

# Heterogeneity of Large Macromolecular Complexes Revealed by 3D Cryo-EM Variance Analysis

Wei Zhang,<sup>1</sup> Marek Kimmel,<sup>2</sup> Christian M.T. Spahn,<sup>3</sup> and Pawel A. Penczek<sup>1,\*</sup>

<sup>1</sup>The University of Texas—Houston Medical School, Department of Biochemistry and Molecular Biology, Houston, TX 77030, USA

<sup>2</sup>Rice University, Department of Statistics, Houston, TX 77005, USA

<sup>3</sup>Institut für Medizinische Physik und Biophysik, Charité—Universitätsmedizin Berlin, 10117 Berlin, Germany

\*Correspondence: [Pawel.A.Penczek@uth.tmc.edu](mailto:Pawel.A.Penczek@uth.tmc.edu)

DOI 10.1016/j.str.2008.10.011

## SUMMARY

Macromolecular structure determination by cryo-electron microscopy (EM) and single-particle analysis are based on the assumption that imaged molecules have identical structure. With the increased size of processed data sets, it becomes apparent that many complexes coexist in a mixture of conformational states or contain flexible regions. We describe an implementation of the bootstrap resampling technique that yields estimates of voxel-by-voxel variance of a structure reconstructed from the set of its projections. We introduce a highly efficient reconstruction algorithm that is based on direct Fourier inversion and that incorporates correction for the transfer function of the microscope, thus extending the resolution limits of variance estimation. We also describe a validation method to determine the number of resampled volumes required to achieve stable estimate of the variance. The proposed bootstrap method was applied to a data set of 70S ribosome complexed with tRNA and the elongation factor G. The proposed method of variance estimation opens new possibilities for single-particle analysis, by extending applicability of the technique to heterogeneous data sets of macromolecules and to complexes with significant conformational variability.

## INTRODUCTION

Cryo-electron microscopy (cryo-EM), together with digital image processing, is a well-established method for structure determination of large macromolecular complexes (>200 kDa). The underlying assumption of single-particle reconstruction is that the macromolecules are isolated, randomly oriented, and have identical structure. If this is the case, the images obtained using the electron microscope are 2D parallel beam projections of the same 3D object with unknown orientations. After selection of projection images from electron micrographs, the orientation parameters are determined using alignment procedures (Penczek et al., 1992, 1994; van Heel et al., 2000), and the 3D density distribution is calculated using a 3D reconstruction algorithm (Haraux and van Heel, 1986; Penczek et al., 2004; Radermacher, 1992). Because

of the need to preserve macromolecules in vitreous ice, the electron dose is limited to a minimum, and the data have very low signal-to-noise ratios (SNRs) (<1.0). This is overcome by inclusion of a very large number,  $10^4$ – $10^6$ , of projection images. Averaging of this amount of data in 3D space results in detailed maps of complexes studied, at a resolution reaching  $\sim 5$  Å, thus allowing backbone tracing (Ludtke et al., 2008; Schüler et al., 2006). The quality of 3D cryo-EM density maps determined using single-particle reconstruction is adversely affected by imperfections in the data, which we will broadly refer to as noise. The sources of noise fall into three categories (Penczek et al., 2006): (1) additive background noise unrelated to the data that originates from the solvent or the background carbon film, (2) alignment errors, and (3) noise due to conformational variability of the specimen imaged, or non-stoichiometry of ligand binding. Although the presence of the first category noise manifests itself in a uniform variance in the reconstructed map, the other two yield nonuniform and structure-dependent distribution of the variance.

Traditionally, the quality of EM maps is evaluated using Fourier techniques (Frank et al., 1981; Saxton and Baumeister, 1982). The “resolution” of an EM map is determined by randomly splitting the available set of 3D projection images into halves, calculating a 3D reconstruction for each subset, and calculating a cross-correlation coefficient between Fourier transforms of the two objects as a function of spatial frequency. This yields the so-called Fourier Shell Correlation (FSC) curve. The FSC is also a measure of a distribution of SNR in the Fourier space, the Spectral SNR (SSNR) (Penczek, 2002; Unser et al., 1987). Resolution is reported as a spatial frequency limit beyond which the SSNR drops below a selected level—for example, equal to 1. Finally, it is also possible to calculate approximate values of Fourier space variance, which provides an additional measure of reliability of the map (Penczek, 2002).

Although Fourier techniques proved to be very useful in assessing global quality of the map, in many cases information about local, real-space, reliability of the map is much more valuable. This information is given by a voxel-by-voxel real-space variance of 3D map which, if available, is helpful in (1) assessment of nonlocality of errors due to alignment errors, conformational heterogeneity of complexes, or substoichiometric ligand binding; (2) determination of reliability of small features in the map that can be artifacts induced by alignment procedures; and (3) assisting the user in docking the known structural domains into EM density maps and alerting to the possibility of multiple solutions.

Previously, we have laid out the foundations for calculation of the voxel-by-voxel variance of a structure computed as a 3D

reconstruction from the set of 2D projections, based on very simple premises (Penczek et al., 2006). We assume that voxels in the reconstructed 3D map can be considered weighted sums of pixels in 2D projections. If the number of 2D projections is large, the estimates of variances and covariances can be calculated using a variant of the bootstrap technique (Davison and Hinkley, 1997). Specifically, a new set of  $N$  2D projections is selected with replacement from a given set of  $N$  projections. In the new set, some of the original projections will appear more than once, and others will be omitted. This selection process is repeated  $B$  times and for each new set of projections a corresponding 3D bootstrap volume is calculated. The voxel-by-voxel (bootstrap) variance  $S^{*2}$  of these volumes is calculated yielding an estimate of the distribution variance  $\sigma^2$  (Hansen et al., 1953):

$$E(S^{*2}) = \frac{\sigma^2}{N}. \quad (1)$$

In the extensive set of tests included in Penczek et al. (2006), we demonstrated that the bootstrap variance (equation 1) contains a number of components of which the variance of the background noise is most dominant. However, the average level of the background variance  $S_{Back}^2$  can be independently estimated using samples of background noise from micrographs or from the spherical region encompassing the reconstructed structure, so that we can isolate the variance component due to variability of the structure:

$$E(S^{*2} - S_{Back}^2) = \frac{\sigma_{Struct}^2}{N}. \quad (2)$$

In the absence of a convincing method for distinguishing the variance arising from alignment errors,  $\sigma_{Struct}^2$  contains both components; however, for samples with conformational heterogeneity the misalignment variance is comparatively small.

So far, the bootstrap method for estimation of the 3D variance in real space (Penczek et al., 2006) has been used for several systems, including the 70S ribosome (Penczek et al., 2006), a transcription factor (Grob et al., 2006), the Ku70–Ku80 dimer (Rivera-Calzada et al., 2007), the 40S ribosomal subunit (Gilbert et al., 2007), and the anaphase-promoting complex/cyclosome (Ohi et al., 2007), and has been shown useful in detecting sample heterogeneity. However, the method is far from becoming routine in cryo-EM structural studies, because key steps in the analysis require interventions of the researcher. Moreover, in the original approach, the Contrast Transfer Function (CTF) effects of the electron microscope were ignored, so the 3D variance is limited in resolution. On the other hand, the demand for powerful methodology for the estimation of the local error of cryo-EM maps is increasing with the development and use of advanced methods for automated docking of molecular models into the cryo-EM densities. Without proper estimation of the local error, it is difficult to prevent overinterpretation of the data and to decide the reliability of the resulting pseudo-atomic models.

Here, we present significant improvements of the technique that respond to demands posed by particularities of cryo-EM data. In the 3D reconstruction algorithm, we included correction for the CTF of the microscope and added compensation for the uneven distribution of sampling points in reciprocal space. Importantly, we developed methods to correct for the normalization

errors in projection data and to determine the number of bootstrap volumes that should be generated to obtain a reliable estimate of the variance. The new version of the method is validated using a heterogeneous complex of the elongation factor G (EF-G) and the endogenous tRNA bound to the *Thermus thermophilus* 70S ribosome (Connell et al., 2007). In a previous study, the 3D variance map for a similar ribosomal complex obtained by the old method resulted in high localized variance caused by the presence or absence of ligands. The improved bootstrap method described here is also capable of detecting a small rotation between the two ribosomal subunits, demonstrating the superiority of the new approach.

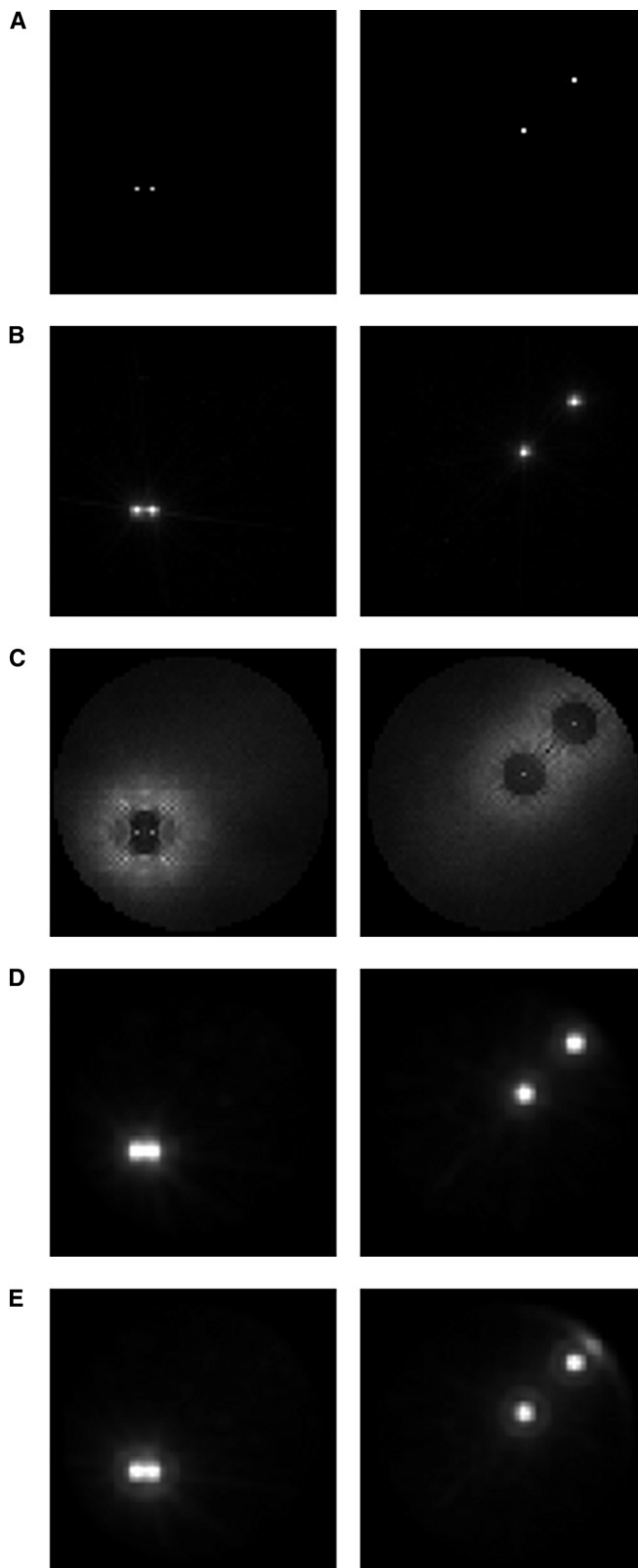
## RESULTS

### Improvement of Resolution of the Variance Map by CTF Correction

First, we illustrate improvement in the resolution of the variance map for a 3D structure calculated from its electron microscope projections that is due to inclusion of the CTF correction by constructing a model case and conduct bootstrap variance calculations with and without CTF correction and compare the results.

The test model comprised four one-pixel objects with amplitudes randomly varying according to normal distribution  $N(1,1)$ . The coordinates of these four points were (23, 27, 27), (27, 27, 27), (42, 42, 47), and (55, 55, 47), respectively, and they were placed in a box sized  $75^3$  voxels. Then we selected a set of 1328 quasi-evenly distributed projection directions. For each direction, we generated a test structure with randomly adjusted amplitudes of embedded points and generated computationally its 2D projection using Eulerian angles of selected projection direction. Finally, we modified all 2D projections by CTFs, assuming physical pixel size of 4.88 Å, microscope voltage of 300 kV, spherical aberration of 2.0 mm, amplitude contrast of 0.1, and with defocus selected with equal probability from one of the three settings: 2.25 μm, 3.00 μm, and 4.00 μm.

We calculated two 3D variance maps for our set of simulated 2D projections: one with CTF correction included in the 3D reconstruction algorithm and the other without. For the calculation of each map, we generated 1024 bootstrap volumes, applied low-pass filtration using Butterworth filter with cut-off frequency  $0.12 \text{ Å}^{-1}$  or  $0.03 \text{ Å}^{-1}$  (for reference the first zero of CTF appears at frequency  $0.035 \text{ Å}^{-1}$ ), and computed the variance map for all these cases. In each case, the sample correlation coefficients between the variance map obtained from odd and even bootstrap volumes was  $\sim 0.95$ , indicating that the selected number of bootstrap volumes are sufficient to yield robust results. The variance map calculated using the CTF correction filtered at  $0.12 \text{ Å}^{-1}$  (Figure 1B) reproduces almost perfectly the variance of original structures (Figure 1A) with the expected loss of resolution that is due to low-pass filtration of bootstrap volumes, which is necessary in order to suppress high-frequency reconstruction artifacts. However, the resolution of the CTF-corrected variance map reaches the theoretical limit of the reconstructed map, as the distance between two first points is only four pixels. In contrast, the variance map computed without CTF correction filtered at  $0.12 \text{ Å}^{-1}$  (Figure 1C) contained very strong artifacts that would make the detection of “true” variance regions



**Figure 1. Bootstrap Variance Maps Calculated Using Simulated Data Modified by the CTF**

We show only selected z-slices; the left column is  $z = 27$ , and the right column is  $z = 47$ . Contrast within each slice was adjusted independently, so the intensities do not reflect absolute values in respective slices.

impossible. The strong artifacts can be suppressed by applying very strong low-pass filtration  $0.03 \text{ \AA}^{-1}$  (Figure 1E); however, under such a strong low-pass filtration, the two points very close to each other become indistinguishable.

#### Validation of the Bootstrap Variance Calculation Method

The variance analysis was applied to a data set of the *T. thermophilus* 70S ribosome complexed with the elongation factor EF-G. EF-G was stalled using the nonhydrolyzable GTP-analog GMPPNP. The complex also contains endogenous tRNA in the P/E site (Connell et al., 2007). The occupancy of ribosomes by EF-G has been estimated to be 60%–70%, on the basis of a centrifugal binding assay. Occupancy in this range has been considered in the past to be high enough to calculate single-particle reconstructions without taking sample heterogeneity into account. Indeed, at a low-to-intermediate resolution ( $>10 \text{ \AA}$ ), EF-G was directly visible in a structure obtained from the complete data set. The subsequent multireference alignment revealed a subset of EF-G-containing particles that yielded a  $7.3 \text{ \AA}$  structure (Connell et al., 2007). Nevertheless, it became apparent that the heterogeneity of the sample is quite complex and combines compositional heterogeneity (i.e., the presence or absence of EF-G and tRNA) with conformational heterogeneity (i.e., the ratchet-like subunit rearrangement [RSR] or the movement of the L1 protuberance). Therefore, we chose this data set for more detailed analysis using the new version of the bootstrap variance calculation.

The images of the 70S•EF-G•GMPPNP complex were collected using a Tecnai F30 G2 Polara EM operated at 300 kV at a defocus range from  $2.5 \text{ }\mu\text{m}$  to  $4.7 \text{ }\mu\text{m}$  (Connell et al., 2007). The original window size was  $300 \times 300$  pixels, and pixel size was  $1.26 \text{ \AA}$ . For application of the new bootstrap method, the images were decimated four-fold, resulting in the window size  $75 \times 75$  with pixel size  $5.04 \text{ \AA}$ . We first applied the renormalization procedure to the entire available data set of 362,361 projection images; after three iterations, scaling factors stabilized and resolution improved from  $11.0 \text{ \AA}$  to  $10.6 \text{ \AA}$ . To make the analysis manageable and to avoid the bias of variance estimation by uneven distribution of projections, we randomly selected a subset of 21,000 images distributed as evenly as possible. The resolution of the resulting 3D map was  $16.0 \text{ \AA}$  (corresponding to spatial frequency  $0.0625 \text{ \AA}^{-1}$ ). Using this reduced data set, we generated 10,240 bootstrap volumes, applied to them low-pass filter with cutoff frequency  $0.04 \text{ \AA}^{-1}$ , and computed the variance map. This restrictive filtration was selected to reflect the fact that the number of unique projections within each bootstrap sample is less than 21,000, further reducing the resolution. However, bootstrap volumes still contained information beyond the first zero of the CTF, which for the processed data set was at  $0.03 \text{ \AA}^{-1}$  for the furthest defocus setting. On the basis of the histogram of the voxel values in the map and the average values of the density

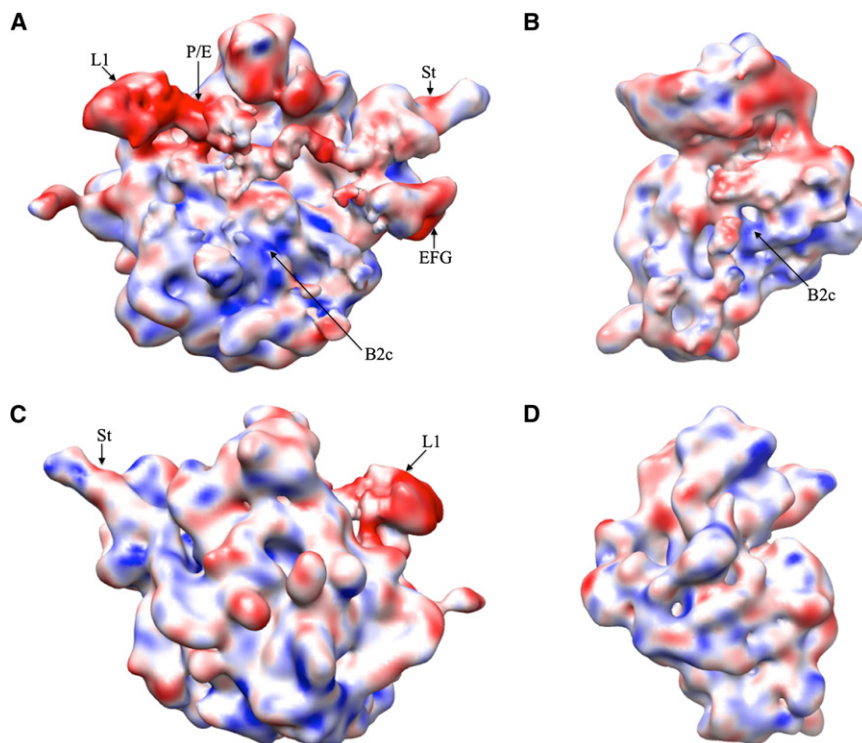
(A) Variance map of the test structures.

(B) Variance map obtained with CTF correction and filtered at  $0.12 \text{ \AA}^{-1}$ .

(C) Variance map obtained without CTF correction and filtered at  $0.12 \text{ \AA}^{-1}$ .

(D) Variance map obtained with CTF correction and volumes low-pass filtered at  $0.03 \text{ \AA}^{-1}$ .

(E) Variance map obtained without CTF correction and volumes low-pass filtered at  $0.03 \text{ \AA}^{-1}$ .



**Figure 2. Conformational Variability of the 70S•EF-G•GMPPNP Complex Revealed by the Real-Space Bootstrap Variance Analysis of the Reconstructed Cryo-EM Structure** (A–D) The surfaces are color-coded using the values of the bootstrap standard deviation ( $\sigma$ ) 3D map. The red region has the highest variability ( $\sigma > 0.4 \text{ g/cm}^3$ ), whereas the blue region has the lowest variability ( $\sigma < 0.1 \text{ g/cm}^3$ ). The panels show computationally isolated ribosomal subunits from the 3D cryo-EM reconstruction of (A) the large subunit shown from the interface side, (B) the small subunit shown from the interface side, (C) the large subunit shown from the solvent side, and (D) the small subunit shown from the solvent side. Landmarks for the subunits: L1, the L1 protuberance; P/E, the P-E-hybrid site tRNA; EFG, the Elongation Factor G; St, the L7/L12 stalk; B2c, inter-subunit bridge 2c (the center of the rotation).

of the vitreous ice ( $0.92 \text{ g/cm}^3$ ), protein ( $1.36 \text{ g/cm}^3$ ), and RNA ( $1.89 \text{ g/cm}^3$ ), we rescaled the map such that the units of voxel values became grams per cubic centimeter. To visualize the structure variability, we color-coded the surface of the cryo-EM map of the 70S•EF-G•GMPPNP complex, by the level of the standard deviation at a given location on the molecule's surface (Figures 2A–2D).

Regions of high variance can be immediately recognized by this visualization and include the densities corresponding to the EF-G and the P/E tRNA, which is expected as a result of the substoichiometric presence of these ligands. Interestingly, several regions of the ribosome itself also exhibit high variance and are in excellent agreement with known locations of conformational changes (Connell et al., 2007; Schuwirth et al., 2005; Selmer et al., 2006; Spahn et al., 2004; Valle et al., 2003). The L1-protuberance is a dynamic feature of the ribosome, and it moves inward to interact with an E or P/E tRNA. Indeed, strong variance is associated with both, the P/E tRNA and the L1 protuberance. Also, the well-established dynamic behavior of the extended L7/L12 stalk is reflected in our variance map.

The ratchet-like subunit rearrangement (RSR) of the ribosome constitutes a complex conformational change (Frank and Agrawal, 2000). It comprises of rotation of the ribosomal 30S subunit relative to the ribosomal 50S subunit and independent movement of the head of the 30S (Connell et al., 2007; Spahn et al., 2001; Valle et al., 2003). The RSR is facilitated by binding of several translation factors, such as EF-G/eEF2 30S (Connell et al., 2007; Spahn et al., 2004; Valle et al., 2003), IF2 (Allen et al., 2005), or RF3 (Gao et al., 2007), and therefore is expected to occur in the majority of the ribosomes, which carry EF-G•GMPPNP, but not in the 30%–40% vacant ribosomes. The dynamic behavior of the 30S head is immediately obvious

from the 3D variance map. Furthermore, the relative rotation of the two ribosomal subunits manifests itself in radial dependence of the variance: the larger the distance of a density element from the center of rotation, the larger should be the corresponding variance. Such dependence is observed in the 3D variance map; the outside regions of the ribosome generally show a larger variability than the inner core. Strikingly, the region with the lowest variance corresponds to the intersubunit bridge B2c, which has been previously identified to constitute the center of the RSR (Spahn et al., 2004).

The calculation was carried on a Linux cluster. For the data set considered here, it took 10 hr to generate 10,240 bootstrap volumes using 32 nodes of the cluster, while each node was equipped with 2 AMD Opteron 248 CPU (64 bits with 1M cache).

## DISCUSSION

The RSR is a global conformational change, but movements of the individual ribosomal elements, especially these close to the center of the rotation, are not very large. Moreover, in the present data set, pictures of ribosomes exhibiting the RSR and ribosomes in the standard conformation are mixed, which reduces the differences due to misalignment. This is probably the reason why, in our initial attempt to use the bootstrap method (Penczek et al., 2006), we could mainly detect the compositional heterogeneity and strong local changes at the L1 protuberance but not the RSR. In contrast, the RSR can be deduced from the variance map presented here. This clearly demonstrates that the cryo-EM-specific modifications of the bootstrap method introduced here result in more accurate variance map, which allows detecting conformational heterogeneity in the sample studied.

Currently, single-particle analysis tools for validation of structure determination are limited to those that assess self-consistency of image alignment (e.g., FSC). However, even if perfect image alignment could be realized, the resolution of a single-particle reconstruction would ultimately depend on the



composition and on conformational homogeneity of the complex under study. The potential to provide information about structural variability in a macromolecular complex represents one of the most promising, but still not fully realized, aspects of single-particle image analysis. To meet the challenge that structural variability presents to single-particle analysis, we introduced here a method for calculation of variance of the reconstructed structure based on statistical resampling. We demonstrated that the method yields detailed information about variability of the ribosomal complex and thus the bootstrap method is a uniquely valuable tool to provide insight into the mechanism and function of large molecular assemblies.

## EXPERIMENTAL PROCEDURES

### Normalization of Cryo-EM 2D Projection Data

The introduction of an additional step of normalization of the projection data represents an essential improvement of the presented implementation. In electron microscopy, imaging conditions are never exactly the same (e.g., variation of the dose), and even within the same micrograph field the background densities can vary by a significant margin as the result of uneven ice thickness and other factors. The normalization errors will result in a high level of background variance that will distort the true relationship between structure variability and background variability of the solvent. In the proposed renormalization scheme, we took advantage of the fact that the alignment procedures used to establish orientation parameters of projection data utilize correlation functions; thus, they are not adversely affected by the errors in normalization of projection data. Once the approximate projection directions are found, we propose to renormalize the data on the basis of reprojections of the current approximation of the structure according to following steps: (1) for each projection, the structure is reprojected using known orientation parameters; (2) CTF is applied to the projection data, and squared CTF is applied to the reprojection; (3) on the basis of the known defocus, the predefined spatial frequency range is identified, which encompasses the first maximum of the squared CTF, and the scaling factor between projection and reprojection is established using information within the selected frequency range; and (iv) after scaling factors to be applied to all projection data, a corrected structure is computed. The procedure is iterated until there is no further change in scaling factors. The proposed method performs very well for single-particle reconstruction applications, where projection are arbitrarily oriented, so there is always sufficient overlap between projections that have different directions. The algorithm usually converges in two or three steps.

### 3D Reconstruction Algorithm for the Bootstrap Technique

In principle, any reconstruction algorithm can be used within the bootstrap method. In practice, we found the direct Fourier inversion approach to be the best choice for our needs. Direct Fourier methods are based on the central section theorem (Bracewell, 1956), which states that the 2D Fourier transform of a projection of a 3D object is a central section of the 3D Fourier transform of this object. Therefore, a set of 2D Fourier transforms of the projections yields an approximation to 3D Fourier transform of the object and a subsequent numerical 3D inverse Fourier transform yields the structure in real space. Here, we present an improved version of the nearest neighbor (NN) direct inversion algorithm (Penczek et al., 2006). In the new version, we account for the nonuniform distribution of samples on a regular 3D grid and also incorporate corrections for the CTF of the microscope.

In the proposed NN direct inversion reconstruction algorithm, the 2D input projections are first padded with zeroes to four times the size and then 2D Fourier transformed. Next, the 2D samples are accumulated within the target 3D Fourier volume using simple NN interpolation. During the interpolation, CTF correction is applied using Wiener filter methodology. As the 2D projection samples are assigned to nodes of the 3D regular Fourier grid, the following formula is applied, for each Fourier voxel (Penczek et al., 1997):

$$F^k = \frac{\sum_i SSNR_i \cdot CTF_i \cdot G_i^k}{\sum_i SSNR_i \cdot CTF_i^2 + 1} \quad (3)$$

where  $G_i^k$  is the value of a Fourier pixel assigned to the  $k$ th voxel in the  $i$ th projection data,  $CTF_i$  is the value of the transfer function at this location of projection data and  $F^k$  is the CTF-corrected value of the  $k$ th voxel. Here,  $SSNR_i$  has two meanings: it is the SSNR of a Fourier pixel in  $i$ th projection data and it also prevents divisions by very small numbers in locations where all CTFs are almost zero.

After all 2D projections are accumulated in a 3D Fourier volume, a 3D weighting function is constructed and applied to individual voxels of 3D Fourier space in order to account for possible nonuniform distribution of samples. Here, for the sake of efficiency, we employed a concept of the "local density" of sampled points (Bracewell and Riddle, 1967) and designed a weighting function satisfying the following criteria:

1. For a given voxel, if all its neighbors are occupied, its weight is set to 1.0 (the minimum weight resulting from the procedure).
2. For a given voxel, if all its neighboring voxels are empty, the assigned weight has the maximum value.
3. Empty voxels located closer to the vacant voxel contribute more to its weight than those located further.

The proposed heuristic weighting function is:

$$W(x, y, z) = \frac{1}{1 - \alpha \sum_{i,j,k=-n}^n I(x+i, y+j, z+k) \exp[-(i+j+k)\beta]} \quad (4)$$

where

$$I(x, y, z) = \begin{cases} 1 & \text{if } F(x, y, z) = 0 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

$\beta$  and  $n$  are constants whose values (0.2 and 3, respectively) were adjusted such that the rotationally averaged power spectrum of the reconstructed structure matches, as closely as possible, the rotationally averaged power spectrum of the test structure from which the projection data were generated (Figure 3).  $\alpha$  is a constant whose value depends on parameters  $\beta$  and  $n$  and that is adjusted such that first two normalization criteria listed above are fulfilled. The inclusion of weights also improves the fidelity of the NN direct-inversion reconstruction, as demonstrated using the FSC technique (Figure 3).

The reconstruction algorithm developed is particularly well suited for the application in the bootstrap technique, as in our implementation where we precalculate 2D FFTs of real-space padded 2D projection data and store them on a computer disk. After resampling, the selected projections are inserted into a 3D Fourier volume, the weights (equation 4) are calculated and applied, the 3D inverse FFT is computed, and the region of interest is windowed out. Thus, a computationally intensive step of preparation of 2D projections is performed only once.

### Determination of the Number of Bootstrap Volumes Required

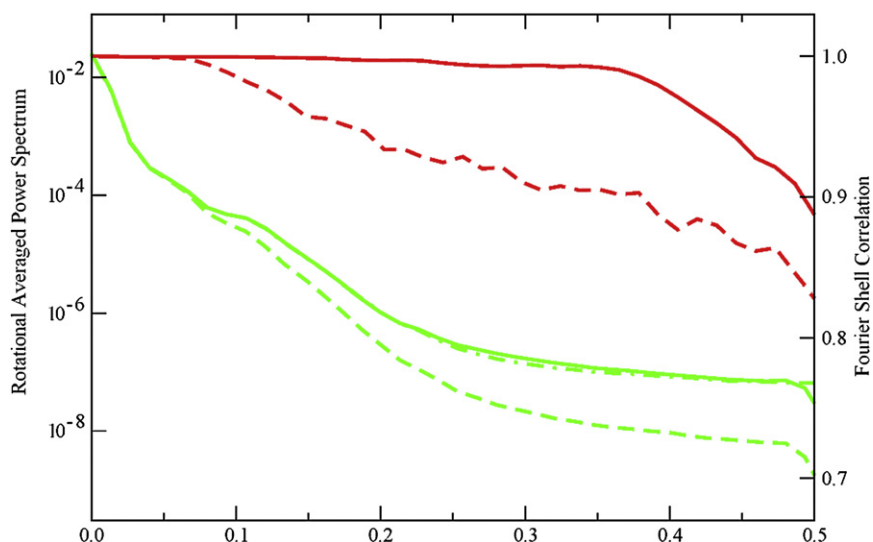
In order to determine whether the bootstrap method converges to an acceptable estimate of the variance with the increasing number of generated bootstrap volumes  $B$ , a method is needed to evaluate the reliability of the computed variance map. Here, we propose to use the sample correlation coefficient ( $\rho_B$ ) of two bootstrap variance volumes calculated from the set of bootstrap volumes randomly divided into halves.

In the Supplemental Data (available online), we show that the expectation value of  $\rho_B$  is:

$$E[\rho_B] \approx \left[ 1 + \frac{\frac{4}{B} \bar{m}_2^2}{S^2(\sigma_{\text{struct}}^2) + \frac{\bar{m}_4 - \bar{m}_2^2}{N}} \right]^{-1} \quad (6)$$

where  $\sigma_{\text{struct}}^2$  is given by equation (2),  $S^2(\sigma_{\text{struct}}^2)$  is calculated as the among-pixel sample variance of  $\sigma_{\text{struct}}^2$ , and  $\bar{m}_4$  and  $\bar{m}_2^2$  are second and fourth moments of the statistical distribution of the voxel values ( $\bar{m}_4 > \bar{m}_2^2 > 0$ ).

On the basis of equation (6), we conclude that  $\rho_B$  increases monotonically with  $B$ , with the rate of the increase decreasing with  $B$ . Therefore, a sensible criterion for termination of the bootstrap process should be based not so much on the attained value of  $\rho_B$  as on the rate of its increase. It also follows that, for a given  $B$ , increasing the size of the data set  $N$  results in a decrease of  $\rho_B$ . Therefore, in order to obtain an estimate of variance as reliable as that



**Figure 3. Improvement of the Fidelity of the Fourier NN Direct Inversion 3D Reconstruction Technique due to Inclusion of Weights Accounting for Uneven Distribution of Sampling Points in Fourier Space**

We generated a set of evenly spaced single axis-tilt projections of a model structure (a modified 70S from *Escherichia coli* placed within cube  $75^3$  voxels). To ensure that gaps existed between sampling points in Fourier space, we set the angular step to  $3.1^\circ$ . Red lines: FSCs between original and reconstructed volume without weights (dashed) and with weights (solid). Green lines: rotationally averaged power spectra of 3D reconstructions calculated without weights (dashed), with weights (solid), and of the model structure (dashed dotted).

obtained for a smaller set, it is necessary to compute a larger number of bootstrap samples.

The fact that calculation performed on a smaller data set converges in a smaller number of steps warrants explanation, as a naive conclusion would be that a smaller data set is preferable for bootstrap calculations. However, the variance estimated using bootstrap technique asymptotically approaches with  $B$  the variance of the finite sample from which the data is drawn—that is, the sample variance, not the variance of the original distribution. In order to improve the estimate of the latter, one has to increase the size of the sample  $N$ . This can be seen from the expression for the correlation coefficient ( $\rho_\gamma$ ) of the variance of original distribution (which in our case is called the variance of structure  $\sigma_{\text{struct}}^2$ ) and the bootstrap variance. The expectation of  $\rho_\gamma$  equals:

$$E(\rho_\gamma) \cong \sqrt{\frac{S^2(\sigma_{\text{struct}}^2)}{S^2(\sigma_{\text{struct}}^2) + \frac{\bar{m}_4 - \bar{m}_2^2}{N} + \frac{2\bar{m}_2^2}{B}}} \quad (7)$$

For derivation, see the [Supplemental Data](#). As expected,  $\rho_\gamma$  increases monotonically with  $N$ , which means that the accuracy of the bootstrap variance increases with  $N$ . Also, for a finite  $N$ ,  $\rho_\gamma$  is always less than 1, which means that for a finite sample one cannot obtain error-free variance of the statistical distribution from which the input data (projections) originated, no matter how many bootstrap volumes  $B$  calculated. In conclusion, a larger data set yields a more accurate result and is preferable for bootstrap calculation.

The bootstrap method described here was implemented and parallelized in the single particle reconstruction software package SPARX (Hohn et al., 2007).

## SUPPLEMENTAL DATA

Supplemental Data include Supplemental Experimental Procedures and can be found with this article online at [http://www.cell.com/structure/supplemental/S0969-2126\(08\)00416-4](http://www.cell.com/structure/supplemental/S0969-2126(08)00416-4).

## ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (grant R01 GM 60635 to P.A.P.), the DFG (grants SFB 740 TP A3 and TP Z1 to C.M.T.S.), by the European Union 3D-EM Network of Excellence, and by the European Union and Senatsverwaltung für Wissenschaft, Forschung und Kultur Berlin (UltraStructureNetwork, Anwenderzentrum to C.M.T.S.).

Received: August 27, 2008

Accepted: October 8, 2008

Published: December 9, 2008

## REFERENCES

- Allen, G.S., Zavialov, A., Gursky, R., Ehrenberg, M., and Frank, J. (2005). The cryo-EM structure of a translation initiation complex from *Escherichia coli*. *Cell* 121, 703–712.
- Bracewell, R.N. (1956). Strip integration in radio astronomy. *Aust. J. Phys.* 9, 198–217.
- Bracewell, R.N., and Riddle, A.C. (1967). Inversion of fan-beam scans in radio astronomy. *Astrophys. J.* 150, 427–434.
- Connell, S.R., Takemoto, C., Wilson, D.N., Wang, H., Murayama, K., Terada, T., Shirouzu, M., Rost, M., Schuler, M., Giesebrecht, J., et al. (2007). Structural basis for interaction of the ribosome with the switch regions of GTP-bound elongation factors. *Mol. Cell* 25, 751–764.
- Davison, A.C., and Hinkley, D.V. (1997). *Bootstrap methods and their application* (New York: Cambridge University Press).
- Frank, J., and Agrawal, R.K. (2000). A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature* 406, 318–322.
- Frank, J., Verschoor, A., and Boublik, M. (1981). Computer averaging of electron micrographs of 40S ribosomal subunits. *Science* 214, 1353–1355.
- Gao, H., Zhou, Z., Rawat, U., Huang, C., Bouakaz, L., Wang, C., Cheng, Z., Liu, Y., Zavialov, A., Gursky, R., et al. (2007). RF3 induces ribosomal conformational changes responsible for dissociation of class I release factors. *Cell* 129, 929–941.
- Gilbert, R.J.C., Gordiyenko, Y., von der Haar, T., Sonnen, A.F.-P., Hofmann, G., Nardelli, M., Stuart, D.I., and McCarthy, J.E.G. (2007). Reconfiguration of yeast 40S ribosomal subunit domains by the translation initiation multifactor complex. *Proc. Natl. Acad. Sci. USA* 104, 5788–5793.
- Grob, P., Cruse, M.J., Inouye, C., Peris, M., Penczek, P.A., Tjian, R., and Nogales, E. (2006). Cryo-electron microscopy studies of human TFIIID: conformational breathing in the integration of gene regulatory cues. *Structure* 14, 511–520.
- Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953). *Sample survey methods and theory, Volume 2* (New York: Wiley).
- Harauz, G., and van Heel, M. (1986). Exact filters for general geometry three dimensional reconstruction. *Optik* 73, 146–156.
- Hohn, M., Tang, G., Goodyear, G., Baldwin, P.R., Huang, Z., Penczek, P.A., Yang, C., Glaeser, R.M., Adams, P.D., and Ludtke, S.J. (2007). SPARX, a new environment for Cryo-EM image processing. *J. Struct. Biol.* 157, 47–55.
- Ludtke, S.J., Baker, M.L., Chen, D.H., Song, J.L., Chuang, D.T., and Chiu, W. (2008). De novo backbone trace of GroEL from single particle electron cryomicroscopy. *Structure* 16, 441–448.

- Ohi, M.D., Feoktistova, A., Ren, L.P., Yip, C., Cheng, Y.F., Chen, J.S., Yoon, H.J., Wall, J.S., Huang, Z., Penczek, P.A., et al. (2007). Structural organization of the anaphase-promoting complex bound to the mitotic activator Slp1. *Mol. Cell* 28, 871–885.
- Penczek, P., Radermacher, M., and Frank, J. (1992). Three-dimensional reconstruction of single particles embedded in ice. *Ultramicroscopy* 40, 33–53.
- Penczek, P.A. (2002). Three-dimensional spectral signal-to-noise ratio for a class of reconstruction algorithms. *J. Struct. Biol.* 138, 34–46.
- Penczek, P.A., Chao, Y., Frank, J., and Spahn, C.M.T. (2006). Estimation of variance in single particle reconstruction using the bootstrap technique. *J. Struct. Biol.* 154, 168–183.
- Penczek, P.A., Grassucci, R.A., and Frank, J. (1994). The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy* 53, 251–270.
- Penczek, P.A., Zhu, J., Schröder, R., and Frank, J. (1997). Three-dimensional reconstruction with contrast transfer function compensation from defocus series. *Scanning Microsc. Suppl.* 11, 147–154.
- Penczek, P.A., Renka, R., and Schomberg, H. (2004). Gridding-based direct Fourier inversion of the three-dimensional ray transform. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 21, 499–509.
- Radermacher, M. (1992). Weighted back-projection methods. In *Electron Tomography*, J. Frank, ed. (New York: Plenum), pp. 91–115.
- Rivera-Calzada, A., Spagnolo, L., Pearl, L.H., and Llorca, O. (2007). Structural model of full-length human Ku70-Ku80 heterodimer and its recognition of DNA and DNA-PKcs. *EMBO Rep.* 8, 56–62.
- Saxton, W.O., and Baumeister, W. (1982). The correlation averaging of a regularly arranged bacterial envelope protein. *J. Microsc.* 127, 127–138.
- Schüler, M., Connell, S.R., Lescoute, A., Giesebrecht, J., Dabrowski, M., Schroeer, B., Mielke, T., Penczek, P.A., Westhof, E., and Spahn, C.M.T. (2006). Structure of the ribosome-bound cricket paralysis virus IRES RNA. *Nat. Struct. Mol. Biol.* 13, 1092–1096.
- Schuwirth, B.S., Borovinskaya, M.A., Hau, C.W., Zhang, W., Vila-Sanjurjo, A., Holton, J.M., and Cate, J.H. (2005). Structures of the bacterial ribosome at 3.5 Å resolution. *Science* 310, 827–834.
- Selmer, M., Dunham, C.M., Murphy, F.V., Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R., and Ramakrishnan, V. (2006). Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* 313, 1935–1942.
- Spahn, C.M., Gomez-Lorenzo, M.G., Grassucci, R., Jorgensen, R., Andersen, G.R., Beckmann, R., Penczek, P.A., Ballesta, J.P., and Frank, J. (2004). Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. *EMBO J.* 23, 1008–1019.
- Spahn, C.M.T., Blaha, G., Agrawal, R.K., Penczek, P., Grassucci, R.A., Trieber, C.A., Connell, S.R., Taylor, D.E., Nierhaus, K.H., and Frank, J. (2001). Localization of the ribosomal protection protein Tet(O) on the ribosome and the mechanism of tetracycline resistance. *Mol. Cell* 7, 1037–1045.
- Unser, M., Trus, B.L., and Steven, A.C. (1987). A new resolution criterion based on spectral signal-to-noise ratios. *Ultramicroscopy* 23, 39–51.
- Valle, M., Zavialov, A., Sengupta, J., Rawat, U., Ehrenberg, M., and Frank, J. (2003). Locking and unlocking of ribosomal motions. *Cell* 114, 123–134.
- van Heel, M., Gowen, B., Matadeen, R., Orlova, E.V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R., Schatz, M., and Patwardhan, A. (2000). Single-particle electron cryo-microscopy: towards atomic resolution. *Q. Rev. Biophys.* 33, 307–369.