Journal of Structural Biology **128,** 82–97 (1999)
Article ID jsbi.1999.4174, available online at http://www.idealibrary.com on IDEAL®

EMAN: Semiautomated Software for High-Resolution Single-Particle Reconstructions

Steven J. Ludtke, Philip R. Baldwin, and Wah Chiu

Verna and Marrs McLean Department of Biochemistry, National Center for Macromolecular Imaging, Baylor College of Medicine, Houston, Texas 77030

Received June 7, 1999, and in revised form August 10, 1999

We present EMAN (Electron Micrograph ANalysis), a software package for performing semiautomated single-particle reconstructions from transmission electron micrographs. The goal of this project is to provide software capable of performing single-particle reconstructions beyond 10 Å as such high-resolution data become available. A complete single-particle reconstruction algorithm is implemented. Options are available to generate an initial model for particles with no symmetry, a single axis of rotational symmetry, or icosahedral symmetry. Model refinement is an iterative process, which utilizes classification by model-based projection matching. CTF (contrast transfer function) parameters are determined using a new paradigm in which data from multiple micrographs are fit simultaneously. Amplitude and phase CTF correction is then performed automatically as part of the refinement loop. A graphical user interface is provided, so even those with little image processing experience will be able to begin performing reconstructions. Advanced users can directly use the lower level shell commands and even expand the package utilizing EMAN's extensive image-processing library. The package was written from scratch in C++ and is provided free of charge on our Web site. We present an overview of the package as well as several conformance tests with simulated data. © 1999 Academic Press

Key Words: **electron cyromicroscopy; CTF; contrast transfer function; single-particle reconstruction; image processing**

INTRODUCTION

Over the past decade, an experimental/computational technique known as single-particle analysis has been growing rapidly in popularity. In this technique, transmission electron microscopy is used to address problems that are difficult to approach with traditional crystallographic methods. A variety of statistical and image processing techniques are applied to a large number of images of identical molecules to produce a three-dimensional structure. This technique offers several advantages over crystallographic methods. First, the macromolecule is typically embedded in vitreous ice, which preserves its native state in a biologically relevant conformation. Second, it is ideally suited to address molecules that are problematical for X-ray crystallography. The majority of X-ray structures are for proteins smaller than 200 kDa. Electron cryomicroscopy, however, is ideal for large macromolecules and assemblies in the range of hundreds to thousands of kilodaltons. Finally, single-particle analysis allows functional states to be addressed relatively easily (Gabashvili *et al.,* 1999; Orlova *et al.,* 1996). When the macromolecule consists of a mixed population of functional states and is still somewhat structurally heterogeneous, statistical techniques can, in principle, be applied to separate particles in the desired state from the mixed population.

For macromolecules that can be purified to chemical and structural homogeneity and are amenable to recording high-resolution images, the biggest limiting factor in applying single-particle methods is data processing. Data collection and scanning can be accomplished relatively rapidly, but once the data are ready for processing, analysis may proceed for weeks before it is even known whether the quality and quantity of data are sufficient to achieve the desired resolution. While several excellent software packages exist that allow single-particle and other reconstruction procedures to be performed (Frank *et al.,* 1996; Schroeter and Bretaudiere, 1996; van Heel *et al.,* 1996; Whittaker *et al.,* 1995), they are designed primarily for experienced users performing reconstructions at intermediate resolutions. EMAN is designed to make this technique more accessible to inexperienced users and provide tools necessary to efficiently process large amounts of data for high-resolution reconstructions.

EMAN is a software package designed specifically for single-particle reconstructions. Its design and

1047-8477/99 $30.00

implementation are based on three general principles. The first is to improve processing efficiency, to make processing the thousands to hundreds of thousands of particles required for reconstructions beyond 10 Å feasible. The second is to make single-particle processing more accessible to inexperienced users, while still providing the tools necessary to allow advanced users to perform custom processing. The third is to introduce a robust method for performing CTF correction with minimal effort on the part of the user. This article describes the design rationales and features of EMAN and demonstrates its applicability with simulated data.

3D RECONSTRUCTIONS IN EMAN

Performing a reconstruction in EMAN is a three-stage process. First, the particles must be selected from scanned micrographs or CCD frames. Second, the boxed-out particles are used to generate a preliminary 3D model. Finally, the preliminary model is used as the starting point for the main refinement loop, which is iterated until the refinement converges. EMAN also allows the CTF of the microscope to be corrected semiautomatically. In this case, CTF parameters are determined and phase correction is performed before the refinement loop. Amplitude corrections are performed automatically as part of the refinement loop.

Particle Selection

Before a reconstruction can be performed, the individual particles must be located in the raw micrographs or CCD frames. EMAN includes *boxer,* a graphical program for manual or semiautomatic particle selection. This program displays the overall micrograph in one window and the boxed particles in another window. The boxed particles are updated immediately when a new particle is selected or an old box is moved. Individual boxes may be adjusted or deleted with a simple mouse click. Contrast inversion is provided to aid in manually locating particles in low-contrast images. For very large or high-resolution micrographs, *boxer* allows the micrograph to be split into several regions of equal size, reducing the memory requirements of the program.

Boxer also incorporates a semiautomatic selection procedure. A number of automatic particle selection methods have been proposed (Frank and Wagenknecht, 1984; Harauz and Fong-Lochovsky, 1989; Lata *et al.,* 1995; Martin *et al.,* 1997; Thuman-Commike and Chiu, 1995; van Heel, 1982). Several of these techniques are based on the technique of generating a rotationally averaged reference image and locating particles by cross-correlating the reference with the entire micrograph. *Boxer* uses a similar technique with a few additional refinements. Since projections of particles in different orientations may have dramatically different appearances, *boxer* uses multiple references to ensure accurate boxing. The user manually selects several particles from the micrograph, which are then rotationally averaged and used as templates. Particles are then located by cross-correlating each template image with the entire micrograph. The individual cross-correlation maps, each from a different template, are then combined by selecting the maximum value at each pixel location from the set of correlation maps. The combined cross-correlation map will then contain peaks for each putative particle in the micrograph. Since some particles may contain multiple peaks, this map is then low-pass filtered to a resolution equivalent to ½ the box size. This minimizes problems with the same particle being multiply selected. Instead, this ensures that peaks that are too close to each other are averaged and a single particle will be identified. A peak-searching algorithm then extracts the location of all recognized particles from this map with peaks above some threshold value.

The only remaining difficulty is determination of this threshold value. Rather than attempting to do this automatically, *boxer* presents the user with a set of threshold sliders. One slider sets the basic peak selection threshold, and two additional threshold values allow particles with excessive or insufficient mean contrast to be excluded. The second pair of sliders is used primarily for eliminating areas of contamination. As the user varies the sliders, the particle boxes in a 1k × 1k section of the micrograph are updated interactively. Once all three thresholds are at appropriate levels, the entire micrograph is autoboxed. While the selection routine in the current version does a relatively good job, there is still room for improvement. Particles that are nonspherical will be selected less accurately than nearly spherical particles. This procedure is likely to be improved considerably in future releases.

Preliminary Model Generation

EMAN's refinement procedure is model based; that is, it requires an initial 3D model to use as a starting point for refinement. The quality of the initial model required to allow the refinement loop to converge is a function of many parameters, including the symmetry of the model, the signal-to-noise ratio in the individual particles, and the degree of spherical asymmetry in the model. Generally speaking, if the data have sufficient signal, the refinement loop will converge even if a very poor initial model is used. Nonetheless, some sort of model is required, and the better the model is, the faster the refinement procedure will converge. Many methods for generating

initial models have been invented by many different groups over the years (Baker and Cheng, 1996; Crowther, 1971; Frank, 1989; van Heel *et al.,* 1996). One solution to this problem is for the user to select representative particle views and generate a model by manually placing geometric objects (cylinders, spheres, etc.) in a 3D map. Of course, this process requires considerable effort and ability on the part of the user and may bias the reconstruction toward an incorrect result. To avoid this, EMAN contains several symmetry-specific routines for automatically generating initial models.

Naturally, in many cases the symmetry is not initially known. In general, the actual symmetry of the molecule must be imposed during the reconstruction or symmetry-breaking artifacts will occur due to high noise levels present in the individual particle images. For example, consider performing a reconstruction of a molecule with a twofold rotational symmetry without imposing this symmetry. If no noise is present, every particle will be assigned one of two Euler angles with equal probability. However, if noise is present, it will dominate the alignment choice, combining constructively to produce symmetry breaking artifacts in the final reconstruction. Generally, with twofold symmetry, this will cause one half of the protein to have much higher mean density than the other half. The lower the noise levels are in the individual particle images, the less this effect will be observed, but in general, the correct symmetry must be applied to obtain a reliable reconstruction from noisy data. The section below, entitled "Symmetry Determination" discusses one method for evaluating the symmetry of a model with unknown symmetry.

For asymmetric particles, those with C_2 symmetry or those for which the routines with imposed symmetry do not work well, a generic model-generating routine is provided. This routine begins by generating a set of reference-free class averages. That is, particles that appear to be similar to one another are grouped together, and then the particles within each group are mutually aligned and averaged. This generates a class average for each group, which should represent one characteristic view of the particle. This routine has several user-defined parameters, which may be adjusted to obtain the best possible distribution of class averages. Assuming the distribution of particle orientations is sufficiently diverse, the set of averages will represent most of the possible particle orientations. Several of these averages are then selected manually for use in generating a 3D model. A Fourier common-lines routine is used to determine the relative orientations of all of the selected averages, which are then combined to generate a 3D model.

For objects known to have icosahedral symmetry, EMAN uses a very fast initial model generation routine. This routine searches the complete particle set for particles with the best five-, three-, and two-fold symmetries. There is, of course, no guarantee that particles with these precise views will exist in the data set, but this is generally not a problem. For each symmetry, EMAN will use the particles that come closest to having the desired symmetry. The quality factor for each particle is measured by calculating the dot product between the particle and itself after an appropriate rotation. Once the particles for each symmetric axis have been determined, each group of particles is mutually aligned and averaged to generate three characteristic class averages. These three views are then used to build a preliminary 3D model. This model is, naturally, very noisy and visually may appear quite different from the final refined model. It is, however, usually sufficient for use as an initial model for refinement.

For objects with C_n symmetry (a single axis of rotational symmetry), where $n > 2$, a procedure similar to the icosahedral procedure is used. First, the particles are searched for views with good C_n symmetry. Then a second search is performed for particles with a mirror or pseudo-mirror symmetry and a poor C_n symmetry. That is, particles that have both a good mirror symmetry and the worst possible C_n symmetry are located. If we consider the projections with C_n symmetry to be "top" views of the particle, then this second group represents possible side views of the particle. Macromolecules with even n will have a true mirror symmetry. Macromolecules with odd n will still typically have a pseudo-mirror symmetry in the side views. Once top-view particles and side-view particles have been located, they are aligned and averaged to generate class averages of the top and one side view of the particle. Since the group of side-view particles may contain a variety of side views, one dominant side view from the set is selected by the alignment routine. A preliminary 3D model is then constructed from these two orthogonal views. As expected, the quality of this model will be very poor; however, we have found it to be sufficient to achieve convergence in several test cases with real data and one test with simulated data. While this method is very fast, it is not robust. It may not provide an adequate starting model in every case. However, it is sufficiently fast that it is worth trying for any molecule with a known or suspected C_n symmetry.

Naturally, other software packages can also be used at this point to generate an initial model. EMAN can read models in a variety of formats, so other packages that provide, for example, multivariate statistical analysis techniques for low-symmetry

particles (Frank *et al.,* 1996; van Heel *et al.,* 1996) or Fourier common-lines techniques for icosahedral particles (Crowther, 1971) may be used to generate an initial model for subsequent refinement in EMAN.

Refinement Loop

As with other structural techniques like NMR and X-ray crystallography, the experimental data in electron cryomicroscopy cannot be directly inverted to generate an optimal 3D model. In cases like this, the most commonly used technique is iterative refinement. In this technique, a rough preliminary model is iteratively refined against the data. True convergence is achieved when the model remains unchanged for several successive iterations. In EMAN, we use a less restrictive definition of convergence, including the noise level of our initial data in the definition. In this definition, convergence is achieved when the FSC (Fourier shell correlation) between successive iterations stabilizes. This is not to say that the FSC between iterations must fall below a specific value. Indeed, if the data are sufficiently noise-free, the FSC may remain above 0.5 at all spatial frequencies. Rather, we require that the FSC curve between successive iterations cease to improve, no matter what numerical values the curves contain. Small, resolution-dependent variations between iterations due to high noise levels are unavoidable, even when the model has stabilized. However, this fact can be used to our advantage. Once our limited definition of convergence is satisfied, any additional changes between models are due solely to noise present in the images. This makes the FSC between successive iterations a rough measure of the resolution of the model. The point at which the FSC curve begins to fall is the point at which noise begins to strongly affect the reconstruction. Typically, the resolution at which the FSC curve falls halfway to its minimum value will correspond roughly to the final resolution of the model as determined by a *t*-test. This is not a robust measure of resolution, but is useful for preliminary estimates as the refinement progresses.

While many of the individual techniques used in EMAN are conceptually similar to those used by SPIDER (Frank *et al.,* 1996) and IMAGIC (van Heel *et al.,* 1996), the overall refinement algorithm, outlined in Fig. 1, is different than that used by either of these packages. EMAN uses particle classification, but the classes generated in the refinement loop are not reference-free, as they generally are in an IMAGIC reconstruction. The projection matching routine used by EMAN to classify particles is similar

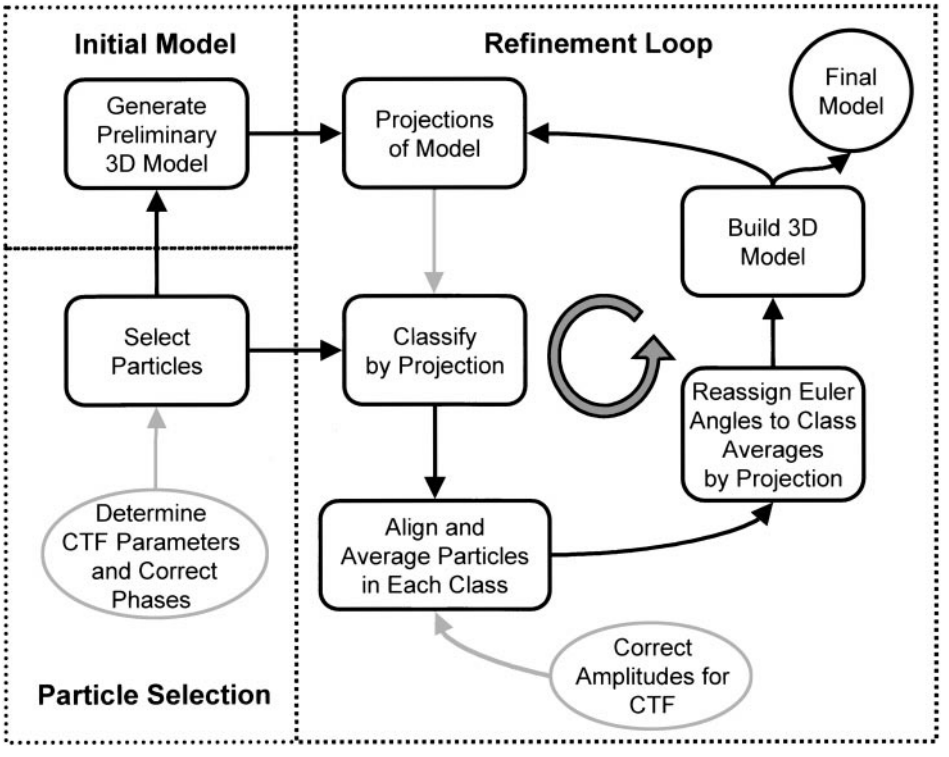

FIG. 1. A flow diagram outlining the reconstruction process in EMAN. Reconstruction occurs in three phases: particle selection, initial model generation, and model refinement. The boxed particles are used to generate a preliminary model, which is then refined iteratively using the same set of particles. The refinement loop is iterated until convergence is achieved. CTF correction (in grey) is optional. Grey line from projections indicates use as references only.

to one commonly used Euler-angle determination technique in SPIDER.

The refinement loop is outlined in the right half of Fig. 1. The refinement process begins by generating a set of N projections with uniformly distributed orientations. These projections are used as references for classification of the raw particle data. Each particle will be associated with one of the projections, resulting in N classes of raw particle images. Several techniques for classification are provided, but the most robust is performed by *classesbymra*. This program translationally and rotationally aligns each particle to each reference image and then calculates a dot product, which is used to rate the similarity of the particle to each reference.

A minimum number of projections are required to provide sufficient angular sampling for a reconstruction at a given resolution. This minimum number of projections can be estimated in various ways, but the well-known $\pi D/d$ approximation is adequate (where D is the size of the particle and d is the desired resolution). In EMAN, the number of projections is selected by the user and may be increased to improve the homogeneity of individual particle classes and provide better angular resolution. This allows the user to vary continuously between the extreme with high angular resolution and very little averaging and the extreme with very poor angular resolution, but with very consistent, low-noise class averages.

The next step is generation of class averages. The particles within a class must be mutually aligned and then averaged. While each particle is grouped with the projection it most closely resembles, this does not mean all of the particles necessarily represent usable data for this orientation. Particles may be partially denatured, in a different functional state, radiation damaged, or simply in a patch of bad ice. If a large angular spacing is used between projections, individual "good" particles may also look considerably different due to small differences in orientation. We do not wish to eliminate particles in this second class, since they represent true data for this macromolecule, and the "blurring" effect generated when they are averaged is expected. This algorithm attempts to determine which particles are good and which are bad as they are mutually aligned. For this procedure to succeed there should be a reasonable number of images in the class ($\sim$10 or more) and more than half of the particles must be good. Identification of bad particles is accomplished through a simple iterative procedure. First, each particle is aligned to the class average that was used to generate the class. The particles are then averaged together to generate an initial class average. Next, each particle is aligned to this initial average. Once all of the particles have been aligned to the average, a histogram of similarity between the average and each aligned particle is generated. The images represented by the tail of this histogram are the least similar to the class average and are not used in generating the new average. They are, however, allowed to participate in the next iteration. This process is iterated several times. The number of iterations and how similar a particle must be to the average to be included in the new average are user-defined parameters. Eventually a self-consistent class average is produced. Note that the reference projection is used only to help speed the convergence and is used only in the first round of alignment. The final average may actually appear considerably different than the initial projection the class was based on. In fact, this difference is the primary reason the overall refinement loop converges so rapidly. If CTF correction is enabled, amplitude corrections are performed automatically as the particles within a class are averaged together (see CTF and Envelope Function Corrections).

Generally, the particle classification routine does an excellent job of assigning particles to the correct class. However, in some cases the low signal-to-noise ratio in the electron micrographs will lead to misassignment of some particles. In most cases this is not a problem, and the incorrectly assigned particles will be eliminated by the alignment/averaging procedure above. In cases where one orientation is strongly preferred the problem may be more severe. We begin by considering the case where the orientations are uniformly distributed. In this case, the correctly assigned particles in each class will outnumber the incorrectly assigned particles. However, if one orientation is strongly preferred, with say 100 times more particles in a particular orientation, the misassigned particles may actually outnumber the correctly assigned particles within some classes. The alignment/averaging algorithm will still produce a consistent average in this case, but the resulting average will represent the preferred orientation rather than the correct orientation. This problem is relatively rare, even in cases where a preferred orientation exists, but it will have serious consequences if it does occur. To deal with this potential problem, once the class averages have been generated, their Euler angles may be redetermined by projection matching. This assignment will be more accurate than the initial classification since the class averages have a much higher signal-to-noise ratio than the individual particles.

The class averages with assigned Euler angles are then used to construct a new 3D model for the next round of refinement. EMAN uses a Fourier space reconstruction routine, which is extremely fast and quite robust for most systems. This is a conceptually

simple method. Since the orientation of each class average is known, each class average is simply inserted into the Fourier volume in the correct orientation, passing through the center of the volume. Each inserted class average is weighted by the number of particles used to generate it. When all of the class averages have been averaged into the Fourier volume, an inverse Fourier transform yields the reconstructed model in real space.

Once the new 3D model has been generated, the overall process is iterated. The convergence of the refinement process can be observed by examining the FSC between the 3D models (Harauz and Heel, 1986) generated in each iteration. This curve will stabilize once the refinement loop has converged. It will also generally give a good estimate of the resolution of the 3D models as described above. An additional resolution measurement can be made by performing a simple two-way T test: the raw particles are split in half and each half is reconstructed separately. There are different definitions for the FSC cut-off that defines the resolution of the model, but we prefer the more conservative 0.5 FSC cut-off (Böttcher *et al.,* 1997). This method may underestimate the reliability of the model somewhat, since the quality of the class averages in EMAN depends nonlinearly on the number of particles used to generate each class. That is, dividing the number of particles in a class by 2 may generate a class average that due to alignment limitations, may be worse than simply using half the particles would suggest.

Symmetry Determination

A prerequisite for determining an accurate structure using single-particle analysis is knowledge of the particle's symmetry. If there is no prior information on the symmetry of the particle, EMAN provides a procedure for making a preliminary symmetry determination. The raw particle data are classified by generating a set of rotational invariants and using a standard k-means iterative classification procedure. The particles in each class are then mutually aligned and averaged, using the same alignment routine described above in the refinement loop. This produces a set of reference-free class averages. Generally, visual inspection of these class averages will give a good idea of the symmetry of the model. For a quantitative assessment, a routine is provided to examine the symmetry quality of each class average for each possible simple rotational symmetry, C_2–C_{14}. Multivariate statistical analysis can also be applied at this point to provide some additional insight.

In particularly difficult cases, rather than applying this technique to high-resolution, low-contrast data, far from focus micrographs can be used instead. These data will have a very high signal-to-noise ratio at low resolution, but contain little or no usable high-resolution information. These data can be used in the above procedure and may even be carried through a complete 3D reconstruction. The goal of this exercise is to rapidly generate a low-resolution model with a high degree of confidence. This model will generally be sufficient to determine the symmetry. The symmetry may be broken if a high-resolution reconstruction is performed later, but for a new particle with very little initial structural information, this technique provides a good place to start.

Once the symmetry has been tentatively determined, a 3D reconstruction is performed with this symmetry imposed. Two methods are then used to test the accuracy of the applied symmetry. First, projections of the final model are compared with the corresponding class averages. If the assigned symmetry is too high, several class averages will not match any of the projections. If the assigned symmetry is too low, say C_3, when the true symmetry is C_6, duplicates of each class will appear in the set of projections. In the second test, the symmetry in the refinement loop is relaxed, and the refinement is run for several additional iterations. If the symmetry assignment was accurate, the refined model will remain essentially unchanged. Generally, some small symmetry-breaking mass movement will occur in this process, but the overall appearance of the model should still maintain its original symmetry. If the symmetry assignment was incorrect, the model will undergo dramatic changes in the first or second iteration after the symmetry is relaxed. Since refinement progresses much more rapidly and requires fewer particles with some applied symmetry, this procedure is more robust than the alternate method of starting with no applied symmetry and then applying the predicted symmetry in the second step.

CTF and Envelope Function Corrections

The goal of single-particle reconstruction is to regenerate the correct three-dimensional structure of a molecule based on two-dimensional projections of the molecule. Unfortunately, the images generated by electron microscopes are not true projections of the specimen. They suffer from a set of artifacts including the CTF and the envelope function of the microscope (Erickson and Klug, 1970; Hanszen, 1971). In addition, noise is present from a variety of sources. If we assume that astigmatism and drift are negligible, these effects are all isotropic. The CTF in particular can cause serious artifacts in a 3D reconstruction. Without CTF correction, the model produced by a reconstruction may contain significant local mass displacements. The high-resolution struc-

ture will generally be severely distorted, and the lack of low-resolution amplitude correction can cause effects such as causing a solid object to appear hollow. The severity and type of these effects, of course, depend on many factors, the most significant of which is the defocus of the data used in the reconstruction. Nonetheless, if a reconstruction that is truly representative of the real structure is desired, CTF correction is crucial.

CTF correction has been routine in electron crystallography and helical reconstructions for some time (Amos *et al.,* 1982; Crowther *et al.,* 1996; Henderson *et al.,* 1986; Jeng *et al.,* 1989; Mimori *et al.,* 1995), but it is only now becoming widespread in single-particle work. This is principally due to the difficulty in performing this correction on noncrystallographic data. Several groups have proposed and/or performed CTF correction on single-particle data to various degrees in the past (Böttcher *et al.,* 1997; Conway *et al.,* 1997; Zhou *et al.,* 1998; Zhu *et al.,* 1997), but the existing correction software is still quite difficult to use.

CTF correction can be performed with varying degrees of completeness, ranging from simple truncation of the data at the first zero crossing of the CTF to complete amplitude and phase correction performed simultaneously on multiple data sets. The EMAN reconstruction procedure incorporates complete CTF amplitude and phase correction that is nearly fully automated. A graphical utility, *ctfit,* allows the user to determine the CTF parameters for the particles in each micrograph. Multiple micrographs with a range of defocus values should be used to sample Fourier space as uniformly as possible. *Ctfit* stores the CTF parameters in the header of each particle and performs phase correction. The complete set of particles from the entire set of micrographs is then combined into a single file for reconstruction. CTF amplitude correction is then performed transparently as part of the reconstruction procedure. To provide a complete description of how EMAN performs CTF correction the mathematical methodology will be discussed in some detail.

The envelope function and CTF are best examined in Fourier space. At this point in the discussion, we will ignore the effects of drift and astigmatism, which cause asymmetries in the CTF and envelope functions. The Fourier transform of the true 3D structure of the molecule is called the structure factor of the molecule, $\overline{F}(s, \theta, \phi)$, where the overbar indicates a complex valued function. As discussed earlier, in Fourier space, a projection of the 3D structure is represented by a slice passing through the origin in Fourier space. This means CTF and envelope functions that are circularly symmetric in the images are also spherically symmetric when extended to 3D. This is a crucial point if we wish to correct for the CTF during a reconstruction. The final goal of the reconstruction is to produce $\overline{F}(s, \theta, \phi)$. The data measured in a transmission electron microscope can be described by

$$\overline{M}(s, \theta, \phi) = C(s)E(s)\overline{F}(s, \theta, \phi),$$

where $\overline{M}(s, \theta, \phi)$ is the measured data, $C(s)$ is the CTF, $E(s)$ is the envelope function, and $\overline{N}(s, \theta, \phi)$ represents random noise with a consistent spectral amplitude profile. If we wish to obtain $\overline{F}(s, \theta, \phi)$, we must first know $C(s)$ and $E(s)$. $\overline{N}(s, \theta, \phi)$ cannot be subtracted directly, since all we know is its mean intensity. We rely on averaging to reduce the mean noise level, which we monitor by examining the signal-to-noise ratio as the reconstruction progresses. To determine $C(s)$ and $E(s)$, it is more convenient to examine the rotationally averaged power spectra

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

where $N(s)^2$ is the mean noise intensity and $F(s)^2$ is the one-dimensional structure factor of the molecule (rotationally averaged power spectrum). To perform the 3D correction, we must determine $C(s)$ and $E(s)$ for each micrograph. To properly weight the data between micrographs, we must also know $N(s)$.

Parameter Determination

We know $M(s)$ for each micrograph and $C(s)$, $E(s)$, and $N(s)$ can all be parameterized based on theoretical or empirical models as shown below. In two-dimensional crystals or helical arrays, $N(s)$ can be determined independently by examining the background between the crystallographic peaks or layer lines. In single-particle processing, the signal is distributed continuously throughout Fourier space rather than in discrete peaks, making this approach impossible. However, since $C(s)$ varies sinusoidally, information about $N(s)$ can be obtained from the zero crossings of $C(s)$. Even with this fact, the overall fitting problem is impossible for a single data set without first determining $F(s)$.

The functional forms used for $C(s)$, $E(s)$, and $N(s)$ are

$$C(s) = A(\sqrt{1 - C_A^2}\sin(\gamma) + C_A\cos(\gamma)),$$

$$\text{where } \gamma = 2\pi\left(\frac{C_s\lambda^3 s^4}{4} + \frac{\Delta Z\lambda s^2}{2}\right), 0 \le C_A \le 1$$

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

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

This gives a total of eight parameters used for fitting: A, C_A, ΔZ, B, and n_{1-4}. The model for $N(s)$ is completely empirical and encompasses a wide range of different physical effects, including incoherent scattering, film noise, and scanner noise. This model has worked well with data from four different microscopes (JEOL 1200EX, 4000EX, 2010F, and 3000SFF), but could potentially require modification for other microscopes. In particular, we anticipate microscopes with energy filters to produce a substantially different noise distribution. $C(s)$ is provided by the well-known weak phase approximation with amplitude contrast corrections (Erickson and Klug, 1970). A five-term theoretical expression exists for $E(s)$, encompassing microscope parameters as well as effects such as specimen movement. In practice, however, a simple Gaussian is experimentally indistinguishable from this aggregate expression in most cases. It is also worth noting that there are currently no parameters to compensate for astigmatism. *Ctfit* is capable of measuring astigmatism angle and defocus difference, but no corrections are currently made for this effect. As single-particle reconstructions approach atomic resolution, however, the astigmatism present even in excellent micrographs may become significant. EMAN can be easily modified to compensate for this, but a considerable speed penalty would be imposed. In addition, specimen charging will induce apparent defocus changes and astigmatism within a single micrograph, and beam tilt will cause resolution-dependent phase shifts. These effects are ignored at present, beyond avoiding particles in micrograph areas with obvious charging.

As mentioned above, before $C(s)$, $E(s)$, and $N(s)$ can be determined accurately from individual micrographs, we must determine $F(s)^2$. Once this has been accomplished, parameter determination becomes a straightforward fitting problem. We will discuss three methods for determining $F(s)^2$. The first method is to perform a solution X-ray scattering experiment on the macromolecule under study (Schmid *et al.,* 1999; Thuman-Commike *et al.,* 1999). This provides an isotropically averaged one-dimensional structure factor that will be nearly identical to $F(s)^2$ above. There may be discrepancies due to the differences in X-ray and electron scattering cross sections for specific charge distributions (Mitsuoka *et al.,* 1999). In addition, the solution scattering experiment will provide a structure factor that is completely isotropic. If the orientations of the single particles in the electron cryomicroscopy experiment are not distributed uniformly, the structure factors may contain noticeable differences. Even in this case, however, the structure factor will generally be sufficient to determine $C(s)$, $E(s)$, and $N(s)$.

The second method for determining $F(s)^2$ is to take one or more micrographs of the sample in ice deposited on a layer of continuous carbon. By selecting areas of the micrograph containing only continuous carbon, the CTF parameters can be determined directly by assuming a one-dimensional structure factor for the carbon film (approximated as a constant for low to intermediate resolutions). These parameters can then be used to determine the structure factor of the macromolecule in the same micrograph. That is, we determine $C(s)$ and $E(s)$ from the carbon film alone and then use these parameters with the macromolecule data in the same micrograph to determine $F(s)^2$ of the macromolecule. Of course, this $F(s)^2$ will be very inaccurate near the zeros of the CTF in an individual micrograph, so an average $F(s)^2$ must be determined by combining the information from several micrographs at different defocus settings.

The final method for determining $F(s)^2$ is the most general and requires no additional data collection, but is the most difficult to implement. $F(s)^2$ is a function solely of the macromolecule being studied. That is, it is the portion of the data that is constant when microscope parameters such as defocus are varied. When data taken at multiple defocuses are fit simultaneously, this fact provides an additional constraint, which makes the fit feasible. As long as the distribution of particle orientations remains fairly consistent, any micrographs of the same protein will do. In the most general case, this would still be an underdetermined fitting problem. However, the oscillatory nature of the CTF combined with a monotonically decreasing envelope function provides the necessary additional constraints to make the fitting problem converge. After the fit has been performed once, with several micrographs, the resulting $F(s)^2$ can be used to determine $C(s)$, $E(s)$, and $N(s)$ for any number of additional micrographs. While relative amplitudes and B-factors can be determined, there is always an arbitrary overall scaling factor and an arbitrary overall B-factor that cannot be determined without a known reference. Currently this technique requires some manual fitting to be performed, but a fully automated solution may be possible. Despite the difficulties, we have successfully applied this method to several problems in which results from one of the other methods were also available, and the $F(s)^2$ determined by the two methods matched extremely well.

Performing the Corrections

Once the parameters have been determined, the actual CTF correction is performed when aligned 2D single-particle images are averaged to generate a class average. The averaging is performed in Fourier space. The basic method is to perform a weighted

average of the images, where the weights vary with spatial frequency. We wish to choose the weights such that data are used optimally; that is, the signal-to-noise ratio is maximized in the final average image at all spatial frequencies. The parameters for $C(s)$, $E(s)$, and $N(s)$ provide a measure of signal-to-noise ratio as a function of spatial frequency for each individual particle. Since $F(s)$ is the same in all of the data for a given particle, we can eliminate it and define the relative signal-to-noise ratio for each particle as $R_n(n) = [C_n(s)^2 E_n(s)^2]/[N_n(s)^2]$, where the subscript, n, denotes particle number. To simplify the expression, we assume that $N_n(s)$ is approximately the same between exposures. The solution in which $N_n(s)$ is allowed to vary is somewhat more complex, but is easily derived. Note that some inaccuracy in $N_n(s)$ will not have a significant effect on the accuracy of the results; it will simply cause slight over- or underweighting of certain particles. In practice, this effect is negligible for data collected on a single microscope. The final class average is expressed as a simple linear combination of the individual aligned images

$$T(s, \theta) = \sum_n k_n(s) M_n(s, \theta),$$

where s and θ represent polar coordinates in Fourier space, and $k_n(s)$ are the weighting coefficients to be determined. The signal-to-noise ratio of the class average, $T(s,\theta)$ is then just

$$R_T(s, \theta) = \sum_n k_n(s) R_n(s, \theta).$$

We wish to determine $k_n(s)$ such that $R_T(s)$ is maximized at all s, $\Sigma_n C_n(s) E_n(s) k_n(s) = 1$ at all s and that the CTF and envelope functions are corrected. The result of this maximization is

$$k_n(s) = \frac{1}{C_n(s)E_n(s)} \frac{R_n(s)}{\sum_m R_m(s)} = \frac{C_n(s)E_n(s)}{\sum_m C_m(s)^2 E_m(s)^2}.$$

That is, for an optimal class average, the weighting coefficients, k, are proportional to $R_n(s)$, the relative signal-to-noise ratio within each image. This methodology makes optimal use of the available information in all of the images. When an image contains no information at a particular spatial frequency, it does not contribute to the final image. In addition, this technique is relatively insensitive to small inaccuracies in the CTF model and/or parameters. The typical effect of fitting inaccuracies would be slightly over- or underweighting a particular image when averaging. Since all the images represent the same structure, this simply causes a slight reduction in the statistical definition of the result.

TESTING

Each routine in EMAN has been subjected to a variety of tests, with both real and simulated data. Discussing all of these tests would make this paper prohibitively long, so we will limit this discussion to two of the most fundamental tests. First, a basic test of the 3D projection/reconstruction routines at high resolution is described, since this procedure is somewhat novel for the single-particle community. Second, a complete reconstruction test using simulated data is presented. EMAN has successfully reproduced and, in fact, improved several reconstructions using real data as well. Again, to limit the complexity of this paper, these results will be published separately. All of the data and specific program parameters used in the tests described here are available for download with the EMAN distribution.

EMAN uses a direct Fourier space reconstruction routine, whereas most reconstruction packages use filtered real-space methods (Frank *et al.,* 1996; van Heel *et al.,* 1996). The direct Fourier algorithm offers several advantages over typical real-space methods, the most important of which is speed. In many other reconstruction schemes, construction of a 3D model is a time-limiting step. The described Fourier algorithm is one to two orders of magnitude faster than typical back-projection methods.

To verify the robustness of EMAN's Fourier algorithm, a simple projection/reconstruction test in the presence of high noise levels was performed. A 3D electron density map of a trimer of the outer shell protein of the bluetongue virus (VP7) was generated from PDB data (1BVP). The PDB model contains a dimer of trimers (Grimes *et al.,* 1995). The electron density map of a single trimer was generated at 3-Å resolution. Two thousand five hundred projections of this model were generated using real-space projection with trilinear interpolation. Noise was then added to each projection at typical levels for electron cryomicroscopy. Figure 2 shows representative projections with and without noise and a plot of the mean signal-to-noise ratio in the noisy particles. No CTF or envelope function was applied in the first test. All 2500 projections with predefined Euler angles were then used to generate a 3D model. A Fourier space reconstruction routine was employed with $2 \times 2 \times 2$ voxel Gaussian interpolation and 20% zero padding. For quantitative comparison, the FSC between the original and the reconstructed models was calculated as a function of spatial frequency (Fig. 3, black line). Naturally, with the added noise, a perfect reconstruction is not possible. Some noise will always be present in the final model. To provide

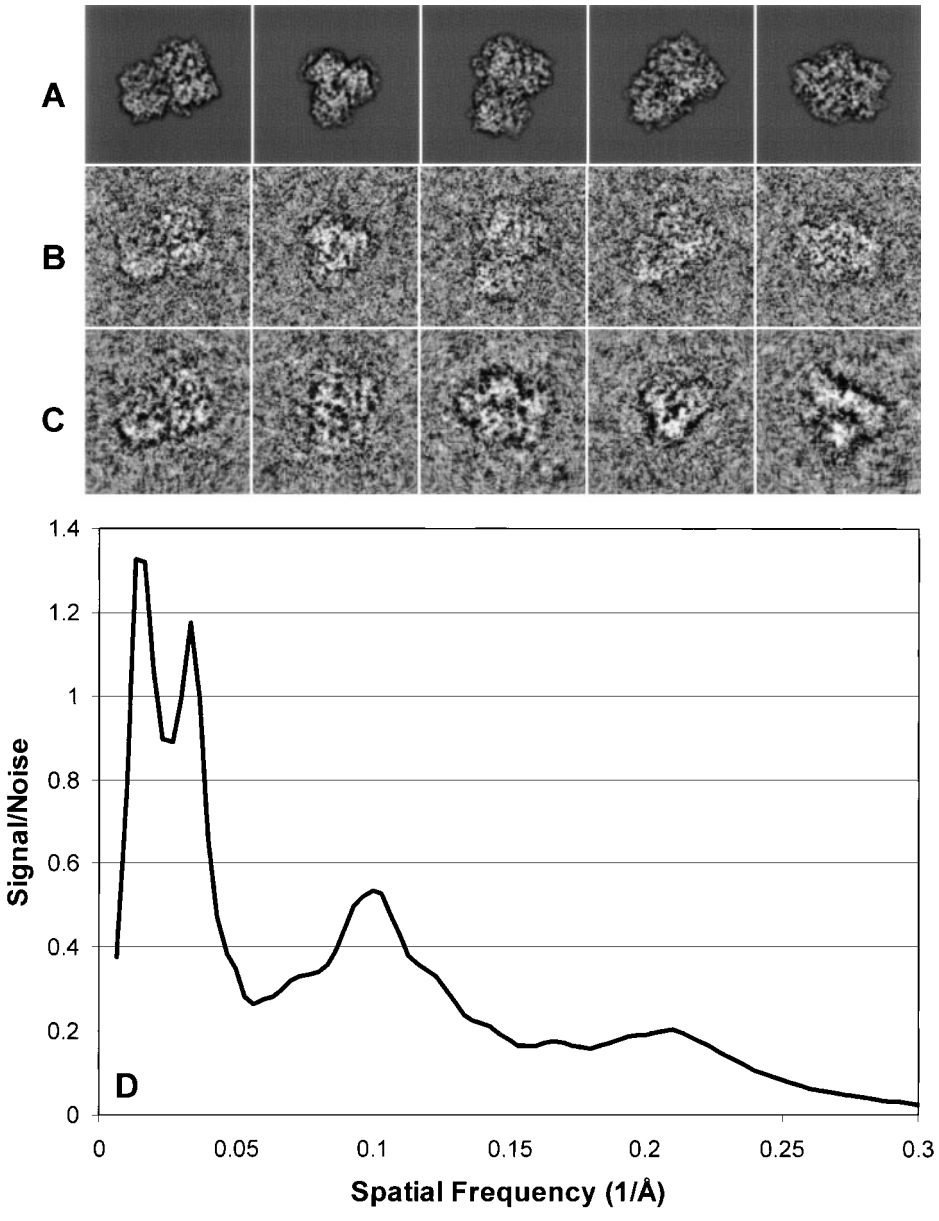

FIG. 2. Particles used for both of the conformance tests described in the text. (A) A few raw projections of the bluetongue virus capsid trimer from PDB data. (B) The same projections with added noise selected from those used in the first test. (C) Representative projections with applied CTF, envelope function, and noise. Each particle shown represents a different simulated defocus. These are a few of the particles used in the second test. (D) The mean signal-to-noise ratio in the individual images shown in B. The curve represents the structure factor of the macromolecule divided by the applied noise level.

a standard for comparison, the signal-to-noise ratio of the original particles was used to calculate the maximum FSC that could be obtained by an ideal reconstruction algorithm (Fig. 3, gray line). While there is a one-to-one relationship between FSC and signal-to-noise ratio, the relationship contains an integral with no analytical solution, so the conversion is calculated numerically. Note that this calculation does not account for the numerical errors induced by the trilinear interpolation used in generating projections, so this curve actually slightly overestimates the best possible FSC. Even without accounting for this affect, the Fourier space reconstruction routine performs extremely well. Also shown in Fig. 3 are two representative views of the original and reconstructed models at 4-Å resolution. Clearly, the two models are nearly indistinguishable.

In the second test, the overall reconstruction routine, including CTF correction, was tested with simulated data. The same BTV trimer was used as a

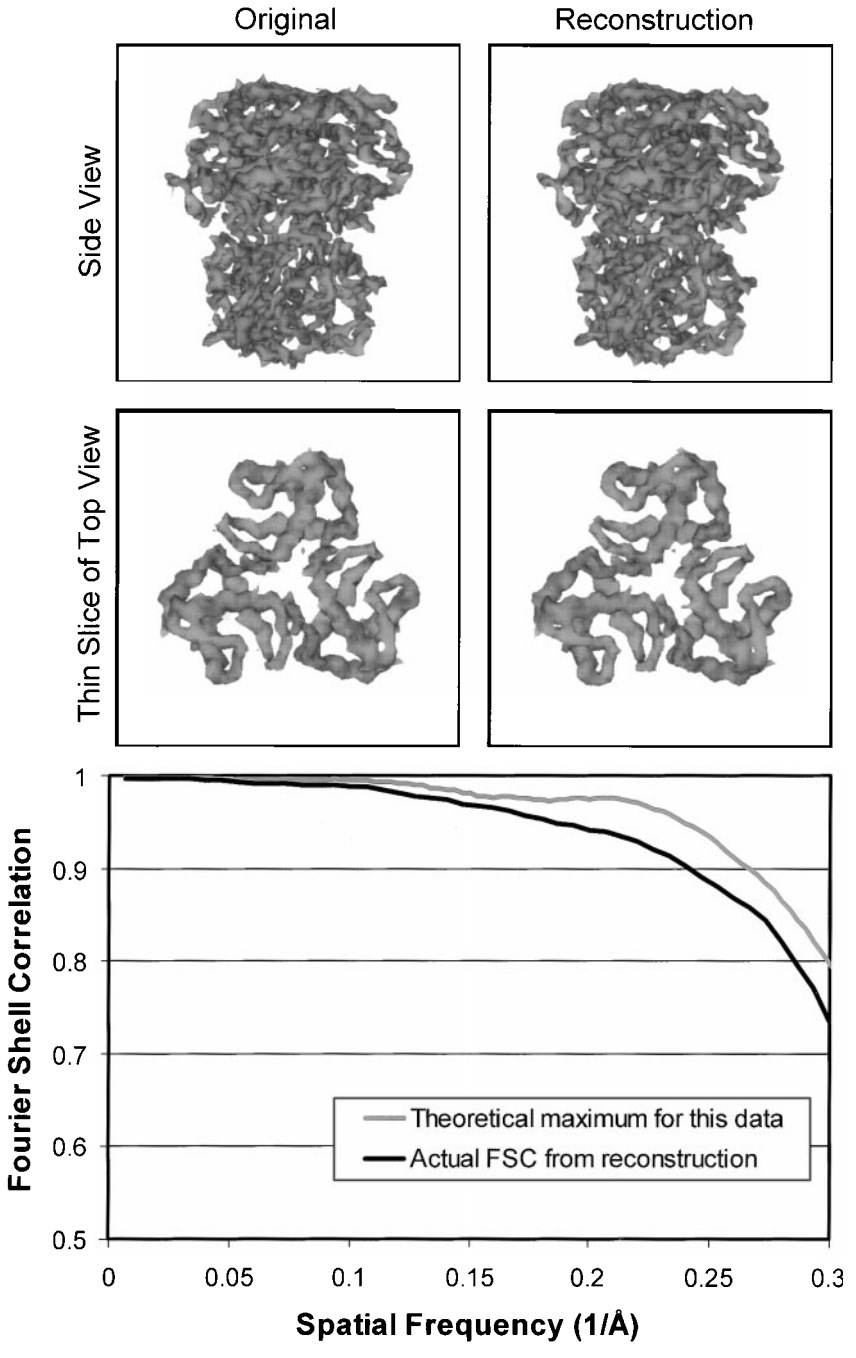

FIG. 3. Results of the first test. A side view and a thin slice of the top view are shown for the original and the reconstructed models. The plot shows the FSC between the reconstruction and the original model. Also shown is the maximum FSC possible with an ideal reconstruction routine with the given data set at the noise level shown in Fig. 2.

test model. In this test, five sets of 500 random projections were generated. To make the test realistic, we used CTF and noise parameters taken from experimental measurements made on a JEOL 2010F electron microscope with a field emission gun operated at 200 keV with an objective lens spherical aberration coefficient of 0.5 mm. Parameters were determined from a measurement of carbon film fit using *ctfit*. These parameters were then applied to each of the five data sets, with each set simulating a different defocus. Figure 2C shows representative projections with noise and applied CTF at various defocus settings. A reconstruction was then performed as described above. An initial model was generated using the simple symmetry search algorithm for a single three-fold rotational symmetry.

This model was then refined against the complete data set with CTF correction. Figure 4 shows the results after zero to three refinement iterations. The FSC indicates a final resolution of ~5 Å, close to the theoretical limit with the applied envelope function and noise levels in the original particles. The structures appear nearly identical at this resolution.

INDIVIDUAL PROGRAMS

EMAN contains a number of individual programs that are useful for general purpose electron microscopy image processing. All of the necessary processing steps in the reconstruction procedure are embedded in a small number of straightforward commands. A graphical user interface is provided to allow rapid analysis of the results and provides instructions to take a novice user through the reconstruction systematically. For advanced users, lower level commands, as well as a complete set of C++ classes, are provided to allow custom processing in cases where the general reconstruction procedures are insufficient. For a typical macromolecule, a refined 3D structure can be generated from boxed out particles with a sequence of three commands. Aside from evaluating results, no user interaction is required during the refinement procedure.

Despite its relative ease of use, generating a final structure and verifying its correctness are not en-

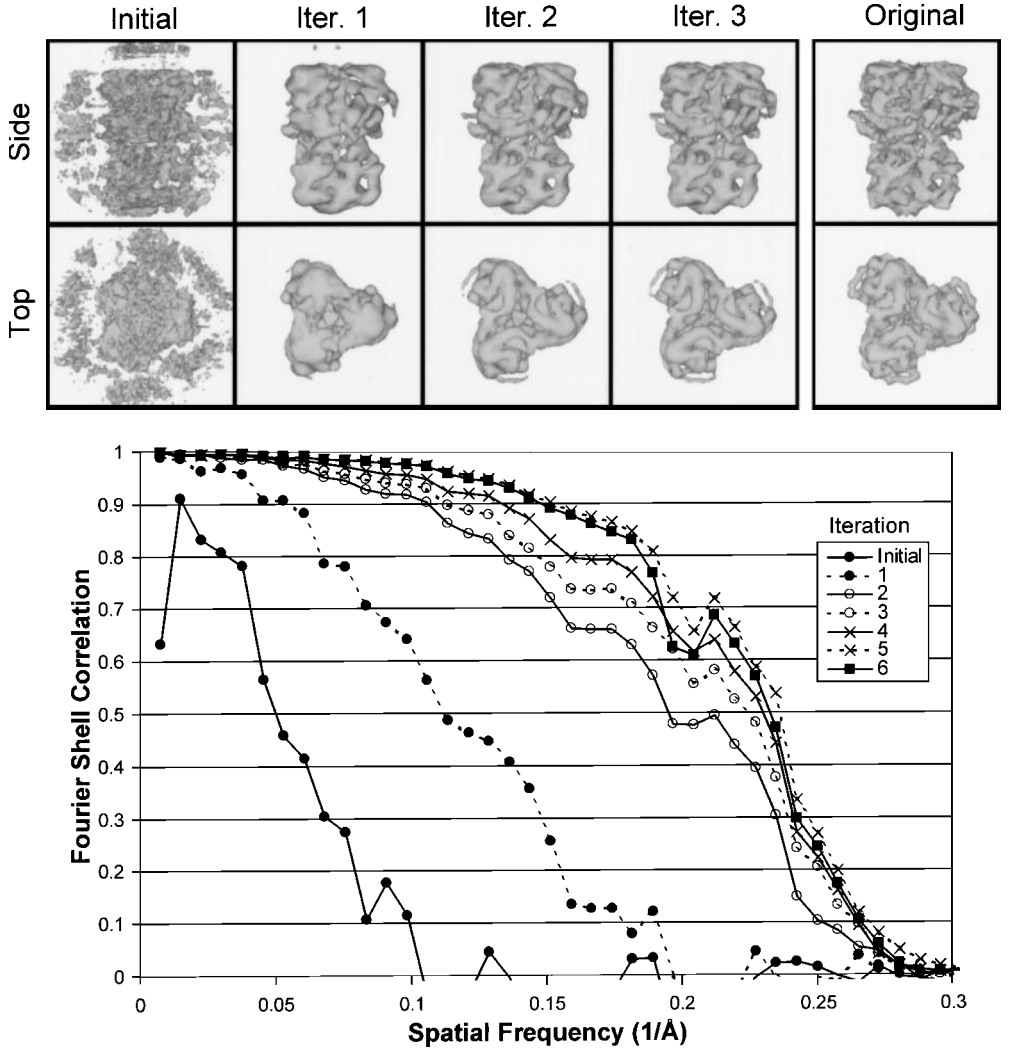

FIG. 4. Results of the second test. In this test, a complete reconstruction was performed with 2500 particles of simulated data. The plot shows the FSC between the model after each iteration and the "true" model used to generate the original particle data. The plot demonstrates that the refinement converged by the fourth iteration with a final resolution of ~4.5 Å. Representative views of the model after each iteration are shown above the plot. The leftmost views represent the preliminary model generated by the automatic symmetry search algorithm described in the text. The rightmost model is the true model. Models after iterations 1–3 are shown in the middle. All models were smoothed to 6 Å with a Gaussian filter.

tirely trivial. A novice user may, for example, overestimate the quality of an early model of a new macromolecule or be unaware of the pitfalls that may exist in making initial symmetry assumptions. Even with fairly automated software, user experience will play a role in generating a robust model. In the future, EMAN will be expanded to include additional routines to ease the analysis procedure and provide additional user guidance. The initial version of the software is usable by novices, but they should expect a considerable amount of trial and error before they begin obtaining robust models.

The graphical user interface in EMAN is packaged in several separate programs. Screenshots of two graphical programs are shown in Fig. 5. The main program, called *eman,* allows graphical browsing of images and command history. Every time an EMAN command-line program is executed, a log-file entry is made, recording the exact command parameters, command run-time, and accessed files. This is performed automatically with no additional effort on the part of the user. This command history can be interactively browsed and searched with *eman.* This provides a mechanism for recording the history of a reconstruction, as well as allowing the user to easily rerun commands with small variations. *Eman* also contains a complete image browser and specific tools for evaluating the results of a reconstruction. Individual 2D or 3D image files can be viewed directly in a window in *eman* with interactive brightness, contrast, and inversion controls. Additionally, when intermediate files from a reconstruction are viewed, *eman* provides a variety of useful features including comparison of projections with class averages and individual images in a class with reference projections.

As mentioned above, EMAN contains a program called *ctfit,* which allows parameters necessary for CTF correction to be interactively determined. This program also acts as a complete CTF simulator for determining good defocus values for future microscope sessions and comparing parameters from different microscopes. It also has the ability to apply CTF, envelope function, and/or noise to a set of images to generate simulated particles for testing purposes. With a 3D model, this can be used to predict how a given particle will appear visually on a microscope in ice.

Helixhunter and *sheethunter* are nongraphical programs that facilitate the location of alpha helices and beta sheets in 4- to 8-Å reconstructions. *Proc3d* is a 3D processing program that can perform most common image processing operations on volume data, such as filtering, centering a model, calculating a difference map, and scaling. In addition, it can take pairs of models and calculate Fourier shell phase difference, Fourier shell correlation coefficients, *R*-factors, etc. *Align3d* is provided to automatically perform rotational alignment of two 3D models.

For isosurface visualization, an excellent solution is Vis5D (http://www.ssec.wisc.edu/~billh/vis5d.html), a freely available program written by the Space Science and Engineering Center at the University of Wisconsin, Madison. EMAN includes *mrc2v5d* and *mrc2v5dt,* utilities to convert 3D MRC models to the format used by Vis5D. Vis5D provides extensive tools for visually comparing models, generating contour slices, creating small animations, and dozens of other useful features.

EMAN is provided completely free of charge with full C++ source. No modification is necessary for compilation on SGIs or Linus based machines. It was written with portability in mind and should compile with relatively little effort on other Unix based machines. Graphical programs were written using the freely available QT toolkit (http://www.troll.no/), which compiles on virtually all Unix based machines. Fourier transforms are performed with the free FFTW library (http://www.fftw.org/), which provides performance comparable to that of machine-specific optimized code (http://www.fftw.org/benchfft/). While the MRC and IMAGIC file formats are used by default, EMAN can also read and write a variety of other formats, including SPIDER, TIFF (currently only 8 bit), GIF, etc. Machine byte order swapping (MSB first vs LSB first) is handled automatically by the image reading functions. Users are encouraged to modify our examples and write their own utility programs. Any useful programs may be submitted to the NCMI for inclusion in future releases of the package. Complete documentation, sample data, and the software itself can be obtained on the NCMI Web site: http://ncmi.bcm.tmc.edu/.

DISCUSSION AND CONCLUSIONS

EMAN was written with the ambitious goal of eventually generating 3- to 5-Å resolution reconstructions from single-particle data. The largest problem in generating such a high-resolution reconstruction is the rapid increase of the number and size of particles such as reconstruction requires. Clearly, the amount of image data scales as the square of the reciprocal of the resolution. That is, the individual particles for a 5-Å reconstruction will be twice the size and contain four times as much data as the particles used for a 10-Å reconstruction. Not only do the individual particles become larger, but the required angular accuracy scales linearly with the reciprocal of the resolution. Since there are two Euler angles, this means the number of reference projections used for Euler angle determination also

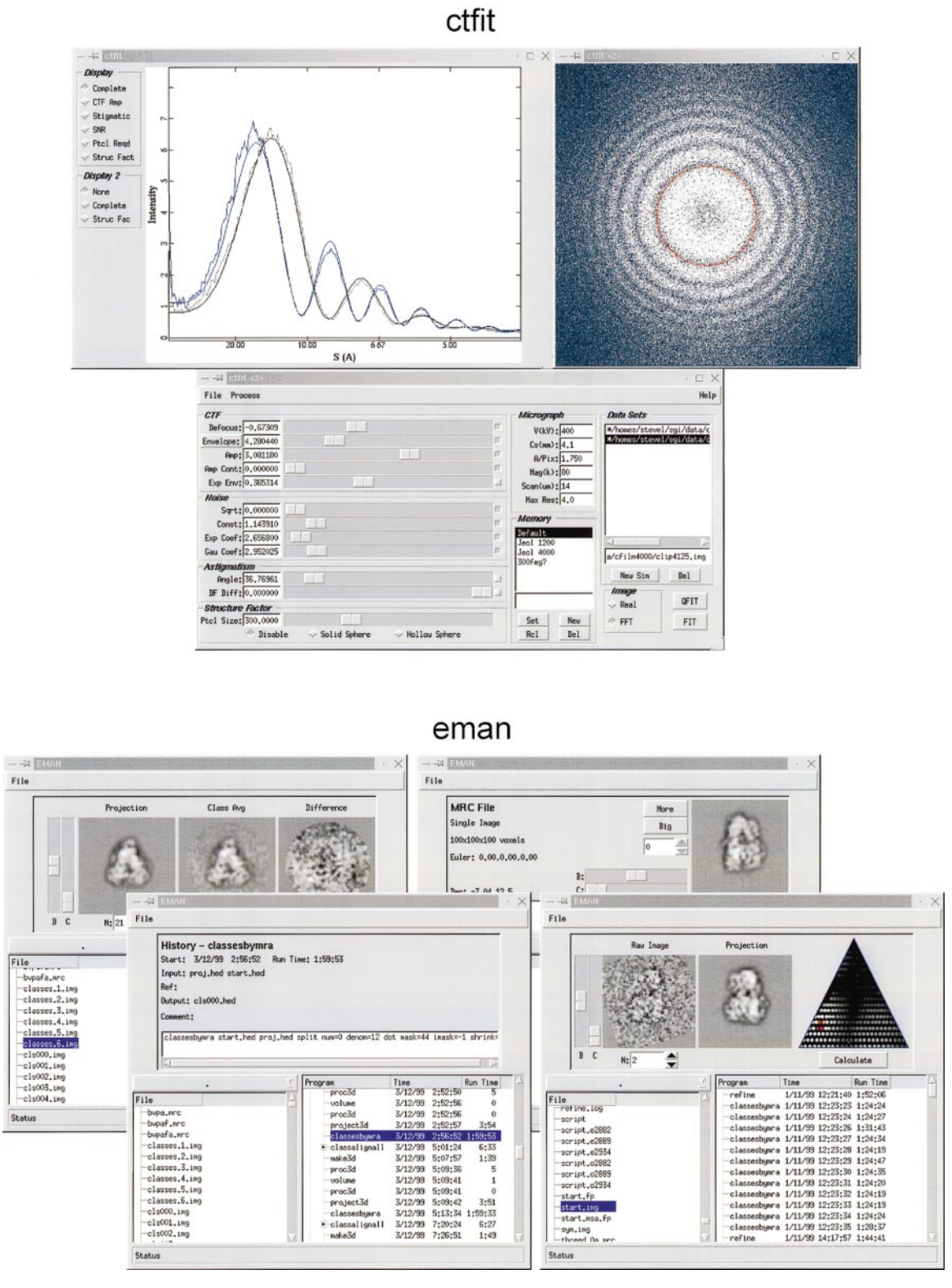

FIG. 5. Sample screenshots of two of the graphical programs included with the EMAN package. *Ctfit* is shown performing a fit of two carbon film images. *Eman* is shown in four of its many different display modes.

scales as the square of the reciprocal of the resolution. This factor does not cause an increase in the required data, but does cause an increase in the processing time for a reconstruction.

In addition to these two factors, we must also take into account the increase in number of particles required due to the low signal-to-noise ratio in the data. There is a direct correspondence between FSC between two independent data sets and signal-to-noise ratio. This means that the number of required particles is inversely proportional to the signal-to-noise ratio at the desired resolution. Since the envelope function causes the signal-to-noise ratio to fall off as e^{-2Bs^2}, the number of particles required rises very rapidly beyond $s = \sqrt{1/B}$. This makes the quality of the microscope imaging system extremely important. For example, with images with $B = 200$ Å^2, going from 20- to 10-Å resolution will require a factor of 20 increase in the number of particles. However, if $B = 100$ Å^2, the same resolution improvement will require only a factor of 4 increase in particle count. In addition, the signal-to-noise ratio is also proportional to the structure factor of the particle. Aside from localized peaks, the structure factor of most macromolecules falls off rapidly from 100 to 10 Å, requiring a further increase in the number of particles (Thuman-Commike *et al.,* 1999).

Combining these factors, we immediately see that processing becomes a limiting factor very rapidly at high resolution. Even if we discount the structure factor, a 5-Å reconstruction with $B = 50$ Å^2 data will require approximately 300 times more computational time than a 10-Å reconstruction with the same data quality. If the 10-Å reconstruction required, for example, a day of processing time, the 5-Å reconstruction would take nearly a year of continuous processing. Improving B to 25 Å^2 would reduce this time to ~2–3 months. Even so, improvements in processing efficiency and availability of parallel processing are critical for higher resolution reconstructions to be achieved.

EMAN is already performing reconstructions in 24 h that previously required a month of user-intensive work. Several novel techniques for rapid projection matching promise another order of magnitude in speed improvement in the near future. In addition, all of the time-limiting steps in the reconstruction have been parallelized for use on shared-memory supercomputers. Single-particle reconstruction is ideally suited to parallel processing, since in most steps the particles or classes can be treated independently, allowing very coarse-grained parallelization. Parallel code is currently implemented using the pthreads library supported by virtually all UNIX machines. Support for distributed processing on clusters of machines should exist by the next release.

One assumption that is frequently made when approximating the feasibility of high-resolution reconstructions is that data with an arbitrarily small signal-to-noise ratio can be recovered if a sufficient number of particles are averaged together. Unfortunately, this is not the case. When the signal-to-noise ratio falls below a particle specific value, alignment routines will no longer be able to accurately align the data at high resolution. If the alignment is not sufficiently good, the data at the corresponding resolution will no longer add coherently, no matter how many data are provided. A very rough estimate, based on our experience with EMAN and various microscopes, any data with a signal-to-noise ratio less than ~0.02 will be unrecoverable. With this requirement and typical dose conditions of ~10 e^-/Å^2, we estimate that $B \approx 2d^2$ provides a reasonable estimate of the largest B capable of providing a reconstruction at a given resolution, d, given an unlimited number of particles. This approximation, of course, ignores the structure factor of the macromolecule, which can have a significant effect. *Ctfit* provides the tools necessary to make a more precise estimate for a specific macromolecule on a specific macromolecule on a specific microscope.

Another factor to consider in performing high-resolution reconstructions is CTF correction. When collecting data for reconstructions beyond 10 Å, it is virtually impossible to avoid multiple oscillations of the CTF. Pictures would have to be taken so close to focus that there would be virtually no contrast in the images, and low-resolution information would be almost completely absent. A more reasonable approach is to combine further from focus images, which provide much greater total signal-to-noise ratio, with a robust CTF correction algorithm. However, even with a robust algorithm, there are limits to the defocus that can be successfully used at a given resolution. When the defocus becomes such that multiple CTF oscillations occur from one pixel to the next at high spatial frequency in Fourier space, reliable corrections can no longer be made. In most cases, this turns out to be a relatively modest requirement. For example, for the 200-keV scope used in earlier examples, defocuses as high as 1.5 μm (first zero of the CTF at ~20 Å) should still produce data correctable to 5–6 Å.

In summary, EMAN provides the tools necessary for achieving the next level of resolution in single-particle reconstructions, while also allowing low-resolution reconstructions to be performed rapidly and routinely. As data with $B = 50$ Å^2 or better become available, single-particle reconstructions at below 10 Å resolution should become routine. As resolution improves, EMAN will continue to be im-

proved to provide an up-to-date tool for single-particle reconstructions.

We thank Joanita Jakana and Irina Serysheva for providing much of the experimental data used when writing and testing the package. This project was supported in part by Grants P41RR02250, F32AR08474 (to S. J. Ludtke), NLM2T15LM07093 (to P. R. Baldwin), and RO1AR41729, by the Robert A. Welch Foundation, and by the W. M. Keck Center for Computational Biology.

REFERENCES

Amos, L. A., Henderson, R., and Unwin, P. N. (1982) Three-dimensional structure determination by electron microscopy of two-dimensional crystals, *Prog. Biophys. Mol. Biol.* **39,** 183–231.

Baker, T. S., and Cheng, R. H. (1996) A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy, *J. Struct. Biol.* **116,** 120–130.

Böttcher, B., Wynne, S. A., and Crowther, R. A. (1997) Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy, *Nature* **386,** 88–91.

Conway, J. F., Cheng, N., Zlotnick, A., Wingfield, P. T., Stahl, S. J., and Steven, A. C. (1997) Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy, *Nature* **386,** 91–94.

Crowther, R. A. (1971) Procedures for three-dimensional reconstruction of spherical viruses by Fourier synthesis from electron micrographs, *Philos. Trans. R. Soc. London B* **261,** 221–230.

Crowther, R. A., Henderson, R., and Smith, J. M. (1996) MRC image processing programs, *J. Struct. Biol.* **116,** 9–16.

Erickson, H. P., and Klug, A. (1970) The Fourier transform of an electron micrograph: Effects of defocusing and aberrations, and implications for the use of underfocus contrast enhancement, *Philos. Trans. R. Soc. London B* **261,** 105–118.

Frank, J. (1989) Image analysis of single macromolecules, *Electron Microsc. Rev.* **2,** 53–74.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., and Leith, A. (1996) SPIDER and WEB: Processing and visualization of images in 3D electron microscopy and related fields, *J. Struct. Biol.* **116,** 190–199.

Frank, J., and Wagenknecht, T. (1984) Automatic selection of molecular images from electron micrographs, *Ultramicroscopy* **12,** 169–176.

Gabashvili, I. S., Agrawal, R. K., Grassucci, R., and Frank, J. (1999) Structure and structural variations of the *Escherichia coli* 30 S ribosomal subunit as revealed by three-dimensional cryo-electron microscopy, *J. Mol. Biol.* **286,** 1285–1291.

Grimes, J., Basak, A. K., Roy, P., and Stuart, D. (1995) The crystal structure of bluetongue virus VP7, *Nature* **373,** 167.

Hanszen, K. J. (1971) The optical transfer theory of the electron microscope: Fundamental principles and applications, *in* Barer, R., and Cosslett, V. E. (Eds.), Advances in Optical and Electron Microscopy, pp. 1–84, Academic Press, New York.

Harauz, G., and Fong-Lochovsky, A. (1989) Automatic selection of macromolecules from electron micrographs by component labelling and symbolic processing, *Ultramicroscopy* **31,** 333–344.

Harauz, G., and Heel, M. V. (1986) Exact filters for three-dimensional reconstruction of an object from projections, *Optik* **73,** 146–156.

Henderson, R., Baldwin, J. M., Downing, K. H., Lepault, J., and Zemlin, F. (1986) Structure of purple membrane from *Halobacterium halobium:* Recording, measurement and evaluation of electron micrographs at 3.5 Å resolution, *Ultramicroscopy* **19,** 147–178.

Jeng, T. W., Crowther, R. A., Stubbs, G., and Chiu, W. (1989) Visualization of alpha-helices in tobacco mosaic virus by cryo-electron microscopy, *J. Mol. Biol.* **205,** 251–257.

Lata, K. R., Penczek, P., and Frank, J. (1995) Automatic particle picking from electron micrographs, *Ultramicroscopy* **58,** 381–391.

Martin, I. M. B., Marinescu, D. C., Lynch, R. E., and Baker, T. S. (1997) Identification of spherical virus particles in digitized images of entire electron micrographs, *J. Struct. Biol.* **120,** 146–157.

Mimori, Y., Yamashita, I., Murata, K., Fujiyoshi, Y., Yonekura, K., Toyoshima, C., and Namba, K. (1995) The structure of the R-type straight flagellar filament of *Salmonella* at 9 Å resolution by electron cryomicroscopy, *J. Mol. Biol.* **249,** 69–87.

Mitsuoka, K., Hirai, T., Murata, K., Miyazawa, A., Kidera, A., Kimura, Y., and Fujiyoshi, Y. (1999) The structure of bacteriorhodopsin at 3.0 Å resolution based on electron crystallography: Implication of the charge distribution, *J. Mol. Biol.* **286,** 861–882.

Orlova, E. V., Serysheva, II, van Heel, M., Hamilton, S. L., and Chiu, W. (1996) Two structural configurations of the skeletal muscle calcium release channel, *Nat. Struct. Biol.* **3,** 547–552.

Schmid, M. F., Sherman, M. B., Matsudaria, P., Tsuruta, H., and Chiu, W. (1999) Scaling structure factor amplitudes in electron cryomicroscopy using X-ray solution scattering, *J. Struct. Biol.* **128,** 51–57.

Schroeter, J. P., and Bretaudiere, J. P. (1996) SUPRIM: Easily modified image processing software, *J. Struct. Biol.* **116,** 131–137.

Thuman-Commike, P. A., and Chiu, W. (1995) Automatic detection of spherical particles from spot-scan electron microscopy images, *J. Micros. Soc. Am.* **1,** 191–201.

Thuman-Commike, P. A., Tsuruta, H., Greene, B., Prevelige, P. E., King, J., and Chiu, W. (1999) Solution X-ray scattering based estimation of electron cryomicroscopy imaging parameters for reconstruction of virus particles, *Biophys. J.* **76,** 2249–2261.

van Heel, M. (1982) Detection of objects in quantum-noise-limited images, *Ultramicroscopy* **8,** 331–342.

van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R., and Schatz, M. (1996) A new generation of the IMAGIC image processing system, *J. Struct. Biol.* **116,** 17–24.

Whittaker, M., Carragher, B. O., and Milligan, R. A. (1995) PHOELIX: A package for semi-automated helical reconstruction, *Ultramicroscopy* **58,** 245–259.

Zhou, Z. H., Jakana, J., Rixon, F. J., and Chiu, W. (1998) 9 Å structure of HSV-1 B capsid, *in* 14th International Congress on Electron Microscopy, 4th ed. pp. 291–292, Cancun, Mexico.

Zhu, J., Penczek, P. A., Schroder, R., and Frank, J. (1997) Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: Procedure and application to the 70S *Escherichia coli* ribosome, *J. Struct. Biol.* **118,** 197–219.