Journal of Structural Biology 173 (2011) 261–270

Contents lists available at ScienceDirect

Journal of Structural Biology

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

Conformational dynamics of supramolecular protein assemblies

Do-Nyun Kim, Cong-Tri Nguyen [1], Mark Bathe *

Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

ARTICLE INFO

Article history:
Received 6 July 2010
Received in revised form 11 September 2010
Accepted 14 September 2010
Available online 18 September 2010

Keywords:
Electron Microscopy Data Bank
Normal mode analysis
Finite element method

ABSTRACT

Supramolecular protein assemblies including molecular motors, cytoskeletal filaments, chaperones, and ribosomes play a central role in a broad array of cellular functions ranging from cell division and motility to RNA and protein synthesis and folding. Single-particle reconstructions of such assemblies have been growing rapidly in recent years, providing increasingly high resolution structural information under native conditions. While the static structure of these assemblies provides essential insight into their mechanism of biological function, their dynamical motions provide additional important information that cannot be inferred from structure alone. Here we present an unsupervised computational framework for the analysis of high molecular weight protein assemblies and use it to analyze the conformational dynamics of structures deposited in the Electron Microscopy Data Bank. Protein assemblies are modeled using a recently introduced coarse-grained modeling framework based on the finite element method, which is used to compute equilibrium thermal fluctuations, elastic strain energy distributions associated with specific conformational transitions, and dynamical correlations in distant molecular domains. Results are presented in detail for the ribosome-bound termination factor RF2 from *Escherichia coli*, the nuclear pore complex from *Dictyostelium discoideum*, and the chaperonin GroEL from *E. coli*. Elastic strain energy distributions reveal hinge-regions associated with specific conformational change pathways, and correlations in collective molecular motions reveal dynamical coupling between distant molecular domains that suggest new, as well as confirm existing, allosteric mechanisms. Results are publically available for use in further investigation and interpretation of biological function including cooperative transitions, allosteric communication, and molecular mechanics, as well as in further classification and refinement of electron microscopy based structures.

1. Introduction

Single-particle reconstructions of supramolecular protein assemblies deposited in the publically accessible Electron Microscopy Data Bank (EMDB, http://www.emdatabank.org/) have been growing rapidly in recent years, representing a total of approximately 250 distinct structures in 2009 (Tagari et al., 2002; Henrick et al., 2003). The EMDB covers a range of supramolecular assemblies including viruses as the dominant class, and RNA binding proteins and protein kinases as major subclasses (Fig. 1). Recent growth of the EMDB parallels early growth of the Protein Data Bank (PDB), which has developed to include tens of thousands of protein crystal structures since its inception in 1971 (Bernstein et al., 1977; Berman et al., 2000). While the static structure of proteins provides invaluable insight into their biological function, their conformational dynamics often play an additional important role in understanding their function mechanistically (Xu and Sigler, 1998; Conway et al., 2001; Zhang et al., 2000).

Normal mode analysis (NMA) has proven to be an effective computational approach to investigate biologically relevant collective motions about a representative ground-state structure, or ensemble thereof (Cui and Bahar, 2006). The primary advantage of NMA over molecular or Brownian dynamics is its relative computational efficiency, which is a result of the harmonic approximation of atomic motions about the ground-state conformation, as well as the neglect of explicit solvent degrees of freedom. Computational efficiency is further enhanced in NMA by using coarse-grained modeling approaches that reduce the number of protein degrees of freedom, which has been essential to facilitating the analysis of high molecular weight protein assemblies. Popular approaches include the Rotational Translational Block (RTB) procedure (Tama et al., 2000), which requires atomic coordinates for the underlying protein structure, the Gaussian (Suezaki and Go, 1975; ben-Avraham, 1993; Tirion, 1996; Haliloglu et al., 1997) and Elastic Network Models (ENM) (Bahar and Rader, 2005), the

* Corresponding author. Address: 77 Massachusetts Avenue, Building NE47, Room 223, Cambridge, MA 02139, USA. Fax: +1 617 324 7554.
E-mail addresses: kingddo@mit.edu (D.-N. Kim), cnguyen8@illinois.edu (C.-T. Nguyen), mark.bathe@mit.edu (M. Bathe).
[1] Present address: Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, 156 Davenport Hall, 607 Matthews Avenue, Urbana, IL 61801, USA.

1047-8477/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2010.09.015

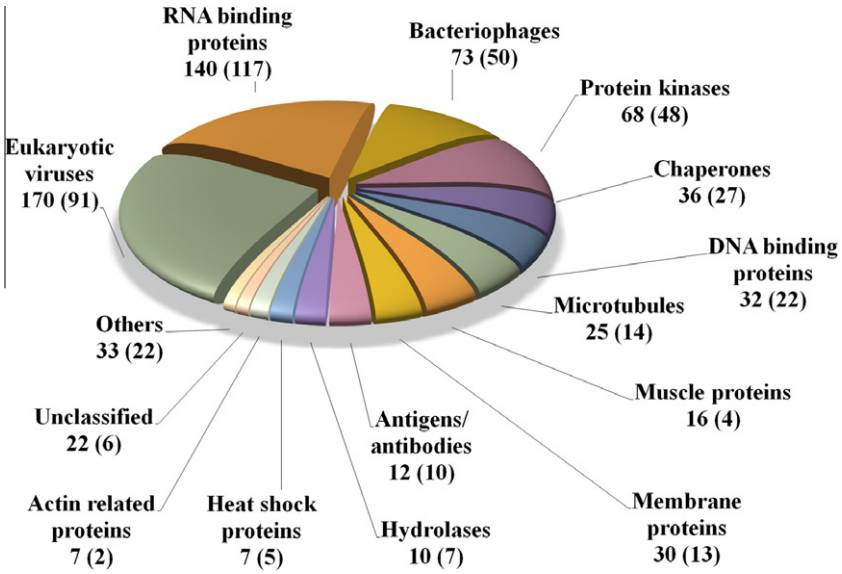

Fig. 1. Classification of EMDB entries according to biological function. Numbers in parentheses denote successfully analyzed structures deposited at http://www.cdyn.org.

Rigid-Cluster Model (Kim et al., 2003, 2005; Schuyler and Chirikjian, 2005), and more recently the Finite Element Method (FEM) (Bathe, 2008). The FEM is a well established numerical procedure with solid theoretical foundations that has been developed over several decades to be applied to a broad range of continuum and molecular-level dynamical phenomena (Phillips et al., 2002; Bathe, 1996; Shenoy et al., 1998). The FEM provides a natural framework for the computation of conformational dynamics and mechanics of high molecular weight proteins and their assemblies based on EM reconstructions because the model is defined using a closed molecular surface, which is naturally provided by single-particle EM reconstructions. In the FE framework employed here, proteins are modeled as homogeneous isotropic elastic bodies characterized by a mean mass density and elastic stiffness, which has been shown to reproduce quantitatively atomic-level protein fluctuations and correlations computed using all-atom NMA (Bathe, 2008). Accurate prediction of atomic motions using the FEM is attributed to its preservation of detailed molecular shape, even for high molecular weight protein assemblies.

While several data banks and servers (Alexandrov et al., 2005; Suhre and Sanejouand, 2004; Wako et al., 2004; Yang et al., 2005; Hollup et al., 2005; Lindahl et al., 2006) exist to disseminate publically the conformational dynamics of protein structures deposited in the PDB, similar data banks do not exist at present for the EMDB. Such a data bank would support both further computational analyses to gain insight into the biological function of high molecular weight protein assemblies lacking atomic structure, as well as potentially serve as a basis set for classification in single-particle reconstruction (Brink et al., 2004). Toward this end, here we establish an unsupervised computational framework to analyze the conformational dynamics of structures deposited in the EMDB and store them in a publically accessible online data bank.[2] In this framework, the molecular surface of EMDB entries are computed and validated for computation of NMs using the FEM (Bathe, 2008). NMs may be used to calculate conformational properties including root-mean-square fluctuations (RMSFs) of the molecules in thermal equilibrium, elastic strain energy densities corresponding to biologically relevant conformational changes, and correlations in collective dynamical motions that may relate to cooperative or allosteric mechanisms. Individual NM shapes and frequencies are provided together with the molecular models used to perform the analyses, which may be used in further FE-based analyses of dynamical and mechanical response (Fig. 2). Results for the ribosome-bound termination factor RF2 from *Escherichia coli*, the nuclear pore complex from *Dictyostelium discoideum*, and the bacterial chaperonin GroEL from *E. coli* are presented in detail here to illustrate the utility of the foregoing results.

2. Materials and methods

Structures in the EMDB are analyzed using an automated procedure that consists of several distinct computational steps (Fig. 3A): (1) retrieval of the EM density map; (2) molecular surface computation and discretization; (3) discretized molecular surface evaluation and repair; (4) FE model generation and NMA; and (5) results processing. The EMDB is monitored regularly to determine when new structures suitable for conformational dynamics analysis are deposited. To date, the preceding analysis approach has been applied to 681 EMDB entries with 453 entries solved successfully. The remaining entries are excluded from the analysis because 55 entries are on hold by the EMDB, 31 entries are tomograms, 10 entries do not provide contour levels or molecular weights from which to determine the molecular surface, 87 entries consist of disconnected multiple bodies, and 45 entries have surface meshes that could not be repaired using the current approach (Fig. 3B and Table S1). Proteins are classified according to their biological function by title and sample name keywords provided by the EMDB (Fig. 1).

2.1. Molecular surface computation and discretization

Computation and discretization of the molecular surface in step 2 is performed using the marching cubes algorithm (Lorensen and Cline, 1987) implemented in Chimera (Pettersen et al., 2004). The triangulated surface is subsequently exported in OBJ format, a geometry definition file format originally developed by Wavefront Technologies, Inc., Santa Barbara, CA. To define the molecular surface, the suggested contour level provided in the EMDB entry is used unless no such contour level is provided. In this case, the molecular weight is used instead (four entries), where the contour level corresponding to the given molecular volume assuming a

[2] Results are available at http://www.cdyn.org.