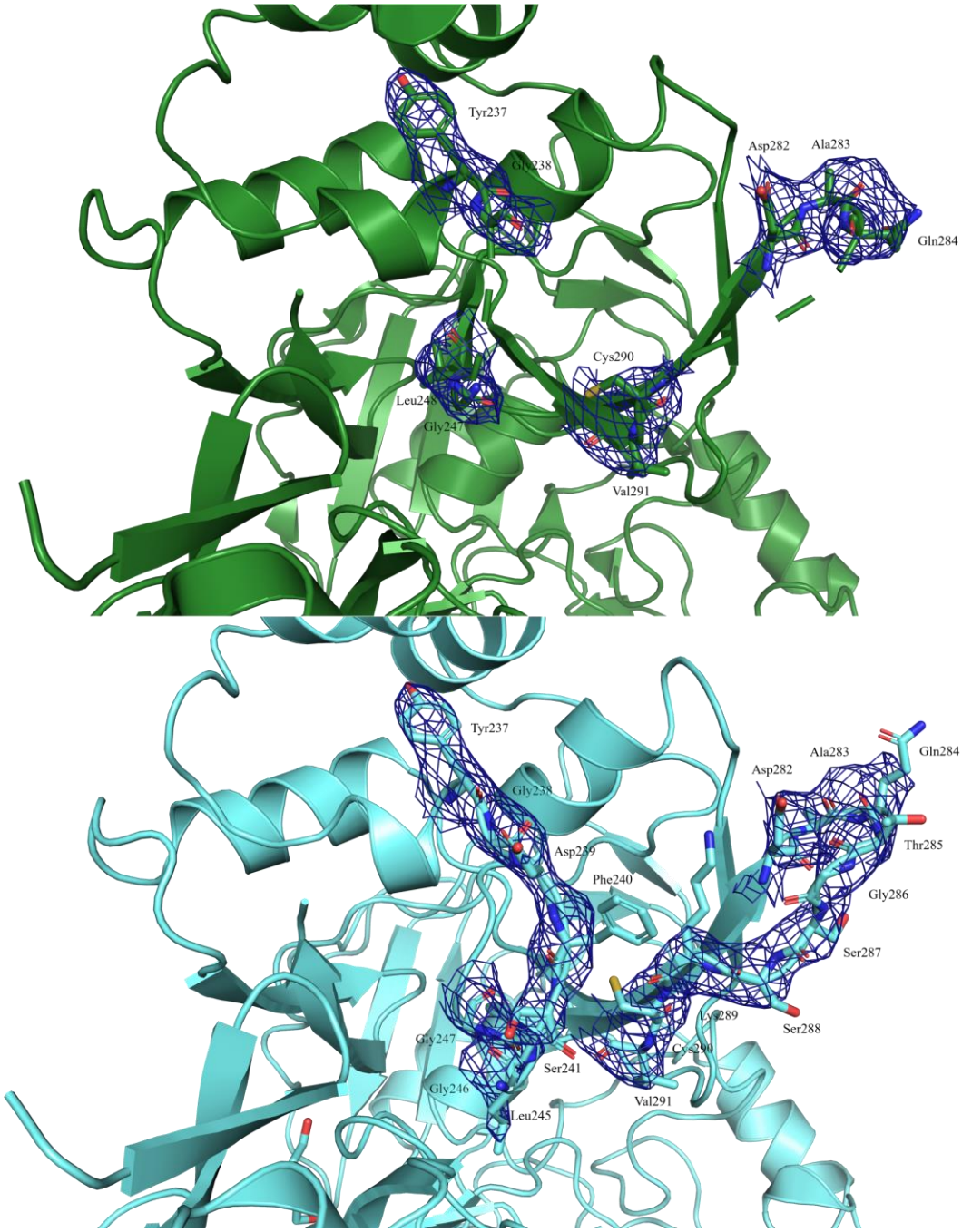
Supplementary Figure 1C: Cryo-EM map validation analysis. FSC-Q values projected on the atomic model. Red indicates atoms that may be associated with noise (possible overfitting) and blue indicates poorly fitted atoms or areas with low resolvability.

2) Use of improved structural models using new refinement methods collected from the Coronavirus Structural Task Force

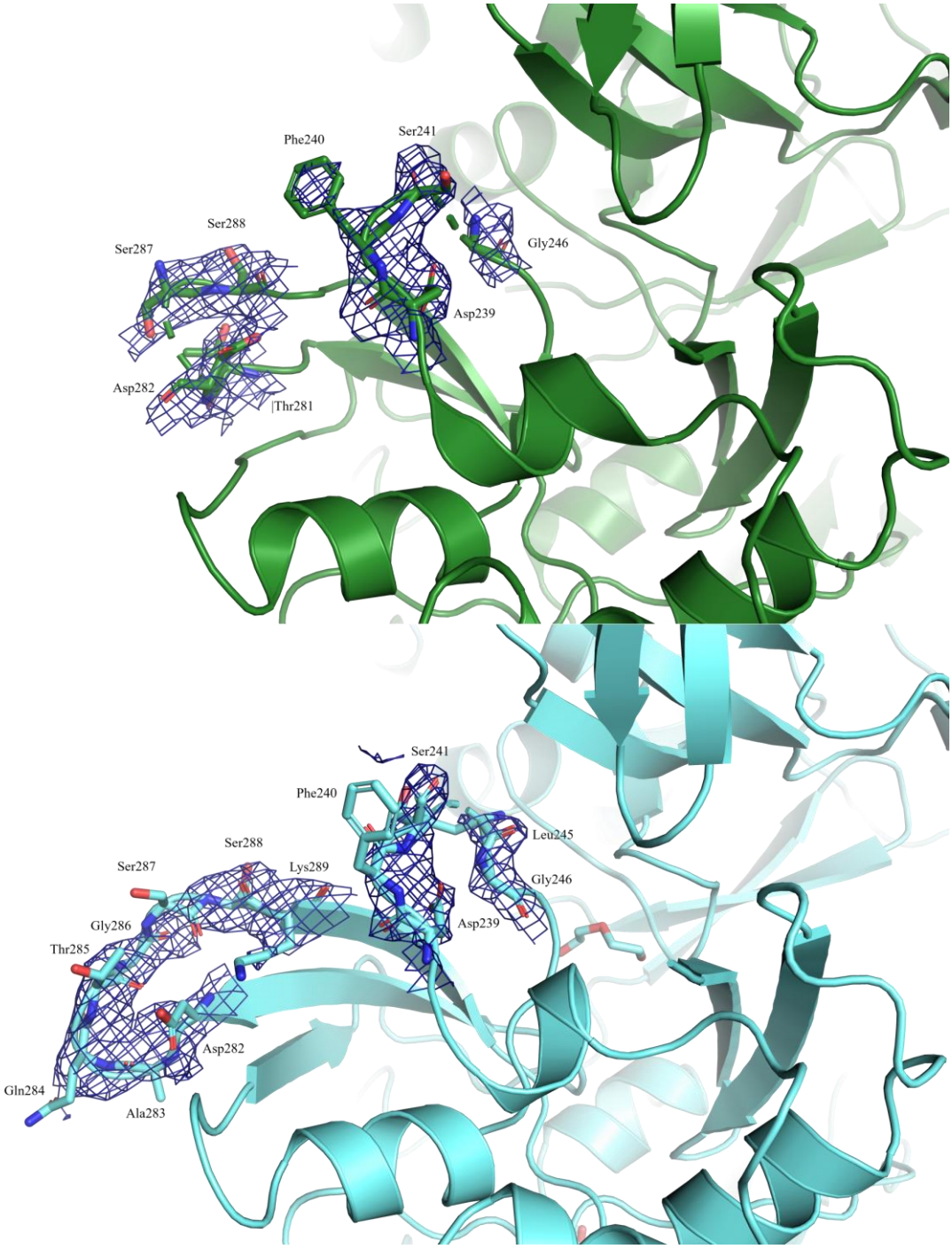
Many downstream users of structural models are not necessarily structural biologists. As a result they may not be familiar with the metrics used to define a well validated structure, or how to pick the best structure for their experiment given multiple structures of the same protein. As work surrounding COVID-19 is of such paramount importance, and hundreds of SARS-CoV-2 structures have now been deposited in the PDB, the Coronavirus Structural Task Force have worked to make selecting a structure simpler for downstream users by curating a repository of available structural models (https://github.com/thorn-lab/coronavirus_structural_task_force). The structures in the repository are updated every Wednesday in the "pdb" folder of the repository, where they are arranged into subfolders by protein, virus name and PDB ID. Within the individual PDB ID folders a "readme.md" file serves as starting point: it summarizes the motivation behind experimentally determining the structure, sequence information, different validation metrics, whether it has been re-processed and/or re-refined by the Task Force manually and other useful links related to the protein, such as potential glycosylation sites, PDB-KB and, of course, 3DBionotes-WS. The "virus name" folder contains a sequence alignment between the NCBI proteome and each PDB deposited for this protein. It also contains a heatmap and/or text file of RMSD between all models from this virus (SARS-CoV or SARS-CoV-2) for this protein. This may help users to pick a suitable PDB from the structures at hand. Within a specific "PDB ID" folder the user will find the original model and reflections, a validation folder with auspex plots, summaries from MolProbity (Williams *et al.*, 2018) and PDB-REDO (Joosten *et al.*, 2014), and in some cases, a folder with a manually reprocessed structure named after the program which was used e.g. REFMAC (Murshudov *et al.*, 2011) or ISOLDE. Any models re-refined by the Task Force will also have a text file detailing which changes were made. These metrics are being incorporated in 3DBionotes-WS, tagging every available entry with a colored badge for every type of validation or refinement. Currently, PDB-Redo, Isolde and RefMac tags are shown, as these produce a new structure in pdb format that is ready to display in 3DBionotes-WS. Other methods providing different types of reports will be incorporated to the annotations, in the right hand side panel. In the following, we illustrate this with an example from the repository, PDB ID 2OZK.

A lot can be learned about the function of proteins in SARS-CoV-2 from related viruses such as SARS-CoV and MERS-CoV, as many proteins from these viruses share common sequences and structures and can be used as an initial guide for simulated drug design efforts. However, data processing software has improved greatly since the 2002 SARS-CoV and the 2012 MERS-CoV outbreaks and older structures can often be improved by reprocessing with the latest software. For example, a structure of the SARS-CoV endoribonuclease (PDB ID: 2OZK) published in 2007 (Joseph *et al.*, 2007) was reprocessed by the Coronavirus Structural

Task Force, and both the model and the fit to the data could be improved: Running 2OZK through the MolProbity server showed the original model to have 169 poor side chain rotamers (16.02%), 28 Ramachandran Outliers (2.37%), and an overall MolProbity score in the 60th percentile when compared to other structures of similar resolution. The model also presents with a R_{work} of 24.3% and R_{free} of 30.0% and a high-resolution limit of 2.90 Å. Following re-refinement in REFMAC5 and model re-building in Coot (Emsley *et al.*, 2010), poor rotamers were reduced to 51 (4.74%), Ramachndran outliers to 16 (1.33%), and the overall MolProbity score improved to the 96th percentile. This means that the model now adheres better to what we know about molecular geometry (bond lengths and angles etc.), and fits the processed X-ray diffraction data better. After re-refinement the final R_{work}/R_{free} was 21.05% and 25.6%, respectively, a drop of roughly 5% compared to the original. In addition to these more general improvements, several missing residues could be added to build missing loops between residues C237-C241, C245-247 and C282-C289 (Supp. Fig. 2A). Missing residues were also present in the same regions of chain D, however only residues D282-D289 were rebuilt in this chain as no new density was observed following re-refinement and addition of 86 water molecules and 4 polyethylene glycol molecules (Supp. Fig. 2B). Overall, the new model is likely much better suited to be used downstream in molecular dynamics, as a template to solve new endoribonuclease structures from SARS-CoV and SARS-CoV-2, and for structure-based drug design efforts.

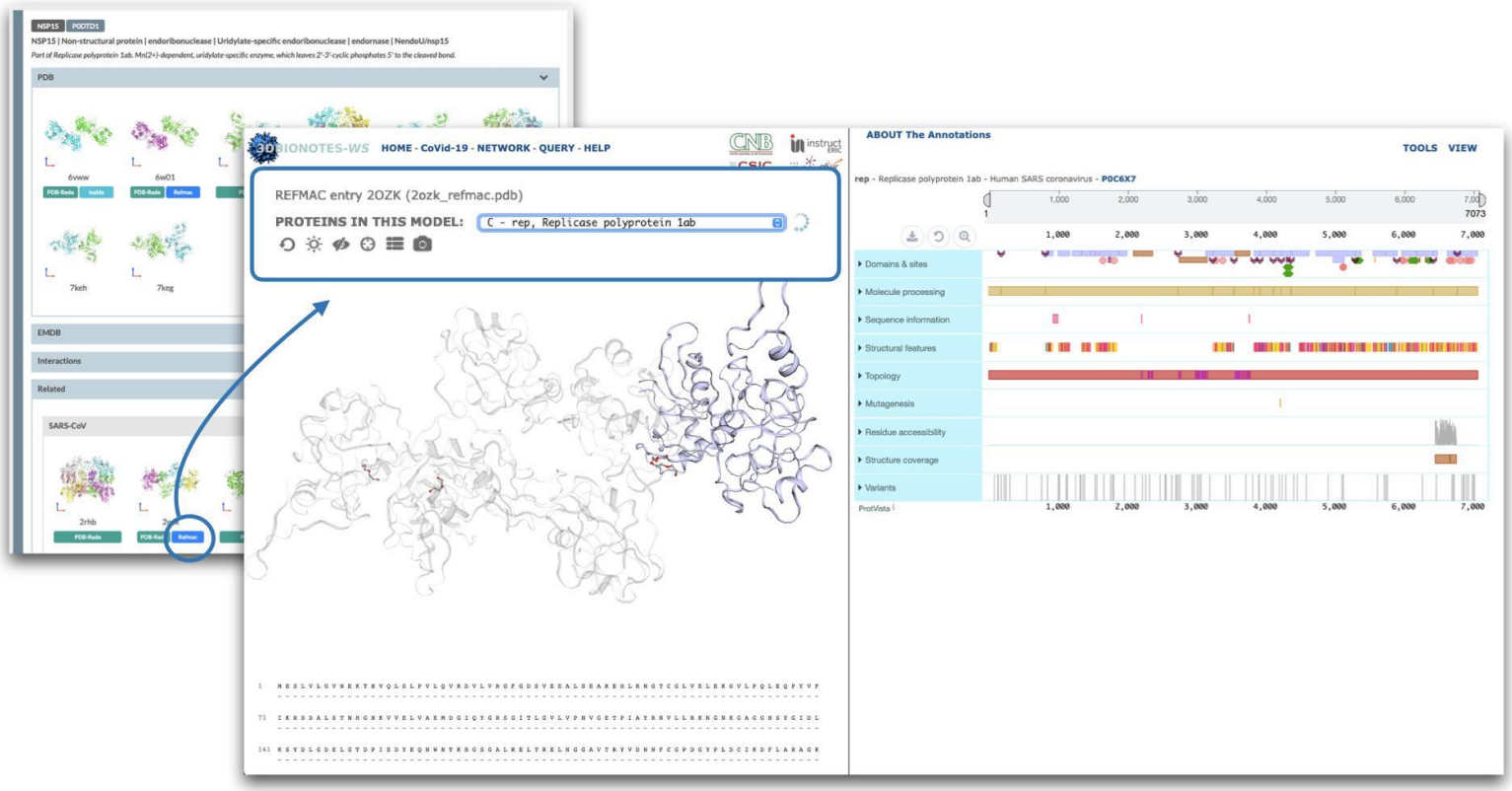


Supplementary Figure 2A: Chain C of 2OZK before (green) and after (blue) reprocessing. The model is represented as a cartoon with dashed lines representing missing residues. The original and rebuilt missing residues (C237-C241, C245-247, and 282-298) are represented as sticks where rebuilt with electron density maps contoured to 1σ.



Supplementary Figure 2B: Chain D of 2OZK before (green) and after (blue) reprocessing. The model is represented as a cartoon with dashed lines representing missing residues. The original and rebuilt missing residues (234-247 and 282-298) are represented as sticks where rebuilt with electron density maps contoured to 1σ.

The re-refined structure of 2OZK can be studied within 3DBionotes-WS clicking in the blue RefMac tag for the corresponding entry in the landing page (Sup. Fig. 2C). Also, it is publicly available in the Coronavirus Structural Task Force Repository: https://github.com/thorn-lab/coronavirus_structural_task_force/tree/master/pdb/endornase/SARS-CoV/2ozk/. In addition to the repository, the Coronavirus Structural Task Force has written literature reviews for each SARS-CoV-2 protein which has had structural data published (NMR, X-ray crystallography or Cryo-EM), background information about the validation pipeline itself, general advice on how to select a structure for downstream use, as well as blog posts on its homepage, www.insidecorona.net.



Supplementary Figure 2C: In the landing page (background) different colored tags are added for every entry with validation and refinement data, currently for PDB-Redo, Isolde and RefMac. When clicking on these tags, the refined model is loaded onto the 3D viewer (front, left). Here the RefMac re-modeling for PDB 2OZK is selected and loaded into the 3D viewer. The functionality of the annotations panel (right, front) remains the same, as this is based on the sequence of the protein (UniProt entry) of the selected chain.

3) Analysis of SARS-CoV-2 Spike variant D614G

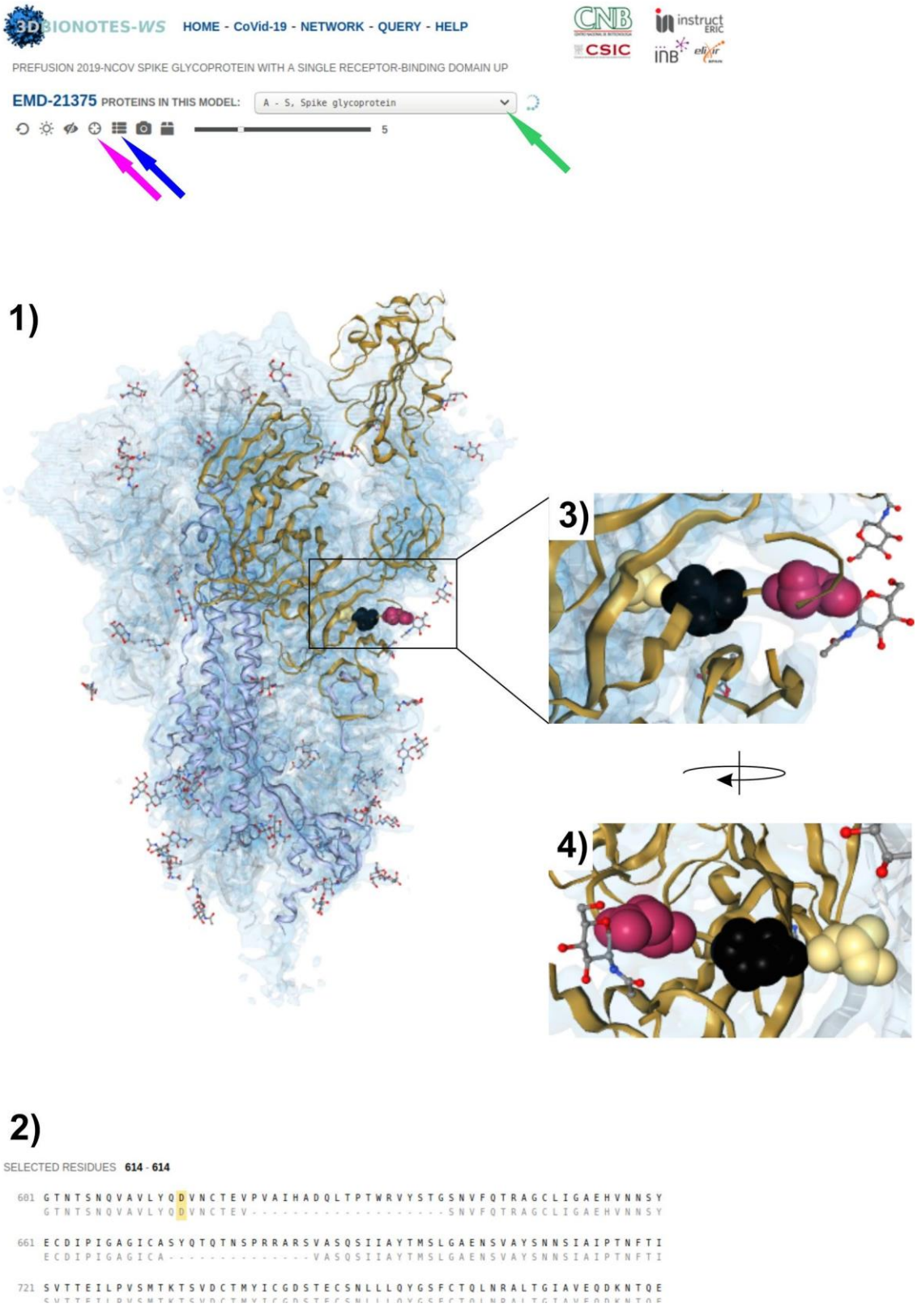
Since 3DBionotes-WS also integrates information of SARS-CoV-2 protein variants from the China National Center for Bioinformation (CNCB), we can take advantage of 3DBionotes-WS web viewer to characterize them structurally. In one of these recently typified variants (Korber et al., 2020), the Spike form G614 has replaced the form D614 originally identified in China, becoming dominant in Europe and later in North-America, Oceania and Asia. Due to its prevalence in several geographic areas, this single amino acid change has been hypothesized to confer a selective advantage. Despite the experimental evidence associating this variant with higher levels of viral RNA in COVID-19 patients and with greater pseudovirus titers from in vitro experiments, thus suggesting more infectivity, no significant correlation has been observed between the D614G mutation and disease severity. Although there exist many open questions to clarify the functional role and the relevance of this mutation in transmissibility and evolution of the pandemic (Korber et al., 2020), we selected this variant as an example to illustrate the potential of 3DBionotes-WS to inspect the structural significance of a single amino acid replacement.

The set of protein variants can be displayed by clicking the right bottom track “Variants" of the 3DBionotes-WS annotations panel, allowing the user to select each individual variant by a mouse click. In this particular example, the mutation of the Aspartic acid 614 by Glycine. A small pop up window (Sup. Fig. 3B-5) details the position and the amino acid replacement of the selected variant. At the same time, a vertical yellow bar intersects the graphical displays of the different features. One of them details the phenotypic effect of this mutation (Sup. Fig. 3B-4). The original un-mutated residue appears depicted in black (same color of the annotation figure) in the 3D viewer on the left side panel (Sup. Fig. 3A-1) and highlighted in the protein sequence alignment at the page bottom, near one of the un-interpreted parts of the map (Sup. Fig. 3A-2). Considering the two units of the structure, S1 (golden) and S2 (blue), the Asp 614 residue is part of the S1 unit. This fact can be corroborated at the pop up window shown by clicking the N-terminal fragment generated by the molecule processing (Sup. Fig. 3B-2). In the selected structure, chain A in PDB ID 6VSB, S1 unit expands from amino acid 13 to 685, thus including the variant position 614. Looking in detail at its neighboring residues (zoomed in Sup. Fig. 3A-3 and 3A-4), we can detect one of them (golden) in direct contact with amino acid 614, and another one (purple), slightly separated. These two residues are particularly relevant due to the potential structure implications of D614G replacement, which might alter its interaction with proximal residues.

The first adjoining residue stands for T859 of the neighboring protomer B, connected by a hydrogen bond with the residue D614 of protomer A. This intra-molecule interaction can be anticipated at the top right panel, which shows the functional mapping annotations covering

protein-protein interactions (PPI). The pop up window (Sup. Fig. 3B-1) details three amino acids depicted in blue (614-616), thus containing the mutation 614, involved in Spike homotrimer interactions among protomers A, B and C. The D614G replacement suppresses the hydrogen bond between adjacent protomers, thus weakening the interaction between S1 and S2 units, which reduces the Spike prefusion state stability (Yurkovetskiy et al., 2020). At the same time, the shortening of the new hydrogen bond between G614 and A647 within the same protomer contributes to strengthening each individual protomer. All these structural changes seem to increase the flexibility of RBDs and, then, the probability of having RBDs with open conformation, which might confer higher infectivity to D614G mutant viruses. In addition, since the T859 amino acid is quite close to its protomer fusion peptide (816-855), the D614G mutation could also affect the dynamics of this relevant structural element.

The second neighboring amino acid, separated by two residues, is N616, which is part of one of the attachment sites of N-glycan chains. The proximity between residues 614 and 616 is observed in the 3DBionotes-WS right panel since the yellow line is quite close to one glycosylation site annotated as post-translational modification (PTM). Details and further references about this polysaccharide linking site appear in the respective pop up window (Sup. Fig. 3B-3). The mutation D614G could somehow disturb the nearby glycosylation site.



Supplementary Figure 3A: 3D viewer for chain A of the atomic structure 6VSB, derived from the EMD-21375 map. **1.** Atomic structure inserted in the density map. The chain A region S1 is shown in gold whereas the region S2 appears in blue. **2.** Sequence alignment of the chain A of the SARS-CoV-2 spike atomic structure and the UniProt ID P0DTC2. **3.** Picture zoom detailing residues D614 (black)

and N616 (purple) from protomer A, and T859 (golden) from protomer B. **4.** Previous image rotated 180°.

Supplementary Figure 3B: Feature annotations section in 3DBionotes-WS. **1.** Blue box to annotate three residues (D614-N616) involved in direct interaction among spike protomers according to the mapping of SARS inter-protomer interacting residues in the sequence of SARS-CoV-2. **2.** Golden box standing for the spike region S1. **3.** Residue N616 annotated as N-glycosylation site. **4.** Yellow box to indicate the phenotype described for D614G mutation. **5.** Genomic variant D614G annotation.

Guide to color and highlight residues and sequences (Sup. Fig. 3A-1):

1. Before starting, ensure that chain A of the Spike glycoprotein is selected in the section “Proteins in this model” (green arrow in Sup. Fig. 3A).
2. Select the variant D614G in the “Variants” track (Sup. Fig. 3B-5). The un-mutated residue should appear in black in Sup. Fig. 3A-1.
3. Click on the “Label selected annotation” icon (magenta arrow in Sup. Fig. 3A).
4. Select the Glycosylation modification residue 616 in the “PTM” track (Sup. Fig. 3B-3). The glycosylated amino acid N should appear in pink in Sup. Fig. 3A-1.
5. Click again on the “Label selected annotation” icon (magenta arrow in Sup. Fig. 3A).
6. Select the chain B of the Spike glycoprotein in “Proteins in this model” section (green arrow in Sup. Fig. 3A).
7. Select the residue T859 in the sequence bottom left panel (Sup. Fig. 3A-2). This amino acid should appear in golden color in Sup. Fig. 3A-1.
8. Click again on the “Label selected annotation” icon (magenta arrow in Sup. Fig. 3A).
9. Select again the chain A of the Spike glycoprotein in “Proteins in this model” section (green arrow in Sup. Fig. 3A).
10. Hide text annotations in “Display selected annotations” icon (blue arrow of Sup. Fig. 3A). The pop up window “Current Labels” will open. Then click in the first box of each label to hide text annotations. Only the second respective boxes should remain selected.
11. Finally, select the first fragment of sequence (before the first cleavage site that splits S1 from S2 region) in the “Molecule processing” track (Sup. Fig. 3B-2). The selected S1 region should appear in golden color in Sup. Fig. 3A-1.

4) Study of drug screening on SARS-CoV-2 main protease (NSP5)

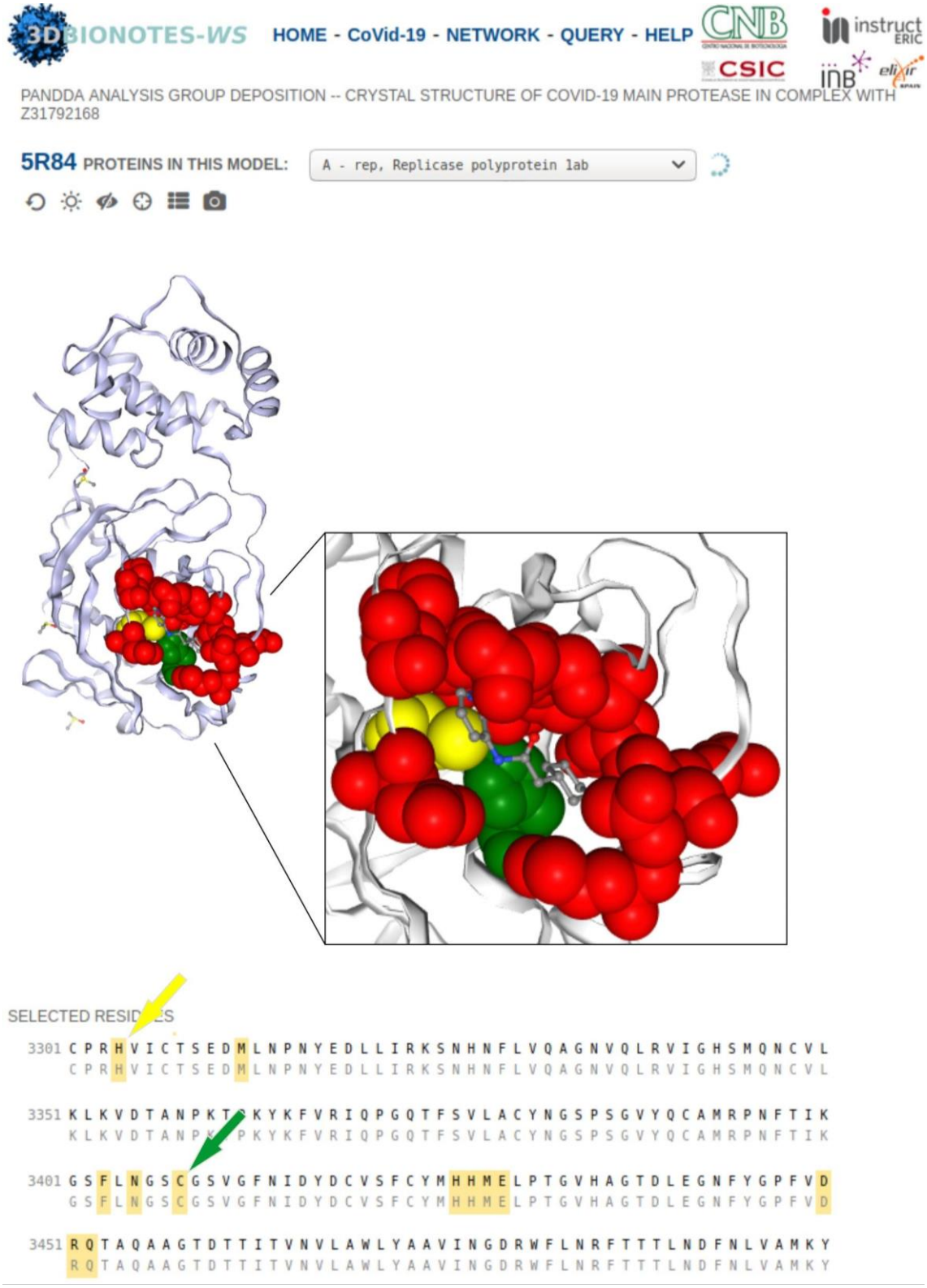
Given the urgent need for antivirals to treat COVID-19 patients, at least while a reliable vaccine is available, different strategies to develop effective therapeutic drugs are being explored in the worldwide ongoing efforts. One of them relies on the SARS-CoV-2 main (3C-like) proteinase, a relevant drug target candidate due to its indispensable role in the virus polyprotein cleavage, since it is the starting point in viral replication, and to the absence of human homologous proteins. The search for an effective inhibitor of this small viral enzyme (35 KD) has thus resulted in close to 300 PDB entries (22-March-2021) since the beginning of the COVID-19 outbreak, 85% of them displaying the structure of the proteinase complexed with a ligand inhibitor. The XChem facility of Diamond Light Source beamline I04-1 (https://www.diamond.ac.uk/Instruments/Mx/Fragment-Screening.html) has contributed with 120 structures obtained by X-ray crystallography using different fragment libraries and the PanDDA analysis method (Pearce et al., 2017).

Since one of the upgrades in the COVID-19 edition of 3DBionotes-WS involves the display in the landing page of all known SARS-CoV-2 proteins, we can take advantage of this improvement to study the PDB entries containing the main protease complexed with different ligands from Diamond's screening. Scrolling down the landing page, we reach the respective 3C-like proteinase (NSP5) section (Sup. Fig. 4A) including 4 subsections that compile the available atomic models of the apoprotein (upper box: PDB), the protease interacting with other proteins or ligands (bottom box: Interactions), SARS-CoV and other related proteins, and computational models, respectively. The first box includes the ground-state model used in the PanDDA analysis method (PDB ID 5R8T, blue arrow), and the second box, which allows filtering the structures by experiment (orange arrow), contains the remaining PDB entries for the proteinase complexed with protein inhibitors in the Diamond experiments. Although the native protease is catalytically active as a symmetric homodimer, those crystal structures have been solved as one protomer per asymmetric unit.

Clicking any of those ligand interaction entries drives to the opening of the structure together with the common annotations panel of 3DBionotes-WS. As an example, we have selected the entry 5R84 (Sup. Fig. 4A, magenta arrow). The picture of the main protease appears complexed with the drug Z31792168 (2-cyclohexyl-~{N}-pyridin-3-yl-ethanamide) (Sup. Fig. 4B).

Protein-drug interaction residues are highlighted as spheres surrounding the structure of the drug. Particularly, yellow and green spheres stand for His41 and Cys145, respectively, the catalytic dyad of the active site. Remaining interacting residues are depicted in red. All those residues are also remarked in the sequence alignment shown at the bottom of the Sup. Fig. 4B that displays yellow and green arrows pointing to the catalytic residues. The annotations panel also highlights all residues interacting with the drug Z31792168 (Sup. Fig. 4C-2, red arrow). Respective annotation boxes of the catalytic dyad are pointed again by yellow and green arrows. As we can see, many of the residues involved in the protein-ligand interaction are almost the same that can be found in the remaining Diamond's drug screening experiments (grey boxes). Moreover, almost the same amino acids were inferred as putative binding sites by mapping in the protease SARS-CoV-2 sequence the ligand interaction residues observed in homologous coronavirus structures (Sup. Fig. 4C-1). As expected, many of these sites also include the catalytic dyad. This active site also appears annotated in the track "Domains and sites" (Sup. Fig. 4C-3), remarked with yellow and green arrows.

Supplementary Figure 4A: 3DBionotes-WS COVID-19 edition landing page showing the section of SARS-CoV-2 main protease that includes the first two subsections, PDB and Interactions, assigned to apoprotein structures and interactions with other proteins or ligands, respectively. Blue arrow points to the PanDDA ground-state model and the magenta one to the structure binding the drug Z31792168. Orange arrow points to the filter that allows to select the view of structures coming from a particular experiment.



Supplementary Figure 4B: 3DBionotes-WS 3D viewer of the atomic structure PDB entry 5R84 that binds the drug Z31792168. The model view shows the drug binding site zoomed. All the residues are depicted as red spheres except the catalytic dyad (Hys41, in yellow, and Cys145, in green). The sequence alignment between 5R84 PDB entry and Uniprot P0DTD1 is displayed below the model.

Drug binding sites are highlighted in yellow. Arrows yellow and green point to residues H41 and C145, respectively.

Supplementary Figure 4C: 3DBionotes-WS annotations display of the atomic structure PDB entry 5R84 that binds the drug Z31792168. **1)** Piled up annotations panel of functional mapping for ligands. Three examples are shown below expanding this panel. **2)** Piled up annotations panel of Diamond Light Source's fragment screening experiment. Four examples are expanded below. Drug binding sites are displayed as grey boxes except for the drug (Z31792168, red arrow) bound to the current structure in the 3D viewer (Sup. Fig. 4B), which shows red boxes. Again, arrows yellow and green point to residues H41 and C145, respectively. **3)** Piled up annotations panel of Domains & sites. Two cases, domains and the active site, are expanded below. The active site residues H41 and C145 are also pointed with yellow and green arrows as indicated before.

References:

Emsley,P. et al. (2010) Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr., 66, 486–501.

Joosten,R.P. et al. (2014) The PDB_REDO server for macromolecular structure model optimization. IUCrJ, 1, 213–220.

Joseph,J.S. et al. (2007) Crystal structure of a monomeric form of severe acute respiratory syndrome coronavirus endonuclease nsp15 suggests a role for hexamerization as an allosteric switch. J. Virol., 81, 6700–6708.

Korber,B. et al. (2020) Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell, 182, 812–827.e19.

Melero,R. *et al.* (2020) Continuous flexibility analysis of SARS-CoV-2 spike prefusion structures. *IUCrJ*, **7**.

Murshudov,G.N. *et al.* (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 355–367.

Pearce,N.M. *et al.* (2017) A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nat. Commun.*, **8**, 15123.

Pintilie,G. *et al.* (2020) Measurement of atom resolvability in cryo-EM maps with Q-scores. *Nat. Methods*, **17**, 328–334.

Ramírez-Aportela,E. *et al.* (2019) : a new deep-learning- and aspect-based local resolution method for electron-microscopy maps. *IUCrJ*, **6**, 1054–1063.

Ramírez-Aportela,E. *et al.* (2021) FSC-Q: A CryoEM map-to-atomic model quality validation based on the local Fourier Shell Correlation. *Cold Spring Harbor Laboratory*, 2020.05.12.069831.

Williams,C.J. *et al.* (2018) MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.*, **27**, 293–315.

Yurkovetskiy,L. *et al.* (2020) Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell*.