Protein-protein Interfaces Integrated into Interaction Networks: Implications on Drug Design

Gozde Kar, Guray Kuzu, Ozlem Keskin and Attila Gursoy

Center for Computational Biology and Bioinformatics and College of Engineering, Koc University, Rumelifeneri Yolu, 34450 Sariyer Istanbul, Turkey

Abstract: The growing perception that diseases are often consequences of multiple molecular abnormalities rather than being the result of a single defect highlights the importance of network-centric view in therapeutic approaches. Protein interaction networks may contribute to understanding of disease, assist in drug design and discovery. Here, we review some recent advances in disease-associated protein interaction networks taking a structural approach. We first describe structural aspects of protein-protein interactions and properties of protein interfaces as related to drug design; we address protein interactions in a network perspective; in particular, we illustrate how integrating protein interfaces onto interaction networks can guide the identification of selective drug targets or drugs targeting multiple proteins in a network.

Keywords: Protein-protein interfaces, protein interaction networks, drug design.

1. INTRODUCTION

Protein-protein interactions have a key role in regulating many biological processes, cellular and signaling pathways. The dysfunction of these pathways due to the alterations in protein-protein interactions may lead to several diseases such as cancer and neurological disorders. Therefore, protein-protein interactions are widely considered as drug targets in disease states [1-3]. However, more recently, network-based approaches have gained importance in drug discovery with the comprehension that diseases are complex; a disease phenotype reflects several pathobiological processes that interact in a complex network [4]. The high interconnectivity in disease-associated networks suggests that it is better to target entire cellular pathways rather than single proteins [5].

Another point that makes the network-level endeavors valuable is the so-called polypharmacology, which has emerged as a new paradigm in drug discovery stating that one drug is more likely to bind to multiple distinct targets [6]. Following on from this principle, a drug may function in several different pathways; thus while being relevant in simultaneous treatment of several diseases, it might also create undesired side-effects. Earlier works suggest that even weak bindings to multiple targets may have profound effect on the biological system [7-9]. To understand such phenotypic responses, it is crucial to investigate the binding behavior of drugs. Drugs targeting protein-protein interactions, ultimately head protein interfaces where two protein chains come into contact. Consequently, understanding the details and principles of protein interfaces is immensely essential to develop a better strategy in drug design [1, 10]. Obviously, one step further is to combine such structural details with the interaction networks to improve drug design concept; and that is the primary focus of this review.

Here, we first describe the physical and chemical aspects of protein-protein interactions important for drug design: focusing on hot spots, pockets, cavities at protein interfaces. Next, we review some recent network-level approaches in disease, drug design and discovery. We discuss the advantages of mapping protein interface structures into interaction networks, specifically in drug repurposing and predicting side-effects. Finally, we provide examples of therapeutic implications on drug design.

2. PHYSICAL AND CHEMICAL ASPECTS OF PROTEIN-PROTEIN INTERACTIONS IMPORTANT FOR DRUG DESIGN

Biological reactions occur through the proteins, and the interactions among them play a crucial role. Protein interactions are bio-physical phenomena. Although the medium is full of numerous molecules at different sizes, proteins to be interacted find each other and the interaction occurs. Proteins are interacting through their surfaces with the help of the shape and biochemical complementarity regarding the flexibility of the molecules and the environmental conditions [11-15]. Hence, protein interactions are special reactions rather than random processes.

Interaction Occurs Through the Interface

The region where the two molecules are contacting is called the binding site, or considering both sides, the interface. Understanding the protein structures is very important to identify these regions. If the structures of contacting proteins are known, it is quite easy to determine the interface. Interfacial residues are usually found by calculation of close residues from two sides based on the distance in the three dimensional space [16-18] or with the help of accessible surface area calculations [18-20].

Many studies investigate properties of interactions to enlighten protein interaction phenomena [13, 14, 16, 19-31]. The stability of the interaction is provided with binding forces; hydrogen bonds, salt bridges, electrostatic interactions and hydrophobic attractions [32, 33]. Disulphide bonds are also but rarely seen between binding proteins. These attractions determine also the specificity of the interaction [33]. For example, obligate complexes rely usually on hydrophobic attractions and transient complexes on hydrogen bonds and salt bridges [13, 19, 21, 33] although obligate interfaces are not more hydrophobic than transient ones [34, 35]. Understanding such details of the complexes is crucial in drug design; many signaling events, which are known to be disease-related, are regulated by transient protein-protein interactions [36, 37].

Considering the composition of the residues, we know that they usually differ between obligate and transient, or homo- and hetero-complexes [13, 19, 21, 33, 38-42]. Hydrophobic residues like Methionine, Phenylalanine, Proline, Alanine, Aspartic acid and Leucine (but not Isoleucine and Glycine) are commonly found in homodimer interfaces; whereas hydrophilic residues like Tryptophan, Cysteine, Histidine, Glutamine, Asparagine, Tyrosine and
Serine (but not Threonine) are dominant in heterodimer interfaces [21, 42].

Interfaces are structurally similar to the cores of globular proteins [43], and many studies have shown that protein binding process is very similar to the protein folding [26-28, 44]. The structural shape of interfaces is planar and well packed, but also differs with respect to type of the interaction [19, 45]. For example, relatively large surface areas are observed in homodimeric complexes compared to heterodimeric complexes [21, 42].

**Structural Features of Interfaces**

Proteins interact biophysically. Therefore, fundamental determinants in protein interactions are the shape and physicochemical complementarity [11, 13, 46, 47]. Protein surfaces are full of pockets, crevices and indentations [48-51]. Some pockets which usually exist before binding are filled with the complementary protein like the key and lock model, when they associate during the interaction [52]. This is also valid for protein-ligand interactions. Size and shape of pockets are principally considered to design molecules or peptides like drugs. Therefore, structural information of the protein has a great importance in drug design. The number of small pockets on the protein surface is not very few [16, 19, 21], on the contrary, much more than expected before. Besides, a cavity that does not pre-exist can be formed upon interacting with a small molecule [53]. In the lack of cavities, it becomes more difficult to inhibit a protein-protein interaction with a small molecule due to the flat large surface area.

In addition to the shape and size of the cavity, physicochemical properties of residues as well as their distribution in the cavity are important for the interaction. Thus, biochemical or electrochemical complementarity should also be considered in drug design. For example, due to the hydrophobicity of the interface, a drug to inhibit the protein-protein interaction should be designed to be hydrophobic [1, 54-57]. But this property makes the drug less soluble, less cell permeable and less selective for its targets [54-56].

Another parameter to characterize the interface is amino acid frequency [40, 58, 59]. The amino acid propensity of the interface is similar to interior of the protein [23, 60]. The interface is usually rich of aromatic and hydrophilic residues, like Cysteine, Tyrosine, Phenylalanine and Tryptophan [23, 61]. Positively charged Arginine and Histidine also prefer to be at interface compared to both surface and core [62]. Thrionine, Proline, Lysine, Glutamic Acid and Alamine are least commonly found in the interface [63]. Ofran and Rost [40] have predicted the type of the interaction with 63-100% accuracy rate just by using amino acid composition and residue-contact preferences. In another study, frequency of amino acids in transient and obligate complexes is analyzed. It is found that Glycine is more frequently seen in transient interfaces rather than the surfaces; whereas Glycine frequency is the same in interface and on surface of obligate complexes [61]. In the same study, pairwise contact preferences of the amino acid types are also investigated. Cystine-Cystine shows the highest preference of all possible contacts. This can be due to its ability to form disulfide bond. It is also found that this pair is more frequent in obligate interfaces than in transient interfaces. Moreover, acidic and basic amino acids are observed to be contacted mostly with other types of amino acids, but rarely with other amino acids with similar physicochemical properties. However, Histidine is found as an exception. It prefers to interact with itself but not with acidic amino acids. Furthermore, nonpolar amino acids prefer to contact with other nonpolar amino acids.

The other steric property of protein-protein interfaces is that they generally have an accessible surface area in the range of 1200-2000 Å² [13, 64]. This property is commonly used to distinguish biological interfaces from non-biological ones. Non-biological interfaces are formed between proteins which do not come together and contact each other actually in their physiological states [65].

Non-biological interactions are also called as crystal packing. These proteins are experimentally crystallized together; but it is mostly due to the enforcement by the crystallographic packing environment in the experiment. Non-biological interactions mostly have smaller accessible surface areas, which are about 400-600 Å² [13, 66]. However, it is not trivial to distinguish biological and non-biological interactions just by considering the buried surface area. There are some examples of non-biological interactions with larger surface areas, which can be more than 2000 Å² [67-70], and some biological interactions, like protein-small molecule interactions (300-1000 Å²), have smaller contact areas [71, 72]. Therefore, computational approaches to predict biological interactions utilize a combination of interface properties, like surface area, residue conservation and amino acid composition [73-77]. Although significant success has been obtained in identification of biological interfaces or distinguishing biological and non-biological interactions, distinct properties of protein interfaces have not been certainly identified yet [21, 53, 65].

**Interfaces are Conserved Throughout the Evolution**

Interacting proteins find each other specifically and selectively in a crowded medium. Therefore, proteins should identify each other through complementary interaction sites. This requires the conservation of the same (or similar) residues in the right orientation for both partners of the interaction [61]. To preserve the interaction, residues in the active or binding sides of proteins resist undergoing a change through the evolution [78, 79]. It is proved in many studies that amino acids in interfaces are more conserved than on rest of the protein surfaces [61, 73, 80-87]. Conservation analysis of residues can be used to predict protein binding sites. The analyses can be based on the sequence [3, 88, 89], or the structure [87, 90, 91], or both sequence and structure [88, 92, 93]. When obligate and transient complexes are compared, obligate complexes are found to have more conserved interfaces than the transient complexes [61, 86]. This supports that proteins of obligate complexes co-evolve, or proteins of transient complexes have higher ability to adapt to a change in their partners [86]. Moreover, analyses in the contact preferences of residues show that, residues prefer to contact with the other residues having a similar conservation grade [61]. In other words; highly conserved residues prefer to contact with highly conserved residues, and much variable residues prefer to contact with variable residues on the other side. Conserved residues are not randomly distributed and thus, some residues are more crucial than the others for the interaction [94]. If a contacting residue pair has an important role for the interaction, the residues are both conserved; if not, the residues are variable at the same grade [61].

The structures of interfaces are more conserved than the global structures of the proteins [95-97]. Even if their global structures and functions are different, proteins can interact through interfaces with similar architectures [16, 98]. Structures of the interfaces have been clustered based on the domain or the whole structure. The Conserved Domain Database (CDD) provides protein domains conserved in evolution, and they are extracted via multiple sequence alignments [99, 100]. Schroeder and coworkers [101] have found 6,000 distinct types of interfaces by clustering domain interfaces. Sali and his group [102] have shown that proteins in the same SCOP families have similar binding architecture. Aloy and coworkers [103] have classified domain-based interactions of known three-dimensional structure and developed 3did web server. PPIClust provides clusters of similar 3D interface patterns in protein complexes [104]. Gao and Školknick have found that structural space of protein-protein interfaces is close to complete and clustered interfaces of dimers into roughly 1,000 distinct types [105].

A structurally non-redundant dataset of protein-protein interfaces can be defined as three groups: Type I, Type II and Type III [16, 106]. Type I is the most common one. It includes interacting
proteins with similar global structures. Type II contains proteins with similar interfaces but their global structures and functions are different. These structures are examples for the conservation of interface motifs even in the absence of global structural similarity [45, 107]. In Type III, only one side of the interface is similar and the other side is somewhat different. Hub proteins are mostly examples of this type.

**Some Residues are Energetically More Important in the Interface: Hot Spots**

All contacting residues do not equally contribute to the binding energy. Regardless of the size of the binding site, only a few residues are responsible for the majority of the total binding energy [94, 108-110]. These residues are called “hot spots”. Hot spots can be identified experimentally by Alanine Scanning Mutagenesis [81, 108]. In this method, each residue is mutated to an Alanine; and if a significant drop is observed in the energy due to the mutation, the residue is assigned as a hot spot. Hot spots can have different physicochemical properties; for example, they can be hydrophobic or polar [53, 86, 111]. Arginine, Tyrosine and Tryptophan are frequently found as hot spots; whereas, Leucine, Serine, Threonine, Valine [109, 112] are rarely. Methionine is rarely [109] or frequently [86] found as hot spot by different studies.

Hot spots are buried and tightly packed in the three-dimensional space [94]. They are located near the center of the interface and away from the solvent [10, 109]. However, they are usually found in discontinuous, discrete highly packed regions in interfaces. These clusters are called “hot regions” [94]. Furthermore, there is a strong correlation between structurally conserved residues and hot spots [86, 94, 113-115]. There are many computational studies based on their physical, biological and/or evolutionary features for the hot spot prediction. These approaches analyze combinations of hot spot features, like conservation, physicochemical properties, residue propensity, sequence profiles, accessible surface area, and contribution to binding energy [3, 92, 116-122]. Hot spots are also significant for drug design [123]. Since they are responsible for key contact potentials of the interactions, drugs are aimed to target these residues [10, 124-127]. The importance of hotspot residues in drug binding is exemplified in (Fig. 1): the antagonist L-685,818 targets to protein-protein interface of FKBP12-T beta R-I complex and binds to FKBP12 through its hot spot residues (Fig. 1C) and hence prevents its binding to T beta R-I (Fig. 1B).

### 3. PROTEIN INTERACTION NETWORKS IN DISEASE AND DRUG DISCOVERY

With the increase in availability of human protein interaction data [128], considerable attention has been directed towards studying the networks related to human disease [129]. In particular, protein interaction networks are inferred to investigate the topological properties of disease-associated genes. Wachi et al [130], for the first time, studied the human interaction map for the analysis of

---

cancer and found that squamous cell lung cancer genes are highly connected and central, hence, display the same topological properties of essential genes. In another study, cancer proteins were found to have, on average, twice as many interaction partners as non-cancer proteins [131]. Moving beyond cancer, in a study of human inherited ataxia disease, most of the ataxia-associated proteins were shown to form an interconnected sub-network in the protein interaction network suggesting that these proteins may function in a cooperative manner [132]. Considering the disease-associated proteins in OMIM Morbid Map [133], it was shown that they have more protein-protein interactions than do non-disease proteins [134]. Goh et al. [135] constructed the first human disease network by linking disease that share disease-associated genes and differently from previous studies, demonstrated that most disease-associated genes are non-essential and are not encoded by hub proteins. In addition to identifying topological properties of disease-associated genes, network-based approaches are also assessed to predict new potential disease-associated genes (see [4] for a recent review).

As aforementioned, a disease state is often the result of a complex combination of molecular events, which consequently implies the complexity of therapeutic approaches. In addition, biological systems are robust; they can often bypass a perturbation and maintain their functions through different mechanisms such as back-up circuits and fail-safe mechanisms [136]. Hence, when selecting the putative drug targets, it is crucial to consider their positioning in the network [5]. Analysis of the network properties of known human drug targets showed that they tend to occur at middle-degree to low-degree nodes, i.e. less connected nodes in protein-protein interaction network [137]. This finding implies that targeting less prominent nodes in a protein interaction network can create fewer side effects and increase synergetic efficacy for combinatorial drugs [137]. Ranking the proteins by topological properties in the human protein-protein interaction network was also shown to be useful in one target identification [138]. Another approach, which seems to make major contributions to drug discovery lately, is to include structural details of proteins into networks and discussed in the following section.

Protein-protein interaction databases provide an invaluable resource to study protein-protein interactions at a network level. These databases grow rapidly as the number of interaction-detection experiments, genome sequencing experiments and proteins with solved structures increase. Among these, the Database of Interacting Proteins (DIP) [139], the Biomolecular Interaction Network Database (BIND - converted to Proteomics Standard Initiative-Molecular Interaction, PSI-MI 2.5) [140], the BioGRID General Repository for Interaction Datasets [141], IntAct [142], the Human Protein Reference Database (HPRD) [143], the Munich Information Center for Protein Sequences (MIPS) [144], the Human Protein Interaction Database (HPID) [145], the database of Protein Structural Interactome map (PSibase) [146] and the database of Protein Domain Interactions (DOMINE) [147] list experimentally determined protein-protein interactions. From these, PSibase and DOMINE include the binding site information, i.e. they indicate where two proteins interact and hence can assist in structural studies in disease.

4. INCLUDING PROTEIN STRUCTURAL INFORMATION INTO INTERACTION NETWORKS: ADVANTAGES IN DRUG DESIGN

The pioneering work of Aloy and Russell [148] illustrated how three dimensional protein structures can be used to infer molecular details of interactions in a network. Following this, structured networks are utilized to understand diseases. Davewal et al. [149] constructed a network related to pancreatic cancer by combining known interactions and structure-based interaction predictions. They predicted 40 novel interactions that are specific to pancreatic cancer. In one study, cancer-associated signaling pathways and their physical protein-protein interactions are analyzed with the goal of providing insights into three-dimensional structure-function relationship [150]. In another study, protein interface structures are integrated to human cancer protein interaction network and interface analysis revealed that cancer-related proteins have smaller, more planar, more charged and less hydrophobic binding sites than non-cancer proteins, which may indicate low affinity and high specificity of the cancer-related interactions [151]. Additionally, integrating interfaces into networks can provide timing of proteins’ interactions; whether they are simultaneous or exclusive [152-154].

Besides inferring the nature of disease-related interactions and their role in the network, comparison of protein-protein interfaces across an interactome can also assist in identifying drug targets or drugs targeting multiple proteins to block parallel pathways in a network [155]. With structural analysis drug binding pockets can be identified and compared with binding pockets of other proteins in the network, which could eventually lead to discovering candidates for drug-targetable protein-protein interactions [156]. Since the number of distinct binding motifs is limited in nature [157], structurally different proteins can share similar interface architectures [16]. Therefore, even though two proteins are structurally different and belong to different pathways in the interactome, if their binding pockets are similar, a drug can bind to both. While this is possible, binding also depends on the chemical nature of the drug ligand itself [158]. It was proposed that the increase in charged residues in a designed peptide drug can provide conformational flexibilities and enhance the propensity to bind multiple proteins [155].

Although there occurs some limitations such as incompleteness of protein interaction networks and structural data, combining protein interfaces with interaction networks could guide identification of all target proteins that are influenced by a drug, either positively or negatively. Considering the position and role of these target proteins in the network could help to predict side-effects or to discover a treatment of a new disease. The latter is called drug repurposing, which means finding new uses of old drugs [159], is gradually gaining popularity since de novo drug design is laborious and very costly [160]. One example is entacapone, which targets catechol-O-methyltransferase and is used in the treatment of Parkinson’s disease [161]. Based on binding site similarity, Kinnings et al. [161] identified a new target of entacapone: enoyl-acyl carrier protein reductase. This new target, which is important in the fatty acid biosynthesis in tuberculosis, is experimentally verified [161] and inhibiting this protein by entacapone offers a way to treat infections such as tuberculosis. Another widely known example is the celecoxib (Celebrex), which is a nonsteroidal anti-inflammatory drug that is known to bind cyclooxygenase-2. Weber et al. [162] showed that celecoxib also binds to a totally unrelated protein; carbonic anhydrase, which has a structurally similar binding site to that of cyclooxygenase-2. This finding reveals a new role of celecoxib in the treatment of glaucoma and possibly for cancer [162]. Provided that protein structures and binding sites are known in a disease-associated network, predicting druggable proteins and the resulting effects will be less demanding. This concept is visualized in (Fig. 2) with the celecoxib example.

5. CASE STUDY: A PROMISCUOUS DRUG: IMATINIB (GLEEVEC)

The concept that one drug binding to multiple distinct targets [6] is known as polypharmacology or drug promiscuity. “One drug-one target” paradigm erodes [6, 163] and several drugs are observed to be promiscuous [164]. One of them is imatinib (also known as Gleevec), which is a milestone in small-molecule drug discovery and is successful in cancer therapy [165]. Imatinib is first designed as a drug targeting Brc-Abl tyrosine kinase, which causes chronic myelogenous leukemia [166]. Later on, it is realized that imatinib is not entirely specific; it can bind to tyrosine kinases other than Abl such as KIT and PDGFRα [166]. Subsequently, imatinib has also
been approved for the treatment of gastrointestinal stromal tumors, which is caused by mutations in KIT or PDGFRα [166]. Binding sites of Abl and KIT to imatinib are similar (Fig. 3). Imatinib seems to have potential for use in a variety of other diseases; for example, more recently, a non-kinase target; NQO2 is identified for imatinib [167] (Fig. 3). NQO2 is a cytoplasmic flavoprotein functioning in the cellular response to oxidative stress [167] and its inhibition by imatinib can lead to anti-inflammatory effects [165]. Comparison of the complex structures of imatinib with Abl and NQO2 shows that conformation of imatinib and binding nature differ: imatinib interacts with NQO2 through hydrophobic interactions without any hydrogen bonds, whereas it uses six hydrogen bonds and several van der Waals interactions in binding to Abl kinase [167, 168]. Imatinib adopts an extended conformation in binding to Abl and that is quite different from the compact conformation seen in the NQO2 structure. Consequently, to identify potential targets of a drug, its distinct conformations are also needed to be taken into account. Known side effects of imatinib include muscle cramps, edema, diarrhea and bone-narrow toxicity [166] although their cause is currently unclear. However, yielding adverse effects is an expected outcome since imatinib is more likely to have several yet unknown targets as well and appears to attack multiple pathways, such as cell cycle and many signaling pathways according to KEGG pathway database annotations [169]. Clearly, a network-centric approach with inclusion of structural and dynamic details will assist in understanding the biological and kinetic profile of promiscuous drugs.

6. CONCLUSIONS

Many promising drug candidates fail the last clinical phases since mechanisms of the disease pathways they target are yet unclear or off-target effects cannot be predicted by in-vitro models [5]. Network-centric approaches offer a way to focus on entire cellular pathways, and hence provide a global perspective in drug design. Protein three-dimensional structure information is further required to examine atomic details of protein-ligand interactions, in particular, to understand how proteins interact with drugs and the resulting functional effect of these interactions. Although structural coverage of the proteome is limited, the structures deposited in the Protein Data Bank has been increasing at a fast rate [170], which will assist in structural studies. Obviously, for drug design studies, finding and analyzing the binding sites of proteins can reveal important clues. We know that some residues (hot spots) in the interfaces are more critical than others to the stability of the complex. As exemplified in this study for the case of FKBP12, probably, a drug ligand is more likely to interact with its target protein utilizing the hot spot residues of the target. Since similar binding sites usually bind the same drug ligands, in a disease-associated structural
network, identifying which binding sites of which proteins show similarity could reveal possible targets of the drugs and ultimately, the possible positive or negative side effects.

REFERENCES


Fig. (3). A promiscuous drug; imatinib binding to different targets. Imatinib binds and inhibits ABL tyrosine kinase and hence is used in the treatment of chronic myelogenous leukemia (left panel). Imatinib also can bind to another kinase: KIT which has a similar binding site to that of ABL (mainly DFG motif is utilized through interacting with imatinib) (right panel). KIT is targeted by imatinib in the treatment of gastrointestinal stromal tumors. Latterly discovered non-kinase target of imatinib is NQO2 (the panel below) although the resulting effect of this interaction is not clear. In this case, imatinib adopts a different conformation, i.e. more compact, and binds to cofactor FAD of NQO2 making mostly hydrophobic contacts.
Protein-protein Interfaces Integrated into Interaction Networks


[95] Illergard K, Ardell DH, Elofsson A. Structure is three to ten times more conserved than sequence—a study of structural response in protein cores. Proteins 2009; 77: 499-508.


Protein-protein Interfaces Integrated into Interaction Networks


