doi:10.1017/S1431927617004718
Microsc. Microanal. 23 (Suppl 1), 2017

New Strategies for Improving CryoEM Single Particle Analysis in EMAN2.2

Steven J. Ludtke[1], Tunay Durmaz[1], Muyuan Chen[2] and James M. Bell[2]

[1.] Verna and Marrs McLean Dept. of Biochemistry, Baylor College of Medicine, Houston, TX, USA.
[2.] Graduate Program in Structural and Computational Biology & Molecular Biophysics, Baylor College of Medicine, Houston, TX, USA.

Single Particle Analysis (SPA) is now considered to be a mature field, and achieving near-atomic resolution for rigid molecules which behave well on the grid is a straightforward process. However, even in 2016 over 50% of published structures were at worse than 6 Å resolution, and over 35% failed to achieve 10 Å. This doesn't even consider the substantial number of unpublished failures. While some of these issues must be addressed by improved specimen preparation, there is still room for new software strategies to have a substantial impact. We present two new methodologies available in EMAN2 [1] in detail, and discuss other new concepts briefly: first, a strategy for improving the high resolution detail in structures containing domains with significant flexibility; second a new strategy for quantitative elimination of bad particles and qualitative assessment of overall data set quality.

When refining a 3-D structure where one or more domains contain significant flexibility, there are two canonical approaches: to refine the full data set to achieve the best structure in the non-flexible region or to subdivide the data into multiple populations to try and resolve the variability itself. The difficulty with the first approach is that the flexible domain influences the alignment of the particles, and thus, depending on geometry, degrades the quality of the rigid domain. A related issue is that typically, maps are filtered uniformly, such that the visible detail is the same across the entire map, underfiltering the flexible domains, and overfiltering the rigid domains. The second issue can be handled through local resolution estimation[2] and local filtration. To address the broader problem, we have integrated our own local resolution and local filtration process into the standard EMAN2 iterative refinement loop. This progressively de-emphasizes flexible domains during alignment, and produces a map with near-optimal local resolvability. The effect is also beneficial for flexible domains, as it pulls together the fragmentary density produced by under-filtration, and makes the overall conformation of those domains more interpretable (Fig 1.)

The problem of quantitatively identifying and eliminating bad particles from a 3-D reconstruction while minimizing human bias remains an important issue for the field. A "bad" particle generally refers to anything that does not constructively contribute to the structure. Unfortunately, such particles fall into at least 2 categories, which are not generally distinguished among: particles that are not biochemically the particle of interest (denatured, or something else entirely) and particles in a different conformational state. One commonly practiced methodology, that of performing 3-D classification multiple times, each time excluding particles associated with structures the scientist dislikes, contains an unhealthy level of subjectivity, especially when very large fractions of the data are discarded. Numerous methods have been developed to quantify particle quality, for example, by comparing particles to preliminary map projections[3]. Successful discrimination generally requires very high image contrast (e.g. large viruses).

We introduce a new technique making use of several related parameters that has demonstrated good separation even for relatively small objects. The Fourier Ring Correlation (FRC) between a particle and a 3-D projection is quite noisy, but a weighted integral of this curve yields a good metric for particle

orientation determination. When trying to use this parameter for identifying bad particles, the FRC does not provide good statistical separation unless the particles have very high contrast. This is largely due to the fact that there are correlations and anticorrelations across different resolution ranges in the FRC. Ice thickness, defocus, buffer composition and other factors all influence the relative values. On the other hand, if the putative particle is not a particle, then we don't expect any such correlations to occur, which helps make separation possible. Fig. 2 Shows the simplest of 2-D comparison for the beta-galactosidase data from the recent CryoEM Map Challenge. Despite the low mass of the particle, clear separation is visible, and refinements performed on the particles from the 2 lobes of this plot clearly show that, as expected, the useful data for high resolution refinement is contained in the upper right lobe.

These are just two of the methodologies which continue to emerge in this field to tackle the problems associated with structural variability and the oft-challenging problem of preparing optimal specimens, and imaging from optimal regions on the grid.

References:

[1] G. Tang *et al.*. J Struct Biol. 157 (2007) 38.
[2] A. Kucukelbir *et al.*. Nat Methods. 11 (2014) 63.
[3] X. Zhang *et al.*. Proc Natl Acad Sci U S A. 105 (2008) 1867.
[4] The authors acknowledge funding from the NIH, grants R01GM080139 and P41GM103832.

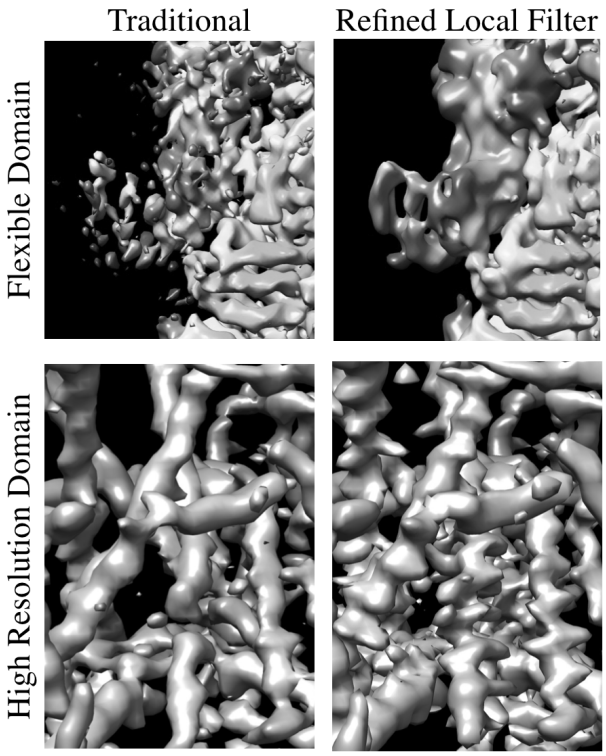

Figure 1 – Example of iterative local filtration applied to a ~4.5 Å Cryo-EM map. Note the improved sidechains in the helices (bottom), and the improved density in the flexible domain (top). Note different scale top vs bottom.

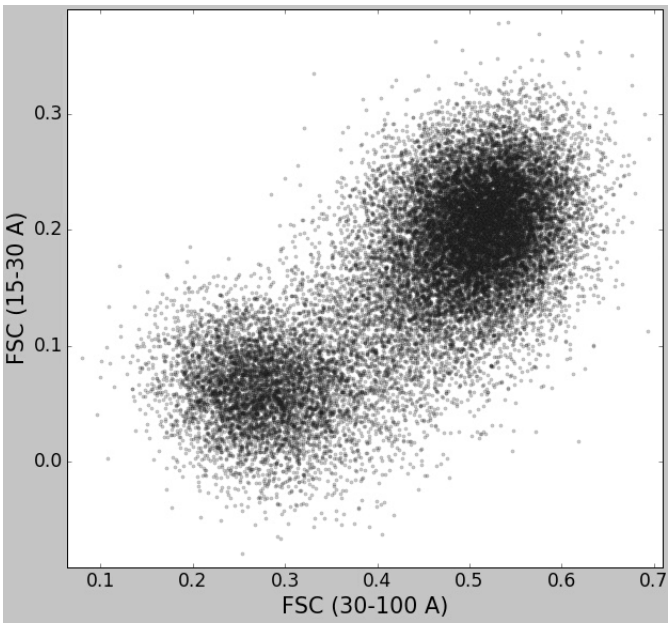

Figure 2 – Each point in this plot represents one putative beta galactosidase particle. The lower left domain represents the "bad" particles. A similar bimodal distribution has been observed over a wide range of specimens.

https://doi.org/10.1017/S1431927617004718 Published online by Cambridge University Press