Progress in Biophysics and Molecular Biology xxx (2014) 1–10

Contents lists available at ScienceDirect

Progress in Biophysics and Molecular Biology

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

Original research

An overview of recent advances in structural bioinformatics of protein–protein interactions and a guide to their principles

Govindarajan Sudha [a], Ruth Nussinov [b, c, **], Narayanaswamy Srinivasan [a, *]

[a] Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India
[b] Cancer and Inflammation Program, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., National Cancer Institute, Frederick, MD 21702, USA
[c] Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Protein–protein interactions
Protein–protein complexes
Homomeric proteins
Oligomeric proteins
Oligomers
Structure
Evolution
Interaction
Function
Conformation

ABSTRACT

Rich data bearing on the structural and evolutionary principles of protein–protein interactions are paving the way to a better understanding of the regulation of function in the cell. This is particularly the case when these interactions are considered in the framework of key pathways. Knowledge of the interactions may provide insights into the mechanisms of crucial 'driver' mutations in oncogenesis. They also provide the foundation toward the design of protein–protein interfaces and inhibitors that can abrogate their formation or enhance them. The main features to learn from known 3-D structures of protein–protein complexes and the extensive literature which analyzes them computationally and experimentally include the interaction details which permit undertaking structure-based drug discovery, the evolution of complexes and their interactions, the consequences of alterations such as post-translational modifications, ligand binding, disease causing mutations, host pathogen interactions, oligomerization, aggregation and the roles of disorder, dynamics, allostery and more to the protein and the cell. This review highlights some of the recent advances in these areas, including design, inhibition and prediction of protein–protein complexes. The field is broad, and much work has been carried out in these areas, making it challenging to cover it in its entirety. Much of this is due to the fast increase in the number of molecules whose structures have been determined experimentally and the vast increase in computational power. Here we provide a concise overview.

1. The classical view of protein–protein interactions

Understanding biological systems requires detailed knowledge of cellular events at the detailed molecular level. This level includes the physical interactions between macromolecules such as DNA, RNA and proteins and between these and their environment, including lipids, ions and second messengers, such as cAMP. Here we focus on protein–protein interactions which are responsible for carrying out diverse processes in living systems. Structural and mechanistic features of protein–protein interactions may be best understood using the three-dimensional structures of the proteins and their complexes. The structural database provides rich data both of static crystal structures and their ensembles in solutions by NMR. Protein ensembles can also be glimpsed from collections of crystal structures of the same protein, however in different bound and unbound states and crystal forms. Even though the crystal environment captures only the state favored under specific crystallization conditions, these static structures still provide crucial information on the nature of the protein–protein interactions. A vast majority of heterocomplexes with known 3D structures are heterodimers (Fig. 1). Therefore, there is a need to study the 3D structures of higher order heteromers, which often form the functional multiprotein assemblies in the cell. Structural bioinformatics of protein–protein interactions, which deals with the analysis of known 3D structures, has provided detailed information on the underlying principles of structure, function and dysfunction, and evolution of protein–protein complexes.

Proteins that are stable only in a protein–protein complex form and remain together throughout their functional life time are

* Corresponding author. Cancer and Inflammation Program, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., National Cancer Institute, Frederick, MD 21702, USA.
** Corresponding author.
E-mail addresses: gsudha@mbu.iisc.ernet.in (G. Sudha), nussinor@helix.nih.gov (R. Nussinov), ns@mbu.iisc.ernet.in, srinivasan@ncbs.res.in (N. Srinivasan).

http://dx.doi.org/10.1016/j.pbiomolbio.2014.07.004
0079-6107/

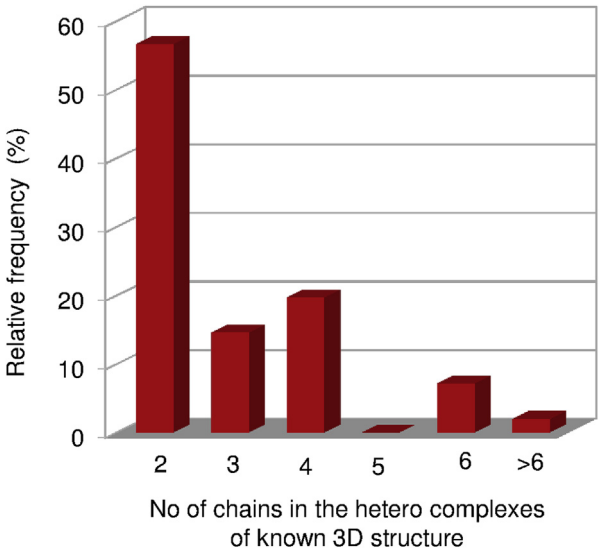

Fig. 1. Distribution of number of chaints in the heterocomplexes of known 3-D structure: The histogram shows the distribution of number of chaints in the heterocomplexes of known structure available in the Protein Data Bank (PDB (Berman et al., 2000)). The data to generate this figure corresponds to the number of chains in the biological units as presented in the PDB.

termed as 'permanent' protein–protein complexes. On their own, these proteins are typically disordered; that is, they exist in a range of conformational states, with none of these having a sufficiently stable conformation to be captured in crystalline form. Protein–-protein complexes that interact with their partner for a brief period of time to carry out a specific function and are stable in their free form are termed 'transient' (Nooren and Thornton, 2003a). On average, there are differences in the structures and chemical characteristics of interfaces between permanent and transient protein–protein complexes (De et al., 2005).

The evolution of the interfaces was suggested to be slower for permanent protein–protein complexes than for transient complexes (Mintseris and Weng, 2005). Transient protein–protein interfaces show higher residue conservation than rest of the tertiary structural surface (Choi et al., 2009; Mintseris and Weng, 2005; Valdar and Thornton, 2001). Physicochemical and geometrical characterization of protein interfaces have been extensively studied that are different from the rest of the surface (Jones and Thornton, 1996) (De et al., 2005; Jones et al., 2000; Lo Conte et al., 1999; Sonavane and Chakrabarti, 2008). Differences in interfacial features have also been observed between permanent and transient protein–protein complexes. Interface size (small interfaces in transient protein–protein complexes *versus* large interfaces in permanent complexes), area, polarity (polar interfaces in transient protein–protein complexes *versus* non-polar interfaces in permanent complexes), shape complementarity, conformational changes upon binding, residue interface propensities and residue contacts have served as distinguishing features to predict and classify permanent and transient protein–protein complexes (Ansari and Helms, 2005; Bahadur et al., 2003; Block et al., 2006; De et al., 2005; Jones and Thornton, 1996; Keskin et al., 2008; Levy and Pereira-Leal, 2008; Mintseris and Weng, 2003; Nooren and Thornton, 2003b; Zhu et al., 2006).

A protein–protein interface can be divided into core and rim which are buried in the interface and remain accessible to the solvent, respectively (Bahadur et al., 2003). Interestingly, the core and the rim differ in their amino acid composition and conservation (Janin et al., 2008). Another important approach to interface residue classification is based on the contributions to interaction energy. The subset of interface residues that serve as major contributors to binding energy in protein–protein interfaces (>2 kcal/mol) have been termed hot-spot residues (Bogan and Thorn, 1998).

Analysis of a large number of 3-D structures of protein–protein complexes revealed that, in general homologous protein–protein complex structures are conserved (Aloy et al., 2003). However, interfaces of distantly-related homologous proteins are usually not topologically equivalent (Rekha et al., 2005). Further detailed analysis showed that spatial orientations of interacting proteins with respect to each other in some of the homologous protein–-protein complexes differ (Kim et al., 2006). Studies also showed that interactions between proteins could often be predicted successfully if the proteins have high sequence similarities with proteins, which are known to interact with each other (Levy and Pereira-Leal, 2008; Mika and Rost, 2006). Studies also showed that structurally similar interfaces can bind proteins with different binding site structures and different functions (Tsai et al., 1996). This is accommodated through conserved interactions at similar interface locations, despite having different partners (Keskin and Nussinov, 2007). Even if the overall structures of the interacting chains are different, interface similarity may exist (Keskin and Nussinov, 2005). While protein–protein interfaces are typically highly specific, there appear to be proteins with 'promiscuous' binding characteristics (Schreiber and Keating, 2011). One way to achieve specificity is by utilizing different hotspot residues in the protein interfaces (Gretes et al., 2009). Alternatively, different conformations in the ensemble may be selected (Ma et al., 1999; Tsai et al., 1999a, b). Clusters of interacting residues have been observed in protein–protein interfaces and cooperative interactions between residues in a cluster generate binding affinity and specificity (Reichmann et al., 2005). Below, we briefly discuss recent and emerging views in structural bioinformatics of protein–protein interactions.

2. Recent and emerging views on protein–protein interactions

2.1. Protein–protein complexes are multifaceted

A grasp of the structural and evolutionary principles of protein–protein interactions is essential to understand the roles of proteins in the cell. Degeneracy is observed not only at the level of protein folds but also at the level of protein–protein interface structures. This is due to the structural constraints of packing of secondary structural elements at the interface and functional constraints (Gao and Skolnick, 2010). Using available 3-D structures of protein–protein complexes, interfaces have been clustered and it was proposed that the repertoire of structures of interfaces is limited (Cukuroglu et al., 2014). However, surprisingly the conservation of interfaces in evolutionarily-related protein–protein complexes does not always take place (Zhang et al., 2010), which suggests that interfaces are tuned for specific interactions, which then lead to specific cellular pathways. Alternate binding modes in homologous protein–protein complexes have been observed, with the interfaces not entirely topologically equivalent (Fig. 2). (Hamp and Rost, 2012; Kundrotas and Vakser, 2013), and there are examples of proteins which can bind to different proteins with non-equivalent locations (Martin, 2010). There seems to be evolutionary 'plasticity' in homologous protein–protein interfaces which are manifested as different types of interface contacts especially those involving polar residues (Andreani et al., 2012). 'Plasticity' reflects the presence of proteins as conformational ensembles, with different conformations being selected followed by minor induced fit optimization (Csermely et al., 2010). At the same time, we also

Please cite this article in press as: Sudha, G., et al., An overview of recent advances in structural bioinformatics of protein–protein interactions and a guide to their principles, Progress in Biophysics and Molecular Biology (2014), http://dx.doi.org/10.1016/j.pbiomolbio.2014.07.004