CryoEM Map-Model Scores: From Average Density to Q-scores

Grigore Pintilie[1] and Wah Chiu[2]

[1]1. Department of Bioengineering, James H. Clark Center, Stanford University, Stanford, California, 94305 USA, United States, [2]School of Medicine, Stanford University, Stanford, California, United States

Michael Rossmann was a pioneer in combining X-ray crystallography and cryoEM maps. In his early papers in this area, he described how to score a model fitted to a map by looking at the density value at each atom; the density value for the atom was obtained by averaging the values at the nearest grid points in the map. Another measure he described was to consider whether the density value at the atom's position was negative, and hence not resolved [1, 2]. Prof. Rossmann wrote and maintained code to calculate such scores, while searching for good fits of a model in a map, part of a method called *EMFit*. He was happy to share and explain how to run the code to interested researchers, including ourselves; we had many conversations and 'remote' sessions (which were then less common), where he showed us how to run the code and how it worked. The methods became very useful to us as we also explored the combination of X-ray structures and cryoEM maps.

The scores devised by Prof. Rossmann influenced and inspired scores still used today. The 'atom in negative density' score is similar to the atom inclusion score, used in the PDB validation reports [3]. The average density score became commonly used in other software such as UCSF Chimera [4]. Other scores used in fitting atomic models to cryoEM maps had already been proposed, e.g. cross-correlation [5]. Such cross-correlation scores typically consider groups of atoms, simulating a map from them that aims to look like a cryoEM map at a given resolution; they depend on various other parameters, e.g. scattering factors and desired resolution. In contrast, the atom inclusion and average density scores ask a simpler and perhaps more effective question about each atom. As we conceived Q-scores, we asked a similar question: can a score apply to each atom individually, and represent to what degree it is resolved [6]. It also borrowed from the cross-correlation formulation, to calculate a number between -1 and 1, with 1 indicating the atom is resolved as it would be in the highest atomic resolution seen to date (Figure 1).

A useful property of Q-scores that emerged from their formulation is that, on average, they correlate strongly to the FSC-estimated resolution of the map (Figure 2). This has not been seen yet for any other generic (i.e. applicable to any atom), cross-correlation based, real-space score [7]. One benefit of this property is the ability to answer the question 'what is a good score?'. In the case of Q-scores, a 'good' value can be determined from the estimated resolution of the map: we simply plug the estimated resolution of the map in the equation for the regression line. When the Q-score of atoms fitted or built into the maps is at or above this expected value, we know this part of the map is resolved and modeled as expected.

Looking back, this journey had a lot to do with the inspiration initially provided by Prof. Rossmann. He asked the question, perhaps more indirectly, about whether a single atom is resolved or not. The question stayed with us for many years as we worked with different scores, until we developed the Q-score in the same vein. Prof. Rossmann also set a great example to develop and maintain code that can be shared and used by others in the field [8], and to take every opportunity to discuss it and potentially inspire others to make use of it or to take it even farther.
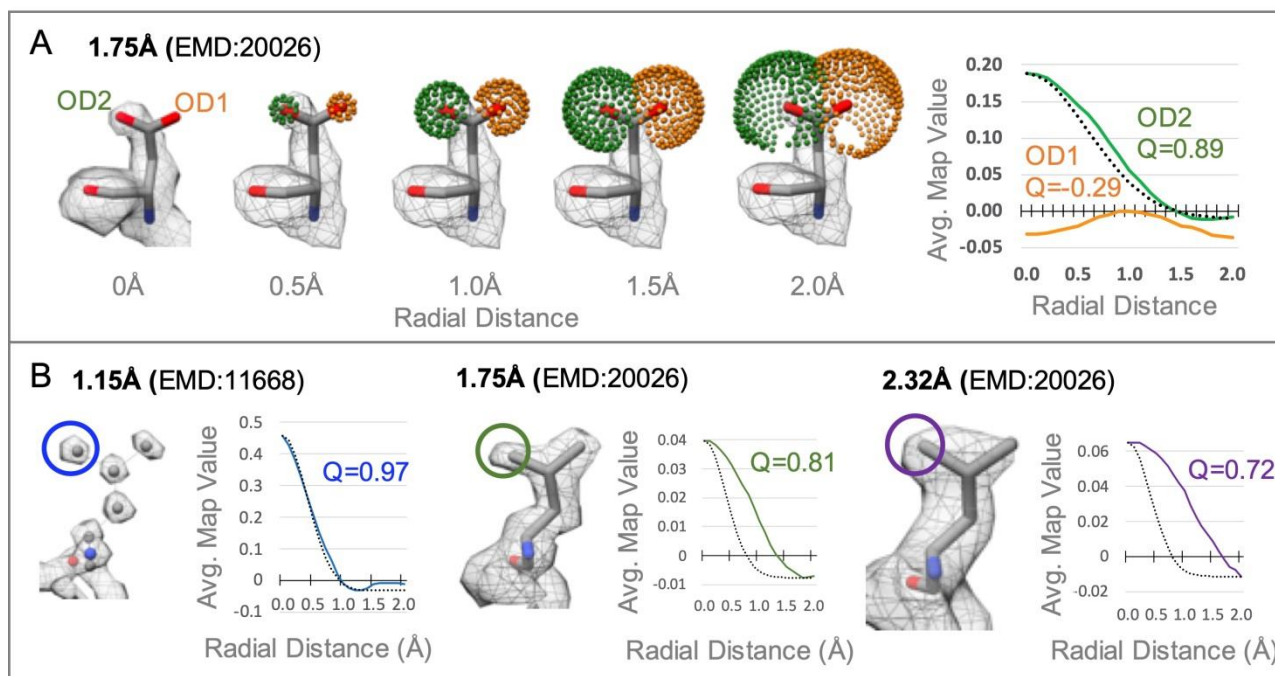
Figure 1. Figure 1. Illustration of atomic profiles and Q-scores around atoms fitted to cryoEM maps at different resolutions. In (A), the atomic profiles around two of the atoms OD1 and OD2 are plotted. The atomic profiles are obtained by distributing points on spheres with varying radii centered on each atom's position; only points on each sphere closer to the atom being considered than to other atoms in the model are used. Map values are averaged at each radial distance. For OD1 the atomic profile has a higher starting point and falls off much like a Gaussian function (plotted in black dotted line), hence it has a higher Q-score (0.89). On the other hand, for the OD1 atom, the atomic profile is mainly flat; hence a low Q-score of -0.29 is obtained. In (B), atomic profiles and Q-scores for atoms at various resolutions are shown. The reference Gaussian the atomic profiles are compared to is plotted with a black dotted line. Q-scores are highest when the atomic profiles match this reference Gaussian and lower when they are less like it.
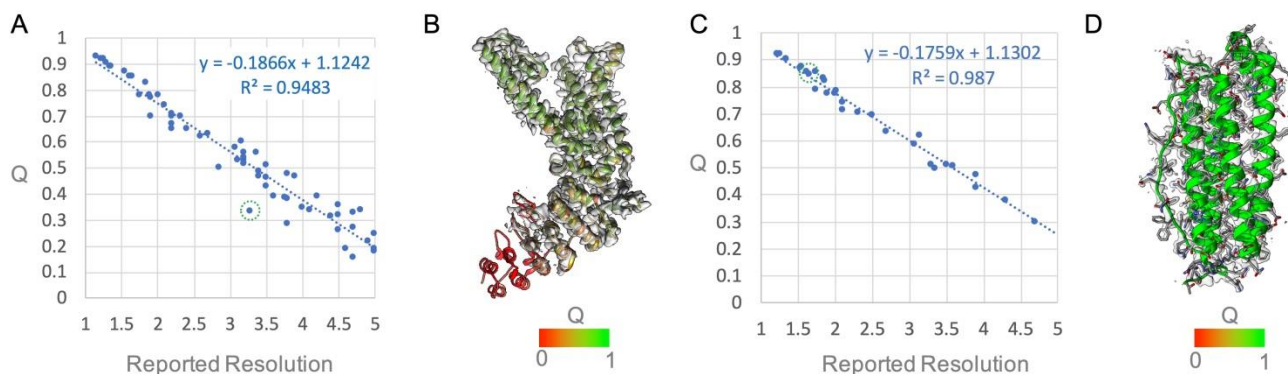

Figure 2. Figure 2. Plots of average Q-scores vs reported resolution for maps and models in the EMDB. In (A), 55 diverse maps and models are used. The average Q-score for each map and model correlates strongly to the reported resolution as shown by linear regression (R2=0.95). In (B) the map and model corresponding to a point which falls below the fitted regression line is shown (EMDB:5778); the Q-score is lower than expected because parts of the model are not resolved very well in the map (shown with red in the ribbon display). In (C), only maps and models of Apoferritin are used; an even stronger correlation between the average Q-score and reported resolution is obtained (R2=0.99). This is because, as shown in (D), maps of Apoferritin tend to have more homogeneous resolvability of atomic features throughout.

References

[1]   M. G. Rossmann, "Fitting atomic models into electron-microscopy maps," *Acta Crystallogr. D Biol. Crystallogr.*, vol. 56, no. Pt 10, pp. 1341–1349, Oct. 2000.

[2]   M. G. Rossmann, R. Bernal, and S. V. Pletnev, "Combining electron microscopic with x-ray crystallographic structures," *J. Struct. Biol.*, vol. 136, no. 3, pp. 190–200, Dec. 2001, doi: 10.1006/jsbi.2002.4435.

[3]   I. Lagerstedt *et al.*, "Web-based visualisation and analysis of 3D electron-microscopy data from EMDB and PDB," *J. Struct. Biol.*, vol. 184, no. 2, pp. 173–181, Nov. 2013, doi: 10.1016/j.jsb.2013.09.021.

[4]   T. D. Goddard, C. C. Huang, and T. E. Ferrin, "Visualizing density maps with UCSF Chimera," *J. Struct. Biol.*, vol. 157, no. 1, pp. 281–287, Jan. 2007, doi: 10.1016/j.jsb.2006.06.010.

[5]   P. L. Stewart, S. D. Fuller, and R. M. Burnett, "Difference imaging of adenovirus: bridging the resolution gap between X-ray crystallography and electron microscopy," *EMBO J.*, vol. 12, no. 7, pp. 2589–2599, Jul. 1993.

[6]   G. Pintilie, K. Zhang, Z. Su, S. Li, M. F. Schmid, and W. Chiu, "Measurement of atom resolvability in cryo-EM maps with Q -scores," *Nat. Methods*, pp. 1–7, Feb. 2020, doi: 10.1038/s41592-020-0731-1.

[7]   C. L. Lawson *et al.*, "Cryo-EM model validation recommendations based on outcomes of the 2019 EMDataResource challenge," *Nat. Methods*, vol. 18, no. 2, Art. no. 2, Feb. 2021, doi: 10.1038/s41592-020-01051-w.

[8]   Pintilie, Grigore, *MapQ, DOI: 10.5281/zenodo.3595455 https://github.com/gregdp/mapq*. 2019.