CHAPTER IV

DISCUSSION

4.1 PPI PREDICTION ALGORITHM

Protein domains are evolutionarily conserved regions and form units of interaction. Although our algorithm derives data from three organisms, we were able to predict PPIs across a diverse set of 109 organisms represented in the test sets. This shows that domain interactions are largely universal and not species dependent. We discuss some of the examples from our positive predictions. The well known examples include interactions of protein kinases (Pfam ID: PF00069) with cyclin N-terminal (Pfam ID: PF00134) and cyclin C-terminal domain containing proteins (Pfam ID: PF02984) (Zheng et al 2000) and interactions between RING finger domain (C3HC4, Pfam ID: PF00097) containing proteins and ubiquitin conjugating enzymes (Pfam ID: PF00179) (Brown et al 1999). One of the smallest members of the Inhibitor of Apoptosis Protein family, the survivin protein plays a role in apoptosis and cell mitosis. It has a Baculovirus Inhibitor Repeat (BIR) domain PF00653. During mitosis it is associated with regulators of cytokinesis including Aurora Kinase B, a protein with the kinase domain PF00069 (Fu et al 2007). One of the PPIs from Intact test set is an interaction between survivin and Aurora kinase B. Our algorithm successfully identifies this interaction. Yet another example is the interaction of protein kinase with Heat shock protein 20 (Hsp20) (Pfam ID: PF00011). In vascular muscle, Hsp20s are shown to be a target for Protein Kinase A (Beall et al 1997). Another positive prediction is the interaction of Hsp20s with Bcl2-associated athanogene (BAG) proteins.
BAG proteins form a family of co-chaperones which interact with heat shock proteins through a specific structural domain known as BAG domain (Pfam Id: PF02179). BAG3, a BAG domain containing protein interacts with Hsp20 (Rosati et al 2011). Another interesting example is a self interaction between CBS domains (Pfam ID: PF00571). CBS domains are ubiquitous domains found in cystathione beta synthases, inosine monophosphate dehydrogenases, voltage gated chloride channels amongst others. They are implicated in a wide variety of cellular functions. CBS domains are mostly found in pairs wherein they associate with each other through their beta sheets in a pseudodimeric arrangement to form a CBS pair or the Bateman domain (Zhangs et al 1999; Kemp 2004). Other examples of our predictions for self-interactions amongst domains include Leucine Rich Repeats (LRRs), 6 Phospho fructo 2-kinase and Biopterin. LRRs are present in a number of proteins with diverse array of functions which include immune responses, extracellular matrix interactions, bacterial pathogenicity, disease resistance and signaling (Park et al 2008). LRRs form a stable structural scaffold for PPI (Kobe & Kajava 2001). Decorin and biglycan are proteoglycans with LRRs in their protein core. They participate in a number of interactions with extra cellular matrix proteins. The crystal structure of dimeric protein core of decorin reveals that dimerization occurs by apposition of the concave sides of the LRR domains (Scott et al 2004). Crystal structure of biglycan shows a similar pattern of dimerization and also provides evidence that dimerization is essential for folding and stability of Class I Small LRR proteoglycans (Scott et al 2006). This self-interaction between LRR domains is predicted by our algorithm.

Inhibiting PPI interfaces between host and pathogens/viruses is of therapeutic interest. Therefore any insights into host pathogen interactions will be of value in therapeutic drug discovery processes. Some examples are available in the test sets. The capsid protein derived from the gag protein rearranges to form the cone shaped core of the Human Immunodeficiency
Virus 1 (HIV-1) virion (Vajdos et al 1997). About 200 copies of the cytoplasmic protein Cyclophilin A is packaged into the virion and this assembly is facilitated through the interaction of Cyclophilin A with the capsid domain of viral gag protein (Franke et al 1994; Thali et al 1994). The interaction between Cyclophilin A (Pfam PF00160) and gag protein of HIV-1 (Pfam PF00098, PF00540, PF00607, PF08705) in the IntAct test set was accurately predicted by us.

**Positive domain associations values in our DDA matrices $T_{org}$ and $T_{ext}$**

We chose to find positive DDA from our $T_{org}$ and $T_{ext}$ using the same threshold criterion of value $t \geq 0.01$. We find from our DDA, about 1% of the domains are positively associated with each other. Details about such domain pairs are provided in Table A1.1. These can be used directly for predicting PPI between single domain protein pairs. We illustrate with two examples as to how DDA information from this study can be utilized. From our DDA information for single domain proteins we find that PF00079 has a positive correlation with PF00240 and therefore we predict a PPI between serpin B7 (UniProt ID: O75635, Pfam ID: PF00079) and ubiquitin like protein 5 (UniProt ID: Q9BZL1, Pfam ID: PF00240). Serpins are serine protease inhibitors which play an important role in biological processes like complement activation, fibrinolysis and apoptosis. The expression studies of serpin 2a in macrophages activated by incubation with bacterial products shows that the serpin 2a is conjugated to an ubiquitin homolog known as IFN-stimulated gene ISG15. The expression of serpin 2a and ISG15 is also induced by incubation with LPS or through *in vivo* infection with BCG (Hamerman et al 2002). Another positive domain correlation in our DDA is between Pfam domains PF00125 and PF00956. Based on this we can predict an interaction between histone 2A protein (UniProt ID: A0PK90, Pfam ID: PF00125) and Nucleosome Assembly Protein (NAP) (UniProt ID: Q96NT1,
Pfam ID: PF00956). NAPs participate in the intracellular transport of histones. They form both homomers and heteromers and can bind to histones. NAP can also form a ternary complex involving histones and p300. During cell cycle progression NAPs are associated with core histones and cyclin B and they undergo nucleocytoplasmic shuttling (Shikama et al 2000). This shows that the predicted interactions are accurate. In this manner, single domain PPI can easily be assessed by using data from Table A1.1.

4.2 **PROTEIN-PROTEIN DOCKING STUDIES OF PREDICTED DDAs**

Using the predicted DDAs, protein-protein docking study was performed to validate the interactions of Serpin B7 and ubiquitin like protein 5, lim domain containing protein and its interacting partners WW domain, protein kinase domain. The expression studies of serpin 2a in macrophages activated by incubation with bacterial products shows that the serpin 2a is conjugated to an ubiquitin homolog known as IFN-stimulated gene ISG15. The expression of serpin 2a and ISG15 is also induced by incubation with LPS or through *in vivo* infection with BCG (Hamerman et al 2002).

Protein-protein interaction showed positive association between lim and WW domain and lim and protein kinase domain. The lim domain consists of two zinc-finger motifs which direct protein-protein interactions. This domain is identified in diverse number of proteins and it plays a vital role in gene regulation, tumour formation and many others (Zheng & Zhao 2007). The WW domain acts as a protein interacting module and mediates protein-protein interactions through its proline rich interacting motif. WW domain-containing proteins are involved in a variety of biological processes which includes transcription, apoptosis, differentiation, receptor signalling, splicing and ubiquitination (Ingham et al 2005; Salah & Aqeilan 2011). The association of lim domain containing protein and the protein with WW
domain is possible since the WW domain acts as a protein interacting module and provides a platform for multiprotein networks. It is also reported that WW domain interactions regulate the Hippo signalling pathway in tumorigenesis.

The Hippo pathway regulates tissue growth by regulating cell growth, cell proliferation, differentiation and apoptosis. The Hippo pathway consists of a kinase cascade core, which is activated by a mechanism that is not fully established. The interaction of WW domain with lim domain may inhibit its activity in the kinase cascade, thereby resulting in anti-apoptosis and affect the Hippo signalling pathway resulting in positive regulation of cell proliferation (Salah & Aqeilan 2011). The interaction between lim domain containing protein and kinase domain protein showed positive association. Kuroda et al (1996) identified a novel protein with three lim domains which are associated with protein kinase domain and suggested LIM domain as one of the targets of PKC. The interaction of lim domain with protein kinase probably provides the clue for understanding the undefined roles of LIM domain-containing proteins.

4.3 DOCKING AND MD SIMULATION OF AAT AND CASPASE-3

Human alpha 1 antitrypsin is a member of serpin superfamily that inhibits serine proteases as well cysteine proteases like caspases. The structure of AAT is composed of 3 beta sheets and 7-9 alpha helices and the RCL (Belorgey et al 2007). The RCL of inhibitory serpins contains the cleavage site of the protease. In the native form serpins are in metastable conformation which is called as the stressed state and when it interacts with proteases serpins undergo conformational changes and form a cleaved state. These conformational changes are mediated by the RCL. Inhibitory mechanism of AAT involves the conformational changes S → R transition. Based on the crystal structures of the two states a model for the conversion of S->R state
has been proposed (Whisstock et al 2000). MD simulation was performed to study the conformational changes that take place during S->R transition.

In the uncleaved AAT, the RCL is more flexible and has high fluctuations. Other than RCL the S5C and gate region has high mobility. The gate region is a beta turn which links the two sheets S3C and S4C. This shows that the uncleaved AAT is free to interact with the protease. Literature reports showed that these regions play a major role during the insertion of RCL (Bottomley et al 2001). Interestingly, it is observed that in the uncleaved AAT the proximal hinge region has high mobility whereas, in the cleaved AAT the distal hinge has high fluctuations. While the protein undergoes S to R transition the hinge region plays an important role in the conformational change of the RCL by providing mobility. The proximal hinge undergoes conformational changes during the insertion of RCL in to the A beta sheet during S->R transition. The shutter and breach regions help in the sheet opening and allow the insertion of the RCL. On analyzing the distance between the sheets S5A and S3A, the difference in distance showed that it may facilitate the sheet opening. Kass et al (2012) investigated the wild type AAT and Z variant of AAT which showed that the breach region at the top of beta sheet A was closed in wild type AAT whereas it is opened in Z variant AAT.

**MD simulation of AAT- caspase-3 complex**

Administration of AAT prevented T1D in NOD mice while over expression of AAT resulted in reduced insulitis and prevented the development of hyperglycemia in NOD mice (Zhang et al 2007). Although AAT is a serine protease inhibitor, studies have shown that it can inhibit aspartate proteases like caspases. AAT protects beta cells from apoptosis. They were able to demonstrate that AAT directly inhibits and completely abolishes caspase-3 activity. Thus this association confers anti apoptotic or a
protective role for AAT in T1D. Since caspase-3 is a common player in T1D and Type-2 diabetes (T2D), AAT has a potential role in the treatment of T1D and T2D. In the current study we sought to understand the association of AAT with caspase-3 at a molecular level using molecular docking and simulations. Our results indicate that AAT interacts with caspase-3. The interaction involves the active site of caspase-3 and the RCL group of AAT. Further, we find that this interaction is stable in the molecular dynamics simulation performed for 50 ns.

4.4 DOMAIN COMBINATIONS

Analysis of domain combinations shows that there seems to be a preference for certain domains to co-exist in nature. This domain combination preference may be driven in evolution through a necessity to co-exist for functional or structural integrity of the protein. The fact that certain domains can only exist in combination with certain specific domains and not as singletons suggests that these behave like composite domains which are probably functionally related. Further, analysis of proteins with 2 distinct domains and their order has revealed that order in which these two domains occur does not seem to affect the protein function. Terminii anchoring preferences of domains reflect their positional preference driven perhaps by function. A comprehensive survey of UniprotKB has revealed the presence of a number of Rosetta stone sequences which has a utility in functional annotation. The correlation of the domains in Rosetta sequences with domain association data from our algorithm further confirms that our algorithm has captured true domain domain associations found in nature.