Robust and efficient likelihood-based docking of models into cryo-EM reconstructions

Randy J Read¹, Claudia Millán¹,², Airlie J McCoy¹ and Thomas C Terwilliger³,⁴

¹Cambridge Institute for Medical Research, University of Cambridge, The Keith Peters Building, Hills Road, Cambridge CB2 0XY, UK; ²SciBite Ltd, BioData Innovation Centre, Wellcome Genome Campus, Hinxton, Cambridge CB10 1DR, UK; ³New Mexico Consortium, Los Alamos, NM 87544, USA; ⁴Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Email: rjr27@cam.ac.uk

Keywords: cryo-EM, likelihood, docking

Fast, reliable docking of models into cryo-EM maps requires understanding of the errors in the maps and the models. Likelihood-based approaches to errors have proven to be powerful and adaptable in experimental structural biology, finding applications in both crystallography and cryo-EM. Indeed, previous crystallographic work on the errors in structural models is directly applicable to likelihood targets in cryo-EM.

In this work, we have derived likelihood targets in Fourier space to characterise, based on the comparison of half-maps, the direction and resolution-dependent variation in the strength of signal and noise in the data. Because the signal depends on local features, the signal and noise are preferably analysed in local regions of the cryo-EM reconstruction.

The signal and error parameters deduced from the comparison of half-maps are then used, in likelihood targets, to measure the agreement with docked atomic models. The likelihood-based rotation function used in crystallography in Phaser [1] can be employed to establish plausible orientations in a docking search. A phased likelihood translation function then yields scores for the placement and rigid-body refinement of oriented models.

For crystallographic molecular replacement, we have previously shown that expected log-likelihood-gain (eLLG) scores can be predicted from the anticipated quality of the model and the quality and completeness of the data, then used to inform the search strategy [2]. We show here that similar eLLG scores can be computed, in advance of the search calculation, for docking models into cryo-EM reconstructions. These scores are used to devise optimal strategies for choices of the resolution of data and the size of search volumes.

Tests demonstrate that the new procedure is fast, robust and effective at placing models into cryo-EM maps. Fig. 1 shows an example in which one chain from the crystal structure of the membrane domain of E. coli respiratory complex 1 (PDB entry 3rko [3]) can be docked correctly into the cryo-EM reconstruction of the whole complex (PDB entry 7nyu, EMDB entry 12654 [4]), in a region where the local resolution is about 11 Å.

Figure 1. Chain L of 3rko (X-ray structure) docked into poorly-ordered cryo-EM density. The docked model is shown in magenta, with chains L and M of the deposited structure (7nyu) in dark and light blue respectively.


This research was supported by the Wellcome Trust (grant 209407/Z/17/Z to RJR) and the National Institutes of Health USA (grant GM063210 to TCT and RJR).