Journal of Structural Biology 202 (2018) 129–141

Contents lists available at ScienceDirect

Journal of Structural Biology

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

Allosteric effects in bacteriophage HK97 procapsids revealed directly from covariance analysis of cryo EM data☆

Nan Xu[a,1], David Veesler[b], Peter C. Doerschuk[c,*], John E. Johnson[d]

[a] School of Electrical and Computer Engineering, Cornell University, United States
[b] Department of Biochemistry, University of Washington, United States
[c] Meinig School of Biomedical Engineering and School of Electrical and Computer Engineering, Cornell University, Phillips Hall Room 305, Ithaca, NY 14853, United States
[d] Department of Integrative Structural and Computational Biology, The Scripps Research Institute, United States

ARTICLE INFO

Keywords:
Bacteriophage HK97
Allosteric effects
Cryo electron microscopy
Covariance analysis
Maximum likelihood reconstruction

ABSTRACT

The information content of cryo EM data sets exceeds that of the electron scattering potential (cryo EM) density initially derived for structure determination. Previously we demonstrated the power of data variance analysis for characterizing regions of cryo EM density that displayed functionally important variance anomalies associated with maturation cleavage events in Nudaurelia Omega Capensis Virus and the presence or absence of a maturation protease in bacteriophage HK97 procapsids. Here we extend the analysis in two ways. First, instead of imposing icosahedral symmetry on every particle in the data set during the variance analysis, we only assume that the data set as a whole has icosahedral symmetry. This change removes artifacts of high variance along icosahedral symmetry axes, but retains all of the features previously reported in the HK97 data set. Second we present a covariance analysis that reveals correlations in structural dynamics (variance) between the interior of the HK97 procapsid with the protease and regions of the exterior (not seen in the absence of the protease). The latter analysis corresponds well with hydrogen deuterium exchange studies previously published that reveal the same correlation.

1. Introduction

Cryo electron microscopy (cryo EM) records image data of individual instances of the particle. Therefore, when the instances are not identical, the heterogeneity of the instances can be characterized. Even when atomic resolution is achieved, heterogeneity can still be important as in the Flock House Virus example of Supplemental Material Section B. Heterogeneity is often thought of at two levels, discrete heterogeneity, such as individual instances lacking some components, and continuous heterogeneity, such as flexibility. In many situations, accounting for heterogeneity is necessary in order to improve the resolution of the reconstruction.

Continuous heterogeneity has been described as an important issue in multiple review articles over a range of years, e.g., Taylor and Glaeser (2008, p. 221) to Bai et al. (2015, p. 55). However, in Ludtke (2016), S. J. Ludtke (the primary developer of the popular EMAN2 (Tang et al., 2007; Ludtke et al., 1999) system) lists only resampling as a method for characterizing continuous heterogeneity. Furthermore, in Scheres (2016), S. H. W. Scheres (the primary developer of the popular RELION (Scheres, 2012; Scheres, 2012) system) only describes methods for dealing with discrete heterogeneity where every particle in a particular class is identical, which can be extended to some forms of continuous heterogeneity when subregions of the structure move as rigid bodies (Scheres, 2016, Sections 4.4–4.6) but not in the absence of such rigid motion. Therefore, new methods for characterizing continuous heterogeneity are needed.

A large on-line continuously-updated repository of references concerning heterogeneity is available (3DEMMethods, Continuously updated). This paper is concerned with continuous statistical heterogeneity of particles that would otherwise be grouped together into a single class (Section 2.2.3) and how symmetry can be applied to the statistics rather than to the particles themselves (Section 2.2.4). Other investigators have studied problems related to continuous statistical heterogeneity. With enough computing, many systems can compute the statistics of the electron scattering intensity at each voxel in the 3-D reconstruction by resampling (Penczek et al., 2011; Spahn and Penczek, 2009; Zhang et al., 2008; Penczek et al., 2006; Simonetti et al., 2008). Resampling is the process of randomly selecting images from the image stack and using the randomly selected images to compute a reconstruction. Repeating this process many times results in many

☆ NX and PCD are grateful to NSF 1217867 for funding.
* Corresponding author.
E-mail address: pd83@cornell.edu (P.C. Doerschuk).
[1] Current address: School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, United States.

https://doi.org/10.1016/j.jsb.2017.12.013
Received 4 October 2017; Received in revised form 24 December 2017; Accepted 27 December 2017
Available online 10 January 2018

reconstructions and the statistics at each voxel can then be computed by standard statistics. Other investigators estimate the covariance of the voxel values from the sample covariance of the images (Liao et al., 2015; Liao and Frank, 2010; Katsevich et al., 2015). Resolution is limited, e.g., $16 \times 16 \times 16$ in Liao et al. (2015), it is not clear how to account for the measurement noise that contaminates experimental images, and the orientations estimated from the images are assumed to be correct when estimating the covariance of the voxel values. Other investigators base their analysis on having an atomic or pseudo-atomic structure from which normal modes can be computed (Jin et al., 2014). While such structures are increasingly available from cryo EM itself, this type of approach is less directly based on the cryo EM images than our approach (Wang et al., 2013; Zheng et al., 2012), it is not clear how to account for the measurement noise, and the computations are extensive, taking 2, 4, and 6 min to analyze an image of size 128×128 pixels using 1, 2, or 3 normal modes on a Dual Intel Xeon X5472 processor running at 3.00 GHz.

Other approaches characterize heterogeneity in terms of statistics, although still without considering symmetry. Without first computing a covariance matrix, the algorithm of Tagare et al. (2015) sequentially computes the principal components of the structure. The electron scattering intensity is represented as a mean plus a linear combination of orthogonal vectors where the weights are independent Gaussian random variables with zero mean and unit variance (Tagare et al., 2015, Eq. 1). The number of vectors is typically less than or equal to five (Tagare et al., 2015, following Eq. 14). This approach has provided interesting information about large-scale motion in ribosomes. However, the virus problem of this paper needs finer scale information that is more spatially localized than a principal component that spans the entire particle and need to incorporate the symmetry of the particle.

A different approach is to mask out variable regions in the structure which can result in higher resolution in the non-variable regions (Rawson et al., 2016). This approach is different from what is proposed in this paper because the focus of Rawson et al. (2016) is on preventing variability from decreasing resolution in other regions of the particle while the focus of this paper is on characterizing the nature of the variability.

In addition, there are other ideas related to but different from continuous heterogeneity. For instance, the method of Frank and Ourmazd (Dashti et al., 2014; Frank and Ourmazd, 2016) does not characterize continuous heterogeneity but rather is a very sophisticated manifold-based method for interpolating among the many structures that can be computed from certain datasets such as the ribosome dataset of Dashti et al. (2014). If the Brownian ratchet ideas reviewed in Moore (2012) are correct, then the characterization of the continuous heterogeneity of the ribosome has to have a statistical character, which is not provided by interpolation methods, because the Brownian ratchet description is itself statistical.

In this paper we further develop ideas for characterizing continuous heterogeneity via moments of the electron scattering intensity (Gong et al., 2016). Let the electron scattering intensity at the 3-D location $\mathbf{x}$ be denoted by $\rho(\mathbf{x})$. The first moment is the average of $\rho(\mathbf{x})$ over all the particles, which is denoted by $\bar{\rho}(\mathbf{x})$. The deviation is the difference between the electron scattering intensity and the first moment, which is denoted by $\delta(\mathbf{x})$. The second moment is the average of the product of the deviation at two different 3-D locations $\mathbf{x}_1$ and $\mathbf{x}_2$, which is denoted by $C_\rho(\mathbf{x}_1,\mathbf{x}_2)$. The advance relative to Gong et al. (2016) is that $\mathbf{x}_1$ and $\mathbf{x}_2$ can be different locations[2]. The enabling advance is in the statistical description of the collection of particles and existing descriptions and the new description, along with a review of the literature, are presented in Section 2. This advance allows the use of these tools to investigate allosteric interactions at spatially distant locations.

[2] When $\mathbf{x}_1 = \mathbf{x}_2$ what is computed is a space-varying variance which is a 3-D cube while when $\mathbf{x}_1 \neq \mathbf{x}_2$ what is computed is the complete space-varying covariance which is a 6-D cube.

HK97 is a dsDNA bacteriophage for which Virus Like Particles (VLPs) are available including two VLPs that trap the particle in the Prohead I stage of maturation (Veesler et al., 2014). One of the VLPs, denoted by $\text{PhI}^{\text{Pro}-}$, lacks the virally-encoded maturation protease that is essential for going beyond the Prohead I stage while the other VLP, denoted by $\text{PhI}^{\text{Pro}+}$, has the protease but the protease is defective so the particle is again trapped at the Prohead I stage. Using standard cryo EM analysis (Veesler et al., 2014), it was not possible to detect a difference between $\text{PhI}^{\text{Pro}+}$ and $\text{PhI}^{\text{Pro}-}$ even though $\text{PhI}^{\text{Pro}+}$ has approximately 100 copies of the protease positioned on the inner surface of the peptide capsid. Using an earlier version of the heterogeneity characterization method described in this paper (Gong et al., 2016), it was possible to distinguish the $\text{PhI}^{\text{Pro}+}$ and $\text{PhI}^{\text{Pro}-}$ particles. However, only variance information was available so it was not possible to investigate how heterogeneity in spatially-distant positions was related. As is described in the previous paragraph, in this paper a generalization of Gong et al. (2016) is described such that covariance information can be determined and therefore the relationship between heterogeneity at different spatial locations can be studied. We use this tool to provide evidence of allosteric interactions between the protease in the δ domain of the capsid and the surface of the capsid.

2. Theoretical methods

2.1. Symmetry

While many particles have no symmetry, e.g., ribosomes, cryo EM has made a large contribution to structural virology and many virus particles or Virus Like Particles (VLPs) have icosahedral symmetry. A particle with electron scattering intensity $\rho(\mathbf{x})$ has icosahedral symmetry if it is unchanged under each of $N_g = 60$ rotational symmetry operators which are either 5-fold, 3-fold, or 2-fold rotational symmetries for which the axes of rotation intersect in a single point at the center of the particle. Let R_β be the βth 3×3 rotation operator for $\beta \in \{1,\ldots,N_g\}$. Then for $\rho(\mathbf{x})$ to be unchanged under the symmetry operators means that $\rho(R_\beta^{-1}\mathbf{x}) = \rho(\mathbf{x})$ for all three-dimensional vectors $\mathbf{x}$ and for all $\beta \in \{1,\ldots,N_g\}$. Symmetry under a group of rotations other than the icosahedral group simply means that the number of matrices (N_g) and the values of the matrices (R_β) change.

2.2. Ensembles of particles

In a crystal used for X-ray crystallography, the particles must be similar in order for a crystal to form and the particle-particle interactions in the crystal, which are generally not present in the particle's natural environment, make the particles even more uniform. Each piece of data, i.e., a diffraction amplitude, is a function of all of the particles. In contrast, in the vitreous film used for cryo EM, the particles interact with the water-air interfaces, in some instances leading to orientational preferences, but the particles do not interact strongly with each other. Each piece of data, i.e., a boxed image, is a function of just one of the particles.

The non-interacting nature of the particles in the cryo EM specimen makes it natural to consider at least four types of particle ensemble. Cartoons for each type are shown in Fig. 1.

2.2.1. Homogeneous ensemble (Fig. 1(a))

Each particle is identical. This is the original point of view in early work such as Adrian et al. (1984, 1981, 1988). We will describe this as a *homogeneous* ensemble. If icosahedral symmetry is present, then $\rho(R_\beta^{-1}\mathbf{x}) = \rho(\mathbf{x})$ for all three-dimensional vectors $\mathbf{x}$ and for all $\beta \in \{1,\ldots,N_g\}$.

2.2.2. Discretely heterogeneous ensemble (Fig. 1(a))

Each particle belongs to one of a finite set of classes and within each class every particle is identical. This point of view is widely represented in current software such as EMAN (Ludtke et al., 1999), EMAN2 (Tang