Archives of Biochemistry and Biophysics 554 (2014) 65–75

Contents lists available at ScienceDirect

Archives of Biochemistry and Biophysics

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

Review

# Evolution of protein interactions: From interactomes to interfaces

Jessica Andreani *, Raphael Guerois *

*CEA, iBiTecS, F-91191 Gif sur Yvette, France*
*Université Paris-Sud & CNRS, UMR 8221, F-91191 Gif sur Yvette, France*

ARTICLE INFO

*Article history:*
Received 10 February 2014
and in revised form 28 April 2014
Available online 20 May 2014

*Keywords:*
Co-evolution
Protein–protein interaction
Interactome
Interface
Computational methods
Bioinformatics
Prediction
Docking

ABSTRACT

Protein–protein interactions lie at the heart of most cellular processes. Many experimental and computational studies aim to deepen our understanding of these interactions and improve our capacity to predict them. In this respect, the evolutionary perspective is most interesting, since the preservation of structure and function puts constraints on the evolution of proteins and their interactions. However, uncovering these constraints remains a challenge, and the description and detection of evolutionary signals in protein–protein interactions is currently a very active field of research. Here, we review recent works dissecting the mechanisms of protein–protein interaction evolution and exploring how to use evolutionary information to predict interactions, both at the global level of the interactome and at the detailed level of protein–protein interfaces. We first present to what extent protein–protein interactions are found to be conserved within interactomes and which properties can influence their conservation. We then discuss the evolutionary and co-evolutionary pressures applied on protein–protein interfaces. Finally, we describe how the computational prediction of interfaces can benefit from evolutionary inputs.


## Introduction

The cell is a crowded environment where many proteins, nucleic acids and small molecules, keep encountering each other [31,101,177] and interact in specific ways to perform a wide range of biological functions. In particular, protein–protein interactions (PPIs)[1] are involved in many, if not all, cellular processes and are thus a topic of major interest in order to understand the complexity and diversity of living systems.

Many approaches have been designed to tackle this question and a number of databases have been developed, providing a wealth of insights about protein–protein interactions, their evolution and their organization (Table 1). Two main fields of action can be distinguished: experimental characterization of interactions on various scales and computational methods. The former keep generating large quantities of data and the latter are needed both to help towards a better interpretation of this abundant material and to complement experiments by predictions. The overall PPI landscape is extremely complex: the estimated number of human interactions ranges from 130,000 [158] to 650,000 [144], while the number of experimentally identified interactions gathered from several of the largest databases lies around 60,000 [174] and might be overestimated due to errors and biases in the reported interactions [20,50,134,158,159]. Protein–protein interactions occur at the scale of the crowded cell, where the set of possible non-specific interactions is theoretically much larger than the set of specific, biologically functional, desired interactions [52]. Promiscuous interactions are partly limited through spatial and temporal regulation [75,138] and maintaining functional interactions imposes constraints on the evolution of PPIs which occurs in these crowded conditions.

In these various respects, evolution of proteins and protein–protein interactions provides essential clues for a better understanding of PPIs. However, given the complexity and the long timescales of evolutionary mechanisms, signals reflecting the structural and functional constraints applied by evolutionary pressures remain difficult to detect. Our aim is to present a general overview of how the evolutionary perspective can help in understanding different levels of PPI network organization and in predicting PPIs by computational means.

Evolution of proteins is slow, because most amino acid substitutions (around 98%) are forbidden at any given time owing to their deleterious effects on protein structure, function or expression

* Address: CEA, iBiTecS, UMR 8221, F-91191 Gif sur Yvette, France.
*E-mail addresses:* jessica.andreani@mines.org (J. Andreani), guerois@cea.fr (R. Guerois).

[1] *Abbreviations used:* PPIs, protein–protein interactions; PRMs, peptide recognition modules; iRMSDs, interface root mean square deviations; TCS, two-component signaling; DCA, direct coupling analysis.

http://dx.doi.org/10.1016/j.abb.2014.05.010
0003-9861/

**Table 1**
Protein–protein interaction databases. Non-exhaustive list of protein–protein interaction databases covering the whole range of information levels, from functional associations to structural details and evolutionary information: (1) PPI without necessary physical or direct contact (genetic interaction, co-expression, co-localization); (2) PPI comprising direct physical contact and indirect interaction within the same macromolecular complex; (3) PPI involving a direct physical contact with information about the interface energetics; (4) PPI involving a direct physical contact with details of the interface structure; (5) PPI database for direct physical contacts with evolutionary information.

| PPI database | Presented information | Level(s) of detail | Most recent references |
|---|---|---|---|
| STRING | Predicted functional protein associations | 1,2 | [147] |
| BioGRID | Literature-curated genetic and protein interactions | 1,2 | [18] |
| IntAct | Curated molecular interaction database | 2 | [63] |
| MIPS | Molecular interaction database(initially for yeast, extended to mammals and plants) | 2 | [104] |
| MINT | Literature-based, curated PPIs | 2 | [90] |
| HPRD | PPIs and post-translational modifications | 2 | [44] |
| ASEdb | Interface hot spot prediction based on alanine scanning energetics | 3 | [151] |
| SKEMPI | Literature-based binding energy changes for complexes with known 3D structures | 3 | [107] |
| SCOWLP | Classification of high-resolution 3D complexes based on SCOP domains | 4 | [149] |
| PIBASE | Structurally defined interfaces between SCOP and CATH domains | 4 | [23] |
| Docking benchmark | PPI database with known 3D structures of bound complex and unbound components | 4 | [55] |
| 3did | Classification of domain-domain interactions based on PFAM | 4 | [111] |
| 3Dcomplex | Hierarchical classification of complexes (based on SCOP/PFAM) | 4 | [87] |
| IBIS | Inferred interaction sites from homologous partners | 4,5 | [139] |
| PRISM | Clustering of interfaces based on structure similarity and evolution | 4,5 | [67] |
| 3D interologs | 3D domain mapping against homologous interfaces with known 3D structure | 4,5 | [91] |
| ProtCID | Clustered interacting domains with known 3D structure based on PFAM | 4,5 | [167] |
| InterEvol | Non-redundant interfaces with known 3D structure, including structural interologs and ortholog sequence alignments | 4,5 | [34] |
| MMDB+VAST | Structural similarities between macromolecular complexes | 4,5 | [95] |
| iPfam | Protein family and domain interactions in the PDB | 4,5 | [36] |
| KBDOCK | Spatial classification of 3D protein domain family interactions | 4,5 | [43] |

[124]. However, over a long time (around 3.5 billion years since the last universal common ancestor), almost all positions stand a chance to undergo substitutions, following changes in other positions. As proteins evolve in rugged fitness landscapes, the tolerated substitutions at one moment depend heavily on whatever mutations have occurred previously, owing to possible compensations [124].

Evolution thus takes many detours, as illustrated by epistasis, which corresponds to non-additive interactions between mutations: the fitness effect of one mutation depends on the state of other loci. As a consequence, some mutations apparently neutral (with no immediate effect on fitness) at one point in evolution can actually be "permissive", i.e., allow the protein to tolerate subsequent mutations which would otherwise have been deleterious and which lead to differences in phenotype. The notion of epistasis has recently received a lot of attention and was proposed as a reconciliation between the neutral and selectionist theories of evolution [162]: within the framework of epistasis, neutral mutations prepare the ground for later selection and adaptation. At the molecular level, reconstruction of ancient proteins showed that epistasis can drive changes in ligand specificity [116]. Such examples point to epistasis as a major factor affecting long-term protein evolution, either at the system level or at the level of the elementary interfaces. Experimental evidence accumulates, illustrating how epistatic phenomena shaped the topology of interaction networks and the fine details of binding interfaces and many of the studies presented in this review can be interpreted through that general framework.

In this review, we will first illustrate the evolutionary properties of interactomes (PPI networks) and protein–protein interactions. Next, we will explore how evolutionary pressures can shape protein–protein interfaces, with a special focus on the structural dimension of interfaces. Finally, we will explain how evolutionary constraints can be used to improve the computational prediction of protein–protein interfaces.

## Evolution of interactomes and protein–protein interactions

*Are interactions overall conserved between interactomes?*

The question of how well protein–protein interactions are conserved between different species has received a lot of interest. Because many difficulties remain in the experimental detection and validation of interactome data, it would be especially interesting to know when we can confidently transfer a PPI from a species in which it has been confirmed to another species. The notion of "interologs" (pairs of interactions between homologous proteins, e.g. A–B and A′–B′ if A and A′, B and B′ are two pairs of homologous proteins) was first introduced by Vidal and co-workers in 2000 [163]. This notion is illustrated in Fig. 1 (panels A and B). Many studies have attempted to quantify the success rate for transferring interactions across species [10,12,13,40,88,100,105,126,128,140,157,163,170]. These studies show that protein–protein interactions are conserved to some extent, but that there is some "rewiring" between the proteins in the PPI networks, i.e., interactions are gained while others are lost during evolution. However, the quantitative conclusions of different studies can be somewhat diverse: for instance, estimated PPI rewiring rates span quite a large spectrum ranging from the slowest at $2.2.10^{-6}$ rewiring interactions per million years [140] to the fastest ones at nearly $10^{-3}$ [88].

Several causes can explain these differences in quantitative estimates. One major reason for discrepancies is linked to the datasets used by each study. The results can depend strongly on the date of the study, the species compared and the PPIs taken into consideration: the detected PPIs depend strongly on experimental techniques and conditions, as evidenced by the small coverage between different high-throughput studies [19].

Another example of how different studies can yield different estimates highlights other possible factors that can account for such variations: the fraction of human PPIs expected to be conserved between human and yeast ranges from close to zero [40]