Available online at www.sciencedirect.com
ScienceDirect
Journal of Structural Biology 156 (2006) 432–441
Journal of Structural Biology
www.elsevier.com/locate/yjsbi

Automated segmentation of molecular subunits in electron cryomicroscopy density maps

Matthew L. Baker [a,*], Zeyun Yu [b], Wah Chiu [a], Chandrajit Bajaj [b]

[a] National Center for Macromolecular Imaging, Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX 77030, USA
[b] Center of Computational Visualization, Department of Computer Sciences and Institute of Computational Engineering and Sciences, The University of Texas at Austin, Austin, TX 78712, USA

Received 23 March 2006; received in revised form 25 May 2006; accepted 26 May 2006
Available online 14 June 2006

Abstract

Electron cryomicroscopy (cryoEM) is capable of imaging large macromolecular machines composed of multiple components. However, it is currently only possible to achieve moderate resolution at which it may be possible to computationally extract the individual components in the machine. In this work, we present application details of an automated method for detecting and segmenting the components of a large machine in an experimentally determined density map. This method is applicable to object with and without symmetry and takes advantage of global and local symmetry axes if present. We have applied this segmentation algorithm to several cryoEM data sets already deposited in EMDB with various complexities, symmetries and resolutions and validated the results using manually segmented density and available structures of the components in the PDB. As such, automated segmentation could become a useful tool for the analysis of the ever-increasing number of structures of macromolecular machines derived from cryoEM.
© 2006 Elsevier Inc. All rights reserved.

Keywords: Segmentation; Electron cryomicroscopy; Macromolecular complexes; Structure

1. Introduction

Structural biology of macromolecular machines has become a crucial tool for understanding its mechanism of action, and in many instances leads to further function exploration. Large machines usually undergo motion and/or conformational changes in order to carry out a specific biological process (Alberts, 1998; Alberts and Miake-Lye, 1992). These large machines are made up of multiple components, ranging from one molecule repeated several times (e.g. GroEL) to tens of non-equivalent molecules (e.g. ribosome). The study of these machines has become more tractable due to the rapid development of proteomics for identifying and purifying ensembles of macromolecules (Gavin et al., 2002; Sali, 2003).

Despite limitations in resolution, electron cryomicroscopy (cryoEM) [1] can image macromolecular machines. One of the greatest advantages of cryoEM is the ability to analyze whole complexes in near native conditions as well as functionally important states. Even in the absence of high-resolution information, analysis of cryoEM structures may reveal significant insight into macromolecular composition, secondary structure and, in some instances, produce pseudo-atomic models of the components (Bottcher et al., 1997; Conway et al., 1997; Zhou et al., 2001).

In analyzing density maps of multi-component machines derived from cryoEM, one of the first steps is to define the boundaries of the individual molecular components (Chiu et al., 2005). These individual components may in fact form higher-ordered structures, such as an asymmetric unit in a

* Corresponding author. Fax: +1 713 798 1625.
E-mail address: mbaker@bcm.edu (M.L. Baker).

[1] *Abbreviations used:* cryoEM, electron cryomicroscopy; 3D, three-dimensional; RDV, rice dwarf virus.

1047-8477/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2006.05.013

virus. Defining the boundary and segmenting the individual components at different resolutions represents a particularly interesting challenge. While at low resolution, individual subunit boundaries may not be resolvable, making the segmentation nearly impossible. Conversely, at moderate resolutions, the molecular boundary of the components is more likely to be visible. The interactions between the neighbor molecules and the residual noise may, however, result in poor resolvability of the components. Even in the event a high-resolution structure is known for one or more components, it is possible that the same components may differ structurally in different functional states.

To date, segmentation has been accomplished mainly by interactively and visually tracing the regions of interest in density maps. This is typically done by examining and defining the regions of interest in a 2D slice of the 3D density map; numerous visualization packages have incorporated these tools such as Amira (TGS, San Diego, CA), IMOD (Boulder Laboratory for 3D Electron Microscopy of Cells, Boulder, CO) and the SAIL package developed for Iris Explorer (NAG, Dowber Grove, IL). Such manual segmentation is tedious and often ambiguous even with the most successful visualization interfaces. Despite the prevalence of these tools, few automated methods for segmentation are available (3D watershed method (Volkmann, 2002), normalized graph cut and eigenvector analysis (Frangakis and Forster, 2004; Frangakis and Hegerl, 2002), fast marching method (Bajaj et al., 2003; Frangakis and Forster, 2004; Frangakis and Hegerl, 2002; Volkmann, 2002)) and as such, segmentation of volumetric density is still considered one of the most difficult tasks in structure interpretation in complex machines.

In this work, we describe the application of an automated segmentation algorithm (Yu and Bajaj, 2005) to three-dimensional (3D) density maps derived from single-particle cryoEM. This algorithm includes three major computational steps: the detection of critical points of the volumetric density, the detection of global and local symmetry axes if present, and boundary segmentation of all subunits or domains. The resulting segmentation of the subunits is capable of producing molecular components with high fidelity, and requiring only minimal manual refinement.

2. Methods

2.1. CryoEM density maps and X-ray coordinates

To validate the segmentation procedures, we obtained structural data from public structural databases. Individual cryoEM density maps were obtained from EBI's electron microscopy database, EMDB. These density maps include the 9.5 Å resolution P22 mature phage (Jiang et al., 2003) (EMDB ID: 1101), the 23 Å resolution P22 tail machine (Tang et al., 2005) (EMDB ID: 1119), the 6.8 Å resolution Rice Dwarf Virus (RDV) (Zhou et al., 2001) (EMDB ID: 1060), 9 Å resolution *Escherichia coli* 70S ribosome (Valle et al., 2003) (EMDB ID: 1056) and the 6 Å, 11.5 Å and 25 Å resolution GroEL (Ludtke et al., 2004; Ludtke et al., 2001; Sewell et al., 2004) (EMDB IDs: 1081, 1080 and 1095, respectively). Corresponding X-ray structures were obtained from the Protein Data Bank (P22 tail spike PDB ID: 1TYU, RDV PDB ID: 1UF2, 50S ribosome PDB ID: 1FFK, 30S ribosome PDB ID: 1IBM, GroEL ID: 1GRL).

2.2. Segmentation

As described, manual segmentation of 3D density maps can be a tedious and subjective process, especially when the resolution is only marginally high enough to discern the boundaries between subunits. As such, we have developed an automatic and objective computational procedure for asymmetric subunit detection of complexes. This approach is a variant of the well-known fast marching method (Malladi and Sethian, 1998; Sethian, 1996; Sethian, 1999), in which a contour is initialized from a pre-chosen seed point and allowed to grow until a certain stopping condition is reached. The traditional fast marching method is designed for single-object segmentation. In order to segment multiple objects, like the molecular components found in macromolecular machine in a cryoEM density map, a seed for each of the components must be chosen. However, assigning only one seed to each object (i.e., component) may be problematic, as demonstrated previously, where a *re-initialization* scheme was proposed (Yu and Bajaj, 2005). This approach consists of three steps: (a) detection of the critical (seed) points; (b) classification of critical points; (c) multi-seeded fast marching method.

The critical points of a scalar map (i.e. cryoEM density map), in general, include three types: maximal, minimal and saddle. Only the maximal critical points are of interest in this approach, which represent the high-density features in the 3D density map and can be simply computed from the local maxima of a given scalar map. These critical points are regarded as the seed points in the fast marching method (Malladi and Sethian, 1998; Sethian, 1996; Sethian, 1999). In principal, the seed points for the fast marching procedure could be generated by other methods (k-means, random, etc.); however, only maximal critical points were considered in this work. Generally, the number of seed points in a map will be much larger than the number of subunits of interest, ranging from hundred to tens-of-thousands depending on the density map itself. As such, each component will be assigned multiple seed points instead of just one.

Complicating the accurate assignment of critical points is noise, which is present in the cryoEM density maps. In visualizing and analyzing cryoEM density maps, a pre-filtering process is generally applied to eliminate noise (Chiu et al., 2005). Linear (e.g. Gaussian filtering) or non-linear (Perona and Malik, 1990; Weickert, 1998) filters often destroy some weak features and eliminate some critical points. In this work, we have used a method based on *gradient vector diffusion* (Bajaj and Yu, 2005; Yu and Bajaj,

Download English Version:

https://daneshyari.com/en/article/2829612

Download Persian Version:

https://daneshyari.com/article/2829612

Daneshyari.com