Methods 59 (2013) 372–381

Contents lists available at SciVerse ScienceDirect

Methods

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

Advances in integrative modeling of biomolecular complexes

Ezgi Karaca, Alexandre M.J.J. Bonvin *

Bijvoet Center for Biomolecular Research, Faculty of Science - Chemistry, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

ARTICLE INFO

Article history:
Available online 23 December 2012
Communicated by Peter Schuck

Keywords:
Hybrid methods
Restraints
Information-driven docking
HADDOCK
Computational structural biology

ABSTRACT

High-resolution structural information is needed in order to unveil the underlying mechanistic of biomolecular function. Due to the technical limitations or the nature of the underlying complexes, acquiring atomic resolution information is difficult for many challenging systems, while, often, low-resolution biochemical or biophysical data can still be obtained. To make best use of all the available information and shed light on these challenging systems, integrative computational tools are required that can judiciously combine and accurately translate sparse experimental data into structural information. In this review we discuss the current state of integrative approaches, the challenges they are confronting and the advances made regarding those challenges. Recent developments are underpinned by noteworthy application examples taken from the literature. Within this context, we also position our data-driven docking approach, HADDOCK that can integrate a variety of information sources to drive the modeling of biomolecular complexes. Only a synergistic combination of experiment and modeling will allow us to tackle the challenges of adding the structural dimension to interactomes, shed "atomic" light onto molecular processes and understand the underlying mechanistic of biomolecular function. The current state of integrative approaches indicates that they are poised to take those challenges.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Proteins and their intricate network of interactions are the mainstay of any cellular process. Dissecting their interaction networks at atomic detail is therefore invaluable, as this will pave the route to a mechanistic understanding of biological function. Atomic detail (high-resolution) information about structure and dynamics of biomolecular complexes is typically acquired by classical experimental methods such as X-ray crystallography and NMR spectroscopy. Compared to other structural biology methods, these are the most accurate ones. They are, however, faced with many challenges, especially when the macromolecular systems under study become very large, comprise flexible or unstructured regions, exist in very tiny amounts, are membrane associated, or when their constituents interact only transiently. In the last decade another method, cryo-EM has emerged as a promising alternative for (high-resolution) structure determination. Its advantage over classical techniques is that it does not require high sample concentration [1,2], leading routinely to medium resolutions in the 8–20 Å range [3]. But rarely the resolution gets better than 8–6 Å, which could only be obtained so far for highly symmetric and stable complexes [4–6].

The number of known 3D structures of macromolecular complexes is considerably smaller than the amount of documented protein–protein interaction data [7,8]. Technical limitations of high-resolution methods and other problems mentioned above hamper closing this growing gap in a rapid manner. As a rescue strategy, structural biologists often resort to using different types of biochemical and biophysical experiments that can quickly provide accurate low-resolution information even for challenging systems. Most of the time, however, the collected data are rather sparse and/or of limited information content. These limitations call for integrative computational tools, like for example docking, that can, using some kind of physical model, judiciously combine and accurately translate sparse experimental data into structural information [9–11].

Currently, integrative modeling is the best strategy when conventional structural methods fail. Using such an integrative approach should reduce the downside features of both experimentation and modeling. From an experimentalist point of view, integrative modeling is beneficial since it can generate new hypothesis to drive experiments, which can significantly speed up the structure determination process and/or increase our understanding of biological function [10–12]. It is also advantageous for modelers, as incorporating experimental data into the modeling

Abbreviations: 3D, Three-Dimensional; AFM, Atomic Force Microscopy; CCS, Collision Cross Section; EM, Electron Microscopy; EPR, Electron Paramagnetic Resonance; FRET, Förster resonance energy transfer; IM, Ion Mobility; RMSD, Root Mean Square Deviation; MD, Molecular Dynamics; MS, Mass Spectrometry; NMR, Nuclear Magnetic Resonance; NOE, Nuclear Overhauser Effect; PCS, Pseudocontact Shifts; PRE, Paramagnetic Relaxation Enhancement; SA, Simulated Annealing; STEM, Scanning Transmission Electron Microscopy.

* Corresponding author. Fax: +31 (0)30 2537623.
E-mail address: a.m.j.j.bonvin@uu.nl (A.M.J.J. Bonvin).

1046-2023/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ymeth.2012.12.004

can accelerate the computational search and greatly help to overcome the shortcomings of *ab initio* modeling, such as high rates of false positives and difficulties in assessing the accuracy of the generated models [13,14].

Integrative methods have most of the time been developed with the drive of dissecting a specific system. Recent examples include successful characterization of a wide range of challenging systems, varying from flexible dimers to whole cells, based on different combinations of X-ray, NMR, cryo-EM, Electron Tomography and SAXS data [15–18]. All of these are important milestones in the field of integrative modeling, however, being mainly application-oriented or system-specific, their general applicability still has to be demonstrated [17]. There is a small number of generic integrative modeling approaches and these are the main focus of this review. We discuss them in detail in the following sections. In the final section, we concentrate on our data-driven docking approach, HADDOCK, and position it within the current state of generic integrative modeling methods by presenting some application examples.

2. Translating sparse data into 3D structures

2.1. Sources of low-resolution information

There are various types of biophysical and biochemical experimental techniques that can quickly provide low-resolution structural information. Assuming that the stoichiometry and composition of the macromolecular complex is known, these can provide useful insight into binding sites, distances between specific pair or groups of atoms, orientation between molecules and/or globular shape of the complex. The most frequently used data and their information content are summarized in Table 1.

Chemical Shift Perturbation (CSP), Hydrogen/Deuterium (H/D) exchange, solvent Paramagnetic Relaxation Enhancement (PRE) and chemical footprinting experiments provide information about interacting surfaces [11,12,17,19]. They all determine the binding site based on the alteration of the environment upon complexation. CSP measures the chemical environment changes induced by ligand binding [20–22]. H/D exchange is conducted by monitoring the exchange of labile hydrogens with deuteriums, so that changes in surface accessibility can be detected [23,24]. Solvent PREs are measured by using chemically inert paramagnetic probes as co-solvents that cause relaxation and thus signal attenuation of solvent accessible protons [25]. In chemical footprinting, the non-interacting surface of the complex is exposed to chemical modification, leaving the binding site unaffected [26]. Mutagenesis allows to identify specific residues that are critical for binding [12,27]. Next to those methods, bioinformatics approaches based on sequence/structure conservation [28], comparative patch analysis [29], correlated mutation studies [30], possibly combined with information about surface properties (e.g. curvature, hydrophobicity, charge) [31], can also be used to predict binding sites. All these approaches are built on the idea that conservation of sequence, contacts, patches or a globular structural element can possibly depict a probable interaction site [11,32,33].

Short-range distances between pair of atoms can be obtained by NOE measurements (<5–6 Å) [34,35], which are used together with dihedral angle restraints derived from J-couplings measurements or from chemical shifts analysis in conventional NMR structure calculation [36]. Chemical cross-linking experiments provide another source of distance information [37,38]. In these, functional groups on the surface of biomolecules are cross-linked using reactive chemicals. Residues are cross-linkable, if they are in close proximity and have chemical properties (e.g. Lys side-chains) allowing them to bind covalently to the cross-linking agent. They are usually identified by MS. The measured distance ranges depend on the cross-linker size and flexibility [39,40]. In the presence of paramagnetic ions (e.g. substituted in a metal binding site or attached to the protein via a tag), PRE [15], Pseudocontact Shifts (PCS) NMR [41] measurements or EPR experiments [42] can help to identify long-range distances up to 20–40 Å, depending in the paramagnetic species used and even 80 Å for EPR measurements. PCS, in addition, also contain orientational information. These effects are observed due to magnetic dipolar interactions between nucleus and the unpaired electrons of the paramagnetic center [17,43,44]. FRET experiments provide another source of long-range distance information: the measured distances depend on the separation of the fluorescently labeled residues of the complex [45–47].

Information on the relative orientation of two molecules can be obtained by Residual Dipolar Coupling (RDC) [48] or NMR Relaxation experiments [49]. In conventional NMR structure calculations, this orientational information is often combined with binding site and distance information, from CSP's and NOE's, respectively [19,35]. Lately it has also been frequently used with shape data from SAXS experiments, in order to reduce the inherent degeneracy entailed by the low-resolution shapes [50,51]. Low-resolution shape information can be obtained from SAXS and cryo-EM experiments. SAXS experiments measure the scattering intensity at very low angles, which can be translated into a low-resolution 3D envelope [52]. In addition, the radius of gyration (Rg) of a complex can be extracted from the SAXS data, which is an indicator of the structure compactness [53]. Cryo-EM experiments provide an electron density map with a resolution range typically between 8 Å and 20 Å [1,3]. The molecular maps extracted from cryo-EM experiments are especially useful when the individual structures of a complex's constituents are known, since these can then be fitted into those maps [54]. Finally, IM-MS experiments also provide shape-related information in the form of Collision Cross Sections (CCS). The CCS corresponds to the rotationally-averaged molecular area to which the buffer gas can collide; it offers thus information on the overall size and conformation of the complex [55–57].

For further information and a more detailed description of the various types of experimental data, please refer to the comprehensive review of *Melquiond* and *Bonvin* on data-driven modeling [12].

Table 1
Sources of low-resolution data, classified based on their information content.

	Data type	Experimental technique
Binding site	Chemical shift perturbations[a]	NMR
	H/D exchange[a]	NMR, MS
	Solvent PRE[a]	NMR
	Mutagenesis[a]	Biochemistry
	Chemical footprinting[a]	Biochemistry
	Conservation (correlated mutations, comparative patch analysis)[a]	Bioinformatics predictions
Distance	NOE distances[b]	NMR
	Chemical cross-links[b]	MS
	PRE[b]	NMR
	Correlated mutations[c]	Biochemistry
	PCS[b]	NMR
	Distances (distribution)[d]	FRET
	Distances from EPR[d]	EPR
Orientation	Residual dipolar couplings[e]	NMR
	Relaxation anisotropy[e]	NMR
Shape	Collision Cross Section[f]	IM-MS
	Radius of gyration[f]	SAXS
	Molecular envelope, globular shape[f]	SAXS, cryo-EM

Resolution/ambiguity level for a given source of information:
[a] Residue.
[b] Atom–atom (separation).
[c] Residue–residue (separation).
[d] Label–label (separation).
[e] Inter-monomer and/or bond vector orientations.
[f] Biomolecular complex.

Download English Version:

https://daneshyari.com/en/article/10826050

Download Persian Version:

https://daneshyari.com/article/10826050

Daneshyari.com