doi:10.1017/S1431927618006566
Microsc. Microanal. 24 (Suppl 1), 2018

Using Reconstruction Statistics to Predict the Number of Images Required for Single Particle Analysis

J. Bernard Heymann

National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD, USA

Despite the remarkable successes of cryo-electron microscopy in solving biomolecular structures, significant uncertainties persist. One practical question is how much data are needed to achieve near-atomic resolution in single particle analysis (SPA). It has a complicated answer, because it depends on the quality of the data, the accuracy of the alignment of the particle images, and the heterogeneity of the specimen. These issues impose limits on the detail in reconstructions, leading to diminishing returns on investment in more data. Using established theory, it is possible to estimate the point at which improving reconstruction quality from an existing set of micrographs becomes prohibitively expensive.

The resolution in a single particle reconstruction is determined as a cutoff in a statistical calculation such as the Fourier shell correlation (FSC) or the spectral signal-to-noise ratio (SSNR). The latter has various formulations in the literature [1-3]. They are related to the following, where the reconstruction SSNR is a function of the spatial frequency, s, and the number of asymmetric units, n (i.e., number of particles times the symmetry order):

$$\alpha_r(n,s) = \frac{n}{2Ds}\langle\alpha(s)\rangle e^{-\frac{B}{2}s^2}$$

The reconstruction SSNR increases with the average particle SSNR, $\langle\alpha(s)\rangle$, and decreases with the particle size, D. The final term is a Gaussian decay function characterized by a B-factor that includes the effects of imaging variability (such as movement), alignment error, and structural variation. Estimating the B-factor is key to developing an expected value for the number of particle images needed, particularly at high resolution.

As a test case, I used the publicly available data set of ß-galactosidase [4] and processed it with the Bsoft package [5]. I adapted the nearest-neighbor, frequency space reconstruction algorithm in Bsoft to calculate separate signal and noise contributions, including all D2 symmetry-related orientations (Figure 1a). The signal can be compared to a reference curve calculated from the atomic structure (Figure 1a, black curve) to model the decay and determine the B-factor (Figure 1b). Choosing a target SSNR of 1/6 (equivalent to a Fourier shell correlation cutoff of 0.143), I then calculated the number of particles needed to reach a desired resolution (Figure 2a). The number of particles required increases steadily up to ~1000, beyond which it escalates dramatically. Whereas the slope of the log-log plot in the first part is reasonable (~3), it increases to >10 in the second part. This means that an enormous number of particles (≫20000 or 80000 asymmetric units) need to be collected to advance the resolution much beyond the 3.4 Å of the final reconstruction (Figure 2b). This however assumes that the imaging details and sample quality remains the same. The analysis here only indicates what can be done with a given set of data and whether it is productive to pursue further processing. If not, the user may consider alternative actions, such as optimizing the sample preparation and imaging conditions [6].

References:

[1] M Unser, *et al*, Ultramicroscopy **30** (1989), p. 429.
[2] PB Rosenthal and R Henderson, J Mol Biol **333** (2003), p. 721.
[3] HY Liao and J Frank, Structure **18** (2010), p. 768.
[4] A Bartesaghi, *et al*, Proc Natl Acad Sci U S A **111** (2014), p. 11709.
[5] JB Heymann, Protein Sci **27** (2018), p. 159.
[6] This work was supported by the Intramural Research Program of NIAMS.

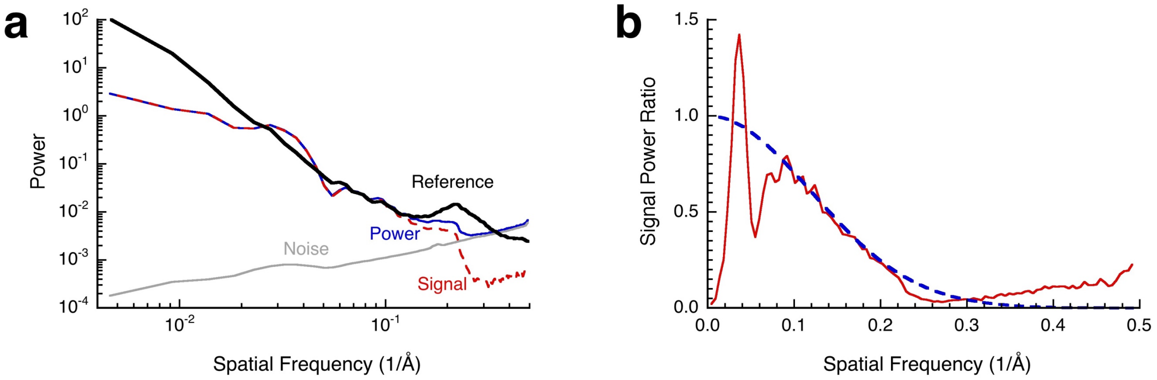

Figure 1. (a) Contributions of the signal (red) and noise (gray) to the radial power spectrum (blue) of a reconstruction of ß-galactosidase, compared to a reference spectrum calculated from an atomic model (black). (b) The ratio of the signal to reference power (red) was calculated from the red and black curves in panel a. A Gaussian decay curve (dashed blue) was fitted to give a B-factor of 70 Å^2.

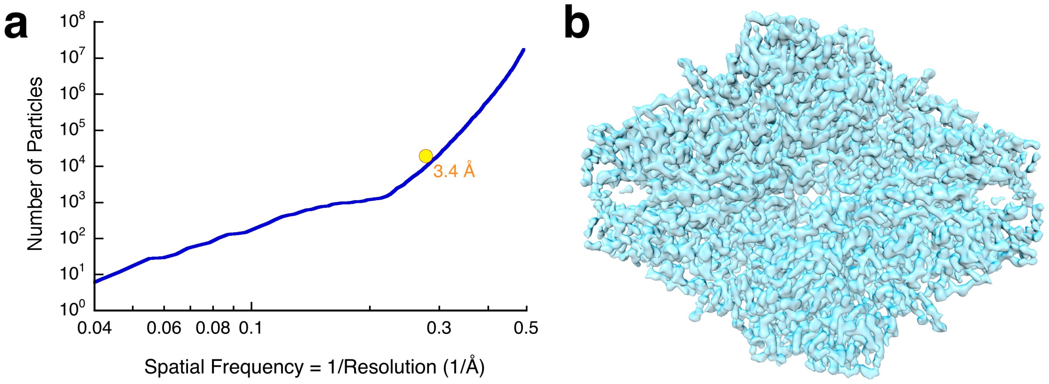

Figure 2. (a) The number of ß-galactosidase particles required to achieve a specific resolution for an SSNR, $\alpha_r(n,s) = 1/6$ (FSC = 0.143). The average particle SSNR, $\langle\alpha(s)\rangle$, was modeled as the reference curve (Figure 1a) as signal, and with unit noise power. The other parameters are: D = 217 Å, B = 70 Å^2. Also indicated (yellow dot) is the actual resolution calculated from two independent reconstructions. Note the change in slope indicating a sharp increase in the number of particles to reach higher resolutions.
(b) Isosurface rendering of a reconstruction from ~20000 particle images.

https://doi.org/10.1017/S1431927618006566 Published online by Cambridge University Press