Journal of Structural Biology 183 (2013) 377–388

Contents lists available at SciVerse ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Likelihood-based classification of cryo-EM images using FREALIGN

Dmitry Lyumkis [a,1], Axel F. Brilot [b,1], Douglas L. Theobald [b], Nikolaus Grigorieff [b,c,*]

[a] National Resource for Automated Molecular Microscopy, Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA
[b] Department of Biochemistry, Rosenstiel Basic Medical Sciences Research Center, Brandeis University, MS029, 415 South Street, Waltham, MA 02454, USA
[c] Howard Hughes Medical Institute, Brandeis University, MS029, 415 South Street, Waltham, MA 02454, USA

ARTICLE INFO

Article history:
Received 24 April 2013
Received in revised form 3 July 2013
Accepted 9 July 2013
Available online 17 July 2013

Keywords:
Electron microscopy
Maximum likelihood
Classification
Single particle
Protein structure

ABSTRACT

We describe an implementation of maximum likelihood classification for single particle electron cryo-microscopy that is based on the FREALIGN software. Particle alignment parameters are determined by maximizing a joint likelihood that can include hierarchical priors, while classification is performed by expectation maximization of a marginal likelihood. We test the FREALIGN implementation using a simulated dataset containing computer-generated projection images of three different 70S ribosome structures, as well as a publicly available dataset of 70S ribosomes. The results show that the mixed strategy of the new FREALIGN algorithm yields performance on par with other maximum likelihood implementations, while remaining computationally efficient.

1. Introduction

Electron cryo-microscopy (cryo-EM) is a versatile technique to visualize the three-dimensional (3D) structure of macromolecules and their assemblies. Since its inception in the 1970s and 80s (Adrian et al., 1984; Taylor and Glaeser, 1974), methods for sample preparation, microscopy instrumentation and image processing have all been significantly improved. Research in the past decade has focused especially on developing the single-particle technique, which can be applied to non-crystalline, isolated molecules. The single-particle technique has recently been applied to a dataset of 80S ribosomes to obtain a resolution of about 4.5 Å (Bai et al., 2013). The data was recorded with a direct electron detector that registers electrons directly rather than via a scintillator. This new detector technology is also capable of recording movies that capture sample motions occurring during beam exposure, allowing for motion correction (Brilot et al., 2012; Campbell et al., 2012). The work on the 70S ribosome further highlighted the importance of new image processing algorithms, most notably the application of Bayesian statistics and the related maximum likelihood (ML) approach for the estimation of 3D structures from the noisy images obtained using low-dose imaging. These methods are less susceptible to initial model bias and over-refinement, especially at low signal-to-noise ratio (SNR) (Sigworth, 1998; Sigworth et al., 2010).

Originally introduced for the processing of images of viruses and the ribosome (Doerschuk and Johnson, 2000; Provencher and Vogel, 1988; Vogel and Provencher, 1988), ML was later developed as a more general approach for single particle structure estimation (Scheres et al., 2005a,b; Sigworth, 1998). A general difficulty to overcome when implementing ML and Bayesian methods is the representation of the noise in cryo-EM data. Implementation of a Gaussian prior for the noise is straightforward and leads to efficient maximization algorithms. Therefore, current implementations assume that data points (pixels) are independent of each other either in real space or reciprocal space (Scheres, 2010). Assuming independence in real space, one cannot accommodate correlations due to the contrast transfer function (CTF) of the electron microscope. Conversely, assuming independence in reciprocal space prohibits real-space masking of noise surrounding each particle to boost overall SNR. The RELION software (Scheres, 2012b) implements an ML approach with a hierarchical prior to impose smoothness in the estimation of single particle structures from cryo-EM images. It performs most calculations in Fourier space, thus taking CTF effects into account. Using RELION, the authors demonstrate superior results compared to most other software that does not make use of the ML formalism. Specifically, they show improved resolution in the final 3D reconstruction, reduced refinement artifacts such as inflated resolution estimation (overfitting), and better separation of structurally distinct particles in heterogeneous datasets (mixtures).

The FREALIGN software (Grigorieff, 2007) used in the present study is designed to refine single particle reconstructions when an initial reconstruction at lower resolution is already available.

* Corresponding author. Address: Brandeis University, MS029, 415 South Street, Waltham, MA 02454, USA. Fax: +1 (781) 736 2419.
E-mail address: niko@grigorieff.org (N. Grigorieff).
[1] These authors contributed equally.

1047-8477/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jsb.2013.07.005

While FREALIGN is not based on ML or Bayesian statistics, it also implements a series of measures to reduce overfitting and overestimation of resolution (Chen et al., 2009; Grigorieff, 2007; Sindelar and Grigorieff, 2012; Stewart and Grigorieff, 2004), leading to a refinement performance similar to RELION. However, unlike RELION, FREALIGN cannot currently be used for classification of heterogeneous datasets. Here, we describe an extension of FREALIGN that allows classification using an ML approach that is related to that described previously (Scheres, 2012a). The new algorithm works together with the previously implemented FREALIGN refinement scheme and is therefore computationally more efficient than implementations that are entirely based on ML. We apply the new algorithm to a computer-generated dataset of 70S ribosomes, as well as an experimental (Baxter et al., 2009) dataset, both of which contain a heterogeneous mixture of conformationally and compositionally variable single particles.

2. Theory

Our new algorithm is an ML procedure: we seek to find the values of our model parameters that maximize the probability of our data. We will first describe a simple likelihood model for EM data from a single structure, connecting our data model to earlier work, and then expand the model to include data generated from mixtures of multiple different structures.

2.1. A Gaussian statistical data model

Following seminal work by Sigworth (Sigworth, 1998), we assume a multivariate Gaussian model for our EM image data. We consider each image to be a two-dimensional (2D) projection of a randomly translated and rotated 3D structure, to which independent white Gaussian noise has been added to each pixel. This data model can be represented by:

$$\mathbf{X}_i = \mathrm{P}(\phi_i, \mathbf{A}) + \sigma_i G_i \tag{1}$$

where $\mathbf{X}_i$ is the ith image in a dataset of N images, $\phi_i = \{\alpha, \beta, \gamma, x, y\}_i$ is a set of transformation parameters (α, β, γ particle Euler angles and x,y coordinates) for image $\mathbf{X}_i$, P is the projection operator to produce a transformed 2D projection of $\mathbf{A}$ according the Euler angles, coordinates ϕ_i and determined CTF (we assume that image defocus is determined in an independent step using, for example the CTFFIND3 software (Mindell and Grigorieff, 2003) that is not part of this formalism), σ_i is the standard deviation of the noise in the image, and G_i is a "noise image" of independent pixels with values sampled from a standard Gaussian distribution (zero mean and unit variance).

2.2. The likelihood function

Given the data model in (1), the probability of observing a single image $\mathbf{X}_i$ is given by the Gaussian probability density function (PDF):

$$\mathrm{p}(\mathbf{X}_i|\phi_i, \Theta) = \left(\frac{1}{\sqrt{2\pi}\sigma_i}\right)^M \exp\left[-\frac{\|\mathbf{X}_i - \mathrm{P}(\phi_i, \mathbf{A})\|^2}{2\sigma_i^2}\right] \tag{2}$$

where M is the number of pixels in image $\mathbf{X}_i$ (in practice, M includes only pixels within an appropriate mask that is often applied to the particle image to reset image densities to a constant value outside the mask), $\|\mathbf{X}\|^2 = \mathbf{X}^T\mathbf{X}$ denotes the squared Frobenius norm of $\mathbf{X}$ (the squared inner product of $\mathbf{X}$, so that $\|\mathbf{X}_i - \mathrm{P}(\phi_i, \mathbf{A})\|^2$ is the sum of squared differences between each pixel in $\mathbf{X}$ and the transformed $\mathbf{A}$), $\Theta = \{\mathbf{A}, \sigma\}$ is a set of parameters associated with structure $\mathbf{A}$, and $\sigma = \{\sigma_i, \ldots, \sigma_N\}$ is vector of σ_i for all images. Here Θ trivially contains only the true structure $\mathbf{A}$ and σ, but Θ will be expanded later with more complex models. The notation for the conditional probability $\mathrm{p}(x|y)$ is read as "the probability of x, given parameter y", and both x and y may be scalars, vectors, or matrices.

Because our error model assumes statistical independence of pixels and images, the joint PDF for a set of images is simply given by the product of PDFs for each individual image:

$$\mathrm{p}(\mathbf{X}|\phi, \Theta) = \prod_i^N \mathrm{p}(\mathbf{X}_i|\phi_i, \Theta) \tag{3}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^{MN} \prod_i^N \sigma_i^{-M} \exp\left[-\frac{\|\mathbf{X}_i - \mathrm{P}(\phi_i, \mathbf{A})\|^2}{2\sigma_i^2}\right] \tag{4}$$

where now $\{\mathbf{X}, \phi\}$ are each vectors of length N.

When we have fixed, observed data $\mathbf{X}_i$, and the parameters ϕ in Eq. (3) are unknown, the probability $\mathrm{p}(\mathbf{X}|\phi, \Theta)$ is a function only of the parameters, and Eq. (3) is called the *likelihood* of the parameters. The term "likelihood function" is used to emphasize that in the likelihood function the variables are the parameters, in contrast to the PDF in which the data values are the variables. In the method of maximum likelihood, the objective is to choose the parameter values that maximize the likelihood, i.e., we find the parameter values that assign the highest possible probability to our observed data. Maximizing the likelihood over the parameters is equivalent to maximizing the log of the likelihood $\ell(\Theta|\mathbf{X})$, which is usually more convenient to work with. The log-likelihood for Eq. (3) is:

$$\ell(\Theta, \phi|\mathbf{X}) = -\frac{MN}{2}\ln(2\pi) - \frac{1}{2}\sum_i^N \frac{\|\mathbf{X}_i - \mathrm{P}(\phi_i, \mathbf{A})\|^2}{\sigma_i^2} - M\sum_i^N \ln\sigma_i \tag{5}$$

Often the maximum likelihood estimate of a parameter can be found analytically by taking the first derivative of the log-likelihood with respect to the parameter, setting it to zero, and solving for the unknown parameter.

2.3. Cross-correlation maximization and least squares as ML

After some algebraic rearrangement, it can be shown that maximizing the log-likelihood (5) over Θ and ϕ with a constant σ is equivalent to maximizing either.

$$\sum_i^N \sum_j^N [\mathrm{R}(\phi_i, \mathbf{X}_i)^T \mathrm{R}(\phi_j, \mathbf{X}_j)] \tag{6}$$

or

$$\sum_i^N [\mathrm{R}(\phi_i, \mathbf{X}_i)^T \hat{\mathbf{A}}] \tag{7}$$

where

$$\hat{\mathbf{A}} = \sum_i^N \mathrm{R}(\phi_i, \mathbf{X}_i) \tag{8}$$

and $\mathrm{R}(\phi, \mathbf{X})$ is a "back-projection" operator that puts the 2D image back into 3D. These equations can be recognized as restatements of the classical least-squares solution to finding the optimal image registration (Frank et al., 1988), based on maximizing cross-correlations. Thus, when assuming a simple Gaussian model with a common σ for all pixels and images, ML is equivalent to least-squares. If each image is further assumed to have its own respective σ_i, then the sums in Eqs. (6)–(8) are simply weighted by the inverse of the σ_i^2, a procedure equivalent to weighted least-squares.

2.4. Nuisance parameters and hierarchical priors

In most scientific estimation problems, certain parameters are of central interest while other parameters are only used as a means to an end. In statistics these "uninteresting" parameters are called *nuisance parameters*. Sigworth (Sigworth, 1998) treated the EM transformation variables (e.g., the rotations and translations that align images) as nuisance parameters, since they are only used transiently to estimate the parameter of interest, the reference structure. When SNR is low, the estimates of nuisance parameters can be highly uncertain. Since ultimately we do not care about the particular values of nuisance parameters, it would be useful to somehow account for, and perhaps mitigate, the uncertainty in their values.

A general statistical method for dealing with nuisance parameters is to treat them as random variables with their own PDF. For example, Sigworth recognized that the image transformation variables could be considered to be random variables themselves, and he proposed a bivariate Gaussian distribution for the *x*,*y* coordinate (translation) transformation variables:

$$\mathrm{p}(\phi_i|\sigma_x,\sigma_y,\hat{x},\hat{y}) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left[-\frac{\|x_i-\hat{x}\|^2}{2\sigma_x^2}-\frac{\|y_i-\hat{y}\|^2}{2\sigma_y^2}\right] \tag{9}$$

where $\{\hat{x},\hat{y},\sigma_x,\sigma_y\}$ are the means and standard deviations of the *x*,*y* coordinates. Model parameters for the Euler angles can also be introduced if their distribution is non-uniform.

A PDF for parameters is called a *prior*. In this case Eq. (9) is specifically referred to as a *hierarchical prior*, since we now have a statistical model with a hierarchy of distributions — a PDF for the data, given certain parameters, supplemented by a higher level PDF for some of the parameters. The parameters of the hierarchical prior (e.g., σ_x and σ_y in Eq. (9)) may be called hierarchical parameters, to distinguish them from the parameters of the pure likelihood function.

Given a hierarchical prior for the ϕ_i parameters, the likelihoods in Eqs. (2) and (4) can then be augmented to construct an *extended likelihood* function $\mathrm{p}(\mathbf{X}_i,\phi_i|\Theta)$ by multiplying the normal likelihood by the hierarchical prior:

$$\mathrm{p}(\mathbf{X}_i,\phi_i|\Theta) = \mathrm{p}(\mathbf{X}_i|\Theta,\phi_i)\mathrm{p}(\phi_i|\sigma_x,\sigma_y,\hat{x},\hat{y}) \tag{10}$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma_i}\right)^M \exp\left[-\frac{\|\mathbf{X}_i-\mathrm{P}(\phi_i,\mathbf{A})\|^2}{2\sigma_i^2}\right]\mathrm{p}(\phi_i|\sigma_x,\sigma_y,\hat{x},\hat{y}) \tag{11}$$

where the $\Theta = \{\mathbf{A},\sigma,\hat{x},\hat{y},\sigma_x,\sigma_y\}$ is the augmented set of all model parameters associated with reference structure **A**. Note that Eqs. (10) and (11) correspond to 3D versions of Eqs. (11) and (12) of Sigworth, using our notation. The full hierarchical joint likelihood of a set of images is thus:

$$\mathrm{p}(\mathbf{X},\phi|\Theta) = \left(\frac{1}{\sqrt{2\pi}}\right)^{MN}\prod_i^N\left(\sigma_i^{-M}\exp\left[-\frac{\|\mathbf{X}_i-\mathrm{P}(\phi_i,\mathbf{A})\|^2}{2\sigma_i^2}\right]\right)\mathrm{p}(\phi_i|\sigma_x,\sigma_y,\hat{x},\hat{y}) \tag{12}$$

with corresponding log-likelihood:

$$\ln[\mathrm{p}(\mathbf{X},\phi|\Theta)] = -\frac{MN}{2}\ln(2\pi) - \frac{1}{2}\sum_i^N\frac{\|\mathbf{X}_i-\mathrm{P}(\phi_i,\mathbf{A})\|^2}{\sigma_i^2} - M\sum_i^N\ln\sigma_i + \sum_i^N\ln[\mathrm{p}(\phi_i|\sigma_x,\sigma_y,\hat{x},\hat{y})]. \tag{13}$$

Other hierarchical priors can be added in a similar fashion to describe the distribution of other parameters, for example defocus (Chen et al., 2009) and magnification. The form of the additional distributions is often assumed to be Gaussian. Other authors may refer to an extended likelihood as a *regularized likelihood, penalized likelihood,* or a *hierarchical likelihood*. An extended likelihood as in Eq. (11) is also a *joint likelihood*, as it is equivalent to the joint PDF of the data and the parameters given the hyperparameters.

Given a hierarchical statistical model and a corresponding extended likelihood, there are several different ways to proceed with parameter estimation. When the hyperparameters of the hierarchical prior distributions are estimated from the data using variants of ML methodology, such techniques are referred to as extended likelihood or *empirical Bayes*. There are two main ML variants: (a) to maximize the extended likelihood directly, and (b) to maximize the marginal likelihood, in which the nuisance parameters have been integrated out.

2.5. Maximization of the joint extended likelihood

The extended likelihood can be maximized over all unknown parameters simultaneously, including both the parameters and the hyperparameters in the optimization. In practice, this is usually done using an iterative algorithm, in which each parameter is maximized in turn, conditional on the current optimal values of all other parameters. However, any multi-parameter optimization method may be used.

Maximization of the joint extended likelihood aims to find the joint point estimates of the "best" values for all parameters simultaneously. However, this method may not work well when the hierarchical prior is diffuse or multimodal. Direct maximization of the joint likelihood works best when the prior PDF for the hyperparameters is smooth and highly peaked.

2.6. Maximization of the marginal likelihood

Alternatively, when the nuisance parameters are highly uncertain, it may be desirable to completely eliminate them from the analysis, while taking into account the uncertainty in their values. This is accomplished by integrating them out of the extended likelihood, resulting in a marginal likelihood function. For example, we can eliminate ϕ_i from the extended likelihood function in Eqs. (10) and (11) by integrating over its distribution:

$$\mathrm{p}(\mathbf{X}_i|\Theta) = \int_{\phi_i} p(\mathbf{X}_i|\Theta,\phi_i)\mathrm{p}(\phi_i|\sigma_x,\sigma_y,\hat{x},\hat{y})\,d\phi_i \tag{14}$$

which results in a marginal PDF that is independent of ϕ_i. This is the approach taken by Sigworth (Sigworth, 1998), where he integrates out the transformation parameters and maximizes the marginal likelihood function over **A** and σ .

In practice there are several choices for accomplishing the marginalization. In the simplest cases an analytical solution can be obtained. Usually we are not so lucky and must resort to numerical methods such as brute force integration, the Expectation–Maximization algorithm, or some combination of the two (Scheres, 2012a; Sigworth, 1998).

2.7. Expectation–Maximization of the marginal likelihood

The Expectation–Maximization algorithm (normally abbreviated as EM, but we will avoid that here) finds the parameter values that maximize the marginal distribution using a mathematical trick that only requires the (non-integrated) joint likelihood. In its most general form, the algorithm cycles between two steps: (a) the "expectation step", in which one finds the expected logarithm of the joint likelihood function, where the expectation is taken over the nuisance parameters (e.g., ϕ in Eq. (11)), conditional on the current values of the other parameters and the data, and

(b) the "maximization step", in which one maximizes the expected log likelihood function found in (a), as usual, over the other (non-nuisance) parameters of interest (e.g., **A** in Eq. (11)). While it may not be obvious why the Expectation–Maximization algorithm works, it can be shown that the algorithm increases the marginal likelihood at each step, and thus it is guaranteed to find a local maximum of the marginal likelihood.

Often the expectation in the first step can be determined analytically. However, in the present case no analytical solutions exist for the expectations of Eq. (13) needed to maximize the marginal likelihood in (14), and so they must be found by numerical integration, making the procedure computationally expensive (Scheres et al., 2005a,b; Sigworth, 1998). Expectation–Maximization of the marginal likelihood delivers superior performance compared to maximizing the joint likelihood in Eq. (12) when the likelihood function is multimodal or does not display clear peaks due to a low SNR. However, as explained below, in most practical cases we expect the likelihood function to exhibit a single peak close to the correct particle alignment parameters ϕ_i. If so, the model parameters that maximize the joint likelihood function (12) will also approximately maximize the marginalized likelihood (14).

2.8. Extended likelihood, hierarchical priors, and Bayesian methods

The extended likelihood is reminiscent of Bayesian methodology, in which the likelihood is multiplied by a prior for each parameter. However, there is an extremely important methodological and ideological difference. In Bayesian methods, the values of the prior parameters are assumed constants, based on prior knowledge. In contrast, in the ML methods discussed here, the hyperparameters of the prior are unknown parameters that ultimately are estimated from the data via a maximization procedure.

A related Bayesian approach seeks the model parameters with the maximum probability given the data and any prior information or assumptions (known as maximum *a posteriori* probability or MAP). However, like all Bayesian methods, MAP requires explicit Bayesian prior distributions with fixed parameters that are independent of the observed data. Scheres (Scheres, 2012a) has described an extended likelihood procedure which includes a hierarchical zero-mean Gaussian prior on the reciprocal space voxels of each image (which can be considered a "smoothing" technique that regularizes the voxel values). While Scheres describes this method as MAP, the parameters of his hierarchical prior are in fact estimated from the data by maximizing the extended likelihood function, similar to Sigworth's method (Sigworth, 1998) and others, including our ML classification method discussed below. Hence, Scheres' methodology is also a variant of empirical Bayes, which is purely likelihood based.

2.9. Mixtures of particles with distinct structures

We now extend this EM data model to the case where the sample contains a heterogeneous mixture of structures in different, distinct conformations. We assume that there are K classes of structures that may be observed in a dataset of images. We imagine that the molecule under consideration adopts a particular conformation $\mathbf{A}_k$ with probability π_k, and then an image of this structure is generated according to the data model presented in Eq. (1). Each particular class will have its own respective set of parameters, e.g., ϕ_{ik}, and Θ_k (however, note that we presently assume that a given image has the same σ_i regardless of which class it is in). Each image also has an additional integer "indicator" parameter z_i that holds the value of the index of the structural class $\mathbf{A}_k$ that generated the image. For example, if the ith image belongs to class 1, then $z_i = 1$. The likelihood of an individual image, given that it is from class k, is thus:

$$\begin{aligned}\mathrm{p}(\mathbf{X}_i, \phi_{ik}|\Theta_k, z_i = k) \\ = \left(\frac{1}{\sqrt{2\pi}\sigma_i}\right)^M \exp\left[-\frac{\|\mathbf{X}_i - \mathrm{P}(\phi_{ik}, \mathbf{A}_k)\|^2}{2\sigma_i^2}\right] \mathrm{p}(\phi_{ik}|\Theta_k, z_i = k)\end{aligned} \tag{15}$$

where

$$\mathrm{p}(\phi_{ik}|\Theta_k, z_i = k) = \frac{1}{2\pi\sigma_{xk}\sigma_{yk}} \exp[-\frac{\|x_{ik} - \hat{x}_k\|^2}{2\sigma_{xk}^2} - \frac{\|y_{ik} - \hat{y}_k\|^2}{2\sigma_{yk}^2}]. \tag{16}$$

Initially, imagine that we have a set of N data images $\mathbf{X}_i$ from K different classes, and that we know to which class each image belongs (i.e., we know the value of z_i). The joint probability of a single image being from class k and having a specific orientation, translation, and pixel values, is:

$$\mathrm{p}(z_i = k, \mathbf{X}_i, \phi_{ik}|\Theta_k) = \mathrm{p}(\mathbf{X}_i, \phi_{ik}|\Theta_k, z_i = k)\mathrm{p}(z_i = k) \tag{17}$$

where it can be seen that

$$\mathrm{p}(z_i = k) = \pi_k. \tag{18}$$

The total likelihood of this set of images is:

$$\mathrm{p}(z = k, \mathbf{X}, \phi|\Theta) = \prod_i^N \mathrm{p}(z_i = k, \mathbf{X}_i, \phi_{ik}|\Theta_k). \tag{19}$$

Note that we can equivalently write Eq. (19) as

$$\mathrm{p}(z, \mathbf{X}, \phi|\Theta) = \prod_i^N \prod_k^K \mathrm{p}(z_i = k, \mathbf{X}_i, \phi_{ik}|\Theta_k)^{I(z_i=k)} \tag{20}$$

where $I(z_i = k)$ in the exponent is an indicator function whose value is 1 if $z_i = k$ and is 0 otherwise. If the image does not belong to class k, then $I(z_i = k) = 0$ and the total product in Eq. (20) is left unchanged. Careful inspection will verify that Eq. (20) reduces to Eq. (19) when all $I(z_i = k)$ are known integers. However, the central problem of classification is that we do *not* know to which class each image belongs.

2.10. Expectation–Maximization for classifying particle mixtures

For a mixture of different particles, the joint log-likelihood is thus given by the logarithm of Eq. (20):

$$\ell(\Theta, z, \phi|\mathbf{X}) = \sum_{i=1}^N \sum_{k=1}^K I(z_i = k) \ln[\pi_k \mathrm{p}(z_i = k, \mathbf{X}_i, \phi_{ik}|\Theta_k)] \tag{21}$$

where Θ contains all parameters for all classes, including $\mathbf{A}_1 \ldots \mathbf{A}_k$ that represent the underlying structures for each class. The log-likelihood in (21) is also a hierarchical likelihood, as π_k can be considered a hierarchical prior for both the z_i indicator values and the indicator function $I(z_i = k)$. Of course, in practice we do not know the values of z_i and $I(z_i = k)$, and so they are unknown parameters. Directly maximizing (21) is difficult, and the z_i may be considered to be nuisance parameters that are only used as a means to find the k reference structures. Hence, we would prefer to integrate them out and maximize the marginal likelihood $\mathrm{p}(\mathbf{X}, \phi|\Theta)$:

$$\mathrm{p}(\mathbf{X}, \phi|\Theta) = \int_z \mathrm{p}(\mathbf{X}, \phi, z|\Theta)\, dz \tag{22}$$

We cannot accomplish this integration analytically, but it is relatively easy to maximize the marginal likelihood in (22) with the Expectation–Maximization algorithm. We simply find the expectation of the joint log-likelihood (21) with respect to z, and maximize that instead.

A natural choice for the target function to maximize for the particle refinement with respect to model k is the probability distribution $\mathrm{p}(\mathbf{X}_i, \phi_{ik}|\Theta_k)$ (Eq. (2), (Sigworth, 1998)). The target function used in FREALIGN, which we will use in our present study, is a weighted correlation coefficient CC_w (Eq. (17) in Stewart and Grigorieff, 2004) modified with the hierarchical prior $\mathrm{p}(\phi_{ik}|\Theta_k)$ in analogy to Eq. (21) in (Sigworth, 1998). Therefore (Chen et al., 2009),

$$\phi_{ik}^m = \mathrm{argmax}\{ \|\mathbf{X}_i\| \|\mathrm{P}(\phi_{ik},\mathbf{A}_k)\| \, \mathrm{CC}_w(\mathbf{X}_i,\Theta_k,\phi_{ik}) + \sigma_i^2 \ln \mathrm{p}(\phi_{ik}|\Theta_k) \}. \tag{23}$$

2.11. Algorithm

Classification is performed using Expectation–Maximization with fixed alignment parameters ϕ_{ik}^m found in the refinement. The expected probability of a particle i belonging to class k is given by

$$q_{ik} = \mathrm{p}(z_i = k|\Theta,\mathbf{X}) = \frac{\mathrm{p}(\mathbf{X}_i,\phi_{ik}^m|\Theta_k)\pi_k}{\sum_{k=1}^{K}\mathrm{p}(\mathbf{X}_i,\phi_{ik}^m|\Theta_k)\pi_k} \tag{24}$$

which will be set to $1/K$ as the starting estimate. We will refer to these probabilities as particle occupancies. Other parameters are updated in each classification cycle as

$$\pi_k = \frac{1}{N}\sum_{i=1}^{N} q_{ik} \tag{25}$$

$$\mathbf{A}_k = \frac{\sum_i^N \frac{q_{ik}}{\sigma_i^2}\mathrm{R}(\phi_i,\mathbf{X}_i)}{\sum_{i=1}^N \frac{q_{ik}}{\sigma_i^2}} \tag{26}$$

$$\hat{x}_k = \frac{1}{\sum_{i=1}^N q_{ik}}\sum_{i=1}^{N} q_{ik}x_{ik}\,, \quad \hat{y}_k = \frac{1}{\sum_{i=1}^N q_{ik}}\sum_{i=1}^{N} q_{ik}y_{ik} \tag{27}$$

$$\sigma_{xk}^2 = \frac{1}{\sum_{i=1}^N q_{ik}}\sum_{i=1}^{N} q_{ik}(x_{ik}-\hat{x}_k)^2\,,$$
$$\sigma_{yk}^2 = \frac{1}{\sum_{i=1}^N q_{ik}}\sum_{i=1}^{N} q_{ik}(y_{ik}-\hat{y}_k)^2 \tag{28}$$

$$\sigma_i^2 = \frac{1}{M}\sum_{i=1}^{K} q_{ik}\|X_i - \mathrm{P}(\phi_i,\mathbf{A}_k)\|^2 \tag{29}$$

where R is the reconstruction operator used in FREALIGN (a Fourier inversion algorithm, Grigorieff, 2007). Rounds of non-ML parameter refinement to update ϕ_{ik} using Eq. (23) can be run between rounds of ML classification to benefit from the improved class averages.

In the initial stages of the refinement, a single reference can be used that represents the average of the mixture present in the dataset. For multi-reference refinement, the desired number of seeds can be generated either by calculating reconstructions from randomly sampled subsets of the data (Penczek et al., 2006; Spahn and Penczek, 2009), or by supplying reconstructions that represent known aspects of the conformations present in the dataset (supervised classification, van Heel and Stoffler-Meilicke, 1985). The number of required seeds is somewhat arbitrary, which is a recognized problem for K-means classification schemes like the one described here. Different numbers of seeds can be tested to find the largest number that produces classes with new features (Scheres, 2010).

2.12. White noise assumption

In applying our ML method to EM data, we model the real space noise as independent, constant variance white Gaussian noise (Sigworth, 1998). However, it is well known that CTF effects and other factors affecting image amplitudes (envelopes) can potentially introduce strong correlations among neighboring pixels in real space, which could invalidate the white noise approximation. We make three major assumptions about the images that help validate the white noise model: (1) The contrast in an image is unrelated to the actual single particle density (signal) and is due to background generated by the sample support layer (ice and/or carbon), shot noise, and noise introduced by the detector/scanner (Baxter et al., 2009; Zeng et al., 2007). (2) Detector MTF effects are small in the resolution range of interest or can be corrected (Zeng et al., 2007). (3) Of the three sources of noise (i.e., background, shot noise, and detector noise), the shot noise is the most dominant, which should largely eliminate CTF induced correlations. This approximation appears to be valid for typical low-dose images of particles embedded in ice (Baxter et al., 2009), where shot noise is relatively large.

The effect of the white noise assumption can be gauged by inspecting the $M \times M$ Pearson correlation matrix **C** of the images. For white noise, the correlation matrix **C** is equivalent to the identity matrix, with ones on the diagonal and zero cross-correlations elsewhere. In principle, we could apply a more complicated Gaussian statistical model in which we assume an arbitrary correlation structure, where the correlations are estimated from the data

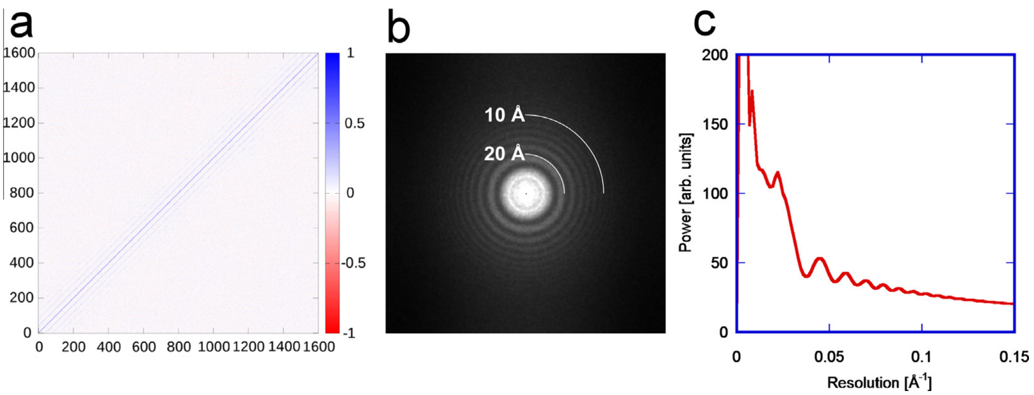


Fig.1. Pixel correlation in a typical cryo-EM image. (a) Correlation matrix calculated using 40 × 40 pixel patches excised from 200 cryo-EM images of 70S ribosomes. The images were recorded at 200 kV with an underfocus of 2.8 μm and digitized with a pixel size of 2.82 Å/pixel. Lines parallel to the diagonal are visible that indicate correlations between pixels that are close to each other in the image. The correlations are due to modulations of part of the noise by the contrast transfer faction (CTF) of the microscope and other effects (envelopes) leading to non-white noise. (b) Thon ring pattern (Thon, 1966) calculated from 857 ribosome images, corresponding to the imaging conditions in (a). (c) Rotational average of the pattern in (b).

simultaneously with the other model parameters. In this augmented statistical treatment, the σ is effectively multiplied by the reduced determinant of the correlation matrix $|\mathbf{C}|^{1/2M}$, which ranges from zero to one (corresponding to completely correlated data and non-correlated data, respectively). By ignoring correlations, then, one underestimates the actual sigma by a factor of $|\mathbf{C}|^{1/2M}$. In the case of the real ribosome data analyzed in this paper (see Results, Fig. 1a), the reduced determinant of the correlation matrix ranges from 0.916 to 0.928, depending on the defocus. These values are close to one, indicating largely uncorrelated data, and only a minor overestimation of the σ, by roughly 7–9%. Taking all these factors into account, our white noise approximation appears appropriate.

2.13. Justification for maximizing the joint likelihood instead of the marginal likelihood

As noted previously, the first exponential in Eq. (11) can usually be approximated by a delta function once the reconstructions $\mathbf{A}_k$ are reasonably close to their corresponding underlying structures (Scheres, 2010). This will be true even if all reconstructions are set to a single reconstruction representing the average of the different underlying structures, provided the differences between the underlying structures are relatively small. The latter condition is fulfilled if the differences will not significantly affect the strong low-resolution signal (typically at 20 Å and lower), which follows the molecular transform of the particle and exceeds the signal at higher resolution usually by several orders of magnitude (Rosenthal and Henderson, 2003). We therefore propose a strategy in which initial refinement of a single reconstruction is performed using our established high-resolution (non-ML) protocols implemented in FREALIGN (Chen et al., 2009; Grigorieff, 2007) to a point where no further improvement is observed.

3. Implementation in FREALIGN

To implement the ML approach described in Section 2, we modified FREALIGN to calculate $\ln[p(\mathbf{X}_i, \phi_{ik}|\Theta_k)]$ for each particle i and each class k. This can be done by running K instances of FREALIGN, each using its own set of alignment parameters ϕ_{ik} and model parameters Θ_k. In a second step, particle occupancies q_{ik} and mixture frequencies π_k are updated using Eqs. (24) and (25). This second step is done using a small additional piece of software called CALC_OCC. This implementation therefore allows straightforward parallelization of the refinement of K classes. The computational load is proportional to the number of particles N and the number of classes K. Furthermore, because maximization of Eq. (23) is done per particle, only one reference reconstruction $\mathbf{A}_k$ and image $\mathbf{X}_i$ have to be kept in memory at a time. One cycle of alignment parameter refinement, classification and reconstruction of the experimental 70S dataset (see below; 10,000 particles with a box size of 130 × 130 pixels and four classes) took about 12 min of CPU time using an Intel Xeon X5690 3.47 GHz CPU. The memory requirements were about 550 MB.

4. Results

4.1. Quantification of pixel correlations in experimental images

As explained in the Theory section, our implementation of the ML classification algorithm is based on the assumption that the noise present in different pixels of an image is uncorrelated. In practice, however, CTF modulations and various slowly varying functions (envelopes) describing the decay of image and noise amplitudes towards high resolution will introduce pixel correlations. To quantify the amount of pixel correlation and its spatial distribution, we calculated correlation matrices **C** for experimental cryo-EM images of 70S *Escherichia coli* ribosome (Baxter et al., 2009) recorded at different defoci. We used sets of $N = 200$ images measuring 40 × 40 pixels that were excised from the ribosome images in areas that excluded signal from the particles. The images were recorded at 200 kV and digitized with a pixel size of 2.82 Å/pixel. The pair-wise pixel correlations were then calculated as

$$c_{mn} = \frac{\sum_i^N (x_{i,m} - \hat{x}_m)(x_{i,n} - \hat{x}_n)}{\sqrt{\sum_i^N (x_{i,m} - \hat{x}_m)^2 \sum_i^N (x_{i,n} - \hat{x}_n)^2}} \tag{30}$$

where c_{mn} are the elements of the correlation matrix **C**, indicating the correlation coefficient between pixel locations m and n, i indicates the image in the set, $x_{i,m}$ are the pixel values and $\hat{x}_m$ are the pixel means calculated for the set. For white noise, the expectation value for c_{mn} is zero for $m \neq n$, leading to a correlation matrix that is similar to the identity matrix (except for random fluctuations affecting the correlation coefficients).

Fig. 1 displays a correlation matrix calculated using images with an underfocus of 2.8 µm. The matrix displays a pattern of lines running parallel to the diagonal and separated by 40 pixels. These lines are due to the conversion of the 2D images into one-column vectors which leads to a 40 pixel periodicity in physical pixel distances. Therefore, these lines describe correlations between nearest neighbors, second nearest neighbors and so on, that arise from the CTF modulations and amplitude envelope. Lines up to about the fourth nearest neighbor are discernible. The average nearest neighbor correlation (average correlations seen in the line nearest the diagonal) observed in Fig. 1a is about 0.28, dropping to 0.12 for the second nearest neighbor. In between the regularly spaced diagonals the correlation is much smaller and fluctuates around zero. Therefore, apart from a few off-diagonal lines, the matrix approximates an identity matrix that would be expected for white noise. The reduced determinant of the correlation matrix in Fig. 1a is 0.924, close to 1, the value for an identity matrix. We also calculated reduced determinants for images with an underfocus of 2.1 µm and 3.5 µm, giving 0.928 and 0.916, respectively. Therefore, the observed correlation matrix suggests that our assumption of white noise is justified. Repeating the correlation analysis with image patches that contained part of ribosomes (signal) reproduced essentially the same results as in Fig. 1a. This is expected as these images were not aligned with each other and, therefore, the signal in different images is not correlated.

To show that the shot noise in the ribosome dataset is essentially white, we also calculated an average power spectrum of 857 ribosome particle images with a defocus of 2.8 µm (Fig. 1b) and plotted the rotational average in Fig. 1c. Except at very low resolution (below 100 Å) where image contrast is affected by inelastic scattering, the noise power at the CTF zeros follows a slowly decaying function, again justifying our assumption of white noise. Background noise added by the embedding ice is affected by the CTF and has been estimated for this data to have about 70% of the variance of the signal produced by the ribosomes. Except at very low resolution, therefore, the CTF-modulated portion of the noise is only a small fraction of the total noise, consistent with the small effect the CTF modulations have on the correlation matrix (Fig. 1a).

The influence of the observed local correlations is further reduced in practical applications of our algorithm because they will most strongly affect the small, high-resolution details in the images. The SNR in a typical cryo-EM image of a single particle is lowest at high resolution, obscuring the fine details of the signal due to the particle. The correlations introduced by CTF and amplitude envelope will therefore affect the likelihood function (12) more or less independently of the signal. The small errors that result from these correlations will change the terms on the right side

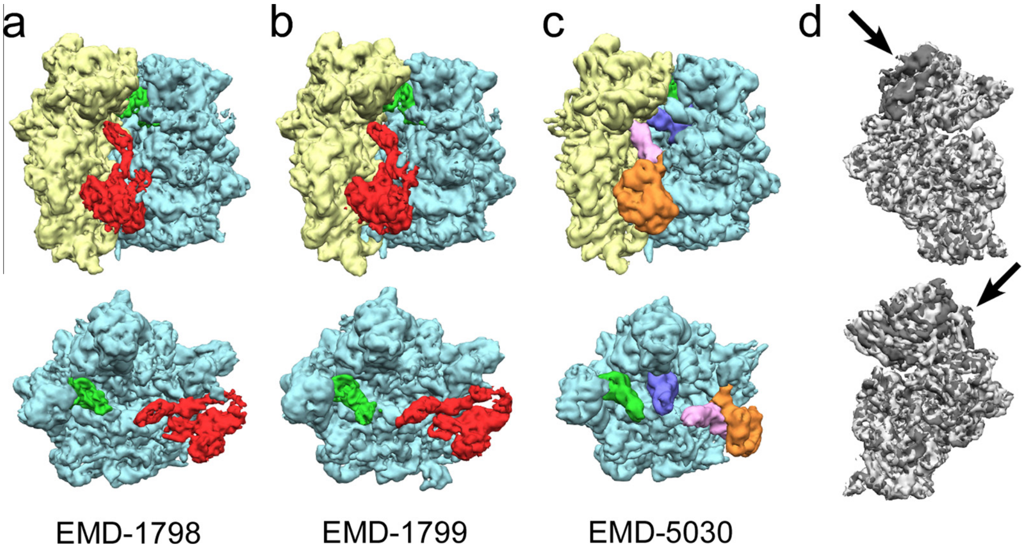


Fig.2. Models used to generate simulated data. Simulated 2D images were generated from three previously determined EM maps that were deposited to the EMDB (Lawson, 2010; Lawson et al., 2011): (a–c) (top) views from the inter-subunit cleft of the complete starting maps and (bottom) from the top of the ribosome with the 30S subunit computationally removed (a) EMD-1798 (50S subunit is blue, 30S subunit yellow, EF-G red, and P/E-site tRNA of the PRE state green), (b) EMD-1799 (50S subunit is blue, 30S subunit yellow, EF-G red, and P/E-site tRNA of the POST state green), and (c) EMD-5030 (50S subunit is blue, 30S subunit yellow, EF-Tu orange, A/T tRNA pink, P-site tRNA blue, and E-site tRNA green). (d) EMD-1798 (light grey) and EMD-1799 (dark grey) contain minor differences, the majority of which can be attributed to slight ratcheting of the 30S subunit with respect to the 50S subunit, and swiveling of the head when aligned to the 30S body. Arrows indicate the differences that are evident in surface rendered maps of the 30S subunit when the two are aligned to each other (top – view from the mRNA entry site; bottom – view from the 50S subunit). Density maps are represented as isosurfaces by UCSF Chimera (Pettersen et al., 2004).

on Eq. (24) in a similar way, cancelling out in the calculation of the occupancies. The other equations (Eqs. (25)–(29)) used to update the model Θ in each iteration are largely unaffected by the correlations, since they do not directly depend on the likelihood function.

4.2. Classification of a simulated 70S ribosome dataset

To quantitatively assess the performance of our new algorithm, we used it to analyze computer-generated image data that mimic experimental data of heterogeneous ribosome populations (Baxter et al., 2009). We selected three previously determined maps of 70S ribosomes – EMD-1798 (Ratje et al., 2010) (Fig. 2a), EMD-1799 (Ratje et al., 2010) (Fig. 2b), and EMD-5030 (Schuette et al., 2009) (Fig. 2c). EMD-1798 and EMD-1799 both contain the elongation factor EF-G, but differ conformationally by a slight ratcheting of the 30S body and swiveling of the head subunit, the extent of which can be appreciated with their overlays in Fig. 2d. EMD-5030 differs from the previous two compositionally by the presence of EF-Tu in place of EF-G, as well as both an A-site and P-site tRNA. Detailed procedures for generating the simulated data are described in Section 6. The general sequence of steps was described previously (Baxter et al., 2009) and is shown in Fig. 3. It is intended to simulate the effects of background noise (an image noise component that is unrelated to the object, yet CTF-dependent), shot noise, and digitization noise.

We analyzed five different datasets containing 10,000 particle images each that differed in their SNR to assess the robustness of the FREALIGN ML approach, including SNR values significantly below those typically observed in an experiment. Previously, an SNR of ~0.05 was estimated for experimentally obtained 70S ribosomes (Baxter et al., 2009). Thus, the final SNRs for the five analyzed datasets were 0.100, 0.050, 0.025, 0.013, and 0.006, respectively (Fig. 3e–i). For each of the five datasets, three different classification schemes were performed: (1) classification only, whereby the Euler angles and shifts were fixed according to their true values (the values used to generate the projections); (2) particle alignment and classification, starting with the true Euler angles and shifts, but allowing changes during refinement; and (3) particle alignment and classification starting with perturbed Euler angles and shifts. The perturbations had a Gaussian distribution with standard deviations of 2.5° and 4 Å for Euler angles and shifts, respectively, and reduced the initial resolution of the starting 3D maps (calculated using the perturbed parameters) to 15–20 Å. This last test was intended to evaluate the ability of the algorithm to simultaneously refine particle alignment and classification (occupancy) parameters. This design enabled us to assess the performance of the algorithm under conditions whereby an increasingly large parameter space must be simultaneously searched during 3D classification.

For each SNR (Fig. 3e–i) and for each of the three classification schemes, 100 refinement iterations of FREALIGN were performed (Fig. 4). To assess the convergence of each run, we monitored the particle composition of the resulting classes after each iteration, as well as the mean occupancy change per particle per class. Plots depicting the variation of total particle compositions of the classes and mean particle occupancy changes with consecutive iterations, as well as values for the final particle composition of each output model with respect to the EMDB map from which they originate, are all displayed in Fig. 4. At an SNR of 0.100, the algorithm could readily recover the true classification parameters and reconstruct the correct starting maps with >99% accuracy, regardless of the classification scheme employed (Fig. 4a–c). At an SNR of 0.050, the algorithm accurately recovered the true classification parameters for each map with >95% certainty when the alignment parameters started with the correct values (Fig. 4d and e). However, starting from perturbed alignment parameters, the algorithm failed to simultaneously refine Euler angles, shifts, and classification parameters for the two smaller classes (Fig. 4f). When the alignment parameters were correct and did not require large adjustments, then even at a lower SNR of 0.025 it was possible to recover true classification parameters with reasonable accuracy (compare Fig. 4g and h with Fig. 4i), although the convergence behavior was somewhat more complicated (e.g., Fig. 4g). At still

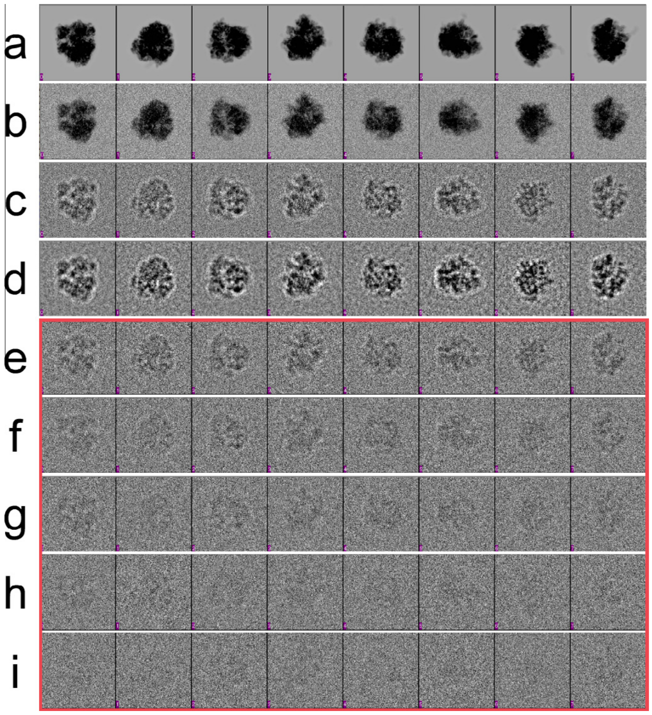


Fig.3. Generation of simulated data. (a) Projections were randomly calculated from each of the three previously determined EM maps, then randomly shifted. (b) Noise was added to the combined dataset such that the SNR was brought down to 1.400, which was followed by (c) application of a CTF in the range of 2 – 4 µm underfocus, and (d) an experimentally determined envelope. Finally, a last layer of noise was added, bringing the final SNR down to (e) 0.100, (f) 0.050, (g) 0.025, (h) 0.013, and (i) 0.006. The datasets used for 3D parameter refinement and classification are outlined in red.

lower SNRs, recovering all three classes became increasingly difficult using our particular dataset and refinement strategy (see Section 6) (Fig. 4j–o). While the SNR values used in these last two datasets were well below experimentally observed values (Baxter et al., 2009), the structure with the biggest difference compared with the other two structures (EMD-5030) could still be recovered (Fig. 4j–m).

We repeated the simulation for all SNRs and refinement strategies with initial particle Euler angles and shifts that were determined *ab initio* using a common lines approach applied to class averages, followed by conventional refinement of the alignment parameters (see Section 6). These simulations represent an experimental situation in which nothing is known about the structure of the particle or their alignments. Results are shown in Supplementary Fig. 1. Both the classification-only and alignment and classification approaches performed as good (SNRs 0.100, 0.013, and 0.006) or better (SNRs 0.050 and 0.025) than in the case where perturbed Euler angles and shifts were supplied (Supplementary Fig. 1, compare columns 1/2 with 3).

4.3. Classification of an experimental 70S ribosome dataset

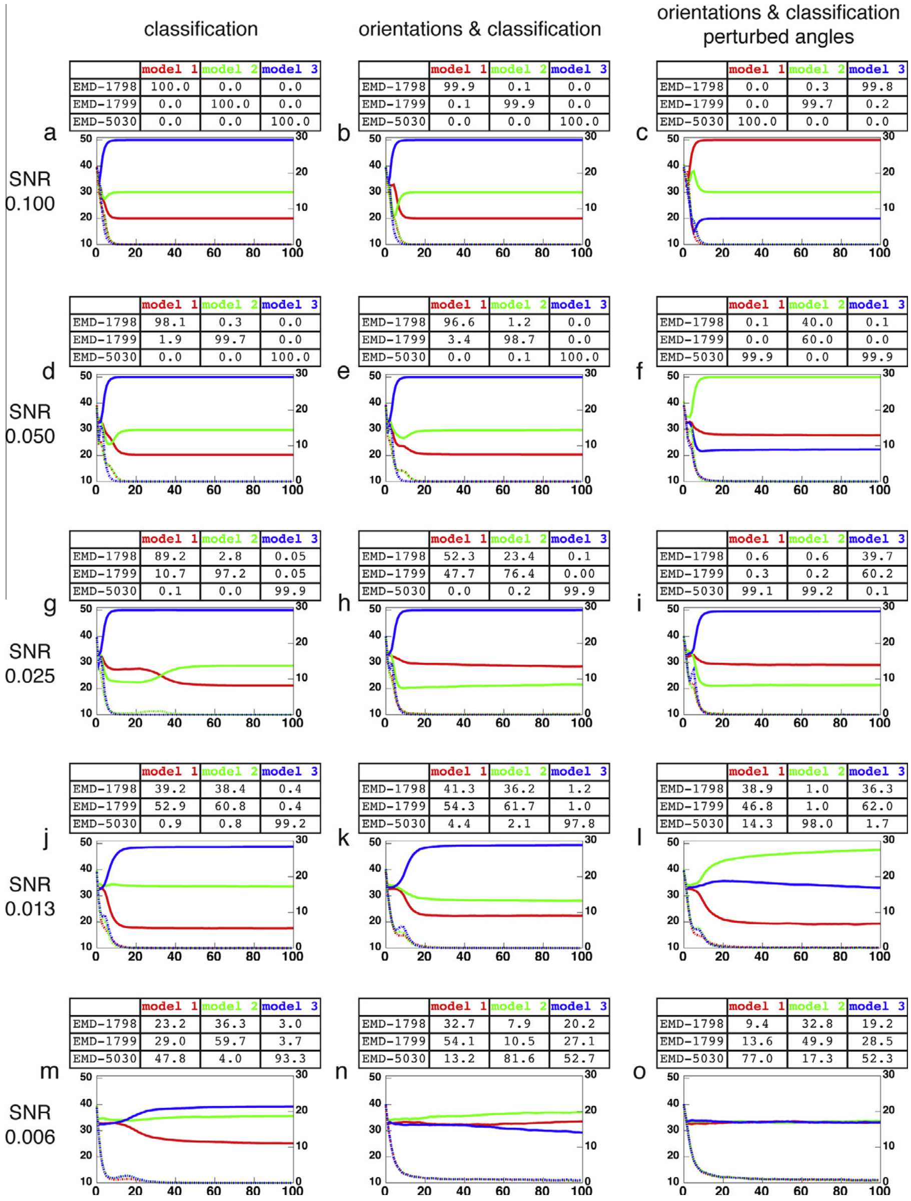
As a second test, we used a publicly available dataset (Baxter et al., 2009) consisting of 10,000 images of 70S *E. coli* ribosomes with bound tRNAs and EF-G co-factor. This dataset was also used to test RELION (Scheres, 2012a,b) which produced three distinct classes - one class containing 70S bound to EF-G and a single tRNA in the E site (about 20 Å resolution), the second class containing 70S bound to three tRNAs and no EF-G (again, about 20 Å resolution), and a small third class (about 7% of the particles) containing the 50S large ribosomal subunit (about 30 Å resolution). The first class (70S with tRNA and EF-G) was duplicated in the analysis which used four classes ($K = 4$, Eq. (15)). Fig. 5 shows the corresponding analysis using the new FREALIGN algorithm, again assuming $K = 4$ classes. For the processing with FREALIGN, particles were masked with a circular mask with a radius of 142 Å. We performed 35 cycles of alignment including data out to 17 Å, followed by 65 cycles of ML classification at 16 Å resolution.

Our analysis detected four distinct classes – three that are equivalent to those detected by RELION and one additional class containing 70S bound to two tRNAs (in the A and P sites) and a weaker density in the E site. The partial presence of E-site density suggests that this additional class is not entirely homogeneous but still contains a mixture of particles. While the 70S-EF-G bound class contains about 50% of the particles in agreement with the previous analyses, the 70S-EF-G devoid class appears to contain 70S complexes with two and three tRNAs at roughly equal proportion. The resolution of all classes except the small 50S class was estimated to be about 15 Å which we confirmed by comparison with an atomic model (Fig. 5). For the 50S class (10% of the particles), we obtained a resolution of only about 40 Å. The resolution of this class is clearly limited by the number of member particles.

5. Discussion

Structural heterogeneity is found in most macromolecular complexes, albeit at varying degrees. While dynamic machines such as the spliceosome exhibit conformational and compositional variability that has prevented reconstruction of most of the splicing intermediates at a resolution higher than about 15 Å (Luhrmann and Stark, 2009), the variability in ribosomes is better understood and can be controlled more easily through careful biochemistry (Frank and Gonzalez, 2010). The highest resolution reconstructions are currently being obtained for icosahedral viruses (Grigorieff and Harrison, 2011), implying low variability compared with many asymmetric particles. The reduced variability is presumably partly the result of the viruses' symmetrical architecture which must restrain the conformational freedom of the subunits within these assemblies. To achieve similarly high resolution with asymmetrical assemblies, computational classification has to be applied to accommodate heterogeneity that cannot be reduced further through biochemical means (Bai et al., 2013). In some cases, such as the ribosome, analysis of the conformational variability may also lead to a deeper mechanistic understanding (Fischer et al., 2010; Mulder et al., 2010).

The presence of particles with multiple conformations or compositions increases the size of the dataset required to resolve each particle class at a given resolution. Therefore, although near-atomic resolution can now be achieved with only a few tens of thousands of particle images (Bai et al., 2013), a mixture of four or five different particle conformations would require four to five times as many images. This number increases still further if some of the classes are much smaller than others, and therefore, more particles need to be imaged to obtain a sufficient number for the smallest class to be reconstructed at the desired resolution. The new algorithm implemented in FREALIGN combines computational speed with the superior convergence behavior of a maximum likelihood approach. Computational efficiency will help make the processing of datasets containing hundreds of thousands of particle images more feasible. FREALIGN is therefore ideally suited for refinement, classification and 3D reconstruction of large heterogeneous datasets at high resolution. Its performance in our test on a small ribosome dataset (10,000 particle images) is on par with that of RELION (Scheres, 2012a,b), as demonstrated by the results shown in Fig. 5.



	model 1	model 2	model 3
EMD-1798	100.0	0.0	0.0
EMD-1799	0.0	100.0	0.0
EMD-5030	0.0	0.0	100.0

	model 1	model 2	model 3
EMD-1798	99.9	0.1	0.0
EMD-1799	0.1	99.9	0.0
EMD-5030	0.0	0.0	100.0

	model 1	model 2	model 3
EMD-1798	0.0	0.3	99.8
EMD-1799	0.0	99.7	0.2
EMD-5030	100.0	0.0	0.0

	model 1	model 2	model 3
EMD-1798	98.1	0.3	0.0
EMD-1799	1.9	99.7	0.0
EMD-5030	0.0	0.0	100.0

	model 1	model 2	model 3
EMD-1798	96.6	1.2	0.0
EMD-1799	3.4	98.7	0.0
EMD-5030	0.0	0.1	100.0

	model 1	model 2	model 3
EMD-1798	0.1	40.0	0.1
EMD-1799	0.0	60.0	0.0
EMD-5030	99.9	0.0	99.9

	model 1	model 2	model 3
EMD-1798	89.2	2.8	0.05
EMD-1799	10.7	97.2	0.05
EMD-5030	0.1	0.0	99.9

	model 1	model 2	model 3
EMD-1798	52.3	23.4	0.1
EMD-1799	47.7	76.4	0.00
EMD-5030	0.0	0.2	99.9

	model 1	model 2	model 3
EMD-1798	0.6	0.6	39.7
EMD-1799	0.3	0.2	60.2
EMD-5030	99.1	99.2	0.1

	model 1	model 2	model 3
EMD-1798	39.2	38.4	0.4
EMD-1799	52.9	60.8	0.4
EMD-5030	0.9	0.8	99.2

	model 1	model 2	model 3
EMD-1798	41.3	36.2	1.2
EMD-1799	54.3	61.7	1.0
EMD-5030	4.4	2.1	97.8

	model 1	model 2	model 3
EMD-1798	38.9	1.0	36.3
EMD-1799	46.8	1.0	62.0
EMD-5030	14.3	98.0	1.7

	model 1	model 2	model 3
EMD-1798	23.2	36.3	3.0
EMD-1799	29.0	59.7	3.7
EMD-5030	47.8	4.0	93.3

	model 1	model 2	model 3
EMD-1798	32.7	7.9	20.2
EMD-1799	54.1	10.5	27.1
EMD-5030	13.2	81.6	52.7

	model 1	model 2	model 3
EMD-1798	9.4	32.8	19.2
EMD-1799	13.6	49.9	28.5
EMD-5030	77.0	17.3	52.3

Fig.4. Classification of the simulated 70S ribosome dataset using $K = 3$ classes. Tests were performed using different levels of noise (arranged vertically) and different classification schemes (arranged horizontally). Each run was initiated with randomized classification parameters, producing three maps at iteration 0 that contained an approximately equal and random particle occupancy distribution, such that any differences among them was essentially random. Five different levels of noise were tested, corresponding to a final SNR of (a–c) 0.100, (d–f) 0.050, (g–i) 0.025, (j–l) 0.013, and (m–o) 0.006. For each SNR level, three independent runs of FREALIGN were performed starting with (a, d, g, j, and m) the correct Euler angles and shifts and disabling their refinement, (b, e, h, k, and n) the correct Euler angles and shifts and enabling parameter refinement, and (c, f, i, l, and o) perturbed Euler angles and shifts, such that the resolution of the starting models at iteration 0 was between 15–20 Å. For each plot, the solid lines (left *y*-axis plotted against the *x*-axis) represent the total particle occupancy for each iteration within output model 1 (red), model 2 (green), and model 3 (blue). Dotted lines (right *y*-axis plotted against the *x*-axis) represent the mean occupancy change per particle per model for each iteration with regard to output model 1 (red), model 2 (green), and model 3 (blue). Both sets of lines are plotted against 100 performed iterations. The tables above each plot describe the particle composition of each output model at iteration 100 (expressed as a percentage) with respect to the original map in the EMDB. Thus, for example, at a SNR of 0.025 and for the perturbed alignment parameters and classification run (panel i), final models 1 and 2 generated by FREALIGN are both primarily composed of particles that originated from EMD-5030 (>99%), whereas model 3 is composed of particles representing the remaining data, and this remainder is divided approximately in accordance with its original 60%/40% distribution within the full 10,000 particle dataset.

In addition to the three classes obtained by RELION, a fourth class was detected that contains no EF-G and only two tRNAs (in the A and P sites). This fourth class is closely related to the class containing three tRNAs as it still shows some weak density in the third tRNA position. While our new algorithm was able to detect this fourth class, it was not capable of completely separating the particles between all the classes. This is presumably due to the relatively small difference between the two- and three-tRNA containing classes. Our new algorithm did not perform as well as RELION on the smallest class present in this dataset – the 50S subunit. The quality of the 50S map at about 40 Å is clearly worse than the 30 Å map obtained using RELION. In terms of its total signal in the dataset, the 50S class is the weakest due to the smallest number of class members and the lower molecular mass of 50S compared with the 70S particles. Therefore, in the limit of very low SNR, maximization of the marginal likelihood as implemented in RELION shows superior convergence. RELION also performs a more comprehensive and computationally expensive orientation search with each iteration

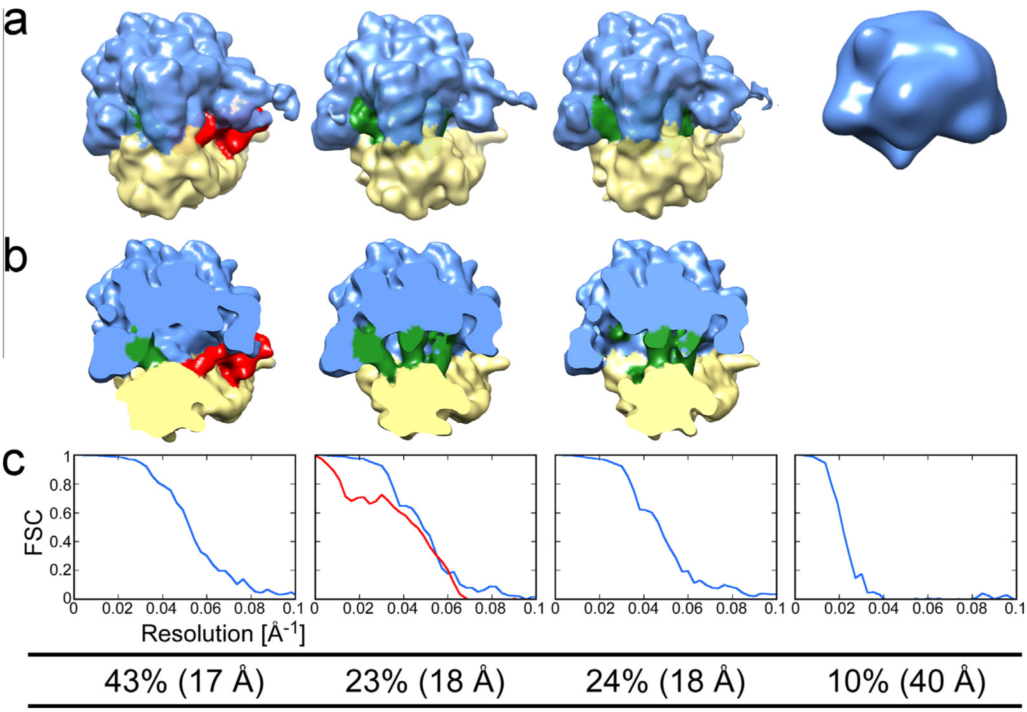


Fig.5. Classification of the experimental 70S ribosome dataset using $K = 4$ classes. Density for EF-G is shown in red, the 50S and 30S subunits are shown in blue and yellow, respectively, and the tRNAs are shown in green. Views with transparent surfaces to show density due to the tRNAs and EF-G are shown in (a), and cut-way views at a higher density threshold to highlight internal density features are shown in (b). Fourier Shell Correlation (FSC) curves for each class are shown in (c) to indicate the resolution of each map. The first two classes represent 70S ribosomes with and without bound EF-G and were also obtained in previous analyses of this test dataset (Elad et al., 2008; Elmlund et al., 2010; Liao and Frank, 2010; Scheres, 2012a; Shatsky et al., 2010). For the second class, a second FSC curve is shown in red, indicating the resolution as estimated using atomic models (PDB codes 2gy9 and 2gya, see Section 6). The last class represents the 50S subunit and was also observed in a classification performed by RELION (Scheres, 2012a,b). The class shown in third position from the left was only observed in some of the previous studies (Elad et al., 2008; Elmlund et al., 2010) and represents a 70S ribosome with tRNAs bound in the A and P sites. There is weaker density for a third tRNA in the E site, indicating that this class still contains some "contaminating" particles from the second class from the left. Density maps are represented as isosurfaces by UCSF Chimera (Pettersen et al., 2004).

that will further increase its success in correctly aligning 50S particles, which cannot be aligned correctly in the initial stages of the refinement when the reference structures represent mostly 70S particles.

6. Methods

6.1. Generation of simulated data

We selected three previously determined cryo-EM maps of 70S ribosomes – EMD-1798 (Ratje et al., 2010) (Fig. 2a), EMD-1799 (Ratje et al., 2010) (Fig. 2b), and EMD-5030 (Schuette et al., 2009) (Fig. 2c). To reduce noise in these maps, binary masks enveloping each volume were generated, then Gaussian low-pass filtered to soften the edges, and multiplied by the corresponding map using the IMAGIC-5 suite (van Heel et al., 1996). Map EMD-1798 was low-pass filtered at 8 Å resolution using a Gaussian filter, and the other two maps were amplitude scaled against the filtered EMD-1798 map using DIFFMAP (http://grigoriefflab.janelia.org). All subsequent procedures were automatically performed using the "create synthetic dataset" functionality implemented within Appion (Lander et al., 2009), the details of which are described below. Projections using random Euler angles and shifts were generated using the project3d and proc2d functions of EMAN (Ludtke et al., 1999). We selected a pixel size of 2.52 Å/pixel, a box size of 160 × 160 pixels and assumed an acceleration voltage of 200 kV. 2000, 3000, and 5000 projections of EMD-1798, EMD-1799, and EMD-5030, respectively were randomly distributed within a 10,000 particle dataset (Fig. 3a). Gaussian distributed noise was added using proc2d (Fig. 3b) to achieve an SNR (variance ratio of signal and noise) of 1.400 (Baxter et al., 2009). The initial addition of noise was then followed by a CTF that was randomly applied to each particle in the range of 2–4 μm underfocus using ace2correct (a variation of ACE1, (Mallick et al., 2005)) (Fig. 3c), and an experimentally obtained (Voss et al., 2010) envelope function (Fig. 3d). Finally the noise was adjusted to five different levels based on the adjusted mean and standard deviation values of the dataset after CTF and envelope application, corresponding to a final SNR of 0.100, 0.050, 0.025, 0.013, and 0.006, respectively (Fig. 3e–i). All particles were normalized to mean and standard deviation values of 0 and 1, respectively, prior to alignment parameter refinement and classification.

6.2. Alignment parameter refinement and 3D classification of simulated data using FREALIGN

For each final level of noise (Fig. 3e–i) and for each of the three classification schemes, 100 iterations of FREALIGN were performed (Fig. 4). Each run was initiated with randomized classification parameters that were obtained with the RSAMPLE program (distributed with FREALIGN). This produced three maps at iteration 0 that contained an approximately equal and random particle occupancy distribution, such that any differences between them were random. The resolution during refinement and classification was always limited to 20 Å. A mask, corresponding to a particle radius of 160 Å was applied and changes in occupancy between iterations were reduced to 50% to improve convergence. Otherwise, the refinement followed standard procedures in FREALIGN (Grigorieff, 1998, 2007).

6.3. Determination of ab initio Euler angle and shifts for 3D classification of the synthetic 70S ribosome dataset

To determine the Euler angles from scratch we utilized the following strategy. First, we used the 0.025 SNR dataset (due to its intermediate level of noise) that contains a mixture of EMD-1798, EMD-1799, and EMD-5030 particles in randomized order to generate reference-free class averages using the ISAC method (Yang et al., 2012). This resulted in 265 ''accounted'' class averages that were further processed using an automated procedure based on the angular reconstitution (common-lines) algorithm implemented in IMAGIC (van Heel et al., 1996). An initial model was obtained that was then used as a reference for projection-matching using Xmipp (Scheres et al., 2008). The Euler angles from the last iteration of projection-matching were converted to Frealign, followed by 5 cycles of alignment parameter refinement. The final angles from this procedure were used as starting points for iterative classification as described above.

6.4. Estimating the resolution of a 3D reconstruction using atomic models

To obtain an unbiased resolution estimate of one of the 3D reconstructions obtained from the experimental 70S dataset (Baxter et al., 2009), we modeled the reconstruction using atomic models for the 30S subunit and three tRNAs (PDB code 2gy9), and 50S subunit (PDB code 2gya). The models were aligned with the 3D reconstruction from the test dataset using UCSF Chimera (Pettersen et al., 2004), converted into a single density map using Bsoft (Heymann and Belnap, 2007). The 3D reconstruction was masked using an envelope derived from a 60-Å low-pass filtered version of the map to remove noise from the solvent surrounding the particle. An FSC curve (FSC_{full}) was then calculated between the masked map and map generated by Bsoft using EMAN's proc3d program (Ludtke et al., 1999). To generate a curve that corresponds to an FSC curve calculated between two half datasets (FSC_{half}), we converted the curve calculated between the model and the map using

$$FSC_{half} = \frac{FSC_{full}^2}{2 - FSC_{full}^2}. \tag{31}$$

Acknowledgments

We are grateful to the Frank group for providing their ribosome test dataset to the cryo-EM community. The work was supported by National Institutes of Health Grants P01 GM62580 (awarded to N.G.), R01 GM094468 (awarded to D.L.T.), P41 GM103310 (supporting D.L.) and an NSERC Fellowship (awarded to A.B.). The synthetic ribosome dataset used in this work can be downloaded at http://nramm.scripps.edu. The new version of Frealign is available for download at http://grigoriefflab.janelia.org.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jsb.2013.07.005.

References

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

Bai, X.C., Fernandez, I.S., McMullan, G., Scheres, S.H., 2013. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. eLife 2, e00461.

Baxter, W.T., Grassucci, R.A., Gao, H., Frank, J., 2009. Determination of signal-to-noise ratios and spectral SNRs in cryo-EM low-dose imaging of molecules. J. Struct. Biol. 166, 126–132.

Brilot, A.F., Chen, J.Z., Cheng, A., Pan, J., Harrison, S.C., Potter, C.S., Carragher, B., Henderson, R., Grigorieff, N., 2012. Beam-induced motion of vitrified specimen on holey carbon film. J. Struct. Biol. 177, 630–637.

Campbell, M.G., Cheng, A., Brilot, A.F., Moeller, A., Lyumkis, D., Veesler, D., Pan, J., Harrison, S.C., Potter, C.S., Carragher, B., Grigorieff, N., 2012. Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. Structure 20, 1823–1828.

Chen, J.Z., Settembre, E.C., Aoki, S.T., Zhang, X., Bellamy, A.R., Dormitzer, P.R., Harrison, S.C., Grigorieff, N., 2009. Molecular interactions in rotavirus assembly and uncoating seen by high-resolution cryo-EM. Proc. Natl. Acad. Sci. USA 106, 10644–10648.

Doerschuk, P.C., Johnson, J.E., 2000. Ab initio reconstruction and experimental design for cryo electron microscopy. IEEE Trans. Inf. Theory 46, 1714–1729.

Elad, N., Clare, D.K., Saibil, H.R., Orlova, E.V., 2008. Detection and separation of heterogeneity in molecular complexes by statistical analysis of their two-dimensional projections. J. Struct. Biol. 162, 108–120.

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

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

Frank, J., Gonzalez Jr., R.L., 2010. Structure and dynamics of a processive Brownian motor: the translating ribosome. Annu. Rev. Biochem. 79, 381–412.

Frank, J., Radermacher, M., Wagenknecht, T., Verschoor, A., 1988. Studying ribosome structure by electron microscopy and computer-image processing. Methods Enzymol. 164, 3–35.

Grigorieff, N., 1998. Three-dimensional structure of bovine NADH:ubiquinone oxidoreductase (complex I) at 22 A in ice. J. Mol. Biol. 277, 1033–1046.

Grigorieff, N., 2007. FREALIGN: high-resolution refinement of single particle structures. J. Struct. Biol. 157, 117–125.

Grigorieff, N., Harrison, S.C., 2011. Near-atomic resolution reconstructions of icosahedral viruses from electron cryo-microscopy. Curr. Opin. Struct. Biol. 21, 265–273.

Heymann, J.B., Belnap, D.M., 2007. Bsoft: image processing and molecular modeling for electron microscopy. J. Struct. Biol. 157, 3–18.

Lander, G.C., Stagg, S.M., Voss, N.R., Cheng, A., Fellmann, D., Pulokas, J., Yoshioka, C., Irving, C., Mulder, A., Lau, P.W., Lyumkis, D., Potter, C.S., Carragher, B., 2009. Appion: an integrated, database-driven pipeline to facilitate EM image processing. J. Struct. Biol. 166, 95–102.

Lawson, C.L., 2010. Unified data resource for cryo-EM. Methods Enzymol. 483, 73–90.

Lawson, C.L., Baker, M.L., Best, C., Bi, C., Dougherty, M., Feng, P., van Ginkel, G., Devkota, B., Lagerstedt, I., Ludtke, S.J., Newman, R.H., Oldfield, T.J., Rees, I., Sahni, G., Sala, R., Velankar, S., Warren, J., Westbrook, J.D., Henrick, K., Kleywegt, G.J., Berman, H.M., Chiu, W., 2011. EMDataBank.org: unified data resource for CryoEM. Nucleic Acids Res. 39, D456–464.

Liao, H.Y., Frank, J., 2010. Classification by Bootstrapping in Single Particle Methods. Proc. IEEE Int. Symp. Biomed. Imaging 2010, 169–172.

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

Luhrmann, R., Stark, H., 2009. Structural mapping of spliceosomes by electron microscopy. Curr. Opin. Struct. Biol. 19, 96–102.

Mallick, S.P., Carragher, B., Potter, C.S., Kriegman, D.J., 2005. ACE: automated CTF estimation. Ultramicroscopy 104, 8–29.

Mindell, J.A., Grigorieff, N., 2003. Accurate determination of local defocus and specimen tilt in electron microscopy. J. Struct. Biol. 142, 334–347.

Mulder, A.M., Yoshioka, C., Beck, A.H., Bunner, A.E., Milligan, R.A., Potter, C.S., Carragher, B., Williamson, J.R., 2010. Visualizing ribosome biogenesis: parallel assembly pathways for the 30S subunit. Science 330, 673–677.

Penczek, P.A., Yang, C., Frank, J., Spahn, C.M., 2006. Estimation of variance in single-particle reconstruction using the bootstrap technique. J. Struct. Biol. 154, 168–183.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612.

Provencher, S.W., Vogel, R.H., 1988. Three-dimensional reconstruction from electron micrographs of disordered specimens. I. Method. Ultramicroscopy 25, 209–221.

Ratje, A.H., Loerke, J., Mikolajka, A., Brunner, M., Hildebrand, P.W., Starosta, A.L., Donhofer, A., Connell, S.R., Fucini, P., Mielke, T., Whitford, P.C., Onuchic, J.N., Yu, Y., Sanbonmatsu, K.Y., Hartmann, R.K., Penczek, P.A., Wilson, D.N., Spahn, C.M., 2010. Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. Nature 468, 713–716.

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

Scheres, S.H., 2010. Classification of structural heterogeneity by maximum-likelihood methods. Methods Enzymol. 482, 295–320.

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

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

Scheres, S.H., Valle, M., Carazo, J.M., 2005a. Fast maximum-likelihood refinement of electron microscopy images. Bioinformatics 21 (Suppl. 2), ii243–ii244.

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

Scheres, S.H., Nunez-Ramirez, R., Sorzano, C.O., Carazo, J.M., Marabini, R., 2008. Image processing for electron microscopy single-particle analysis using XMIPP. Nat. Protoc. 3, 977–990.

Schuette, J.C., Murphy, F.V.t., Kelley, A.C., Weir, J.R., Giesebrecht, J., Connell, S.R., Loerke, J., Mielke, T., Zhang, W., Penczek, P.A., Ramakrishnan, V., Spahn, C.M., 2009. GTPase activation of elongation factor EF-Tu by the ribosome during decoding. EMBO J. 28, 755–765.

Shatsky, M., Hall, R.J., Nogales, E., Malik, J., Brenner, S.E., 2010. Automated multi-model reconstruction from single-particle electron microscopy data. J. Struct. Biol. 170, 98–108.

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

Sigworth, F.J., Doerschuk, P.C., Carazo, J.M., Scheres, S.H., 2010. An introduction to maximum-likelihood methods in cryo-EM. Methods Enzymol. 482, 263–294.

Sindelar, C.V., Grigorieff, N., 2012. Optimal noise reduction in 3D reconstructions of single particles using a volume-normalized filter. J. Struct. Biol. 180, 26–38.

Spahn, C.M., Penczek, P.A., 2009. Exploring conformational modes of macromolecular assemblies by multiparticle cryo-EM. Curr. Opin. Struct. Biol. 19, 623–631.

Stewart, A., Grigorieff, N., 2004. Noise bias in the refinement of structures derived from single particles. Ultramicroscopy 102, 67–84.

Taylor, K.A., Glaeser, R.M., 1974. Electron diffraction of frozen, hydrated protein crystals. Science 186, 1036–1037.

Thon, F., 1966. Zur Defokussierungsabhängigkeit des Phasenkontrastes bei der elektronenmikroskopischen Abbildung. Z. Naturforsch. 21a, 476–478.

van Heel, M., Stoffler-Meilicke, M., 1985. Characteristic views of *E. coli* and *B. stearothermophilus* 30S ribosomal subunits in the electron microscope. EMBO J. 4, 2389–2395.

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

Vogel, R.H., Provencher, S.W., 1988. Three-dimensional reconstruction from electron micrographs of disordered specimens. II. Implementation and results. Ultramicroscopy 25, 223–239.

Voss, N.R., Lyumkis, D., Cheng, A., Lau, P.W., Mulder, A., Lander, G.C., Brignole, E.J., Fellmann, D., Irving, C., Jacovetty, E.L., Leung, A., Pulokas, J., Quispe, J.D., Winkler, H., Yoshioka, C., Carragher, B., Potter, C.S., 2010. A toolbox for ab initio 3-D reconstructions in single-particle electron microscopy. J. Struct. Biol. 169, 389–398.

Yang, Z., Fang, J., Chittuluru, J., Asturias, F.J., Penczek, P.A., 2012. Iterative stable alignment and clustering of 2D transmission electron microscope images. Structure 20, 237–247.

Zeng, X., Stahlberg, H., Grigorieff, N., 2007. A maximum likelihood approach to two-dimensional crystals. J. Struct. Biol. 160, 362–374.

Supplementary Material

for

Likelihood-based classification of cryo-EM images using FREALIGN

Dmitry Lyumkis[1], Axel F. Brilot[2], Douglas L. Theobald[2], Nikolaus Grigorieff[2,3]*

[1] National Resource for Automated Molecular Microscopy, Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

[2] Department of Biochemistry, Rosenstiel Basic Medical Sciences Research Center, Brandeis University, MS029, 415 South Street, Waltham, MA 02454, USA

[3] Howard Hughes Medical Institute, Brandeis University, MS029, 415 South Street, Waltham, MA 02454, USA

ab initio Euler assignment
3D classification

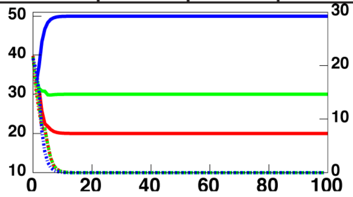
ab initio Euler assignment
3D orientations & classification

perturbed true angles
3D orientations & classification

SNR
0.100

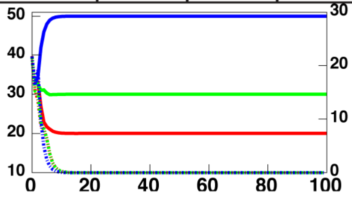
a

	model 1	model 2	model 3
EMD-1798	99.9	0.1	0.0
EMD-1799	0.1	99.9	0.0
EMD-5030	0.0	0.0	100.0



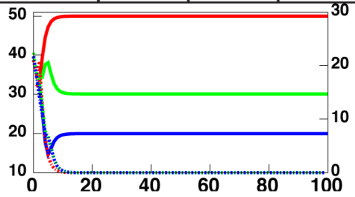
b

	model 1	model 2	model 3
EMD-1798	99.9	0.1	0.0
EMD-1799	0.1	99.9	0.0
EMD-5030	0.0	0.0	100.0



c

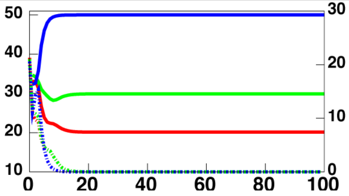
	model 1	model 2	model 3
EMD-1798	0.0	0.3	99.8
EMD-1799	0.0	99.7	0.2
EMD-5030	100.0	0.0	0.0



SNR
0.050

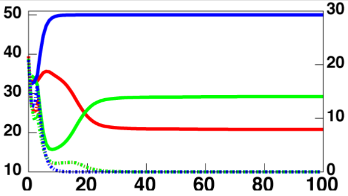
d

	model 1	model 2	model 3
EMD-1798	97.6	1.2	0.0
EMD-1799	2.4	98.8	0.0
EMD-5030	0.0	0.0	100.0



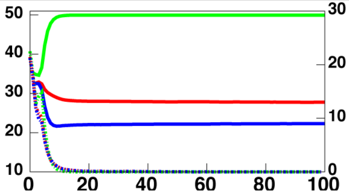
e

	model 1	model 2	model 3
EMD-1798	95.3	0.6	0.0
EMD-1799	4.7	99.4	0.0
EMD-5030	0.0	0.0	100.0



f

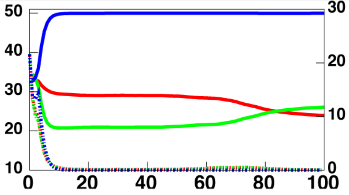
	model 1	model 2	model 3
EMD-1798	0.1	40.0	0.1
EMD-1799	0.0	60.0	0.0
EMD-5030	99.9	0.0	99.9



SNR
0.025

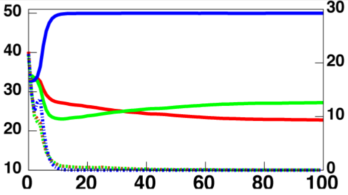
g

	model 1	model 2	model 3
EMD-1798	78.2	5.0	0.05
EMD-1799	21.8	94.9	0.05
EMD-5030	0.0	0.1	99.9



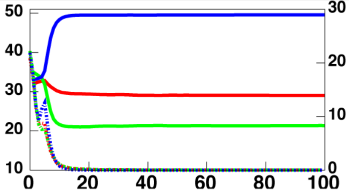
h

	model 1	model 2	model 3
EMD-1798	80.7	5.9	0.1
EMD-1799	19.1	94.0	0.00
EMD-5030	0.2	0.1	99.9



i

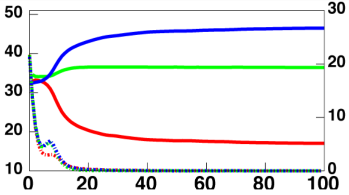
	model 1	model 2	model 3
EMD-1798	0.6	0.6	39.7
EMD-1799	0.3	0.2	60.2
EMD-5030	99.1	99.2	0.1



SNR
0.013

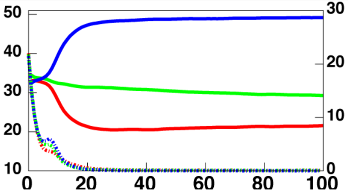
j

	model 1	model 2	model 3
EMD-1798	36.3	36.3	1.3
EMD-1799	40.7	62.3	0.8
EMD-5030	23.0	1.4	97.9



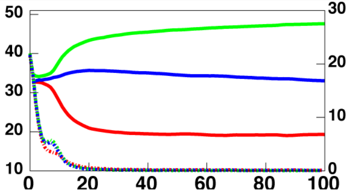
k

	model 1	model 2	model 3
EMD-1798	36.6	39.3	1.2
EMD-1799	56.5	59.2	1.0
EMD-5030	6.9	1.5	97.8



l

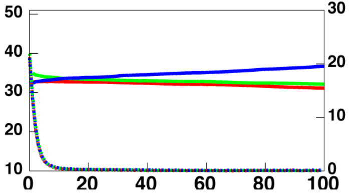
	model 1	model 2	model 3
EMD-1798	38.9	1.0	36.3
EMD-1799	46.8	1.0	62.0
EMD-5030	14.3	98.0	1.7



SNR
0.006

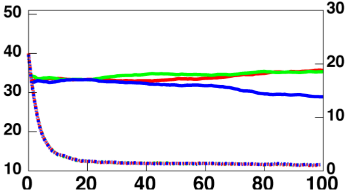
m

	model 1	model 2	model 3
EMD-1798	23.3	24.6	13.2
EMD-1799	31.5	43.6	16.7
EMD-5030	45.2	31.8	70.1



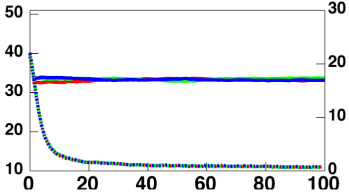
n

	model 1	model 2	model 3
EMD-1798	11.7	21.5	27.2
EMD-1799	14.7	34.3	43.5
EMD-5030	73.6	44.2	29.3



o

	model 1	model 2	model 3
EMD-1798	9.4	32.8	19.2
EMD-1799	13.6	49.9	28.5
EMD-5030	77.0	17.3	52.3



Supplemental Figure 1: Comparison of 3D classification starting with *ab initio* assigned alignment parameters and with perturbed true parameters.

All classification runs were performed as in Fig. 4; tables describe the particle composition of each output model at iteration 100, and graphs describe classification trajectories. (a,d,g,j,m) classification-only refinement with Frealign using *ab initio* assigned Euler angles. (b,e,h,k,n) alignment and classification refinement with Frealign using *ab initio* assigned Euler angles and shifts. (c,f,i,l,o) alignment and classification refinement with Frealign starting from alignment parameters that were slightly perturbed from their true values (these panels are identical to panels (c,f,i,l,o) in Fig. 4 and are shown for the purpose of direct comparison).