Journal of Structural Biology 187 (2014) 76–83

Contents lists available at ScienceDirect

Journal of Structural Biology

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

CryoEM and image sorting for flexible protein/DNA complexes

Seth A. Villarreal, Phoebe L. Stewart *

Department of Pharmacology and Cleveland Center for Membrane and Structural Biology, Case Western Reserve University, 10900 Euclid Ave., Cleveland, OH 44106, United States

ARTICLE INFO

Article history:
Received 7 October 2013
Received in revised form 13 November 2013
Accepted 5 December 2013
Available online 9 December 2013

Keywords:
Cryo-electron microscopy
CryoEM
Single particle reconstruction
Flexibility
Heterogeneity
Eigenimage sorting
Protein/DNA complexes

ABSTRACT

Intrinsically disordered regions of proteins and conformational flexibility within complexes can be critical for biological function. However, disorder, flexibility, and heterogeneity often hinder structural analyses. CryoEM and single particle image processing techniques offer the possibility of imaging samples with significant flexibility. Division of particle images into more homogenous subsets after data acquisition can help compensate for heterogeneity within the sample. We present the utility of an eigenimage sorting analysis for examining two protein/DNA complexes with significant conformational flexibility and heterogeneity. These complexes are integral to the non-homologous end joining pathway, and are involved in the repair of double strand breaks of DNA. Both complexes include the DNA-dependent protein kinase catalytic subunit (DNA-PKcs) and biotinylated DNA with bound streptavidin, with one complex containing the Ku heterodimer. Initial 3D reconstructions of the two DNA-PKcs complexes resembled a cryoEM structure of uncomplexed DNA-PKcs without additional density clearly attributable to the remaining components. Application of eigenimage sorting allowed division of the DNA-PKcs complex datasets into more homogeneous subsets. This led to visualization of density near the base of the DNA-PKcs that can be attributed to DNA, streptavidin, and Ku. However, comparison of projections of the subset structures with 2D class averages indicated that a significant level of heterogeneity remained within each subset. In summary, image sorting methods allowed visualization of extra density near the base of DNA-PKcs, suggesting that DNA binds in the vicinity of the base of the molecule and potentially to a flexible region of DNA-PKcs.

1. Introduction

There is growing recognition of the importance of structural flexibility and intrinsic disorder in proteins for performing essential biological functions (Babu et al., 2012; Mittag et al., 2010; Uversky, 2011). Conformational flexibility is an issue that experimental structural biology is beginning to address. Often flexible regions are absent from crystal structures. NMR methods are being developed to quantify intrinsically disordered proteins and provide ensemble descriptions of flexible regions (Jensen et al., 2013). In general, cryoEM single particle image reconstruction requires a significant level of structural homogeneity in the dataset. Small intrinsically disordered regions within a symmetrical assembly, such as an icosahedral virus, can be modeled with molecular dynamics simulations (Flatt et al., 2013). Inherently flexible complexes such as the ribosome, which stochastically fluctuates among multiple conformations, can be induced into intermediate states, which can be studied by cryoEM and X-ray crystallography (Frank, 2012). Various software approaches are being developed for dealing with heterogeneity within cryoEM datasets. These include a bootstrapping approach (Penczek et al., 2006), a Bayesian statistics approach (Scheres, 2012), and several methods based on Multivariate Statistical Analysis (MSA) (Orlova and Saibil, 2010; van Heel et al., 2000).

CryoEM single particle studies of macromolecular complexes can be complicated by flexibility within the components, heterogeneity in the composition of the complex, and low signal-to-noise ratios. The first two factors lead to a loss of density for flexible regions through averaging and the last factor can result in overfitting and refinement on noise artifacts. The so-called gold standard refinement approach provides a degree of protection from overfitting and the subsequent misinterpretation of the resulting structure (Scheres and Chen, 2012). This refinement approach involves processing half datasets separately (e.g., odd and even particle images), calculating separate maps for the odd and even particle images, and using these half maps for additional rounds of refinement while keeping the odd and even datasets independent (odd map with odd particle images, and even map with even particle images). The resolution is still estimated by comparison of the odd and even maps as normal, but the odd and even maps are not influenced by each other during refinement. In this manner, random features based solely on noise within either half map should not be reinforced in the merged full map, overfitting

* Corresponding author. Fax: +1 216 368 1300.
E-mail addresses: pls47@case.edu, phoebe.stewart@case.edu (P.L. Stewart).

1047-8477/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jsb.2013.12.002

is limited, and more realistic estimates are obtained for the resolution.

Components within the non-homologous end joining (NHEJ) pathway have to be adaptable enough to act on a wide variety of deleterious DNA end substrate configurations. In addition to repairing deleterious breaks, this pathway is also critical for repairing programmed double-strand DNA breaks necessary for generation of diversity in the adaptive immune system. The DNA-dependent protein kinase catalytic subunit (DNA-PKcs) has a central regulatory role in NHEJ repair and is a large (469 kDa) serine/threonine protein kinase belonging to the phosphatidylinositol-3-OH kinase (PI3K)-related protein family. DNA-PKcs functions together with Ku, which is a heterodimer of Ku70 and Ku86. When a double-strand break occurs, Ku recognizes and binds broken DNA ends. Next DNA-PKcs is recruited to the break site by the Ku–DNA complex via interaction with the C-terminal region of Ku86. The interaction between DNA-PKcs and DNA leads to autophosphorylation of DNA-PKcs, which is presumed to regulate the access of the DNA end to other NHEJ components. DNA-PKcs has predicted disordered regions, and we speculate that NHEJ complexes may use intrinsically disordered regions to enable recognition of a diverse array of damaged ends during the repair process. In addition, conformational flexibility may facilitate recruitment of other factors in the NHEJ pathway, including enzymes with nuclease, polymerase and ligase activities (Lieber, 2010).

Numerous structural studies have provided only moderate resolution structures of DNA-PKcs (Chiu et al., 1998; Leuther et al., 1999; Rivera-Calzada et al., 2005; Sibanda et al., 2010; Williams et al., 2008). A crystal structure of human DNA-PKcs at 6.6 Å resolution revealed helices throughout the molecule but did not permit fitting of the primary sequence into the density (Sibanda et al., 2010). Coordinates were only determined for 46% of the backbone residues and these coordinates were not assigned to specific residues within the sequence. Loops between secondary structural elements are missing. In addition, since specific residues were not assigned there is a possibility that large contiguous regions of protein are missing from the crystal structure due to flexibility. Sibanda et al. proposed that the kinase domain is localized within the head region after showing a good superposition for a region in the head of the 6.6 Å resolution crystal structure with the atomic structure of the homologous PI3K-γ kinase (Walker et al., 1999). Crystallization of DNA-PKcs was performed in complex with a C-terminal fragment of Ku86 (Ku80ct194, aa 539–732), but it was not possible to identify the Ku80ct194 fragment within the overall structure. EM studies have revealed the head and the base of the molecule (Chiu et al., 1998; Leuther et al., 1999; Rivera-Calzada et al., 2005; Williams et al., 2008), as well as a large central channel with a helical protrusion recessed from the opening of the channel that has been proposed as a possible DNA binding site (Williams et al., 2008).

The EM and X-ray structures of DNA-PKcs differ mainly in the size of the base of the molecule (Supplemental Fig. S1), which is composed in part by conformationally flexible HEAT repeats (Sibanda et al., 2010). The base of the EM structure appears substantially larger than that of the X-ray structure. The two structures might represent different conformations, with the cryoEM structure representing the more prevalent conformation in solvent and with the crystal structure missing significant regions due to flexibility.

Crystal structures of DNA-free and DNA-bound forms of Ku have been determined using full-length Ku70 and a truncated form of Ku86 (aa1–565), missing 167 C-terminal residues (Walker et al., 2001). The DNA-binding core of Ku is formed by Ku70 and the majority of Ku86, excluding the extended C-terminal portion of Ku86. Double-stranded DNA binds within a preformed ring of the Ku DNA-binding core. NMR structures have been determined for portions of the C-terminal region of Ku86, including 592–709aa (Harris et al., 2004) and 566–710aa (Zhang et al., 2004). Neither of the NMR studies revealed the structure of the final ~20 residues at the C-terminus of Ku86, which are known to mediate a specific interaction with DNA-PKcs (Falck et al., 2005; Gell and Jackson, 1999). A small angle X-ray scattering (SAXS) analysis of full-length Ku with and without DNA indicates that the Ku86 C-terminal region forms a flexible arm that extends up to ~100 Å away from the DNA-binding core (Hammel et al., 2010).

Here we present the results of an eigenimage sorting analysis applied to cryoEM datasets of two NHEJ complexes including DNA-PKcs, double-stranded DNA, and Ku. Our original goal was to localize the binding site on DNA-PKcs for double-stranded DNA within these complexes. The samples were formed with DNA biotinylated at one end and complexed with streptavidin, so as to block multiple DNA-PKcs and Ku proteins from associating with a single DNA molecule. Considerable structural flexibility within DNA-PKcs, as well as heterogeneity within the NHEJ complexes, hampered our efforts to localize the DNA binding site. However, eigenimage sorting allowed us to constrain the region of DNA-PKcs where DNA and Ku are likely to bind.

2. Results

2.1. Initial structures of DNA-PKcs/DNA and DNA-PKcs/Ku/DNA

DNA-PKcs/DNA and DNA-PKcs/Ku/DNA complexes were formed with DNA biotinylated at one end, allowing one streptavidin tetramer to bind, and limiting access of DNA-PKcs to only one end of the DNA. The length of the double-stranded DNA molecules differed in the two complexes with 18 bp in the DNA-PKcs/DNA complex and a longer 35 bp DNA molecule in the DNA-PKcs/Ku/DNA complex to allow for binding both Ku and DNA-PKcs. CryoEM datasets were collected for DNA-PKcs/DNA and DNA-PKcs/Ku/DNA (35,176 and 19,825 particle images, respectively). Image processing of the DNA-PKcs/DNA and DNA-PKcs/Ku/DNA datasets was performed with the 3D Fourier Space software package (Benlekbir et al., 2012) which incorporates the so-called gold standard refinement process. This resulted in structures for the two complexes that resemble uncomplexed DNA-PKcs (Fig. 1). The resolution of these structures measured 15.0 and 21.4 Å, respectively, using the Fourier Shell Correlation (FSC) 0.5 threshold (Supplemental Fig. S2).

For comparison, we re-processed the cryoEM dataset of DNA-PKcs with 3D Fourier Space using the portion of the dataset collected with a single magnification (284,173 particles). With gold standard refinement of this partial dataset, we calculated a DNA-PKcs structure at 13.6 Å resolution by the FSC 0.5 threshold (Fig. 1A, Supplemental Fig. S2). All three structures showed similar overall features, albeit at different resolutions. No clear density for DNA, streptavidin or Ku was observed in the structures of the complexes. There appeared to be no large conformational changes in DNA-PKcs when in complex with DNA or Ku/DNA; however, we cannot rule out small conformational changes. Both complex structures included >99% of the particle images, and it is possible that they represent heterogeneous populations with multiple conformations or partial DNA or Ku occupancy. The homogeneity of the complex samples was assessed by negative-stain EM (Supplemental Fig. S3). Visual inspection of particle images and class averages led to an estimate of ~30% of each population forming the desired complex.

During refinement of DNA-PKcs/DNA and DNA-PKcs/Ku/DNA, we noted that the resolutions indicated by the FSC curves were influenced by the density mask used. 3D Fourier Space allows the calculation of a custom mask for each complex with a user-specified density threshold. For consistency, we calculated