Structure
Ways & Means

PRIME: Probabilistic Initial 3D Model Generation for Single-Particle Cryo-Electron Microscopy

Hans Elmlund,[1,*] Dominika Elmlund,[1] and Samy Bengio[2]
[1]Department of Structural Biology, Stanford University Medical School, Fairchild Building, 1st Floor, 299 Campus Drive, Stanford, CA 94305-5126, USA
[2]Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA
*Correspondence: hael@stanford.edu
http://dx.doi.org/10.1016/j.str.2013.07.002

SUMMARY

Low-dose electron microscopy of cryo-preserved individual biomolecules (single-particle cryo-EM) is a powerful tool for obtaining information about the structure and dynamics of large macromolecular assemblies. Acquiring images with low dose reduces radiation damage, preserves atomic structural details, but results in low signal-to-noise ratio of the individual images. The projection directions of the two-dimensional images are random and unknown. The grand challenge is to achieve the precise three-dimensional (3D) alignment of many (tens of thousands to millions) noisy projection images, which may then be combined to obtain a faithful 3D map. An accurate initial 3D model is critical for obtaining the precise 3D alignment required for high-resolution (<10 Å) map reconstruction. We report a method (PRIME) that, in a single step and without prior structural knowledge, can generate an accurate initial 3D map directly from the noisy images.

INTRODUCTION

Understanding molecular mechanisms of biologica macromolecules requires structural information at the highest possible resolution. The transmission electron microscope is a powerful tool for obtaining information about the structure and dynamics of large macromolecular assemblies (molecular weight of 200 kDa to 70 MDa) at intermediate to atomic resolution (3–30 Å; Fischer et al., 2010; Frank, 2006; Ludtke et al., 2004; van Heel et al., 2000; Wiedenheft et al., 2011).

In single-particle cryo-EM, a solution of biologic molecules is frozen (Adrian et al., 1984) and imaged with a low electron dose (5–15 electrons/Å^2). This reduces radiation damage (Knapek and Dubochet, 1980; Unwin and Henderson, 1975), preserves the atomic structures, but compromises the signal-to-noise ratio (SNR) of the individual two-dimensional (2D) projection images. Production of a good initial three-dimensional (3D) model ab initio from noisy images is difficult, and therefore, many high-quality cryo-EM maps are generated by alignment to initial 3D models obtained by crystallography (Fischer et al., 2010; Rosenthal and Henderson, 2003; Scheres, 2012b; Scheres et al., 2007a; Zhang et al., 2008) or small-angle X-ray scattering (Wiedenheft et al., 2011).

Standard single-particle reconstruction methodologies involve multiple steps, including grouping of images based on similarity, generation of averages with improved SNR, calculation of an initial low-resolution 3D model, and map refinement by repeated cycles of individual particle image alignment and volume reconstruction (Baker et al., 2010; Elmlund et al., 2010; Elmlund and Elmlund, 2012; Frank, 2006; Hohn et al., 2007; Ludtke et al., 1999; Scheres et al., 2008; Shaikh et al., 2008; Tang et al., 2007; van Heel et al., 1996). If the initial model does not faithfully represent the imaged molecule, the final map may be biased toward an incorrect result (Elmlund, 2010).

The use of randomized orientations as a starting point for 3D alignment has been a routine procedure for icosahedral virus reconstruction for more than a decade. More recently, it was proposed as an objective approach to single-particle reconstruction of asymmetrical molecules (Sanz-García et al., 2010; Yan et al., 2007). The initial model is projected into 2D reference images in uniformly distributed orientations. The orientations of the experimental images are determined based on matching to the reference images. The approach is deterministic in that the highest scoring—although not necessarily correct, owing to low SNR and poor model quality—orientation is assigned to each image. A new volume is calculated and the process is repeated until convergence. This approach requires extensive restarting from different random seeds. Because an erroneous map cannot be distinguished from a correct one when the structure is unknown, external validation (Henderson et al., 2011) is required.

"Weighted" orientation assignment has been shown to overcome bias introduced by a starting model that does not faithfully represent the images when applied to 2D alignment via maximum likelihood estimation (ML; Sigworth, 1998). The improved search behavior of ML algorithms is explained by the reduced capacity of the statistical model. Capacity refers to the size of the set of functions that the statistical model (the reconstruction) can be tuned to represent. Three-dimensional ML approaches, inspired by Sigworth's pioneering work, apply continuous probability distribution estimation to determine the distribution of orientations (Scheres, 2012a; Scheres et al., 2005, 2007a, 2007b). "Weighted" orientation assignment can be viewed as means for exploiting the angular information redundancy of the 3D Fourier transform at low resolution. Exploiting this redundancy alone, however, is not sufficient to accomplish ab initio 3D reconstruction in our hands—the search behavior

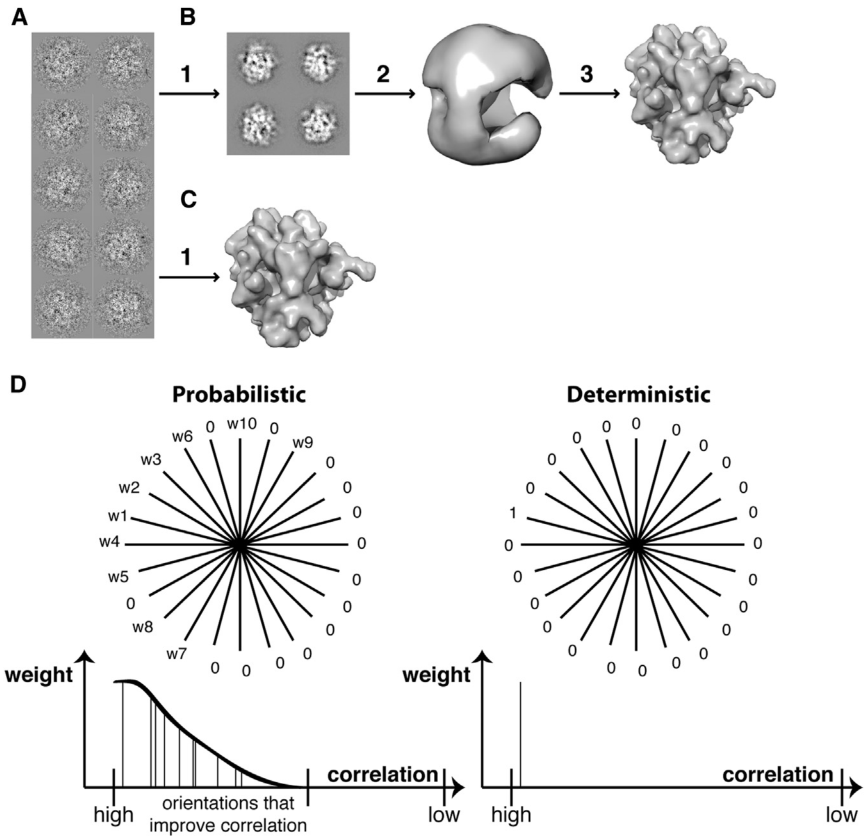

Figure 1. Schematic Representation of the Image Processing Workflow

(A) Individual ribosome cryo-EM images.

(B) The workflow of standard single-particle image processing: (1) grouping of images based on similarity and generation of averages with improved SNR, (2) generation of a preliminary low-resolution model, and (3) map refinement by repeated cycles of individual particle image alignment and volume reconstruction.

(C) The workflow of our approach: (1) generation of an accurate 3D representation of the imaged molecule in a single optimization step.

(D) Schematic representation of probabilistic versus deterministic image alignment for one single particle image. The radial lines represent 3D orientations in the discrete search space. Our probabilistic alignment assigns each particle image to a range of high-scoring orientations with weights. Sampling a weight distribution for all orientations that improve the correlation with randomly positioned delta functions generates sparse orientation weights during search. Deterministic image alignment assigns each particle image a single orientation, assumed to be the correct one.

of the algorithm must be improved as well. The 3D ML approaches developed have not been shown to converge without the use of an accurate initial 3D model (Scheres, 2012b; Scheres et al., 2007a).

We report a probabilistic initial 3D model generation procedure for single-particle cryo-EM (PRIME) that, in a single step and without prior structural knowledge, can generate an accurate initial 3D map directly from EM images. In all tested cases, the procedure converges to the correct map whether a 3D reconstruction of an unrelated molecule, a random model, or random noise is used for initialization.

Methods

Our image processing workflow is illustrated schematically and compared with a typical single-particle reconstruction protocol in Figure 1. The initial random model is projected into 2D images in evenly distributed orientations and correlations between each particle image and the reference images are calculated. Standard methods assign the single highest scoring orientation to each image. In contrast, our method assigns a range of high-scoring orientations (Figure 1D), the contribution of each image to the 3D reconstruction being determined by weight factors that are proportional to the correlation. The alignment of each image is based on the average of many likely alignments. Multiple copies of the same image are included in the 3D reconstruction by sampling all orientations that improve the correlation and assigning them different weights. Thus, the 3D reconstruction becomes a weighted average of the images assigned to each orientation. The computational demands of one iteration of our procedure correspond (on average) to about one iteration of standard deterministic 3D alignment (Baker et al., 2010; Frank et al., 1996; Harauz and Ottensmeyer, 1984; Ludtke et al., 1999; Penczek et al., 1994; Radermacher, 1994; Sanz-García et al., 2010; Shaikh et al., 2008; van Heel et al., 1996).

Probabilistic Orientation Search

Define a configuration S

$$S = (\{\psi_1, \theta_1, \varphi_1\}_1, \ldots, \{\psi_n, \theta_n, \varphi_n\}_n) \in \Omega \qquad \text{(Equation 1)}$$

of the $3n$ rotational variables of the optimization problem, where n is the number of images. The origin shift parameters are not included in the formalism for clarity. The search space Ω consists of all possible orientations and all possible combinations of orientations for the n images.

We use stochastic hill climbing to explore the search space Ω and find the alignment that gives the best 3D model, as judged by the global score given by Equation 4 below. For the data sets of 5,000 images analyzed here (see Results) this corresponds to a search of more than 25,000 parameters. While basic hill climbing always chooses the steepest uphill move, stochastic hill climbing randomly selects among the possible uphill moves (Russel and Norvig, 2003). This reduces the tendency to get trapped in local optima due to greedy acceptance of neighboring moves. The stochastic search introduces large fluctuations capable of rearranging major parts of a system. A neighborhood $\Gamma(S)$ maps S to S'

$$S \rightarrow S' \in \Gamma(S) \qquad \text{(Equation 2)}$$

to facilitate local search. The neighborhood $\Gamma(S)$ is defined by keeping all but the i:th particle image orientation fixed and varying the orientation of the i:th particle image (one-exchange neighborhood). Stochastic hill climbing scans the search space by sequentially updating a single configuration S. The global score (Equation 4, below) is a weighted average of individual scores λ_i. An individual score λ_i is defined as the correlation between the i:th image and a projection of the reconstruction in orientation $\{\psi_i, \theta_i, \phi_i\}$. The correlation λ_i is calculated in real space, assuming statistically normalized images (zero pixel average and unit variance)

$$\lambda_i = \frac{\int_1^{n_{pix}} dj\left(X_j^{(i)} Y_j^{(k)}\right)}{\sqrt{\int_1^{n_{pix}} dj\left(X_j^{(i)}\right)^2 \int_1^{n_{pix}} dj\left(Y_j^{(k)}\right)^2}}, \qquad \text{(Equation 3)}$$

where $X^{(i)}$ is the i:th particle image, $Y^{(k)}$ is the k:th reference image, and n_{pix} is the number of pixels in the polar image representation.

For clarity, consider only the 3D rotational variables of the alignment problem. Our optimization algorithm maximizes the goal function g

$$g = \int_1^n di \int_1^p dj \int_0^{2\pi} d\theta \lambda_{ij}^{\theta} w_{ij}^{\theta}, \qquad \text{(Equation 4)}$$

where p is the number of projection directions, λ_{ij}^{θ} denotes the cross-correlation coefficient (Equation 3) between the i:th particle image and the j:th reference image for in-plane rotation θ, and w_{ij}^{θ} denotes the strength of association (weight) between particle i and orientation j, θ. We are looking for the n distributions of orientation weights that maximize g so that most weights are zero (sparse distribution). Determination of the number of nonzero weights is based on the selected resolution limit, the molecular radius, and the angular resolution of the discrete 3D orientation search space (as described below).

The reconstruction algorithm is resolution-limited using a low-pass filter with a cosine tapering edge. A subset of high-scoring (feasible) orientations R_i is defined and its size r is equal to the number of orientations per particle image included in the weighting scheme. The size of the feasible subset of orientations r is defined based on the low-pass limit l_{lp} (in angstroms) by calculating the average number of projection directions within d radians from any projection direction of the search space, where

$$d = \arctan\left(\frac{l_{lp}}{r_{mol}}\right) \qquad \text{(Equation 5)}$$

and r_{mol} is the molecular radius (in angstroms). The size of the feasible subset of orientations r decreases with increasing low-pass limit resolution and increasing molecular radius. The process is initialized with a random model—a spherical density with a radius approximately equal to the particle radius. The projection directions of the discrete search space are generated by taking a spiraling path on the unit sphere to generate p orientations (Saff and Kuijlaars, 1997). The reference images are generated by low-pass filtering the current reconstruction, projecting it, and transforming the resulting projection images into polar coordinates (Cong et al., 2003; Penczek et al., 1994; Radermacher, 1994). The i:th particle image is low-pass filtered, shifted over a discrete grid of translations, and transformed into polar coordinates. In-plane transformations of the i:th particle image are defined with respect to reference images by searching the neighborhood $\Gamma(S)$ (projection matching).

As described above, λ_{ij}^{θ} denotes the cross-correlation between the i:th particle image and the j:th reference image for in-plane rotation θ. A subset $Q_i \in \Omega$ of feasible orientations is defined by requiring $\lambda_{ij}^{\theta} \geq \lambda_{i,best}$. Selecting r orientations that satisfy $\lambda_{ij}^{\theta} \geq \lambda_{i,best}$ from $Q_i \in \Omega$ for inclusion in the weighting scheme searches in-plane rotations and projection directions. Feasible orientations are selected hierarchically. First, r projection directions are selected randomly (using the highest scoring in-plane rotation for each projection direction). Next, one in-plane rotation per projection direction is selected randomly from the set of in-plane rotations that improve the correlation. In summary, new feasible orientations are accepted with a probability determined by the uniform distribution over all orientations that improve the correlation. If no orientations improve the correlation, the r highest scoring orientations are included in the weighting scheme. The indeterminism of the correlation function defines a probabilistic rule for moving from the current alignment to the new neighbor. Alignments that are more robust toward uncertainties in the reconstruction are found by modeling this uncertainty in the set of "acceptable" alignments. Probabilistic orientation assignment accounts for the large errors in the goal function landscape introduced by noise. The i:th particle image's feasible subset of orientations is denoted $R_i \subset Q_i$. Our weighting scheme is independent of any image formation model or assumption of noise statistics. An alignment S is now defined in a search space of orientation weights, where every particle image is assigned one weight factor per discrete orientation. The correlation values in the feasible subset are normalized to the interval [0,1] to allow better discrimination between the orientations

$$\eta_{ij}^{\theta} = \frac{exp\{(\lambda_{ij}^{\theta} - \lambda_{min})/\Delta\lambda\} - 1}{exp(1) - 1}, \qquad \text{(Equation 6)}$$

where λ_{min} is the minimum correlation and $\Delta\lambda = \lambda_{max} - \lambda_{min}$ is the difference between the maximum and minimum correlation in the feasible subset. The orientation weight w_{ij}^{θ} for particle i projection direction j and in-plane rotation θ is calculated

$$w_{ij} = \begin{cases} \dfrac{\exp\left(\eta_{ij}^{\theta}\right)}{\int_{R_i} dj \exp\left(\eta_{ij}^{\theta}\right)} & if\ j, \theta \in R_i\ and\ \lambda_{ij}^{\theta} > 0 \\ 0\ else \end{cases}. \qquad \text{(Equation 7)}$$

The variance of the orientation weight distribution depends on the map quality and the SNR of the images. Large reconstruction errors or high noise flattens the orientation weight distribution and spreads a single image over a large region of the search space, accounting for the indeterminism of the cross-correlation function. In contrast, an image with high SNR aligned to an accurate reconstruction centers the orientation weight distribution energy on a few closely lying orientations. All orientations outside the feasible subset are assigned a weight of zero. We call this scheme a sparse orientation weighting approach, because most of the orientations receive a weight of zero. When the orientation weights have been calculated for all particle images, a volume is reconstructed using direct Fourier inversion (Elmlund and Elmlund, 2012). The reconstruction is obtained by gridding every particle image i to the Fourier volume in all orientations of its feasible subset of orientations R_i, while multiplying the particle Fourier transform with its corresponding orientation weight. When all images have been gridded, the Fourier volume is normalized and a new reference volume is generated by reverse Fourier transformation. The probabilistic volume reconstruction algorithm is described in detail in the Supplemental Experimental Procedures. A flowchart for the method is presented in Figure 2.

RESULTS

To demonstrate the robustness of our approach toward initialization, we performed alignment of 5,000 cryo-EM images of the ribosome (Frank, 2009), using an initial 3D model of an unrelated molecule—RNA polymerase II—scaled to the size of a ribosome. The progress of the reconstruction is depicted in Figure 3A (Movies S1, S2, and S3 available online). After a few iterations, the erroneous initial model is transformed into the characteristic ribosome shape. We concluded that the algorithm in this case eliminates bias introduced by an initial 3D model that does not faithfully represent the imaged molecule.

We repeated the ribosome reconstruction process three times, using different random 3D models for initialization (Figure 3B). The final maps agreed to a resolution higher than 13 Å, as measured by the $FSC = 0.143$ criterion (Rosenthal and Henderson, 2003; Table 1).

A set of 5,000 GroEL images was selected randomly from a larger data set of ~50,000 images (Stagg et al., 2008). The alignment was initialized with a random model without assuming any point-group symmetry. The process converged after 25 iterations (Figure 3C). We compared our map with a previously obtained GroEL cryo-EM map (Ludtke et al., 2004) that had been low-pass filtered to the corresponding resolution (Figure 3C, lower panel). The maps agreed to a resolution of 12.2 Å as measured by the $FSC = 0.143$ criterion.

A set of 5,000 beta-galactosidase images was selected randomly from a larger data set of ~40,000 images (courtesy of R. Henderson). The alignment was initialized with a random model without assuming any point-group symmetry. The process converged after 25 iterations. We repeated the alignment, using random noise for initialization. The correlation coefficient between the two unfiltered volumes, generated by initialization from different starting points, was 0.95 (Figure S1). To validate the reconstructions, we docked the available X-ray structure (Protein Data Bank [PDB] code 3VD3) to one of the maps (Figure 3D). The progress of the reconstruction for one of the asymmetric reconstruction runs is depicted in Movie S4.

Finally, we tested the robustness of our algorithm on simulated data and concluded that an accurate initial 3D map can be reconstructed from as few as 1,000 very noisy images (SNR = 0.01; Figure 4).

DISCUSSION

Standard projection image alignment or projection matching is often erroneously classified as an exhaustive search procedure.

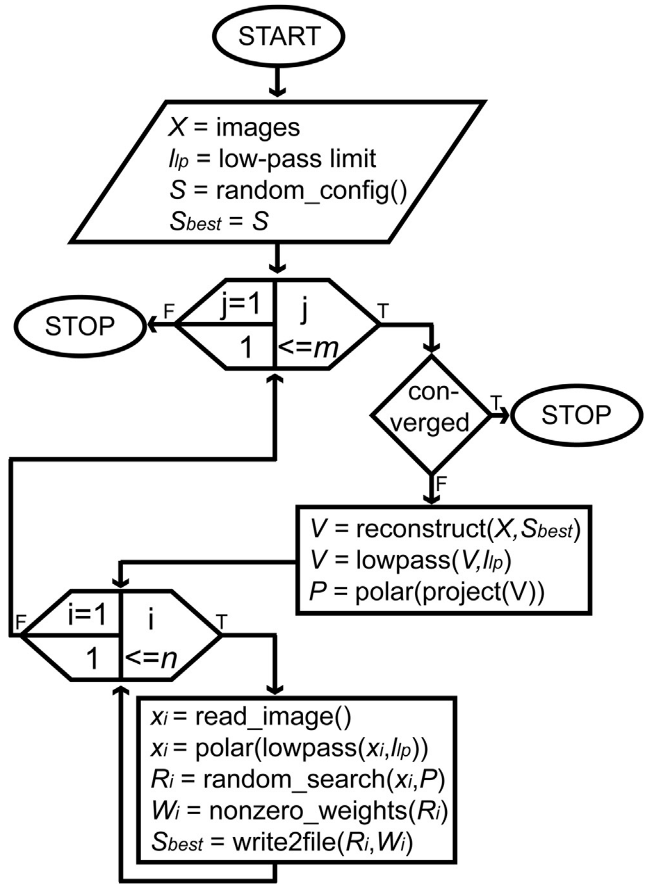

Figure 2. Flowchart for the Method

A low-pass limit l_{lp} and CTF-corrected images X are inputted, where "CTF-corrected images" refers to images corrected using the simple heuristic of binary phase flipping, adopted in numerous image-processing packages (Frank et al., 1996; Ludtke et al., 1999; Tang et al., 2007; van Heel et al., 1996). A configuration S is initialized randomly and S_{best} is set to S. The top loop controls the maximum number of iterations m. The process is repeated until little further change occurs in the most likely orientations. The j:th iteration reconstructs a volume, low-pass filters it, projects it in all projection directions of the neighborhood $\Gamma(S)$, and calculates polar reference images. The second loop is over particle image indices i. The particle image is read, low-pass filtered, and transformed into polar coordinates. The stochastic search first evaluates the correlation $\lambda_{i,best}$ for the current best orientation $\{\psi_{i,best}, \theta_{i,best}, \phi_{i,best}\}$, then searches orientations in random order to define a subset $R_i \subset Q_i$ of feasible orientations that satisfy $\lambda_{ij}^{\theta} \geq \lambda_{i,best}$ and stops when r orientations have been found. All nonzero orientation weights are calculated and the new best configuration S_{best} is written to file.

Scanning through the entire search space would mean evaluating all possible orientations and all possible combinations of orientations. A systematic search of all possible image alignments is computationally intractable even for a few hundred images (time estimates would be given in CPU years).

Standard projection matching is a greedy local search procedure. It uses a one-exchange neighborhood, defined by searching all orientations for *one* particle image. This is done for all particle images, followed by update of the model. The distinction between exploring the neighborhood exhaustively and exploring the entire search space exhaustively is important. Greedy local

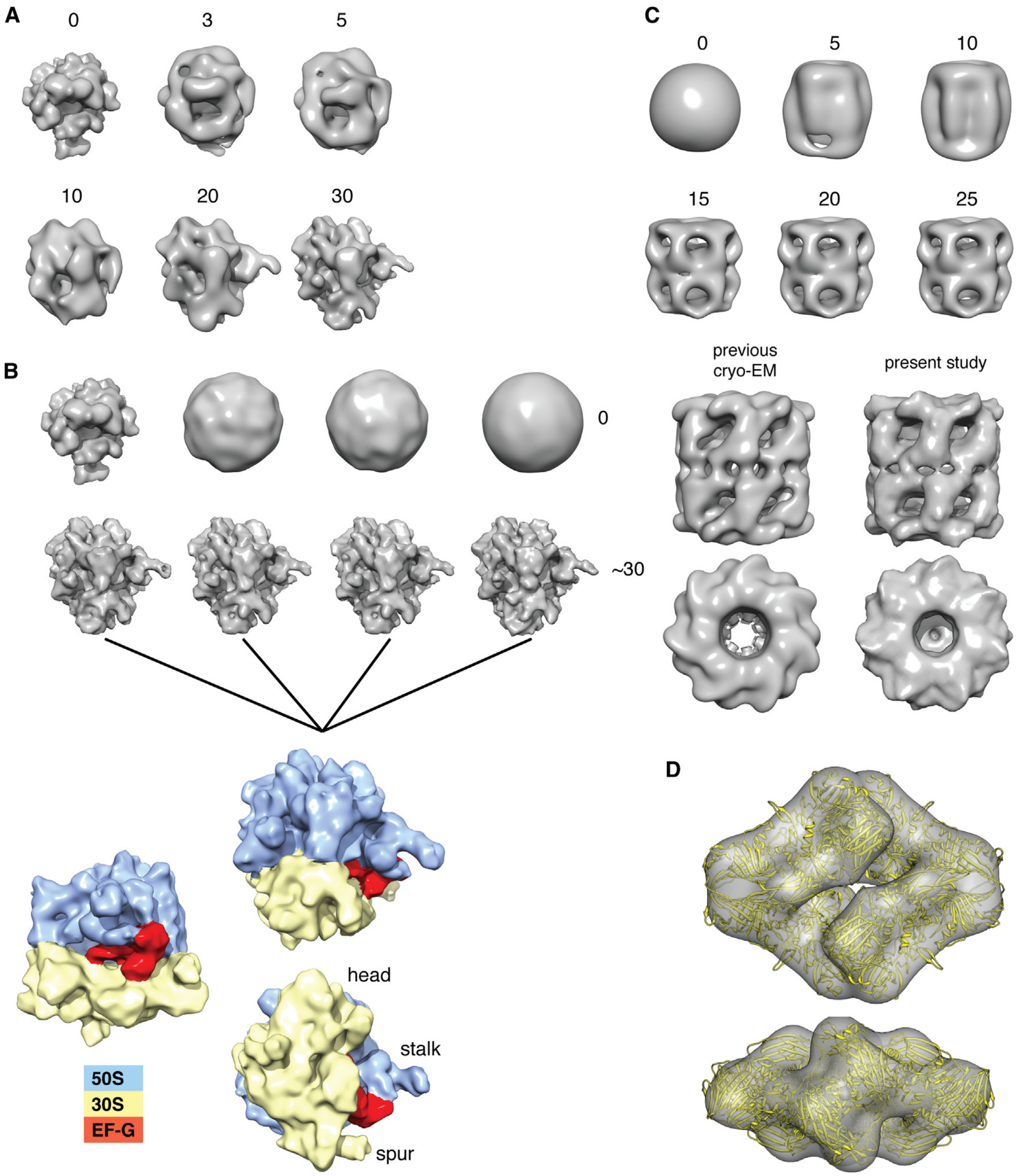

Figure 3. Initial 3D Model Generation from Cryo-EM Images of Molecules with Known Structure

(A) Alignment of ribosome cryo-EM images using an unrelated molecule—RNA polymerase II—as an initial 3D model. The progress of the reconstruction of the asymmetric ribosome structure is shown for one viewing angle (iteration number is indicated). The erroneous initial 3D model (at iteration number 0) was transformed into the characteristic ribosome shape. See also Movies S1, S2, and S3.

(B) One view of our four different ribosome maps (middle panel), generated by initialization from different starting points (top panel), and their average (bottom panel). Quaternary structure regions are indicated in the average map.

(C) Initial 3D model generation from GroEL cryo-EM images. One view of the d7 symmetric structure is shown for different rounds (iteration number is indicated). No a priori assumption was made about the d7 point-group symmetry of the molecule. Bottom panel: two views of our GroEL map (right) compared with a previously obtained cryo-EM map of GroEL (left, EMDB accession code: 1081), low-pass filtered to the corresponding resolution. The maps agreed to a resolution of 12.2 Å as measured by the *FSC* at 0.143.

(D) Two views of one of our beta-galactosidase reconstructions. No a priori assumption was made about the d2 point-group symmetry of the molecule. Docking of the available X-ray structure (PDB code 3VD3) produced an excellent fit.

See also Figure S1 and Movie S4.

Table 1. Resolution at FSC 0.143 between Our Docked Ribosome Reconstructions, Generated by Initialization from Different Starting Points

Resolution at *FSC* = 0.143	1	2	3	4
1	N/A	9.1	8.9	12.8
2	9.1	N/A	9.0	11.7
3	8.9	9.0	N/A	9.7
4	12.8	11.7	9.7	N/A

N/A, not applicable.

search algorithms guarantee convergence to a local optimum. There are no other approximation algorithms that give better convergence guarantees from a mathematical standpoint, although the search behavior of simple hill climbing algorithms (e.g., projection matching) can be improved upon with other methods. A common strategy is to begin with a greedy local search algorithm and introduce modifications to improve its search behavior. Classical examples include greedy randomized adaptive local search (GRASP) (Feo and Resende, 1995), tabu search (Glover, 1990), and simulated annealing (Kirkpatrick et al., 1983). Here, we synergize projection matching with stochastic hill climbing (Russel and Norvig, 2003) and soft optimization to produce an algorithm that has linear time complexity with respect to the number of images and diversified search behavior.

Many design decisions have to be made when developing an optimization algorithm. Why are we using a direct search procedure rather than a gradient-based optimization method, like steepest descent or conjugate gradients? If capable of exploring the search space, gradient-based methods would give orders of magnitude faster convergence. In our hands, gradient-based methods fail to produce acceptable solutions to the problem.

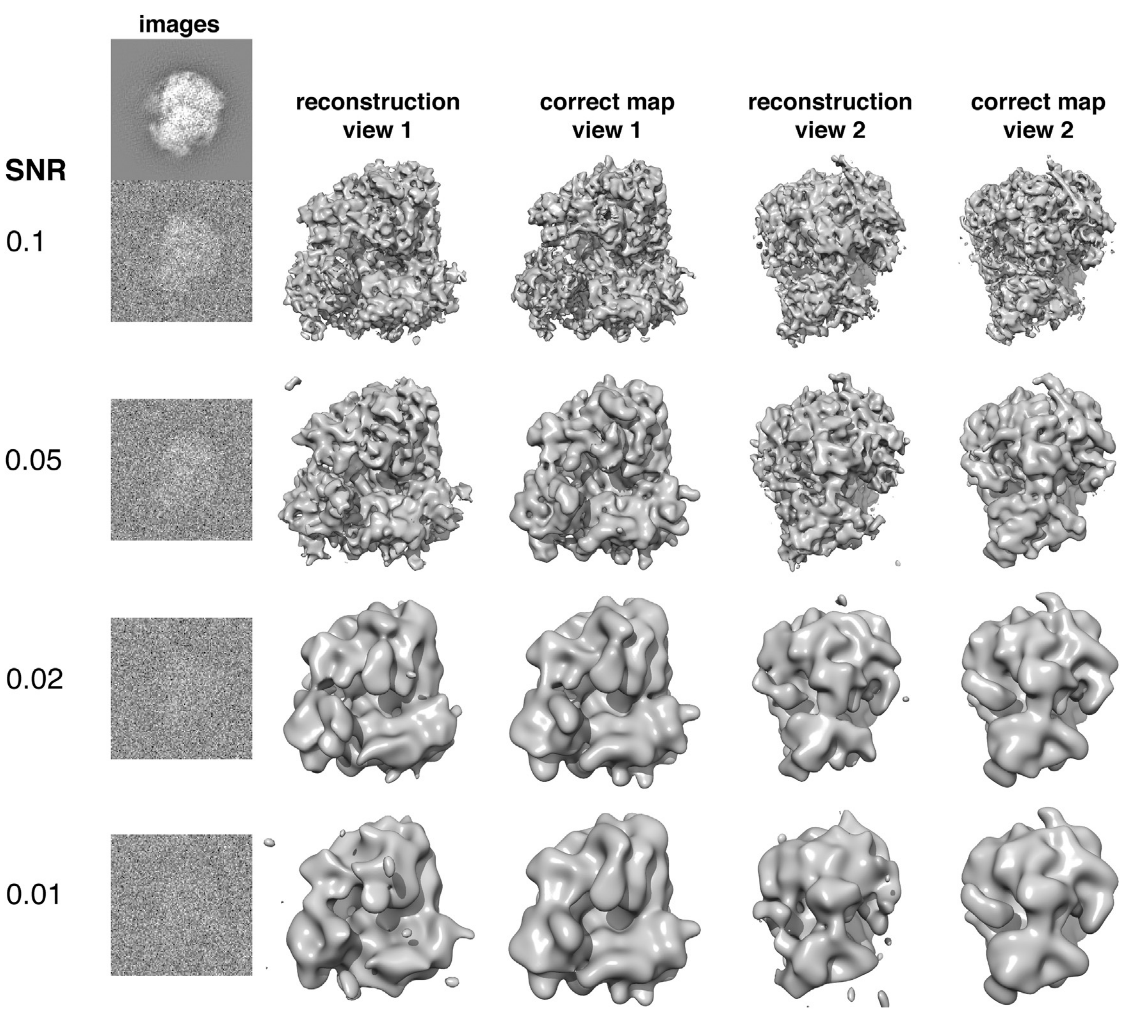

Figure 4. Initial 3D Model Generation from 1,000 Simulated Ribosome Images with Different SNRs
We generated simulated images by adding Gaussian noise to randomly oriented projections of the map (EMDB code 2275). The SNR = 0.1 images were reconstructed to 10 Å, the SNR = 0.05 images to 15 Å, the SNR = 0.02 images to 25 Å, and the SNR = 0.01 images to 30 Å. Two views of our 3D reconstructions are compared with the corresponding views of the correct map, low-pass filtered to the corresponding resolution. The *FSC* between the correct density map and our docked maps showed correlation higher than 0.143 at the applied low-pass limit for all three reconstructions.

That gradient-based optimization methods fail indicates either that the goal function is nondifferentiable or that the goal function landscape is nonconvex. A nonconvex goal function landscape could be explained by two image properties: the many rotational autocorrelation peaks and the low SNR obtained in the low-dose regime. The errors in the goal function landscape increase as the SNRs of the images decrease. Gradient-based optimizers become trapped in poor local optima due to the noise. We therefore use a direct search-based optimization method that does not require gradient information.

Because of the low SNR of cryo-EM images it is risky to greedily attempt to find a single best (deterministic) orientation for each image. "Weighted" orientation assignment—the hallmark of ML refinement (Loh and Elser, 2009; Scheres, 2012a; Scheres et al., 2007a, 2007b; Sigworth, 1998)—refers to the determination of weight distributions, relating known orientations in the search space with the unknown orientation of the particle images. The expectation maximization algorithm, commonly used for ML estimation, is a fixed-point iteration algorithm or hill climber (Dempster et al., 1977). The expectation maximization algorithm guarantees improvement of the likelihood in every iteration (Bishop, 2006), but it is a local search method.

Our algorithm is not an expectation maximization algorithm. We introduce the concept of a feasible subset of orientations and apply stochastic local search for search space exploration. We sacrifice the guarantee of improving the goal function in every iteration in favor of increased search diversification, which forces the procedure to evaluate many more feasible orientations and combinations of orientations without introducing significant computational overheads, as compared to standard projection matching. The expectation maximization algorithm, the standard projection-matching algorithm, and our algorithm are all local search procedures that, from a mathematical standpoint, guarantee convergence to a local optimum. In practice, however, the convergence rate and the quality of the local optimum obtained may vary considerably between different local search procedures.

We approach the cryo-EM image alignment problem by discretizing the orientation space and applying an optimization method that determines a matrix of weights that describes the likelihood that image *i* adopts orientation *j.* The alignment of each image is based on the average of many feasible (or likely) alignments. In physics and probability theory, methods of this kind are referred to as mean field theories or self-consistent field theories. Mean field approaches similar to our method have been used for prediction of protein side chain conformations (Koehl and Delarue, 1994) and processing of coherent X-ray free electron laser diffraction images (Loh and Elser, 2009). Our algorithm operates in real space (the Fourier transform is only used to generate projections of the reconstruction) and the background pixel values are not used to determine any statistics. Masking of images and volumes is therefore part of the procedure, which is critical for noisy images.

In all tested cases, our single-particle reconstruction protocol generates accurate 3D maps directly from the noisy cryo-EM images, without prior 2D image alignment, clustering, averaging, a priori assumptions about the structure or its symmetry, or reseeding.

EXPERIMENTAL PROCEDURES

CTF Correction

We use the simple heuristic of binary phase flipping for CTF correction (Frank et al., 1996; Ludtke et al., 1999; Tang et al., 2007; van Heel et al., 1996). CTF correction by phase flipping only corrects the resolution-dependent CTF phase inversions, disregarding the damping of Fourier amplitudes with increasing resolution. Phase-flipped images are noisier than images corrected with more accurate approaches, such as the reciprocal space adaptive Wiener filter (Penczek, 2010). Here, we are concerned with the problem of reconstructing accurate low-resolution maps ab initio from very noisy images. More sophisticated approaches for dealing with the CTF will be implemented in the future.

Description of Data Sets

The ribosome images were recorded on film, using an FEI TF20 electron microscope with a field-emission gun operating at 200 kV (Frank, 2009). Richard Henderson provided the beta-galactosidase images, recorded with a direct electron detector, using an FEI Polara electron microscope with a field-emission gun operating at 300 kV. The national resource for automated molecular microscopy (NRAMM), Scripps (NIH P41 program RR17573) provided the GroEL images, recorded with an ordinary CCD detector, using an FEI F20 electron microscope with a field-emission gun operating at 200 kV.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Probabilistic Volume Reconstruction, one figure, and four movies and can be found with this article online at http://dx.doi.org/10.1016/j.str.2013.07.002.

ACKNOWLEDGMENTS

H.E. was supported by grants from the Wenner-Gren and Carl Trygger foundations. D.E. was supported by grants from the Swedish Research Council (VR). D.E. and H.E. were supported by Roger Kornberg's NIH grants AI-21144 and GM-36659. Calculations were done at the Biox3 center for distributed computing at Stanford University. Thanks to Pawel Penczek, Duane Loh, Filipe Maia, and Alexis Rohou for idea-shaping discussions and coding assistance.

Received: May 24, 2013
Revised: July 7, 2013
Accepted: July 8, 2013
Published: August 6, 2013

REFERENCES

Adrian, M., Dubochet, J., Lepault, J., and McDowall, A.W. (1984). Cryo-electron microscopy of viruses. Nature *308*, 32–36.

Baker, M.L., Zhang, J., Ludtke, S.J., and Chiu, W. (2010). Cryo-EM of macromolecular assemblies at near-atomic resolution. Nat. Protoc. *5*, 1697–1708.

Bishop, C.M. (2006). Pattern Recognition and Machine Learning (New York: Springer).

Cong, Y., Kovacs, J.A., and Wriggers, W. (2003). 2D fast rotational matching for image processing of biophysical data. J. Struct. Biol. *144*, 51–60.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data Via Em Algorithm. J Roy Stat Soc B Met *39*, 1–38.

Elmlund, D. (2010). Towards unbiased 3d reconstruction in single-particle cryo-electron microscopy. In The Royal Institutue of Technology (Palo Alto: The Royal Institutue of Technology), pp. 1–38.

Elmlund, D., and Elmlund, H. (2012). SIMPLE: Software for ab initio reconstruction of heterogeneous single-particles. J. Struct. Biol. *180*, 420–427.

Elmlund, D., Davis, R., and Elmlund, H. (2010). Ab initio structure determination from electron microscopic images of single molecules coexisting in different functional states. Structure *18*, 777–786.

Feo, T.A., and Resende, M.G.C. (1995). GREEDY RANDOMIZED ADAPTIVE SEARCH PROCEDURES. J. Glob. Optim. *6*, 109–133.

Fischer, N., Konevega, A.L., Wintermeyer, W., Rodnina, M.V., and Stark, H. (2010). Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. Nature *466*, 329–333.

Frank, J. (2006). Three-dimensional Electron Microscopy of Macromolecular Assemblies, *Volume 2* (New York: Oxford University Press, Inc.).

Frank, J. (2009). http://www.ebi.ac.uk/pdbe/emdb/singleParticledir/SPIDER_FRANK_data/.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y.H., 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.

Glover, F. (1990). Artificial-Intelligence, Heuristic Frameworks and Tabu Search. Manage Decis Econ *11*, 365–375.

Harauz, G., and Ottensmeyer, F.P. (1984). Direct 3-Dimensional Reconstruction for Macromolecular Complexes from Electron-Micrographs. Ultramicroscopy *12*, 309–320.

Henderson, R., Chen, S., Chen, J.Z., Grigorieff, N., Passmore, L.A., Ciccarelli, L., Rubinstein, J.L., Crowther, R.A., Stewart, P.L., and Rosenthal, P.B. (2011). Tilt-pair analysis of images from a range of different specimens in single-particle electron cryomicroscopy. J. Mol. Biol. *413*, 1028–1046.

Hohn, M., Tang, G., Goodyear, G., Baldwin, P.R., Huang, Z., Penczek, P.A., Yang, C., Glaeser, R.M., Adams, P.D., and Ludtke, S.J. (2007). SPARX, a new environment for cryo-EM image processing. J. Struct. Biol. *157*, 47–55.

Kirkpatrick, S., Gelatt, C.D., Jr., and Vecchi, M.P. (1983). Optimization by simulated annealing. Science *220*, 671–680.

Knapek, E., and Dubochet, J. (1980). Beam damage to organic material is considerably reduced in cryo-electron microscopy. J. Mol. Biol. *141*, 147–161.

Koehl, P., and Delarue, M. (1994). Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. J. Mol. Biol. *239*, 249–275.

Loh, N.T.D., and Elser, V. (2009). Reconstruction algorithm for single-particle diffraction imaging experiments. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. *80*, 026705.

Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. J. Struct. Biol. *128*, 82–97.

Ludtke, S.J., Chen, D.H., Song, J.L., Chuang, D.T., and Chiu, W. (2004). Seeing GroEL at 6 A resolution by single particle electron cryomicroscopy. Structure *12*, 1129–1136.

Penczek, P.A. (2010). Image restoration in cryo-electron microscopy. Methods Enzymol. *482*, 35–72.

Penczek, P.A., Grassucci, R.A., and Frank, J. (1994). The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. Ultramicroscopy *53*, 251–270.

Radermacher, M. (1994). Three-dimensional reconstruction from random projections: orientational alignment via Radon transforms. Ultramicroscopy *53*, 121–136.

Rosenthal, P.B., and Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. J. Mol. Biol. *333*, 721–745.

Russel, S.J., and Norvig, P. (2003). Artificial Intelligence: A Modern Approach, Second Edition (Upper Saddle River, New Jersey: Prentice Hall).

Saff, E.B., and Kuijlaars, A.B.J. (1997). Distributing many points on a sphere. Mathematical Intelligencer *19*, 5–11.

Sanz-García, E., Stewart, A.B., and Belnap, D.M. (2010). The random-model method enables ab initio 3D reconstruction of asymmetric particles and determination of particle symmetry. J. Struct. Biol. *171*, 216–222.

Scheres, S.H.W. (2012a). A Bayesian view on cryo-EM structure determination. J. Mol. Biol. *415*, 406–418.

Scheres, S.H.W. (2012b). RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. *180*, 519–530.

Scheres, S.H.W., Valle, M., Nuñez, R., Sorzano, C.O.S., Marabini, R., Herman, G.T., and Carazo, J.M. (2005). Maximum-likelihood multi-reference refinement for electron microscopy images. J. Mol. Biol. *348*, 139–149.

Scheres, S.H.W., Gao, H.X., Valle, M., Herman, G.T., Eggermont, P.P.B., Frank, J., and Carazo, J.M. (2007a). Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. Nat. Methods *4*, 27–29.

Scheres, S.H.W., Núñez-Ramírez, R., Gómez-Llorente, Y., San Martín, C., Eggermont, P.P.B., and Carazo, J.M. (2007b). Modeling experimental image formation for likelihood-based classification of electron microscopy data. Structure *15*, 1167–1177.

Scheres, S.H., Núñez-Ramírez, R., Sorzano, C.O., Carazo, J.M., and Marabini, R. (2008). Image processing for electron microscopy single-particle analysis using XMIPP. Nat. Protoc. *3*, 977–990.

Shaikh, T.R., Gao, H., Baxter, W.T., Asturias, F.J., Boisset, N., Leith, A., and Frank, J. (2008). SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. Nat. Protoc. *3*, 1941–1974.

Sigworth, F.J. (1998). A maximum-likelihood approach to single-particle image refinement. J. Struct. Biol. *122*, 328–339.

Stagg, S.M., Lander, G.C., Quispe, J., Voss, N.R., Cheng, A., Bradlow, H., Bradlow, S., Carragher, B., and Potter, C.S. (2008). A test-bed for optimizing high-resolution single particle reconstructions. J. Struct. Biol. *163*, 29–39.

Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J. (2007). EMAN2: an extensible image processing suite for electron microscopy. J. Struct. Biol. *157*, 38–46.

Unwin, P.N.T., and Henderson, R. (1975). Molecular structure determination by electron microscopy of unstained crystalline specimens. J. Mol. Biol. *94*, 425–440.

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.

van Heel, M., Gowen, B., Matadeen, R., Orlova, E.V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R., Schatz, M., and Patwardhan, A. (2000). Single-particle electron cryo-microscopy: towards atomic resolution. Q. Rev. Biophys. *33*, 307–369.

Wiedenheft, B., Lander, G.C., Zhou, K.H., Jore, M.M., Brouns, S.J.J., van der Oost, J., Doudna, J.A., and Nogales, E. (2011). Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature *477*, 486–489.

Yan, X., Dryden, K.A., Tang, J., and Baker, T.S. (2007). Ab initio random model method facilitates 3D reconstruction of icosahedral particles. J. Struct. Biol. *157*, 211–225.

Zhang, W., Kimmel, M., Spahn, C.M.T., and Penczek, P.A. (2008). Heterogeneity of large macromolecular complexes revealed by 3D cryo-EM variance analysis. Structure *16*, 1770–1776.

