



## A Toolbox for *ab initio* 3-D reconstructions in single-particle electron microscopy

Neil R. Voss<sup>a,1</sup>, Dmitry Lyumkis<sup>a,1</sup>, Anchi Cheng<sup>a</sup>, Pick-Wei Lau<sup>a</sup>, Anke Mulder<sup>a</sup>, Gabriel C. Lander<sup>a</sup>, Edward J. Brignole<sup>a</sup>, Denis Fellmann<sup>a</sup>, Christopher Irving<sup>a</sup>, Erica L. Jacovetty<sup>a</sup>, Albert Leung<sup>a</sup>, James Pulokas<sup>a</sup>, Joel D. Quispe<sup>a</sup>, Hanspeter Winkler<sup>b</sup>, Craig Yoshioka<sup>a</sup>, Bridget Carragher<sup>a</sup>, Clinton S. Potter<sup>a,\*</sup>

<sup>a</sup> National Resource for Automated Molecular Microscopy and Department of Cell Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

<sup>b</sup> Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306, USA

### ARTICLE INFO

#### Article history:

Received 3 September 2009

Received in revised form 2 December 2009

Accepted 3 December 2009

Available online 16 December 2009

#### Keywords:

Transmission electron microscopy (TEM)

Single-particle analysis

Initial model

3-D reconstruction

50S ribosomal subunit

### ABSTRACT

Structure determination of a novel macromolecular complex via single-particle electron microscopy depends upon overcoming the challenge of establishing a reliable 3-D reconstruction using only 2-D images. There are a variety of strategies that deal with this issue, but not all of them are readily accessible and straightforward to use. We have developed a “toolbox” of *ab initio* reconstruction techniques that provide several options for calculating 3-D volumes in an easily managed and tightly controlled workflow that adheres to standard conventions and formats. This toolbox is designed to streamline the reconstruction process by removing the necessity for bookkeeping, while facilitating transparent data transfer between different software packages. It currently includes procedures for calculating *ab initio* reconstructions via random or orthogonal tilt geometry, tomograms, and common lines, all of which have been tested using the 50S ribosomal subunit. Our goal is that the accessibility of multiple independent reconstruction algorithms via this toolbox will improve the ease with which models can be generated, and provide a means of evaluating the confidence and reliability of the final reconstructed map.

© 2009 Elsevier Inc. All rights reserved.

### 1. Introduction

The “initial model problem” in single-particle electron microscopy (EM) refers to the difficulty of obtaining an accurate preliminary three-dimensional representation of a sample exclusively from two-dimensional images. The addition of a spatial dimension requires further knowledge of the relative angular orientation of the 2-D projections, termed Euler angles, such that the latter can be properly oriented and back-projected to assemble a 3-D structure. The task of determining Euler angles has been actively addressed with a wide range of techniques, but each has drawbacks and limitations. A crucial caveat for the construction of *any* initial model is that a map will emerge, regardless of its correctness. Discriminating between accurate vs. inaccurate representations of a sample can be an arduous task, often requiring additional biochemical and/or structural analysis. Furthermore, noise and reference bias in the 2-D alignment can produce over-fitting, leading to general skepticism of preliminary models resulting from a single

3-D reconstruction package (Stewart and Grigorieff, 2004). With this in mind, we have developed a standardized toolbox of initial model generators that include most available software packages and that implements uniform conventions, with the ultimate goal of converging at a single, refined structure.

In transmission electron microscopy, the internal features present in a 3-D object are transferred onto 2-D micrographs, and so it has long been known that the structure of the object can be reconstructed from its corresponding 2-D projections ((Crowther et al., 1970); for reviews, see (Crowther and Klug, 1975; Frank, 1981)). Such reconstructions are often based on the so-called central slice theorem, which states that a 2-D projection of a 3-D object represents a 2-D slice through the center of the 3-D density distribution in Fourier space. Utilization of this theorem will assign orientations to 2-D projections, which would then form the 3-D object. An alternative to computationally determining Euler angles is to use physical tilting of the specimen inside the microscope to provide constraints on the relative angles between 2-D projection views. Random-conical tilt reconstructions (Radermacher et al., 1986), orthogonal tilt reconstructions (Leschziner and Nogales, 2006), and tomographic reconstructions (Frank, 2006) all use variations of this approach.

Angular reconstitution, a real-space version of the central slice theorem, is one example of a common lines based algorithm used

\* Corresponding author. Address: The Scripps Research Institute, 10550 North Torrey Pines Road, CB-129, La Jolla, CA 92037, USA. Fax: +1 858 784 9090.

E-mail address: [cpotter@scripps.edu](mailto:cpotter@scripps.edu) (C.S. Potter).

<sup>1</sup> Authors contributed equally to this work.

for *ab initio* angular assignment. The algorithm assumes particles are distributed with random orientations and identifies intersecting 1-D lines from different 2-D projections (van Heel, 1987). Initially, three 2-D projections, preferably corresponding to orthogonal views of the 3-D object, are chosen as starting points to fix the selected images in a single orientation with respect to each other. This is then followed by a brute-force search, whereby the 1-D lines for each consecutive projection are compared to all 1-D lines of those projections whose orientation is already determined. An alternative Fourier-space version of the central slice theorem (Ludtke et al., 1999) uses a cross common line approach that searches for the orientation and center parameters to minimize the mean phase differences between intersections of 2-D planes (i.e. common lines) in Fourier space. In practice, the low signal-to-noise ratio (SNR) of raw images usually limits the use of both methods to class averages, the combinatorial sum of aligned, identical single particles, which have significantly higher signal-to-noise ratios. Additionally, factors such as conformational heterogeneity, misalignment, misclassification, and preferred orientation (for a review, see van Heel et al. (2000)) contribute to the difficulty of obtaining a reliable *ab initio* reconstruction using common lines approaches, particularly for objects devoid of internal symmetry.

Tomography is the most straightforward technique used to directly obtain a 3-D structure from a set of 2-D projections, and involves collecting an entire series of tilted images at a given angular increment (for a comprehensive review see Frank (2006)). Though not always deemed as such, it is in fact the quintessential “single-particle” reconstruction method, since a separate tomogram and the resulting 3-D structure can be uniquely obtained for every macromolecular complex lying on the grid, avoiding the averaging of individual particles in conventional single-particle EM. A crucial drawback, however, is that physical tilting of the grid beyond  $\sim 60^\circ$  in either direction leads to a rapid increase in the effective specimen thickness, thus limiting electron contribution to image formation, in practice resulting with a consequent missing wedge of information. Furthermore, because each tomogram represents a collection of images corresponding to a single region on the grid, the electron dose must be significantly lowered to prevent radiation damage to the sample. The lower SNR often eliminates the possibility of image alignment on individual particles as is done in conventional single-particle reconstructions and, unless fiducial markers (e.g. electron dense gold clusters) are introduced, the alignment is instead performed at a much-less accurate global scale of the micrographs, ultimately lowering the overall resolution of the resulting structure.

Alternative *ab initio* reconstruction techniques involve the collection of only two images from the same region of the grid, each acquired at a different tilt relative to the electron beam. In both random-conical tilt (RCT) (Radermacher et al., 1986) and orthogonal tilt reconstruction (OTR) (Leschziner and Nogales, 2006) methodologies, the two angles provide different views of identical particles, constraining the geometric relation between them. RCT requires image pairs, one untilted and the other tilted to  $45^\circ$  or more. Untilted particle images are aligned and classified, putting identical views in a single in-plane orientation. The corresponding tilted particle images, whose relative geometry is established, are used to back-project and reconstruct the 3-D volume. The inherently random in-plane orientation of identical views allows the corresponding tilted images to be back-projected in a conical fashion, thereby reducing the missing wedge to a missing cone. Elimination of missing information can be achieved with OTR, which differs from RCT in that both image pairs are tilted and collected at  $-45^\circ$  and  $+45^\circ$  inside the microscope, providing the equivalent of a  $90^\circ$  tilt angle compared to RCT. The method is complicated by the 2-D alignment of tilted data and requires large numbers of particles exhibiting no preferred orientation on the grid. In the

case of a single orientation, OTR is unable to create a reliable model, as the particle tilt pairs for any given class average would not fill the entire Fourier space.

The method of choice for creating *ab initio* reconstructions often depends on the availability of a particular algorithm within the EM processing packages utilized in a lab. While it makes sense to use a variety of approaches on a completely unknown structure, in practice, incompatible data and parameter file formats, along with an absence of established mathematical conventions makes this process challenging and encourages package lock-in (reviewed in (Carragher and Smith, 2008)). To this end, we have developed a toolbox of *ab initio* reconstruction techniques that streamlines and largely automates most of the intermediate steps for each of the most commonly used algorithms. The toolbox guides the user through the steps of creating *ab initio* reconstructions using the Appion pipeline (Lander et al., 2009), thus making it straightforward to obtain reconstructions using a variety of orthogonal techniques. Below we provide a brief introduction to the overall architecture of the “initial model” pipeline and provide examples of how each of the major *ab initio* reconstruction methods has been implemented. The basic functionality of each of these methods was validated using five distinct 50S large ribosomal subunit datasets, and the results for the reconstruction of the 50S subunit are described and discussed.

## 2. Methods

### 2.1. Sample preparation

The 50S ribosomal samples were prepared from *Escherichia coli* MRE600 cells as described previously (Bunner et al., 2008). Briefly, ribosomal subunits were isolated as 70S particles from cells, dissociated into individual subunits by dialysis, and separated over sucrose gradient. For EM sample preparation, 50S samples at a concentration of  $24 \mu\text{M}$  in buffer (50 mM Tris-HCl, pH 7.8, 10 mM  $\text{MgCl}_2$ , 100 mM  $\text{NH}_4\text{Cl}$ , 6 mM  $\beta\text{ME}$ ) were diluted 1:200 for vitreous ice and 1:1000 for negative stain.

Carbon coated and C-flat (Protochips, Inc.) grids were cleaned using a Gatan Solarus plasma cleaner (5 s, 25%  $\text{O}_2$ , 75% Ar). For negatively-stained specimens, a  $3 \mu\text{L}$  drop of 50S sample was applied to the grid followed by 2% Uranyl acetate using a deep-staining protocol (Ohi et al., 2004). For carbon sandwich specimens, an additional layer of thin carbon was floated onto the grid. 50S ribosomal subunits preserved in vitreous ice were prepared on C-flats overlaid with an additional thin layer of carbon and samples were vitrified using an FEI Vitrobot.

### 2.2. Data collection

All data collection was performed using a transmission electron microscope (FEI Tecnai F20) operating at 120 kV. Four major datasets were collected for the 50S ribosomal subunit. RCT (Dataset 1) and tomography (Dataset 2) datasets were collected with 50S subunits embedded in deep stain. Common lines *ab initio* reconstructions (Dataset 3) and projection-matching refinements (Dataset 4) were carried out on 50S subunits embedded in vitreous ice. All images were recorded with a Tietz F415  $4\text{k} \times 4\text{k}$  pixel CCD camera ( $15 \mu\text{m}$  pixel) and were collected at  $50,000\times$  magnification (0.163 nm pixel size) using the Leginon data collection software (Suloway et al., 2005, 2009; Yoshioka et al., 2007).

### 2.3. The Appion image-processing architecture

All *ab initio* methods are implemented within the Appion pipeline (Lander et al., 2009). Written as a set of python wrappers for a

wide array of existing processing software, Appion provides inter-package compatibility, while maintaining a generic, intuitive interface meant to guide the user through each step of image analysis. It is connected to a centralized database that stores, tracks, and links all the relevant input and output parameters, beginning with the creation of a micrograph and continuing through to the final 3-D reconstruction. A web-based graphical user interface (AppionWeb) provides the means by which to launch processing jobs, while graphical and tabular summaries display all relevant output at the end of each run (Supplementary Fig. 1). Unless relevant to the result, intermediate steps in the processing are largely automated (Supplementary Fig. 2 provides one such example). This general framework keeps track of experiments and all metadata related to the creation of *ab initio* reconstructions.

#### 2.4. Standardized conventions

The realization of inter-package compatibility depends on the standardization of both mathematical and lexical conventions in EM. The alignment routine illustrates the problem at hand. First, the direction of translations and rotations are often defined differently among various packages (Table 1). Second, the order of operations in carrying out the transformation also varies. For example, in order to place a particle in a specific orientation, Xmipp first shifts, then mirrors, and finally rotates each image, whereas SPIDER shifts, rotates, then mirrors. IMAGIC, on the other hand, gives the option of either shifting or rotating first, specified by the user. Notably, the outcome depends on the order of operations, while the bookkeeping is further complicated by the exact geometry of the coordinate system in use. Even lexical definitions have suffered from such inconsistencies, as in the case where “clustering”, “classification,” and “data analysis” can all refer to an identical operation. We have adhered to the conventions suggested by Heymann et al., (Heymann et al., 2005) which allow us to inter-convert parameters between packages and maintain a single, standardized format.

#### 2.5. Particle alignment and classification

An integral component of most *ab initio* 3-D reconstruction approaches is the initial alignment and classification of 2-D images. We have incorporated a variety of alignment and classification tools into the pipeline, including methods from Xmipp (Pascual-Montano et al., 2001; Sorzano et al., 2004; Scheres et al., 2005a,b), IMAGIC (van Heel et al., 1996) and SPIDER (Frank et al., 1996), to perform reference-free and reference-based particle alignment, feature analysis, and particle clustering. With six different alignment routines and four different classification routines currently in the pipeline, it is already possible to perform 24 inde-

pendent particle classifications. An example of the available methods along with input and output parameters for the pipeline is shown in Fig. 1. As always, the pipeline conforms to the general principles of the Appion architecture, employing a standardized format that allows for inter-package compatibility.

#### 2.6. Initial model calculation and *ab initio* reconstruction within Appion

Conventional approaches for creating *ab initio* reconstructions and supplying initial models for structural refinement have been largely automated within the Appion infrastructure. In certain cases, minor changes have been made to either improve the algorithm or simply facilitate its integration into Appion. Each approach is schematically summarized in Fig. 2, with some additional details explained below.

##### 2.6.1. Pre-existing models

Appion allows for the use of any pre-existing model as a starting point for 3-D reconstruction. The PDB uploader converts PDB coordinates (Berman et al., 2000) into volume data with a user-specified pixel size, box size, and resolution. An additional option is use of the “biological unit” as defined in the PDB, important for structures with internal symmetry (e.g. viruses). An EMDB uploader performs a largely similar task from the electron microscopy database (Tagari et al., 2002). Finally, a manual uploader allows users to incorporate any custom initial model into the pipeline, a tool that we have found to be particularly valuable for visiting scientists, arriving with an initial model on a thumb drive in their pocket.

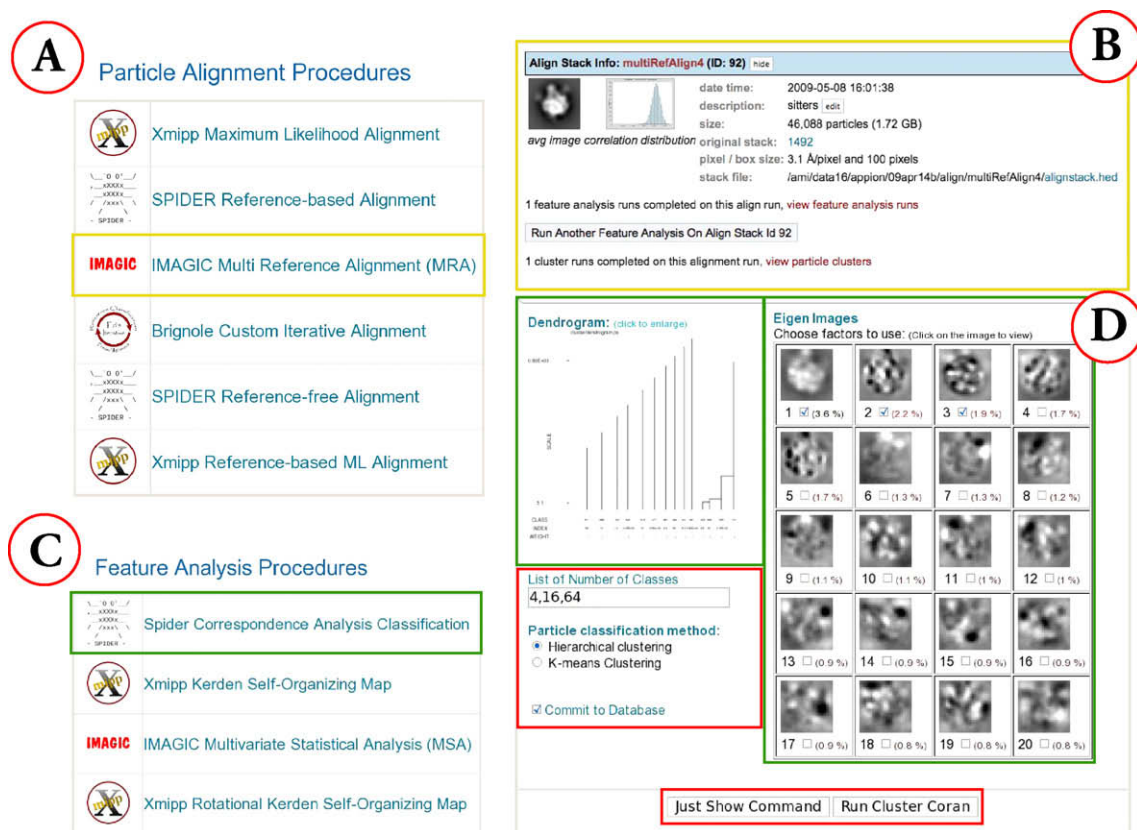
##### 2.6.2. Angular reconstitution

An illustrative summary of the method is shown in Supplementary Fig. 3. It begins with class averages that have been pre-processed through iterative alignment and classification using the Appion pipeline (Supplementary Fig. 3A). The user selects three initial class averages, ideally representing different axial views of the 3-D object, and inputs particle-specific parameters into the web-based form (Supplementary Fig. 3B). The 3d0 batch script then performs Euler angle assignments for the entire set of class averages, error-based sorting, 3-D reconstruction, and calculation of the Fourier-shell correlation (FSC). As expected, the three chosen class averages have a large influence on the result, so it is imperative to run multiple initializations to obtain a consistent final model. The 3d0 summary page (Supplementary Fig. 3C) aids in this process, allowing the user to quickly assess the qualitative and quantitative accuracy of the resulting density prior to launching an iterative 3-D refinement. This final step (Supplementary Fig. 3D) applies multi-reference alignment and multivariate

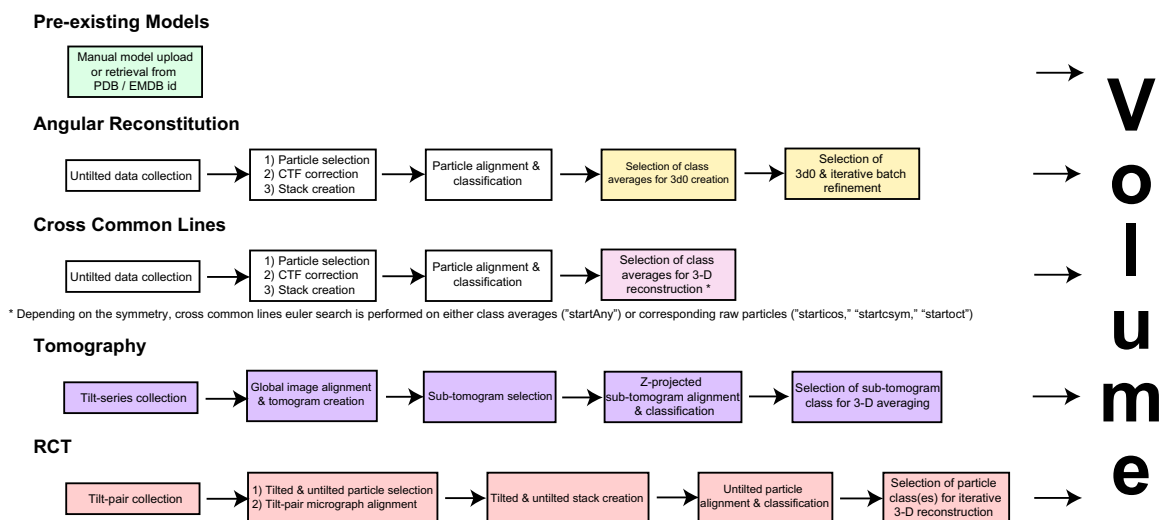
**Table 1**

Varying conventions among different EM software packages used in *ab initio* reconstructions. Results here corresponds to the particle transformation viewed in EMAN's V2 viewer (Ludtke et al., 1999). When given a positive angle in the specific processing package, the in-plane rotation defines the direction in which a 2-D image is turned according to the viewer. Similarly, positive X and Y translations move the image according to the first and second shift direction, respectively. The Euler system refers to the axes of rotation for 3-D volumes. The file format refers to the main format used by the package.

Program	Convention			
	In-plane rotation	Shift direction (x-y)	Euler system	Primary file format
Heymann et al.	Counterclockwise	Right-up	ZYZ	–
Bsoft	Counterclockwise	Right-up	ZYZ	Many
Xmipp	Clockwise	Right-up	ZYZ	SPIDER single
SPIDER	Counterclockwise	Right-up	ZYZ	SPIDER
EMAN1	Counterclockwise	Right-up	ZXZ	MRC, IMAGIC
EMAN2	Counterclockwise	Right-up	ZXZ	HDF5
IMAGIC	Clockwise	Down-right	ZYZ	IMAGIC
Python-Numpy	Counterclockwise	Right-up	–	–



**Fig. 1.** The alignment and classification pipeline. (A) Routines from different packages are available for raw particle alignment. (B) Summary page showing output for a stack of raw particles aligned with IMAGIC multi-reference alignment. The total sum from the aligned stack and a graphical correlation distribution are additionally displayed. (C) An aligned stack can be subjected to one of four feature analyses, as implemented in each of the processing packages. (D) A package-specific feature analysis summary form is displayed along with the further option of particle clustering into class averages.



**Fig. 2.** *Ab Initio* reconstructions and initial models within Appion: Each methodology for acquiring an initial 3-D structure, as implemented within the Appion framework, is illustrated with a step-by-step scheme, wherein each box represents the major stage at which user intervention is required. Colored boxes are specific to the particular methodology employed, whereas white boxes represent generic procedures in single-particle image analysis.

statistical analysis (Frank and van Heel, 1982; van Heel, 1984) to the raw particles to generate improved averages for each iteration, with the rest of the steps remaining largely identical to 3d0 generation. (Supplementary Fig. 2 shows output structures at various iterations of the refinement).

### 2.6.3. Cross common lines

A set of class averages from the alignment and classification pipeline is selected by the user. Depending on the presumed symmetry of the particles, an EMAN cross common-line search and 3-D reconstruction (Ludtke et al., 1999) is launched ("startAny" for C1



symmetry, “startcsym” for Cn symmetry, “starticos” for icosahedral symmetry, or “startoct” for octahedral symmetry).

#### 2.6.4. Tomography

An image stack is created from a tilt-series, and all data is converted to the standardized file formats and map orientations of IMOD (Kremer et al., 1996). The IMOD correlation-based fiducial-less global alignment algorithm then aligns the micrographs to each other, allowing for the creation of a full tomogram through 3-D back-projection. Alternatively, as is the case of this study, an external program (Winkler and Taylor, 2006) can be used to create the full tomogram, which is then uploaded back into the Appion database. A z-projection image of the tomogram is generated, displayed, and stored in the database during the upload, becoming available for further processing steps. Sub-tomograms are then extracted either manually or with the use of one of the automatic particle-picking algorithms already in the pipeline. These individual sub-tomograms are generally crude representations of the object, and it is best to perform multiple sub-tomogram averaging to reduce the effects of the missing wedge and improve the SNR. The z-axis projections of the individual sub-tomograms, each corresponding to a single macromolecular particle, can be readily processed using the conventional 2-D alignment and classification available in the pipeline. The sub-tomograms from a well-ordered projection class are then selected and centered along the z-axis, wherein the algorithm minimizes artifacts from misalignments and back-projections using thresholding criteria. The individual sub-tomograms are subsequently transformed according to their alignment and averaged to provide a 3-D model.

#### 2.6.5. RCT

Tilted data collection, particularly for vitreous ice specimens, is difficult and tedious, and was recently addressed by a method that uses a feature-based tracking algorithm to make automated data collection possible (Yoshioka et al., 2007). We have also addressed another rate-limiting step with the implementation of the TiltPicker program, which automates the particle matching process and determines all essential image tilt parameters (Voss et al., 2009). The matched particle coordinates are then used to create both a tilted and untilted stack, and the latter is processed using the Appion alignment and classification pipeline. The alignment parameters allow for the assignment of Euler angles to all tilted particle images, which are then back-projected into a volume for each class. A subsequent iterative centering operation uses cross-correlation of raw, tilted particles against corresponding volume projections to produce a refined *ab initio* reconstruction. An example of the output from the RCT pipeline is shown in Supplementary Fig. 4.

#### 2.6.6. Projection-matching refinement

All models were refined using EMAN projection-matching (Ludtke et al., 1999), wherein raw particles are aligned to projections covering the complete asymmetric unit, and the orientation of the best matched projection is assigned to each particle. We used 12 iterations of projection-matching, decreasing to an angular projection step size of 4°. Approximately 70% of raw particles were kept in each round.

#### 2.7. Synthetic data

An 80,172 particle synthetic dataset (Dataset 5) was used to validate all results. Its creation is summarized in Supplementary Fig. 5. Projections of the 50S ribosomal subunit (Matadeen et al., 1999) were created using EMAN project3d program with 68 iterations at an angular increment of 4° (1179 particles per iteration). The particles were randomly shifted, rotated, and flipped, and the SNR was reduced to 0.1. A contrast transfer function with a Gauss-

ian distribution of defocus values was applied to each particle using a mean defocus of  $-1.5\ \mu\text{m}$  and a standard deviation of  $0.4\ \mu\text{m}$ . This was followed by the application of an experimentally determined envelope function, and an additional layer of noise, bringing down the SNR to 0.05, in agreement with real EM data (Baxter et al., 2009). The applied contrast transfer function was corrected, but with the introduction of a small standard error ( $\sigma = \pm 0.04\ \mu\text{m}$ ) in an attempt to simulate the error associated with automated CTF estimation. Finally, the raw particle stack was filtered to maintain information between 5 and 600 Å, respectively.

### 3. Results

Initial models were generated using a variety of methods and then used as a starting points for a projection-matching refinement of an 82,575 particle dataset (Dataset 4) of 50S subunits embedded in vitreous ice and a synthetic 80,172 particle dataset (Dataset 5) created from computational projections of 50S subunits. The results are summarized in Fig. 3 and Supplementary Fig. 6 and described in more detail below. All processing was completed using algorithms incorporated into the Appion pipeline. A full textual description of the processing required to obtain each initial model is available in Supplementary material.

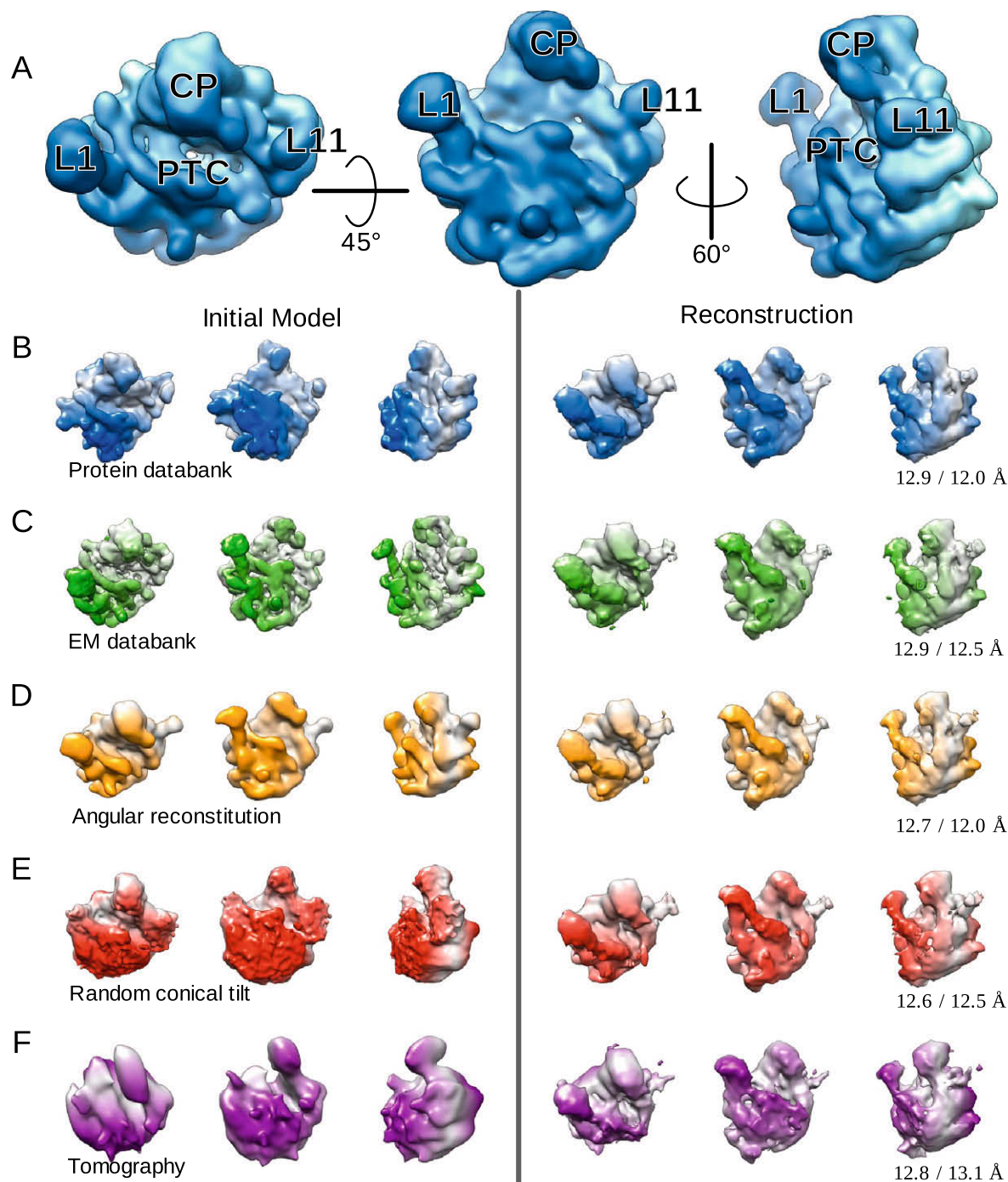
#### 3.1. Pre-existing models

The use of a similar structure available in one of the public databases provides the most straightforward means of generating an initial model. This might work well when the structure of interest is expected to vary from an existing model by only a minor conformational or structural change. Models of the 50S ribosomal subunits were uploaded from the PDB (PDB id 1JJ2 (Klein et al., 2001)) and EMDB (EMDB id 1019 (Matadeen et al., 1999)) databases. Both were low-pass filtered to 20 Å prior to use. During processing of the data, we determined that the pixel size of the EMDB structure needed to be adjusted by  $\sim 20\%$  (from 1.55 nm/pixel to 1.3 nm/pixel) so that it would correspond with the X-ray PDB maps and our experimentally acquired data. Although the PDB model (Fig. 3B, left) is missing the L1 protein (left arm) and most of the L11 and L7/12 proteins (right arm), this information is accurately recovered in the final reconstruction (Fig. 3B, right) after only two iterations. The EMDB initial model, once adjusted to the correct scale, also results in a refined structure that reproduces all of the major structural features of the 50S subunit (Fig. 3C).

#### 3.2. Common lines

The model from angular reconstitution (Fig. 3D, left) provides an example of the common lines methodology as a means to a high-resolution reconstruction in and of itself, attaining a resolution of 16.2 Å by FSC<sub>0.5</sub> prior to projection-matching refinement. Careful comparison with the EMDB structure shows that most major proteins and RNA densities are in agreement. The major exception is the L11 arm, a protein that is crucial for tRNA loading and is known to be conformationally variable (Kavran and Steitz, 2007). When compared to how the methodology is conventionally used, the major difference with our implementation of angular reconstitution lies in the high level of intrinsic automation. An essentially identical structure to that of the EMDB was obtained with only a few key steps requiring user intervention, namely the selection of initial class averages and an appropriate 3d0 model for batch refinement.

A second model was constructed using cross common lines (Fig. 4F, left; Fig. 5, bottom left). Inaccurate in some of its

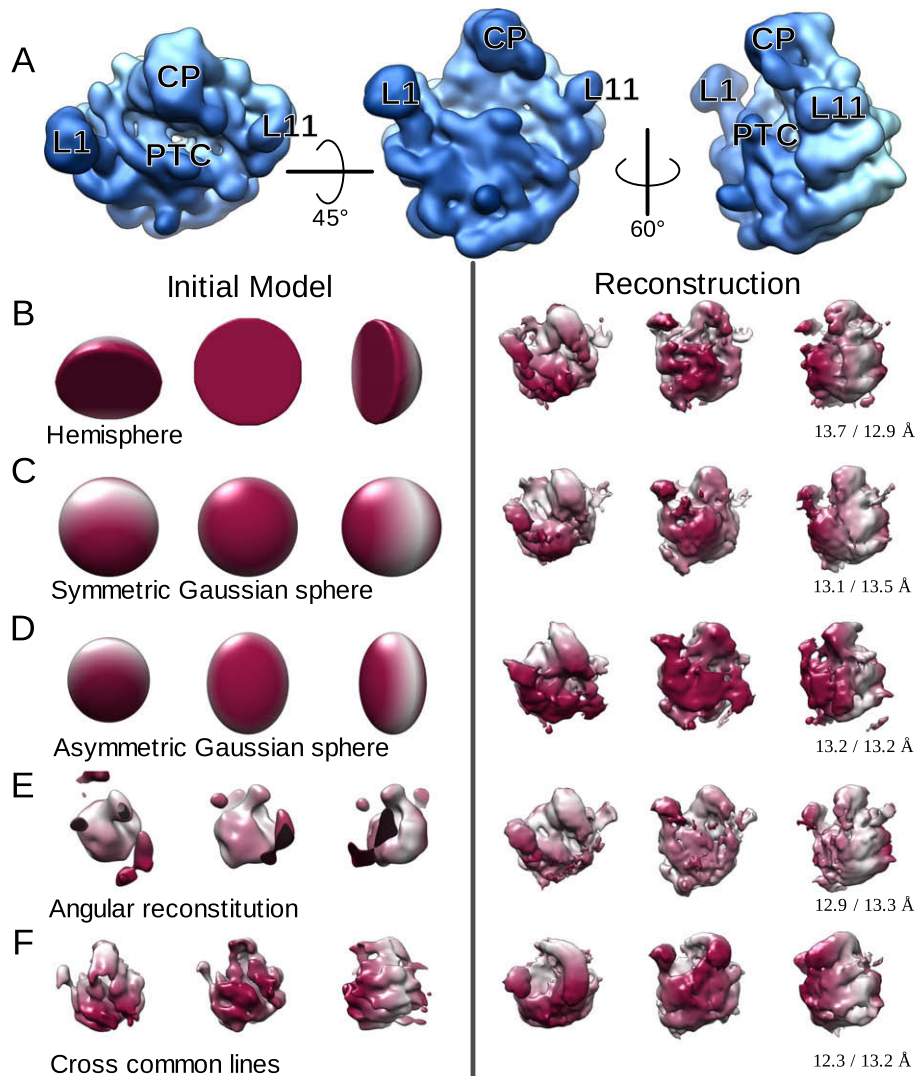


**Fig. 3.** Initial models and final reconstructions of the 50S ribosomal subcomplex. (A) EMDB ribosomal model showing relative rotations for each view with an emphasis on defining features such as the L1 and L11 protein arms, the central protuberance (CP), and the tRNA cleft with the protein transferase center (PTC). (B–F) Rotated views of the ribosomal subunit are shown for each initial model (left) and final reconstruction (right). The best resolution is given by FSC<sub>0.5</sub> and RMeasure (Sousa and Grigorieff, 2007), respectively. All parameters are summarized within [Supplementary material](#).

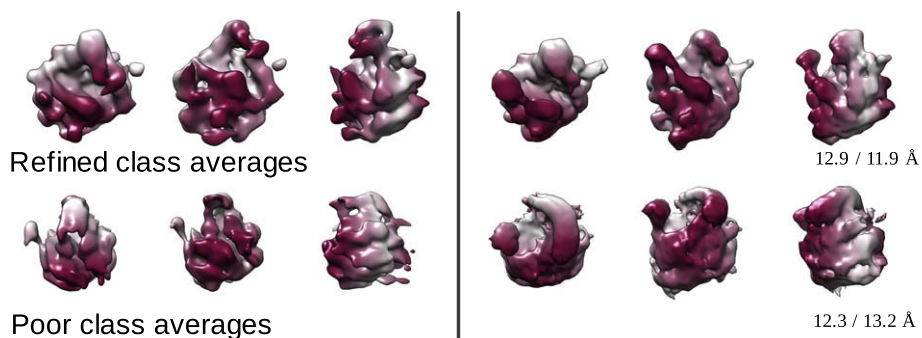
predominant features, the volume converged on a structure with an abnormally elongated central protuberance and an absence of the central cleft. One interesting observation was that upon substituting the input class averages by those produced in the last iteration of angular reconstitution batch refinement, a significantly cleaner and more accurate structure was attained (Fig. 5, top left). This suggests that *ab initio* reconstructions from common lines rely much more on proper classification and 2D alignment of the raw particles rather than the particular algorithm used for Euler angle determination.

### 3.3. Tomography

The presented tomographic model is averaged from 93 particles and provides an example of the minimum features required to converge to an accurate 3-D structure (Fig. 3F, left) (in our earlier attempt, a model averaged from only 22 particles belonging to one class with a limited in-plane rotation, and hence a limited Euler angle distribution, [Fig. 6B] also converged to an equally accurate refinement). In the case of the 50S subunit, it is merely the presence of appropriate object dimensions and a well-defined central



**Fig. 4.** Inaccurate reconstructions from initial models. All parameters for refinement were identical to those used to generate the reconstructions in Fig. 3. (A) EMDB model as in Fig. 3A. Iterative projection-matching refinement was performed on (B) a hemisphere, (C) symmetric Gaussian sphere, (D) asymmetric Gaussian sphere, (E) poorly reconstructed model from angular reconstitution using a preferred orientation dataset, and (F) cross common lines using a non-preferred orientation dataset. Convergence was not achieved for any model.

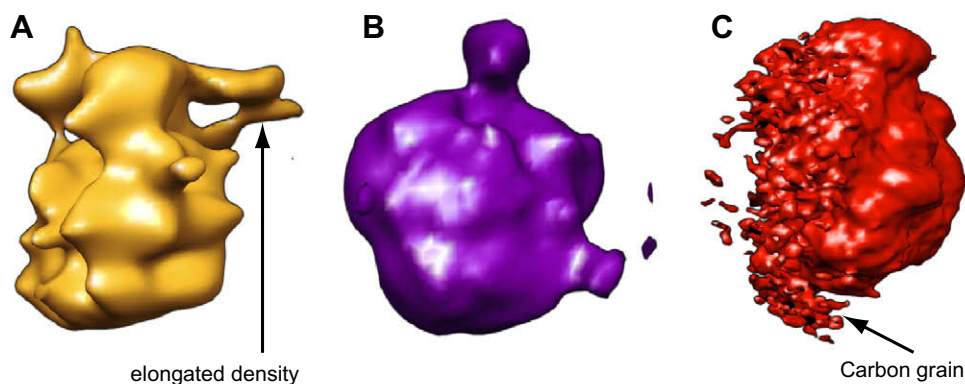


**Fig. 5.** Influence of particle alignment on a common lines Euler search. All parameters for refinement were identical to those used to generate the reconstructions in Fig. 3. Refined and iteratively aligned class averages created from the 5th iteration of angular reconstitution batch refinement were input into a cross common lines Euler search to obtain an initial model which readily converged to an accurate structure (top). Both initial model and final reconstruction are compared to those from Fig. 4F (bottom), which did not converge.

protuberance. These particular requirements need to be stressed, since Gaussian spheres and poorly reconstructed common lines models generate various structural aberrations within the final

volumes and did not converge to a single, defined structure (Fig. 4). The beginning of a tRNA cleft, though visible in the initial model, was arguably unnecessary for proper alignment and





**Fig. 6.** Undesired features found in the models generated by the various methods for *ab initio* 3-D reconstructions. (A) Side view of the 50S ribosomal subunit with elongated density is shown, created using automated 3d0 construction with angular reconstitution. (B) Frontal view showing the generally lower resolution of tomography, obtained using only 22 particles from an earlier attempt. (C) Side view of the RCT reconstruction showing the carbon grain upon which the ribosome lies in its preferred orientation on negatively-stained grids.

classification, as its details are not well pronounced. This result further illustrates the idea that it is the low-resolution information that dictates the accuracy of alignment and classification in a projection-matching algorithm. That said, convergence may not be attainable with a more challenging sample, particularly one smaller in size or exhibiting conformational heterogeneity.

#### 3.4. RCT

From 2-D analyses of the dataset, it became evident that the 50S subunit resides almost exclusively in a single orientation on the grid (Fig. 6C). While facilitating the 2-D alignment and classification, this also eliminates the ability to reduce the effects of the missing cone by creating multiple reconstructions from different class averages. The sole resulting model from the RCT dataset achieved a 28.2 Å resolution by FSC<sub>0.5</sub> (Fig. 3E, left). Despite the missing cone of information, it displays a well-defined tRNA cleft containing the protein transferase center (PTC), central protuberance (CP), and both L1 and L11 arms. In the presence of alternative orientations of the ribosomal subunit, the missing cone effect could be reduced, as this would allow averaging from multiple 3-D reconstructions. Nevertheless, the structural details provided all the necessary information to readily converge at a refined 3-D reconstruction.

All raw data (stacks and initial models) can be downloaded from <http://ami.scripps.edu/experiment>.

#### 4. Discussion

The advantages and disadvantages of each approach are illustrated in the resulting initial models obtained for the 50S large ribosomal subunit. Angular reconstitution is best suited to situations where particles are randomly orientated on the grid. We were not able to obtain a reliable initial model with this method when particles resided in highly preferred orientations (Fig. 4E, left). In general, unless 2-D projections are available in orientations relatively close to the three axial views, determining a correct C1 structure is difficult and potentially misleading. In our case, prior to its improvement with successive batch refinement, the initial 3d0 generated by angular reconstitution displayed missing features and extended density (Fig. 6A). In the case of tomography, an earlier initial model was of substantially lower resolution than provided by some of the other techniques (Fig. 6B). While this was partly remedied by simply increasing the number of particles and averaged sub-tomograms within the final volume (comparing

Fig. 3F, left, and Fig. 6B), the global nature of tomogram alignment and the missing cone will still likely limit the resolution that can be achieved. This effect is less pronounced in the RCT reconstructions, since the preferred orientation of the 50S subunit lying on the carbon support film provides a nearly ideal scenario for the technique. Densities corresponding to specific RNA domains were present (but not highly pronounced) in our *ab initio* RCT structure, appropriately matching to RNA regions in the PDB or EMDB models. The effect of the interaction of the particle with the carbon substrate is also clear in this reconstruction, as evidenced by the carbon grain corresponding to the support layer on which the ribosomal subunit rests in its preferred state (Fig. 6C). At a lower threshold, the carbon substrate is clearly defined as a plane of noise along the face of the crown view. Without multi-volume averaging, this presents a drawback and might lead to structural mischaracterization if the method is used alone. Finally, while we were not able to generate a reliable model using orthogonal tilt reconstruction for the data presented in this paper (the result of an exclusively preferred orientation on negatively-stained grids and the difficulty of collecting OTR data for vitrified samples), the method has also been incorporated into the Appion pipeline. In the absence of preferred orientation, OTR provides an attractive additional method for initial model generation, eliminating the missing cone and the need for multi-volume averaging.

Our final refined reconstructions shown in Fig. 3 share identical features and demonstrate that it is possible to converge to a single 3-D structure using several independent approaches. This data allows for one of several conclusions to be drawn. Either (1) the method for projection-matching refinement always converges to a single result regardless of the input initial model, or (2) the initial models are in fact telling a similar story. We found good reason to support the second conclusion, in particular from our observation of initial models that did not converge to the correct structure. Fig. 4 provides five such examples. In all cases the output volumes are neither consistent with each other, nor with the refinements from Fig. 3. Each contains chunks of improperly allocated density and displays various degrees of irregularities, some of which potentially lie in areas of local minimum values. While we acknowledge that many such structures are often salvageable through additional iterative refinement (especially with a robust sample like the 50S subunit), the conclusion remains unchanged – “good” initial models readily lead to a single, definitive result, whereas “bad” initial models do not. Not all models provide appropriate starting points for refinement, and even well-tested methods can sometimes produce unreliable structures.



The variable degree of correctness within each *ab initio* reconstruction raises the critical question of how to distinguish an accurate from an inaccurate representation of the object. A recent discussion of this issue argued that a direct reconstruction method such as RCT can be used to verify models calculated using angular reconstitution (Cheng et al., 2006). Provided that the 2-D alignment and classification necessary for RCT is correctly performed on a large enough number of particles, such an approach would often resolve the issue. From our observations of structural convergence, several other points are evident. For one, inaccuracy can be assessed by the presence of random features within the 3-D model, whose relative orientation changes from one reconstruction to the next. This was the case with non-converged reconstructions (Fig. 4) and is particularly relevant to common lines, where angular assignment depends fundamentally upon the quality of alignment and classification. Inaccuracies in the latter will almost certainly propagate errors into subsequent reconstructions (Fig. 5). On the other hand, accurate models should share similar low-resolution features. For the 50S subunit, these features are not overly complex, as evidenced through both tomographic initial models. For other macromolecules, this may not be the case, and it is the identification of such features that often presents a significant challenge to the microscopist.

Regardless of the specific approach used to distinguish the “good” from the “bad” initial model, a single concept underlies its general motivation – the use of independent techniques to assess structural agreement. Verifying any *ab initio* reconstruction via several independent methodologies should therefore be standard procedure, one that is greatly simplified with the availability of automated data collection and a streamlined toolbox of integrated analysis techniques. With this in mind, our own goal is to continue to add routines to the Appion pipeline and we encourage other groups to provide their own contributions. Appion is freely available under the Apache Open Source License, Version 2.0. Software can be downloaded from <http://www.appion.org>.

## Acknowledgments

We thank Dr. James Williamson and Anne E. Bunner (The Scripps Research Institute, La Jolla, CA) for providing purified 50S ribosomal subunits and Dr. Teresa Ruiz (University of Vermont, Burlington, VT) for thin carbon used in specimen preparation. We thank Dr. Andres Leschziner for technical support in implementing OTR within the Appion pipeline. We also thank Dr. William Young (The Scripps Research Institute, La Jolla, CA) for the computational support of Appion within the TSRI Garibaldi cluster. This project was primarily funded by grants from the National Institute of Health (NIH) through the National Center for Research Resources (NCRR) P41 program (Grants RR17573 and RR023093).

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jsb.2009.12.005](https://doi.org/10.1016/j.jsb.2009.12.005).

## References

- 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. *Journal of Structural Biology* 166, 126–132.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein data bank. *Nucleic Acids Research* 28, 235–242.
- Bunner, A.E., Trauger, S.A., Siuzdak, G., Williamson, J.R., 2008. Quantitative ESI-TOF analysis of macromolecular assembly kinetics. *Analytical Chemistry* 80, 9379–9386.
- Carragher, B., Smith, R., 2008. Software tools for molecular microscopy. *Journal of Structural Biology* 163, 224–228.
- Cheng, Y., Wolf, E., Larvie, M., Zak, O., Aisen, P., Grigorieff, N., Harrison, S.C., Walz, T., 2006. Single particle reconstructions of the transferrin-transferrin receptor complex obtained with different specimen preparation techniques. *Journal of Molecular Biology* 355, 1048–1065.
- Crowther, R.A., Klug, A., 1975. Structure analysis of macromolecular assemblies by image reconstruction from electron micrographs. *Annual Review of Biochemistry* 44, 161–182.
- Crowther, R.A., DeRosier, D.J., Klug, A., 1970. The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 317, 319–340.
- Frank, J., 1981. Three-dimensional reconstruction of single molecules. *Methods in Cell Biology* 22, 325–344.
- Frank, J., 2006. *Electron tomography – methods for three-dimensional visualization of structures in the cell*, second ed. Springer, New York, 445.
- Frank, J., van Heel, M., 1982. Correspondence analysis of aligned images of biological particles. *Journal of Molecular Biology* 161, 134–137.
- Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., Leith, A., 1996. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *Journal of Structural Biology* 116, 190–199.
- Heymann, J.B., Chagoyen, M., Belnap, D.M., 2005. Common conventions for interchange and archiving of three-dimensional electron microscopy information in structural biology. *Journal of Structural Biology* 151, 196–207.
- Kavran, J.M., Steitz, T.A., 2007. Structure of the base of the L7/L12 stalk of the *Haloarcula marismortui* large ribosomal subunit: analysis of L11 movements. *Journal of Molecular Biology* 371, 1047–1059.
- Klein, D.J., Schmeing, T.M., Moore, P.B., Steitz, T.A., 2001. The kink-turn: a new RNA secondary structure motif. *EMBO J.* 20, 4214–4221.
- Kremer, J.R., Mastronarde, D.N., McIntosh, J.R., 1996. Computer visualization of three-dimensional image data using IMOD. *Journal of Structural Biology* 116, 71–76.
- 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. *Journal of Structural Biology* 166, 95–102.
- Leschziner, A.E., 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. *Journal of Structural Biology* 153, 284–299.
- Ludtke, S.J., Baldwin, P.R., Chiu, W., 1999. EMAN: semiautomated software for high-resolution single-particle reconstructions. *Journal of Structural Biology* 128, 82–97.
- Matadeen, R., Patwardhan, A., Gowen, B., Orlova, E.V., Pape, T., Cuff, M., Mueller, F., Brimacombe, R., van Heel, M., 1999. The *Escherichia coli* large ribosomal subunit at 7.5 Å resolution. *Structure* 7, 1575–1583.
- Ohi, M., Li, Y., Cheng, Y., Walz, T., 2004. Negative staining and image classification – powerful tools in modern electron microscopy. *Biological Procedures Online* 6, 23–34.
- Pascual-Montano, A., Donate, L.E., Valle, M., Brcena, M., Pascual-Marqui, R.D., Carazo, J.M., 2001. A novel neural network technique for analysis and classification of EM single-particle images. *Journal of Structural Biology* 133, 233–245.
- Radermacher, M., Wagenknecht, T., Verschoor, A., Frank, J., 1986. A new 3-D reconstruction scheme applied to the 50S ribosomal subunit of *E. coli*. *Journal of Microscopy* 141, RP1–RP2.
- Scheres, S.H.W., Valle, M., Carazo, J.-M., 2005a. Fast maximum-likelihood refinement of electron microscopy images. *Bioinformatics* 21, ii243–ii244.
- Scheres, S.H.W., Valle, M., Nunez, R., Sorzano, C.O.S., Marabini, R., Herman, G.T., Carazo, J.-M., 2005b. Maximum-likelihood multi-reference refinement for electron microscopy images. *Journal of Molecular Biology* 348, 139–149.
- Sorzano, C.O.S., Marabini, R., Velazquez-Muriel, J., Bilbao-Castro, J.R., Scheres, S.H.W., Carazo, J.M., Pascual-Montano, A., 2004. XMIPP: a new generation of an open-source image processing package for electron microscopy. *Journal of Structural Biology* 148, 194–204.
- Sousa, D., Grigorieff, N., 2007. *Ab initio* resolution measurement for single particle structures. *Journal of Structural Biology* 157, 201–210.
- Stewart, A., Grigorieff, N., 2004. Noise bias in the refinement of structures derived from single particles. *Ultramicroscopy* 102, 67–84.
- Suloway, C., Pulokas, J., Fellmann, D., Cheng, A., Guerra, F., Quispe, J., Stagg, S., Potter, C.S., Carragher, B., 2005. Automated molecular microscopy: the new Legimon system. *Journal of Structural Biology* 151, 41–60.
- Suloway, C., Shi, J., Cheng, A., Pulokas, J., Carragher, B., Potter, C.S., Zheng, S.Q., Agard, D.A., Jensen, G.J., 2009. Fully automated, sequential tilt-series acquisition with Legimon. *Journal of Structural Biology* 167, 11–18.
- Tagari, M., Newman, R., Chagoyen, M., Carazo, J.-M., Henrick, K., 2002. New electron microscopy database and deposition system. *Trends in Biochemical Sciences* 27, 589.
- 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, 111–124.
- van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., Schatz, M., 1996. A new generation of the IMAGIC image processing system. *Journal of Structural Biology* 116, 17–24.
- van Heel, M., Gowen, B., Matadeen, R., Orlova, E.V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R., Schatz, M., Patwardhan, A., 2000. Single-particle electron cryo-

- microscopy: towards atomic resolution. *Quarterly Reviews of Biophysics* 33, 307–369.
- Voss, N.R., Yoshioka, C.K., Radermacher, M., Potter, C.S., Carragher, B., 2009. DoG Picker and TiltPicker: Software tools to facilitate particle selection in single particle electron microscopy. *Journal of Structural Biology* 166, 205–213.
- Winkler, H., Taylor, K.A., 2006. Accurate marker-free alignment with simultaneous geometry determination and reconstruction of tilt series in electron tomography. *Ultramicroscopy* 106, 240–254.
- Yoshioka, C., Pulokas, J., Fellmann, D., Potter, C.S., Milligan, R.A., Carragher, B., 2007. Automation of random conical tilt and orthogonal tilt data collection using feature-based correlation. *Journal of Structural Biology* 159, 335–346.