JMB

A 11.5 Å Single Particle Reconstruction of GroEL Using EMAN

Steven J. Ludtke[1], Joanita Jakana[1], Jiu-Li Song[2], David T. Chuang[2] and Wah Chiu[1]*

[1]*National Center for Macromolecular Imaging Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, One Baylor Plaza, Houston TX 77030, USA*

[2]*Department of Biochemistry University of Texas Southwestern Medical School Dallas, TX 75390, USA*

Single-particle analysis has become an increasingly important method for structural determination of large macromolecular assemblies. GroEL is an 800 kDa molecular chaperone, which, along with its co-chaperonin GroES, promotes protein folding both *in vitro* and in the bacterial cell. EMAN is a single-particle analysis software package, which was first publicly distributed in 2000. We present a three-dimensional reconstruction of native naked GroEL to ~11.5 Å performed entirely with EMAN. We demonstrate that the single-particle reconstruction, X-ray scattering data and X-ray crystal structure all agree well at this resolution. These results validate the specific methods of image restoration, reconstruction and evaluation techniques implemented in EMAN. It also demonstrates that the single-particle reconstruction technique and X-ray crystallography will yield consistent structure factors, even at low resolution, when image restoration is performed correctly. A detailed comparison of the single-particle and X-ray structures exhibits some small variations in the equatorial domain of the molecule, likely due to the absence of crystal packing forces in the single-particle reconstruction.

© 2001 Academic Press

Keywords: single-particle reconstruction; GroEL; structure; electron cryomicroscopy; X-ray scattering

*Corresponding author

Introduction

Over recent years, single-particle reconstruction has become an increasingly popular technique for structural analysis of large proteins and macromolecular assemblies.[1] To achieve accurate reconstructions at higher resolutions, accurate image restoration by correcting the contrast-transfer function (CTF) and envelope function is required. In addition, the number of required particles increases rapidly with increasing resolution.[2] The EMAN software suite was introduced in 2000 to approach these issues.[3]

The prokaryotic chaperonin GroEL is a molecular chaperone which, along with its co-chaperonin GroES, promotes protein folding both *in vitro* and in the bacterial cell.[4–6] GroEL is a double-ring complex with two heptameric rings of identical 57 kDa subunits stacked back-to-back[7,8], for a total

mass of ~800 kDa. Electron cryomicroscopy has been a powerful approach in deciphering the mechanism by which GroEL/GroES mediate protein-folding.[9–11] So far, most of the published structures have been determined only at relatively low resolution (20-30 Å). Though a number of crystal structures of GroEL have been reported (e.g. 1AON, 1DER, 1GRL, 1OEL, 1DKD), single-particle techniques are still being applied to examine bound ligands and functional states in more detail.

We have performed a single-particle reconstruction of native, naked GroEL completely within the EMAN software environment with no reference to previous X-ray or EM structures. Our result demonstrates the robustness of the single-particle reconstruction technique in general and validates the specific reconstruction and CTF/envelope-function correction methodologies implemented in EMAN.

Results

Electron cryomicroscopy and X-ray scattering

Two independent types of experimental data were collected for this study: single-particle data

---

acquired by electron cryomicroscopy (Figure 1) and X-ray solution scattering data (Figure 2). Though the X-ray crystal structure of GroEL has been determined, it was not referenced in any way during any part of the reconstruction procedure. The X-ray solution scattering data were used in the context of the mathematical model describing the CTF of the microscope, to make the CTF and envelope-function parameter determination as accurate as possible (as described below). The X-ray scattering experiment provided an isotropic (one-dimensional) structure factor intensity, which was then used in conjunction with *CTFIT*'s eight-parameter model to perform a fit of the single particle power spectrum from each micrograph. Rings were visible in the power spectrum to ~8 Å^{-1} in the closest-to-focus micrograph and ~12 Å^{-1} in the furthest-from-focus micrograph. Figure 3 shows an example of a fit performed on the data from one micrograph. The discrepancies at low resolution are most likely due to preferred orientation in the EM data, as described below. This fact is supported by the accurate agreement of the structure factor of the final single-particle reconstruction, and demonstrates that accurate CTF and envelope-function parameters can be determined even in the presence of preferred orientations.

Three-dimensional reconstruction and refinement

The reconstruction was performed using ~5000 particles taken from five micrographs ranging from 1 to 2.4 μm underfocus on a 400 keV electron cryomicroscope. Approximately 1700 particles were discarded in the final reconstructed model. This elimination process is performed automatically during the reconstruction by comparing each particle to others in approximately the same orientation, and keeping only the most self-consistent data.

Figure 4 shows the sequence in the iterative refinement of the GroEL structure. The top of this Figure shows the Fourier-shell correlation (FSC) between successive models in the refinement procedure. This sequence is used to judge when the refinement has converged. Due to the high noise levels present in the individual particles, true convergence, in which two successive models are identical, will never be reached. Pseudo-convergence, where the FSC curve from one iteration to the next is essentially unchanged, is the criteria used to terminate the refinement process. The bottom half of this Figure shows the reconstructions generated in refinement iterations one to four. Each model is CTF and envelope-function corrected, filtered and masked. The structure has nearly converged after four or five iterations.
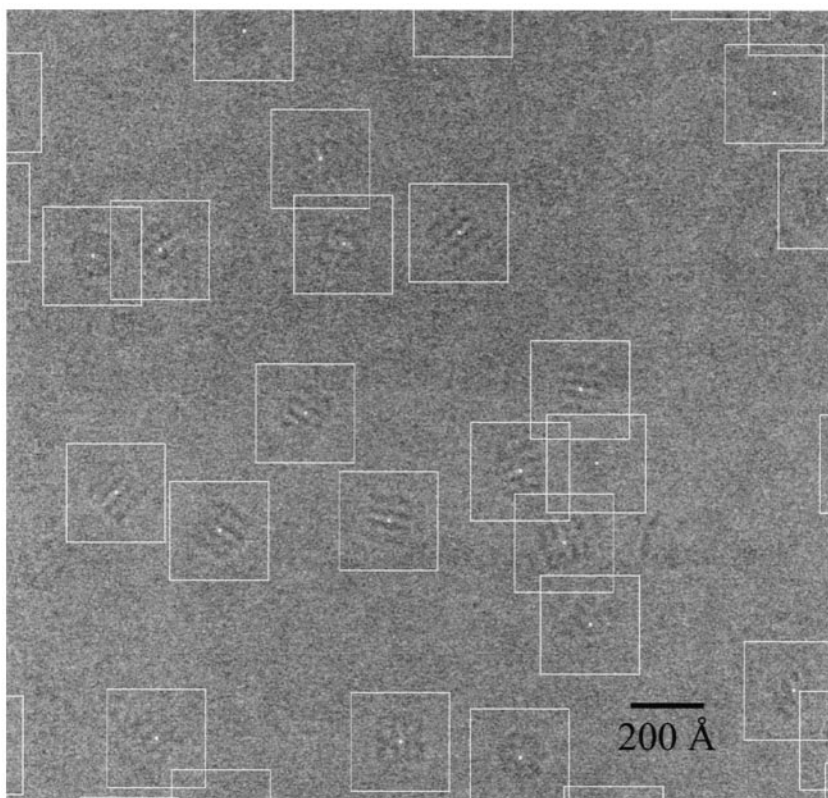

Figure 1. Typical area of a GroEL micrograph at ~2 μm underfocus. This is from one of the five micrographs used in the 11.5 Å reconstruction. Particles have been boxed out for clarity.
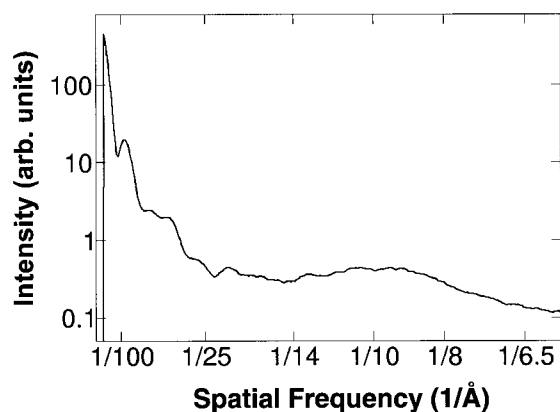
Figure 2. The corrected background-subtracted X-ray solution scattering curve of GroEL collected at SLAC.
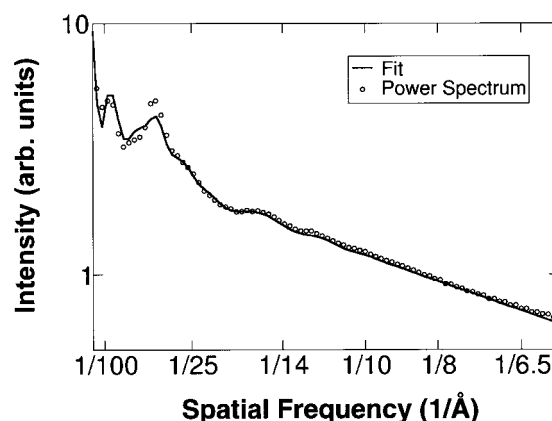

Figure 3. Typical fit of the averaged power spectrum for the particles from one micrograph to the CTF and envelope-function model including the X-ray scattering curve.

The final 3-D model, obtained after eight rounds of iterative refinement is shown in Figure 5. The single-particle reconstruction was filtered during CTF and envelope-function correction of the 2-D images as described below. In order to validate our structure, we compared our reconstruction with the X-ray crystallographic model, which was blurred to a similar resolution using a simple Gaussian filter with a half-width of 11.5 Å in Fourier space (Figure 5). This filter is a rough approxi-

mation to the Wiener filter applied to the 2-D images during reconstruction.

Validation of 3-D reconstruction with X-ray crystallographic and scattering data

The 1-D structure factor of the single-particle reconstruction is compared to both the X-ray solution scattering curve and the curve generated
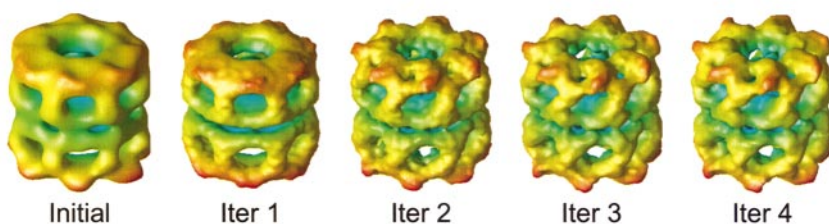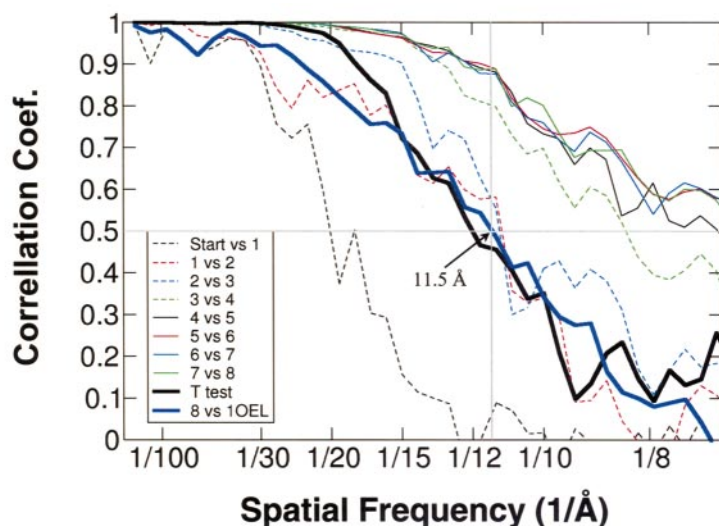

Figure 4. Refinement sequence. (Top) Plot of FSC between each 3-D model and the model from the previous iteration. The thick black line represents a simple T-test, i.e. the FSC between models generated with only the even-numbered particles *versus* one generated with only the odd-numbered particles. The thick blue line represents the FSC between the 3-D model after eight iterations of refinement and the X-ray crystal structure of GroEL in this state. (Bottom) The initial 3-D model and the models after refinement rounds 1-4.
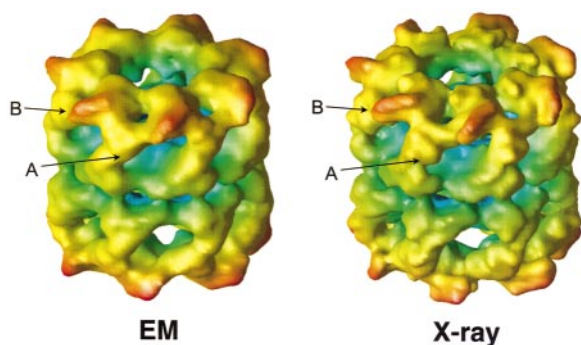
Figure 5. The final, refined 3-D model after eight rounds of refinement compared to the X-ray crystal structure blurred with a Gaussian filter to ~12 Å resolution. Feature A points to an area of slight variability in the reconstruction in the intermediate domain. This variability implies structural heterogeneity in the protein in solution in this region. Feature B is the density corresponding to a two-helix bundle, which can almost be resolved at this resolution.
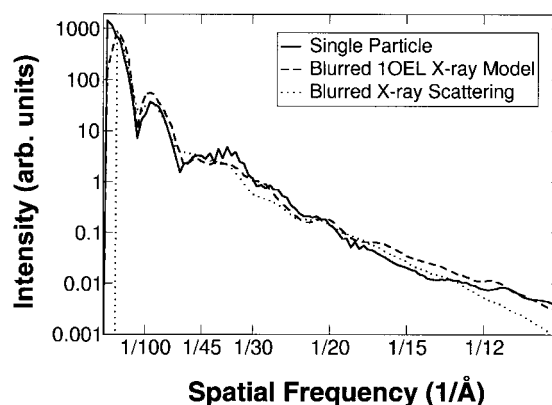

Figure 6. The power spectrum of the CTF-corrected, Wiener-filtered single-particle reconstruction, a 11.5 Å Gaussian-filtered 3-D model of 1OEL, and the X-ray scattering curve Gaussian filtered to 11.5 Å. Note that some of the differences in these curves are attributed to differences in filtering, as explained in the text.

from the X-ray crystallographic reconstruction (Figure 6). Both of the X-ray curves were multiplied by a Gaussian filter with a half-width of 11.5 Å to roughly simulate the filtering applied to the EM model. All three curves are found to match extremely well, even at low resolution. The small discrepancy between the curves at higher resolution can be attributed to the difference between the Wiener filter used in the single-particle reconstruction and the Gaussian filter used on the X-ray data. The slight discrepancy between the two X-ray curves is likely due to Fourier aliasing during generation of the 1-D structure factor from the X-ray crystallographic reconstruction.

The FSC is used to compare the similarity of the models as a function of resolution. The thick black line in Figure 4 shows the FSC calculated by performing a two-way T-test on the single-particle data, as described below. The thick blue line represents the FSC between the final single-particle model and the X-ray crystallographic reconstruction. Using the 0.5 FSC criteria, both curves agree that the single particle resolution is approximately 11-12 Å.

Discussion

Refinement and CTF/envelope-function correction

The basic refinement scheme is the same as previously described.[3] Some improvements have been made to the CTF correction algorithms as discussed below. Differences in CTF and envelope-function correction techniques are what most distinguishes EMAN from other available single-particle reconstruction software. In EMAN, each

particle potentially has an independent set of CTF parameters. The data-set containing all images at all defocuses is processed together as a single set. CTF amplitude correction is performed automatically as class-averages are generated.

The latest version of the correction routine also incorporates a true Wiener filter, making use of the measured contrast (signal-to-noise ratio) as a function of spatial frequency in the raw images. Absolute contrast scaling for the Wiener filter is provided by the CTF model including the solution X-ray scattering intensity. However, the structure factor of the final reconstruction is derived from the electron micrographs, which were corrected using a mathematical model of the CTF/envelope function and not by applying the X-ray solution scattering data directly to the final reconstruction. Use of the X-ray scattering data in the Wiener filter term does not impose any features from X-ray scattering on the reconstructed model due to the highly non-linear nature of this filter. This is an important distinction, since this process provides an important cross-validation of the mathematical model used for CTF correction.

Our methodology is substantially different from those used in most other single-particle reconstruction methodologies. For example, in SPIDER, particles are separated into groups of similar defocuses and processed to generate individual 3-D models for each defocus group.[12] The 3-D reconstructions are then merged and CTF corrected. A Wiener filter is also incorporated into the mathematics used for CTF correction, but a single average contrast value is used at all spatial frequencies. In this specific case,[12] envelope-function correction and filtering caused an apparent mismatch between the X-ray scattering curve and the

structure factor of the final single-particle reconstruction, which was then corrected. Here, we demonstrate that a good match between the different structural techniques is possible, and that these different techniques do, in fact, agree quite well (Figure 6).

Data orientation distribution

Figure 7 illustrates the distribution of Euler angles among the particles in the data-set determined during refinement. Clearly, GroEL exhibits a strong orientation preference. A vast majority of the particles (∼75 %) are within ten degrees of the equator of the asymmetric triangle, indicating side-views of GroEL. A significant, but much smaller, number of particles is found in or near the top, 7-fold symmetric, view. Fortunately, a uniform distribution of particles around the equator provides a complete description of the particle. That is, an equatorial distribution is sufficient to fill Fourier space with no missing cone. This preferred orientation does, however, lead to an asymmetry in the resolution of the 3-D model. With such an axial distribution of particles, the averaging along the symmetric axis in Fourier space is better than the averaging along the other two axes. This leads to a slightly anisotropic resolution in the final model. The vertical resolution (variations in the direction of the symmetric axis), will be slightly higher than
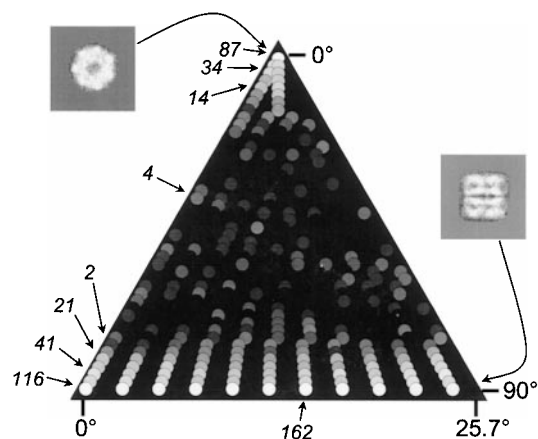

Figure 7. A plot representing the Euler angle distribution of classified particles within the asymmetric triangle. Note that the triangle is stretched horizontally for better visualization, which makes the horizontal and vertical projection distribution appear non-uniform. The brightness of each point indicates the number of particles used in the class-average in that orientation on a log scale. For clarity, several points are labeled with the number of particles found. Clearly, GroEL exhibits preferred orientations with side and top-views being the most common by far. However, since there is a uniform distribution along the equator of the asymmetric triangle, the class-averages represent a complete description of the model, that is, there is no missing cone in Fourier space. This distribution does produce a non-isotropic resolution.

the radial resolution (radially out from this axis). Unfortunately, due to statistical noise in the FSC curves themselves, independently estimating the vertical and horizontal resolutions cannot be performed with any accuracy.

Resolution evaluation

The T-test performed to estimate the resolution of the reconstruction deserves some discussion, since there are a number of different ways this test can be performed. The most complete T-test requires performing completely independent reconstructions on the two halves of the data, including the generation of independent starting models. This test not only measures the resolution due to the noise level of the final model, but also investigates the reproducibility of the structure itself and its dependence on choice of starting model. These are really two independent issues. In the simplified T-test used here, the two models were generated using the same reference from the previous iteration, so they do not test the initial model dependence of the refinement. We address this issue separately. This simpler test investigates the amount of noise present in the final model.

Note also that this T-test will tend to somewhat underestimate the resolution of the model generated with all of the data, since the final model was generated with twice the data of either of the T-test models. In a traditional student T-test, the data are split into two, three, four, etc. groups. A resolution test is done for each of these groups, and the final resolution is interpolated for the case where the data are not split at all. Unfortunately, when particles are aligned for a class-average, the quality of the alignment is sensitive to the noise level of the final average. This means that the resolution/quality of a reconstruction degrades faster than linearly with the number of particles. If the particle count falls too low, a reconstruction may become impossible. This makes splitting the data into more than two groups difficult in many cases.

We also consider the question of dependence on the starting model. A simple argument shows that this should not be a significant effect. For instance, the final model contains substantial high-resolution features. Since the starting model was heavily low-pass filtered and contained no detailed features, these features must be derived entirely from the data. However, there is a counterargument that a flawed algorithm could generate random artifacts, and iteratively amplify them into pseudo high-resolution features. To counter this argument in the case of EMAN, we performed two complete reconstructions from the full data-set. In the first case we started with the *ab-initio* starting model produced from the raw data. In the second case, we used the X-ray crystallographic as a starting model for the reconstruction. The refinement starting with the X-ray model, in fact, converged to a

model virtually identical with the *ab-initio* reconstruction at the measured resolution.

Filtering the 3-D map

A Wiener filter is included in the CTF/envelope-function correction equations used to generate class averages in 2-D. This process can be viewed as first performing a "perfect" correction to the data, which will amplify the structure factors to their "true" predicted values, then, second, applying a Wiener filter to damp out the resulting over-amplified high-frequency noise. When these filtered class-averages are combined to form a 3-D model, no additional filtration is required in 3-D. The data are already optimally filtered, with the two caveats discussed below.

EM *versus* X-ray structures

Since the FSC analysis of the X-ray model *versus* the EM model gives a resolution very close to that determined by the T-test on the EM data alone, it is reasonable to claim not only that the EM model is accurate to this resolution, but also that any differences between the EM and X-ray structures at this resolution must contain a small fraction of the total molecular mass. This also provides a concrete demonstration that the automasking procedure outlined below does not lead to an inappropriate overestimation of resolution, since the X-ray model was not masked at all.

In terms of the overall structural similarity, the X-ray model leaves a strong impression of having more detail than the EM model (Figure 5). Careful inspection shows that this is not the case. For instance, we can examine the density corresponding to the two-helix bundle labeled B in Figure 5. At 11.5 Å resolution, neither our map nor the equivalent map from X-ray crystallography can resolve the neighboring helices. Nevertheless, it is clear that the amount of visible separation of densities between the helices is similar in the two maps. There are two reasons for the apparent difference in level of detail as an overall structure. First, only a Gaussian blur was performed on the X-ray model, and the EM reconstruction has phase errors as well. This "phase blurring" is responsible for much of the difference in overall appearance between the models. Second, EMAN's use of class-averages means that particles in slightly different orientations are averaged together. To achieve optimal signal-to-noise ratio and resolution, the angular spacing between class-averages must be selected carefully. If too large a value is selected, considerable blurring will occur, and it will be impossible to go beyond an angle-limited resolution. If too small a value is used, the number of particles in each class average becomes small enough that the iterative particle alignment within each average will converge poorly, reducing the resolution of the final reconstruction. The value selected for this model was a near optimal selec-tion, meaning there is a slight amount of rotational blurring in the final model, but the resolution is close to the best value that can be achieved with this data. This could be improved, but only at the cost of increasing the noise level.

Most of the high-resolution protrusions that seem to be missing in the EM model are still present, just smeared slightly. There is, however, one exception. In the intermediate domain, labeled A, we have observed substantial mass displacements. Even when the model has converged, substantially more variation occurs between iterations in the intermediate domain than in other regions of the model. This is likely due to real structural variations in the protein in solution. With single-particle analysis it is possible to perform statistical classification and generate an ensemble of 3-D models representative of the range of motions in this domain, but doing so will require substantially more data. Since these changes are near the limiting resolution of this reconstruction, we cannot make a more definitive statement about this effect until we have produced a much larger data-set and a higher-resolution model. It should also be noted that this effect occurs even when the X-ray model is used as a starting model for refinement, confirming that it is not an initial model-dependent effect.

It was previously noted in the X-ray structure (1OEL) that the seven subunits were not strictly identical. This fact is not responsible for the observed variations, since these changes are too small to be observed at this resolution. The observed variations in the intermediate domain are much larger, and cannot be accounted for by considering the averaging imposed by the full D7 symmetry applied in this reconstruction. This region of GroEL is known to undergo substantial motion when triggered by nucleotide binding.[13] While this preparation was designed to be nucleotide-free, it is quite possible that there is some natural conformational variability in this domain. This is likely to be the source of the instability we noticed in this region during iterative refinement. Clearly this motion is functionally significant. Once we have achieved ~8 Å resolution and have a much larger data-set to work with, we should be able to perform some statistical separations on the data and better interpret this conformational variation.

Conclusion

In summary, we have confirmed that single-particle analysis can produce reliable 3-D maps at intermediate resolutions. We have also demonstrated that even at low resolution, agreement between electron microscopy, X-ray scattering and X-ray crystallographic structure factors is possible. Future work will focus on improving the resolution of this map to the point where individual helices can be accurately resolved, and on examin-

ing functional states and ligand binding. While much work at lower resolution has been performed by several groups,[9,10,14,15] higher resolution will allow a much more accurate fitting of X-ray model domains into the EM model, and investigate if the slight structural differences between the EM and X-ray models are real. Our results have rigorously demonstrated our methodologies to be reliable for intermediate-resolution structures of large macro-molecular machines, which can be reliably used to understand various functional and mechanistic questions through structural analysis.

Materials and Methods

Expression and purification of GroEL

Chaperonin GroEL was purified according the method described previously.[16] Briefly, the pGroESL plasmid that overexpressed *Escherichia coli* GroEL and GroES was transformed into ES[ts] CG-712 cells, which were grown at 37 °C under chloramphenicol selection to an A_{600} of 0.6. The expression of GroEL was induced by IPTG over-night at 37 °C. Cell lysates were prepared by sonication in a lysis buffer containing 1 mM phenylmethylsulfonyl fluoride and 1 mM benzamidine. After an initial 30 %-60 % $(NH_4)_2SO_4$ fractionation, the protein sample was treated with 10 mM Mg-ATP at 37 °C for two hours, followed by DEAE-Sepharose column chromatography. The GroEL fractions collected from the ion-exchange column in the absence of Mg-ATP were concentrated and purified on a Sephacryl S-400HP column with Mg-ATP also omitted from the column buffer. Eluted GroEL fractions were pooled, concentrated and further purified on a Reactive Red column (2.6 cm × 70 cm), with GroEL recovered in 20 mM Tris-HCl (pH 7.5), 50 mM $MgCl_2$ in the flow-through fractions. Concentrations of GroEL were determined spectrophotometrically using the published extinction coefficient[17] of 1.22×10^4 $M^{-1}cm^{-1}$.

X-ray scattering

X-ray solution scattering experiments were conducted at beam line 4-2 of the Stanford Synchrotron Radiation Laboratory as previously described.[18] GroEL at concentrations of 45 mg/ml and 3 mg/ml were measured at two detector lengths. The resulting data were background-subtracted using buffer blanks and corrected using established procedures.[2] Data from both length scales were combined into a single scattering profile covering the range from 400 to 4.4 Å (Figure 2).

Electron cryomicroscopy

Microscope grids were prepared with carbon holey film and glow-discharged immediately before freezing. GroEL in solution at ~1 mg/ml was frozen across the holes by rapid plunging into liquid ethane. Imaging was performed on a JEOL 4000 microscope with a LaB$_6$ gun and a Gatan liquid nitrogen specimen holder at 400 keV. A condenser aperture of 100 μm, a spot size of 3 and an objective lens aperture of 70 μm were used. A nominal magnification of 50,000× with a dose of ~30e$^-$/Å^2 was used to record the image onto Kodak 163 film developed for 12 minutes in Kodak D10 developer at 20 °C. Micrographs were then scanned on a Zeiss SCAI scanner at 7 μm/pixel. A 2 × 2 median filter was applied to

provide 14 μm/pixel, providing the final sampling of 2.7 Å/pixel.

Particle selection and CTF/envelope-function parameter determination

Particles were selected using a new automatic particle selection procedure implemented in the latest version of EMAN, which is more accurate for particles with non-rotationally symmetric projections total; 4916 particles were selected from five micrographs ranging from 1 to 2.4 μm underfocus. A section of a typical micrograph is shown in Figure 1. CTF and envelope-function parameters were determined using standard techniques in *ctfit*, the CTF-determination program in EMAN. The power spectrum was calculated for each particle in a single micrograph, and these power spectra were first rotationally averaged then averaged together (Figure 3). The slightly modified version of the original CTF model[3] is:

$$M^2(s) = F^2(s)CTF^2(s)E^2(s) + N^2(s)$$

where:

$$CTF(s) = -A(\sqrt{1 - Q^2}\sin(\gamma) + Q\cos(\gamma))$$

$$\gamma = -2\pi\left(\frac{C_s\lambda^3 S^4}{4} - \frac{\Delta Z\lambda s^2}{2}\right)$$

$$E(s) = e^{-Bs^2}$$

$$N^2(s) = n_1 e^{n_2 S + n_3 s^2 + n_4\sqrt{s}}$$

$M^2(s)$ is the power spectrum of the particles, $F^2(s)$ is the structure-factor intensity from X-ray scattering, λ is the electron wavelength, s is spatial frequency and A, Q, ΔZ, B and n_{1-4} are the fit parameters. This includes the correction of a misprint in the definition of N^2 in the original publication.[2,3] In addition, we are now using the more common convention that the defocus is positive for underfocus.[19] The X-ray scattering structure factor was used only twice in the reconstruction: first, to provide a more accurate fit of the CTF/envelope-function parameters for each micrograph (Figure 3), and second, to provide a properly scaled signal-to-noise ratio for the Wiener filter used to reduce high-frequency noise.

CTF-phase flips were corrected immediately, before any reconstruction was performed. Amplitude CTF correction was performed in a later step during the generation of class-averages, as described previously,[3] with the addition of a Wiener filter term as discussed below.

The experimental B-factors[2] of the five micrographs ranged from 100 Å^2 in the closest-to-focus image to 250 Å^2 in the furthest-from-focus image. Amplitude contrast was ~10 % in all images, although the effects of the preferred orientation of the particles on the low-resolution structure factor cause some uncertainty in this value. The match of the structure factor of the final 3-D model to the X-ray scattering structure factor indicates that the selected parameters were sufficiently accurate for accurate CTF correction, even at low spatial frequencies. These values are consistent with values obtained for other specimens on this microscope.

Initial model

The 3-D reconstruction is performed in two stages: initial model generation and model refinement. In this case, the initial model was generated using a novel technique based on the rotational symmetry of the model. Briefly, in this approach, we began with the assumption that the particle had at least one n-fold rotational axis with n to be determined. Each particle was subjected to a self-symmetry evaluation, and the 100 particles with the best n-fold rotational symmetry, and the 100 particles with the best mirror symmetry and the poorest n-fold rotational symmetries were selected. The best n-fold particles represent the ''top''-view of the particle and the mirror particles represent ''side''-views of the particle. Note that for particles with an odd-numbered symmetry, side-views would not necessarily have true mirror symmetry, but generally one of the side-views will have a pseudo mirror-symmetry which is adequate to distinguish it from other views.

The 100 particles for each view were then mutually aligned and averaged together utilizing the standard EMAN class-averaging technique.[3] This technique excludes a fraction of the particles based on their dissimilarity to the generated average. Though the self-symmetry search is somewhat inaccurate due to high noise levels, this process insures that the class averages are as self-consistent as possible. It also forces the side-views, which may be a set mixed with several different side-views, to produce a single, relatively consistent average (Figure 7).

Once the two class averages representing two orthogonal views were produced, a 3-D model was constructed from just these two views. This model was extremely noisy and inaccurate, but had the approximate size and shape of GroEL. To help improve the model for use in refinement, the symmetry used for the particle search was imposed on the 3-D model, and it was filtered to low resolution. An automatic masking algorithm was also applied to the 3-D model to eliminate some of the artifacts produced by building a model from only two views.

Refinement and CTF/envelope-function correction

A number of improvements have been made to the iterative refinement procedure used by EMAN since the original publication[3], so we briefly describe the current methodology. The overall refinement technique remains the same as previously described, that is, particles are classified by comparison to projections of earlier models. Classified particles are then mutually aligned, corrected for CTF and envelope function, and averaged to generate a set of class-averages, which are combined to form a new 3-D model. This process is then iterated until convergence is achieved. The number of projections generated for classification is determined based on the size of the particle and the desired resolution. The class-averages are Fourier transformed and inserted into the 3-D Fourier volume using three-pixel Gaussian interpolation.[3] The 3-D model is then produced by performing an inverse Fourier transform on this volume.

The interpolation performed in direct Fourier inversion produces some undesirable artifacts in the reconstructed model. To improve agreement between the 3-D model and the class-averages, the model is iteratively modified in real-space. Projections of the 3-D model are generated corresponding to each class-average. The difference between each class-average and projection is then used to adjust the densities in the 3-D model. The process is iterated several times until changes become negligible.

CTF/envelope-function correction is performed as part of the class-averaging procedure. The corrections are performed as previously described[3] except for the addition of a Wiener filter term. This expression is mathematically derived to be the optimal correction in a least-squares sense under a few reasonable assumptions. The modified amplitude-correction equation is (expanded for clarity):

$$T(s, \theta) = \left(\sum_i \frac{1}{|CTF_i(s)E_i(s)|} \cdot \frac{R_i(s)}{R(s)} \cdot M_i(s, \theta) \right) \cdot \frac{F^2(s)R(s)}{1 + F^2(s)R(s)}$$

where

$$R_i(s) = \frac{CTF_i^2(s)E_i^2(s)}{N_i^2(s)}$$

$$R(s) = \sum_i \frac{CTF_i^2(s)E_i^2(s)}{N_i^2(s)}$$

$T(s,\theta)$ represents the filtered, CTF/envelope-function corrected class-average, $N_i(s)$ represents the fitted noise in each particle and $M_i(s,\theta)$ represent the ith aligned individual particle images. Three corrections are applied in the equation above, the first represents the actual CTF correction, i.e. dividing by the CTF and envelope functions. The second weights the contributions of each particle by relative contrast (signal-to-noise ratio), $R_i(s)$, providing the mathematically optimal combination of the data from multiple micrographs. The third is a traditional Wiener filter.[20] This filter will take a noisy, but otherwise ''perfect'' image and filter it to produce an image which matches the true image as well as possible in a least-squares sense when the spectral signal-to-noise ratio is known. The R_i-values represent the relative contrast in each image. Note that the X-ray scattering structure factor, $F(s)$, could be incorporated into the $R(s)$ values, to provide absolute values for the signal-to-noise ratios, but as they are used, the $F(s)$ would cancel out anyway. The Wiener filter expression, however, relies on $F(s)$ to provide an accurate absolute contrast. Note that this filter term is not imposing any of the features present in $F(s)$ on the results, it is simply used to filter out the high-frequency noise that would otherwise be present in T. This term could be replaced with a Gaussian or other predefined filter function, which would completely eliminate the dependence on $F(s)$, which may not be generally available.

While the Wiener filter performs optimal filtration on the 2-D class-averages, when the class-averages are combined to form a 3-D model, additional averaging occurs, increasing the contrast relative to the 2-D class-averages. For truly optimal filtration in 3-D, the 2-D filter would have to be removed, and the Wiener filter would have to be recalculated from the 3-D model. Fortunately, this effect is limited, due to the fact that such averaging is most significant at low resolution where the contrast is already significantly larger than 1. The averaging is much less significant at high spatial frequences, and thus the over-filtration is a minor effect, and not presently corrected. In addition, since the filtration is performed in 2-D, the effective filtration in 3-D is not spherically symmetric.

This effect is also countered by a flaw in our estimate of the signal-to-noise ratio. We assume in our estimate that all particles in the class-average are exactly identical except for noise. In reality there will be some inaccuracy in alignment, as well as some structural variation between molecules. These effects mean we may be slightly overestimating the signal-to-noise ratio of the class-average, particularly at high resolution, and thus under-filtering the class-averages. This effect is, in fact, visible for GroEL, as the class-averages have a somewhat grainy appearance that would not be present in a Wiener filter with a perfect estimate of signal-to-noise ratio. This effect does not affect the accuracy of the reconstructions, only the level of filtration applied for visualizing the result.

3-D map masking

An additional algorithm is used to eliminate many high-density artifacts outside the 3-D model. These artifacts are often caused by high particle density or ice contamination in the micrograph. They are eliminated through the use of an iterative automasking algorithm. The algorithm begins by including a spherical ball of density centered on the model. It then iteratively locates all densities connected to this sphere above a specified density threshold. Typically this threshold is ~1/2 the density threshold used for isosurface rendering. This avoids the possibility of the mask impinging on the visualized isosurface. Once all the connected mass has been located, a user-defined number of 1 voxel thick layers is iteratively included in the mask. Typically three or four layers (or ~10 Å) are sufficient to make sure the edge of the mask occurs at a distance from the model where the mean density is nearly zero. Not only does this help to insure no desired low densities, which are actually part of the model, are accidentally excluded, but also, since the mask occurs in regions of near zero density, it also reduces any improper correlations when two such models are compared in a resolution test. It should be emphasized that this mask is not fixed for a given model, but is regenerated every time a new 3-D model is generated. This further reduces the possibility of improper correlations between masks causing an overestimation of resolution. Since the mask extends to the zero mean-density region outside the model, a sharp mask can safely be used. In the iterative refinement, this 3-D mask is also used to provide 2-D masks for the class-averaging procedure. The additional shells added outside the mask provide adequate opportunity for the model to grow in any given direction during iterative refinement in cases where the initial model is not very accurate.

Determining handedness of the model is impossible using only untilted micrographs, since micrographs represent projections of protein density. Since the handedness of GroEL is well established, we simply assigned the known handedness of the model during alignment rather than performing an additional experiment.

Acknowledgments

References

1. Nogales, E. & Grigorieff, N. (2001). Molecular machines: putting the pieces together. *J. Cell. Biol.* **152**, F1-F10.
2. Saad, A., Ludtke, S. J., Jakana, J., Rixon, F. J., Tsuruta, H. & Chiu, W. (2001). Fourier amplitude decay of electron cryo-microscopic images of single particles and effects on structure determination. *J. Struct. Biol.* **133**, 32-42.
3. Ludtke, S. J., Baldwin, P. R. & Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82-97.
4. Sigler, P. B., Xu, Z., Rye, H. S., Burston, S. G., Fenton, W. A. & Horwich, A. L. (1998). Structure and function in GroEL-mediated protein folding. *Annu. Rev. Biochem.* **67**, 581-608.
5. Xu, Z. & Sigler, P. B. (1998). GroEL/GroES: structure and function of a two-stroke folding machine. *J. Struct. Biol.* **124**, 129-41.
6. Martin, J. & Hartl, F. U. (1997). Chaperone-assisted protein folding. *Curr. Opin. Struct. Biol.* **7**, 41-52.
7. Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L. & Sigler, P. B. (1994). The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature,* **371**, 578-586.
8. Braig, K., Adams, P. D. & Brunger, A. T. (1995). Conformational variability in the refined structure of the chaperonin GroEL at 2.8 Å resolution. *Nature Struct. Biol.* **2**, 1083-1094.
9. Chen, S., Roseman, A. M., Hunter, A. S., Wood, S. P., Burston, S. G., Ranson, N. A. *et al.* (1994). Location of a folding protein and shape changes in GroEL-GroES complexes imaged by cryo-electron microscopy. *Nature,* **371**, 261-264.
10. Carazo, J. M., Marco, S., Abella, G., Carrascosa, J. L., Secilla, J. P. & Muyal, M. (1991). Electron microscopy study of GroEL chaperonin: different views of the aggregate appear as a function of cell growth temperature. *J. Struct. Biol.* **106**, 211-220.
11. White, H. E., Chen, S., Roseman, A. M., Yifrach, O., Horovitz, A. & Saibil, H. R. (1997). Structural basis of allosteric changes in the GroEL mutant Arg197 → Ala. *Nature Struct. Biol.* **4**, 690-694.
12. Gabashvili, I. S., Agrawal, R. K., Spahn, C. M., Grassucci, R. A., Svergun, D. I., Frank, J. & Penczek, P. (2000). Solution structure of the *E. coli* 70S ribosome at 11.5 Å resolution. *Cell,* **100**, 537-549.
13. Ma, J., Sigler, P. B., Xu, Z. & Karplus, M. (2000). A dynamic model for the allosteric mechanism of GroEL. *J. Mol. Biol.* **302**, 303-313.
14. Roseman, A. M., Chen, S., White, H., Braig, K. & Saibil, H. R. (1996). The chaperonin ATPase cycle: mechanism of allosteric switching and movements of substrate-binding domains in GroEL. *Cell,* **87**, 241-251.

15. Langer, T., Pfeifer, G., Martin, J., Baumeister, W. & Hartl, F. U. (1992). Chaperonin-mediated protein folding: GroES binds to one end of the GroEL cylinder, which accommodates the protein substrate within its central cavity. *EMBO J.* **11**, 4757-4765.

16. Chuang, J. L., Wynn, R. M., Song, J. L. & Chuang, D. T. (1999). GroEL/GroES-dependent reconstitution of alpha2 beta2 tetramers of humanmitochondrial branched chain alpha-ketoacid decarboxylase. Obligatory interaction of chaperonins with an alpha beta dimeric intermediate. *J. Biol. Chem.* **274**, 10395-10404.

17. Burston, S. G., Ranson, N. A. & Clarke, A. R. (1995). The origins and consequences of asymmetry in the chaperonin reaction cycle. *J. Mol. Biol.* **249**, 138-152.

18. Tsuruta, H., Brennan, S., Rek, Z. U., Irving, T. C., Tompkins, W. H. & Hodgson, K. O. (1998). A wide bandpass multilayer monochromator for biological small angle scattering and diffraction. *J. Appl. Phys.* **31**, 672-682.

19. Erickson, H. P. & Klug, A. (1970). The Fourier transform of an electron micrograph: effects of defocusing and aberrations, and implications for the use of underfocus contrast enhancement. *Phil. Trans. Roy. Soc. ser. B*, **261**, 105-118.

20. Press, W. H., Vetterling, W. T., Teukolsky, S. A. & Flannery, B. P. (1992). *Numerical Recipes in C*, pp. 547-549, Press Syndicate of the University of Cambridge, Cambridge.

Edited by W. Baumeister