



Published in final edited form as:

Structure. 2008 December 10; 16(12): 1770–1776. doi:10.1016/j.str.2008.10.011.

Heterogeneity of large macromolecular complexes revealed by 3-D cryo-EM variance analysis

Wei Zhang^a, Marek Kimmel^b, Christian M.T. Spahn^c, and Pawel A. Penczek^{a,*}

^aThe University of Texas – Houston Medical School, Department of Biochemistry and Molecular Biology, 6431 Fannin, MSB 6.218, Houston, TX 77030, USA

^bRice University, Department of Statistics, 6100 Main St., MS 138, Houston, TX 77005, USA

^cInstitut für Medizinische Physik und Biophysik, Charité - Universitätsmedizin Berlin, Ziegelstr. 5-9, 10117 Berlin, Germany

Abstract

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 datasets it becomes apparent that many complexes coexist in a mixture of conformational states or contain flexible regions. As the cryo-EM data is collected in form of projections of imaged molecules, the information about variability of reconstructed density maps is not directly available. To address this problem, we describe a new 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 introduced a novel highly efficient reconstruction algorithm that is based on direct Fourier inversion and which 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 dataset of 70S ribosome complexed with tRNA and the elongation factor G. The variance map revealed regions of high variability: the L1 protein, the EF-G and the 30S head and the ratchet-like subunit rearrangement. The proposed method of variance estimation opens new possibilities for single particle analysis, by extending applicability of the technique to heterogeneous datasets of macromolecules, and to complexes with significant conformational variability.

Keywords

single particle reconstruction; bootstrap; electron microscopy; ribosome; elongation cycle

1. 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

*Correspondence e-mail: Pawel.A.Penczek@uth.tmc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

obtained using the electron microscope are 2-D parallel beam projections of the same 3-D object with unknown orientation. After selection of projection images from electron micrographs the orientation parameters are determined using alignment procedures [1-3] and the 3-D density distribution is calculated using a 3-D reconstruction algorithm [4-6]. Due to 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 ratio (SNR) (<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 3-D space results in detailed maps of complexes studied, at a resolution reaching $\sim 5 \text{ \AA}$, thus allowing backbone tracing [7,8]. The quality of 3-D 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 [9]: (1) additive background noise unrelated to the data that originates from the solvent or the background carbon film, (2) alignment errors, (3) noise due to conformational variability of the specimen imaged, or non-stoichiometry of ligand binding. Whereas the presence of the first category noise manifests itself in a uniform variance in the reconstructed map, the other two yield non-uniform and structure-dependent distribution of the variance.

Traditionally, the quality of EM maps is evaluated using Fourier techniques [10,11]. The 'resolution' of a EM map is determined by randomly splitting the available set of 3-D projection images into halves, calculating a 3-D 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) [12, 13]. 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 [13].

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 3-D map which, if available, is helpful in (i) assessment of non-locality of errors, due to alignment errors, conformational heterogeneity of complexes or substoichiometric ligand binding, (ii) determination of reliability of small features in the map which can be artifacts induced by alignment procedures, (iii) 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 3-D reconstruction from the set of 2-D projections, based on very simple premises [9]. In this latter work, we assume that voxels in the reconstructed 3-D map can be considered weighted sums of pixels in 2-D projections. If the number of 2-D projections is large, the estimates of variances and covariances can be calculated using a variant of the bootstrap technique [14]. Specifically, a new set of N 2-D 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, other will be omitted. This selection process is repeated B times and for each new set of projections a corresponding 3-D bootstrap volume is calculated. The voxel-by-voxel (bootstrap) variance S^{*2} of these volumes is calculated yielding an estimate of the distribution variance σ^2 [15]:

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

In the extensive set of tests included in [9] we demonstrated that the bootstrap variance (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 \overline{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:

$$E\left(\mathbf{S}^{*2} - \overline{S}_{Back}^2\right) = \frac{\sigma_{Struct}^2}{N}. \quad (2)$$

In the absence of a convincing method for distinguishing the variance arising from alignment errors, σ_{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 3-D variance in real space [9] has been used for several systems, including the 70S ribosome [9], a transcription factor [16], the Ku70–Ku80 dimer [17], the 40S ribosomal subunit [18], and the anaphase-promoting complex/cyclosome [19], and has been shown useful in detecting sample heterogeneity. However, the method is far from becoming routine in cryo-EM structural studies, as 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 3-D 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 over-interpretation 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 3-D reconstruction algorithm we included correction for the CTF of the microscope and we added compensation for the uneven distribution of sampling points. 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 *T. thermophilus* 70S ribosome [20]. In a previous study, the 3D variance map for a similar ribosomal complex obtained by the old method resulted in high 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.

2. Methods

2.1 3-D 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 [21] which states the 2-D Fourier transform of a projection of a 3-D object is a central section of the 3-D Fourier transform of this object. Therefore, a set of 2-D Fourier transforms of the projections yields an approximation to 3-D Fourier transform of the object and a subsequent numerical 3-D inverse Fourier transform yields the structure in real space. Here we present an improved version of

the nearest neighbor (NN) direct inversion algorithm [9]. In the new version, we account for the non-uniform distribution of samples on a regular 3-D grid and also incorporate corrections for the contrast transfer function (CTF) of the microscope.

In the proposed NN direct inversion reconstruction algorithm, the 2-D input projections are first padded with zeroes to four times the size and then 2-D Fourier transformed. Next, the 2-D samples are accumulated within the target 3-D Fourier volume using simple NN interpolation. During the interpolation, CTF correction is applied using Wiener filter methodology. As the 2-D projection samples are assigned to nodes of the 3-D regular Fourier grid, the following formula is applied, for each Fourier voxel [22]:

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

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 2-D projections are accumulated in a 3-D Fourier volume, a 3-D weighting function is constructed and applied to individual voxels of 3-D 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 [23] 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:

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

where

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

β 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 (Fig.1). α is a constant whose value depends on parameters β and n and which is adjusted such that first two normalization criteria listed above are fulfilled. The final step of the reconstruction algorithm entails windowing out, in real space, the relevant section of the volume. The inclusion of weights improves the fidelity of the NN direct-inversion reconstruction, as demonstrated using the FSC technique (Fig.1).

The reconstruction algorithm developed is particularly well suited for the application in the bootstrap technique, as in our implementation we pre-calculate 2-D FFTs of real-space padded 2-D projection data and store them on a computer disk. After resampling, the selected projections are inserted into a 3-D Fourier volume, the weights (4) are calculated and applied, the 3-D inverse FFT is computed and the region of interest is windowed out. Thus, a computationally intensive step of preparation of 2-D projections is performed only once. In effect, on an AMD Opteron 2.4GHz processor computation of one bootstrap volume using 21,000 projections sized 75^2 pixels requires 235 seconds.

2.2 Normalization of cryo-EM 2-D 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 due to uneven ice thickness and other factors. The normalization errors will result in a high level of background variance that will distort the true relation 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 based on reprojections of the current approximation of the structure according to following steps: (i) for each projection, the structure is reprojected using known orientation parameters, (ii) CTF is applied to the projection data and squared CTF is applied to the reprojection, (iii) based on the known defocus a 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, (iv) after scaling factors to 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.

2.3 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 (ρ_B) of two bootstrap variance volumes calculated from the set of bootstrap volumes randomly divided into halves.

In the Supplemental Material A we show that the expectation value of ρ_B is:

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

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

Based on (6) we conclude that ρ_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 ρ_B as on the rate of its increase. It also follows that, for a given B , increasing the size of the dataset N results in a decrease of ρ_B . Therefore, in order to obtain an estimate of variance as reliable as that 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, i.e., 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 (ρ_γ) of the variance of original distribution (which in our case is called the variance of structure σ_{struct}^2) and the bootstrap variance. The expectation of ρ_γ equals:

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

(for derivation see Supplemental Material A). As expected, ρ_γ increases monotonically with N , which means that the accuracy of the bootstrap variance increases with N . Also, for a finite N ρ_γ 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 [24].

3. Results

First we provide test results to illustrate improvement in the resolution of the variance map for a 3-D structure calculated from its electron microscope projections that is due to inclusion of the CTF correction. We construct a model case and conduct bootstrap variance calculations with and without CTF correction and compare the results. Second, we provide results of the application of the bootstrap variance calculation method to the cryo-EM dataset of 70S ribosome. We demonstrate that increased stability of the variance map yields unique insight into conformational variability of the complex.

3.1 The bootstrap variance calculation with CTF correction

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 753 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 we generated computationally its 2-D projection using Eulerian angles of selected projection direction. Finally, we modified all 2-D projections by CTFs assuming physical pixel size of 4.88 Å, microscope voltage 300 kV, spherical aberration 2.0 mm, amplitude contrast 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 3-D variance maps for our set of simulated 2-D projections: one with CTF correction included in the 3-D 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 Å⁻¹ and 0.03 Å⁻¹ (for reference the first zero of CTF appears at frequency 0.035 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 ~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 Å⁻¹ (Fig.2-b) reproduces almost perfectly the variance of original structures (Fig.2-a) with the expected loss of resolution that is due to low-pass filtration of bootstrap volumes that 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 Å⁻¹ (Fig.2-c) contains very strong artifacts that would make the detection of “true” variance regions impossible. The strong artifacts can be suppressed by applying very strong low-pass filtration 0.03 Å⁻¹ (Fig.2-e), however under such a strong low-pass filtration, the two points very close to each other become indistinguishable.

3.2 Application of the bootstrap variance calculation method to the 70S ribosome cryo-EM dataset

The variance analysis was applied to a data set of the *Thermus thermophilus* 70S ribosome complexed with the elongation factor EF-G. EF-G was stalled using the non-hydrolyzable GTP-analogue GMPPNP. The complex also contains endogeneous tRNA in the P/E site [20]. The occupancy of ribosomes by EF-G has been estimated to be 60-70% based on 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 Å) EF-G was directly visible in a structure obtained from the complete data set. The subsequent multi-reference alignment revealed a subset of EF-G-containing particles that yielded a 7.3 Å structure [20]. 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 dataset 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 300KV at a defocus range from 2.5 μm to 4.7 μm [20]. The original window size was 300 × 300 pixels and pixel size was 1.26 Å. For application of the new bootstrap method, the images were decimated fourfold resulting in the window size 75×75 with pixel

size 5.04 Å. We first applied the renormalization procedure to the entire available dataset of 362,361 projection images; after three iterations scaling factors stabilized and resolution improved from 11.0 Å to 10.6 Å. 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 3-D map was 16.0 Å (corresponding to spatial frequency 0.0625 Å⁻¹). Using this reduced dataset we generated 10,240 bootstrap volumes, applied to them low-pass filter with cutoff frequency 0.04 Å⁻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 dataset was equal to 0.03 Å⁻¹ for the furthest defocus setting. Based on the histogram of the voxel values in the map and on the average values of the density of the vitreous ice (0.92 g/cm³), protein (1.36 g/cm³), and RNA (1.89 g/cm³), we rescaled the map such that the units of voxel values became g/cm³. In order 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 (Fig.3a-d).

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 due to the sub-stoichiometric presence of these ligands. Interestingly, also several regions of the ribosome itself exhibit high variance and these locations are in excellent agreement with known locations of conformational changes [20,25-28]. 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 [29]. It comprises of rotation of the ribosomal 30S subunit relative to the ribosomal 50S subunit and independent movement of the head of the 30S [20,28,30]. The RSR is facilitated by binding of several translation factors such as EF-G/eEF2 30S [20,27, 28], IF2 [31] or RF3 [32] 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 inter-subunit bridge B2c, which has been previously identified to constitute the center of the RSR [27].

The calculation was carried on a Linux cluster. For the dataset considered here, it takes 10 hours to generate 10,240 bootstrap volumes using 64 CPUs of the cluster.

4. Discussion

We have developed a new version of the bootstrap method of estimating variance in a 3-D density map of a macromolecular complex reconstructed from a set of its cryo-EM projections. We built on the foundation of the general principle introduced in [9] and we introduced improvements necessary for the method to account for peculiarities of cryo-EM data and thus fully realize the potential of the method. This was achieved by the addition to the direct Fourier inversion reconstruction algorithm of the correction for the CTF of the microscopy and by accounting for highly nonuniform coverage of 3-D Fourier space by sampling points, as

typically encountered in EM data. These improvements eliminated computational limitations to the resolution of variance maps that can be calculated for cryo-EM data; indeed, as demonstrated in tests on simulated data the resolvability of the variance map can reach the theoretical limit of the reconstruction algorithm. The new version of the algorithm was implemented on distributed memory multiprocessing clusters of computers, allowing for rapid calculation of large number of resampled volumes, as required by the bootstrap technique. A major contribution is establishment of the method for the determination of the number of bootstrap volumes that have to be computed in order to obtain a reliable variance map. We demonstrated that the accuracy of the variance map increases monotonically with the number of bootstrap volumes, but the rate of improvement decreases, which yields a practical criterion for determining their acceptable minimal number. We also demonstrated that increasing the number of bootstrap volumes will not result in the variance of the distribution that governs variability of studied macromolecules, but merely in the variance of the final sample as given by the available set of projection images. Thus, a more accurate variance map can be obtained only by increasing the number of cryo-EM projection images, which in turn calls for increased number of bootstrap volumes that have to be computed.

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 [9] 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 this 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by grants from the NIH R01 GM 60635 (to PAP), the DFG (SFB 740 TP A3 and TP Z1 to CMTS), 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 CMTS).

References

1. Penczek P, Radermacher M, Frank J. Three-dimensional reconstruction of single particles embedded in ice. *Ultramicroscopy* 1992;40:33–53. [PubMed: 1580010]

2. Penczek PA, Grassucci RA, Frank J. The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy* 1994;53:251–270. [PubMed: 8160308]
3. van Heel M, Gowen B, Matadeen R, Orlova EV, Finn R, Pape T, Cohen D, Stark H, Schmidt R, Schatz M, Patwardhan A. Single-particle electron cryo-microscopy: towards atomic resolution. *Quarterly Reviews of Biophysics* 2000;33:307–369. [PubMed: 11233408]
4. Harauz G, van Heel M. Exact filters for general geometry three dimensional reconstruction. *Optik* 1986;73:146–156.
5. Radermacher, M. Weighted back-projection methods. In: Frank, J., editor. *Electron Tomography*. New York: Plenum; 1992. p. 91-115.
6. Penczek PA, Renka R, Schomberg H. Gridding-based direct Fourier inversion of the three-dimensional ray transform. *Journal of the Optical Society of America A* 2004;21:499–509.
7. Schüler M, Connell SR, Lescoute A, Giesebrecht J, Dabrowski M, Schroeer B, Mielke T, Penczek PA, Westhof E, Spahn CMT. Structure of the ribosome-bound cricket paralysis virus IRES RNA. *Nature Structural and Molecular Biology* 2006;13:1092–1096.
8. Ludtke SJ, Baker ML, Chen DH, Song JL, Chuang DT, Chiu W. De novo backbone trace of GroEL from single particle electron cryomicroscopy. *Structure* 2008;16:441–448. [PubMed: 18334219]
9. Penczek PA, Chao Y, Frank J, Spahn CMT. Estimation of variance in single particle reconstruction using the bootstrap technique. *Journal of Structural Biology* 2006;154:168–183. [PubMed: 16510296]
10. Saxton WO, Baumeister W. The correlation averaging of a regularly arranged bacterial envelope protein. *Journal of Microscopy* 1982;127:127–138. [PubMed: 7120365]
11. Frank J, Verschoor A, Boublik M. Computer averaging of electron micrographs of 40S ribosomal subunits. *Science* 1981;214:1353–1355. [PubMed: 7313694]
12. Unser M, Trus BL, Steven AC. A new resolution criterion based on spectral signal-to-noise ratios. *Ultramicroscopy* 1987;23:39–51. [PubMed: 3660491]
13. Penczek PA. Three-dimensional Spectral Signal-to-Noise Ratio for a class of reconstruction algorithms. *Journal of Structural Biology* 2002;138:34–46. [PubMed: 12160699]
14. Davison, AC.; Hinkley, DV. *Bootstrap methods and their application*. New York: Cambridge University Press; 1997.
15. Hansen, MH.; Hurwitz, WN.; Madow, WG. *Sample survey methods and theory*. Vol. 2. New York: Wiley; 1953.
16. Grob P, Cruse MJ, Inouye C, Peris M, Penczek PA, Tjian R, Nogales E. Cryo-electron microscopy studies of human TFIID: Conformational breathing in the integration of gene regulatory cues. *Structure* 2006;14:511–520. [PubMed: 16531235]
17. Rivera-Calzada A, Spagnolo L, Pearl LH, Llorca O. Structural model of full-length human Ku70-Ku80 heterodimer and its recognition of DNA and DNA-PKcs. *EMBO Reports* 2007;8:56–62. [PubMed: 17159921]
18. Gilbert RJC, Gordiyenko Y, von der Haar T, Sonnen AF-P, Hofmann G, Nardelli M, Stuart DI, McCarthy JEG. Reconfiguration of yeast 40S ribosomal subunit domains by the translation initiation multifactor complex. *Proceedings of the National Academy of Sciences of the United States of America* 2007;104:5788–5793. [PubMed: 17389391]
19. Ohi MD, Feoktistova A, Ren LP, Yip C, Cheng YF, Chen JS, Yoon HJ, Wall JS, Huang Z, Penczek PA, Gould KL, Walz T. Structural organization of the anaphase-promoting complex bound to the mitotic activator Slp1. *Molecular Cell* 2007;28:871–885. [PubMed: 18082611]
20. Connell SR, Takemoto C, Wilson DN, Wang H, Murayama K, Terada T, Shirouzu M, Rost M, Schuler M, Giesebrecht J, Dabrowski M, Mielke T, Fucini P, Yokoyama S, Spahn CM. Structural basis for interaction of the ribosome with the switch regions of GTP-bound elongation factors. *Molecular Cell* 2007;25:751–764. [PubMed: 17349960]
21. Bracewell RN. Strip integration in radio astronomy. *Austr J Phys* 1956;9:198–217.
22. Penczek PA, Zhu J, Schröder R, Frank J. Three-dimensional reconstruction with contrast transfer function compensation from defocus series. *Scanning Microscopy Supplement* 1997;11:1–10.
23. Bracewell RN, Riddle AC. Inversion of fan-beam scans in radio astronomy. *Astrophysical Journal* 1967;150:427–434.

24. Hohn M, Tang G, Goodyear G, Baldwin PR, Huang Z, Penczek PA, Yang C, Glaeser RM, Adams PD, Ludtke SJ. SPARX, a new environment for Cryo-EM image processing. *Journal of Structural Biology* 2007;157:47–55. [PubMed: 16931051]
25. Schuwirth BS, Borovinskaya MA, Hau CW, Zhang W, Vila-Sanjurjo A, Holton JM, Cate JH. Structures of the bacterial ribosome at 3.5 Å resolution. *Science* 2005;310:827–834. [PubMed: 16272117]
26. Selmer M, Dunham CM, Murphy FV, Weixlbaumer A, Petry S, Kelley AC, Weir JR, Ramakrishnan V. Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* 2006;313:1935–1942. [PubMed: 16959973]
27. Spahn CM, Gomez-Lorenzo MG, Grassucci R, Jorgensen R, Andersen GR, Beckmann R, Penczek PA, Ballesta JP, Frank J. Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. *EMBO Journal* 2004;23:1008–1019. [PubMed: 14976550]
28. Valle M, Zavialov A, Sengupta J, Rawat U, Ehrenberg M, Frank J. Locking and unlocking of ribosomal motions. *Cell* 2003;114:123–134. [PubMed: 12859903]
29. Frank J, Agrawal RK. A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature* 2000;406:318–322. [PubMed: 10917535]
30. Spahn CMT, Blaha G, Agrawal RK, Penczek P, Grassucci RA, Trieber CA, Connell SR, Taylor DE, Nierhaus KH, Frank J. Localization of the ribosomal protection protein Tet(O) on the ribosome and the mechanism of tetracycline resistance. *Molecular Cell* 2001;7:1037–1045. [PubMed: 11389850]
31. Allen GS, Zavialov A, Gursky R, Ehrenberg M, Frank J. The cryo-EM structure of a translation initiation complex from *Escherichia coli*. *Cell* 2005;121:703–712. [PubMed: 15935757]
32. Gao H, Zhou Z, Rawat U, Huang C, Bouakaz L, Wang C, Cheng Z, Liu Y, Zavialov A, Gursky R, Sanyal S, Ehrenberg M, Frank J, Song H. RF3 induces ribosomal conformational changes responsible for dissociation of class I release factors. *Cell* 2007;129:929–941. [PubMed: 17540173]

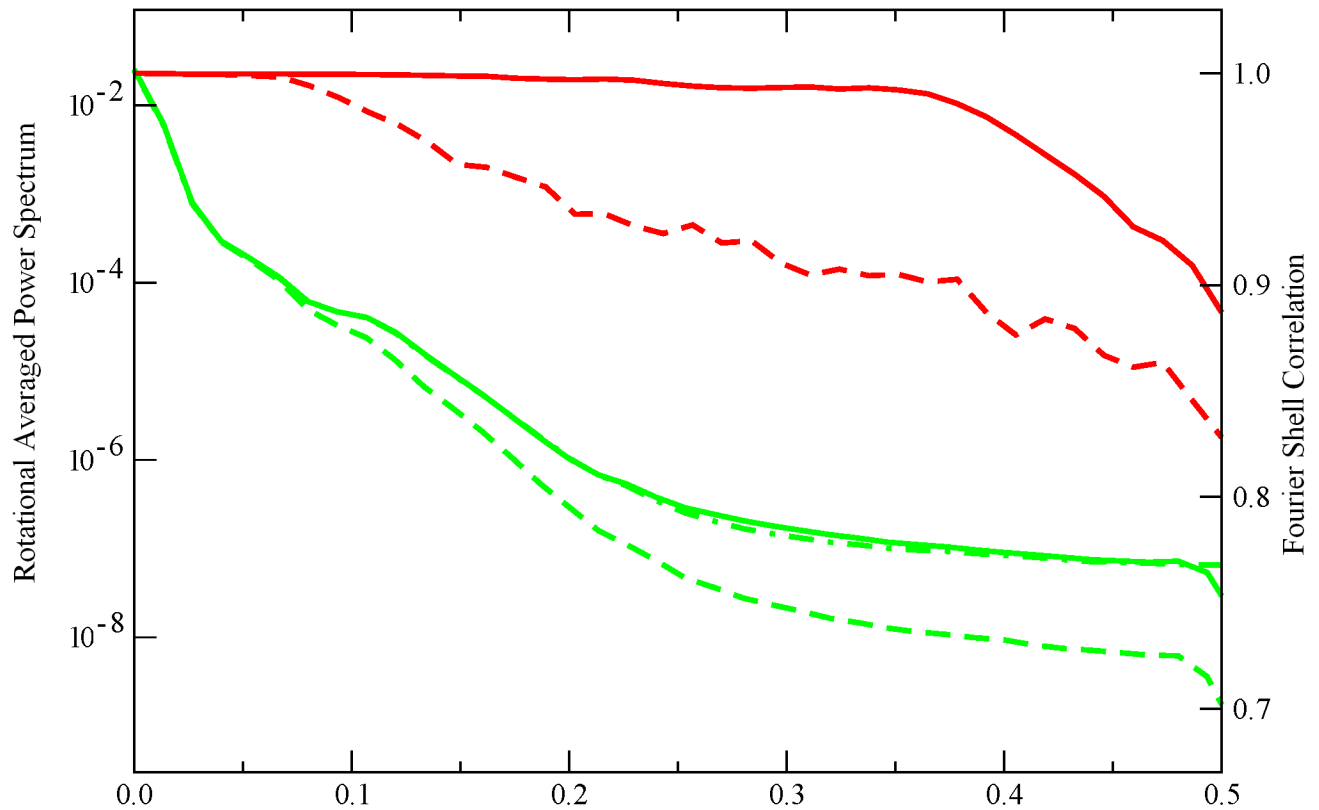


Figure 1.

Improvement of the fidelity of the Fourier NN direct inversion 3-D 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 *E. 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° . 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).

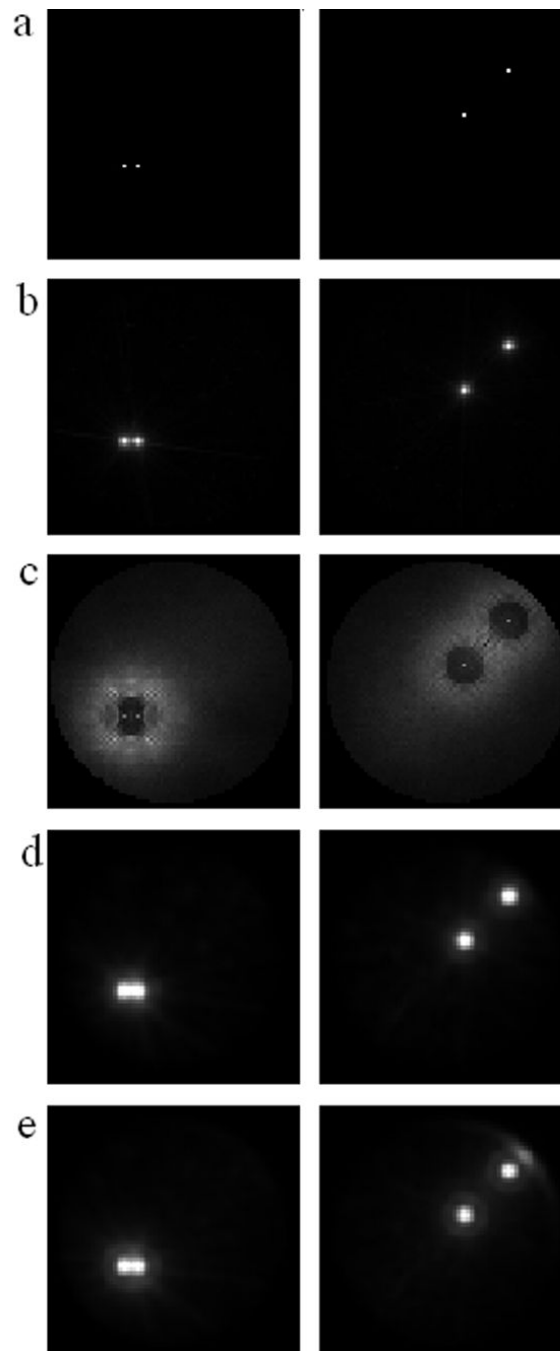
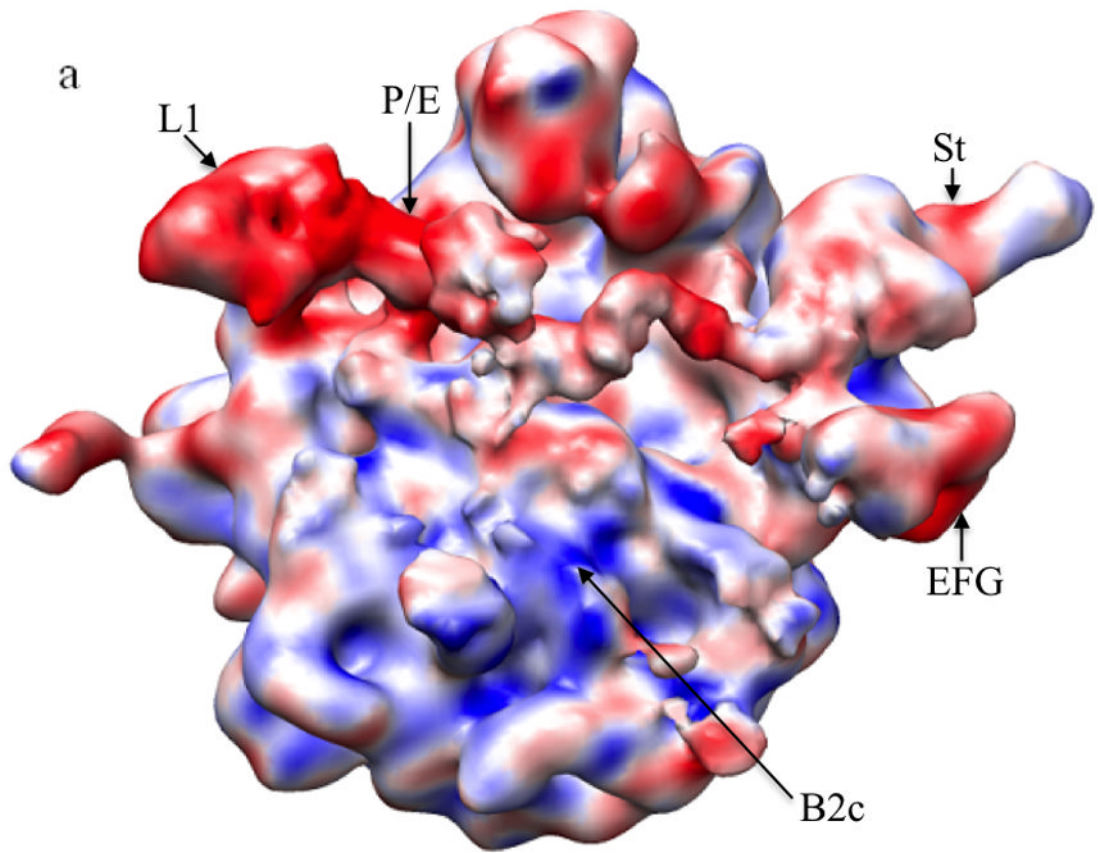
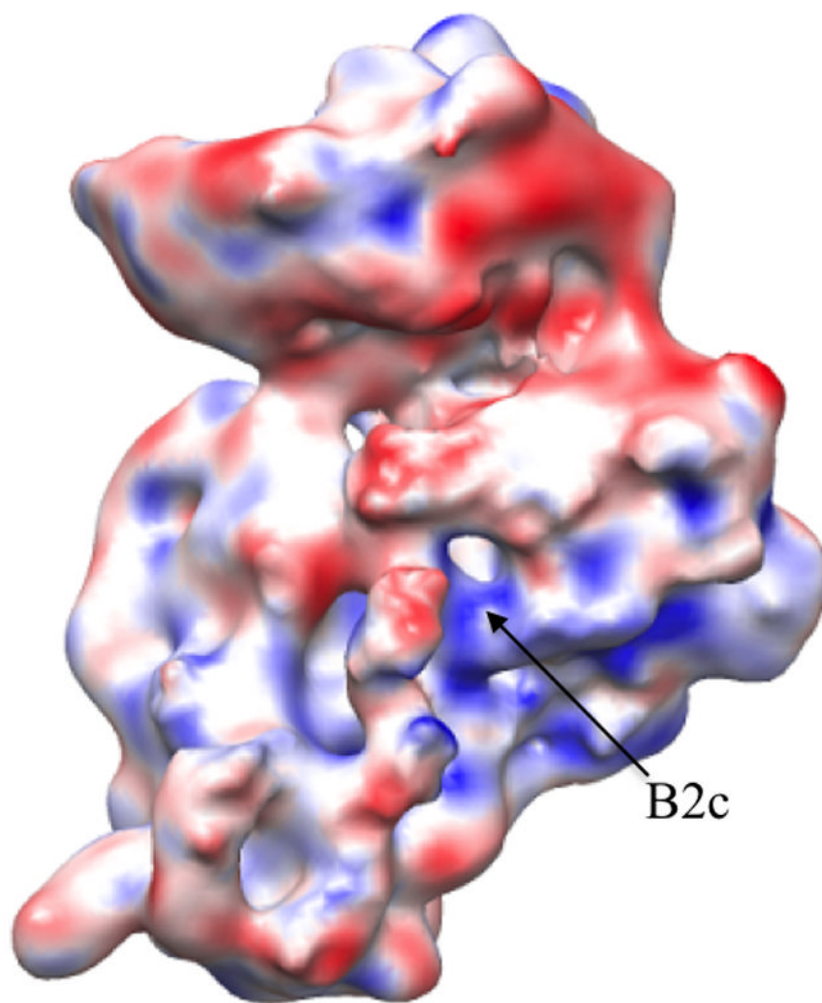


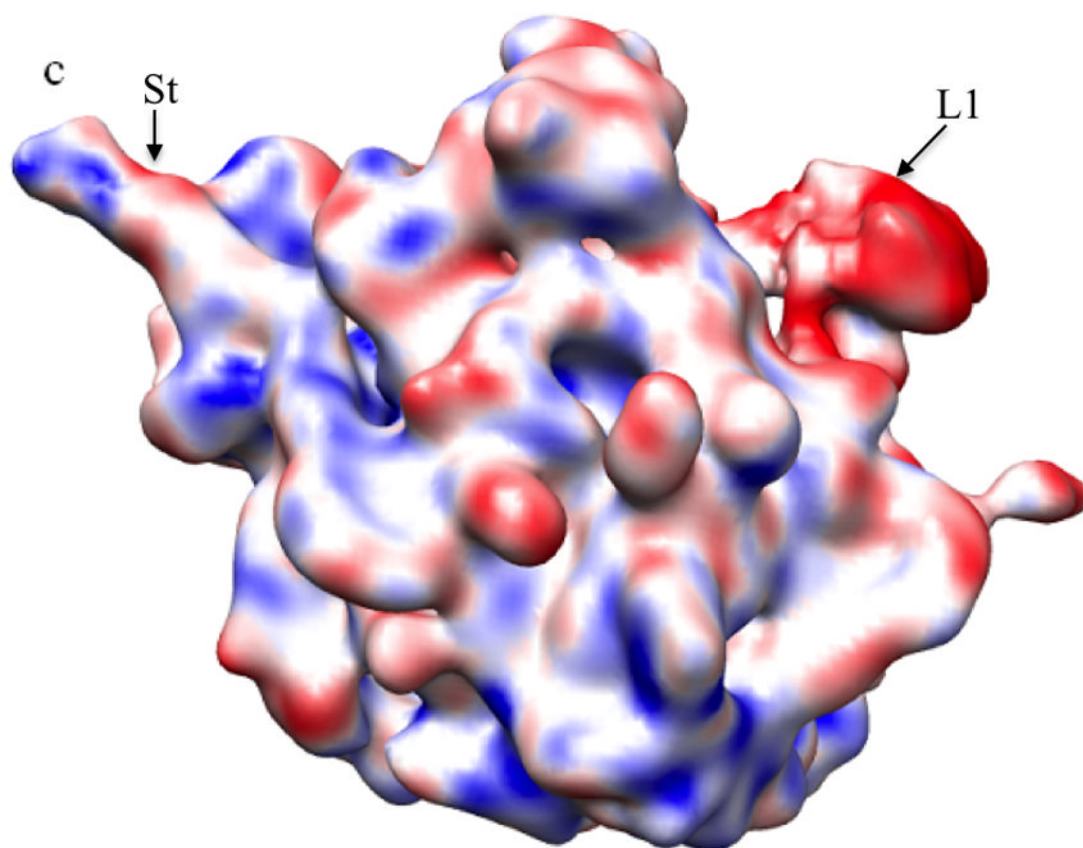
Figure 2.

Bootstrap variance maps calculated using simulated data modified by the CTF. We show only selected z-slices, the left column is $z=27$, the right column is $z=47$. Contrast within each slice was adjusted independently, so the intensities do not reflect absolute values in respective slices. (a) Variance map of the test structures, (b) variance map obtained with CTF correction, (c) variance map obtained without CTF correction, (d) variance map obtained with CTF correction under strong low-pass filtration, (e) variance map obtained without CTF correction under strong low-pass filtration

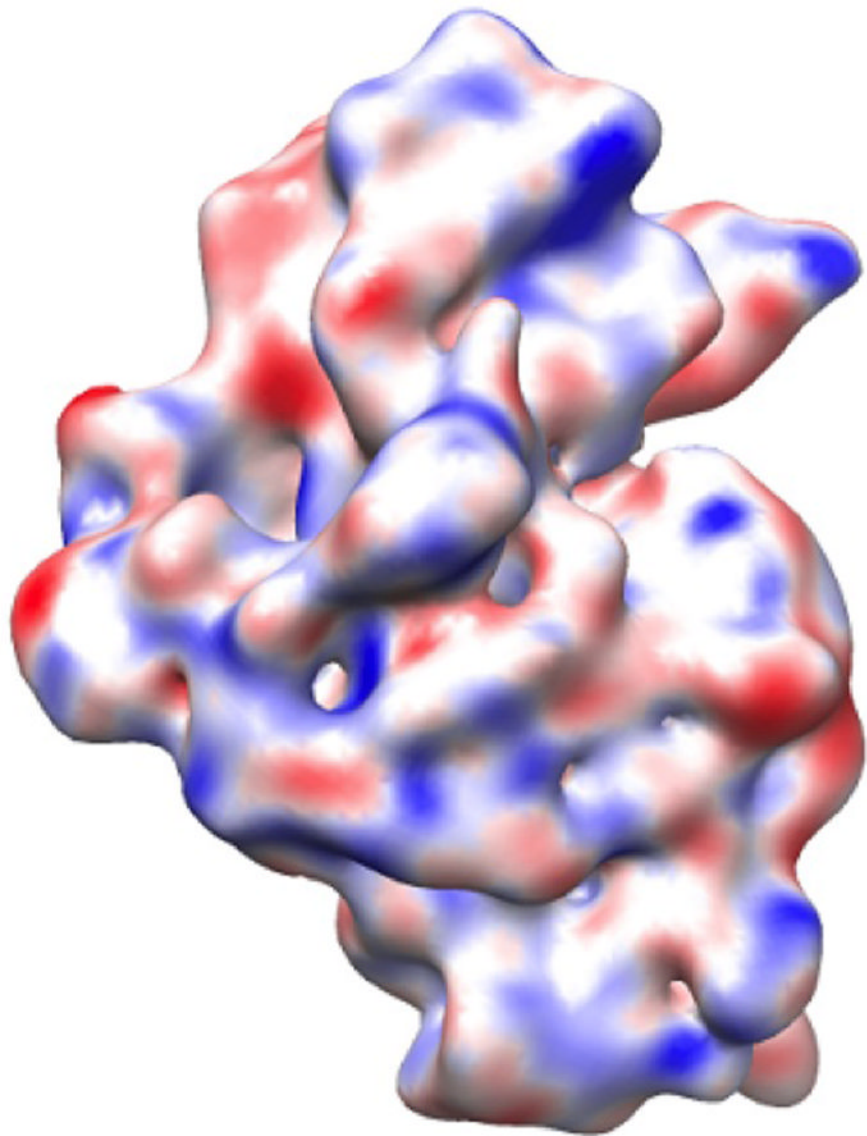


b





d

**Figure 3.**

Conformational variability of the 70S•EF-G•GMPPNP complex revealed by the real-space bootstrap variance analysis of the reconstructed cryo-EM structure. The surfaces are color-coded using the values of the bootstrap standard deviation (σ) 3-D map. The red region has the highest variability ($\sigma > 0.4 \text{ g/cm}^3$), while 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).