The Phantom in the Noise and Validation of 3D EM Reconstructions

J. Bernard Heymann

National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, 50 South Dr, Bethesda, MD 20892

Single particle reconstructions derived from electron micrographs yield increasing detail of biological molecules and assemblies. However, it is well-known that alignment of pure noise images against a reference can recover that reference on reconstruction [1]. The question is how to distinguish a reconstruction from noise from one from legitimate particle images. Figure 1A shows the estimated resolutions of reconstructions from projection images with defined signal-to-noise ratios (SNR_{imp}). Comparison with reconstructions from aligned gaussian noise images provides a lower bound for the SNR that can be tolerated (open circles in Figure 1A). Increasing the number of images suppresses noise, giving a relationship between the product of SNR with the number of images and the resolution obtained (Figure 1B). Therefore, given the number of images contributing to a reconstruction with a defined SNR, and considering alignment errors (see below), a particular resolution should be expected. Any excess in images used in a reconstruction may signal validity issues.

While the results in Figure 1 are suggestive, a better modeling of the relationship between reconstruction quality and the relevant influences has been developed [2-4] and is used here with some modifications. The SSNR (spectral SNR) of the reconstruction at a reciprocal voxel-distance of k is a function of the number of images used, n_p, the symmetry order, n_s, the average signal-to-noise ratio in the images, and deviations in rotational (view) and translational alignment, σ_v and σ_t, respectively:

$$SSNR(k) = \frac{n_p n_s}{2k} \frac{\left\langle |F(k)|^2 \right\rangle}{\left\langle |N(k)|^2 \right\rangle} e^{-2\left(\sigma_v^2 + \sigma_t^2\right)\left(\frac{k}{D}\right)^2} \quad (1)$$

where D is the size of an image (edge length in pixels). The Fourier shell correlation is calculated as:

$$FSC(k) = \frac{SSNR(k)}{SSNR(k)+1} \quad (2)$$

A reasonable correspondence was found using these equations for reconstruction from noise-imposed images with perfect alignment (Figure 2A). The coefficients for rotational and translational alignment were evaluated from the FSC curves of reconstructions from noise-free images with imposed alignment errors (Figure 2B). Application of this model to real cases requires the determination of the SNR in images and an accurate estimate of the errors in alignment. While difficult, it should be possible to get approximations that would allow the development of a tool for validation.

References:
[1] M van Heel, PNAS **110** (2013) E4175-4177.
[2] HY Liao and J Frank, Structure **18** (2010) 768-775.
[3] PB Rosenthal and R Henderson, J Mol Biol **333**(2003) 721-745.
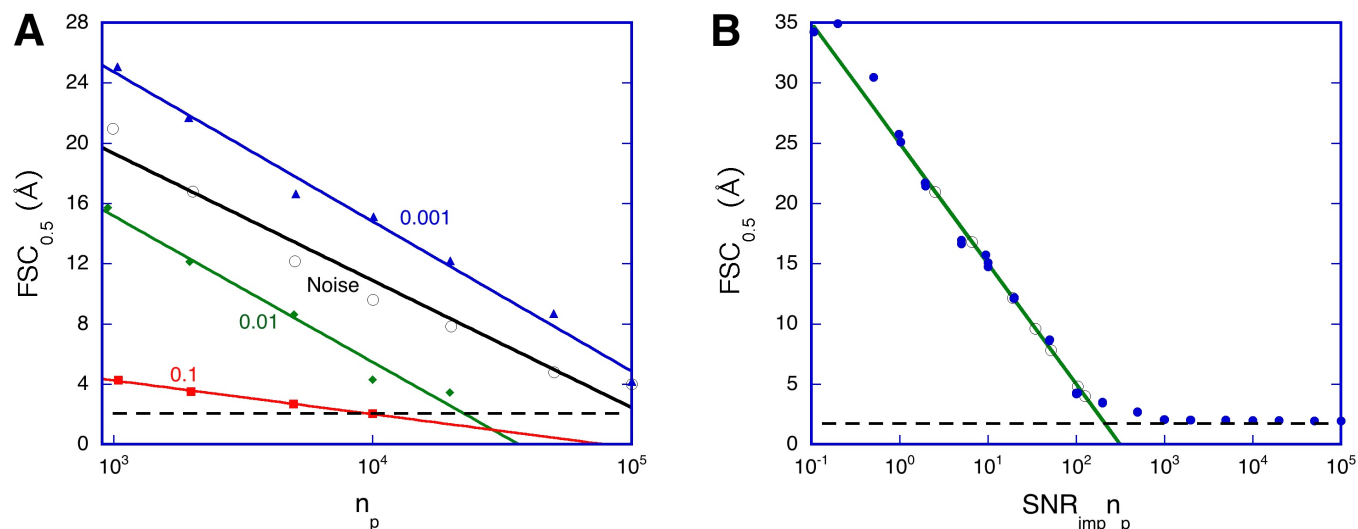[4] M Unser *et al,* J Struct Biol **149** (2005) 243-255.

Figure 1: Comparison of the resolution estimates obtained for reconstructions from pure noise images (open circles) with projections with imposed noise (filled symbols). (All calculations were done with Bsoft: http://bsoft.ws) (A) A map was calculated from the coordinates of proteinase K (PDB: 3DYB) with dimensions 80x80x80 and sampled at 1 Å/voxel. Gaussian noise images were generated and aligned to this map using information up to the Nyquist frequency, two half-reconstructions calculated and the FSC curves determined (open circles). Projections of the map were generated, gaussian noise imposed, two half-reconstructions calculated and the FSC curves determined (filled symbols). The fitted lines are logarithmic functions to show the trends, but are not applicable here in terms of a specific theory. (B) The same resolution estimates as in (A) as a function of the product between the imposed SNR and the number of images. The line is given by $FSC_{0.5} = 25 - 10 \log (SNR_{imp}n_p)$. The resolution estimates of the pure noise reconstructions are shown on the fitted line (open circles), giving an apparent average SNR = 0.0027 (StDev ~ 0.0009).
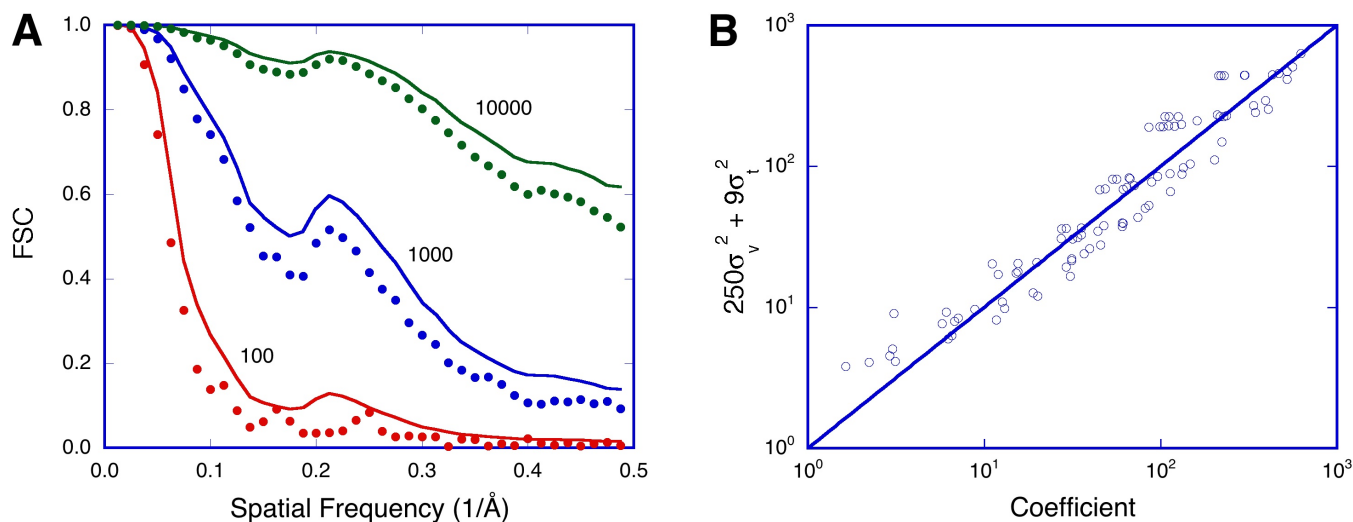

Figure 2: (A) Estimated FSC curves using equations 1 and 2 for $SNR_{imp} = 0.1$ and different numbers of images with perfect alignment. (B) Estimating the coefficients for rotational and translational error by fitting FSC curves for reconstructions calculated from 1000 noise-free images with gaussian-distributed deviations from correct alignments, giving a combined coefficient for rotation and translational errors.