nature research

https://doi.org/10.1038/s41592-020-00990-8

Supplementary information

Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction

In the format provided by the authors and unedited

Supplementary Information for

Non-uniform refinement: Adaptive regularization improves single particle cryo-EM reconstruction

The follow sections provide supplementary information concerning potential over-fitting of regularization parameters, and the sensitivity of the non-uniform refinement algorithm to variations of default parameter values and filter kernels. We also include more details concerning the masking used during refinement in the experiments, and we include results from one more experimental dataset, namely, a 90 kDa membrane protein with no soluble domain.

Over-fitting regularization parameters

As mentioned in Discussion, local resolution estimates or local statistical tests have commonly been used to adaptively filter 3D maps. While these methods are generally satisfactory as one-time post-processing steps for visualization, in our experience they can lead to severe over-fitting when used iteratively within refinement as a substitute for regularization. During iterative refinement, small mis-estimations of local resolution at a few locations, e.g., due to noise in local resolution estimates, cause subtle over- or under-fitting, leaving slight density variations. Over multiple iterations of refinement, these errors can produce strong erroneous density that contaminate particle alignments and the local estimation of resolution itself, creating a vicious cycle. Supplementary Figure 1 provides an example of such over-fitting.

The resulting artefacts (e.g., streaking and spikey density radiating from the structure) are particularly prevalent in datasets with junk particles, structured outliers, or small particles that are already difficult to align. To mitigate these problems, the approach we advocate couples an implicit resolution measure to a particular choice of local regularizer, with optimization explicitly designed to control model capacity and avoid over-fitting of regularizer parameters.

Parameter sensitivity in Non-Uniform Refinement

Here we examine the sensitivity of non-uniform refinement to the adaptive window factor (the AWF parameter), which controls the spatial support used in estimating the regularizer parameters, and the family of filters used for the adaptive regularization. We consider values of the AWF parameter between 1 and 10. And we consider a number of low-pass filter kernels, including different orders of Butterworth filter, an isotropic Gaussian kernel (the parameter of which is the standard deviation), and a Sinc filter (the parameter of which is the cut-off frequency beyond which amplitudes are reduced to zero). We consider these variations in estimating the density map from the STRA6-CaM dataset. Beyond the parameters specified, all other parameters of the algorithm were held fixed at the default values used in all other experiments in the paper.

Supplementary Figure 2 shows FSC curves for density maps computed from STRA6-CaM data for different values of AWF (left), and different types of filter kernel (right). The results demonstrate that non-uniform refinement is not highly sensitive to these parameter choices on this dataset. For all experiments in the paper we used

default choices of AWF=3 and an 8th-order Butterworth filter. We find these choices produce good results across many datasets, including STRA6-CaM. It is possible that one could vary these choices for each dataset to obtain marginal improvements in reconstruction resolution. Nevertheless, one advantage of the non-uniform refinement algorithm, as shown in these plots, is that manual tuning is not required to achieve high quality reconstructions in practice.

Masks used during refinement

In conventional uniform refinement algorithms it is common practice to apply a real-space mask to the 3D reference density at each iteration of refinement. This plays an essential role during refinement in removing noise from the solvent regions of the map, thereby facilitating more accurate particle image alignments to the 3D map, and hence the overall resolution of the final 3D map. In the experiments in this paper we used the standard dynamic masking routine in *cryoSPARC* for both uniform and non-uniform refinements. This ensures a fair comparison between the two algorithms. For this dynamic masking routine, a soft, shaped mask is generated at each iteration by thresholding the 3D density map, followed by dilation and then application of soft padding. Supplementary Figure 3 shows examples of the masks used for uniform refinement of the STRA6-CaM, PfCRT and $Na_v1.7$ channel reconstructions. In all cases, the threshold is set to 20% of the maximum density, with dilation of 6Å and soft cosine padding of 8Å.

90 kDa Membrane Protein with no soluble region.

In addition to the protein complexes discussed in the main body of the paper, here we include results on a C3 symmetric membrane protein with a molecular weight of 90 kDa and no soluble domain, and no Fabs bound. The protein mass is surrounded by a large detergent micelle, making the pose estimation of particle images difficult. The dataset, courtesy of Oliver Clarke, contains 42,740 particle images (pixel size $1.05\mathring{A}$, box size 200 pixels, C3 symmetry yielding 128,200 asymmetric units). It is part of an on-going, unpublished study, used here with permission, but we do not name the protein or show the entire 3D map.

Supplementary Figure 4A shows FSC curves from uniform and non-uniform refinements, with improvements in global 3D map resolution from $3.9\mathring{A}$ to $3.6\mathring{A}$. One can also see that some local regions improve in quality more than the numerical resolutions might suggest. Supplementary Figure 4C shows densities for three transmembrane α-helices, with two (left, center) being closer to the periphery of the protein near the micelle, and one (right) being more central. The peripheral α-helices show significant improvement in the resolved detail of the 3D map, allowing tracing of backbone and side-chains in the non-uniform refinement map. By comparison, model building would be difficult with the corresponding uniform refinement map. The improvement of the central α-helix is modest, and was chosen to show that some parts of the uniform refinement map are resolved at its nominal $3.9\mathring{A}$ resolution. Supplementary Figure 4B shows large differences in particle alignments between uniform and non-uniform refinement.

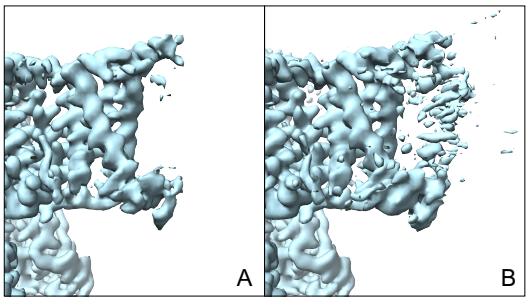

Figure 1: Side view of a peripheral transmembrane region of the $Na_V1.7$ channel showing that naive forms of local filtering during iterative refinement can lead to significant over-fitting. **A**: Unsharpened density map at the 8^{th} iteration of non-uniform refinement as presented in this work. **B**: Unsharpened density map at the 8^{th} iteration of a simpler iterative refinement scheme where local resolution estimation and local filtering are performed at each iteration. Both maps are thresholded at the same level. The map in **B** has acquired clear artefacts of over-fitting, with incorrect high-resolution density dominating the peripheral transmembrane region of the protein. This artefactual density continues to strengthen over subsequent iterations, contaminating local resolution estimates and leading to errors in particle alignments.

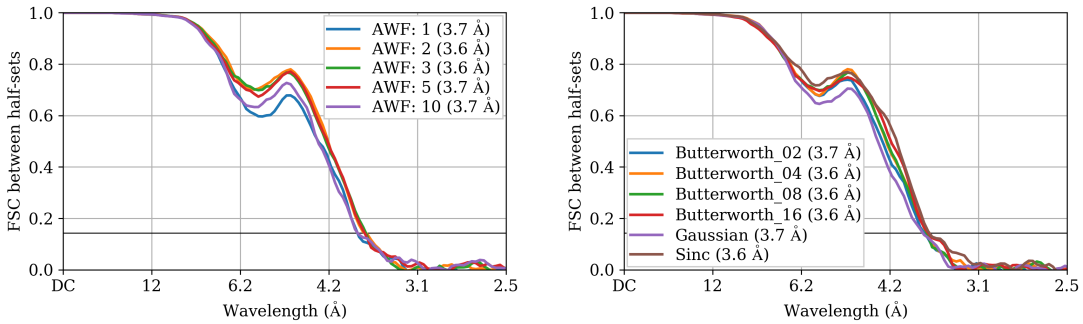

Figure 2: Gold-Standard FSC curves for non-uniform refinement reconstructions computed on the STRA6-CaM [**?**] dataset. **Left:** Curves for reconstructions computed with varying values of Adaptive Window Factor (AWF). Values of 2, 3, and 5 yield very similar resolution and FSC curves, while values of 1 and 10 show slight decrease in resolution. All results are computed with an 8th-order Butterworth filter. **Right:** Curves for reconstructions computed with varying types of filter using in adaptive filtering. Gaussian and low-order Butterworth filters perform slightly worse compared to high-order Butterworth (8th and 16th order) and Sinc filters. All results are computed with AWF=3.

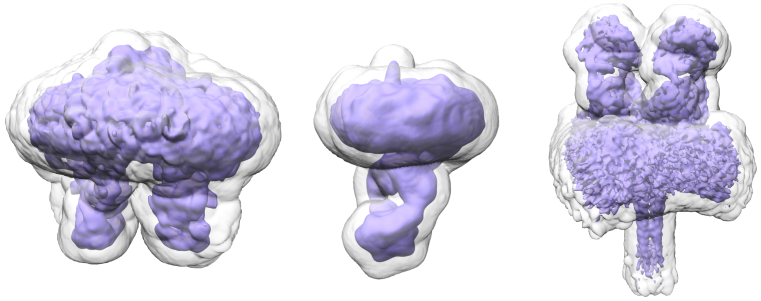

Figure 3: 3D density maps for STRA6-CaM (left), PfCRT (middle), and the $Na_v1.7$ channel (right) shown with spatial masks used during 2D-3D alignment during refinement iterations. In standard cryoSPARC dynamic masking, a soft, shaped mask is generated at each iteration by thresholding the 3D density map, followed by dilation and then application of soft padding.

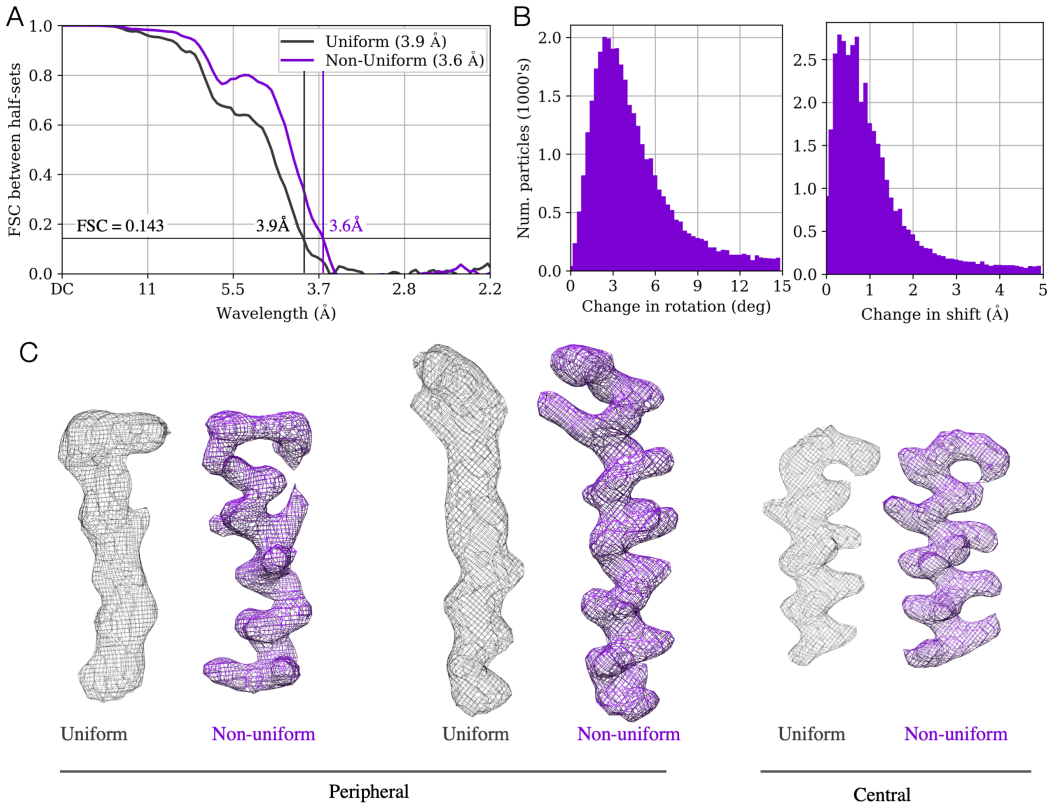

Figure 4: **C3 Symmetric Membrane Protein:** Results of uniform and non-uniform refinement on a dataset of 42,740 particle images of a C3 symmetric 90 kDa membrane protein with no soluble domains (data courtesy of Oliver Clarke). **A**: FSC curves computed using the same mask show numerical improvement from $3.9\mathring{A}$ to $3.6\mathring{A}$, indicating improved global average resolution. **B**: Histograms of change in particle alignments between uniform and non-uniform refinement. Optimal regularization in non-uniform refinement improves the ability to align particles over iterations, causing these changes. **D**: Density map detail for transmembrane α-helices. Two (left, center) are peripheral, near the micelle. One (right) is central within the protein. Both maps are filtered using the corresponding FSC curve, sharpened with the same B-factor of $-180\mathring{A}^2$, no local filtering or sharpening is used, and thresholds are set to keep enclosed volume constant. Density differences are thus due to algorithmic rather than visualization differences. Non-uniform (purple) map has clear density for backbone and side-chains while for peripheral α-helices, helical pitch is barely resolved in uniform (grey) map density. The central α-helix (right) improves less dramatically.