Progress in Biophysics and Molecular Biology xxx (2014) 1–9

Contents lists available at ScienceDirect

Progress in Biophysics and Molecular Biology

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

Original research

Prediction and redesign of protein–protein interactions

Rhonald C. Lua [a], David C. Marciano [a], Panagiotis Katsonis [a], Anbu K. Adikesavan [a], Angela D. Wilkins [a, c], Olivier Lichtarge [a, b, c, *]

[a] Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA
[b] Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX 77030, USA
[c] Computational and Integrative Biomedical Research Center, Baylor College of Medicine, Houston, TX 77030, USA

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Evolutionary trace
Protein–protein interaction networks
Functional sites
Molecular evolution
Functional annotation
Protein design

ABSTRACT

Understanding the molecular basis of protein function remains a central goal of biology, with the hope to elucidate the role of human genes in health and in disease, and to rationally design therapies through targeted molecular perturbations. We review here some of the computational techniques and resources available for characterizing a critical aspect of protein function – those mediated by protein–protein interactions (PPI). We describe several applications and recent successes of the Evolutionary Trace (ET) in identifying molecular events and shapes that underlie protein function and specificity in both eukaryotes and prokaryotes. ET is a part of analytical approaches based on the successes and failures of evolution that enable the rational control of PPI.

1. Introduction

Protein–protein interactions (PPI) are ubiquitous and underlie the assembly of macromolecular machines, mediate signaling pathways in cellular networks (Jones and Thornton, 1996; Schadt, 2009; Yamada and Bork, 2009) and control cell-to-cell communication (Sattentau and Moore, 1993). Thus, characterizing PPI sites is an essential step towards: 1) identifying drug targets and mimicking them to create potential therapeutics (Moellering et al., 2009; Wells and McClendon, 2007), 2) engineering and modifying protein activity (Zorn and Wells, 2010), and 3) interpreting the impact of allelic variations. To achieve these goals, a key step is to identify which proteins bind together and which amino acids mediate binding affinity and interaction specificity.

PPIs are diverse but they can be classified broadly into permanent (obligate, oligomeric, tight, more stable) and transient (relatively weak) interactions, which present different levels of co-expression and co-evolution (Jones and Thornton, 1996; Nooren and Thornton, 2003a,b; Mintseris and Weng, 2005; La et al., 2013). The interacting proteins of permanent complexes display higher tendencies to be co-expressed and co-localized than transiently-interacting proteins (Brown and Jurisica, 2007). Moreover, the amino acid residues that mediate permanent interactions evolve at a slower rate (they are more conserved) than residues involved in transient interactions (Mintseris and Weng, 2005). Therefore, it is expected to be more challenging to recognize the transient than the permanent interactions, because of the weaker signs of evolutionary pressure across the interface. However, this difficulty is mitigated by the fact that even the interface residues of the weak transient homodimers are generally more conserved than surface residues (Nooren and Thornton, 2003a,b).

A possible third category of PPI (or a special subcategory of transient interactions) is one mediated by disordered regions or unstructured segments of proteins (Babu et al., 2012; Tompa and Fuxreiter, 2008; Brown et al., 2011). Most eukaryotic proteins have both structured and disordered regions (Babu et al., 2012), and some proteins, such as the tumor suppressor p53, use disorder to achieve conformational heterogeneity and bind to a multitude of different protein partners (Hsu et al., 2012). Alternative splicing events and posttranslational modifications, which commonly occur in disordered regions, also contribute to distinct partner binding (Hsu et al., 2013). Disordered proteins generally evolve more rapidly than ordered proteins (Brown et al., 2011), although functional residues have been detected in disordered regions using evolutionary conservation (Nguyen Ba et al., 2012). The structured protein partner of the disordered protein showed substantially

* Corresponding author. Baylor College of Medicine, Department of Molecular and Human Genetics, One Baylor Plaza, BCM225, Houston, TX 77030, USA. Tel.: +1 713 798 5646.
E-mail addresses: lichtarge@bcm.edu, sweaver@bcm.edu (O. Lichtarge).

http://dx.doi.org/10.1016/j.pbiomolbio.2014.05.004
0079-6107/

higher conservation in binding regions as compared to non-binding regions (Hsu et al., 2012).

Direct elucidation of protein–protein interfaces comes from crystallographic and NMR complexes (Berman et al., 2000) and are compiled into catalogs of interfaces (Schlessinger et al., 2006; Huang and Honda, 2006; de Beer et al., 2014; Krissinel and Henrick, 2007). Mutational studies can probe the energetic, catalytic or dynamic role of interface residues (Jackson and Fersht, 1993), but only a small fraction of them has been studied, due to the high cost, length of time required, and the growing number of interfaces (Hall, 2007; Liolios et al., 2006). Therefore, it is imperative to invent, refine, and update computational techniques that bypass experimental mutational assays by analyzing sequence and structure information.

In this review, we discuss how computational approaches can help in identifying PPI (See Fig. 1). Specifically, Section **2.1** focuses on predicting which proteins associate functionally or interact physically. Section **2.2** shows how to predict binding sites on protein surfaces, by using protein sequences and evolution data. Section **2.3** moves on to predicting pairs of interacting residues using new promising methods that rely on coevolution principles. Section **2.4** discusses *ab initio* and template-based docking, and how it can benefit from predictions of binding sites and interacting residues described in Sections **2.2** and **2.3**. Section **2.5**, in turn, focuses on narrowing predictions of specificity and binding determinants to the ones relevant to the studied interaction. Section **3** illustrates recent applications of using such computational methods to identify, modulate and inhibit PPIs. The main application case focuses on the attempts to solve the puzzle of the long sought RecA–LexA PPI sites.

2. Current methods for PPIs

2.1. Finding and establishing links between proteins ("Which proteins interact?")

In order to characterize protein–protein interfaces, the knowledge of which proteins physically interact is critical. Computational biology often transfers functional information from well-understood proteins to lesser-known ones using the concept of homology (Tatusov et al., 1997). Similarity searches (Mount, 2007) or shared domains (Aloy and Russell, 2006) can point to proteins in which the query of interest likely shares similar binding partners. This has become a common practice and has been applied in organizing PPI networks (Brown and Jurisica, 2007; Huang et al., 2004; Persico et al., 2005). However, homology transfer can be unreliable for interactions in phylogenetically distant species and should be used carefully (Lewis et al., 2012).

A complementary approach is to identify the proteins that are concurrently present or absent across large numbers of species. This co-occurrence inferred from phylogenetic profiling suggests a biological connection (Pellegrini et al., 1999; Tatusov et al., 2000; Schneider et al., 2013). The similarity of phylogenetic profiles can be assessed by assigning to each protein a vector encoding the patterns of presence or absence of that protein throughout many species. By finding matching or similar vectors, we can hypothesize which proteins interact. The resolution is expected to be low because disentangling physical and functional associations can be problematic (Kensche et al., 2008) but, in conjunction with multiple types of data, this approach can be useful (Snel and Huynen, 2004; Kim and Subramaniam, 2006). Gene co-expression is used in a

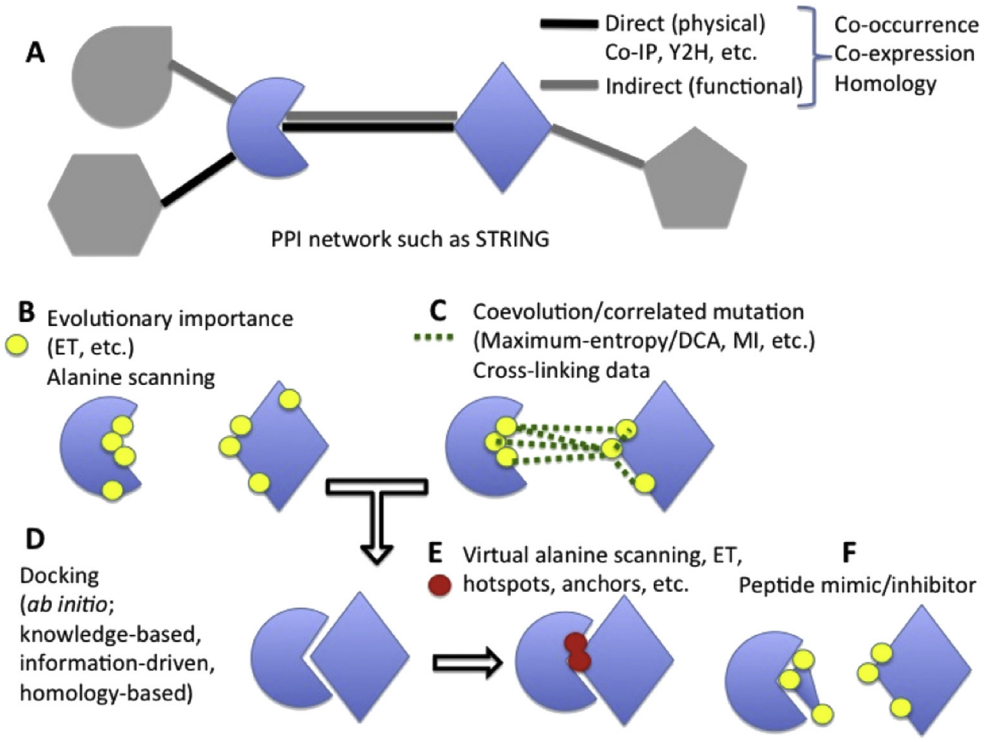

Fig. 1. Computational characterization of PPI that also serves as an outline for much of this article. **A.** Databases of PPI networks allow us to answer the question "Which proteins interact?" directly or functionally. **B.** Computational predictions of interacting/binding sites on one protein surface from sequence analysis using ET (yellow circles) can complement experimental data such as alanine scanning mutagenesis. **C**. Coevolution from simultaneous analysis of two protein partners known to directly interact can complement biochemical and cross-linking data. Dashed lines indicate coevolving or significantly correlated residues. **D**. Prediction of PPI interfaces by docking. **E.** Knowledge of the protein–protein complex enables refinement of computational predictions of key residues (red circles). **F**. Design of peptide to mimic and inhibit native interaction. Key binding and specificity determinants (yellow circles) are built into a scaffold that enhances other desirable features such as solubility or helical propensity. Note that characterization of PPI is not necessarily followed in the sequential order **A**–**F**.

Please cite this article in press as: Lua, R.C., et al., Prediction and redesign of protein–protein interactions, Progress in Biophysics and Molecular Biology (2014), http://dx.doi.org/10.1016/j.pbiomolbio.2014.05.004