Large-Scale Integration of Protein Structural Data into Protein-Protein Interaction Networks

by

Gözde Kar Makinacı

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
Chemical and Biological Engineering

Koç University

October, 2012

Koç University

Graduate School of Sciences and Engineering

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

Gözde Kar Makinacı

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. Özlem Keskin

_____

Prof. Attila Gürsoy

_____

Prof. Türkan Haliloğlu

_____

Assoc. Prof. Halil Kavaklı

_____

Assist. Prof. Nathan Lack

_____

Assist. Prof. Nurhan Özlü

_____

Assoc. Prof. Uğur Sezerman

Date _____

ABSTRACT

Protein-protein interaction networks provide a global picture of cellular function and biological processes. Structural prediction and modeling of protein-protein interactions at the network level is crucial; it helps in assigning protein function, elucidating functional mechanisms, and drug discovery. In the first part of this dissertation, we illustrate the importance of integrating protein structural information into interaction networks, particularly in identification of selective drug targets or drugs targeting multiple proteins. Next, we present a large, proteome-scale strategy that predicts protein associations based on interface structural motifs, to analyze human ubiquitination pathway. Substrate ubiquitination is mediated by the interactions between E2 enzymes and E3 ligases. Although these E2 and E3 proteins function in a concerted manner, the principles of selectivity between them are still not entirely understood. Our method allows elucidation of which E3s interact with which E2s and how they interact with each other. Interface analysis of E2-E3 complexes reveals important clues for inferring the specificity of the interactions. In the last part, the focus is directed towards studying circadian clock regulation in p53-deficient background. p53-deficient cells are prone to tumorigenesis and cancer. However, upon circadian clock disruption by *Cry* knockout, they show an increased sensitivity to apoptosis by genotoxic agents and hence are protected from the early onset of cancer. We aim to elucidate how apoptotic signals are activated in *p53*-null cells upon *Cry* knockout by combining experiments with a large-scale computational approach. In particular, we perform a large-scale integration of microarray expression profiles with protein-protein interaction networks. As such, we observe that the expressions of several apoptotic genes are increased upon *Cry* knockout in *p53*-null cells and a minor amount of genes would promote cell survival leading to, in overall, a shift towards cell apoptosis. In addition, we highlight the pathways that intersect with circadian clock and illustrate how these pathways response to circadian clock disruption. Our findings would assist in identifying targets in treatment of cancers associated with p53-deficiency.

ÖZET

Protein-protein etkileşim ağları hücre fonksiyonunu ve biyolojik prosesleri global bir çerçevede sunmaktadır. Protein-protein etkileşimlerinin ağ düzeyinde yapısal tahmini ve modellenmesi önemlidir; protein fonksiyonu ve fonksiyonel mekanizmaları anlamaya ve ilaç bulguları için yardımcı olur. Bu tezin ilk bölümünde, protein yapısal bilgisinin etkileşim ağlarına entegre edilmesinin önemini, özellikle selektif ilaç hedeflerinin veya birden fazla proteini hedef alan ilaçların belirlenmesindeki önemini gösteriyoruz. Bunun sonrasında, insan ubikuitin biyolojik yolunu analiz etmek için, protein ilişkilerini ara yüzey yapısal motiflere bağlı olarak tahmin eden geniş, proteom düzeyinde bir strateji sunuyoruz. Substrat ubikuitin eklenmesi E2 enzimi ve E3 ligazlarinin etkileşimleri doğrultusunda yönetilir. E2 ve E3 proteinleri beraber uygun bir şekilde fonksiyon gösterse de, aralarındaki seçicilik halen tam olarak anlaşılamamıştır. Metodumuz hangi E3lerin hangi E2ler ile etkileşime girdiğini ve nasıl etkileştiklerini aydınlatmaya olanak sağlanmaktadır. E2-E3 komplekslerinin ara yüzey analizleri, bu etkileşimlerin bağlanma özgünlüğü hakkında önemli ipuçları ortaya çıkarmaktadır. Son bolümde, odağımızı p53 eksikliğinde biyolojik saat regülasyonunu çalışmaya yönlendiriyoruz. p53 geni eksik olan hücreler tümör ve kanser oluşumuna yatkındırlar. Bununla birlikte, bu hücreler, biyolojik saat Cry geninin hücreden çıkarılması ile bozulunca, genotoksik maddelerle hücre ölümüne daha duyarlı hale gelmekte ve böylece kanserin erken başlamasından korunmaktadırlar. Deneyleri büyük düzeyde hesaplamalı yöntemlerle birleştirerek, p53 ve Cry geni olmayan hücrelerde, hücre ölümü sinyallerinin nasıl aktive edildiğini açıklamak istiyoruz. Spesifik olursak, gen ekspresyon profillerini protein-protein etkileşim ağları ile birleştiriyoruz. Böylelikle, p53 geni barındırmayan hücrelerde Cry genini de hücreden çıkarınca apoptotik genlerin ekspresyonlarının arttığını, çok az sayıda genin hücre yaşamını desteklediğini, bütünde hücre ölümüne doğru bir yönelme gözlemliyoruz. Ek olarak, biyolojik saat ile kesişen biyolojik yolları vurguluyoruz ve bunların biyolojik saat bozulmasına nasıl tepki verdiklerini gösteriyoruz. Bulgularımız, p53-eksikliği ile bağlantılı kanser oluşumunun tedavisi ve hedef belirlemelerinde yardımcı olacaktır.

ACKNOWLEDGEMENT

First and foremost, I would like to thank my supervisors, Prof. Özlem Keskin and Prof. Attila Gürsoy for their encouragement, friendship and continuous support during my PhD. It has been a great chance to work in such a productive research group. Their intuition and experiences have contributed significantly to my academic background and I am truly indebted to them a lot.

I am really thankful to Prof. Ruth Nussinov who greatly motivated, inspired and guided me with her endless experience and energy. I am also very thankful to Assoc. Prof. Halil Kavaklı for his great support and guidance in preparing this dissertation. I would like to thank my other thesis committee members Prof. Burak Erman and Prof. Türkan Haliloğlu for their critical reading and useful comments and thesis jury members Assist. Prof. Nathan Lack, Assist. Prof. Nurhan Özlü and Assoc. Prof. Uğur Sezerman for their valuable time. I would also like to thank the Scientific and Technological Research Council of Turkey (TUBITAK) for their financial support during my PhD study.

I would like to thank my past officemates Nurcan Tunçbağ, Cengiz Ulubaş and Emre Güney and my current officemates Billur Engin, Ece Özbabacan, Güray Kuzu, Engin Çukuroğlu and Emine Güven for all their support and good times together.

Finally, I thank my family for their patience and continuous support throughout the PhD years.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

PDB	Protein Data Bank
ASA	Accessible Surface Area
PPI	Protein – Protein Interaction
RMSD	Root Mean Square Deviation
Prism	Protein Interactions by Structural Matching

Chapter 1

INTRODUCTION

Protein-protein interactions have a key role in regulating several biological processes, cellular and signaling events. With the increase in availability of protein interaction data, considerable attention has been directed towards studying the networks of these interactions which provides a global picture of the protein function. To further understand the details of the interactions in the network, in particular, to address the question of "*how do proteins interact*", the knowledge of protein three-dimensional structures is crucial. Although there has been an exponential increase in protein structural information, the gap between the number of known interactions obtained from high-throughput experiments and structurally known protein complexes is still large. Constructing the global structural network of the functional proteome, therefore, has been a challenging goal in structural systems biology [1]. Towards this aim, computational approaches for predicting the functional associations of proteins are immensely useful. Such prediction algorithms [2] would provide the binding details of the interactions at a network level.

With the advance of technologies such as whole genome sequencing, highthroughput experiments and expression profiling, vast amounts of data, covering different aspects of cellular physiology, have emerged [3, 4]. One such example is transcriptomics data (microarray-based genome-wide expression profiles), which provides essential insights into the understanding of the molecular mechanisms underlying normal and dysfunctional biological processes. While these kinds of omics data provide unprecedented views of cellular components in the biological systems [4], they are frequently difficult to interpret due to the overwhelming dimensions of data and the noise inherent in the biological and experimental systems [5]. One way to overcome such problems is to combine these large and heterogeneous data sets. For example, analysis of gene expression data can be improved by its integration with protein interaction networks giving more insight into actual biological networks.

This dissertation, primarily focuses on large-scale integration of protein structural and gene expression profiling data into protein-protein interaction networks. Structural analysis of protein interactions at the molecular and the proteome levels are performed and modeling of protein associations towards construction of structural protein interaction networks at large scale are studied. Crosstalk between pathways are investigated through analysis of microarray gene expression profiles together with protein interaction networks. The presented findings, ultimately, assist in functional genomics and drug design studies.

The outline of this dissertation is as follows:

In Chapter 2, an extended and recent literature review focusing on the protein-protein interaction networks is presented. This chapter includes the corresponding works related to physical and chemical aspects of protein-protein interactions, construction of structural protein interaction networks and integration of microarray gene expression data into protein interaction networks.

In Chapter 3, the importance of protein structural information in drug design and discovery is highlighted through presenting several examples. Here, first, structural models of ubiquitination pathway complexes related to disease are visualized and the importance of incorporating hot spot information in drug design studies is illustrated. Next, the concept of allostery and population shift in drug discovery is introduced and structural modeling of allosteric protein associations in signaling is demonstrated.

In Chapter 4, ubiquitination, which is crucial for many cellular processes such as protein degradation, DNA repair, transcription regulation, and cell signaling, is studied. The interactions between E2 and E3 proteins in the human proteome are structurally modeled and thus a human E2-E3 structural protein interaction network is constructed. The accuracy of the predictions is 76% indicating that the predicted E2-E3 interactions are in agreement with validated functional E2-E3 pairs. Interface analysis of the E2-E3 complexes illuminates the critical residues contributing to specificity in E3 binding.

Chapter 5 is designed to introduce large-scale integration of gene expression profiles with protein-protein interaction networks. Here, the focus is directed towards circadian clock pathway and the aim is to elucidate the pathways under clock regulation. The outcome of an additional clock disruption on p53-deficient cells are investigated by analyzing pathways related to cell death and growth. The crosstalk proteins are highlighted

by extracting the critical paths connecting circadian clock pathway to apoptotic and anti-apoptotic processes.

Finally, Chapter 6 presents a discussion of the results, major conclusions and contributions of this dissertation.

Chapter 2

LITERATURE REVIEW

In this chapter, firstly, a comprehensive review of the studies related to physical and chemical aspects of protein-protein interactions is presented. Next, the most recent review of structural modeling of the protein interactions and constructing structural protein interaction networks are provided. Finally, the widespread use of microarray gene expression profiling data and its integration with protein interaction networks are reviewed.

2.1 Physical and Chemical Aspects of Protein-Protein Interactions

Biological reactions occur through the proteins; and the interactions among them play a crucial role. Protein interactions are biophysical phenomena. Although the medium is full of numerous molecules at different sizes, proteins to be interacted find each other and the interaction occurs. Proteins are interacting through their surfaces with the help of the shape and biochemical complementarity regarding the flexibility of the molecules and the environmental conditions [6-10]. Hence, protein interactions are special reactions rather than random processes.

2.1.1 Interaction Occurs through the Interface

The region where the two molecules are contacting is called the binding site, or considering both sides, the interface. Understanding the protein structures is very important to identify these regions. If the structures of contacting proteins are known, it is quite easy to determine the interface. Interfacial residues are usually found by calculation of close residues from two sides based on the distance in the three dimensional space [11-13] or with the help of accessible surface area calculations [13-15].

Many studies investigate properties of interactions to enlighten protein interaction phenomena [8, 9, 11, 14-26]. The stability of the interaction is provided with binding forces; hydrogen bonds, salt bridges, electrostatic interactions and hydrophobic attractions [21, 27]. Disulphide bonds are also but rarely seen between binding proteins. These

attractions determine also the specificity of the interaction [27]. For example, obligate complexes rely usually on hydrophobic attractions and transient complexes on hydrogen bonds and salt bridges [8, 14, 16, 27] although obligate interfaces are not more hydrophobic than transient ones [28, 29]. Understanding such details of the complexes is crucial in drug design; many signaling events, which are known to be disease-related, are regulated by transient protein-protein interactions [30, 31].

Considering the composition of the residues, we know that they usually differ between obligate and transient, or homo- and hetero-complexes [8, 14, 16, 27, 32-36]. Hydrophobic residues like Methionine, Phenylalanine, Proline, Alanine, Aspartic acid and Leucine (but not Isoleucine and Glycine) are commonly found in homodimer interfaces; whereas hydrophilic residues like Tryptophan, Cysteine, Histidine, Glutamine, Asparagine, Tyrosine and Serine (but not Threonine) are dominant in heterodimer interfaces [16, 36].

Interfaces are structurally similar to the cores of globular proteins [37], and many studies have shown that protein binding process is very similar to the protein folding [21-23, 38]. The structural shape of interfaces is planar and well packed, but also differs with respect to type of the interaction [14, 39]. For example, relatively large surface areas are observed in homodimer complexes compared to heterodimer complexes [16, 36].

2.1.2 Structural Features of Interfaces

Proteins interact biophysically. Therefore, fundamental determinants in protein interactions are the shape and physicochemical complementarity [6, 8, 40, 41]. Protein surfaces are full of pockets, crevices and indentations [42-45]. Some pockets which usually exist before binding are filled with the complementary protein like the key and lock model, when they associate during the interaction [46]. This is also valid for protein-ligand interactions. Size and shape of pockets are principally considered to design molecules or peptides like drugs. Therefore, structural information of the protein has a great importance in drug design. The number of small pockets on the protein surface is not very few [11, 14, 16]; on the contrary, much more than expected before. Besides, a cavity that does not pre-exist can be formed upon interacting with a small molecule [47]. In the lack of cavities, it becomes more difficult to inhibit a protein-protein interaction with a small molecule due to the flat large surface area.

In addition to the shape and size of the cavity, physicochemical properties of residues as well as their distribution in the cavity are important for the interaction. Thus, biochemical or electronic complementarity should also be considered in drug design. For example, due to the hydrophobicity of the interface, a drug to inhibit the protein-protein interaction should be designed to be hydrophobic [48-52]. But this property makes the drug less soluble, less cell permeable and less selective for its targets [48-50].

Another parameter to characterize the interface is amino acid frequency [34, 53, 54]. The amino acid propensity of the interface is similar to interior of the protein [18, 55]. The interface is usually rich of aromatic and hydrophilic residues, like Cysteine, Tyrosine, Phenylalanine and Tryptophan [18, 56]. Positively charged Arginine and Histidine also prefer to be at interface compared to both surface and core [57]. Threonine, Proline, Lysine, Glutamic Acid and Alanine are least commonly found in the interface [58]. Ofran and Rost [34] have predicted the type of the interaction with 63-100% accuracy rate just by using amino acid composition and residue-contact preferences. In another study, frequency of amino acids in transient and obligate complexes is analyzed. It is found that Glycine is more frequently seen in transient interfaces rather than the surfaces; whereas Glycine frequency is the same in interface and on surface of obligate complexes [56]. In the same study, pairwise contact preferences of the amino acid types are also investigated. Cystine-Cystine shows the highest preference of all possible contacts. This can be due to its ability to form disulfide bond. It is also found that this pair is more frequent in obligate interfaces than in transient interfaces. Moreover, acidic and basic amino acids are observed to be contacted mostly with other types of amino acids, but rarely with other amino acids with similar physicochemical properties. However, Histidine is found as an exception. It prefers to contact with itself but not with acidic amino acids. Furthermore, nonpolar amino acids prefer to contact with other nonpolar amino acids.

The other steric property of protein-protein interfaces is that they generally have an accessible surface area in the range of 1200-2000 $Å^2$ [8, 59]. This property is commonly used to distinguish biological interfaces from non-biological ones. Non-biological interfaces are formed between proteins which do not come together and contact each other actually in their physiological states [60]. Non-biological interactions are also called as crystal packing. These proteins are experimentally crystallized together; but it is mostly due to the enforcement by the crystallographic packing environment in the experiment.

Non-biological interactions mostly have smaller accessible surface areas, which are about 400-600 Å^2 [8, 61]. However, it is not trivial to distinguish biological and non-biological interactions just by considering the buried surface area. There are some examples of non-biological interactions with larger surface areas, which can be more than 2000 Å^2 [62-65], and some biological interactions, like protein-small molecule interactions (300-1000 Å^2), have smaller contact areas [66, 67]. Therefore, computational approaches to predict biological interactions utilize a combination of interface properties, like surface area, residue conservation and amino acid composition [68-72]. Although significant success has been obtained in identification of biological interfaces or distinguishing biological and non-biological interactions, distinct properties of protein interfaces have not been certainly identified yet [16, 47, 60].

2.1.3 Interfaces are Conserved throughout the Evolution

Interacting proteins find each other specifically and selectively in a crowded medium. Therefore, proteins should identify each other through complementary interaction sites. This requires the conservation of the same (or similar) residues in the right orientation for both partners of the interaction [56]. In order to preserve the interaction, residues in the active or binding sides of proteins resist undergoing a change through the evolution [73, 74]. It is proved in many studies that amino acids in interfaces are more conserved than on rest of the protein surfaces [56, 68, 75-82]. Conservation analysis of residues can be used to predict protein binding sites. The analyses can be based on the sequence [83-85], or the structure [82, 86, 87], or both sequence and the structure [83, 88, 89]. When obligate and transient complexes are compared, obligate complexes are found to have more conserved interfaces than the transient complexes [56, 81]. This supports that proteins of obligate complexes co-evolve, or proteins of transient complexes have higher ability to adapt to a change in their partners [81]. Moreover, analyses in the contact preferences of residues show that, residues prefer to contact with the other residues having a similar conservation grade [56]. In other words; highly conserved residues prefer to contact with highly conserved residues, and much variable residues prefer to contact with variable residues on the other side. Conserved residues are not randomly distributed and thus, some residues are more crucial than the others for the interaction [90]. If a contacting residue pair has an

important role for the interaction, the residues are both conserved; if not, the residues are variable at the same grade [56].

The structures of interfaces are more conserved than the global structures of the proteins [91-93]. Even if their global structures and functions are different, proteins can interact through interfaces with similar architectures [11, 94]. Structures of the interfaces have been clustered based on the domain or the whole structure. The Conserved Domain Database (CDD) provides protein domains conserved in evolution, and they are extracted via multiple sequence alignments [95, 96]. Schroeder and coworkers [97] have found 6,000 distinct types of interfaces by clustering domain interfaces. Sali and his group [98] have shown that proteins in the same SCOP families have similar binding architecture. Aloy and coworkers [99] have classified domain-based interactions of known three-dimensional structure and developed 3did web server. PPiClust provides clusters of similar 3D interface patterns in protein complexes [100]. Gao and Skolnick have found that structural space of protein-protein interfaces is close to complete and clustered interfaces of dimers into roughly 1,000 distinct types [101].

A structurally non-redundant dataset of protein-protein interfaces can be defined as three groups: Type I, Type II and Type III [1, 11]. Type I is the most common one. It includes interacting proteins with similar global structures. Type II contains proteins with similar interfaces but their global structures and functions are different. These structures are examples for the conservation of interface motifs even in the absence of global structural similarity [39, 102]. In Type III, only one side of the interface is similar and the other side is somewhat different. Hub proteins are mostly examples of this type.

2.1.4 Some Residues are Energetically more Important in the Interface: Hot Spots

All contacting residues do not equally contribute to the binding energy. Regardless of the size of the binding site, only a few residues are responsible for the majority of the total binding energy [90, 103-105]. These residues are called “hot spots”. Hot spots can be identified experimentally by Alanine Scanning Mutagenesis [76, 103]. In this method, each residue is mutated to an Alanine; and if a significant drop is observed in the energy due to the mutation, the residue is assigned as a hot spot. Hot spots can have different physicochemical properties; for example, they can be hydrophobic or polar [20, 47, 81]. Arginine, Tyrosine and Tryptophan are frequently found as hot spots; whereas, Leucine,

Serine, Threonine, Valine [104, 106] are rarely. Methionine is rarely [104] or frequently [81] found as hot spot by different studies.

Hot spots are buried and tightly packed in the three dimensional space [90]. They are located near the center of the interface and away from the solvent [104, 107]. However, they are usually found in discontinuous, discrete highly packed regions in interfaces. These clusters are called “hot regions” [90]. Furthermore, there is a strong correlation between structurally conserved residues and hot spots [81, 90, 108-110]. There are many computational studies based on their physical, biological and/or evolutionary features for the hot spot prediction. These approaches analyze combinations of hot spot features, like conservation, physicochemical properties, residue propensity, sequence profiles, accessible surface area, and contribution to binding energy [84, 88, 111-117]. Hot spots are also significant for drug design [118]. Since they are responsible for key contact potentials of the interactions, drugs are aimed to target these residues [38, 107, 119-121].

2.2 Structural Protein Interaction Networks

The protein-protein interactions can be experimentally identified by several methods such as yeast two-hybrid [122], phage display [123], protein arrays [124], and affinity purification [125] techniques. The experimental databases catalog the data gained by these techniques. These databases grow rapidly as the number of interaction–detection experiments, genome sequencing experiments and proteins with solved structures increase. Among these, the Database of Interacting Proteins (DIP) [126], the Biomolecular Interaction Network Database (BIND) [127], the BioGRID General Repository for Interaction Datasets [128], IntAct [129], the Human protein reference database **(**HPRD) [130], the Munich Information Center for Protein Sequences (MIPS) [131], the Human Protein Interaction Database **(**HPID) [132], the database of Protein Structural Interactome map (PSIbase) [133] and the database of protein domain interactions **(**DOMINE) [134] list experimentally determined protein-protein interactions. From these, PSIbase and DOMINE include the binding site information, i.e. they indicate where two proteins interact and hence can assist in structural studies.

The networks of protein-protein interactions can provide understanding of the global organization of cellular processes; however, they still lack structural and chemical characteristics of each interaction. At this point, knowledge of protein structural

information is a crucial asset. This concept is illustrated in **Figure 2.1**; the figure displays a schematic representation of a classical network and a structural interaction network. The protein labeled P1 has two different binding regions, B11 and B12: through the B12 region, the protein interacts with P3; and through the B11 region it can interact with three partners. Thus, the interactions of P1 with P2, P4 and P6 exclude each other; that is, they cannot occur at the same time. On the other hand, the interaction with P3 is simultaneously possible. In **Figure 2.1C**, three possible interactions are displayed. This figure indicates that P1 can be used in three different complexes. This information cannot be obtained from the abstract network shown in **Figure 2.1A**.

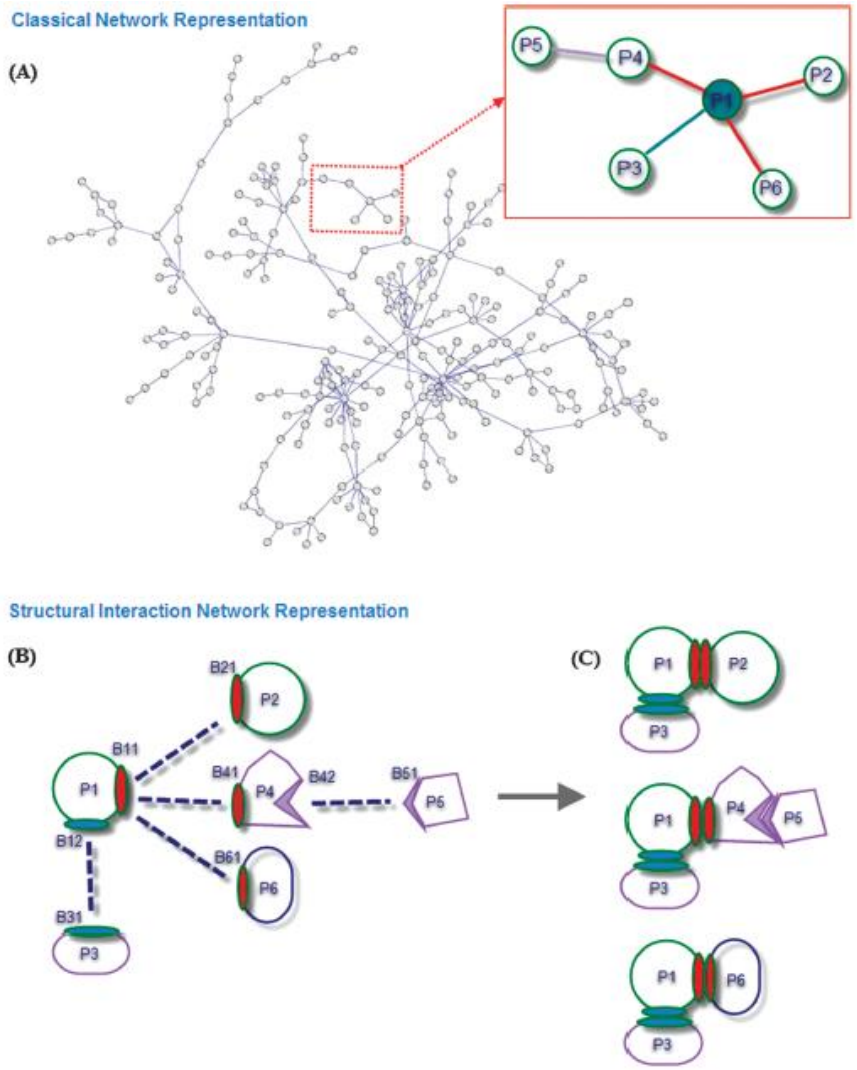

Figure 2.1 Structural Interaction Network Representation

Protein structures can be determined by X-ray crystallography, NMR spectroscopy [135] and cryo-electron microscopy (EM) [136] at several resolutions. X-ray crystallography is the most widely used technique and gives a static information about protein structure. NMR data is obtained in solution where many structural, thermodynamic and kinetic properties can be analyzed. This method is limited by the size of the protein complex. Due to the crystallization problems of transient complexes, they are usually underrepresented in PDB. Cryo-EM method is useful for visualization of the transient complexes although detailed positions of the subunits in a protein complex are not distinguishable.

Among the structural network studies, the pioneering work of Aloy and Russell [137] illustrated how three dimensional protein structures can be used to infer molecular details of interactions in a network. In another study, Kim et al. have combined structural modeling with network analysis. They have mapped the interactions to known structures of interfaces and distinguished the interfaces of each interaction. They have classified the network hubs as single- or multi-interface hubs. The former was found to have at most two distinct binding interfaces and are enriched in signaling proteins, whereas the latter has more than two binding interfaces and tend to be members of large and stable complexes [138].

Structured networks are also utilized to understand diseases. Dawelbait et al. [139] constructed a network related to pancreatic cancer by combining known interactions and structure-based interaction predictions. They predicted 40 novel interactions that are specific to pancreatic cancer. In one study, cancer-associated signaling pathways and their physical protein-protein interactions are analyzed with the goal of providing insights into three-dimensional structure-function relationship [140]. In another study, protein interface structures are integrated to human cancer protein interaction network and interface analysis revealed that cancer-related proteins have smaller, more planar, more charged and less hydrophobic binding sites than non-cancer proteins, which may indicate low affinity and high specificity of the cancer-related interactions [141]. Additionally, integrating interfaces into networks can provide timing of proteins' interactions; whether they are simultaneous or exclusive [60, 142, 143].

Although currently there occurs some limitations such as incompleteness of protein interaction networks and structural data, with the complete knowledge of pairwise protein-

protein interactions and the structures of protein associations, it will unquestionably be possible to construct high-quality structural protein interaction networks, which will lead to a better understanding of molecular mechanisms of living systems.

2.3 Structural Modeling of Protein Associations

Structural modeling of protein-protein interactions can be achieved via docking or template-based strategies.

Docking is the procedure to find the best bound state for given 3D structures of two (or more) proteins. Considering that protein surfaces are flexible and there are many potential ways in which proteins can interact, the docking problem is difficult. There are several docking algorithms so far [144-149]. Such algorithms first obtain the candidate conformations with a fast search algorithm, and then rank these conformations by a high quality scoring function to find the near native model. For scoring functions, geometric and chemical complementarity, electrostatic, van der Waals forces or knowledge-based potentials are frequently assessed [150]. However, scoring functions are not fully optimized yet. In addition, flexibility of the proteins should be considered while searching for the native state of the protein complexes. Refinement algorithms are therefore developed to re-assess the rigid-body docking solutions and re-rank the modeled associations [151-153]. Although docking at large-scale is computationally very expensive, in the first large-scale docking effort, Aloy and his collegues have managed to model 3000 putative protein complexes for yeast protein network [154].

Template-based structural modeling methods have gained more attention with the increase in availability of structural data. Using a known protein complex as the template, such approaches can utilize sequence or structural similarity to model protein-protein associations. There are homology-based and interface-based strategies. In one study, the sequence homologues of the known protein complexes are searched and predicted interaction between homolog proteins are scored the using empirical potentials derived from known protein interactions [137]. If the score of a predicted complex is high enough, the homologous protein pair associates in a similar way with the template complex. This method is used to model protein complexes in yeast interaction network and fibroblast growth factor/receptor system. Then, a web server, InterPreTS, is designed to predict protein interactions using this method for a given set of protein sequences using Blast2

search tool to find the homologues [155]. Another homology-based method is named as Multiprospector, which is based on multimeric-threading [156] and uses a template library consisting of protein complexes. In the first step of the algorithm, each target sequence is assigned to a protein structure in the template library. In the second step, multi-chain threading is applied and each target protein pair is assigned to a group of quaternary structures. The quality of the predictions is evaluated by the interfacial potentials and Z-scores. Native interactions are discriminated from artificial ones with accuracy of 90% [157]. In another study, structural similarity of overall protein structures is considered and domain information is integrated to search putative protein complexes. According to matching of the domains, protein interactions are predicted and scored using statistical potentials that are derived from side chain-side chain contacts. The method was shown to distinguish the non-native contacts with an accuracy of 0.99 [158].

In nature, the number of protein folds is limited [159]. Since folding and binding are similar processes, the number of distinct interface motifs is also limited in nature [160]. Therefore, protein pairs can interact via similar interface architectures even though their global folds are different [11, 94]. Based on this origin, PRISM algorithm, which is the first strategy to model protein associations based on interface motifs, is developed [20, 142, 161]. The rationale is: if two complementary partners of a template interface are similar to the surface of two target proteins, these two proteins are principally interact with each other using this template architecture. This method is used to generate a structural interaction map of cancer proteins [141] and to show the multi-face nature of the hub proteins [60]. Following this idea, similar algorithms such as ISearch, which uses domain-domain interfaces as the template, are developed [102, 162, 163]. These template-based methods decrease the solution space by limiting the possible orientations. Therefore, compared to docking strategies, such methods are computationally much faster and efficient especially on a large-scale. As the diversity in template sets increases, such methods will become more popular and useful in the future.

2.4 Microarray Gene Expression Profiling Data

Microarray experiments generate a wealth of expression data providing important insights into several biological processes. Using microarrays, expression patterns that differ between diseased and healthy samples can be identified and biomarkers can be labeled

[164]. However, there are some limitations: from the biological standpoint, mRNA expression level does not always correlate with the protein levels [165]; and from the technical standpoint, there would be experimental noise due to the differences in experimental setup [165, 166]. To overcome such problems, integration of microarray data with additional data sources can be very helpful. Protein-protein interaction networks together with microarray data were illustrated to be useful in interpreting gene expression data by improving sample classification [167, 168] and detection of differentially expressed genes [169-171]. In their pioneering study, Ideker and collegues have proposed a method to detect active subnetworks in a subset of the profiled samples [172]. Later on, this approach has been improved by several groups [173-176] and utilized to identify subnetworks co-expressed across all samples and to extract functionally coherent co-expressed gene sets [5]. Main objective has been to derive biologically interesting subnetworks of interpretable size from large-scale protein-protein interaction data. Since cellular function or a phenotype is regulated by several proteins rather than being a result of a single gene [177], combined analysis of expression profiles and protein-protein interactions allows the detection of previously unknown dysregulated modules in interaction networks, which will be especially immensely useful in disease states to identify functional associations of disease-related genes.

Chapter 3

IMPORTANCE OF PROTEIN STRUCTURAL INFORMATION IN DRUG DESIGN AND DISCOVERY

Protein-protein interactions have a key role in regulating many biological processes, cellular and signaling pathways. The dysfunction of these pathways due to the alterations in protein-protein interactions may lead to several diseases such as cancer and neurological disorders. Therefore, protein-protein interactions are widely considered as drug targets in disease states [51, 84, 178]. In this chapter, we highlight the advantages of inclusion of protein structural details and illustrate how structural modeling and binding site analysis would assist in drug design and discovery. In the last section, we introduce the concept of allostery and provide structural models of allosteric signaling proteins.

3.1 Including Protein Structural Information into Interaction Networks: Advantages in Drug Design

Structural networks have been recently utilized to understand diseases. Dawelbait et al. [139] constructed a network related to pancreatic cancer by combining known interactions and structure-based interaction predictions. They predicted 40 novel interactions that are specific to pancreatic cancer. In one study, cancer-associated signaling pathways and their physical protein-protein interactions are analyzed with the goal of providing insights into three-dimensional structure-function relationship [140]. In another study, protein interface structures are integrated to human cancer protein interaction network and interface analysis revealed that cancer-related proteins have smaller, more planar, more charged and less hydrophobic binding sites than non-cancer proteins, which may indicate low affinity and high specificity of the cancer-related interactions [141]. Additionally, integrating interfaces into networks can provide timing of proteins' interactions; whether they are simultaneous or exclusive [60, 142, 143].

Besides inferring the nature of disease-related interactions and their role in the network, comparison of protein-protein interfaces across an interactome can also assist in identifying

drug targets or drugs targeting multiple proteins to block parallel pathways in a network [179]. With structural analysis drug binding pockets can be identified and compared with binding pockets of other proteins in the network, which could eventually lead to discovering candidates for drug-targetable protein-protein interactions [180]. Since the number of distinct binding motifs is limited in nature [160], structurally different proteins can share similar interface architectures [11]. Therefore, even though two proteins are structurally different and belong to different pathways in the interactome, if their binding pockets are similar, a drug can bind to both. While this is possible, binding also depends on the chemical nature of the drug ligand itself [181].

Although there occurs some limitations such as incompleteness of protein interaction networks and structural data, combining protein interfaces with interaction networks could guide identification of all target proteins that are influenced by a drug, either positively or negatively. Considering the position and role of these target proteins in the network could help to predict side-effects or to discover a treatment of a new disease. The latter is called drug repurposing, which means finding new uses of old drugs [182], is gradually gaining popularity since *de novo* drug design is laborious and very costly [183]. Provided that protein structures and binding sites are known in a disease-associated network, predicting druggable proteins and the resulting effects will be less demanding. Here, this concept is visualized in **Figure 3.1** by an example; the celecoxib (Celebrex), which is a nonsteroidal anti-inflammatory drug that is known to bind cyclooxygenase-2. Weber et al. [184] showed that celecoxib also binds to a totally unrelated protein; carbonic anhydrase, which has a structurally similar binding site to that of cyclooxygenase-2. This finding reveals a new role of celecoxib in the treatment of glaucoma and possibly for cancer [184].

3.2 Importance of Hotspot Residues in Drug Design and Discovery

Drugs targeting protein-protein interactions, ultimately head protein interfaces where two proteins come into contact. Understanding the details and principles of protein interfaces is, therefore, immensely essential to develop efficient strategies in drug design. At protein interfaces there are some critical residues that account for the majority of the binding energy called hot spots [104]. A hot spot is defined as a residue that, when mutated to alanine, leads to a dramatic decrease in binding free energy ($\Delta\Delta G_{binding} > 2$ kcal/mol) [76, 104]. Since these residues are more critical than others to the stability of the complex,

targeting the hot spots may lead to improved inhibition of protein-protein interactions and to designation of novel compounds.

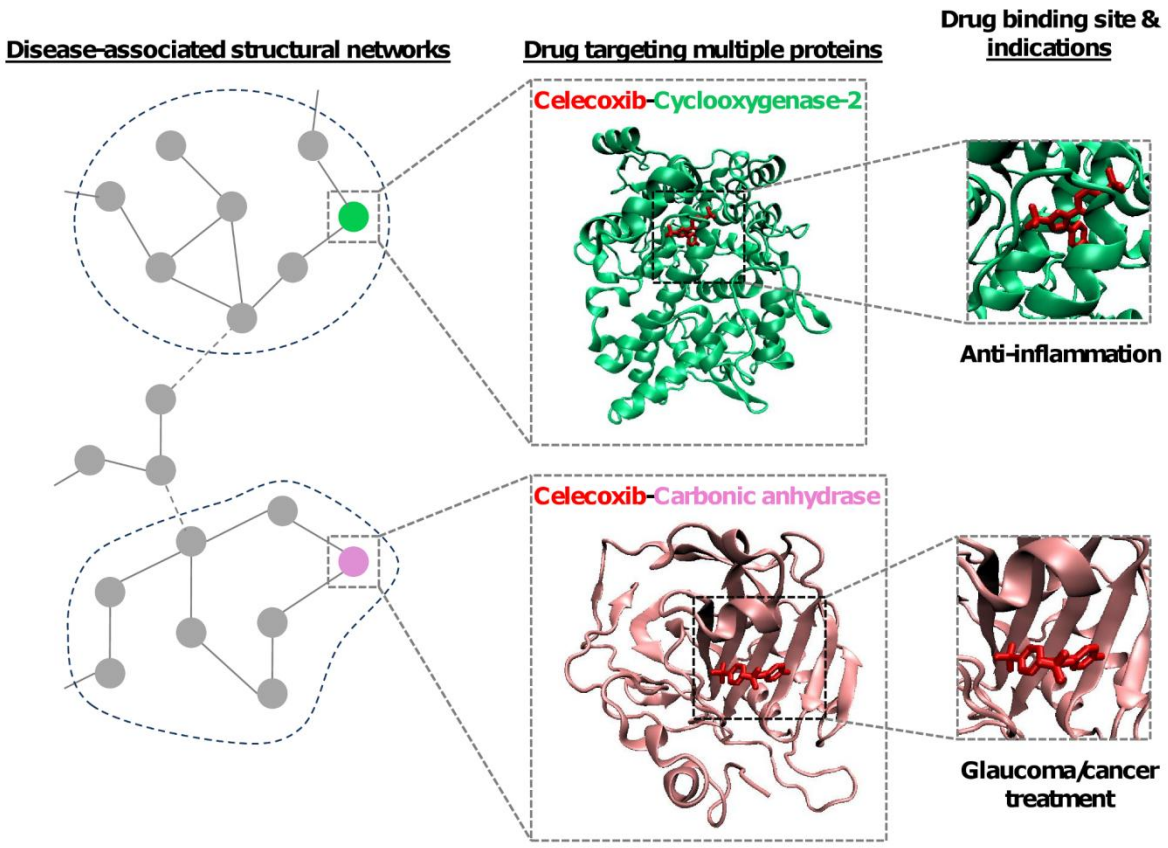

Figure 3.1 Protein structures are integrated into disease-associated networks. Identifying protein interfaces and binding pockets in disease-associated structural network can assist in finding drug targets and role of these proteins in disease. As an example, a nonsteroidal anti-inflammatory drug, which is known to target multiple proteins, is visualized. It is interesting that although cyclooxygenase-2 (in white color, PDB code: 3l1n) and carbonic anhydrase (in gray color, PDB code: 1oq5) are structurally different proteins; all-alpha and all-beta structure, respectively, their binding sites in contact with celecoxib share structural and physico-chemical similarities [184]. Latterly found interaction with carbonic anhydrase reveals the new role of celecoxib in the treatment of glaucoma and possibly for cancer [184].

Here, in **Figure 3.2**, we present an example illustrating how structural modeling of E2-E3 associations and identifying hot spot residues in human ubiquitination pathway would be beneficial to drug discovery. c-Cbl is an E3 ubiquitin ligase that attenuates signaling through poly- or monoubiquitination of several activated receptor tyrosine kinases (such as Flt-3, c-kit, and M-CSF) and other tyrosine kinases of the Scr family [185]. The RING

domain has a central role in c-Cbl function because its deletion or disruption abolishes the function of c-Cbl [186]. From its RING domain, c-Cbl binds to an E2 and mediates ubiquitin transfer from E2 to the target substrates. Structure of c-Cbl interacting with one of its E2 partners, UBE2L3, is deposited in Protein Data Bank (PDB) (pdb code: 1fbv) [187]. Two loop regions, L1 and L2 of UBE2L3 are in contact with the α-helix and zinc-chelating loops of the RING domain. The central F63 residue in loop L1, P97 and A98 in loop L2 of UBE2L3 mediate its binding to c-Cbl. The c-Cbl linker region interacts with α-helix 1 of UBE2L3 [187]. c-Cbl ligase-UBE2L3 complex showing the critical binding regions (α-helix 1, loop L1 and L2) of UBE2L3 is visualized in **Figure 3.2A.**

c-Cbl is also reported to interact with another E2, UBE2D1, and ubiquitinate epidermal growth factor receptor [188], although the structure of the complex is not available as yet. Using Prism algorithm, we modeled c-Cbl-UBE2D1 complex based on a known interface of another E2-E3 complex; UBE2D2-CNOT4 (pdb code: 1ur6:AB) [189]. We observe that UBE2D1 (pdb code: 2c4p:A) binds to c-Cbl (pdb code: 1fbv:A) mainly through its α-helix 1, loop L1 and L2 regions. Earlier work suggested that mutations in the c-Cbl RING domain are associated with acute myelogenous leukemia and myeloproliferative neoplasms and the impairing the degradation of tyrosine kinases is an important mechanism in cancer [190]. In particular, on the RING domain of c-Cbl, the substitution of cysteine or arginine residues at position 384 (C384Y), 404 (C404S), and 420 (R420Q) are observed in patients with acute myelogenous leukemia [190]. Additionally, Ile383 and Trp408 of c-Cbl were shown to have a critical role in E2 binding [187] and the mutation of Trp408 to alanine reduces c-Cbl's affinity for the E2 and eliminates its ubiquitin-ligase activity in vitro [191]. Interface analysis of our c-Cbl-UBE2D1 model indicates that these critical residues; I383, C384, C404 and W408, correspond to computational hotspots predicted by Hotpoint server [192], implying their importance in binding and function of c-Cbl. Such structural modeling of E2-E3 associations could help in understanding E2-E3 selectivity and discovering drug candidates targeting E3s. **Figure 3.2B** illustrates the modeled c-Cbl-UBE2D1 complex and the critical residues in binding.

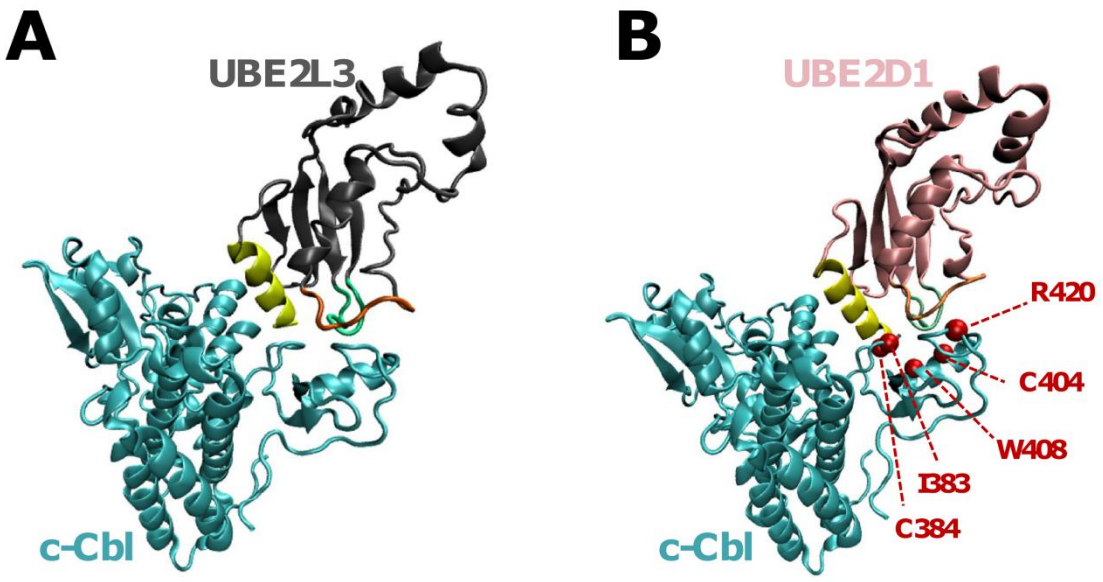

Figure 3.2 Structural Modeling of c-Cbl E3 ubiquitin ligase-E2 interactions
(A) c-Cbl interacts with UBE2L3 obtained from PDB (pdb code: 1fbv:AC). α-helix 1, loop L1 and loop L2 regions of UBE2L3 are displayed in yellow, green and orange color, respectively. Two loop regions, L1 and L2 of UBE2L3 are in contact with the α-helix and zinc-chelating loops of the RING domain. The central F63 residue in loop L1, P97 and A98 in loop L2 of UBE2L3 mediate its binding to c-Cbl. The c-Cbl linker region interacts with α-helix 1 of UBE2L3 [187]. **(B)** Modeled c-Cbl-UBE2D1 complex. Although the interaction was reported earlier [188], binding details and the structure of the complex was not available. c-Cbl residues labeled in red color was observed to be critical: C384, C404, R420 were mutated in patients with acute myelogenous leukemia [190]. Additionally, Ile383 and Trp408 of c-Cbl were shown to have a critical role in E2 binding [187]. These residues correspond to computational hotspots predicted by Hotpoint server indicating the usefulness of structural modeling of E2-E3 associations

3.3 Allosteric Proteins in Signaling Pathways

Allostery is a cooperative event, up- or down-regulating protein activity. From the functional standpoint, the key role of allosteric events is to increase binding selectivity at the target site [193]: binding to even slightly different allosteric effectors or at different allosteric sites can enhance specificity. From the pharmacological standpoint, allosteric effects can adversely affect protein function: disease-related mutations often lie on major allosteric routes [194, 195]. The efficacy of drugs that bind residues on major propagation pathways can be expected to be higher. Hence, identification of major pathways in the ensemble is an important goal.

Signaling pathways are complex and dynamic [196], and are important for identifying possible therapeutic targets. Ras proteins mediate signaling pathways that control cell growth and differentiation. They act as molecular switch by cycling between active

guanosine triphosphate (GTP)-bound and inactive guanosine diphosphate (GDP)-bound states. They are activated by nucleotide exchange factors Son of sevenless (Sos) and Ras guanine nucleotide releasing factor 1 (RasGRF1) upon conversion of GDP-bound Ras to GTP-bound. Their large conformational changes are best described by the population shift model rather than induced fit [197]. In addition to the active and inactive states highly populated intermediates are also sampled [198]. In its active state, Ras can interact with effectors in signaling cascades, whereas inactive Ras cannot [199]. In the mitogen-activated protein kinase (MAPK) signaling pathway, Ras is first activated by Sos, then it binds Raf protein kinases (A-Raf, B-Raf and C-Raf). Although there are binding studies on Ras interaction with C-Raf [200, 201], little is known on binding of B-Raf to Ras. Since B-Raf is important in many cancer types, details of its interactions are essential for drug design. Here, using the active and inactive conformations, we search for possible interactions of Ras with B-Raf using Prism [20]. Results indicate that active Ras can favorably bind to B-Raf while inactive Ras cannot due to steric effects. **Figure 3.3A** illustrates the conformational change upon activation. **Figure 3.3B** displays the Prism-predicted Ras - B-Raf interaction. When binding to B-Raf, Ras uses switch I (residues 30-38) and switch II (residues 60 to 76) regions which correspond to the most significant conformational change. The figure indicates that without this conformational change, inactive Ras - B-Raf interaction is unfavorable.

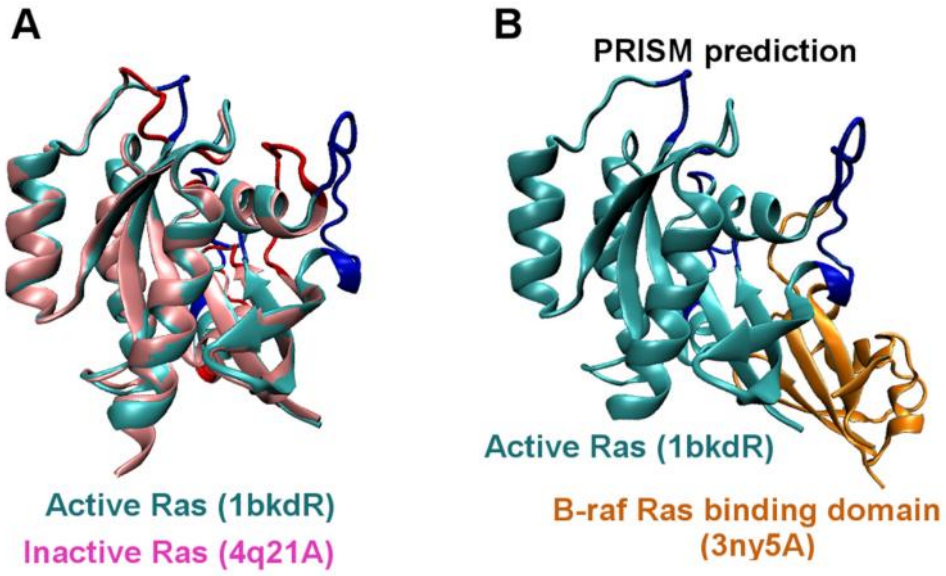

Figure 3.3 Allosteric Ras protein in MAPK signaling pathway. **(A)** Visualization of the conformational change in Ras protein upon activation by Sos. Inactive and active allosteric Ras protein structures (obtained from the PDB) are shown in pink and cyan color, respectively. The superposition is based on matched residues with the distance between superimposed C^{α} atoms ≤ 2 Å. The conformational changes (unmatched residues) are highlighted in red and blue, respectively, for inactive and active Ras. Conformational changes correspond mostly to residues from switch I (residues 30-38) and switch II (residues 60 to 76) Ras regions. **(B)** The interaction between activated Ras (PDB code: 1bkdR) and Ras binding domain of B-Raf (3ny5A) is predicted by Prism [20]. Binding site corresponds to switch I and switch II regions. B-Raf can bind to activated Ras favorably whereas it cannot bind to the inactive structure.

Chapter 4

HUMAN PROTEOME-SCALE STRUCTURAL MODELING OF E2-E3 INTERACTIONS EXPLOITING INTERFACE MOTIFS

Ubiquitination is crucial for many cellular processes such as protein degradation, DNA repair, transcription regulation and cell signaling. Ubiquitin attachment takes place via a sequential enzymatic cascade involving ubiquitin-activation (by E1 enzymes), ubiquitin-conjugation (by E2 enzymes), and ubiquitin substrate-tagging (by E3 enzymes). E3 ligases mediate ubiquitin transfer from E2s to substrates and as such confer substrate specificity. Although E3s can interact and function with numerous E2s, it is still unclear how they choose which E2 to use. Identifying all E2 partners of an E3 is essential for inferring the principles guiding E2 selection by an E3. Here we model the interactions of E3 and E2 proteins in a large, proteome-scale strategy based on *interface structural motifs*, which allows elucidation of 1) *which* E3s interact with which E2s in the human ubiquitination pathway; and 2) *how* they interact with each other. Interface analysis of E2-E3 complexes reveals that loop L1 of E2s is critical for binding; the residue in the sixth position in loop L1 is widely utilized as an interface hot spot and appears indispensible for E2 interactions. Other loop L1 residues also confer specificity on the E2-E3 interactions: HECT E3s are in contact with the residue in the second position in loop L1 of E2s; but this is not the case for the RING finger type E3s. Our modeled E2-E3 complexes illuminate how slight sequence variations in E2 residues may contribute to specificity in E3 binding. These findings may be important for discovering drug candidates targeting E3s, which have been implicated in many diseases.

4.1 Introduction

Protein modification by ubiquitin (ubiquitination) is a key mechanism for controlling many cellular processes such as protein degradation, DNA repair, signal transduction, transcription, immunity, endocytosis and cell death. Ubiquitination is achieved by a

sequential enzymatic cascade of ubiquitin-activating (E1), ubiquitin-conjugating (E2), and ubiquitin-ligating (E3) enzymes [202]. In the human genome, there are only two known E1s [203], which are conserved across different species while there are multiple E2s and many more E3s. Ubiquitin is activated by an ATP-dependent E1 and transferred to E2. E3s recognize specific target substrates and catalyze the ubiquitin transfer from the E2 to the substrate. Ubiquitin can be conjugated to the substrate as a monomer on one (monoubiquitination) or more substrate lysines (multiubiquitination) or as a polymer (polyubiquitination) by forming ubiquitin chains [204]. There are three types of E3 ligases based on the protein families they belong to: Homologs of E6AP Carboxy Terminus (HECT), Really Interesting New Gene (RING), and the UFD2 homology (U-box) family proteins. HECT is a domain of ~350 amino acids. It has a bilobal structure, in which the N-terminal lobe contains the E2-binding site, and the C-terminal lobe confers catalytic activity. The conserved Cys residue in the C-terminal lobe forms thioester bonds with ubiquitin [205]. The RING finger domain is formed by a short motif rich in histidine and cysteine residues that coordinate zinc atoms in a cross-brace structure, characterized by a central α-helix and variable-length loops separated by small beta strands [206]. The U-box domain constitutes a relatively small family of E3s and is similar to the structure of the RING domain with the exception that it lacks the conserved histidine and cysteine residues [207]. The ubiquitination mechanisms for these three E3 classes differ. For HECT domain E3s, the ubiquitin is first transferred from E2 to the active site residue of HECT E3, with subsequent transfer from E3 to the substrate protein. For RING and U-box type E3s, ubiquitin is directly transferred from E2 to the substrate without an E3 intermediate linkage. The E2 family is characterized by the presence of a highly conserved 150-200 amino acid catalytic core domain, which consists of four α-helices, a short 3_{10} helix and a four-stranded, antiparallel β-sheet [208, 209] (**Figure 4.1A**). The beta sheet forms a central region bordered by helices and there are two loop regions, L1 and L2 that are located C-terminal to β-strands 1 and 3 [209-211]. These loop regions show a high level of flexibility and are involved in E3 selection and binding [209]. E2 families are distinguished by minor sequence differences in the core and amino and/or carboxyl terminal extension domains [204]. Based on these, they are classified into four groups. Class I E2s consist only of the catalytic core domain. In addition to the core domain, Class II and Class III E2s contain

amino terminal and carboxyl terminal extensions, respectively. Class IV E2s contain both amino and carboxyl terminal extensions [211].

Currently, there is a limited number of known E2s (~40 in human); however the number of known E3s is increasing rapidly (~500 or more have been proposed to exist in human) [212-214], which suggests that one E2 can recognize several different E3s [215]. E2 and E3 proteins are known to function combinatorially; however, the principles determining the E2-E3 selectivity are unclear and in many cases it is not known which E3s interact with which E2s [216]. The first structural clues of E2-E3 interaction specificity were obtained from the crystal structure of the E3 ligase c-Cbl RING and UBE2L3 [187] (**Figure 4.1B**). Two loop regions, L1 and L2 of UBE2L3 are in contact with the α-helix and zinc-chelating loops of the RING domain. The central F63 residue in loop L1, P97 and A98 in loop L2 of UBE2L3 mediate its binding to c-Cbl. In addition, the c-Cbl linker region interacts with α-helix 1 of UBE2L3 [187]. In another study, CNOT4 RING finger binding to the ubiquitin-conjugating enzyme UBE2D2 was found to be highly selective [217]. Charge-alteration of residues E49 of CNOT4 and K63 of UBE2D2 result in altered-specificity of a functional E2–E3 enzyme pair [217]. In one experimental approach, two-hybrid experiments were performed to identify E2 partners of Brca1 ubiquitin ligase. Brca1 was found to interact with multiple different E2s and to possess unique ubiquitin-transfer properties depending on the E2 used [218, 219]. In another study, *in vivo* cross-linking methods were used to identify HRD1 (HMG-CoA reductase degradation) ubiquitin ligase interactions [220]. Two large-scale studies have recently addressed this E2-E3 identification problem: Wijk et al. [221] performed a global yeast-two hybrid screen and uncovered over 300 high quality interactions; Markson et al. [222] combined yeast two-hybrid screens with homology modeling methods to generate a map of human E2-E3 RING interactions.

Although these studies identified new E2-E3 pairs, the nature and structural details of the interactions in the ubiquitin system are lacking. Here, we aim to model the human E2-E3 interactions on a large, proteome-scale and to obtain an insight into their interaction specificity. To carry out this study, we have used Prism [20, 142, 161], which employs a highly efficient strategy to predict protein associations based on interface structural motifs. The Prism rationale argues that if any two proteins contain regions on their surfaces that are similar to complementary partners of a known interface, in principle these two proteins can interact with each other through these regions. This knowledge-based strategy, which

utilizes structural and evolutionary similarity, is made more physical and biologically relevant by including flexibility and energetic assessment in the modeling. This is achieved by using FiberDock [153], a flexible docking refinement server. Using Prism, we have constructed a human structural E2-E3 network consisting of 107 predicted interactions among 22 E2s and 16 E3s. 36% of our predicted interactions were reported in earlier studies as interacting pairs; however, how they interact has been unclear. We first observed that E3 proteins could interact with multiple E2s and likewise E2 proteins could interact with multiple E3s, which is expected. However, analysis of the modeled interfaces of E2-E3 putative complexes revealed some structurally conserved residues which are present in almost all interfaces and as such are likely to be indispensible for E2 binding. Comparison of the E2-HECT domain E3 and E2-RING domain E3 interfaces suggests that the E2 loop L1 residues confer specificity in binding to different E3s. The structural E2-E3 network in this study, together with interface analysis, provides a resource for future studies of ubiquitination and E2-E3 selectivity.

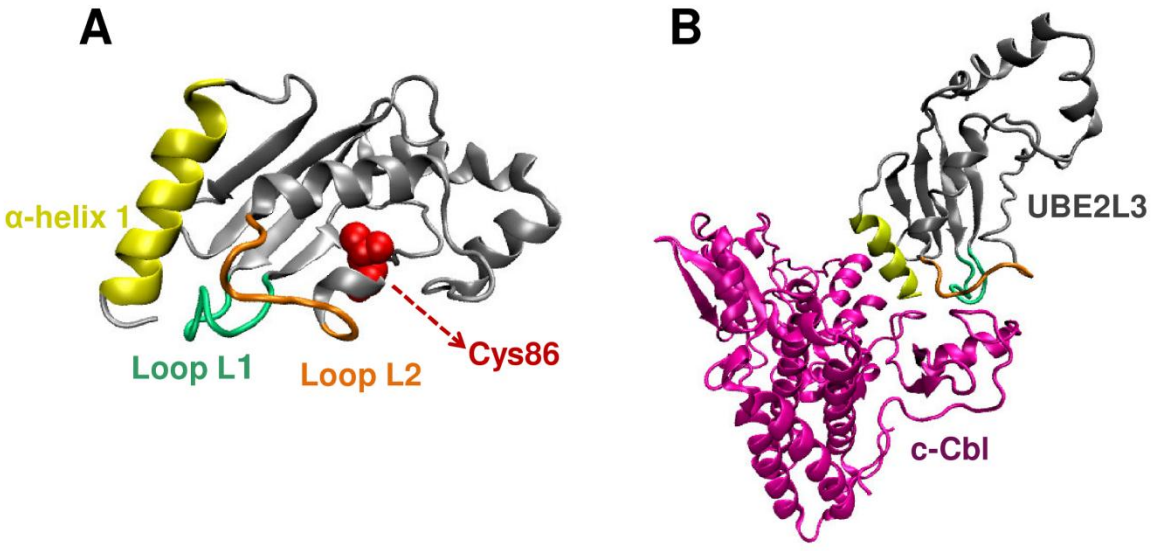

Figure 4.1 Structural representation of E2 and E3 proteins.(A) A member of E2 family, UBE2D1 (PDB code:2c4p:A) is visualized. The catalytic core domain consists of 150 residues. α-helix 1 and highly flexible regions; loop L1 and L2, which are involved in E3 selection and binding, are displayed in yellow, green and orange color, respectively. Conserved catalytic cysteine residue, to which ubiquitin is attached, is shown as red spheres. (B) The interaction between an E3 ligase c-Cbl RING and UBE2L3 is shown. Loop L1 and L2 of UBE2L3 interact with α-helix and zinc-chelating loops of the c-Cbl RING domain and α-helix 1 of UBE2L3 is in contact with c-Cbl linker region [187].

4.2 Methodology

4.2.1 Template And Target Dataset

In this study, we predict and model complexes based on the known interfaces in a template dataset. To construct the template dataset, we extract all known E2-E3 complexes in the ubiquitination pathway from the PDB [223]. There are 9 available E2-E3 complex structures which are listed in Supplementary Material.

The target dataset contains the E2 and E3 protein structures among which we want to uncover possible interactions. The list of ubiquitin ligases (E3) and ubiquitin conjugating enzymes (E2) related to the human ubiquitination pathway are obtained from the KEGG database [224] and available 3D structures are extracted from the PDB. There are 24 E2 proteins, 20 RING finger type E3, 9 HECT type E3 and 3 U-box type E3 proteins with three dimensional protein structures. Among these, Prism algorithm predicts interactions between 22 E2 and 16 E3 proteins.

4.2.2 The Prediction Algorithm

The prediction algorithm is composed of four consecutive steps (**Figure 4.2**): extraction of the surface of target proteins, structural alignment, collision check and flexible refinement. In the first step, surface regions of the target proteins are extracted using Naccess [225] based on the relative accessible surface area of the residues. If the relative surface accessibility of a residue is more than 15%, then it is labeled a surface residue. In the second step, each interface in the template dataset is split into its chain components. Using the MultiProt engine [226], our algorithm searches whether the target surfaces are structurally similar to complementary partners of a template interface. At least 40% of the residues of template chains should be matched to the target surface residues to pass to the next step. If the template chains contain less than 50 residues, this threshold is 60%. In addition to the structural similarity, evolutionary similarity between the template and target surfaces is assessed; at least one hot spot in each template partner should match with the target surface. In the third step, each target protein is transformed onto the corresponding template interface to form the complex structure. After the transformation, the colliding residues are checked; if two partners have more than five spatially-colliding residues, then the match is eliminated. At this step, side chain collisions are not considered and left to the last step for a more accurate treatment. The last step is the flexible refinement (backbone

and side chain) of the predicted complexes using FiberDock [153] to resolve the steric clashes and rank the predicted complexes according to their calculated global energy. FiberDock uses both low and high frequency normal modes and models the backbone and side-chain movements according to the binding van der Waals forces between the receptor and ligand. We assign the larger protein chain to be the receptor and the smaller to be the ligand. In the side chain optimization, only clashing interface residues are considered as movable and 20% of clashes between the side-chain atoms are allowed. Following the refinement process, FiberDock calculates a global energy for each predicted complex. The putative complex structures with a global energy less than -10 kcal/mol are included in the human structural E2-E3 interaction network.

4.2.3 Interface Analysis

Interface regions of putative E2-E3 complexes are analyzed using the Hotpoint server [192], which uses an efficient method to determine computational hot spots based on conservation, solvent accessibility and statistical pairwise residue potentials of the interface residues. Hot spots are critical residues at the interface which account for the majority of the binding free energy [104]. Structure-based multiple sequence alignment of human E2s are obtained from Christensen et al. [219] and interface and computational hot spot residues are displayed. Computational mutagenesis analysis is performed using FoldX algorithm [227]. First, the putative complex structures are subjected to an optimization procedure using the repair function of FoldX. Next, quantative estimations of the binding affinities of the wild-type and mutants are obtained by the PositionScan function.

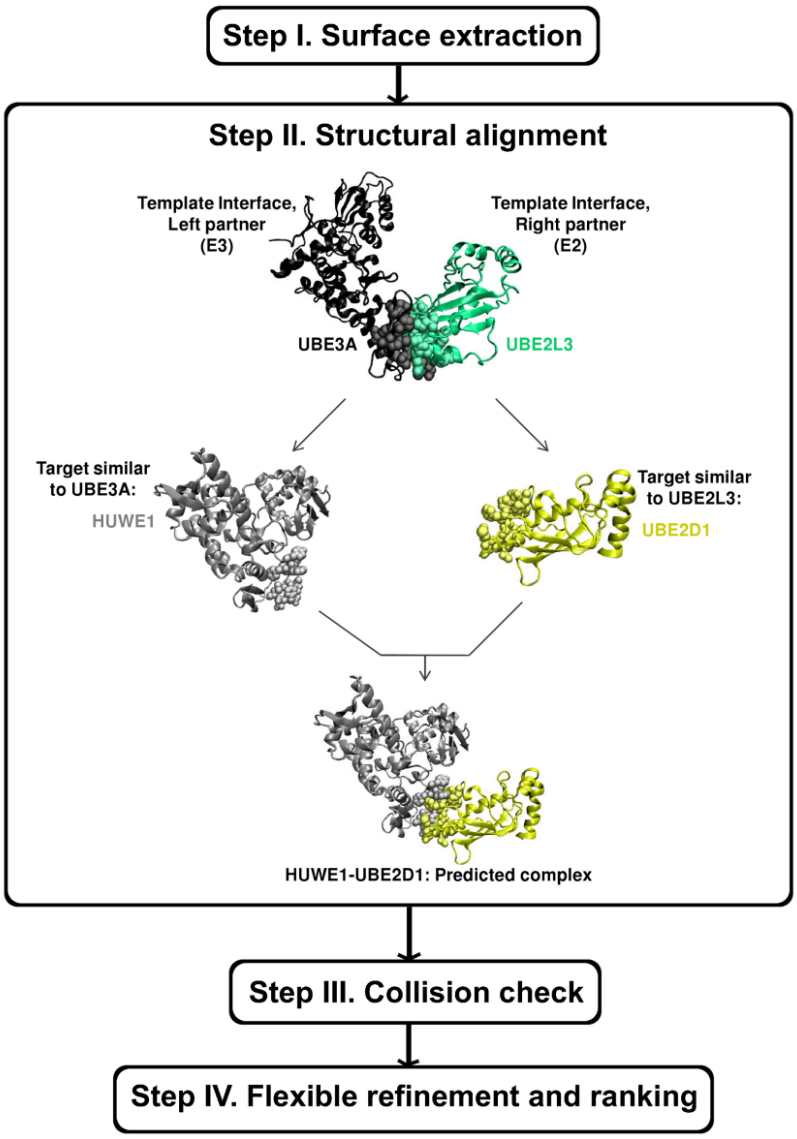

Figure 4.2 Schematical illustration of the Prism algorithm. In the first step, surface regions of target proteins are extracted. In the second step, each template interface is split into the complementary partner chains and the partners are structurally aligned with target protein surfaces by the MultiProt engine [226]. This concept is visualized here: Ube3a-UBE2L3 (PDB code: 1c4z:AD) is one of the complexes in our template interface set. When target proteins are structurally aligned with Ube3a and UBE2L3, an E3 ligase Huwe1 (PDB code: 3g1n:A) is found to have a structurally similar binding region to that of Ube3a. Likewise, one of the E2 proteins; UBE2D1 (PDB code: 2c4p:A) has a similar binding region to that of UBE2L3. Then, in principle, these two proteins Huwe1 and UBE2D1 can interact with each other through these regions. Predicted complexes are assessed in the third step: target proteins are transformed onto the template interface and if the residues of the target partners collide, these pairs are eliminated. In the last step, to obtain biologically more relevant interactions, flexible refinement of the rigid docking solutions of MultiProt is performed using FiberDock [153] and the predicted complexes are ranked according to the calculated global energy.

4.3 Results

Interactions between ubiquitin-conjugating (E2) and ubiquitin-ligating (E3) enzymes are essential for ubiquitination (**Figure 4.3)**. Here we computationally model E2-E3 interacting pairs in the human proteome on a broad scale to obtain the interaction network between all E2s and E3s whose structures have been determined, and in particular, to obtain structural insight into E2-E3 selectivity based on the large set. Modeling the structural network provides data not only relating to which E2s interact with which E3s; but also to how they interact. Analysis of these interactions may yield patterns which distinguish between the E2 and E3 interaction classes.

In the following sections, first we explain how we constructed the human structural E2-E3 interaction network and present its topological characteristic. Next we provide an analysis of E2-E3 interfaces. We then describe the patterns that we observed and hypothesize that they may play a role in E2-E3 specificity.

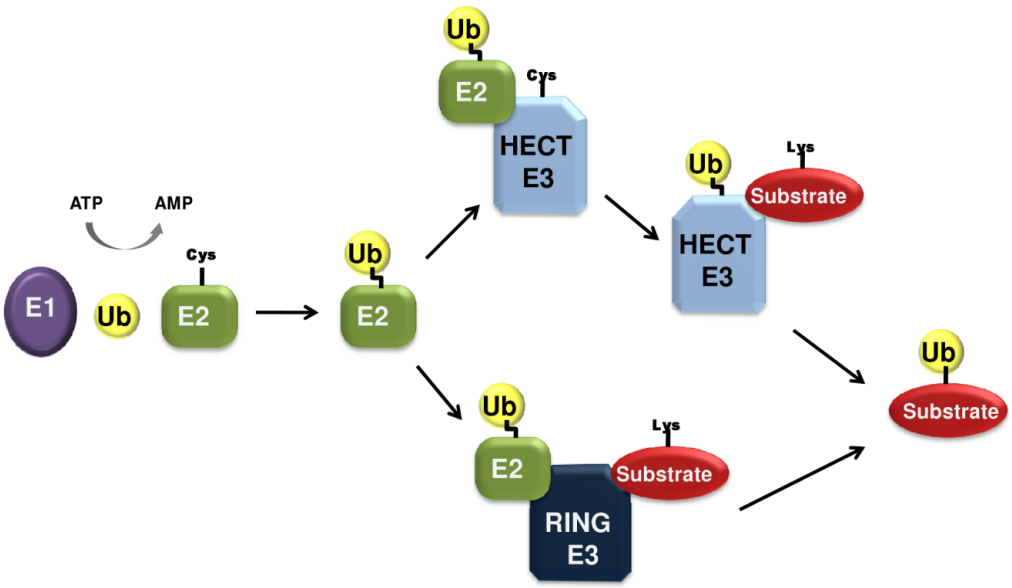

Figure 4.3 The ubiquitination mechanism. The ubiquitination process starts with the activation of ubiquitin by an E1 enzyme which then transfers it to the E2 protein. E2 together with ubiquitin bind to E3 protein ligase which interacts with the target substrate through another region. E3 ligase assists in ubiquitin transfer from E2 onto the substrate. The ubiquitin transfer mechanism depends on the E3 type; for HECT domain E3s, ubiquitin is first transferred from E2 to the active site residue of HECT E3, with subsequent transfer from E3 to the substrate protein whereas for RING and U-box type E3s, ubiquitin is directly transferred from E2 to the substrate without an E3 intermediate linkage.

4.3.1 Constructing a Human Structural E2-E3 Interaction Network

Incorporating protein structural information into networks is crucial for understanding the details of the interactions. Here, using the available structural proteome, we construct a human structural E2-E3 interaction network. First we compile the available structures for E2 and E3 proteins in the human ubiquitination pathway; there are 24 E2 proteins, 20 RING finger type E3, 9 HECT type E3 and 3 U-box type E3 proteins. To uncover possible interactions among these E2 and E3 proteins, we use our efficient prediction algorithm Prism [20, 142, 161]. Prism has proven useful in our previous studies in the construction of pathways [60] and characterization of interactions [30] and networks [141]. On a docking benchmark [228], Prism method is validated: it efficiently finds high-quality models for 87 out of 88 benchmark complexes and their binding regions [229]. Prism exploits structural and evolutionary similarities, and a template interface dataset which consists of known protein-protein interfaces. Based on these, it predicts potential interactions between target proteins. In this study, the known E2-E3 interfaces from the PDB constitute the template interfaces. **Figure 4.2** presents a schematic representation of the Prism algorithm which is composed of four consecutive steps. In the first step, the surface regions of the target proteins are extracted. In the second step, each template interface is split into the complementary partner chains and the partners are structurally aligned with the surfaces of the target proteins by the MultiProt engine [226]. 40% of the residues of template chains should match the target surfaces to pass to the next step. Besides the structural similarity, evolutionary similarity between target surface and template interface is assessed: at least one hot spot in each template partner should correctly match with the target surface. In the third step, target proteins are transformed onto the template interface and if the residues of the target partners collide (interpenetrate), then these pairs are eliminated. The last step is flexible refinement of the rigid docking solutions of MultiProt to remove steric clashes, refine and rank the predicted complexes according to the global energy using FiberDock [153]. FiberDock considers both side chain and backbone flexibility.

The resulting network, which consists of Prism predictions, contains 107 interactions between 38 proteins and they are listed in Supplementary Material. There are 22 E2 proteins, 9 RING-finger domain E3, 5 HECT-domain E3 and 2 U-box domain E3 proteins. Among these 22 E2s and 16 E3s, the number of all possible E2-E3 pairs is 352 (i.e. 22x16), out of which Prism predicts 107 to be interacting and leaves the remaining 245 as

unfavorable E2-E3 pairs. Out of 352 possibilities, 51 pairwise E2-E3 interactions were already reported in earlier studies. Although information related to non-interacting E2-E3 pairs is limited, a detailed literature survey showed that 24 E2-E3 pairs were reported to be negative. We recover 76% (39/51) of the known pairs through 107 predicted interactions verifying 36% (39/107) of the network. Of the 24 E2-E3 interactions reported as negative, 75% (18/24) are labeled as unfavorable by Prism. On the basis of known E2-E3 interaction data, the accuracy of our predictions is 76% ((39+18)/(51+24)), which indicates that the predicted E2-E3 interactions are in agreement with validated functional E2-E3 pairs. The Prism method depends on the coverage of known interface architectures and at the time of this study, there are only 9 E2-E3 known complex structures in the PDB. Because the size and coverage of the PDB increases exponentially, we expect the prediction efficiency of Prism to increase.

4.3.2 Topological Characteristic of the Structural E2-E3 Interaction Network

In terms of network topology, since the network is structural and the structural database is incomplete, it is not a typical scale-free protein-protein interaction network and does not follow a power law distribution; however, it is a connected network with an average degree of 5.6. As we noted above, we observe that E3s interact with multiple E2s and likewise E2 proteins interact with multiple E3s. Some of the E3s and E2s have many interaction partners raising the possibility that they are ‘adaptable’ and multi-functional. Among the E2s, UBE2D (UbcH5) family proteins show the highest number of interactions, consistent with previous findings that UbcH5 is active with most E3s [219]. These E2s are involved in degradation of misfolded and short-lived proteins and their orthologs Ubc4 an Ubc4 in *Saccharomyces cerevisiae* have been shown to be functionally redundant [230]. Such redundancy among UBE2D E2s would be a physiological advantage because it protects the ubiquitination networks from genetic perturbation [222]. Another highly interacting E2 in our network is UBE2N, which is involved in Lys63 ubiquitin chain assembly [231]. In contrast, E2s such as UBE2F, UBE2S, UBE2G1 and UBE2Q1/2 are involved in very few interactions. In comparison to previous E2-E3 RING interaction networks constructed by yeast two-hybrid screens [221, 222], we encounter similar topological trends: E2s showing a high number of E3 interactions are reported to be UBE2N, UBE2D and UBE2E families [221, 222], while UBE2F, UBE2S and UBE2G1 have only a few interactions [222]. Thus,

it appears that while some E2s such as the UBE2D family may function in the majority of the ubiquitination events in human cells [222], others such as UBE2F, which is known to conjugate ubiquitin-like protein NEDD8, are less important in maintaining the integrity of the ubiquitination network [221].

Different patterns of substrate ubiquitination lead to different substrate fates. Monoubiquitination can regulate DNA repair, endocytosis and gene expression, whereas polyubiqutination through Lys48 ubiquitin chains generally results in proteasomal degradation, and Lys63-linked ubiquitin chains can function in endocytosis and signaling [232]. Although mechanisms that control lysine selection in substrates are not clearly identified, structural aspects of E2-E3 pairs and their binding to the substrate appear to be important [233]. A RING E3 can utilize different E2s that have different linkage specificities [234], and its activity is more likely to depend on the nature of the E2s which are present [219]. However, for HECT E3s, the chain linkages of the substrate are usually determined by the E3 itself [235]. In our network, we observe that highly interacting RING and HECT E3s would mediate mono- and poly-ubiquitination, indicating that they play roles in a broad range of cellular processes. Chain linkage properties of E2s and E3s (if available) are listed in Supplementary Material. We find that Ube3a, a HECT protein ligase, for which several substrates are identified including p53, Mcm7 and cell cycle regulator Cdkn1b, has 17 E2 partners. Ube3a specifically favors Lys48-linked polyubiquitination [235] and thus mediates degradation of cytoplasmic misfolded proteins [236]. As a multi-functional protein ligase, it is conceivable that Ube3a can interact with several ubiquitin-conjugating E2s, which are all able to catalyze Lys48-linked chains. We also predict Ube3a to be interacting with Sumo-conjugating enzyme UBE2I, which was reported earlier [237], although the sumoylation function of Ube3a was not clearly identified. Ube3a mediates the ubiquitination of Pml (promyelocytic leukemia protein) tumor suppressor [238], which is also known to undergo sumoylation required for nuclear body formation [239]. Thus, in addition to ubiquitination of Pml, Ube3a may function as an E3 mediating the sumoylation of Pml. Another highly interacting HECT E3 is Smurf2, which can induce degradation of receptor-regulated R-Smads and R-Smad bound partners [240], mediating TGFβ signaling via ubiquitination of the receptors or inhibitory Smad7 adaptor [241]. Smurf2 is widely expressed in tissues. In addition to its ubiquitin-conjugating function, we predict that it interacts with Sumo- and Nedd8-conjugating E2s,

UBE2I and UBE2M, respectively. In the TGFβ signaling pathway, Smad4 is known to be sumoylated in the presence of UBE2I, which results in the redistribution of Smad4 to subnuclear speckles [242]. Thus, Smurf2 may also interact with UBE2I in this pathway. For Huwe1, a recently identified HECT E3 implicated in the regulation of cell proliferation, apoptosis and DNA damage response [243-245], most of the E2 partners are not yet identified. We predict that UBE2C, -D1, -D2, -D3, J2 and -L6 could be possible E2 partners. HECT E3s usually prefer to interact with E2s containing a phenylalanine at position 62 (by UBE2D1 numbering). We observe that all predicted E2 partners of Huwe1 contain a phenylalanine at this position except UBE2C, which has a tyrosine residue. Since both phenylalanine and tyrosine are hydrophobic residues [108], and since UBE2C is similar to UBE2D1 in forming multiple different ubiquitin linkages [246], we hypothesize that UBE2C would be a possible partner of HECT E3s. Considering the RING E3s, c-Cbl is a highly interacting ligase with 12 E2 partners in our network. c-Cbl, which targets numerous substrates, has both monoubiquitination (facilitating endocytosis [247], lysosomal degradation [248]) and polyubiquitination activities. In our network, in addition to UBE2D and UBE2L family interactions as described in the literature, and other polyubiquitin catalyzing E2s, c-Cbl is also found to interact with UBE2B, which may catalyze mono-ubiquitination of substrates targeted by c-Cbl. For another RING E3, Brca1 (Breast cancer type 1 susceptibility protein), although multiple different E2 partners, possessing unique ubiquitination activities, were identified in a yeast two-hybrid study by Christensen et al. [219], Prism predicts only three interactions (with UBE2B, -I and –N) and the latter two were reported before. Brca1 represents a special group of E3 enzymes which function as a heterodimeric complex with Bard1 [218]. Brca1 binds E2s and is an active E3 ligase only in association with Bard1. Christensen et al. [219] tested the Brca1-Bard1 RING domains for E2 interactions and several E2 partners: UBE2D1, -D2, -D3, -E1, -E2, -I, -K, -L3 and –N were detected. However, yeast two-hybrid screening with isolated Brca1 was shown to yield different E2 interactions from screening with fused Brca1-Bard1 dimers [221]. The reason we observe a limited set of E2s and many E2s as unfavorable for Brca1 is that we only considered a single Brca1 domain. A detailed analysis of the E2-E3 interactions is given in the next section.

4.3.3 Analysis of the E2-E3 Interaction Interfaces

The principles of selectivity between E2 and E3 proteins are still largely unclear. The factors that play a role in determing which E2 would bind E3 include (i) the substrate physiological/chemical state; that is, whether its concentration is too high, or it is damaged, etc; (ii) the local concentration of the E2, which is affected by for example, its subcellular localization or distribution in tissues; and (iii) the E2 conformational state, which is determined by for example its post-translational modifications [249, 250]. All factors change dynamically with the fluctuating cellular conditions. Here we focus on preferred interactions from the structural standpoint. Below, we present an analysis of the interfaces of the E2-E3 complexes.

4.3.3.1 General Trends in E2-E3 Binding

Analysis of the interfaces of E2-E3 pairs is expected to reveal which residues are conserved among E2s interacting with a common E3 and which residues differ. Such analysis may provide clues to specificity. For each E2-E3 pair in our network, we analyzed the interacting interfaces and predicted hotspot residues at the interfaces using the Hotpoint server [192]. The critical residues at the interface account for the majority of the binding energy are called hot spots [104]. The Hotpoint server determines computational hot spot residues based on conservation, solvent accessibility and statistical pairwise residue potentials of the interface residues. The predicted hot spots are observed to match with the experimental hot spots with an accuracy of 70% [192, 251].

The E2 proteins which have been characterized so far are known to recognize E3s through the L1 and L2 loops and the N-terminal α-helix 1 on the E2 surface [252]. In particular, loop L1 has been shown to be critical in E3 binding. Loop L1 residues phenylalanine at position 62 and lysine at position 63 of UBE2D2 have been observed to bind to RING E3s c-Cbl [187] and CNOT4 [215]. In our network, E2 interface residues also mostly correspond to the α-helix 1, loop L1 and L2 regions. Considering all of the E2 interfaces, there seem to be some structurally conserved residues in E3 binding. We observe that the sixth residue of loop L1 (phenylalanine 62 by UBE2D2 numbering) is a critical residue in E3 binding which frequently acts as a hotspot. In addition, the fifth residue in α-helix 1 (arginine 5 by UBE2D2 numbering) and the Ser-Pro-Ala motif in loop L2 (UBE2D2 labeling) mostly participate in the E3 binding. Earlier work has also shown

the importance of the Ser-Pro-Ala motif in E2s binding to U-box E3 CHIP [253] and RING E3 Brca1 [219]. These residues are found in almost all E2 interfaces in our network, which suggests that they could be indispensible for E2 binding to an E3. To investigate whether these residues are structurally conserved among the E2 family, we used the ProBIS algorithm [254], which detects structurally conserved regions of a protein by local structural alignments through ~23000 non-redundant PDB structures based on geometry and physicochemical properties. UBE2D1 (PDB code:2c4p) is the query protein. 57 similar structures are found. The α-helix 1 residues Arg5 and Glu9, loop L1 residues Tyr60, Pro61, Pro64 and Pro65, loop L2 residues Trp 93 and Ser94 are observed to show high structural conservation. The Phe62 residue in loop L1, which we frequently observe in the E2-E3 interface in our netwok, is found to be moderately conserved. Thus, E2 family proteins can employ these conserved residues to interact with different E3s. Interface and computational hotspot residues on α-helix 1, loop L1 and L2 regions of E2s are provided for all E2-E3 interactions in the Supplementary Material.

For HECT E3s, which are composed of two subdomains connected by a flexible peptide linker, the C-terminal lobe contains the catalytic cysteine and the N-terminal lobe contains the E2 binding region [205]. For Huwe1, in the E2 binding region (residues 4150-4200), ten important residues for E2 binding were identified: F4153, G4156, L4157, Y4159, L4160, Y4170, L4172, V4178, Y4206 and C4211 [243]. Our Huwe1-E2 models indicates that Huwe1 utilizes most of these residues when interacting with its E2 partners. In particular, 6 of these residues (F4153, L4157, L4160, Y4170, L4172 and Y4206) are identified as hotspot by Hotpoint server. Similarly, for other HECT E3s - Nedd4l, Smurf2, Ube3a and Wwp1 - our interface modeling and computational hotspots are in good agreement with experimental data obtained in previous works [243, 255, 256]. From the RING E3 perspective, for c-Cbl, Trp408 and Ile383 residues were shown to have a critical role in E2 binding [187]. The side chain of Trp408 is exposed to the solvent in the E2 binding cleft and hyrophobic residues are often found in an equivalent position in other RING E3s [234]. In each of our c-Cbl-E2 models, both Trp408 and Ile383 are labeled as hotspots by the Hotpoint server, indicating their importance in binding. Another study showed that mutation of Trp408 to alanine reduces c-Cbl's affinity for the E2 and eliminates its ubiquitin-ligase activity in vitro [191]. For the Traf6-UBE2N complex structure, seven residues within the RING domain of Traf6, Glu69, Pro71, Ile72, Leu74,

Met75, Ala101 and Pro106, were found to contribute significantly to E2 binding. [257] Among these, Ile72 and Leu74 were completely buried at the interface and contribute the most surface areas [257]. In our Traf6-E2 models, these residues reside at the interface although their contribution to binding differs, probably depending on the E2 partner present. Most of these residues are labeled as hotspots (especially Ile72 and Leu74). A comparison of HECT and RING E3s interacting with E2s is given in the next section.

4.3.3.2 HECT E3s Utilize Loop L1 of E2s Distinctively From the RING Finger Type E3s

As we noted above, ubiquitin transfer mechanisms differ among the E3 families. Here, we investigate whether this distinctive mechanistic feature, i.e. of ubiquitin forming an intermediate with the HECT domain before being transferred to the substrate, is also observed in HECT E3-E2 interaction interfaces. The 16 E3s in the putative E2-E3 complexes that were predicted by Prism belong to different families: Rbx1, Rbx2, Mdm2, Xiap, Trim37, c-Cbl, Brca1, Traf6 and Birc3 are RING finger type; Huwe1, Nedd4l, Wwp1, Smurf2 and Ube3a are HECT type; and Ube4a and Ube4b are of the U-box type. These E3s appear to share the same E2s; however, different from the RING finger and U-box type E3s, HECT-type E3s utilize the second position residue in loop L1 of E2s. The identity of this residue in loop L1 is not conserved among E2s, but its contacts with HECT E3s seem to be conserved structurally and chemically. One example shown in **Figure 4.4** is UBE2L3, an E2 which is involved in degradation of many proteins. We find that HECT E3s Nedd4l, Smurf2, Ube3a and RING finger E3s Traf6, c-Cbl and Rbx2 interact with the same E2, UBE2L3, using a shared E2 binding site. For Ube3a and c-Cbl, the complex structures with UBE2L3 are already available in the PDB, whereas for Nedd4l, Smurf2 and Traf6 the structural complexes are not known, although these interactions were reported in earlier works [258-260]. We modeled UBE2L3 interactions with Nedd4l and Smurf2 based on a known E2-E3 interface between UBE2D2 and Nedd4l (pdb code: 3jw0:AC) and the interaction with Traf6 is based on a known interface between the UBE2D1 and CHIP proteins (pdb code: 2oxq:AC). The interaction with Rbx2 is predicted by Prism. **Figure 4.5** lists the interface and hotspot residues of UBE2L3.

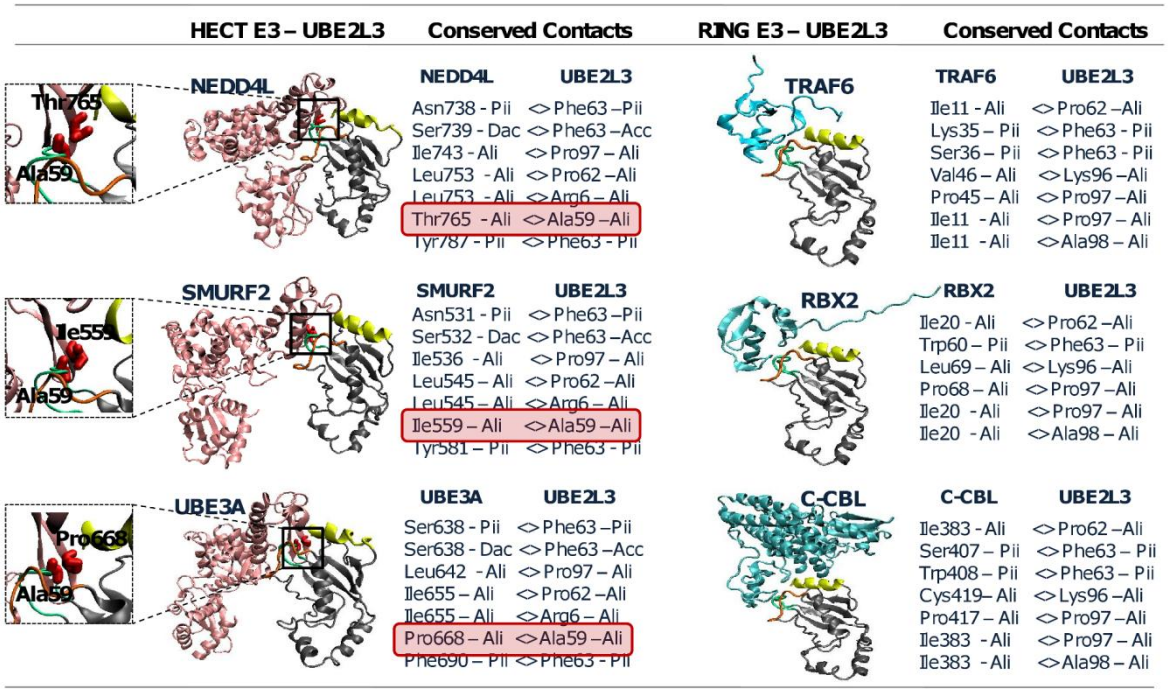

Figure 4.4 Interactions of UBE2L3 with HECT E3s and RING-finger E3s are visualized. UBE2L3, HECT E3s and RING E3s are displayed in gray, pink and blue color, respectively. UBE2L3's binding regions are α-helix 1 (in yellow), loop L1 (in green) and loop L2 (in orange). On these regions, there are seven structurally conserved chemical contacts, which are identified using MAPPIS. [261] Among these, loop L1 residues Pro62 and Phe63 and loop L2 residue Pro97 appear to be conserved through all interactions of UBE2L3, whereas loop L1 second position residue Ala59 (in red color) is observed only in HECT E3-UBE2L3 interfaces. Hydrophobic aliphatic residue contacts of Ala59 are visualized in the zoomed representation in the left panel. MAPPIS abbreviations of physico-chemical properties of residue contacts are: Ali: aliphatic hydrophobic property, Pii: aromatic property, Dac: hydrogen bond donor and acceptor, and Acc: hydrogen bond acceptor.

When we compare the chemical contacts of these interactions using MAPPIS [261], we observe seven structurally conserved contacts for HECT E3s among five UBE2L3 residues: Arg6, Ala59, Pro62, Phe63 and Pro97; and seven structurally conserved contacts for RING E3s among five UBE2L3 residues: Pro62, Phe63, Lys96, Pro97 and Ala98. Among these, loop L1 residues Pro62 and Phe63 and loop L2 residue Pro97 appear to be conserved through all interactions of UBE2L3. These residues mostly make hydrophobic aliphatic and aromatic (π) contacts with E3 residues according to the MAPPIS [261] classification (see **Figure 4.4**) and thus they seem to be important in binding. Earlier work has supported this observation; Phe63 of UBE2L3 was shown to be recognized by both

HECT and RING E3 ligases; c-Cbl and Ube3a [187]. Comparison of the chemical contacts of HECT and RING E3s (left and right panel of **Figure 4.4**, respectively) reveals that the second position residue Ala59 in loop L1 of UBE2L3 appears to be utilized solely when interacting with HECT E3s. A previous study on E2 binding of Ube3a also reported this second position loop L1 residue as being important in HECT E3-E2 binding [262]. We observe that this alanine residue makes hydrophobic aliphatic contacts with Thr765 of Nedd4l and Ile559 of Smurf2 and Pro668 of Ube3a. When MAPPIS [261] is applied to known E2-HECT E3 (UBE2D2-Nedd4l, UBE2L3-Ube3a), E2-RING E3 (UBE2D2-Not4hp, UBE2L3-c-Cbl, and UBE2N-Traf6) and E2-U-box E3 (UBE2D3-Ube4b, UBE2D1-STIP1) complexes in the PDB, similar trends are observed: loop L1 second position residue contacts are only employed when interacting with HECT E3s. Although the residue type is not conserved among other E2s, for example, UBE2D2 utilizes threonine and UBE2N uses glutamic acid at the second position in loop L1, the contacts of these residues through HECT E3s are structurally conserved. Interestingly, this observation holds among almost all E2-HECT E3 interactions which might be associated with the distinctive ubiquitin transfer mechanism of HECT E3s. This unique residue contact might play a role in mediating the allosteric communication between E3 binding site and E3 active site, and thus make the ubiquitin transfer from E2 to HECT E3 possible.

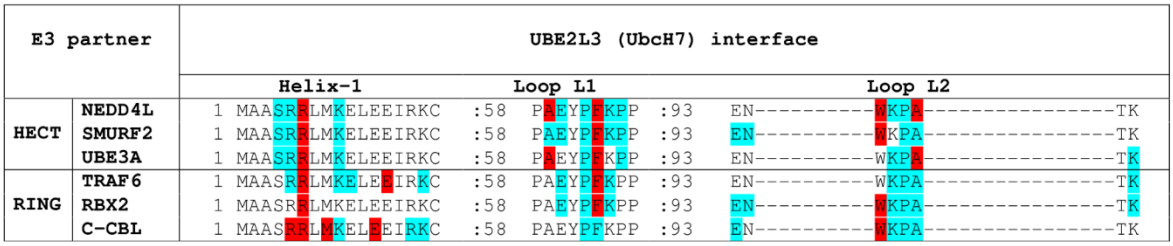

Figure 4.5 Interface and computational hotspot residues of UBE2L3 interacting with E3s. Interface residues and hotspot residues of UBE2L3 are shown in cyan and red color, respectively. Comparison of HECT type E3s (Nedd4l, Smurf2 and Ube3a) and RING finger type E3s (Traf6, Rbx2 and c-Cbl) reveals that loop L1 of UBE2L3 is utilized differently. UBE2L3 uses the second position residue in loop L1 frequently as hotspot while interacting with HECT E3s.

4.3.3.3 Sequence Variations in E2 Residues May Contribute Specificity to E3 Binding: The UBE2E Example

The UBE2E family E2 proteins UBE2E1 and UBE2E2 are highly similar both in sequence and structure (97% sequence similarity and rmsd: 0.54 Å). The Prism results indicate that UBE2E1 and UBE2E2 share two E3 partners, Xiap and C-Cbl. Their remaining E3 partners are different: UBE2E1 specifically interacts with Mdm2 and Birc3 whereas UBE2E2 prefers to interact with Wwp1, Rbx2 and Ube3a. Earlier, van Wijk et al. [221] reported that UBE2E2 has shown an E3 interaction pattern distinct from the profile of UBE2E1. Since UBE2E1 and UBE2E2 are quite similar, we investigate which residues would be responsible for this interaction specificity. **Figure 4.6A** displays the interface and computational hotspot residues on UBE2E1 and UBE2E2 which bind to different E3s. The residues in the binding regions of UBE2E1 and UBE2E2 are identical except for three: the residue in the 12^{th} position in the α-helix 1 (Asp58 in UBE2E1 corresponding to Glu12 in UBE2E2), the residue in the first position in loop L1 (Thr103 in UBE2E1 corresponding to Ser57 in UBE2E2) and the residue in the third position in loop L1 (Glu105 in UBE2E1 corresponding to Asp59 in UBE2E2). Interestingly, these residues are not utilized at the common interface of E3 partners Xiap and C-Cbl. In this way, these E3s are likely to interact with both UBE2E1 and UBE2E2. In contrast, the residue at the third position in loop L1 seems to be crucial in ensuring the interaction specificity of the UBE2E family. For UBE2E2, this Asp59 favors specific binding to Wwp1, Rbx2 and Ube3a whereas for UBE2E1, the same position residue Glu105 plays a role in Mdm2 binding. Another specific binding of UBE2E1 is with Birc3. For this interaction, UBE2E1 utilizes an α-helix 1 residue, Leu44, which does not occur in the UBE2E2 sequence. A previous study [221] reported the alteration of E3-interaction specificity of E2 UBE2N due to the mutation of its two helix 1 residues. This UBE2N mutant failed to interact with more than half of its E3 partners and gained some new interactions instead [221]. Similarly, for the UBE2E family, the mutation of loop L1 residues Glu105 and Asp59 for UBE2E1 and UBE2E2, respectively, would probably alter the E3-interaction specificity. To investigate the importance of the contribution of these residues to the binding energy of the complexes, we performed a computational mutagenesis analysis using FoldX algorithm [227], whose force-field is based on emprical energy terms correlated with experimental ΔΔG measurements [263]. Using FoldX, loop L1 residues Glu105 (UBE2E1) and Asp59

(UBE2E2) are mutated to alanine and quantative estimations of the binding affinities of the wild-type and mutants are obtained. We observe that mutations of these residues in UBE2E1 and UBE2E2 that interact with the same E3s (Xiap and C-Cbl) do not affect the binding affinities significantly (ΔΔG=0.3 kcal/mol on average). In contrast, for the specific E3 interactions of UBE2E1 and UBE2E2 (with Mdm2, Birc3, Rbx2, Ube3a and Wwp1), mutations have a destabilizing effect (ΔΔG=1.2 kcal/mol on average). The most destabilizing is the mutation of loop L1 sixth position residue phenylalanine, which is identified as a computational hotspot by the Hotpoint server (ΔΔG=1.9 kcal/mol on average). In **Figure 4.6B** left panel, the interactions of UBE2E1 and UBE2E2 with their common partner Xiap are illustrated. Xiap, which mediates the degradation of many proteins such as Caspase-3 and Smac, has been recently reported to interact with both UBE2E1 [264] and UBE2E2 [222]. A specific interaction of UBE2E2 with Ube3a is displayed on the right panel of **Figure 4.6B**. We observe that Asp59 (shown in cyan color) is utilized as an interface residue and its mutation is significantly destabilizing (ΔΔG=2.6 kcal/mol). Consequently, while with mutation of this residue UBE2E1 and UBE2E2 may still interact with Xiap and C-Cbl, they will most probably fail to interact with other E3s. These findings illustrate how slight sequence variations in E2 residues may contribute to the specificity of E3 binding.

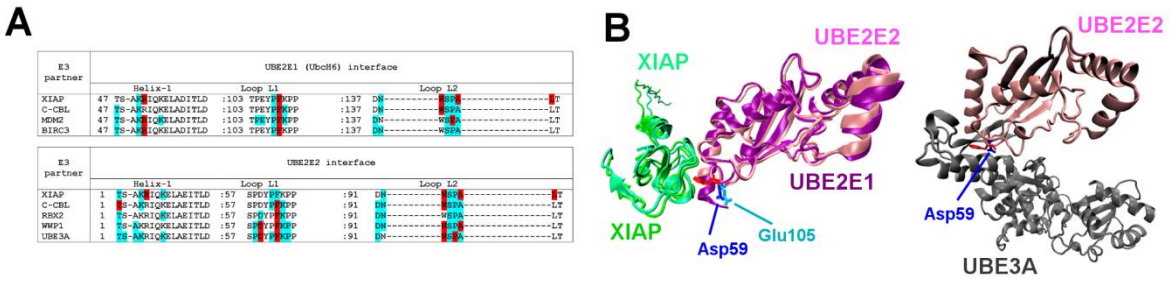

Figure 4.6 Interface and computational hot spot residues of UBE2E family. (A) For the common interaction partners, Xiap and c-Cbl, interface residues on UBE2E1 and UBE2E2 are similar (first two rows). The major contribution to specificity might be inferred via the third position loop L1 residue of UBE2E1 and UBE2E2 (Glu105 and Asp59, respectively). (B) On the left panel, Xiap interactions with UBE2E1 and UBE2E2 are displayed (superimposed onto each other). Computational mutagenesis analysis indicate that mutation of Glu105 (blue color) and Asp59 (cyan color) of UBE2E1 and UBE2E2, respectively, does not affect binding affinities when interacting with Xiap (ΔΔG=0.1 kcal/mol). On the right panel, UBE2E2 specifically interacts with Ube3a utilizing loop L1 Asp59 residue as an interface residue and its mutation is significantly destabilizing (ΔΔG=2.6 kcal/mol). For all interactions of UBE2E family, phenylalanine (red color) at sixth position in loop L1 appears to contribute significantly to binding.

4.3.3.4 E2 Selection Based on Substrate and Subcellular Localization

E2 selection by an E3 may be affected or determined by the substrate that will be degraded since the substrate should be in close proximity to the E2 during the ubiquitin transfer. Subcellular localization of E2 and E3 enzymes may also affect the interaction preferences. To identify which E2-E3 interactions are more likely to occur in our network, we classified the E2 and E3 genes in our network according to their subcellular localization as defined by Gene Ontology [265] cellular component terms. We observe that many E2 and E3 enzymes are found both in the nucleus and in the cytoplasm and some are also found at the plasma membrane. The classification of E2-E3 interactions according to the cellular components, i.e. the interactions in the cytoplasm, nucleus and plasma membrane, is illustrated in **Figure 4.7.** The interactions appear to be localization-specific; for example, although Mdm2 ligase is ubiquitous (existing everywhere in the cell), its interactions are dependent on the localization of the E2. One substrate of Mdm2 is p53 tumor suppressor protein, which localize to either the nucleus or cytoplasm and can be ubiquitinated and degraded by Mdm2 [266]. In our network, to ubiquitinate p53, Mdm2 can select UBE2E1 or UBE2D1 in the nucleus, while in the cytoplasm it can additionally prefer UBE2D3, UBE2G2 and UBE2L6. Interactions with the UBE2D family and UBE2G2 were reported earlier [221]. Mdm2 can also be found at the plasma membrane, where it was shown to bind to Beta-arrestin 2 (βarr2), which is a regulatory protein playing a central role in the endocytosis of most G-protein-coupled receptors [267], to drive its ubiquitylation [268]. According to our network, at the plasma membrane, Mdm2 can only interact with UBE2D3 to ubiquitinate βarr2.

Other than Mdm2, Huwe1 and Ube3a (also called E6-AP) are known to ubiquitinate p53, targeting it for degradation as well [243, 269]. To perform this specific function, Mdm2, Huwe1 and Ube3a most likely select similar E2s. Consequently, common E2 partners for these E3s would reveal which E2s would be selected in a substrate-specific manner. In our network, common E2 partners for Mdm2, Huwe1 and Ube3a are UBE2D family E2s (UBE2D1, UBE2D2, UBE2D3) and UBE2L6. Interactions between Ube3a and these E2s are experimentally known [258, 262, 270, 271] without E2-E3 complex structures. Since Huwe1 is a recently identified E3 ligase, many of its interaction partners are not known. We observe that most of the interface residues on E2s are shared indicating that these E2s can interact with only one E3 at a time. When we compare the global

energies of the putative E2-E3 complexes computed by FiberDock, we see that Mdm2 binds most favorably to UBE2D1 (global energy: -26.14 kcal/mol) in the cytoplasm and the nucleus whereas Ube3a prefers UBE2L6 (global energy: -45.32 kcal/mol) in the cytoplasm and UBE2D1 (global energy: -24.9 kcal/mol) in the nucleus. The most favorable E2 partner for Huwe1 is UBE2D2 (global energy: -23.38 kcal/mol), for which cellular component information is not available. The calculated Fiberdock energies for these E2-E3 complexes are listed in **Table 4.1**. For each E3 interacting with E2s, the binding energies seem to be close to each other, which is consistent with earlier studies that reported similar binding affinities for a common E3 interacting with multiple E2s [219, 237, 272]. Depending on which substrate is ubiquitinated and where it is ubiquitinated in the cell, E2 selection of E3s may vary. Our E2-E3 models provide a source for future investigation of selectivity.

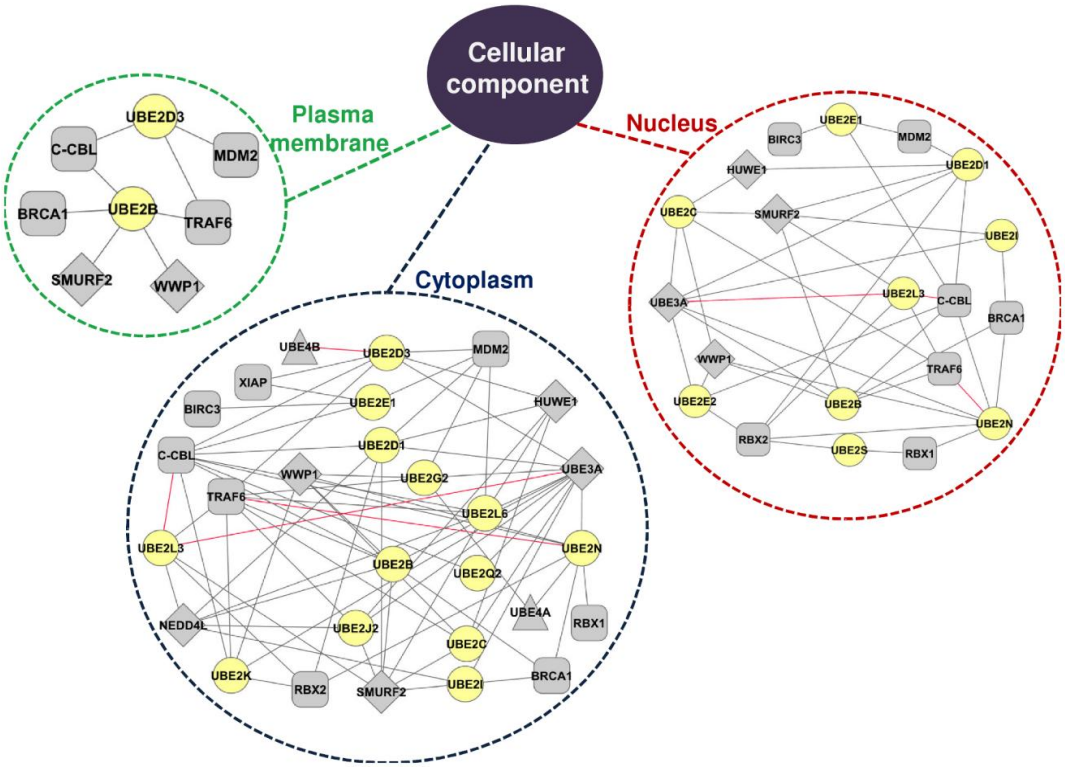

Figure 4.7 Classification of the human structural E2-E3 interaction network according to the GO [265] cellular component terms. There are 72, 41 and 8 E2-E3 interactions in cytoplasm, nucleus and plasma membrane classification, respectively. Some interactions are likely to occur ubiquitously whereas others appear to be localization-specific. E2 proteins are displayed in yellow color and E3 proteins are in gray. E3s are visualized according to the domain types; RING-finger domain, HECT-domain and U-box domain E3s are in rectangular, diamond and triangular shape, respectively. Red color edges are the E2-E3 interactions for which the complex structures are available in the PDB.

Table 4.1 Interactions and binding energies of E2-E3 complexes mediating degradation of the same substrate, p53 protein.
E3 proteins Mdm2, Huwe1 and Ube3a interact with E2 proteins UBE2D1, UBE2D2, UBE2D3 and UBE2L6 with different binding energies (kcal/mol) computed by Fiberdock. The most favorable partners for Mdm2, Huwe1 and Ube3a are UBE2D1, UBE2D2 and UBE2L6, respectively, for which the binding energies are underlined below.

	E2 proteins			
E3 proteins	**UBE2D1**	**UBE2D2**	**UBE2D3**	**UBE2L6**
MDM2	-26.1	-21.9	-15.7	-10.4
HUWE1	-10.6	-23.4	-11.5	-14.9
UBE3A	-24.9	-29.3	-32.3	-45.3

4.4 Concluding Remarks

Substrate ubiquitination is mediated by the interactions between E2 enzymes and E3 ligases. Although these E2 and E3 proteins function in a concerted manner, the principles of selectivity between them are still not entirely understood. Here we address E2-E3 interactions in the human proteome by taking a structural approach, and computationally model E2-E3 complexes based on interface structural motifs. By exploiting the available structural proteome and the powerful Prism algorithm [20, 142, 161], we construct a human structural E2-E3 interaction network. The results indicate which E2s are likely to interact with which E3s and how they could interact. Analysis of the modeled interfaces of E2-E3 pairs elucidates binding patterns: some residues are structurally conserved among E2 proteins and appear to be essential for all E2-E3 interactions, whereas others, particularly in loop L1, appear to play important roles in E3 selectivity. Further, classification of E2 and E3 enzymes according to their subcellular localization would reveal which E2-E3 pairs are more likely to occur in each cellular component and how the E2 preference may vary depending on the substrate.

Several E3s have been implicated in disease, including cancer and neurodegenerative diseases [273]. Recently various HECT-E3s have emerged as important regulators of cancer development [274]. RING-finger E3s are classified as either tumor suppressors or oncoproteins and are often overexpressed in cancer [273]. Therefore, in disease processes, to block the activity of E3s, inhibitory molecules can be developed which target E3-E2 interface or E3-ubiquitylation substrate interface. In this study, we focused on predicting

and analyzing E2-E3 pairs in the human ubiquitination pathway. We believe that the structural E2-E3 network and the E2-E3 interface data in this study provide a resource for future studies of ubiquitination and E2-E3 selectivity, especially in discovering drug candidates targeting E3s.

Chapter 5

ACTIVATION OF APOPTOTIC SIGNALS THROUGH CIRCADIAN CLOCK DISRUPTION

Circadian clock regulates many biochemical pathways at the cellular level by generating ~24-hour oscillations in the physiology and behavior of the organism known as circadian rhythms. A recent study indicated that disruption of circadian clock with Cryptochrome (*Cry*) knockout in mouse with *p53*-null background decrease cancer incidence and increase life span due to increased sensitivity of cells to apoptosis by genotoxic agents and hence protects *p53*-null mice from the early onset of cancer. Here, we combine gene expression analysis with protein-protein interaction data to investigate the altered pathways due to *Cry* knockout in *p53*-null background. We focus on pathways related to cellular growth and death and provide a comparison of $p53^{-/-}$ and $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ mice fibroblast cells at the gene expression level. The results reveal that *Cry* knockout in a p53-null cells leads to upregulation of several pro-apoptotic genes and downregulation of those related to cell survival. In particular, we observe that circadian clock pathway would intersect with several pathways including apoptosis, Wnt, Notch, p53 and insulin signaling. The connection between circadian clock disruption and apoptotic events appears to be regulated by GSK3β, which is a key component of several signaling pathways. NF-κB activators such as TLR2, NOTCH1 and TNFSF12, regulated by GSK3β, are downregulated which would fail to activate NF-κB making *p53*-null cells more sensitive to apoptosis in *Cry*-null background. GSK3β would also mediate the alterations in Wnt, insulin and MAPK signaling leading to the suppression of cell survival. These findings are important for understanding the pathways/processes that would be affected upon *Cry* knockout as well as for identifying targets in treatment of cancers associated with *p53*-deficiency.

5.1 Introduction

The circadian clock generates cyclical changes in the physiology and behavior of organisms with a periodicity of ~24 hours, known as circadian rhythms [275]. The central

commanding center of circadian rhythms is located in the suprachiasmatic nuclei (SCN) in the hypothalamus. The SCN coordinates the activity of peripheral clocks, which operate in all cells in human and mice. The mammalian circadian oscillatory mechanism is achieved through transcription-translation feedback loops through the clock elements such as E-box, D-box, and ROR/REV-ERB binding elements (RORE). Among these regulatory sequences, the E-box is known to be the most important in the molecular oscillatory system. The transcriptional activators CLOCK and BMAL1 proteins heterodimerize in the cytoplasm and translocate to the nucleus where they bind to E box elements located in the protomer region of various genes and induce the transcription of circadian clock genes up to 10% of the genome including three periods (*Per1*, *Per2* and *Per3*) and two cryptochromes (*Cry1* and *Cry2*) [276]. PERs and CRYs bind to each other and upon dimerization along with casein kinase Iε, they inhibit the CLOCK-BMAL1 complex and hence their own transcription [277-279]. The time delay between transcription of PERs and CRYs and their action as repressors is a key factor in generating circadian rhythms and circadian phenotype [279]. Circadian clock pathway is visualized in **Figure 5.1** schematically.

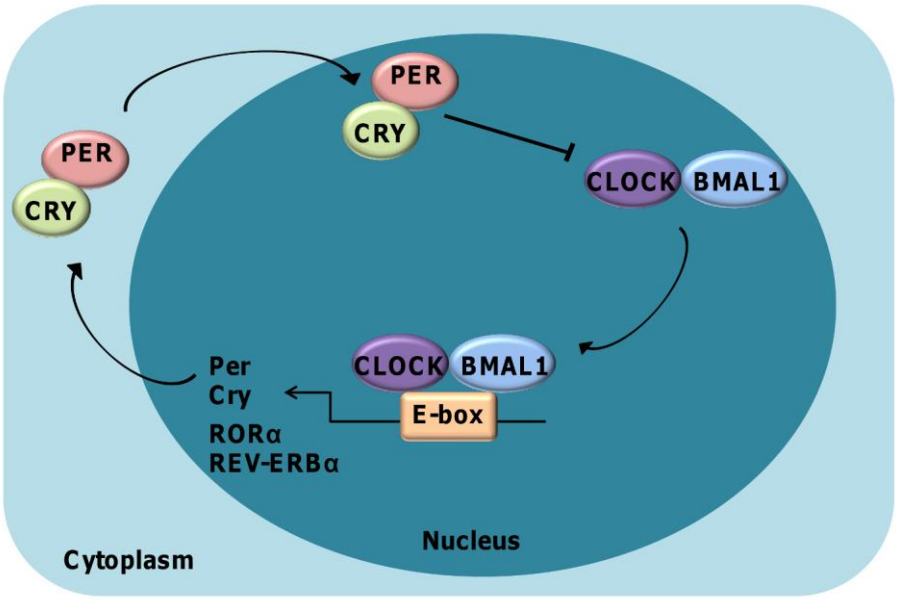

Figure 5.1 A simple visualization of circadian clock pathway. The transcriptional activators CLOCK and BMAL1 proteins bind to each other and translocate to the nucleus where they bind to E-box elements located in the protomer region of genes. They induce transcription of several genes including Per, Cry, RORα and REV-ERBα. PER and CRY proteins bind to each other in the cytoplasm and translocate to the nucleus to inhibit the CLOCK-BMAL1 complex and hence their own transcription [276-279].

Circadian clock disruption by environmental and genetic factors results in a number of pathological conditions [280-282] and predisposes humans and mice to cancer [280, 282]. However, the evidence regarding the occurrence of cancer due to clock disruption is somewhat conflicting; although disruption of the clock by *Per2* mutation is observed to predispose mice to cancers [283], disruption by *Cry* knockout [284] or *Clock* mutation [285] does not. Therefore, rather than the disruption of circadian clock itself, the mechanism by which the clock is disrupted seems to be more important. Recent studies indicate that clock is in communication with different signal transduction mechanisms specifically regulating the metabolism of organisms. To elucidate the mechanisms behind clock disruption and to understand the functions of clock genes, it is crucial to identify the pathways by which circadian clock genes communicate. Previously it was shown that circadian clock and cell cycle pathways interface at a number of critical points. For example, BMAL1-CLOCK complex positively regulates WEE1 antimitotic kinase which regulates G2/M transition and is responsible for circadian variation in liver regeneration [286]. In addition, BMAL1-CLOCK complex represses transcription of *c-Myc* gene, which is a critical regulator of cell cycle progression [287]. Mechanisms of circadian control for other cell cycle-related genes such as cyclins and cyclin-dependent kinases (*Cdk2*, *Cdc2*) are currently unknown. Likewise, the expressions of several apoptosis-related genes, such as *p53*, *Mdm2*, *Apaf1*, *Dapk1*, *p63* and *Bcl2*, are changed in circadian mutants [288], however, mechanism of circadian regulation is unknown. Recently, Ozturk et al. [289] investigated the effect of the *Cry* knockout on carcinogenesis in a mouse strain prone to cancer because of a *p53* mutation. Contrary to the expectation that clock disruption by *Cry* knockout would further increase cancer risk, they found that *Cry* knockout protects *p53* mutant mice from the early onset of cancer and extends their median lifespan about ~50%, in part by sensitizing *p53* mutant cells to apoptosis in response to genotoxic stress [289]. More recently, it was observed that increased sensitivity to apoptosis after UV irradiation would be due to the high levels of *p73*, a functional homolog of *p53*, in $p53^{KO}$ Cry^{DKO}cells [290]. In addition, *Cry* knockout is shown to enhance tumor necrosis factor α (TNFα)-initiated apoptosis by interfering with NF-κB signaling pathway through the GSK3β kinase [291].

In this study, we performed a large-scale integration of microarray expression profiles with protein-protein interaction networks to investigate how cellular processes are affected

in *p53*-null mice cells upon *Cry* knockout. We mainly focus on the processes related to cell growth and death and obtain associated pathways; cell cycle, apoptosis and p53 signaling from KEGG pathway database [224]. We construct the so-called "crosstalk networks" of circadian clock and cell growth and death pathways, enriching these by physical and functional interactions using STRING search tool [292]. Combining gene expression profiles (fold changes $p53^{-/-}$ $Cry1^{-/-}Cry2^{-/-}$ / $p53^{-/-}$) with protein-protein interaction crosstalk network data reveals differentially regulated genes and pathways and how apoptotic signals would be activated in *p53*-null cells upon *Cry* knockout. We observe that expressions of several apoptotic genes are increased upon *Cry* knockout in $p53^{-/-}$ cells and a minor amount of genes would promote cell survival leading to, in overall, a shift towards cell apoptosis.

5.2 Methodology

5.2.1 Cell Culture Maintenance and Total RNA Preparation

We used mouse skin fibroblasts from 2 age matched $p53^{-/-}$ and triple knockout C57BL/6 mouse [289] for our microarray analyses. Culture medium was removed from exponentially growing cells, set aside, and then cells were washed once with warm PBS before placement under a GE germicidal lamp emitting primarily 254 nm UV light (UV-C) connected with a digital timer [293]. After receiving the 10 $J/m^2$of UV radiation, culture medium was added back to the cells, which were subsequently placed back into the cell culture incubator for the indicated length of time. A UV-C sensor (UV Products, San Gabriel, CA) was used to calibrate the fluence rate of the incident light. Mouse skin fibroblasts were rinsed twice with PBS, trypsinized, suspended in medium, and collected by centrifugation. Cell pellets were washed twice with cold PBS and resuspended in RLT lysis buffer (QIAGEN, Valencia, CA). Total RNA was prepared with RNeasy mini kit (QIAGEN) as suggested by the manufacturer. DNA contamination was avoided by treating all the samples with Qiagen on-column DNase digestion. The quality and quantity of the total RNA sample were determined using an Agilent Bioanalyzer (Bioanalyzer 2100, Agilent Technologies) and NanoDrop® ND-1000 Spectrophotometer **(**NanoDrop Technologies), respectively.

5.2.2 Whole Mouse Genome Oligo Microarray Hybridization and Data Analysis

We used whole mouse genome oligo microarrays (Agilent Technologies) for gene expression analysis. Both the cDNA labeling and microarray hybridization were followed

as recommended by the manufacturer (Agilent two-color microarray-based gene expression analysis). After hybridization, microarray slides were scanned using the Agilent Scanner (Model B) located in the University of North Carolina Genomics Core Facility. We used GeneSpring version GX 7.3.1 software (Agilent Technologies, Inc) for data analysis. The results represent an average of two biological replicates for each of the experiments.

5.2.3 Identification of Differentially Expressed Genes

Fold changes (the ratio of gene expression *p53-Cry1-Cry2* knockout to *p53* knockout mice cells after 4-hours of UV-induction) follow a symmetric distribution when converted to log2 values. The genes with expressions greater than $\mu+2\sigma$ (mean plus two times standard deviation) are assigned as "up-regulated" and similarly, the genes with expressions lower than $\mu-2\sigma$ (mean minus two times standard deviation) are assigned as "down-regulated". There are 1059 differentially expressed genes at the 95% confidence level ($\alpha=0.05$) out of which 447 genes are up-regulated and 612 genes are down-regulated.

5.2.4 Construction of Crosstalk Network of Circadian Clock and Cell Growth & Death Pathways

The genes related to cell growth and death processes (apoptosis, p53 signaling and cell cycle pathways) and circadian clock pathway genes are obtained from KEGG database [224]. Since cell cycle pathway interacts closely with MAPK signaling, Wnt signaling and TGFβ signaling pathways according to KEGG annotations, the genes in these pathways are also considered. Thus, six crosstalk networks (circadian-apoptosis, circadian-p53, circadian-cell cycle, circadian-MAPK, circadian-Wnt, circadian-TGFβ) are constructed using STRING search tool [294], which provides physical and functional protein-protein associations derived from experiments, databases, genomic context and co-expression studies. While constructing the crosstalk networks, only highest-confidence interactions (STRING score>0.9) are considered and only first partners of seed gene list are included; i.e. network depth is 1.

5.2.5 Extraction of Subnetworks

Since we analyze the gene expression change upon *Cry1-Cry2* mutation in *p53*-knockout mice cells, first, we removed *p53*, *Cry1* and *Cry2* and their corresponding protein-protein

interactions from each crosstalk network. Then, circadian pathway genes, differentially expressed genes and Caspase 3 gene, which plays a central role in the execution-phase of cell apoptosis, are selected and subnetworks, which show the interactions among these genes, are extracted from each crosstalk network by using BiNoM Cytoscape plugin [295]. Given the selected genes, BiNoM connects them by shortest paths and extract a subnetwork, which provides a practical way to analyze changes in gene expressions at a network-level. BiNoM applies a agglomerative hierarchical clustering algorithm to build subnetworks; the distance between the selected nodes is the minimal linkage applied to the base distance to satisfy the triangular inequality and the maximal length of the shortest paths are assigned to obtain non-intersecting subnetworks.

5.3 Results

Studies have showed that circadian clock regulates many pathways in particular metabolic processes, however, the exact mechanisms are unclear [296, 297]. Therefore, considering the alterations of gene expressions within the scope of their interactions in cellular processes is an asset in understanding the details of regulation mechanisms. Ozturk et al. indicated that there is a relationship between the apoptosis and circadian clock, particularly between p53 and circadian clock. They have shown that the Cry mutation protects p53 mutant mice from the early onset of cancer and extends their median lifespan approximately 50% [289]. In this study, we want to define pathways and mechanisms for understanding how the absence of the *Cry* genes delays the onset of the cancer and its relation to p53-mediated pathways using both computational and experimental approaches. Therefore, we conducted a microarray experiment using fibroblast cell lines of the $p53^{-/-}$ and $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$. In particular, we compared the gene expression profiles of two conditions: $p53^{-/-}$ and $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ primary mouse skin fibroblasts.

We integrate microarray gene expression profiles with protein-protein interaction networks. We observed that Cry knockout would closely affect cellular processes related to cell growth and death; cellular growth factors are suppressed while genes related to cell death are activated. In the following sections, we first present an analysis of differentially expressed genes. Then, we take a detailed look into the interactions of these genes in a network concept: we map them onto crosstalk networks, which include interactions

between circadian pathway and pathways related to cellular growth and death, and thus being enriched representations of these pathways.

5.3.1 Differentially Expressed Genes

The genes with a significant change in expression upon *Cry* knockout (fold ratio $p53^{-/-}$ $Cry1^{-/-}Cry2^{-/-}$/ $p53^{-/-}$) after 4-hours of UV-induction are labeled as differentially expressed. Of 1059 differentially expressed genes, 447 are upregulated and 612 are downregulated given as Supplementary Material. From these, ~30% of them have pathway information according to KEGG pathway database. To identify the biological processes they take place in and to observe their distribution in apoptotic processes, we extract their gene ontology (GO) annotations [298]. We see that 298 out of 447 upregulated genes and 391 out of 612 downregulated genes have GO biological process annotations. Among 689 annotated genes, 10% (70 out of 689) are related to "cell death" processes. A detailed investigation of these genes showed that 29 of them are apoptotic (positively regulating cell death), 33 are anti-apoptotic (negatively regulating cell death) and 8 are either apoptotic or anti-apoptotic depending on the cellular condition. Our analysis indicated that apoptotic genes are mostly upregulated (20 out of 29), whereas, anti-apoptotic genes are downregulated (26 out of 33). In **Table 5.1**, up- and down-regulated genes annotated in cell death are listed together with their functions in this biological process.

5.3.2 Mapping Differentially Expressed Genes onto Crosstalk Networks

Apoptosis and p53 signaling pathways are related to cell death [224]. Pathway related to cell growth is cell cycle pathway which interacts closely with MAPK signaling, Wnt signaling, and TGFβ signaling pathways [224]. To see the interactions between these pathways and the circadian clock pathway, "crosstalk networks" are constructed by assembling the protein-protein interactions of these pathway genes using STRING search tool. Such analysis yielded six crosstalk networks: circadian-apoptosis, circadian-p53 signaling, circadian-cell cycle, circadian-MAPK signaling, circadian-Wnt signaling and circadian-TGFβ signaling networks. Then, we map differentially expressed genes onto these crosstalk networks to detect alterations upon *Cry* knockout. Among them, circadian crosstalk networks with apoptosis, p53 signaling and Wnt signaling are observed to be affected at most; hence we focus on these three crosstalk networks and below present our findings.

Table 5.1 Up- and down-regulated genes annotated in cell death by Gene Ontology
Of 1059 differentially expressed genes, 689 genes are annotated by Gene Ontology biological process terms. Among 689 annotated genes, 10% (70 out of 689) are related to "cell death" process. 29 of these positively regulate cell death (apoptotic) and 33 of them induces a negative regulation on cell death (anti-apoptotic), while eight of these genes can be either apoptotic or anti-apoptotic. We observe that apoptotic genes are mostly upregulated (20 out of 29), whereas, anti-apoptotic genes are downregulated (26 out of 33).

	Up-regulated genes	**Down-regulated genes**
Apoptotic (29 genes)	*Adipoq, Aldh1a2, Cd28, Col4a3, Ctsc, Fgl2, Gas1, Gzmc, Gzme, Mmp9, Pglyrp1, Ptprj, Rarb, Scin, Serpina3g, Serpinb9, Sfrp1, Sfrp2, Sfrp4, Wt1*	*Aatk, Dab2ip, Dbc1, Dmrt2, Igf2r,ORF9, Sox7,Unc5a,Unc5b*
Anti-apoptotic (33 genes)	*Adamts20, Angpt4, Cxcr7, Il1rn, Krt18, Mdk, Nefl*	*Adm, Alb1, Aqp1, Axin2, Cd59b, Comp, Cxcr4, Dhrs2, Dnajc10, E2f2, Fkhl18, Foxc1, Foxc2, Foxo1, Furin, Hc, Krt20, Lrp5, Naip2, Nod1, Notch1, Rel, Tgfb2, Tnfsf12, Ung, Wfs1*
Either apoptotic or anti-apoptotic (8 genes)	*Ccl2, Cdh13, Clu*	*Cck, Lyz, Lyzs, Map3k1, S100b*

5.3.2.1 Inactivation of Wnt Signaling upon Cry Knockout

One of the pathways that clock communicate is Wnt signaling pathway which regulates cell cycle plays an important role in cell growth, development, cell proliferation and oncogenesis [299-301]. Aberrant activation of Wnt pathway and nuclear accumulation of β-catenin result in tumor progression [302]. Association between Wnt signaling pathway and apoptosis has been established as well: Wnt signaling pathway regulates apoptosis through a variety of mechanisms [303]; through β-catenin, GSK3β-NF-κB, c-JUN N-terminal kinase signaling [304, 305]. Thus, to gain an insight of how *Cry* knockout in *p53*-knockout cells would affect Wnt signaling pathway and to see whether a crosstalk occurs between Wnt signaling and enhanced apoptosis upon *Crys* knockout, first, we constructed a cross-talk network consisting of interactions among Wnt signaling pathway and circadian clock pathway genes (see Methods for details). Then, to observe the changes in gene

expression upon *Cry* knockout at a network level in detail, a sub-network is extracted from the cross-talk network (**Figure 5.2A**).

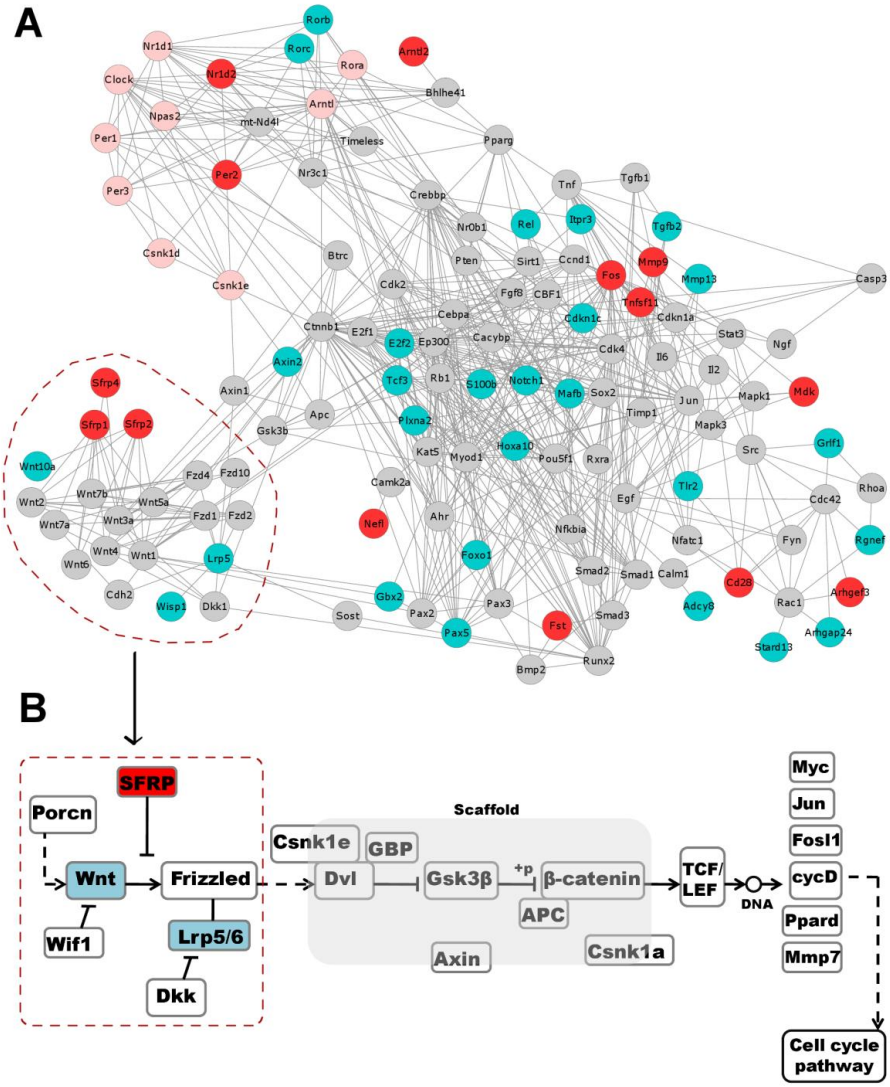

Figure 5.2 Wnt-Circadian subnetwork. (A) Circadian clock pathway genes are shown in pink color. Up-regulated and down-regulated genes upon *Cry* knockout after 4-hours of UV-irradiation are shown in red and blue color, respectively. Gray colored genes are non-differentially expressed genes either belonging to Wnt pathway or acting as linkers connecting circadian pathway genes to Wnt pathway genes. (B) The interactions between Wnt signaling genes are detailed according to KEGG pathway annotations. When *Cry* is knockout in *p53*$^{-/-}$ cells, the expression levels of the Wnt ligand; WNT10A and the Wnt receptor; LRP5 decreased more than to half and the Wnt antagonists; SFRP1, SFRP2 and SFRP4, which bind to Wnt proteins to prevent signal activation, show approximately two-fold increase in expression. No significant expression change in the scaffold complex is observed, although *Axin1* levels increased to 1.4-fold. *Axin2*, which provides a negative feedback loop to Wnt signaling, is downregulated. Upon transduction of a Wnt signal, first transcription of *Axin2* gene is induced via the β-catenin/TCF/LEF pathway, then it can function in the destruction complex to inhibit β-catenin expression [294]. Thus, with the decrease in Wnt signaling, *Axin2* level decreases.

This Wnt-circadian subnetwork (**Figure 5.2A**) consists of the interactions among the genes that show differential expression upon *Cry* knockout, circadian clock pathway genes and Caspase 3 gene. There are 124 genes and 568 interactions with an average degree of 9.1. We observe that expression levels of Wnt ligands and antagonists change significantly with the *Cry* knockout leading to inactivation of Wnt signaling (**Figure 5.2B** showing a directed network representation of Wnt signaling as a simplified version of KEGG pathway database [224]). In particular, expression levels of Wnt ligands are decreased by half. The expression levels of the Wnt receptor; LRP5 is also decreased significantly and the Wnt antagonists; SFRP1, SFRP2 and SFRP4, which bind to Wnt proteins to prevent signal activation, show approximately two-fold increase in expression. Upregulation and overexpression of SFRP proteins have been associated with increased apoptosis before [306, 307]. Additionally, blockade of Wnt signaling was shown to induce apoptosis in human colorectal cancer cells and as such may have a therapeutic role in the treatment of this disease [308].

The resulting downregulation of Wnt signaling upon *Cry* knockout is more likely to indicate that circadian clock and Wnt signaling pathways should intersect at critical points, which would consequently induce increased sensitivity of $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ cells to apoptosis. In fact, very recently, it was shown that *Cry* knockout in $p53^{-/-}$ cells leads to hyperphosphorylation of GSK3β and inactivation of GSK3β, which fails to activate the antiapoptotic effect of NF-κB, making cells more sensitive to TNFα-induced apoptosis [291] (The concept is visualized in **Figure 5.3A**). Our network-based analysis reveals that two NF-κB activators; NOTCH1 and TLR2, which are known to be regulated by GSK3β [309, 310], are down-regulated. TLR2 (toll-like receptor 2) was reported to activate NF-κB and impart an anti-apoptotic effect in stressed cardiac myocytes [311]. NOTCH1 being a NF-κB regulator [312], functions in Notch signaling pathway involved in cell proliferation and apoptosis [313]. Downregulation of NOTCH1 was reported to reduce NF-κB activity and thus inhibit cell growth and induce apoptosis in pancreatic cancer cells [312].

Looking at the downstream of NF-κB, in our Wnt-circadian subnetwork, we see that one of the downregulated Wnt ligands; WNT10A is a NF-κB target and its expression is induced by TNFα [314, 315]. Considering these all and expanding the finding presented recently [291], it is conceivable that GSK3β plays a central regulatory role and with *Cry* knockout it is inactivated leading to downregulation of NF-κB activators, which in turn

fails to activate NF-κB making cells more sensitive to TNFα-induced apoptosis. Then, with the decrease of NF-κB activity, Wnt signaling could become inactivated as well suppressing the cell growth and proliferation and enhancing apoptotic tendency. This observation is visualized in **Figure 5.3B**. Ample evidence shows the multifaceted nature and key roles of GSK3β in mediating signals among circadian clock and metabolic processes (visualized in **Figure 5.3B**). In the absence of Wnt signal, GSK3β acts with AXIN and APC proteins and mediates phosphorylation of β-catenin targeting it for degradation by proteosome, thus preventing anti-apoptotic function of β-catenin. It can also inhibit another survival-promoting transcription factor; HSF1 (heat shock factor 1), reducing its expression that can facilitate apoptosis [316]. GSK3β could also be the key protein connecting circadian signals to insulin signaling: it was observed to phosphorylate IRS1, thus impairing insulin action [317]. IRS proteins stimulate the activity of MAPK families such as MAPK13 (p38), that are involved in cell survival, differentiation and proliferation [318]. We observe that both IRS1 and MAPK13 levels are downregulated in *Cry* knockout animals. IRS proteins are known to bind the p110 catalytic subunit of the phosphatidylinositol 3-kinase (PI3K), which activates the kinase PDK1. PDK1 phosphorylate AKT resulting in its activation and AKT can phosphorylate GSK3β inhibiting its activity. Earlier work by Wu et al. underlines the multifunctional behavior of GSK3β and suggests that there exist three pools of GSK3β in cells: one pool is associated with AXIN regulated by Wnt, another pool is regulated by PI3K-AKT pathway and an additional AXIN-independent pool that is also regulated by Wnt [319]. In this study, we do not observe changes in PI3K-AKT pathway however our results imply that circadian clock disruption due to *Cry* knockout could affect Wnt signaling, insulin signaling, and MAPK13 signaling pathways and GSK3β appears to be the central regulatory protein in these processes.

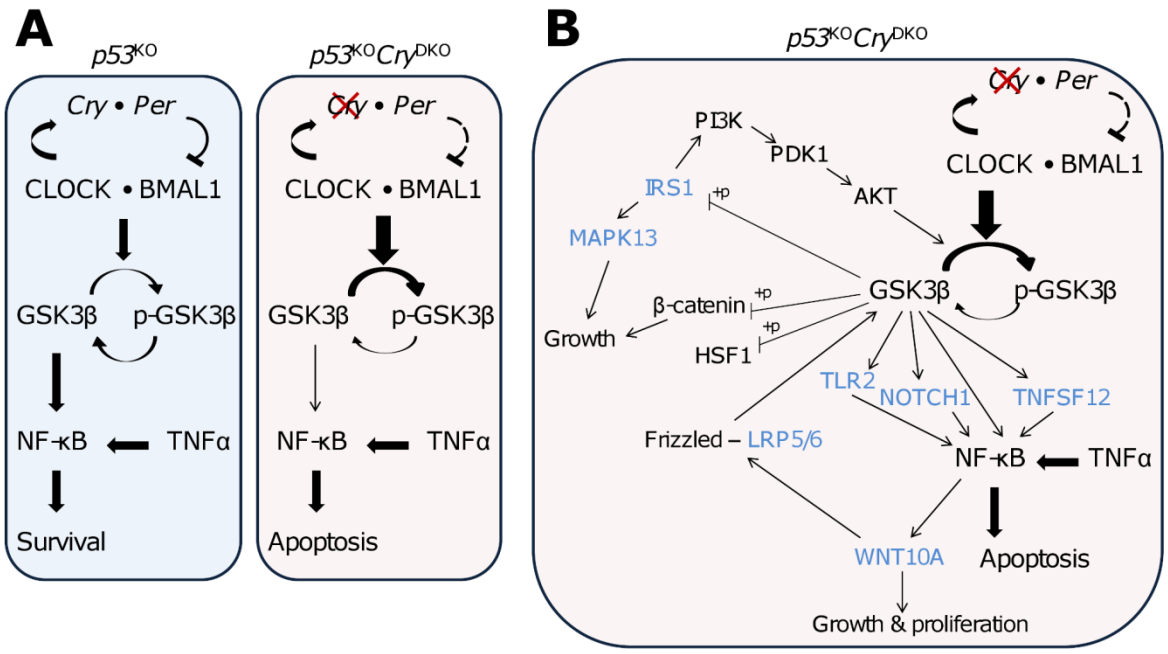

Figure 5.3 Representation of the relationship between circadian clock and apoptotic pathways. **(A)** A very recent study of Lee et al. [291] showed that *Cry* knockout in *p53* knockout cells resulted in hyperphosphorylation of GSK3β. Becoming inactive, GSK3β would fail to activate NF-κB, which makes cells more sensitive to apoptosis [291]. **(B)** Our network-based analysis combined with gene expression profiles reveals several genes which appear to play crucial roles in apoptosis. Following the finding of Lee et al. [291], we observe that NF-κB activators; TLR2, NOTCH1 and TNFSF12, which are known to be regulated by GSK3β, are downregulated (shown in blue color), hence would fail to activate NF-κB and enhance cell apoptosis. In addition, downstream target of NF-κB; WNT10A is downregulated, which could lead to a decrease in cell growth and proliferation. With the downregulation of WNT10A and LRP5, GSK3β could mediate the phosphorylation of β-catenin targeting it for degradation by proteosome, thus inhibiting the anti-apoptotic function of β-catenin. It can also inhibit another survival-promoting transcription factor; HSF1 (heat shock factor 1), reducing its expression that can facilitate apoptosis. GSK3β could also be the key protein connecting circadian signals to insulin signaling: it can phosphorylate IRS1, thus impairing insulin action. IRS proteins can stimulate the activity of MAPK13 (p38). Downregulation of both IRS1 and MAPK13 upon *Cry* knockout imply that GSK3β could play a key role through circadian clock to insulin and MAPK signaling and consequently suppressing cell growth.

Besides circadian and Wnt signaling genes, the subnetwork contains genes playing roles in cell development, proliferation and differentiation processes. For example, FOS (or C-FOS) gene being a proto-oncogene, which is up-regulated with *Cry* knockout according to our data, is known to play a critical role in proliferation, tumorigenesis, tumor invasion, and metastasis [320]. Continuous expression of FOS was demonstrated to precede apoptosis [321] and activation or overexpression of FOS is related to induction of apoptosis [322-325]. Another up-regulated gene is matrix metallopeptidase 9 (MMP9); a protease that can degrade or proteolytically modify the extracellular matrix components,

including collagens, laminin, and proteoglycans. Through proteolysis of these molecules, MMP proteins can up- or down-regulate apoptosis depending on the relative concentrations, tissue specificity and balance between MMP proteins and tissue inhibitors of metalloproteinases (TIMP) [326, 327]. Pro-apoptotic functions of MMP9 are observed before: it may trigger neuronal cell death [328] and its activation can induce apoptosis in human monocytic cells via extracellular release of TNFα and a soluble Fas ligand [329]. In our subnetwork, MMP9 is displayed to be in contact with its inhibitor; TIMP1, which shows no change in expression upon *Cry* knockout. Thus, elevated levels of MMP9 expression is more likely to be related with an increase in its pro-apoptotic activity, which triggers cell apoptosis. Other downregulated genes such as PLXNA, PAX5, GBX2 and RGNEF are related with cell growth and differentiation: PLXNA2 plays role in axon guidance and invasive growth; PAX5 functions in B-cell differentiation and neural development; GBX2 acts as a transcription factor for cell pluriptency and differentiation; and RGNEF is a Rho-guanine nucleotide exchange factor regulating signaling pathways downstream of integrins and growth factor receptors. Downregulation of these genes would suppress cell growth and differentiation processes as well.

5.3.2.2 Alterations in Apoptosis and p53 Signaling Pathways upon Cry Knockout

There are two major pathways of apoptosis: the extrinsic (death receptor pathway; FAS and other TNFR superfamily members and ligands) and the intrinsic (mitochondria-associated) pathways, which are linked and molecules in one pathway can influence the other [330]. To reveal how apoptotic pathway proteins and their interactions would be affected by the circadian clock disruption, circadian-apoptosis crosstalk network is constructed and combined with gene expression profiles. Analysis of differentially expressed genes and their protein-protein interactions indicate the importance of those in apoptosis pathway. We observe that apoptosis pathway interacts with many other pathways, alterations in which would consequently trigger or inhibit apoptosis.

In the apoptosis-circadian subnetwork, we show that there exist a number of critical paths combining CLOCK-BMAL1 complex with other cellular processes such as NF-κB signaling, MAPK signaling and TNFα-mediated signals (**Figure 5.4**). To label the so-called “critical paths”, we start from CLOCK-BMAL1 complex and label the interactions until reaching the proteins of differentially expressed genes in the subnetwork, giving

priority to regulatory interactions (such as activation, inhibition, phosphorylation events) if available and to highly interacting proteins.

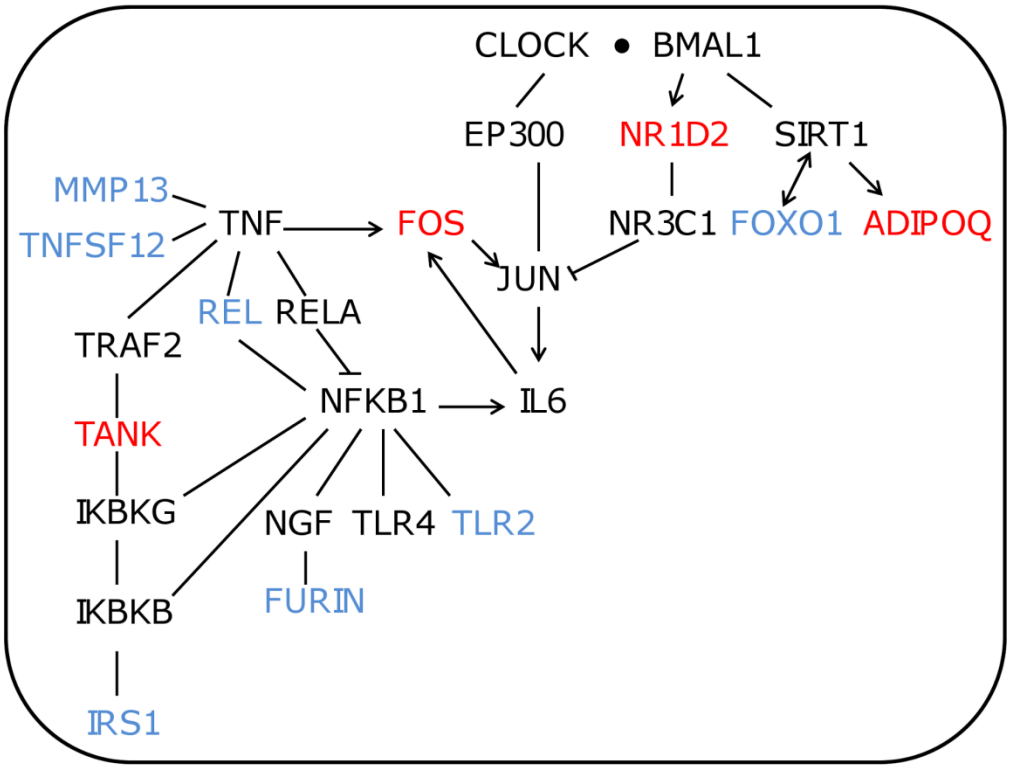

Figure 5.4 Critical paths in apoptosis-circadian subnetwork. From CLOCK-BMAL1 complex to proteins of differentially expressed genes, the key proteins connecting these are displayed and the interactions are annotated whenever possible; i.e. as activation or inhibition. Protein products of upregulated and downregulated genes are colored in red and blue, respectively. Highly interacting proteins are TNF and NFKB1 (NF-κB).

Proteins shown in **Figure 5.4** constitute the key proteins in apoptosis-circadian subnetwork and they frequently act in multiple cellular pathways. One of these is E1A binding protein P300; EP300 that associates with CLOCK to regulate positively clock gene expression [331]. EP300 is also a cofactor for JUN protein [332], which plays a key role in the induction of apoptosis [333]. JUN and JUN-regulated genes are known to be activated by FOS [334], which is found to be upregulated in our study (as discussed in the previous section). A very central protein in this subnetwork is TNF (TNFα), which mediates a variety of biological activities including cell proliferation and apoptosis. The specific response to TNF depends on cell type [335] and TNF signal transduction is mediated by TRAF2, which activates NF-κB. Traf-interacting protein; TANK forms a complex with

TRAF2 [336]. We observe that TANK is upregulated upon *Cry* knockout. Indeed, overexpression of TANK is known to inhibit TRAF2-mediated NF-κB activation [337], which would consequently enhance apoptosis. A member of the tumor necrosis factor (TNF) superfamily of cytokines; TNFSF12, which mediates NF-κB activation in several different cell types [338-342] is observed to be downregulated. TNFSF12 protein is also known to bind GSK3β, which is a key mediator of NF-κB pathway [343]. TNF binds to MMP13; matrix metalloprotease 13, which is a regulator of cell survival and downregulated upon *Cry* knockout. Downregulation of MMP13 is shown to reduce tumor growth in mouse osteosarcoma; the primary malignant cancer of bone [344]. In addition, MMP13 inhibition resulted in reduced cell growth in melanoma cells [345]. TNF and NF-κB signals also intersect with insulin signaling (see **Figure 5.4**). IRS1 (insulin receptor substrate 1), which is a cytoplasmic adaptor protein mediating insulin dependent mitogenesis and regulating glucose metabolism, is frequently associated with tumor growth and proliferation [346]. IRS1 is downregulated upon *Cry* knockout and consistent with this finding, earlier works indicate that suppression of IRS1 expression promotes apoptosis in breast carcinoma cells [347], whereas the overexpression of IRS1 confers resistance to TGF-β-induced cell death in hepatocellular carcinoma cells [348]. In **Figure 5.4**, we see that Nuclear receptor subfamily 1 group D member 2 Gene; NR1D2, which is known to be activated by BMAL1 (ARNTL), is upregulated. With the *Cry* knockout, BMAL1 expression increases which, in turn, would lead to an increase in NR1D2 levels. NR1D2 can form a complex with NR3C1 (nuclear receptor subfamily 3, group C, member 1 Gene), which is a receptor for glucocorticoids. NR3C1 can affect inflammatory responses and cellular proliferation and is a potent inhibitor of JUN activity [349]. Another protein binding to CLOCK-BMAL1 complex is SIRT1; a protein deacetylase that is required for high-magnitude circadian transcription of several core clock genes such as *Bmal1*, *Per2* and *Cry1* [350]. SIRT1 binds to CLOCK-BMAL1 complex and promotes the deacetylation and degradation of PER2 [350]. In **Figure 5.4**, SIRT1 is shown to activate FOXO1 and Adiponectin (ADIPOQ) proteins, which are downregulated and upregulated, respectively. Although the functional consequences of the interactions between FOXO1 and SIRT1 is not clearly understood [351], SIRT1 was found to deacetylate FOXO1 and enable activation of FOXO1 transcription and conversely, FOXO1 is a positive transcriptional regulator of SIRT1 [351]. Sirt1 also increases adiponectin transcription in adipocytes

[352]. Adiponectin, which is upregulated upon *Cry* knockout, is a cytokine that is expressed abundantly in adipose tissue [353-355] and has been shown to inhibit cell proliferation and induce apoptosis in leukemia cells [356] and to suppress tumor growth in mice [357]. It is also identified as a novel growth inhibitor in prostate cancer cells [358]. SIRT1 was found to be an effective inhibitor of NF-κB signaling as well [359], which makes it a central protein that could play crucial roles in relating circadian clock to other cellular processes. Apoptosis-circadian subnetwork is given in Supplementary Material.

We should note that the roles of some genes in cancer progression and metastasis are controversial: for example, for CCL2, an inflammatory chemokine, its levels correlate with tumor progression in human breast cancer [360, 361], whereas in pancreatic cancer, high levels of CCL2 invoke tumor destruction [362]. Then, the specific roles of such genes may be different depending on the cellular context and the tumor type. With the *Cry* knockout, we observe that the expression level of CCL2 is upregulated. The list of all differentially expressed genes upon *Cry* knockout in apoptosis-circadian crosstalk network is provided in **Table 5.2**.

If these genes serve such crucial functions to increase apoptosis in tumor cells through circadian rhythm, one would expect that they are either modulating clock or their expression levels are under the clock control. To check whether changes in those genes would affect circadian rhythms, we referred to the study of Zhang et al. [363], in which they conducted a genome-wide RNAi screen to identify clock genes and their modifiers through Biogps (http://biogps.org/). They found nearly 1000 genes whose knockdown resulted in reduced circadian rhythm amplitude and hundreds of genes whose knockdown led to long or short period length of oscillation or increased amplitude [363]. According to their findings, we observe that many of the differentially expressed genes in apoptosis-circadian crosstalk network (**Table 5.2**) could affect circadian rhythms upon their knockout and they are labeled in **Table 5.2**. Hence, these genes could have regulatory effects on circadian pathway and could be important in regulation of physiology and behavior in the whole organism. For example, NR1D2 is upregulated upon *Cry* knockout. Considering the results of Zhang et al. [363], knockout of NR1D2 would result in both period and amplitude change in circadian rhythms. Additionally, upon *Clock* mutation, its expression is observed to change significantly. NR1D2 is already known to regulate Bmal1 expression [364] and its expression levels are under the clock control [363]. Similarly, a member of

the tumor necrosis factor (TNF) superfamily of cytokines; TNFSF12, which is observed to be downregulated upon *Cry* knockout in this study, appears to affect circadian oscillations when knockout [363] and its expression changes significantly with *Clock* mutation as well [363].

Table 5.2 List of differentially expressed genes in apoptosis-circadian crosstalk network. Up- and down-regulated genes upon *Cry* knockout in *p53*-deficient cells are listed. Third column shows the fold change ratio of gene expressions when *Cry* is mutated (expression ratio of $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ cells / $p53^{-/-}$ cells). In the fourth column, the corresponding effects of expression change on apoptosis (either enhance or inhibit apoptosis) are provided. In the last column, the genes whose knockdown could affect circadian oscillations (in terms of a change in amplitude and/or a change in period length) are labelled according to the observations of Zhang et al. [363].

Gene symbol	Gene name	Fold-change ratio	Enhance (+) or inhibit (-) apoptosis & Ref.	Affect circadian rhythms (+) or not(-) as observed by Zhang et al. [363]
Up-regulated genes				
Adipoq	Adiponectin, C1Q and collagen domain containing	1.9	+ [356]	-
Arntl2	Aryl hydrocarbon receptor nuclear translocator-like 2	2.2		+
Ccl2	Chemokine (C-C motif) ligand 2	2.4	+ [362] & - [360, 361]	+
Cd14	Cd14 antigen	1.9		+
Cd28	Cd28 antigen	1.8		
Fos	FBJ osteosarcoma oncogene	1.9	+ [321-325]	-
Icam1	Intercellular adhesion molecule 1	2.4		+
Il1rn	Interleukin 1 receptor antagonist	2.9	- [365]	
Mmp9	Matrix metallopeptidase 9	3.2	+ [328]	+
Nr1d2	Nuclear receptor subfamily 1, group D, member 2	2.4		+
Nr5a2	Nuclear receptor subfamily 5, group A, member 2	1.9		+
Per2	Period homolog 2	3.4	+ [366]	+
Tank	TRAF family member-associated NF-κB activator	1.9	+ [337]	+
Down-regulated genes				
Adcy8	Adenylate cyclase 8	0.3		+
Axin2	Axis inhibition protein 2	0.4	+ [294]	
Irs1	Insulin receptor substrate 1	0.4	+ [347]	+
Foxo1	Forkhead box O1	0.5	+ [367]	+
Furin	Furin (paired basic amino acid cleaving enzyme)	0.5	+ [367]	+

Lpl	Lipoprotein lipase	0.4		+
Map3k1	Mitogen-activated protein kinase 1	0.5	+ [368] & - [369]	+
Mmp13	Matrix metallopeptidase 13	0.3	+ [344]	-
Prkar2b	Protein kinase, cAMP dependent regulatory, type II beta	0.5		
Rel	Reticuloendotheliosis oncogene	0.5		+
Rorb	RAR-related orphan receptor beta	0.4		+
Rorc	RAR-related orphan receptor gamma	0.3		+
Tlr2	Toll-like receptor 2	0.5	+ [309]	-
Tnfsf12	Tumor necrosis factor ligand superfamily, member 12	0.4	+ [338-342]	+

We also investigate the effects of circadian disruption on p53 signaling. p53 activation is induced by several stress signals, including DNA damage, oxidative stress and activated oncogenes and results in transcription of p53-regulated genes and three major outputs; cell cycle arrest, cellular senescence or apoptosis. The critical paths in p53-circadian subnetwork (shown in **Figure 5.5**) include TGFβ, Akt and Notch signaling. One of the key genes in **Figure 5.5** is EP300, (as mentioned above) acting as a connector gene interacting with both CLOCK and other crucial regulators such as NOTCH1, TCF3 and VEGFA. EP300 functions as a transcriptional coactivator for NOTCH1 [370], whose downregulation was shown to inhibit cell growth and induce apoptosis [312]. EP300 also associates with the transcription factor TCF3 and enhances TCF3-mediated transcription activity [371]. TCF3 is downregulated upon *Cry* knockout and earlier work indicated that reduction and inactivation of TCF3 was associated with apoptosis [372, 373]. In **Figure 5.5**, we see that growth factor signals are downregulated. One important regulator of cell growth is TGFB2 (transforming growth factor beta 2). Inhibition of the TGFB2 expression was reported to be an effective approach for malignant tumor therapy [374]. Following TGFB2 path, TNFSF11 (tumor necrosis factor ligand superfamily member 11), which is a receptor activator of NF-κB ligand, is upregulated with the *Cry* knockout. TNFSF11 appear to have both apoptotic and anti-apoptotic effects, probably dependent on the cell type [375]. Overexpression of TNFSF11 has been associated with excessive bone resorption in multiple myeloma [376] and it regulates osteoclast differentiation and provides a pro-survival signal [377-379]. On the contrary, TNFSF11 can suppress cell proliferation and induce apoptosis through TRAF6-dependent mechanism [375].

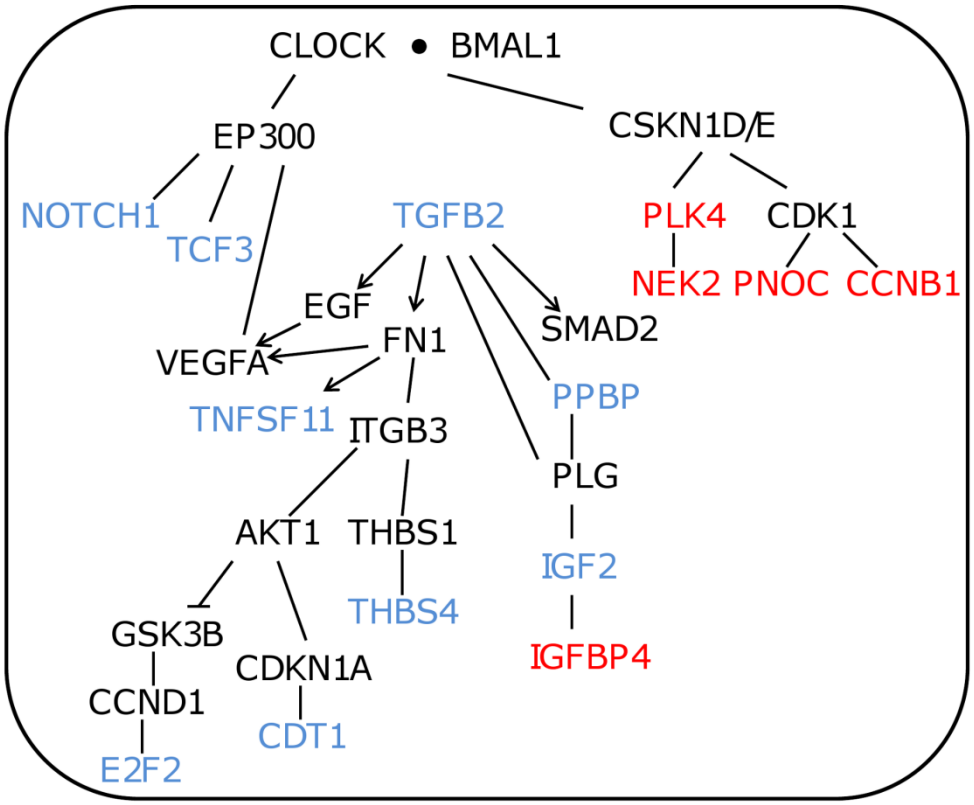

Figure 5.5 Critical paths in p53-circadian subnetwork. From CLOCK-BMAL1 complex to proteins of differentially expressed genes, the key proteins connecting these are displayed and the interactions are annotated whenever possible; i.e. as activation or inhibition. Protein products of upregulated and downregulated genes are colored in red and blue, respectively.

Other genes acting in cell proliferation; E2F2 (E2F transcription factor 2) and THBS4 (thrombospondin 4) are downregulated. Downregulation of E2F2 could lead to apoptosis since its normal expression levels can promote cell survival by reduction of the expression of the pro-apoptotic genes *Caspase-6* and *Apaf-1* [380]. CDT1 (chromatin licensing and DNA replication factor 1) was observed to be overexpressed and cause tumor formation in the absence of *p53* in mice cells [381]. With *Cry* knockout in $p53^{-/-}$ cells, CDT1 is downregulated, which can contribute to enhanced apoptosis. We observe that insulin signaling is also affected (shown in **Figure 5.5**): insulin-like growth factor 2; IGF2 is downregulated. Reducing the expression of IGF2 was shown to suppress the protection against apoptosis in pancreatic islet β-cells [382]. The expression of insulin-like growth factor binding protein 4; IGFBP4 is increased. Overexpression of IGFBP4 was found to

delay the growth of malignant prostate epithelial cells and enhance the sensitivity of these cells to apoptosis [383]. A similar effect was detected in colorectal cancer cells as well; IGFBP4 induction resulted in increased apoptosis and decreased proliferation of these cells [384]. In **Figure 5.5**, CSNK1E (casein kinase 1 epsilon gene) and CSNK1D (casein kinase 1 delta gene) are shown as clock connectors. These kinases are both central components of the circadian clock acting as a negative regulator of circadian rhythmicity by phosphorylating PER1 and PER2. They can also phosphorylate a large number of proteins. They can bind to several kinases such as PLK4 and CDK1 and would regulate many cellular processes.

Although aforementioned genes have pro-apoptotic effects, changes in p53-circadian crosstalk network also indicate the existence of pro-growth (and survival) signals. For example, PAX1 (paired box gene 1), which is a transcription factor, is upregulated with *Cry* knockout. However, upregulation of PAX1 appears to be related with cell survival since it is frequently expressed in tumor cell lines including breast, ovarian, lung and colon cancer and often required for cancer cell survival [385]. It is conceivable that some genes would promote survival upon *Cry* knockout in $p53^{-/-}$ cells as well, however, in overall; the balance tends to shift toward cell apoptosis. p53 signaling-circadian subnetwork is given in Supplementary Material. The list of all differentially expressed genes upon *Cry* knockout in p53-circadian crosstalk network is provided in **Table 5.3**. We should note that according to the study of Zhang et al. [363], most of the differentially expressed genes in p53-circadian crosstalk network (**Table 5.3**) could affect the circadian rhythms when they are knockout. The possible effect would be a change in amplitude or length of the period. These genes are labeled in **Table 5.3**.

Table 5.3 List of differentially expressed genes in p53-circadian crosstalk network. Up- and down-regulated genes upon *Cry* knockout in *p53*-deficient cells are listed. Third column shows the fold change ratio of gene expressions when *Cry* is mutated (expression ratio of $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ cells / $p53^{-/-}$ cells). In the last column, the corresponding effects of expression change on apoptosis (either enhance or inhibit apoptosis) are provided. In the last column, the genes whose knockdown could affect circadian oscillations (in terms of a change in amplitude and/or a change in period length) are labelled according to the observations of Zhang et al. [363].

	Gene symbol	Gene name	Fold-change ratio	Enhance (+) or inhibit (-) apoptosis & Ref.	Affect circadian rhythms (+) or not(-) as observed by Zhang et al. [363]
		Up-regulated genes			
	Arntl2	Aryl hydrocarbon receptor nuclear translocator-like 2	2.2		+
	Ccnb1	Cyclin B1	2.0		+
	Fos	FBJ osteosarcoma oncogene	1.9	+ [321-325]	-
	Igfbp4	Insulin-like growth factor binding protein 4	1.8	+ [383, 384]	+
	Nek2	Serine/threonine-protein kinase Nek2	1.8		
	Nr1d2	Nuclear receptor subfamily 1, group D, member 2	2.4		+
	Nr5a2	Nuclear receptor subfamily 5, group A, member 2	1.9		+
	Pax1	Paired box gene 1	2.7	- [385]	+
	Per2	Period homolog 2	3.4	+ [366]	+
	Plk4	Serine/threonine-protein kinase PLK4	1.8		+
	Pnoc	Prepronociceptin	1.8		
	Serping1	Serine peptidase inhibitor, clade G, member 1	2.1		
	Tnfsf11	Tumor necrosis factor ligand superfamily, member 11	2.3	+ [375] & - [377-379]	
		Down-regulated genes			
	Axin2	Axis inhibition protein 2	0.4	+ [294]	
	Cdkn1c	Cyclin-dependent kinase inhibitor 1C	0.3		
	Cdt1	Chromatin licensing and DNA replication factor 1	0.5	+ [381]	
	E2f2	E2F transcription factor 2	0.5	+ [380]	+
	Igf2	Insulin-like growth factor 2	0.5	+ [382]	
	Irs1	Insulin receptor substrate 1	0.4	+ [347]	+
	Foxo1	Forkhead box O1	0.5	+ [367]	+
	Mcm6	DNA replication licensing factor MCM6	0.4		+
	Notch1	Notch gene homolog 1	0.5	+ [312]	
	Ppbp	Pro-platelet basic protein	0.5		

	Rorb	RAR-related orphan receptor beta	0.4		+
	Rorc	RAR-related orphan receptor gamma	0.3		+
	Tcf3	Transcription factor 3	0.5	+ [372, 373]	+
	Tgfb2	Transforming growth factor, beta 2	0.5	+ [374]	
	Thbs4	Thrombospondin 4	0.3		-

5.3.3 Potential Effects of *Cry* Knockout in Wild-Type p53 Cells

In this study, we analyze potential effects of *Cry* knockout in *p53*-deficient primary mouse skin fibroblasts. Disruption of *Cry* function was suggested to improve the efficacy of chemotherapy for tumors with *p53* mutation by enhancing cell apoptosis [290]. However, in wild-type *p53* cells, *Cry* knockout may not produce similar effects. Very recently, Destici et al. [386] investigated the DNA damage sensitivity of $Cry1^{-/-}|Cry2^{-/-}$ primary mouse skin fibroblasts and interestingly they observed that the absence of *Cry* genes does not affect the cell-autonomous DNA damage response upon exposure of primary cells to genotoxic agents. We looked over their gene expression data (Wild-type vs. $Cry1^{-/-}|Cry2^{-/-}$ primary mouse skin fibroblasts) to examine whether similar changes in gene expression are observed for these cells as well. Only a few differentially expressed genes in our list match with theirs that are labeled as differentially expressed and up/down-regulation behavior appears to be similar. Only one of these is related to apoptotic process: EIF5A2, which is an mRNA-binding protein involved in translation elongation and has pro-apoptotic effects, is identified as up-regulated in both cases. For our differentially expressed genes critical in activating cell apoptosis in *p53*-deficient cells, we find no significant change in expression of them in the data of Destici et al. [386]. *Per2* gene is found to be up-regulated for both of the cases, which is an expected outcome of *Cry* knockout. Overexpression of *Per2* gene was found to induce cancer cell apoptosis before [366]. However, the same study emphasized that in non-tumorigenic cells, *Per2* overexpression neither inhibits cell growth nor induces apoptosis [366]. Consequently, *Cry* knockout in non-tumorigenic cells would possibly result in disruption of circadian rhythms but may not enhance cell apoptosis. Consistent with the study of Destici et al. [386] mentioned above, absence of *Cry* genes seems to affect apoptotic processes only in the context of an additional *p53* deficiency. This observation indicates the importance of targeting *Cry* gene in therapeutic strategies for treating *p53*-deficient tumors.

5.4 Concluding Remarks

Circadian clock is in communication with different signal transduction mechanisms that regulate the metabolism of organisms. Hence, a disruption in clock would lead to several alterations in cellular mechanisms. In this study, we perform a large-scale analysis to understand how a circadian clock disruption by *Cry* knockout would affect cellular pathways in *p53*-deficient mouse cell lines. We integrate gene expression profiles (as comparison of two conditions: $p53^{-/-}$ and $p53^{-/-}Cry1^{-/-}Cry2^{-/-}$ primary mouse skin fibroblasts) with protein-protein interaction networks. Analyzing the differentially expressed genes within the framework of their protein-protein interactions in the networks assists in identifying critical genes and pathways that are likely to be under circadian clock regulation. Very recently, GSK3β was observed to be a key regulatory protein connecting circadian clock to NF-κB pathway and with *Cry* knockout GSK3β would become inactive and fail to activate NF-κB, which makes *p53*-deficient cells more sensitive to apoptosis [291]. Our large-scale pathway based approach reveals other NF-κB activators, which are known to be regulated by GSK3β, and which could play crucial roles in promoting apoptosis upon *Cry* knockout. In addition, we observe that the expression of several apoptotic and anti-apoptotic genes change upon *Cry* knockout in *p53* knockout cells: apoptotic genes are found to be mostly up-regulated whereas anti-apoptotic genes are down-regulated. In overall; the balance tends to shift toward cell apoptosis.

In comparison with the gene expression profiles of a recent experiment of *Cry* knockout in wild-type *p53* cells [386], we see that there occur no significant change in the expression of these apoptotic and anti-apoptotic genes listed in this study. This finding is consistent with their argument that the absence of *Cry* genes does not affect DNA damage response upon exposure of primary cells to genotoxic agents in wild-type *p53* cells. Therefore, such alterations in apoptotic and survival pathways upon *Cry* knockout would emerge only in the context of an additional *p53* deficiency. Since most of the tumors contain a mutation in *p53* gene, circadian clock disruption through *Cry* knockout would improve the efficacy of treatments of cancers associated with *p53* deficiency.

Chapter 6

CONCLUSION

The main focus of this dissertation has been the analysis of protein-protein interactions and their pathways in the guidance of additional "omics" data. Ubiquitination pathway and circadian clock pathway are studied integrating protein structural data and microarray gene expression data at a large-scale, respectively.

At proteome-level, we apply a combinatorial docking approach to predict functional association of proteins in human ubiquitination pathway. This approach (implemented as Prism algorithm) is based on the origin that the number of protein-protein interface motifs is limited in nature, therefore, same interface motifs can be used by interacting pairs repeatedly even though proteins have structurally different global folds. Thus, known interface architectures can be used to model the complexes between two target proteins, independent of their global structures. By exploiting the available structural proteome and Prism algorithm, we construct a structural E2-E3 interaction network with a prediction accuracy of 76%, indicating the predictions are in good agreement with functional E2-E3 pairs. We also perform an analysis at a molecular-level, in particular, elucidate binding patterns: some residues are structurally conserved among E2 proteins and appear to be essential for all E2−E3 interactions, whereas others, particularly in loop L1, appear to play important roles in E3 selectivity. We believe that the structural E2-E3 network and the E2-E3 interface data in this study provide a resource for future studies of ubiquitination and E2-E3 selectivity, especially in discovering drug candidates targeting E3s.

In the future, in addition to our E2-E3 modeling, the interactions of E3s with substrates can be examined and modeled to obtain a more complete picture of the ubiquitination pathway. Obtaining the three-dimensional model of the E2-E3-substrate complexes can provide crucial information about the ubiquitin system and could give an opportunity for understanding diseases associated with the mutations/alterations in E3 ligases.

Integration of protein structural information at a large-scale can also contribute significantly to drug design studies. Analysis of protein interfaces would provide insight into interaction behavior and specificity. Hot spots are particularly the targets of pharmaceutical agents and crucial for interaction specificity. We illustrate the correspondence of computational hot spot residues to experimental critical residues in binding related to disease providing several examples.

Combining microarray gene expression profiling data with protein-protein interaction networks would reveal important aspects of protein function and regulation. We provide large-scale integration of microarray data with circadian clock pathway and pathways related to cell growth and apoptosis. We focus on p53-deficient background and investigate how an additional circadian clock disruption would affect the apoptotic pathways. Our findings reveal the critical genes that are more likely to function as crosstalk genes regulating the information flow between circadian clock and apoptotic pathways. This study would ultimately assist in identifying targets in treatment of cancers associated with p53-deficiency.

As a future work, the results from our network-based microarray data analysis can be further assessed by experiments. For a set of critical genes that appear to regulate apoptotic processes upon circadian clock disruption, real-time polymerase chain reaction (qPCR) experiments can be performed to obtain more accurate measurements of their transcription. Since the mRNA and protein levels in the cell are often not correlated, applying the western blot technique can provide an efficient means for detecting the levels of specific proteins in the absence of p53 and Cry genes.

In overall, considerable information is gained towards protein recognition and function and we believe that this work will contribute to further studies of functional genomics and drug design.

Appendix A

APPENDIX

A.1. Webservers, Softwares, Tools, Databases

A.1.1. HOTPOINT

Hotpoint predicts hot spots in protein interfaces using an empirical model. The empirical model incorporates a few simple rules consisting of occlusion from solvent and total knowledge-based pair potentials of residues. The prediction model is computationally efficient and achieves high accuracy of 70%. The input to the HotPoint server is a protein complex and two chain identifiers that form an interface. The server provides the hot spot prediction results, a table of residue properties and an interactive 3D visualization of the complex with hot spots highlighted. This web server can be used for analysis of any protein-protein interface which can be utilized by researchers working on binding sites characterization and rational design of small molecules for protein interactions. HotPoint is accessible at http://prism.ccbb.ku.edu.tr/hotpoint [192].

A.1.2. NACCESS

Naccess calculates the accessible surface areas of the molecules rolling a solvent probe on the desired molecule. The default value for the radius of the solvent is 1.4 Å. The path gained by the center of the probe gives the accessible surface area. In addition to the accessible surface area, the output file of the Naccess gives also relative accessible area for each individual residue. Relative accessibility can be described as the percent accessibility of a residue relative to the accessibility of it in the extended conformation. If relative accessibility is larger than 5% then, this residue is identified as surface residue.

A.1.3. MULTIPROT

Multiprot is fully automated software which identifies multiple structural alignments of a given set of protein structures. Structural alignment method is based on the Geometric

Hashing Algorithm which detects common parts of the given structures in all possible ways. This is a sequence-order and directionality independent algorithm. Multiprot considers only C^{α} atoms and in the output file, the matched residue pairs, number of them and the RMSD value between these residues are provided. The algorithm does not force all residues to participate in the alignment; in contrast, it searches the best scored partial alignment for the given structures. It has a sequence order independent feature that makes Multiprot appropriate for protein interfaces analysis. Multiprot is used both in clustering part and in cluster type separation part [226].

A.1.4. CytoScape

Cytoscape is molecular interaction network visualization software which also intergrates biological information such as gene expression profiles, GO annotations etc. Additional features like network analyzer, functional enrichment generator, and additional file format support can be installed as plugins. Cytoscape user can visualize the protein – protein interaction network or other networks by loading .sif file which contains pairwise interaction information. Network visualization properties such as node shape, color, edge shape, color etc. can be defined by the user. It has also various filtering and selection tools. The more, the resulting graph can be organized several layouts such as hierarchical layout, spring embedded layout, circular layout etc. Here, we used Cytoscape for visualization of functional interaction network of PDB. Cytoscape is downloadable through the web page http://www.cytoscape.org/.

A.1.5. VMD

VMD is a molecule visualization and analysis tool. Biological systems such as proteins, nucleic acids, lipid bilayer assemblies, etc. can be visualized by the help of VMD. VMD can read standard Protein Data Bank (PDB) files and display the contained structure. It can be used also to animate and analyze the trajectory of molecular dynamics (MD) simulations, and can interactively manipulate molecules being simulated on remote computers (Interactive MD) [387].

A.1.6. FiberDock

FiberDock is a flexible refinement program for docking [153]. It models both side-chain and backbone flexibility and performs rigid body optimization on the ligand orientation. The movements of the backbone and side-chain are modeled according to the binding van der Waals forces between the receptor and ligand. The method uses both low and high frequency normal modes and therefore it can model both global and local conformational changes. After refining all the docking solution candidates, the refined models are re-ranked according to an energy function. FiberDock is available at http://bioinfo3d.cs.tau.ac.il/FiberDock/.

BIBLIOGRAPHY

1. Keskin O, Gursoy A, Ma B, Nussinov R: Principles Of Protein-Protein Interactions: What Are The Preferred Ways For Proteins To Interact? *Chem Rev* 2008, 108(4):1225-1244.

2. Shoemaker Ba, Panchenko Ar: Deciphering Protein-Protein Interactions. Part Ii. Computational Methods To Predict Protein And Domain Interaction Partners. *Plos Comput Biol* 2007, 3(4):E43.

3. Ideker T, Bafna V, Lemberger T: Integrating Scientific Cultures. *Mol Syst Biol* 2007, 3:105.

4. Joyce Ar, Palsson Bo: The Model Organism As A System: Integrating 'Omics' Data Sets. *Nat Rev Mol Cell Biol* 2006, 7(3):198-210.

5. Ulitsky I, Shamir R: Identifying Functional Modules Using Expression Profiles And Confidence-Scored Protein Interactions. *Bioinformatics* 2009, 25(9):1158-1164.

6. Grunberg R, Leckner J, Nilges M: Complementarity Of Structure Ensembles In Protein-Protein Binding. *Structure* 2004, 12(12):2125-2136.

7. Lawrence Mc, Colman Pm: Shape Complementarity At Protein/Protein Interfaces. *J Mol Biol* 1993, 234(4):946-950.

8. Lo Conte L, Chothia C, Janin J: The Atomic Structure Of Protein-Protein Recognition Sites. *J Mol Biol* 1999, 285(5):2177-2198.

9. Sheinerman Fb, Norel R, Honig B: Electrostatic Aspects Of Protein-Protein Interactions. *Curr Opin Struct Biol* 2000, 10(2):153-159.

10. Zhang Q, Sanner M, Olson Aj: Shape Complementarity Of Protein-Protein Complexes At Multiple Resolutions. *Proteins* 2009, 75(2):453-467.

11. Keskin O, Tsai Cj, Wolfson H, Nussinov R: A New, Structurally Nonredundant, Diverse Data Set Of Protein-Protein Interfaces And Its Implications. *Protein Sci* 2004, 13(4):1043-1055.

12. Gong S, Yoon G, Jang I, Bolser D, Dafas P, Schroeder M, Choi H, Cho Y, Han K, Lee S *Et Al*: Psibase: A Database Of Protein Structural Interactome Map (Psimap). *Bioinformatics* 2005, 21(10):2541-2543.

13. Gadkari Ra, Varughese D, Srinivasan N: Recognition Of Interaction Interface Residues In Low-Resolution Structures Of Protein Assemblies Solely From The Positions Of C(Alpha) Atoms. *Plos One* 2009, 4(2):E4476.

14. Jones S, Thornton Jm: Principles Of Protein-Protein Interactions. *Proc Natl Acad Sci U S A* 1996, 93(1):13-20.

15. Jones S, Marin A, Thornton Jm: Protein Domain Interfaces: Characterization And Comparison With Oligomeric Protein Interfaces. *Protein Eng* 2000, 13(2):77-82.

16. Jones S, Thornton Jm: Analysis Of Protein-Protein Interaction Sites Using Surface Patches. *J Mol Biol* 1997, 272(1):121-132.

17. Keskin O, Ma B, Rogale K, Gunasekaran K, Nussinov R: Protein-Protein Interactions: Organization, Cooperativity And Mapping In A Bottom-Up Systems Biology Approach. *Phys Biol* 2005, 2(2):S24-35.

18. Chakrabarti P, Janin J: Dissecting Protein-Protein Recognition Sites. *Proteins* 2002, 47(3):334-343.

19. Chothia C, Janin J: Principles Of Protein-Protein Recognition. *Nature* 1975, 256(5520):705-708.

20. Aytuna As, Gursoy A, Keskin O: Prediction Of Protein-Protein Interactions By Combining Structure And Sequence Conservation In Protein Interfaces. *Bioinformatics* 2005, 21(12):2850-2855.

21. Xu D, Tsai Cj, Nussinov R: Hydrogen Bonds And Salt Bridges Across Protein-Protein Interfaces. *Protein Eng* 1997, 10(9):999-1012.

22. Tsai Cj, Nussinov R: Hydrophobic Folding Units At Protein-Protein Interfaces: Implications To Protein Folding And To Protein-Protein Association. *Protein Sci* 1997, 6(7):1426-1437.

23. Tsai Cj, Xu D, Nussinov R: Structural Motifs At Protein-Protein Interfaces: Protein Cores Versus Two-State And Three-State Model Complexes. *Protein Sci* 1997, 6(9):1793-1805.

24. Negi Ss, Braun W: Statistical Analysis Of Physical-Chemical Properties And Prediction Of Protein-Protein Interfaces. *J Mol Model* 2007, 13(11):1157-1167.

25. Gruber J, Zawaira A, Saunders R, Barrett Cp, Noble Me: Computational Analyses Of The Surface Properties Of Protein-Protein Interfaces. *Acta Crystallogr D Biol Crystallogr* 2007, 63(Pt 1):50-57.

26. Headd Jj, Ban Ye, Brown P, Edelsbrunner H, Vaidya M, Rudolph J: Protein-Protein Interfaces: Properties, Preferences, And Projections. *J Proteome Res* 2007, 6(7):2576-2586.

27. De S, Krishnadev O, Srinivasan N, Rekha N: Interaction Preferences Across Protein-Protein Interfaces Of Obligatory And Non-Obligatory Components Are Different. *Bmc Struct Biol* 2005, 5:15.

28. Teyra J, Pisabarro Mt: Characterization Of Interfacial Solvent In Protein Complexes And Contribution Of Wet Spots To The Interface Description. *Proteins* 2007, 67(4):1087-1095.

29. Rodier F, Bahadur Rp, Chakrabarti P, Janin J: Hydration Of Protein-Protein Interfaces. *Proteins* 2005, 60(1):36-45.

30. Kar G, Keskin O, Gursoy A, Nussinov R: Allostery And Population Shift In Drug Discovery. *Curr Opin Pharmacol* 2010, 10(6):715-722.

31. Ozbabacan Se, Engin Hb, Gursoy A, Keskin O: Transient Protein-Protein Interactions. *Protein Eng Des Sel* 2011, 24(9):635-648.

32. Guharoy M, Chakrabarti P: Secondary Structure Based Analysis And Classification Of Biological Interfaces: Identification Of Binding Motifs In Protein-Protein Interactions. *Bioinformatics* 2007, 23(15):1909-1918.

33. Mintseris J, Weng Z: Structure, Function, And Evolution Of Transient And Obligate Protein-Protein Interactions. *Proc Natl Acad Sci U S A* 2005, 102(31):10930-10935.

34. Ofran Y, Rost B: Analysing Six Types Of Protein-Protein Interfaces. *J Mol Biol* 2003, 325(2):377-387.

35. Vacic V, Uversky Vn, Dunker Ak, Lonardi S: Composition Profiler: A Tool For Discovery And Visualization Of Amino Acid Composition Differences. *Bmc Bioinformatics* 2007, 8:211.

36. Zhanhua C, Gan Jg, Lei L, Sakharkar Mk, Kangueane P: Protein Subunit Interfaces: Heterodimers Versus Homodimers. *Bioinformation* 2005, 1(2):28-39.

37. Tuncbag N, Gursoy A, Guney E, Nussinov R, Keskin O: Architectures And Functional Coverage Of Protein-Protein Interfaces. *J Mol Biol* 2008, 381(3):785-802.

38. Haliloglu T, Keskin O, Ma B, Nussinov R: How Similar Are Protein Folding And Protein Binding Nuclei? Examination Of Vibrational Motions Of Energy Hot Spots And Conserved Residues. *Biophys J* 2005, 88(3):1552-1559.

39. Keskin O, Nussinov R: Similar Binding Sites And Different Partners: Implications To Shared Proteins In Cellular Pathways. *Structure* 2007, 15(3):341-354.

40. Veselovsky Av, Ivanov Yd, Ivanov As, Archakov Ai, Lewi P, Janssen P: Protein-Protein Interactions: Mechanisms And Modification By Drugs. *J Mol Recognit* 2002, 15(6):405-422.

41. Tsuchiya Y, Nakamura H, Kinoshita K: Discrimination Between Biological Interfaces And Crystal-Packing Contacts. *Advances And Applications In Bioinformatics And Chemistry* 2008, 2008(1):99 - 113.

42. Kuhn D, Weskamp N, Schmitt S, Hullermeier E, Klebe G: From The Similarity Analysis Of Protein Cavities To The Functional Classification Of Protein Families Using Cavbase. *J Mol Biol* 2006, 359(4):1023-1044.

43. Binkowski Ta, Naghibzadeh S, Liang J: Castp: Computed Atlas Of Surface Topography Of Proteins. *Nucleic Acids Res* 2003, 31(13):3352-3355.

44. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J: Castp: Computed Atlas Of Surface Topography Of Proteins With Structural And Topographical Mapping Of Functionally Annotated Residues. *Nucleic Acids Res* 2006, 34(Web Server Issue):W116-118.

45. An J, Totrov M, Abagyan R: Pocketome Via Comprehensive Identification And Classification Of Ligand Binding Envelopes. *Mol Cell Proteomics* 2005, 4(6):752-761.

46. Li X, Keskin O, Ma B, Nussinov R, Liang J: Protein-Protein Interactions: Hot Spots And Structurally Conserved Residues Often Locate In Complemented Pockets That Pre-Organized In The Unbound States: Implications For Docking. *J Mol Biol* 2004, 344(3):781-795.

47. Veselovsky Av, Archakov Ai: Inhibitors Of Protein-Protien Interactions As Potential Drugs. *Curr Comput Aided Drug Des* 2007, 3(1):51-58.

48. Arkin M: Protein-Protein Interactions And Cancer: Small Molecules Going In For The Kill. *Curr Opin Chem Biol* 2005, 9(3):317-324.

49. Lipinski Ca, Lombardo F, Dominy Bw, Feeney Pj: Experimental And Computational Approaches To Estimate Solubility And Permeability In Drug Discovery And Development Settings. *Adv Drug Deliv Rev* 2001, 46(1-3):3-26.

50. Walters Wp, Ajay, Murcko Ma: Recognizing Molecules With Drug-Like Properties. *Curr Opin Chem Biol* 1999, 3(4):384-387.

51. Chene P: Drugs Targeting Protein-Protein Interactions. *Chemmedchem* 2006, 1(4):400-411.

52. Bradford Jr, Needham Cj, Bulpitt Aj, Westhead Dr: Insights Into Protein-Protein Interfaces Using A Bayesian Network Prediction Method. *J Mol Biol* 2006, 362(2):365-386.

53. Larsen Ta, Olson Aj, Goodsell Ds: Morphology Of Protein-Protein Interfaces. *Structure* 1998, 6(4):421-427.

54. Nguyen Mn, Rajapakse Jc, Duan K: Amino Acid Features For Prediction Of Protein-Protein Interface Residues With Support Vector Machines. *Proceedings Of Evobio* 2007, 2007:187 - 196.

55. Zhanhua C, Gan Jg, Lei L, Mathura Vs, Sakharkar Mk, Kangueane P: Identification Of Critical Heterodimer Protein Interface Parameters By Multi-Dimensional Scaling In Euclidian Space. *Front Biosci* 2005, 10:844-852.

56. Von Eichborn J, Gunther S, Preissner R: Structural Features And Evolution Of Protein-Protein Interactions. *Genome Inform* 2010, 22:1-10.

57. Baussand J, Camproux Ac: Deciphering The Shape And Deformation Of Secondary Structures Through Local Conformation Analysis. *Bmc Struct Biol* 2011, 11:9.

58. Neuvirth H, Raz R, Schreiber G: Promate: A Structure Based Prediction Program To Identify The Location Of Protein-Protein Binding Sites. *J Mol Biol* 2004, 338(1):181-199.

59. Horton N, Lewis M: Calculation Of The Free Energy Of Association For Protein Complexes. *Protein Sci* 1992, 1(1):169-181.

60. Tuncbag N, Kar G, Gursoy A, Keskin O, Nussinov R: Towards Inferring Time Dimensionality In Protein-Protein Interaction Networks By Integrating Structures: The P53 Example. *Mol Biosyst* 2009, 5(12):1770-1778.

61. Henrick K, Thornton Jm: Pqs: A Protein Quaternary Structure File Server. *Trends Biochem Sci* 1998, 23(9):358-361.

62. Janin J: Specific Versus Non-Specific Contacts In Protein Crystals. *Nat Struct Biol* 1997, 4(12):973-974.

63. Wodak Sj, Mendez R: Prediction Of Protein-Protein Interactions: The Capri Experiment, Its Evaluation And Implications. *Curr Opin Struct Biol* 2004, 14(2):242-249.

64. Norel R, Petrey D, Wolfson Hj, Nussinov R: Examination Of Shape Complementarity In Docking Of Unbound Proteins. *Proteins* 1999, 36(3):307-317.

65. Gray Jj, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl Ca, Baker D: Protein-Protein Docking With Simultaneous Optimization Of Rigid-Body Displacement And Side-Chain Conformations. *J Mol Biol* 2003, 331(1):281-299.

66. Cheng Ac, Coleman Rg, Smyth Kt, Cao Q, Soulard P, Caffrey Dr, Salzberg Ac, Huang Es: Structure-Based Maximal Affinity Model Predicts Small-Molecule Druggability. *Nat Biotechnol* 2007, 25(1):71-75.

67. Smith Rd, Hu L, Falkner Ja, Benson Ml, Nerothin Jp, Carlson Ha: Exploring Protein-Ligand Recognition With Binding Moad. *J Mol Graph Model* 2006, 24(6):414-425.

68. Valdar Ws, Thornton Jm: Conservation Helps To Identify Biologically Relevant Crystal Contacts. *J Mol Biol* 2001, 313(2):399-416.

69. Zhu H, Domingues Fs, Sommer I, Lengauer T: Noxclass: Prediction Of Protein-Protein Interaction Types. *Bmc Bioinformatics* 2006, 7:27.

70. Bernauer J, Bahadur Rp, Rodier F, Janin J, Poupon A: Dimovo: A Voronoi Tessellation-Based Method For Discriminating Crystallographic And Biological Protein-Protein Interactions. *Bioinformatics* 2008, 24(5):652-658.

71. Liu S, Li Q, Lai L: A Combinatorial Score To Distinguish Biological And Nonbiological Protein-Protein Interfaces. *Proteins* 2006, 64(1):68-78.

72. Liu Q, Li J: Propensity Vectors Of Low-Asa Residue Pairs In The Distinction Of Protein Interactions. *Proteins* 2010, 78(3):589-602.

73. Shakhnovich Be, Dokholyan Nv, Delisi C, Shakhnovich Ei: Functional Fingerprints Of Folds: Evidence For Correlated Structure-Function Evolution. *J Mol Biol* 2003, 326(1):1-9.

74. Torrance Jw, Bartlett Gj, Porter Ct, Thornton Jm: Using A Library Of Structural Templates To Recognise Catalytic Sites And Explore Their Evolution In Homologous Families. *J Mol Biol* 2005, 347(3):565-581.

75. Caffrey Dr, Somaroo S, Hughes Jd, Mintseris J, Huang Es: Are Protein-Protein Interfaces More Conserved In Sequence Than The Rest Of The Protein Surface? *Protein Sci* 2004, 13(1):190-202.

76. Thorn Ks, Bogan Aa: Asedb: A Database Of Alanine Mutations And Their Effects On The Free Energy Of Binding In Protein Interactions. *Bioinformatics* 2001, 17(3):284-285.

77. Pupko T, Bell Re, Mayrose I, Glaser F, Ben-Tal N: Rate4site: An Algorithmic Tool For The Identification Of Functional Regions In Proteins By Surface Mapping Of Evolutionary Determinants Within Their Homologues. *Bioinformatics* 2002, 18 Suppl 1:S71-77.

78. Grishin Nv, Phillips Ma: The Subunit Interfaces Of Oligomeric Enzymes Are Conserved To A Similar Extent To The Overall Protein Sequences. *Protein Sci* 1994, 3(12):2455-2458.

79. Ouzounis C, Perez-Irratxeta C, Sander C, Valencia A: Are Binding Residues Conserved? *Pac Symp Biocomput* 1998:401-412.

80. Bartlett Gj, Porter Ct, Borkakoti N, Thornton Jm: Analysis Of Catalytic Residues In Enzyme Active Sites. *J Mol Biol* 2002, 324(1):105-121.

81. Ma B, Elkayam T, Wolfson H, Nussinov R: Protein-Protein Interactions: Structurally Conserved Residues Distinguish Between Binding Sites And Exposed Protein Surfaces. *Proc Natl Acad Sci U S A* 2003, 100(10):5772-5777.

82. Konc J, Janezic D: Protein-Protein Binding-Sites Prediction By Protein Surface Structure Conservation. *J Chem Inf Model* 2007, 47(3):940-944.

83. Watson Jd, Laskowski Ra, Thornton Jm: Predicting Protein Function From Sequence And Structural Data. *Curr Opin Struct Biol* 2005, 15(3):275-284.

84. Gonzalez-Ruiz D, Gohlke H: Targeting Protein-Protein Interactions With Small Molecules: Challenges And Perspectives For Computational Binding Epitope Detection And Ligand Finding. *Curr Med Chem* 2006, 13(22):2607-2625.

85. Wang B, Ge Ls, Huang D, Wong Hs: Prediction Of Protein-Protein Interacting Sites By Combining Svm Algorithm With Bayesian Method. *Icnc 2007* 2007, 2:329 - 333.

86. Laskowski Ra, Watson Jd, Thornton Jm: Protein Function Prediction Using Local 3d Templates. *J Mol Biol* 2005, 351(3):614-626.

87. Laskowski Ra, Watson Jd, Thornton Jm: Profunc: A Server For Predicting Protein Function From 3d Structure. *Nucleic Acids Res* 2005, 33(Web Server Issue):W89-93.

88. Kortemme T, Baker D: A Simple Physical Model For Binding Energy Hot Spots In Protein-Protein Complexes. *Proc Natl Acad Sci U S A* 2002, 99(22):14116-14121.

89. Chung Jl, Wang W, Bourne Pe: Exploiting Sequence And Structure Homologs To Identify Protein-Protein Binding Sites. *Proteins* 2006, 62(3):630-640.

90. Keskin O, Ma B, Nussinov R: Hot Regions In Protein--Protein Interactions: The Organization And Contribution Of Structurally Conserved Hot Spot Residues. *J Mol Biol* 2005, 345(5):1281-1294.

91. Illergard K, Ardell Dh, Elofsson A: Structure Is Three To Ten Times More Conserved Than Sequence--A Study Of Structural Response In Protein Cores. *Proteins* 2009, 77(3):499-508.

92. Valas Re, Yang S, Bourne Pe: Nothing About Protein Structure Classification Makes Sense Except In The Light Of Evolution. *Curr Opin Struct Biol* 2009, 19(3):329-334.

93. Zhang Qc, Petrey D, Norel R, Honig Bh: Protein Interface Conservation Across Structure Space. *Proc Natl Acad Sci U S A* 2010, 107(24):10896-10901.

94. Tsai Cj, Lin Sl, Wolfson Hj, Nussinov R: A Dataset Of Protein-Protein Interfaces Generated With A Sequence-Order-Independent Comparison Technique. *J Mol Biol* 1996, 260(4):604-620.

95. Marchler-Bauer A, Anderson Jb, Deweese-Scott C, Fedorova Nd, Geer Ly, He S, Hurwitz Di, Jackson Jd, Jacobs Ar, Lanczycki Cj *Et Al*: Cdd: A Curated Entrez Database Of Conserved Domain Alignments. *Nucleic Acids Res* 2003, 31(1):383-387.

96. Marchler-Bauer A, Panchenko Ar, Shoemaker Ba, Thiessen Pa, Geer Ly, Bryant Sh: Cdd: A Database Of Conserved Domain Alignments With Links To Domain Three-Dimensional Structure. *Nucleic Acids Res* 2002, 30(1):281-283.

97. Kim Wk, Henschel A, Winter C, Schroeder M: The Many Faces Of Protein-Protein Interactions: A Compendium Of Interface Geometry. *Plos Comput Biol* 2006, 2(9):E124.

98. Korkin D, Davis Fp, Sali A: Localization Of Protein-Binding Sites Within Families Of Proteins. *Protein Sci* 2005, 14(9):2350-2360.

99. Stein A, Ceol A, Aloy P: 3did: Identification And Classification Of Domain-Based Interactions Of Known Three-Dimensional Structure. *Nucleic Acids Res* 2011, 39(Database Issue):D718-723.

100. Aung Z, Tan Sh, Ng Sk, Tan Kl: Ppiclust: Efficient Clustering Of 3d Protein-Protein Interaction Interfaces. *J Bioinform Comput Biol* 2008, 6(3):415-433.

101. Gao M, Skolnick J: Structural Space Of Protein-Protein Interfaces Is Degenerate, Close To Complete, And Highly Connected. *Proc Natl Acad Sci U S A* 2010, 107(52):22517-22522.

102. Martin J: Beauty Is In The Eye Of The Beholder: Proteins Can Recognize Binding Sites Of Homologous Proteins In More Than One Way. *Plos Comput Biol* 2010, 6(6):E1000821.

103. Clackson T, Wells Ja: A Hot Spot Of Binding Energy In A Hormone-Receptor Interface. *Science* 1995, 267(5196):383-386.

104. Bogan Aa, Thorn Ks: Anatomy Of Hot Spots In Protein Interfaces. *J Mol Biol* 1998, 280(1):1-9.

105. Shulman-Peleg A, Shatsky M, Nussinov R, Wolfson Hj: Spatial Chemical Conservation Of Hot Spot Interactions In Protein-Protein Complexes. *Bmc Biol* 2007, 5:43.

106. Moreira Is, Fernandes Pa, Ramos Mj: Hot Spots--A Review Of The Protein-Protein Interface Determinant Amino-Acid Residues. *Proteins* 2007, 68(4):803-812.

107. Arkin Mr, Wells Ja: Small-Molecule Inhibitors Of Protein-Protein Interactions: Progressing Towards The Dream. *Nat Rev Drug Discov* 2004, 3(4):301-317.

108. Hu Z, Ma B, Wolfson H, Nussinov R: Conservation Of Polar Residues As Hot Spots At Protein Interfaces. *Proteins* 2000, 39(4):331-342.

109. Ahmad S, Keskin O, Mizuguchi K, Sarai A, Nussinov R: Ccrxp: Exploring Clusters Of Conserved Residues In Protein Structures. *Nucleic Acids Res* 2010, 38(Web Server Issue):W398-401.

110. Halperin I, Wolfson H, Nussinov R: Protein-Protein Interactions; Coupling Of Structurally Conserved Residues And Of Hot Spots Across Interfaces. Implications For Docking. *Structure* 2004, 12(6):1027-1038.

111. Gao Y, Wang R, Lai L: Structure-Based Method For Analyzing Protein-Protein Interfaces. *J Mol Model* 2004, 10(1):44-54.

112. Guerois R, Nielsen Je, Serrano L: Predicting Changes In The Stability Of Proteins And Protein Complexes: A Study Of More Than 1000 Mutations. *J Mol Biol* 2002, 320(2):369-387.

113. Huo S, Massova I, Kollman Pa: Computational Alanine Scanning Of The 1:1 Human Growth Hormone-Receptor Complex. *J Comput Chem* 2002, 23(1):15-27.

114. Rajamani D, Thiel S, Vajda S, Camacho Cj: Anchor Residues In Protein-Protein Interactions. *Proc Natl Acad Sci U S A* 2004, 101(31):11287-11292.

115. Moreira Is, Fernandes Pa, Ramos Mj: Hot Spot Occlusion From Bulk Water: A Comprehensive Study Of The Complex Between The Lysozyme Hel And The Antibody Fvd1.3. *J Phys Chem B* 2007, 111(10):2697-2706.

116. Guney E, Tuncbag N, Keskin O, Gursoy A: Hotsprint: Database Of Computational Hot Spots In Protein Interfaces. *Nucleic Acids Res* 2008, 36(Database Issue):D662-666.

117. Ofran Y, Rost B: Protein-Protein Interaction Hotspots Carved Into Sequences. *Plos Comput Biol* 2007, 3(7):E119.

118. Acuner Ozbabacan Se, Gursoy A, Keskin O, Nussinov R: Conformational Ensembles, Signal Transduction And Residue Hot Spots: Application To Drug Discovery. *Curr Opin Drug Discov Devel* 2010, 13(5):527-537.

119. Brenke R, Kozakov D, Chuang Gy, Beglov D, Hall D, Landon Mr, Mattos C, Vajda S: Fragment-Based Identification Of Druggable 'Hot Spots' Of Proteins Using Fourier Domain Correlation Techniques. *Bioinformatics* 2009, 25(5):621-627.

120. Ma B, Nussinov R: Trp/Met/Phe Hot Spots In Protein-Protein Interactions: Potential Targets In Drug Design. *Curr Top Med Chem* 2007, 7(10):999-1005.

121. Wells Ja, Mcclendon Cl: Reaching For High-Hanging Fruit In Drug Discovery At Protein-Protein Interfaces. *Nature* 2007, 450(7172):1001-1009.

122. Uetz P, Giot L, Cagney G, Mansfield Ta, Judson Rs, Knight Jr, Lockshon D, Narayan V, Srinivasan M, Pochart P *Et Al*: A Comprehensive Analysis Of Protein-Protein Interactions In Saccharomyces Cerevisiae. *Nature* 2000, 403(6770):623-627.

123. Landgraf C, Panni S, Montecchi-Palazzi L, Castagnoli L, Schneider-Mergener J, Volkmer-Engert R, Cesareni G: Protein Interaction Networks By Proteome Peptide Scanning. *Plos Biol* 2004, 2(1):E14.

124. Macbeath G, Schreiber Sl: Printing Proteins As Microarrays For High-Throughput Function Determination. *Science* 2000, 289(5485):1760-1763.

125. Bauer A, Kuster B: Affinity Purification-Mass Spectrometry. Powerful Tools For The Characterization Of Protein Complexes. *Eur J Biochem* 2003, 270(4):570-578.
126. Xenarios I, Salwinski L, Duan Xj, Higney P, Kim Sm, Eisenberg D: Dip, The Database Of Interacting Proteins: A Research Tool For Studying Cellular Networks Of Protein Interactions. *Nucleic Acids Res* 2002, 30(1):303-305.
127. Bader Gd, Betel D, Hogue Cw: Bind: The Biomolecular Interaction Network Database. *Nucleic Acids Res* 2003, 31(1):248-250.
128. Stark C, Breitkreutz Bj, Reguly T, Boucher L, Breitkreutz A, Tyers M: Biogrid: A General Repository For Interaction Datasets. *Nucleic Acids Res* 2006, 34(Database Issue):D535-539.
129. Hermjakob H, Montecchi-Palazzi L, Lewington C, Mudali S, Kerrien S, Orchard S, Vingron M, Roechert B, Roepstorff P, Valencia A *Et Al*: Intact: An Open Source Molecular Interaction Database. *Nucleic Acids Res* 2004, 32(Database Issue):D452-455.
130. Peri S, Navarro Jd, Kristiansen Tz, Amanchy R, Surendranath V, Muthusamy B, Gandhi Tk, Chandrika Kn, Deshpande N, Suresh S *Et Al*: Human Protein Reference Database As A Discovery Resource For Proteomics. *Nucleic Acids Res* 2004, 32(Database Issue):D497-501.
131. Schoof H, Spannagl M, Yang L, Ernst R, Gundlach H, Haase D, Haberer G, Mayer Kf: Munich Information Center For Protein Sequences Plant Genome Resources: A Framework For Integrative And Comparative Analyses 1(W). *Plant Physiol* 2005, 138(3):1301-1309.
132. Han K, Park B, Kim H, Hong J, Park J: Hpid: The Human Protein Interaction Database. *Bioinformatics* 2004, 20(15):2466-2470.
133. Gong S, Yoon G, Jang I, Bolser D, Dafas P, Schroeder M, Choi H, Cho Y, Han K, Lee S *Et Al*: Psibase: A Database Of Protein Structural Interactome Map (Psimap). *Bioinformatics* 2005, 21(10):2541-2543.
134. Raghavachari B, Tasneem A, Przytycka Tm, Jothi R: Domine: A Database Of Protein Domain Interactions. *Nucleic Acids Res* 2008, 36(Database Issue):D656-661.
135. Zuiderweg Er: Mapping Protein-Protein Interactions In Solution By Nmr Spectroscopy. *Biochemistry* 2002, 41(1):1-7.
136. Henderson R: Realizing The Potential Of Electron Cryo-Microscopy. *Q Rev Biophys* 2004, 37(1):3-13.

137. Aloy P, Russell Rb: Interrogating Protein Interaction Networks Through Structural Biology. *Proc Natl Acad Sci U S A* 2002, 99(9):5896-5901.

138. Kim Pm, Lu Lj, Xia Y, Gerstein Mb: Relating Three-Dimensional Structures To Protein Networks Provides Evolutionary Insights. *Science* 2006, 314(5807):1938-1941.

139. Dawelbait G, Winter C, Zhang Y, Pilarsky C, Grutzmann R, Heinrich Jc, Schroeder M: Structural Templates Predict Novel Protein Interactions And Targets From Pancreas Tumour Gene Expression Data. *Bioinformatics* 2007, 23(13):I115-124.

140. Huang Yj, Hang D, Lu Lj, Tong L, Gerstein Mb, Montelione Gt: Targeting The Human Cancer Pathway Protein Interaction Network By Structural Genomics. *Mol Cell Proteomics* 2008, 7(10):2048-2060.

141. Kar G, Gursoy A, Keskin O: Human Cancer Protein-Protein Interaction Network: A Structural Perspective. *Plos Comput Biol* 2009, 5(12):E1000601.

142. Tuncbag N, Gursoy A, Nussinov R, Keskin O: Predicting Protein-Protein Interactions On A Proteome Scale By Matching Evolutionary And Structural Similarities At Interfaces Using Prism. *Nat Protoc* 2011, 6(9):1341-1354.

143. Gursoy A, Keskin O, Nussinov R: Topological Properties Of Protein Interaction Networks From A Structural Perspective. *Biochem Soc Trans* 2008, 36(Pt 6):1398-1403.

144. Andrusier N, Mashiach E, Nussinov R, Wolfson Hj: Principles Of Flexible Protein-Protein Docking. *Proteins* 2008, 73(2):271-289.

145. Gray Jj: High-Resolution Protein-Protein Docking. *Curr Opin Struct Biol* 2006, 16(2):183-193.

146. Halperin I, Ma B, Wolfson H, Nussinov R: Principles Of Docking: An Overview Of Search Algorithms And A Guide To Scoring Functions. *Proteins* 2002, 47(4):409-443.

147. De Vries Sj, Van Dijk M, Bonvin Am: The Haddock Web Server For Data-Driven Biomolecular Docking. *Nat Protoc* 2010, 5(5):883-897.

148. Lesk Vi, Sternberg Mj: 3d-Garden: A System For Modelling Protein-Protein Complexes Based On Conformational Refinement Of Ensembles Generated With The Marching Cubes Algorithm. *Bioinformatics* 2008, 24(9):1137-1144.

149. Cheng Tm, Blundell Tl, Fernandez-Recio J: Pydock: Electrostatics And Desolvation For Effective Scoring Of Rigid-Body Protein-Protein Docking. *Proteins* 2007, 68(2):503-515.

150. Moont G, Gabb Ha, Sternberg Mj: Use Of Pair Potentials Across Protein Interfaces In Screening Predicted Docked Complexes. *Proteins* 1999, 35(3):364-373.

151. Andrusier N, Nussinov R, Wolfson Hj: Firedock: Fast Interaction Refinement In Molecular Docking. *Proteins* 2007, 69(1):139-159.

152. Li L, Chen R, Weng Z: Rdock: Refinement Of Rigid-Body Protein Docking Predictions. *Proteins* 2003, 53(3):693-707.

153. Mashiach E, Nussinov R, Wolfson Hj: Fiberdock: Flexible Induced-Fit Backbone Refinement In Molecular Docking. *Proteins* 2010, 78(6):1503-1519.

154. Mosca R, Pons C, Fernandez-Recio J, Aloy P: Pushing Structural Information Into The Yeast Interactome By High-Throughput Protein Docking Experiments. *Plos Comput Biol* 2009, 5(8):E1000490.

155. Aloy P, Russell Rb: Interprets: Protein Interaction Prediction Through Tertiary Structure. *Bioinformatics* 2003, 19(1):161-162.

156. Lu L, Lu H, Skolnick J: Multiprospector: An Algorithm For The Prediction Of Protein-Protein Interactions By Multimeric Threading. *Proteins* 2002, 49(3):350-364.

157. Lu H, Lu L, Skolnick J: Development Of Unified Statistical Potentials Describing Protein-Protein Interactions. *Biophys J* 2003, 84(3):1895-1901.

158. Davis Fp, Braberg H, Shen My, Pieper U, Sali A, Madhusudhan Ms: Protein Complex Compositions Predicted By Structural Similarity. *Nucleic Acids Res* 2006, 34(10):2943-2952.

159. Chothia C: Proteins. One Thousand Families For The Molecular Biologist. *Nature* 1992, 357(6379):543-544.

160. Aloy P, Bottcher B, Ceulemans H, Leutwein C, Mellwig C, Fischer S, Gavin Ac, Bork P, Superti-Furga G, Serrano L *Et Al*: Structure-Based Assembly Of Protein Complexes In Yeast. *Science* 2004, 303(5666):2026-2029.

161. Ogmen U, Keskin O, Aytuna As, Nussinov R, Gursoy A: Prism: Protein Interactions By Structural Matching. *Nucleic Acids Res* 2005, 33(Web Server Issue):W331-336.

162. Gunther S, May P, Hoppe A, Frommel C, Preissner R: Docking Without Docking: Isearch--Prediction Of Interactions Using Known Interfaces. *Proteins* 2007, 69(4):839-844.

163. Jung Sh, Hyun B, Jang Wh, Hur Hy, Han Ds: Protein Complex Prediction Based On Simultaneous Protein Interaction Network. *Bioinformatics* 2010, 26(3):385-391.

164. Wouters Bj, Lowenberg B, Delwel R: A Decade Of Genome-Wide Gene Expression Profiling In Acute Myeloid Leukemia: Flashback And Prospects. *Blood* 2009, 113(2):291-298.

165. Jansen R, Greenbaum D, Gerstein M: Relating Whole-Genome Expression Data With Protein-Protein Interactions. *Genome Res* 2002, 12(1):37-46.

166. Klebanov L, Yakovlev A: How High Is The Level Of Technical Noise In Microarray Data? *Biol Direct* 2007, 2:9.

167. Chuang Hy, Lee E, Liu Yt, Lee D, Ideker T: Network-Based Classification Of Breast Cancer Metastasis. *Mol Syst Biol* 2007, 3:140.

168. Rapaport F, Zinovyev A, Dutreix M, Barillot E, Vert Jp: Classification Of Microarray Data Using Gene Networks. *Bmc Bioinformatics* 2007, 8:35.

169. Wei P, Pan W: Incorporating Gene Networks Into Statistical Tests For Genomic Data Via A Spatially Correlated Mixture Model. *Bioinformatics* 2008, 24(3):404-411.

170. Ma X, Lee H, Wang L, Sun F: Cgi: A New Approach For Prioritizing Genes By Combining Gene Expression And Protein-Protein Interaction Data. *Bioinformatics* 2007, 23(2):215-221.

171. Li D, Liu W, Liu Z, Wang J, Liu Q, Zhu Y, He F: Princess, A Protein Interaction Confidence Evaluation System With Multiple Data Sources. *Mol Cell Proteomics* 2008, 7(6):1043-1052.

172. Ideker T, Ozier O, Schwikowski B, Siegel Af: Discovering Regulatory And Signalling Circuits In Molecular Interaction Networks. *Bioinformatics* 2002, 18 Suppl 1:S233-240.

173. Nacu S, Critchley-Thorne R, Lee P, Holmes S: Gene Expression Network Analysis And Applications To Immunology. *Bioinformatics* 2007, 23(7):850-858.

174. Rajagopalan D, Agarwal P: Inferring Pathways From Gene Lists Using A Literature-Derived Network Of Biological Relationships. *Bioinformatics* 2005, 21(6):788-793.

175. Guo Z, Wang L, Li Y, Gong X, Yao C, Ma W, Wang D, Zhu J, Zhang M, Yang D *Et Al*: Edge-Based Scoring And Searching Method For Identifying Condition-Responsive Protein-Protein Interaction Sub-Network. *Bioinformatics* 2007, 23(16):2121-2128.

176. Cabusora L, Sutton E, Fulmer A, Forst Cv: Differential Network Expression During Drug And Stress Response. *Bioinformatics* 2005, 21(12):2898-2905.

177. Qiu Yq, Zhang S, Zhang Xs, Chen L: Detecting Disease Associated Modules And Prioritizing Active Genes Based On High Throughput Data. *Bmc Bioinformatics* 2010, 11:26.

178. Fry Dc: Protein-Protein Interactions As Targets For Small Molecule Drug Discovery. *Biopolymers* 2006, 84(6):535-552.

179. Keskin O, Gursoy A, Ma B, Nussinov R: Towards Drugs Targeting Multiple Proteins In A Systems Biology Approach. *Curr Top Med Chem* 2007, 7(10):943-951.

180. Sugaya N, Ikeda K, Tashiro T, Takeda S, Otomo J, Ishida Y, Shiratori A, Toyoda A, Noguchi H, Takeda T *Et Al*: An Integrative In Silico Approach For Discovering Candidates For Drug-Targetable Protein-Protein Interactions In Interactome Data. *Bmc Pharmacol* 2007, 7:10.

181. Xie L, Bourne Pe: Structure-Based Systems Biology For Analyzing Off-Target Binding. *Curr Opin Struct Biol* 2011, 21(2):189-199.

182. Chong Cr, Sullivan Dj, Jr.: New Uses For Old Drugs. *Nature* 2007, 448(7154):645-646.

183. Dimasi Ja, Hansen Rw, Grabowski Hg: The Price Of Innovation: New Estimates Of Drug Development Costs. *J Health Econ* 2003, 22(2):151-185.

184. Weber A, Casini A, Heine A, Kuhn D, Supuran Ct, Scozzafava A, Klebe G: Unexpected Nanomolar Inhibition Of Carbonic Anhydrase By Cox-2-Selective Celecoxib: New Pharmacological Opportunities Due To Related Binding Site Recognition. *J Med Chem* 2004, 47(3):550-557.

185. Swaminathan G, Tsygankov Ay: The Cbl Family Proteins: Ring Leaders In Regulation Of Cell Signaling. *J Cell Physiol* 2006, 209(1):21-43.

186. Blake Tj, Shapiro M, Morse Hc, 3rd, Langdon Wy: The Sequences Of The Human And Mouse C-Cbl Proto-Oncogenes Show V-Cbl Was Generated By A Large Truncation Encompassing A Proline-Rich Domain And A Leucine Zipper-Like Motif. *Oncogene* 1991, 6(4):653-657.

187. Zheng N, Wang P, Jeffrey Pd, Pavletich Np: Structure Of A C-Cbl-Ubch7 Complex: Ring Domain Function In Ubiquitin-Protein Ligases. *Cell* 2000, 102(4):533-539.

188. Umebayashi K, Stenmark H, Yoshimori T: Ubc4/5 And C-Cbl Continue To Ubiquitinate Egf Receptor After Internalization To Facilitate Polyubiquitination And Degradation. *Mol Biol Cell* 2008, 19(8):3454-3462.

189. Kar G, Keskin O, Nussinov R, Gursoy A: Human Proteome-Scale Structural Modeling Of E2-E3 Interactions Exploiting Interface Motifs. *J Proteome Res* 2012, 11(2):1196-1207.

190. Makishima H, Cazzolli H, Szpurka H, Dunbar A, Tiu R, Huh J, Muramatsu H, O'keefe C, Hsi E, Paquette Rl *Et Al*: Mutations Of E3 Ubiquitin Ligase Cbl Family Members Constitute A Novel Common Pathogenic Lesion In Myeloid Malignancies. *J Clin Oncol* 2009, 27(36):6109-6116.

191. Joazeiro Ca, Wing Ss, Huang H, Leverson Jd, Hunter T, Liu Yc: The Tyrosine Kinase Negative Regulator C-Cbl As A Ring-Type, E2-Dependent Ubiquitin-Protein Ligase. *Science* 1999, 286(5438):309-312.

192. Tuncbag N, Keskin O, Gursoy A: Hotpoint: Hot Spot Prediction Server For Protein Interfaces. *Nucleic Acids Res* 2010, 38(Web Server Issue):W402-406.

193. Ma B, Tsai Cj, Pan Y, Nussinov R: Why Does Binding Of Proteins To Dna Or Proteins To Proteins Not Necessarily Spell Function? *Acs Chem Biol* 2010, 5(3):265-272.

194. Datta D, Scheer Jm, Romanowski Mj, Wells Ja: An Allosteric Circuit In Caspase-1. *J Mol Biol* 2008, 381(5):1157-1167.

195. Del Sol A, Tsai Cj, Ma B, Nussinov R: The Origin Of Allosteric Functional Modulation: Multiple Pre-Existing Pathways. *Structure* 2009, 17(8):1042-1050.

196. Lavrik In: Systems Biology Of Apoptosis Signaling Networks. *Curr Opin Biotechnol* 2010, Doi:10.1016/J.Copbio.2010.07.001.

197. Grant Bj, Gorfe Aa, Mccammon Ja: Large Conformational Changes In Proteins: Signaling And Other Functions. *Curr Opin Struct Biol* 2010, 20(2):142-147.

198. Grant Bj, Gorfe Aa, Mccammon Ja: Ras Conformational Switching: Simulating Nucleotide-Dependent Conformational Transitions With Accelerated Molecular Dynamics. *Plos Comput Biol* 2009, 5(3):E1000325.

199. Filchtinski D, Sharabi O, Ruppel A, Vetter Ir, Herrmann C, Shifman Jm: What Makes Ras An Efficient Molecular Switch: A Computational, Biophysical, And Structural Study Of Ras-Gdp Interactions With Mutants Of Raf. *J Mol Biol* 2010, 399(3):422-435.

200. Hiwasa T, Kasama M, Nakadai T, Sawada T, Sakiyama S: Loss Of Raf-1-Binding Activity Of V-Ha-Ras By The Deletion Of Amino Acid Residues 64-72 And 143-151. *Cell Signal* 1996, 8(5):393-396.

201. Zhang Xf, Settleman J, Kyriakis Jm, Takeuchi-Suzuki E, Elledge Sj, Marshall Ms, Bruder Jt, Rapp Ur, Avruch J: Normal And Oncogenic P21ras Proteins Bind To The Amino-Terminal Regulatory Domain Of C-Raf-1. *Nature* 1993, 364(6435):308-313.

202. Hershko A, Ciechanover A: The Ubiquitin System. *Annu Rev Biochem* 1998, 67:425-479.

203. Jin J, Li X, Gygi Sp, Harper Jw: Dual E1 Activation Systems For Ubiquitin Differentially Regulate E2 Enzyme Charging. *Nature* 2007, 447(7148):1135-1138.

204. Pickart Cm: Mechanisms Underlying Ubiquitination. *Annu Rev Biochem* 2001, 70:503-533.

205. Rotin D, Kumar S: Physiological Functions Of The Hect Family Of Ubiquitin Ligases. *Nat Rev Mol Cell Biol* 2009, 10(6):398-409.

206. Borden Kl: Ring Domains: Master Builders Of Molecular Scaffolds? *J Mol Biol* 2000, 295(5):1103-1112.

207. Ohi Md, Vander Kooi Cw, Rosenberg Ja, Chazin Wj, Gould Kl: Structural Insights Into The U-Box, A Domain Associated With Multi-Ubiquitination. *Nat Struct Biol* 2003, 10(4):250-255.

208. Hofmann Rm, Pickart Cm: In Vitro Assembly And Recognition Of Lys-63 Polyubiquitin Chains. *J Biol Chem* 2001, 276(30):27936-27943.

209. Burroughs Am, Jaffee M, Iyer Lm, Aravind L: Anatomy Of The E2 Ligase Fold: Implications For Enzymology And Evolution Of Ubiquitin/Ub-Like Protein Conjugation. *J Struct Biol* 2008, 162(2):205-218.

210. Michelle C, Vourc'h P, Mignon L, Andres Cr: What Was The Set Of Ubiquitin And Ubiquitin-Like Conjugating Enzymes In The Eukaryote Common Ancestor? *J Mol Evol* 2009, 68(6):616-628.

211. Winn Pj, Religa Tl, Battey Jn, Banerjee A, Wade Rc: Determinants Of Functionality In The Ubiquitin Conjugating Enzyme Family. *Structure* 2004, 12(9):1563-1574.

212. Semple Ca: The Comparative Proteomics Of Ubiquitination In Mouse. *Genome Res* 2003, 13(6b):1389-1394.

213. Von Arnim Ag: A Hitchhiker's Guide To The Proteasome. *Sci Stke* 2001, 2001(97):Pe2.

214. Wong Br, Parlati F, Qu K, Demo S, Pray T, Huang J, Payan Dg, Bennett Mk: Drug Discovery In The Ubiquitin Regulatory Pathway. *Drug Discov Today* 2003, 8(16):746-754.

215. Dominguez C, Bonvin Am, Winkler Gs, Van Schaik Fm, Timmers Ht, Boelens R: Structural Model Of The Ubch5b/Cnot4 Complex Revealed By Combining Nmr, Mutagenesis, And Docking Approaches. *Structure* 2004, 12(4):633-644.

216. Weissman Am: Themes And Variations On Ubiquitylation. *Nat Rev Mol Cell Biol* 2001, 2(3):169-178.

217. Winkler Gs, Albert Tk, Dominguez C, Legtenberg Yi, Boelens R, Timmers Ht: An Altered-Specificity Ubiquitin-Conjugating Enzyme/Ubiquitin-Protein Ligase Pair. *J Mol Biol* 2004, 337(1):157-165.

218. Christensen De, Brzovic Ps, Klevit Re: E2-Brca1 Ring Interactions Dictate Synthesis Of Mono- Or Specific Polyubiquitin Chain Linkages. *Nat Struct Mol Biol* 2007, 14(10):941-948.

219. Christensen De, Klevit Re: Dynamic Interactions Of Proteins In Complex Networks: Identifying The Complete Set Of Interacting E2s For Functional Investigation Of E3-Dependent Protein Ubiquitination. *Febs J* 2009, 276(19):5381-5389.

220. Gardner Rg, Shearer Ag, Hampton Ry: In Vivo Action Of The Hrd Ubiquitin Ligase Complex: Mechanisms Of Endoplasmic Reticulum Quality Control And Sterol Regulation. *Mol Cell Biol* 2001, 21(13):4276-4291.

221. Van Wijk Sj, De Vries Sj, Kemmeren P, Huang A, Boelens R, Bonvin Am, Timmers Ht: A Comprehensive Framework Of E2-Ring E3 Interactions Of The Human Ubiquitin-Proteasome System. *Mol Syst Biol* 2009, 5:295.

222. Markson G, Kiel C, Hyde R, Brown S, Charalabous P, Bremm A, Semple J, Woodsmith J, Duley S, Salehi-Ashtiani K *Et Al*: Analysis Of The Human E2 Ubiquitin Conjugating Enzyme Protein Interaction Network. *Genome Res* 2009, 19(10):1905-1911.

223. Berman Hm, Westbrook J, Feng Z, Gilliland G, Bhat Tn, Weissig H, Shindyalov In, Bourne Pe: The Protein Data Bank. *Nucleic Acids Res* 2000, 28(1):235-242.

224. Kanehisa M, Goto S: Kegg: Kyoto Encyclopedia Of Genes And Genomes. *Nucleic Acids Res* 2000, 28(1):27-30.

225. Hubbard S, Thornton J: 'Naccess', Computer Program. In.: Department Of Biochemistry And Molecular Biology, University College London.; 1993.

226. Shatsky M, Nussinov R, Wolfson Hj: A Method For Simultaneous Alignment Of Multiple Protein Structures. *Proteins* 2004, 56(1):143-156.

227. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L: The Foldx Web Server: An Online Force Field. *Nucleic Acids Res* 2005, 33(Web Server Issue):W382-388.

228. Hwang H, Pierce B, Mintseris J, Janin J, Weng Z: Protein-Protein Docking Benchmark Version 3.0. *Proteins* 2008, 73(3):705-709.

229. Tuncbag N, Keskin O, Nussinov R, Gursoy A: Fast And Accurate Modeling Of Protein-Protein Interactions By Combining Template-Interface-Based Docking With Flexible Refinement. *Proteins* 2012, 80(4):1239-1249.

230. Seufert W, Jentsch S: Ubiquitin-Conjugating Enzymes Ubc4 And Ubc5 Mediate Selective Degradation Of Short-Lived And Abnormal Proteins. *Embo J* 1990, 9(2):543-550.

231. Deng L, Wang C, Spencer E, Yang L, Braun A, You J, Slaughter C, Pickart C, Chen Zj: Activation Of The Ikappab Kinase Complex By Traf6 Requires A Dimeric Ubiquitin-Conjugating Enzyme Complex And A Unique Polyubiquitin Chain. *Cell* 2000, 103(2):351-361.

232. Passmore La, Barford D: Getting Into Position: The Catalytic Mechanisms Of Protein Ubiquitylation. *Biochem J* 2004, 379(Pt 3):513-525.

233. Sadowski M, Sarcevic B: Mechanisms Of Mono- And Poly-Ubiquitination: Ubiquitination Specificity Depends On Compatibility Between The E2 Catalytic Core And Amino Acid Residues Proximal To The Lysine. *Cell Div* 2010, 5:19.

234. Deshaies Rj, Joazeiro Ca: Ring Domain E3 Ubiquitin Ligases. *Annu Rev Biochem* 2009, 78:399-434.

235. Wang M, Pickart Cm: Different Hect Domain Ubiquitin Ligases Employ Distinct Mechanisms Of Polyubiquitin Chain Synthesis. *Embo J* 2005, 24(24):4324-4333.

236. Mishra A, Godavarthi Sk, Maheshwari M, Goswami A, Jana Nr: The Ubiquitin Ligase E6-Ap Is Induced And Recruited To Aggresomes In Response To Proteasome Inhibition And May Be Involved In The Ubiquitination Of Hsp70-Bound Misfolded Proteins. *J Biol Chem* 2009, 284(16):10537-10545.

237. Eletr Zm, Kuhlman B: Sequence Determinants Of E2-E6ap Binding Affinity And Specificity. *J Mol Biol* 2007, 369(2):419-428.

238. Louria-Hayon I, Alsheich-Bartok O, Levav-Cohen Y, Silberman I, Berger M, Grossman T, Matentzoglu K, Jiang Yh, Muller S, Scheffner M *Et Al*: E6ap Promotes The Degradation Of The Pml Tumor Suppressor. *Cell Death Differ* 2009, 16(8):1156-1166.

239. Muller S, Matunis Mj, Dejean A: Conjugation With The Ubiquitin-Related Modifier Sumo-1 Regulates The Partitioning Of Pml Within The Nucleus. *Embo J* 1998, 17(1):61-70.

240. Izzi L, Attisano L: Regulation Of The Tgfbeta Signalling Pathway By Ubiquitin-Mediated Degradation. *Oncogene* 2004, 23(11):2071-2078.

241. Ebisawa T, Fukuchi M, Murakami G, Chiba T, Tanaka K, Imamura T, Miyazono K: Smurf1 Interacts With Transforming Growth Factor-Beta Type I Receptor Through Smad7 And Induces Receptor Degradation. *J Biol Chem* 2001, 276(16):12477-12480.

242. Lee Ps, Chang C, Liu D, Derynck R: Sumoylation Of Smad4, The Common Smad Mediator Of Transforming Growth Factor-Beta Family Signaling. *J Biol Chem* 2003, 278(30):27853-27863.

243. Zhao X, Heng Ji, Guardavaccaro D, Jiang R, Pagano M, Guillemot F, Iavarone A, Lasorella A: The Hect-Domain Ubiquitin Ligase Huwe1 Controls Neural Differentiation And Proliferation By Destabilizing The N-Myc Oncoprotein. *Nat Cell Biol* 2008, 10(6):643-653.

244. Hall Jr, Kow E, Nevis Kr, Lu Ck, Luce Ks, Zhong Q, Cook Jg: Cdc6 Stability Is Regulated By The Huwe1 Ubiquitin Ligase After Dna Damage. *Mol Biol Cell* 2007, 18(9):3340-3350.

245. Adhikary S, Marinoni F, Hock A, Hulleman E, Popov N, Beier R, Bernard S, Quarto M, Capra M, Goettig S *Et Al*: The Ubiquitin Ligase Hecth9 Regulates Transcriptional Activation By Myc And Is Essential For Tumor Cell Proliferation. *Cell* 2005, 123(3):409-421.

246. Kirkpatrick Ds, Hathaway Na, Hanna J, Elsasser S, Rush J, Finley D, King Rw, Gygi Sp: Quantitative Analysis Of In Vitro Ubiquitinated Cyclin B1 Reveals Complex Chain Topology. *Nat Cell Biol* 2006, 8(7):700-710.

247. Jiang X, Sorkin A: Epidermal Growth Factor Receptor Internalization Through Clathrin-Coated Pits Requires Cbl Ring Finger And Proline-Rich Domains But Not Receptor Polyubiquitylation. *Traffic* 2003, 4(8):529-543.

248. Masson K, Heiss E, Band H, Ronnstrand L: Direct Binding Of Cbl To Tyr568 And Tyr936 Of The Stem Cell Factor Receptor/C-Kit Is Required For Ligand-Induced Ubiquitination, Internalization And Degradation. *Biochem J* 2006, 399(1):59-67.

249. Pichler A, Knipscheer P, Oberhofer E, Van Dijk Wj, Korner R, Olsen Jv, Jentsch S, Melchior F, Sixma Tk: Sumo Modification Of The Ubiquitin-Conjugating Enzyme E2-25k. *Nat Struct Mol Biol* 2005, 12(3):264-269.

250. Sadowski M, Mawson A, Baker R, Sarcevic B: Cdc34 C-Terminal Tail Phosphorylation Regulates Skp1/Cullin/F-Box (Scf)-Mediated Ubiquitination And Cell Cycle Progression. *Biochem J* 2007, 405(3):569-581.

251. Tuncbag N, Gursoy A, Keskin O: Identification Of Computational Hot Spots In Protein Interfaces: Combining Solvent Accessibility And Inter-Residue Potentials Improves The Accuracy. *Bioinformatics* 2009, 25(12):1513-1520.

252. Ye Y, Rape M: Building Ubiquitin Chains: E2 Enzymes At Work. *Nat Rev Mol Cell Biol* 2009, 10(11):755-764.

253. Xu Z, Kohli E, Devlin Ki, Bold M, Nix Jc, Misra S: Interactions Between The Quality Control Ubiquitin Ligase Chip And Ubiquitin Conjugating Enzymes. *Bmc Struct Biol* 2008, 8:26.

254. Konc J, Janezic D: Probis Algorithm For Detection Of Structurally Similar Protein Binding Sites By Local Structural Alignment. *Bioinformatics* 2010, 26(9):1160-1168.

255. Ogunjimi Aa, Briant Dj, Pece-Barbara N, Le Roy C, Di Guglielmo Gm, Kavsak P, Rasmussen Rk, Seet Bt, Sicheri F, Wrana Jl: Regulation Of Smurf2 Ubiquitin Ligase Activity By Anchoring The E2 To The Hect Domain. *Mol Cell* 2005, 19(3):297-308.

256. Verdecia Ma, Joazeiro Ca, Wells Nj, Ferrer Jl, Bowman Me, Hunter T, Noel Jp: Conformational Flexibility Underlies Ubiquitin Ligation Mediated By The Wwp1 Hect Domain E3 Ligase. *Mol Cell* 2003, 11(1):249-259.

257. Yin Q, Lin Sc, Lamothe B, Lu M, Lo Yc, Hura G, Zheng L, Rich Rl, Campos Ad, Myszka Dg *Et Al*: E2 Interaction And Dimerization In The Crystal Structure Of Traf6. *Nat Struct Mol Biol* 2009, 16(6):658-666.

258. Nuber U, Schwarz S, Kaiser P, Schneider R, Scheffner M: Cloning Of Human Ubiquitin-Conjugating Enzymes Ubch6 And Ubch7 (E2-F1) And Characterization Of Their Interaction With E6-Ap And Rsp5. *J Biol Chem* 1996, 271(5):2795-2800.

259. Wiesner S, Ogunjimi Aa, Wang Hr, Rotin D, Sicheri F, Wrana Jl, Forman-Kay Jd: Autoinhibition Of The Hect-Type Ubiquitin Ligase Smurf2 Through Its C2 Domain. *Cell* 2007, 130(4):651-662.

260. Geetha T, Jiang J, Wooten Mw: Lysine 63 Polyubiquitination Of The Nerve Growth Factor Receptor Trka Directs Internalization And Signaling. *Mol Cell* 2005, 20(2):301-312.

261. Shulman-Peleg A, Shatsky M, Nussinov R, Wolfson Hj: Multibind And Mappis: Webservers For Multiple Alignment Of Protein 3d-Binding Sites And Their Interactions. *Nucleic Acids Res* 2008, 36(Web Server Issue):W260-264.

262. Huang L, Kinnucan E, Wang G, Beaudenon S, Howley Pm, Huibregtse Jm, Pavletich Np: Structure Of An E6ap-Ubch7 Complex: Insights Into Ubiquitination By The E2-E3 Enzyme Cascade. *Science* 1999, 286(5443):1321-1326.

263. Tokuriki N, Stricher F, Serrano L, Tawfik Ds: How Protein Stability And New Functions Trade Off. *Plos Comput Biol* 2008, 4(2):E1000002.

264. Dynek Jn, Goncharov T, Dueber Ec, Fedorova Av, Izrael-Tomasevic A, Phu L, Helgason E, Fairbrother Wj, Deshayes K, Kirkpatrick Ds *Et Al*: C-Iap1 And Ubch5 Promote K11-Linked Polyubiquitination Of Rip1 In Tnf Signalling. *Embo J* 2010, 29(24):4198-4209.

265. Harris Ma, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C *Et Al*: The Gene Ontology (Go) Database And Informatics Resource. *Nucleic Acids Res* 2004, 32(Database Issue):D258-261.

266. Yu Zk, Geyer Rk, Maki Cg: Mdm2-Dependent Ubiquitination Of Nuclear And Cytoplasmic P53. *Oncogene* 2000, 19(51):5892-5897.

267. Pierce Kl, Lefkowitz Rj: Classical And New Roles Of Beta-Arrestins In The Regulation Of G-Protein-Coupled Receptors. *Nat Rev Neurosci* 2001, 2(10):727-733.

268. Shenoy Sk, Mcdonald Ph, Kohout Ta, Lefkowitz Rj: Regulation Of Receptor Fate By Ubiquitination Of Activated Beta 2-Adrenergic Receptor And Beta-Arrestin. *Science* 2001, 294(5545):1307-1313.

269. Yamamoto Y, Huibregtse Jm, Howley Pm: The Human E6-Ap Gene (Ube3a) Encodes Three Potential Protein Isoforms Generated By Differential Splicing. *Genomics* 1997, 41(2):263-266.

270. Anan T, Nagata Y, Koga H, Honda Y, Yabuki N, Miyamoto C, Kuwano A, Matsuda I, Endo F, Saya H *Et Al*: Human Ubiquitin-Protein Ligase Nedd4: Expression, Subcellular Localization And Selective Interaction With Ubiquitin-Conjugating Enzymes. *Genes Cells* 1998, 3(11):751-763.

271. Hatakeyama S, Jensen Jp, Weissman Am: Subcellular Localization And Ubiquitin-Conjugating Enzyme (E2) Interactions Of Mammalian Hect Family Ubiquitin Protein Ligases. *J Biol Chem* 1997, 272(24):15085-15092.

272. Brzovic Ps, Keeffe Jr, Nishikawa H, Miyamoto K, Fox D, 3rd, Fukuda M, Ohta T, Klevit R: Binding And Recognition In The Assembly Of An Active Brca1/Bard1 Ubiquitin-Ligase Complex. *Proc Natl Acad Sci U S A* 2003, 100(10):5646-5651.

273. Nalepa G, Rolfe M, Harper Jw: Drug Discovery In The Ubiquitin-Proteasome System. *Nat Rev Drug Discov* 2006, 5(7):596-613.

274. Bernassola F, Karin M, Ciechanover A, Melino G: The Hect Family Of E3 Ubiquitin Ligases: Multiple Players In Cancer Development. *Cancer Cell* 2008, 14(1):10-21.

275. Reppert Sm, Weaver Dr: Coordination Of Circadian Timing In Mammals. *Nature* 2002, 418(6901):935-941.

276. Hirayama J, Sassone-Corsi P: Structural And Functional Features Of Transcription Factors Controlling The Circadian Clock. *Curr Opin Genet Dev* 2005, 15(5):548-556.

277. Hughes M, Deharo L, Pulivarthy Sr, Gu J, Hayes K, Panda S, Hogenesch Jb: High-Resolution Time Course Analysis Of Gene Expression From Pituitary. *Cold Spring Harb Symp Quant Biol* 2007, 72:381-386.

278. Ueda Hr: Systems Biology Of Mammalian Circadian Clocks. *Cold Spring Harb Symp Quant Biol* 2007, 72:365-380.

279. Bell-Pedersen D, Cassone Vm, Earnest Dj, Golden Ss, Hardin Pe, Thomas Tl, Zoran Mj: Circadian Rhythms From Multiple Oscillators: Lessons From Diverse Organisms. *Nat Rev Genet* 2005, 6(7):544-556.

280. Fu L, Lee Cc: The Circadian Clock: Pacemaker And Tumour Suppressor. *Nat Rev Cancer* 2003, 3(5):350-361.

281. Levi F, Schibler U: Circadian Rhythms: Mechanisms And Therapeutic Implications. *Annu Rev Pharmacol Toxicol* 2007, 47:593-628.

282. Stevens Rg: Circadian Disruption And Breast Cancer: From Melatonin To Clock Genes. *Epidemiology* 2005, 16(2):254-258.

283. Fu L, Pelicano H, Liu J, Huang P, Lee C: The Circadian Gene Period2 Plays An Important Role In Tumor Suppression And Dna Damage Response In Vivo. *Cell* 2002, 111(1):41-50.

284. Gauger Ma, Sancar A: Cryptochrome, Circadian Cycle, Cell Cycle Checkpoints, And Cancer. *Cancer Res* 2005, 65(15):6828-6834.

285. Antoch Mp, Gorbacheva Vy, Vykhovanets O, Toshkov Ia, Kondratov Rv, Kondratova Aa, Lee C, Nikitin Ay: Disruption Of The Circadian Clock Due To The Clock Mutation Has Discrete Effects On Aging And Carcinogenesis. *Cell Cycle* 2008, 7(9):1197-1204.

286. Matsuo T, Yamaguchi S, Mitsui S, Emi A, Shimoda F, Okamura H: Control Mechanism Of The Circadian Clock For Timing Of Cell Division In Vivo. *Science* 2003, 302(5643):255-259.

287. Akhtar Ra, Reddy Ab, Maywood Es, Clayton Jd, King Vm, Smith Ag, Gant Tw, Hastings Mh, Kyriacou Cp: Circadian Cycling Of The Mouse Liver Transcriptome, As Revealed By Cdna Microarray, Is Driven By The Suprachiasmatic Nucleus. *Curr Biol* 2002, 12(7):540-550.

288. Khapre Rv, Samsa We, Kondratov Rv: Circadian Regulation Of Cell Cycle: Molecular Connections Between Aging And The Circadian Clock. *Ann Med* 2010, 42(6):404-415.

289. Ozturk N, Lee Jh, Gaddameedhi S, Sancar A: Loss Of Cryptochrome Reduces Cancer Risk In P53 Mutant Mice. *Proc Natl Acad Sci U S A* 2009, 106(8):2841-2846.

290. Lee Jh, Sancar A: Circadian Clock Disruption Improves The Efficacy Of Chemotherapy Through P73-Mediated Apoptosis. *Proc Natl Acad Sci U S A* 2011, 108(26):10668-10672.

291. Lee Jh, Sancar A: Regulation Of Apoptosis By The Circadian Clock Through Nf-{Kappa}B Signaling. *Proc Natl Acad Sci U S A* 2011, 108(29):12036-12041.

292. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P *Et Al*: The String Database In 2011: Functional Interaction Networks Of Proteins, Globally Integrated And Scored. *Nucleic Acids Res* 2011, 39(Database Issue):D561-568.

293. Gaddameedhi S, Kemp Mg, Reardon Jt, Shields Jm, Smith-Roe Sl, Kaufmann Wk, Sancar A: Similar Nucleotide Excision Repair Capacity In Melanocytes And Melanoma Cells. *Cancer Res* 2010, 70(12):4922-4930.

294. Jho Eh, Zhang T, Domon C, Joo Ck, Freund Jn, Costantini F: Wnt/Beta-Catenin/Tcf Signaling Induces The Transcription Of Axin2, A Negative Regulator Of The Signaling Pathway. *Mol Cell Biol* 2002, 22(4):1172-1183.

295. Zinovyev A, Viara E, Calzone L, Barillot E: Binom: A Cytoscape Plugin For Manipulating And Analyzing Biological Networks. *Bioinformatics* 2008, 24(6):876-877.

296. Lamia Ka, Papp Sj, Yu Rt, Barish Gd, Uhlenhaut Nh, Jonker Jw, Downes M, Evans Rm: Cryptochromes Mediate Rhythmic Repression Of The Glucocorticoid Receptor. *Nature* 2011, 480(7378):552-556.

297. Zhang Ee, Liu Y, Dentin R, Pongsawakul Py, Liu Ac, Hirota T, Nusinow Da, Sun X, Landais S, Kodama Y *Et Al*: Cryptochrome Mediates Circadian Regulation Of Camp Signaling And Hepatic Gluconeogenesis. *Nat Med* 2010, 16(10):1152-1156.

298. Ashburner M, Ball Ca, Blake Ja, Botstein D, Butler H, Cherry Jm, Davis Ap, Dolinski K, Dwight Ss, Eppig Jt *Et Al*: Gene Ontology: Tool For The Unification Of Biology. The Gene Ontology Consortium. *Nat Genet* 2000, 25(1):25-29.

299. Arias Am, Brown Am, Brennan K: Wnt Signalling: Pathway Or Network? *Curr Opin Genet Dev* 1999, 9(4):447-454.

300. Cadigan Km, Nusse R: Wnt Signaling: A Common Theme In Animal Development. *Genes Dev* 1997, 11(24):3286-3305.

301. Peifer M, Polakis P: Wnt Signaling In Oncogenesis And Embryogenesis--A Look Outside The Nucleus. *Science* 2000, 287(5458):1606-1609.

302. Gerstein Av, Almeida Ta, Zhao G, Chess E, Shih Ie M, Buhler K, Pienta K, Rubin Ma, Vessella R, Papadopoulos N: Apc/Ctnnb1 (Beta-Catenin) Pathway Alterations In Human Prostate Cancers. *Genes Chromosomes Cancer* 2002, 34(1):9-16.

303. Pecina-Slaus N: Wnt Signal Transduction Pathway And Apoptosis: A Review. *Cancer Cell Int* 2010, 10:22.

304. Crowder Rj, Freeman Rs: Glycogen Synthase Kinase-3 Beta Activity Is Critical For Neuronal Death Caused By Inhibiting Phosphatidylinositol 3-Kinase Or Akt But Not For Death Caused By Nerve Growth Factor Withdrawal. *J Biol Chem* 2000, 275(44):34266-34271.

305. Li F, Chong Zz, Maiese K: Winding Through The Wnt Pathway During Cellular Development And Demise. *Histol Histopathol* 2006, 21(1):103-124.

306. Zhou Z, Wang J, Han X, Zhou J, Linder S: Up-Regulation Of Human Secreted Frizzled Homolog In Apoptosis And Its Down-Regulation In Breast Tumors. *Int J Cancer* 1998, 78(1):95-99.

307. Ko J, Ryu Ks, Lee Yh, Na Ds, Kim Ys, Oh Ym, Kim Is, Kim Jw: Human Secreted Frizzled-Related Protein Is Down-Regulated And Induces Apoptosis In Human Cervical Cancer. *Exp Cell Res* 2002, 280(2):280-287.

308. He B, Reguart N, You L, Mazieres J, Xu Z, Lee Ay, Mikami I, Mccormick F, Jablons Dm: Blockade Of Wnt-1 Signaling Induces Apoptosis In Human Colorectal Cancer Cells Containing Downstream Mutations. *Oncogene* 2005, 24(18):3054-3058.

309. Martin M, Rehani K, Jope Rs, Michalek Sm: Toll-Like Receptor-Mediated Cytokine Production Is Differentially Regulated By Glycogen Synthase Kinase 3. *Nat Immunol* 2005, 6(8):777-784.

310. Guha S, Cullen Jp, Morrow D, Colombo A, Lally C, Walls D, Redmond Em, Cahill Pa: Glycogen Synthase Kinase 3 Beta Positively Regulates Notch Signaling In Vascular Smooth Muscle Cells: Role In Cell Proliferation And Survival. *Basic Res Cardiol* 2011, 106(5):773-785.

311. Frantz S, Kelly Ra, Bourcier T: Role Of Tlr-2 In The Activation Of Nuclear Factor Kappab By Oxidative Stress In Cardiac Myocytes. *J Biol Chem* 2001, 276(7):5197-5203.

312. Wang Z, Zhang Y, Li Y, Banerjee S, Liao J, Sarkar Fh: Down-Regulation Of Notch-1 Contributes To Cell Growth Inhibition And Apoptosis In Pancreatic Cancer Cells. *Mol Cancer Ther* 2006, 5(3):483-493.

313. Artavanis-Tsakonas S, Rand Md, Lake Rj: Notch Signaling: Cell Fate Control And Signal Integration In Development. *Science* 1999, 284(5415):770-776.

314. Li J, Peet Gw, Balzarano D, Li X, Massa P, Barton Rw, Marcu Kb: Novel Nemo/Ikappab Kinase And Nf-Kappa B Target Genes At The Pre-B To Immature B Cell Transition. *J Biol Chem* 2001, 276(21):18579-18590.

315. Kirikoshi H, Sekihara H, Katoh M: Up-Regulation Of Wnt10a By Tumor Necrosis Factor Alpha And Helicobacter Pylori In Gastric Cancer. *Int J Oncol* 2001, 19(3):533-536.

316. He B, Meng Yh, Mivechi Nf: Glycogen Synthase Kinase 3beta And Extracellular Signal-Regulated Kinase Inactivate Heat Shock Transcription Factor 1 By Facilitating The

Disappearance Of Transcriptionally Active Granules After Heat Shock. *Mol Cell Biol* 1998, 18(11):6624-6633.

317. Eldar-Finkelman H, Krebs Eg: Phosphorylation Of Insulin Receptor Substrate 1 By Glycogen Synthase Kinase 3 Impairs Insulin Action. *Proc Natl Acad Sci U S A* 1997, 94(18):9660-9664.

318. Greene Mw: Circadian Rhythms And Tumor Growth. *Cancer Lett* 2012, 318(2):115-123.

319. Wu D, Pan W: Gsk3: A Multifaceted Kinase In Wnt Signaling. *Trends Biochem Sci* 2010, 35(3):161-168.

320. Wagner Ef, Eferl R: Fos/Ap-1 Proteins In Bone And The Immune System. *Immunol Rev* 2005, 208:126-140.

321. Smeyne Rj, Vendrell M, Hayward M, Baker Sj, Miao Gg, Schilling K, Robertson Lm, Curran T, Morgan Ji: Continuous C-Fos Expression Precedes Programmed Cell Death In Vivo. *Nature* 1993, 363(6425):166-169.

322. Day Ml, Zhao X, Wu S, Swanson Pe, Humphrey Pa: Phorbol Ester-Induced Apoptosis Is Accompanied By Ngfi-A And C-Fos Activation In Androgen-Sensitive Prostate Cancer Cells. *Cell Growth Differ* 1994, 5(7):735-741.

323. Preston Ga, Lyon Tt, Yin Y, Lang Je, Solomon G, Annab L, Srinivasan Dg, Alcorta Da, Barrett Jc: Induction Of Apoptosis By C-Fos Protein. *Mol Cell Biol* 1996, 16(1):211-218.

324. Zanella Cl, Timblin Cr, Cummins A, Jung M, Goldberg J, Raabe R, Tritton Tr, Mossman Bt: Asbestos-Induced Phosphorylation Of Epidermal Growth Factor Receptor Is Linked To C-Fos And Apoptosis. *Am J Physiol* 1999, 277(4 Pt 1):L684-693.

325. Zhang X, Zhang L, Yang H, Huang X, Otu H, Libermann Ta, Dewolf Wc, Khosravi-Far R, Olumi Af: C-Fos As A Proapoptotic Agent In Trail-Induced Apoptosis In Prostate Cancer Cells. *Cancer Res* 2007, 67(19):9425-9434.

326. Mannello F, Gazzanelli G: Tissue Inhibitors Of Metalloproteinases And Programmed Cell Death: Conundrums, Controversies And Potential Implications. *Apoptosis* 2001, 6(6):479-482.

327. Sternlicht Md, Werb Z: How Matrix Metalloproteinases Regulate Cell Behavior. *Annu Rev Cell Dev Biol* 2001, 17:463-516.

328. Wang X, Lee Sr, Arai K, Tsuji K, Rebeck Gw, Lo Eh: Lipoprotein Receptor-Mediated Induction Of Matrix Metalloproteinase By Tissue Plasminogen Activator. *Nat Med* 2003, 9(10):1313-1317.

329. Tamura F, Nakagawa R, Akuta T, Okamoto S, Hamada S, Maeda H, Kawabata S, Akaike T: Proapoptotic Effect Of Proteolytic Activation Of Matrix Metalloproteinases By Streptococcus Pyogenes Thiol Proteinase (Streptococcus Pyrogenic Exotoxin B). *Infect Immun* 2004, 72(8):4836-4847.

330. Igney Fh, Krammer Ph: Death And Anti-Death: Tumour Resistance To Apoptosis. *Nat Rev Cancer* 2002, 2(4):277-288.

331. Curtis Am, Seo Sb, Westgate Ej, Rudic Rd, Smyth Em, Chakravarti D, Fitzgerald Ga, Mcnamara P: Histone Acetyltransferase-Dependent Chromatin Remodeling And The Vascular Clock. *J Biol Chem* 2004, 279(8):7091-7097.

332. Lee Js, See Rh, Deng T, Shi Y: Adenovirus E1a Downregulates Cjun- And Junb-Mediated Transcription By Targeting Their Coactivator P300. *Mol Cell Biol* 1996, 16(8):4312-4326.

333. Maggirwar Sb, Ramirez S, Tong N, Gelbard Ha, Dewhurst S: Functional Interplay Between Nuclear Factor-Kappab And C-Jun Integrated By Coactivator P300 Determines The Survival Of Nerve Growth Factor-Dependent Pc12 Cells. *J Neurochem* 2000, 74(2):527-539.

334. Glauser Da, Schlegel W: Sequential Actions Of Erk1/2 On The Ap-1 Transcription Factor Allow Temporal Integration Of Metabolic Signals In Pancreatic Beta Cells. *Faseb J* 2007, 21(12):3240-3249.

335. Relaix F, Wei Xj, Wu X, Sassoon Da: Peg3/Pw1 Is An Imprinted Gene Involved In The Tnf-Nfkappab Signal Transduction Pathway. *Nat Genet* 1998, 18(3):287-291.

336. Cheng G, Baltimore D: Tank, A Co-Inducer With Traf2 Of Tnf- And Cd 40l-Mediated Nf-Kappab Activation. *Genes Dev* 1996, 10(8):963-973.

337. Kaye Km, Devergne O, Harada Jn, Izumi Km, Yalamanchili R, Kieff E, Mosialos G: Tumor Necrosis Factor Receptor Associated Factor 2 Is A Mediator Of Nf-Kappa B Activation By Latent Infection Membrane Protein 1, The Epstein-Barr Virus Transforming Protein. *Proc Natl Acad Sci U S A* 1996, 93(20):11085-11090.

338. Polek Tc, Talpaz M, Darnay Bg, Spivak-Kroizman T: Tweak Mediates Signal Transduction And Differentiation Of Raw264.7 Cells In The Absence Of Fn14/Tweakr. Evidence For A Second Tweak Receptor. *J Biol Chem* 2003, 278(34):32317-32323.

339. Han S, Yoon K, Lee K, Kim K, Jang H, Lee Nk, Hwang K, Young Lee S: Tnf-Related Weak Inducer Of Apoptosis Receptor, A Tnf Receptor Superfamily Member, Activates Nf-Kappa B Through Tnf Receptor-Associated Factors. *Biochem Biophys Res Commun* 2003, 305(4):789-796.

340. Saitoh T, Nakayama M, Nakano H, Yagita H, Yamamoto N, Yamaoka S: Tweak Induces Nf-Kappab2 P100 Processing And Long Lasting Nf-Kappab Activation. *J Biol Chem* 2003, 278(38):36005-36012.

341. Kawakita T, Shiraki K, Yamanaka Y, Yamaguchi Y, Saitou Y, Enokimura N, Yamamoto N, Okano H, Sugimoto K, Murata K *Et Al*: Functional Expression Of Tweak In Human Hepatocellular Carcinoma: Possible Implication In Cell Proliferation And Tumor Angiogenesis. *Biochem Biophys Res Commun* 2004, 318(3):726-733.

342. Tran Nl, Mcdonough Ws, Savitch Ba, Sawyer Tf, Winkles Ja, Berens Me: The Tumor Necrosis Factor-Like Weak Inducer Of Apoptosis (Tweak)-Fibroblast Growth Factor-Inducible 14 (Fn14) Signaling System Regulates Glioma Cell Survival Via Nfkappab Pathway Activation And Bcl-Xl/Bcl-W Expression. *J Biol Chem* 2005, 280(5):3483-3492.

343. De Ketelaere A, Vermeulen L, Vialard J, Van De Weyer I, Van Wauwe J, Haegeman G, Moelans I: Involvement Of Gsk-3beta In Tweak-Mediated Nf-Kappab Activation. *Febs Lett* 2004, 566(1-3):60-64.

344. Ma O, Cai Ww, Zender L, Dayaram T, Shen J, Herron Aj, Lowe Sw, Man Tk, Lau Cc, Donehower La: Mmp13, Birc2 (Ciap1), And Birc3 (Ciap2), Amplified On Chromosome 9, Collaborate With P53 Deficiency In Mouse Osteosarcoma Progression. *Cancer Res* 2009, 69(6):2559-2567.

345. Meierjohann S, Hufnagel A, Wende E, Kleinschmidt Ma, Wolf K, Friedl P, Gaubatz S, Schartl M: Mmp13 Mediates Cell Cycle Progression In Melanocytes And Melanoma Cells: In Vitro Studies Of Migration And Proliferation. *Mol Cancer* 2010, 9:201.

346. Mardilovich K, Pankratz Sl, Shaw Lm: Expression And Function Of The Insulin Receptor Substrate Proteins In Cancer. *Cell Commun Signal* 2009, 7:14.

347. Cesarone G, Garofalo C, Abrams Mt, Igoucheva O, Alexeev V, Yoon K, Surmacz E, Wickstrom E: Rnai-Mediated Silencing Of Insulin Receptor Substrate 1 (Irs-1) Enhances Tamoxifen-Induced Cell Death In Mcf-7 Breast Cancer Cells. *J Cell Biochem* 2006, 98(2):440-450.

348. Tanaka S, Wands Jr: Insulin Receptor Substrate 1 Overexpression In Human Hepatocellular Carcinoma Cells Prevents Transforming Growth Factor Beta1-Induced Apoptosis. *Cancer Res* 1996, 56(15):3391-3394.

349. Yang-Yen Hf, Chambard Jc, Sun Yl, Smeal T, Schmidt Tj, Drouin J, Karin M: Transcriptional Interference Between C-Jun And The Glucocorticoid Receptor: Mutual Inhibition Of Dna Binding Due To Direct Protein-Protein Interaction. *Cell* 1990, 62(6):1205-1215.

350. Asher G, Gatfield D, Stratmann M, Reinke H, Dibner C, Kreppel F, Mostoslavsky R, Alt Fw, Schibler U: Sirt1 Regulates Circadian Clock Gene Expression Through Per2 Deacetylation. *Cell* 2008, 134(2):317-328.

351. Xiong S, Salazar G, Patrushev N, Alexander Rw: Foxo1 Mediates An Autofeedback Loop Regulating Sirt1 Expression. *J Biol Chem* 2011, 286(7):5289-5299.

352. Qiao L, Shao J: Sirt1 Regulates Adiponectin Gene Expression Through Foxo1-C/Enhancer-Binding Protein Alpha Transcriptional Complex. *J Biol Chem* 2006, 281(52):39915-39924.

353. Maeda K, Okubo K, Shimomura I, Funahashi T, Matsuzawa Y, Matsubara K: Cdna Cloning And Expression Of A Novel Adipose Specific Collagen-Like Factor, Apm1 (Adipose Most Abundant Gene Transcript 1). *Biochem Biophys Res Commun* 1996, 221(2):286-289.

354. Hu E, Liang P, Spiegelman Bm: Adipoq Is A Novel Adipose-Specific Gene Dysregulated In Obesity. *J Biol Chem* 1996, 271(18):10697-10703.

355. Scherer Pe, Williams S, Fogliano M, Baldini G, Lodish Hf: A Novel Serum Protein Similar To C1q, Produced Exclusively In Adipocytes. *J Biol Chem* 1995, 270(45):26746-26749.

356. Yokota T, Oritani K, Takahashi I, Ishikawa J, Matsuyama A, Ouchi N, Kihara S, Funahashi T, Tenner Aj, Tomiyama Y *Et Al*: Adiponectin, A New Member Of The Family Of Soluble Defense Collagens, Negatively Regulates The Growth Of Myelomonocytic Progenitors And The Functions Of Macrophages. *Blood* 2000, 96(5):1723-1732.

357. Brakenhielm E, Veitonmaki N, Cao R, Kihara S, Matsuzawa Y, Zhivotovsky B, Funahashi T, Cao Y: Adiponectin-Induced Antiangiogenesis And Antitumor Activity Involve Caspase-Mediated Endothelial Cell Apoptosis. *Proc Natl Acad Sci U S A* 2004, 101(8):2476-2481.

358. Bub Jd, Miyazaki T, Iwamoto Y: Adiponectin As A Growth Inhibitor In Prostate Cancer Cells. *Biochem Biophys Res Commun* 2006, 340(4):1158-1166.

359. Kwon Hs, Brent Mm, Getachew R, Jayakumar P, Chen Lf, Schnolzer M, Mcburney Mw, Marmorstein R, Greene Wc, Ott M: Human Immunodeficiency Virus Type 1 Tat Protein Inhibits The Sirt1 Deacetylase And Induces T Cell Hyperactivation. *Cell Host Microbe* 2008, 3(3):158-167.

360. Azenshtein E, Luboshits G, Shina S, Neumark E, Shahbazian D, Weil M, Wigler N, Keydar I, Ben-Baruch A: The Cc Chemokine Rantes In Breast Carcinoma Progression: Regulation Of Expression And Potential Mechanisms Of Promalignant Activity. *Cancer Res* 2002, 62(4):1093-1102.

361. Saji H, Koike M, Yamori T, Saji S, Seiki M, Matsushima K, Toi M: Significant Correlation Of Monocyte Chemoattractant Protein-1 Expression With Neovascularization And Progression Of Breast Carcinoma. *Cancer* 2001, 92(5):1085-1091.

362. Monti P, Leone Be, Marchesi F, Balzano G, Zerbi A, Scaltrini F, Pasquali C, Calori G, Pessi F, Sperti C *Et Al*: The Cc Chemokine Mcp-1/Ccl2 In Pancreatic Cancer Progression: Regulation Of Expression And Potential Mechanisms Of Antimalignant Activity. *Cancer Res* 2003, 63(21):7451-7461.

363. Zhang Ee, Liu Ac, Hirota T, Miraglia Lj, Welch G, Pongsawakul Py, Liu X, Atwood A, Huss Jw, 3rd, Janes J *Et Al*: A Genome-Wide Rnai Screen For Modifiers Of The Circadian Clock In Human Cells. *Cell* 2009, 139(1):199-210.

364. Guillaumond F, Dardente H, Giguere V, Cermakian N: Differential Control Of Bmal1 Circadian Transcription By Rev-Erb And Ror Nuclear Receptors. *J Biol Rhythms* 2005, 20(5):391-403.

365. Dewberry Rm, King Ar, Crossman Dc, Francis Se: Interleukin-1 Receptor Antagonist (Il-1ra) Modulates Endothelial Cell Proliferation. *Febs Lett* 2008, 582(6):886-890.

366. Hua H, Wang Y, Wan C, Liu Y, Zhu B, Yang C, Wang X, Wang Z, Cornelissen-Guillaume G, Halberg F: Circadian Gene Mper2 Overexpression Induces Cancer Cell Apoptosis. *Cancer Sci* 2006, 97(7):589-596.

367. Bult Cj, Eppig Jt, Kadin Ja, Richardson Je, Blake Ja: The Mouse Genome Database (Mgd): Mouse Biology And Model Systems. *Nucleic Acids Res* 2008, 36(Database Issue):D724-728.

368. Yujiri T, Sather S, Fanger Gr, Johnson Gl: Role Of Mekk1 In Cell Survival And Activation Of Jnk And Erk Pathways Defined By Targeted Gene Disruption. *Science* 1998, 282(5395):1911-1914.

369. Widmann C, Gerwins P, Johnson Nl, Jarpe Mb, Johnson Gl: Mek Kinase 1, A Substrate For Devd-Directed Caspases, Is Involved In Genotoxin-Induced Apoptosis. *Mol Cell Biol* 1998, 18(4):2416-2429.

370. Oswald F, Tauber B, Dobner T, Bourteele S, Kostezka U, Adler G, Liptay S, Schmid Rm: P300 Acts As A Transcriptional Coactivator For Mammalian Notch-1. *Mol Cell Biol* 2001, 21(22):7761-7774.

371. Bradney C, Hjelmeland M, Komatsu Y, Yoshida M, Yao Tp, Zhuang Y: Regulation Of E2a Activities By Histone Acetyltransferases In B Lymphocyte Development. *J Biol Chem* 2003, 278(4):2370-2376.

372. Jiang Y, Zhou Xd, Liu Yk, Huang Xw, Zhao Y, Xue Q, Sun Rx, Chen J, Wu X: Antisense Tcf Inhibits The Neoplastic Growth Of Liver Cancer Cells. *J Cancer Res Clin Oncol* 2004, 130(11):671-678.

373. Zhang Z, Hartmann H, Do Vm, Abramowski D, Sturchler-Pierrat C, Staufenbiel M, Sommer B, Van De Wetering M, Clevers H, Saftig P *Et Al*: Destabilization Of Beta-Catenin By Mutations In Presenilin-1 Potentiates Neuronal Apoptosis. *Nature* 1998, 395(6703):698-702.

374. Schlingensiepen Kh, Schlingensiepen R, Steinbrecher A, Hau P, Bogdahn U, Fischer-Blass B, Jachimczak P: Targeted Tumor Therapy With The Tgf-Beta 2 Antisense Compound Ap 12009. *Cytokine Growth Factor Rev* 2006, 17(1-2):129-139.

375. Bharti Ac, Takada Y, Shishodia S, Aggarwal Bb: Evidence That Receptor Activator Of Nuclear Factor (Nf)-Kappab Ligand Can Suppress Cell Proliferation And Induce Apoptosis Through Activation Of A Nf-Kappab-Independent And Traf6-Dependent Mechanism. *J Biol Chem* 2004, 279(7):6065-6076.

376. Roux S, Meignin V, Quillard J, Meduri G, Guiochon-Mantel A, Fermand Jp, Milgrom E, Mariette X: Rank (Receptor Activator Of Nuclear Factor-Kappab) And Rankl Expression In Multiple Myeloma. *Br J Haematol* 2002, 117(1):86-92.

377. Lacey Dl, Timms E, Tan Hl, Kelley Mj, Dunstan Cr, Burgess T, Elliott R, Colombero A, Elliott G, Scully S *Et Al*: Osteoprotegerin Ligand Is A Cytokine That Regulates Osteoclast Differentiation And Activation. *Cell* 1998, 93(2):165-176.

378. Malyankar Um, Scatena M, Suchland Kl, Yun Tj, Clark Ea, Giachelli Cm: Osteoprotegerin Is An Alpha Vbeta 3-Induced, Nf-Kappa B-Dependent Survival Factor For Endothelial Cells. *J Biol Chem* 2000, 275(28):20959-20962.

379. Fuller K, Wong B, Fox S, Choi Y, Chambers Tj: Trance Is Necessary And Sufficient For Osteoblast-Mediated Activation Of Bone Resorption In Osteoclasts. *J Exp Med* 1998, 188(5):997-1001.

380. Ebelt H, Zhang Y, Kampke A, Xu J, Schlitt A, Buerke M, Muller-Werdan U, Werdan K, Braun T: E2f2 Expression Induces Proliferation Of Terminally Differentiated Cardiomyocytes In Vivo. *Cardiovasc Res* 2008, 80(2):219-226.

381. Seo J, Chung Ys, Sharma Gg, Moon E, Burack Wr, Pandita Tk, Choi K: Cdt1 Transgenic Mice Develop Lymphoblastic Lymphoma In The Absence Of P53. *Oncogene* 2005, 24(55):8176-8186.

382. Cornu M, Yang Jy, Jaccard E, Poussin C, Widmann C, Thorens B: Glucagon-Like Peptide-1 Protects Beta-Cells Against Apoptosis By Increasing The Activity Of An Igf-2/Igf-1 Receptor Autocrine Loop. *Diabetes* 2009, 58(8):1816-1825.

383. Damon Se, Maddison L, Ware Jl, Plymate Sr: Overexpression Of An Inhibitory Insulin-Like Growth Factor Binding Protein (Igfbp), Igfbp-4, Delays Onset Of Prostate Tumor Formation. *Endocrinology* 1998, 139(8):3456-3464.

384. Durai R, Yang Sy, Sales Km, Seifalian Am, Goldspink G, Winslet Mc: Increased Apoptosis And Decreased Proliferation Of Colorectal Cancer Cells Using Insulin-Like Growth Factor Binding Protein-4 Gene Delivered Locally By Gene Transfer. *Colorectal Dis* 2007, 9(7):625-631.

385. Muratovska A, Zhou C, He S, Goodyer P, Eccles Mr: Paired-Box Genes Are Frequently Expressed In Cancer And Often Required For Cancer Cell Survival. *Oncogene* 2003, 22(39):7989-7997.

386. Destici E, Oklejewicz M, Saito S, Van Der Horst Gt: Mammalian Cryptochromes Impinge On Cell Cycle Progression In A Circadian Clock-Independent Manner. *Cell Cycle* 2011, 10(21):3788-3797.

387. Humphrey W, Dalke A, Schulten K: Vmd: Visual Molecular Dynamics. *J Mol Graph* 1996, 14(1):33-38, 27-38.

VITA

Gözde Kar was born in Bursa, Turkey, on June 7, 1983. She received the B.Sc. Degree in Chemical Engineering from Boğaziçi University in 2006 and M.Sc. Degree in Computational Science and Engineering from Koç University in 2008. She received the Ph.D. degree from Koc University in Chemical and Biological Engineering in 2012. From September 2006 to September 2012 she worked as teaching and research assistant at Koç University.

Her research mainly focuses on structural modeling and analysis of protein-protein interactions using computational methods. She has published articles in prestigious journals such as PLOS Computational Biology, Journal of Proteome Research, Molecular Biosystems, Current Opinion in Pharmacology and Briefings in Bioinformatics.