BIOINFORMATICS ORIGINAL PAPER Vol. 26 no. 19 2010, pages 2406–2415
doi:10.1093/bioinformatics/btq456

Structural bioinformatics Advance Access publication August 11, 2010

A Bayesian method for 3D macromolecular structure inference using class average images from single particle electron microscopy

Navdeep Jaitly[1,*], Marcus A. Brubaker[1], John L. Rubinstein[2,3] and Ryan H. Lilien[1,4,*]

[1]Department of Computer Science, University of Toronto, [2]Molecular Structure and Function Program, The Hospital for Sick Children Research Institute, [3]Department of Biochemistry and [4]Banting and Best Department of Medical Research, University of Toronto, Toronto, ON, USA

Associate Editor: Burkhard Rost

ABSTRACT

Motivation: Electron cryo-microscopy can be used to infer 3D structures of large macromolecules with high resolution, but the large amounts of data captured necessitate the development of appropriate statistical models to describe the data generation process, and to perform structure inference. We present a new method for performing *ab initio* inference of the 3D structures of macromolecules from single particle electron cryo-microscopy experiments using class average images.

Results: We demonstrate this algorithm on one phantom, one synthetic dataset and three real (experimental) datasets (ATP synthase, V-type ATPase and GroEL). Structures consistent with the known structures were inferred for all datasets.

Availability: The software and source code for this method is available for download from our website: http://compbio.cs.toronto.edu/cryoem/

Contact: ndjaitly@cs.toronto.edu; lilien@cs.toronto.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on March 30, 2010; revised on July 27, 2010; accepted on August 5, 2010

1 INTRODUCTION

Single particle electron cryo-microscopy has achieved remarkable success in determining 3D structures of macromolecules at resolutions close to atomic scales (Cheng and Walz, 2009; Frank, 2006; Zhou, 2008). Some of the protein structures recently determined at high resolutions using this modality include ϵ15 bacteriophage at 4.5 Å, GroEL at 4.2 Å and cytoplasmic polyhedrosis virus and rotavirus at 3.8 Å (Jiang *et al.*, 2008; Ludtke *et al.*, 2008; Yu *et al.*, 2008; Zhang *et al.*, 2008). These experiments used a large amount of data, knowledge of the symmetry of these structures, and an approximate initial model of the structure to achieve this high resolution. However, successful 3D inference of asymmetric structures and symmetric structures with unknown symmetries still remains a challenging task in the absence of an initial model. In this article, we present a general method for creating an initial model (typically referred to as *ab initio* 3D inference) that can be used to create high-resolution structures. Before we describe our method, we first briefly describe the typical data analysis steps in performing 3D inference using single particle images. The reader is referred elsewhere for a detailed review of the various analysis steps in the data-processing pipeline (Fernandez *et al.*, 2006; Thuman-Commike, 2001; van Heel *et al.*, 2000).

Electron micrographs are electron-optic images of the biological sample recorded on a charge-coupled device or on film. These images contain projections of hundreds to thousands of copies of the macromolecule in different and unknown orientations (which may or may not be uniform over the space of rotations). However, the projections are convoluted with the point spread function of the microscope and corrupted by noise. Before a 3D structure can be estimated, these large micrograph images must be preprocessed to select the subsections of the images that correspond to the projections of the individual macromolecules (or particles) under study. Extensive research has been conducted on achieving this goal, called particle picking, using different techniques (Potter *et al.*, 2004). Once these particle images have been extracted, 3D structure determination may be performed in one of the two main ways, depending on the availability of a prior low-resolution model of the macromolecule. With an initial model, the particle images that have been picked are used to generate a higher resolution model through a process called refinement. To do so, the orientations of the noisy particle images with respect to the initial model are first determined, and then these orientations are used to generate a high-resolution model. Several methods have been developed to determine the orientations of particle images with respect to a model. In the projection matching technique (Harauz and Ottensmeyer, 1984a, b; Penczek *et al.*, 1994), projections of the model at different orientations are computed and the particle images are matched to these projections to determine their orientation. Frealign (Grigorieff, 1998, 2007) uses a signal-to-noise ratio weighted correlation coefficient between the Fourier transform of the particle images and the corresponding central slice of the Fourier transform of the 3D model to determine the orientation of particles. Maximum likelihood-based procedures have also been developed for aligning particles to multiple 2D (Scheres *et al.*, 2005a, b) and 3D references (Scheres *et al.*, 2007). These methods have the added advantage that the refinement of the 2D references or 3D models is performed simultaneously with the alignment.

*To whom correspondence should be addressed.

 © The Author 2010. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 31, 2016

In the absence of an initial model, the noisy particle images are first classified into groups, and a representative image, called a class average image, is built for each group from its members. These class average images are assumed to capture the more frequently observed views of the structure. Over 25 years of research has gone into methods for creating high signal-to-noise class average images. These include multivariate statistical analysis (MSA; van Heel and Frank, 1981; van Heel, 1984), the 'alignment through classification' method (Harauz *et al.*, 1988), Bayesian modeling with Gibbs sampling (Samsó *et al.*, 2002), and maximum likelihood methods using single (Sigworth, 1998) and multiple references (Scheres *et al.*, 2005b). The higher signal-to-noise class average images can then be oriented relative to each other by the common lines method (Penczek *et al.*, 1996) or by angular reconstitution (van Heel *et al.*, 1997; van Heel, 1987), and 3D structure determination can be performed using the Fourier Slice Theorem.

The inference of a correct structure from the class averages is complicated by the existence of spurious models, which may explain the observed data. Experimental and analytical methods have been developed to make this inference more robust. In the Random Conical Tilt method (Radermacher *et al.*, 1987), the alignment problem is simplified by acquiring micrographs as tilt pairs. The tilt geometry is used to provide constraints, which make alignments more robust. Tomography (Walz *et al.*, 1997) and the orthogonal tilt-method (Leschziner and Nogales, 2006) similarly improve robustness through the use of multiple images taken at different tilts. In some data analysis methods (Zheng *et al.*, 2002), an assumption of known symmetry is used to prevent incorrect models from being generated by the 3D reconstruction. However, these are, by their nature, restricted to symmetric systems where the symmetry is previously known. Recently, a new method was proposed for *ab initio* 3D reconstruction, which bypasses the use of class averages by using particle images directly (Sanz-García *et al.*, 2010). This method performs 3D reconstruction by iteratively refining an initial model built from particles that were assigned random orientations.

In this article, we present a probabilistic method for performing *ab initio* inference of a 3D model from class average images. Our method defines a prior probability distribution of orientations of particles in the micrographs, and a probability distribution of errors and attempts to fit model parameters that best explain the observed data. In practice, this probabilistic method works analogously to back-projection (Harauz and Ottensmeyer, 1984a) in that it starts with a random model that is iteratively refined. However, it does so under a formal probabilistic model that allows for prior information to be incorporated in a principled manner. In addition, our method attempts to avoid local minima in the iterative procedure. These local minima can cause methods to produce non-optimal solutions. We address this point by using a deterministic annealing (Rose *et al.*, 1990) procedure in the model fitting process. We assume higher variance in the data at the start of the model fitting process and then slowly reduce the assumed variance to move towards a correct, globally optimal model.

This method is similar to that of Scheres *et al.* (2005a), Scheres *et al.* (2007) and Sigworth (1998) in that an observed image is modeled as a multivariate Gaussian with a mean described by the projection of the 3D model. However, these methods were meant for 2D refinement using a reference image (Sigworth, 1998), alignment and 2D averaging with multiple reference images (Scheres *et al.*, 2005b), and refinement using multiple 3D reference models (Scheres *et al.*, 2007), while our method is used for *ab initio* 3D structure inference. As these methods start with approximate solutions (e.g. initial 3D models), which are refined, these methods only need to find local solutions. However, *ab initio* inference needs to find a global solution, for which we use deterministic annealing.

Our use of annealing to find a global minima is not unique in this problem space. Ogura (Ogura and Sato, 2006) employed simulated annealing to perform 3D inference from class averages using a modified cross correlation score. Similarly, Iba *et al.* (2003) used genetic algorithms with a common-lines-based score to find global minima in the space of orientations of class average images. However, our approach is different from these approaches; they use annealing to find a single optimal orientation of images while we probabilistically average over (i.e. marginalize out) the orientation variables in our model.

Our contribution, thus, lies in the application of annealing to find a good minima with a probabilistic model, the use of a Bayesian prior for the distribution of Coulomb density in space, and for incorporation of knowledge of orientations into the model fitting. We show the application of our method to determine 3D structures from class averages of five systems—an asymmetric phantom structure with projections in random orientations, a synthetic dataset with projections of Ribosome in random orientations, two asymmetric, experimentally observed structures (bovine mitochondrial ATP synthase and V-type ATPase) with projections along the equatorial axis and an experimentally observed symmetric structure (GroEL) with projections generally along the top and side views.

2 METHODS

In this section, we describe the model used, the method used to fit the model to the data, and the datasets.

2.1 Bayesian model for electron cryo-microscopy data

Our model has two main components: the parameterization used to represent Coulomb density in space and the probability model to calculate the likelihood of the observed data. These will now be discussed in turn.

2.1.1 Data model We assume that the 3D space over which we are trying to determine the Coulomb density is a cube of side N, centered at the origin, which is split into N^3 voxels by dividing each side (corresponding to x, y and z directions) into N equal length intervals. The Coulomb density at a vertex of arbitrary index, i, where $1 \le i \le N^3$, is the parameter α_i. Thus, the unknown parameters for the model are $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \ldots, \alpha_{N^3})$. The Coulomb density (interchangeably referred to as intensity) at any point (x, y, z) in the cube is obtained by tri-linear interpolation over the vertices of the voxel the point lies in. That is,

$$d(x,y,z) = \sum_{i \in \mathcal{N}(x,y,z)} \alpha_i \left(1 - \frac{(N-1)}{2}|x - p_i^x|\right) \cdot \left(1 - \frac{(N-1)}{2}|y - p_i^y|\right)\left(1 - \frac{(N-1)}{2}|z - p_i^z|\right) \tag{1}$$

where $p_i = (p_i^x, p_i^y, p_i^z)$ corresponds to the coordinates of the i-th vertex, and $\mathcal{N}(x,y,z)$ is the set of vertices of the voxel containing the point (x,y,z). It can be seen that the projection of this model onto a plane through the origin, described by a rotation matrix, $\mathbf{R}$ can be calculated as

$$P(x,y;\mathbf{R},\boldsymbol{\theta}) = \sum_i \alpha_i c_i(x,y;\mathbf{R}) \tag{2}$$

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 31, 2016

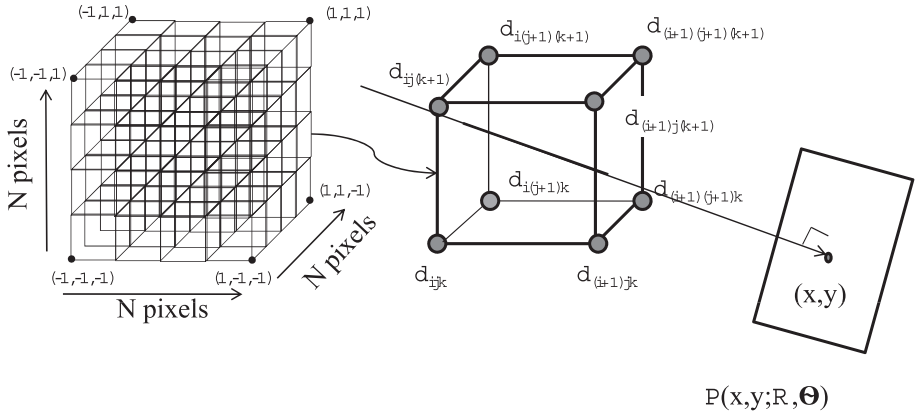

Fig. 1. Model used to represent the Coulomb density and to compute the projection of a macromolecule. The density of a macromolecule is modeled over a 3D grid that is composed of N^3 voxels. The Coulomb density of a point in the interior of a grid is computed as a tri-linear interpolation over the densities at pixels on the vertices of the voxel it belongs to. The intensity at a point in a projection taken at rotation, **R**, is the line integral over the grid along the ray perpendicular to the plane of projection, defined by rotation matrix **R**. As a result, the projection of the model on a point can be computed as a weighted linear sum of contributions of pixels at the borders of voxels. The weights, $c_i(x,y;\theta,\mathbf{R})$, from a pixel, i, in the model, can be calculated by ray-tracing.

where $c_i(x,y;\mathbf{R})$ is the contribution of vertex i to pixel (x,y) in the projection, $\mathbf{P}(\boldsymbol{\theta},\mathbf{R})$ (note that the c_i's do not depend on the parameters $\boldsymbol{\theta}$)[1]. The c_i's can be calculated by taking the line integral along the line parallel to the z-axis (defined by the rotation matrix, **R**) that passes through pixel (x,y) in the image (see Fig. 1).

2.1.2 Probability model The likelihood of observing an image, **I**, from the model, $\boldsymbol{\theta}$, under rotation, **R**, is calculated using the assumption that each pixel of the image is a Gaussian random variable with mean equal to the value of the corresponding pixel in the projection, $\mathbf{P}(\boldsymbol{\theta},\mathbf{R})$, of the model under rotation, i.e.

$$p(\mathbf{I}|\mathbf{R},\boldsymbol{\theta}) = \mathbf{N}\left(\mathbf{I}|\mathbf{P}(\boldsymbol{\theta},\mathbf{R}),\sigma^2\right) \tag{3}$$

where σ is assumed to be the SD of the noise.

Marginalizing over rotation matrices, **R**, with prior probability distribution $p(\mathbf{R})$, we can compute the probability of observing an image, **I**, as

$$p(\mathbf{I}|\boldsymbol{\theta}) = \int p(\mathbf{I}|\mathbf{R},\boldsymbol{\theta})p(\mathbf{R})d\mathbf{R} \tag{4}$$

Note that in our model, we assume that all images are centered correctly and translations do not need to be marginalized out. In our experience, the centering of images performed by software packages used to create class average images alleviates the need to do so. However, this assumption can easily be changed and translation added into the model above. Then, given M images, the probability of observing them is

$$p(\mathbf{I}_1,\mathbf{I}_2,\cdots\mathbf{I}_M|\boldsymbol{\theta}) = \prod_{j=1}^{M} p(\mathbf{I}_j|\boldsymbol{\theta}) \tag{5}$$

since the probability of observing an image is independent of the probability of observing the other images given the model. We can now add priors, $p(\boldsymbol{\theta})$, for our parameters. Given images, $\mathbf{I}_{1..M}$, the posterior probability for the parameters, $\boldsymbol{\theta}$, is

$$p(\boldsymbol{\theta}|\mathbf{I}_{1..M}) \propto p(\boldsymbol{\theta})\prod_{j=1}^{M} p(\mathbf{I}_j|\boldsymbol{\theta}) \tag{6}$$

[1]We follow the convention that matrices and vectors are in bold font $(\text{e.g.}\boldsymbol{\theta},\mathbf{P}(\boldsymbol{\theta},\mathbf{R}))$, while their individual elements $(\text{e.g.}P(x,y;\boldsymbol{\theta},\mathbf{R}))$ are not in bold font.

The prior we use penalizes changes in intensity between neighbouring grid points. More formally, the probability of a set of values $\boldsymbol{\theta} = (\alpha_1,\alpha_2,\ldots,\alpha_{N^3})$ is

$$p(\boldsymbol{\theta}) \propto \exp\left(-\lambda \sum_{(i,j)\in\text{Neighbours}(\boldsymbol{\theta})} (\alpha_i - \alpha_j)^2\right) \tag{7}$$

where λ is a smoothing parameter that controls how strongly changes are penalized. This prior can encourages smooth densities when the observations are inconsistent and effectively interpolates the density when data is missing or ambiguous.

2.2 Model fitting

Starting with an artificially high noise parameter, σ, a constant smoothing parameter λ (described in the previous section), a constant L_1-regularization parameter, β, and a random initial model, the model is fit to the data using a Quasi-Newton optimization (β and details of the optimization method are described in the next section). Subsequently, the noise is reduced deterministically, and the fit model is used as the initial seed in the next iteration of data fitting.

The cycle of deterministic annealing is repeated as long as the model inferred at the end of a cycle is better than the model inferred at the end of the previous cycle. See Figure 2 for an overview of the process.

2.2.1 Quasi-Newton optimization with orthant-wise limited-memory Quasi-Newton The model is fit by a limited memory Quasi-Newton optimization method called orthant-wise limited-memory Quasi-Newton (OWL-QN) method (Andrew and Gao, 2007) using a C implementation available at http://www.chokkan.org/software/liblbfgs/. This method uses the gradient of a function to perform a Quasi-Newton minimization of the function under L_1 regularization. L_1 regularization is a standard penalty function in model fitting that attempts to shift the values of the parameters towards 0, and thus prevents them from taking spurious values in order to explain the data, effectively reducing the degrees of freedom. In this manner, only parameters that are strongly supported by the data are able to achieve large absolute values. The amount of penalty applied is controlled by the regularization parameter β, with higher values of β causing a larger penalty, and hence a larger shift of values towards 0.

In our model, we attempt to minimize the negative log posterior probability function of the parameters $\boldsymbol{\theta}$ with OWL-QN in order to find the model that generates the observed data with the highest probability.

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 31, 2016

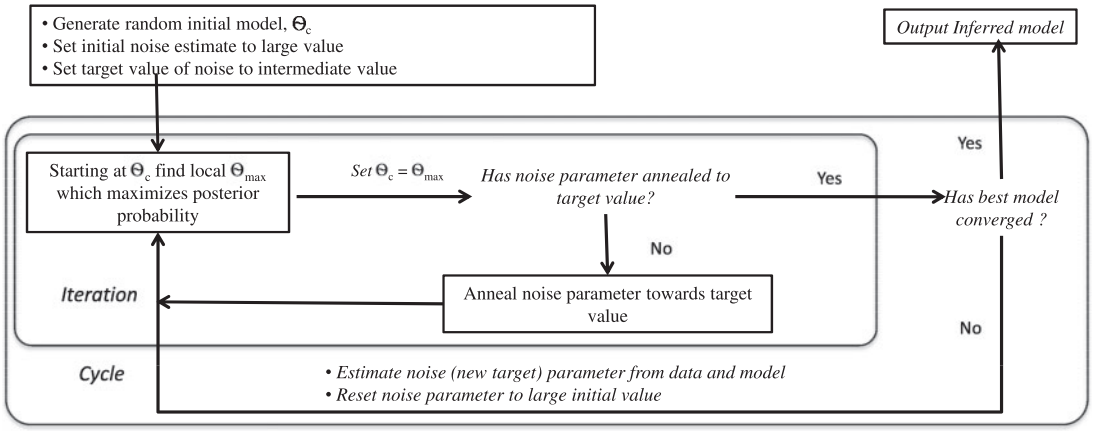

Fig. 2. Overview of our method for structure inference. The final structure is inferred by multiple cycles of deterministic annealing. In each cycle, the noise parameter, σ, is annealed from a high value to a target value of noise over multiple steps. In each of these steps, a MAP estimate of the parameters is found and used as the starting model for the subsequent step. In the first cycle of annealing, the target noise value is an intermediate value; in subsequent cycles, the target noise value is chosen from the difference between the data and the projections from the model inferred in the previous cycle.

The derivative of the log of the posterior probability, $\log p(\boldsymbol{\theta}|\mathbf{I}_{1..M})$, w.r.t. α_i, required by OWL-QN can be calculated as

$$\frac{\partial}{\partial\alpha_i}\log p(\boldsymbol{\theta}|\mathbf{I}_{1..M}) = \frac{1}{p(\boldsymbol{\theta})}\frac{\partial}{\partial\alpha_i}p(\boldsymbol{\theta}) + \sum_{k=1}^{M}\frac{1}{p(\mathbf{I}_k|\boldsymbol{\theta})}\frac{\partial}{\partial\alpha_i}p(\mathbf{I}_k|\boldsymbol{\theta}) \tag{8}$$

The individual terms in the sum in Equation (8) can be calculated using Equations (2) and (4)

$$\frac{\partial}{\partial\alpha_i}p(\mathbf{I}_k|\boldsymbol{\theta}) = \int\left[\frac{\partial}{\partial\alpha_i}p(\mathbf{I}_k|\mathbf{R},\boldsymbol{\theta})\right]p(\mathbf{R})\,\mathrm{d}\mathbf{R} \tag{9}$$

where $\frac{\partial}{\partial\alpha_i}p(\mathbf{I}_k|\mathbf{R},\boldsymbol{\theta})$ is

$$\frac{\partial}{\partial\alpha_i}p(\mathbf{I}_k|\mathbf{R},\boldsymbol{\theta}) = C\exp\left(-\sum_{(x,y)}\frac{(P(x,y;\mathbf{R},\boldsymbol{\theta})-\mathrm{I}_k(x,y))^2}{2\sigma^2}\right) \cdot \sum_{(x,y)}(\mathrm{I}_k(x,y)-P(x,y;\mathbf{R},\boldsymbol{\theta}))\,c_i(x,y;\mathbf{R}) \tag{10}$$

where $C\ (=(2\pi)^{-\frac{N_p}{2}}\sigma^{-N_p-2})$ is the Gaussian normalization constant and N_p the number of pixels in an image. The derivative of the prior $p(\boldsymbol{\theta})$ w.r.t. α_i is elementary and is not shown here.

It can be seen that no obvious closed form solution exists for the likelihood [Equation (5)] and the gradients [Equation (9)] because of the difficulty in performing integration over rotation matrices, $\mathbf{R}$. Hence, these must be calculated numerically by using an approximation to the integral over rotations. To perform this numerical approximation, we sample Q rotations, $\mathbf{R}_{1..Q}$, from the distribution $p(\mathbf{R})$ that represents our prior belief of how rotations are distributed for a dataset, and compute the averages over these rotations, i.e.

$$p(\mathbf{I}_k|\boldsymbol{\theta}) \approx \frac{1}{Q}\sum_{j=1}^{Q}p(\mathbf{I}_k|\mathbf{R}_j,\boldsymbol{\theta}) \tag{11}$$

$$\frac{\partial}{\partial\alpha_i}p(\mathbf{I}_k|\boldsymbol{\theta}) \approx \frac{1}{Q}\sum_{j=1}^{Q}\left[\frac{\partial}{\partial\alpha_i}p(\mathbf{I}_k|\mathbf{R}_j,\boldsymbol{\theta})\right] \tag{12}$$

Note that we choose to use a single set of rotations for all observed images. Without prior information about the orientations of the molecules, this distribution can be uniform. Alternatively, this sampling can also be tailored to reflect the actual distribution of orientations of a molecule, such as rotations only along a side/equatorial view or rotations that are distributed along the top and side/equatorial views. Such a biased sampling of rotations may occur for structures where mass is distributed more along one direction (the dominant axis) than others. We show examples of all these rotation schemes in our datasets below.

In the above approximations, the most computationally demanding step is the calculation of the projection, $\mathbf{P}(\mathbf{R},\boldsymbol{\theta})$, of the model at a given rotation, $\mathbf{R}$. However, because the same set of rotations are shared for each observed image each projection of the model needs to be calculated only once and can be used to calculate the individual terms for all the images, $\mathbf{I}_{1..M}$ in Equations (11) and (12). Thus, for each rotation matrix, the calculation is linear in the number of pixels per image and the number of images.

The rotation matrices used for the approximations can be generated using prior beliefs. For a uniform distribution of rotations, we divide the unit sphere into equal sections using geodesic grids (Sahr *et al.*, 2003), and then for each point on this sphere, we perform rotations around the new z-axis. Note that we did not create rotations by taking periodic values of ψ, θ and ϕ Eulerian angles because this leads to non-uniform sampling, with more rotation matrices being sampled on the poles of a viewing sphere. For the equatorial/side view distribution of rotations, we choose a grid of equally spaced ϕ angles between 0 and 360, with θ angle equal to 90 and ψ angle equal to 90 (the macromolecules were assumed to be oriented correctly in the plane of the image), under the ZYZ Euler angle convention. Similarly, for distributions consisting of top and equatorial/side views, we added rotations with equally spaced ϕ angles between 0 and 360, ψ angles equal to 0, and θ angles equal to 0 and 180 to the set of rotations generated for equatorial/side view distributions. Thus, by computing the derivatives of the approximation we are able to perform OWL-QN optimization of the negative likelihood function to obtain approximate MAP estimates for α's, for a given noise parameter, σ, prior parameter, λ, and L_1 regularization parameter, β. The values of λ and β were set at 10^{-4} and 10^{-4}, respectively, for all the datasets.

2.2.2 Deterministic annealing The noise parameter is annealed exponentially from a high initial value of 20 down to a low final value of 0.3 over 20 steps in the first round of optimization. A value of 20 was assumed to be a reasonable starting point because it was much higher than the maximum intensity in any pixel in any of the images, which were scaled linearly such that the maximum intensity was a value of 1. Similarly, 0.3 was assumed to be a reasonably low noise value to anneal to in the first iteration. In subsequent rounds of optimization, the parameter is re-annealed from 20 down to a final value, which is the SD of noise calculated from the model fitting in the last round of optimization. The cycle of annealing

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 31, 2016

is stopped when the improvement in the log likelihood of the best model of an annealing cycle over the log likelihood of the best model of the previous cycle is less than a user specified value, ϵ (typically chosen to be 1 for our test cases). See Figure 2 for a flowchart representation of the process.

3 DATASETS AND RESULTS

3.1 Datasets

Figure 3 shows views of the five test structures we attempted to infer from class average images and the class average images used for the inference. The five datasets were a phantom structure, bovine mitochondrial ATP synthase, GroEL, V-type ATPase and Ribosome.

3.1.1 Phantom structure The phantom structure used in this article was first described in Baker and Rubinstein (2008). It consists of a long cylindrical core with a peripheral helical coil that runs along the side of the cylindrical object, and a handle-like structure on one side. Fifty noiseless projections of size 32×32 were created along random orientations, which were then used for inference. For this article, to reflect the uniform distribution of projections in space, the approximations to Equations (6) and (8) were calculated using uniformly distributed rotations as described earlier.

3.1.2 ATP synthase The ATP synthase dataset used here was described in Rubinstein *et al.* (2003). In that study, 5984 particles extracted from the experimental data were used to create 20 class average images with MSA (van Heel and Frank, 1981; van Heel, 1984). Of these, nine reasonable class average images were selected using the presence of mirror-image pairs as an indicator of the validity of a class average. We down-sampled these images from 128×128 to 32×32 for computational speed. The class average images all shared only a single common-line because the dominant orientations of ATP synthase represent views along the side (equatorial axis). As a result, the typical approach of angular reconstitution based on common-lines was not applicable, and an alternative approach based on registration points was used for structure determination. Thus, this is an interesting structure for any general *ab initio* inference method. For this dataset, to reflect the fact that the class average images were known to have been side/equatorial views, we computed the approximations to Equations (6) and (8) along rotations generated from an equatorial distribution as described in Section 2.2.1.

3.1.3 GroEL The GroEL dataset described in Stagg *et al.* (2008) was used as a third test system. The particle picked dataset was obtained from the Automated Molecular Group at the Scripps Research Institute at http://ami.scripps.edu/experiment/index.php. The particles were then split automatically into 20 class average images using the maximum likelihood method described in Scheres *et al.* (2005a) and implemented in the Xmipp system (Sorzano *et al.*, 2004). Of these, seven class average images that appeared qualitatively cleaner than the rest (on the basis of noise) were chosen to perform the *ab initio* inference. As with ATP synthase, we down-sampled these images to 32×32 to reduce the computational time required for inference of the 3D structure. The GroEL dataset is of interest to us because it represents a symmetric structure, but we do not explicitly use symmetry in the inference. To reflect our belief in the orthogonality of the two sets of views, we computed the approximations to Equations (6) and (8) along orientations that represented a mix of top and side views (without assigning any of the images to any particular view). The rotation matrices for these views were generated as described in Section 2.2.1.

3.1.4 Ribosome Synthetic data for Ribosome was downloaded from the Xmipp website (Sorzano *et al.*, 2004). The dataset contained 20 000 images of size 130×130. We used these images to create 50 class average images using Xmipp, as we had done for the GroEL particles. These were down-sampled to 32×32. For the Ribosome dataset, no assumption was made of the orientations, similar to the phantom dataset above.

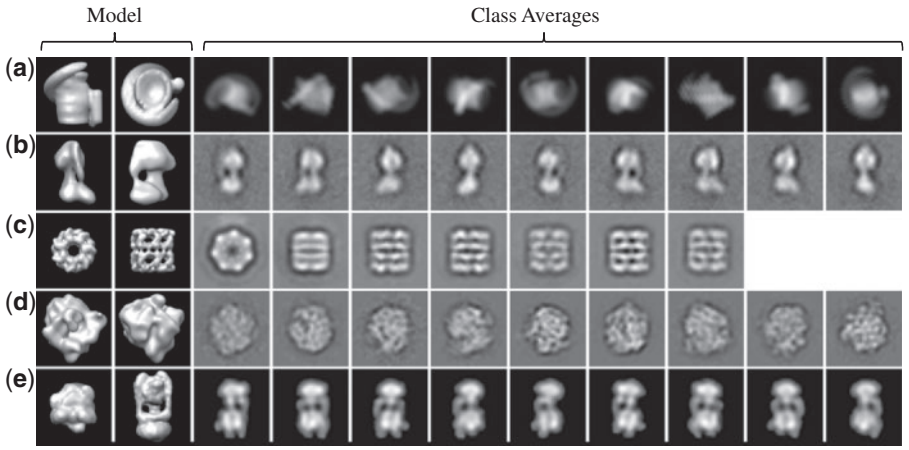

Fig. 3. Datasets used in our study. (**a**) Row 1: two orthogonal views of a phantom structure which was first demonstrated in Baker and Rubinstein (2008) and 9 of the 50 projections that were used in our *ab initio* inference. (**b**) Row 2: two side views of bovine mitochondrial ATP synthase inferred in Rubinstein *et al.* (2003) and the nine class average images (calculated from experimental data) that were used in our *ab initio* inference. (**c**) Row 3: top view and side view of GroEL inferred in Stagg *et al.* (2008) and the seven class average images (calculated from experimental data) that were used in our *ab initio* inference. (**d**) Two opposite views (front and back) of a low-resolution Ribosome structure and 9 of the 50 projections that were used in our *ab initio* inference. (**e**) Top and side view of V-type ATPase and 9 of the 10 projections used in our *ab initio* inference.

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 31, 2016

3.1.5 V-type ATPase The V-type ATPase dataset used here was described in Lau and Rubinstein (2010). In that study 19 825 particles extracted from experimental data were used to create class averages with MSA (van Heel and Frank, 1981; van Heel, 1984). Of these, 10 class average images were selected using the presence of mirror-image pairs. Similar to the above datasets these class averages were down-sampled to 32×32. We used the same rotation prior (equitorial views) that was used for ATP synthase.

3.2 Results

Figure 4 shows the inferred structures for the five test datasets at different points of the annealing cycles. Animations of the smoothed inferred structures and corresponding views of original structures can be seen in the .avi files in the Supplementary Material.

3.2.1 Phantom dataset: a structure that was qualitatively comparable to the original phantom structure (on the basis of the presence of cylindrical stem, peripheral helix and handle) was obtained towards the last two iterations of the first cycle, see Figure 4a. The earlier iterations presented a spherical object with intensity decaying away from the center of the model. This corresponds to a model built with images assigned almost equal probability in all orientations. It is possible that the annealing could have been performed in fewer steps as most of the intermediate steps changed the structure qualitatively little.

3.2.2 ATP synthase A structure that had the main qualitative features of a central and a peripheral stalk of ATP synthase was inferred by the 20th iteration of the first annealing cycle, as shown in Figure 4b. Earlier iterations showed a model that appeared rotationally symmetric around the main axis as a result of the assignment of equal probabilities to all equatorial views. Further refinement of the model was achieved in the subsequent annealing cycles. A detailed comparison of Figures 3 and 4 reveals that the model inferred by our approach has the opposite handedness compared to the model from Rubinstein *et al.* (2003).

3.2.3 GroEL GroEL also produced a qualitatively accurate model by the end of the first cycle of annealing, which can be seen in Figure 4c. In addition, it took four annealing cycles to converge to the final model. The final model has the important characteristics of GroEL, including a central channel and an approximate 7-fold symmetry. The 7-fold symmetry is more obvious from the side views than it is from the top view. (See Supplementary Material animation file GroELTopAndSide.avi). Interestingly, some of the intermediate models for GroEL exhibit better symmetry, probably indicating some overfitting in the later iterations and cycles (compare iterations 10 and 15 against iteration 20 in Fig. 4c). The algorithm's discovery of the 7-fold symmetry arises without its explicit specification in the inference.

3.2.4 Ribosome Ribosome produced a qualitatively accurate model by the end of the first cycle of annealing. However, in the subsequent cycles, the model looks less similar to the reference model, and the measured resolution falls (possible explanations are discussed in Section 4).

3.2.5 V-type ATPase: V-type ATPase also produced a qualitatively accurate model by the end of the first cycle.

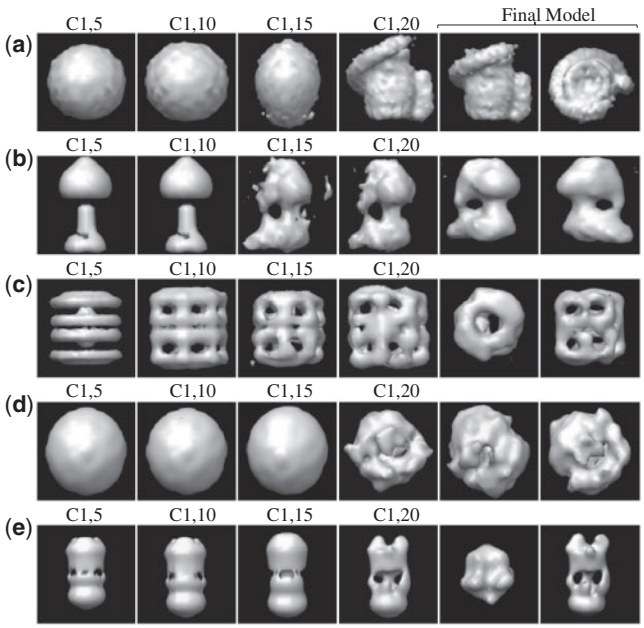

Fig. 4. Model fitting over different annealing cycles and iterations. This figure shows the inferred model at different annealing cycles and iterations. The notation above the images represents the cycle number, and iteration number in the cycle. (**a**) The phantom structure reaches a reasonable model by the end of the first cycle of annealing. Shown are the inferred structures after iterations 5, 10, 15 and 20 of the first cycle [C1,(5,10,15,20)] and two views of the final model after one cycle. (**b**) A reasonable structure is inferred for ATP synthase by the end of the first cycle of annealing. The final model was achieved after six cycles. (**c**) A reasonable structure is inferred for GroEL by the end of the first cycle of annealing. The final model was achieved after four cycles. (**d**) A reasonable structure is inferred for ribosome by the end of the first cycle. However, the quality of model degrades in the subsequent cycles. (**e**) A reasonable structure is achieved for V-type ATPase by the end of the first cycle.

As was seen with ATP synthase, earlier iterations produced more cylindrically symmetric models, while the later iterations resulted in a model similar to that seen in Lau and Rubinstein (2010).

3.2.6 ATP synthase and GroEL under no assumptions of rotations Structure determination was also performed for ATP synthase and GroEL by sampling rotations from a uniform distribution rather from than their known distributions. Structures inferred with this method can be seen as animated structures in the Supplementary Material (see files ATPUniform.avi and GroELUniform.avi). It can be seen that GroEL produces a reasonable structure while the ATP reconstruction lacks some detail. This is probably because only 9 class averages were available for ATP synthase, while 50 class averages were used for GroEL.

3.2.7 Assessment of resolution We attempted to measure the resolution of the models through the use of the Fourier Shell Correlation (FSC) criterion (Harauz and van Heel, 1986; van Heel and Schatz, 2005). The typical approach of dividing the available set of particle images into two sets was not applicable here because we had only a small set of class average images. Instead, we used

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 31, 2016

the models available from previous reconstructions (or the known models) as reference models, and computed the FSC curves between our inferred models and these reference models. This method is appropriate, because we expect our algorithm to generate low-resolution models compared to the high resolutions these models were reconstructed at.

At FSC of 0.3, resolutions of 48.9, 40.7, 41.8, 96.1 and 50.6 Å were calculated for the phantom structure, ATP synthase (under a prior of equitorial views), GroEL (under a prior of top and side views), ribosome and V-type ATPase, respectively (FSC curves can be found in the Supplementary Material). Here, we assumed a pixel size of 11.2 to account for the down-sampling from particle sizes of 2.8 Å. The model inferred after the first cycle of annealing with ribosome was significantly better (43.8 Å) than the model inferred after the last cycle of annealing with the ribosome. The animated file provided for ribosome shows the model inferred after the first cycle. Possible reasons for the low quality of model inferred for ribosome in the final cycle are suggested in the Section 4.

3.2.8 Run time Experiments were performed on a Linux Ubuntu 9.04 64-bit OS with a 2.66 GHz Intel Core® i7 processor. Three-dimensional inference for ATP synthase, GroEL and V-type ATPase under the assumed equitorial and top view priors took ∼2 h for four to six cycles of annealing. Inference for the phantom dataset, ribosome, and GroEL under no assumptions of rotations, with 50 class averages each, took approximately a week of computation for four to six cycles of annealing (with each cycle taking one to two days). ATP synthase without any assumptions of rotations and nine class averages took ∼6 days.

4 DISCUSSION

We have shown above that our method performs reasonably well in discovering 3D models from class average images for five different systems. Here, we give insight into our motivations for the choices made in different components of the algorithm and provide some guidance on how *ab initio* models may be created for novel systems using our method.

4.1 Data and likelihood model

The first design decision was the choice of the representation used to express an arbitrary structure. We are attempting to create a representation for a structure with continuous Coulomb density using a finite number of parameters. The resulting representation needs to possess some intuitive properties, such as gradual transition of Coulomb density from one vertex to another in a voxel. Our use of tri-linear interpolation permits us to represent the continuous Coulomb density of the protein structure using a finite number of parameters. In addition, the Coulomb density transitions smoothly between vertices within a voxel (although the derivative of the intensity at the edge of a voxel may be discontinuous) and the projection in any direction takes into account the length of the ray (or amount of matter) passing through a voxel in a given orientation. In contrast, Scheres *et al.* (2007) uses blobs that are a specific generalization of Kaiser-Bessel functions (Lewitt, 1992), centered at the edges of the voxels. These basis functions are spherically symmetric functions with limited support, and have the advantage that their projections can be computed easily and rapidly. One possible disadvantage that blobs may have is that in a representation with blobs centered on voxel edges, densities can arise which have values in the middle of the voxel that are lower or higher than values at any of the edges; tri-linear interpolation on voxels does not suffer from this property. A formal comparison of a pixel- and a blob-based representation suggests that a pixel-based representation may offer better resolution (Lee, 1993). However, in practice, blobs perform as well as pixel-based representations, at much faster speeds, as shown in Marabini *et al.* (1998). In future work, we will explore that advantage of using this alternative representation.

Our tri-linear model also allows for easy transition to higher resolution models using a previously fit, low-resolution model. For example, if a model was fit using a $32 \times 32 \times 32$ grid, it could be used as a seed for a higher resolution model with a $128 \times 128 \times 128$ grid. This is done by re-sampling the $32 \times 32 \times 32$ grid using the tri-linear interpolation on internal points.

One theoretical disadvantage of our representation is that the derivative of the density is discontinuous along the edges of voxels. It is unclear if other interpolation methods (e.g. tri-quadratic) would give rise to significantly different results.

4.2 Prior distribution and regularization

The other main component of our model is the prior probability distribution representing our belief about the likelihood of different densities. The prior probability distribution is important because the space of possible 3D densities is generally going to be underconstrained given the relatively small number of observations (i.e. class averages) available for a typical dataset. For example, with the ATP synthase dataset we had nine class average images with 1024 pixels each, while we had approximately 32 000 unknown parameters. To handle this sparsity of information, the prior captures our beliefs about the plausible distributions of the Coulomb density, providing additional constraints in the model fitting. Further, any choice of prior must not be so restrictive that valid models are penalized.

A suitable assumption in modeling a physical structure is that the amount of change in Coulomb density from one voxel to another is low over the complete set of voxels. The prior we chose is based on this assumption. Our value of $\lambda = 10^{-4}$ performed well for all five systems, leading to solutions that retain important features without excess noise. Increasing λ causes the model to have a lower resolution, as high-frequency changes in intensity of voxels are removed, while decreasing it causes the final model to have more spurious changes of intensity from one voxel to the next. It may be that different structures require different values of λ for correct inference. A value of λ that would be appropriate for a large set of structures may be chosen by an analysis of the structures available in the EMDB (http://www.ebi.ac.uk/pdbe/emdb/).

Typically, spurious models fit with only a few class averages will have a large number of voxels assigned small density values. This is because such models can satisfy the constraints that are enforced by the observed pixels of the few class averages, without many significant violations. Conversely, such models rarely have very large values because they would lead to projections that significantly violate some constraints. Thus, we also enforce the assumption that voxels are effectively zero unless there is strong evidence to the contrary, thereby limiting the degrees of freedom of the model.

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 31, 2016

While we do not explicitly model this as a prior, it is taken into account in the fitting process with the use of an L_1 regularization. Our chosen value of $\beta = 10^{-4}$ (see Section 2.2.1) seems to perform reasonably for all five systems. Higher values of β cause solutions to be sparse and also affect resolution. With more images, a higher value of β may be used because the data is able to overpower the prior. It must be pointed out here that the ability to use the same parameters across all datasets relies strongly on our normalization procedure, which normalizes class average images by scaling the maximum intensity over all images down (or up) to 1.

4.3 Model fitting

The use of a OWL-QN for fitting data presents several advantages. It is well known that Quasi-Newton methods are able to, at least partially, account for the curvature in parameter space by using low-rank approximations to the Hessian of the parameter space (Nocedal and Wright, 2006). This leads to faster optimization as compared to a simple approach such as gradient descent. The use of line-searches in these procedures also obviates the need for specifying parameters, such as learning rate, which additionally complicate any fitting procedure that depends on gradient descent. The additional advantage of using OWL-QN over other Quasi-Newton procedures, such as L-BFGS, is that it specifically incorporates L_1-regularization in the optimization process.

Probably, the most important component for the success of this method is the use of deterministic annealing of the noise parameter, σ. It was seen in our experiments that starting with a random model and a value of σ close to the real value gave arbitrary, incorrect models. This is because small values of σ created many steep, local optima in the objective function, causing the model to remain close to its initial value. In this scenario, the parameters changed very little from the starting parameters, getting stuck in a local minima which were poor in explaining the data. In contrast, a high values of σ results in a smoother objective function and fewer local minima, allowing the space of models to be more easily explored. As the σ is reduced, more and more features of the real object emerge, and correct orientations contribute more to the model. An exponentially decaying annealing schedule allows us to reach an average structure quickly and spend more time in the refinement stages with lower σ.

In the first cycle of annealing, we anneal the noise estimate to an intermediate value, since an accurate estimate of the noise is not available at the start of this cycle (a value of 0.3 worked reasonably for all five of our test cases). We then compute the root mean square deviation between the projections of the models and the images at the end of the cycle and use it as the target noise SD to anneal down to in the subsequent cycle. As a result, by the end of our training we are annealing to the expected, true noise value. Other than allowing us to recover from local minima, the repeated cycles of annealing are, thus, also useful in annealing to a correct value of noise.

The above method of estimation of final noise is a local estimate, because of the inter-dependence between the final model inferred and the value of noise used in the inference. When the number of class averages is low, and a low value of variance is used, overfitting can result, which reinforces an even lower variance in the next iteration or cycle. As a result, it was seen that the final GroEL model inferred in the last cycle of annealing had less obvious 7-fold symmetry, than some of the models in the intermediate annealing cycles. Similarly, the ribosome structure inferred at the end of the first cycle is more accurate than that inferred at the end of the second and subsequent cycles.

4.4 General observations and conclusions

Our experiment with the ATP synthase dataset provided insight into the characteristic of noise. In our model, we assumed that each class average image had the same noise variance. However, we observed that different ATP synthase class average images from previous work (Rubinstein *et al.*, 2003) had statistically different noise levels. Inspection of the data revealed that different class average images had been computed using different numbers of particle images. As a result, the variance of noise was lower for class average images built from more particle images and higher for those built with fewer particle images. In our approach, we recreated the class average images by using an equal number of particles to try and achieve equal variance for each class average image. An alternate approach would be to allow each class average to have its own variance, and to define a prior on the variance parameters. The variance parameters would then be included in the set of parameters being optimized by OWL-QN. Such an approach would replace the deterministic annealing schedule, with an annealing schedule dependent on the data. We hope to explore this approach further in future work.

The resolution of sampling over rotation space can significantly affect the resolution of the inferred structure and the computational time needed to fit a model. By using an appropriate prior distribution of rotations, we can get a better model faster than by using a completely uniform distribution of rotations because time is spent in sampling relevant rotations only. For example, in the case of ATP synthase, a much larger sample of rotations from a uniform distribution is required, compared to that from an equatorial distribution to achieve similar granularity of the relevant rotations; this granularity also results in a large number of irrelevant rotations which do not contribute to the model, and in fact can cause spurious reconstructions when an insufficient number of class average images are available. We expect that the insufficient sampling of rotations and low number of class averages was the reason why the structure inferred for ATP under uniformly distributed rotations lacked sufficient detail (see Supplementary Material file ATPUniform.avi). Thus, wherever possible, rotations should be distributed according to an appropriate prior distribution.

The resolution of the reconstruction should also be affected by the down-sampling of the images, which we performed for computational efficiency. In 3D reconstruction with raw particle images this can have a significant impact, when there are hundreds to thousands of particle images. However, the resolution of an initial model inferred from a small number of class averages is expected to be low even if larger images are used. By downsampling images by 4 times, we have reduced the number of pixels in images by 16 times, but we have also reduced the number of parameters by 64 times. Thus, it is expected that this down-sampling should possibly have less than a proportional effect on the resolution, in our inference procedure.

A major limitation of our method is the inability to judge the accuracy of the inferred model even when the correct model is given. Here, we used the FSC to assess the accuracy of the model, by comparing it to a known structure, or a high-resolution structure. Nevertheless, these numbers are biased in a non-obvious way by the prior which enforces continuity between pixels and the L1

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 31, 2016

regularization that favors solutions with large number of zero valued voxels. Thus, our reported resolutions should only be interpreted as crude approximations. Our model does, however, allow for a comparison between the accuracy of different models on the basis of the posterior probability of a model, which reflects the consistency between the experimental data and the model. Given two inferred structures, it is possible to state which one is better on the basis of higher posterior probability. However, this approach requires that the assumptions made by the model are valid. For example, we referred to the need to recreate class average images for ATP synthase to ensure that each of the class average images had the same noise SD. As such, proper model comparison may only be possible when our method creates models from raw particles where these assumptions are more likely to be valid.

We conclude by pointing out that the model we have defined permits sophisticated modifications in the future, by the virtue of its probabilistic nature. Such modifications include modern sampling techniques that allow for better sampling of the space of orientations as the model fitting proceeds, and annealed sampling procedures that explore modes more efficiently than our deterministic annealing technique. These techniques could additionally be applied to raw particles rather than class averages. Moreover, with an improved instrument noise model, it should be possible to compare 3D structures inferred from different methods just on the basis of the likelihoods. Lastly, even though we did not enforce symmetry for any of the structures in this article, our probabilistic method does allow us to incorporate prior knowledge of symmetry through the use of a prior that is symmetric.

We have released the source code of our program at http://compbio.cs.toronto.edu/cryoem/ to allow others to apply this software to their own data. A parallel implementation is being developed to apply the method to raw particle images.

ACKNOWLEDGEMENTS

We acknowledge Abraham Heifets, Izhar Wallach, Maria Mirza, Lindsay Baker, Fernando Flores-Mangas, Leonid Sigal and Iain Murray for valuable discussions and Zongyi Yang for developing a parallel implementation of the program. The GroEL dataset was obtained from the National Resource for Automated Molecular Microscopy which is supported by the National Institutes of Health through the National Center for Research Resources P41 program (RR17573).

Funding: Portions of this research were funded by grants to RHL from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Bill and Melinda Gates Foundation and to JLR from the Canadian Institutes of Health Research (MOP81294).

Conflict of Interest: none declared.

REFERENCES

Andrew,G. and Gao,J. (2007) Scalable training of L_1-regularized log-linear models. In *ICML'07: Proceedings of the 24th International Conference on Machine Learning*. ACM, New York, NY, pp. 33–40.

Baker,L.A. and Rubinstein,J.L. (2008) Angle determination for side views in single particle electron microscopy. *J. Struct. Biol.*, **162**, 260–270.

Cheng,Y. and Walz,T. (2009) The advent of near-atomic resolution in single-particle electron microscopy. *Ann. Rev. Biochem.*, **78**, 723–742.

Fernandez,J.J. *et al.* (2006) Image processing and 3-d reconstruction in electron microscopy. *IEEE Signal Process. Mag.*, **23**, 84–94.

Frank,J. (2006) *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State.* 2nd edn. Oxford University Press, New York, NY, USA.

Grigorieff,N. (1998) Three-dimensional structure of bovine nadh:ubiquinone oxidoreductase (complex I) at 22 Å 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.

Harauz,G. and Ottensmeyer,F.P. (1984a) Direct three-dimensional reconstruction for macromolecular complexes from electron micrographs. *Ultramicroscopy*, **12**, 309–319.

Harauz,G. and Ottensmeyer,F.P. (1984b) Nucleosome reconstruction via phosphorus mapping. *Science*, **226**, 936–940.

Harauz,G. and van Heel,M. (1986) Exact filters for general geometry three dimensional reconstruction. *Optik*, **73**, 146–156.

Harauz,G. *et al.* (1988) Statistical image analysis of electron micrographs of ribosomal subunits. In Harry,F.N.J. and Kivie,M. (eds) *Ribosomes*, Vol. 164 of *Methods in Enzymology*. Academic Press, USA, pp. 35–49.

Iba,H. *et al.* (2003) Inference of Euler angles for single-particle analysis by means of evolutionary algorithms. *Biosystems*, **72**, 43–55.

Jiang,W. *et al.* (2008) Backbone structure of the infectious ϵ-15 virus capsid revealed by electron cryomicroscopy. *Nature*, **451**, 1130–1134.

Lau,W.C.Y. and Rubinstein,J.L. (2010) Structure of intact Thermus thermophilus V-type ATPase by cryo-EM reveals organization of the membrane-bound VO motor. *Proc. Natl Acad. Sci. USA*, **107**, 1367–1372.

Lee,D. (1993) Cramer-rao bound and image representation in emission tomography. In *Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record.* Vol. 3, San Francisco, CA, USA, pp. 1458–1462.

Leschziner,A.E. and Nogales,E. (2006) The orthogonal tilt reconstruction method: an approach to generating single-class volumes with no missing cone for ab initio reconstruction of asymmetric particles. *J. Struct. Biol.*, **153**, 284–299.

Lewitt,R.M. (1992) Alternatives to voxels for image representation in iterative reconstruction algorithms. *Phys. Med. Biol.*, **37**, 705.

Ludtke,S.J. *et al.* (2008) De novo backbone trace of GroEL from single particle electron cryomicroscopy. *Structure*, **16**, 441–448.

Marabini,R. *et al.* (1998) 3d reconstruction in electron microscopy using art with smooth spherically symmetric volume elements (blobs). *Ultramicroscopy*, **72**, 53–65.

Nocedal,J. and Wright,S.J. (2006) *Numerical Optimization*. 2nd edn. Springer, New York, USA.

Ogura,T. and Sato,C. (2006) A fully automatic 3D reconstruction method using simulated annealing enables accurate posterioric angular assignment of protein projections. *J. Struct. Biol.*, **156**, 371–386.

Penczek,P.A. *et al.* (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.

Penczek,P.A. *et al.* (1996) A common-lines based method for determining orientations for N > 3 particle projections simultaneously. *Ultramicroscopy*, **63**, 205–218.

Potter,C.S. *et al.* (2004) Automated particle selection for cryo-electron microscopy. *J. Struct. Biol.*, **145**, 1–2.

Radermacher,M. *et al.* (1987) Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50s ribosomal subunit of escherichia coli. *J. Microsc.*, **146**, 113–136.

Rose,K. *et al.* (1990) A deterministic annealing approach to clustering. *Pattern Recognit. Lett.*, **11**, 589–594.

Rubinstein,J.L. *et al.* (2003) Structure of the mitochondrial ATP synthase by electron cryomicroscopy. *EMBO J.*, **22**, 6182–6192.

Sahr,K. *et al.* (2003) Geodesic discrete global grid systems. *Cartogr. Geogr. Inf. Sci.*, **30**, 121–134.

Samsó,M. *et al.* (2002) A Bayesian method for classification of images from electron micrographs. *J. Struct. Biol.*, **138**, 157–170.

Sanz-García,E. *et al.* (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. *et al.* (2005a) Fast maximum-likelihood refinement of electron microscopy images. *Bioinformatics*, **21** (Suppl. 2), 243–244.

Scheres,S.H.W. *et al.* (2005b) Maximum-likelihood multi-reference refinement for electron microscopy images. *J. Mol. Biol.*, **348**, 139–149.

Scheres,S.H.W. *et al.* (2007) Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nat. Methods*, **4**, 27–29.

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 31, 2016

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

Sorzano,C.O. *et al.* (2004) XMIPP: a new generation of an open-source image processing package for electron microscopy. *J. Struct. Biol.*, **148**, 194–204.

Stagg,S.M. *et al.* (2008) A test-bed for optimizing high-resolution single particle reconstructions. *J. Struct. Biol.*, **163**, 29–39.

Thuman-Commike,P.A. (2001) Single particle macromolecular structure determination via electron microscopy. *FEBS Lett.*, **505**, 199–205.

van Heel,M. and Frank,J. (1981) Use of multivariate statistics in analysing the images of biological macromolecules. *Ultramicroscopy*, **6**, 187–194.

van Heel,M. and Schatz,M. (2005) Fourier shell correlation threshold criteria. *J. Struct. Biol.*, **151**, 250–262.

van Heel,M. *et al*. (1997) Angular reconstitution in three-dimensional electron microscopy: historical and theoretical aspects. *Scanning Microsc.*, 195–210.

van Heel,M. *et al*. (2000) Single-particle electron cryo-microscopy: towards atomic resolution. *Q. Rev. Biophys.*, **33**, 307–369.

van Heel,M. (1984) Multivariate statistical classification of noisy images (randomly oriented biological macromolecules). *Ultramicroscopy*, **13**, 165–183.

van Heel,M. (1987) Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy*, **21**, 111–123.

Walz,J. *et al*. (1997) Tricorn protease exists as an icosahedral supermolecule in vivo. *Mol. Cell*, **1**, 59–65.

Yu,X. *et al*. (2008) 3.88 Å structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. *Nature*, **453**, 415–419.

Zhang,X. *et al*. (2008) Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. *Proc. Natl Acad. Sci. USA*, **105**, 1867–1872.

Zheng,Y. *et al*. (2002) 3-d maximum likelihood reconstructions of viruses from cryo electron microscope images and parallel computation. In *Proceedings of the 2002 International Conference on Image Processing (ICIP 2002), Rochester, New York, USA, September 22–25, 2002*. Vol. 2, IEEE, pp. II-617–II-620.

Zhou,Z.H. (2008) Towards atomic resolution structural determination by single-particle cryo-electron microscopy. *Curr. Opin. Struct. Biol.*, **18**, 218–228.

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 31, 2016