Qeios

Peer Review

Review of: "CryoSAMU: Enhancing 3D Cryo-EM Density Maps of Protein Structures at Intermediate Resolution with Structure-Aware Multimodal U-Nets"

Shiraz S Kaderuppan[1]

1. Newcastle University, Newcastle upon Tyne, United Kingdom

The manuscript describes an innovative approach to enhance 3D cryo-EM density maps of protein structures by adopting a deep neural network (DNN) approach to do this. Interestingly, the study seeks to integrate conventional 3D cryo-EM density maps with ESM-IF1-based structural embeddings, thereby providing contextual information to the models and naming this approach CryoSAMU.

Nonetheless, I have some queries about the manuscript as it is currently presented, and hopefully, the authors would be able to address these accordingly:

1. What batch size was applied to the Expected & Target 3D images (*exp.3D-images* & *tgt.3D-images*) for training the U-Net DNN model in Figure 1b? Did the authors consider the effect of varying the batch size and possibly the learning rate to improve the prediction accuracy of their proposed models? If so, it would be good to see how the authors arrived at this size of 64 x 64 x 64 for individual voxels, and if this was an optimal value to employ for the voxel-dimension hyperparameter in this context.
2. The authors mentioned the use of a 3D U-Net model in their proposed algorithm (CryoSAMU). This seems to be an interesting approach, and it would be good if they could illustrate the detailed architecture of this 3D U-Net (such as the block sizes in each layer), and (in particular) how the input data is handled in the 3rd dimension (Figure 1a just seems to show a 2D model architecture). Were the *exp.3D-images* fed as individual slices along the Z-plane into the U-Net before being concatenated into the final *pred.3D-images* or reconstructed (enhanced) maps? How was the 3D

 doi.org/10.32388/YULHYE

convolution used to handle the input images in this context – can the authors probably indicate the equation of the 3D convolution operation, and perhaps explain its mode of operation in a paragraph?

3. According to the authors, structure embeddings were generated using a frozen pretrained ESM-IF1 model. Can the authors explain how they validated the embeddings generated using this model, and if native environmental effects on protein folding *de novo* (as may be elucidated via the use of the Chou-Fasman index) were accounted for in the generation of these structural embeddings? Additionally, can information derived from these embeddings also be incorporated into the loss functions when training the DNN models to improve model accuracy?
4. A recent publication in *Nature Communications* (https://doi.org/10.1038/s41467-023-44593-1) highlights the use of a different model (TEMPy-ReFF) for resolving cryo-EM density maps. It would be good if the authors of this study could possibly include a comparison of their proposed DNN framework (CryoSAMU) against TEMPy-ReFF, in relation to the metrics which were being utilized in this study (i.e., RSCC, FSC, etc.).
5. Figure 3 - enlarge the images of the violin plots (perhaps rearrange them into a 2 x 2 array) as they are too small to be analyzed. Please also account for the horizontal/lateral dimension across the bulge of the violins – what does this bulge represent?
6. Table 1 – please include the B-factor type (local or global) where these are being used in the comparative algorithms (i.e., Deposit, Autosharpen, DeepEMhancer, EMready).
7. Page 17 of the PDF file – there's a typo error ('prefroms') which should rightly be 'performs'. Please correct the typo.

Actual rating: 3.5 stars (I can't select this from the above)

Declarations

Potential competing interests: No potential competing interests to declare.