Minireview

Single particle macromolecular structure determination via electron microscopy

Pamela A. Thuman-Commike*

QED Labs, 1190 S. Bascom Ave., Suite 208, San Jose, CA 95128, USA

Received 25 June 2001; revised 27 July 2001; accepted 27 July 2001

First published online 22 August 2001

Edited by Amy M. McGough

Abstract Three-dimensional structure determination of macromolecules and macromolecular complexes is an integral part of understanding biological functions. For large protein and macromolecular complexes structure determination is often performed using electron cryomicroscopy where projection images of individual macromolecular complexes are combined to produce a three-dimensional reconstruction. Single particle methods have been devised to perform this structure determination for macromolecular complexes with little or no underlying symmetry. These computational methods generally involve an iterative process of aligning unique views of the macromolecular images followed by determination of the angular components that define those views. In this review, this structure determination process is described with the aim of clarifying a seemingly complex structural method. © 2001 Federation of European Biochemical Societies. Published by Elsevier Science B.V. All rights reserved.

Key words: Computer image processing; Electron microscopy; Macromolecular complex; Single particle; Three-dimensional structure

1. Introduction

Structural characterization of macromolecules and macromolecular complexes is a useful component in understanding complex biological processes. Electron microscopy is increasingly used to perform three-dimensional structural studies on a wide range of macromolecules [1–3]. Depending on the underlying symmetry of the macromolecule being studied, a variety of data processing methods can be performed including those aimed at objects forming helical filaments [4,5], icosahedrons [6,7], two-dimensional crystals [8,9], and single particles with little or no symmetry [10,11].

Single particle techniques have shown a dramatic increase in usage over the past decade including the recent determination of asymmetric ribosomal subunits to 7.5 and 11 Å resolution [12,13]. A large range of macromolecules and macromolecular complexes are amenable to single particle structural studies due to limited requirements in underlying symmetry and macromolecular size. Generally, macromolecules larger than 500 kDa [14] can be studied with these techniques. Fur-

*Fax: (1)-408-295 8348.

thermore, although the absence of symmetry introduces some difficulties in data processing, even macromolecules that are asymmetric can be studied using this method if all macromolecules are in the same conformation. These three-dimensional structural studies of macromolecules and macromolecular complexes provide a variety of structural information including morphological characterization, comparisons of conformational changes between various states, and characterizations of components in cellular pathways. Recent single particle structural studies have provided insight into many biological processes including DNA break repair [15], muscle contraction [16,17], nucleic acid metabolism [18], nucleocytoplasmic transport [19,20], protein folding [21], and protein synthesis/RNA translocation [12,13]. This review provides an introduction to these structural methods with the aim of making these methods understandable to research scientists that may not otherwise undertake such studies.

2. Single particle structure determination

Single particle structure determination typically begins with preparation of a homogeneous sample that is placed on an electron microscope grid and prepared for imaging. In electron cryomicroscopy, the grid is rapidly frozen in liquid ethane to preserve the three-dimensional structure in vitreous ice while providing the contrast necessary for imaging [22]. Alternatively, for low resolution or preliminary investigations where sample quality and suitability are unknown the sample may be embedded in a stain such as uranyl acetate prior to imaging. Following imaging, the data are assessed for quality by visual examination of the micrographs. During this examination the micrographs are screened for imaging defects and to ensure the imaging conditions are suitable to obtain the desired resolution of the structural study. Acceptable micrographs are digitized and subjected to further computational quantitative assessment to verify data quality [23].

Once images have been obtained, scanned, and quality assessed, computer image processing begins. The ultimate goal of this data processing is to determine the three-dimensional structure of the object. To achieve this goal the relative orientation of each particle must be determined with respect to a reference orientation of the original three-dimensional object. Fig. 1 illustrates the main steps performed during data processing and Fig. 2 provides a schematic of the iterative data processing cycle performed during single particle structure determination.

E-mail address: pam@qedlabs.com (P.A. Thuman-Commike).



Fig. 1. Depiction of single particle structure determination. A: Cartoon representation of a sample micrograph with three unique views of the object under study. B: Nine selected particle images from the micrograph shown in A. The particle images were carefully selected so the center of each object is near the center of the image. C: Aligned particle images. Each of the nine particle images now positioned to a common view. D: Classified particle images. The three unique views that were present in the original micrograph have resulted in three particle image classes. E: Determined orientations. In this cartoon example, the orientations correspond to the side, front, and top of the object. In a real example classes. If angular reconstitution was used then averages of all of the images in each class determined in D would be used for orientation determination. Alternatively, if the single axis tilt method was used then each class would result in a different reconstruction using the corresponding particles in the tilted micrograph. F: Three-dimensional reconstruction. Combination of the oriented images to determine a single three-dimensional reconstruction. Note that this example depicts a cartoon representation to facilitate understanding, actual electron microscopy images are projection images rather than simple outlines of the object under study and require a large number of unique views. G: Sample of a real electron cryomicroscopy image. Shown is a digital micrograph of DNA-dependent protein kinase [15] with the inset depicting one individual selected particle image. Figure reproduced from [15] with permission of primary author and publisher.



Fig. 2. Schematic of iterative data processing cycle for macromolecular structure determination. using electron microscopy. Given suitable images, an iterative data processing cycle is performed aimed at determining the orientation of individual particles as they lie on the electron microscope grid during imaging. The process is iterative in that once a preliminary structure is determined the process repeats until a final high quality structure at the desired resolution is obtained. During each iteration of the data processing cycle, the structure obtained in the previous cycle is used to assist in the orientation determination. Data processing completes when the structure is determined to a resolution that allows the underlying biological questions to be addressed and structural analysis begins. An initial structure is determined using the pathway outlined by the solid lines. Given the initial structure the pathways outlined by the dashed lines can then be used in combination with the solid pathways to improve the quality and resolution of the reconstruction by determining more orientations and by improving the accuracy of previously determined orientations.

Briefly, orientation determination proceeds via one of three possible methods. Although the details of the methods differ each follows a similar data processing sequence. First, particle selection is performed to identify individual particles within the micrograph. Particle alignment is then performed to modify the rotational and translational position of the particles within the image such that all particles are positioned similarly within their individual images. Next, the particles are classified to identify groups of particles that represent the same views of the macromolecule. The most appropriate orientation determination method is then applied to the preprocessed images and once orientations have been determined for a suitable number of particles three-dimensional reconstruction is performed. Initially, the three-dimensional reconstruction is a rough estimate of the underlying macromolecular structure. This estimate is then used to improve the previously determined particle orientations and to determine additional particle orientations. Following each iteration of alignment, classification, orientation determination, and reconstruction the resolution of the structure is assessed and once the resolution is suitable to address the biological question then data

processing is complete. The following sections detail the specifics of these data processing steps.

Note that some of the computations associated with this data processing are performed in Fourier space and use the Fourier transform. Electron microscopy images depict the density of the imaged object. These densities are represented by pixels in the digitized image in what is known as image space or real space. Fourier space is an alternative mathematical representation of an image that describes the image using phases and amplitudes rather than the densities associated with the real space pixels of an electron microscopy image. A Fourier transform is the mechanism by which the Fourier representation of an image is calculated. An inverse Fourier transform reverses the transformation converting the Fourier image back to real space pixels.

2.1. Particle selection

In order to determine the orientation of individual particles data processing must be performed on each particle as a separate image rather than working on the entire micrograph (Fig. 1A). Thus, the first data processing task is to specify the position of each individual particle within the micrograph in a process known as particle selection (Fig. 1B). Particle selection can be performed using manual or semi-automated computer programs. In manual particle selection, the micrograph is displayed on the screen and the user interactively denotes the location of the particles with the mouse. Semiautomated particle selection facilitates this process in that the software attempts to determine the particle locations rather than requiring the locations be manually specified. Because electron microscopy images are noisy (Fig. 1G), semi-automated particle selection typically requires user interaction to remove any selected areas that do not correspond to particles.

2.2. Particle alignment

Particle alignment is the process of repositioning each particle within its image such that after alignment all particles are positioned similarly (Fig. 1C). Specifically, particle alignment involves locating the correct in-plane rotation and x-/y-translation that places all particles in the same relative position. The goal of this step is to determine these rotations and translations accurately so that when the images are all viewed together key structural features appear in the same position.

Alignment is complicated by the low contrast and noise present in electron microscopy images. Several methods have been devised to overcome these obstacles. The simplest approach compares a reference against each particle image forcing the particle into a position that closely resembles the reference [10]. Unfortunately, because of the noise present in electron microscopy images this approach can result in a bias towards the appearance of the reference image. In fact, comparisons with random noise and a reference image have been shown to yield an alignment that resembles the reference image [11].

To overcome such biases, variations of the basic reference based alignment are used. These methods align the particles using multiple rounds of comparisons with differing references to reduce the bias. Two such methods are commonly used: the reference-free alignment method [24] and the multi-reference alignment method [25]. Unless a previous structure is known these methods are preferable to basic reference based alignment because they are less likely to bias the data. No matter what alignment method is used, however, it is important to remember the hazards of biasing the data by the reference. For unknown structures, ensuring that the correct structure has been obtained can require processing more than one independent data set to verify all data sets produce the same structural result.

2.3. Particle classification

Particle classification is the task of identifying similar views of the same object or different objects into classes (Fig. 1D). Classification allows segregation of different particle types and particle orientations within a micrograph and provides a means to reduce noise by averaging particle images with similar views. Classification proceeds by identification of common features in the data. Multivariate statistical analysis (MSA) is used to perform this step by compressing the large quantities of image data present in electron microscopy images into eigen images that combine common image features [26,27]. Given the eigen images, common features are identified by an operation called cluster analysis. See [28,29] for details on the mathematical basis of the MSA method, eigen image analysis, and cluster analysis.

Two important issues in classification are the accuracy of image alignment prior to classification and the specified number of classes to be identified during cluster analysis. Accurate classification requires that the common features in a set of particles be identified. Thus, alignment increases the accuracy of classification by positioning each particle similarly. Alignment, however, can be difficult without references generated from a previous structure. Consequently, the first round of alignment and classification may not be highly accurate but as data processing progresses subsequent alignment improvements result in more accurate classification. Similarly, it can be difficult to estimate the number of unique views present in the data. To overcome this, a small number of classes (i.e. views) are initially assumed and as structure determination progresses the number is increased based on the statistical results of classification and the quality of the identified classes.

2.4. Orientation determination

Depending on the available data, single particle orientation determination (Fig. 1E) can be performed by following one of three general approaches: single axis tilt [30,31], angular reconstitution [32], and projection matching [33]. The single axis tilt method requires that the particles in the imaged data be positioned in a single (or very small number) of preferred orientations. Angular reconstitution places no restriction on the type of data available as it determines orientations using properties of the projection images. Last, projection matching requires a previously obtained three-dimensional model of the object under study and uses comparisons with that model to determine particle orientations. The following sections describe each of these orientation determination methods in detail.

2.4.1. Single axis tilt. The single axis tilt method [30,31] is used when the macromolecule displays a single or small number of preferred orientations during imaging. When this occurs, images are collected in pairs. In the first image the microscope stage is not tilted, while in the second image the microscope stage has a relatively high tilt. Because the particles all have a preferred orientation, the 0° tilt image will show similar views of the particles while the highly tilted image provides a distribution of orientations. Data processing aligns the particles in the 0° tilt image to a common position. Next, to ensure that the particles do not form more than one preferred orientation and to separate any particles not in that preferred orientation classification is typically performed. Last, following alignment and the optional classification, the tilt angle, applied during imaging, and the translational and rotational movements, determined during alignment, are applied to the corresponding tilted particle images to determine their relative orientations.

2.4.2. Angular reconstitution. The angular reconstitution orientation determination method [32] relies on the central projection theorem to determine orientations. The central projection theorem [34] states that the center of the Fourier transforms of any two-dimensional projection images of the same object intersect at the center of three-dimensional Fourier space. Restated, this theorem states that Fourier transforms of projection images, which are planes in Fourier space. From

geometry we know that when two planes intersect they form a line. Consequently, the central projection theorem provides that pairs of electron microscopy projection images of the same object have overlapping pieces (i.e. common lines) of data. Application of this theorem to single particle structure determination implies that if the overlapping data present in two images can be located then the corresponding relative orientations between those two particles can be computed. Because the amount of repeated data within sets of images is small and the electron microscopy images contain noise, image alignment and classification are performed on the data to facilitate orientation determination. The particles in the resulting classes are then averaged together producing a composite image with less noise than the individual particle images. It is these class averaged images for which orientations are determined by searching for the overlapping data as described in [32].

2.4.3. Projection matching. Projection matching [33] can be applied to any set of particles for which there exists a previously determined structure or model three-dimensional structure of the macromolecule. This approach is used either when a previously determined model of the structure under study exists or during later iterations of orientation determination using the reconstruction obtained in the previous round of orientation determination. The basic concept of projection matching is to compare projection images computed from the previously obtained structure, for which the orientations are known, with the actual electron microscopy images. Generally, the comparisons use correlation between the computed projection image and the particle image although a variety of such comparisons have been applied [33,35].

2.5. Three-dimensional reconstruction

Following orientation determination, all of the individual particle image data are combined in a computational task known as three-dimensional reconstruction that computes the three-dimensional structure of the object (Fig. 1F). This task is generally performed with a method known as weighted back-projection [36]. Weighted back-projection is a process similar to the inverse Fourier transform. A two-dimensional inverse Fourier transform determines the real space two-dimensional image from a two-dimensional Fourier space image. A weighted back-projection determines the real space three-dimensional density from a series of two-dimensional Fourier space images using a series of transformations. Reconstruction is performed to a specified resolution as determined by the number of unique particle image views and imaging conditions. For moderate resolution structures (20-35 Å resolution) no additional processing is required but for higher resolution structures (beyond ~ 15 Å resolution) corrections are generally applied to the reconstruction to account for degradations applied during imaging. These correction methods are applied in Fourier space using knowledge of the imaging conditions [37,38].

2.6. Iterative data processing cycle

Although each of the orientation determination approaches described have different data requirements and use different data processing schemes they all follow the common data processing pathway shown in Fig. 2. This is an iterative process aimed at improving the three-dimensional reconstruction with each round of data processing. The alignment preprocessing step can be easily biased by the type of reference image used, classification is influenced by the accuracy of the alignment, orientation determination relies on proper classification, and reconstruction requires a large number of unique and accurately determined orientations. Thus, iteration is important to improve the accuracy of each of these stages. In addition, the data must be examined at each step of data processing to ensure cumulative errors are not propagated. Details on the mechanisms by which some of these potential problems are observed and addressed can be found in the articles describing the theory of these methods such as [24,25,28–33,36]. Iteration continues until the structure is accurately determined to the desired resolution. Accuracy can be measured using a number of statistical safeguards output by each individual computational step. The resolution itself depends on the parameters used during imaging, the number of unique views, and the correctness of the determined orientations. The number of unique evenly distributed views depends on the size of the object, the desired resolution, and the amount of noise present. Theoretically, determination of a structure D Å in diameter at a resolution d requires a minimum of D/d images [39]. In practice, noise and imaging conditions greatly increase this number. Typical moderate resolution (20–35 Å) single particle structures use thousands of individual particle images and higher resolution structures use well over 10000 individual particle images in the final structure. Resolution assessment ideally compares independently determined data sets that have not been compared to one another at any point in the data processing [40]. Such independent comparisons reduce potential bias that may have occurred in the alignment and classification stages while allowing quantitative assessment of resolution similarity.

3. Structural analysis

Given a three-dimensional reconstruction of the desired resolution the task of structural interpretation begins. Structural analysis is aimed at using the structure to answer specific biological questions. The specific actions performed in this analysis depend on the structure being analyzed and the questions being addressed. Visualization of the three-dimensional structure is most commonly performed by viewing a shaded surface representation of the structure allowing overall morphology to be readily observed (Fig. 3A, top row). Depending on the visualization software package used the displayed structure can be rotated, zoomed, shaded, color coded, and dissected into pieces (Fig. 3A, middle row). Difference maps are another form of structural analysis (Fig. 3B). The most common usage of difference maps is to compare the structures of two different biochemical states of the same macromolecular complex. Computational dissection of a structure into smaller pieces or sub-components is a structure analysis tool that can be used to highlight or investigate a particular aspect of a structure (Fig. 3A, bottom row). Last, structural analysis often includes plots and/or generation of two-dimensional images formatted to describe specific aspects of a complex structure.

4. Summary

The electron microscopy single particle structure determination methods described in this review have been developed by



Fig. 3. Structural analysis via three-dimensional visualization. A: Shown is a structural visualization. of the bacteriophage SPP1 portal protein [46]. The top row shows the surface representation of the object from two different views. In this type of visualization, the displayed surface is determined by calculating the appropriate mass of the structure given a threshold value [47]. The middle row shows additional views computationally manipulated to reveals areas otherwise hidden from view. The bottom row shows a computationally extracted component of the structure. Notice that detailed annotations and other visual pointers greatly increase the ability to understand the key concepts the structure reveals. Figure reproduced from [46] with permission of primary author and publisher. B: Shown are two stereo views of the 30S ribosome subunit bound to translational initiation factor IF3 with positive (magenta) and negative (blue) differences identified [48]. The stereo representation allows more accurate viewing of three-dimensional features than a single non-stereo view. The left and right views provide a surface representation of the ribosome and the associated difference map features identifying the location of IF3. A difference map between an object A and another object B is obtained by subtracting one object from another. Positive differences, obtained by subtracting B from A, show features that are present in A but not in B, while negative differences so that new features present in the object as well as missing features can be located. Figure reproduced with permission of primary author and publisher from [48]. Copyright (1999) National Academy of Sciences, USA.

a number of researchers over the past 30 years. Software packages specifically designed for single particle structure determination that are commonly used by the structural biology community are IMAGIC [41] and SPIDER [42]. In addition, SUPRIM [43] and EM [44] are general purpose electron microscopy software packages that can be used for many or all of the data processing steps necessary for single particle structure determination, and EMAN has recently been developed [45]. The broad range of macromolecules and macromolecular complexes that can be studied using the single particle method is driving the need for general biomedical research scientists to perform structural studies. The existing software packages fo-

cus primarily on the needs of structural biologists and electron microscopists rather than biomedical scientists unfamiliar with the structural methods. That is, they provide the methods and assume the user has detailed knowledge available on how to use those methods or access to a structural biologist who has previously used the software package. Given the many potential pitfalls, biases, and difficulties in single particle structure determination, further advancement of this structural method, to a more general population of biomedical research scientists, requires that single particle software packages be aimed at allowing non-structural biologists to determine single particle structures on their own. Acknowledgements: I thank Dr. Justine Malinski and Mr. Alan Commike for critically reviewing the manuscript.

References

- Stowell, M., Miyazawa, A. and Unwin, N. (1998) Curr. Opin. Struct. Biol. 8, 595–600.
- [2] Kühlbrandt, W. and Williams, K. (1999) Curr. Opin. Chem. Biol. 3, 537–543.
- [3] Tao, Y. and Zhang, W. (2000) Curr. Opin. Struct. Biol. 10, 616– 622.
- [4] DeRosier, D.J. and Moore, P.B. (1970) J. Mol. Biol. 52, 355-369.
- [5] Stewart, M. (1988) J. Electron Microsc. Tech. 9, 325–358.
- [6] Crowther, R.A. (1971) Phil. Trans. R. Soc. Lond. B. 261, 221– 230.
- [7] Thuman-Commike, P.A. and Chiu, W. (2000) Micron 31, 687– 711.
- [8] Unwin, P.N.T. and Henderson, R. (1975) J. Mol. Biol. 94, 425– 440.
- [9] Amos, L.A., Henderson, R. and Unwin, P.N.T. (1982) Prog. Biophys. Mol. Biol. 39, 183–231.
- [10] Frank, J. and Radermacher, M. (1992) Ultramicroscopy 46, 241– 262.
- [11] van Heel, M., Winkler, H., Orlova, E. and Schatz, M. (1992) Scann. Microsc. 6 (Suppl.), 23–42.
- [12] Matadeen, R. et al. (1999) Structure 7, 1575–1583.
- [13] Gabashvili, I., Agrawal, R., Spahn, C., Grassucci, R., Svergun, D., Frank, J. and Penczek, P. (2000) Cell 100, 537–549.
- [14] Henderson, R. (1995) Q. Rev. Biophys. 28, 171-193.
- [15] Chiu, C., Cary, R., Chen, D., Peterson, S. and Stewart, P. (1998)
 J. Mol. Biol. 284, 1075–1081.
- [16] Wagenknecht, T., Grassucci, R., Frank, J., Saito, A., Inui, M. and Fleischer, S. (1989) Nature 338, 167–170.
- [17] Orlova, E.V., Serysheva, I.I., van Heel, M., Hamilton, S.L. and Chiu, W. (1996) Nature Struct. Biol. 3, 547–552.
- [18] San Martin, C., Radermacher, M., Wolpensinger, B., Engel, A., Miles, C., Dixon, N. and Car-azo, J.M. (1998) Structure 6, 501– 509.
- [19] Akey, C. and Radermacher, M. (1993) J. Cell Biol. 122, 1-19.
- [20] Hinshaw, J.E., Carragher, B.O. and Milligan, R.A. (1992) Cell 69, 1133–1141.
- [21] Roseman, A.M., Chen, S., White, H., Braig, K. and Saibil, H.R. (1996) Cell 87, 241–251.
- [22] Dubochet, J., Lepault, J., Freeman, R., Berriman, J.A. and Homo, J.C. (1982) J. Microsc. 128, 219–237.

- [23] Zhou, Z.H., Hardt, S., Wang, B., Sherman, M.B., Jakana, J. and Chiu, W. (1996) J. Struct. Biol. 116, 216–222.
- [24] Penczek, P., Radermacher, M. and Frank, J. (1992) Ultramicroscopy 40, 33–53.
- [25] van Heel, M. and Stöffler-Meilicke, M. (1985) EMBO J. 4, 2389– 2395.
- [26] van Heel, M. and Frank, J. (1981) Ultramicroscopy 6, 187– 194.
- [27] Frank, J. and van Heel, M. (1982) J. Mol. Biol. 161, 134-137.
- [28] van Heel, M. (1989) Optik 82, 114–126.
- [29] Frank, J. (1990) Q. Rev. Biophys. 23, 281-329.
- [30] Radermacher, M., Wagenknecht, T., Verschoor, A. and Frank, J. (1987) J. Microsc. 146, 113–136.
- [31] Radermacher, M. (1988) J. Electron Microsc. Tech. 9, 359-394.
- [32] van Heel, M. (1987) Ultramicroscopy 21, 111-124.
- [33] Penczek, P.A., Grassucci, R.A. and Frank, J. (1994) Ultramicroscopy 53, 251–270.
- [34] DeRosier, D.J. and Klug, A. (1968) Nature 217, 130–134.
- [35] Baker, T.S. and Cheng, H. (1996) J. Struct. Biol. 116, 120-130.
- [36] Radermacher, M. (1992) in: Electron Tomography: 3D Imaging with the Transmission Electron Microscope (Frank, J., Ed.), pp. 91–115, Plenum Press, New York.
- [37] Kübler, O., Hahn, M. and Seredynski, J. (1978) Optik 51, 171– 188.
- [38] Frank, J. and Penczek, P. (1995) Optik 98, 125-129.
- [39] Crowther, R.A., DeRosier, D.J. and Klug, A. (1970) Proc. R. Soc. Lond. 317, 319–340.
- [40] Grigorieff, N. (2000) Acta Crystallogr. D 56, 1270-1277.
- [41] van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R. and Schatz, M. (1996) J. Struct. Biol. 116, 17–24.
- [42] Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M. and Leith, A. (1996) J. Struct. Biol. 116, 190–199.
- [43] Schroeter, J. and Bretaudier, J. (1996) J. Struct. Biol. 116, 131– 137.
- [44] Hegerl, R. (1996) J. Struct. Biol. 116, 30-34.
- [45] Ludtke, S., Baldwin, P. and Chiu, W. (1999) J. Struct. Biol. 128, 82–97.
- [46] Orlova, E., Dube, P., Beckmann, E., Zemlin, F., Lurz, R., Trautner, T., Tavares, P. and van Heel, M. (1999) Nature Struct. Biol. 6, 842–846.
- [47] Conway, J.F., Trus, B.L., Booy, F.P., Newcomb, W.W., Brown, J.C. and Steven, A.C. (1996) J. Struct. Biol. 116, 200–208.
- [48] McCutcheon, J., Agrawal, R., Philips, S., Grassucci, R., Gerchman, S., Clemons, W., Ramakrishnan, V. and Frank, J. (1999) Proc. Natl. Acad. Sci. USA 96, 4301–4306.