Computational and Structural Biotechnology Journal 13 (2015) 492–503

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

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

Integrative Modeling of Macromolecular Assemblies from Low to Near-Atomic Resolution

Xiaojun Xu, Chunli Yan, Robert Wohlhueter, Ivaylo Ivanov *

Department of Chemistry, Center for Diagnostics and Therapeutics, Georgia State University, Atlanta, GA 30302, USA

ARTICLE INFO

Article history:
Received 1 June 2015
Received in revised form 9 August 2015
Accepted 13 August 2015
Available online 3 September 2015

Keywords:
Integrative modeling
Hybrid modeling
Molecular dynamics flexible fitting (MDFF)
Small angle X-ray scattering (SAXS)
Electron microscopy (EM)
DNA replication and repair

ABSTRACT

While conventional high-resolution techniques in structural biology are challenged by the size and flexibility of many biological assemblies, recent advances in low-resolution techniques such as cryo-electron microscopy (cryo-EM) and small angle X-ray scattering (SAXS) have opened up new avenues to define the structures of such assemblies. By systematically combining various sources of structural, biochemical and biophysical information, integrative modeling approaches aim to provide a unified structural description of such assemblies, starting from high-resolution structures of the individual components and integrating all available information from low-resolution experimental methods. In this review, we describe integrative modeling approaches, which use complementary data from either cryo-EM or SAXS. Specifically, we focus on the popular molecular dynamics flexible fitting (MDFF) method, which has been widely used for flexible fitting into cryo-EM maps. Second, we describe hybrid molecular dynamics, Rosetta Monte-Carlo and minimum ensemble search (MES) methods that can be used to incorporate SAXS into pseudoatomic structural models. We present concise descriptions of the two methods and their most popular alternatives, along with select illustrative applications to protein/nucleic acid assemblies involved in DNA replication and repair.

Contents

1. Introduction

The structures of complex biological assemblies command considerable attention, since critical cellular activities are more often than not carried out by such assemblies rather than by a single macromolecular component. A high-resolution structural model of an assembly is often crucial to understanding its function; and biological mechanisms can be deduced from a detailed view of the structure and interactions of components in an assembly. Structures at atomic resolution are usually obtained through X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy. However, the size and flexibility of macromolecular assemblies often pose technical difficulties, confounding structural elucidation and impeding mechanistic exploration by conventional methods. Cryo-electron microscopy (cryo-EM) is one of the most promising techniques

* Corresponding author at: Georgia State University, P.O. Box 3965, Atlanta, GA 30302, USA. Tel.: +1 404 413 5529; fax: +1 404 413 5505.
E-mail address: iivanov@gsu.edu (I. Ivanov).

http://dx.doi.org/10.1016/j.csbj.2015.08.005

for elucidating larger macromolecular complexes but until recently it was only capable of generating structural models at resolutions of 8–20 Å [1] – substantially lower than routine X-ray crystallography. Better resolution (3.5–4.5 Å) was reported only for complexes with high symmetry and stability [2–4]. Not until very recently have the advances in high-resolution image-capturing hardware [5] and image-processing technology [6] enabled cryo-EM to yield near-atomic resolution maps [7,8]. With the new technologies, the structure of a mammalian TRP channel, TRPV1, was successfully determined at a resolution of 3.4 Å, for the first time reaching side-chain resolution for a membrane protein without crystallization [9,10]. In 2014, the success of cryo-EM was boosted by many other explorations, resulting in 3.0–5.0 Å resolution structures of β-galactosidase [11], membrane proteins [12–14] and ribosomal machineries [15] and leading to the notion of "resolution revolution" in single particle cryo-EM [16]. Recently, Campbell et al. reported a cryo-EM reconstruction of 2.8 Å for the 700 kDa *Thermoplasma acidophilum* 20S proteasome [17]. Furthermore, in 2015 the Subramaniam group at the National Cancer Institute further refined a β-galactosidase EM structure to an unprecedented 2.2 Å resolution [18], whereby the authors were able to identify densities of structural water molecules and ions, and demonstrated it is rather the intrinsic flexibility of the target molecule/complex and the quality of the specimen than the image-capturing or processing technologies that prevented achieving resolution close to 2 Å by cryo-EM. Apart from the breakthrough of near-atomic resolution, cryo-EM offers significant advantages in not requiring the high concentration of protein/complex that X-ray crystallography demands [19]. Nor does it require preparation of macroscopic crystals, since individual complexes are preserved in a frozen hydrated state on an EM grid. Thus, cryo-EM visualizes a structure more akin to that "in solution", and probably of more relevance to in vivo conditions [19]. Given all of these exciting developments, cryo-EM stands poised to overtake X-ray crystallography and play an even more prominent role in the visualization of macromolecular complexes.

Other technologies also generate spatial envelopes of biological molecules or assemblies e.g. negative stain electron microscopy (EM) and small angle X-ray scattering (SAXS), while detailed interaction profiles are accessible through methodologies like chemical footprinting, cross-linking, fluorescence resonance energy transfer (FRET), mass spectrometry (MS), proteomics studies, and so on [20,21]. Though both shape and interactions often contribute to modeling a complex, the results from these methods are largely heterogeneous and dispersed in the literature. Therefore, an integrative modeling approach capable of combining these heterogeneous data and translating them into a uniform structural representation would be valuable in advancing our understanding of the relevant biological functions of these assemblies. Incorporating information from such diverse approaches may in fact lead to a highly useful model in less time and effort than by the conventional means of X-ray crystallography or NMR spectroscopy. And this may be the only means of arriving at a useful model. Moreover, the resulting model may be more useful to experimentalists, in that, by consolidating diverse experimental data, it may generate new hypotheses directly amenable to experimental tests. A notable example of the power and utility of integrative modeling methods was given by an elegant study by Alber et al., which elucidated the architecture of the nuclear pore complex (NPC) using a combination of diverse high-quality proteomic and structural data [22]. The advance was made possible by an integrative modeling platform IMP. IMP provides software tools to represent almost any conceivable combination of experimental data (e.g. relative positions of protein domains, mutational data on residue contacts, shape information from SAXS envelopes, EM densities and symmetry information). This data could even be of a type not normally used for structure determination or ambiguous in terms of structural interpretation. This diverse data is subsequently converted to spatial restraints, which collectively determine a scoring function. A structural ensemble is then generated and analyzed, which optimally satisfies the scoring function. The considerable freedom to mix and match modules in IMP allows the seamless construction of new hybrid modeling protocols. The major advantage of IMP lies in the flexible nature of the code, written as a software framework – a collection of independent modules in C++ and Python. IMP also provides interfaces for developers to introduce new scoring functions, sampling schemes, analysis methods, model representations and integrative modeling applications [23].

To start integrative modeling, all relevant data from different lines of experimental, physical, bioinformatics, and statistical studies have to be pooled together for close examination. Upon a proper choice of the resolution with which the system of interest will be defined in the model, the applicable data that were collected in the first stage would have to be translated into spatial restraints on part or all of the system. For example, a residue–residue contact can be incorporated by applying a harmonic constraint on the distance between these two residues, and a cryo-EM density map can be used to generate a 3D-grid based function to bias the system being modeled to evolve toward it. To sample these constrained functions all together, various methods can then be applied, such as molecular dynamics (MD), Monte Carlo (MC), Brownian dynamics, and docking. In the end, an ensemble of models is generated for analysis and refinement toward a final model. Recent successes in implementing integrative modeling include a variety of systems, utilizing experimental data from X-ray, NMR, cryo-EM and SAXS [20]. These successes have contributed many innovative insights into biomolecular assemblies, and generated much interest in the approach. Karca et al. have comprehensively reviewed how different types of experimental data can be translated into restraints, suggesting four categories of restraints e.g. binding sites, distance, orientation, and shape, operating at a high level of abstraction [21]. When no high-resolution experimental structure (or structures from closely homologous organisms) are available, cryo-EM maps can still be used for secondary structure element identification using computational tools such as SSHunter [24], ab initio protein modeling using EM-fold [25], de novo protein structure prediction using RosettaCM [26,27]. In this review we concentrate on cryo-EM- and SAXS-based integrative modeling using atomistic MD simulation.

DNA replication and repair are fundamentally important biological processes and involve multiple protein-DNA complexes. The detailed structures of many of these complexes, however, are difficult to obtain through X-ray or NMR studies, due to their large size and intrinsic flexibility. Meanwhile, a great number of related experimental results, including X-ray crystal structures, biochemistry and biophysical signatures of various components, are accessible. This extensive body of information provides a favorable scenario in which to apply the integrative modeling approach. The modeling of the human Rad9–Hus1–Rad1/FEN1/DNA ternary complex [28] is reviewed here to illustrate the MDFF method [29] utilizing a negative stain EM density map. Other applications, in which the conformational space of ubiquitinated and/or SUMOylated Proliferating Cell Nuclear Antigen (PCNA) is explored, are also presented as a guide to incorporating experimental SAXS data into a hybrid modeling protocol [30,31].

2. Methods

2.1. Molecular Dynamics Flexible Fitting

Although the resolution of current cryo-EM methodology is generally not comparable to that of X-ray crystallography [1], cryo-EM is routinely capable of providing coarse structural information on macromolecular complexes, and in a biologically more realistic environment, perhaps even capturing different functional states [32]. Combining atomistic detail from crystal structures with a cryo-EM density map provides complementarity and enhances the model construct that might be deduced from each set of data alone. Methods developed for fitting atomic structures into cryo-EM maps can be divided generally into rigid-body docking and flexible fitting. Rigid-body docking (also often called rigid-body fitting), refers to the process of placing the atomic