Computational and Structural Biotechnology Journal 14 (2016) 385–390

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

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

Cryo-electron Microscopy Analysis of Structurally Heterogeneous Macromolecular Complexes

Slavica Jonić *

IMPMC, Sorbonne Universités - CNRS UMR 7590, UPMC Univ Paris 6, MNHN, IRD UMR 206, 75005, Paris, France

ARTICLE INFO

Article history:
Received 9 August 2016
Received in revised form 4 October 2016
Accepted 11 October 2016
Available online 14 October 2016

Keywords:
Cryo-electron microscopy
Single particle analysis
Macromolecular complexes
Heterogeneity
Structure
Flexibility
Conformational changes
Dynamics

ABSTRACT

Cryo-electron microscopy (cryo-EM) has for a long time been a technique of choice for determining structure of large and flexible macromolecular complexes that were difficult to study by other experimental techniques such as X-ray crystallography or nuclear magnetic resonance. However, a fast development of instruments and software for cryo-EM in the last decade has allowed that a large range of complexes can be studied by cryo-EM, and that their structures can be obtained at near-atomic resolution, including the structures of small complexes (e.g., membrane proteins) whose size was earlier an obstacle to cryo-EM. Image analysis to identify multiple coexisting structures in the same specimen (multiconformation reconstruction) is now routinely done both to solve structures at near-atomic resolution and to study conformational dynamics. Methods for multiconformation reconstruction and latest examples of their applications are the focus of this review.

1. Introduction

Recent instrumental and methodological developments for cryo-electron microscopy (cryo-EM) [1–19] made that the structures of macromolecular complexes are now often determined at subnanometer and near-atomic resolutions [20–41]. The most exciting results in terms of resolution and size of solved structures are currently being obtained with the latest-generation cryo-electron microscopes equipped with direct electron detectors (registering electrons directly rather than via a scintillator and recording movies allowing for correction of the specimen motion occurring during beam exposure) and software for automated collection of images, in combination with the use of advanced image analysis methods and high performance computing platforms [42–44].

First structures at near-atomic resolution were obtained for large complexes with high symmetry such as icosahedral-symmetry viruses [20,21]. However, several works have recently shown that cryo-EM can be used for near-atomic resolution of structures of small complexes (170–500 kDa) with low symmetry [22,27,41] or no symmetry [26,32], where the best resolution (1.8 Å) was obtained for 334 kDa glutamate dehydrogenase [40]. Bartesaghi and collaborators have pointed out that, rather than imaging technologies or image-processing methods, the major bottleneck to a routine cryo-EM determination of structures at resolutions close to 2 Å is currently the preparation of specimens of adequate quality that takes into account intrinsic protein flexibility [27]. Regarding larger complexes, subnanometer resolution is currently often achieved [24,25,28,30,36] and near-atomic resolution is becoming more and more frequent [23,29,31,33–35,37–39].

Three-dimensional (3D) reconstruction from heterogeneous sets of images normally results in low-resolution density maps. Thus, data heterogeneity analysis to isolate images of complexes of similar molecular compositions and conformations is a usual prerequisite to structural determination at high resolution. Biochemical procedures can usually be optimized so that the majority of complexes in the specimen, if not all of them, have the same molecular composition. However, the same composition rarely means the same conformation, due to the flexibility of complexes. Thus, conformational heterogeneity of specimens is usually analyzed by image analysis and classification methods. The reconstruction of different coexisting structures from the same sample will here be referred to as multiconformation reconstruction. It involves a classification strategy that assigns the particles having similar structures (similar molecular compositions and similar conformations) to the same class of particles. Multiconformation reconstruction is used to obtain high-resolution structures and provides insights into conformational dynamics of macromolecular complexes. Multiconformation reconstruction methods will be reviewed here together with the latest examples of their applications.

* IMPMC-UMR 7590, Université Pierre & Marie Curie, Case Courrier 115, 4 Place Jussieu, 75005 Paris, France.
E-mail address: Slavica.Jonic@impmc.upmc.fr.

http://dx.doi.org/10.1016/j.csbj.2016.10.002

Image classification in multiconformation reconstruction can be supervised or unsupervised. Supervised classification aims at sorting images into classes based on information on expected conformational states (prior knowledge about the distribution of states), and its use is limited to studying systems where this prior information is available. The majority of methods reviewed here belong to the group of unsupervised classification methods whose aim is to find actual conformational states without a prior knowledge about the distribution of states. Due to computational limitations, the majority of available multiconformation reconstruction methods assume specimens with relatively few different conformations of complexes (usually, less than 10) and restrained compositional heterogeneity. They also take into account that biochemical preparation of specimens is usually optimized to reduce the number of different structures coexisting in the same specimen. These methods are sometimes referred to as discrete conformational heterogeneity methods. They differ in the number of required initial 3D models (0, 1, or several) (Fig. 1), but a more important difference among these methods is whether they analyze heterogeneity at image level (in a 2D space) or at volume level (in a 3D space). Thus, these methods will here be grouped in two groups (2D and 3D heterogeneity analysis methods) and reviewed in two separate sections.

Development of methods for analyzing quasicontinuums of conformational states is an active field of research that will here be only briefly discussed (Outlook section). These methods will be fully reviewed in a separate publication.

2. 2D Heterogeneity Analysis Methods

In this section are reviewed methods that perform 3D reconstruction of different structures identified by analyzing structural heterogeneity at the level of images. Some of these methods use 3D starting models to determine the orientation of images while the other methods, referred to as ab initio methods, use no prior structural information.

2.1. Orientation Determination Without a Starting 3D Model

The orientation of images can be determined based on the central section theorem [45]. This theorem states that the Fourier transform of a 2D projection is a plane intersecting the origin of the 3D object's Fourier transform and that this plane is parallel to the projection plane [45,46]. Any two non-parallel 2D projections of the same 3D object will therefore share a common line in Fourier space. Thus, the orientation of images can be determined by determining the relative orientation of common lines between the 2D Fourier transforms of images [47,48]. The 3D model of the object obtained using images and the determined orientation is referred to as ab initio 3D model.

If the given set of images is heterogeneous, the images have to be sorted into structurally homogeneous subsets (image sorting) and 3D geometrical relationships among the images have to be determined (image orienting). When using no prior 3D model, image sorting and orienting can be performed in two separate steps or simultaneously. In the two-step approach proposed in [49], image orienting is preceded by a classification of images in classes of similar orientations (orientation classes) and a classification of each orientation class in classes of similar structures (image sorting), and both classifications are based on 2D multivariate statistical analysis (MSA) [50,51]. This approach, here referred to as nonsimultaneous sorting and orienting, has been particularly efficient in separating small and large particle images or images of ligand bound and unbound complexes [49,52,53]. In the approach for simultaneous sorting and orienting proposed in [54], all 6 parameters (3 Euler angles, 2 shifts, and structure assignment) are considered simultaneously for all images by solving a multidimensional optimization problem and common line correlations in Fourier space [54]. The larger the expected number of different structures, the more complex is the optimization problem to solve. So far, this approach was only used to separate two conformational states, such as open/closed and ligand bound/unbound states [54,55].

The main problem with the methods in this group is their low robustness to noise. They are thus usually used with 2D average images that have a higher signal-to-noise ratio (SNR) than individual images [53,55]. Also, their applications in studies with more than two conformational states have not yet been demonstrated.

2.2. Orientation Determination Using a Starting 3D Model

Methods in this group aim at facilitating recognition of structural variability by minimizing orientational variability. They assume that dissimilarities between images corresponding to different structures are larger than those between images corresponding to the same structure but having slightly different angular directions.

The orientational variability is minimized by determining the orientation of images with respect to a preliminary 3D model that is usually obtained by combining images from the entire heterogeneous data set. Images assigned to the same projection direction are then sorted in clusters by analyzing discrepancies between common lines [56–58] or between entire images or their regions [59–62]. Clusters in each projection direction are labeled (different structures are assigned to different clusters) and those with the same label in different projection directions are combined in the same 3D reconstruction. Cluster labeling is a difficult task and the labeling approaches are usually not trivial. For instance, in [57], distinct cluster averages corresponding to a selected view (the view selected visually as the view showing the highest variability) and presumably representing different conformers are used as conformational references for the conformational assignment of cluster averages in all other orientations based on the highest cross-correlation of common lines between the cluster averages and the conformational references. On the contrary, the approach proposed in [58] considers all cluster averages simultaneously instead of selecting a single representative view and defining conformational references, by computing all pairwise similarities between the cluster averages based on cross-correlation of common lines.

The preliminary 3D model should have good quality and a potential model bias should be considered. In Fig. 1, these methods are referred to as 2D variance analysis methods to be distinguished from the 3D variance analysis methods that also use an initial 3D model to orient images but analyze heterogeneity at volume level (classification based on 3D variance analysis that is described below).

The 2D heterogeneity analysis methods have also been used with globally homogenous data sets to select the most self-consistent subset of particles for a high-resolution 3D reconstruction. For instance, a procedure involving MSA-based classification of images, ab initio reconstruction, and iterative refinement has recently resulted in the first subnanometer-resolution structures of the complete portal-phage tail interface that mimic the states before and after DNA release during phage infection [63].

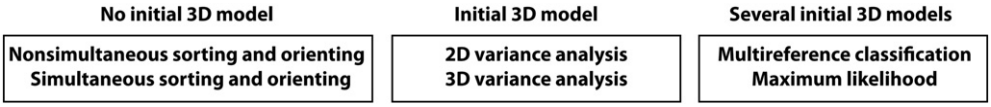

Fig. 1. Multiconformation reconstruction methods reviewed here.