A NOVEL STRUCTURAL PROTEIN-PROTEIN INTERACTION NETWORK MODEL:
Its Applications on Drug Off-Target Prediction and Genotype-Phenotype Linkage

by

Hatice Billur Engin Aras

A Thesis Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Doctor of Philosophy
in
Computer Engineering

Koç University

November, 2013

Koç University
Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a doctoral dissertation by

Hatice Billur Engin Aras

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____
Prof. Attila Gürsoy

_____
Prof. Özlem Keskin

_____
Assist. Prof. Alkan Kabakçıoğlu

_____
Assoc. Prof. Engin Erzin

_____
Assoc. Prof. Öznur Özkasap

_____
Prof. Türkan Haliloğlu

_____
Prof. Uğur Sezerman

Date _____

ABSTRACT

Network descriptions and analyses are important tools in systems biology; they are powerful in abstracting the complex relationships inside cells and between them, and they often provide clues for drug discovery. In the first part of this dissertation, we introduce a structural network model that we call "Protein Interface and Interaction Network (P2IN)", which is the integration of protein–protein interface structures and protein interaction networks. This interface-based network organization clarifies which protein pairs have structurally similar interfaces and which proteins may compete to bind the same surface region.

Next, we propose a new network attack strategy, "The Interface Attack", based on protein–protein interface motifs. Similar interface architectures can occur between unrelated proteins. Consequently, in principle, a drug that binds to one has a certain probability of binding to others. The interface attack strategy simultaneously removes from the network all interactions that consist of similar interface motifs. This strategy is inspired by network pharmacology and allows inferring potential off-targets. We built the P2IN with the p53 signaling network and performed network robustness analysis. We show that (1) "hitting" frequent interfaces (a set of edges distributed around the network) might be as destructive as eliminating high degree proteins (hub nodes), (2) frequent interfaces are not always topologically critical elements in the network, and (3) interface attack may reveal functional changes in the system better than the attack of single proteins. As a case study, we tried to detect the off-targets of some CDK6 binding drugs. We found that drugs blocking the interface between CDK6 and CDKN2D may also affect the interaction between CDK4 and CDKN2D.

Lastly, we describe how we use protein interactions and the structural knowledge on interacting surfaces of proteins (interfaces) in predicting the genotype-phenotype relationship. We built the phenotype specific sub-networks of protein-protein interactions (PPIs) involving the relevant genes responsible for lung and brain metastasis from primary tumor in breast cancer. First, we selected the PPIs most relevant to metastasis causing genes (seed genes), by using the "guilt-by-association" principle. Then, we modeled structures of the interactions whose complex forms are not available in Protein Databank. Finally, we mapped mutations to interface structures (real and modeled), in order to spot the interactions that might be manipulated by these mutations. Functional analyses performed on these sub-networks revealed the potential relationship between immune system, infectious diseases and lung metastasis progression, but this connection was not observed significantly in the brain metastasis. Besides, structural analyses showed that some PPI interfaces in both metastasis sub-networks are originating from microbial proteins, which in turn were mostly related with cell adhesion. Cell adhesion is a key mechanism in metastasis; therefore these PPIs may be involved in similar molecular pathways that are shared by infectious disease and metastasis. Finally, by mapping the mutations and amino acid variations on the interface regions of the proteins in the metastasis sub-networks we found evidence for some mutations to be involved in the mechanisms differentiating the type of the metastasis.

ÖZET

Ağ açıklamaları ve analizleri sistem biyolojisi için önemli araçlardır ; hücre içindeki ve hücreler arasındaki karmaşık ilişkileri özetlemekte güçlüdürler ve genellikle ilaç keşfi için ipuçları sağlayabilirler. Bu tezin ilk bölümünde, "Protein Arayüzey ve Etkileşim Ağı (P2IN)" olarak isimlendirdiğimiz yapısal bir ağ modelini tanıttık. Bu ağ protein-protein arayüzey yapılarının ve protein etkileşim ağlarının entegrasyonundan oluşmaktadır. Bu arayüzey bilgisine bağlı ağ organizasyonu hangi protein çiftlerinin yapısal olarak benzer arayüzeylere sahip olduğunu ve hangi proteinlerin aynı yüzey bölgesine bağlanmak için rekabet ettiğine açıklık getirmektedir.

Daha sonra, protein-protein arayüzey motiflerine dayanan yeni bir ağ saldırı stratejisi önerdik, "Arayüzey Saldırısı". Benzer arayüzey mimarileri ilintisiz protein çiftleri arasında oluşabilirler. Bu nedenle, prensip olarak, birine bağlanan bir ilacın belli bir oranda diğerlerine de bağlanma olasılığı vardır. Arayüzey Saldırısı, benzer arayüzey motiflerinden oluşan tüm etkileşimleri ağdan aynı anda kaldırır. Bu strateji ağ farmakolojisinden ilham almıştır ve potansiyel “dış-hedefler” ’in tahminine izin verir. Biz p53 sinyal ağının P2IN’ini inşa ettik ve ağda sağlamlık analizleri gerçekleştirdik. Biz (1) sıkça gözlemlenen arayüzeylerin (ağın çeşitli yerlerine dağılmış kenarlar) saldırılara hedef alınmasının, yüksek dereceli proteinlerin (hub düğümleri) ortadan kaldırılması kadar yıkıcı olabileceğini (2) sıkça gözlemlenen arayüzeylerin ağda her zaman topolojik olarak kritik noktalarda bulunmadığını (3) Arayüzey Saldırısının sistemdeki fonksiyonel değişiklikleri, tek tek proteinleri hedef alan saldırılardan daha iyi ortaya çıkarabildiğini gösterdik. “Dış-hedef” tespiti örnek çalışmasında, CDK6 ve CDKN2D arasındaki arayüzeyi engelleyen ilaçların, CDK4 ve CDKN2D arasındaki etkileşimi de etkileyebileceğini bulduk.

Son olarak da, genotip-fenotip ilişkisinin tahmini için, protein etkileşimleri ve bu etkileşimlerin üç-boyutlu yapısının nasıl kullanıldığını açıkladık. Meme kanserinde primer tümörün akciğer ve beyin metastazına yol açması ile ilintili fenotipe özel protein etkileşim alt-ağları inşa ettik. İlk olarak, "işbirliği-ile-suçluluk" prensibini kullanarak, metastaza neden olan genlerle (tohum gen) en çok ilişkide bulunan protein etkileşimlerini seçtik. Daha sonra, kompleks halleri Protein Bilgi Bankası’ nda bulunmayan etkileşimlerin yapılarını modelledik. Son olarak, mutasyonlar tarafından manipüle edilmiş olabilecek etkileşimleri bulmak için, arayüzey yapıları üzerinde mutasyonları işaretledik. Bu alt ağlarda yapılan fonksiyonel analizler bağışıklık sistemi, enfeksiyon hastalıkları ve akciğer metastazı arasındaki potansiyel ilişkiyi ortaya çıkarmıştır, ama bu bağlantı beyin metastazı için kayda değer bir şekilde gözlenmemiştir. Bunun yanı sıra, yapısal analizler her iki metastaz alt-ağı içindeki protein etkileşim arayüzlerinin mikrobiyal protein kaynaklı olduğunu gösterdi. Bahsi geçen bu protein etkileşimlerinin hücre yapışması ile ilintili olduğu gözlemlendi. Hücre yapışması metastaz için önemli bir mekanizmadır; bu nedenle bu protein etkileşimleri bulaşıcı hastalık ve metastaz tarafından paylaşılan benzer moleküler yolaklarla ilgili olabilirler. Son olarak da, metastaz alt-ağlarındaki proteinlerin arayüzey bölgelerine amino asit varyasyonlarını eşleyerek, bazı mutasyonların metastaz türünün ayırt mekanizmalarına dahil olduğuna dair ipuçları bulduk.

ACKNOWLEDGEMENT

It's hard but you know it's worth the fight...

Thank you God for letting me fulfill my childhood dream of becoming a scientist.

Earning my PhD degree has been the most challenging pursuit of the first 30 years of my life. The best and worst moments of this journey have been shared with many people. I know words wont be enough to show my gratitude to them, but here is my acknowledgement page.

My first debt of gratitude must go to my advisors Prof. Attila Gürsoy and Prof. Özlem Keskin. I wish to express my deepest appreciation to them. They have taught me; both consciously and unconsciously how good science is done. I appreciate all their contributions of time, ideas and funding to make my PhD experience productive. I'm also thankful to them for the excellent examples they have provided as successful scientists. They will always remain dear to me.

Besides my advisors, I would like to thank the rest of my thesis committee members Assist. Prof. Alkan Kabakçıoğlu, Assoc. Prof. Engin Erzin, and Assoc. Prof. Öznur Özkasap for their critical reading and useful comments and thesis jury members Prof. Uğur Sezerman and Prof. Türkan Haliloğlu for their valuable time.

I would like to thank TUBITAK for their financial support during my PhD study (Research Grant Numbers: 109T343 and 109E207).

I offer my deep appreciation to Prof. Ruth Nussinov for her precious helps during my PhD and for sharing her endless experiences with us. I would like to thank Prof. Baldo Oliva and his lab members for the six months I have been a visiting researcher in Barcelona. I am deeply grateful to Besray Ünal, Emre Güney, Milica Pavlovic and Simone Ringlau for their exceptional friendship and our Barcelona days.

I would like to thank my lab-mate Ece Acuner-Özbabacan for her friendship and support; I would feel so lonely without her. I would like to thank my lab members Alper Başpınar, Deniz Demircioğlu, Doğuş Doğru, Emine Güven, Engin Çukuroğlu, Güray Kuzu, Selin Karagülle, Serena Muratçıoğlu and Tayfun Tümkaya. My special thanks goes to Nurcan Tunçbağ, she never got fed up with my questions and I learned a lot from her. I would like to thank Cengiz Ulubaş , Gözde Kar, Osman Yoğurtçu, Özge Şensoy and Sefer Baday. Distance was never an excuse for our friendship. I would like to thank my best confidant Mert Subaşı for being the most patient listener ever.

I would never be able to survive PhD without the fun times we had together with my high school friends Aybike Özden, Ceren Hayran, Duygu Ergelen, Merve Tolunay and Meral Turanlı. I feel so lucky to have them in my life. Thank you girls!

I want to thank my family for always believing in me and supporting me through my degree. Especially, the patience and understanding shown by my mum and dad is highly appreciated. I know at times, my temper is particularly trying. Teşekkürler Anne! Tesekkürler Baba! I would like to thank my brothers Yıldırım and Ömer Engin, they are so precious to me. I would also like to thank my new sisters Merve Engin, Banu Aras and Şebnem Tanrıkut. I offer my deepest gratitude to Nurcan and Ersin Aras, my parents in law, for honestly sharing my feelings these days.

My husband, Alp Aras, whose encouragement, sacrifices and love made all the difference and helped me finish this journey. He already has my heart so I will just give him a heartfelt "thank you".

I dedicate this work to my family. I hope that it makes you proud…

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

AIGL	Average Inverse Geodesic Length
BMSN	Brain Metastasis Sub-Network
GCS	Giant Component Size
HN	Hub Node
LMSN	Lung Metastasis Sub-Network
MD	Maximal Damage
MFI	Most Frequent Interface
MIM	Molecular Interaction Map
P2IN	Protein Interface and Interaction Network
PDB	Protein Data Bank
PIN	Protein Interaction Network
PPI	Protein-Protein Interaction
PRISM	Protein Interactions by Structural Matching

Chapter 1

INTRODUCTION

Proteins usually team up together to function in several biological processes. Other proteins with which it interacts often regulate the function and activity of a protein. Bearing in mind that protein interactions are the basis for all cellular processes, building a protein-protein interaction network, a map of physical interactions between proteins, is a very essential step in understanding the complex mechanism in living systems. Networks help abstract the complex relationships inside cells and between them. While data are incomplete, and the approaches may not have matured, network descriptions and tools gradually become commonly used [1].

Conventional protein interaction networks provide binary information: whether two proteins interact or not. In order to grasp the information on "how protein couples interact?" the knowledge on three-dimensional structure of proteins is essential. In the last decade the molecular details of interactions have started to be integrated in protein interaction network. They provided mechanistic information about the regulatory mechanisms of protein interactions such as "mutually exclusive interactions" [2] and the location preferences of the mutations on the interfaces [3].

Another use of networks is its topological properties such as hubs, betweenness, modules, etc. Network topology determines the information flow. Information flow and robustness analyses are used to locate essential components.

This dissertation mainly focuses on integrating structural knowledge to protein-protein interaction networks and utilizing this additional information in solving problems like drug off-target prediction and genotype-phenotype mapping. Interface structures of protein-protein interactions are modeled in atomic level and cancer related structural protein interaction networks are built. Sequence variations and drug-protein interactions are also integrated (separately) into these networks for answering specific questions. Presented methods and results may be in service to cancer bioinformatics, drug-protein interactions and systems biology.

The outline of this dissertation is as follows:

In Chapter 2, an extensive literature review is provided. This section starts with reviewing the literature on protein interactions, protein interfaces and structural protein interaction networks. Then, tells about the network robustness studies applied on biological networks. Afterwards, the network-based strategies in polypharmacology are reviewed from a systems biology perspective. Finally, the attempts to shed light in the molecular mechanisms of metastasis via systems biology approaches are reviewed.

In Chapter 3, a novel structural protein-protein interaction network model, based interface structure similarity is proposed. This new network model is named as "Protein Interface and Interaction Network" (P2IN). P2IN is explained in detail through several cancer related structural protein-protein interaction network examples.

In Chapter 4, the use of P2INs in drug off-target prediction problem is presented. For this purpose the network attacks are used to depict the effects of drugs in p53 signaling network. We found that drugs blocking the interface between CDK6 and CDKN2D may also affect the interaction between CDK4 and CDKN2D. The methodology for locating similar interfaces on a network of protein interactions is described and the interface attack, which is a novel attack strategy based on similar interface concept, is defined.

In Chapter 5, the employment of P2INs for bridging the gap between genotype and phenotype is shown. The phenotype specific structural networks for brain and lung metastasis of breast cancer are built. By mapping the sequence variations on these structural networks, SNPs related with each phenotype are revealed in case studies.

This dissertation ends with a chapter discussing the results and main conclusions.

Chapter 2

LITERATURE REVIEW

In this chapter, a comprehensive review of the studies related to protein-protein interaction networks and their applications to poly-pharmacology and metastasis is presented. First, protein-protein interactions and their types, protein-protein interfaces and structural protein-protein interaction networks are described. Then, network robustness studies are reviewed. Later, network-based strategies in poly-pharmacology are presented. Finally, the main contributions of this thesis work are described after attempts to understand the molecular mechanisms behind metastasis via systems biology approaches are reviewed.

2.1. Protein-Protein Interactions

Proteins rarely act alone. When two or more proteins team up together, protein-protein interactions occur. Other proteins with which it interacts often regulate the function and activity of a protein. PPIs are the basis for all cellular processes.

Protein interactions have a great structural and functional diversity. Large macromolecular complexes, such as ribosome, are highly stable and permanent whereas dynamic and transient interactions are key components in signaling and regulatory networks [4-7]. Protein-protein interactions (PPIs) can be classified based on their composition, affinity and life time [8, 9] as: i) homo- and hetero-oligomeric complexes, ii) non-obligate and obligate complexes and iii) transient and permanent complexes, respectively.

2.1.1. Homo-oligomeric and hetero-oligomeric complexes

These groups of complexes are differentiated based on their compositions such that if a PPI occurs between identical chains, it is said to form a homo-oligomer whereas if the PPI takes place among non-identical chains then it forms a hetero-oligomer complex. Homo-oligomers are symmetric and provide a good scaffold for stable macromolecules. For example, a chaperonin protein is formed by seven GroEL proteins associating as a homo-heptamer to form a cylinder and seven GroES proteins

cap one side of this cylinder [10]. The cylindrical region is an example of a homo-oligomer, whereas the GroEL/GroES complex is a supramolecule of hetero-oligomers. The stability of hetero-oligomers can vary and form a basis to gather different proteins that cooperate in a single macromolecule. For example, a/b tubulins form a stable dimer and these dimers form long protofilaments, which are constituents of microtubules [11].

2.1.2. Obligate and non-obligate complexes

The key point for differentiation between these two groups is affinity. If the constituents (protomers, monomers) of a complex are unstable on their own *in vivo* then this is an obligate interaction whereas the components of non-obligate interactions can exist independently. As an obligate complex example, Ku proteins, which are involved in DNA repair, are shown to bind DNA as obligate homodimers [12]. On the other hand, signaling protein complexes are good non-obligate interaction examples, due to their transient nature. After contributing to the propagation of a signal, they are dissociated into the stable constituent proteins. For example, H-Ras protein, which is a G protein, has a key role in controlling the cell growth and differentiation signaling pathways. It interchangeably forms non-obligate complexes with guanosine triphosphatase (GTPase) activating proteins (GAPs) (acceleration of GDP-bound state of H-Ras – switch OFF) and guanine nucleotide-exchange factors (GEFs) (acceleration of GTP-bound state of H-Ras - switch ON), when the cell is resting and when activated in response to stimuli, respectively [13].

2.1.3. Transient and permanent complexes

These groups of interaction types are discriminated based on the lifetime (or stability) of the complex. Permanent interactions are usually very stable and irreversible (e.g. IL-5 cytokine dimer (PDB ID: 3b5k) [14]). However, the components of the transient interactions associate and dissociate temporarily *in vivo* [8, 14-18]. a/b tubulin dimer is an example of an obligate/permanent complex whereas the dimers of a/b dimers are transient and non-obligatory providing the dynamic nature to microtubules in cell division, cargo transportation and cytoskeleton [19]. Non-obligate interactions are predominantly transient [17], with a few examples of permanent (**Figure 2.1**), but

obligate interactions are usually permanent in nature [8]. It should be noted that, permanent and obligate terms are used interchangeably in the literature.

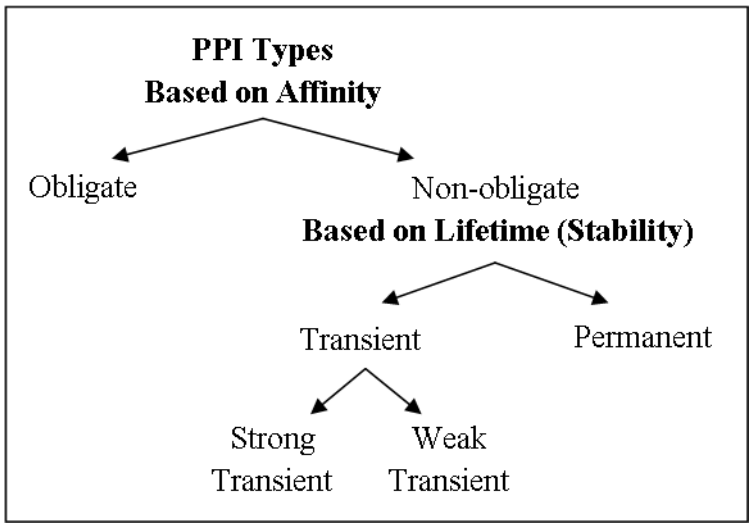

Figure 2.1. Relation of protein-protein interaction types based on affinity and stability [20]. Non-obligate interactions are transient but there are some examples of permanent non-obligate interactions such as enzyme-inhibitor interactions (e.g. thrombin-rhodniin inhibitor interaction)

2.2. Protein-Protein Interfaces

Proteins interact through their interfaces. Interfaces involve interacting residues that are coming from two different chains, along with neighboring residues [21]. Once the crystal structure of a protein–protein complex is present, investigating the atomic properties of the protein–protein interface is possible. However, if the complex structure is missing it is quite easy to predict the interface based on protein structures. Interfacial residues may be located by calculation based on the distance in the three-dimensional space [22, 23] or accessible surface area[24, 25].

2.3. Structural Protein-Protein Interaction Networks

Building a PPI network, a map of physical interactions between proteins, is a very essential step in understanding the complex mechanism in living systems. In PPI

networks nodes represent proteins and the undirected edges denote physical contact between two proteins. The complete collection of all PPIs within the cell is called "interactome"[26].

Until very recently structural information had nothing to do with PPI networks. However, in the last decade the molecular details of interactions have provided mechanistic information about the regulatory mechanisms of proteins. In 2006, Kim et al. added a dimension to PPI networks through structural modeling [2]. They described the interactions of two proteins that are binding to a common partner via the same binding site as "mutually exclusive interactions". According to their study these two interactions could not happen simultaneously. They constructed structural yeast PPI network with 873 proteins and 1269 interactions, 438 of which are mutually exclusive.

Moreover, Yang et al.[27] introduced SAPIN; a framework for the structural analysis of PPI networks. SAPIN was identifying the protein regions involved in interactions and provided template structures and identified the compatible and mutually exclusive interactions.

Later, Patrick Aloy and his co-workers provided structural details at atomic resolution for over 12,000 PPIs in 8 model organisms through the integration of interaction data from the main pathway repositories [28].

Wang et al. [3], built a human structural interaction network by combining PPI data and homology modeling. Using this structural network, they showed that, for corresponding diseases, in-frame mutations had a tendency to occur on the interface regions of the interacting proteins.

Kar et al. [29] built a structural network of cancer related human protein-protein interactions. In this network interactions were replaced by interfaces, coming from either known or predicted complexes. They investigated the topological properties of cancer network and performed a detailed analysis of the interfaces in this network. Their results 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. Besides,

they claimed that cancer-related proteins tend to interact with their partners through distinct interfaces, corresponding mostly to multi-interface hubs.

2.4. Network Robustness Studies

An attack on a network is executed in order to disrupt the information flow locally or globally, to disable a pathway or to destroy the network as a whole. An attack implies deletion or attenuation of an edge or a node of the network [30]. In chapter 4 of this dissertation we utilized network attacks to depict the effects of drugs in a protein-protein interaction network and to develop a drug off-target detection method. In this section you will find a brief review on the network attack types.

2.4.1. Node Attack

A node attack on the network removes edges focused at a single node. There are two different node attacks [30]: complete knockout **(Figure 2.2.a)** and partial knockout (**Figure 2.2.b).** Complete knockout refers to removing a node with all of its edges; partial attack involves removing randomly selected half of the edges of a node. Complete node attacks are commonly used attack strategies. The targets of these attacks vary according to the network topology. Complex networks were believed to be randomly linked [31] until Barabasi *et al*. discovered a common topology [32]. This discovery introduced scale-free networks into network theory. While in a random network nodes have roughly the same number of edges, in a scale-free network there are many nodes with a small number of edges and a few nodes (hubs) with a large number of connections. Random node attacks may be destructive to networks that are randomly linked, whereas scale-free networks are highly robust under these attacks. Scale-free networks are defenseless upon few vital node removals [33]. Accordingly, targeting hub nodes is a preferred approach in network attacks[34-37]. Detaching those nodes, which have many neighbors, will disrupt the information flow.

Partial knockout was performed by Agoston *et al.[30]* on *E. coli* and *S. cerevisiae* networks. They removed randomly half of the edges of target node or attenuated all the edges of the node. This study suggested that partial weakening of a small number of nodes (3- 5) might have a stronger effect than completely removing a selected node;

in both cases the most damaging nodes were selected. Zhang *et al*. [38] questioned whether this result is a general concept for complex networks and retested all attack strategies on the Barabasi-Albert (BA) scale-free network [32] and the Erdös-Renyi (ER) random network [31]. They confirmed that multi target partial attacks may disturb complex networks more than single target complete attacks and ER random networks are more resistant to multi target partial attacks than the BA networks.

2.4.2. Edge Attack

An edge attack removes one or multiple edges from the network, where the edges do not have to be incident to a node. Depending on the network topology, attacking a high betweenness edge may damage the system more than attacking a hub node with many edges. Thus, deleting a number of edges scattered in different regions of the network might be a more efficient attack strategy than targeting a node[30]. This attack is a 'distributed attack' (**Figure 2.2.c).**

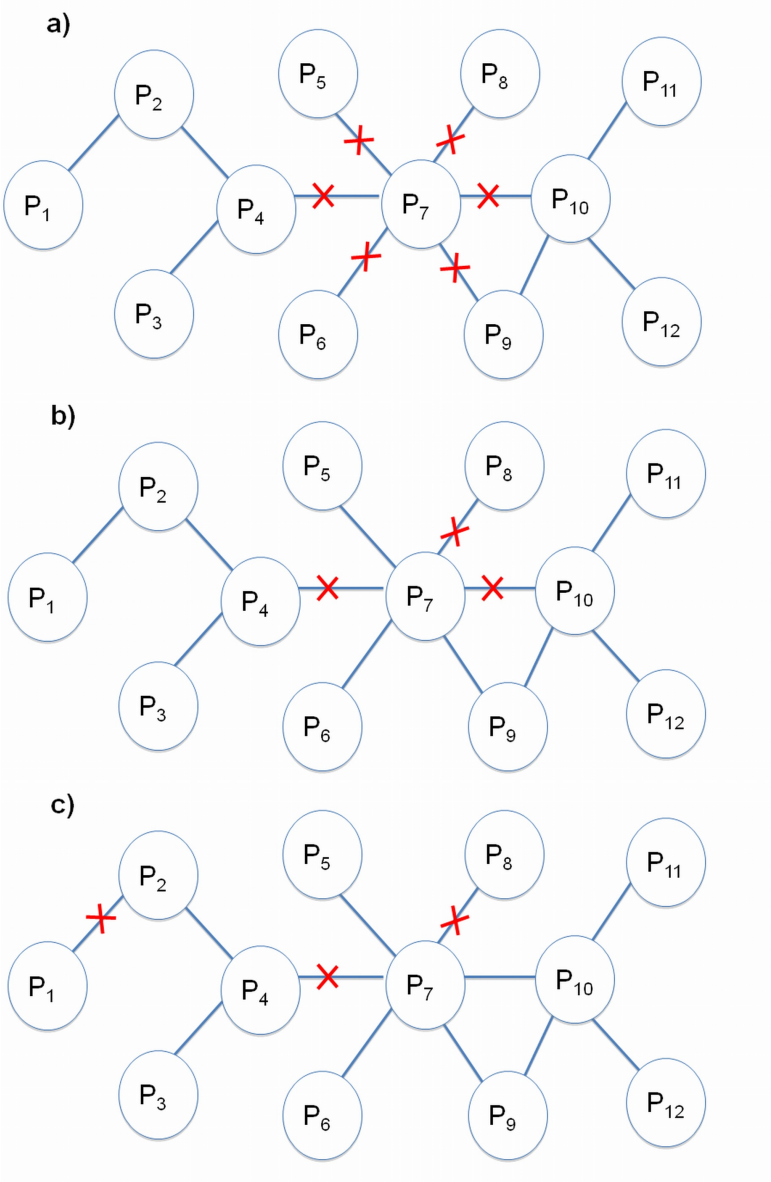

Figure 2.2. Network Attacks [39]. (a) Complete knockout (b) and partial knockout target a hub node. (c) Distributed attack.

2.5. Network-Based Strategies in Polypharmacology

In the 4th chapter of this dissertation we proposed a network attack strategy that is inspired by polypharmacology and protein-protein interface targeting drugs. In this section you will find a broad revision of the network-based strategies in the Polypharmacology and gain a systems biology view on drug discovery. This section

will provide a basis for the research we performed on drug off-target prediction via a network-based model.

2.5.1. Poly-Pharmacology

Poly-pharmacology searches for lead compounds that bind to multiple targets, and introduces a new concept of network pharmacology, which enlarges the 'drugome' [40]. It builds upon systems biology and drug discovery [41] aiming to treat diseases through multiple targets, which can be both a drug with several targets or a number of drugs with distinct targets. Poly-pharmacology describes and advocates consideration of a "many-to-many" relation between a ligand-protein couple, in contrast to a dominant "one drug-one target" drug design paradigm [42]. The novel computational approaches to poly-pharmacology has been reviewed recently by Xie et al. [43].

2.5.2. PPI Targeting Drugs

Design of drugs that disrupt PPIs is known to be notoriously difficult. This is for two reasons: protein-protein interfaces have a more flat surface when compared to enzymes and usually do not have grooves which can serve as binding pockets [44]. The pockets on protein-protein interfaces are typically smaller than those in protein-ligand interactions [45] and difficult to drug. However, now it is becoming increasingly possible to overcome these handicaps, and PPI inhibitors are gaining importance as a class of drug targets [46]. One of the most important findings was that interface regions usually contain clusters of residues, which are key contributors to the binding energy. These smaller regions of the interaction surface constitute "hot spots" [47-49]. Different studies showed that small molecules target hotspots on the protein-protein interfaces [49, 50]. Hot spots on an interface may be predicted via HotPoint [51], HSPred [52], KFC [53], and additional servers. Moreover, currently allosteric drugs that bind elsewhere and lead to conformational changes in the interface appear increasingly feasible. Nonetheless, although designing drugs for disrupting protein-protein interactions has surged, some such drugs have long been in existence. Protease inhibitors are well-known marketed examples of this drug class [54]. The number of PPI targeting drugs is rising; examples include inhibitors targeting IL-2 [55], MDM2 [56, 57], BCL-2/BCL-XL[58], XIAP [59] and VLA-4 [59]. There are a number of

reviews [44, 54, 60-65], which investigate PPI inhibiting drugs, and these provide a more extensive list.

Pockets at the active sites of enzymes are typically stable, with high population times. Recently attempts to target PPI have also focused on detection and targeting of transient dynamic pockets, which may be stabilized upon drug binding. Such grooves can be found in PPIs [21] and elsewhere on protein surfaces and can serve as orthosteric and allosteric binding sites. Transient pockets occur often [66]; the question is their size and population time. Furthermore, the surface of interacting proteins is flexible and some disordered proteins can only be solved upon interaction with their partners [67]. Their flexibility implicates formation of transient pockets, which are very useful for inhibitor design [68]. A number of drugs have been reported to stick to these transient pockets on the surface of protein interfaces [69]. Metz et al. [70] proposed a tool that locates transient pockets of PPIs on the basis of geometry, and molecular dynamics simulation protocols are also being developed toward this aim. Further validation of the presence of at least small pockets comes from a computational analysis of crystal structures. This study observed that among 18 protein-protein complexes, 16 contain pre-existing pockets in their unbound structures [71].

A number of clinical therapies are based on humanized monoclonal antibodies which disrupt PPI. These therapies have high specificity and low toxicity; however, they also have some deficiencies such as lack of cell/blood–brain barrier permeability, and poor oral bioavailability. Thus, humanized monoclonal antibodies therapies may not be broadly applicable to PPI inhibitor design [60] in the near future.

The data on PPI inhibitors have been compiled in databases and literature. One of these, 2P2I [72], is a database which provides structural data for a collection of protein-protein interfaces with known inhibitors. TIMBAL [73] is also a database where one can find small molecules disrupting PPIs. Furthermore, Sali et. al. spotted “bi-functional positions” of proteins (overlapping ligand and protein binding sites) by aligning homologous proteins. They pointed out the significant number of proteins that

have such bi-functional positions and released the collection of structurally characterized modulators of protein interactions at http://pibase.janelia.org [74].

2.5.3. Drugs targeting multiple proteins

Side effects are one of the main reasons for drug failure [75]. In the last 10 years nearly 20 drugs have been banned from the market for causing severe side effects [76]. Adverse effects can be caused by the inherent mechanism of action of a drug, by toxic metabolites following drug degradation and by unpredictable side effects due to "off-targets" drug hits. A number of studies highlight the promiscuity as a common attribute of drugs. Yildirim *et al*. [77] constructed a drug-target network from 4252 drugs targeting 394 human proteins. They found out that among 890 drugs 788 had at least one common target and in that network the average number of target proteins was 1.8 per drug. This result revealed the fact that new drugs generally target known druggable proteins and that the number of drugs targeting others is low in the market. However a more recent study by Mestres *et al*. [78], updated the average number of target proteins per drug as 6.3, which points to the high tendency of drugs to be multi-targeted. Paolini *et al*. [79] searched for the extent of promiscuity in the global pharmacological space and they also observed that among 276,122 active drug compounds, 35% hit multiple targets. The data compiled in the drug-target databases also indicates this many-to-many behavior. Some of these databases are listed in the **Table 2.1.** Among the multi-target drug examples, there are a number of kinase inhibitors, which operate by affecting multiple targets [80, 81], steroidal anti-inflammatory drugs (NSAIDs), salicylate, metformin or Gleevec™ [82] and the anticancer drug lenalidomide [43]. Several multi-target drugs were also discovered by chance [83].

Table 2.1. List of some drug-target databases.

Server / Database	Explanation	Web Site
DrugBank **[84]**	Detailed drug data with comprehensive drug target information	http://www.drugbank.ca/
TTD **[85]**	Therapeutic Target Database	http://xin.cz3.nus.edu.sg/group/ttd/ttd.asp

Stitch **[86]**	Chemical-Proteins Interactions database	http://stitch.embl.de/
PDSP Ki **[87]**	Capabilities of drugs binding to molecular targets	pdsp.cwru.edu/pdsp.php
PDTD **[88]**	Potential Drug Target Database	http://www.dddc.ac.cn/pdtd/
Wombat-PK **[89]**	Clinical Pharmacokinetics and Drug Target Information	http://www.sunsetmolecular.com/
BindingDB **[90]**	measured binding affinities of protein targets and small molecules	http://www.bindingdb.org/
PDBBind **[91]**	a collection of experimentally measured binding affinity data	www.pdbbind.org
KeggDrug **[92]**	Drug information resource of approved drugs	http://www.genome.jp/kegg/drug/
PubChem **[93]**	Biological activities of small molecules	http://pubchem.ncbi.nlm.nih.gov/
PharmGKB **[94]**	pharmacogenomics knowledge resource	http://www.pharmgkb.org/
DART **[95]**	Drug adverse reaction and target database	http://xin.cz3.nus.edu.sg/group/drt/dart.asp
SuperTarget **[96]**	A resource for analyzing drug-target interactions	http://bioinformatics.charite.de/supertarget
Promiscuous**[97]**	A resource of protein-protein and protein-drug interactions	http://bioinformatics.charite.de/promiscuous/
CTD **[98]**	Comparative Toxicogenomics Database	http://ctd.mdibl.org
sc-PDB **[99]**	Database of Druggable Binding Sites from the Protein Data Bank	http://bioinfo-pharma.u-strasbg.fr/scPDB/

There are examples suggesting that targeting multiple proteins simultaneously may be successful, such as non-steroidal anti-inflammatory drugs (NSAIDs); antidepressants; multi-target kinase inhibitors and anticancer drugs [100]. Drug combinations

('cocktails') may bind at different sites on the same protein; or to multiple different proteins. Examples include the three drugs combination used to treat HIV infection, which is composed of reverse-transcriptase and protease inhibitors [101] and the drug combinations known as "CHOP", which is used in the treatment of non-Hodgkin's lymphoma [102]. Another synergistic drug combination example is Cytarabine and Aplidin **(Figure 2.3)**, used for enhancing their antitumor activities in leukaemia and lymphoma models [103]. Cytarabine is an anticancer drug used in the treatment of patients with leukemia [104] and Aplidin is another, which activates EGFR, Src, JNK and p38MAPK [105] and inhibits VEGF [106]. When the system-wide effect of Aplidin is investigated, it is observed to activate the death receptor of Fas ligands [107]. This outcome of Aplidin may be due to the activation of the JNK/p38 MAPK pathway [108]. In turn, Fas ligand activates the receptor-mediated extrinsic cascade of apoptosis [109]. In addition, Cytarabine increases cellular stress by inhibiting DNA repair and RNA synthesis and drops the MCL1 level which leads to activation of CASPs [110]. Finally CASPs trigger apoptosis via the mitochondrial intrinsic cascade [109].

Due to drug combinations' side effects and at the same time enhanced treatment potential, detecting new drug cocktails and understanding their underlying mechanisms are important tasks. There is an increasing number of publications in this area, including in a recent work [111] a "drug cocktail network" built to investigate existing drug cocktails and to identify new ones. The authors note that drugs in a cocktail tend to interact with same partners and share common therapeutic effects. Another example is the computational method Zhao et al. [112] used for inferring new drug combinations. They combined molecular and pharmacological properties of drugs for this purpose and looked for feature patterns enriched in drug combinations. 69 % of their method's predictions were reported by literature. They also proposed some clues for combinatorial therapies. "Combinatorial Drug Assessment" [113] is an alternative tool for combinatorial drug discovery, which uses gene expression profiling and multiple signaling pathways. Lastly, Wang et al. [114] considered drug combinations in a genetic interaction network and the associated human pathways.

They observed that drug combinations alter functionally-correlated pathways and have a smaller influence range in the genetic interaction networks.

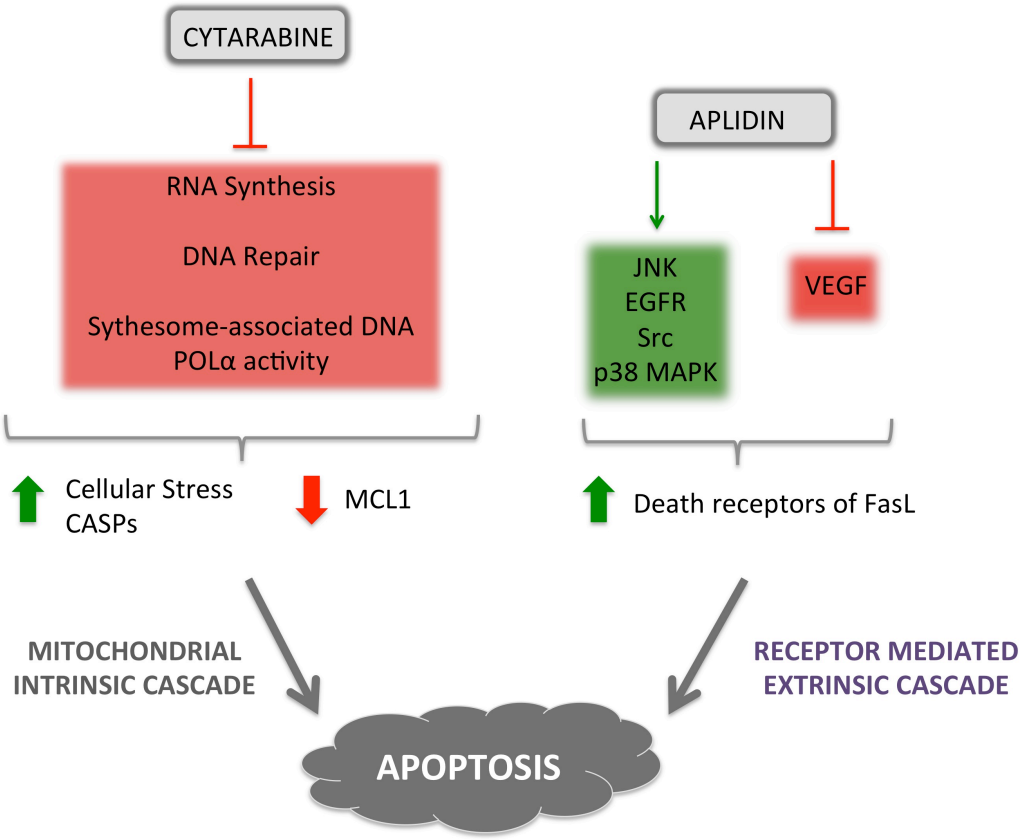

Figure 2.3. Synergistic drug combination of Cytarabine and Aplidinenhances antitumor activities [115].

2.5.4. A Systems Biology View

Biological systems are governed by physical and functional interactions. Systems biology simulates and orchestrates the molecules to optimally adapt the organism response to its environment. Diseases disturb the network; 'good' drugs restore the network to its 'proper' desired state [116]. Network descriptions and analyses are important tools in systems biology; they are powerful in abstracting the complex relationships inside cells and between them, and they often provide clues for drug discovery [117]. While data are incomplete, and the approaches may not have

matured, network descriptions and tools are gradually becoming common place [1].

The human protein-protein interaction (PPI) network is huge with approximately 130,000 binary interactions between proteins [118], and is expected to be far larger, with around 650,000 PPIs [119]. Protein-protein interaction networks are of vast importance in medicine [120, 121]. From the drug development standpoint, we would expect it to have critical components, which would make enticing drug targets. Network topology may help, because drug targets are usually not arbitrarily located on the protein interaction networks [122]. Drugs that perturb topologically critical nodes (such as highly connected nodes) have increased risk of causing lethality [35], while blocking the targeted function. This is likely to be the reason why marketed drugs do not generally target high degree nodes [123]. An 'ideal' drug target would have fewer neighbors while being located at some strategic point of the human disease network [101]. Such a target may be a non-vital bridging node [124]. It may disturb the information flow and the disease process while not causing serious side effects. For complex diseases like cardiovascular disease, central nervous system disorders, cancer, Alzheimer and aging, a network perspective is critically important. These require consideration of the global map of protein interactions and estimation of the expected outcome on the multiple, inter-connected pathways. As we describe below, in some diseases that are resistant to drug therapy [125], network-based strategy, where another protein in the same pathway is targeted may suggest alternative targets that may lead to the sought outcome [126, 127].

When enriched by high-throughput data, networks may model possible responses to a drug or optimum combination of drugs for reaching a desired outcome. On the other hand, because such data are derived from population of cells, its accuracy for specific environments and physiological states may be compromised. Networks may be analyzed using mathematical models such as Flux Balance Analysis [72, 128-130], differential equations [131], Petri Nets [132], Integer Linear Programming [133] and Boolean logic gates [134].

Another use of networks is its topological properties such as hubs, betweenness, modules, etc. Network topology determines the information flow. Information flow and robustness analyses are used to locate essential components. These algorithms are utilized to find perturbed proteins by hypothetical drugs [135] or for locating optimum drug targets that have little influence on other functions, apart from the intended one [136-139]. A key question is how to choose an efficient combination of multiple drug targets; especially those that while not key players in central pathways, ensure information flow among the network elements [124].

2.5.5. The advantages and handicaps of modeled protein-protein interactions in mono- and poly-pharmacology

Key requirements in drug discovery are the availability of protein structures and their interactions; the pathways in which they are located and the pathway cross-talk; and how similar are the binding sites to which they bind to those of other proteins in the cell. Modeling protein interactions can help by predicting which proteins interact, which permits the construction of more complete pathways and the cellular network. These may allow prediction of how targeting a specific protein can affect the entire system. The PRISM server [140, 141] is one of the tools which predicts the interacting protein couples and their interface structures. Further, because the modeled interactions provide also the information on how the proteins interact, they allow prediction of which partners interact through the same binding site. Thus, if a particular protein is targeted, this may abolish the competitive binding at the same shared site, driving the system in a certain direction, which the structural network may forecast. Such predictions may be particularly powerful for multi-molecular complexes, which are prone to toxic side effects. If the drugs target certain PPIs, the structural network may suggest the other PPI which share similar motifs [39]; and as such, may also be affected by the drug, which may also lead to toxic side effects.

Networks may be used to explain side effects of multi-target drugs. Xie et al. [142] studied the side effects of torcetrapib, which is an inhibitor of cholesteryl ester transfer protein (CETP). Torcetrapib was a proposed treatment for cardiovascular

disease and was in clinical trials. The authors compared all ligand-binding sites in all available protein structures, with the pockets on torcetrapib and created an off-target binding network. They combined their study with biological pathways and found the likely reasons for the effects of torcetrapib on blood pressure.

Propagation of the effects of drugs in the network may be observed for orthosteric and allosteric drugs [143, 144]. In the case of orthosteric drugs, which block the protein active site, the protein is impaired and its function is abolished; in the case of allosteric drugs, the modulating effects of drugs propagate through the protein and, through the protein-protein interactions, across the pathways. However, the effects are likely to be strongest in proteins sharing the same complex [144].

A key handicap of modeled structural networks is that they provide a static view of cell, and of the proteins. Yet, the cellular network is highly dynamic; proteins associate and dissociate. This is challenging to model, because the affinities of their interactions which are typically measured in solution, do not necessarily reflect the *in vivo* environment, where the affinities at one binding site are affected by prior allosteric events at different sites, for example, binding of other partners or post-translational modifications. They also may not account for the presence of co-factors; and fluctuations in the environment. An additional challenging problem is protein dynamics; protein structures fluctuate, and the distributions of their conformational ensembles change dynamically, which affects the binding site conformations and drug binding [143]. Accounting for dynamics in the proteins and across the pathways and the network is an extremely challenging problem. This is because it both necessitates detailed experimental data and highly demanding computational requirements. To date, modeling on the network scale is not able to fully address these problems [117]. However, for specific proteins, on the local scale, Nuclear Magnetic Resonance (NMR) and molecular dynamic simulations may be able to provide some clues.

Despite these shortcomings, single- and multi-drug pharmacology can benefit from the modeled structural proteome. Predictions are able to provide leads and hypotheses, which can then be validated by experiment.

2.6. Understanding the Molecular Mechanisms behind Metastasis via Systems Biology Approaches

2.6.1. Breast Cancer Metastasis

According to American Cancer Society, breast cancer is the second most common cause of cancer death among women [145]. Around 5-10% of breast cancer cases arise from gene mutations. The mutations on BRCA1, BRCA2, p53, PTEN, STK11, CHEK2, ATM, BRIP1 and PALB2 genes may be named as examples [146, 147]. Although the death rate of patients decreased with mammographic screenings and systemic adjuvant therapies [148], the breast cancer is pointed out to be the leading reason of death among women with the age of 40-59 [149, 150].

Metastasis is the mechanism that causes the distant spread of cancer [151]. As our diagnosing and treating ability of cancer advances, the fatality is moving towards metastatic phase [152]. Metastasis is the primary reason of death in cancer patients [153]. As well, the death cause of a breast cancer patient is most of the times is the metastasis in another organ, not the primer tumor. A better understanding of the molecular mechanism of the metastatic process may help to improve the clinical methods for approaching to the disease.

Breast cancer is considered to have a distinct metastatic pattern[154]. The lungs and bones are common breast cancer metastasis sites [155]. Besides most of the central nervous system metastasis originate from lung cancer (40-50%), which is followed by breast cancer (20-30%) [156]. Up to 20-40% of the patients with adult systemic malignancies grow brain metastasis [157, 158]. Brain metastasis is predicted to have 200,000 cases in us [159], which is 10 times more than the primary brain tumors [160].

2.6.2. Systems Biology Approaches to Understand Metastasis

In the recent years, numerous studies have been trying to shed light on molecular mechanisms of metastasis. Some of them are: oncogene activation with new experimental methods [161], identifying organ specific metastasis [155, 162], the identification of genes associated with metastases [151, 154, 163-165] and discovery

of pathways playing role in metastasis [166]. Besides, a series of studies in different laboratories revealed the required transcription factors for starting the process of metastasis, programming the biological changes in the cell [167-173].

Likewise, recently published gene expression profiles of breast carcinomas [174-176] have attracted wide interest in this regard. DNA-microarray studies demonstrated that primary breast tumors developing metastasis can be distinguished from tumors that do not metastasize, using gene expression profiles[148].

Massagué and his co-workers published several papers about breast cancer metastasis in the last decade, and in particular two of them studied the metastases of breast cancer towards brain and lung. One article [155] identified18 genes that mediate breast cancer to lung metastasis, and the other [177] classified 17 genes that mediate breast cancer to brain metastasis. They used differential expression analysis to identify these genes.

Genes related with metastasis are usually biologically related with each other [151]. For this reason, analysis of individual genes does not provide solid results about the metastatic process. Network formation and analyses are important tools for systems biology, providing a powerful abstraction of intracellular complex relationships. Most common diseases such as diabetes, schizophrenia, hypertension and cancer, are also believed to be caused by multiple genes (multi-genic) [178]. Recently, genes that have the potential to be involved with several diseases are uncovered through the integration of functional information of proteins and the protein interaction network [179-181]. Interactions in the sub-networks generally indicates functional signaling cascades, metabolic pathways or molecular complexes, which gives an idea about the cause or the result of the disease (phenotype) [121]. Protein interaction networks were also used to predict genes involved in breast cancer metastasis, and to identify the disease-related sub-networks [180, 182].

On the other hand, structural data can be very useful for explaining the molecular mechanisms leading to disease when used in conjunction with information about the mutation responsible for the disease [183]. For instance, Wang and colleagues [3] investigated the molecular mechanisms underlying complex genotype-phenotype relationships by integrating large-scale PPI data, mutation knowledge and atomic level

three-dimensional (3D) protein structure information available in RCSB Protein Databank (PDB) [184]. They revealed that the in-frame mutations are augmented on the disease related proteins' interaction interfaces. Similarly, David et. al. [185] combined structural data of proteins/protein-complexes and non-synonymous single nucleotide polymorphisms (nsSNPs) and they investigated the location of nsSNPs for creating a database. They have observed that disease-causing nsSNPs that occur on the protein surface prefer to be located on the protein-protein interfaces.

2.6.3. Association Between Metastasis, Infectious Diseases and Immune System

Previous studies highlighted the resemblances in cellular and molecular mechanisms of invasion between metastasis and infectious diseases [186-189]. Besides, in a recent study, Haile et al. hypothesized that metastasis process and pathogens should be utilizing the same pathways [190]. Liu et al. also mentioned that certain pathogens, activated immune cells and tumor cells may be sharing same tactics to spread in the body [191].

Metastasis, which is believed to be relatively impossible to treat completely [192], is mostly resilient to standard treatments, thus the attempts to develop a treatment for metastasis with engineered bacteria is getting many of the researchers' attention. Recently, Hayashi et al. [193] proposed a targeted therapy for metastasis with a genetically-modified strain of Salmonella typhimurium. They claim that their approach is promising for curing metastasis without the need of chemotheraphy. Moreover, in 2004 Yu et al. [194], showed that bacteria injected into living animals are able to find and replicate in metastases. *Escherichia coli,* cytosolic vaccinia and three attenuated pathogens (*Vibrio cholerae*, *Salmonella typhimurium*, and *Listeria monocytogenes*) all entered tumors and replicated. Authors remarked the "tumor-finding" ability of bacteria and viruses (engineered to transport multiple genes) might be used for diagnosing and curing cancer. Another example of bacteria used for targeting metastasis is the use of Salmonella in conjunction with the endogenous angiogeneic thrombospondin-1 (TSP-1) that has been caused the inhibition of melanoma growth and metastasis in B16F10 melanoma models[195, 196]. Highly site-

specific adherens of bacteria makes them promising for tumor specific treatments, yet it is not easy.

In 2010 Dallo et al [197] published a very interesting article suggesting that bacteria under SOS may evolve anticancer phenotypes targeting metastatic cells. Disturbed with the drugs, bacteria can be stimulated to stick to and to occupy cancer cells so that bacteria survive the drug attack.

The cancer-fighting immune system mechanisms are similar to those fighting bacteria [198], such as Toll-like receptors [199]. Besides, the bacteria settlement in tumor sites may activate the immune cells in host and may demolish the immunosuppressive phenotype of tumor microenvironment [200]. Plus, cancer appears to develop similar maneuvers to bacteria (masking cells to avoid discovery, release of immunomodulators to collapse the immune system and misleading the immune system by sending fake messages), for overcoming immune system **[198]**. In fact this is not the only common feature cancer cells share with bacteria colonies. They also acquire more basic survival tactics that have been evolved by bacteria; speedy reproduction to make the cell number, creating variation in the populations and having continuous communication among cells. Moreover, cancer and bacteria are alike in the case of drug resistance. Bacteria gets resistant to antibiotic treatments after a while, that is also what happens to cancer after frequent anticancer drug treatment **[201-203]**. Additionally, the mysterious quiescence and strike back characteristic of cancer also seems to be evolved by bacteria and used by cancer **[198]**. Cancer may reappear after it had not been identified by examinations and blood tests for an indefinite amount of time. Equivalent state may be observed on bacteria before sporulation and subsequent germination.

2.7. Main Contributions

The success of the bioinformatics approaches is restricted by the availability/reliability of the data. With the completion of the Human Genome Project and other genome sequencing projects, our understanding of the molecular biology accelerated astonishingly in the last couple of years. However the interactome level large-scale

structural knowledge is far from being complete. In order to address this problem a number of structural PPI prediction algorithms have been developed, one of which is PRISM. With the help of PRISM, we focused on increasing the structural knowledge on PPIs and integrating this knowledge to PPI networks in this dissertation. We provided structural predictions for the architecture of interfaces of several PPIs through out this thesis.

We combined the experimental data and the modeled structural networks to build cellular pathways, and suggest which specific pathways are likely to be affected by a drug or a genetic variation happening on a PPI interface. We worked on several cancer related pathways, built their structural protein interaction networks and utilized the structural information on these networks in solving problems like drug off-target prediction and genotype-phenotype mapping. The structural networks models we provided will serve as a foundation for the future cancer bioinformatics, structural and functional genomics research.

We made use of network descriptions to find ways through which protein interactions can help single- and multi-target drug discovery efforts. Such structural networks may facilitate structure-based drug design; forecast side effects of drugs; and suggest how the effects of drug binding can propagate in multi-molecular complexes and pathways.

The methods introduced in this dissertation may be applied on larger datasets and the outcomes may be validated via experiments. Deepening the analysis on structural networks with such attempts may reveal important futures about structural proteomics.

Chapter 3

A NOVEL STRUCTURAL NETWORK MODEL

This chapter presents a new network model, which we name "Protein Interface and Interaction Network (P2IN)". Similar network models were used by our group previously to analyze interface properties of cancer-related proteins [204] and topological properties of hubs [205]. This new model introduces structural information into protein interaction networks (PINs). This representation illustrates which proteins may compete for the same binding site on a protein, and all protein pairs with structurally similar interface architectures.

3.1 Protein-Protein Interface Motifs and "Similar Interfaces" Concept

The 3D structures of the protein-protein complexes and their interfaces are obtained through the application of the Protein Interactions by Structural Matching (PRISM) method [140, 141, 206]. An interface is the contact region between two interacting proteins. In our studies we assume that interfaces consist of PDB chains. Interface templates are the known structures of protein complexes. These structures of interacting proteins are derived from Protein Databank (PDB) [207].

PRISM bioinformatics tool predicts possible interactions, and how the interaction partners connect structurally, based on geometrical comparisons of the template structures and the target structures. The algorithm has four steps. First, the surfaces of all target proteins are extracted. Second, using the MultiProt engine[208], the surfaces of the target proteins are structurally aligned with known interfaces (templates) obtained from the PDB. In this step PRISM checks whether any surface region of the monomers is structurally similar to one of the complementary chains of the template interfaces, disregarding the order of the residues in the protein chain. Third, it places the two chains that are structurally similar to the template interface onto the template complex. This leads to a putative complex. The fourth step involves flexible refinement of the putative complexes by FiberDock [209, 210]. This resolves steric clashes and ranks the predicted protein complexes by their energies. Combining geometric complementarity with docking tools makes the prediction more physical.

In recent years, the PRISM algorithm was applied on various signaling pathways and reasonable structural models of the unknown interactions were obtained [39, 211, 212]. PRISM was able to model the structure of protein complexes in human proteome-scale E2-E3 interactions with 76% accuracy [211] and in human apoptosis pathway with 78% accuracy [212]. Besides, the prediction performance of PRISM algorithm was recently analyzed on standard docking benchmarks, and found to be comparable to other rigid docking strategies, however considerably more efficient (see Tuncbag et al. [213]).

An interface template consists of two chains of a PDB structures. The template is named with the combination of PDB ID and chain names. For example, the template interface named "1YWK-AC" template is originating from A and C chains of structure with the "1YWK" PDB ID.

PRISM finds the similarity scores between the surface of each target in our datasets and each side of a PDB template (a template has two sides, i.e. the two complementary surfaces in the complex, in cyan and magenta, **Figure 3.1**, top line). From this output, it predicts the interface (**Figure 3.1**, bottom line). For instance, take the target protein pair in **Figure 3.1**, "TAF1" and "CDK4", and template interface "1BLX-AB"; if "TAF1" has a region on its surface which is similar to the binding site on one chain of "1BLX-AB" and "CDK4" on the second chain, then they are predicted to interact similar to the interface "1BLX-AB". This means that the binding sites of proteins "TAF1" and "CDK4" are similar to those of the protein chains of interface "1BLX-AB".

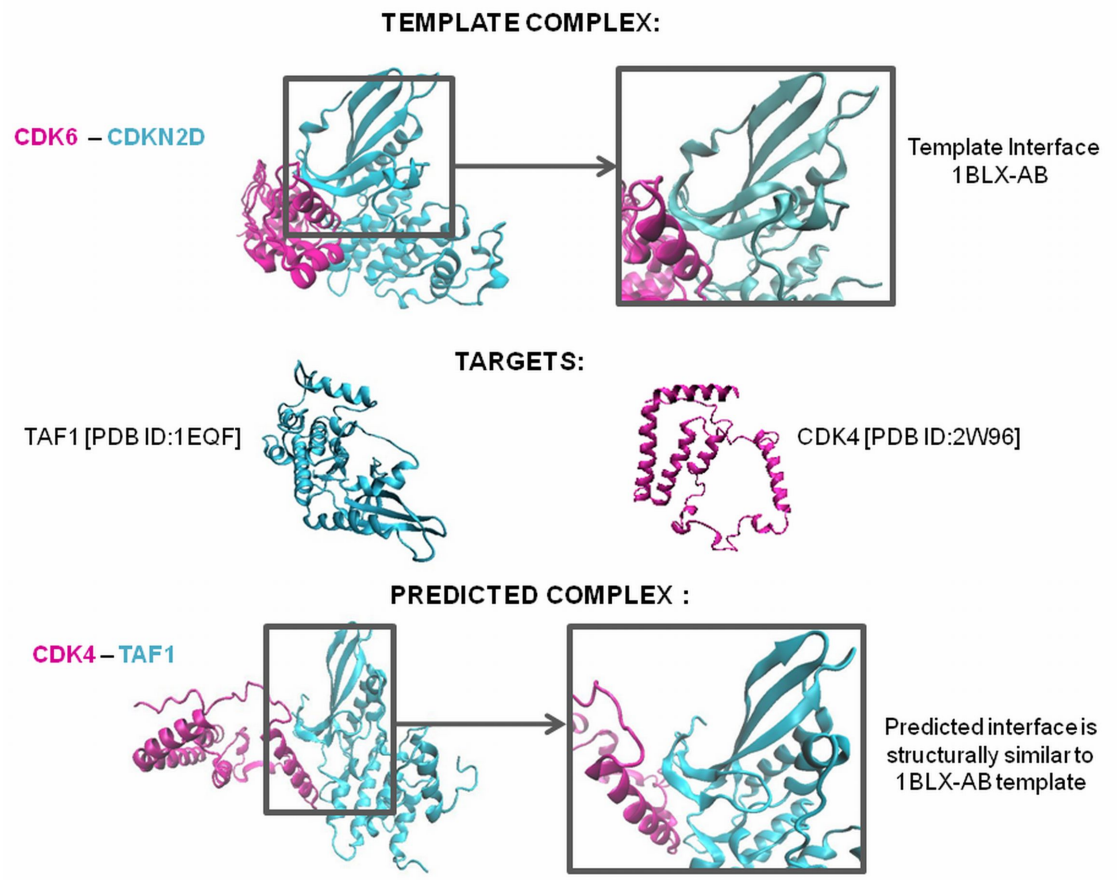

Figure 3.1. Interface Structure Prediction for Interacting Target Proteins [39]. Interface information is obtained from the "Protein Interactions by Structural Matching" (PRISM) server. PRISM searches for spatial motif similarity on target proteins' surfaces using geometric complementarity and considers evolutionary conservation of hot spots based on a non-redundant protein-protein interfaces template dataset derived from the PDB. Its prediction principle is to compare both sides of a template interface with surface regions of any given two monomers, and if they are similar these two proteins are predicted to interact with each other via this interface region. In the above example the CDK6 [PDB:1BLX-A] and CDKN2D [PDB:1BLX-B] complex is derived from PDB and the target proteins CDK4 and TAF1 are found to be interacting via an interface structurally similar to 1BLX-AB interface. CDK 4 and TAF1 are predicted to be interacting via 1BLX-AB interface.

3.2. Protein Interface and Interaction Network (P2IN) Model

PINs give binary information relating to whether two proteins communicate. Being enriched with structural information, P2IN is a more physical and realistic version of PIN. Unlike the PINs whose nodes are proteins and the interactions are the connecting edges, P2IN have interface information linked to its edges and each protein in the network has a 3D structure. Interactions between the proteins are represented by edges going through the interfaces of the two chains (**Figure 3.2**). Similar interfaces may exist between different protein pairs and the same protein pair may interact through different interfaces[214-216].

A P2IN is an undirected graph, G, that describes the interface architecture of PPIs. The edges (E) of this network happen on a set of proteins (V) and each edge is labeled with an interface name (ℓ). This undirected graph G = (V, E, ℓ) consists of a set of nodes (V), labels (ℓ) and edges (E $\subseteq$ VxVx ℓ). For example in Figure 3.2 node, edge and interface sets are as follows; V = {"CDK6", "CDKN2D"}, E = {("CDK6", "CDKN2D", "1blxAB")}, ℓ= {"1blxAB"}.

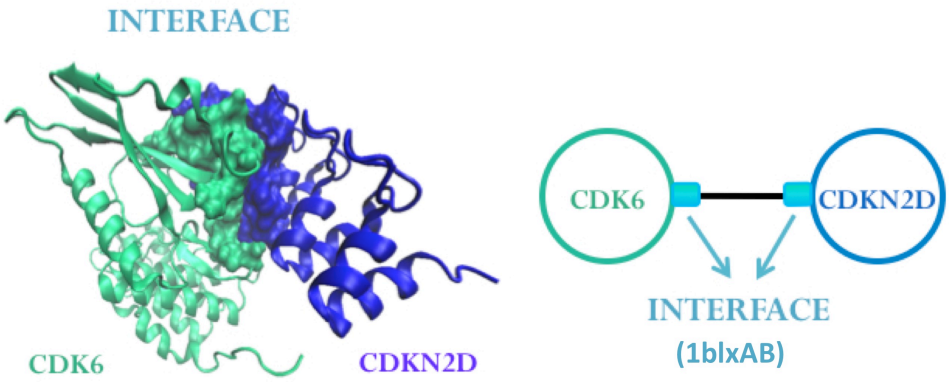

Figure 3.2. The P2IN Representation [39]. Interactions between proteins are represented by the edges going through the interfaces whose two chains represent the binding site regions of the proteins.

P2IN is capable of providing structural details that a PIN is not able to describe. Some of these details are exemplified in **Figure 3.3** : different protein pairs interacting via the same interface (CDK6 – CDKN2D and CDK4 – TAF1 interact via same interface);

a protein pair interacting using different interfaces (CDKN2D and CDK4) and multiple proteins competing to bind the same region on a protein (RAD51, CCNE1 and HDAC1 going for the same binding site on CDK6). This additional knowledge may allow identification of interactions which cannot take place simultaneously (**Figure 3.4**). Partners of a protein interacting with the same binding site cannot coexist. In addition, since ligands tend to bind proteins that have similar binding sites[142, 217, 218], locating protein pairs that interact via similar interfaces may help to predict additional, off-targets of these drugs. Thus, P2IN might be one step closer to mimicking systems-wise drugs effects [219].

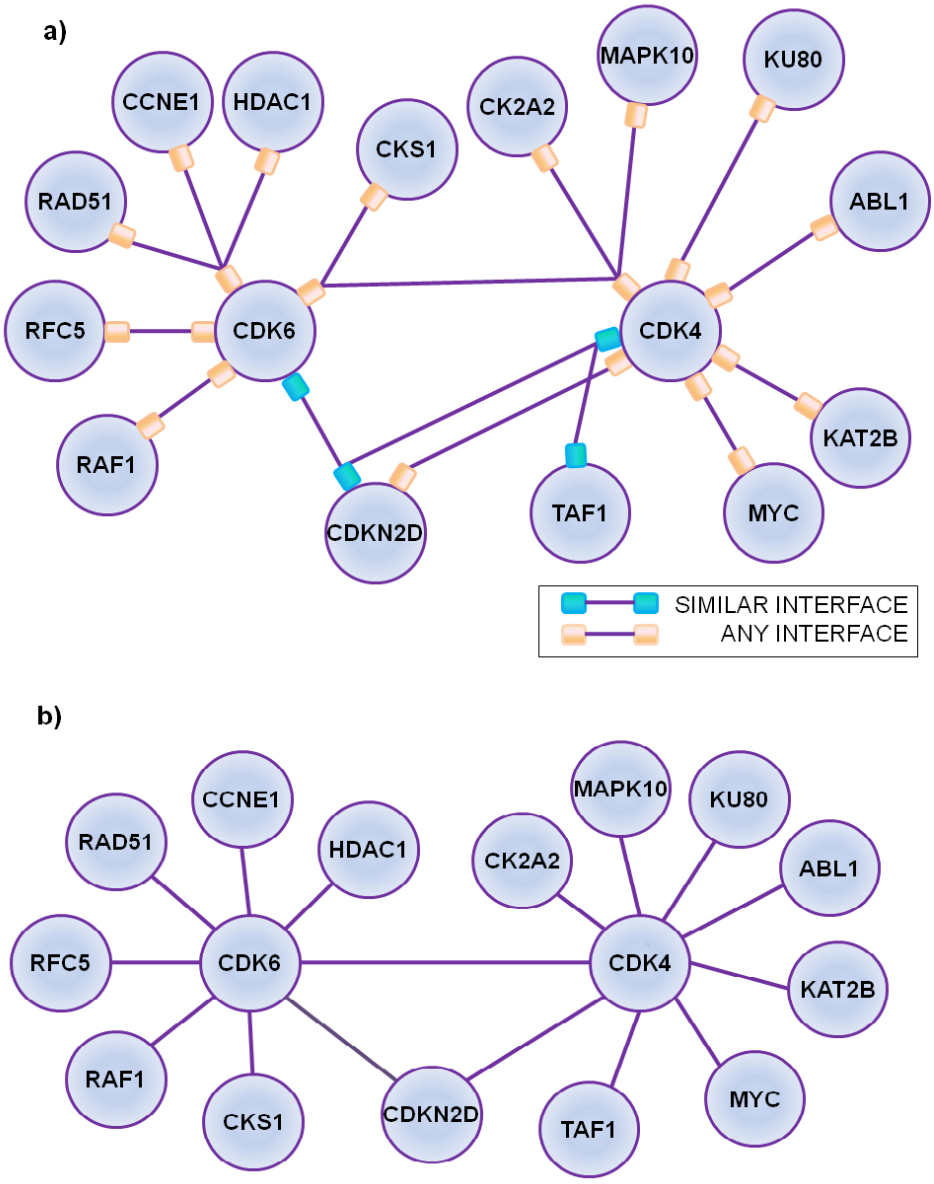

Figure 3.3. Protein – Protein Interactions and Interface Networks (P2IN) versus Protein-Protein Interaction Network (PIN) [39]. (a) A subset of PRISM predictions represented with P2IN and (b) its PIN counterpart. In P2IN the same interface may exist between different protein pairs (CDK6 – CDKN2D; CDK4 – TAF1 interact via same interface) and the same protein pair may interact using different interfaces (CDKN2D and CDK4). Moreover many proteins may compete to bind the same binding on a protein (RAD51, CCNE1 and HDAC1 bind the same site on CDK6). PIN's are not capable of depicting such structural information of protein interactions.

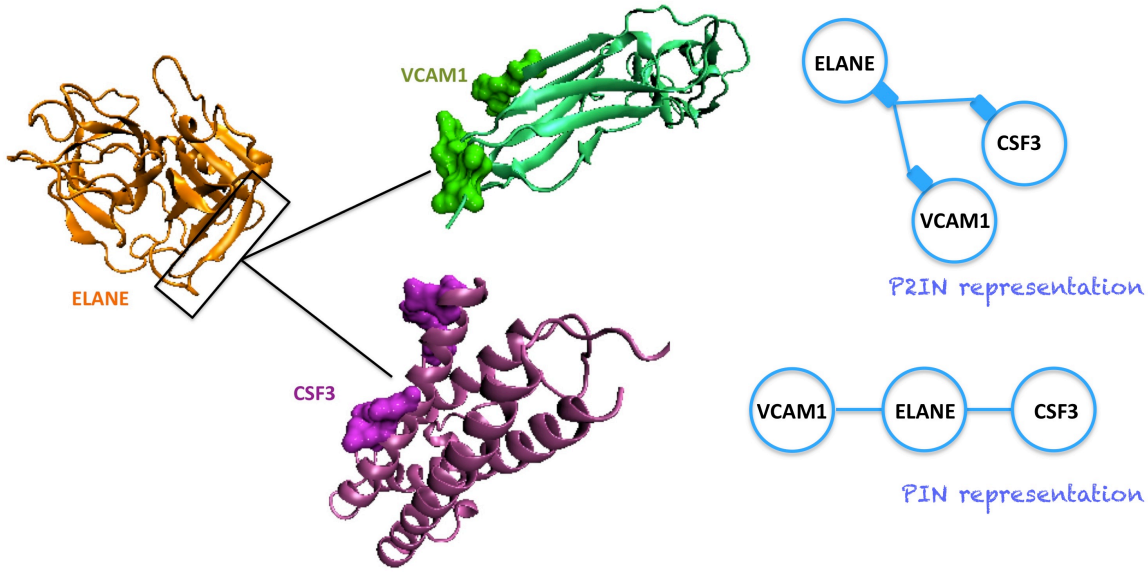

Figure 3.4. CSF3 and VCAM1 proteins competing to bind ELANE via the same binding site. Graphical representation of this phenomenon through P2IN and PIN.

3.3. Cancer Related P2INs

Through out my Ph.D. studies I dealt with the structural modeling of cancer related protein interactions and constructed a number of cancer-related P2INs. In this section I described the construction of 4 cancer related P2INs in detail. The p53 centered network, the IL10 centered network and lung/brain metastasis (derived from breast cancer) networks.

3.3.1. P53 Centered Network

The p53 tumor suppressor is a center of a protein interaction network. Under cellular stress, it is a key factor in the decision between cell cycle progression or apoptosis [36]. Stress signals may be due to failures in DNA replication, chromosome segregation and cell division [220]. Malfunction of p53 causes uncontrolled growth [221]. p53 is inactivated in more than 50% of human cancers [222, 223]. We constructed the p53 signaling P2IN using the PRISM [224, 225] predictions for this signaling pathway. Our network has 251 interactions among 81 proteins (please refer to **Table A.1** for the list of PRISM interaction predictions for p53 network). 46 different types of interface structures are observed for these interactions. 26 out of the

251 are present in Kohn's molecular interaction map (MIM) [226]; 59 are in PPI databases such as HPRD [227], Mint [228], IntAct [229], Reactome [230], BioGrid [231], Pathway Commons [232] and NCI-Nature PID [233]. 66 interaction predictions are directly experimentally validated and there is evidence in the STRING [234] database for 90 of the interactions predicted by PRISM. Overall, 104 interactions out of 251 are validated experimentally or via STRING.

3.3.2. IL10 Centered Protein-Protein Interaction Network

Inflammation by innate immunity is the first line of defense against pathogenic infections [235]. It is also involved in all phases of cancer development, including tumor initiation, promotion and metastatic dissemination [236-238]. Interleukin-10 (IL-10), identified by Mosmann and colleagues in 1989 [239], is an anti-inflammatory cytokine. It restricts the immune response to pathogens and prevents damage to the host. It is secreted by a number of immune cells and has diverse effects on many of the cell-types in the immune system.

We constructed an IL-10 centered human structural protein-protein interaction network that consists of the first and second-degree neighbors of IL-10. Each node represented a protein and each edge represented the interaction between the two proteins it connects. This network is composed of 49 proteins and 70 interactions between them (**Table A.2** and **Figure 3.5**). Among these 70 interactions only 2 (IL-10-IL-10RA and APOE-LRP1) were present in the PDB in a complex form. Accordingly, we predicted the interfaces by using PRISM and 40 additional interactions were modeled (**Figure 3.5** the edges highlighted with pink). As a result we increased the available structural interface data from 2 to 42 (**Table A.2**).

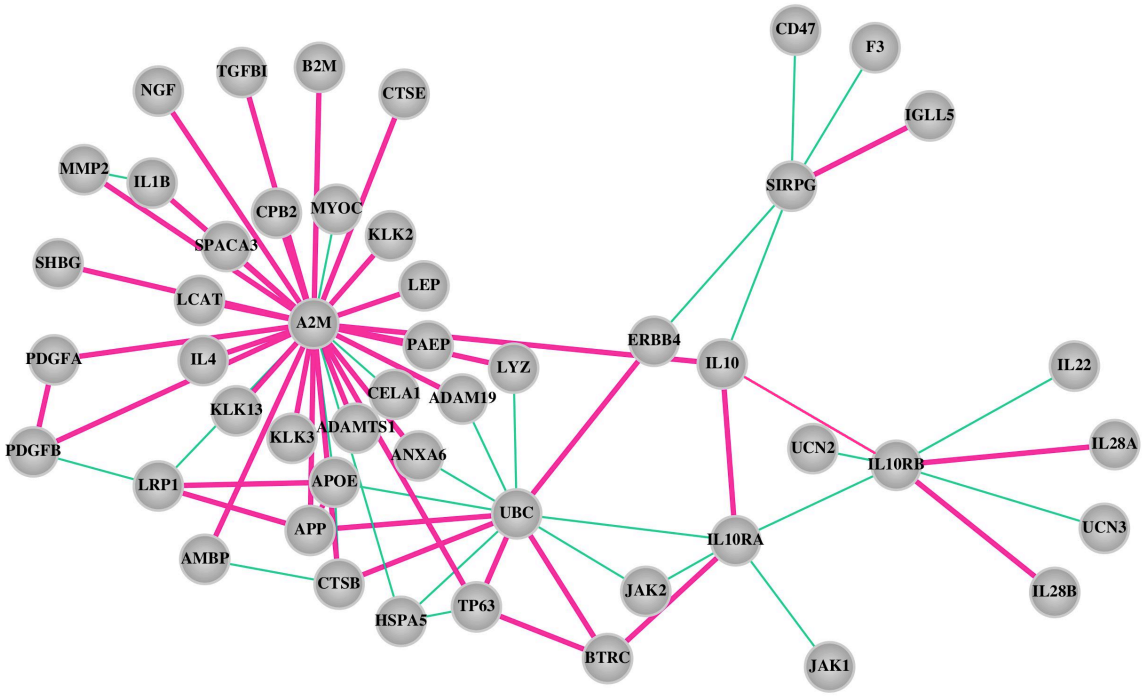

Figure 3.5. The IL-10 centered P2IN. There are 49 proteins and 70 interactions in this network and only 2 of the interactions have structural data in a complex form in the PDB. We modeled the interfaces for 40 additional interactions. Thus there are 42 interactions with interface models (edges highlighted in pink). The remaining 28 edges (out of 70) could not be modeled and are shown in cyan.

3.3.3. Lung and Brain Metastasis P2INs of Breast Cancer

In order to understand the molecular mechanism of the brain/lung metastasis of breast cancer patients, we have generated lung and brain metastatic breast cancer sub-networks by finding the most relevant edges to the seed genes identified by Massagué and his co-workers [155, 177].

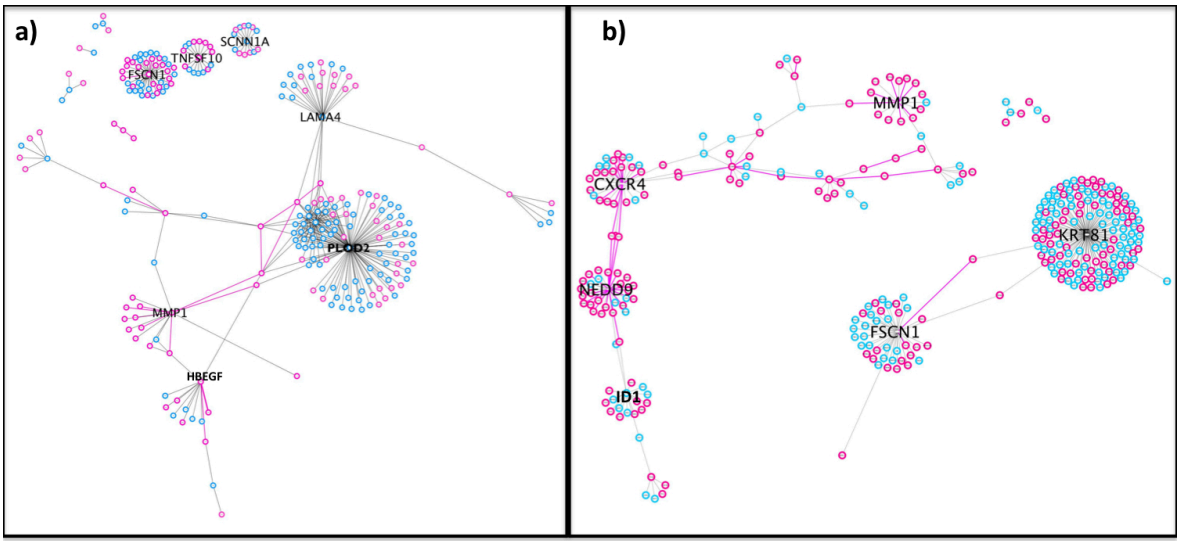

Figure 3.6. The a) brain and b) lung metastasis P2INs of breast cancer. The nodes that has structural information and the edges that has interface model are highlighted in pink.

First, we built a comprehensive human PPI network, by combining the available PPI data from various databases. Then we ranked all the interactions of this network according to their relevance to genes that are known to be mediating breast cancer to brain and lung metastasis. Subsequently, we formed two distinct metastasis PPI sub-networks from high ranked interactions. We obtained a brain metastasis sub network (BMSN) with 255 nodes and 335 edges (**Figure 3.6.a**), and a lung metastasis sub network (LMSN) with 322 nodes and 327 edges (**Figure 3.6.b** and **Table A.3**). Please refer to Chapter 5 for the details of the PPI networks' constructions.

The BMSN has 58 interactions with known 3D structures for both partners. LMSN has 102 such interactions. PRISM modeled 18 out of 58 interactions as a binary complex in the BMSN (see pink edges in **Figure 3.6.a**). For the LMSN, 50 out of 102 interactions were modeled (see pink edges in **Figure 3.6.b** and **Tables A.3 - 3.1**).

Table 3.1. Edges in both metastasis sub-networks. BMSN has 335 edges, among which 58 are connecting two proteins with 3D structures. Thus, only 58 of them may be modeled by PRISM. PRISM predicted 18 of them. Besides, LMSN has 327 interactions. Among them, 102 are connecting two proteins that have 3D structures. PRISM preformed predictions for 50 of those 102 edges.

	BRAIN	LUNG
Number of Edges	335	327
Edges that may be Modeled	58	102
Edges Modeled	18	50

3.4. Methodology

3.4.1. Preparation of the Datasets for Interface Predictions

PRISM uses template based prediction approach, and needs the 3D structure of the queried proteins. It cannot make estimation for a protein, which does not have a 3D structure. Accordingly if an edge is not connecting two proteins whose 3D structures are available, PRISM will not be able to find results for that edge. The details of preparation of the datasets for interface predictions are described in detailed in the following paragraphs (**Figure 3.7**).

The PPIs in a P2IN may be mined from a number of PPI databases (such as DIP[240], MIPS[241], HPRD[242], BIND[243], IntAct[244], MINT[245] and BioGRID [246]). Once the set of PPIs is determined, the structural knowledge related with the interacting proteins need to be gathered.

We downloaded the complete Uniprot database [247] in order to perform automated mapping of protein names to PDB structures. The main problem in this phase is that a protein may have multiple PDB IDs or its 3D structure might not be known. It is possible to have a protein which does not have any structural information, as well as a protein which has multiple PDB structures regarding a specific region on it.

So there might be a number of redundant PDB chains regarding a specific region of a protein. We used TM-align [248] in order to eliminate redundancy of similar structures corresponding to the same interface. Accordingly, we clustered PDBs that have a TM-score greater than 0.5 and an RMSD score smaller than 2.5A. Then we chose one representative, the structure that has the best resolution and the longest chain length, for each group of PDBs that describe the same region.

For the cases in which there were no structural data available, we employed the I-TASSER server [249] for generating the homology models and selected the top 5 models generated by the server.

Finally, we took the structural knowledge of both interaction partners and cross product them to translate protein interaction pairs to protein chain interaction pairs. For example, if there is an interaction between P1 and P2. Additionally, if P1 has non-redundant PDB chains PDB1 and PDB2, and P2 has non-redundant PDB chains PDB3 and PDB4; interaction between P1 and P2 will be translated to PDB1-PDB3, PDB1-PDB4, PDB2-PDB3 and PDB2-PDB4.

As described, the structural counter part of each PPI is obtained. The final input data for PRISM analysis consists of chain level interaction information of proteins.

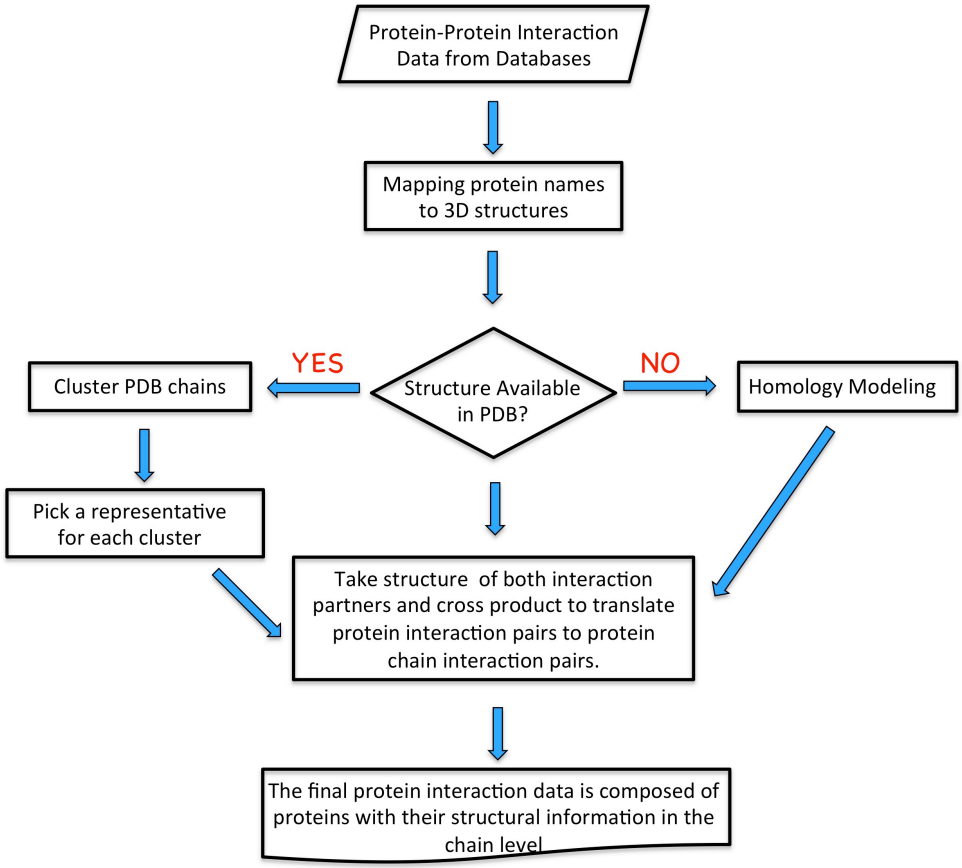

Figure 3.7. Flowchart of the preparation of the datasets for the interface analysis of Prism server.

3.4.2. Constructing Protein Interface and Interaction Network (P2IN)

The first step of building a P2IN is to gather raw data of protein interactions and their 3D structures. Protein interactions are collected from the literature and databases; the 3D structure of the interfaces is obtained from application of PRISM [224, 225]. There may be more than one possible template interface for one interaction pair; in such a case, there is more than one possible binding site between two proteins. All possibilities are considered, and every matching interface template is included in the interface and interaction networks. Proteins whose interface sites cannot be predicted by the PRISM server are discarded. This decreases the number of proteins and interactions.

3.4.2.1 P53 Centered P2IN

We studied the interactions between the proteins in the p53 signaling pathway. The list of proteins that are involved in this pathway was compiled from the literature [226] and databases by Tuncbag *et al*. [213]. Among these proteins, 85 had 3D structures in the PDB. The interaction and interface data is obtained from PRISM predictions. We used 1037 template interfaces that were extracted from the PDB [250] for the prediction process. The resulting interface predictions with energies lower than -10 are accepted.

PRISM predicted 251 interactions among 81 proteins and there are 46 different interface structures in the network. The number of proteins dropped from 85 to 81, since PRISM did not infer interactions for some proteins. If we were to link each protein in the network to other proteins, we would end up with ~3300 edges. PRISM infers 251 interactions out of those 3300 possibilities and 41% of those predictions are already known. Furthermore, in the generated p53 P2IN, there are 15 PPIs, which have PDB structures in complex form. PRISM was able to predict 13 of those interfaces correctly (**Table 3.2**).

Table 3.2. PRISM Predictions for 15 Interactions with Available PDB Structures. There are 15 interaction predictions in the p53 P2IN, which have PDB structures in complex form. Out of these interactions, PRISM made 13 correct predictions.

Predicted Interaction	Prediction Status	PDB ID
CDK2 - CKS1B	CORRECTLY PREDICTED	1BUH
CDK2 - CCNE1	CORRECTLY PREDICTED	1W98
CDK2 - CCNB1	INCORRECT PREDICTION	2JGZ
CDKN1B - CCNA2	CORRECTLY PREDICTED	1JSU
CDKN2D - CDK6	CORRECTLY PREDICTED	1BLX
MYC - MAX	CORRECTLY PREDICTED	1NKP
RAF1 - RAP1A	CORRECTLY PREDICTED	1C1Y
RELA - NFKBIA	INCORRECT PREDICTION	1NFI
RPA1 - RPA2	CORRECTLY PREDICTED	1L1O
RPA1 - RPA3	CORRECTLY PREDICTED	1L1O
RPA2 - RPA3	CORRECTLY PREDICTED	1L1O
SKP1 - SKP2	CORRECTLY PREDICTED	2AST
SKP2 - CKS1B	CORRECTLY PREDICTED	2AST
TFDP2 - E2F4	CORRECTLY PREDICTED	1CF7
TP53 - TP53BP2	CORRECTLY PREDICTED	1YCS

3.4.2.2. IL-10 Centered P2IN

We used the String server [251] for selecting the first and second-degree neighbors of IL-10. Only interactions with experimental evidence and confidence score larger than 0.4 (the default confidence value) were considered. There were 4 first-degree and 45 second-degree neighbor proteins of IL-10. Overall, we had 50 proteins comprising the IL-10 centered protein-protein interaction network.

We checked the structural data available for the 50 target proteins. We encountered 958 PDB [252] chains for 39 of the 50 proteins and for the remaining 11, we built homology models (except IGHV3-6, whose sequence information could not be found) (**Table A.4**). We employed the I-TASSER server [249] for generating the homology models and selected the top 5 models generated by the server.

We reduced the redundancy of similar interface architectures for each protein, using TM-align [248]. We classified PDB structures that have template modeling (TM)-scores larger than 0.5 and RMSD under 2.5Å. We assigned a representative PDB

structure for each similar structure group and ended up with 127 representative structures for the 39 proteins. The final IL-10 centered network is composed of 49 proteins (IGHV3-6 protein not included due to lack of structural data) and 70 interactions.

3.4.2.3. Lung and Brain Metastasis P2INs of Breast Cancer:

We searched for the 3D structural information of the proteins of lung metastasis sub-networks (LMSN) and brain metastasis sub-networks (BMSN) via the PDB. Brain metastasis network has 255 proteins and for 117 of them we found 1612 PDB structures. On the other hand, LMSN has 322 proteins and for 182 proteins we found 2712 PDB structures. In BMSN there are 58 interactions connecting proteins with known structure stored in the PDB (these interactions can be modeled with PRISM) and in LMSN there are 102 such interactions. This means that, we could only make models for these edges.

We eliminated redundancy of similar structures corresponding to the same interface using TM-align[248]. Accordingly, we grouped PDBs that have a TM-score greater than 0.5 and an RMSD score smaller than 2.5A. We chose one representative for each group of PDBs that describe the same region. We ended up with 255 PDB structures for 117 proteins of the BMSN, and with 414 PDB structures for 182 proteins of the LMSN.

In this experiment we have used 7922 interface templates (mined in 2006 from PDB) [140]. We filtered the PRISM results by considering only the interaction predictions with an energy value lower than 0. For each interface model PRISM structurally compares 2 PDB chains (target chains) to all 7922 interface templates. PRISM made multiple predictions for some of the interactions; we used the models with the lowest free binding energies.

Chapter 4
P2IN PRACTICES FOR DRUG OFF-TARGET PREDICTION

4.1. Network Attacks may Imply Effects of Drugs

The interface attack strategy proposed in this work focuses on protein-protein interface motifs. Currently protein–protein interfaces are increasingly becoming targets in drug discovery [253] [67], and it was suggested that the high flexibility of monomers may lead to overlooking small highly populated pockets that may occur when in the complex form [67]. Finding small-molecule drugs that hit protein–protein interactions is still highly challenging [49, 254-257]. Although generally interfaces of PPIs (~1500 - 3000 $Å^2$) are larger than protein-small molecule interactions (~300 - 1000 $Å^2$), an optimized small molecule may bind with an affinity comparable to that of the native partner protein or peptide [49].

Our interface attack is inspired by interface motifs and by multi-target drugs. Since drugs may disrupt protein interactions which have structurally similar interfaces, we aim to develop a strategy which may take a first step toward prediction of the outcome of disabling a set of structurally similar interactions in protein-protein interaction networks (PINs). Our study is the first to target interfaces in a network attack. A few successful PPI drugs on the market [256] such as tirofiban targeting the integrins (cardiovascular conditions) [258]; and maraviroc targeting CCR5–gp120 interactions (HIV) [259], and several new drugs entering Phase II clinical trials [260], suggest that protein interfaces can be druggable.

4.2. Interface Attack: A New Network Attack Strategy

Here we propose an attack strategy which is based on the expectation that PPI-targeting drugs may disrupt a number of protein-protein interactions which have structurally similar interfaces. Interface attack is the graphical representation of this strategy and removes interactions with similar interfaces from the network (**Figure 4.1**).

Interface attack is a kind of distributed attack, since it targets one or more interactions between protein pairs. However, instead of selecting random edges or the ones which

lead to the most damage, structurally similar interfaces are targeted. Interface attack is a knowledge-based distributed attack.

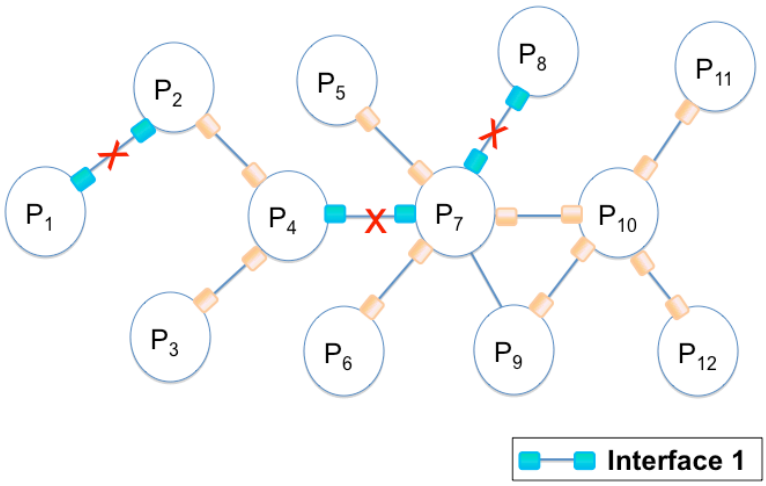

Figure 4.1. Interface Attack [39]. Interface attack hits the set of edges, which interact via structurally similar interfaces (marked with red crosses). When the interaction between P1 and P2 is targeted, the interactions between P4 and P7; P7 and P8 are also hit, since they all interact through interface 1.

4.3. P2INs may Help in Identifying Predicting Druggable Protein Interfaces and Drug Off-Targets

This section describes a case study for off-target prediction application on the interfaces of p53 P2IN. CDK6 is a regulator of cell cycle progression and affects the activity of tumor suppressor protein RB which inhibits it and keeps the cell growing in G1 phase. Inactivation through phosphorylation by CDK leads to cell cycle progression. Some CDK6 inhibitors that block the G1/S transition of cell are listed in **Table 4.1**. The drugs in this table have 3D structures in complex with CDK6 [207].

Table 4.1. List of CDK6 Inhibitors.

INHIBITOR NAME	RESOURCE	PUBCHEM ID [261]	PDB ID
Aminopurvalanol [262]	PDB	6914609	2F2C
PD-0332991 [263]	TTD[264], PDB	5330286	2EUF
CHEBI: 792519 [265]	PDB	49800099	3NUP

CHEBI: 792520 [265]	PDB	49800100	3NUX
Fisetin[266]	Uniprot[267]	5281614	1XO2

CDKN2D is a cyclin dependent kinase inhibitor, which forms a stable complex with CDK6 (**Figure 4.2.a**). The drugs listed in the table (Aminopurvalanol, PD-0332991, CHEBI: 792519, CHEBI: 792520 and Fisetin) seem to interfere with CDK6 and CDKN2D interface, when the CDK6–CDKN2D complex is superimposed on CDK6 and drug complexes present in PDB (**Figure 4.2.b – 4.2.c, Figure A.1**). The crystal structure of CDK6 and CDKN2D interface is available (PDB ID: 1BLX, chains: A, B[268]. 1BLX is a complex between human CDK6 and mouse CDKN2D. The same complex is also available for human CDK6 and human CDKN2D (PDB ID: 1BI8, chains: A, B) [269]. We considered the mouse and human CDKN2D as homologs, with 87% sequence similarity and 0.41 RMSD and used the 1BLX complex in this study since it has a better X-ray resolution). PRISM predicts an interaction between CDK4 and CDKN2D, with a structurally similar interface to the CDK6-CDKN2D interface. The interaction of CDK4 and CDKN2D is detected by *in vitro* and *in vivo* assays[270], but the 3D structure of their complex is unavailable. The interface attack by the five drugs blocking the interaction of CDK6-CDKN2D may disturb the CDK4-CDKN2D interaction.

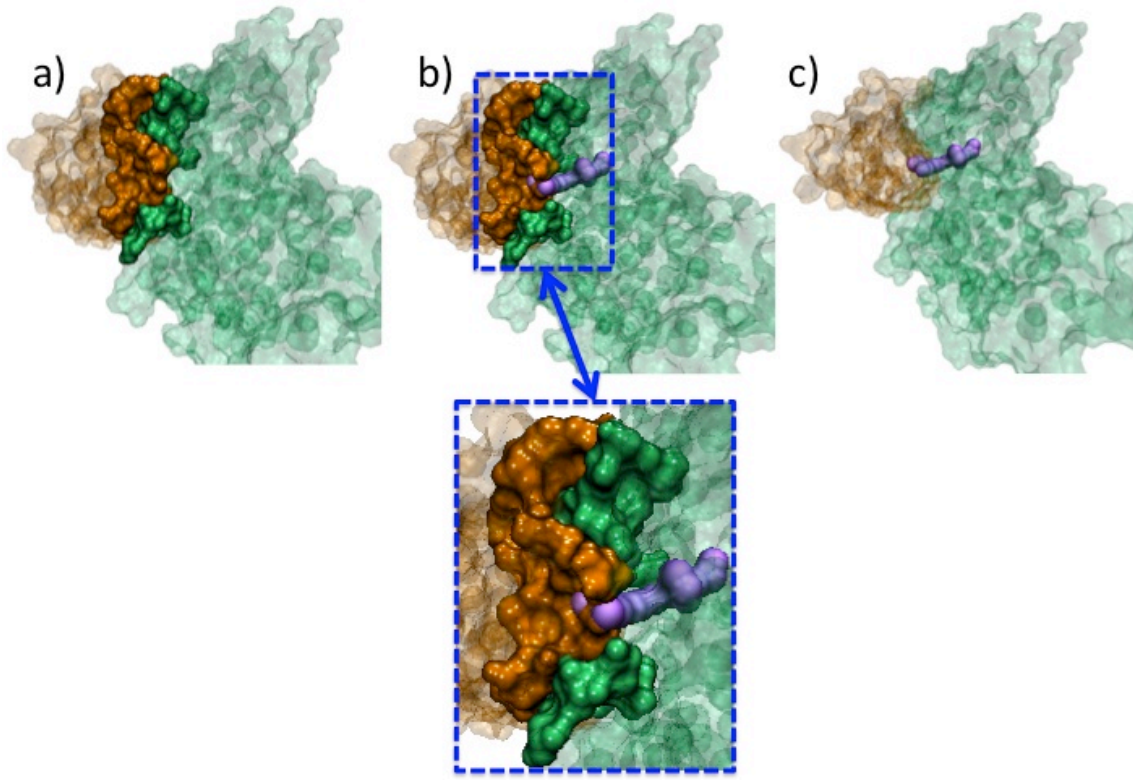

Figure 4.2. The CDK6 (green) - CDKN2D (orange) Complex and CHEBI: 792520 (purple) Interference [39]. (a) The interface of CDKN2D - CDK6 is from PDB ID:1BLX. (b,c) In the PDB, CHEBI: 792520 has a 3D structure in complex with CDK6 (PDB ID: 3NUX). When CDK6 proteins of 3NUX and 1BLX are superimposed, CHEBI: 792520 interferes with the CDK6 and CDKN2D interface. These two figures are predicted outcomes; no structural data are available.

Using the HotPoint server [271], we identified the computational hotspots of CDK4, CDK6 and CDKN2D. When the interfaces with CDKN2D are superimposed by using Multiprot engine [208], CDK4 (obtained from PRISM predictions) and CDK6 (obtained from the PDB) have a number of identical hotspots (**Figure 4.3**). CDKN2D interacts with them via the same surface area. Lastly, we found that the hotspot (CDK6 residue Ile19) that is closest to the ligand binding region on CDK6, is also present on the binding region of CDK4 (residue Ile12) (**Figures 4.4-4.5**). These drugs are also close to hotspots Gln98, and Asp97 on CDK4, and Gln103 (hotspot), Asp102 (non-hotspot) on CDK6 (**please refer to Figures A.2 – A.3).** These residues overlap when CDK4 (PDB ID: 2W96, chain: B) and CDK6 (PDB ID: 1BLX, chain: A) are

superimposed with Multiprot engine (RMSD: 1.28 Å). At this point we propose that CDK4 may be an off-target candidate for drugs targeting CDK6. In order to see how alike the binding pockets of CDK4 and CDK6 are, we superimposed the ligand binding sites using VMD[272] (**Figure A.4**). The results revealed that CDK4 has a binding pocket which is similar to that of CDK6, with RMSD 0.87 Å.

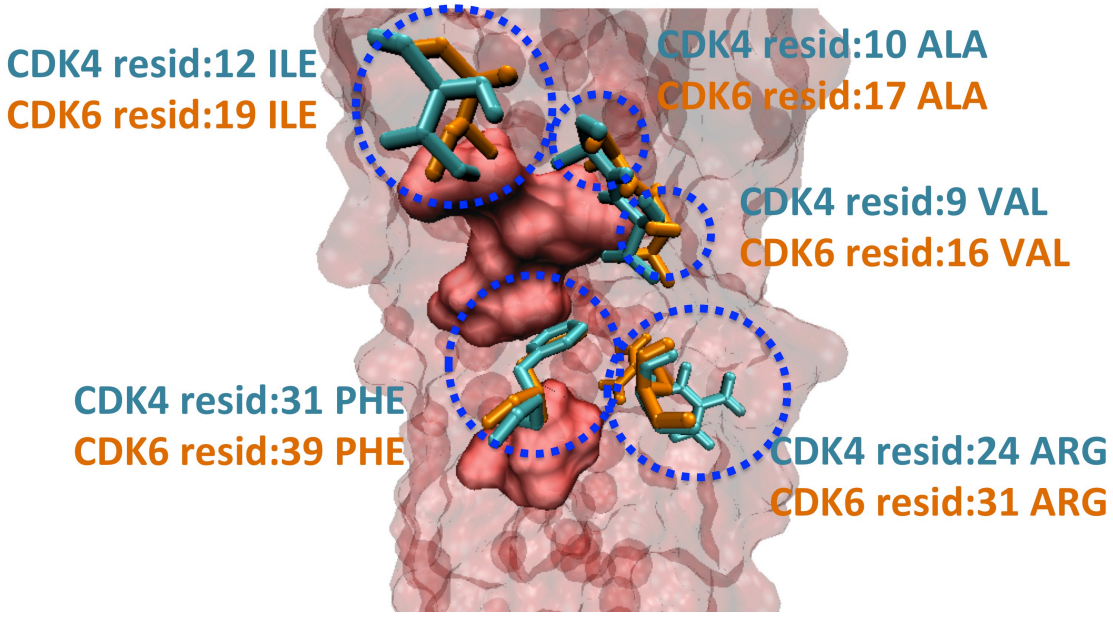

Figure 4.3. Hotspots of CDK4-CDKN2D and CDK6-CDKN2D Interfaces [39]. The predicted hotspots of CDK4 (cyan) and CDK6 (orange) proteins are represented with "Licorice" and the hotspots of CDKN2D are drawn as a (red) surface, using VMD [272]. The red, transparent body in the background is also CDKN2D protein. CDK4 and CDK6 have a number of identical hotspots, when their interfaces with CDKN2D are superimposed.

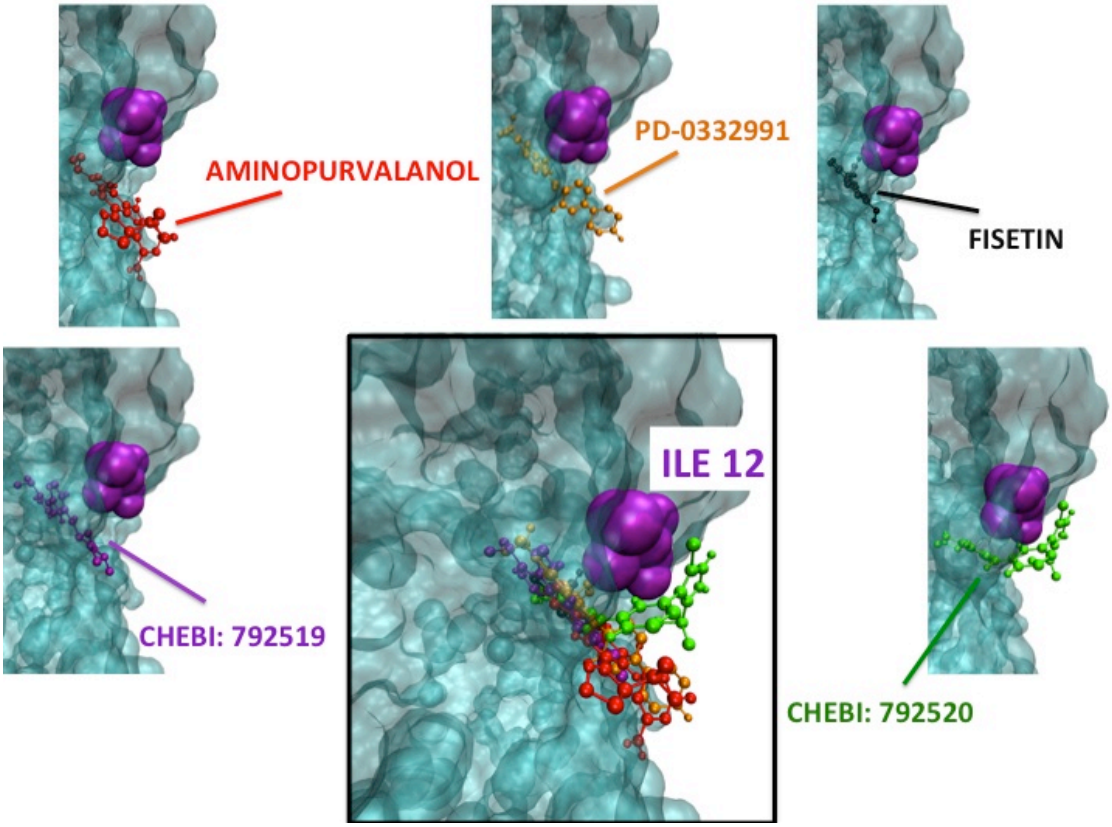

Figure 4.4. CDK4 Docking Simulations [39]. AutoDock [273] is used to dock the drugs (Aminopurvalanol, PD-0332991, CHEBI: 792519, CHEBI: 792520 and Fisetin) to candidate off target CDK4. The hotspot (CDK6 residue Ile19) that is closest to the ligands' binding region on CDK6, is also present on the binding region of CDK4 (residue Ile12).

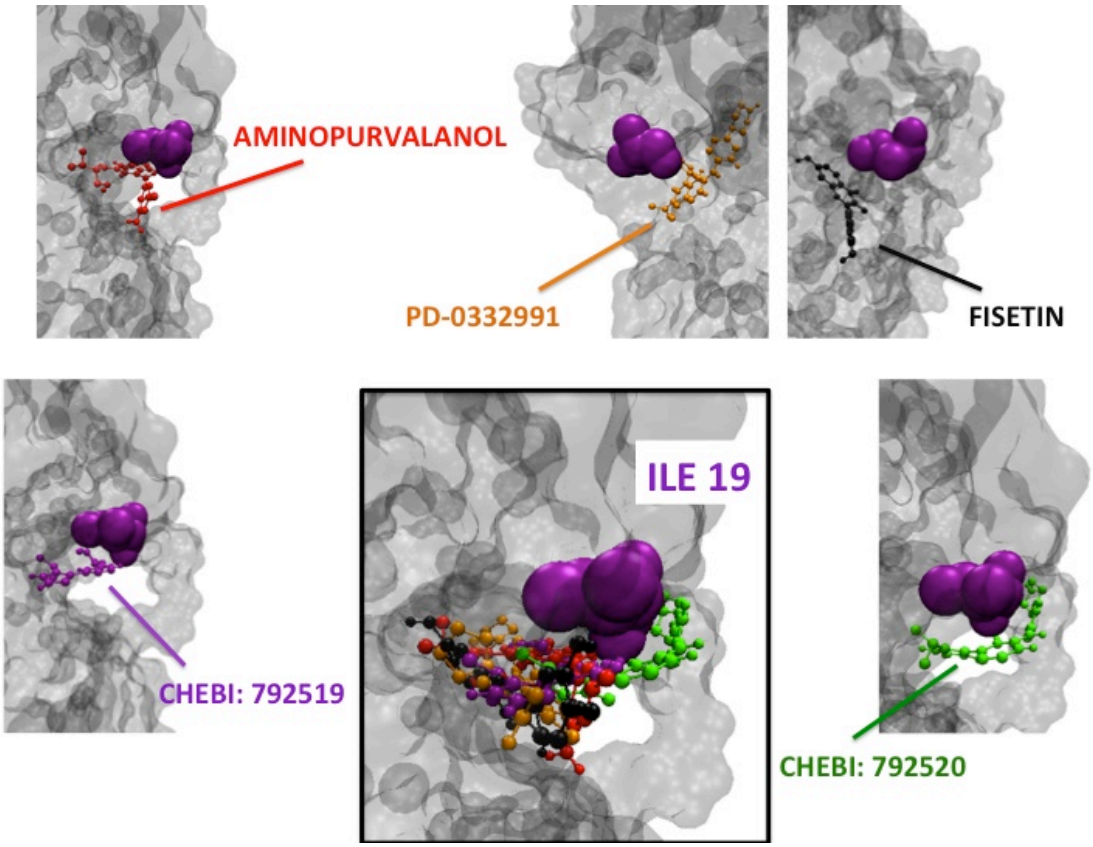

Figure 4.5. CDK6 Docking Simulations [39]. AutoDock[273] is used to dock the mentioned drugs (Aminopurvalanol, PD-0332991, CHEBI: 792519, CHEBI: 792520 and Fisetin) to primary target CDK6. The hotspot (CDK6 residue Ile19) that is closest to the ligands' binding region **on CDK6 is also present on the binding region of CDK4 (residue 12).**

Docking simulations may suggest if a ligand is capable of binding to a protein. AutoDock[273] is used to dock these drugs to candidate off-target CDK4 (**Figure 4.4**) and primary target CDK6 (**Figure 4.5**). As shown in **Table 4.2**, the binding free energies between CDK4 and the drugs are promising; they are comparable to the binding energies between CDK6 and its inhibitors. The listed energies are the lowest binding free energies of the most populated clusters. The RMSD values of superimpositions of the best poses of each drug molecule docked to CDK4 compared to CDK6 are also provided in **Table 4.2** (**Figure A.5**). These findings strengthen our proposition that CDK4 is an off-target for the drugs targeting CDK6.

Table 4.2. AutoDock [273] Results. Results given in terms of the lowest binding energy of the largest conformational clusters are in the first two rows. The RMSD values of superimpositions of the best poses of each drug molecule docked to CDK4 compared to CDK6 are in the last row.

	PD-0332991	**Fisetin**	**Aminopurvalanol**	**CHEBI: 792520**	**CHEBI: 792519**
CDK4	-8.22 kcal/mol	-7.59 kcal/mol	-5.97 kcal/mol	-7.55 kcal/mol	-6.51 kcal/mol
CDK6	-8.05 kcal/mol	-6.75 kcal/mol	-7.69 kcal/mol	-6.81 kcal/mol	-6.18 kcal/mol
RMSD	0.57 Å	0.68 Å	0.89 Å	1.83 Å	1.92 Å

Lastly, we searched for the inter-relationship between CDK6 inhibitors and CDK4 in the literature. We found that PD-0332991 has been designed to turn off both CDK4 and CDK6[274]. Moreover, SuperTarget states that CDK4 is a target of CHEBI: 792520 [275]. Accordingly, we are able to verify two of our off-target predictions. To conclude, we may now suggest that CDK6 binding drugs that block the interface between CDK6 and CDKN2D, may also bind to CDK4 and disrupt the interaction between CDK4 and CDKN2D. Therefore, when CDK6-CDKN2D interaction is hit in the interface attack, we may also break the interaction between CDK4 and CDKN2D.

4.4. Biological Consequences of Interface Attack versus Complete Node Attack

Networks of protein interactions are vital tools for explaining a series of events in the cell which may be triggered by a drug. A drug which inhibits protein-protein interactions may be represented in the network by removing the respective edges. To foresee the effects of a drug designed to inhibit all the interactions of a single protein, one can simply remove this node from the network and investigate the changes. For making an accurate functional analysis, we need all known protein interactions in the p53 pathway. We constructed a p53 network which, regardless of the structural availability, contains all known protein interactions and proteins. We simulated the changes in the network when subject to node and interface attacks. We partitioned the network using the “Affinity Propagation” algorithm [276]. This clustering algorithm

determines the representative examples (examplars) of the graph and then partitions the network according to these examplars.

We mapped the experimentally validated PRISM interface predictions of the p53 pathway on Kohn's MIM [226] as the starting point for constructing an experimentally validated network of protein interactions enriched with interfaces. We obtained a p53 PIN with 109 nodes and 227 edges. We expanded this network with the 66 PRISM predicted interfaces that were experimentally validated (26 interactions present in Kohn's MIM, 33 additional interactions from various experimental databases). We gathered a network of 115 nodes and 269 edges. Recall that there were a number of proteins from databases other than Kohn's MIM in our PRISM target. As a result the number of nodes also increased (**Figure 4.6**). The clusters generated by the Affinity Propagation algorithm are shown using pie charts (**Figure 4.7** top row). Clusters are named according to the highest degree node of that partition.

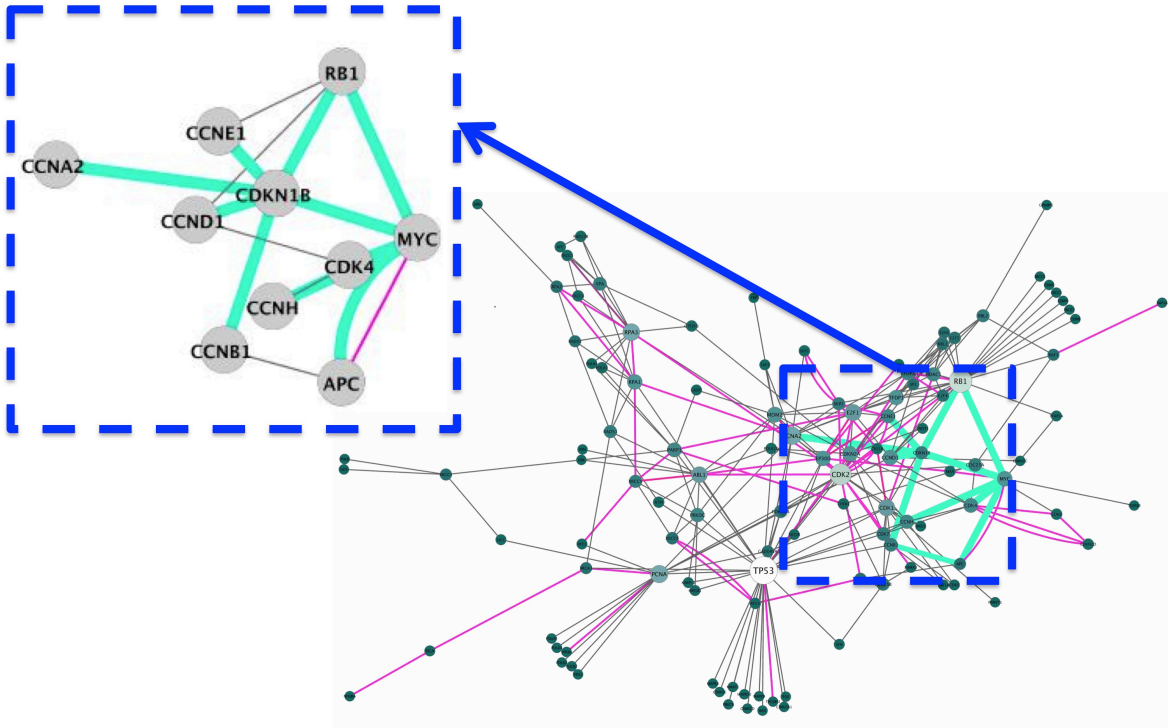

Figure 4.6. Structurally Enriched MIM Attacked Based on the 1jsuBC Interface [39]. Experimentally validated edges of p53 P2IN mapped on the Kohn's MIM [226]. The edges with interface structures are shown in pink color and the edges with 1jsuBC

interface is highlighted in green. In the close-up figure edges with 1jsuBC interfaces are also can be seen in green.

When the 1jsuBC interface (template interface is between the CCNA2 and CDKN1B proteins) is attacked, 11 edges are removed from the network. Six of these are around the CDKN1B node. Therefore, this node is completely removed from the network by the 1jsuBC interface attack, in addition to the removal of 5 edges around other nodes. One can see that this attack causes the cluster, with the RB1 hub node, to get significantly bigger (please refer to the slices of RB1cluster in the top and middle rows of **Figure 4.7**). RB1 now has a greater influence on the network. MYC is no more the hub node of a cluster (red slice present in the top row of **Figure 4.7** dissapears in the middle row of **Figure 4.7**) and the cluster of CDK2 enlarges from 9% of the nodes of network to 16% (**Figure 4.7** middle row). A complete node attack targeting the CDKN1B protein, means breaking all of this node's interactions detaching it from the network. PRISM predicts that all 6 interactions of CDKN1B have a similar structure to 1jsuBC interface. Thus, to block all of the interactions of CDKN1B, a drug has to attack the 1jsuBC interface, which affects 5 more edges in the network. However, in the case of complete node attack on CDKN1B, only edges of this node are discarded from the network. We do not observe a significant change in the sizes of the clusters following complete node attack (see top and bottom rows of **Figure 4.7**).

The changes observed after the interface attack appear reasonable. During the 1jsuBC interface attack, CDKN1B is removed from the network, CDK2 cluster gets bigger and the influence of this protein on other nodes increases. Since CDKN1B has inhibitory activity on some CDK2 complexes[277], this change is expected. Once MYC is not a hub in a cluster, the RB1 cluster expands. In the presence of MYC, the RB1 transcription is suppressed and MYC activates a set of miRNAs, which in turn inhibit the translation of RB1 [278]. Finally, CDKN1B and RB1 are tumor-suppressors. The RB1 cluster gets bigger when CDKN1B loses all of its interactions, possibly suggesting that RB1 may be involved in an alternative pathway.

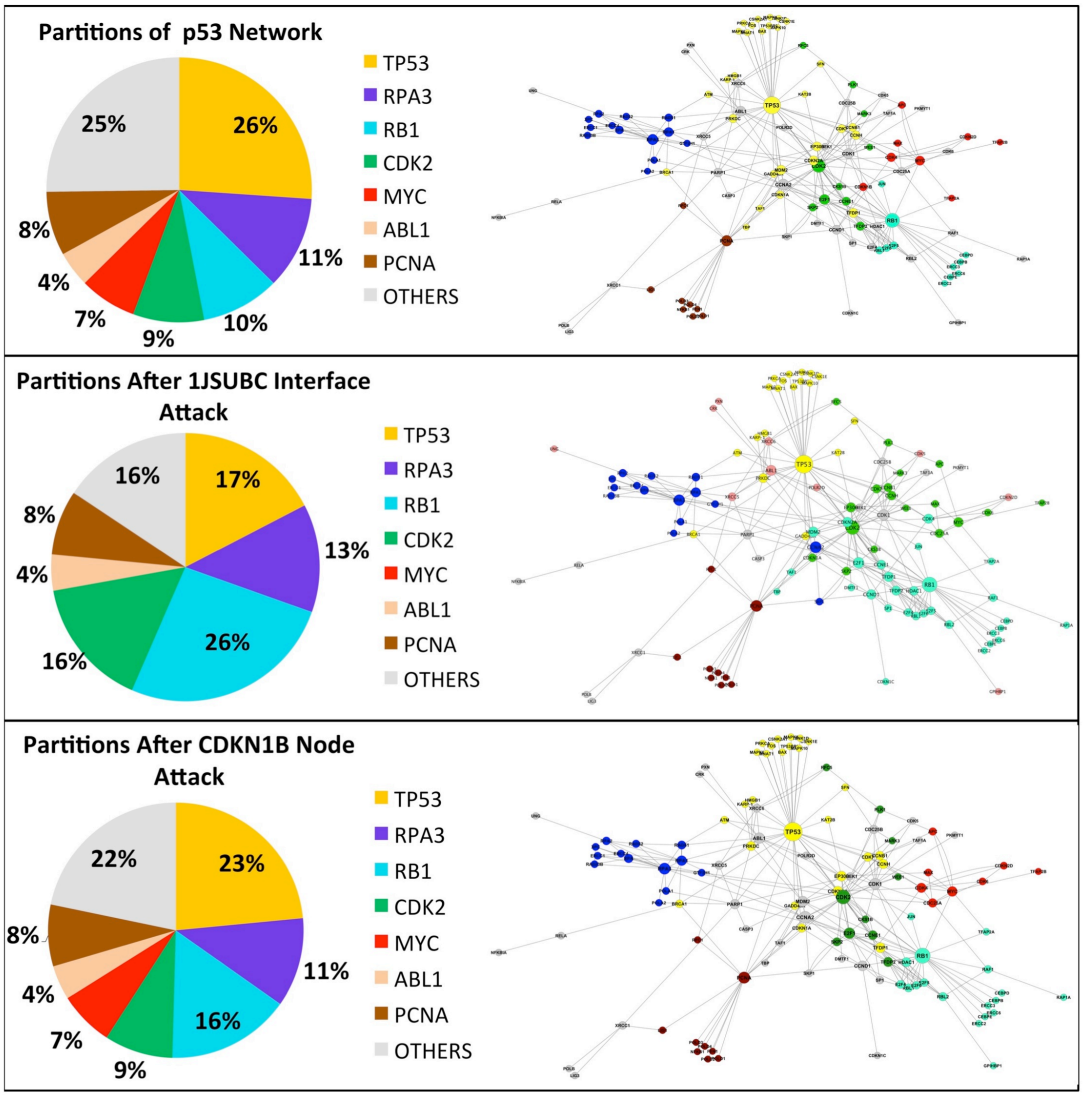

Figure 4.7. Pie Charts of Clusters Generated with Affinity Propagation Algorithm [39]. In the pie charts each slice represents a cluster and they are named with the clusters' hub nodes. Percentages of the slices are the ratio of the node number in the corresponding cluster to the total number of nodes in the network. (Top row) Clusters of the network generated by mapping the experimentally validated PRISM predictions of p53 pathway onto Kohn's MIM. (Middle row) The clusters after 1jsuBC interface attack. (Bottom Row) The clusters after CDKN1B node attack.

4.5. Network Attack Scenarios Applied to P53 P2IN and Changes in the Network Robustness

P53 P2IN is a small sub-network, and it does not have a scale-free architecture. The average number of interfaces per node is 3.24 in this P2IN. Besides, its clustering coefficient is 0.197, its network diameter (the largest distance between two nodes) is 7, its network radius (the minimum distance, among the non-zero distances, between two nodes) is 4, its characteristic path length (the expected distance between two connected nodes) is 2.672 and its average number of neighbors (average connectivity of a node) is 5.926.

The robustness of a network relates to its ability to withstand the damage caused by attacks. It can be expressed by topological parameters. The most commonly used robustness parameters are the average inverse geodesic length (AIGL) [34, 279] and the giant component size [34] (GCS). To monitor the change in the connectedness of the nodes in the system, we use both.

For the p53 P2IN survivability analysis, several attack types and target selection strategies are used. These attack scenarios refer to partial or complete knockout of hub nodes and deletion of multiple edges that are scattered around the network. At each step a new target is hit and the topological parameters are recalculated until the network is left without interactions.

4.5.1. Hub Node Attack

A hub is the highest degree node of the network; it is the node that has the largest number of interactions. This attack type targets the largest degree node of the network. Hitting this element also affects its interacting partners and causes a serious disturbance in the network communication.

4.5.2. Frequent Interface Attack

In P2IN, the number of occurrences of each interface type is known. In this strategy the most frequently observed interface is selected as the target of interface attack.

4.5.3. Maximal Damage Strategy

The maximal damage strategy is a greedy algorithm, which was studied by Agoston *et al. [30]*. It hits the component that will harm the network the most in each attack. This

tactic may be used in both node and edge attack types. Removing multiple edges that are selected according to the maximal damage target selection strategy is a kind of a distributed attack. It targets the node or interface that is expected to cause the greatest possible harm.

4.5.4. Frequent Interface Attack is as Harmful as Complete Hub Knockout and it is a More Realistic Scenario

Breaking an edge can be considered as the graphical representation of a drug blocking the interaction of two proteins. If we were to map node-targeted attacks (complete or partial knockout) to a drug mechanism, it would be a “magic bullet"; even if a drug would specifically bind to one protein, in most of the cases it may not obstruct all of its interactions. It seems that complete knockout is rarely observed in realistic drug action. The common “similar binding sites should recognize similar ligands” strategy[280], motivated us to develop the interface attack.

Complete/partial hub node attacks and interface attacks based on their frequencies of occurrence are performed on the p53 P2IN. In **Figure 4.8** the change in the network robustness is plotted according to AIGL and GCS. The x-axis stands for the number of attacks, while the y-axis is the AIGL or GCS values during the attacks. A drop in AIGL or GCS of the network correlates with the damage caused to the system. The plots show that attacking the most frequent interface in the p53 signaling network is at least as harmful as complete removal of the hub nodes from the network. Thus, rather than targeting a well connected protein, which is more likely to be essential [35], we may target edges that have similar interface structures.

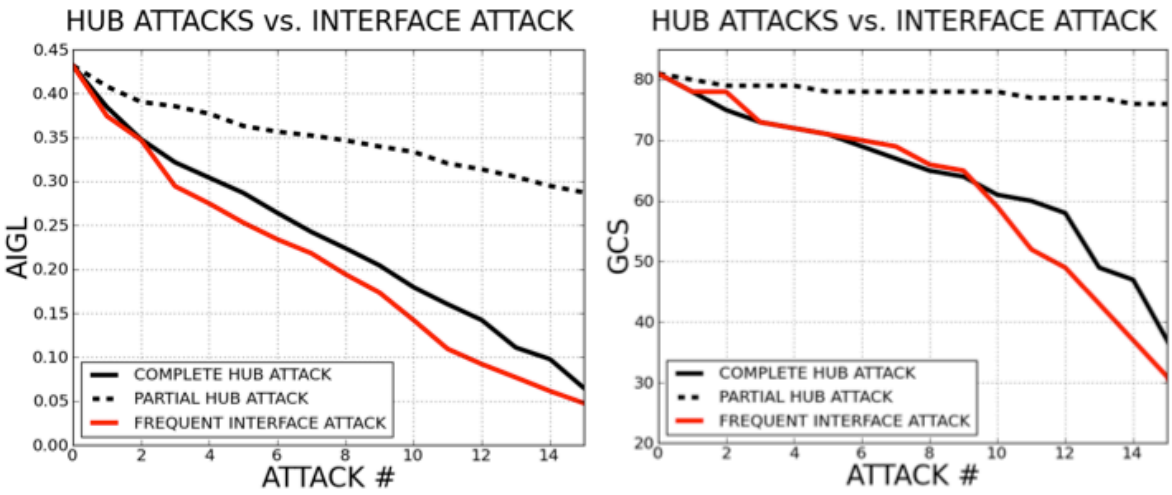

Figure 4.8. Hub Node Attack versus Most Frequent Interface Attack [39]. The figure plots of the damage to the network following 15 successive complete hub node attacks, partial hub node attacks and frequent interface attacks (for AIGL (on the left) and GCS (on the right) topological parameters). The results suggest that the most frequent interface attack and complete hub knockout lead to roughly the same damage, while the effect of the partial hub knockout is to a lesser extent.

The most frequent interface (PDB ID: 1JSU, chains: B, C) is observed 46 times in the p53 network. 21 of these predictions are validated experimentally or present in the STRING database. If there was a drug designed to disturb one of these 46 interactions, not just that particular edge, but all 46 interactions could be hit. This interface is not focused around a hub node, however many high degree nodes of the p53 P2IN utilize it (**Figure 4.9).** During a possible attack some of the hub nodes will also be partially affected. Hence, building the interface and interaction network of a biological system may provide us such insights and may be helpful for drug development.

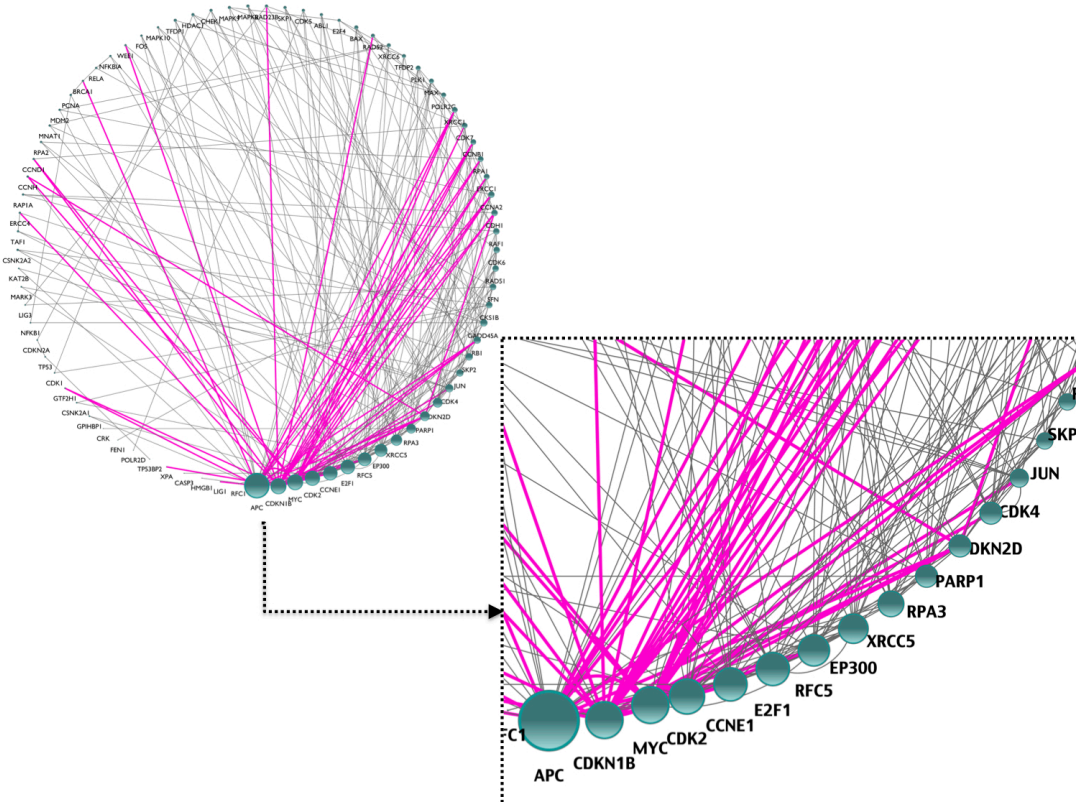

Figure 4.9. Degree sorted circular layout of p53 P2IN. Most frequent interface 1jsuBC is utilized by the edges highlighted in pink. The node sizes are proportional to their degrees.

4.5.5. Interface Attack is not as Harmful as Distributed Attack when Maximal Damage Strategy is Applied

Agoston *et al*. [30] showed that rather than removing a node completely from the network, one could inflict similar damage by removing a number of edges distributed around the network. They chose the most destructive edges.

Interface attack is a kind of distributed attack, but it chooses the target edge set based on interface similarity. We performed distributed attacks and interface attacks on the p53 P2IN. In this experiment we followed a maximal damage target selection strategy, by selecting the most damaging edges (distributed attack) or the most damaging interface in successive attacks. The comparison of the damage caused by distributed attack and interface attack is plotted in **Figure 4.10**. The x-axis is the number of edges removed during attacks and the y-axis the change in the network GCS and AIGL. It is

clear that distributed attack harms the network more than interface attack. However, comparison of interface attack and distributed attack is not straightforward, since distributed attack selects edges one by one, while interface attack chooses between sets of edges. This is why distributed attack is so harmful and is nearly the optimal attack strategy for collapsing the network. However, interface attack seems to be physically more suitable for simulating the impact of multi-target drugs on the network, since the interactions affected by multi-target drugs are not always the most harmful.

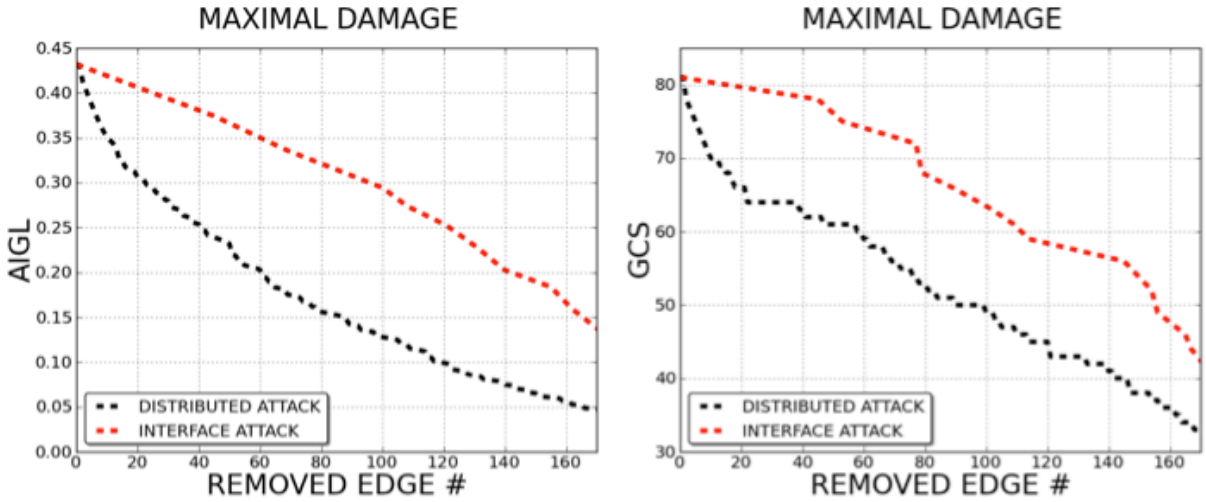

Figure 4.10. Maximal Successive Damage Strategy on Distributed and Interface Attack [39]. Damage in the network (both according to AIGL (on the left) and GCS (on the right) topological parameters) is monitored, under successive attacks. Distributed attack and interface attack are executed using the maximal damage strategy. The number of edges removed from the network in each attack is parallel to the harm attacks cause on the network. It is obvious that distributed attack is the most harmful strategy.

4.5.6. Frequent Interfaces are not Observed on Topologically Critical Interactions

When random edge attacks are compared with frequent interface attacks (**Figure 4.11**) according to the change in giant component sizes, the most frequent interface attack is less harmful to the p53 P2IN than random edge attacks.

However, when the attacks are performed on randomly selected interfaces, we observe that on average they harm the network more than random edge attacks. Consequently, random interface attacks are more harmful to the network than frequent interface attacks; that is, a frequent interface is less likely to hit topologically critical elements of the network. This makes the network more resistant to failures.

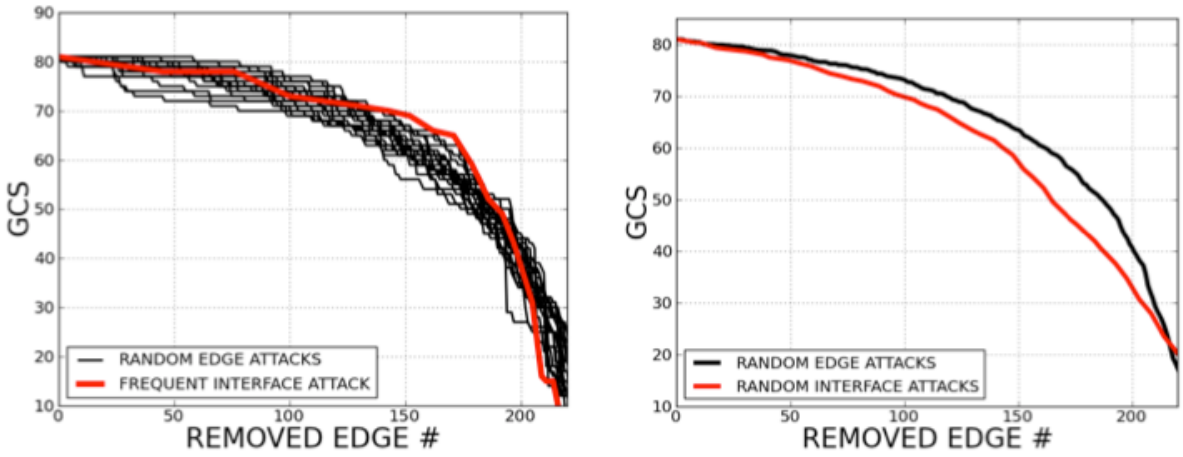

Figure 4.11. Random Edge Attacks versus Interface Attacks [39]. The most frequent interface attack is relatively less harmful to the p53 P2IN than random edge attacks (on the left). However, the average of random interface attacks harms the network more than the average of random edge attacks (on the right). Consequently, random interface attacks give more harm to the network than frequent interface attacks.

4.6. Methodology

4.6.1. Docking Parameters

For adding polar hydrogens, assigning Gasteiger charges and drawing grid boxes AutoDockTools 1.5.4[273] was used. Binding affinities were calculated with AutoGrid version 4. Lamarckian genetic algorithm (trials of 50 dockings, population size of 150, and maximum number of generations of 27000) was used to do the docking experiments using AutoDock 4.0 [273].

4.6.2. Clustering Algorithm

We partitioned the network according to the "Affinity Propagation" algorithm [276] with the help of Clustermaker plugin [281] of Cytoscape [282].

4.6.3. Mapping the experimentally validated PRISM interface predictions of p53 pathway on the Kohn's MIM

Kohn's MIM has some nodes that do not have a protein counterpart, or some nodes correspond to multiple proteins. Before constructing the PIN, we updated nodes in Kohn's MIM by removing or expanding some of them (**Table A.5**). If a node was replaced with multiple proteins, the number of interactions automatically increased. We searched the String database for validating the new edges and picked the ones which were coming from experiments or databases. For example, the "CDK4-6" node corresponds to three proteins (CDK4 – CDK5 – CDK6). In the original map there was an interaction between "CDK7" and "CDK4-6". The "CDK7" interactions with CDK4 and CDK5 are validated, but not with CDK6. The full list of interactions can be found in **Table A.6**.

4.6.4. Robustness Measures

AIGL is the sum of the inverses of all shortest paths, divided by the number of possible node combinations. The definition is given in Equation 1. The notation used is as follows:

ℓ = average geodesic length

n = number of nodes

i, j = proteins

d_{ij} = distance between proteins i and j

If there is no path connecting nodes i and j, the distance between them is set to infinity. Some studies use the average geodesic length but we preferred to use AIGL.

Even after several attacks, AIGL will not be equal to infinity, because if there is no navigable route between i and j, $\frac{1}{\text{dij}} = 0$.

$$\ell^{-1} = \frac{1}{(n)(n-1)} \sum_{i \neq j} \frac{1}{d_{ij}} \quad \text{(Eq. 1)}$$

GCS is the number of nodes in the network's largest connected sub-graph and it may give important clues about the collapsing mechanism of network under attacks.

NetworkX [283], a Python language software package, was used for the damage simulations on p53 centered network.

Chapter 5

P2IN PRACTICES FOR LINKING GENOTYPE TO PHENOTYPE

In this chapter, on behalf of understanding the molecular mechanism of the brain/lung metastasis of breast cancer patients, we have generated lung and brain metastatic breast cancer sub-networks by finding the most relevant edges to the seed genes identified by Massagué and his co-workers [155, 177]. Then, we enriched these networks with structural information of 3D structural models of known protein-complexes and predicted its protein-protein interfaces. We have analyzed the protein-protein interfaces commonly employed in these sub-networks and observed that interactions of microbial origin played an important role. We also investigated the mutations happening on the most relevant proteins of the breast cancer metastasis sub-networks. (**Figure 5.1**)

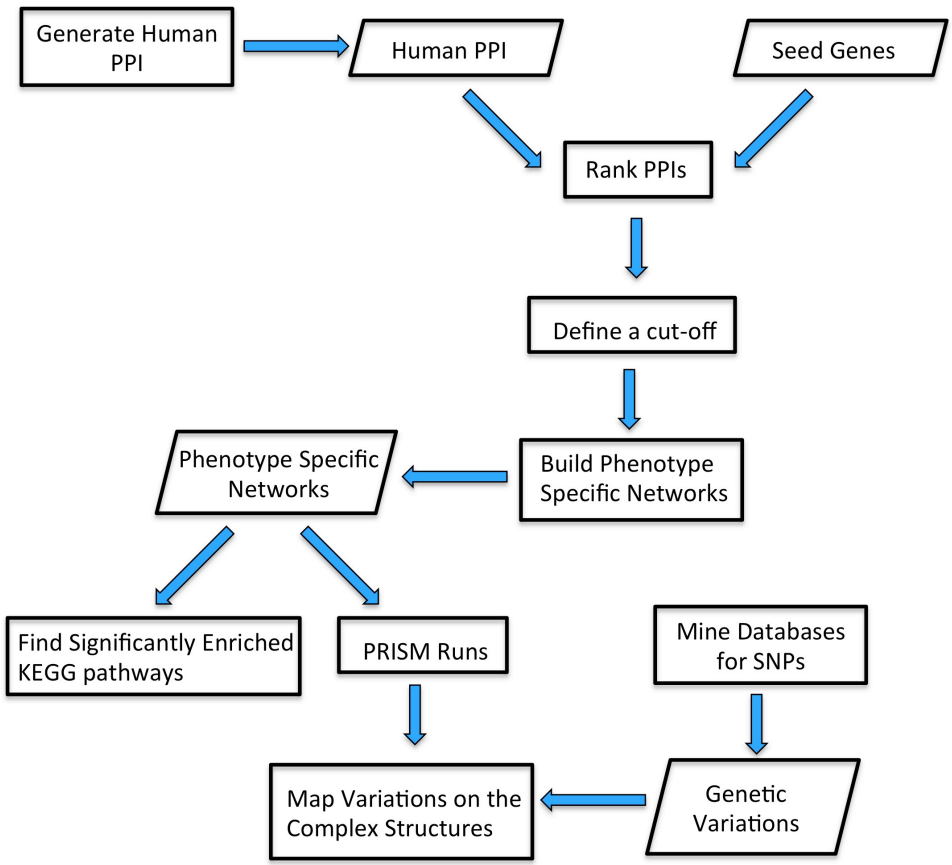

Figure 5.1. Flow chart of the bioinformatics pipeline designed for genotype-phenotype mapping.

5.1. P2INs of Lung and Brain Metastasis Driven from Breast Cancer

5.1.1. Identifying Brain & Lung Metastatic Breast Cancer Sub-networks and Their Functional Annotations

We have built a comprehensive human PPI network that consisted of 11,123 proteins and 149,931 interactions. We ranked each PPI in the network, according to its relevance to the seed nodes causing breast cancer metastasis, using GUILD (Genes Underlying Inheritance Linked Disorders) network-based prioritization tool [284].

We defined a score threshold and discarded interactions below the threshold based on the following reasoning: 1) we need two comparable sets of nodes and edges for brain and lung metastasis, where the topology may be different but not the size; 2) predicting the interface structures of interacting proteins is a highly time-consuming step, therefore we needed to reduce the network to a limited sub-network of small but highly relevant edges (i.e. less than 500) for each metastasis under study.

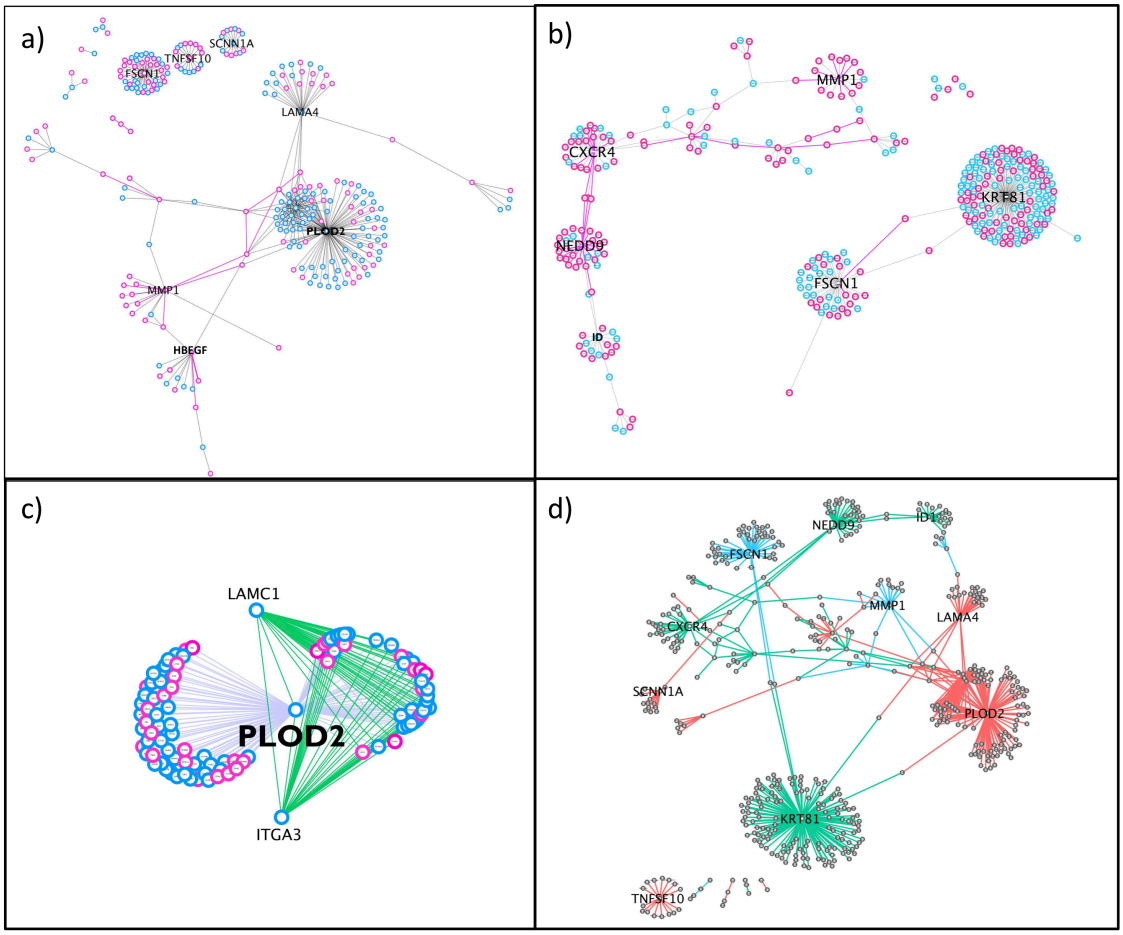

Figure 5.2. The BMSN and the LMSN networks [285]. We obtained a) the BMSN and b) the LMSN by choosing the edges of human PPI network with GUILD Score higher than 0.178. The proteins that have PDB structures are highlighted in pink, plus the edges that have complexes modeled by PRISM are also in pink color. c) PLOD2 cluster (the first-degree neighbors of PLOD2) from the BMSN d) BMSN and LMSN merged as a one big network. There are 84 common proteins and 71 common PPIs (blue edges). The edges that are only present in LMSN are shown with green and the edges that are only present in BMSN are shown with pink.

We plotted the number of edges versus their scores to select the best cut-off (see **Figure A.6**). We observed a dramatic rise in the number of interactions (and also nodes), between scores 0.15 and 0.18 for the punctuation of brain and lung metastasis (**Figure A.7 and Table 5.1**). Accordingly, we selected 0.178 as the common GUILD cut-off score to generate both sub-networks**.** This cutoff yielded a brain metastasis BMSN with 255 nodes and 335 edges **(Figure 5.2.a)**, and a lung metastasis LMSN with 322 nodes and 327 edges **(Figure 5.2.b)**.

Table 5.1. The number of edges and nodes of metastasis networks according to Guild Scores.

	BRAIN METASTASIS		LUNG METASTASIS	
CUTOFF VALUES	**#OF NODES**	**#OF EDGES**	**#OF NODES**	**#OF EDGES**
Score 0.140	276	5382	354	7085
Score 0.170	255	4220	322	328
Score 0.178	255	335	322	327

Although we used all proteins of both sub-networks (BMSN and LMSN) in our analyses, we tracked down the evidence for the expressions of the genes that coded the proteins in both sub-networks in breast tissue. We found that 87% of the genes in the LMSN (280 out of 322, see **Table A.8)** and 93% in the BMSN (238 out of 255, see **Table A.9)** are expressed in breast tissue.

We used ClueGo [286] to find significant KEGG pathways in BMSN (**Table 5.2)** and LMSN (**Table 5.5)**. Each pathway in KEGG belongs to a class according to KEGG Orthology (KO) [287]. Then we mapped each KEGG pathway to its KEGG class. Subsequently, we calculated the percentages of observed KEGG classes (**Figure 5.3**).

We found out that “Transport and Catabolism Cellular Processes” and “Replication and Repair Genetic Information Processing” classes contain the most abundant significant pathways in BMSN “Infectious Diseases”, “Cancer” and “Immune System” were the classes of most abundant pathways in the LMSN.

Table 5.2. The KEGG pathways enriched (P<0.05) in brain metastasis network with respect to ClueGO p-value are listed in this table.

PATHWAY NAME	ClueGo PValue (after Bonferroni Correction)	KEGG Class
path:hsa04142 Lysosome	5.47E-14	Cellular Processes; Transport and Catabolism
path:hsa05222 Small cell lung cancer	4.86E-06	Human Diseases; Cancers
path:hsa04210 Apoptosis	3.50E-05	Cellular Processes; Cell Growth and Death
path:hsa03430 Mismatch repair	0.001638707	Genetic Information Processing; Replication and repair
path:hsa04145 Phagosome	2.47E-04	Cellular Processes; Transport and Catabolism
path:hsa04640 Hematopoietic cell lineage	0.001806729	Organismal Systems; Immune System
path:hsa04960 Aldosterone-regulated sodium reabsorption	0.042397972	Organismal Systems; Excretory system
path:hsa05146 Amoebiasis	0.009476797	Human Diseases; Infectious Diseases
path:hsa03460 Fanconi anemia pathway	0.027029128	Genetic Information Processing; Replication and repair

Table 5.3. The KEGG pathways enriched (P<0.05) in lung metastasis network with respect to ClueGO p-value are listed in this table.

PATHWAY NAME	ClueGo PValue (after Bonferroni Correction)	KEGG Class
path:hsa04062 Chemokine signaling pathway	8.53E-16	Organismal Systems; Immune System
path:hsa03010 Ribosome	1.41E-08	Genetic Information Processing; Translation
path:hsa04670 Leukocyte transendothelial migration	1.65E-07	Organismal Systems; Immune System
path:hsa05100 Bacterial invasion of epithelial cells	7.15E-06	Human Diseases; Infectious diseases
path:hsa04512 ECM-receptor interaction	8.72E-05	Environmental Information Processing; Signaling Molecules and Interaction

path:hsa04012 ErbB signaling pathway	1.16E-04	Environmental Information Processing; Signal transduction
path:hsa04722 Neurotrophin signaling pathway	1.20E-04	Organismal Systems; Nervous system
path:hsa04530 Tight junction	8.94E-04	Cellular Processes; Cell Communication
path:hsa04520 Adherens junction	0.00293406	Cellular Processes; Cell communication
path:hsa05142 Chagas disease (American trypanosomiasis)	0.004287851	Human Diseases; Infectious diseases
path:hsa05160 Hepatitis C	0.004273745	Human Diseases; Infectious Diseases
path:hsa05212 Pancreatic cancer	0.010078577	Human Diseases; Cancers
path:hsa04660 T cell receptor signaling pathway	0.006243754	Organismal Systems; Immune system
path:hsa05162 Measles	0.004967513	Human Diseases; Infectious Diseases
path:hsa05131 Shigellosis	0.016765074	Human Diseases; Infectious diseases
path:hsa05220 Chronic myeloid leukemia	0.014252803	Human Diseases; Cancers
path:hsa04910 Insulin signaling pathway	0.020275511	Organismal Systems; Endocrine System
path:hsa05213 Endometrial cancer	0.027536913	Human Diseases; Cancers
path:hsa05120 Epithelial cell signaling in Helicobacter pylori infection	0.0372481	Human Diseases; Infectious diseases
path:hsa04350 TGF-beta signaling pathway	0.042778509	Environmental Information Processing; Signal transduction
path:hsa04720 Long-term potentiation	0.044536399	Organismal Systems; Nervous system

According to the functional analysis we have observed a functional link between lung metastasis of breast cancer, infectious diseases and immune system. Although, BMSN was also significantly enriched in some pathways that are governed by “Immune system” and “Infectious Diseases”, these two classes were not covering the most abundant pathways. It is interesting that immune system and infectious diseases seem to play an important role in lung metastasis, while transport and catabolism seem to play a major role for brain metastasis. Indeed, lung tissue is in contact with the environment, being likely prepared for infection, while brain is separated of circulating blood by the blood-brain barrier and it requires metabolic processes to transport and catabolize glucose. Still, these results are obtained for networks which expression is produced mostly in breast.

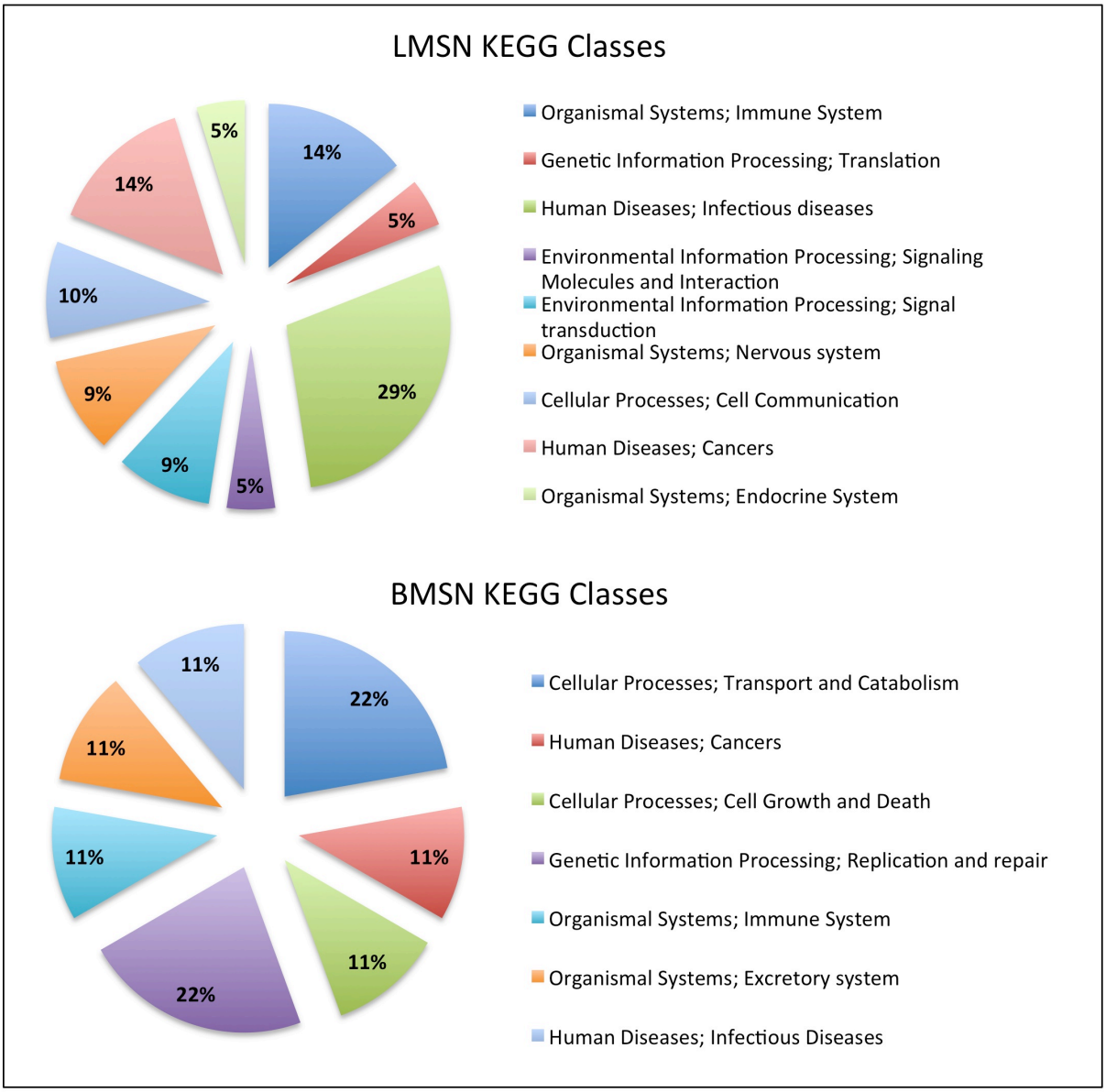

Figure 5.3. The percentages of KEGG classes observed in LMSN and BMSN [285].

5.1.2. Structural Analysis of the Metastasis Sub-Networks

The network representation of PPIs provides information about the sets of interacting proteins (i.e. whether two proteins bind or do not bind and the number of interactions a protein can have). Introducing structural knowledge to PPI networks adds an extra dimension of data to the representation. When we know how proteins are interacting structurally, we can detect multiple proteins trying to bind the same region on a protein

surface. This extra knowledge may help us realize which interactions cannot happen concurrently. Besides, there may be protein pairs interacting via similar interface architectures. A drug targeting on any of these PPIs will have a high probability of targeting the others as well [39, 115], since ligands have tendency to bind to similar binding sites [288-290]. Moreover, knowing the interface region of two proteins helps us to check whether mutations of these proteins occur in the interface or not.

Among the PPIs of the BMSN, only 4 of them had 3D structural data of the binary complex in PDB. Similarly, for LMSN, only 2 PPIs were found with the structure of the binary complex in PDB **(**see **Table 5.4)**. In order to increase the structural coverage of interactions of our sub-networks, it is necessary to use modeling. We used PRISM [140, 141, 206] in order to predict, assign and model the structure of the interface of protein-pairs in the BMSN and LMSN (see **Methods** for the details).

Table 5.4. Interactions available in PDB. In PDB 4 of the PPIs of brain metastasis network had 3D structural data in their complex forms. Similarly, only 2 were found for lung metastasis network.

	Protein Name	**Protein Name**	**PDB ID of The Complex**
BMSN	TNFRSF10B	TNFSF10	1D0G, 1D4V, 1DU3
	ITGA5	ITGB1	3VI4, 3VI3
	MMP1	TIMP1	2J0T
	CSF3	CSF3R	2D9Q
LMSN	MMP1	TIMP1	2J0T
	CXCL12	CXCR4	2K03, 2K04, 2K05

PRISM produces template-based predictions and it models the structure of an interaction based on the known 3D structure of two interacting proteins. The BMSN has 58 interactions with known 3D structures for both partners. LMSN has 102 such interactions. PRISM modeled 18 out of 58 interactions as a binary complex in the BMSN (see **Figure 5.2.a**). For the LMSN, 50 out of 102 interactions were modeled (see **Figure 5.2.b)**.

We should note that PRISM can model an interaction using structurally different interface templates or can use the same template interface to model different interacting protein pairs. Besides, a protein may be embodied with different chains (as identified in the PDB) or domains describing different portions or protein-states (i.e. due to post-transcriptional modifications). Therefore, the interaction between two proteins can imply more than one interface region (i.e. produced by two or more pairs of domains) that may or may not occur at the same time. This would explain the causes for multiple interface predictions. On the other hand, template interfaces can be assigned to several interactions, some of them being common for different sub-networks or highly frequent in some sub-network. This arises a particular interest because it can explain a phenotype but also has implications on the putative use of drugs disrupting a particular set of interactions. As a consequence, for BMSN we obtained 32 predictions for 18 PPIs coming from 28 interface templates. Therefore, the average template interface frequency in BMSN is 1.14 (32/28). For LMSN, we obtained 99 predictions for 50 interactions and 75 out of 99 corresponded to different template interfaces. Thus, the average template interface frequency for LMSN is 1.32 (99/75). The numbers of occurrences of interfaces in both metastasis networks are shown in **Table A.10**.

We studied the common template interfaces in the BMSN and LMSN. We observed top 3 high frequency template interfaces in the LMSN: 1) 2b8nAB 8 times, the interface extracted from the homodimer Glycerate kinase, putative. 2) 1jogCD 5 times, the interface extracted from the homodimer Uncharacterized protein HI_0074. 3) 2a6aAB 4 times, the interface extracted from the homodimer Peptidase M22 glycoprotease. We observed 4 template interfaces with less frequency (only in 2 PPIs) in the BMSN: 1) 2b8nAB (as for LMSN), 2) 1nqlAB, taken from the interface between EGFR-EGF, 3) 1qjcAB the interface extracted from the homodimer phosphopantetheine adenylyltransferase and 1moxAC (the interface between EGFR-TGFA). Interestingly, the 2b8nAB template interface is the most frequent interface in both sub-networks (see **Figure 5.4** and **Figure 5.5**). Details of the most frequent interface templates can be found in **Table A.11** and **Table A.12**. We observed that the three most common interface templates in LMSN are all coming from bacterial proteins.

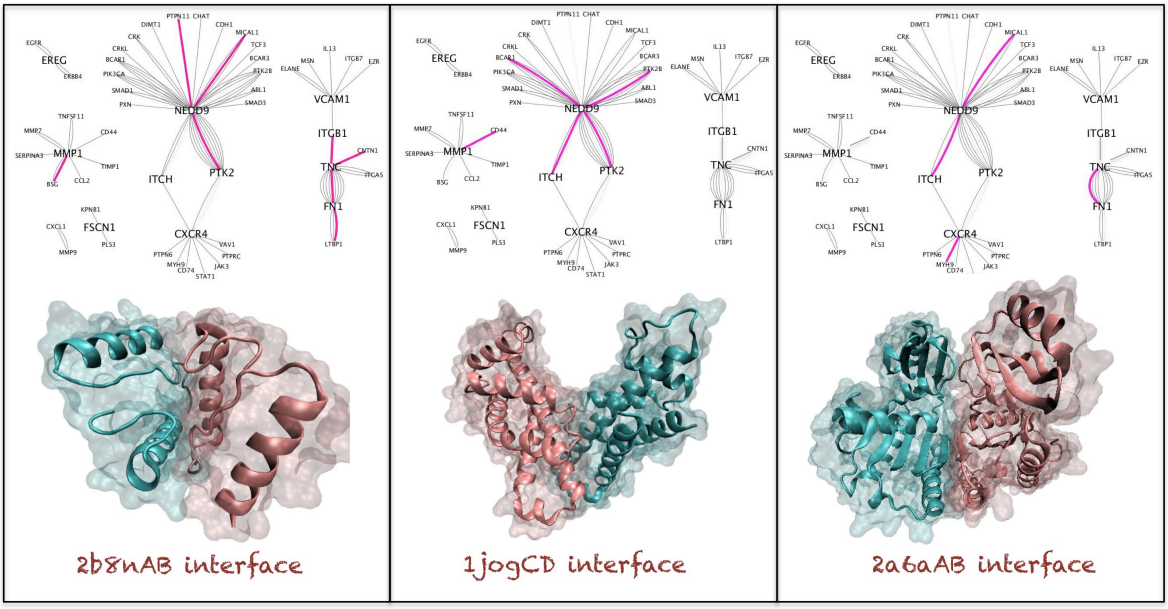

Figure 5.4. Commonly observed interfaces of lung metastasis network [285]. In this figure structural sub-networks are also included. In these sub-networks only the interactions that have PRISM modeled complex structures are present. Each node represents a protein that has 3D structure and each edge stands for a distinct model between two proteins. The relevant template interfaces are represented with pink edges in these structural sub-networks.

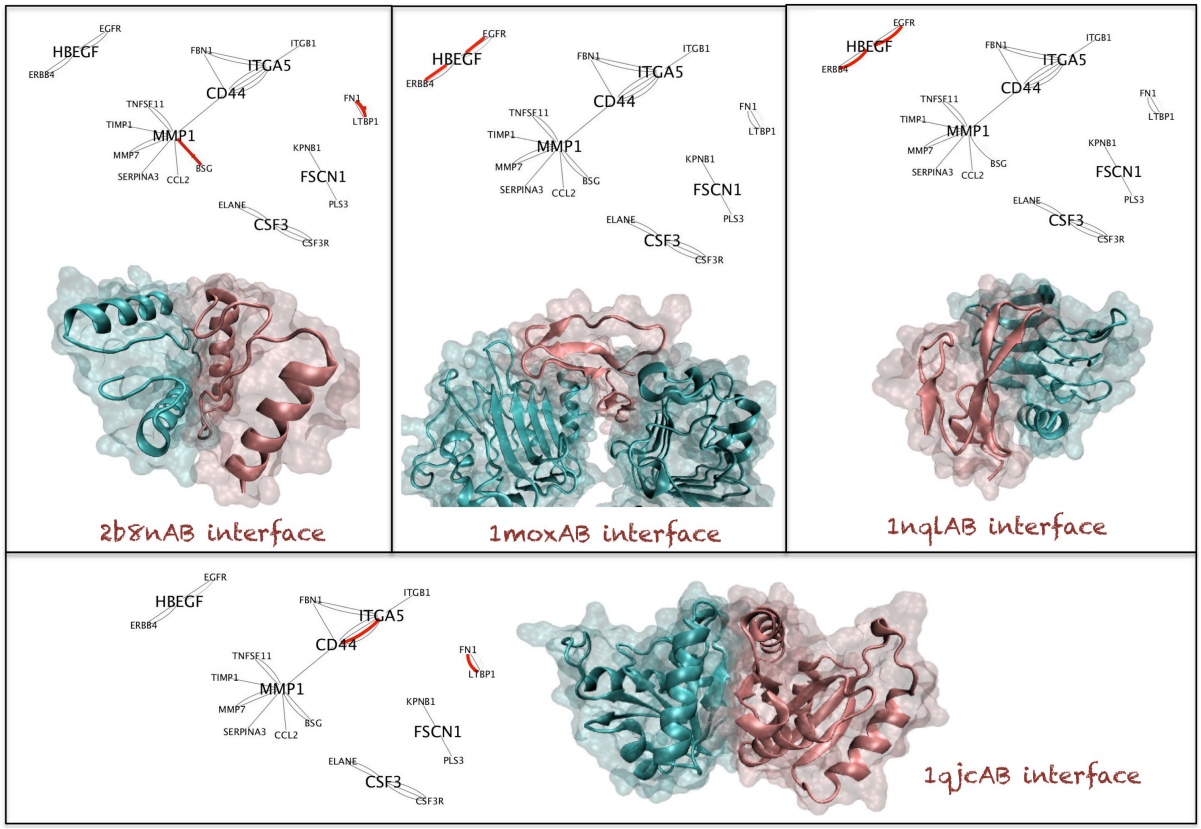

Figure 5.5. Commonly observed interfaces of brain metastasis network [285]. Legend for the sub-networks is the same as in **Figure 5.4**.

Then we studied the source organisms of all the template interfaces used in our sub-networks. We used 28 different template interfaces (**Table A.10** for modeling the complexes in BMSN. Each template interface consists of 2 chains, thus there are 56 template interface chains utilized for the predictions. Among them, 30 template interface chains are originating from microbes (bacteria/virus) (see **Figure 5.6, bottom sketch**). The probability of observing 30 or more microbial chains in a randomly selected set of 56 template interface chains is not significant (p-value = 0.09). Likewise, there were 150 template interface chains (75 template interfaces see **Table A.10**) used for the modeling of LMSN's complexes. 78 out of 150 template interface chains are coming from microbes (see **Figure 5.6, top sketch**). Observing 78 or more template interface chains found in microbes in a randomly selected set of 150 is significant (p-value=0.024). Thus, metastasis protein complexes may be mimicking microbial interface architectures to form complexes, although only for LMSN this feature is significant.

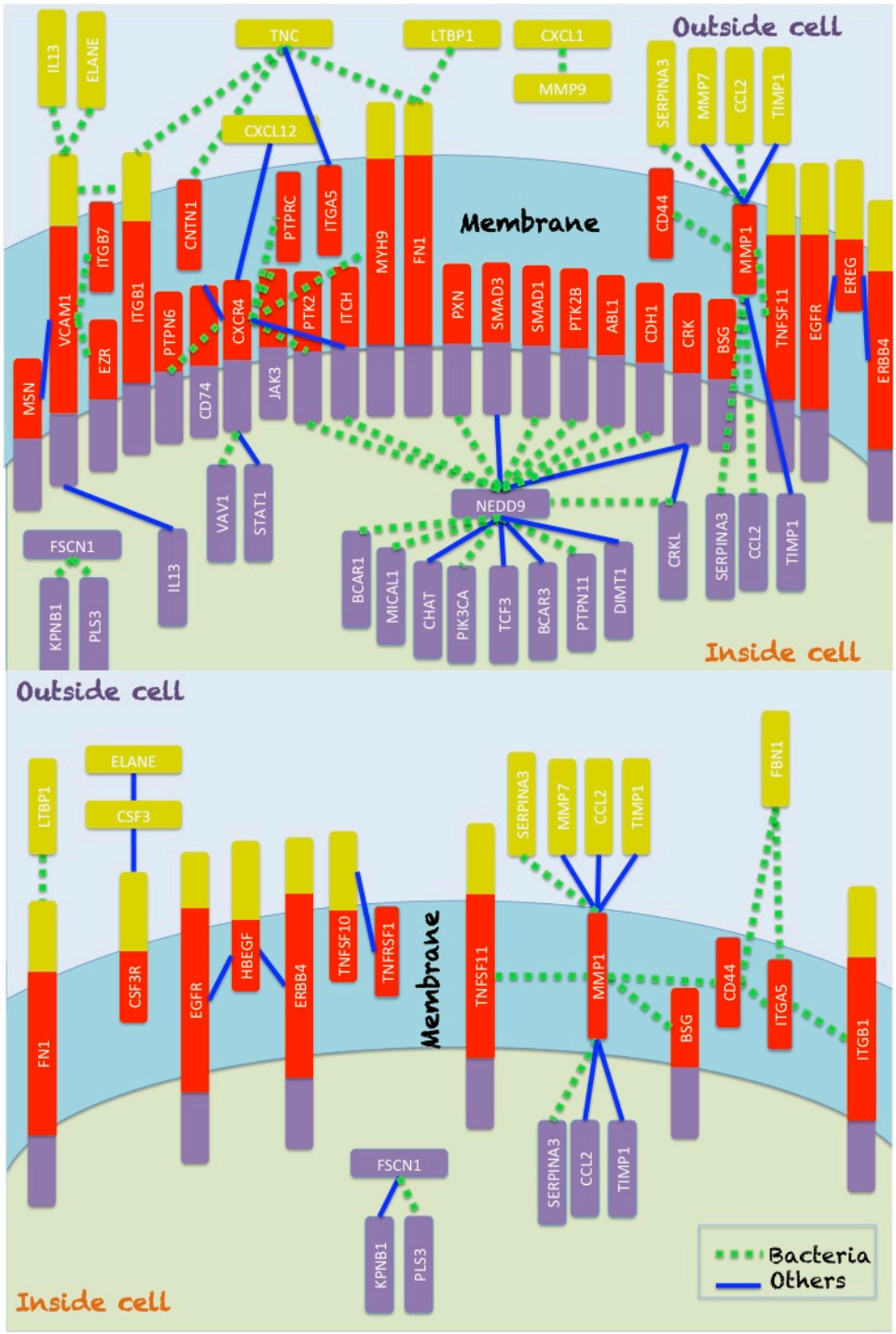

Figure 5.6. The "subcellular locations" depiction of lung metastasis structural sub-network (top sketch) and brain metastasis structural sub-network (bottom sketch). Proteins, which are only in membrane, are shown in red, which are only observed in extracellular region are in yellow and the ones only in intracellular region are purple. The proteins, which can be present in multiple regions of the cell has multiple colors (e.g., EGFR), which is present in intracellular & extracellular regions

and membrane. The green dashed edges are the interactions, which have similar architecture to bacteria/virus interfaces.

Then we investigated the interactions modeled with templates of protein interactions found in microbes. 53% of the models are coming from microbial origin in BMSN (**Figure 5.7, left**) and 59% of the models are coming from microbes in LMSN (**Figure 5.7, right**). Again, the protein complexes in LMSN, utilize more interface templates with microbial origin than the ones in brain network.

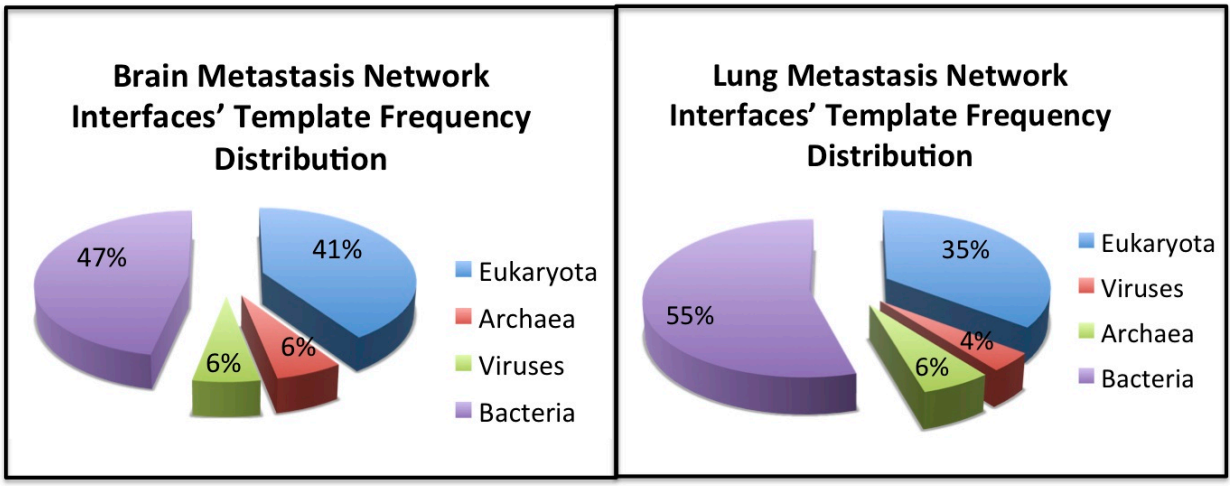

Figure 5.7. Percentages of source organisms [285]. We considered the interfaces' number of observations in the networks. 53% of the modeled complexes use microbial template interfaces in BMSN and this percentage is 59% in LMSN.

There are 14 proteins in BMSN whose interactions are modeled via templates originating from microbes. Seven out of these 14 proteins (**Table A.13)** are actually known to be involved in host-pathogen interactions. For LMSN this ratio is 14/40 (**Table A.14)**. These proteins have binding sites similar to microbial interfaces and some of them are observed to be involved in the host-pathogen protein-protein interactions. This finding suggests that these metastasis related proteins might be involved in mechanisms shared by metastasis and infectious diseases.

Likewise, except 1nqlAB and 1moxAC templates, all the common interfaces observed in both metastasis sub-networks are coming from bacteria. The human proteins in our networks, which are using these frequent templates, have mostly cell adhesion

biological process. Moreover, 50 % of all the proteins modeled with microbial templates in our sub-networks are related with cell adhesion (**Tables A.14 – A.15**). Besides, in BMSN, 25% of the proteins modeled with non-microbial interface predictions are related with cell adhesion. Finally, 21% of the proteins in the LMSN use non-microbial interface architecture (an interface other than microbial interfaces) to interact. Cell adhesion molecules play a significant role in cancer metastasis [291, 292]. Those molecules use mechanisms of cell adhesion for creating metastasis in another organ [293]. Proteins using bacterial interface architectures for interacting with other proteins may be reproducing the adhesion ability of the bacterial proteins.

Moreover, both functional analysis discussed above and the structural analysis suggest a relationship between pathogens, immune system and metastasis. Pathogens may be triggering some mechanisms that lead to metastasis of a primary breast cancer tumor or vice-versa, metastasis may create the proper environment for bacteria invasion.

Actually, previous studies highlighted the resemblances in cellular and molecular mechanisms of invasion between metastasis and infectious diseases [186-189]. Besides, in a recent study, Haile et al. hypothesized that metastasis process and pathogens should be utilizing the same pathways [190]. Liu et al. also mentioned that certain pathogens, activated immune cells and tumor cells may be sharing same tactics to spread in the body [191]. These findings reinforce our functional and structural analyses results.

5.1.3. Overview of the Lung/Brain Metastasis Sub-networks

Network representation of the proteins and their interactions provides a systems level abstraction. Via network representation we may identify the proteins that are central and important. Hubs, proteins with a high number of interactions, are the vulnerable points of scale-free networks and are very important. As expected the hub proteins in the LMSB and BMSN are actually the protein products of the seed genes mentioned earlier. However, not all of the seed genes' products are hubs in these two networks (**Table 5.5**). In BMSN PLOD2, HBEGF, MMP1, LAMA4, FSCN1, TNFSF10 and SCNN1A are the hubs (**Figure 5.2.a**), whereas in LMSN KRT81 (KRTHB1), FSCN1,

ID1, NEDD9, CXCR4, VCAM1 and MMP1 are the hubs (**Figure 5.2b**). Consequently, these seed genes are more critical from a systems point of view.

Table 5.5. Metastasis seed genes. 18 genes [155] that mediate breast cancer to lung metastasis, and 17 genes [177] that mediates breast cancer to brain metastasis. (*) Implies the genes, whose protein products are hubs in the metastasis sub-networks.

LUNG METASTASIS SEEDS	BRAIN METASTASIS SEEDS
MMP1*	MMP1*
RARRES3	RARRES3
FSCN1*	FSCN1*
ANGPTL4	ANGPTL4
LTBP1	LTBP1
PTGS2	PTGS2
KYNU	SEPP1
TNC	LAMA4*
C10orf116	PLOD2*
CXCL1	COL13A1
CXCR4*	SCNN1A*
KRTHB1* (KRT81)	RGC32
VCAM1	PELI1
LY6E	TNFSF10*
EREG	B4GALT6
NEDD9*	HBEGF*
MAN1A1	CSF3
ID1*	

These hubs are mostly not in direct interaction with each other, consequently the topology of both networks consist of a number of node clusters (a seed gene and its interaction partners). Please refer to **Figures A.8** and **A.9** for the significantly enriched KEGG pathways in each cluster.

Furthermore, there are 2 hub nodes, LAMC1 and ITGA3, in BMSN that are not seed genes. They became hub nodes in the network because of the their interactions with PLOD2's interaction partners (shown with green edges in **Figure 5.2.c**). PLOD2 cluster (the first degree neighbors of PLOD2) is shown in **Figure 5.2.c**. They have a very high potential of being major players in brain metastasis formation. In fact,

ITGA3 is down regulated in metastatic medulloblastoma tumors and claimed to be allowing metastatic tumors to spread more eagerly [294].

There are 84 common proteins and 71 common PPIs (blue edges in **Figure 5.2.d**) in both metastasis networks. There are PPIS present only in LMSN (green edges in **Figure 5.2.d**) and only in BMSN (pink edges in **Figure 5.2.d**). As one can see from **Figure 5.2.d,** FSCN1 and MMP1 are two hubs that are common to both metastasis sub-networks, thus they are not very helpful in differentiating two metastasis types. On the other hand, the interactions of PLOD2, the highest ranked protein in BMSN, are only present in BMSN. Similarly, KRT81 is the highest ranked protein in LMSN and its interactions are only present in LMSN. These two proteins may be playing key roles in the related metastasis types.

In **Figure 5.2.a** and **Figure 5.2.b** the proteins that have PDB structures are shown with pink nodes, while the proteins that don't have PDB structures are shown with blue nodes. Most of the hub nodes do not have PDB structures, thus we couldn't make further structural analyses for them. The edges that are modeled with PRISM are shown in pink in **Figure 5.2.a** and **Figure 5.2.b**.

In **Figure 5.4** and **Figure 5.5** the most frequently observed template interfaces in LMSN and BMSN are depicted. In these figures structural sub-networks are also included. In these sub-networks only the interactions that have PRISM modeled complex structures are present. Each node represents a protein that has 3D structure and each edge stands for a distinct model between two proteins. The relevant template interfaces are represented with pink edges in these structural sub-networks. According to structural sub-network of lung metastasis NEDD9 is a hub protein with multiple interface architectures on different regions of its surface (please see **Figure A.10** for the first three most frequently observed interfaces of LMSN mapped on NEDD9). Actually, NEDD9 has multiple domains like SH3 domain, SH2 domain and C-terminal domain containing a HLH motif that it uses for its interactions [295-297]. Right now there is only one PDB structure available in PDB (PDB ID: 2L81) that contains the SH3 and the SH2 domains. Accordingly our predictions are limited with these two domains.

5.2. Genetic Variations on The Protein Interfaces

There are 6 proteins that are present in both metastasis sub-networks and have at least one different interaction partner in each network. We wanted to find out whether the reason why these proteins are changing partners is related with genetic variations. By mapping the mutations on the proteins' 3D structures we may see if the mutation is on the interface region and if the mutated residue is a hotspot, which may intensely affect the interaction strength.

We have PRISM models for 12 interactions that these 6 proteins are involved in (**Table 5.6**). These 12 interactions are happening between 13 proteins. By using the genetic variation data in UNIPROT and COSMIC we made further investigations for them. There are 386 genetic variations taking place on the mentioned 13 proteins; 251 variations on the surface, 135 variations in the core. Among these 386 genetic variations, only 28 of them are happening on the interface regions. Even in recent publications it is mentioned that in-frame mutations [3] and disease causing SNPs [185] have a tendency to occur on protein-protein interfaces we have not encountered this phenomenon **(Tables A.17 and A.18).** However, if we had a larger protein set, this result might have been different. Plus the structural information we have on interfaces is very limited, most probably we are missing some additional interfaces. Thus the genetic variations mapped on the surface region may be coinciding with interfaces as well.

Table 5.6. List of proteins that exist in both metastasis network and the different interactions they make in each metastasis network.

PROTEIN	BRAIN NETWORK INTERACTION PARTNERS	LUNG NETWORK INTERACTION PARTNERS
ELANE	CSF3	VCAM1
EGFR	HBEGF	EREG
ITGA5	ITGB1 CD44 FBN1	TNC
ERBB4	HBEGF	EREG
CD44	FBN1 ITGA5 MMP1	MMP1
FN1	-	TNC

Two of the interactions that have genetic variations on their interface regions are discussed further as case studies below.

5.2.1. EGFR and ERBB4

The EGFR and ERBB4 proteins interact with HBEGF in BMSN, whereas they interact with EREG in LMSN. In fact HBEGF is a gene known to have a role in brain metastasis of breast cancer [177], while EREG is a gene known to be mediating lung metastasis of breast cancer **[155]**. The structural models of these interactions are not available in PDB, but we have PRISM predictions for these complexes. HBEGF is predicted to interact with EGFR and ERBB4 via the same binding site on its surface, and this is also the case for EREG (**Figure 5.8**).

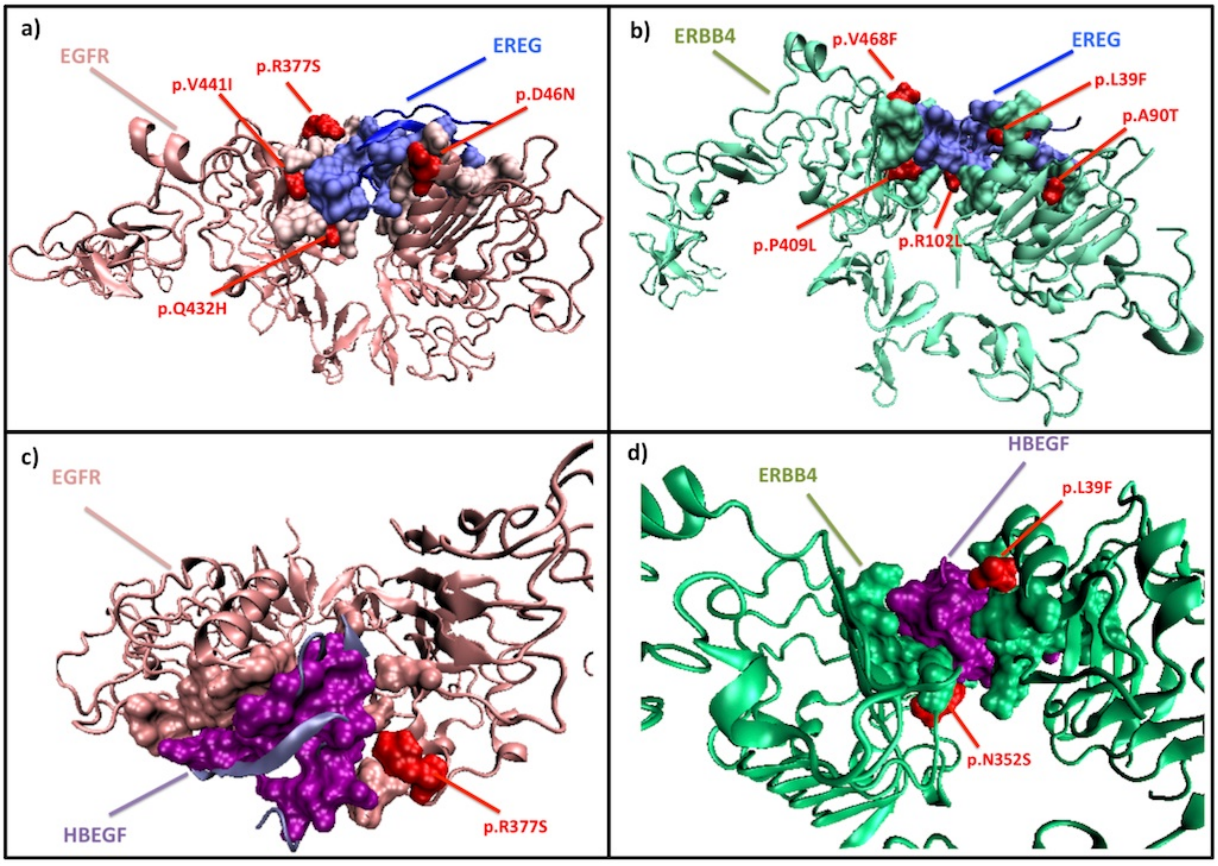

Figure 5.8. The PRISM predictions for a) EREG (blue) – EGFR (pink), b) EREG (blue) – ERBB4 (green) interaction, c) HBEGF (purple) - EGFR (pink) interacrion and d) HBEGF (purple – ERBB4 (green)) interaction [285]. We have discovered multiple genetic variations happening on these interfaces.

Both EREG and HBEGF are growth factors that may be integrated to the membrane and can also be present in the extracellular space. EGFR binds EGF family members via its L1 (between residues 1-151) and L2 (between residues 312-481) domains [298]. The interface residues modeled with PRISM on EGFR (interfaces with HBEGF and EREG) are lying in these domains. Similar to EGFR; ERBB4 binds to EGF family members via its L1 and L2 domains (between residues 27-198 and 324-517 [299]). Most of the interface residues of ERBB4 modeled by PRISM are coinciding with these domains as well. Plus, the EGF-like domain (between residues 20-208) of HBEGF is known to have an important role in binding to EGFR [300]. The predicted interface residues for HBEGF are taking place in its EGF-like domain. EREG's C-terminal (between residues 96-106) is suggested to be involved with its binding to ErbB receptors[301]. The C-terminus of EREG is in the interface model produced by PRISM.

There are a number of EGFR complexes, one ERBB4 complex and one HBEGF complex available in PDB, while there are no EREG complexes. When we compare our model's interface residues with the binding sites of the available PDB complexes, we see that they are all overlapping (see **Tables A.19, A.20, A.21 and A.22**).

Position 102 in the amino-acid sequence of EREG acquires a SNP (p.R102L) in some cancer patients (derived from COSMIC database). This amino acid is on the interface region of EREG-ERBB4 interactions. Plus, this residue lies in the C-terminal of EREG that is known to be essential for its interactions with ErbB receptors. Moreover, ERBB4 acquires 5 different mutations that coincide with its interfaces. These mutations are observed in cancer patients (derived from COSMIC). Genetic variations p.L39F, p.A90T, p.409L and p.V468F mutations are coinciding with ERBB4-EREG interactions. Furthermore p.L39F and p.N352S mutations are coinciding with ERBB4-HBEGF interaction. Additionally, 4 mutations of EGFR derived from COSMIC database are coinciding with its interactions. While p.D46N, p.Q432H and p.V441I are affecting EGFR-EREG interaction, p.R377S mutation is affecting both EGFR-EREG and EGFR-HBEGF interactions.

These mutations may be making the mentioned interactions stronger or weaker but they are most probably changing the functions of the EREG, HBEGF, EGFR and

ERBB4 proteins (**Figure 5.8**). Besides, there may be a relationship between the metastasis progression and these mutations.

5.2.2. ELANE (ELA2)

ELANE interacts with CSF3 in BMSN, while it is switching its interaction partner to VCAM1 in LMSN. CSF3 is a seed gene in BMSN **[177]**, while VCAM1 is a seed gene in LMSN **[155]**. The structural models of these interactions are not available in PDB, but we have PRISM predictions for these complexes.

ELANE has variants that coincides with its interfaces (p.V98L, p.V101L, p.V101M and p.S126L (derived from UNIPROT)). The variances in the amino acid 101, which are polymorphisms, coincide with one of the hotspots of the interface region between ELANE and CSF3 and the variances in the amino acids 98 (polymorphism) and 126 (unclassified variation) are inside the interface region of VCAM1 on ELANE (**Figure 5.9**). These amino-acid variances may be affecting the interactions of ELANE with CSF3 and VCAM1. As a result, these amino acid variations may be related with metastasis progression in breast cancer patients.

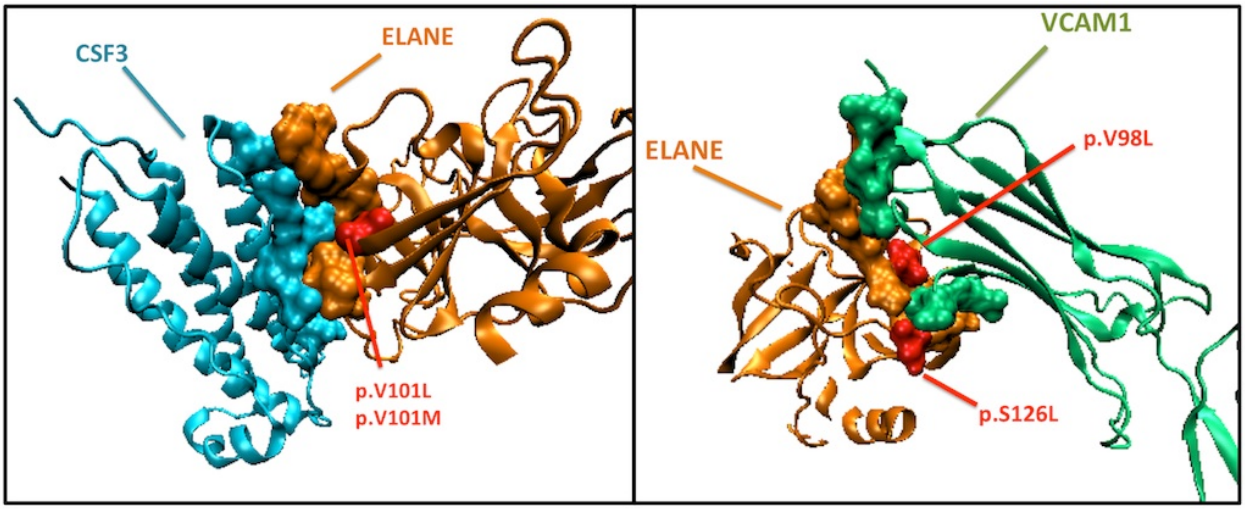

Figure 5.9. The PRISM predictions for ELANE (orange) - VCAM1 (green) and ELANE - CSF3 (blue) interaction [285]. The amino acids 98, 101, 126 (red amino acids) on ELANE have genetic variations. Amino acid 101 is a hotspot in the CSF3 – ELANE interface, moreover amino acids 98 and 126 are part of the ELANE – VCAM1 interface.

5.3. Methodology

5.3.1. The Human PPI Network

Experimental data on protein interactions are spread among multiple databases. Even if the data in these databases partially overlap, the reliability of data differs because of the variations in the experimental techniques and the organisms used. In addition, information of the same protein can be stored with different designations in different databases. Therefore, all the available data should be queried properly and matches should be combined to form a comprehensive human PPI network. We made use of BIANA [302] (Biological Integration And Network Analysis) bioinformatics tool in order to form human PPI network. BIANA gathered PPI data from various databases and dealt with mapping between the different identifiers. We combined DIP[240], MIPS[241], HPRD[242], BIND[243], IntAct[244], MINT[245] and BioGRID [246] databases (all downloaded on May, 2011). Interactions and protein information were integrated with BIANA assuming that two proteins from different databases were the same if they had the same UNIPROT Accession, amino acid sequence, or Entrez Gene Identifier.

5.3.2. The Sub-networks Implicated in Lung and Brain Metastatic Breast Cancer

We used GUILD, a network-based disease-gene prioritization tool [284] to identify the sub-networks implicated in the two phenotypes of our interest: 1) breast cancer metastasis in lung, and 2) breast cancer metastasis in brain. GUILD package includes several methods of "guilt-by-asssociation" to prioritize a list of candidate genes associated with a phenotype. Guilt-by-association approaches are based on a set of genes associated with a phenotype, named seeds, and the tendency that other genes associated with the same phenotype will interact with the seeds. We took 18 genes that mediate breast cancer to lung metastasis **[155]**, 17 genes mediating breast cancer to brain metastasis [177] identified by Massagué and his co-workers and used them as seeds for each phenotype (**Table 5.5**).

We employed the NetCombo algorithm in GUILD using the default parameters as in [284] to rank all the proteins of the major component of the human PPI network. This algorithm combines the algorithms of NetScore, NetZcore and NetShort. The scores

were different for proteins produced by genes associated with brain metastasis than those associated with lung metastasis. Therefore, two different sub-networks were considered with the proteins associated with lung or brain metastasis and their interactions.

GUILD scored only the nodes (proteins/genes) but not the edges (PPIs) and gene-gene associations), therefore we needed to transfer the score of the nodes into the edges. Thus, we defined the score of the edge as the average of the scores of its nodes (the values of these scores lie between 0 and 1). We selected a common threshold cut-off on the score of the edges to set up the sub-networks of brain and lung metastasis with similar size.

We used HPRD [242], UNIPROT [295, 296] and TIGER[303] databases for checking the expression of genes in breast tissue.

The average node degree is 2.6 for BMSN and 2 for LMSN. Nodes with 12 or more edges are considered to be hubs.

We have used VMD [272] for visualizing protein structures and for network visualizations we have used Cytoscape [282].

5.3.3. Functional Analysis of Brain and Lung Metastatic Networks

We used the ClueGo [286], a Cytoscape [282] plugin, designed for biological interpretation of gene sets. The significance (enrichment) analysis was performed with right-sided hyper-geometric testing with a Bonferroni step down P-value correction factor. KEGG pathways used for the calculations are downloaded in 24.05.2012. P-values smaller than 0.05 were considered significant.

5.3.4. Structural Analysis of Brain and Lung Metastatic Networks

We have used “uniprot_sprot.dat” (downloaded in November of 2012, from UNIPROT’s ftp server) for detecting the source organisms of the PDB chains used for modeling the protein complexes in both metastasis networks.

For significance testing, we have calculated the p-value of a hyper-geometric distribution using the R package[304]. P-values smaller than 0.05 were considered significant. Please refer to **Table A.23** for the numbers we have used for calculations.

Every protein-protein interface consists of two chains. The 7922 template interfaces used in the experiments, consist of 15844 template chains. Among them the source organism of 11255 were available in "uniprot_sprot.dat" and 4918 were coming from microorganisms (bacteria/virus).

The protein interfaces that are available in PDB are clustered according to their structural similarity. These clusters are provided in PRINT database which can be accessed from the http://prism.ccbb.ku.edu.tr/interface/ address. We mentioned these structurally similar protein clusters as PRINT clusters all through the text.

While detecting the source organisms of the template interfaces, we have taken into account all the interfaces, not only the representative interfaces (in each PRINT cluster). Besides, we have used the biological process and the molecular functions listed in UNIPROT database for our analyses.

In **Figure 5.6** the interfaces coming from pathogens are presented with green dashed lines. We have used "Cellular Component Ontology"[305] in order to find out the locations of the molecules. If an interface has PDB's coming from both eukaryotes and pathogens in its cluster, we have counted it as a pathogenic interface.
We have made use of UNIPROT and HPIDB[306] databases to mine the knowledge on the host-pathogen relationships of the related proteins. We have checked whether the proteins are known to be interacting with pathogens or not (**Tables 5.10** and **5.11**).

5.3.5. Genetic Variations on Interface Surfaces

We obtained the available point mutations related with cancer from COSMIC [307] database and humsavar.txt of UNIPROT database. UNIPROT [295, 296] provides the variants of a protein's amino-acid sequence. These variations can be polymorphisms, variations between strains, isolates or cultivars, disease-associated mutations or RNA editing events. Both databases provide detailed information about the mutations, as well as the mutated residue numbers. Then, we mapped these point

mutations to the interface regions of interacting proteins in the metastasis sub-networks (BMSN and LMSN). We used the PDBSWS database [308] for the PDB and Uniprot residue-level alignment.

We used Naccess [309] for determining the surface and core residues. Naccess computes the atomic accessible area by rolling a probe (typically with the same radius as water (1.4 Angstroms)) around the Van der Waal's surface of macromolecule. It employs the Lee & Richards method [310], whereby a probe of given radius is rolled around the surface of the molecule, and the path traced out by its centre is the accessible surface.

For the statistical calculations of location preferences of genetic variations we used fisher's (exact) test and two-tailed P-value for statistical significance (P-value smaller than 0.05 was considered statistically significant) as described in David et al.'s [185] article. We used the R package[304] for the statistical calculations.

Hot spots are the residues that contribute more to the binding free energy with respect to other residues in the protein-protein interface. We have used HotPoint [271] for hot spot predictions. This webserver calculates the hot spots in protein interfaces using an empirical model with 70% accuracy.

Chapter 6

CONCLUSION

The main focus of this dissertation has been the integration of structural knowledge to protein-protein interaction networks and utilizing this additional information in solving drug off-target prediction and genotype-phenotype mapping problems.

We proposed a new network representation (P2IN), which introduces the structures of protein interfaces into the PINs. In addition to providing the binary information of whether two proteins interact with each other, the P2IN also provides information on the structure of the complex that they form. Through out my PhD studies we have built a number of cancer related P2INs and increased the structural knowledge on protein interactions of these networks. The accuracy of these P2INs is limited with the completeness of protein interactions, reliability of homology models and availability of template interface structures, however with the exponential growth of the number of protein complexes in PDB and the discovery of pairwise protein interactions building high-quality P2INs will be possible.

P2IN representation allows us to propose a new attack strategy, interface attack, of hitting edges between protein pairs that interact via structurally similar interfaces rather than nodes. We generated the signaling network of the p53 P2IN and tested its robustness to various attacks. Both node and edge attacks are performed. The interface attack is found to be as destructive as hub node attacks; however, it is not as harmful as distributed attack that targets maximal edges. A drug that disturbs a frequent interface type may be as destructive as a drug targeting a high degree protein, suggesting the usefulness of considering the frequency of interface motifs during drug development. We discovered that some drugs (Aminopurvalanol, PD-0332991, CHEBI:792519, CHEBI:792520 and Fisetin) binding to CDK6, disrupt its interaction with CDKN2D. We applied our interface attack strategy to this case and found that drugs blocking this interface may also affect the interaction between CDK4 and CDKN2D. CDK4 also appears an off-target for drugs binding to CDK6. This example illustrates the promise in our strategy as a first step in indentifying potential off-target drug hits. Finally, we provided a case study of a comparison between node and interface attacks. Challenging next steps are accounting for

molecular flexibility. Proteins are highly dynamic, and structure-based drug discovery requires detailed structural treatment to uncover transient pockets which are unlikely to be observed in the static crystal snapshots and rigid docking. Nonetheless, systems-wide outcome involving possible off-targets of a drug is an important consideration, and eventually would need to be integrated with detailed structural investigation in attempts to forecast potential side effects. Here, our concept of interface attack exploits structural motifs. It is inspired by network pharmacology, an emerging paradigm in drug discovery.

In the future this drug off-target prediction approach may be exerted on the complete human interactome. Working in larger scale would provide a more complete view of the pathways affected by a given drug. Besides, while building the P2IN, considering both the bound and unbound states of a protein will also increase the reliability of our off-target prediction approach. In addition, experimental verifications could be very useful to prove the credibility of our prediction method.

We combined PPI networks, protein-protein interface structure and genetic variations together at the systems level to explain genotype-phenotype relationships. We have built two networks of proteins playing roles in different breast cancer metastasis and tried to explain the mechanisms behind metastasis process.

We built a comprehensive human PPI network, by combining the available PPI data from various databases. Then we ranked all the interactions of this network according to their relevance to genes that are known to be mediating breast cancer to brain and lung metastasis. Subsequently, we formed two distinct metastasis PPI sub-networks from high ranked interactions. Next, we introduced structural knowledge to metastasis PPI sub-networks. Only a small proportion of our protein complexes were available in PDB. We modeled the interface structures of PPIs by using PRISM tool. Knowing the interface structure between two proteins and the residue numbers on the interface surface, allowed us checking whether the mutations are located in the interfaces or not.

We preformed functional analysis on metastasis sub-networks and observed that the proteins engaged in LMSN are enriched in “Infectious Diseases”, “Cancer” and “Immune System” KEGG classes. This correlation pinpoints a relationship between pathogens, immune system and lung metastasis. This may be due to the fact that, brain is a better-

protected area than the lung, due to the blood-brain barrier and being less exposed to outside world compared to lung. Besides, the protein complexes in LMSN utilize more interface templates found in PPIs in microbes than BMSN. This finding reinforces our conclusion about the relationship between lung metastasis progression and pathogens. Furthermore, we saw that in both metastasis sub-networks the proteins using microbial interface architectures are mostly related with cell adhesion. Cell adhesion is a very important mechanism for metastasis and our findings suggest that there may be some mechanistic commonalities, such as cell adhesion, between pathogens and metastatic cancer cells employed during cell invasion. Actually, most of these proteins have interactions with proteins of pathogens themselves.

We provided structural predictions for the architecture of interfaces of interactions between EGFR-EREG, EGFR-HBEGF, ERBB4-EREG, ERBB4-HBEGF, ELANE-CSF3 and ELANE-VCAM1. Moreover, we have discovered some genetic variations happening on these interfaces which are most probably related with the metastasis progression of breast cancer patients.

For future studies our metastasis network models may provide a foundation and may also be helpful for finding escape pathways of breast cancer metastasis. In our results, there is a group of SNPs that are nominated to be related with specific metastasis types, they could be validated with experiments. This would increase the impact of our predictions.

We have utilized P2INs in predicting drug off-targets and linking genotype to phenotype. However these networks have the potential to be used for answering several other questions as well. Deepening the analysis on these networks may reveal important futures about structural proteomics. Besides, the visualization of P2INs may be improved in the future, which may enable us to acquire structural data intuitively from the network representation.

In the overall, significant information is gained towards the protein interactions in the systems level. Integration of structural knowledge into protein interaction networks helped us answer many questions by providing additional information on “how protein couples interact”. We believe that this work will serve functional and structural genomics, cancer bioinformatics and drug design.

APPENDIX

Supplementary Figures:

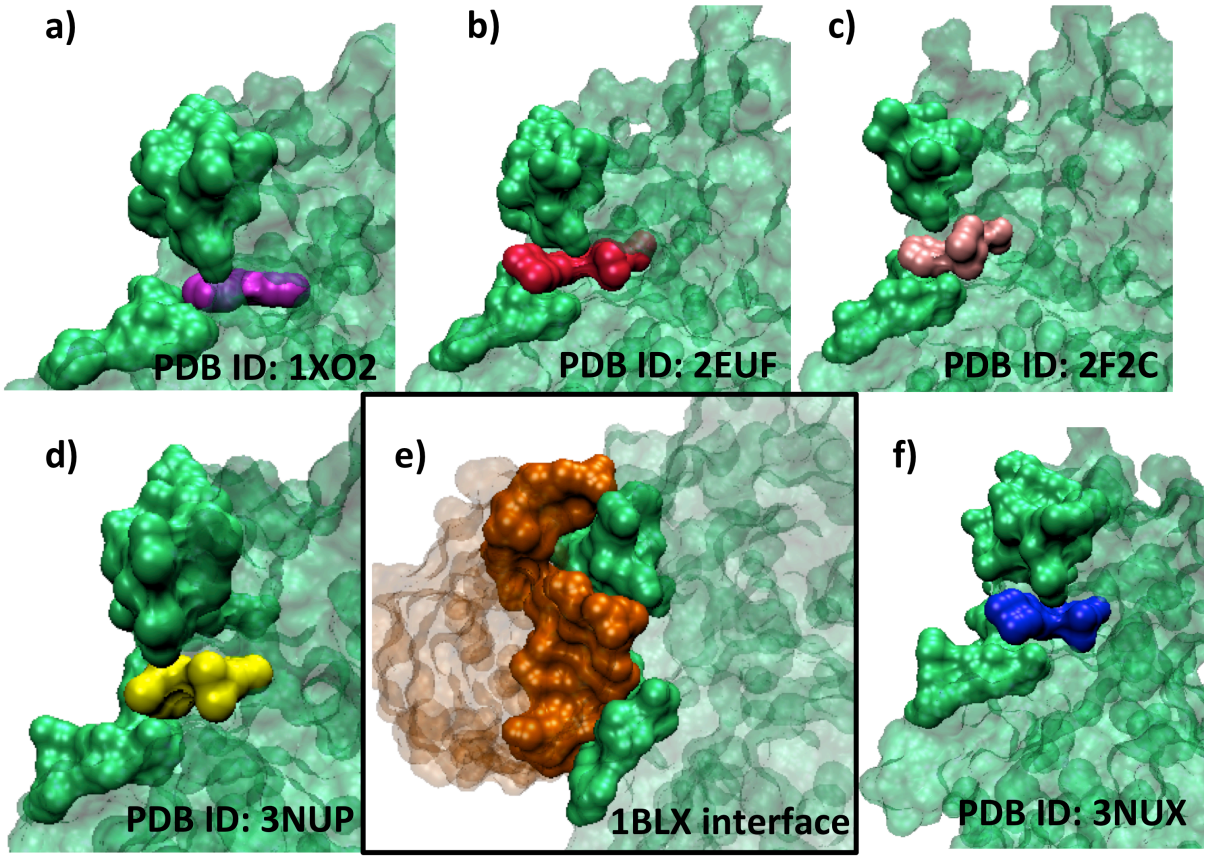

Figure A.1. CDK6 binding site (one chain of CDK6–CDKN2D interface) highlighted on CDK6-drug complexes present in PDB [39]. In each figure CDK6 structure is the transparent green body and the binding site on CDK6 is the opaque green one a) Fisetin-CDK6 complex b) PD-0332991-CDK6 complex, c) Aminopurvalanol-CDK6 complex, d) CHEBI: 792519-CDK6 complex, e) CDKN2D-CDK6 complex, f) CHEBI: 792520-CDK6 complex. The structures of CDK6 are not exactly the same in each PDB (please refer to **Table A.7** for the RMSD values of CDK6 structures), as a consequence we didn't perform a superimposition between CDK6-drug complexes and CDKN2D-CDK6 complex like we did in the **Figure 4.2**.

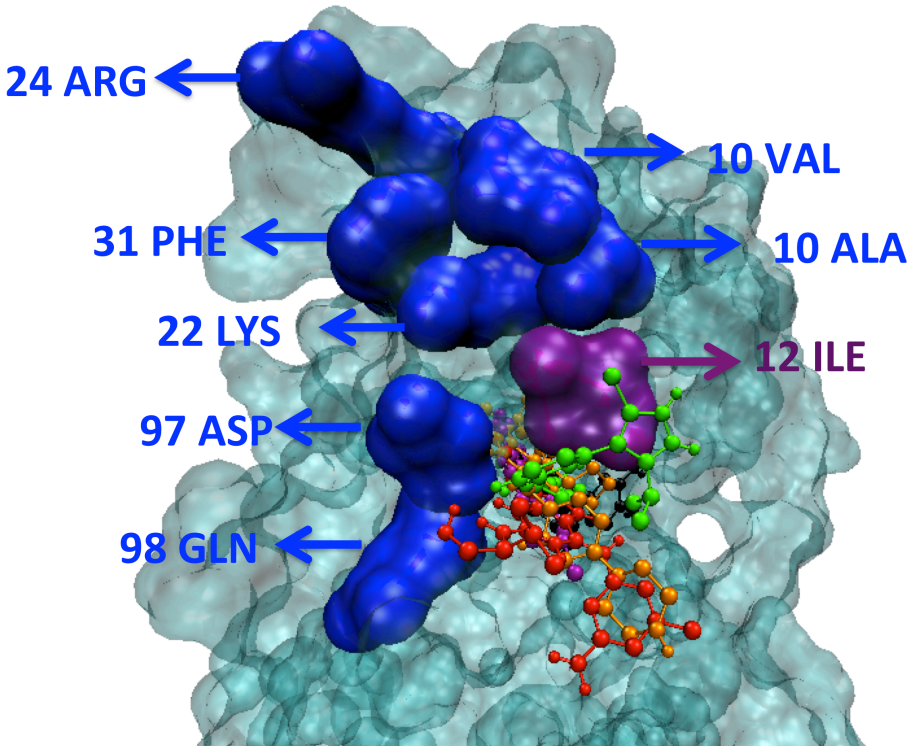

Figure A.2. The hotspots of CDK4 (dark blue surfaces), CDK4 structure (cyan transparent body) and the drugs (balls and sticks) docked on CDK4 can be seen all together in this figure [39]. The drugs are close to hotspots 12 ILE, 98 GLN and 97 ASP.

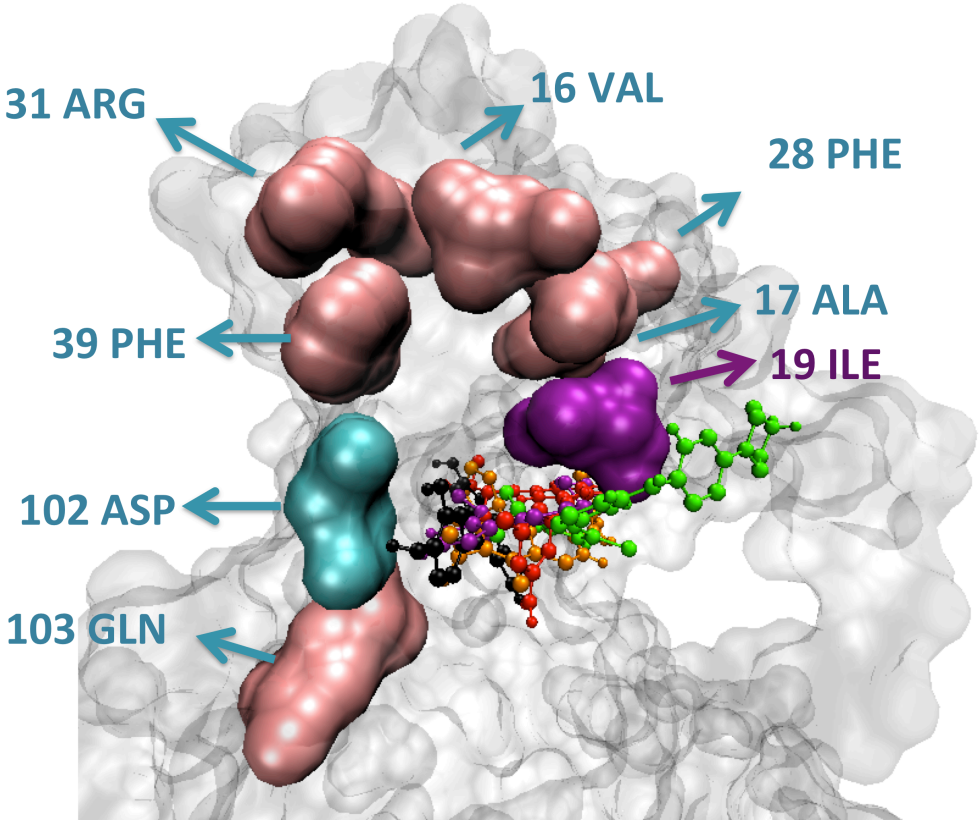

Figure A.3. The hotspots of CDK6 (pink surfaces), CDK6 structure (gray transparent body) and the drugs (balls and sticks) docked on CDK6 can be seen all together in this figure [39]. The drugs are close to hotspots 19 ILE, 103 GLN and non-hotspot residue 102 ASP (cyan surface).

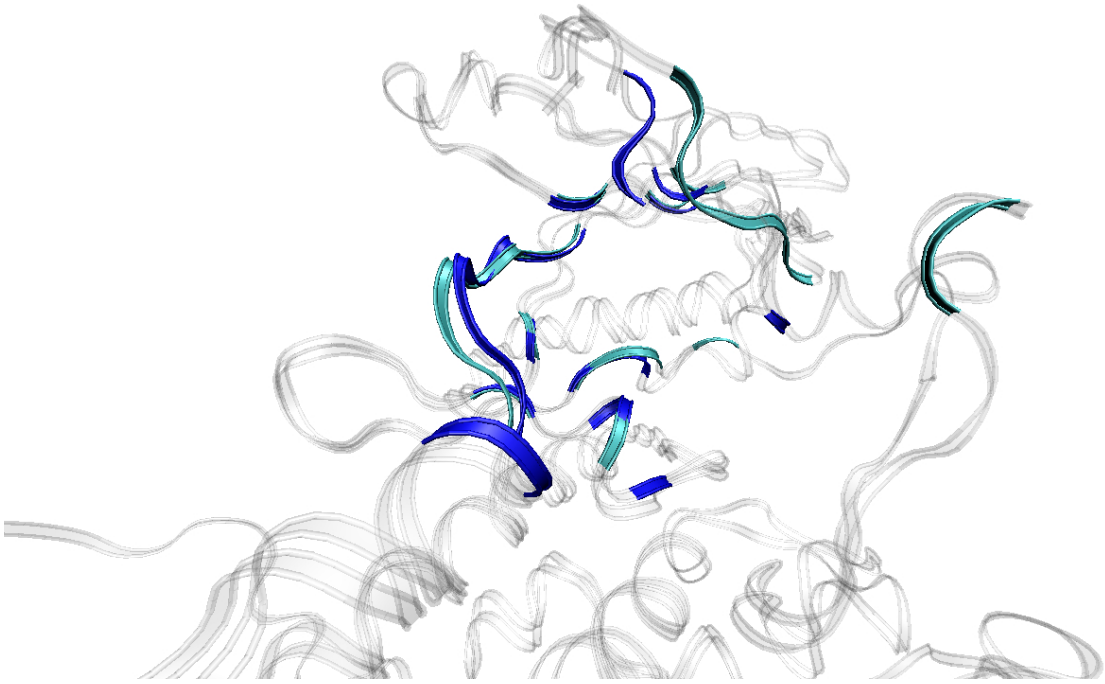

Figure A.4. Superimposition of pockets of CDK6 (cyan) and CDK4 (dark blue) using VMD visualization tool [39].

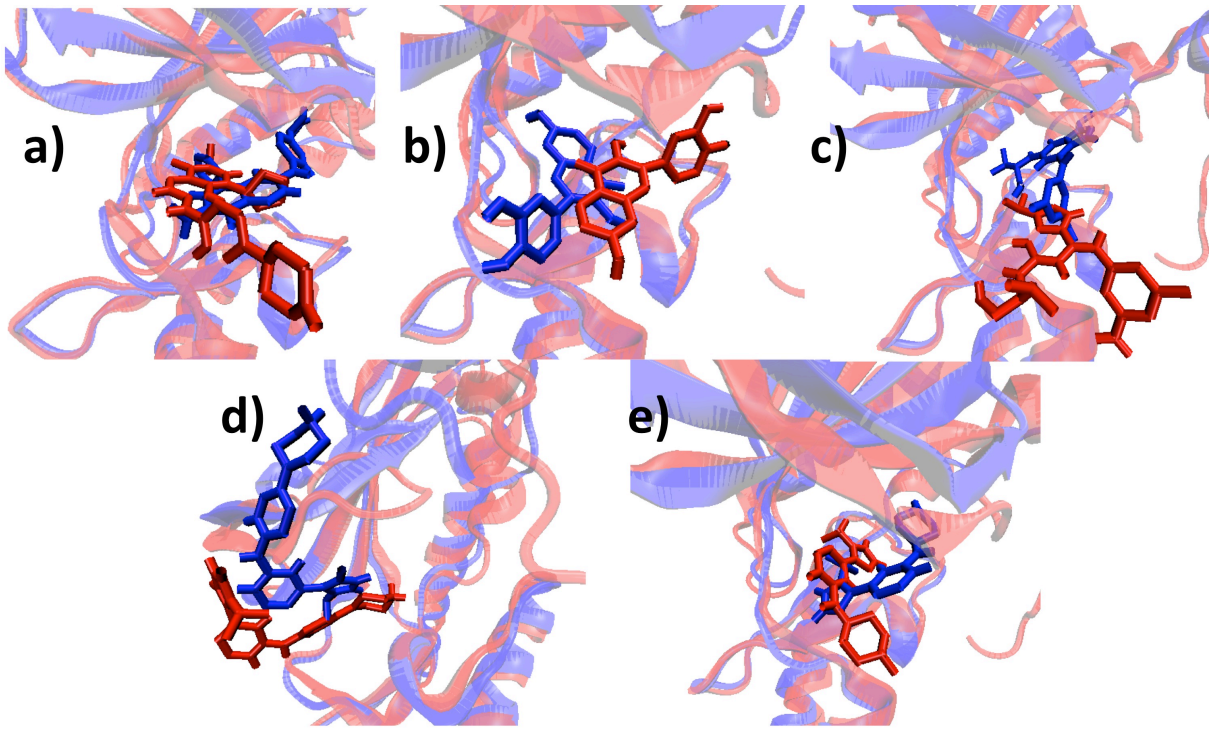

Figure A.5. Superimposition of pockets of CDK6 (dark blue), CDK4 (red) and the drugs docked to them [39]. The blue ligands are docked on CDK6 and the red ones are docked on CDK4. The ligands in the figures are: a) PD-0332991 b) Fisetin c) Aminopurvalanol d) CHEBI: 792520 e) CHEBI: 792519.

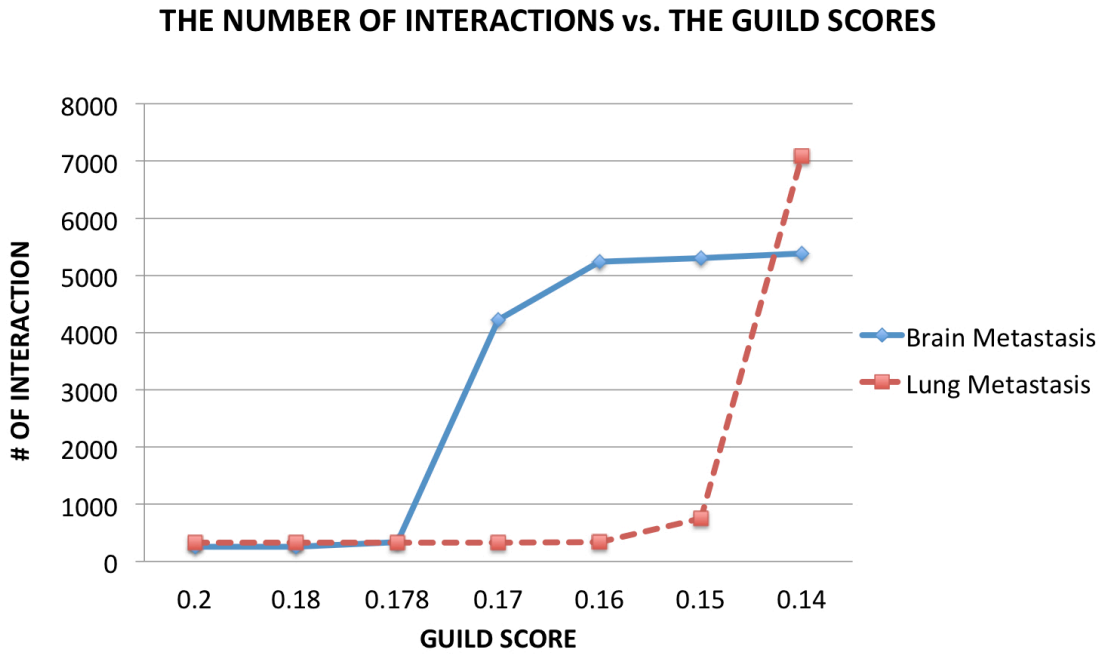

Figure A.6. This graph shows the increase in the number of interactions, as the number of GUILD score gets smaller [285].

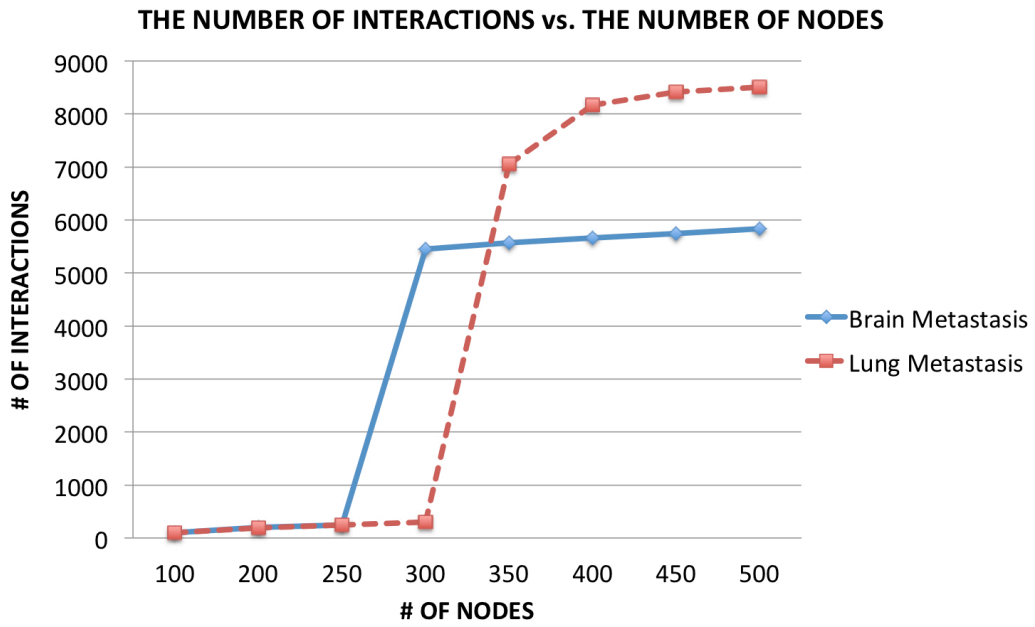

Figure A.7. This graph shows the increase in the number of interactions, as the number of nodes gets bigger [285].

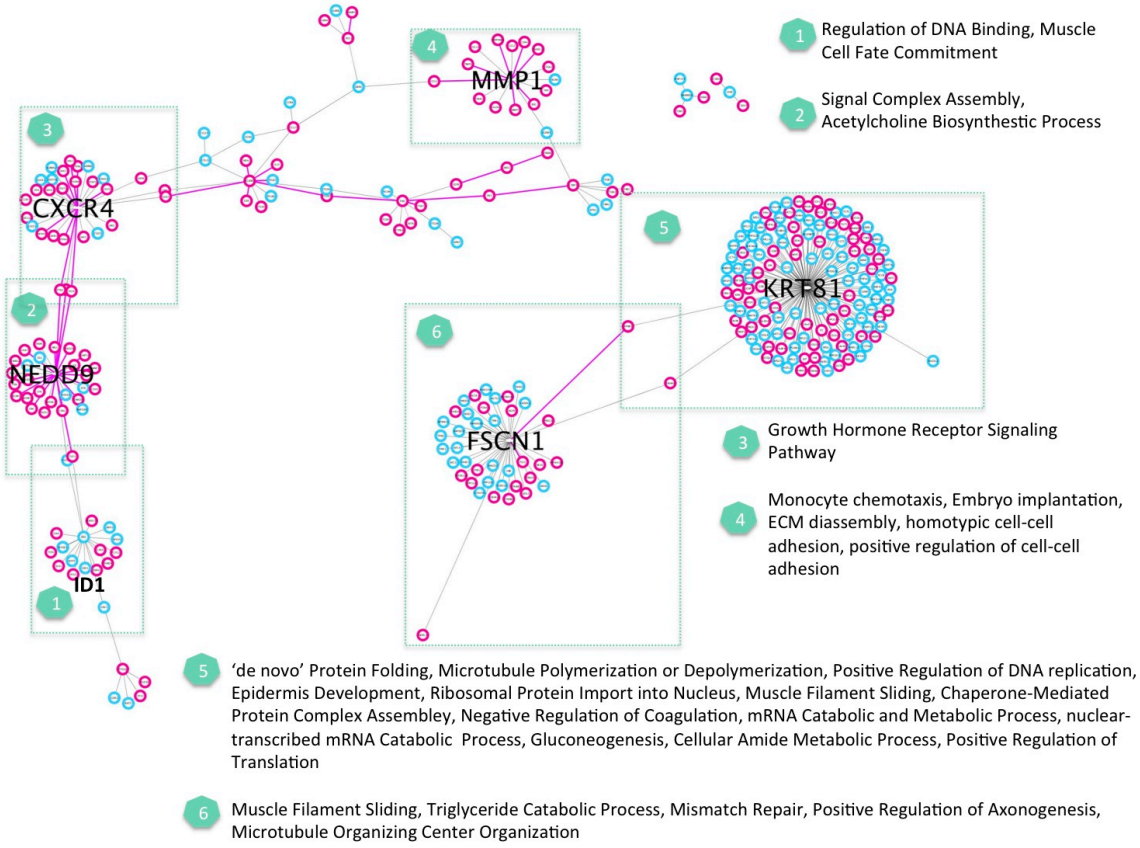

Figure A.8. The significantly enriched KEGG pathways in the clusters of LMSN.

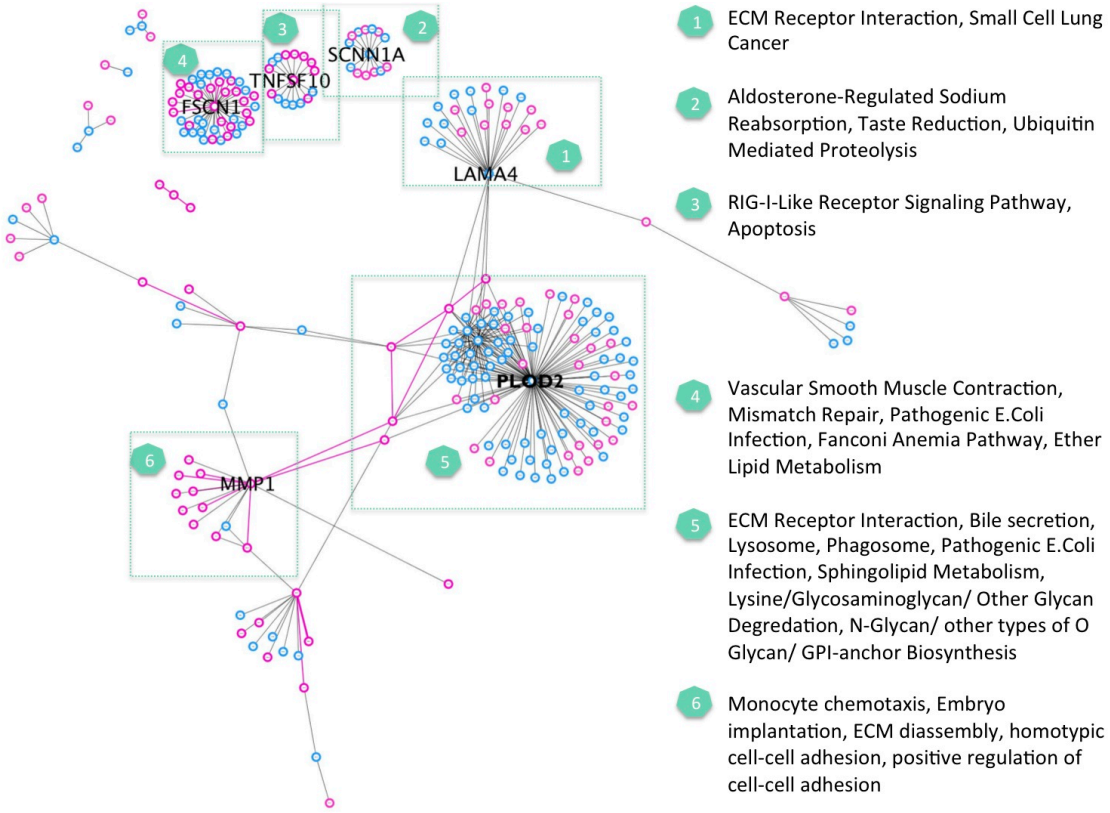

Figure A.9. The significantly enriched KEGG pathways in the clusters of BMSN.

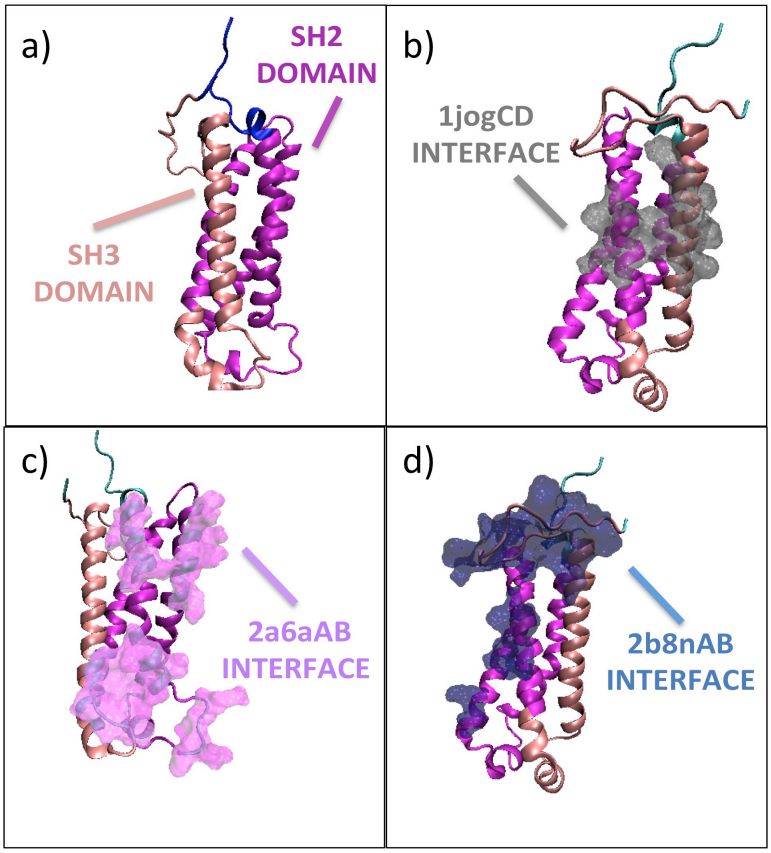

Figure A.10. NEDD9 is a hub protein with multiple interface architectures on different regions of its surface. a) There is only one PDB structure available in PDB (PDB ID: 2L81) that contains the SH3 and the SH2 domains. The first three most frequently observed interfaces (b) 1jogCD interface, c) 2a6aAB interface and d) 2b8nAB interface) of LMSN mapped on NEDD9.

Supplementary Tables:

Table A.1. List of PRISM Interaction Predictions for p53 Network. Out of 251 PRISM interaction predictions, 26 are present in Kohn's map, 59 are present in various PPI databases and 90 are present in STRING database. 104 interactions are validated totally.

PPI Predictions	Interface Template	Validation	PPI Predictions	Interface Template	Validation
APC - E2F1	1gl2AD		ERCC4 - SKP2	2astBC	
APC - CCNH	1vf6BD		ERCC4 - CDKN1B	1jsuBC	
APC - TFDP1	1gl2AD		FOS - E2F1	1gl2AD	STRING
APC - CCNA2	1jsuBC		FOS - RFC5	1gl2AD	
APC - JUN	1gl2AD		GADD45A - RAP1A	1c1yAB	
APC - MAPK10	1pq1AB		GADD45A - CDKN2D	1gveAB	
APC - BRCA1	1jsuBC		GADD45A - NFKB1	1gveAB	STRING
APC - RFC5	1jsuBC		GADD45A - SFN	1gveAB	

APC - XRCC6	1rkeAB		GADD45A - BAX	1gveAB	STRING
APC - CCNE1	1rkeAB		GADD45A - MNAT1	1gveAB	
APC - POLR2G	1cxzAB		GADD45A - APC	1gveAB	
APC - RB1	1jsuBC		GADD45A - PLK1	1gveAB	
APC - GTF2H1	1jsuBC		GADD45A - MAPK8	1lqbAB	STRING
APC - EP300	1g4yBR		GPIHBP1 - APC	1pq1AB	
APC - LIG1	1jsuBC		GPIHBP1 - CDK2	1rkeAB	
APC - PARP1	1jsuBC		GTF2H1 - CDKN1B	1jsuBC	
APC - EP300	1hx1AB		HDAC1 - XRCC5	1n62DF	STRING
APC - CCNB1	1jsuBC	KOHN'S MAP	HDAC1 - CDK1	1unlAD	BIOGRID
APC - RPA1	1jsuBC		HDAC1 - CDK6	1unlAD	STRING
APC - KAT2B	1rkeAB		JUN - POLR2G	2ahmCG	
APC - MAX	1gl2AD		JUN - E2F1	1gl2AD	STRING
APC - MDM2	1rkeAB		JUN - EP300	2ahmCG	BIOGRID
APC - MYC	1gl2AD	MINT	JUN - XRCC5	2ahmCG	
APC - XRCC1	1jsuBC		JUN - RAD51	2ahmCG	
APC - HMGB1	1g4yBR		JUN - SFN	2ahmCG	
APC - FOS	1gl2AD		JUN - MARK3	2ahmCG	
APC - MYC	1jsuBC	MINT	LIG3 - SFN	1e8oCD	
APC - RAD52	1pq1AB		LIG3 - RAD23B	1tf0AB	
BAX - XRCC5	1nw9AB		MAPK8 - RPA3	1lqbAB	
BAX - RAF1	1wmhAB	STRING	MAPK9 - E2F4	2btfAP	
BAX - EP300	1rkeAB	STRING	MDM2 - CCNA2	1nw9AB	KOHN'S MAP
CCNA2 - XRCC5	2ahmCG		MYC - CCNE1	1jsuBC	STRING
CCNA2 - EP300	1rkeAB	STRING	MYC - RAD52	1pq1AB	
CCNB1 - RPA3	1quqCD		MYC - MAX	3ezeAB	KOHN'S MAP
CCNB1 - CCNE1	1gveAB		MYC - CCNB1	1jsuBC	STRING
CCND1 - MYC	1jsuBC	STRING	MYC - RB1	1jsuBC	BIOGRID
CCND1 - CDKN1B	1jsuBC	KOHN'S MAP	MYC - CDH1	1jsuBC	STRING
CCNE1 - RPA3	1quqCD		MYC - CCNA2	1jsuBC	STRING
CCNE1 - MAPK8	1w36CD		MYC - CDKN1B	1jsuBC	INTACT
CCNE1 - MNAT1	1gveAB	STRING	MYC - E2F1	1gl2AD	STRING
CCNE1 - CDK5	1unlAD		PARP1 - XRCC5	2ahmCG	BIOGRID
CCNE1 - CASP3	1gveAB	STRING	PARP1 - RAF1	1wmhAB	STRING
CCNE1 - CDK5	1gveAB		PARP1 - CDKN1B	1jsuBC	STRING
CCNE1 - EP300	1rkeAB	KOHN'S MAP	PARP1 - SKP2	3ezeAB	
CCNE1 - CHEK1	1unlAD		PARP1 - CHEK1	1unlAD	MINT
CCNH - MYC	1jsuBC	BIOGRID	PARP1 - CCNH	1jsuBC	
CDK1 - CKS1B	1buhAB	KOHN'S MAP	PARP1 - E2F1	2ahmCG	INTACT
CDK2 - PLK1	1rkeAB	INTACT	PARP1 - NFKBIA	3ezeAB	STRING
CDK2 - ABL1	1rkeAB	INTACT	PARP1 - CDH1	1jsuBC	
CDK2 - APC	1rkeAB		PARP1 - MYC	1jsuBC	STRING
CDK2 - MAPK9	2btfAP		PCNA - BRCA1	1xkpAC	NCI-NATURE PID
CDK2 - SFN	1rkeAB		PCNA - CDKN1B	1pq1AB	STRING
CDK2 - CCNE1	1rkeAB	KOHN'S MAP	PCNA - NFKB1	1h9sAB	PATHWAY COMMONS
CDK2 - CCNE1	1unlAD	KOHN'S MAP	PLK1 - RPA3	1quqCD	
CDK2 - RB1	1nw9AB	KOHN'S MAP	PLK1 - CDH1	1lqbAB	STRING
CDK2 - TFDP2	1rkeAB	KOHN'S MAP	POLR2D - CDK5	1gveAB	
CDK2 - SKP2	1unlAD	KOHN'S MAP	POLR2G - MAPK9	2btfAP	
CDK2 - EP300	1rkeAB	KOHN'S MAP	POLR2G - TFDP1	2btfAP	
CDK2 - RPA3	1rkeAB	KOHN'S MAP	POLR2G - RPA3	2btfAP	
CDK2 - BAX	1rkeAB	STRING	POLR2G - E2F4	2btfAP	
CDK2 - E2F1	1e8oCD	KOHN'S MAP	RAD23B - TFDP2	1fxtAB	
CDK2 - CCNB1	1oiyBC	BIOGRID	RAD23B - CDKN1B	1jsuBC	
CDK2 - CKS1B	1buhAB	KOHN'S MAP	RAD51 - XRCC5	1rkeAB	STRING
CDK2 - XRCC5	1oiyBC		RAD51 - SFN	3ezeAB	

CDK2 - MARK3	1oiyBC	INTACT	RAD51 - SFN	1gveAB	
CDK4 - CDKN2D	1blxAB	BIOGRID	RAF1 - RAP1A	1c1yAB	KOHN'S MAP
CDK4 - CSNK2A2	1buhAB		RAF1 - RAD52	1lqbAB	
CDK4 - XRCC5	1rkeAB		RAF1 - SKP2	1wmhAB	
CDK4 - CDKN2D	1e8oCD	BIOGRID	RAF1 - FEN1	1wywAB	
CDK4 - MYC	1jsuBC	NCI-NATURE PID	RAF1 - RAD51	1wmhAB	
CDK4 - MAPK10	1buhAB		RAP1A - CDH1	1c1yAB	
CDK4 - KAT2B	1a9nAB		RB1 - WEE1	1unlAD	
CDK4 - ABL1	1gveAB		RB1 - PARP1	1gh6AB	STRING
CDK4 - CDK6	1buhAB	CELL-MAP	RB1 - RAD23B	1tf0AB	
CDK6 - CKS1B	1buhAB		RELA - SFN	1rkeAB	STRING
CDK6 - CCNE1	1unlAD	STRING	RELA - BRCA1	1xkpAC	BIOGRID
CDK6 - RAD51	1unlAD		RELA - NFKBIA	1oy3CD	BIOGRID
CDK6 - RAF1	1wywAB		RFC1 - XRCC5	1nw9AB	BIOGRID
CDK7 - CCND1	1xg2AB	STRING	RFC5 - RAD51	1gveAB	
CDK7 - RB1	1unlAD	STRING	RFC5 - RAD51	1t08AB	
CDK7 - RAD52	1lqbAB		RFC5 - CDK6	1l0oBC	
CDK7 - MYC	1jsuBC	STRING	RFC5 - MYC	1jsuBC	
CDK7 - HDAC1	1unlAD		RFC5 - PLK1	1gveAB	MINT
CDK7 - CDKN2D	1blxAB		RFC5 - EP300	1rkeAB	
CDK7 - CKS1B	1buhAB		RFC5 - CSNK2A2	1gveAB	
CDKN1B - RAD52	1jsuBC		RFC5 - XRCC6	1gveAB	INTACT
CDKN1B - CCNE1	1jsuBC	BIOGRID	RFC5 - MAX	1gl2AD	
CDKN1B - WEE1	1jsuBC	STRING	RFC5 - CDKN1B	1jsuBC	
CDKN1B - CDH1	1jsuBC	STRING	RFC5 - MYC	1gl2AD	
CDKN1B - CCNB1	1jsuBC	INTACT	RFC5 - XRCC6	1rkeAB	INTACT
CDKN1B - CCNA2	1jsuBC	KOHN'S MAP	RFC5 - CSNK2A1	1gveAB	
CDKN1B - RB1	1jsuBC	PATHWAY COMMONS	RFC5 - CCNE1	1gveAB	
CDKN2A - TP53	1lqbAB	BIOGRID	RFC5 - E2F1	1gl2AD	
CDKN2A - CHEK1	1blxAB		RPA1 - RPA2	1l1oEF	BIOGRID
CDKN2D - MAPK8	1blxAB		RPA1 - XRCC5	1rkeAB	MINT
CDKN2D - CDH1	1rypMN		RPA1 - RPA3	1quqCD	BIOGRID
CDKN2D - CDKN1B	1jsuBC	STRING	RPA1 - E2F1	1e8oCD	
CDKN2D - CCNA2	1e8oCD		RPA1 - CCNA2	1lqbAB	BIOGRID
CDKN2D - CCNB1	1e8oCD		RPA1 - CDKN1B	1jsuBC	
CDKN2D - CDK6	1blxAB	BIOGRID	RPA2 - RPA3	1quqCD	BIOGRID
CDKN2D - CHEK1	1blxAB		SKP1 - RPA3	1quqCD	
CDKN2D - MAPK9	1blxAB		SKP1 - E2F1	2c38SV	
CKS1B - MAPK10	1buhAB		SKP2 - SKP1	2astAB	KOHN'S MAP
CRK - RPA3	1quqCD		SKP2 - CCNE1	1gveAB	BIOGRID
CSNK2A1 - CKS1B	1buhAB		SKP2 - CKS1B	2astBC	BIOGRID
CSNK2A2 - CKS1B	1buhAB		SKP2 - SKP1	1fs2AD	KOHN'S MAP
E2F1 - SFN	1e8oCD		SKP2 - CDK5	1unlAD	
E2F1 - CDH1	1e8oCD		TAF1 - ERCC4	1lkyAB	
E2F1 - RPA2	2c38SV	STRING	TAF1 - RPA3	1quqCD	
E2F1 - EP300	2ahmCG	KOHN'S MAP	TAF1 - CDK4	1blxAB	
E2F1 - NFKBIA	2ahmCG	STRING	TFDP1 - XRCC5	1rkeAB	
E2F1 - EP300	1gl2AD	KOHN'S MAP	TFDP1 - EP300	1rkeAB	
E2F1 - XRCC5	2ahmCG		TFDP2 - RB1	1j2jAB	KOHN'S MAP
E2F1 - WEE1	1e8oCD		TFDP2 - EP300	1rkeAB	
E2F1 - MAX	1gl2AD		TFDP2 - E2F4	1cf7AB	KOHN'S MAP
E2F4 - XRCC5	1rkeAB		TP53 - TP53BP2	1ycsAB	BIOGRID
EP300 - MAX	2ahmCG	BIOGRID	XPA - CDKN1B	1jsuBC	
EP300 - ABL1	1rkeAB	MINT	XRCC1 - CCNE1	1gveAB	
ERCC1 - MYC	1jsuBC		XRCC1 - EP300	1h9sAB	
ERCC1 - EP300	1rkeAB		XRCC1 - CDKN1B	1jsuBC	

ERCC1 - JUN	2ahmCG		XRCC1 - PLK1	1gveAB	
ERCC1 - RPA3	1quqCD	KOHN'S MAP	XRCC1 - MYC	1jsuBC	
ERCC1 - MAX	2ahmCG		XRCC1 - CDH1	1tueFG	
ERCC1 - APC	1hx1AB		XRCC5 - ABL1	1rkeAB	KOHN'S MAP
ERCC1 - CDKN1B	1jsuBC		XRCC6 - MDM2	1rkeAB	STRING
			XRCC6 - MNAT1	1gveAB	

Table A.2. PPIs and Interface Templates used for modeling the interactions in the IL-10 centered network

PPI	Interface Templates	PPI	Interface Templates
PDGFA-PDGFB	1vppVW 1n7fAB	SIRPG-F3	-
IL10-IL10RA	1qjcAB 1zdnAB	CTSB-A2M	1wurAE
UBC-TP63	1jo0AB	IL10RA-UBC	-
UBC-ERBB4	1u2eAC	A2M-LCAT	1lnuDF 1mu4AB 1yrlAC 1u2eAC
SIRPG-ERBB4	-	ANXA6-A2M	1hkxMN
UBC-JAK2	-	MMP2-IL1B	-
A2M-KLK13	1h7sAB 1jmjAB 1m2oAB 1yllAB 1mzhAB 1ns5AB 1bvsAB 1g8tAB	A2M-PAEP	1vbfAB 1f9aAF 1lehAB 1epaAB 2fa1AB 1yqdAB 1bebAB 1y9iAD 1okjAB 2a1bAB 1pbiAB 2a0sAB 1mu4AB 1mzhAB 1jieAB 1aorAB
A2M-AMBP	1wwhBD 1xsvAB	IL10RA-IL10RB	-
APP-APOE	2ffjAB 1a49AB 1zydAB 1ylmAB 1nh0AB 1qjcAB 1iokAB 1l8wAC 1t6uAF	B2M-A2M	1oh0AB 1ix1AB 1tr0AB 1xsvAB
IL10RA-JAK1	-	APOE-UBC	-
HSPA5-UBC	-	A2M-LEP	1mu4AB 2scpAB 1f74AC 1hkxMN
IL10RB-IL10	-	IL10-A2M	1mwqAB
SIRPG-IL10	-	A2M-CPB2	2b8nAB
APP-LRP1	1vr0BC 1dxxAD 1pl4AD 1fr3AB 1k9jAB 1g8tAB	PDGFB-A2M	2a0sAB 2b8nAB 1z8hCD 1wb1BD 1hssAB
LRP1-A2M	-	AMBP-CTSB	-
A2M-KLK3	2a0sAB 1jmjAB	APOE-CTSB	-
APOE-LRP1	1ylmAB	A2M-SPACA3	1cmbAB
IL10RB-IL22	-	UCN2-IL10RB	-
IGLL5-SIRPG	1t92AB	A2M-ADAM19	1wovAB 2fa1AB
LYZ-UBC	-	IL10RA-JAK2	-
IL10RB-UCN3	-	PDGFA-A2M	1yqhAB 1cqkAB
APOE-A2M	-	A2M-IL1B	1t92AB
A2M-APP	1v8pEF 1mn8AB 1jogCD 1nbqAB 1dxxAD 1fftAC 1jyaAB 1qorAB 1q7sAB 1iw8AB 1bo1AB 1a49AB 2b8nAB 2scpAB 1a96AB 2a1bAB 2a74DF	A2M-KLK2	1f9nBE 1oh0AB 1rd5AB 1pbiAB 2b8nAB 2b8tCD 1uypAE 1vjlAB 1jmjAB
SIRPG-CD47	-	NGF-A2M	1tqjCD 1jieAB
A2M-LYZ	2b8nAB 1aorAB	IL28B-IL10RB	2b99CE 1c2pAB
A2M-MMP2	1jogCD	ADAM19-UBC	-
A2M-HSPA5	-	TGFBI-A2M	1wdjAC 1oh0AB
CTSE-A2M	1zh8AB 1vmfAB	IL28A-IL10RB	2b99CE
LRP1-PDGFB	-	MYOC-A2M	-

UBC-APP	1qorAB 1barAB	BTRC-UBC	1mu4AB
A2M-ADAMTS1	1gveAB 2aalEF	A2M-TP63	1ns5AB
A2M-IL4	1mu4AB	TP63-HSPA5	-
BTRC-IL10RA	1wurAE	SHBG-A2M	1f9aAF 1z8hCD 1m4rAB
BTRC-TP63	2astAB	UBC-ANXA6	-
CELA1-A2M	-	UBC-CTSB	2a0sAB

Table A.3. PPIs and Interface Templates used for modeling the interactions in the brain/lung metastasis networks.

BMSN		LMSN	
PPI	**Interface Templates**	**PPI**	**Interface Templates**
PLS3-FSCN1	2bo4CD	NEDD9-PTK2B	3ezeAB 1jogCD 1zdnAB 1tb3AD 1fiuAB 1pbiAB 1t3uAB
BSG-MMP1	2b8nAB 1xx9CD	ITCH-CXCR4	2btfAP
MMP1-TNFSF11	1zdnAB 1g8tAB	PLS3-FSCN1	2bo4CD
LTBP1-FN1	1ywkAC 2b8nAB 1qjcAB	MICAL1-NEDD9	2b8nAB 1qjcAB 2a6aAB
FSCN1-KPNB1	1tueAH	NEDD9-PTPN11	2b8nAB
MMP7-MMP1	1qiaCD 1b3dAB	TNC-CNTN1	2b8nAB
CD44-FBN1	1oh0AB	FN1-TNC	2b8nAB 1u6iAF 1oh0AB 1u2eAC 1p65AB 1wb1BD 1yllAB 2a6aAB 1xmzAB
MMP1-CCL2	1nh0AB	TNC-ITGB1	2b8nAB
ERBB4-HBEGF	1moxAC 1nqlAB	BSG-MMP1	2b8nAB 1xx9CD
EGFR-HBEGF	1moxAC 1nqlAB	MYH9-CXCR4	2a6aAB
ITGA5-ITGB1	1kkmAB	NEDD9-CRKL	1zuwAC
FBN1-ITGA5	1kamAB 2btfAP	MMP1-TNFSF11	1zdnAB 1g8tAB
CSF3-CSF3R	1jzmAB 2b99CE 1cd9AB	LTBP1-FN1	1ywkAC 2b8nAB 1qjcAB
MMP1-SERPINA3	1jyaAB	NEDD9-BCAR1	1y0eAB 1jogCD 1y9iAD 1tljAB
MMP1-CD44	1jogCD	MSN-VCAM1	1xedAC
CSF3-ELA2	1gveAB 1jflAB	VCAM1-ELA2	1x8dAB 1a49AB
CD44-ITGA5	1eq2GJ 1qjcAB 1rd5AB 1okjAB	NEDD9-CRK	1vi6AB 1gveAB
MMP1-TIMP1	1bqqMT	PTPN6-CXCR4	1u0kAB
RNF135-RARRES3	-	VCAM1-EZR	1twjCD
LAMC1-PLOD2	-	FSCN1-KPNB1	1tueAH
ITGA3-PLOD2	-	NEDD9-ABL1	1t6uAF 1wmhAB
ITGA5-PLOD2	-	TNC-ITGA5	1symAB 1q5cAB 1f6fBC
PLOD2-FBN1	-	NEDD9-SMAD3	1sj1AB
PLOD2-CD44	-	CXCR4-JAK3	1s96AB
ITGB1-PLOD2	-	VCAM1-IL13	1qorAB
PLOD2-SETD3	-	MMP7-MMP1	1qiaCD 1b3dAB
PLOD2-AADACL1	-	NEDD9-DIMT1L	1pe0AB
PLOD2-PPT1	-	NEDD9-BCAR3	1p60AB
PLOD2-PH-4	-	NEDD9-SMAD1	1o60AB 1t6uAF
CD276-PLOD2	-	EREG-ERBB4	1nqlAB 1moxAC
TREML2-PLOD2	-	MMP1-CCL2	1nh0AB
TXNDC15-PLOD2	-	NEDD9-PIK3CA	1nh0AB 2erbAB 1v8pEF 1p5qAC
GLT25D1-PLOD2	-	CXCR4-STAT1	1n1bAB
BTD-PLOD2	-	PTK2-CXCR4	1mzhAB 1vr0BC 2bo4CD
SGSH-PLOD2	-	EREG-EGFR	1moxAC 1nqlAB

SEL1L-PLOD2	-	PTPRC-CXCR4	1l1yAD
SIAE-PLOD2	-	VCAM1-ITGB1	1kkmAB
C21orf29-PLOD2	-	NEDD9-PTK2	1k2fAB 2b8nAB 1jogCD 1yw0AD 1c4zAD 1xqcAB 1on2AB 1gveAB 1um0CD 1y9iAD
ACP2-PLOD2	-	NEDD9-TCF3	1jzmAB
ATP1B3-PLOD2	-	MMP1-SERPINA3	1jyaAB
DNASE2-PLOD2	-	NEDD9-ITCH	1jogCD 1rkeAB 2a6aAB
FUT11-PLOD2	-	MMP1-CD44	1jogCD
CD109-PLOD2	-	VCAM1-ITGB7	1jd1AB
CLN5-PLOD2	-	NEDD9-PXN	1j2rCD
PLOD2-PCYOX1	-	CXCR4-CD74	1iieAB
PLOD2-P4HA1	-	CDH1-NEDD9	1iawAB
PLOD2-EGFL11	-	NEDD9-CHAT	1gveAB
PLOD2-CLPTM1	-	CXCR4-VAV1	1fr3AB
PLOD2-FKBP9	-	MMP9-CXCL1	1e8oCD 1djrDE
PLOD2-C1orf85	-	MMP1-TIMP1	1bqqMT
PLOD2-BTN2A1	-	RNF135-RARRES3	-
PLOD2-SLC3A2	-	CCL17-VCAM1	-
ECE1-PLOD2	-	CCL22-VCAM1	-
C20orf3-PLOD2	-	ITGAD-VCAM1	-
COL5A1-PLOD2	-	KRT81-MYH2	-
PLOD2-SUMF1	-	KRT81-PGAM2	-
PLOD2-PODXL2	-	KRT81-KRT34	-
PLOD2-TPBG	-	KRT81-VSIG8	-
KIAA0090-PLOD2	-	KRT81-LRRC15	-
PLOD2-GAA	-	KRT81-USP15	-
DPP7-PLOD2	-	KRT81-FABP4	-
ADNP-PLOD2	-	KRT81-YWHAE	-
PLOD2-NPC1	-	KRT81-ADSL	-
PLOD2-MAN2B1	-	KRT81-CSTF1	-
ATP6AP1-PLOD2	-	KRT81-KPNB1	-
STT3B-PLOD2	-	KRT81-RNF40	-
CD97-PLOD2	-	KRT81-TUT1	-
PIGS-PLOD2	-	KRT81-LSM6	-
PLOD2-LYPLA3	-	KRT81-LSM2	-
STT3A-PLOD2	-	KRT81-C19orf50	-
STCH-PLOD2	-	KRT81-PRPF4	-
PLOD2-PIGT	-	KRT33B-KRT81	-
LAMP2-PLOD2	-	KRT81-KRT31	-
ERO1L-PLOD2	-	KRT81-PRSS1	-
NUP210-PLOD2	-	KRT81-PRPF3	-
PLXNB2-PLOD2	-	KRT85-KRT81	-
SCARB1-PLOD2	-	KRT81-IRS4	-
PLOD3-PLOD2	-	KRT81-ECH1	-
PLOD2-CD36	-	KRT81-FAM33A	-
LNPEP-PLOD2	-	KRT81-RBM25	-
LAMP1-PLOD2	-	KRT81-GRPEL1	-
SSR1-PLOD2	-	KRT81-MCCC1	-
PLOD2-HYOU1	-	KRT81-PCCA	-
HEXA-PLOD2	-	KRT81-KRT32	-
GNS-PLOD2	-	KRT81-MRPS27	-
LAMA5-PLOD2	-	KRT81-R3HCC1	-
PLOD2-M6PR	-	KRT81-KRT78	-
PLOD2-NOMO1	-	KRT81-KRT16	-
PLOD1-PLOD2	-	KRT81-DCD	-
CTSD-PLOD2	-	KRT81-CDCA4	-
NCSTN-PLOD2	-	KRT81-PCCB	-

PLOD2-L1CAM	-	KRT81-KRT2	-
PLOD2-ITGAV	-	KRT81-PRDX4	-
PSAP-PLOD2	-	KRT81-S100A9	-
PLOD2-VPS37C	-	KRT81-KRT6B	-
TXNDC10-PLOD2	-	KRT81-PC	-
PLOD2-GBA	-	KRT81-C18orf24	-
PLOD2-SCARB2	-	KRT81-PRDX3	-
TPP1-PLOD2	-	KRT81-KRT14	-
PLOD2-IGF2R	-	KRT81-SERTAD4	-
GUSB-PLOD2	-	KRT81-KRT5	-
CLU-PLOD2	-	KRT81-BAG2	-
FAM107B-PLOD2	-	KRT81-SNF1LK2	-
FKBP10-PLOD2	-	KRT81-PPP2R2B	-
PLOD2-ASAH1	-	KRT81-TUBA1B	-
PLOD2-A2M	-	KRT81-DDX6	-
SERPINA1-PLOD2	-	KRT81-HCFC2	-
PLOD2-STOM	-	KRT81-RECQL4	-
PLOD2-LGALS3BP	-	KRT81-PPP4C	-
PLOD2-BSG	-	KRT81-RBM7	-
TUBB4-PLOD2	-	KRT81-KRT1	-
PLOD2-PIGR	-	KRT81-PPIH	-
CRTAP-PLOD2	-	KRT81-RPL30	-
PLOD2-LEPRE1	-	KRT81-LSM4	-
PLOD2-MARCKS	-	KRT81-SKIV2L2	-
PLOD2-UNC84B	-	KRT81-KRT10	-
FOLR1-PLOD2	-	KRT81-KRT9	-
DSG1-PLOD2	-	KRT81-TCP1	-
BCAN-MMP1	-	KRT81-LSM8	-
CCL13-MMP1	-	KRT81-SELENBP1	-
PLOD2-TFRC	-	KRT81-CCT7	-
TUBB-PLOD2	-	KRT81-RPLP1	-
RPA2-PLOD2	-	KRT81-CCT8	-
DYNLL1-PLOD2	-	KRT81-RPLP2	-
TFPI-MMP1	-	KRT81-CCT6A	-
CCL7-MMP1	-	KRT81-HIST1H1E	-
CMA1-MMP1	-	KRT81-TUFM	-
ITGA2-MMP1	-	KRT81-TMPO	-
COL2A1-MMP1	-	KRT81-NUFIP2	-
LAMC1-LAMA4	-	KRT81-CALM1	-
ITGA3-LAMA4	-	KRT81-CAPZA1	-
ITGA5-LAMA4	-	KRT81-CCT4	-
FBN1-LTBP1	-	KRT81-CALML3	-
FBN2-LTBP1	-	KRT81-CCT2	-
ITGB1-LAMA4	-	KRT81-PRDX1	-
HBEGF-CD44	-	KRT81-MRPL12	-
MMP1-IGFBP3	-	KRT81-CCT3	-
HBEGF-MMP7	-	KRT81-EEF1B2	-
LAMC3-LAMA4	-	KRT81-PPP2CA	-
COL13A1-NID2	-	KRT81-CDC42	-
LAMB2-LAMA4	-	KRT81-MATR3	-
HBEGF-LTBP3	-	KRT81-PPP2R1B	-
LTBP1-IGFBP3	-	KRT81-TUBB2C	-
HBEGF-CD9	-	KRT81-CKB	-
TGM1-RARRES3	-	KRT81-HSPA9	-
ITGB3-LAMA4	-	KRT81-VIM	-
ITGA1-COL13A1	-	KRT81-CCT5	-
HBEGF-CD82	-	KRT81-ALDOA	-
COL13A1-NID1	-	KRT81-PPP2CB	-

SPARC-COL13A1	-	KRT81-HNRNPL	-
FN1-COL13A1	-	KRT81-HNRPH1	-
LAMA4-ATF7IP	-	KRT81-PPP2R1A	-
HSPG2-COL13A1	-	KRT81-SKIL	-
HBEGF-FBLN1	-	KRT81-JUP	-
ITGB5-LTBP1	-	KRT81-S100A7	-
ZHX1-LAMA4	-	KRT81-JUN	-
UNC119-LAMA4	-	KRT81-HSPD1	-
PTN-LAMA4	-	KRT81-MCM5	-
LAMA4-MEF2C	-	KRT81-IQGAP1	-
LAMA4-C1orf103	-	KRT81-RPL14	-
RIF1-LAMA4	-	KRT81-HSPA5	-
BRD7-LAMA4	-	KRT81-MEPCE	-
UBR1-LAMA4	-	KRT81-RPL21	-
LAMA4-TP53	-	KRT81-RPL19	-
LAMA4-MED31	-	KRT81-RPL4	-
SUMO2-LAMA4	-	KRT81-RPS8	-
COPS6-LAMA4	-	KRT81-EEF1A1	-
LAMA4-APC	-	KRT81-RPS6	-
TUBB2A-LAMA4	-	LGALS7-KRT81	-
GAPDH-LAMA4	-	KRT81-HSP90AA1	-
EEF1A1-LAMA4	-	KRT81-RPL6	-
LAMB1-LAMA4	-	KRT81-TUBB	-
HBEGF-S100A4	-	KRT81-RPS14	-
ATN1-LTBP1	-	ANXA2-KRT81	-
TGM2-LTBP1	-	KRT81-RPL24	-
HBEGF-ZBTB16	-	KRT81-RPS11	-
HBEGF-BAG1	-	KRT81-RPS27L	-
HBEGF-NRD1	-	KRT81-DDX3X	-
FSCN1-DNAJB9	-	KRT81-RPL23	-
PTGS2-NUCB1	-	KRT81-MYH4	-
TP53-PTGS2	-	KRT81-PRPF31	-
SEPP1-EGFR	-	KRT81-ACTB	-
FSCN1-EDIL3	-	KRT81-KRT15	-
FSCN1-RAB1A	-	KRT81-YBX1	-
PTGS2-CAV1	-	KRT81-RPS4X	-
FSCN1-NDEL1	-	KRT81-MYL6	-
FSCN1-KIAA1949	-	KRT81-RPLP0	-
FSCN1-NDE1	-	KRT81-RPL7A	-
PTGS2-BAT5	-	KRT81-HSPA1A	-
COPS5-PTGS2	-	KRT81-PSMD11	-
TNFRSF10D-TNFSF10	-	KRT81-C1QBP	-
FSCN1-PAFAH1B2	-	KRT81-RPL17	-
FSCN1-PAFAH1B1	-	KRT81-HNRPF	-
NGFR-FSCN1	-	KRT81-S100A8	-
FSCN1-PAFAH1B3	-	KRT81-RAN	-
SEPP1-PTK2	-	ITGA8-TNC	-
FSCN1-PMS1	-	CXCL1-DARC	-
FSCN1-MTMR15	-	CCL13-MMP1	-
TPM2-FSCN1	-	FSCN1-DNAJB9	-
FSCN1-FANCG	-	ITGA9-VCAM1	-
FSCN1-TPM1	-	DPP4-CXCR4	-
FSCN1-FANCD2	-	FSCN1-YWHAE	-
FSCN1-PRKCA	-	CXCR4-ADRBK2	-
PPP1CB-FSCN1	-	ID1-ELSPBP1	-
FSCN1-PPP1CA	-	TGM1-RARRES3	-
TMOD3-FSCN1	-	CCL7-MMP1	-
SASS6-FSCN1	-	ELSPBP1-NEDD9	-

FSCN1-FANCI	-	FSCN1-RAB1A	-
PRKCD-FSCN1	-	BCAN-MMP1	-
FSCN1-CAPZA2	-	FSCN1-NDEL1	-
FSCN1-PPP1R12A	-	FSCN1-EDIL3	-
FSCN1-C19orf21	-	FSCN1-KIAA1949	-
FSCN1-RFC3	-	FSCN1-NDE1	-
FSCN1-CORO1C	-	ID1-MYF6	-
FSCN1-RAPGEF2	-	FSCN1-PAFAH1B2	-
ACTA1-FSCN1	-	ITGA9-TNC	-
FSCN1-CTNNB1	-	FSCN1-PAFAH1B1	-
FSCN1-MLH1	-	NGFR-FSCN1	-
FSCN1-PMS2	-	FSCN1-PMS1	-
FSCN1-YWHAZ	-	FSCN1-MTMR15	-
FSCN1-PCNA	-	ELA2-CXCR4	-
FSCN1-YWHAE	-	FSCN1-FANCG	-
FSCN1-UBB	-	FSCN1-PAFAH1B3	-
FSCN1-RFC4	-	PRKCD-FSCN1	-
FSCN1-ACTC1	-	FSCN1-PRKCA	-
TNFRSF11B-TNFSF10	-	ITGA4-VCAM1	-
SCNN1A-SCNN1G	-	FSCN1-FANCD2	-
ACP5-TNFSF10	-	TPM2-FSCN1	-
HECW1-SCNN1A	-	FSCN1-FANCI	-
TNFRSF10C-TNFSF10	-	FSCN1-CTNNB1	-
SCNN1A-SCNN1B	-	PPP1CB-FSCN1	-
IER3-TNFSF10	-	FSCN1-RAPGEF2	-
TNFRSF10A-TNFSF10	-	ACTA1-FSCN1	-
TNFRSF10B-TNFSF10	-	FSCN1-PPP1CA	-
OTUD7B-TNFSF10	-	FSCN1-PMS2	-
WWP2-SCNN1A	-	FSCN1-RFC3	-
CFLAR-TNFSF10	-	FSCN1-TPM1	-
CASP10-TNFSF10	-	TFPI-MMP1	-
FADD-TNFSF10	-	FSCN1-MLH1	-
SNX3-SCNN1A	-	CXCR4-CXCL12	-
DKFZP564O0523-TNFSF10	-	FSCN1-CAPZA2	-
WWP1-SCNN1A	-	FSCN1-YWHAZ	-
STX1A-SCNN1A	-	TMOD3-FSCN1	-
CASP8-TNFSF10	-	SASS6-FSCN1	-
NEDD4L-SCNN1A	-	FSCN1-ACTC1	-
CUL3-TNFSF10	-	FSCN1-UBB	-
SQSTM1-TNFSF10	-	FSCN1-PCNA	-
PELI1-IRAK4	-	FSCN1-PPP1R12A	-
SCNN1A-NEDD4	-	FSCN1-C19orf21	-
TSG101-SCNN1A	-	FSCN1-CORO1C	-
USP10-SCNN1A	-	FSCN1-RFC4	-
ITCH-SCNN1A	-	CXCL1-IL8RA	-
UBE2I-SCNN1A	-	CMA1-MMP1	-
IRAK2-PELI1	-	ID1-MYF5	-
PELI1-IRAK1	-	CTSG-VCAM1	-
HRAS-RARRES3	-	CXCR4-CCR5	-
C13orf15-CDC2	-	ID1-ELK4	-
BCAN-MMP7	-	CXCR4-SDC4	-
ITGA3-LAMC1	-	MYOG-ID1	-
ITGA5-LAMC1	-	ID1-TCF3	-
ITGA5-ITGA3	-	PTPN11-CXCR4	-
LAMC1-FBN1	-	NEDD9-SH2D3C	-
ITGA3-FBN1	-	CXCR4-CTSG	-
LAMC1-CD44	-	ID2-NEDD9	-

ITGA3-CD44	-	ID1-IFI16	-
TFPI-MMP7	-	NCAN-TNC	-
ITGB1-LAMC1	-	ITGA2-MMP1	-
ITGB1-ITGA3	-	ID1-HES1	-
LAMC1-SETD3	-	JAK2-CXCR4	-
LAMC1-SEL1L	-	CXCR4-SOCS3	-
LAMC1-AADACL1	-	DOCK3-NEDD9	-
LAMC1-PPT1	-	MMP1-IGFBP3	-
LAMC1-PH-4	-	CXCR4-STAT2	-
LAMC1-ACP2	-	STAT5B-CXCR4	-
LAMC1-PCYOX1	-	RAPGEF1-NEDD9	-
LAMC1-P4HA1	-	CXCR4-SOCS1	-
LAMC1-ATP1B3	-	MYOD1-ID1	-
CD276-LAMC1	-	NEDD9-PTPN12	-
TREML2-LAMC1	-	COL2A1-MMP1	-
TXNDC15-LAMC1	-	CXCR4-CD4	-
GLT25D1-LAMC1	-	JAK1-CXCR4	-
BTD-LAMC1	-	STAT3-CXCR4	-
SGSH-LAMC1	-	GNA13-CXCR4	-
SIAE-LAMC1	-	ID1-CAV1	-
C21orf29-LAMC1	-	TRIP6-NEDD9	-
DNASE2-LAMC1	-	ARRB2-CXCR4	-
FUT11-LAMC1	-	ID1-TCF12	-
CD109-LAMC1	-	ID1-ELK1	-
CLN5-LAMC1	-	GNAI1-CXCR4	-
LAMC1-EGFL11	-	LYN-NEDD9	-
LAMC1-CLPTM1	-	LCK-NEDD9	-
LAMC1-FKBP9	-	FYN-NEDD9	-
LAMC1-C1orf85	-	CXCL1-IL8RB	-
LAMC1-BTN2A1	-	NCK1-NEDD9	-
LAMC1-SLC3A2	-	ID1-TCF4	-
LAMC1-ECE1	-	CXCR4-HSPA8	-
LAMC1-C20orf3	-	ID1-RUNX1T1	-
FBN2-FBN1	-	ITGB6-TNC	-
LAMC1-COL5A1	-	SMAD2-NEDD9	-
LAMC1-SUMF1	-	FBN2-LTBP1	-
LAMC1-PODXL2	-	ZYX-NEDD9	-
LAMC1-TPBG	-	ID1-CASK	-
KIAA0090-LAMC1	-	ID1-IKBKG	-
LAMC1-GAA	-	ID1-PSMD4	-
DPP7-LAMC1	-	PTPRB-TNC	-
LAMC1-ADNP	-	LTBP1-IGFBP3	-
LAMC1-NPC1	-	EGFR-TNC	-
ITGA3-SETD3	-	ITGB5-LTBP1	-
ITGA3-CD276	-	PTGS2-NUCB1	-
ITGA3-TREML2	-	PTGS2-CAV1	-
ITGA3-TXNDC15	-	TGM2-LTBP1	-
ITGA3-GLT25D1	-	FBN1-LTBP1	-
ITGA3-BTD	-	ATN1-LTBP1	-
ITGA3-SGSH	-	PTGS2-BAT5	-
ITGA3-SEL1L	-	TP53-PTGS2	-
ITGA3-AADACL1	-	COPS5-PTGS2	-
ITGA3-PPT1	-	HRAS-RARRES3	-
ITGA3-PH-4	-	CD247-LY6E	-
ITGA3-SIAE	-	FCGR2B-LY6E	-
ITGA3-C21orf29	-	CCL17-DARC	-
ITGA3-ACP2	-	DPP4-CCL22	-
ITGA3-PCYOX1	-	CCR8-CCL17	-

ITGA3-P4HA1	-	CCL17-CCR4	-
ITGA3-ATP1B3	-	CCL22-CCL19	-
ITGA3-EGFL11	-	CCL22-CCR4	-
ITGA3-DNASE2	-	CCL2-DARC	-
ITGA3-FUT11	-	ITGA8-NPNT	-
ITGA3-CD109	-		
ITGA3-CLPTM1	-		
ITGA3-CLN5	-		
ITGA3-FKBP9	-		
ITGA3-C1orf85	-		
ITGA3-BTN2A1	-		
LAMC1-MAN2B1	-		
ITGA3-SLC3A2	-		

Table A.4. A list of Proteins in the IL-10 Centered Protein-Protein Interaction Network. The Distance from IL-10 column provides the degree of contiguity of the proteins to IL-10 protein. For example, if a protein is a first-degree neighbor of IL-10, its distance from IL-10 is 1.

Protein Name	Protein Name	Distance from IL10	Source of Structural Data
A2M	Alpha-2-macroglobulin	1	PDB
ADAM19	Disintegrin and metalloproteinase domain-containing protein 19	2	Homology Modeling
ADAMTS1	A disintegrin and metalloproteinase with thrombospondin motifs 1	2	PDB
AMBP	Alpha-1 microglycoprotein	2	PDB
ANXA6	Annexin A6	2	PDB
APOE	Apolipoprotein E	2	PDB
APP	Amyloid beta A4 protein	2	PDB
B2M	Beta-2-microglobulin	2	PDB
BTRC	F-box/WD repeat-containing protein 1A	2	PDB
CD47	Leukocyte surface antigen CD47	2	PDB
CELA1	Chymotrypsin-like elastase family member 1	2	Homology Modeling
CPB2	Carboxypeptidase B2	2	PDB
CTSB	Cathepsin B	2	PDB
CTSE	Cathepsin E	2	PDB
ERBB4	Receptor tyrosine-protein kinase erbB-4	2	PDB
F3	Tissue factor	2	PDB
HSPA5	78 kDa glucose-regulated protein	2	PDB

IGHV3-6	Ig heavy chain V region 3-6	2	N/A
IGLL5	Immunoglobulin lambda-like polypeptide 5	2	Homology Modeling
IL10	Interleukin-10	0	PDB
IL10RA	Interleukin-10 receptor subunit alpha	1	PDB
IL10RB	Interleukin-10 receptor subunit beta	1	PDB
IL1B	Interleukin-1 beta	2	PDB
IL22	Interleukin-22	2	PDB
IL28A	Interleukin 28A	2	Homology Modeling
IL28B	Interferon lambda-3	2	PDB
IL4	Interleukin 4	2	PDB
JAK1	Tyrosine-protein kinase JAK1	2	PDB
JAK2	Tyrosine-protein kinase JAK2	2	PDB
KLK13	Kallikrein-13	2	Homology Modeling
KLK2	Kallikrein-2	2	Homology Modeling
KLK3	Prostate-specific antigen	2	PDB
LCAT	Phosphatidylcholine-sterol acyltransferase	2	Homology Modeling
LEP	Leptin	2	PDB
LRP1	Prolow-density lipoprotein receptor-related protein 1	2	PDB
LYZ	Lysozyme	2	PDB
MMP2	72 kDa type IV collagenase	2	PDB
MYOC	Myocilin	2	Homology Modeling
NGF	Beta-nerve growth factor	2	PDB
PAEP	Glycodelin	2	Homology Modeling
PDGFA	Platelet-derived growth factor subunit A	2	PDB
PDGFB	Platelet-derived growth factor subunit B	2	PDB
SHBG	Sex hormone binding globulin	2	PDB
SIRPG	Signal-regulatory protein gamma	1	PDB
SPACA3	Sperm acrosome membrane-associated protein 3	2	Homology Modeling
TGFBI	Transforming growth factor-beta-induced protein ig-h3	2	PDB
TP63	Tumor protein 63	2	PDB
UBC	Polyubiquitin-C	2	PDB
UCN2	Urocortin-2	2	PDB
UCN3	Urocortin-3	2	PDB

Table A.5. Protein List of Kohn's MIM. Kohn's molecular interaction map (MIM) has some nodes that do not have a protein counterpart, or some nodes correspond to multiple proteins. We updated Kohn's MIM's nodes by removing or expanding some of them.

Kohn's Original Nodes	Kohn's Nodes Updated
14_3_3	SFN
Abl	ABL1
APC	APC
Bax	BAX
BRCA1	BRCA1
Casp3	CASP3
CycA	CCNA2
CycB	CCNB1
CycD	CCND1
CycE	CCNE1
CycH	CCNH
E-cad	CDH1
Cdk1	CDK1
Cdk2	CDK2
Cdk4-6	CDK4, CDK5, CDK6
Cdk7	CDK7
p16	CDKN2A
p19ARF	CDKN2A
Chk1	CHEK1
Cks1	CKS1B
Crk	CRK
E2F1-2-3	E2F1
E2F4	E2F4
ERCC1	ERCC1
XPF	ERCC4
Fos	FOS
HDAC1	HDAC1
DP1-2	TFDP1, TFDP2
JNK	MAPK8, MAPK9, MAPK10
MAPK	MAPK8, MAPK9, MAPK10
FEN-1	FEN1
C-TAK1	MARK3
HMG	HMGB1
HR23B	RAD23B
Jun	JUN
Mdm2	MDM2
Gadd45	GADD45A
Ligase_1	LIG1
Ligase_3	LIG3

Max	MAX
Myc	MYC
RPA	RPA1, RPA2, RPA3
CK2	CSNK2A1, CSNK2A2
p27	CDKN1B
p300	EP300
p36MAT1	MNAT1
p53	TP53
PARP	PARP1
pCAF	KAT2B
PCNA	PCNA
Plk1	PLK1
pRb	RB1
Rad51	RAD51
Rad52	RAD52
Raf1	RAF1
Ras	RAP1A
RF-C	RFC1
RPase_2	POLR2D
Skp1	SKP1
Skp2	SKP2
TAFII250	TAF1
TFIIH	GTF2H1
Wee1	WEE1
XPA	XPA
XRCC1	XRCC1
Ku70	XRCC6
Ku80	XRCC5
AP2	TFAP2A, TFAP2B, TFAP2C
ATM	ATM
Cdc25A	CDC25A
Cdc25C	CDC25B
C-EBP	CEBPA, CEBPB, CEBPC, CEBPD, CEBPE,CEBPG
CK1d-k	CSNK1D, CSNK1E
CSB	ERCC6
HBP1	GPIHBP1
Dpase_a	POLA1, POLA2
DMP1	DMTF1
Dpase_b	POLB
Dpase_d	POLD1, POLD2, POLD3, POLD4
dsDNA	-
E2F6	E2F6
DNA-PK	PRKDC
Histones	-
Karp-1	KARP-1
E2F5	E2F5

p57	CDKN1C
p68	-
Paxillin	PXN
PKC	PRKCA, PRKCB
Rep_fork	-
RHA	-
SL1	TAF1A
Sp1	SP1
ssb	-
ssDNA	-
U-glyc	UNG
Myt1	PKMYT1
p107	RBL1
p130	RBL2
XPB	ERCC3
XPC	XPC
XPD	ERCC2
p21	CDKN1A
TBP	TBP

Table A.6. The Updated Interactions' List of Kohn's MIM. If a node was replaced with multiple proteins, the number of interactions automatically increased. We searched STRING database for validating the new edges and picked the ones, which were coming from high throughput experiments or databases.

Interaction	String Prediction Method	Interaction	String Prediction Method
TFAP2A ppi MYC	EXPERIMENTS	TFDP1 ppi E2F6	EXPERIMENTS
TFAP2B ppi MYC	EXPERIMENTS	TFDP2 ppi E2F6	EXPERIMENTS
TFAP2A ppi RB1	EXPERIMENTS	TFDP1 ppi RBL1	EXPERIMENTS
CSNK1D ppi TP53	EXPERIMENTS	TFDP2 ppi RBL1	EXPERIMENTS
CSNK1E ppi TP53	EXPERIMENTS	TFDP1 ppi RBL2	EXPERIMENTS
TP53 ppi PRKCA	EXPERIMENTS	TFDP2 ppi RBL2	EXPERIMENTS
PARP1 ppi POLA1	EXPERIMENTS	TFDP1 ppi CCNA2	DATABASES
PARP1 ppi POLA2	EXPERIMENTS	TFDP2 ppi CCNA2	DATABASES
PCNA ppi POLD1	EXPERIMENTS	TFDP1 ppi CDK2	DATABASES
PCNA ppi POLD2	EXPERIMENTS	TFDP2 ppi CDK2	DATABASES
PCNA ppi POLD3	EXPERIMENTS	TP53 ppi TFDP1	EXPERIMENTS
PCNA ppi POLD4	EXPERIMENTS	RB1 ppi TFDP1	EXPERIMENTS
RB1 ppi CEBPB	EXPERIMENTS	RB1 ppi TFDP2	EXPERIMENTS
RB1 ppi CEBPD	EXPERIMENTS	TP53 ppi MAPK8	EXPERIMENTS
RB1 ppi CEBPE	EXPERIMENTS	TP53 ppi MAPK9	EXPERIMENTS
RPA1 ppi POLA1	EXPERIMENTS	TP53 ppi MAPK10	EXPERIMENTS
RPA3 ppi POLA1	DATABASES	TP53 ppi RPA1	EXPERIMENTS
RPA3 ppi POLA2	DATABASES	CDK2 ppi RPA3	DATABASES

RPA2 ppi UNG	EXPERIMENTS	CCNA2 ppi RPA3	EXPERIMENTS
CDC25A ppi CDK4	DATABASES	TP53 ppi RPA1	EXPERIMENTS
CDC25A ppi CDK6	DATABASES	RAD51 ppi RPA1	EXPERIMENTS
CDK4 ppi CCND1	EXPERIMENTS	RAD51 ppi RPA3	DATABASES
CDK4 ppi CDKN2A	EXPERIMENTS	RAD52 ppi RPA1	EXPERIMENTS
CDK4 ppi CDKN2A	EXPERIMENTS	RAD52 ppi RPA2	EXPERIMENTS
CDK6 ppi CDKN2A	EXPERIMENTS	RAD52 ppi RPA3	EXPERIMENTS
CDK7 ppi CDK4	EXPERIMENTS	RPA3 ppi RAD23B	DATABASES
CDK7 ppi CDK5	EXPERIMENTS	RPA1 ppi ERCC4	EXPERIMENTS
CCNH ppi CDK4	DATABASES	RPA3 ppi ERCC4	DATABASES
CCNH ppi CDK5	EXPERIMENTS	RPA3 ppi ERCC1	DATABASES
TFDP1 ppi E2F1	EXPERIMENTS	RPA3 ppi GTF2H1	DATABASES
TFDP2 ppi E2F1	EXPERIMENTS	XPA ppi RPA1	EXPERIMENTS
TFDP1 ppi E2F4	EXPERIMENTS	XPA ppi RPA3	DATABASES
TFDP2 ppi E2F4	EXPERIMENTS	XPA ppi RPA2	EXPERIMENTS
TFDP1 ppi E2F5	EXPERIMENTS	XPC ppi RPA3	DATABASES
TFDP2 ppi E2F5	EXPERIMENTS	TP53 ppi CSNK2A1	EXPERIMENTS

Table A7. RMSD values of CDK6 structures. We highlighted the RMSD values higher than 2.5 with red.

		CHAIN 2					
CHAIN 1	**RMSD**	**1BLX_A**	**1XO2_B**	**2EUF_B**	**2F2C_B**	**3NUP_A**	**3NUX_A**
	1BLX_A	-	2.67	2.8	2.82	0.78	0.92
	1XO2_B	2.67	-	0.88	1.06	1.69	1.77
	2EUF_B	2.8	0.88	-	0.9	1.68	1.78
	2F2C_B	2.82	1.06	0.9	-	1.95	1.98
	3NUP_A	0.78	1.69	1.68	1.95	-	0.4
	3NUX_A	0.92	1.77	1.78	1.98	0.4	-

Table A.8. We have tested the evidence of the presence of the genes of lung metastasis sub-network in different databases.

Gene	Breast Tissue Source of Evidence	Gene	Breast Tissue Source of Evidence	Gene	Breast Tissue Source of Evidence	Gene	Breast Tissue Source of Evidence	Gene	Breast Tissue Source of Evidence
ABL1	HPRD	DCD	TIGER	ITGB6	TIGER	PAFAH1B1	HPRD	RPLP0	TIGER
ACTA1	HPRD	DDX3X	HPRD	ITGB7	TIGER	PAFAH1B2	TIGER	RPLP1	TIGER
ACTB	HPRD	DDX6	HPRD	JAK1	TIGER	PAFAH1B3	TIGER	RPLP2	UNIPROT
ACTC1	HPRD	DIMT1L	TIGER	JAK2	TIGER	PC	TIGER	RPS11	TIGER
ADRBK2	HPRD	DNAJB9	TIGER	JAK3	-	PCCA	TIGER	RPS14	TIGER
ADSL	HPRD	DOCK3	TIGER	JUN	HPRD	PCCB	TIGER	RPS27L	TIGER
ALDOA	TIGER	DPP4	TIGER	JUP	HPRD	PCNA	HPRD	RPS4X	TIGER
ANXA2	HPRD	ECH1	TIGER	KIAA1949	TIGER	PGAM2	UNIPROT	RPS6	HPRD
ARRB2	TIGER	EDIL3	-	KPNB1	UNIPROT	PIK3CA	TIGER	RPS8	TIGER
ATN1	TIGER	EEF1A1	TIGER	KRT1	TIGER	PLS3	TIGER	RUNX1T1	TIGER
BAG2	TIGER	EEF1B2	TIGER	KRT10	TIGER	PMS1	TIGER	S100A7	HPRD
BAT5	TIGER	EGFR	HPRD	KRT14	HPRD	PMS2	TIGER	S100A8	HPRD
BCAN	TIGER	ELA2	-	KRT15	HPRD	PPIH	TIGER	S100A9	TIGER
BCAR1	HPRD	ELK1	TIGER	KRT16	TIGER	PPP1CA	TIGER	SASS6	-
BCAR3	TIGER	ELK4	TIGER	KRT2	UNIPROT	PPP1CB	TIGER	SDC4	TIGER
BSG	HPRD	ELSPBP1	-	KRT31	-	PPP1R12A	TIGER	SELENBP1	TIGER
C18orf24	TIGER	ERBB4	HPRD	KRT32	-	PPP2CA	HPRD	SERPINA3	TIGER
C19orf21	TIGER	EREG	TIGER	KRT33B	-	PPP2CB	TIGER	SERTAD4	TIGER
C19orf50	TIGER	EZR	HPRD	KRT34	-	PPP2R1A	TIGER	SH2D3C	-
C1QBP	TIGER	FABP4	TIGER	KRT5	HPRD	PPP2R1B	HPRD	SKIL	TIGER
CALM1	TIGER	FAM33A	TIGER	KRT6B	HPRD	PPP2R2B	TIGER	SKIV2L2	TIGER
CALML3	UNIPROT	FANCD2	TIGER	KRT78	-	PPP4C	TIGER	SMAD1	TIGER
CAPZA1	TIGER	FANCG	TIGER	KRT81	UNIPROT	PRDX1	HPRD	SMAD2	TIGER
CAPZA2	TIGER	FANCI	TIGER	KRT85	-	PRDX3	TIGER	SMAD3	TIGER
CASK	HPRD	FBN1	TIGER	KRT9	-	PRDX4	TIGER	SNF1LK2	TIGER
CAV1	TIGER	FBN2	TIGER	LCK	TIGER	PRKCA	HPRD	SOCS1	-
CCL13	-	FCGR2B	TIGER	LGALS7	-	PRKCD	HPRD	SOCS3	TIGER
CCL17	HPRD	FN1	HPRD	LRRC15	TIGER	PRPF3	TIGER	STAT1	HPRD
CCL19	TIGER	FSCN1	HPRD	LSM2	TIGER	PRPF31	UNIPROT	STAT2	HPRD
CCL2	TIGER	FYN	TIGER	LSM4	UNIPROT	PRPF4	TIGER	STAT3	HPRD
CCL22	-	GNA13	TIGER	LSM6	TIGER	PRSS1	-	STAT5B	TIGER
CCL7	-	GNAI1	TIGER	LSM8	TIGER	PSMD11	TIGER	TCF12	HPRD
CCR4	-	GRPEL1	TIGER	LTBP1	TIGER	PSMD4	TIGER	TCF3	TIGER
CCR5	-	HCFC2	TIGER	LY6E	UNIPROT	PTGS2	HPRD	TCF4	TIGER
CCR8	-	HES1	TIGER	LYN	TIGER	PTK2	HPRD	TCP1	TIGER
CCT2	TIGER	HIST1H1E	-	MATR3	UNIPROT	PTK2B	UNIPROT	TFPI	TIGER
CCT3	TIGER	HNRNPL	-	MCCC1	TIGER	PTPN11	TIGER	TGM1	TIGER
CCT4	TIGER	HNRPF	TIGER	MCM5	HPRD	PTPN12	TIGER	TGM2	HPRD
CCT5	TIGER	HNRPH1	TIGER	MEPCE	UNIPROT	PTPN6	TIGER	TIMP1	TIGER
CCT6A	TIGER	HRAS	HPRD	MICAL1	TIGER	PTPRB	HPRD	TMOD3	HPRD
CCT7	TIGER	HSP90AA1	HPRD	MLH1	UNIPROT	PTPRC	HPRD	TMPO	TIGER
CCT8	TIGER	HSPA1A	TIGER	MMP1	-	PXN	HPRD	TNC	HPRD
CD247	HPRD	HSPA5	HPRD	MMP7	TIGER	R3HCC1	TIGER	TNFSF11	GO:0033598 mammary gland epithelial cell
CD4	TIGER	HSPA8	HPRD	MMP9	HPRD	RAB1A	TIGER		
CD44	HPRD	HSPA9	TIGER	MRPL12	TIGER	RAN	TIGER		
CD74	HPRD	HSPD1	HPRD	MRPS27	TIGER	RAPGEF1	TIGER	TP53	HPRD
CDC42	HPRD	ID1	TIGER	MSN	TIGER	RAPGEF2	TIGER	TPM1	HPRD
CDCA4	TIGER	ID2	TIGER	MTMR15	-	RARRES3	TIGER	TPM2	TIGER
CDH1	HPRD	IFI16	TIGER	MYF5	-	RBM25	UNIPROT	TRIP6	TIGER
CHAT	-	IGFBP3	TIGER	MYF6	-	RBM7	TIGER	TUBA1B	-
CKB	TIGER	IKBKG	UNIPROT	MYH2	-	RECQL4	TIGER	TUBB	TIGER
CMA1	TIGER	IL13	-	MYH4	UNIPROT	RFC3	TIGER	TUBB2C	TIGER
CNTN1	TIGER	IL8RA	-	MYH9	TIGER	RFC4	HPRD	TUFM	TIGER
COL2A1	TIGER	IL8RB	-	MYL6	TIGER	RNF135	UNIPROT	TUT1	TIGER
COPS5	TIGER	IQGAP1	TIGER	MYOD1	-	RNF40	TIGER	UBB	TIGER
CORO1C	TIGER	IRS4	-	MYOG	-	RPL14	TIGER	USP15	TIGER
CRK	TIGER	ITCH	TIGER	NCAN	TIGER	RPL17	TIGER	VAV1	-
CRKL	HPRD	ITGA2	TIGER	NCK1	TIGER	RPL19	TIGER	VCAM1	TIGER
CSTF1	TIGER	ITGA4	TIGER	NDE1	TIGER	RPL21	TIGER	VIM	HPRD
CTNNB1	HPRD	ITGA5	HPRD	NDEL1	TIGER	RPL23	TIGER	VSIG8	-
CTSG	-	ITGA8	-	NEDD9	TIGER	RPL24	TIGER	YBX1	TIGER
CXCL1	-	ITGA9	TIGER	NGFR	HPRD	RPL30	UNIPROT	YWHAE	TIGER
CXCL12	TIGER	ITGAD	-	NPNT	TIGER	RPL4	TIGER	YWHAZ	TIGER
CXCR4	TIGER	ITGB1	HPRD	NUCB1	HPRD	RPL6	TIGER	ZYX	TIGER
DARC	TIGER	ITGB5	HPRD	NUFIP2	TIGER	RPL7A	TIGER		

Table A.9. We have tested the evidence of the presence of the genes of brain metastasis sub-network in different databases

Gene	Breast Tissue Source of Evidence	Gene	Breast Tissue Source of Evidence	Gene	Breast Tissue Source of Evidence	Gene	Breast Tissue Source of Evidence	Gene	Breast Tissue Source of Evidence
A2M	HPRD	CRTAP	TIGER	ITCH	TIGER	PAFAH1B2	TIGER	SQSTM1	TIGER
AADACL1	UNIPROT	CSF3	HPRD	ITGA1	TIGER	PAFAH1B3	TIGER	SSR1	TIGER
ACP2	TIGER	CSF3R	-	ITGA2	TIGER	PCNA	HPRD	STCH	TIGER
ACP5	HPRD	CTNNB1	HPRD	ITGA3	TIGER	PCYOX1	TIGER	STOM	TIGER
ACTA1	HPRD	CTSD	HPRD	ITGA5	HPRD	PELI1	TIGER	STT3A	TIGER
ACTC1	HPRD	CUL3	HPRD	ITGAV	HPRD	PH-4	TIGER	STT3B	TIGER
ADNP	UNIPROT	KFZP564O052	TIGER	ITGB1	HPRD	PIGR	TIGER	STX1A	-
APC	TIGER	DNAJB9	TIGER	ITGB3	TIGER	PIGS	TIGER	SUMF1	TIGER
ASAH1	TIGER	DNASE2	TIGER	ITGB5	HPRD	PIGT	TIGER	SUMO2	TIGER
ATF7IP	UNIPROT	DPP7	TIGER	KIAA0090	TIGER	PLOD1	TIGER	TFPI	TIGER
ATN1	TIGER	DSG1	-	KIAA1949	TIGER	PLOD2	TIGER	TFRC	HPRD
ATP1B3	TIGER	DYNLL1	TIGER	KPNB1	TIGER	PLOD3	TIGER	TGM1	TIGER
ATP6AP1	TIGER	ECE1	HPRD	L1CAM	TIGER	PLS3	TIGER	TGM2	HPRD
BAG1	HPRD	EDIL3	-	LAMA4	TIGER	PLXNB2	HPRD	TIMP1	TIGER
BAT5	TIGER	EEF1A1	TIGER	LAMA5	TIGER	PMS1	TIGER	TMOD3	HPRD
BCAN	TIGER	EGFL11	-	LAMB1	HPRD	PMS2	TIGER	TNFRSF10A	TIGER
BRD7	TIGER	EGFR	HPRD	LAMB2	HPRD	PODXL2	TIGER	TNFRSF10B	TIGER
BSG	HPRD	ELA2	-	LAMC1	HPRD	PPP1CA	TIGER	TNFRSF10C	TIGER
BTD	TIGER	ERBB4	HPRD	LAMC3	TIGER	PPP1CB	TIGER	TNFRSF10D	TIGER
BTN2A1	TIGER	ERO1L	TIGER	LAMP1	TIGER	PPP1R12A	TIGER	TNFRSF11B	TIGER
C13orf15	UNIPROT	FADD	TIGER	LAMP2	HPRD	PPT1	TIGER	TNFSF10	HPRD
C19orf21	TIGER	FAM107B	TIGER	LEPRE1	TIGER	PRKCA	HPRD	TNFSF11	(GO:0033598) mammary gland epithelial cell
C1orf103	TIGER	FANCD2	TIGER	LGALS3BP	HPRD	PRKCD	HPRD		
C1orf85	TIGER	FANCG	TIGER	LNPEP	TIGER	PSAP	HPRD		
C20orf3	TIGER	FANCI	TIGER	LTBP1	TIGER	PTGS2	HPRD		
C21orf29	-	FBLN1	HPRD	LTBP3	TIGER	PTK2	HPRD	TP53	HPRD
CAPZA2	TIGER	FBN1	TIGER	LYPLA3	TIGER	PTN	TIGER	TPBG	TIGER
CASP10	HPRD	FBN2	TIGER	M6PR	TIGER	RAB1A	TIGER	TPM1	HPRD
CASP8	HPRD	FKBP10	TIGER	MAN2B1	TIGER	RAPGEF2	TIGER	TPM2	TIGER
CAV1	TIGER	FKBP9	UNIPROT	MARCKS	TIGER	RARRES3	TIGER	TPP1	TIGER
CCL13	-	FN1	HPRD	MED31	TIGER	RFC3	TIGER	TREML2	-
CCL2	TIGER	FOLR1	TIGER	MEF2C	TIGER	RFC4	HPRD	TSG101	HPRD
CCL7	-	FSCN1	HPRD	MLH1	UNIPROT	RIF1	TIGER	TUBB	TIGER
CD109	TIGER	FUT11	-	MMP1	-	RNF135	UNIPROT	TUBB2A	TIGER
CD276	TIGER	GAA	TIGER	MMP7	TIGER	RPA2	TIGER	TUBB4	TIGER
CD36	UNIPROT	GAPDH	TIGER	MTMR15	-	S100A4	HPRD	TXNDC10	TIGER
CD44	HPRD	GBA	TIGER	NCSTN	TIGER	SASS6	-	TXNDC15	-
CD82	TIGER	GLT25D1	TIGER	NDE1	TIGER	SCARB1	TIGER	UBB	TIGER
CD9	HPRD	GNS	TIGER	NDEL1	TIGER	SCARB2	UNIPROT	UBE2I	TIGER
CD97	TIGER	GUSB	TIGER	NEDD4	HPRD	SCNN1A	TIGER	UBR1	TIGER
CDC2	HPRD	HBEGF	TIGER	NEDD4L	TIGER	SCNN1B	TIGER	UNC119	TIGER
CFLAR	TIGER	HECW1	-	NGFR	HPRD	SCNN1G	TIGER	UNC84B	HPRD
CLN5	TIGER	HEXA	TIGER	NID1	TIGER	SEL1L	TIGER	USP10	TIGER
CLPTM1	TIGER	HRAS	HPRD	NID2	TIGER	SEPP1	TIGER	VPS37C	TIGER
CLU	HPRD	HSPG2	TIGER	NOMO1	TIGER	SERPINA1	TIGER	WWP1	TIGER
CMA1	-	HYOU1	TIGER	NPC1	TIGER	SERPINA3	TIGER	WWP2	TIGER
COL13A1	TIGER	IER3	TIGER	NRD1	TIGER	SETD3	HPRD	YWHAE	TIGER
COL2A1	TIGER	IGF2R	TIGER	NUCB1	HPRD	SGSH	TIGER	YWHAZ	TIGER
COL5A1	HPRD	IGFBP3	TIGER	NUP210	TIGER	SIAE	TIGER	ZBTB16	HPRD
COPS5	TIGER	IRAK1	TIGER	OTUD7B	TIGER	SLC3A2	TIGER	ZHX1	TIGER
COPS6	TIGER	IRAK2	TIGER	P4HA1	TIGER	SNX3	TIGER		
CORO1C	TIGER	IRAK4	HPRD	PAFAH1B1	HPRD	SPARC	HPRD		

Table A.10. The frequency of interfaces in both metastasis networks.

Interface Template Name	Frequency in Lung Metastasis Network	Frequency in Brain Metastasis Network
2b8nAB	8	2
2a6aAB	4	0
1nqlAB	2	2

Interface Template Name	Frequency in Lung Metastasis Network	Frequency in Brain Metastasis Network
1xedAC	1	0
1l1yAD	1	0
1on2AB	1	0

1qjcAB	2	2	1e8oCD	1	0
1moxAC	2	2	1a49AB	1	0
1jogCD	5	1	1iieAB	1	0
1gveAB	3	1	1fr3AB	1	0
1oh0AB	1	1	1djrDE	1	0
1bqqMT	1	1	1p65AB	1	0
1zdnAB	2	1	1symAB	1	0
2bo4CD	2	1	1c4zAD	1	0
1tueAH	1	1	1y0eAB	1	0
1xx9CD	1	1	1x8dAB	1	0
1g8tAB	1	1	1p5qAC	1	0
1b3dAB	1	1	1u6iAF	1	0
1jflAB	0	1	1u0kAB	1	0
1okjAB	0	1	1wb1BD	1	0
1qiaCD	1	1	1twjCD	1	0
1eq2GJ	0	1	1xqcAB	1	0
2b99CE	0	1	1tljAB	1	0
1kkmAB	1	1	1f6fBC	1	0
2btfAP	1	1	1p60AB	1	0
1jyaAB	1	1	1pbiAB	1	0
1cd9AB	0	1	1yw0AD	1	0
1jzmAB	1	1	1iawAB	1	0
1nh0AB	2	1	1v8pEF	1	0
1rd5AB	0	1	1qorAB	1	0
1kamAB	0	1	1rkeAB	1	0
1ywkAC	1	1	1o60AB	1	0
1t6uAF	2	0	1jd1AB	1	0
1y9iAD	2	0	1vr0BC	1	0
1pe0AB	1	0	1n1bAB	1	0
1j2rCD	1	0	1fiuAB	1	0
1u2eAC	1	0	3ezeAB	1	0
1wmhAB	1	0	1k2fAB	1	0
1tb3AD	1	0	1um0CD	1	0
1sj1AB	1	0	1vi6AB	1	0
1zuwAC	1	0	1q5cAB	1	0
1xmzAB	1	0	1s96AB	1	0
2erbAB	1	0	1yllAB	1	0
1mzhAB	1	0	1t3uAB	1	0

Table A.11 Most frequently used interfaces while modeling the interactions of lung metastasis network.

Template Interface		2b8nAB	1jogCD	2a6aAB
Proteins of the Complex		Glycerate kinase, putative	Uncharacterized protein HI_0074	Peptidase M22 glycoprotease
PRINT Cluster Size		1	17	1
# of PPIs Modelled		8	5	4
# of Proteins Using This Interface		11	7	7
Origin		Thermotoga Maritima bacteria	Eukaryote and Bacteria (**Table A.24**)	Thermotoga Maritima bacteria
Common Biological Processes		-	oxygen transportation (**Table A.24**)	hydrolase and protease
Common Molecular Functions		enzymatic activities like kinase, oxidoreductase, transferase	-	-
Proteins in the Metastasis Network	**Common Biological Processes**	cell adhesion (**Table A.25**)	cell adhesion, angiogenesis, host-virus interaction, immunity (**Table A.25**)	cell adhesion, cell shape and host-virus interaction (**Table A.25**)
	Common Molecular Functions	enzymatic activities (**Table A.26**).	enzymatic activities (**Table A.26**)	- (**Tables A.26**)

Table A.12. Most frequently used interfaces while modeling the interactions of brain metastasis network.

Template Interface	2b8nAB	1qjcAB	1nqlAB	1moxAC
Proteins of the Complex	Glycerate kinase, putative	coaD	EGFR-EGF	EGFR-TGFA
PRINT Cluster Size	1	7	1	4
# of PPIs Modelled	2	2	2	2
# of Proteins Using This Interface	4	4	3	3
Origin	Thermotoga Maritima Bacteria	E. Coli and Thermatoga Maritime Bacteria	Homo Sapiens	Homo Sapiens
Common Biological Processes	-	Coenzyme A biosynthesis	-	-

Common Molecular Functions		enzymatic activities like kinase, oxidoreductase, transferase	nucleotidyltransferase and transferase	Developmental Protein Kinase Receptor Transferase Tyrosine-protein kinase Growth Factor	Developmental Protein Kinase Receptor Transferase Tyrosine-protein kinase Growth Factor Mitogen
Proteins in the Metastasis Network	**Common Biological Processes**	- (**Tables A.27**)	cell adhesion (**Tables A.27**)	Apoptosis Lactation Transcription Transcription Regulation (**Table A.27**)	Apoptosis Lactation Transcription Transcription Regulation (**Table A.27**)
	Common Molecular Functions	- (**Tables A.28**)	receptor (**Tables A.28**)	Developmental Protein Kinase Receptor Transferase Tyrosine-protein kinase Growth Factor Activator (**Table A.28**)	Developmental Protein Kinase Receptor Transferase Tyrosine-protein kinase Growth Factor Activator (**Table A.28**)

Table A.13. Host-pathogen knowledge on proteins that use pathogenic interface architectures in brain metastasis network.

PROTEIN	RELATIONSHIP	SOURCE
LTBP1	-	-
TNFSF11	-	-
SERPINA3	-	-
BSG	-	-
FBN1	-	-
PLS3	-	-
FSCN1	-	-
FN1	interaction with bacteria	[311]
CCL2	involvement in mycobacterium tuberculosis susceptibility	[312]
MMP1	Host-virus interaction	UNIPROT
ITGB1	Host-virus interaction	UNIPROT
KPNB1	Host-virus interaction	UNIPROT
CD44	HIV1 downregulates	HPIDB
ITGA5	Host-virus interaction	UNIPROT

Table A.14. Host-pathogen knowledge on proteins that use pathogenic interface architectures in lung metastasis network.

PROTEIN	RELATIONSHIP	SOURCE
ABL1	HIV1 downregulates	HPIDB
BCAR1	-	-
BSG	-	-
CCL2	involvement in mycobacterium tuberculosis susceptibility	[312]
CD44	HIV1 downregulates	HPIDB
CDH1	-	-
CNTN1	-	-
CRKL	-	-
CXCL1	-	-
CXCR4	Host-virus interaction	UNIPROT
ELA2 (elane)	associated with Hendra virus	HPIDB
EZR	-	
FN1	interaction with bacteria	[311]
FSCN1	-	
IL13	-	
ITCH	Host-virus interaction	UNIPROT
ITGB1	Host-virus interaction	UNIPROT
ITGB7	Host cell receptor for virus entry	UNIPROT
JAK3	-	
KPNB1	Host-virus interaction	UNIPROT
LTBP1	-	
MICAL1	-	
MMP1	Host-virus interaction	UNIPROT
MMP9	-	
MYH9	-	
NEDD9	-	
PIK3CA	associated with influenza A virus	HPIDB
PLS3	-	
PTK2	-	
PTK2B	-	
PTPN11	-	
PTPN6	-	
PTPRC	defense response to virus	UNIPROT
PXN	-	

SERPINA3	-	
SMAD1	-	
TNC	-	
TNFSF11	-	
VAV1	-	
VCAM1	Host-virus interaction	UNIPROT

Table A.15. Proteins in brain metastasis sub-network with PRISM interface predictions

PROTEIN	BIOLOGICAL FUNCTION	USING PATHOGEN INTERFACE ARCHITECTURE
LTBP1	-	YES
SERPINA3	-	YES
BSG	-	YES
FBN1	-	YES
PLS3	-	YES
FSCN1	-	YES
MMP1	-	YES
KPNB1	-	YES
FN1	cell adhesion	YES
CCL2	cell adhesion	YES
ITGB1	cell adhesion	YES
CD44	cell adhesion	YES
ITGA5	cell adhesion	YES
TNFSF11	positive regulation of homotypic cell-cell adhesion	YES
CSF3R	cell adhesion	NO
HBEGF	-	NO
EGFR	cell-cell adhesion	NO
ELANE	-	NO
ERBB4	-	NO
MMP7	-	NO
TIMP1	-	NO

Table A.16. Proteins in lung metastasis sub-network with PRISM interface predictions

PROTEIN	BIOLOGICAL FUNCTION	USING PATHOGEN INTERFACE ARCHITECTURE

CDH1	cell adhesion	YES
CNTN1	cell adhesion	YES
TNC	cell adhesion	YES
BCAR1	cell adhesion	YES
NEDD9	cell adhesion	YES
PTK2B	cell adhesion	YES
PXN	cell adhesion	YES
ABL1	cell adhesion	YES
FN1	cell adhesion	YES
CCL2	cell adhesion	YES
ITGB7	cell adhesion	YES
ITGB1	cell adhesion	YES
CD44	cell adhesion	YES
ITGA5	cell adhesion	YES
VCAM1	cell adhesion	YES
MYH9	cell-cell adhesion	YES
EZR	leukocyte cell-cell adhesion	YES
PTPRC	negative regulation of cell adhesion involved in substrate-bound cell migration	YES
PTK2	negative regulation of cell-cell adhesion, positive regulation of cell adhesion	YES
TNFSF11	positive regulation of homotypic cell-cell adhesion	YES
PTPN11	regulation of cell adhesion mediated by integrin	YES
BSG	-	YES
CRKL	-	YES
CXCL1	-	YES
IL13	-	YES
JAK3	-	YES
MMP9	-	YES
SMAD1	-	YES
FSCN1	-	YES
LTBP1	-	YES
MICAL1	-	YES
PLS3	-	YES
PTPN6	-	YES
SERPINA3	-	YES
VAV1	-	YES
ELA2 (elane)	-	YES
MMP1	-	YES

KPNB1	-	YES
CXCR4	-	YES
ITCH	-	YES
PIK3CA	-	YES
CD74	-	NO
CHAT	-	NO
CRK	cell adhesion	NO
EGFR	cell-cell adhesion	NO
ERBB4	-	NO
EREG	-	NO
SMAD3	-	NO
MMP7	-	NO
MSN	leukocyte cell-cell adhesion	NO
STAT1	-	NO
TCF3	-	NO
TIMP1	-	NO
BCAR3	-	NO
DIMT1L	-	NO

Table A.17. Distribution of the residue numbers and the mutation numbers per protein.

		MUTATIONS				RESIDUE NUMBER			
PROTEIN	PDB_ID; CHAIN ID	SURFACE		CORE	TOTAL	SURFACE		CORE	TOTAL
		INTERFACE	NON-INTERFACE			INTERFACE	NON-INTERFACE		
ELANE	2Z7F_E	4	9	19	32	18	109	91	218
CSF3	1CD9_A	0	2	0	2	8	105	58	171
VCAM1	1VCA_A	0	16	2	18	7	136	56	199
EGFR	3NJP_B	4	57	26	87	36	365	214	615
HBEGF	1XDT_R	0	1	1	2	26	11	4	41
EREG	1K36_A	1	1	0	2	25	14	7	46
ITGA5	3VI3_A	5	16	11	32	46	299	256	601
CD44	2I83_A, 1UUH_A	1	1	3	5	39	58	51	148
TNC	2RB8_A	2	5	2	9	22	45	26	93
FBN1	2W86_A	4	6	2	12	26	90	38	154
ERBB4	3U7U_B	5	69	48	122	34	321	195	550
MMP1	1SU3_B	0	25	20	45	23	219	173	415
FN1	1FNF_A	2	15	1	18	16	246	106	368
SUM		28	223	135	386	326	2018	1275	3619

Table A.18. The total residues numbers/genetic variations observed in different locations and the odds ratio, 95% confidence interval, and the P-value for a two tailed test that OR is different from 1.0.

	# of Residues	Genetic Variations		OR	95 percent CI	P value
Core	1275	135	Core vs. surface	0.99	0.78 - 1.24	0.95

Surface	2344	251	Interface vs surface noninterface	0.76	0480 - 1.15	0.21
Interface	326	28	Core vs. interface	1.26	0.81 - 2.01	0.3
Surface noninterface	2018	223				
Total	3619	386				

Table A.19. Interface residues (Sequence IDs) of HBEGF-EGFR model, the binding site residues of HBEGF protein's complexes available in PDB and the binding site residues of EGFR protein's complexes available in PDB. The interface residues that are overlaping with available binding site residues are in italic, bold fonts.

	IN PDB	**PRISM MODEL**		**IN PDB**			
PDB ID; Chain ID	**1XDT ; R**	**-**		**1NQL; A**	**1MOX; A**	**1IVO; A**	**3NJP; A**
PROTEIN NAME	**HBEGF**	**HBEGF**	**EGFR**	**EGFR**	**EGFR**	**EGFR**	**EGFR**
RESIDUES	112	111	***36***	36	36	36	36
	115	***112***	***37***	37	37	37	37
	122	113	***38***	38	38	38	38
	124	114	***39***	39	39	39	39
	126	***115***	***40***	40	40	40	40
	127	117	***41***	41	41	41	41
	129	***124***	***42***	42	42	42	42
	130	***126***	***69***	46	69	46	46
	131	***127***	***92***	47	93	50	53
	132	***129***	***93***	93	122	69	69
	133	***131***	***122***	114	123	92	93
	134	***132***	***123***	122	125	93	114
	135	133	***125***		126	114	122
	136	***134***	***349***		149	122	123
	137	***139***	***377***		152	123	125
	138	***140***	***379***		349	125	349
	139	***141***	***380***		370	349	370
	140	142	***381***		372	370	372
	141	***143***	***408***		374	372	373
	143	144	***433***		378	372	374
	147	145			379	374	377
					380	375	379

							381	381	380
							406	382	381
							408	406	382
							432	408	406
							433	432	408
							435	433	432
							436	435	433
							439	436	436
							462	439	439
								441	441
								462	462
								464	464
								489	491
									492

Table A.20. Interface residues (Sequence IDs) of EREG-EGFR model and the binding site residues of EGFR protein's complexes available in PDB. The interface residues that are overlaping with available binding site residues are in italic, bold fonts.

	PRISM MODEL		**IN PDB**			
PDB ID; Chain ID	**-**		**1NQL; A**	**1MOX; A**	**1IVO; A**	**3NJP; A**
PROTEIN NAME	**EREG**	**EGFR**	**EGFR**	**EGFR**	**EGFR**	**EGFR**
RESIDUES	71	34	36	36	36	36
	72	35	37	37	37	37
	73	***36***	38	38	38	38
	74	***38***	39	39	39	39
	75	***39***	40	40	40	40
	77	***40***	41	41	41	41
	84	***41***	42	42	42	42
	86	***42***	46	69	46	46
	87	43	47	93	50	53
	89	***46***	93	122	69	69
	91	***69***	114	123	92	93
	92	***93***	122	125	93	114
	93	***122***		126	114	122
	94	***123***		149	122	123

	96	C-Terminus	***125***			152	123	125
	99		***372***			349	125	349
	101		***373***			370	349	370
	104		***374***			372	370	372
	105		***377***			374	372	373
	106		***379***			378	372	374
	107		***380***			379	374	377
	108		***381***			380	375	379
			382			381	381	380
			383			406	382	381
			406			408	406	382
			408			432	408	406
			432			433	432	408
			433			435	433	432
			439			436	435	433
			441			439	436	436
			442			462	439	439
			462				441	441
			489				462	462
							464	464
							489	491
								492

Table A.21. Interface residues (Sequence IDs) of HBEGF-ERBB4 model, the binding site residues of HBEGF protein's complexes available in PDB and the binding site residues of ERBB4 protein's complexes available in PDB. The interface residues that are overlapping with available binding site residues are in italic, bold fonts.

	IN PDB		**PRISM MODEL**		**IN PDB**
PDB ID; Chain ID	**1XDT; R**		**-**		**3U7U; A**
PROTEIN NAME	**HBEGF**		**HBEGF**	**ERBB4**	**ERBB4**
RESIDUES	112		111	33	34
	115		***112***	***34***	35
	122		113	***35***	36
	124		114	36	37
	126		***115***	***37***	38

	127		117	***38***	39
	129		118	***39***	40
	130		***124***	***40***	41
	131		***126***	52	42
	132		***127***	***91***	44
	133		***131***	***120***	51
	134		***132***	121	91
	135		***133***	***125***	111
	136		***134***	***352***	112
	137		***135***	***375***	120
	138		***138***	***376***	121
	139		***139***	***377***	123
	140		***140***	***382***	125
	141		***141***	***383***	148
	143		142	***384***	346
	147		144	***411***	369
			145	***435***	370
			146	***443***	371
			147	***444***	382
					383
					384
					385
					405
					429
					432
					435
					437
					438
					459

Table A.22. Interface residues (Sequence IDs) of EREG-ERBB4 model and the binding site residues of ERBB4 protein's complexes available in PDB. The interface residues that are overlapping with available binding site residues are in italic, bold fonts.

	PRISM MODEL			**IN PDB**
PDB ID; Chain ID	**-**			**3U7U; A**
PROTEIN NAME	**EREG**	**ERBB4**		**ERBB4**
RESIDUES	71	32		34
	72	33		35

	73		***34***		36
	74		***36***		37
	75		***37***		38
	77		***38***		39
	78		***39***		40
	84		***40***		41
	86		***41***		42
	87		***44***		44
	89		52		51
	91		90		91
	93		***91***		111
	94		***120***		112
	96	**C-terminus**	***121***		120
	99		373		121
	100		***375***		123
	101		***376***		125
	102		***377***		148
	104		***383***		346
	105		409		369
	106		***411***		370
	107		***435***		371
	108		***443***		382
			444		383
			465		384
			467		385
			468		405
			494		429
					432
					435
					437
					438
					459

Table A.23. The table for the source organism distribution of template chains, used for modeling the complexes of BMSN and LMSN.

	Lung Met. Net. Template Chains	Brain Met. Net. Template Chains	All Template Chains in the Dataset
Eukaryota	60	22	5822

Archaea	12	4	515
Viruses	6	4	716
Bacteria	72	26	4202
Microbial (Viruses + Bacteria)	78	30	4918
Total Number of Template Chains	150	56	11255

Table A.24. The interfaces in the 1jogCD PRINT cluster.

PDB chains	Protein Name	Organism	Molecular Function	Biological Process
1jog_CD/AB	uncharacterized protein	bacteria	-	-
1i4y_AH/FG/BC/DE: 1i4z_CF/AD/BG/EH : 2hmqCD : 2hmzCD : 1hmdCD : 1hmoCD	Hemerythrin	eukaryota	-	Oxygen transport, Transport
1wwpAB	uncharacterized protein	bacteria	-	-
1wty_BC/AD	uncharacterized protein	bacteria	-	-

Metastasis	Template Interface	Proteins in the Network	cell adhesion	host-virus interaction	angiogenesis	notch signalling pathway	cell shape	cell cycle	Antiviral defense	Apoptosis	Immunity	Innate immunity	Adaptive immunity	Ubl conjugation pathway	acute phase	cel division	growth regulation	mitosis	collagen degredation	protein phosphatase	hydrolase
LUNG	2b8nAB	ITGB1	+	+																	
	2b8nAB, 2a6aAB	TNC	+																		
	2b8nAB	CNTN1	+			+															
	2b8nAB, 2a6aAB	FN1	+		+		+							+							
	2b8nAB	LTBP1	Biological Process Not Listed																		
	2b8nAB, 2a6aAB	MICAL1	Biological Process Not Listed																		
	2b8nAB,1jogCD	PTK2			+																
	2b8nAB, 2a6aAB, 1jogCD	NEDD9	+					+								+	+	+			
	2b8nAB	PTPN11																		+	+
	2b8nAB	BSG	Biological Process Not Listed																		
	2b8nAB,1jogCD	MMP1		+															+		
	2a6aAB	MYH9					+														
	2a6aAB	CXCR4		+																	
	2a6aAB, 1jogCD	ITCH		+					+	+	+	+		+							
	1jogCD	CD44	+																		
	1jogCD	PTK2B			+						+		+								
	1jogCD	BCAR1	+																		
			BIOLOGICAL FUNCTIONS																		

Table A.25. The biological processes of the proteins utilizing the most frequent interfaces of LMSN

Metastasis	Template Interface	Proteins in the Network	Kinase	Oxidoreductase	Transferase	Protease	Metalloprotease	Hydrolase	Monooxygenase	Motor protein	Myosin	Protein phosphatase	Tyrosine-protein kinase	Ligase	Developmental protein	Blood group antigen	Receptor	Transducer	G-protein coupled receptor	Integrin	Host cell receptor for virus entry
LUNG	2b8nAB	ITGB1														+			+	+
	2b8nAB, 2a6aAB	TNC	Molecular Functions Not Listed																		
	2b8nAB	CNTN1	Molecular Functions Not Listed																		
	2b8nAB, 2a6aAB	FN1	Molecular Functions Not Listed																		
	2b8nAB	LTBP1	Molecular Functions Not Listed																		
	2b8nAB, 2a6aAB	MICAL1		+					+												
	2b8nAB,1jogCD	PTK2	+		+								+		+						
	2b8nAB, 2a6aAB, 1jogCD	NEDD9	Molecular Functions Not Listed																		
	2b8nAB	PTPN11						+				+									
	2b8nAB	BSG														+					
	2b8nAB,1jogCD	MMP1				+	+	+													
	2a6aAB	MYH9								+	+										
	2a6aAB	CXCR4															+	+	+		+
	2a6aAB, 1jogCD	ITCH	Molecular Functions Not Listed																		
	1jogCD	CD44														+	+				
	1jogCD	PTK2B	+		+								+								
	1jogCD	BCAR1	Molecular Functions Not Listed																		
			MOLECULAR FUNCTIONS																		

Table A.26. The molecular functions of the proteins utilizing the most frequent interfaces of LMSN

Table A.27. The biological processes of the proteins utilizing the most frequent interfaces of brain metastasis network.

Metastasis	Template Interface	Proteins in the Network	Kinase	Transferase	Protease	Metalloprotease	Hydrolase	Tyrosine-protein kinase	Developmental protein	Activator	Growth Factor	Blood group antigen	Receptor	Integrin
BRAIN	2b8nAB, 1qjcAB	FN1	Molecular Functions Not Listed											
	2b8nAB, 1qjcAB	LTBP1	Molecular Functions Not Listed											
	2b8nAB	BSG									+			
	2b8nAB	MMP1			+	+	+							
	1qjcAB	CD44									+	+		
	1qjcAB	ITGA5										+	+	
	1nqlAB, 1moxAC	ERBB4	+	+				+	+	+			+	
	1nqlAB, 1moxAC	HBEGF								+			+	
	1nqlAB, 1moxAC	EGFR	+	+				+	+				+	
			MOLECULAR FUNCTIONS											

Table A.28. The molecular functions of the proteins utilizing the most frequent interfaces of brain metastasis

Metastasis	Template Interface	Proteins in the Network	cell adhesion	host-virus interaction	angiogenesis	cell shape	Lactation	Transcription	Transcription regulation	Apoptosis	acute phase	collagen degredation
BRAIN	2b8nAB, 1qjcAB	FN1	+		+	+					+	
	2b8nAB, 1qjcAB	LTBP1	Biological Process Not Listed									
	2b8nAB	BSG	Biological Process Not Listed									
	2b8nAB	MMP1		+								+
	1qjcAB	CD44	+									
	1qjcAB	ITGA5	+	+								
	1nqlAB, 1moxAC	ERBB4					+	+	+	+		
	1nqlAB, 1moxAC	HBEGF	Biological Process Not Listed									
	1nqlAB, 1moxAC	EGFR	Biological Process Not Listed									
			BIOLOGICAL PROCESSES									

BIBLIOGRAPHY

1. **Faratian, D., et al.,** ***Systems pathology--taking molecular pathology into a new dimension.*** **Nature reviews. Clinical oncology, 2009. 6(8): p. 455-64.**
2. **Kim, P.M., et al.,** ***Relating three-dimensional structures to protein networks provides evolutionary insights.*** **Science, 2006. 314(5807): p. 1938-41.**
3. **Wang, X., et al.,** ***Three-dimensional reconstruction of protein networks provides insight into human genetic disease.*** **Nature biotechnology, 2012. 30(2): p. 159-64.**
4. **Stein, A., et al.,** ***Dynamic interactions of proteins in complex networks: a more structured view.*** **Febs J, 2009a. 276(19): p. 5390-405.**
5. **Bhattacharyya, R.P., et al.,** ***Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits.*** **Annu Rev Biochem, 2006. 75: p. 655-80.**
6. **Bashor, C.J., et al.,** ***Rewiring cells: synthetic biology as a tool to interrogate the organizational principles of living systems.*** **Annu Rev Biophys, 2010. 39: p. 515-37.**
7. **Bashor, C.J., et al.,** ***Rewiring cells: synthetic biology as a tool to interrogate the organizational principles of living systems.*** **Annual review of biophysics, 2010. 39: p. 515-37.**
8. **Nooren, I.M. and J.M. Thornton,** ***Diversity of protein-protein interactions.*** **Embo J, 2003a. 22(14): p. 3486-92.**
9. **Park, S.H., et al.,** ***Prediction of protein-protein interaction types using association rule based classification.*** **BMC Bioinformatics, 2009. 10: p. 36.**
10. **Braig, K., et al.,** ***The crystal structure of the bacterial chaperonin GroEL at 2.8 A.*** **Nature, 1994. 371(6498): p. 578-86.**
11. **Lowe, J., et al.,** ***Refined structure of alpha beta-tubulin at 3.5 A resolution.*** **J Mol Biol, 2001. 313(5): p. 1045-57.**
12. **Krishna, S.S. and L. Aravind,** ***The bridge-region of the Ku superfamily is an atypical zinc ribbon domain.*** **J Struct Biol, 2010. 172(3): p. 294-9.**
13. **Vetter, I.R. and A. Wittinghofer,** ***The guanine nucleotide-binding switch in three dimensions.*** **Science, 2001. 294(5545): p. 1299-304.**
14. **Nooren, I.M. and J.M. Thornton,** ***Structural characterisation and functional significance of transient protein-protein interactions.*** **J Mol Biol, 2003b. 325(5): p. 991-1018.**
15. **Mintseris, J. and Z. Weng,** ***Atomic contact vectors in protein-protein recognition.*** **Proteins, 2003. 53(3): p. 629-39.**
16. **Block, P., et al.,** ***Physicochemical descriptors to discriminate protein-protein interactions in permanent and transient complexes selected by means of machine learning algorithms.*** **Proteins, 2006. 65(3): p. 607-22.**
17. **Janin, J., R.P. Bahadur, and P. Chakrabarti,** ***Protein-protein interaction and quaternary structure.*** **Q Rev Biophys, 2008. 41(2): p. 133-80.**
18. **Levy, E.D. and J.B. Pereira-Leal,** ***Evolution and dynamics of protein interactions and networks.*** **Curr Opin Struct Biol, 2008. 18(3): p. 349-57.**
19. **Hyams, J.S. and C.W. Lloyd,** ***Microtubules, Modern Cell Biology.*** **Modern Cell Biology. Vol. 13. 1993, New York: Wiley-Liss.**
20. **Ozbabacan, S.E., et al.,** ***Transient protein-protein interactions.*** **Protein engineering, design & selection : PEDS, 2011. 24(9): p. 635-48.**

21. **Keskin, O., et al., *Principles of protein-protein interactions: what are the preferred ways for proteins to interact?* Chemical reviews, 2008. 108(4): p. 1225-44.**
22. **Keskin, O., et al., *A new, structurally nonredundant, diverse data set of protein-protein interfaces and its implications.* Protein science : a publication of the Protein Society, 2004. 13(4): p. 1043-55.**
23. **Gong, S., et al., *PSIbase: a database of Protein Structural Interactome map (PSIMAP).* Bioinformatics, 2005. 21(10): p. 2541-3.**
24. **Jones, S. and J.M. Thornton, *Principles of protein-protein interactions.* Proceedings of the National Academy of Sciences of the United States of America, 1996. 93(1): p. 13-20.**
25. **Jones, S., A. Marin, and J.M. Thornton, *Protein domain interfaces: characterization and comparison with oligomeric protein interfaces.* Protein engineering, 2000. 13(2): p. 77-82.**
26. **Cusick, M.E., et al., *Interactome: gateway into systems biology.* Human molecular genetics, 2005. 14 Spec No. 2: p. R171-81.**
27. **Yang, J.S., et al., *SAPIN: a framework for the structural analysis of protein interaction networks.* Bioinformatics, 2012. 28(22): p. 2998-9.**
28. **Mosca, R., A. Ceol, and P. Aloy, *Interactome3D: adding structural details to protein networks.* Nature methods, 2013. 10(1): p. 47-53.**
29. **Kar, G., A. Gursoy, and O. Keskin, *Human cancer protein-protein interaction network: a structural perspective.* PLoS computational biology, 2009. 5(12): p. e1000601.**
30. **Agoston, V., P. Csermely, and S. Pongor, *Multiple weak hits confuse complex systems: a transcriptional regulatory network as an example.* Physical review. E, Statistical, nonlinear, and soft matter physics, 2005. 71(5 Pt 1): p. 051909.**
31. **P. Erdös, A.R., *On Random Graphs.* Publicationes Mathematicae, 1959. 6: p. 290-297.**
32. **Barabasi, A.L. and R. Albert, *Emergence of scaling in random networks.* Science, 1999. 286(5439): p. 509-12.**
33. **Albert, R., H. Jeong, and A.L. Barabasi, *Error and attack tolerance of complex networks.* Nature, 2000. 406(6794): p. 378-82.**
34. **Holme, P., et al., *Attack vulnerability of complex networks.* Phys Rev E Stat Nonlin Soft Matter Phys, 2002. 65(5 Pt 2): p. 056109.**
35. **Jeong, H., et al., *Lethality and centrality in protein networks.* Nature, 2001. 411(6833): p. 41-2.**
36. **Dartnell, L., et al., *Robustness of the p53 network and biological hackers.* FEBS Lett, 2005. 579(14): p. 3037-42.**
37. **Crucitti P, L.V., Marchiori M, Rapisard A, *Efficiency of Scale-free Networks: Error and Attack Tolerance.* Physica A, 2003(320): p. 622-642.**
38. **ZHANG DM, YIN YP, TAN J, PAN GJ, HE MH, *Multiple Partial Attacks on Complex Networks.* Chinese Physical Society, 2008.**
39. **Engin, H.B., et al., *A strategy based on protein-protein interface motifs may help in identifying drug off-targets.* Journal of chemical information and modeling, 2012. 52(8): p. 2273-86.**
40. **Hopkins, A.L., *Network pharmacology: the next paradigm in drug discovery.* Nature chemical biology, 2008. 4(11): p. 682-90.**

41. **Boran, A.D. and R. Iyengar, *Systems approaches to polypharmacology and drug discovery*. Current opinion in drug discovery & development, 2010. 13(3): p. 297-309.**
42. **Brown, J.B. and Y. Okuno, *Systems biology and systems chemistry: new directions for drug discovery*. Chemistry & biology, 2012. 19(1): p. 23-8.**
43. **Xie, L., S.L. Kinnings, and P.E. Bourne, *Novel computational approaches to polypharmacology as a means to define responses to individual drugs*. Annual review of pharmacology and toxicology, 2012. 52: p. 361-79.**
44. **Arkin, M.R. and J.A. Wells, *Small-molecule inhibitors of protein-protein interactions: progressing towards the dream*. Nature reviews. Drug discovery, 2004. 3(4): p. 301-17.**
45. **Fuller, J.C., N.J. Burgoyne, and R.M. Jackson, *Predicting druggable binding sites at the protein-protein interface*. Drug discovery today, 2009. 14(3-4): p. 155-61.**
46. **Wanner, J., et al., *Druggability assessment of protein-protein interfaces*. Future medicinal chemistry, 2011. 3(16): p. 2021-38.**
47. **Clackson, T. and J.A. Wells, *A hot spot of binding energy in a hormone-receptor interface*. Science, 1995. 267(5196): p. 383-6.**
48. **Acuner Ozbabacan, S.E., et al., *Conformational ensembles, signal transduction and residue hot spots: application to drug discovery*. Current opinion in drug discovery & development, 2010. 13(5): p. 527-37.**
49. **Wells, J.A. and C.L. McClendon, *Reaching for high-hanging fruit in drug discovery at protein-protein interfaces*. Nature, 2007. 450(7172): p. 1001-9.**
50. **Thangudu, R.R., et al., *Modulating protein-protein interactions with small molecules: the importance of binding hotspots*. Journal of molecular biology, 2012. 415(2): p. 443-53.**
51. **Tuncbag, N., O. Keskin, and A. Gursoy, *HotPoint: hot spot prediction server for protein interfaces*. Nucleic acids research, 2010. 38(Web Server issue): p. W402-6.**
52. **Lise, S., et al., *Prediction of hot spot residues at protein-protein interfaces by combining machine learning and energy-based methods*. BMC bioinformatics, 2009. 10: p. 365.**
53. **Darnell, S.J., L. LeGault, and J.C. Mitchell, *KFC Server: interactive forecasting of protein interaction hot spots*. Nucleic acids research, 2008. 36(Web Server issue): p. W265-9.**
54. **Vselovsky, A.V.A., A. I. , *Inhibitors of Protein-Protein Interactions as Potential Drugs*. Current Computer-Aided Drug Design, 2007(3): p. 51-58.**
55. **Braisted, A.C., et al., *Discovery of a potent small molecule IL-2 inhibitor through fragment assembly*. Journal of the American Chemical Society, 2003. 125(13): p. 3714-5.**
56. **Vu, B.T. and L. Vassilev, *Small-molecule inhibitors of the p53-MDM2 interaction*. Current topics in microbiology and immunology, 2011. 348: p. 151-72.**
57. **Canner, J.A., et al., *MI-63: a novel small-molecule inhibitor targets MDM2 and induces apoptosis in embryonal and alveolar rhabdomyosarcoma cells with wild-type p53*. British journal of cancer, 2009. 101(5): p. 774-81.**
58. **Tse, C., et al., *ABT-263: a potent and orally bioavailable Bcl-2 family inhibitor*. Cancer research, 2008. 68(9): p. 3421-8.**

59. Straub, C.S., ***Targeting IAPs as an approach to anti-cancer therapy.*** **Current topics in medicinal chemistry, 2011. 11(3): p. 291-316.**

60. Blazer, L.L. and R.R. Neubig, ***Small molecule protein-protein interaction inhibitors as CNS therapeutic agents: current progress and future hurdles.*** **Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology, 2009. 34(1): p. 126-41.**

61. Silviya D. Furdas, L.C., Wolfgang Sippl and Manfred Jung, ***Inhibition of bromodomain-mediated protein–protein interactions as a novel therapeutic strategy.*** **Med. Chem. Commun., 2012(3): p. 123 - 134.**

62. Busschots, K., et al., ***In search of small molecules blocking interactions between HIV proteins and intracellular cofactors.*** **Molecular bioSystems, 2009. 5(1): p. 21-31.**

63. Buchwald, P., ***Small-molecule protein-protein interaction inhibitors: therapeutic potential in light of molecular size, chemical space, and ligand binding efficiency considerations.*** **IUBMB life, 2010. 62(10): p. 724-31.**

64. Sperandio, O., et al., ***Rationalizing the chemical space of protein-protein interaction inhibitors.*** **Drug discovery today, 2010. 15(5-6): p. 220-9.**

65. Arkin, M.R. and A. Whitty, ***The road less traveled: modulating signal transduction enzymes by inhibiting their protein-protein interactions.*** **Current opinion in chemical biology, 2009. 13(3): p. 284-90.**

66. Arkin, M.R., et al., ***Binding of small molecules to an adaptive protein-protein interface.*** **Proceedings of the National Academy of Sciences of the United States of America, 2003. 100(4): p. 1603-8.**

67. Gao, M. and J. Skolnick, ***Structural space of protein-protein interfaces is degenerate, close to complete, and highly connected.*** **Proc Natl Acad Sci U S A, 2010. 107(52): p. 22517-22.**

68. Eyrisch, S. and V. Helms, ***Transient pockets on protein surfaces involved in protein-protein interaction.*** **Journal of medicinal chemistry, 2007. 50(15): p. 3457-64.**

69. Eyrisch, S. and V. Helms, ***What induces pocket openings on protein surface patches involved in protein-protein interactions?*** **Journal of computer-aided molecular design, 2009. 23(2): p. 73-86.**

70. Metz, A., et al., ***Hot spots and transient pockets: predicting the determinants of small-molecule binding to a protein-protein interface.*** **Journal of chemical information and modeling, 2012. 52(1): p. 120-33.**

71. Li, X., et al., ***Protein-protein interactions: hot spots and structurally conserved residues often locate in complemented pockets that pre-organized in the unbound states: implications for docking.*** **Journal of molecular biology, 2004. 344(3): p. 781-95.**

72. Bourgeas, R., et al., ***Atomic analysis of protein-protein interfaces with known inhibitors: the 2P2I database.*** **PloS one, 2010. 5(3): p. e9598.**

73. Higueruelo, A.P., et al., ***Atomic interactions and profile of small molecules disrupting protein-protein interfaces: the TIMBAL database.*** **Chemical biology & drug design, 2009. 74(5): p. 457-67.**

74. Davis, F.P. and A. Sali, ***The overlap of small molecule and protein binding sites within families of protein structures.*** **PLoS computational biology, 2010. 6(2): p. e1000668.**

75. **Scheiber, J., et al., *Gaining Insight into Off-Target Mediated Effects of Drug Candidates with a Comprehensive Systems Chemical Biology Analysis*. J Chem Inf Model, 2009.**

76. **Scheiber, J., et al., *Mapping adverse drug reactions in chemical space*. J Med Chem, 2009. 52(9): p. 3103-7.**

77. **Yildirim, M.A., et al., *Drug-target network*. Nature biotechnology, 2007. 25(10): p. 1119-26.**

78. **Mestres, J., et al., *Data completeness--the Achilles heel of drug-target networks*. Nature biotechnology, 2008. 26(9): p. 983-4.**

79. **Paolini, G.V., et al., *Global mapping of pharmacological space*. Nature biotechnology, 2006. 24(7): p. 805-15.**

80. **Hopkins, A.L., J.S. Mason, and J.P. Overington, *Can we rationally design promiscuous drugs?* Current opinion in structural biology, 2006. 16(1): p. 127-36.**

81. **Hopkins, A.L., *Network pharmacology*. Nature biotechnology, 2007. 25(10): p. 1110-1.**

82. **Csermely, P., V. Agoston, and S. Pongor, *The efficiency of multi-target drugs: the network approach might help drug design*. Trends in pharmacological sciences, 2005. 26(4): p. 178-82.**

83. **Lange, R.P., et al., *The targets of currently used antibacterial agents: lessons for drug discovery*. Current pharmaceutical design, 2007. 13(30): p. 3140-54.**

84. **Knox, C., et al., *DrugBank 3.0: a comprehensive resource for 'omics' research on drugs*. Nucleic acids research, 2011. 39(Database issue): p. D1035-41.**

85. **Zhu, F., et al., *Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery*. Nucleic acids research, 2012. 40(Database issue): p. D1128-36.**

86. **Kuhn, M., et al., *STITCH 2: an interaction network database for small molecules and proteins*. Nucleic acids research, 2010. 38(Database issue): p. D552-6.**

87. **BL Roth, W.K., S Patel and E Lopez, *The Multiplicity of Serotonin Receptors: Uselessly diverse molecules or an embarrasment of riches?* . The Neuroscientist, 2000. 6: p. 252-262.**

88. **Gao, Z., et al., *PDTD: a web-accessible protein database for drug target identification*. BMC bioinformatics, 2008. 9: p. 104.**

89. **T.I. Oprea, P.B., G. Berellini, M. Olah, K. Fejgin, S. Boyer, *Rapid ADME Filters for Lead Discovery*, in *Molecular Interaction Fields*2006 Wiley-VCH: New York. p. 249-272.**

90. **Liu, T., et al., *BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities*. Nucleic acids research, 2007. 35(Database issue): p. D198-201.**

91. **Wang, R., et al., *The PDBbind database: methodologies and updates*. Journal of medicinal chemistry, 2005. 48(12): p. 4111-9.**

92. **Kanehisa, M., et al., *From genomics to chemical genomics: new developments in KEGG*. Nucleic acids research, 2006. 34(Database issue): p. D354-7.**

93. **Bolton E, W.Y., Thiessen PA, Bryant SH, *PubChem: Integrated Platform of Small Molecules and Biological Activities*, in *Annual Reports in Computational Chemistry*2008, American Chemical Society: Washington, DC.**

94. **McDonagh, E.M., et al.,** ***From pharmacogenomic knowledge acquisition to clinical applications: the PharmGKB as a clinical pharmacogenomic biomarker resource.*** **Biomarkers in medicine, 2011. 5(6): p. 795-806.**

95. **Ji, Z.L., et al.,** ***Drug Adverse Reaction Target Database (DART) : proteins related to adverse drug reactions.*** **Drug safety : an international journal of medical toxicology and drug experience, 2003. 26(10): p. 685-90.**

96. **Hecker, N., et al.,** ***SuperTarget goes quantitative: update on drug-target interactions.*** **Nucleic acids research, 2012. 40(Database issue): p. D1113-7.**

97. **von Eichborn, J., et al.,** ***PROMISCUOUS: a database for network-based drug-repositioning.*** **Nucleic acids research, 2011. 39(Database issue): p. D1060-6.**

98. **Davis, A.P., et al.,** ***The Comparative Toxicogenomics Database: update 2011.*** **Nucleic acids research, 2011. 39(Database issue): p. D1067-72.**

99. **Meslamani, J., D. Rognan, and E. Kellenberger,** ***sc-PDB: a database for identifying variations and multiplicity of 'druggable' binding sites in proteins.*** **Bioinformatics, 2011. 27(9): p. 1324-6.**

100. **Knight, Z.A., H. Lin, and K.M. Shokat,** ***Targeting the cancer kinome through polypharmacology.*** **Nature reviews. Cancer, 2010. 10(2): p. 130-7.**

101. **Fliri, A.F., W.T. Loging, and R.A. Volkmann,** ***Cause-effect relationships in medicine: a protein network perspective.*** **Trends in pharmacological sciences, 2010. 31(11): p. 547-55.**

102. **Dixon, S.J. and B.R. Stockwell,** ***Drug discovery: engineering drug combinations.*** **Nature chemical biology, 2010. 6(5): p. 318-9.**

103. **Humeniuk, R., et al.,** ***Aplidin synergizes with cytosine arabinoside: functional relevance of mitochondria in Aplidin-induced cytotoxicity.*** **Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, U.K, 2007. 21(12): p. 2399-405.**

104. **Pelicano, H., et al.,** ***Targeting Hsp90 by 17-AAG in leukemia cells: mechanisms for synergistic and antagonistic drug combinations with arsenic trioxide and Ara-C.*** **Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, U.K, 2006. 20(4): p. 610-9.**

105. **Cuadrado, A., et al.,** ***JNK activation is critical for Aplidin-induced apoptosis.*** **Oncogene, 2004. 23(27): p. 4673-80.**

106. **Biscardi, M., et al.,** ***VEGF inhibition and cytotoxic effect of aplidin in leukemia cell lines and cells from acute myeloid leukemia.*** **Annals of oncology : official journal of the European Society for Medical Oncology / ESMO, 2005. 16(10): p. 1667-74.**

107. **Gajate, C. and F. Mollinedo,** ***Cytoskeleton-mediated death receptor and ligand concentration in lipid rafts forms apoptosis-promoting clusters in cancer chemotherapy.*** **The Journal of biological chemistry, 2005. 280(12): p. 11641-7.**

108. **Mansouri, A., et al.,** ***Sustained activation of JNK/p38 MAPK pathways in response to cisplatin leads to Fas ligand induction and cell death in ovarian carcinoma cells.*** **The Journal of biological chemistry, 2003. 278(21): p. 19245-56.**

109. **Jia, J., et al.,** ***Mechanisms of drug combinations: interaction and network perspectives.*** **Nature reviews. Drug discovery, 2009. 8(2): p. 111-28.**

110. **de Vries, J.F., et al.,** ***The mechanisms of Ara-C-induced apoptosis of resting B-chronic lymphocytic leukemia cells.*** **Haematologica, 2006. 91(7): p. 912-9.**

111. **Xu, K.J., J. Song, and X.M. Zhao,** ***The drug cocktail network.*** **BMC systems biology, 2012. 6 Suppl 1: p. S5.**

112. **Zhao, X.M., et al.,** ***Prediction of drug combinations by integrating molecular and pharmacological data.*** **PLoS computational biology, 2011. 7(12): p. e1002323.**

113. **Lee, J.H., et al.,** ***CDA: combinatorial drug discovery using transcriptional response modules.*** **PloS one, 2012. 7(8): p. e42573.**

114. **Wang, Y.Y., et al.,** ***Exploring drug combinations in genetic interaction network.*** **BMC bioinformatics, 2012. 13 Suppl 7: p. S7.**

115. **Engin, H.B., et al.,** ***Network-Based Strategies Can Help Mono- and Poly-pharmacology Drug Discovery: A Systems Biology View.*** **Current pharmaceutical design, 2013.**

116. **Azmi, A.S., et al.,** ***Proof of concept: network and systems biology approaches aid in the discovery of potent anticancer drug combinations.*** **Molecular cancer therapeutics, 2010. 9(12): p. 3137-44.**

117. **Keskin, O., et al.,** ***Towards drugs targeting multiple proteins in a systems biology approach.*** **Current topics in medicinal chemistry, 2007. 7(10): p. 943-51.**

118. **Venkatesan, K., et al.,** ***An empirical framework for binary interactome mapping.*** **Nature methods, 2009. 6(1): p. 83-90.**

119. **Stumpf, M.P., et al.,** ***Estimating the size of the human interactome.*** **Proceedings of the National Academy of Sciences of the United States of America, 2008. 105(19): p. 6959-64.**

120. **Ruffner, H., A. Bauer, and T. Bouwmeester,** ***Human protein-protein interaction networks and the value for drug discovery.*** **Drug discovery today, 2007. 12(17-18): p. 709-16.**

121. **Ideker, T. and R. Sharan,** ***Protein networks in disease.*** **Genome research, 2008. 18(4): p. 644-52.**

122. **Ma'ayan, A., et al.,** ***Network analysis of FDA approved drugs and their targets.*** **The Mount Sinai journal of medicine, New York, 2007. 74(1): p. 27-32.**

123. **Hase, T., et al.,** ***Structure of protein interaction networks and their implications on drug design.*** **PLoS computational biology, 2009. 5(10): p. e1000550.**

124. **Hwang, W.C., A. Zhang, and M. Ramanathan,** ***Identification of information flow-modulating drug targets: a novel bridging paradigm for drug discovery.*** **Clinical pharmacology and therapeutics, 2008. 84(5): p. 563-72.**

125. **Kitano, H.,** ***Cancer Robustness and Therapy Strategies*****, in** ***CANCER SYSTEMS BIOLOGY, BIOINFORMATICS AND MEDICINE*****2011, Springer.**

126. **Stein, A., et al.,** ***Dynamic interactions of proteins in complex networks: a more structured view.*** **The FEBS journal, 2009. 276(19): p. 5390-405.**

127. **Pache, R.A., et al.,** ***Towards a molecular characterisation of pathological pathways.*** **FEBS letters, 2008. 582(8): p. 1259-65.**

128. **Kinnings, S.L., et al.,** ***The Mycobacterium tuberculosis drugome and its polypharmacological implications.*** **PLoS Comput Biol, 2010. 6(11): p. e1000976.**

129. **Dasika, M.S., A. Burgard, and C.D. Maranas,** ***A computational framework for the topological analysis and targeted disruption of signal transduction networks.*** **Biophys J, 2006. 91(1): p. 382-98.**

130. **Lee, J.M., E.P. Gianchandani, and J.A. Papin,** ***Flux balance analysis in the era of metabolomics.*** **Brief Bioinform, 2006. 7(2): p. 140-50.**

131. **McAdams, H.H. and L. Shapiro,** ***Circuit simulation of genetic networks.*** **Science, 1995. 269(5224): p. 650-6.**

132. Peleg, M., D. Rubin, and R.B. Altman, *Using Petri Net tools to study properties and dynamics of biological systems*. J Am Med Inform Assoc, 2005. 12(2): p. 181-99.

133. Li, Z., et al., *Detecting drug targets with minimum side effects in metabolic networks*. IET Syst Biol, 2009. 3(6): p. 523-33.

134. Layek, R., et al., *Cancer therapy design based on pathway logic*. Bioinformatics, 2011. 27(4): p. 548-55.

135. Ruths, D., et al., *The signaling petri net-based simulator: a non-parametric strategy for characterizing the dynamics of cell-specific signaling networks*. PLoS computational biology, 2008. 4(2): p. e1000005.

136. Dasika, M.S., A. Burgard, and C.D. Maranas, *A computational framework for the topological analysis and targeted disruption of signal transduction networks*. Biophysical journal, 2006. 91(1): p. 382-98.

137. Sridhar, P., T. Kahveci, and S. Ranka, *An iterative algorithm for metabolic network-based drug target identification*. Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing, 2007: p. 88-99.

138. Sridhar, P., et al., *Mining metabolic networks for optimal drug targets*. Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing, 2008: p. 291-302.

139. Yang, K., et al., *Finding multiple target optimal intervention in disease-related molecular network*. Molecular systems biology, 2008. 4: p. 228.

140. Tuncbag, N., et al., *Predicting protein-protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM*. Nature protocols, 2011. 6(9): p. 1341-54.

141. Ogmen, U., et al., *PRISM: protein interactions by structural matching*. Nucleic acids research, 2005. 33(Web Server issue): p. W331-6.

142. Xie, L., J. Li, and P.E. Bourne, *Drug discovery using chemical systems biology: identification of the protein-ligand binding network to explain the side effects of CETP inhibitors*. PLoS Comput Biol, 2009. 5(5): p. e1000387.

143. Kar, G., et al., *Allostery and population shift in drug discovery*. Current opinion in pharmacology, 2010. 10(6): p. 715-22.

144. Nussinov, R., C.J. Tsai, and P. Csermely, *Allo-network drugs: harnessing allostery in cellular networks*. Trends in pharmacological sciences, 2011. 32(12): p. 686-93.

145. DeSantis, C., et al., *Breast cancer statistics, 2011*. CA: a cancer journal for clinicians, 2011. 61(6): p. 409-18.

146. Pasche, B., *Cancer genetics. Introduction*. Cancer treatment and research, 2010. 155: p. xi-xii.

147. Gage, M., D. Wattendorf, and L.R. Henry, *Translational advances regarding hereditary breast cancer syndromes*. Journal of surgical oncology, 2012. 105(5): p. 444-51.

148. Weigelt, B., J.L. Peterse, and L.J. van 't Veer, *Breast cancer metastasis: markers and models*. Nature reviews. Cancer, 2005. 5(8): p. 591-602.

149. Carroll, J.C., et al., *Hereditary breast and ovarian cancers*. Canadian family physician Medecin de famille canadien, 2008. 54(12): p. 1691-2.

150. Jemal, A., et al., *Cancer statistics, 2009*. CA: a cancer journal for clinicians, 2009. 59(4): p. 225-49.

151. **Yun, J., et al., *Signalling pathway for RKIP and Let-7 regulates and predicts metastatic breast cancer.* The EMBO journal, 2011. 30(21): p. 4500-14.**

152. **Barnholtz-Sloan, J.S., et al., *Incidence proportions of brain metastases in patients diagnosed (1973 to 2001) in the Metropolitan Detroit Cancer Surveillance System.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2004. 22(14): p. 2865-72.**

153. **Hamilton, A. and N.R. Sibson, *Role of the systemic immune system in brain metastasis.* Molecular and cellular neurosciences, 2013. 53: p. 42-51.**

154. **Muller, A., et al., *Involvement of chemokine receptors in breast cancer metastasis.* Nature, 2001. 410(6824): p. 50-6.**

155. **Minn, A.J., et al., *Genes that mediate breast cancer metastasis to lung.* Nature, 2005. 436(7050): p. 518-24.**

156. **Rahmathulla, G., S.A. Toms, and R.J. Weil, *The molecular biology of brain metastasis.* Journal of oncology, 2012. 2012: p. 723541.**

157. **Nathoo, N., et al., *Pathobiology of brain metastases.* Journal of clinical pathology, 2005. 58(3): p. 237-42.**

158. **Landis, S.H., et al., *Cancer statistics, 1998.* CA: a cancer journal for clinicians, 1998. 48(1): p. 6-29.**

159. **Gavrilovic, I.T. and J.B. Posner, *Brain metastases: epidemiology and pathophysiology.* Journal of neuro-oncology, 2005. 75(1): p. 5-14.**

160. **Patchell, R.A., *The management of brain metastases.* Cancer treatment reviews, 2003. 29(6): p. 533-40.**

161. **Bild, A.H., et al., *Oncogenic pathway signatures in human cancers as a guide to targeted therapies.* Nature, 2006. 439(7074): p. 353-7.**

162. **Kang, Y., et al., *A multigenic program mediating breast cancer metastasis to bone.* Cancer cell, 2003. 3(6): p. 537-49.**

163. **Mehrotra, J., et al., *Very high frequency of hypermethylated genes in breast cancer metastasis to the bone, brain, and lung.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2004. 10(9): p. 3104-9.**

164. **Liang, Z., et al., *Silencing of CXCR4 blocks breast cancer metastasis.* Cancer research, 2005. 65(3): p. 967-71.**

165. **Brown, D.M. and E. Ruoslahti, *Metadherin, a cell surface protein in breast tumors that mediates lung metastasis.* Cancer cell, 2004. 5(4): p. 365-74.**

166. **Nguyen, D.X., et al., *WNT/TCF signaling through LEF1 and HOXB9 mediates lung adenocarcinoma metastasis.* Cell, 2009. 138(1): p. 51-62.**

167. **Batlle, E., et al., *The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells.* Nature cell biology, 2000. 2(2): p. 84-9.**

168. **Cano, A., et al., *The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression.* Nature cell biology, 2000. 2(2): p. 76-83.**

169. **Comijn, J., et al., *The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion.* Molecular cell, 2001. 7(6): p. 1267-78.**

170. **Bolos, V., et al., *The transcription factor Slug represses E-cadherin expression and induces epithelial to mesenchymal transitions: a comparison with Snail and E47 repressors.* Journal of cell science, 2003. 116(Pt 3): p. 499-511.**

171. **Yang, J., et al.,** ***Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis.*** **Cell, 2004. 117(7): p. 927-39.**
172. **Hartwell, K.A., et al.,** ***The Spemann organizer gene, Goosecoid, promotes tumor metastasis.*** **Proceedings of the National Academy of Sciences of the United States of America, 2006. 103(50): p. 18969-74.**
173. **Mani, S.A., et al.,** ***Mesenchyme Forkhead 1 (FOXC2) plays a key role in metastasis and is associated with aggressive basal-like breast cancers.*** **Proceedings of the National Academy of Sciences of the United States of America, 2007. 104(24): p. 10069-74.**
174. **van 't Veer, L.J., et al.,** ***Gene expression profiling predicts clinical outcome of breast cancer.*** **Nature, 2002. 415(6871): p. 530-6.**
175. **van de Vijver, M.J., et al.,** ***A gene-expression signature as a predictor of survival in breast cancer.*** **The New England journal of medicine, 2002. 347(25): p. 1999-2009.**
176. **Ramaswamy, S., et al.,** ***A molecular signature of metastasis in primary solid tumors.*** **Nature genetics, 2003. 33(1): p. 49-54.**
177. **Bos, P.D., et al.,** ***Genes that mediate breast cancer metastasis to the brain.*** **Nature, 2009. 459(7249): p. 1005-9.**
178. **Van Heyningen, V. and P.L. Yeyati,** ***Mechanisms of non-Mendelian inheritance in genetic disease.*** **Human molecular genetics, 2004. 13 Spec No 2: p. R225-33.**
179. **Ergun, A., et al.,** ***A network biology approach to prostate cancer.*** **Molecular systems biology, 2007. 3: p. 82.**
180. **Wu, X., et al.,** ***Network-based global inference of human disease genes.*** **Molecular systems biology, 2008. 4: p. 189.**
181. **Lee, I., et al.,** ***Predicting genetic modifier loci using functional gene networks.*** **Genome research, 2010. 20(8): p. 1143-53.**
182. **Chuang, H.Y., et al.,** ***Network-based classification of breast cancer metastasis.*** **Molecular systems biology, 2007. 3: p. 140.**
183. **Kann, M.G.,** ***Protein interactions and disease: computational approaches to uncover the etiology of diseases.*** **Briefings in bioinformatics, 2007. 8(5): p. 333-46.**
184. **Berman, H.M., et al.,** ***The Protein Data Bank.*** **Nucleic acids research, 2000. 28(1): p. 235-42.**
185. **David, A., et al.,** ***Protein-protein interaction sites are hot spots for disease-associated nonsynonymous SNPs.*** **Human mutation, 2012. 33(2): p. 359-63.**
186. **Mareel, M. and A. Leroy,** ***Clinical, cellular, and molecular aspects of cancer invasion.*** **Physiological reviews, 2003. 83(2): p. 337-76.**
187. **Sordat B, P.J.-C., Weiss L,** ***Is there a common definition for invasiveness?*** **Invasion Metastasis 10, 1990: p. 178–192.**
188. **Orozco, E., L. Benitez-Bibriesca, and R. Hernandez,** ***Invasion and metastasis mechanisms in Entamoeba histolytica and cancer cells. Some common cellular and molecular features.*** **Mutation research, 1994. 305(2): p. 229-39.**
189. **Leroy, A., et al.,** ***Metastasis of Entamoeba histolytica compared to colon cancer: one more step in invasion.*** **Invasion & metastasis, 1994. 14(1-6): p. 177-91.**
190. **Haile, S.,** ***Cancer metastasis and in vivo dissemination of tissue-dwelling pathogens: extrapolation of mechanisms and exchange of treatment strategies thereof.*** **Medical hypotheses, 2008. 70(2): p. 375-7.**

191. Shih, K.-J.L.a.N.-Y., *The Role of Enolase in Tissue Invasion and Metastasis of Pathogens and Tumor Cells*. J. Cancer Mol., 2007. 3(2): p. 45-48.
192. Ryan, R.M., J. Green, and C.E. Lewis, *Use of bacteria in anti-cancer therapies*. BioEssays : news and reviews in molecular, cellular and developmental biology, 2006. 28(1): p. 84-94.
193. Hayashi, K., et al., *Cancer metastasis directly eradicated by targeted therapy with a modified Salmonella typhimurium*. Journal of cellular biochemistry, 2009. 106(6): p. 992-8.
194. Yu, Y.A., et al., *Visualization of tumors and metastases in live animals with bacteria and vaccinia virus encoding light-emitting proteins*. Nature biotechnology, 2004. 22(3): p. 313-20.
195. Shiau, A.L., et al., *Prothymosin alpha enhances protective immune responses induced by oral DNA vaccination against pseudorabies delivered by Salmonella choleraesuis*. Vaccine, 2001. 19(28-29): p. 3947-56.
196. Volpert, O.V., J. Lawler, and N.P. Bouck, *A human fibrosarcoma inhibits systemic angiogenesis and the growth of experimental metastases via thrombospondin-1*. Proceedings of the National Academy of Sciences of the United States of America, 1998. 95(11): p. 6343-8.
197. Dallo, S.F. and T. Weitao, *Bacteria under SOS evolve anticancer phenotypes*. Infectious agents and cancer, 2010. 5(1): p. 3.
198. Ben-Jacob, E., D.S. Coffey, and H. Levine, *Bacterial survival strategies suggest rethinking cancer cooperativity*. Trends in microbiology, 2012. 20(9): p. 403-10.
199. Wang, R.F., Y. Miyahara, and H.Y. Wang, *Toll-like receptors and immune regulation: implications for cancer therapy*. Oncogene, 2008. 27(2): p. 181-9.
200. Lee, C.H., *Engineering bacteria toward tumor targeting for cancer treatment: current state and perspectives*. Applied microbiology and biotechnology, 2012. 93(2): p. 517-23.
201. Lage, H., *An overview of cancer multidrug resistance: a still unsolved problem*. Cellular and molecular life sciences : CMLS, 2008. 65(20): p. 3145-67.
202. Lambert, G., et al., *An analogy between the evolution of drug resistance in bacterial communities and malignant tissues*. Nature reviews. Cancer, 2011. 11(5): p. 375-82.
203. Glickman, M.S. and C.L. Sawyers, *Converting cancer therapies into cures: lessons from infectious diseases*. Cell, 2012. 148(6): p. 1089-98.
204. Kar, G., A. Gursoy, and O. Keskin, *Human cancer protein-protein interaction network: a structural perspective*. PLoS Comput Biol, 2009. 5(12): p. e1000601.
205. Gursoy, A., O. Keskin, and R. Nussinov, *Topological properties of protein interaction networks from a structural perspective*. Biochemical Society transactions, 2008. 36(Pt 6): p. 1398-403.
206. Keskin, O., R. Nussinov, and A. Gursoy, *PRISM: protein-protein interaction prediction by structural matching*. Methods in molecular biology, 2008. 484: p. 505-21.
207. Berman, H.M., et al., *The Protein Data Bank*. Nucleic Acids Res, 2000. 28(1): p. 235-42.
208. Shatsky, M., R. Nussinov, and H.J. Wolfson, *A method for simultaneous alignment of multiple protein structures*. Proteins, 2004. 56(1): p. 143-56.

209. **Mashiach, E., R. Nussinov, and H.J. Wolfson,** ***FiberDock: a web server for flexible induced-fit backbone refinement in molecular docking.*** **Nucleic acids research, 2010. 38(Web Server issue): p. W457-61.**

210. **Mashiach, E., R. Nussinov, and H.J. Wolfson,** ***FiberDock: Flexible induced-fit backbone refinement in molecular docking.*** **Proteins, 2010. 78(6): p. 1503-19.**

211. **Kar, G., et al.,** ***Human proteome-scale structural modeling of E2-E3 interactions exploiting interface motifs.*** **Journal of proteome research, 2012. 11(2): p. 1196-207.**

212. **Acuner Ozbabacan, S.E., et al.,** ***Enriching the human apoptosis pathway by predicting the structures of protein-protein complexes.*** **Journal of structural biology, 2012. 179(3): p. 338-46.**

213. **Tuncbag, N., et al.,** ***Fast and accurate modeling of protein-protein interactions by combining template-interface-based docking with flexible refinement.*** **Proteins, 2012. 80(4): p. 1239-49.**

214. **Keskin, O., et al.,** ***A new, structurally nonredundant, diverse data set of protein-protein interfaces and its implications.*** **Protein Sci, 2004. 13(4): p. 1043-55.**

215. **Keskin, O., et al.,** ***Protein-protein interactions: organization, cooperativity and mapping in a bottom-up Systems Biology approach.*** **Phys Biol, 2005. 2(2): p. S24-35.**

216. **Keskin, O. and R. Nussinov,** ***Similar binding sites and different partners: implications to shared proteins in cellular pathways.*** **Structure, 2007. 15(3): p. 341-54.**

217. **Haupt, V.J. and M. Schroeder,** ***Old friends in new guise: repositioning of known drugs with structural bioinformatics.*** **Brief Bioinform, 2011.**

218. **Perot, S., et al.,** ***Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery.*** **Drug Discov Today, 2010. 15(15-16): p. 656-67.**

219. **Hopkins, A.L.,** ***Network pharmacology: the next paradigm in drug discovery.*** **Nat Chem Biol, 2008. 4(11): p. 682-90.**

220. **Harris, S.L. and A.J. Levine,** ***The p53 pathway: positive and negative feedback loops.*** **Oncogene, 2005. 24(17): p. 2899-908.**

221. **Hanahan, D. and R.A. Weinberg,** ***The hallmarks of cancer.*** **Cell, 2000. 100(1): p. 57-70.**

222. **Vogelstein, B., D. Lane, and A.J. Levine,** ***Surfing the p53 network.*** **Nature, 2000. 408(6810): p. 307-10.**

223. **Haupt, S., et al.,** ***Apoptosis - the p53 network.*** **J Cell Sci, 2003. 116(Pt 20): p. 4077-85.**

224. **Tuncbag, N., et al.,** ***Predicting protein-protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM.*** **Nat Protoc, 2011. 6(9): p. 1341-54.**

225. **Ogmen, U., et al.,** ***PRISM: protein interactions by structural matching.*** **Nucleic Acids Res, 2005. 33(Web Server issue): p. W331-6.**

226. **Kohn, K.W.,** ***Molecular interaction map of the mammalian cell cycle control and DNA repair systems.*** **Molecular biology of the cell, 1999. 10(8): p. 2703-34.**

227. **Prasad, T.S., K. Kandasamy, and A. Pandey,** ***Human Protein Reference Database and Human Proteinpedia as discovery tools for systems biology.*** **Methods Mol Biol, 2009. 577: p. 67-79.**

228. **Ceol, A., et al.,** ***MINT, the molecular interaction database: 2009 update.*** **Nucleic Acids Res, 2010. 38(Database issue): p. D532-9.**

229. Aranda, B., et al., *The IntAct molecular interaction database in 2010.* Nucleic Acids Res, 2010. 38(Database issue): p. D525-31.
230. Matthews, L., et al., *Reactome knowledgebase of human biological pathways and processes.* Nucleic Acids Res, 2009. 37(Database issue): p. D619-22.
231. Stark, C., et al., *BioGRID: a general repository for interaction datasets.* Nucleic Acids Res, 2006. 34(Database issue): p. D535-9.
232. Cerami, E.G., et al., *Pathway Commons, a web resource for biological pathway data.* Nucleic Acids Res, 2011. 39(Database issue): p. D685-90.
233. Schaefer, C.F., et al., *PID: the Pathway Interaction Database.* Nucleic Acids Res, 2009. 37(Database issue): p. D674-9.
234. Szklarczyk, D., et al., *The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored.* Nucleic Acids Res, 2011. 39(Database issue): p. D561-8.
235. Guven Maiorov, E., et al., *The structural network of inflammation and cancer: Merits and challenges.* Semin Cancer Biol, 2013. 23(4): p. 243-51.
236. Trinchieri, G., *Cancer and inflammation: an old intuition with rapidly evolving new concepts.* Annu Rev Immunol, 2012. 30: p. 677-706.
237. Trinchieri, G., *Inflammation in cancer: a therapeutic target?* Oncology (Williston Park), 2011. 25(5): p. 418-20.
238. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. 144(5): p. 646-74.
239. Fiorentino, D.F., M.W. Bond, and T.R. Mosmann, *Two types of mouse T helper cell. IV. Th2 clones secrete a factor that inhibits cytokine production by Th1 clones.* J Exp Med, 1989. 170(6): p. 2081-95.
240. Salwinski, L., et al., *The Database of Interacting Proteins: 2004 update.* Nucleic acids research, 2004. 32(Database issue): p. D449-51.
241. Pagel, P., et al., *The MIPS mammalian protein-protein interaction database.* Bioinformatics, 2005. 21(6): p. 832-4.
242. Keshava Prasad, T.S., et al., *Human Protein Reference Database--2009 update.* Nucleic acids research, 2009. 37(Database issue): p. D767-72.
243. Bader, G.D., D. Betel, and C.W. Hogue, *BIND: the Biomolecular Interaction Network Database.* Nucleic acids research, 2003. 31(1): p. 248-50.
244. Kerrien, S., et al., *The IntAct molecular interaction database in 2012.* Nucleic acids research, 2012. 40(Database issue): p. D841-6.
245. Licata, L., et al., *MINT, the molecular interaction database: 2012 update.* Nucleic acids research, 2012. 40(Database issue): p. D857-61.
246. Stark, C., et al., *The BioGRID Interaction Database: 2011 update.* Nucleic acids research, 2011. 39(Database issue): p. D698-704.
247. Apweiler, R., et al., *UniProt: the Universal Protein knowledgebase.* Nucleic acids research, 2004. 32(Database issue): p. D115-9.
248. Zhang, Y. and J. Skolnick, *TM-align: a protein structure alignment algorithm based on the TM-score.* Nucleic acids research, 2005. 33(7): p. 2302-9.
249. Roy, A., A. Kucukural, and Y. Zhang, *I-TASSER: a unified platform for automated protein structure and function prediction.* Nature protocols, 2010. 5(4): p. 725-38.
250. Tuncbag, N., et al., *Towards inferring time dimensionality in protein-protein interaction networks by integrating structures: the p53 example.* Mol Biosyst, 2009. 5(12): p. 1770-8.

251. Franceschini, A., et al., *STRING v9.1: protein-protein interaction networks, with increased coverage and integration*. Nucleic acids research, 2013. 41(Database issue): p. D808-15.

252. Bernstein, F.C., et al., *The Protein Data Bank: a computer-based archival file for macromolecular structures*. Journal of molecular biology, 1977. 112(3): p. 535-42.

253. Arkin, M.R. and A. Whitty, *The road less traveled: modulating signal transduction enzymes by inhibiting their protein-protein interactions*. Curr Opin Chem Biol, 2009. 13(3): p. 284-90.

254. Fry, D.C. and L.T. Vassilev, *Targeting protein-protein interactions for cancer therapy*. J Mol Med (Berl), 2005. 83(12): p. 955-63.

255. Whitty, A. and G. Kumaravel, *Between a rock and a hard place?* Nat Chem Biol, 2006. 2(3): p. 112-8.

256. Fuller, J.C., N.J. Burgoyne, and R.M. Jackson, *Predicting druggable binding sites at the protein-protein interface*. Drug Discov Today, 2009. 14(3-4): p. 155-61.

257. Gonzalez-Ruiz, D. and H. Gohlke, *Targeting protein-protein interactions with small molecules: challenges and perspectives for computational binding epitope detection and ligand finding*. Curr Med Chem, 2006. 13(22): p. 2607-25.

258. Wishart, D.S., et al., *DrugBank: a knowledgebase for drugs, drug actions and drug targets*. Nucleic Acids Res, 2008. 36(Database issue): p. D901-6.

259. Kuritzkes, D., S. Kar, and P. Kirkpatrick, *Fresh From The Pipeline; Maraviroc*. Nature Reviews Drug Discovery, 2008. 7(1): p. 15-16.

260. Domling, A., *Small molecular weight protein-protein interaction antagonists: an insurmountable challenge?* Curr Opin Chem Biol, 2008. 12(3): p. 281-91.

261. Bolton E, W.Y., Thiessen PA, Bryant SH, *PubChem: Integrated Platform of Small Molecules and Biological Activities*, in *Annual Reports in Computational Chemistry*2008, American Chemical Society: Washington, DC.

262. Lu, H. and U. Schulze-Gahmen, *Toward understanding the structural basis of cyclin-dependent kinase 6 specific inhibition*. J Med Chem, 2006. 49(13): p. 3826-31.

263. Baughn, L.B., et al., *A novel orally active small molecule potently induces G1 arrest in primary myeloma cells and prevents tumor growth by specific inhibition of cyclin-dependent kinase 4/6*. Cancer Res, 2006. 66(15): p. 7661-7.

264. Chen, X., Z.L. Ji, and Y.Z. Chen, *TTD: Therapeutic Target Database*. Nucleic Acids Res, 2002. 30(1): p. 412-5.

265. Cho YS, B.M., Brain C, Chen CH, Cheng H, Chopra R, Chung K, Groarke J, He G, Hou Y, Kim S, Kovats S, Lu Y, O'Reilly M, Shen J, Smith T, Trakshel G, Vögtle M, Xu M, Xu M, Sung MJ, *4-(Pyrazol-4-yl)-pyrimidines as Selective Inhibitors of Cyclin-Dependent Kinase 4/6*. Journal of Medicinal Chemistry, 2010. 53 p. 7938–795.

266. Lu, H., et al., *Crystal structure of a human cyclin-dependent kinase 6 complex with a flavonol inhibitor, fisetin*. J Med Chem, 2005. 48(3): p. 737-43.

267. Magrane, M. and U. Consortium, *UniProt Knowledgebase: a hub of integrated protein data*. Database (Oxford), 2011. 2011: p. bar009.

268. Brotherton, D.H., et al., *Crystal structure of the complex of the cyclin D-dependent kinase Cdk6 bound to the cell-cycle inhibitor p19INK4d*. Nature, 1998. 395(6699): p. 244-50.

269. **Russo, A.A., et al.,** ***Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16INK4a.*** **Nature, 1998. 395(6699): p. 237-43.**

270. **Chan, F.K., et al.,** ***Identification of human and mouse p19, a novel CDK4 and CDK6 inhibitor with homology to p16ink4.*** **Mol Cell Biol, 1995. 15(5): p. 2682-8.**

271. **Tuncbag, N., A. Gursoy, and O. Keskin,** ***Identification of computational hot spots in protein interfaces: combining solvent accessibility and inter-residue potentials improves the accuracy.*** **Bioinformatics, 2009. 25(12): p. 1513-20.**

272. **Humphrey, W., A. Dalke, and K. Schulten,** ***VMD: visual molecular dynamics.*** **Journal of molecular graphics, 1996. 14(1): p. 33-8, 27-8.**

273. **Morris, G.M., et al.,** ***AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility.*** **Journal of computational chemistry, 2009. 30(16): p. 2785-91.**

274. **Fry, D.W., et al.,** ***Specific inhibition of cyclin-dependent kinase 4/6 by PD 0332991 and associated antitumor activity in human tumor xenografts.*** **Mol Cancer Ther, 2004. 3(11): p. 1427-38.**

275. **Gunther, S., et al.,** ***SuperTarget and Matador: resources for exploring drug-target relationships.*** **Nucleic Acids Res, 2008. 36(Database issue): p. D919-22.**

276. **Frey, B.J. and D. Dueck,** ***Clustering by passing messages between data points.*** **Science, 2007. 315(5814): p. 972-6.**

277. **Bloom, J. and M. Pagano,** ***Deregulated degradation of the cdk inhibitor p27 and malignant transformation.*** **Seminars in cancer biology, 2003. 13(1): p. 41-7.**

278. **El Baroudi, M., et al.,** ***A curated database of miRNA mediated feed-forward loops involving MYC as master regulator.*** **PLoS One, 2011. 6(3): p. e14742.**

279. **Latora, V. and M. Marchiori,** ***Efficient behavior of small-world networks.*** **Phys Rev Lett, 2001. 87(19): p. 198701.**

280. **Defranchi, E., et al.,** ***Binding of protein kinase inhibitors to synapsin I inferred from pair-wise binding site similarity measurements.*** **PLoS One, 2010. 5(8): p. e12214.**

281. **Morris, J.H., et al.,** ***clusterMaker: a multi-algorithm clustering plugin for Cytoscape.*** **BMC Bioinformatics, 2011. 12(1): p. 436.**

282. **Smoot, M.E., et al.,** ***Cytoscape 2.8: new features for data integration and network visualization.*** **Bioinformatics, 2011. 27(3): p. 431-2.**

283. **Aric A. Hagberg, D.A.S.a.P.J.S.** ***Exploring network structure, dynamics, and function using NetworkX.*** **in** ***Proceedings of the 7th Python in Science Conference (SciPy2008).*** **Aug 2008. Pasadena, CA USA.**

284. **Guney, E. and B. Oliva,** ***Exploiting protein-protein interaction networks for genome-wide disease-gene prioritization.*** **PloS one, 2012. 7(9): p. e43557.**

285. **Engin HB GE, K.O., Oliva B, Gursoy A,** ***Integrating Structure to Protein-Protein Interaction Networks that Drive Metastasis to Brain and Lung in Breast Cancer.*** **PlosOne, In Press.**

286. **Bindea, G., et al.,** ***ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks.*** **Bioinformatics, 2009. 25(8): p. 1091-3.**

287. **Mao, X., et al.,** ***Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary.*** **Bioinformatics, 2005. 21(19): p. 3787-93.**

288. **Xie, L., J. Li, and P.E. Bourne,** ***Drug discovery using chemical systems biology: identification of the protein-ligand binding network to explain the side effects of CETP inhibitors.*** **PLoS computational biology, 2009. 5(5): p. e1000387.**

289. **Haupt, V.J. and M. Schroeder,** ***Old friends in new guise: repositioning of known drugs with structural bioinformatics.*** **Briefings in bioinformatics, 2011. 12(4): p. 312-26.**

290. **Perot, S., et al.,** ***Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery.*** **Drug discovery today, 2010. 15(15-16): p. 656-67.**

291. **Zetter, B.R.,** ***Adhesion molecules in tumor metastasis.*** **Seminars in cancer biology, 1993. 4(4): p. 219-29.**

292. **Bendas, G. and L. Borsig,** ***Cancer cell adhesion and metastasis: selectins, integrins, and the inhibitory potential of heparins.*** **International journal of cell biology, 2012. 2012: p. 676731.**

293. **Walter F., P.B.,** ***Medical Physiology: A Cellular And Molecular Approach*****, ed. Elsevier/Saunders. Vol. p. 1300. 2003.**

294. **MacDonald, T.J., et al.,** ***Expression profiling of medulloblastoma: PDGFRA and the RAS/MAPK pathway as therapeutic targets for metastatic disease.*** **Nature genetics, 2001. 29(2): p. 143-52.**

295. **Keya De Mukhopadhyay, A.G.E., Andrew P. Hinck, Kihoon Yoon, John E. Cornell, Lan Yu, Zhao Liu, Junhua Yang, and LuZhe Sun, 2012.**

296. ***Update on activities at the Universal Protein Resource (UniProt) in 2013.*** **Nucleic acids research, 2013. 41(D1): p. D43-D47.**

297. **Nourry, C., et al.,** ***Direct interaction between Smad3, APC10, CDH1 and HEF1 in proteasomal degradation of HEF1.*** **BMC cell biology, 2004. 5: p. 20.**

298. **Jorissen, R.N., et al.,** ***Epidermal growth factor receptor: mechanisms of activation and signalling.*** **Experimental cell research, 2003. 284(1): p. 31-53.**

299. **Sillitoe, I., et al.,** ***New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures.*** **Nucleic acids research, 2013. 41(Database issue): p. D490-8.**

300. **Shin, S.Y., et al.,** ***The chemical synthesis and binding affinity to the EGF receptor of the EGF-like domain of heparin-binding EGF-like growth factor (HB-EGF).*** **Journal of peptide science : an official publication of the European Peptide Society, 2003. 9(4): p. 244-50.**

301. **Sato, K., et al.,** ***Solution structure of epiregulin and the effect of its C-terminal domain for receptor binding affinity.*** **FEBS letters, 2003. 553(3): p. 232-8.**

302. **Garcia-Garcia, J., et al.,** ***Biana: a software framework for compiling biological interactions and analyzing networks.*** **BMC bioinformatics, 2010. 11: p. 56.**

303. **Liu, X., et al.,** ***TiGER: a database for tissue-specific gene expression and regulation.*** **BMC bioinformatics, 2008. 9: p. 271.**

304. **Team, R.C.,** ***R: A Language and Environment for Statistical Computing*****, 2013, R Foundation for Statistical Computing: Vienna, Austria.**

305. **Ashburner, M., et al.,** ***Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.*** **Nature genetics, 2000. 25(1): p. 25-9.**

306. **Kumar, R. and B. Nanduri,** ***HPIDB--a unified resource for host-pathogen interactions.*** **BMC bioinformatics, 2010. 11 Suppl 6: p. S16.**

307. Forbes, S.A., et al., ***COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer.*** **Nucleic acids research, 2011. 39(Database issue): p. D945-50.**

308. Martin, A.C., ***Mapping PDB chains to UniProtKB entries.*** **Bioinformatics, 2005. 21(23): p. 4297-301.**

309. Hubbard, S.J.T., J. M. n, ***naccess.*** **Department of Biochemistry and Molecular Biology, University College London, 1993.**

310. Lee, B. and F.M. Richards, ***The interpretation of protein structures: estimation of static accessibility.*** **Journal of molecular biology, 1971. 55(3): p. 379-400.**

311. Schwarz-Linek, U., et al., ***Pathogenic bacteria attach to human fibronectin through a tandem beta-zipper.*** **Nature, 2003. 423(6936): p. 177-81.**

312. Flores-Villanueva, P.O., et al., ***A functional promoter polymorphism in monocyte chemoattractant protein-1 is associated with increased susceptibility to pulmonary tuberculosis.*** **The Journal of experimental medicine, 2005. 202(12): p. 1649-58.**

Vita

H. Billur Engin Aras was born in Istanbul, Turkey, on April 13, 1983. She received the B.Sc. Degree in Civil Engineering from Boğaziçi University in 2006 and M.Sc. Degree in Computer Engineering from Sabancı University in 2008. She received the Ph.D. Degree from Koç University in Computer Engineering in 2013. From September 2008 to September 2013 she worked as a teaching and research assistant at Koç University. She had been a visiting researcher in different institutes such as National Cancer Institute (NCI) and University of Pompeu Fabra.

Her research mainly focuses on structural modeling of protein-protein interfaces, their integration to protein-protein interaction networks, systems biology approaches to understand complex diseases and drug off-target prediction using computational methods. She has published articles in prestigious journals such as Protein Engineering Design and Selection, Journal of Chemical Information and Modeling, Current Pharmaceutical Design and PLOS One.

She will continue her academic career as a Postdoctoral Associate in the Department of Medicine at University of California San Diego (UCSD), where she will be focusing on new bioinformatics approaches for the analysis of genome-wide cancer data sets provided by The Cancer Genome Atlas, the UCSD / Moores Cancer Center.