Progress in Biophysics and Molecular Biology 128 (2017) 33–38

Contents lists available at ScienceDirect

Progress in Biophysics and Molecular Biology

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

Integrating computational methods and experimental data for understanding the recognition mechanism and binding affinity of protein-protein complexes

M. Michael Gromiha*, K. Yugandhar

Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, Tamilnadu, India

ARTICLE INFO

Article history:
Received 30 March 2016
Received in revised form
4 January 2017
Accepted 5 January 2017
Available online 7 January 2017

Keywords:
Binding affinity
Protein-protein complex
Recognition mechanism
Binding sites
Non-covalent interactions

ABSTRACT

Protein-protein interactions perform several functions inside the cell. Understanding the recognition mechanism and binding affinity of protein-protein complexes is a challenging problem in experimental and computational biology. In this review, we focus on two aspects (i) understanding the recognition mechanism and (ii) predicting the binding affinity. The first part deals with computational techniques for identifying the binding site residues and the contribution of important interactions for understanding the recognition mechanism of protein-protein complexes in comparison with experimental observations. The second part is devoted to the methods developed for discriminating high and low affinity complexes, and predicting the binding affinity of protein-protein complexes using three-dimensional structural information and just from the amino acid sequence. The overall view enhances our understanding of the integration of experimental data and computational methods, recognition mechanism of protein-protein complexes and the binding affinity.

Contents

1. Introduction

Protein-protein interactions (PPIs) play important roles in most of the cellular processes in life. Protein-protein complexes perform

* Corresponding author.
E-mail address: gromiha@iitm.ac.in (M.M. Gromiha).

http://dx.doi.org/10.1016/j.pbiomolbio.2017.01.001
0079-6107/

diverse functions such as receptors, signaling, molecular switching, ubiquitination and so on (Huang et al., 1999; Hage et al., 1999). The mechanism of protein-protein recognition at molecular level has been investigated through experiments such as surface plasmon resonance (SPR), isothermal titration calorimetry (ITC), fluorescence spectroscopy, spectrophotometric assays, radio ligand binding, and stopped-flow fluorimetry as well as computational methods. These approaches provide information on important interactions influencing the affinity of protein-protein complexes, binding site residues at the interface and thermodynamic parameters for understanding the recognition mechanism (Sudha et al., 2014). On the other hand, PPIs have been studied through construction and large-scale analysis of their interaction networks (Yugandhar and Gromiha, 2016).

The availability of protein-protein complex structures in Protein Data Bank (Rose et al., 2015) enabled researchers to analyze the binding site residues based on their physicochemical, energetic and conformational properties, which are utilized to develop methods for predicting the binding sites as well as understanding the recognition mechanism (Nooren and Thornton, 2003). Goncearenco et al. (2015) investigated the evolutionary roots of binding sites and showed the existence of universal common ancestor in all cellular organisms. On the other hand, the available thermodynamic data on binding such as the dissociation constant and the free energy change upon binding are utilized to relate structural features to understand thermodynamics underlying the binding specificity of protein-protein complexes (Vangone and Bonvin, 2015).

In this review, we focus on two major aspects of PPI studies i.e., (i) influence of important structural features and interactions for understanding the recognition mechanism and (ii) computational methods to relate the experimental binding affinity of protein-protein complexes. In the first part, we survey the importance of sequence and structural features along with non-covalent interactions for the binding of protein-protein complexes and understanding the recognition mechanism. The second part deals with the binding affinity of protein-protein complexes, by relating the affinity with structure based parameters and predicting the affinity from amino acid sequence.

2. Identification of binding site residues

The binding site residues in protein-protein complexes can be directly identified from the three-dimensional structures of protein-protein complexes based on three different aspects (Gromiha, 2010): (i) distance based criterion, (ii) solvent accessibility of residues in free protein and complex and (iii) energy based approach.

2.1. Distance based criterion

This approach computes the distance between all the atoms in the interacting proteins from their X, Y and Z coordinates. In a protein, if the distance between any of the atoms in a residue is less than a cutoff of 4-6 Å (Keskin et al., 2004) with any atom in the partner protein, then the residue is considered as a binding site residue.

2.2. Solvent accessibility of residues upon binding

Accessible surface area (ASA) is calculated by rolling a probe (typically a water molecule with radius 1.4 Å) on the surface of a protein. In a complex, ASA is computed for a residue both in the free and bound forms. If a residue has a difference of ASA >0.1 $Å^2$ between the free protein and the complex, it is identified as a binding site residue (Bahadur et al., 2004).

2.3. Energy based approach

Energy based approach utilizes the interaction energy computed with AMBER potential between all atoms in the pair of proteins. The contribution from the atoms in a residue is summed up to obtain the interaction energy of a residue. Residues which have an interaction energy of less than 1 kcal/mol are treated as binding site residues (Gromiha et al., 2009). The databases, which contain the information about interacting residues, are listed in Yughandar and Gromiha (2017).

2.4. Prediction of binding site residues from three-dimensional structure of a free protein or amino acid sequence

In the absence of complex structures, several methods have been proposed to predict the binding sites from the structure of a free protein or just from the amino acid sequence (Maheshwari and Brylinski, 2015; Yughandar and Gromiha, 2017). The structure based methods mainly utilize surface patches (Jones and Thornton, 1997), solvent accessibility (Chen and Zhou, 2005), residue propensity (Neuvirth et al., 2004), local structural similarity (Jordan et al., 2012), empirical scoring functions (Liang et al., 2006) and energy landscapes (Fernandez-Recio et al., 2004) for identifying the binding sites.

The sequence based methods are based on the occurrence of proline at the flanking segments of interaction sites (Kini and Evans, 1996), hydrophobic moment of sequence stretches (Gallet et al., 2000), phylogenetic motifs (La and Kihara, 2012), multiple sequence alignment and conservation of residues (Shulman-Pelag et al., 2008), position specific scoring matrices (Ofran and Rost, 2007) and predicted secondary structure and solvent accessibility (Ofran and Rost, 2007). Further, a few methods have been developed by considering the sequence of both partners in a complex utilizing conservation score and contact propensity (Ahmad and Mizuguchi, 2011). Although the methods based on PSSMs show better performance than other methods, they are often time consuming and the prediction results mainly depend on the database used for the alignment search. A list of available methods for predicting the binding sites is given in Yugandhar and Gromiha (2017). Further, several methods have been developed for predicting the three-dimensional structures of protein-protein complexes using their unbound proteins, and the details are reviewed in Gromiha et al. (2016).

3. Important features influencing the recognition of protein-protein complexes

Various investigations have been carried out to understand the recognition mechanism of protein-protein complexes such as secondary structure, solvent accessibility, conservation of amino acid residues, non-covalent interactions, binding propensity, contributions due to main chain and side chain atoms etc.

3.1. Sequence and structure based features

Ofran and Rost (2003) showed that the interface residues are dominated by specific amino acid residues and amino acid composition is a good feature for predicting the binding sites. These binding sites are shown to be highly conserved at the interface (Ma et al., 2003; Guharoy and Chakrabarti, 2005). Furthermore, PPIs have been studied in terms of efficient clustering (Aung et al., 2008), conformational changes and docking simulations (Lensink and Méndez, 2008), energetic contributions (Zhu et al., 2008) and so on. Recently, Saranya et al. (2016) analyzed the importance of secondary structure and conformational changes in protein-protein complexes using three-dimensional structures. In addition, the