EMVS: A New Set of Annotations for cryo-EM Quality Validation Reports.

Macias J.R.[1], Ramirez-Aportela E.[1], Vilas J.L.[1], Martinez M.[1], Parra-Perez A.M.[1], Sorzano C.O.S.[1], Carazo J.M.[1].

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

The resolution achieved for single-particle Cryogenic electron microscopy (cryo-EM) maps has increased on average from 14.4 Å to 5.9 Å in just a few years, with many maps in the resolution range that allows building atomic models. The advent of the COVID-19 pandemic has multiplied the efforts to generate key viral structures to enable the early development of vaccines and promising treatments. As a result, an increasing number of new structures are being published and submitted to the Electron Microscopy Data Bank (EMDB) and the Protein Data Bank (PDB). In particular, more than 500 and 1300 SARS-CoV-2 entries have been included in EMDB and PDB, respectively. More than 500 PDB atomic structures just this year.

However, some concerns have been raised to pay particular attention not only to the quantity of data, but also to its quality. In this regard, new methods and tools have been proposed to the community to assess the quality and validate cryo-EM maps and their map-derived atomic models.

With the aim of providing map and model assessment information, in this work we present an additional group of annotations recently incorporated into the 3DBionotes web platform [1] to evaluate the quality of cryo-EM maps and their fitted atomic models (Figure 1). 3DBionotes, created and hosted by the Instruct Image Processing Centre, in the Biocomputing Unit of the National Centre of Biotechnology in Spain (CNB-CSIC), provides an interactive environment with structural and multi-omics data oriented to structural biology analysis. The scores obtained by the different validation methods incorporated in 3DBionotes inform users about how well the structural model is supported by the experimental data. In particular, the local resolution analysis methods MonoRes [2](Vilas et al., 2018), DeepRes [3](Ramírez-Aportela et al., 2019) and BlocRes [4](Cardone et al., 2003) assess the map quality, while FSC-Q [5] (Ramírez-Aportela et al., 2021) and Q-score [6](Pintilie et al ., 2020) evaluate fitting and resolvability of the built atomic model.

3DBionotes is publicly accessible at https://3dbionotes.cnb.csic.es.

References:

[1] Macias J.R. et al. 3DBionotes COVID-19 edition, Bioinformatics, 2021; btab397, https://doi.org/10.1093/bioinformatics/btab397.

[2] Vilas, J. L. et al. MonoRes: automatic and accurate estimation of local resolution for electron microscopy maps. Structure 26, 337–344 e334 (2018).

[3] 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.

[4] Cardone, G., Heymann, J. B. & Steven, A. C. One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. J. Struct. Biol. 184, 226–236 (2013).

[5] Ramírez-Aportela E. et al. (2021) FSC-Q: A CryoEM Map-to-Atomic Model Quality Validation Based on the Local Fourier Shell Correlation. Nat. Commun. 12 (1), 42. 10.1038/s41467-020-20295-w.

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

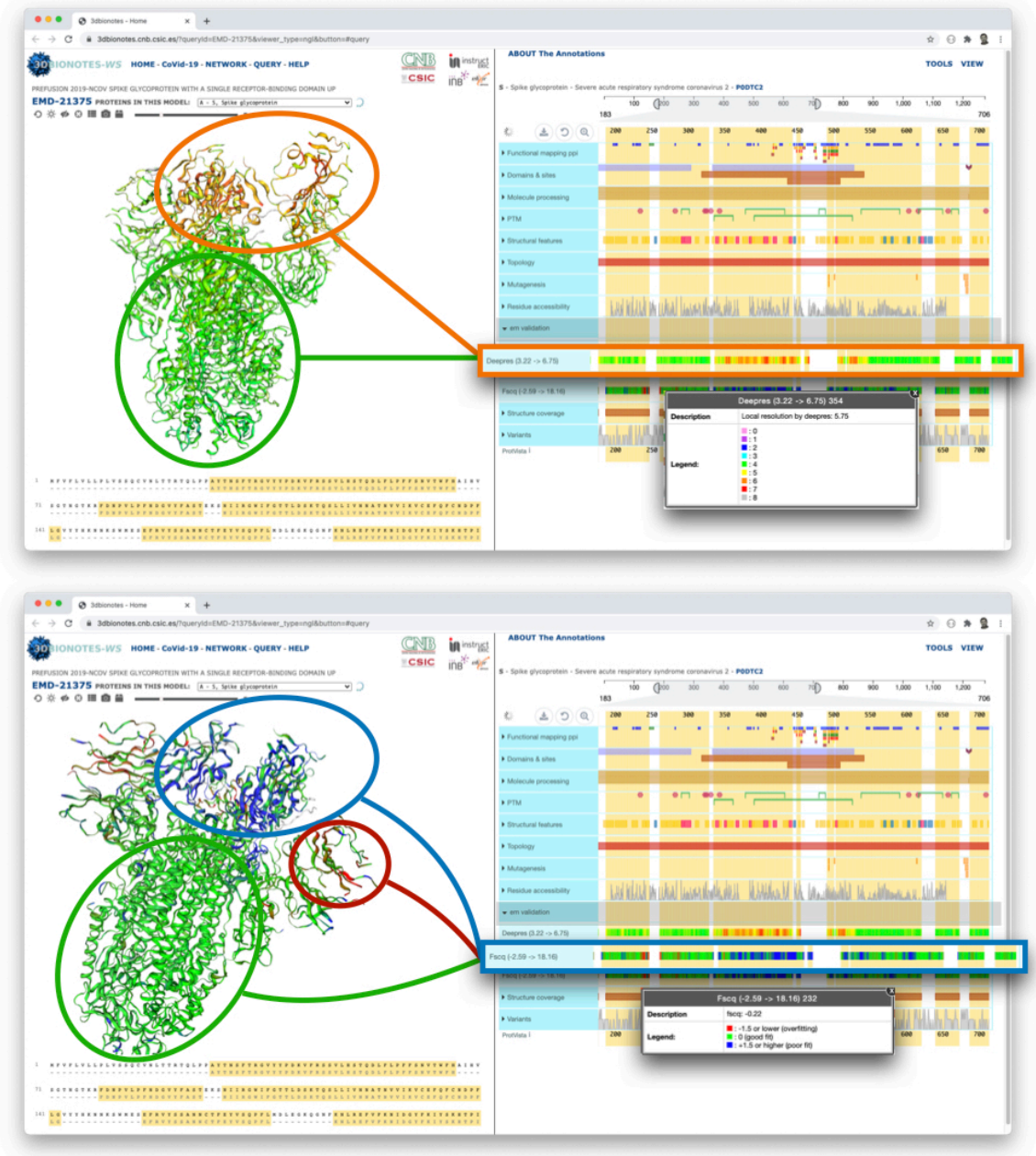

Figure 1. 3DBionotes web interface showing a validation analysis of the SARS-CoV-2 spike glycoprotein in prefusion state with a single receptor-binding domain up (EMD-21375, PDB:6VSB). Among other multi-omics annotations, in the right-hand side panel, the new validation scores are presented in a specific track with colored boxes for each residue of the atomic model. Those values can also be represented directly onto the atomic model in the 3D viewer, on the left-hand side panel, giving the user a better understanding of the local quality for certain areas of the structure. In this case, the more stable central areas of the protein (in green) perform better than the more flexible areas of the RBD domain (in red or blue) with lower local resolution values (DeepRes), possible overfitting and poorly fitted atoms or areas with low resolvability (Fsc-Q).