Biophysical characterization of cytosolic domains of three import receptors from Arabidopsis thaliana

RASHMI REKHA PANIGRAHI (M.Sc.)

School of Chemistry and Biochemistry
University of Western Australia, WA, Australia

July 2014

This thesis is submitted in fulfillment of the requirements for the degree of
Doctor of Philosophy in Biochemistry

Co-ordinating Supervisor: Prof. Alice Vrielink
Co-supervisor: Prof. Jim Whelan
Co-supervisor: Prof. Charlie Bond
Dedicated to

my beloved husband

Senthil
ABSTRACT

Eukaryotic cells are characterized by the presence of several organelles that are surrounded by membranes. The majority of proteins destined to these organelles are nuclear encoded, synthesized in the cytosol and translocated across the membrane aided by specific receptors, into destined locations in the organelles. The receptors involved in protein translocation can be classified into two categories: 1) those that recognize precursor proteins through interaction with their specific presequences and 2) those that recognize cytosolic transporters (chaperones) which are involved in the delivery of precursor proteins to the import machinery for the specific organelle to which they are targeted.

The projects aim to biophysically characterize the cytosolic domains of three import receptors from *Arabidopsis thaliana* and identifying their interactions with their binding partners. The first part of the study focuses on the cytosolic domain of Metaxin, an outer mitochondrial membrane protein. To date literature is not available delineating the exact domain boundaries of the cytosolic region of the protein. Mutational studies aid in identifying the region that can be expressed as a soluble dimer however time and concentration dependent aggregation prohibits further characterization.

The second part of the project involves biophysical characterization of the cytosolic TPR domain of Toc64, a receptor on the outer membrane of the chloroplast. Previous studies with the human homologues of the chaperones suggested that this domain interacts with Hsp70/90 chaperones, with a preference for Hsp90. Using isothermal calorimetry and computational approaches, the energetics and the important residues involved in the interaction is analysed. These studies suggest that the TPR domain recognizes both chaperones with equimolar affinity contradicting previous findings.

Om64, a homologue of Toc64, is TPR containing protein found in the outer mitochondrial membrane and implicated in protein import into the mitochondrion. The cytosolic TPR domain of this protein is shown to interact with the C-terminal region of Hsp90. In the third part of the project, a comparative analysis of the inter-residue interactions between TPR domains of Toc64 and Om64 with the C-terminal regions of Hsp70/Hsp90 is performed using simulation studies. Furthermore, the binding energies of the interactions are calculated. Our results suggest that Toc64 exhibits higher affinity
towards Hsp70 than Hsp90, however Om64 interacts with Hsp70 and Hsp90 with similar affinity. Our findings provide an explanation for the absence of mistargeting of chloroplast precursors into mitochondria.

In summary, this study reveals that the cytosolic receptor domain of Metaxin may require additional components for its stability in vitro. Furthermore, this study gives insights into the specificity of chaperone-receptor interaction and thus protein targeting to specific organelles such as mitochondria and chloroplast at the molecular level.
DECLARATION

The work described in this thesis was carried out between June 2009 and July 2013 in the School of Chemistry and Biochemistry at University of Western Australia. The author carried out the experiments unless otherwise stated. This work has not been submitted, in part or in full, for a higher degree in any other institution.

_______________________________
Rashmi Rekha Panigrahi
STATEMENT OF CONTRIBUTIONS

   **Author Contributions:** RP performed the literature survey and wrote the manuscript. JW and AV are finalizing the manuscript. Overall contribution by RP: 90%

   **Author Contributions:** RP cloned, expressed and purified the protein, and performed all the biophysical and computational studies. AA trained RP with ITC experiments and data analysis. RP, JW and AV all finalized the manuscript. Overall contribution by RP: 85%

   **Author Contributions:** RP designed and completed simulation studies with the help of AV. RP, JW and AV finalized the manuscript. Overall contribution by RP: 90%
ACKNOWLEDGEMENTS

This thesis is an outcome of mentorship and inspiration from several outstanding individuals. I take this opportunity to express my sincere gratitude to them. I am privileged to write this thesis under the guidance of Prof. Alice Vrielink and Prof. Jim Whelan.

To my supervisor Prof. Alice Vrielink, I am deeply indebted for your guidance, wisdom and encouragement, starting from applying for IPRS (in 2008) to submission of the thesis (2014). Your perpetual energy and enthusiasm in the form of support has helped me to shape the work in the current form. I can never forget the lemon cakes that you had specially made for me, the lunch parties at your house and the awesome Rottnest trip. Specially your words: “well done” has always motivated me to go to the next level.

To my co-supervisor Prof. Jim Whelan, my sincere thanks for your invaluable advice and insightful comments. I appreciate all your contributions of time and ideas, to make my PhD thesis productive. Your enthusiasm towards my research work has been motivational for me.

To Prof. M. R. N. Murthy and Prof. H. S. Savithri, I profusely thank you for introducing me to the field of Structural biology, which has now developed as my passion. Thanks for allowing me to work on multiple projects, experiences in which have helped me in my PhD. Working with you has been an invaluable and incredibly rewarding experience. Special thanks for allowing me to use your lab space in the last seven months to write the manuscripts and the thesis.

My research work would not have been possible without the funding support from Prof. Vrielink’s startup grant, SIRF and travel awards provided by UWA and excellent computational facility provided by IVEC. Special thanks to Christopher from IVEC for technical support.

To Chris, Mihwa and Daniel, special thanks for sharing with me the secrets of molecular biology and thanks to all the AV/Bond lab members: Sanjika, Mihwa, Chris, Daniel, Natalie, Amanda, Rohan, Emily, Ben, Gavin, Anandhi, Ash, Caroline, Cameron, Yee Seng, Claudia and Martin, for your help and stimulating discussions.
To the MRN lab members: Geeta, Subashini, Mamta, Yamuna, Shveta, Bharath, Ashutosh, Satya and Ithayaraja, thank you for welcoming me to such a vibrant group.

Great friends are difficult to find, hard to leave and impossible to forget. To Beena, Neha, Emily, Natalie, Wresti, Maram, Gavin, Ash, Kylie and Chris, though we were together for few years in Perth, your affection and the happy moments that we shared has left an inerasable memory. I wish to meet and work with you again in the future.

To my family, though 5000 miles away, you made me feel as if you were always by my side. Thank you for being so supportive.

I am sorry for all the dead bacteria during my work. But, be proud you are part of my thesis and have made it in a publication!
TABLE OF CONTENT

CHAPTER 1 .......................................................................................................................... 1

THE DESIGN AND STRUCTURE OF OUTER MEMBRANE RECEPTORS FROM
PEROXISOMES, MITOCHONDRIA AND PLASTIDS ....................................................... 1
1.1 INTRODUCTION ........................................................................................................... 3
  1.1.1 General outline of protein import into organelles .............................................. 4
  1.1.2 Targeting signals ............................................................................................... 5
1.2 STRUCTURAL OVERVIEW OF OUTER MEMBRANE IMPORT
RECEPTORS ....................................................................................................................... 10
  1.2.1 Outer membrane receptors of peroxisomes .................................................... 10
  1.2.2 Outer membrane receptors of mitochondria .................................................. 20
  1.2.3 Outer membrane receptors of chloroplast ...................................................... 31
1.3 DISCUSSION .................................................................................................................. 37
1.4 FUTURE PERSPECTIVES .......................................................................................... 38
1.5 SCOPE OF THE THESIS ............................................................................................ 40

CHAPTER 2 .......................................................................................................................... 43

BIOPHYSICAL CHARACTERIZATION OF RECOMBINANT PROTEIN
EXPRESSING THE CYTOSOLIC DOMAIN OF Metaxin FROM Arabidopsis
Thaliana ......................................................................................................................... 43
2.1 INTRODUCTION ........................................................................................................... 45
2.2 MATERIALS AND METHODS .................................................................................. 48
  2.2.1 Bioinformatics analysis ..................................................................................... 48
  2.2.2 Peptide synthesis ............................................................................................. 48
  2.2.3 Construction of plasmid expressing cytosolic domain of Mtx ......................... 48
  2.2.4 Over expression and purification of deletion constructs .................................. 49
  2.2.5 Competitive import inhibition assay ................................................................. 50
  2.2.6 Gel filtration chromatography ......................................................................... 51
  2.2.7 Circular dichroism (CD) and thermal shift assay (TSA) .................................. 51
  2.2.8 Dynamic light scattering (DLS) and analytical ultracentrifugation (AUC) .... 52
  2.2.9 Non reducing PAGE and native PAGE analysis ............................................... 53
2.2.10 Limited proteolysis ................................................................. 53
2.2.11 Isothermal titration calorimetry (ITC) ........................................ 54

2.3 RESULTS ........................................................................................................ 54

2.3.1 Bioinformatics analysis of full length Mtx ........................................... 54
2.3.2 Design, cloning and purification of deletion constructs .................. 56
2.3.3 Competitive import inhibition assay .................................................. 58
2.3.4 Size exclusion chromatography .......................................................... 59
2.3.5 Circular dichroism and thermal shift assay ........................................... 60
2.3.6 Dynamic light scattering and analytical ultracentrifugation studies .... 62
2.3.7 Non reducing PAGE and native PAGE analysis .................................. 64
2.3.8 Limited proteolysis .................................................................................. 65
2.3.9 Isothermal titration calorimetry .............................................................. 66
2.3.10 Theoretical modelling ........................................................................... 67

2.4 DISCUSSION .................................................................................................... 68

2.5 FUTURE DIRECTIONS .................................................................................... 69

CHAPTER 3 .......................................................................................................... 73

LIGAND RECOGNITION BY THE TPR DOMAIN OF THE IMPORT FACTOR
TOC64 FROM ARABIDOPSIS THALIANA ......................................................... 73

3.1 ABSTRACT ....................................................................................................... 75

3.2 INTRODUCTION .............................................................................................. 75

3.3 MATERIALS AND METHODS ....................................................................... 78

3.3.1 Expression clone of the TPR domain of AtToc64 ............................... 78
3.3.2 Expression in E. coli ................................................................................. 79
3.3.3 Protein purification .................................................................................. 79
3.3.4 Relative molecular mass estimation by size exclusion chromatography.... 79
3.3.5 Peptide synthesis ...................................................................................... 80
3.3.6 Circular Dichroism studies (CD) ............................................................. 80
3.3.7 Dynamic Light Scattering studies (DLS) ............................................... 81
3.3.8 Analytical Ultra-centrifugation (AUC) .................................................... 81
3.3.9 Isothermal Titration Calorimetry (ITC) .................................................. 81
3.3.10 Molecular dynamics simulation studies ................................................. 82

3.4 RESULTS .......................................................................................................... 84

3.4.1 Purification and characterization of TPR domain of AtToc64 .......... 84
3.4.2 Estimation of polydispersity by DLS and AUC ........................................ 86
CHAPTER 4 .................................................................................................... 113

EXPLORING LIGAND RECOGNITION, SELECTIVITY AND DYNAMICS OF
TPR DOMAINS OF CHLOROPLAST TOC64 AND MITOCHONDRIA OM64
FROM ARABIDOPSIS THALIANA .............................................................. 113

4.1 ABSTRACT ............................................................................................. 115
4.2 INTRODUCTION ..................................................................................... 116
4.3 MATERIALS AND METHODS ................................................................. 120
  4.3.1 Starting structures ............................................................................ 120
  4.3.2 Preparation for simulation ............................................................... 121
  4.3.3 Analysis of trajectory .................................................................... 122
4.4 RESULTS AND DISCUSSION .............................................................. 123
  4.4.1 Comparative analysis of protein–peptide interaction ....................... 125
  4.4.2 An anchor using a carboxylate clamp ................................................. 126
  4.4.3 Key protein-peptide interactions ..................................................... 128
  4.4.4 A cavity for Valine ........................................................................... 130
  4.4.5 Intrapeptide interactions ................................................................. 130
  4.4.6 Binding energy and computational alanine scanning ....................... 132
  4.4.7 Dynamics in the protein-peptide interaction .................................... 134
  4.4.8 Dynamic cross-correlation map (DCCM) ........................................... 134
  4.4.9 Principal component analysis .......................................................... 135
LIST OF TABLES

Table 2-1 Primers used to design metaxin constructs .........................................................49
Table 3-1 Analytical ultracentrifugation ........................................................................87
Table 3-2 Thermodynamic parameters obtained for TPR-Hsp interaction using
isothermal titration calorimetry.........................................................................................89
Table 3-3 Computational alanine scanning for both ligand bound systems.................97
Table 4-1 Computational alanine scanning for the four ligand bound systems ..........132
LIST OF FIGURES

**Figure 1-1** Schematic representation of different eukaryotic targeting signal sequences for different organelles: peroxisomes, mitochondria and chloroplasts. .......................... 7

**Figure 1-2** Schematic representation of receptors from outer membrane of peroxisomes. ........................................................................................................................................... 11

**Figure 1-3** Cartoon representations of different peroxisomal receptors in apo or complex forms available to date. .............................................................................................. 13

**Figure 1-4** Cartoon representations of different peroxisomal receptors in apo or complex forms available to date. .............................................................................................. 16

**Figure 1-5** Schematic representation of receptors from outer membrane of mitochondria. ........................................................................................................................................... 21

**Figure 1-6** Cartoon representations of different outer mitochondrial membrane receptors in either apo or complex forms available to date. .................................................. 25

**Figure 1-7** Schematic representation of receptors from outer membrane of chloroplast. ........................................................................................................................................... 32

**Figure 1-8** Cartoon representations of different receptors on the chloroplast outer membrane known till date. ........................................................................................................... 34

**Figure 2-1** Insertion of β-barrel proteins on the outer membrane by SAM complex. .................................................. 46

**Figure 2-2** Secondary structure and disordered segment predictions of Metaxin. .................................................. 56

**Figure 2-3** Design of deletion constructs of Metaxin. ........................................................................................................ 58

**Figure 2-4** Competition import inhibition assay ........................................................................................................ 59

**Figure 2-5** An analysis of oligomeric nature of Metaxin deletion constructs using size exclusion chromatography ........................................................................................................ 60

**Figure 2-6** Circular dichroism and thermal melt studies of ΔMtx131. ........................................................................... 62

**Figure 2-7** Analysis of non-specific oligomerization in ΔMtx131 using DLS and AUC. ................................................................. 63

**Figure 2-8** Non reducing PAGE analysis of ΔMtx131 with increasing concentration of glutathione. ................................................................. 64
Figure 2-9 Effect of detergents on preventing aggregation of purified \(\Delta\text{Mtx131}\) with time.........................................................................................................................65
Figure 2-10 Limited proteolysis of \(\Delta\text{Mtx131}\) with trypsin............................................66
Figure 2-11 ITC isotherms obtained for interactions of freshly purified \(\Delta\text{Mtx131}\) with different synthetic peptides used as ligands...............................................................67
Figure 2-12 Five models for \(\Delta\text{Mtx131}\) obtained from I-TASSER....................................68
Figure 3-1 Topological model of Toc64 from \textit{Arabidopsis thaliana}............................78
Figure 3-2 Biophysical characterization of \(\text{AtToc64} \_\text{TPR-H6}\) using size exclusion chromatography and circular dichroism..............................................................85
Figure 3-3 Biophysical characterization of \(\text{AtToc64} \_\text{TPR-H6}\) using DLS and AUC ...86
Figure 3-4 Binding isotherms for interaction of \(\text{AtToc64} \_\text{TPR-H6}\) with Hsp70 and Hsp90. ........................................................................................................................................88
Figure 3-5 The ITC binding data for alanine scanning mutagenesis of the peptide interaction with \(\text{AtToc64} \_\text{TPR-H6}\). .................................................................................................90
Figure 3-6 Analysis of the molecular dynamics trajectory obtained after 50 ns of simulation...............................................................................................................................................92
Figure 3-7 Characterization of the average structures obtained after simulation of the complexes........................................................................................................................................................................93
Figure 3-8 Interactions occurring at the protein-peptide interface generated by Ligplot. ........................................................................................................................................................................94
Figure 3-9 The carboxylate clamp ................................................................................95
Figure 3-10 Key intrapeptide interactions........................................................................96
Figure 3-11 Principal component analyses of the three systems..................................99
Figure 4-1 Graphical abstract showing a model for the interactions between C-termini of chaperones Hsp70 and Hsp90 with the TPR domains of two homologues, Toc64 and Om64 from \textit{Arabidopsis thaliana}.........................................................116
Figure 4-2 Analysis of the molecular dynamics trajectories obtained after 50 ns of simulation in Toc64_TPR systems and Om64_TPR systems.................................................................................................................125
Figure 4-3 The carboxylate clamp ................................................................................127
Figure 4-4 Interactions occurring at the protein-peptide interface in \(\text{AtToc64} \_\text{TPR-Hsp} \) systems generated by Ligplot........................................................................................................................................129
Figure 4-5 Interactions occurring at the protein-peptide interface in \(\text{ArOm64} \_\text{TPR-Hsp} \) systems generated by Ligplot.................................................................................................................................................................................................130
Figure 4-6 Key intrapeptide interactions........................................................................131
Figure 4-7 Principal component analyses of AtToc64_TPR-Hsp systems. ............... 135

Figure 4-8 Principal component analyses of AtOm64_TPR-Hsp systems. ............. 136

Figure 4-9 A proposed model for interaction of Toc64 and Om64 with Hsp70 and Hsp90 in Arabidopsis based on computational studies. ................................. 140
Chapter 1

The design and structure of outer membrane receptors from peroxisomes, mitochondria and plastids.
1.1 INTRODUCTION

The origin of the eukaryotic cell likely represents the most important event in the evolution of life on earth, after the evolution of life itself. Over the years several proposals regarding the origin of organelles in eukaryotic cells have been put forward. While the nucleus defines the eukaryotic cell, it is now proposed that the endosymbiosis involving two prokaryotic cells, one of these cells leading to the formation of mitochondria, was the trigger to establish the eukaryotic cell. This was associated with reduction in the size of mitochondrial genome, transfer of its genetic material to the nucleus and hence the proteins involved in the metabolic pathways associated with the organelle were carried out in the cytosol (Timmis et al., 2004). Evidence supports the notion that mitochondrion evolved from an α-proteobacterial progenitor via endosymbiosis with a eukaryotic cell (Gray, 2012). The chloroplast, a specialized plastid, is also believed to have originated from an endosymbiotic event in which an ancestral photosynthetic cyanobacterium was engulfed by a heterotrophic host cell that already contained mitochondria (Cavalier-Smith, 2000, Martin et al., 2002). Both of these organelles contain their own DNA and are surrounded by a double membrane. Although similar theory of endosymbiotic origin was initially proposed for peroxisomes, recently experiment evidence such as targeting of some peroxisomal membrane proteins to endoplasmic reticulum (ER) prior to its targeting to its destined location (Schlüter et al., 2006) suggest an evolutionary link between ER and peroxisome (Gabaldón, 2010, Schlüter et al., 2006).

Both mitochondria and chloroplast only encode a small amount of the proteins required, while peroxisomes are required to import all their proteins. A prerequisite for the transport of the nuclear encoded cytosolically synthesized proteins to the respective organelles was the development of protein-import machinery that specifically import proteins destined for that organelle. This specificity is achieved by the combined action of targeting signals, protein import receptors on the surface of organelles, and may also require cytosolic chaperone factors. Notably despite the approximately 1.45 billion years (Martin and Mentel, 2010) of evolution since the event that gave rise to the mitochondrial endosymbiosis, and approximately 580 million years (Cavalier-Smith,
2002) since the event that gave rise to plastids, it appears that the solutions for targeting proteins to these organelles displays similarities over wide phylogenetic gaps.

1.1.1 General outline of protein import into organelles

Approximately 95% of proteins destined to mitochondria, plastids and peroxisomes are nuclear encoded and post-translationally directed to specific intracellular locations as nascent polypeptide chains. It is noteworthy that while protein import for these organelles can occur post-translationally (Neupert and Herrmann, 2007), as opposed the co-translation mode of targeting to the ER using the signal recognition particle, in the in vivo conditions, the kinetics of import may be co-translational (Neupert, 1997). Also for some proteins, such as fumarase in yeast, it appears that targeting is co-translational (Knox et al., 1998). Protein sorting and translocation requires the presence of specific targeting signals within the amino acid sequence of these polypeptides that enable their initial recognition by surface receptors located on the membrane of the destined organelles. Furthermore, the nascent polypeptide chains are prevented from aggregating through the association with chaperones in an ATP dependent manner. In the mitochondria, the cytosolic mitochondria import stimulation factor (MSF) (Hachiya et al., 1993) along with Hsp70 and Hsp90 chaperones (Young et al., 2003) help in maintaining the nascent preproteins in an import competent conformation. Many transit peptides contain phosphoserine or phosphothreonine residues, which act as binding sites for 14-3-3 protein. Hsp70 and 14-3-3 proteins associate with preproteins destined to the chloroplast to form the “guidance complex” which is needed for efficient import (May and Soll, 2000). Import of PTS1 containing proteins into peroxisomes requires cytosolic Hsp70 (homologue of E. coli DnaK) and Hsp40 (homologue of E. coli DnaJ) (Walton et al., 1994, Hettema et al., 1998). Protein unfolding is not essential for targeting to the peroxisomal matrix (Subramani, 1996b), and both monomeric and oligomeric folded proteins can be imported into the matrix. Hsp70 is thought to stabilize the partially unfolded region containing the PTS to make it accessible to the receptor (Subramani, 1996b). However peroxisomal membrane proteins are cytosolically synthesized and require chaperones to maintain their solubility (Sparkes and Baker, 2002).

In the last few decades, extensive research on organellar import machineries has provided a wealth of information on the various organellar components involved in the
process of protein import, that starts with the recognition of precursor proteins, through translocation across one or two membranes, removal of targeting signal and assembly into a functional protein complex. In this review, we provide an overview of the current state of knowledge on the outer membrane import receptors from the above organelles. For this, we first focus on the signal sequences required for recognition by specific outer membrane receptor proteins of these organelles. Then we will discuss the structural details and biophysical characterizations of the receptors of the organelles available to date. We will conclude highlighting the structural features of the outer membrane receptors yet to be solved.

1.1.2 Targeting signals

1.1.2.1 Peroxisomal targeting signals

Peroxisome targeting signals (PTS) target cytosolically synthesized peroxisomal matrix proteins into the organelle (Figure 1-1). There are two consensus motifs identified. PTS1, which is present at the C-terminus of the majority of matrix proteins, is considered to comprise only the C-terminal tripeptide with the consensus sequence (S/A/C)-(K/R/H)-(L/M), first identified in firefly luciferase (Gould et al., 1987). However recent studies suggest that this motif may vary with respect to length and sequence and may also differ with species (Michels et al., 2005, Neuberger et al., 2003). In addition up to the C-terminal 12 residues are presumed to affect binding affinity of the PTS1-protein to the peroxisomal receptor (Michels et al., 2005). PTS2, the second consensus motif, is comprised of a nonapeptide (R/K)-(L/V/I)-X5-(Q/H)-(L/A) located towards the N terminus of the protein. This differs between species (Petriv et al., 2004). Unlike the PTS1, the PTS2 are cleaved from the cargo protein after import (Tanaka et al., 2008). The protease involved in this process is DEG15 and LON2 in plants (Helm et al., 2007, Lingard and Bartel, 2009) and TYSND1 in mammals (Kurochkin et al., 2007). Removal of the targeting signal is not mandatory for import. Some peroxisomal matrix proteins contain internal targeting signals called PTS3, which remain poorly defined. These proteins use the PTS1 dependent targeting pathway for their import (Elgersma et al., 1995). Also co-import of a cargo proteins has also been observed for peroxisomes, where one protein with a functional PTS1/PTS2 can form a heteromeric complex with a protein lacking a consensus targeting motif, and thus aid in import of the latter into the organelle in plant, yeast and mammalian cells (McNew and Goodman, 1994).
Peroxisomal membrane targeting signals (mPTs) are of two types: mPTS1 and mPTS2 and have been defined in a number of peroxisomal membrane proteins (PMPs). In Pex3p from *Hansenula* (Baerends et al., 1996), *Pichia* (Wiener et al., 1996) and human (*Hs*) (Kammerer et al., 1998), which belongs to type I PMPs, an N terminal transmembrane domain preceded by 3-4 conserved basic residues, has been implicated in targeting (Subramani, 1996a). Mutational studies have shown that the positively charged pentapeptide (R<sub>11</sub>-X-K-K-K<sub>15</sub>) at the N-terminus is essential for correct localization and function of Pex3p in *Hansenula* (Baerends et al., 2000). In type II PMPs such as PMP47 from *Candida*, stop-transfer signal sequences in the transmembrane domain (TMD) ensures stable association to the peroxisome membrane (Sparkes and Baker, 2002). In the case of single spanning PMPs the TMD and a stretch of positively charged lumen oriented residues are necessary for targeting and/or membrane association. In contrast, for multi-spanning membrane proteins, the cooperation of multiple regions of the protein appear important. This added level of complexity in the latter causes difficulty in formulating general rules for targeting and membrane association. Several motifs have been reported and two common ones are (K/R)-(K/R)-X<sub>3,7</sub>-(T/S)-X<sub>2</sub>-(D/E) and Y-X<sub>3</sub>-L-X<sub>3</sub>-P-X<sub>3</sub>-(K/Q/N) (Titorenko and Rachubinski, 2001). mPTS2 follow an indirect route to the peroxisomal membrane through the endoplasmic reticulum (ER) membrane (Waterham et al., 1994) and have been described in plant and yeasts. An example of this is Pex15p from *S. cerevisiae*, which reveals O-linked glycosylation, suggesting passage through ER. Similar is the case with Pex2p and Pex16p from *Y. lipolytica* which have N-linked glycosylation. Further, the development of peroxisome is affected, if ER protein transport is blocked (Mullen et al., 1999). Two motifs have been reported for mPTS2: (R/K)-X-(K/R)-X-(K/R)-X-(L/I)-X<sub>9</sub>-10-(F/Y) and (L/I/V)-X-R-X-(K/R)-X-K-X-(L/I) and are present in the C-terminus (Titorenko and Rachubinski, 2001).
### Figure 1-1

Schematic representation of different eukaryotic targeting signal sequences for different organelles: peroxisomes, mitochondria and chloroplasts.

*The brown color represent hydrophobic regions and those in blue/brown represent positively charged/hydrophobic regions respectively. The grey region represents residues flanking the former. In the luminal targeting peptides: A, N, H and C stand for acidic domain, N-domain, H-domain and C-terminal region respectively.*

### 1.1.2.2 Mitochondrial targeting signals

Mitochondrial targeting signals (MTS) can be found at the N-terminus, C-terminus or internally in protein destined to this organelle (**Figure 1-1**). Although consensus sequences are not observed in MTS, they display distinct physicochemical properties. The mitochondrial precursors destined to matrix and inner membrane contain cleavable N-terminal presequence whereas many outer membrane proteins, polytopic inner membrane proteins and soluble IMS proteins contain internal targeting signals. The N-
terminal classical targeting sequences generally have positively charged residues, which have the propensity to form an amphipathic $\alpha$-helix to be recognized by the mitochondrial outer membrane receptors and are proteolytically processed by mitochondrial processing peptidases (Lithgow and Schneider, 2010). The NMR structure of the presequence from F1,β subunit of ATP-synthase from *N. plumbaginifolia* indicates the formation of helix in membrane mimicking environment (Moberg et al., 2004). Studies in yeast have shown that the positively charged surface of the amphipathic alpha helix interacts with Tom22 and the hydrophobic surface of the same is recognized by Tom20 (Yamano et al., 2008). Further the high resolution structures of Tom20 with presequence from rat aldehyde dehydrogenase also indicated formation of similar secondary structure (Abe et al., 2000). However the structure of yeast mitochondrial processing peptidase with presequence peptide from CoxIV demonstrated that the peptide adopted an extended conformation (Taylor et al., 2001). This suggests that the presequences adopt conformation depending on the receptor they encounter. The C-terminal targeting signals are similar to that of the N-terminal signals (Lee et al., 1999), but are much less frequently seen. About 30% of protein destined to mitochondria lack N-terminal targeting information and contain internal targeting signals which are positively charged segments known as “carrier sequence motifs” (Diekert et al., 1999). These carrier precursors form loops (Neupert and Herrmann, 2007) while traversing the TOM pore and use the translocon inner membrane 22 (Tim22) pathway for insertion into the inner membrane. These targeting elements, in combination with preceding hydrophobic stretches, help in membrane potential dependent integration (Neupert and Herrmann, 2007) into the bilayer of the inner membrane as in membrane-spanning metabolite transporters located in the inner membrane. Proteins destined to the inner membrane space (IMS) have bipartite targeting signals: an N-terminal MTS 15 - 40 residues rich in basic and hydroxylated amino acids, followed by a C-terminal IMS targeting signal. In this case the MTS is cleaved before translocation across the inner membrane. Some IMS destined proteins may contain matrix targeting signals followed by a transmembrane anchor which acts as a stop transfer signal and is eventually cleaved releasing the IMS protein (Stojanovski et al., 2003). The targeting signals of integral membrane proteins with single transmembrane domain, sometimes lie adjacent to or embedded in this domains, which also have one or more adjacent basic residues necessary for targeting (Wattenberg and Lithgow, 2001). Matrix targeting proteins follow the Tim23 pathway. Another type of
IMS targeting signal contains cysteine that can form disulphide bond with IMS receptor Mia40 and another hydrophobic residue in the signal is responsible for specific recognition by the receptor through hydrophobic interactions (Chacinska et al., 2009). The import of precursors for β-barrel proteins on the outer membrane involves the association of sorting and assembly machinery (SAM). These proteins contain β-sorting signals at the C-terminus recognized by the SAM complex and are quite different form alpha helical presequences (Kutik et al., 2008).

1.1.2.3 *Plastid targeting signals*

Proteins destined to the outer membrane of the chloroplast such as Toc75 and Toc159, have N-terminal cleavable chloroplast targeting signals also know as transit peptides (Li and Chen, 1996) (Figure 1-1), however some outer membrane chloroplastic proteins like Toc34 and Toc159 lack these signals (Hofmann and Theg, 2005). Chloroplast targeting signals are composed of predominantly hydroxylated, hydrophobic and positively charged residues and a very low abundance of acidic residues (Zhang and Glaser, 2002) and are cleaved by stromal processing peptidases. Similar to plant mitochondrial presequences, the transit peptides are serine rich. In contrast to mitochondrial presequences, arginine and leucine residues are rarely found at the N-terminus of transit peptides (Bhushan et al., 2006). Both transit peptides and mitochondrial targeting presequences have the potential to be amphipathic and have a helix-coil-helix organization in higher plants, but it is the hydroxylated residues in transit peptides and basic residues in presequences that determine the amphipathicity. Algal transit peptides are shorter than those in higher plants and appear to have one helical region, whereas in higher plants multiple helical regions can exist e.g transit peptide of ferredoxin from higher plants contain two discontinuous helical domains (Bruce, 2000). Transit peptides are generally unstructured in aqueous environment and form random coil, however they form helices upon contacting the membrane. Two plausible interactions of the transit peptides with the membrane lipids have been suggested: ionic interaction between basic residues of transit peptides and anionic phospholipids, and the hydrogen bonding interactions between hydroxyl groups of transit peptides and the galactose head groups of glycolipids (Pilon et al., 1995, Bruce, 2000). This is also associated with hydrogen bond formation between hydroxylated residues of the transit peptides and tightly bound water molecules on the membrane.
interface, which favours the non-bilayer forming tendency monogalactosyldiacylglycerol (MGDG) on the chloroplast membrane, which forms a H$_{11}$ phase conformation necessary for targeting and translocation (Rietveld et al., 1995). Micelle disruption experiments suggest that transit peptides have special affinity to MGDG and DGDG, which are unusual class of lipid specific to the eukaryotic chloroplast membrane that have very high lipid to protein ratio (3:1) (Patron and Waller, 2007). Transit peptides targeting to the stroma contain the stroma targeting domain (STD) which are known to be comprised of three distinct regions: an uncharged N-terminal domain of ~10 residues beginning with MA and terminating with G/P, a central domain rich in S/T and lacking acidic residues and a C-terminal domain rich in arginines, thus can form an amphiphilic $\beta$-strand (Claros et al., 1997). It is hypothesized that interaction of transit peptides with the translocons on the chloroplastic envelopes might be mediated by multiple mechanisms as physiochemical properties (pH, membrane environment) and recognition of specific sequences of the transit peptides (Chotewutmontri et al., 2012). Proteins targeted to the thylakoid have bipartite targeting signals containing information for import across the chloroplast envelope membranes and targeting to the thylakoid. There is a N-terminal STD and a C-terminal LTD (thylakoid luminal targeting domain) (Cline and Henry, 1996). The LTD exhibit hydrophobicity compared to STD and are divided into four regions: extended acidic domain with 12-15 residues (A-domain), positively charged short N-terminal (N-domain), 12-18 residue hydrophobic core (H-region) and polar C-terminal (C-region) (Schnell, 1998). Conserved A-X-A-X motif found in the -3 and -1 position of the targeting signal defines the cleavage site for thylakoid luminal processing peptidase.

1.2 STRUCTURAL OVERVIEW OF OUTER MEMBRANE IMPORT RECEPTORS

1.2.1 Outer membrane receptors of peroxisomes

Peroxisomes are highly dynamic in morphology and metabolism. There are 12 peoxisomal proteins (peroxins) that make the translocon required for the peroxisomal matrix protein import which are translated on free polyribosomes. Import involves four steps: binding of receptor to cargo, docking of receptor-cargo complex on the
peroxisomal membrane, membrane translocation of complex followed by release of cargo and receptor recycling (Figure 1-2) (Brown and Baker, 2008).

![Diagram of peroxisomal receptors](image)

<table>
<thead>
<tr>
<th>Organism</th>
<th>Protein</th>
<th>Domain</th>
<th>Ligand</th>
<th>Structure determination</th>
<th>PDB</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trypanosome Human</td>
<td>Pex5</td>
<td>C-terminal (TPR domain)</td>
<td>pPTS1</td>
<td>Crystallography</td>
<td>1HXI</td>
<td>Kumar et al., 2001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-Do-</td>
<td></td>
<td>2J9Q</td>
<td>Stanley et al., 2007</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-Do-</td>
<td></td>
<td>2COM</td>
<td>Stanley et al., 2006</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-Do-</td>
<td></td>
<td>2C0L</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-Do-</td>
<td></td>
<td>1FCQ</td>
<td>Gatto et al., 2000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-Do-</td>
<td></td>
<td>3C0V</td>
<td>Sampathkumar et al., 2008</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-Do-</td>
<td></td>
<td>3C0V</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-Do-</td>
<td></td>
<td>3C0V</td>
<td></td>
</tr>
<tr>
<td>Yeast</td>
<td>Pex7p</td>
<td>Cytosolic</td>
<td>Pex21pC</td>
<td>Crystallography</td>
<td>3W15</td>
<td>Pan et al., 2013</td>
</tr>
<tr>
<td>Yeast</td>
<td>Pex13</td>
<td>Cytosolic</td>
<td>Pex14</td>
<td>Crystallography</td>
<td>2VR1</td>
<td>Kursula et al. *</td>
</tr>
<tr>
<td>Rattus</td>
<td>Pex14</td>
<td>Cytosolic N-terminal</td>
<td>-</td>
<td>Crystallography</td>
<td>3FF5</td>
<td>Su et al., 2009</td>
</tr>
<tr>
<td>Human</td>
<td>Pex5</td>
<td>Cytosolic</td>
<td>Pex14</td>
<td>SAXS</td>
<td></td>
<td>Shiozawa et al., 2009</td>
</tr>
<tr>
<td>Human</td>
<td>Pex19</td>
<td>Cytosolic C-terminal</td>
<td>-</td>
<td>Crystallography</td>
<td>2W8L</td>
<td>Schueller et al., 2010</td>
</tr>
<tr>
<td>Human</td>
<td>Pex14</td>
<td>Cytosolic</td>
<td>Pex5</td>
<td>NMR</td>
<td>2W84</td>
<td>Neufeld et al., 2009</td>
</tr>
<tr>
<td>Human</td>
<td>Pex14</td>
<td>Cytosolic</td>
<td>Pex19</td>
<td>NMR</td>
<td>2W85</td>
<td>Neufeld et al., 2009</td>
</tr>
<tr>
<td>Human</td>
<td>Pex3p</td>
<td>Cytosolic</td>
<td>Pex19p</td>
<td>Crystallography</td>
<td>3MK4</td>
<td>Schmidt et al., 2010</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3AJB</td>
<td>Sato et al., 2010</td>
</tr>
<tr>
<td>Yeast</td>
<td>Pex4p</td>
<td>Cytosolic N-terminal</td>
<td>Pex22pC</td>
<td>Crystallography</td>
<td>2Y9M</td>
<td>Williams et al., 2011</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4BWF</td>
<td>Williams et al., 2013</td>
</tr>
<tr>
<td>Mouse</td>
<td>Pex1</td>
<td>A-domain</td>
<td>-</td>
<td>Crystallography</td>
<td>1WLF</td>
<td>Shiozawa et al., 2004</td>
</tr>
</tbody>
</table>

Figure 1-2 Schematic representation of receptors from outer membrane of peroxisomes. The bottom panel shows a list of receptors whose structures have been studied. * stands for unpublished data.

### 1.2.1.1 Pex5 and its complexes

Proteins that contain a PTS1 or PTS2 target sequence are recognized by soluble peroxisomal shuttling receptors, Pex5 and Pex7 respectively. While the former is found
in all organisms, the latter is absent in *C. elegans*, which apparently does not utilize this targeting pathway (Michels et al., 2005, Motley et al., 2000). After recognizing the cargo, these shuttling receptors interact with the peroxisomal translocation machinery. The amino acid sequence of Pex5 is highly conserved with an N-terminal peroxisomal docking domain (also involved in recycling the receptors after import) and a C-terminal one with two sets of 3 tetratricopeptide repeats (TPR) essential for to interaction with PTS1 (Lanyon-Hogg et al., 2010). There are two isoforms of the protein: Pex5S and Pex5L (short and long form) found in mammals, of which Pex5L participates in PTS2 import pathway due to an internal 37 residue insertion (Otera et al., 2000). The crystal structure of the C-terminal domain of *Hs*Pex5 with a PTS1 model pentapeptide (YQSKL) (PDB: 1FCH) (Gatto et al., 2000) revealed that this region is made up of two TPR motif triplets (1-3 and 5-7) hinged by a part of TPR motif 4, thus demonstrating a ring like arrangement with a groove in the centre to accommodate the peptide (Figure 1-3A). The signal peptide is recognized by the asparagine residues (Asn\(^{378}\), Asn\(^{489}\), Asn\(^{497}\), Asn\(^{524}\) and Asn\(^{531}\)). The peptide residue, leucine is bound in a hydrophobic pocket, lysine contacts a water molecule which interacts with Glu\(^{379}\) and Glu\(^{348}\) and serine forms hydrogen bonds with two water molecules, one of which also hydrogen bonds to the peptide tyrosine side chain. An Asn\(^{489}\)Lys mutation causes loss of PTS1 import, but does not affect PTS2 import. Further, the crystal structure of *Hs*Pex5C_TPR domain (residues: 315-639) in complex with sterol carrier protein 2 containing the PTS1 sequence as AKL (PDB: 2C0L) (Stanley et al., 2006) suggests that interactions of the C-terminal AKL motif of the ligand with four conserved asparagines (Asn\(^{415}\), Asn\(^{526}\), Asn\(^{534}\) and Asn\(^{561}\)). The two TPR triplet arches are connected by the 7C loop, which connects the seventh TPR and the C-terminal tail (residues: 589-601) and the distorted TPR4 segment to form pseudo circular structure (Figure 1-3B). Comparison to the structure without bound peptide revealed a transition from an open, snail like conformation to a closed, circular conformation in the bound form. The ring opening causes the release of the target, which is caused due to the loop C-terminal to the TPR (called the 7C loop). Conformational changes within the sterol carrier protein cause disassembly of the C-terminal PTS1. The alanine residue of the PTS prevents hydrogen bonding network formation and the preceding residues make van der Waals contacts with the receptor. Further the crystal structures of Pex5 with different PTS1 peptides, namely SKL, SHL and AKL (PDB: 3CV0, 3CVL, 3CVN, 3CVP and 3CVQ) (Sampathkumar et al., 2008), show structural conservation of residues interacting with
the PTS1 backbone whereas those residues not involved in ligand interaction lack conservation. The structure of AtPex5 has been modelled using the HsPex5 structure as a template reveal conservation of interaction with PTS1 however targeting studies indicate subtle species specific differences in the geometry of the binding pocket (Lanyon-Hogg et al., 2010).

**Figure 1-3** Cartoon representations of different peroxisomal receptors in apo or complex forms available to date.

**A.** Crystal structure of HsPex5 bound to PTS1 containing peptide (PDB: 1FCH). The peptide is shown as green sticks. TPRs 1-3 of HsPex5 are shown in teal and TPRs 5-7 are shown in magenta. Important asparagine residues of the protein that interact with the peptide are shown in sticks. Other secondary structural elements of the protein are shown in brown. **B.** Crystal structure of HsPex5 in complex with sterol carrier protein 2 containing PTS1 sequence (AKL) (PDB: 2C0L). The TPRs 1-3 are shown in teal and TPRs 5-7 are shown in magenta. The 7C loop is shown in red. Sterol carrier protein is shown in green and the PTS1 sequence (AKL) is
shown in stick model. Asparagine residues from the TPR interacting with the PTS1 sequence are shown as stick. C. Crystal structure of PTS2 from Fox3p bound to receptor complex of ScPex7 and C-terminal domain of ScPex21 (PDB: 3W15). Pex21 and PTS2 are shown in magenta and cyan respectively. The special N-terminal region which exhibits low similarity with other WD40 motifs (blue) containing the bulge loop (red) and WD 40 motifs (green) of Pex7 are shown. The residues forming the acidic patch are shown in stick model. D. Crystal structure of N-terminal domain of Pex14 from rat (PDB: 3FF5). All the helices are labelled and two putative Pex5p binding sites are shown surrounded by positively charged residues (blue) and the phenylalanine residues (magenta). Site1 is surrounded by F35 and R40, whereas site2 is surrounded by K34, F52,K55 and K56.

1.2.1.2 Pex7 and its complexes

Arabidopsis thaliana contains sixty PTS2-targeted proteins whereas none are found in C. elegans; in S. cerevisiae, this pathway is maintained solely for the import of 3-ketoacyl thiolase. The crystal structure of two PTS2 containing aldolases from L. Mexicana and T. brucei (Chudzik et al., 2000) reveal how PTS2 monomers interact with each other to form a PTS2 dimer. Evidence suggests that PTS2 pathway in plants can aid in import of hetero multimeric complexes (Flynn et al., 1998). Pex7 does not interact with the membrane and requires accessory proteins such as Pex18p and Pex21p in S. cerevisiae (Purdue et al., 1998); Pex20p in Hansenula (Otzen et al., 2005); Pex5 in Arabidopsis (Nito et al., 2002) and Pex5L isoforms in mammals (Einwachter et al., 2001). Sequence alignment of the co-receptors indicate a common di-aromatic W-X₃-F/Y docking motif, found as a single or in multiple occurrences, denoting evolutionary conservation of this functional motif (Einwachter et al., 2001). Pex7 is predicted to belong to the β–transducin-related (WD-40) protein family which contain a distinct N-terminus and a seven-bladed β-propeller with six WD repeats, the core of which begin with Gly-His and end with Trp-Asp. The ends are separated by twenty seven residues mostly hydrophobic, which are conserved and cause the core to fold into three β-sheets separated by turns (Zhang and Lazarow, 1996). The heterotrimeric crystal structure of PTS2 bound to the Pex7p-Pex21p receptor complex from yeast (PDB: 3W15) (Pan et al., 2013) revealed their spatial organization as a complex. PEX7p complexes with the C-terminal of Pex21p (residues 190-288) also termed as Pex21pC, with the help of its hydrophobic bulge loop and the acidic patch at the side of the loop; both are located on the top surface of Pex7p (Figure 1-3C) and the interface residues are conserved in their
homologs. The PTS2 of the N-terminus of Fox3p also termed as Fox3pN, forms an amphipathic $\alpha$-helix that binds into a large cleft with two minor grooves for accepting the side chains of the helix. Binary complexes of Pex7p-Fox3pN, Pex7p-Pex21pC or Pex21pC-Fox3pN were unstable due to the exposure of hydrophobic surfaces. Electrostatic interactions of Pex7p with Pex21pC or Fox3pN are observed on either side of the hydrophobic core. Thus this along with hydrophobic effect stabilizes the complex.

1.2.1.3 **Pex13, Pex14 and their complexes**

The cargo-receptor complex formed using the PTS1/PTS2 pathway is docked at the peroxisomal membrane to PMPs namely Pex13p, Pex14p (Pex17p in yeast) representing a pathway convergence (Rayapuram and Subramani, 2006). Pex14p was the first of these receptor proteins shown to interact with both Pex5p and Pex7p. The difference between the fungal versus plant and mammalian systems is that in the fungal system the receptors in either of the pathways dock independently, contrary to the latter case where Pex5L binds to Pex7 and this complex then interacts Pex14 thus converging the two pathways (Braverman et al., 1998). Cargo free Pex5p has higher affinity for Pex13 than Pex14 and disassembles the Pex14 homooligomer (Itoh and Fujiki, 2006). Pex13 is a PMP with the N-and C-termini exposed to the cytosol; pull down assays using GST (glutathione transferase) show that 100 residues at the N-terminus act as a binding domain for Pex5 and Pex7 in mammals (Otera et al., 2002) and *Arabidopsis* (Mano et al., 2006) respectively and the C-terminal SH3 domain in yeast (Bottger et al., 2000) exhibits similar binding. The crystal structure of SH3 domain of *Sc*Pex13 complexed with a peptide from Pex14 has been reported (PDB: 2V1R), although the details are yet to be published.

Pex14 is transmembrane protein known to form oligomers. The N-terminus is chemically/proteolytically protected and is known to interact with the di-aromatic docking motif of Pex5 (Otera et al., 2002). Yeast Pex14 has two Pex5p binding sites, one at each terminus: 50 residues at the N-terminal and 58 residues at C-terminus which overlaps with the Pex7p binding region (Azevedo and Schliebs, 2006). Pex14 can form homooligomer with the help of a conserved coiled-coil motif (residue 140-278 in mammalian Pex14p) and binds to Pex5p with a stoichiometry of 5-6:1 of Pex14:Pex5p (Gouveia et al., 2000, Shiozawa et al., 2009). The crystal structure of the N-terminal conserved domain (residues 25-70) of mammalian Pex14, reveals a three $\alpha$-helix bundle
with a conserved hydrophobic surface containing two pockets from the side chains of Phe$^{35}$ and Phe$^{52}$ and some positively charged residues (PDB: 3FF5) (Su et al., 2009). The $\alpha_1$ and $\alpha_2$ helices are joined by a short $3_{10}$ helix and are arranged in an antiparallel orientation and the third $\alpha$ helix links the above two antiparallel helices (Figure 1-3D). Alanine mutations of either of the two phenylalanines caused loss of interaction between Pex5p and Pex14. The WXXXF/Y motif of Pex5p, which is speculated to adopt a helical conformation on interaction with its partners, is thought to be recognized by these pockets. A small angle X-ray scattering study of the HsPex5-Pex14-PTS1 multimeric complex showed that the complex retained an elongated shape as Pex5p with minor conformational changes and the unfolded N-terminus could interact with a wide range of proteins (Shiozawa et al., 2009). Furthermore, single particle electron microscopy (Moscicka et al., 2007) of Hansenula Pex5p revealed a tetramer at pH7 and a conformational transition from a closed to an open form upon interaction with Pex20p.

![Figure 1-4 Cartoon representations of different peroxisomal receptors in apo or complex forms available to date.](image-url)
A. Crystal structure of the C-terminal domain of HsPex19p (PDB: 2WL8). The helices are labelled and the side chains of polar residues (blue) and hydrophobic residues (magenta) that are conserved are shown. The exposure of hydrophobic surface residues shown is unique and suggests a plausible functional role. B. Crystal structure of the complex of HsPex3p and HsPex19p peptide (PDB: 3AJB). The residue W104 divides the cavity into two halves, one half interacts provides complementary surface for leucine triad (L18, L21 and L22) from the Pex19p peptide and the other half for the side chain of F29. The residues of the protein (slate blue) forming the cavity and that of the peptide (cyan) are shown in stick model. C. Structure of N-terminal domain of peptide derived from HsPex14p in complex with HsPex5p using NMR (PDB: 2W84). The residues W118 and F122 of Pex5p shown in magenta interact with residues from the Pex14p represented as green sticks. D. Crystal structure of UBC domain of ScPex4p in complex with soluble domain of ScPex22p (PDB: 2Y9M). Residues of Pex22p (wheat color) interacting of Y172 of Pex4p (blue–white color) are shown in stick model.

1.2.1.4 Pex19 and its complexes

Pex19 is a cytosolic peroxin, which serves as a receptor and chaperone for newly synthesized PMPs and recognizes complex mPTS (Sacksteder et al., 2000). The crystal structure of the folded C-terminus of Pex19p containing residues 161-273 (PDB: 2WL8) (Schueller et al., 2010) showed that this region was made up of a three helix bundle ($\alpha_2$-4) with a exposed N-terminal $\alpha_1$ helix. There were four protomers in the asymmetric unit, however no evidence of oligomerization was observed in solution. The hydrophobic face of the $\alpha_1$ helix is oriented towards the center of the domain (Figure 1-4A). The N-terminal region is flexible but contributes to PMP recognition and is sufficient for binding to Pex3p. The C-terminal CAAX motif (A denotes aliphatic residue and X denotes any residue) is the farnesylation site and is important for PMP interaction (Ructäschel et al., 2009) except in yeast. Pex19p has two binding site for Pex3p, the N-terminal site has stronger affinity and the other near the PMP binding site which has weaker affinity. The crystal structures of human Pex3p complexed with N-terminal 44 residues of Pex19p (PDB: 3MK4; 3AJB) reveals that highly specific interaction due to their complementary hydrophobic and solvent inaccessible surfaces (Schmidt et al., 2010; Sato et al., 2010). The Pex3p has a central helix, which runs along the major axis and is surrounded by five other helices. The interaction between the central helix and the surrounding ones are mainly hydrophobic and these residues are conserved. The Pex19p peptide forms an $\alpha$-helix and sits obliquely to the central helix.
and the interaction between the former and Pex3p is mainly van der Waals and/or hydrophobic interactions (Figure 1-4B).

The NMR solution structure of the N-terminal end of Pex14 (residues 16-60) with Pex5 (residues 108-127) and with Pex19 (residues 66-77) (PDB: 2W84 and 2W85 respectively) (Neufeld et al., 2009) indicates that both the ligands adopt an amphipathic α-helical conformation (Figure 1-4C) and they bind to the same site formed by the α1 and α2 helices of Pex14. The binding site is flanked by highly positively surface residues (Arg25, Lys34, Arg40, Lys55 and Lys56) and that of the ligands contain negatively charged patches formed by Asp111, Glu116 and Glu121 in Pex5 and Glu73 in Pex19. Thus charge complementarity is important for the above interaction. However both ligands bind in slightly different orientations. The interactions involve the W-X3-F/Y motif of Pex5 and the F/YF-X3-F motif of Pex19.

1.2.1.5 Pex proteins involved in receptor recycling

Following import of the cargo, the receptors are recycled back to the cytosol. Export of Pex5p to the cytosol require the action of three RING finger peroxins, Pex2p, Pex10p and Pex12p and the RING domain of yeast Pex10p acts as E3 ligase for ubiquitination of Pex5p (Williams et al., 2008). Additionally the ubiquitin-conjugating (E2) enzyme, Pex4p, its membrane anchor, Pex22p (a membrane protein) and ATPases, Pex1p and Pex6p, which are tethered to the membrane by Pex15p in yeast or Pex26 in mammals, are also involved (Kiel et al., 2005). Hence the recycling is an ATP-driven process. Mammalian Pex5p, lacking the first 17 or 110 residues, display import properties however the recycling activity is attenuated. These observations indicate the involvement of the N-terminus in recycling (Costa-Rodrigues et al., 2004). Mammals lack a Pex4-Pex22 complex and the equivalent function is carried out by the cytosolic E2D1/2/3 family, the counterpart of yeast Pex4p (Grou et al., 2008). Pex2p is an integral membrane protein, where the N- and C-termini are exposed to the cytosol and is apparently dimeric as determined by expressing epitope-tagged rat protein in COS-7 cell lines (Harano et al., 1999). The crystal structure of the Pex415-183 and Pex2254-180 complex (PDB: 2Y9M) (Williams et al., 2011) reveals that Pex415-183 comprise of a core domain consisting four antiparallel β-strands (β1-β4), one helix (α2), a 3_{10} helix, N-terminal helix (α1) and a C-terminal helix-turn-helix (α3-α4). Thus Pex4 adopts an ubiquitin conjugating (UBC) fold. Pex2254-180 formed a novel fold as it constituted five
parallel β-strands (β3-β2-β1-β4-β5) tightly sandwiched on either sides by eight helices. The interaction had high binding affinity and the interfacial residues were conserved in Pex4p sequences and the mutation of its Tyr\textsuperscript{172} to alanine abolished the binding (Figure 1-4D). Pex4p required the binding with Pex22p to modify Pex5p and the association with peroxisomal membrane. Further the Cys\textsuperscript{105} was a variation on the highly conserved His-Pro-Asn motif found in E2 enzymes and an unusual disulfide bond was detected between Cys\textsuperscript{105} and Cys\textsuperscript{146} of Pex4p in the crystal structure (PDB: 4BWF) (Williams et al., 2013). This disulfide bond helps in fixing the α2-3 loop to β4-310 helix loop thus allowing the active site in an open conformation, and disruption of the bond caused increase in flexibility of the former loop and the active site. Further the access of the substrate/ubiquitin to the active site cysteine is restricted due to the above disruption thus affecting the ability of the enzyme to transfer ubiquitin to the substrate.

1.2.1.6 Pex15, Pex6 and Pex1

Pex15p is an integral membrane protein with the N-terminus facing the cytosol. The binding of Pex15p with Pex6p is dependent on the AAA cassettes (ATPases associated with a wide range of cellular activities) of the latter which contain the Walker A and B motif for ATP binding and hydrolysis respectively (Birschmann et al., 2003). There are two AAA cassettes in Pex6p namely D1 (residues 421-716) and D2 (residues 704-1030), which have opposing roles where ATP binding with D1 favors Pex6p-Pex15p interaction and ATP hydrolysis by D2 favors the disruption of the complex. It is hypothesized that Pex1p (having AAA cassette) may be involved in this process \textit{in vivo}. Pex1p and Pex6p oligomerization also involves the AAA cassettes (Birschmann et al., 2005). In humans, Pex1p interaction with Pex6p requires the binding of ATP to both the D1 and D2 domains suggesting a conformational change associated with this interaction, however the same interaction in yeast involves no ATP hydrolysis in the D2 domain of Pex1p (Fujiki et al., 2012). The crystal structure of the N-terminal AAA-ATPase domain (residues 3-180) of mouse Pex1p (PDB: 1WLF) (Shiozawa et al., 2004) suggests the presence of two globular subdomains: an N-terminal double-ψ-barrel fold, where two symmetrically interleaved ββαβ-elements giving rise to a pseudo-knotted structure resembling the letter “ψ”; and a C-terminal β-barrel. The two domains are connected by a 4-5 residues linker and a shallow groove between the two lobes which can form a putative substrate binding site. The C-terminal tail-anchored Pex26p
localizes the Pex1p-Pex6p complexes to the peroxisome, where the latter undergo conformational changes upon ATP binding (Nashiro et al., 2011). A homologue of mammalian Pex26, identified in *Arabidopsis*, is APEM9 (Goto et al., 2011), however the structural details are not available.

### 1.2.2 Outer membrane receptors of mitochondria

The Translocon of the outer membrane of mitochondria (TOM) recognizes and translocates preproteins synthesized on the cytosolic ribosomes into or across the mitochondrial outer membrane and forms a 450 kDa multi-subunit complex in yeast (Becker et al., 2008). It is comprised of seven subunits: Tom70 (Tom71), Tom40, Tom22, Tom20, Tom7, Tom6, Tom5 (*Figure 1-5*) (Pfanner and Geissler, 2001). Tom70 acts as a preprotein receptor and co-chaperone that cooperates with Hsp70-Hsp90 aiding in import. While Hsp90 functions in Tom70 dependent import in mammals, Hsp70 aids import in both yeast and mammals (Young et al., 2003). No equivalent homolog to Tom70 has been found in *Arabidopsis*. 
1.2.2.1 Tom70

Yeast Tom70 has an N-terminal transmembrane domain anchored to the outer membrane and a large cytosolic exposed region. The 3.0 Å resolution crystal structure of the latter from yeast Tom70 (PDB: 2GW1) encompassing residues 39 – 617 suggests that the recombinant protein is a crystallographic dimer with two domains comprising a total of eleven TPR motifs (with a total of 26 alpha helices) in each monomer (Wu and Sha, 2006). The dimerization interface is formed by helices 6, 7, 25 and 26 and is stabilized by hydrophobic and van der Waals interactions. The first three TPR motifs, present in the N-terminal domain, are thought to be involved in recognition of Hsp70 by interacting with the C-terminal EEVD motif of the latter (Figure 1-6A). The C-terminal domain, which has a largest conserved region, contains the presequence or mitochondrial targeting signal binding pocket. The flexibility in the C-terminal domain
Chapter 1

is thought to regulate the size of the pocket, which has a dimension of $15 \times 20 \times 20 \, \text{Å}$, which could accommodate an alpha helix. The electron density map for residues 223 - 249 is very poorly defined indicating disorder in this region of the structure. Tom71 has been identified in yeast, as a parologue of Tom70 (53% sequence identity and 70% similarity) and is loosely associated with the Tom complex. This protein plays a role in Mfb1 mitochondrial localization, which is responsible for regulation and connectivity of mitochondrial tubules in yeast (Kondo-Okamoto et al., 2008). The crystal structure of yeast Tom71 (PDB: 3FP3) encompassing residues 1007 – 639, displays that the cytosolic domain is comprises 28 $\alpha$-helices and reveals a difference in the arrangement of the N- and C- terminal domains to that of Tom70 (Li et al., 2009). The N-terminus is swung away from the C-terminus to form an “open” state in contrast to the yeast Tom70 where the two domains are packed to form an elongated molecule to form in a “closed” state (Figure 1-6B). The presequence or mitochondrial targeting signal binding pocket is exposed in the open state with an estimated dimension of $\sim 25 \times 35 \times 20 \, \text{Å}$ and can accommodate peptides with secondary structure. The residues of the presequence binding pocket are conserved between the Tom70/71 proteins indicating that they share similar binding specificity and might recognize internal targeting signals in preproteins (Brix et al., 2000). The closed state prevents preprotein binding. Further the crystal structures of the protein complexed with the C-terminal fragment of Hsp70 (Ssa1) and Hsp90 (PDB: 3FP4 and 3FP2 respectively) suggest that the peptides ligands attain similar conformations and bind to a basic groove located in the N-terminal domain. This interaction locks the protein in an open state thereby increasing the volume of the preprotein binding pocket at the C-terminus and thus facilitating preprotein loading. Tom70, Tom71 and Tom22 are notably absent in plants (Carrie et al., 2010).

However solution small angle X-ray scattering (SAXS) studies (Mills et al., 2009) conducted on the cytosolic domain of ScTom70 (residues 51 - 617) have shown that the protein existed as an elongated monomer contrary to the dimer observed in the crystal structure. The “closed form” observed in the crystal structure was due to the mis-assignment of the chains A and B of the two domains in the crystallographic dimer. Further there was no change in the solution structure upon binding with chaperone. It was observed that binding of the chaperone peptide did not alter the protein’s affinity for the presequence peptide. Modelling studies suggested that chaperone binding was incompatible with dimer formation. Similarly in the case of Tom 71, crystallographic
data for the C-terminus of the Hsp70 bound form (PDB: 3LCA), suggested significant flexibility because of the flexibility in the N- and C-terminal domains (Li et al., 2010). This accounts for the interaction of the protein with a broad range of precursor substrates. It was also observed that the buried surface area between the N- and C-terminal domains was 1464 Å², which was almost 5.7% of the total surface area. Hence it is proposed that the small interaction surface might provide flexibility between the domains.

1.2.2.2 Tom40

Tom40 is an integral membrane protein and forms the key structural component of the protein-conducting channel in the outer mitochondrial membrane. Electron microscopy (EM) studies on Tom40 from Neurospora crassa (Nc) suggests that the pore size of the protein is 2 - 3 nm (Ahting et al., 2001). The NcTom40 forms a two pore structure in the Tom complex, however EM studies suggests that both one and two pore forms are found in isolated Tom40. It has been observed that Tom 6, 7 and 22 are important for stability of the Tom complex. Circular dichroism (CD) and Fourier transform infrared spectroscopy (FTIR) studies suggest that the secondary structure comprises ~31% β-sheet, 22% α-helix and 47% coiled conformations. Tom40 has affinity to non-native polypeptide chains and thus preventing them from aggregation prior to import. Electrophysiology experiments show that Tom40 exhibits a voltage–dependent gating mechanism and removal of protein environment comprising of Tom 6, 7 and 22, weakens the channel activity. Sequence analysis of the voltage dependent ion channel-1 (VDAC1) from human and members of the Tom40 family exhibit of 15-25% identity/40-60% similarity indicating an evolutionary relationship (Zeth, 2010). Though a reliable homology model of Tom40 from human could be built where presence of 19 strands have been predicted, residues 1 – 77 could not be modeled as they do not align to the VDAC structure and are predicted to form random coils. Contrary to VDAC, the presence of a large number of negatively charged residues lining the Tom40 pore and its interface to the IMS is speculated to be due to its affinity towards positively charged presequence peptides. Antibody labeling studies suggests that the N- and C-terminus is towards the IMS. Homology modeling of NcTom40 was performed using the 2.3 Å resolution crystal structure of mouse VDAC1 (PDB: 3EMN) (Ujwal et al., 2008) suggested the presence of 19 β-strands with one putative N-terminal α-helix located
inside the pore, where a single residue in the helix (R49) appears to be conserved and aligns with K12 in hsVDAC-1 (Gessmann et al., 2011). The conserved residues E45 and R49 have been predicted to play a role in channel gating. Conserved residues R213, S222, S234, D248, T270 and D287 are arranged in a line crossing through the pore from the cytosolic to the IMS face, and predicted to be helix anchor regions inside the Tom40 pore. All of these residues except D287 are in contact with the α-helix and the latter is predicted to move within the pore during gating events and also stabilizes the β-barrel structure. Opposite to the helix anchor region is the polar slide, which is similar to the “greasy slide” made up of aromatic residues in bacterial maltoporin (Meyer, 1990) and is shown to be important for substrate binding. The polar slides are thought to be involved in recognition and translocation of precursors by providing a surface thermodynamically favoring the above process. Purified and refolded NcTom40 adopted its β-barrel topology as analyzed by CD studies. Further limited proteolysis followed by MALDI-TOF and ESI-MS/MS analysis suggested that 8 out of 12 sites were located in the loop region, but the sites located in the core were not accessible to the proteases. The assignment of the putative β-strands 1 and 6 was found to be incorrect because of the proteolysis at residues A67, K132 and R66.
Figure 1-6 Cartoon representations of different outer mitochondrial membrane receptors in either apo or complex forms available to date.

A. Crystal structure of Tom70 from yeast (2GW1). The N-terminal domain (yellow) is predicted to bind to C-terminus of Hsp70 and the C-terminal domain is shown in green. The structure shows a closed conformation. B. Crystal structure of yeast Tom71 (PDB: 3FP3). The N-terminal TPR domain (yellow) encompasses the Hsp binding pocket and the C-terminal region (cyan) comprises the preprotein binding pocket. C. Structure of Tom20 in complex with the presequence of aldehyde dehydrogenase presequence from rat obtained by NMR (PDB: 1OM2). The residues (cyan) of Tom20 forming the hydrophobic groove for interaction of the residues of the presequence (pink) are shown. D. Crystal structure of rat ALDH presequence tethered to Tom20 (PDB: 2VIT). The residues forming the hydrophobic groove are shown in cyan, peptide residues interacting with this groove are shown in pink and the cysteine residues used for tethering are shown in the figure.
1.2.2.3 **Tom20**

Tom20 from *Arabidopsis* is C-terminally anchored to the membrane by a single transmembrane segment, whereas a reverse orientation is seen in fungal and mammalian counterparts. The functional domain of rat Tom20 is composed of two acidic regions, a Glu-rich region and a single tetratricopeptide repeat (TPR). TPR domains are composed of a canonical arrangement of antiparallel $\alpha$-helices connected by loops and are mediators of protein-protein interaction. The TPR motif in Tom20 forms a part of the binding site but does not solely define the recognition. The region encompassing residues 51 – 145, was found to be protease resistant and was termed as $\Delta50$Tom20.

The NMR structure of rat $\Delta50$Tom20 in complex with the presequence of aldehyde dehydrogenase (pALDH) composed of residues 12 – 22 (PDB: 1OM2) (Abe et al., 2000) suggests that the core comprises of four $\alpha$-helices with a the fifth helix loosely attached to the core. The bound precursor forms an amphiphilic $\alpha$-helix with its hydrophobic surface interacting with the hydrophobic groove of the Tom20 receptor (Figure 1-6C), while the hydrophilic residues were exposed to the aqueous solvent and the hydrophobic interactions mediate presequence recognition. Tethering of the presequence with a spacer containing a cysteine incorporated at its terminus, to the Cys100 of Tom20 along with removal of the fifth helix aided crystallization (PDB: 2V1S, 2V1T) (Figure 1-6D) (Saitoh et al., 2007). Reduction of the disulfide tether in the crystal state resulted in poor density for the presequence peptide denoting mobility (PDB: 3AWR) (Saitoh et al., 2011). Furthermore the structure suggested that the presequence conformation in the crystal was not due to the disulfide bond but resulted from the noncovalent interactions with the receptor along with the residues of the spacer and not the length of the spacer (PDB: 3AX2). Tethering can pose disadvantages by restricting the dynamics and facilitate artifactual structural conformations. Hence another method to aid in crystallization was considered called molecular stiffening where an intramolecular disulfide bond was introduced into the presequence, which restricts the freedom of the peptide without influencing the bound states. NMR studies showed that the incorporation of two cysteines to form an intramolecular disulfide bond increased the binding affinity by 100 fold without changing the binding mode in solution. This complex could be crystallized under several conditions (PDB: 3AX5).

The solution structure of the receptor domain of AtTom20 (PDB: 1ZU2) (Perry et al., 2006) revealed a presequence binding groove on the concave surface of the receptor. It
comprises two TPR like motifs, which are longer than the more conventional motifs (containing 43 - 44 residues instead of 34 residues) however the interhelical angles are retained. AtTom20 show higher similarity to other TPR containing proteins than Tom20 from rat suggesting lack of common ancestry.

1.2.2.4 Tom22

Tom22 is an integral membrane protein with a central transmembrane segment and is associated with the TOM complex. The cytosol exposed N-terminal region is rich in acidic residues which interacts with Tom20 and thus constitutes the presequence binding site (Mayer et al., 1995). The C-terminus faces the IMS and interacts with Tom40 and Tom7 (Esaki et al., 2004). The isolated cytosolic domain of yeast Tom22 bound to the presequence but did not take up an ordered structure. Yeast Tom22 was essential for the stabilization of assembly of Tom20 with the Tom40 core complex (Yamano et al., 2008). It has been proposed that the primitive TOM complex lacked Tom20 and Tom70, as the yeast mutants lacking these proteins are viable only if the level of Tom22 is maintained (Lithgow et al., 1994). Tom22 in plants lack the N-terminal acidic cytosolic domain present in animals. The cytosolic region is shortened to 5 kDa in ArTom22 and 10.8 kDa in ScTom22 (Maćašev et al., 2004). In plants this region is basic, and is highly unlikely to recognize presequences owing to charge repulsion. NMR studies on cytosolic domain of ArTom22 (residues 3 - 48) revealed that the domain was unstructured which could not be reverted using “refolding” purification methods (Rimmer, 2010, Rimmer et al., 2011). Furthermore no interactions were detected with the presequence alternative oxidase (AOX) via NMR methods suggesting that ArTom22 is not a presequence receptor. Cytosolic domain of ArTom22 interacts with ArTom20 and antagonizes presequence binding thus facilitating cargo release for further import.

1.2.2.5 Small Tom proteins: Tom5, 6 and 7

Tom7 is an integral membrane protein with a Nout-Cin orientation, consisting of a hydrophobic core with predicted β-strands and is highly conserved through of evolution (Hönlinger et al., 1996). Import of porin is significantly affected in a Tom7 knockout mitochondrion. The protein is involved in the lateral movement of porin out of the general import pore, thus playing a role in the later stage beyond the recognition of
preproteins by the cytosolic receptor domains. Tom7 favours the dissociation of subunits of the translocase (Tom20/22) from the Tom40 precursor in the early phase of import and in later stage destabilizes interaction of the same with the imported Tom40 reducing the size of the general import complex (Model et al., 2001). Tom6 is deeply embedded into the membrane. The absence of Tom6 destabilizes the interactions between Tom70, Tom40 and Tom20/22 (Hönlinger et al., 1995). Thus Tom6 and 7 do not interact directly with the preprotein however they influence the stability of the TOM complex in a partially antagonistic fashion thereby promoting the translocase dynamics (Model et al., 2001). Tom5 has a C-terminal membrane anchor (residues 27 - 45) with a N-terminal negatively charged cytosolic domain. It is involved in the transfer of preproteins from Tom22 into the Tom40 import pore. Basic residues flank the hydrophobic segment to interact with the charged head groups of phospholipids. It is required for preprotein insertion in the postreceptor stage in yeast (Dietmeier et al., 1997) and may be involved in targeting precursors of Tom7 (Model et al., 2001) and some translocon inner membrane (Tim) proteins destined to IMS (Kurz et al., 1999). Tom5 was closely associated with Tom40 protecting it from proteolysis. Proteome analysis suggest the presence of these proteins in Arabidopsis, however owing to their small size, it is difficult to determine whether these are orthologs or analogs (Murcha et al., 2013). To date, no three-dimensional structure of Tom7/6/5 is available.

1.2.2.6 SAM complex: Sam50

The sorting and assembly of β-barrel proteins is carried out by the SAM/TOB complex, which might also be responsible for the insertion of the a-helical proteins Tom22, Tom5, Tom6 and Tom7 (Thornton et al., 2010). This 140 kDa complex has been isolated from Neurospora, and is composed of Sam50/Tob55. Sam35/Tob38 and Sam37/Tob37 (Figure 1-5) in a stoichiometry of 1:1:1 calculated using isotope dilution mass spectrometry analysis and blue native gel electrophoresis (Klein et al., 2012). Homology modeling of human Sam50 using MODELLER, has been carried out because of high similarity with TpsB transporter FhaC from Bordetella pertuis. The modeled structure comprises 16 β-strands and one POTRA domain (polypeptide transport asssociated). The fact that there is existence of POTRA domain and 12 residues per β-strand clearly denotes evolutionary relationship with Omp85 (Zeth, 2010). The cytosolically synthesized β-barrel preproteins pass through the TOM pore
into the IMS, where they interact with the small Tims (Habib et al., 2005) and are transferred to the SAM complex. The N-terminal α-helical POTRA domain of Sam50 on the IMS face, is proposed to recognize the β-barrel precursors by serving as substrate binding sites (Habib et al., 2007). It also aids in the release of proteins from the SAM complex in the course of insertion into the membrane (Stroud et al., 2011). The junction between the N-terminal strand1 and the C-terminal strand 16 is not stabilized by a connecting loop or turn. In addition significant rearrangement in Sam50 after binding to the precursors has been reported (Kutik et al., 2008).

1.2.2.7 SAM complex: Sam35

Sam35 is tightly bound and anchored to the mitochondrion by Sam50, and is speculated to interact with cytosolic chaperones to keep the β-barrel precursors in an import competent conformation (Waizenegger et al., 2004). β-sorting signals, present in all β-barrel proteins such as Tom40, are conserved across all eukaryotic families and are necessary and sufficient for selective recognition by the SAM complex (Kutik et al., 2008). Sam35 is a peripheral membrane protein and is shown to recognize the C-terminal β-signal further inducing the opening of the Sam50 channel by increasing in the Sam50 channel conductance, whereas Sam37 assists in the release of substrates from the SAM complex. Interestingly it has been shown that a signal–induced SAM channel can accommodate several β-strands contrary to TOM channel, which leads to the idea that β-barrel precursors are inserted into the membrane not as single strands but as several partially folded elements as in case of biogenesis of Tom40 (Rapaport and Neupert, 1999). It has been speculated that there is sequential integration of antiparallel β-strands inside the Sam50 channel and hence a large intermediate β-barrel is formed which then separates into two new barrels (Klein et al., 2012). Sam35 and Sam37, located on the cytosolic face, provide a protective environment and assist in the assembly of the cytosolic domains of the mitochondrial β-barrel proteins (Chan and Lithgow, 2008).

1.2.2.8 SAM complex: Metaxin

Metaxin is an outer membrane protein in the mitochondria which participates in early stages of protein import, are shown to be important for import of VDACs and
recognizes a variety of precursors, however its role differs considerably between plants, yeast and animals (Lister et al., 2007). It has a cytosolically exposed N-terminal domain, and a C-terminal membrane anchor comprising of two transmembrane helices. The N-terminal region of mouse metaxin shows significant sequence identity to Tom37 from yeast (Armstrong et al., 1997). Sam37, Sam35 and metaxin are predicted to have glutathione S-transferase (GST-N) and GST-C domains. AtMetaxin has sequence similarity with HsMetaxin1. In humans Metaxin1 and Metaxin2 play a role in barrel protein import, which is different from Sam50 (Kozjak-Pavlovic et al., 2007). Construct optimization had been carried out for AtMetaxin, which suggest that N-terminal 131 residues behave as a dimer. This dimer is transformed into higher order aggregates upon storage and increase in concentration (described in chapter 2). Different detergents and buffers failed to yield a stable dimer for structural studies. Due to extensive protein aggregation, in vitro interaction studies were not possible, though the construct was shown to inhibit import using an in vivo competition import assay. Similarity searches against proteins with known structure using HHPre did suggested that these proteins from Candida albicans adopted an overall GST-like fold, however the three-dimensional structures are suggested to be significantly different from their human isoforms. Homology modelling attempts for these proteins have not been successful in Sam37, Sam35 and metaxins as no reliable models could be designed based on template proteins with GST-like fold (Qu et al., 2012).

### 1.2.2.9 Mdm-10

Mdm-10 is also a β-barrel protein, which is not a subunit of the SAM complex but a regulatory interaction partner. It participates specifically in Tom40 assembly and is evolutionarily related to Tom40 (Wideman et al., 2010). The protein is found in yeast but not in higher eukaryotes. The function of this protein remains to be further clarified. Structural modeling of T. melanosporum Mdm10 shows insertions in loops L8-9 and L15-16 not seen in Tom40 sequences. L8-9 insertion is proline rich and is predicted to contain two short β-strands, which is modeled as an extended loop. Furthermore, the N terminus, which is predicted to contain two β-strands, forms part of the region external to the β-barrel lumen. Hence Mdm10 forms 20 - 21 β-strands depending on the choice of porin used as model (Bay et al., 2012). Tom7 regulates the association of this protein with the SAM complex (Yamano et al., 2010).
1.2.2.10 Om64

The outer envelope protein-64kDa also known as Om64 has the same location to that of Tom70 on the outer membrane of plant mitochondria but the two proteins are not orthologs and Om64 does not interact with the Tom40 complex as per blue native PAGE analysis (Lister et al., 2007). It comprises an N terminal transmembrane region followed by a cytosolic 3-TPR repeat domain. It exhibits 67% sequence identity with plastid localized Toc64 and interacts with the C-terminal fragment of cytosolic Hsp70 and Hsp90 (Chew et al., 2004). The structure of the 3-TPR domain of Om64 has been modelled using I-TASSER and the interaction with the Hsp peptides have been studied (our unpublished work). It has been shown using surface plasmon resonance (SPR) studies that Om64 prefers to interact with Hsp70 over Hsp90 (Schweiger et al., 2013). Notably in an Arabidopsis mutant that lack all three functional isoforms of Tom20, deletion of Om64 resulted in embryo lethality, indicating that it can act as a receptor for a wide variety of proteins.

1.2.3 Outer envelope receptors of chloroplast

The translocon of the outer membrane of the chloroplast (TOC) is the primary import apparatus necessary for translocation of proteins destined into this organelle. The TOC core comprises of three primary components: Toc34, Toc75 and Toc159 (Figure 1-7). Blue native PAGE analysis of pea leaf chloroplast suggests that these three components are present in a stoichiometry of 3:3:1 respectively (Kikuchi et al., 2006) or 4:4:1 respectively (Schleiff et al., 2003). Transmission electron microscopy (EM) structural analysis suggests that the pea TOC complex forms a toroidal particle with 13 nm diameter and 10-12 nm height (Schleiff et al., 2003).
Chapter 1

Figure 1-7 Schematic representation of receptors from outer membrane of chloroplast.

The bottom panel shows a list of receptors whose structures have been studied.

1.2.3.1 Toc34

Toc34 is an integral membrane protein with an N-terminal cytosolic domain (residues 1 - 266), a transmembrane region (residues 267 - 283) and a C-terminal tail exposed to the IMS (residues 284 - 310) (Seedorf et al., 1995). The crystal structure of the cytosolic domain of PsToc34 (PDB: 1H65) (Sun et al., 2001) reveals a crystallographic dimer. Each monomer contains a GTP binding domain (G domain) and an α-helical region at the C-terminus. A Mg$^{2+}$ ion has been observed at the nucleotide-binding site along with an extra electron density suggesting a bound GDP molecule (Figure 1-8A). Hydrogen bonding interactions of Arg128 and Arg133 along with the hydrogen bonding interactions of GDP with Arg133 and its hydrophobic interaction with Tyr132 (of the partner monomer) are the key players in dimerization. Mutation of Arg128 caused reduction in GTPase activity. It has been proposed that dimerization is necessary for
mutual activation of GTPases, which has been also observed between signal recognition particle (SRP) and the α-subunit of its receptor (SR) (Powers and Walter, 1995). The catalytic arginine residues called the “arginine fingers”, Arg133 in case of Toc34, are important for GTPase activating proteins (GAP) like function (Ahmadian et al., 1997), were also found in SRP and SR. PsToc34 has two paralogs in the Arabidopsis genome (AtToc33 and AtToc34). The crystal structure of AtToc33 residues 1–251 in complex with Mg$^{2+}$ and GDP/GMPPNP (PDB: 3BB3 and 3BB1 respectively) has been determined and the study proposes a concentration dependent dimerization as detected using analytical ultracentrifugation evidenced by a higher dissociation constant for AtToc33 compared to PsToc34 (Koenig et al., 2008). Although AtToc33 and PsToc34 have similar conformations in the ligand bound states, mutation of the conserved Arg130 (corresponding to Arg133 of PsToc34) to alanine in the former causes a minor reduction in the catalytic rate and this mutant was crystallized as a monomer (residues 2 - 250) (Figure 1-8B) (Yeh et al., 2007). The monomer resembled that of PsToc34 as the two proteins had 59% sequence identity, however structural variations were observed in the region corresponding to the dimer interface of PsToc34. The α4’-helix and the β3’-strand (PsToc34) exist as extended coils in the AtToc33 (R130A). Furthermore, the small 310 helix (α5’) forms an extended α6 helix in the latter monomeric structure. As observed using gel filtration studies, in the presence of exogenous Mg$^{2+}$, Arg130 aided dimerization in freshly prepared AtToc33wt and does not function as arginine finger. This protein is found to be associated with Toc75 (Seedorf et al., 1995). It is speculated that Toc34 modulates the gating properties of Toc75 or in the recognition and presentation of precursors (Keegstra and Froehlich, 1999, Schleiff and Soll, 2000).
1.2.3.2 Toc75

Toc75 is a member of Omp85 superfamily, a group of β-barrel proteins found in gram-negative bacteria, mitochondria and chloroplast. There are three N-terminal POTRA domains facing the cytosol followed by a C-terminal β-barrel pore, however the N and C termini face the IMS. Two topological models have been proposed for Toc75. The first model is based on a hydropathy plot and CD spectroscopy. The model predicted the presence of 16 membrane spanning β-sheets connected by loop made out of ~229 amino acids and has a hydrophilic pore of diameter 0.8 – 0.9 nm (Hinnah et al., 1997). The second model was based on proteolytic digestion of proteoliposomes with amino acid sequencing and modelling (Sveshnikova et al., 2000). This model contradicted the previous model in containing large loops in the termini facing the IMS. However the numbers of predicted strands by both the models were consistent. As no high-resolution structure is available, homology modelling studies of PsToc75-POTRA domains were performed using the crystal structure of FhaC (PDB: 2QDZ) (Dave, 2010). The individual POTRA domain comprises two α-helices packed against three β-strands (β-α-α-β-β fold). Seven inter-β-strand loop domains have been identified as Toc Loop Domains (TLD). The POTRA1 domain is shown to interact directly with the transit peptides however the role of POTRA2-3 is not well understood. Four genes are found
encoding the homologue of \( PsToc75 \) in \( Arabidopsis \) namely \( AtToc75-1, AtToc75-III, AtToc75-IV \) and \( AtToc75-V \) based on their location of different chromosomes (Jackson-Constan and Keegstra, 2001, Eckart et al., 2002) of which evidence at protein level for \( AtToc75-I \) is not available. \( AtToc75-III \) is the functional homologue to \( PsToc75 \) and shares 73% sequence identity (Inoue and Keegstra, 2003). \( AtToc75-IV \) is 44 kDa, shares 60% sequence identity with \( PsToc75 \) and contains 8 \( \beta \)-strands and no POTRA domain (Jackson-Constan and Keegstra, 2001). \( AtToc75-V \) (also called \( AtOEP80 \)) shares 22% sequence identity to \( PsToc75 \), and is hypothesized to be involved in insertion of other \( \beta \)-barrel proteins into the outer membrane. It is predicted to form porin like channels with 16 \( \beta \)-strands and is more closely related to Sam50 of mitochondria than Toc75 (Eckart et al., 2002). Heterologously expressed Toc75 reconstituted with liposomes formed a cation selective channel where the channel properties were affected only by preproteins that mimic the transit peptides (Hinnah et al., 1997). Toc75 shows preference for transit peptides based on charge and conformation and is involved mainly in hydrogen bonding and van der Waals interactions (Hinnah et al., 2002). It is predicted to act as chaperone during the translocation of its substrate (Sánchez-Pulido et al., 2003).

### 1.2.3.3 Toc159

Toc159 (also referred as Toc160) was originally identified as a proteolytic fragment Toc86. Proteolytic sensitivity studies (Chen et al., 2000) on the full length protein was identified to have a tripartite domain structure: an N-terminal acidic domain or the A domain (residues 1 - 598) with a pI of 3.6 containing two repetitive motifs, a GTP binding domain or the G domain (residues 599 - 1062) followed by a C-terminal membrane anchor domain or the M domain (1063 -1499). However a three-dimensional structure is not available to date. CD studies have shown that the A domain of \( AtToc159 \) is disordered and undergoes conformational changes at temperature and pH extremes as observed for intrinsically disordered proteins (Richardson et al., 2009). This domain is analogous to Tom22 and Tom20 and is thought to play a role in electrostatic interactions with the positively charged transit peptides. Furthermore, the N terminal domain consists of motifs mimicking the ATP-binding site of protein kinases, however the function has not been characterized (Bölter et al., 1998). The A- and the G- domains are cytosolic, however the latter is found attached to the outer membrane in the absence
of the M domain, which suggests the possibility of interaction with other subunits of the import complex (Bauer et al., 2000). Toc34 and Toc159 are integral GTPases, which act as primary receptors for preproteins and share homology in their G domains (Andrès et al., 2010) however this homology is not shared with other GTP binding proteins. Similar to Toc34, dimerization is proposed to play an important role in GTPase activation in Toc159. Interestingly, Toc159 interacts with Toc33 in Arabidopsis (Hiltbrunner et al., 2001) and the G domains of Toc159 forms a heterodimer which has been shown to occur in vivo (Rahim et al., 2009). Removal of the G-domain affects the precursor binding and import intermediate formation whereas the translocation remains unaffected (Chen et al., 2000). There are three homologues found in Arabidopsis namely AtToc159, AtToc132 and AtToc120 which share 48%, 37% and 39% sequence identity with Pstoc159 respectively and the latter two are hypothesized to be receptors for non-photosynthetic precursors (Jarvis and Soll, 2001).

1.2.3.4 Toc64

Toc64 is suggested to be transiently associated with the core complex (Schleiff et al., 2003). The consensus model for Pstoc64 topology has been proposed to consist of a N-terminal transmembrane region, an IMS localized amidase domain which is silenced by a point mutation, a charged region and a C-terminal cytosolically exposed 3-tetratricopeptide repeat domain (3-TPR) (Qbadou et al., 2006). The protein traverses the membrane three times and has a N_in-C_out orientation. The 3-TPR domain is a repeat of helix turn helix motif and interacts with the C terminal region of the heat shock proteins which carry the preproteins, thus aiding preprotein delivery to the import complex (Mirus et al., 2009). AtToc64 displays 67% sequence similarity with Pstoc64 and it has been proposed that the N-terminal domain functions as an insertion signal (Lee et al., 2004). The 3-TPR domain of AtToc64 along with the solvation helix (capping helix) has been modelled using I-TASSER and is shown to interact with equal affinity with the C-terminal fragment of human Hsp70/90 using isothermal titration calorimetry studies (Panigrahi et al., 2013) and Arabidopsis using SPR studies (Schweiger et al., 2013).

1.2.3.5 Toc12

Toc12 is another later identified component of the TOC complex, which exposes a
soluble C-terminal domain to the IMS where it interacts with IMS localized ATP-bound Hsp70. Further on hydrolysis of ATP, which is induced by the above interaction, the chaperone interacts with the incoming preprotein at the exit site of the Toc complex in the IMS face. Bioinformatic analysis of PtToc12 has shown that the N-terminus of the protein contains a β-barrel type membrane anchored and the C-terminus exhibits high sequence similarity to the J domain of Dna-J. The J domain in Toc12 stimulates ATPase activity of Hsp70 chaperones. A three dimensional model was built using HsHsp40 as a template (Becker et al., 2004) and subsequent molecular dynamics simulation studies which showed that residues 44 - 60 form a flexible region whereas residues 63 - 103 form a highly structured region. A proline residue present in the loop connecting the first two helices in HsHsp40 is absent in Toc12. In addition a further two cysteines at the end of this loop stabilize the structure by disulfide bond formation. Disruption of this bond leads to disruption of the co-chaperone activity.

1.3 DISCUSSION

A significant insight into the protein import pathway in peroxisome, mitochondria and chloroplast has been obtained by combination of different techniques. Mutational analysis along with biophysical techniques like isothermal titration calorimetry, surface plasmon resonance, mass spectrometry, analytical ultracentrifugation aid in verifying regions important for interaction and the stoichiometry of complexes in vitro, which is further confirmed by in vivo cell based assays with the full-length membrane spanning proteins along with their partners in the complex. Crystallography is most preferred technique, which helps in determining high-resolution protein structures, although the nonspecific protein aggregation hampers crystal formation. The crystal structure of membrane spanning β-barrel proteins aiding in import have not been reported. Apart from the problem of aggregation, these integral membrane proteins may need their partners to stabilize them. Thus one strategy widely used is deletion of membrane spanning region in import receptors as in Tom20, Tom71 etc. Additionally solution structures using NMR is not uncommon when diffracting crystals cannot be obtained as in case of AtTom20, however there is limitation to size of the protein and aggregation. Molecular tethering or tying the signal peptide to the receptor with a linker sequence has also aided in capturing receptors in ligand bound state for structural studies using NMR or crystallography as in Tom20. It is worth mentioning that homology modeling
has significantly contributed to the structural understanding of many β-barrel import receptors such as Tom40. In addition electron microscopy has aided in getting an idea regarding the dimensions of the channels of the complexes as in Tom40.

Thus the task of delineating the structural architecture of protein import receptors is quite challenging. The emerging structural advances promise to explain us how the signal recognized by one domain is transmitted to other receptors in the complex thus facilitating import. Clearly a combination of high-resolution imaging techniques complemented with x-ray crystallography or NMR are needed to understand the function and structure of the receptor complexes in more details.

1.4 FUTURE PERSPECTIVES

In the last decade, with the development of high resolution imaging techniques combined with computer-assisted modelling and mutational studies, progress has been made in understanding the three dimensional structure and dynamics of outer membrane receptors from peroxisome, mitochondria and chloroplast. However many questions remain to be answered.

Although a number of structures of peroxisomal receptors have been obtained, there are no reports on those from plants (Figure 1-2). Two models have been proposed for peroxisomal protein import, one in which subunits of multimeric proteins are imported as monomers into the peroxisomal matrix and the other in which monomers are folded, assembled and targeted as multimers which is seen in the PTS2 pathway in plants (Flynn et al., 1998). Structural information of peroxisomal receptors bound to multimeric proteins could aid in understanding the mode of stabilization of the complex. Similarly structural insight on Pex26 or its counterpart from Arabidopsis APEM9 would throw light on the mechanism involving localization of Pex1-Pex6 complex to the peroxisome. PMPs are targeted to the peroxisomes with the help of Pex19-Pex3 complex, however they are also trafficked via ER. It would be interesting to study the targeting signals and chaperones involved in either of the pathways, thus aiding to gain insight into peroxisomal membrane biogenesis. Although a number of structures of peroxisomal receptors and their complexes have been obtained, there are no reports on those from plants (Figure 1-2). A detailed structure-function study on peroxisome
import pathway in plant might aid in highlighting the differences to the yeast or mammal counterparts and on a broader level help to understand the evolution of adaptation by peroxisomes in plants to biotic and abiotic stress.

Compared to peroxisomal receptors, few reports on structural details are available regarding mitochondrial and chloroplastic outer membrane receptors (Figure 1-5 and Figure 1-7). It has been shown that Tom proteins cannot import their own preproteins and need other Tom receptors to aid in specific import (Pfanner and Geissler, 2001). It is interesting to study the structural determinants that aid receptors differentiate between self and non-self. Small Toms e.g. Tom5, 6 and 7 are known to play role in translocon dynamics, which makes the understanding of their structure and mode of interaction with its partners very essential. Although a number of structures of cytosolic domain of rat Tom20 in complex with rat ALDH presequence have been reported, structural basis of interaction of Sam35 and Sam37 with the precursors of β-barrel proteins is yet to be understood. Further considering the diversity in mitochondrial targeting signals destined to different compartments of mitochondria, it interesting to note that all of them are translocated by the same TOM complex. An in depth investigation might unravel new modes of presequence recognition by different receptors. The domain classification of metaxin shows the presence of putative glutathione binding site in its N-terminus. It is interesting to know if this site is functional and if this aids in presequence recognition. Further the structural and mechanistic details of assembly and insertion of β-barrel proteins into the membrane by SAM complex that involves tightly regulated transversal and lateral pores, is poorly understood.

To date only structure of transit peptide reported is that of algal Ferridoxin using NMR (Lancelin et al., 1994). Structural insight into interaction of chaperones with Toc64 has been recently obtained, however no structural details have been obtained regarding interaction of chloroplast membrane receptors with transit peptides. Although bioinformatics predictions of transit peptides have been carried out extensively, peptide-receptor complex is necessary to validate the residues responsible for recognition and further understand the mechanism of import into chloroplast. Structural details explaining the mode of interaction of Toc159 with preproteins and with Toc33 in Arabidopsis remains to be explained. The homologues of this protein found in Arabidopsis are hypothesized to interact with non-photosynthetic precursors, the
mechanism of which is yet to be established. Thus the structural studies would help us to understand the precise mechanism of protein import.

1.5 SCOPE OF THE THESIS

Mitochondria and chloroplasts are two organelles that coexist in plant cells and their genomes code for a limited number of proteins. The functioning of these organelles is dependent on nuclear encoded, cytosolically synthesized preproteins destined to these organelles, which are recognized by import receptors on the outer membrane of these organelles. These preproteins have targeting signals, which aid in their recognition by specific receptors. Molecular chaperones aid the preproteins to maintain their native unfolded state in cytosol prior to import. Some outer membrane receptors specifically recognize chaperones, when laden with preproteins. Thus specificity of ligand recognition by outer membrane import receptors is important for maintaining the fidelity of protein targeting.

This investigation is conducted to understand the molecular details underlying specificity of interactions between protein import receptors and their respective ligands. To achieve this aim, the cytosolic domains of three import receptors from Arabidopsis thaliana are chosen and biophysical characterization is carried out using experimental and computational methodologies.

Chapter 2 focuses on the biophysical characterization of the cytosolic domain of Metaxin, an import receptor on the outer mitochondrial membrane and a member of the SAM complex. A number of recombinant deletion constructs are designed to delineate the stable soluble cytosolic domain and their subsequent expression in E. coli reveals that the N-terminal 131 residues have the propensity to form dimer. However the dimer state is short lived because of its tendency to form high-order aggregates. Attempts to prevent aggregation have not been successful. Competition import inhibition experiments reveal that the dimer is functional under in vivo experimental conditions, however the aggregation prone nature of the protein prevents ligand binding studies using isothermal titration calorimetry. The above observations suggest that the cytosolic domain of Metaxin might need other receptors from the SAM complex for its in vitro stability.
Chapter 3 focuses on the biophysical characterization of the cytosolic domain of Toc64, an import receptor on the outer membrane of chloroplast. The recombinantly expressed cytosolic domain of the protein is a monomer and bioinformatics analysis shows that it is comprised of three TPR motifs. Reports indicate that this 3-TPR domain interacts with human homologues of C-terminal octapeptides of Hsp70 and Hsp90 chaperones with a preference towards the latter using in vivo studies. In the current study, the above interactions are investigated using isothermal titration calorimetry and molecular dynamics simulation studies, in order to understand the contribution of residues towards recognition and selectivity. Contrary to previous findings, the current study indicates that the C-termini of both the chaperones interact with the TPR domain with similar affinity. Further the C-terminal EEVD motifs of the octapeptides are important for the interaction and mutation of any of these residues to alanine significantly affects binding. Finally a study of protein dynamics upon ligand binding reveals that the terminal helices, H1 and H7 of the TPR domains undergoes extensive motion as compared to the helices constituting the inner surface of the TPR cradle.

Chapter 4 focuses on the biophysical characterization of the cytosolic domain of Om64, an import receptor on the outer membrane of mitochondria with regards to its interaction with Hsp70 and Hsp90 chaperones from Arabidopsis, using computational methods. Toc64 and Om64 are homologs, hence comparative binding studies and energy calculations are performed in order to understand the molecular details underlying specificity for their interactions with Hsp70/90. The TPR domain from Toc64 interacts preferentially with the C-terminal octapeptide of Hsp70 and that of Om64 interacts with both chaperones with similar affinity. Furthermore, the terminal helices, H1 and H7, show motion in the Toc64_TPR-Hsp70 complex, which is higher compared to that observed in the other three complexes. Previous reports indicate that chloroplast proteins can be mis-targeted to the mitochondria, however the opposite phenomenon is not observed. The lack of specificity of Om64 towards Hsp70 and Hsp90 described in the current study explains the phenomenon of mistargeting at a molecular level.
Chapter 2

Biophysical characterization of recombinant protein expressing the cytosolic domain of Metaxin from

*Arabidopsis thaliana*
2.1 INTRODUCTION

The conservation of import apparatus through various eukaryotic lineages has been shown experimentally (Lister et al., 2005). The presence of β-barrel proteins with multiple membrane spanning β-strands found only in the outer membrane of eukaryotic mitochondria and chloroplasts, reflect an evolutionary origin from bacterial ancestors (Neupert and Herrmann, 2007). Some of these proteins include Tom40, Tob55 (also called Sam50 or Tom50) and Voltage-dependent anion-selective channels (VDAC) or porins. Their precursors are imported through the TOM complex and then through the Sorting and Assembly Machinery (SAM) complex (Wiedemann et al., 2003) or Topogenesis of mitochondrial Outer membrane β-barrel (TOB) complex (Paschen et al., 2003).

The important components of the SAM/TOB complex are Sam50 (Tob55/Tom50), Sam37 (Mas37/Tom37) and Sam35 (Tob38/Tom38). Tob55 is the central functioning unit as it is highly conserved (Neupert and Herrmann, 2007) and Sam35, which is solely peripheral, is thought be involved in mediating transient association between the TOB and TOM complexes (Thomas et al., 2004). The assembly function is strongly impaired in a Sam50 deficient mitochondrion while the import function is still retained (Kozjak et al., 2003). Deletion of Sam50 and Sam35 selectively affects the insertion and assembly of barrel proteins (Kozjak et al., 2003, Paschen et al., 2003). So the β-barrel precursor proteins first interact with Tom20 and move through the TOM pore. Small Tim proteins *viz.*, Tim9 and Tim10 guide these precursors to the SAM complex (Figure 2-1). Sam50 has a N-terminal hydrophilic region called the POTRA (polypeptide translocation associated) domain that is involved in recognition of these precursors exposed to the inner membrane space (Sánchez-Pulido et al., 2003). It might then pass these precursors to the membrane embedded C-terminal region (which forms a β-barrel domain with 14-16 trans-membrane β-strands) of Sam50, which facilitates insertion and assembly of these proteins (Neupert and Herrmann, 2007).
The presequence of the nascent preprotein bound to the heat shock protein is recognized by Tom20, which is a member of Tom complex and is passed through the Tom40 pore. In the Inner membrane space (IMS) preproteins are associated with small Tims (Tim8, Tim9, Tim10 and Tim13), which direct them to SAM complex. This complex helps in insertion of β-barrel proteins as shown by the blue dotted arrow. Other members of the TOM complex include Tom5, Tom6, Tom7, and Tom9. Om64 is a protein, which is also known to help in recognition of the C-terminus of heat shock proteins, thus aiding in precursor import through Tom40.

**Figure 2-1** Insertion of β-barrel proteins on the outer membrane by SAM complex.

Metaxin (Mtx) is an outer membrane protein of the mitochondria, involved in the sorting and assembly of β-barrel proteins on the outer membrane similar to the SAM complex explained above (Lister et al., 2007). The gene coding for Mtx was initially discovered to be located between the glucocerebrosidase gene and the thrombospondin gene in mouse on chromosome 3E3-F1 and was identified to be important for embryonic development in mice (Bornstein et al., 1995). In mammals, two isoforms of Mtx have been identified: Mtx1 is a component of the import complex, contains a C-terminal signal anchor domain and shares weak identity with the N-terminus of yeast Tom37 (Abdul et al., 2000) and Mtx2 shares 29% identity with Mtx1 but lacks the transmembrane anchor and is bound to the cytosolic face of the mitochondrial outer membrane by means of its interaction with Mtx1 (Armstrong et al., 1999, Abdul et al., 2000). Mtx1 is the mammalian homologue of Sam37 and Mtx2 is the mammalian homologue of Sam35. They play a role in the biogenesis of TOM40 and VDAC (Kozjak-Pavlovic et al., 2007, Lister et al., 2007). Studies have also shown that Mtx is
also required for tumour necrosis factor α (TNF-α) induced cell death (Armstrong et al., 1997, Wang et al., 2001). Removal of the membrane anchoring domain of Mtx leads to disruption of its mitochondrial localization and is also shown to be required for Mtx to be involved in TNF induced cell death (Wang et al., 2001). Mitofilin, a protein that helps in maintaining mitochondrial cristae morphology located on the inner membrane, has been found, using co-immunoprecipitation studies in human cell lines, to be present in a complex with Mtx1, Mtx2 and SAM50 located on the outer membrane (Xie et al., 2007). However blue-native polyacrylamide gel electrophoresis (BN-PAGE) has suggests that human Mtx is not present as a complex with SAM50 (Kozjak-Pavlovic et al., 2007). These observations suggest diversity in the function of Mtx.

Only one form of Mtx has been found in Arabidopsis, which exhibits 21% sequence similarity with mouse Mtx and 11% similarity with yeast Sam35 (Lister et al., 2007). Hence plant Mtx appears to significantly diverge from that of yeast and animals. Based on sequence analysis, Mtx from Arabidopsis has been shown to possess two hydrophobic segments near the C-terminus which possibly anchors the protein to the membrane (Lister et al., 2007). Co-localization studies indicate that the fusion construct of full-length Mtx with a N-terminal green fluorescence protein was directed into the mitochondria (Lister et al., 2007). Accumulation of VDAC in the cytosol and reduction in the rate of import of VDAC and TOM40 in Mtx deficient plants suggests that it plays a role in β-barrel protein import (Lister et al., 2007). It is also found to interact with certain mitochondrial precursor proteins such as alternative oxidase (AOX), the FAd-subunit of mitochondrial ATP synthase and mitochondrial phosphate carrier protein (PiC) through their targeting signals (Lister et al., 2007). Competition import assays suggested that the region between residues 72 - 112 was essential for the recognition of the AOX signal sequence and import (Lister et al., 2007). Further phylogenetic analysis suggested that Mtx from Arabidopsis belongs to the glutathione S transferase (GST) superfamily and contain an N-terminal GST-N-METAXIN and a C-terminal GST-C-METAXIN conserved domain (Lister et al., 2007). The GST-N-METAXIN belongs to the Thioredoxin superfamily. Based on this information regarding the domain organization and function of Mtx, an analysis of the structural features of Mtx that mediate presequence recognition has been carried out. In this study, we focus on biophysical methods such as circular dichroism (CD), native gel analysis and isothermal titration calorimetry (ITC) to identify the stable and functional cytosolic domain of
Chapter 2

Metaxin from Arabidopsis. Our findings suggest that a recombinant expressed construct containing the N-terminal 131 residues can be purified as a dimer, however it aggregates over time. Though, in vivo import inhibition studies suggest that this Metaxin construct is active, the aggregation hindered in vitro biophysical experiments such as ITC. After extensive experimentation, we deduce that additional factors might be involved in conferring stability to the cytosolic domain of Metaxin and thus aiding in presequence recognition.

2.2 MATERIALS AND METHODS

2.2.1 Bioinformatics analysis

The sequence of full length Mtx was obtained from the UniProt database (Bairoch et al., 2005). Secondary structure prediction were carried out using PSIPRED (Buchan et al., 2010). Disorder prediction was assessed using FoldIndex (Prilusky et al., 2005) and IUPred (Dosztányi et al., 2005). Amino acid based sequence analysis was performed using BLAST (Altschul et al., 1997). Transmembrane region predictions were performed using the HMMTOP algorithm (Tusnády and Simon, 1998). Different parameters of primary structure analysis such as molecular weights and pIs for deletion constructs were computed using ProtParam (Gasteiger et al., 2005). The 3D models were generated using the I-TASSER online server (Zhang, 2008), which generates models using multiple threading alignment procedures. This method avoids bias towards a particular structural model, which would occur if homology modelling methods were used. A confidence score (C-score) and a TM score, used to assess the topological similarity, are assigned to each of the above models.

2.2.2 Peptide synthesis

Synthetic peptides mimicking the soya bean alternative oxidase presequence (SGANRVANTAMFVAKG) and peptide derived from yeast cytochrome c oxidase subunit IV (CoxIV) presequence (SLRQSIRFFK) were designed. These peptides were synthesized and provided at >85% purity by Biomatik (Canada).

2.2.3 Construction of plasmid expressing cytosolic domain of Mtx

A plasmid containing the Mtx gene was obtained from Prof. Jim Whelan at the ARC Centre of Excellence in Plant Energy Biology at the University of Western Australia. A
series of deletion constructs of Mtx were designed in order to delineate the region encompassing the cytosolic domain for further biophysical characterization. *NcoI* and *XhoI* restriction sites were designed at the 5’ and 3’ ends of the forward and reverse primers respectively for PCR amplification and subsequent steps of cloning. Primers used for designing various deletion constructs are shown in Table 2-1. pETM10/pETM11 (Invitrogen), were chosen as suitable bacterial expression vectors as they contain codons to incorporate a N-terminal hexa-histidine tag as well as a kanamycin resistant cassette.

**Table 2-1 Primers used to design metaxin constructs.**

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metaxin Sense_ Nco1</td>
<td>5’-CCATGGAAAGGCGATCAAGAGA-3’</td>
</tr>
<tr>
<td>ΔMetaxin82N_Xho1</td>
<td>5’-GGGCTCGAGTTATATCAATCTTCTTTTGAG-3’</td>
</tr>
<tr>
<td>ΔMetaxin112N_Xho1</td>
<td>5’-CGCCTCGAGTTATAGGTAAGCGCTTCTTCTAGCC-3’</td>
</tr>
<tr>
<td>ΔMetaxin131N_Xho1</td>
<td>5’-GGGGCTCGAGTTATTAATCTGAATAGTAGATTTTCGATG-3’</td>
</tr>
<tr>
<td>ΔMetaxin153N_Xho1</td>
<td>5’-GGGCTCGAGTTATATCTATGCTGTCAAAC-3’</td>
</tr>
<tr>
<td>ΔMetaxin182N_Xho1</td>
<td>5’-GGGCTCGAGTTATACCTTCAAAACGCTT-3’</td>
</tr>
<tr>
<td>ΔMetaxin191N_Xho1</td>
<td>5’-CGACTCGAGTTATATTCGATTTCAAA-3’</td>
</tr>
<tr>
<td>ΔMetaxin230N_Xho1</td>
<td>5’-GGGCTCGAGTTATAGCATAATCGCACAAGATTAC-3’</td>
</tr>
<tr>
<td>ΔMetaxin240N_Xho1</td>
<td>5’-GGGCTCGAGTTATAGCTTGGATGAGAAGCCTGACT-3’</td>
</tr>
</tbody>
</table>

The primer names of the Metaxin deletion constructs represent the number of residues from the N-terminus of the sequence of Metaxin from *Arabidopsis thaliana* followed by the restriction site. In the case of Toc64 and Om64, the primer names represent the domains of the respective proteins followed by their restriction site. The restriction sites in the sequences are shown in bold font and underlined. *NcoI* and *XhoI* restriction sites were present in the sense and antisense primers respectively. The PCR products were cloned into the expression vectors successfully as established by sequencing the final cloned plasmids. Thus the desired deletion constructs with hexa-histidine tags were prepared.

2.2.4 Over expression and purification of deletion constructs

The plasmids containing the deletion constructs were transformed into the BL21 (DE3) strain of E. coli. A pre-inoculum was prepared by inoculating a single colony into 5ml
Luria Broth (LB) with 40 µg/mL kanamycin as the antibiotic and grown overnight in a shaking incubator at 37 °C. This pre-inoculum was used to inoculate 500 ml of LB containing 40 µg/ml of kanamycin. The culture was grown at 37 °C and the optical density (O.D.) at 600 nm was monitored regularly. At an O.D. of 0.5, protein expression was induced by the addition of isopropylthiogalactoside (IPTG) to a final concentration of 1 mM and the culture was further incubated for 4 hours at 37 °C. The bacterial medium was centrifuged at 8000 rpm at 4 °C for 15 minutes and the pellet was resuspended in 50 mL of buffer containing 50 mM Hepes pH 7.5, 500 mM KCl and 10% glycerol and 1% sodium lauroyl sarcosinate. The resuspended pellet was lysed using Emulsiflex C5 high-pressure homogenizer at 14000 psi. The lysed cells were centrifuged at 15000 rpm at 4 °C for 40 minutes. The pellet was discarded and the supernatant incubated overnight with 2mL of Ni-NTA agarose beads at 4 °C. The beads with bound lysate were placed in a column and eluted under gravity. The flowthrough was collected as well as further washes using a various concentrations of imidazole (30 mM, 50 mM and 100 mM) in 50 mM Hepes pH 7.5, 500 mM KCl and 10% glycerol. The histidine tagged protein was eluted from the column at 200 mM and 500 mM imidazole in the above buffer. All fractions were analyzed for purity on a 15% SDS PAGE gel.

2.2.5 Competitive import inhibition assay

An import inhibition assay (Lister et al., 2007) was performed using purified Mtx constructs, which were expressed in E. coli. [35 S]-Met-labeled precursor proteins were synthesized using a rabbit reticulocyte TnT in vitro transcription/translation lysate (Promega). Freshly isolated mitochondria obtained from Prof. Jim Whelan at the ARC Centre of Excellence in Plant Energy Biology at the University of Western Australia, was used for the study. 50 µg of mitochondria were added to 180 µL of ice-cold import master mix (0.3 M sucrose, 50 mM KCl, 10 mM Mops, 5 mM KH2PO4, 0.1 % w/v BSA, 1 mM MgCl2, 1 mM Methionine, 0.2 mM ADP, 0.75 mM ATP, 5 mM succinate, 5 mM DTT, 1 mM GTP and 1 mM NADH, pH 7.5) and incubated on ice for 3 minutes. Five microliters of the radiolabeled precursor protein was mixed with 15 µL of competitor protein (2 µg) and the import reaction was initiated and incubated at room temperature for 15 minutes. The radiolabeled precursor/competitor solution was added to the mitochondrial import master mix and incubated at 26 °C for 10 minutes. The import reaction was halted by incubation on ice and Proteinase K treatment was carried out. Proteolysis was inhibited by the addition of 1 µL of 100 mM phenylmethylsulfonyl
fluoride (PMSF). Mitochondria were reisolated by centrifugation at 20,000 g for 3 min at 4 °C and were resuspended in the SDS PAGE sample buffer. The protein samples were separated by SDS PAGE gel and import of radiolabeled protein was detected.

2.2.6 Gel filtration chromatography

A superdex™75 prep grade size exclusion chromatography column (GE Healthcare) with a bed volume and a void volume (V₀) of 120 mL and 45 mL respectively, was used for subsequent purification. The protein, obtained after Ni-NTA chromatography, was concentrated using an Amicon centrifugal filter unit (10 kDa cutoff) to 10 mg/mL before applying to the size exclusion chromatography column and was eluted using an elution buffer containing 20 mM Hepes pH 8.0, 50 mM KCl and 10 % glycerol. The relative molecular masses of the peaks obtained were calculated using a logarithmic interpolation from a calibration curve made up of gel filtration standards using their elution volumes (Vₑ). Fractions corresponding to non-aggregated protein (not present as the void volume of the column) were pooled and concentrated using the above concentrator according to the requirements needed for subsequent biophysical experiments.

2.2.7 Circular dichroism (CD) and thermal shift assay (TSA)

Far-UV CD spectra corresponding to peptide bond absorption were recorded from 185 to 260 nm at 10 °C using a Jasco-810 spectropolarimeter. Spectra were collected for 10 μM of protein in 10 mM of sodium phosphate buffer pH 7.3 in a Quartz SUPRASIL cuvette (Hellma) with a path length of 1 mm. Measurements were performed with an increment of 0.5 nm an integration time of 4 sec and bandwidth of 2 nm. The signal due to buffer alone was subtracted from that of the protein. The proportions of secondary structures of the protein was estimated from the [θ] values between 190 and 240 nm using the DichroWeb server (Whitmore and Wallace, 2008). The change in molar ellipticity was monitored at 222 nm while varying the temperature from 10 - 80 °C at a rate of 1 °C/minute. The thermally denatured sample was cooled to 10 °C at the same rate, to observe the effect of temperature on the folding of the recombinantly expressed non-aggregated Mtx deletion construct.

Thermal shift assays (Niesen et al., 2007), using the fluorescent dye SYPRO Orange (Invitrogen), were performed to investigate the stability of the non aggregated deletion mutant of Mtx in different buffers conditions. SYPRO Orange enables one to
distinguish between the folded and unfolded state of a protein by binding to the hydrophobic portions of protein followed by a change in the intensity of fluorescence emission. When the protein is thermally denatured, more of the hydrophobic interior is exposed and the fluorophore emission increases. A CFX96 real-time PCR detection system (Bio-Rad) was used for these experiments. A master mix (160 μL) was prepared by adding equal volumes of protein at a concentration of 7 mg/mL and 250X dye. One microliter of the master mix was added to each well of a 96-well PCR plate (Bio-Rad) containing 19 μL of test buffer. Each reaction was performed in triplicate. The test buffers were designed to screen a range of buffers (50mM) at different pHs with/without 50 - 300 mM KCl or NaCl. Lysozyme was used as a positive control. The plate was sealed with Crystal clear sealing tape (Hampton Research) and thermal denaturation curves measured from 4 °C to 90 °C in 0.5 °C/30 second increments. The unfolding reaction was simultaneously monitored using a fluorescence detector setup customized to accommodate the optimal excitation (Ex) and emission (Em) wavelengths for SYPRO Orange (Ex:490/Em:570) (Phillips and de la Peña, 2011).

2.2.8 Dynamic light scattering (DLS) and analytical ultracentrifugation (AUC)

In order to ascertain the oligomeric state of the protein, DLS experiments were carried out using a Zetasizer Nano ZS (Malvern) using manufacturer’s operating protocol. 35 μM of protein in the superdex buffer was used for this study. The measurement was carried out in triplicate and each run consisted of 10 acquisitions.

AUC (sedimentation velocity) experiments were carried out to independently establish the homogeneity of protein molecules in solution and determine their molecular mass (Lebowitz et al., 2002). These experiments were conducted on a Beckman XL-A analytical ultracentrifuge, equipped with a An60-Ti rotor and absorbance optics (Ralston). The experiments were carried out using 70 μM of protein (390 μL) and superdex buffer (400 μL) separately loaded into a double sector centrepiece and built up in the rotor. Protein samples were monitored by UV absorbance at 280 nm in a continuous mode. The sample was centrifuged at 60,000 rpm for 16 hours under a vacuum of less than 100 micron at 20 °C. A total of 300 scans were collected. The sedimentation coefficient distribution function of the macromolecule, c(s) method implemented in the program SEDFIT, wherein the value is obtained from a direct fit to the experimental data (Michel et al., 2006) was used for the analysis. The partial
specific volume (0.74 mL/g), buffer density (1.005 g/mL) and viscosity (0.01002 P) obtained from SEDNTERP were used for the calculations (Laue et al., 1992). This benefits from excellent resolution and sensitivity for characterizing sample homogeneity. A confidence level of p = 0.95 was used while solving the size distributions.

2.2.9 Non reducing PAGE and native PAGE analysis

In order to examine whether the intra or interchain disulphide bonds play a role in oligomerization or stabilization of Mtx constructs, non reducing PAGE analysis was performed. Aliquotes of protein at 1 mg/mL were incubated with increasing concentration of glutathione for 12 hrs. 12% non reducing PAGE was used, the components of which were similar to Laemmeli’s SDS PAGE, without β-ME and DTT.

The effect of detergents on the stabilization of the oligomeric state of the non-aggregated deletion mutant was accessed using 10 % native PAGE. As the pI of the deletion mutant was less than 7, discontinuous native PAGE was used. The components of the native PAGE were similar to that of Laemmeli’s SDS PAGE, without β-ME, DTT and SDS. Detergents including Triton, Chaps, Decylmaltoside (DM), Dodecylmaltoside (DDM), Tridecylmaltoside (TDM), N,N-Dimethyldodecylamine N-oxide (DDAO) and β-octyl glucoside (β-OG) were used at three times their critical micellar concentration (CMC). All the detergents were purchased from Anatrace. Freshly prepared aliquots of each purified construct at a concentration of 0.5 mg/mL were incubated with detergents for three days initially and then again for a period of three weeks in order to assess the effect of detergents on the protein with time. Ovalbumin (Sigma), at the above concentration, was used as a control as it had the same pI as that of the construct. These experiments were carried out in triplicate after independent purifications.

2.2.10 Limited proteolysis

This method is widely used to probe protein folding and identification of stable domain. It is based on the proteolytic susceptibility of exposed and flexible regions when incubated with a relatively low concentration of proteases. Non aggregated Mtx after gel filtration at 1 mg/mL was incubated with trypsin at 1:10,000 (w/w). Aliquotus were removed after 10, 20, 30, 60 and 90 minutes followed by addition of 1 mM PMSF to
stop the reaction. Sample buffer was added to the aliquots and boiled at 95 °C for 5 minutes and were analysed on a 15% SDS PAGE.

2.2.11 Isothermal titration calorimetry (ITC)

Binding studies were performed by ITC to quantitatively investigate the interaction between various peptides and non aggregated Mtx deletion mutant (Freyer and Lewis, 2008). Peptides were dissolved in the superdex buffer at 8 mM concentration and 302 mL were titrated against 1.4 mL of protein at 0.15 mM in a VP-ITC Microcal Instrument (GE Healthcare). Titrations were carried out at 20 °C using 30 injections of 10 µL each injected at intervals of 200 seconds. The non-linear least square curve-fitting algorithm (Microcal Origin) was used for data fitting.

2.3 RESULTS

2.3.1 Bioinformatics analysis of full length Mtx

The sequence of Metaxin was obtained from NCBI (accession number NP_565446). BLAST searches of the Mtx full length amino acid sequence with sequences of proteins from the Protein Data Bank (PDB) showed that Mtx exhibited 33% sequence identity with SH3 domain of human Ras P21, however this identity covered only 11% of the query sequence. Further, a delta-BLAST search was performed and this experiment showed that the sequence of Mtx demonstrated 15% identity with Glutathione S-transferase (PDB: 1AW9) with an e-value of 6e-16. Thus it is anticipated that Mtx shares very low structure homology with the available structures in the PDB. The N-terminal region, which belongs to the GST-N family, is characterized by a diverse group of cytosolic dimeric proteins. The GST fold consists of an N-terminal thioredoxin-fold (TRX) domain and a C-terminal alpha helical domain. This TRX fold class contains the CXXC motif, which has been shown to interact with glutathione (Atkinson and Babbitt, 2009). The presence of this motif in Mtx in the form of CPNC (residues 25 - 28) raises the question whether the interaction with glutathione is possible. The secondary structure prediction using PSIPRED, showed that the protein comprises of predominantly alpha helices and coils (Figure 2-2A). Analysis of the Mtx sequence using FoldIndex and IUPred as predictors of intrinsic disorder indicated that the protein contains stretches of ordered and disordered sequences, although the assignments made by the two programs were substantially different, especially at the N-terminal region (Figure 2-2B-C). Residues 59 -76, 142 – 187, 237 - 284 and 289-315 were predicted to
be disordered by FoldIndex (Figure 2-2B). In case of IUPred, which estimates energy based on inter-residue interactions, residues 248 – 273 and 307 -315 were disordered (Figure 2-2C). Hence the C-terminal region as predicted by both the programs was disordered. According to the UniProt analysis there are two transmembrane helices encompassing residues 195 – 215 and 284 – 304, at the C-terminus (http://www.uniprot.org/uniprot/O64471) while the N-terminal region is predicted to be cytosolic. This agreed with the prediction performed by HMMTOP. No specific domain boundaries could be identified for the full length Mtx using bioinformatics analyses.
Chapter 2

2.3.2 Design, cloning and purification of deletion constructs

Recombinant over-expression of the full-length protein might lead to aggregation of protein as inclusion bodies in *E. coli*. With the aim to identify an intact and well-behaved soluble cytosolic domain and to avoid the complexity associated with the presence of the transmembrane regions of Mtx, a series of deletion constructs were designed based on the secondary structure prediction (Figure 2-3). As the C-terminal region of the protein comprises the transmembrane helices, truncations were designed
from this end and the N-terminal residues were kept intact. Initially clones of ΔMtx112, ΔMtx153 and ΔMtx191 were designed with and without a cleavable TEV site. However complete cleavage of TEV site could not be obtained because the purified constructs aggregated. Henceforth ΔMtx82, ΔMtx131, ΔMtx153, ΔMtx182, ΔMtx230, ΔMtx240 were designed containing N-terminal hexa-histidine tags without the incorporation of a TEV cleavage site. This effectively shortened the purification since TEV cleavage and removal of TEV and the cleaved peptide was no longer necessary and thus preventing aging of the purified protein, which could lead to aggregation. ΔMtx131 was purified as a dimer, which aggregated with time (as discussed below), hence further optimization of the construct was carried out. A number of residues on the either side of ΔMtx131 were deleted/added, to give ΔMtx127, ΔMtx136, ΔMtx138. Interestingly ΔMtx127 behaved similar to ΔMtx131 whereas ΔMtx136 and ΔMtx138 eluted as soluble aggregates. As the sequence of Mtx contained the CXXC motif at the N-terminus, reducing agents (dithiothreitol and β-mercaptoethanol) at concentrations of 5, 10 and 20 mM were incorporated in buffers for all the purification steps. No changes in the oligomerization propensity of ΔMtx131 compared to that purified without reducing agent were observed. Further in vitro studies were carried out using ΔMtx131.
2.3.3 Competitive import inhibition assay

In order to assess activity of the deletion constructs of Mtx expressed in *E. coli* and purified using Ni-NTA affinity chromatography, import inhibition studies were performed. Recombinantly expressed ΔMtx112, ΔMtx131 and ΔMtx191 constructs were used in the competition assay with wild type mitochondria to ascertain whether the recombinant protein is active and the regions of Mtx required to compete with the mitochondrial-located Mtx for the import of radiolabeled AOX and PiC (Figure 2-4). BSA was used as a negative control; wherein no inhibition of import of AOX and PiC
and the presence of matured product was observed by SDS PAGE. All of the Mtx deletion constructs inhibited import and were functionally active. Hence further characterization of each of the deletion mutants were performed.

**Figure 2-4** Competition import inhibition assay

*Competition import inhibition assay of purified recombinant Metaxin deletion constructs with radiolabeled precursors of phosphate carrier protein (PiC) and soya bean alternative oxidase (AOX). Except lane 1, all the samples are loaded in duplicates. Lane 1: no mitochondria; Lane 2 and 3: mitochondria and BSA (negative control); Lanes 4 and 5: with mitochondria and ΔMtx112; Lanes 6 and 7: with mitochondria and ΔMtx131; Lanes 8 and 9: with mitochondria and ΔMtx191.*

2.3.4 Size exclusion chromatography

In order to investigate the oligomeric state of the deletion constructs, size exclusion was performed after Ni-NTA affinity purification. It was observed that all of the deletion mutants of Mtx except ΔMtx131 were high order aggregates in solution as they eluted near the void volume of the superdex™75 column ([Figure 2-5A](#)). These observations did not change when a superdex™200 column was used, which has a higher MW cutoff than the former, suggesting that these constructs were not suitable for further biophysical analysis. ΔMtx131 existed as a mixture of dimers and aggregates. It was also observed that the aggregation peak fraction increased as the concentration of the injected protein increased suggesting that the aggregation is concentration dependent. The dimer fraction was pooled, concentrated and reinjected onto the column. Interestingly only one peak was observed which corresponded to the dimer ([Figure 2-5B](#)) suggesting that the aggregation is time dependent. In order to prevent the aggregation from affecting the biophysical studies, the dimeric form of ΔMtx131,
freshly eluted from the size exclusion chromatography column, was used immediately for further biophysical experiments. Further incubation with 2 mM glutathione did not change the oligomeric state of the protein.

**Figure 2-5** An analysis of oligomeric nature of Metaxin deletion constructs using size exclusion chromatography.

**A.** Overlay of elution profiles of Metaxin deletion constructs (different colored traces) obtained after size exclusion chromatography using a Superdex 75 prep-grade column with a void volume of 45 mL. Different kinds of stars show different oligomeric states. **B.** Dimer fraction of ΔMtx131 reloaded on the column immediately after size exclusion (magenta) and after storage at 10 mg/mL concentration for 1 day. Inset on the left shows the purified ΔMtx131 on a 15% SDS PAGE and that on the right shows Ve/Vo versus LogMW plot for superdex standards. Numbers 1-4 represent Bovine serum albumin, Mr 66,000, Ovalbumin Mr 44,000, Carbonic anhydrase Mr 29,000 and Lysozyme Mr 14,000; Alcohol dehydrogenase, Mr 150,000 was used to check the void volume Red triangle represents value for the non aggregated fraction of ΔMtx131.

### 2.3.5 Circular dichroism and thermal shift assay

To investigate the secondary structure of ΔMtx131, circular dichroism studies were performed. The secondary structure analysis of freshly prepared protein in phosphate buffer gave 24% helices, 27% strands, 20% turns and 29% unstructured regions (**Figure 2-6A**). The thermal melt profile of the protein showed a gradual non-sigmoidal signature, sloping downwards as the temperature increased from 10 – 90 °C suggesting that the protein was either initially very flexible, partially unfolded or a heterogeneous population of folded structures aggregated (**Figure 2-6A inset**). Furthermore, the CD spectrum of the protein before and after thermal denaturation remained identical.
In order to assess the effect of different buffers or salts in preventing aggregation thermal shift assays were performed. Thermal melt profiles with a range of buffers indicated that the initial values of fluorescence for the respective melt curves were high suggesting that the hydrophobic regions of the protein were exposed (Figure 2-6B) and as the temperature increased the protein aggregated due to intermolecular hydrophobic interactions, which caused quenching of fluorescence. Addition of salts at different concentrations to the test buffers did not lead to any significant change in the previously observed profiles. In summary, the ThermoFluor profiles of ΔMtx131 in different buffer and salt conditions were featureless suggesting that the construct exhibited unfolded/aggregated characteristics. These results further supported the thermal melt data obtained by circular dichroism (Figure 2-6A).
Chapter 2

Figure 2-6 Circular dichroism and thermal melt studies of ΔMtx131.

A. Circular dichroism spectrum of ΔMtx131 after analysis using DICHROWEB. The inset shows the non-cooperative thermal melt profile of the protein obtained by CD. B. SYPRO Orange based ThermoFluor assay of ΔMtx131 in buffers of different pH. Lysozyme powder was dissolved water at pH 7.0 and superdex buffer with the dye are used as positive and negative controls respectively. C. CD scan of ΔMtx131 at different temperatures while performing thermal melt. The higher value of positive elipticity seen at higher temperature cannot be considered because of the high value of machine high tension.

2.3.6 Dynamic light scattering and analytical ultracentrifugation studies

ΔMtx131 obtained freshly from size exclusion chromatography in superdex buffer was used for light scattering studies. This protein, at a concentration of 0.5 mg/mL, indicated a polydispersity of 15%, thus was considered to be monodisperse (Figure 2-7A). As size exclusion chromatography revealed time dependent aggregation, the level of polydispersity was re-analysed after storage for 24 hours at 4 °C. The sample revealed a significant level of polydispersity (50.8%) indicating that the protein exists as multiple species in solution (Figure 2-7B). Additionally sedimentation velocity AUC
Biophysical characterization of Metaxin

Experiments on freshly pooled ΔMtx131 dimer suggested that although the dimer fraction was predominant (63.6%), higher order oligomers were also present (Figure 2-7C). These small fractions of oligomers may act as nucleation sites, which can aid in the transition from dimer to a totally aggregated sample with time.

![Figure 2-7 Analysis of non-specific oligomerization in ΔMtx131 using DLS and AUC.](image)

**Figure 2-7** Analysis of non-specific oligomerization in ΔMtx131 using DLS and AUC.

A. DLS profile of 0.5 mg/mL of purified dimeric ΔMtx131 immediately after size exclusion chromatography. B. DLS profile of purified dimeric ΔMtx131, after storage at 4°C for 24 hours. C. Analytical ultracentrifugation profile of 1mg/mL of ΔMtx131 in superdex buffer. The inset shows the sedimentation velocity profile of the protein plotted as a function of absorbance at 230 nm versus the position from the axis of rotation at time interval of 300 sec.
2.3.7 Non-reducing PAGE and native PAGE analysis

In order to understand the effect of glutathione on ΔMtx131, increasing concentrations of glutathione was added to freshly purified protein and the mixture incubated overnight. Non-reducing PAGE analysis of the incubated samples did not reveal any dimer–monomer transition because of breakage of intra or inter chain disulphide bond (Figure 2-8).

Figure 2-8 Non-reducing PAGE analysis of ΔMtx131 with increasing concentration of glutathione. 
Red arrow shows the increasing concentration (mM) of glutathione.

To establish the role of detergents in prevention of the aggregation intermediates, ΔMtx131 dimer incubated with a range of detergents were electrophoresed by native PAGE after three days of storage at 4 °C. The aggregation decreased significantly in the presence of Chaps and Triton as compared to other detergents (Figure 2-A). The experiments were repeated after a prolonged incubation time of 3 weeks with Chaps and Triton only. Native gel analysis showed degradation along with aggregation suggesting that the detergents did not aid in preventing aggregation in ΔMtx131 (Figure 2-B). There might be proteases present along with the purified protein, which the protease inhibitor cocktail did not inhibit, thus causing degradation.
Biophysical characterization of Metaxin

Figure 2-9 Effect of detergents on preventing aggregation of purified ΔMtx131 with time.

A. Effect of detergents on freshly purified ΔMtx131 dimer fraction as visualized on a 10% native PAGE gel after incubation for 3 days. Lanes 1 and 9 are Ovalbumin (calculated pI = 4.5) and previously purified ΔMtx131 (calculated pI = 4.79) respectively. Lanes 2 – 8 consist of freshly purified ΔMtx131 incubated with 3 times CMC of Triton, Chaps, DM, DDM, TDM, DDAO and β-OG respectively. B. Effect of Triton (lane 2) and Chaps (lane 3) on the above dimer after storage for 3 weeks at 4 °C. Ovalbumin (lane 1) is used as a positive control.

2.3.8 Limited proteolysis

The dimer fraction of ΔMtx131 after gel filtration was analysed for the presence of a stable domain using limited proteolysis. The protein was digested using a low concentration of trypsin and no stable domain was observed after 90 minutes of incubation (Figure 2-). This could be due to the fact that the recombinantly expressed protein is not well folded.
Figure 2-10 Limited proteolysis of ΔMtx131 with trypsin.

Lanes with different samples are labelled as T (time) with subscript of time period in minutes. M corresponds to the molecular weight markers.

2.3.9 Isothermal titration calorimetry

ITC experiments were performed to understand the energetics of binding between presequences of AOX/ yeast CoxIV and freshly purified ΔMtx131 dimer (Figure 2-8). No interaction could be detected in the above cases. The concentration of protein could not be increased beyond 0.15 mM as previously concentration dependent aggregation was observed by size exclusion chromatography.
2.3.10 Theoretical modelling

As \( \Delta Mtx131 \) eluted as a dimer and also encompassed the region (residues 72 - 112) predicted to be essential for recognition of presequence (Lister et al., 2007a), modelling studies were performed to obtain a three dimensional structure. \textit{Ab initio} structure prediction of \( \Delta Mtx131 \) using I-TASSER generated 5 models. The C-score and the TM score of the best model predicted were -1.46 and 0.53 ± 0.15 respectively (Figure 2-9). Thus the statistics obtained for the modelled structures were outside acceptable confidence levels. This was likely due to the low sequence identity between Mtx and the sequences of the protein structures used for threading.
2.4 DISCUSSION

Metaxin from *Arabidopsis* is an integral membrane protein as predicted by bioinformatics analysis and has two C-terminal transmembrane helices that anchor the protein to the membrane. *In vivo* competition import inhibition experiments with recombinantly expressed N-terminal deletion constructs of Mtx suggest that the NiNTA purified constructs inhibit the import of AOX and PiC into the mitochondria, thus suggesting that they are functionally active. Size exclusion chromatography studies suggest that most of these deletion mutants form soluble higher order aggregates. Interestingly the size exclusion profile of ΔMtx131 showed the presence of dimer along with these higher order aggregates. The region encompassed by ΔMtx131 belonged to GST_N family and dimerization is predicted to be the characteristic feature of members of this family, which correlated with the above finding. The dimer fraction pooled after size exclusion retained its oligomeric state when reloaded immediately back onto the column. However the same dimer showed aggregation when reloaded after prolonged storage. This observation was confirmed using light scattering analysis. AUC experiments detected a small percentage of higher order oligomers in the freshly pooled dimer sample. Size exclusion chromatography as well as non-reducing PAGE experiments of ΔMtx131 with glutathione, did not show a change in state compared to that observed in its absence. We suggest that these higher order oligomers may act as nucleation for further aggregation with time and that this effect increases with concentration. Addition of detergents or change of buffer components did not help in preventing aggregation as observed from native PAGE and thermal shift experiments.
Although circular dichroism studies of ΔMtx131 showed the presence of secondary structural elements, limited proteolysis of with trypsin at 1:10,000 enzyme-substrate ratio (w/w) did not indicate the presence of any stable folded domain. Interaction analysis between presequences of AOX/CoxIV with ΔMtx131 using ITC did not confirm any interaction to support the results of in vivo import assay. Thus, the aggregation prone nature of ΔMtx131 hindered further in vitro structural and binding analysis. It may be possible that in vivo import inhibition could be a false positive, where the import of precursors were inhibited due to the presence of aggregates composed of partially folded/mis-folded/unfolded Mtx constructs.

As ΔMtx131 had low sequence similarity with protein structures available in the protein structure database, the models generated using I-TASSER had a low C score as well as TM score and were not reliable. Computational predictions indicated that residues 59 - 76 are disordered. Unstructured regions in proteins are both common and biologically important, particularly in modulating molecular recognition processes (Huang and Montelione, 2005, Dunker et al., 2005). Disordered regions can act as recognition elements and can undergo transitions from disorder to order on binding to specific ligands or binding partners (Dyson and Wright, 2005, Dyson and Wright, 2002, Demchenko, 2001). The presence of a disordered region in ΔMtx131 did not affect the biological function in vivo, suggesting that the recombinant protein is biologically active. However, additional factors present in the in vivo import inhibition assay may aid in inhibition of import. The purified recombinant protein tends to aggregate with time and concentration, which adversely affected the in vitro interaction studies. It is therefore difficult to conclude whether the aggregation or the absence of other factors affect the in vitro binding experiments. Interestingly, recombinant proteins can form functional aggregates e.g. aggregates of mitochondrial protein MAVS (Hou et al., 2011) potently activate transcription factor IRF3 to induce type I interferons.

2.5 FUTURE DIRECTIONS

The above biophysical studies were carried out in order to obtain a three dimensional structural insight of the cytosolic receptor domain of Metaxin. European Bioinformatics Institute (EBI)’s InterPro database shows that to date there is no high-resolution structure or biophysical characterization of Metaxin available from any organism
Recombinant expression of deletion constructs of Metaxin from *Arabidopsis* presented a formidable problem of aggregation, which hampered further biophysical or structural characterization. Chloride intracellular channel proteins (CLICs) contain a thioredoxin domain (similar to that of Metaxin) and an all helical C-terminal domain (Littler et al., 2010). Recombinantly purified CLIC1 demonstrates structural plasticity as it can adopt two conformations, one being globular and other forming a channel in the presence of an artificial bilayer. Furthermore, the crystal structure of CLIC1 suggests that the protein does not bind to glutathione non-covalently. Bearing in mind the similarities of the domain organization, it may be worthwhile to explore if the full length form of Metaxin can be purified using recombinant methods and if it can integrate into an artificial bilayer or a detergent micelle as has been observed for CLIC1.

The absence of a stable folded domain in recombinantly expressed Metaxin may suggest that its interaction with other members of the complex is necessary for its stability. A good example is the interaction of the natively unfolded viral protein genome linked (VPg) with the viral encoded movement protein (MP) which provides the specificity for transport of viral RNA from one cell to the other (Chowdhury and Savithri, 2011). Similarly, in the SNARE complex, where the recombinantly expressed proteins Snc1, lacking transmembrane region, and Sec9 is unstructured but on association with Sso1 undergo conformational transition into a stable complex essential for vesicle-membrane fusion, thus implying binding-induced structure formation (Fiebig et al., 1999). Hence, further investigation on other members of SAM complex and their association with Metaxin will aid in understanding the underlying import process. Metaxin could be used as bait in yeast two-hybrid experiments to identify different interacting partners from the SAM complex. Reports suggest that Metaxin2 from human had been identified using this approach (Armstrong et al., 1999).

Receptors have been shown to attained stable conformation on binding to their respective ligands. But these interactions are often weak exhibiting micromolar to millimolar affinities (Saitoh et al., 2011) as observed between cytosolic domain of Tom20 and rat aldehyde dehydrogenase (ALDH) presequence where no crystals could be obtained by mixing the two components (Saitoh et al. 2007). In the absence of diffracting crystals, structural studies using nuclear magnetic resonance (NMR) have
been successful as in Tom20 from *Arabidopsis* (Perry et al., 2006) and the Tom20 complex with rat ALDH presequence (Abe et al., 2000). Another strategy to trap weak or transient interactions, used successfully to gain structural insight into protein-peptide interactions, is the use of linker; such as chemical cross-linking as in case of rat Tom20 and rat ALDH presequence (Saitoh et al., 2007). Introducing a stretch of amino acids as linker between the protein and the peptide might help in stabilizing the interaction. This strategy has been carried out to stabilize complexes of Tom20 with the rat ALDH presequence (Obita et al., 2003) as well as intrinsically disordered neuron specific substrate protein neuromodulin and calmodulin (Chichili et al., 2013).

Thus we speculate that co expression of Metaxin or its deletion constructs with it partners from the SAM complex or the presequences might aid in obtaining a properly folded protein further assisting in biophysical and structural characterization.
Chapter 3

Ligand recognition by the TPR domain of the import factor Toc64 from *Arabidopsis thaliana*
3.1 ABSTRACT

The specific targeting of protein to organelles is achieved by targeting signals being recognised by their cognate receptors. Cytosolic chaperones, bound to precursor proteins, are recognized by specific receptors of the import machinery enabling transport into the specific organelle. The aim of this study was to gain greater insight into the mode of recognition of the C-termini of Hsp70 and Hsp90 chaperones by the Tetratricopeptide Repeat (TPR) domain of the chloroplast import receptor Toc64 from Arabidopsis thaliana (At). The monomeric TPR domain binds with 1:1 stoichiometry in similar micromolar affinity to both Hsp70 and Hsp90 as determined by isothermal titration calorimetry (ITC). Mutations of the terminal EEVD motif caused a profound decrease in affinity. Additionally, this study considered the contributions of residues upstream as alanine scanning experiments of these residues showed reduced binding affinity. Molecular dynamics simulations of the TPR domain helices upon peptide binding predicted that two helices within the TPR domain move backwards, exposing the cradle surface for interaction with the peptide. Our findings from ITC and molecular dynamics studies suggest that ArToc64_TPR does not discriminate between C-termini peptides of Hsp70 and Hsp90.

3.2 INTRODUCTION

Non-globular proteins, which contain repeated structural motifs arranged in tandem, are ubiquitous in nature. Many of these proteins have extended structures with exposed interaction surfaces. The common scaffold is known to interact with a variety of ligands (Grove et al., 2008) and the binding of the ligand does not induce any structural rearrangement (Morlot et al., 2007, Grove et al., 2008). Armadillo repeats (ARM), Ankyrins (Ank), Leucine-rich repeats (LRR), HEAT repeats and Tetratricopeptide repeats (TPR) are some examples of repeat protein (Grove et al., 2008, Boersma and Plückthun, 2011, Zeytuni and Zarivach, 2012). Structural and binding studies of the Hsp70 and Hsp90 organizing protein (Hop) (Brinker et al., 2002, Scheufler et al., 2000) have provided some knowledge of the interactions between the TPR domain and peptides, however, the role of conformational changes in these interactions has not been well characterized.
In this study, we focus on Tetratricopeptide Repeat (TPR) repeats and their versatility with respect to ligand recognition. These structural domains act as interaction scaffolds and mediators of multi-protein complexes and are found in all kingdoms of life (Allan and Ratajczak, 2011). TPR repeats form the 20 most common folds in the Pfam database (Kajander et al., 2005). They consist of multiple repeats of degenerate 34 amino acids, forming the canonical helix-turn-helix fold. Typically proteins with this motif contain 3-16 sequential TPR motifs, arranged in a tandem array. The anti-parallel packing of the helices forms a grooved surface with concave and convex faces, with ligand binding usually occurring within the concave surface (D'Andrea and Regan, 2003). TPR interaction with their ligand is usually specific. This specificity is provided by the unique geometry of the binding pocket and ionic, hydrogen bond and hydrophobic interactions between amino acids residues of the TPR and the ligand. TPR domains span a range of oligomeric states from monomers (Sampathkumar et al., 2008, Scheufler et al., 2000) to higher oligomers (Lunelli et al., 2009, Zeytuni et al., 2011, Zhang et al., 2010). Nuclear protein ssn6 (Schultz et al., 1990), chromatin associated protein CDC23 (Sikorski et al., 1990) and mitotic chromosome disjunction protein nuc2+ (Hirano et al., 1990) were among the first to be identified as TPR containing proteins. Furthermore, proteins with TPR domains play an essential role in the import of proteins into mitochondria (Baker et al., 2007), chloroplast (Mirus et al., 2009) and peroxisomes (Brocard and Hartig, 2006).

Proteins destined to the chloroplast are nuclear encoded, synthesized in the cytosol, and transported as precursor proteins (or preproteins), usually with NH₂-terminal targeting sequence called transit peptides (Li and Chiu, 2010). As both the mitochondria and chloroplast co-exist in plant cells, the sorting of protein between these organelles is a unique aspect of plant cell biology, and thus place a higher degree of stringency of protein targeting and sorting compared to non-plant cells. In chloroplasts, the process of recognition and translocation is initiated by subunits of the multimeric protein import complex called translocon at the outer envelope of chloroplast (Toc). This Toc core is comprised of the channel type, Toc75 (Hinnah et al., 2002), membrane anchored GTPases Toc159 (Smith et al., 2004), Toc34 (Koenig et al., 2008) and an integral membrane protein Toc64 (Sohrt and Soll, 2000). Topology analyses of Toc64 suggest that it consists of an amidase domain and a C-terminal TPR domain. The protein spans the membrane three times and positions the TPR domain facing the cytosol (Qbadou et
al., 2007, Li and Chiu, 2010). The TPR domain of Toc64 is composed of three TPR repeats followed by a solvation or capping helix (Figure 3-S1A) (Cheng et al., 2006). Association of Toc64 with preproteins was found to be chaperone mediated (Qbadou et al., 2007, Qbadou et al., 2006). Toc64 from Pisum sativum was found to interact with the C-terminal peptides of both Hsp70 and Hsp90 from wheat germ lysate, however, a preference was observed for interaction with the C-terminus of Hsp90 from human (Qbadou et al., 2007). In the above interaction studies, Toc64, Toc64_TPR, and the C-terminal peptide of Hsp90 were matrix immobilized. It was also observed from in vivo experiments that the TPR domain of Toc64 exhibited a stronger interaction with Hsp90 whereas the transmembrane region acts as substrate for Hsp70 (Qbadou et al., 2007). Thus, the TPR domain of Toc64 acts as a docking site, preferentially for Hsp90 bound preproteins, whereas Hsp70 was thought to be non-essential for interaction with preproteins (Qbadou et al., 2007, Qbadou et al., 2006). The EEVD motif in the C-terminus of Hsp70 and Hsp90 families is highly conserved in all eukaryotes. Although this motif anchors to the TPR domain by a dicarboxylate clamp mechanism, residues N-terminal to this sequence have also been reported to contribute towards the specificity of interaction through hydrophobic and van der Waals contacts as in case of Hop (Scheufler et al., 2000). The interaction analysis of the EEVD peptide with TPR domain showed a drastic reduction in binding affinity in Hop on removal of the above N terminal residues (Brinker et al., 2002).

In the present study, we characterized the interaction of the TPR domain from AtToc64 with C-terminal peptides from Hsp70 and Hsp90 by using biophysical approaches as a continuation of studies carried out by Qbadou S, et al. (Qbadou et al., 2006). Using isothermal titration calorimetry studies and molecular dynamics simulations, we have investigated the mode of interaction of AtToc64 with C-terminal octapeptides of Hsp70 and Hsp90. AtToc64_TPR interacts with the above octapeptides with similar affinity and 1:1 stoichiometry. Noticeably, point mutations of any of the last five residues of the peptides to alanine cause significant abolishment in peptide binding. Using molecular dynamics, we have delineated the contribution of residues involved in interactions and the dynamics mediating binding events. These studies showed that the Hsp70/90 octapeptide bound conformations are mediated by the outward movement of the terminal helices of the TPR, exposing the inner surface of the cradle for interaction with the peptide.
3.3 MATERIALS AND METHODS

3.3.1 Expression clone of the TPR domain of AtToc64

An expression clone for the TPR domain of AtToc64 was constructed using a cDNA encoding the full length protein (accession number NP_188424, At3g17970). The region encompassing the TPR domain of AtToc64 was composed of residues 474 – 589 (Figure 3-1). PCR amplification of this region used the forward primer (5’- TTACCATGGCCGAGATTGCCAAAGAGAAGGGTAA -3’, NcoI site underlined), and the reverse primer (5’- TTCACTCGAGCTGGAATTTCATCTCTCTCTGC -3’, XhoI site underlined). The PCR amplified product was digested with NcoI and XhoI and inserted into the pETM10 expression vector (Invitrogen), digested with the same restriction enzymes to generate the AtToc64_TPR_pETM10 expression plasmid encoding the TPR domain containing a C-terminal hexa-histidine tag (AtToc64_TPR-H6).

Figure 3-1 Topological model of Toc64 from Arabidopsis thaliana.

A. Domain organization in AtToc64 from the uniprot database (http://www.uniprot.org/uniprot/Q9LVH5#section_name). Different regions shown are helical transmembrane regions (dark grey), cytoplasmic exposed regions shown (light grey), regions exposed to the intermembrane space (pink) and TPR domain (green). The TPR domain of the protein expressed for this study is shown in green with a hexa-histidine tag at the C-terminus. B. Topological model of the domain organization. OM: outer membrane, IMS: Intermembrane space.
3.3.2 Expression in *E. coli*

Overexpression of the TPR sequence was performed in *E. coli* BL21(DE3) (Novagen). 1mg of plasmid DNA was used to transform 50 mL of competent cells. Single colonies were grown in Luria broth medium at 37 °C supplemented with 40 mg/µL of kanamycin (Amresco). Protein expression was induced by the addition of 1mM isopropyl- β-D-thiogalactopyranoside (IPTG). After 12 hours of induction at 20 °C, cells were harvested by centrifugation, and the pellet obtained was used for protein purification.

3.3.3 Protein purification

Recombinant expressed TPR domain was purified using nickel affinity chromatography (Ni-NTA). Briefly, 12 g of pellet was resuspended with 400 ml of lysis buffer (50 mM Hepes, 500 mM KCl, 10% glycerol, pH 7.3). The resuspended pellet was lysed using an Emulsiflex C5 high-pressure homogenizer (Avestin) at 13000 p.s.i. The lysate was clarified by centrifuging at 22,700 x g for 45 min followed by filtering the supernatant through a 0.22 µm membrane. The supernatant was loaded onto a 5 mL HisTrap FFcrude column using an ÄKTA purifier FPLC system (GE Healthcare). The Ni-NTA bound TPR was eluted using a concentration gradient of 40 mM to 1 M of imidazole (pH 7.3) and the concentration of the protein was quantified by measuring the absorbance at 280 nm using a calculated extinction coefficient of 1.09 M⁻¹ cm⁻¹. The protein purity was evaluated using standard SDS-PAGE analysis. The purified protein was concentrated using an Amicon centrifugal filter unit (30 kDa cutoff).

3.3.4 Relative molecular mass estimation by size exclusion chromatography

A superdex 75 prep grade size exclusion chromatography column (GE Healthcare) was equilibrated with superdex buffer (25 mM Hepes, 100 mM KCl, pH 7.3). The column was calibrated using gel filtration standards (Alcohol dehydrogenase, *M*, 150,000: void volume; Bovine serum albumin, *M*, 66,000 (1); Ovalbumin *M*, 44,000 (2), Carbonic anhydrase *M*, 29,000 (3); Lysozyme *M*, 14,000 (4). 2 mL of concentrated protein at 8 mg/mL was applied to the equilibrated column and eluted using the above superdex buffer. The relative molecular masses of the peaks obtained were calculated using a logarithmic interpolation. The elution peak corresponding to non-aggregated *At*Toc64_TPR-H6 was pooled and concentrated in the buffer systems required for various biophysical experiments.
3.3.5 Peptide synthesis

Synthetic C-terminal octapeptides to HsHsp70 (GPTIEEVD) and HsHsp90 (TSRMEEVD) were designed. In addition, shortened versions of these peptides were designed containing only the C-terminal pentapeptide sequences: Hsp70_C5 (IEEVD), Hsp90_C5 (MEEVD). Additionally, alanine mutations of each of the amino acids in the octapeptides were designed representing a library of 16 Ala-substituted peptides. Two random sequences of five residues (ASDTM) and eight residues (DMTSRGTQ) were also designed for negative control experiments. All the above peptides were synthesized and provided at >85% purity by Biomatik (Canada).

3.3.6 Circular Dichroism studies (CD)

Far-UV CD spectra corresponding to peptide bond absorption were recorded from 185 to 260 nm at 10 °C using a Jasco-810 spectropolarimeter. Spectra were collected for 10 μM of protein in 10 mM of sodium phosphate buffer pH 7.3 in a Quartz SUPRASIL cuvette (Hellma) with a path length of 1 mm. Measurements were made with an increment step of 0.5 nm, an integration time of 4 sec per step and a bandwidth of 2 nm. The signal due to buffer alone was subtracted from that of the protein.

The proportions of secondary structures of the protein was estimated from the [Θ] values between 190 and 240 nm using the DichroWeb server (Whitmore and Wallace, 2008) (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) and the CDSSTR algorithm (Compton and Johnson, 1986, Sreerama and Woody, 2000, Manavalan and Johnson, 1987)

The helical content of AtToc64_TPR-H6 was calculated from the molar ellipticity at 208 and 222 nm (Greenfield and Fasman, 1969) using the following equation:

\%α-helix = ([Θ]_{222} + 3000)/(36000 + 3000) \times 100

The change in molar ellipticity was monitored at 222 nm while varying the temperature from 10-80 °C at a rate of 1 °C/min. The thermally denatured sample was cooled to 10 °C at the same rate to observe the effect of temperature on the folding of recombinant expressed AtToc64_TPR-H6.
3.3.7 **Dynamic Light Scattering studies (DLS)**

DLS experiments were carried out with a Zetasizer Nano ZS (Malvern). 35 μM of protein in the superdex buffer was used for this study. Measurements were performed in triplicate consisting of 10 acquisitions per run.

3.3.8 **Analytical Ultra-centrifugation (AUC)**

These experiments were conducted on a Beckman XL-A analytical centrifuge, which was equipped with An60-Ti rotor and absorbance optics (Ralston). The experiments were carried out using 70 μM of protein samples in superdex buffer (390 μL) with superdex buffer (400 μL) as reference separately loaded into a double sector centrepiece and built up in the rotor. Protein samples were monitored continuously by UV absorbance at 280 nm. Prior to the start of centrifugation, the rotor was equilibrated to 20 °C and the vacuum was brought below 10 micron. The rotor was then set to spin at 60,000 rpm for 16 hrs. A total of 300 scans were collected. The acquired data were analysed using the SEDFIT program (Laue et al., 1992). The c(s) method implemented in the program was used for the data analysis, where c(s) is the sedimentation coefficient distribution function of the macromolecule. The physical parameters of the sample solution used for the data analysis were partial specific volume (0.74 mL/g), buffer density (1.005 g/mL) and viscosity (0.01002 P). This provides excellent resolution and sensitivity for characterizing sample homogeneity. A confidence level of p=0.95 was used while solving the size distributions.

3.3.9 **Isothermal Titration Calorimetry (ITC)**

Peptides were dissolved in the superdex buffer at 6 – 14 mM concentration and 302 μL were titrated against 1.4 ml of protein at 0.1 - 0.3 mM in a VP-ITC Microcal Instrument (GE Healthcare). Titrations were carried out at 20 °C using 30 injections of 10 μL each injected at interval of 200 seconds. Injections were continued beyond saturation levels to allow determination of the heats of ligand dilution. The non-linear least square curve-fitting algorithm (Microcal Origin) was used for data fitting. After subtraction of the heat of dilution, three floating variables: stoichiometry (N), binding constant (K_d) and change in enthalpy of interaction (ΔH) were obtained. For subsequent alanine scan experiments the stoichiometry was fixed at N=1.
3.3.10 Molecular dynamics simulation studies

3.3.10.1 Initial Geometry

A reliable model of the TPR domain for *in silico* studies was obtained by submitting the *AtToc64_TPR* protein sequence for automated protein structure modelling using the I-TASSER pipeline (Roy et al., 2011, Roy et al., 2010). A model was constructed using multiple threading alignments, which avoided bias towards a particular structural model as in homology modelling (Venclovas and Margelevicius, 2005). The molecular systems for protein-ligand complexes were built using the high-resolution crystal structure of 1ELW, the TPR domain of Hop in complex with the C-terminal octapeptide of Hsp70 (Scheufler et al., 2000) (obtained from the Protein Data Bank) as templates and a model obtained from the I-TASSER server. The TPR domains from Hop and *AtToc64* exhibited 50% sequence identity. The starting structures were refined in 200 independent FlexPepDock simulations, which consisted of a low-resolution pre-optimization step followed by a high-resolution refinement and high-resolution mode simulations using Rosetta FlexPepDock (London et al., 2011, Raveh et al., 2010). Thus two near native models of protein-peptide complexes were constructed and used for molecular dynamics simulations. In total, three molecular systems were prepared: *AtToc64_TPR* receptor (Apo), *AtToc64_TPR* receptor with C-term Hsp70 (octapeptide) bound (T_C70) and *AtToc64_TPR* receptor with C-term Hsp90 (octapeptide) bound (T_C90). Ramachandran plots (RAMPAGE) (Lovell et al., 2003), of the modelled structures are found in the supplementary materials.

3.3.10.2 Preparation for simulation

Graphics Processing Unit (GPU) accelerated Assisted Model Building with Energy Refinement (AMBER) suite version 12 (Goetz et al., 2012) associated with the latest all-atom ff12SB force field was used for simulation studies (Case et al., 2012). The starting structures were neutralized using Na⁺ and Cl⁻ ions and hydrogen atoms positioned using the *tleap* module from AMBERTOOLS12. The protein was centered in a solvent truncated octahedron box, which was made of TIP3P (3-point charged) triangulated water molecules with a 12 Å cut off in all directions (Jorgensen et al., 1983). The total number of atoms including water molecules was approximately 20,000 across various systems. The systems were minimized using a two-phase energy minimization approach which included 2500 cycles of steepest descent and 2500 cycles...
of conjugate gradient with solute atoms restrained by harmonic potentials with force constants of 50 kcal mol\(^{-1}\) Å\(^2\). This was followed by 5000 steps of unrestrained whole system minimization. 50 ps of density equilibration with weak harmonic restraints of 2 kcal mol\(^{-1}\) Å\(^2\) on the solute molecule was performed followed by restrained equilibration for 500 ps under constant pressure and temperature conditions. All simulations were performed using the SHAKE algorithm (Ryckaert et al., 1977) with constraints on hydrogen-linked bonds (allowing a tolerance of 0.0001). To evaluate long-range electrostatic interactions, the Particle Mesh Ewald (PME) method (Salomon-Ferrer et al., 2013) was used with a cut-off of 9 Å. An integration time step of 2 fs was used to numerically solve Newton’s equations of motion. Langevin dynamics was used to maintain a constant temperature of 300 K throughout the simulations. All the simulations were performed using the PMEMD module in AMBER12. Production runs were carried out for 50 ns in an explicit solvent environment and with an isothermal-isobaric (NPT) ensemble.

### 3.3.10.3 Analysis of trajectory

After the production runs were completed, each of the trajectories were analyzed based on variation in kinetic and potential energies using the *ptraj* program. Root-mean-square deviation (rmsd) and atomic positional fluctuation per residue (rmsf) were analyzed to understand the overall conformational changes throughout the trajectory. The number of hydrogen and other non-bonded bond interactions were calculated using HBPLUS (McDonald and Thornton, 1994). Interactions of arginine or lysine side chains with oxygen atoms of aspartate or glutamate residues which were from 3 – 4 Å in distance and exhibited angles (donor – H – acceptor) less than 90° were not classified as hydrogen bonds, but considered as electrostatic interactions. In addition, the solvent accessible surface area and binding energies for the protein-peptide interactions were calculated using NACCESS (Hubbard and Thornton, 1993) and MM/PB(GB)SA tools (Kollman et al., 2000, Srinivasan et al., 1998) respectively. Computational Alanine scanning was performed with the peptide, in order to understand the contribution of individual amino acid residues towards binding. When calculating the difference in free energies (ΔΔG) between the wild type and the mutants, the results of using Generalized Born calculations (GB) were taken into consideration, as they are well suited for protein-ligand and protein-protein interaction calculations. These values were calculated according to the following equation:

\[
\Delta \Delta G = \Delta G^{\text{wild}} - \Delta G^{\text{mutant}}
\]
For all of the above analyses, the last 5ns of the trajectory was analyzed. The APBS software was used to compute the electrostatic potential surface (Baker et al., 2001). Ligplot was used to map the hydrogen and hydrophobic bonding patterns between the peptide residues and the residues of TPR involved in interaction (Wallace et al., 1995). Visual Molecular Dynamics (Humphrey et al., 1996) was used for visualizing the trajectories of the simulations of the three systems. The dictionary of protein secondary structure (DSSP) (Kabsch and Sander, 1983) was used to assign the secondary structure per amino acid along different trajectories.

Principal Component Analysis (PCA) was also performed in order to substantiate the above results (Amadei et al., 1993). GROMOS 4.5 (Pronk et al., 2013) was used on the protein backbone atoms (Cα) to obtain the collective coordinates for protein motions from covariance matrices in the form of sets of eigen values (amount of motion) and eigen vectors (direction of motion). In this way all the linearly correlated motions were analysed. Correlated atomic motion in the apo and ligand bound forms were obtained by analyzing the dynamical cross correlation map (DCCM) of Cα atoms using Bio3D (Grant et al., 2006, Ichiye and Karplus, 1991). This provided a means to understand the correlation of motions of neighbouring and/or distant residues (Harte et al., 1990). All graphical representations were generated using PyMOL (DeLano, 2002).

3.4 RESULTS

3.4.1 Purification and characterization of TPR domain of ArToc64

Recombinant ArToc64_TPR-H6 (Figure 3-1) was cloned, over-expressed and purified using Ni-NTA affinity chromatography. The purity was accessed by SDS-PAGE where a single band, corresponding to 14.3 kDa, was observed (Figure 3-S1B). The observed molecular mass agreed with that predicted from the sequence of his-tagged ArToc64_TPR. Further size exclusion chromatography was performed which showed that the protein exists predominantly as a dimer calculated from gel filtration protein standards (Figure 3-2A). The dimer fractions were pooled, concentrated and used for further studies. The addition of DTT (2 mM and 5 mM) to the purification buffers did not change the oligomeric state of the protein as analysed using size exclusion
Ligand recognition by the TPR domain of Toc64

chromatography. Based on these results, we hypothesize that the TPR domain is either globular and dimeric or elongated and monomeric.

The TPR repeats are known to adopt an alpha-helical solenoid structure; this is supported by analysis of the protein sequence by PSIPRED (Buchan et al., 2010, Jones, 1999) (Figure 3-2B). Circular dichroism spectra can be used to estimate the regular secondary structural elements in a protein. The spectral signature revealed that protein adopted a predominantly (80%) helical structure (Figure 3-2C), in agreement with that observed in the PSIPRED prediction. Thermal denaturation experiments showed that the protein had a T\text{m} of \(-35\) °C (Figure 3-S1C).

![Figure 3-2 Biophysical characterization of AtToc64_TPR-H6 using size exclusion chromatography and circular dichroism.](image)

**Figure 3-2** Biophysical characterization of AtToc64_TPR-H6 using size exclusion chromatography and circular dichroism.

A. Elution profile of AtToc64_TPR-H6 from size exclusion chromatography using a Superdex 75 prep grade column. The different shaped stars represent different forms of the protein as
described in the figure. The Ve/Vo versus LogMW plot for superdex standards is shown as an inset. Numbers 1 - 4 represent different standard proteins used as described in the methods. Red triangle represents AtToc64_TPR-H6. B. Schematic representation of secondary structure of AtToc64_TPR. C. CD spectrum profile of the protein obtained after analysis with DICHROWEB.

3.4.2 Estimation of polydispersity by DLS and AUC

In order to understand the oligomeric state of the protein, DLS studies were performed. This analysis on a 35 μM solution of freshly purified recombinant ArToc64_TPR-H6 showed that the protein was predominantly monodispersed with a polydispersity value of 13% (Figure 3-3A). Storage of the protein at 4 °C for 24 hours caused an increase in the extent of polydispersity (52%) as compared to the freshly purified sample (Figure 3-4B).

Figure 3-5 Biophysical characterization of ArToc64_TPR-H6 using DLS and AUC.  
A. DLS profile of freshly purified ArToc64_TPR-H6. B. DLS profile of ArToc64_TPR-H6 after 24 hours of storage at 4 °C. C. AUC profile of ArToc64_TPR-H6. Top panel shows the sedimentation velocity profile of the protein plotted as a function of absorbance at 230 nm versus the position from the axis of rotation at time interval of 300 sec. Solid lines represent nonlinear best-fits to a continuous sedimentation distribution model. Bottom panel shows c(s) distribution for the protein at confidence level (P) of 0.95. D. AUC profile of ArToc64_TPR-H6.
incubated with octapeptide of Hsp70 for overnight. E. AUC profile of AtToc64_TPR-H6 incubated with octapeptide of Hsp90 for overnight.

AUC experiments were carried out to independently establish the homogeneity of protein molecules in solution and determine their molecular mass (Lebowitz et al., 2002). The protein existed as a mixture of monomeric and dimeric forms (Figure 3-3C), where the former was predominantly present (Table 3-1). These results also explained the observed sample polydispersity by DLS; this polydispersity was irreversible as neither a second round of gel filtration chromatography or sample dilution succeeded in yielding monomeric species. Additionally, AUC experiments were performed in combination with the two peptides (octapeptides) (Figure 3-3D and E). Interestingly, it was observed that in all cases the monomeric form was predominant (Table 3-1). The result of the AUC studies suggests that the behaviour of AtToc64_TPR-H6 by size exclusion chromatography is likely due to the non-globular, ellipsoid shape and not due to the presence of dimers in solution. The higher oligomers observed in the AUC is likely due to nonspecific self-association of protein molecules with time.

Table 3-1 Analytical ultracentrifugation.

<table>
<thead>
<tr>
<th>Components</th>
<th>Sedimentation Coefficient</th>
<th>Molecular mass (kDa)</th>
<th>Monomer %</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtToc64_TPR</td>
<td>1.60 ± 0.17</td>
<td>15 ± 2</td>
<td>67</td>
</tr>
<tr>
<td>AtToc64_TPR with C-terminal Hsp70 octapeptide</td>
<td>1.90 ± 0.30</td>
<td>**</td>
<td>83</td>
</tr>
<tr>
<td>AtToc64_TPR with C-terminal Hsp90 octapeptide</td>
<td>1.90 ± 0.20</td>
<td>14.9 ± 2</td>
<td>71</td>
</tr>
</tbody>
</table>

Analytical ultracentrifugation experiments were performed at 20 °C in superdex buffer. As the frictional coefficient was high ($f/f_0 = 1.54$) in the Hsp70 bound form hence adequate determination of the molecular mass was not possible. However the sedimentation coefficient obtained suggests that it is a monomer.

3.4.3 ITC characterization of peptide interaction with AtToc64_TPR-H6

In order to understand the energetics of binding between the C-termini of Hsp70/Hsp90 and the TPR domain of AtToc64, binding studies using ITC were performed (Campoy and Freire, 2006, Freyer and Lewis, 2008). The curves obtained were fit using a 1:1
binding model. The binding stoichiometries (n) were found to be 1.2 and 0.94, when protein at 300 μM in the superdex buffer was titrated with 10 mM Hsp70 octapeptide or 14 mM Hsp90 octapeptide solutions respectively at 20 °C. Similar experiments were carried with pentapeptide versions of Hsp70 and Hsp90, wherein n values were 1.2 and 1.04 respectively. Hence in all of the above cases, the thermodynamics of binding events exhibited exothermic behaviour. The heat of reaction per injection was calculated from the area under the peak, which gradually decreased with complex formation and reached the heat of dilution of the respective peptide when the protein was saturated. After ~20 and ~11 injections in the case of Hsp70 and Hsp90 respectively, the binding events reached their saturation. The thermodynamic parameters and the final titration curve (Figure 3-4) were computed, resulting in similar binding affinities for the two octapeptides to TPR. Reducing the length of the peptide to the C-terminal five residues containing the EEVD motif decreased the binding affinity two and half fold in the case of Hsp70 and four fold in the case of Hsp90 (Table 3-2).

Figure 3-4 Binding isotherms for interaction of ArToc64_TPR-H6 with Hsp70 and Hsp90.

The ITC isotherms obtained for the Hsp70 octapeptide binding to ArToc64_TPR-H6 (in green) and the Hsp90 octapeptide binding to ArToc64_TPR-H6 (in magenta). The bottom panels show the curves obtained for titration of the octapeptides of Hsp70/90 into TPR (triangles); the C-
terminal pentapeptides of Hsp70/90 into TPR (diamonds) and random peptides into TPR (crosses).
The Gibb’s free energy (ΔG) of the above binding events were similar in magnitude to each other, ranging from -4.90 to -4.94 kcal/mol. In the case of the octapeptide of Hsp90, a favourable enthalpic contribution to binding was observed (ΔH = -6.18 kcal/mol) with a decrease in entropy (ΔS = -4.33 cal/mol•K). This is characteristic of an enthalpy driven process for binding of the peptide to the protein, where possible changes in conformation of one or both components in the binding event is expected (Diehl et al., 2010). In the case of the octapeptide of Hsp70, the enthalpic contribution to the binding event was quite low (ΔH = -1.0 kcal/mol) accompanied by an increase in entropy (ΔS = 13.2 cal/mol•K). An increase in entropy was also seen in the binding of the pentapeptide of both Hsp70 (ΔS = 9.29 cal/mol•K) and Hsp90 (ΔS = 5.03 cal/mol•K), suggesting an entropy driven event which may be due to a combination of the release of water molecules from the binding site upon ligand binding and a classical hydrophobic effect (Fischer et al.,1999). In summary, a favourable enthalpic binding process, in spite of a decrease in entropy, have aided the binding of the protein with the Hsp90 octapeptide; whereas a favourable entropic process, despite a decrease in enthalpy have aided the binding of protein with Hsp70 octapeptide.

**Table 3-2** Thermodynamic parameters obtained for TPR-Hsp interaction using isothermal titration calorimetry.

<table>
<thead>
<tr>
<th>Systems</th>
<th>N</th>
<th>ΔG (kcal/mol)</th>
<th>ΔH (kcal/mol)</th>
<th>TΔS (kcal/mol)</th>
<th>Kd (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T_C70_C5</td>
<td>0.93 +/-0.13</td>
<td>-4.37</td>
<td>-1.63</td>
<td>2.72</td>
<td>568</td>
</tr>
<tr>
<td>T_C70_C8</td>
<td>1.2 +/-0.12</td>
<td>-4.90</td>
<td>-1.00</td>
<td>3.86</td>
<td>230</td>
</tr>
<tr>
<td>T_C90_C5</td>
<td>1.04 +/-0.13</td>
<td>-4.17</td>
<td>-2.68</td>
<td>1.47</td>
<td>800</td>
</tr>
<tr>
<td>T_C90_C8</td>
<td>0.94 +/-0.04</td>
<td>-4.94</td>
<td>-6.18</td>
<td>-1.26</td>
<td>218</td>
</tr>
</tbody>
</table>

*N* is the stoichiometry; ΔG is the calculated change in Gibb’s free energy; ΔH is the change in enthalpy; ΔS in the change in entropy and Kd is the binding affinity. The reported values of ΔH and Kd have an error of ±5%
3.4.4 Experimental alanine scanning

The EEVD motifs in the peptides have been shown previously to act as an anchor to the protein (Brinker et al., 2002). As both the octapeptides bound to the protein with similar affinity, it was of interest to investigate the contribution of each residue towards the interaction. For this work alanine scanning mutagenesis was carried out on each of the octapeptides used for the study and the binding was characterized by ITC studies (Figure 3-). For this description, the N-terminal Gly residue is termed Gly\(^{117}\), and the following residues are arranged in ascending order, e.g. Pro\(^{118}\), Thr\(^{119}\), Ile\(^{120}\), Glu\(^{121}\), Glu\(^{122}\), Val\(^{123}\), Asp\(^{124}\), in case of Hsp70 and a similar nomenclature for Hsp90 is used. Alanine substitution of Asp\(^{124}\) abolished binding in Hsp70 and caused a significant decrease in case of Hsp90, suggesting the occurrence of electrostatic interactions. A strong contribution of hydrophobic interactions to the binding was hypothesized because of a significant increase in \(K_d\) or the complete loss of binding in the case of Ile\(^{120}\)/Met\(^{120}\) and Val\(^{123}\) respectively. As expected, mutation of Glu\(^{121}\) and Glu\(^{122}\) to Alanine abolished or significantly reduced the binding affinity for Hsp70 and Hsp90 respectively. Mutation of the first four N-terminal residues in both the peptides to Alanine caused a decrease in binding affinity with the exception of the Arg\(^{119}\) in case of Hsp90 (Figure 3-5, Table 3-S1).

![Figure 3-5](image)

**Figure 3-5** The ITC binding data for alanine scanning mutagenesis of the peptide interaction with AtToc64_TPR-H6.

A. Data using the C-terminal octapeptide from Hsp70. B. Data using the C-terminal octapeptide of Hsp90. Asterik (*) suggests that the heat change during binding event is quite low and the signal to noise ratio is high. Hence the \(K_d\) is considered as no binding.
3.4.5 Simulation studies

The quality of the models generated were checked using Ramachandran plot where ~94% of the residues were present in the favoured region (Figure 3-S3-S5). Further the RMSD between the ArToc64_TPR modelled using I-TASSER and the TPR domain of Hop (1ELW) was found to be 1.4 Å (Figure 3-S2). Three systems were simulated based on the isothermal and isobaric (NPT) ensemble for 50 ns each: apo ArToc64_TPR (apo), ArToc64_TPR with the octapeptide of Hsp70 bound (T_C70) and ArToc64_TPR with the octapeptide of Hsp90 bound (T_C90). The RMS fluctuation for each system at the end of the equilibration was minimal. The overall structural stability of the systems throughout the simulation was assessed by calculating the RMSD of the Cα atoms from the appropriate starting structures for each simulation. The average Cα-RMSD were found to be 2.15 Å, 1.38 Å, and 1.39 Å for the apo, T_C70 and T_C90 systems respectively (Figure 3-6A). This suggested that the complexed structures were stable and retained their overall structure during the simulation. Further analysis of the three independent simulations for each of the systems indicated that, in all cases, the cradle topology of the TPR was maintained. An analysis of Cα atomic positional fluctuation (RMSF) for each of the systems showed that the N and C termini exhibited higher temperature factors (B factor). Additionally, higher B factors were also observed for TPR residues 29-33 in the T_C70 system, corresponding to a loop connecting the H2 and H3 helices (Figure 3-B). Finally, the RMSF of the terminal residues were decreased relatively upon ligand binding.
Figure 3-6 Analysis of the molecular dynamics trajectory obtained after 50 ns of simulation.

A. Root mean square deviation (RMSD) plot for the apo TPR receptor (black trace), the Hsp70 C-terminal octapeptide bound form of the receptor (T_70: green trace) and the Hsp90 C-terminal octapeptide bound form of the receptor (T_C90: magenta trace). B. The atomic positional fluctuation (RMSF) plot obtained for each of the above systems. The loop with high B factors in T_C70 is shown with a green arrow.

3.4.6 Analysis of protein-peptide interactions

The concave surface of the TPR cradle was crucial for peptide binding as observed in the average structures (Figure 3-A and B). This is exemplified through intermolecular hydrogen bonding and nonbonding interactions with the residues that line the inner surface of the cradle. Approximately 15 interactions of each type are found in both of the peptide bound forms (Figure 3-C). Notably, two types of hydrogen bonding interactions existed on the protein-peptide interface: sequence specific interactions, which involved the side chains of the peptide residues and sequence independent interactions, which involved the peptide main chain. Intra-residue interactions are also observed and are likely to be necessary to maintain the proper conformation of the peptides necessary for interactions with the protein.
Figure 3-7 Characterization of the average structures obtained after simulation of the complexes.

A. The TPR domain with bound octapeptide from Hsp70 (green) B. The TPR domain with bound octapeptide from Hsp90 (magenta) C. Number of interactions in the T_C70 system (green) and in the T_C90 system (magenta) during the simulations.

3.4.7 Carboxylate clamp

Interaction analysis of the crystal structure of TPR domains from Hop liganded with the C-terminal octapeptide and pentapeptide of Hsp70 and Hsp90 respectively, have shown that the carboxylate moieties of the highly conserved terminal Asp residue of the peptides form electrostatic interactions with the protein residue (Scheufler et al., 2000). This has been referred to as the two-carboxylate clamp. The conserved Asp^{124} in both the complexes are clamped to the TPR by the terminal carboxylate moieties and held in place by a myriad of interactions to the protein (Figure 3-).
Figure 3-8 Interactions occurring at the protein-peptide interface generated by Ligplot.

A. Interactions between the TPR domain and the C-terminal octapeptide of Hsp70; B. Interactions between the TPR domain and C-terminal octapeptide of Hsp90. The peptide is shown in purple bonds and the protein in brown bonds.

In case of the T_C70 system, the two carboxylates are held in place by hydrogen bonding interaction with Lys^5 and Asn^9 of helix H1. Additionally, the main chain carboxylate is stabilized by interaction with Asn^40 of helix H3 whereas Thr^36 of helix H3 and Lys^70 of helix H5 help in stabilizing the side chain carboxylate (Figure 3-A). In the T_C90 system, the terminal carboxylate of Asp^{124} are clamped by hydrogen bonding interactions with Lys^5, Asn^9 of helix H1 and Asn^40 of helix H3 (Figure 3-B). Thus the carboxylate clamp, which is highly conserved in both the systems, anchors the peptides to the TPR.
3.4.8 Other key hydrogen bonding interactions

The hydrogen bonding interactions from the TPR domain to the peptide in each complex are directed predominantly towards side chains, hence exploit sequence-specific features. Residues N terminal to Glu^{121} of the C-term Hsp70 peptide do not display any hydrogen bonding interaction with the TPR. In T_C70 system, Ser^{118} forms hydrogen-bonding interaction with Ser^{108} and Arg^{111} and Met^{120} displays hydrogen bonding interaction with Arg^{104}. Side chain hydrogen bonding interactions of Glu^{121} exist with side chains of Arg^{74}/Arg^{74}, Thr^{77}/Asn^{102} and Arg^{104}/Arg^{104}) in the T_C70/T_C90 systems respectively (Figure 3-). The carboxylate oxygen atoms of Glu^{122} form hydrogen bond acceptors for Lys^{70}, Arg^{74} and Arg^{104} nitrogen containing groups in both the systems and additionally with Arg^{119} in the T_C90 system. Val^{123} acts as hydrogen bond acceptor for Lys^{70}, exclusively in T_C90 system.

3.4.9 A groove for Valine

Among the charged residues of the EEVD motif in the peptides, valine is the only hydrophobic residue. The Val^{123} is held in place by a number of van der Waals and hydrophobic interaction with different residues of the TPR domain, which form a groove. The interacting residues forming van der Waals contacts are Asn^{9} and Asn^{40} in case of T_C70 system and Asn^{40}, Lys^{70} and Arg^{74} in case of T_C90 system. Similarly, Phe^{12} and Ala^{43} in case of T_C70 and Phe^{12} in case of T_C90 form hydrophobic interactions with Val^{123}. 

**Figure 3-9** The carboxylate clamp. 

A. Key interacting residues of the TPR domain with the carboxylate of Asp^{124} of Hsp70 (shown in green bonds); B. Key interacting residues of the TPR domain with the carboxylate of Asp^{124} of Hsp90 (shown in magenta bonds). The oxygen atoms of the carboxylate moieties are labelled. The TPR domain is shown with grey bonds.
3.4.10 Intrapeptide interactions

Intrapeptide interactions play a key role in giving a particular conformation to the peptide. This helps in the proper presentation of the peptide residues, which aid in interaction to the protein (Figure 3-10). In the T_C70 system, the key intrapeptide hydrogen bonding interactions are as below: the main chain nitrogen of Ile\textsuperscript{120} with main chain oxygen of Pro\textsuperscript{118}; the main chain nitrogen of Val\textsuperscript{123} with main chain oxygen of Glu\textsuperscript{121} and the main chain nitrogen of Glu\textsuperscript{122} with side chain oxygen of Glu\textsuperscript{122}. Similarly in the T_C90 system, the key intrapeptide hydrogen bonding interactions are: the amino group of Arg\textsuperscript{119} with the side chain oxygen of Glu\textsuperscript{122} and the main chain nitrogen of Glu\textsuperscript{122} with its own side chain oxygen.

![Key intrapeptide interactions](image)

**Figure 3-10** Key intrapeptide interactions.

A. Intrapeptide interaction in the T_C70 system (coloured green) B. Intrapeptide interaction in the T_C90 system (coloured magenta). The TPR domain is shown as a ribbon diagram in grey.

3.4.11 Computational alanine scanning

Though absolute free energies from simulations are not achievable (Deng and Roux, 2009), the underlying trend caused due to mutation has been reliably captured. Computational Alanine scanning was performed on the trajectories of the two complex systems. Six of the eight residues in the each of the peptides were mutated individually to Alanine. The terminal amino acids could not be mutated by MMPB/GBSA due to limitation of the program. The calculated binding free energies for the Hsp70 and Hsp90 liganded systems were found to be -12.6 and -13.8 kcal/mol respectively. Mutations of Val\textsuperscript{123}, Glu\textsuperscript{122}, Glu\textsuperscript{121}, Met\textsuperscript{120} and Ile\textsuperscript{120} individually to Ala caused a decrease in binding, which was in accordance with our experimental ITC findings (Table 3-3). Alanine mutation of Arg\textsuperscript{119} slightly improved binding as observed in our experiments. In contrast, mutation of Pro\textsuperscript{118} and Ser\textsuperscript{118} to Alanine had only a minimal effect on binding, a result that differed from the ITC results. The slight change in
conformation of the peptides caused by the mutations may have resulted in changes to the binding affinity; these have not been taken into consideration by the computational alanine scanning studies. However, the differences found in the latter case was quite small, and hence the trend of change in binding free energies could be considered acceptable. In all computational studies the secondary structure of the protein is conserved all throughout the simulation and there is very little change between alpha, pi and $3_{10}$ helices (Figure 3-S6).

Table 3-3 Computational alanine scanning for both ligand bound systems.

<table>
<thead>
<tr>
<th></th>
<th>ΔΔG</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hsp70 mutants</strong></td>
<td></td>
</tr>
<tr>
<td>GATIEEVD</td>
<td>-0.39</td>
</tr>
<tr>
<td>GPAIEEVD</td>
<td>-0.04</td>
</tr>
<tr>
<td>GPTAEED</td>
<td>-2.01</td>
</tr>
<tr>
<td>GPTIAEVD</td>
<td>-17.60</td>
</tr>
<tr>
<td>GPTIEAVD</td>
<td>-15.90</td>
</tr>
<tr>
<td>GPTIEEAD</td>
<td>-3.66</td>
</tr>
<tr>
<td><strong>Hsp90 Mutants</strong></td>
<td></td>
</tr>
<tr>
<td>TARMEEVD</td>
<td>-0.14</td>
</tr>
<tr>
<td>TSAMEEVD</td>
<td>1.21</td>
</tr>
<tr>
<td>TSCANEED</td>
<td>-3.08</td>
</tr>
<tr>
<td>TSRMAEVD</td>
<td>-12.00</td>
</tr>
<tr>
<td>TSMEEAVD</td>
<td>-3.37</td>
</tr>
<tr>
<td>TSMMAEAD</td>
<td>-3.97</td>
</tr>
</tbody>
</table>

Point mutations to alanine were performed computationally for each residue of the Hsp70 and Hsp90 octapeptides. Changes in binding free energy (ΔΔG) in kcal/mol is calculated using MMGBSA.

3.4.12 Dynamic cross-correlation map (DCCM)

Dynamic cross-correlation maps suggest that distinct correlation exist between the helices (Figure 3-S7). In the ligand unbound form, the majority of residues show a positive correlation between H1 and H2; H3 and H4; H4 and H5 (Figure 3-S7A). Fewer residues show positive correlation between H5 and H6. Negative correlated motion is observed in H1 with respect to H4 and H6. Also the segment, residues between 40-60, which comprises of a part of H3, a loop and a part of H4 show negative correlation with H7. Similarly, H5 shows negative correlation with H7. Ligand binding reduces all the observed inter helical positive correlations, further enhancing negative correlated
motions (Figure 3-S7B-C). In T_C70, additional negative correlation is seen between H6 and the region consisting of residues 60 – 70 (Figure 3-S7B).

3.4.13 Principal component analysis (Essential dynamics)

Although DCCM, is a good way to analyze motions between the pair of atoms, the complexity of collective principal atomic motion could not be visualized. To understand this, essential dynamics was performed. The overall protein motions were decomposed into a set of eight eigenvectors. The first two eigenvectors (most significant principal components, PC1 and PC2) accounted for ~65% of the total fluctuation. The 3D plot of these two components projected, with the potential energy on the z-axis, is represented as a 2D plot (Figure 3-A-C). The plots depict that there is clearly more conformational space sampled in the case of Apo form. The dominant motions in the apo form are confined to the terminal residues of the H1 and H7 helices (Figure 3-D), which correlate with the rmsd and positional fluctuation plots. The distributions of both the complexes are smaller than that observed in the apo form. In the T_C90 system, the conformational spread is more restricted (Figure 3-C) than observed for the T_C70 system (Figure 3-B). An analysis of the individual principal components in the complexes suggests that protein in the peptide bound forms do not show loss of alpha helicity and suggests that peptide binding may act to stabilize the TPR. The most prominent motions are observed in the H1 and H7 helices with very little motion in concave inner surface of the TPR cradle (Figure 3-D-F).
Ligand recognition by the TPR domain of Toc64

3.5 DISCUSSION

In the present study, we purified recombinant AtToc64_TPR-H6 to homogeneity. Analytical ultracentrifugation experiments estimate that the TPR domain exists predominantly as a monomer. Isothermal calorimetry studies indicate that the domain interacts with the C-terminal regions of Hsp70 and Hsp90 with similar micromolar affinity and 1:1 stoichiometry. Molecular dynamics studies have been performed to provide atomic level descriptions for the protein–peptide interactions. The terminal aspartate in both Hsp70 and Hsp90, is anchored to the TPR by a dicarboxylate clamp, supporting previous findings (Scheufler et al., 2000). Electrostatic potential surface representations of the ligand bound form of AtToc-TPR shows that the peptides are bound to a predominantly positively charged surface within the cradle of the TPR domain. In the case of Hsp70, the N-terminal residues are exposed and not interacting with the protein (Figure 3-S8). Alanine scanning mutations of residues 120 to 124 in both peptides significantly perturb binding to the TPR domain. Intramolecular hydrogen bonding interactions between different residues of the peptides are observed, which
might be necessary for providing suitable conformations to the peptides needed for binding to the receptor. Though hydrogen bonds, electrostatic and hydrophobic contacts exist, van der Waals interactions play a major role in positioning the EEVD motif. Alanine scanning mutation of either Gly\textsuperscript{117} or Pro\textsuperscript{118} using ITC reduces the binding affinity of the peptide, however the MD studies do not show that these residues interact specifically with the TPR domain. A possible explanation for this observation may be that residues contribute to the ideal conformation of the peptide necessary for interaction with the protein or they can prevent sampling of unwanted conformational space by the peptide. Proline commonly adopts a cis conformation. Glycine, as a small and achiral amino acid, can occupy a larger volume of conformational space without unfavourable steric interactions with other amino acids. Because of these unique features, glycine and proline residues often occur in turns and loops (Krieger et al., 2005, Trevino et al., 2007).

Using MMPB/GBSA for computational alanine scanning, the change in binding energy due to mutation of Gly\textsuperscript{117} and Pro\textsuperscript{118} to alanine could not be captured. The plausible reason for the above observation is that in MD experiments, the peptide is initially docked to the protein. The binding energy change upon mutation can be reliably captured for residues of the peptide, which directly interact with the protein. Although residues such as glycine and proline in the Hsp70 C-terminal peptide, do not interact directly with the TPR domain, they are likely to play an important role in providing the required conformation to the peptide for recognition by the TPR domain. Thus the binding energy changes (\(\Delta\Delta G\)) cannot be reliably computed for alanine mutations of residues, which are not directly interacting with the protein.

Unfavorable entropic contributions upon ligand binding are often compensated for by increased dynamic motion in distant regions of the protein (Evans and Bronowska, 2010). Essential dynamics shows motion in the terminal helices upon binding with either peptide. The H1 and H7 helices move backwards in order to expose the inner surface of the cradle for peptide binding. It is known that TPRs have a rigid conformation (Cheng et al., 2006). In contrast, this study provides evidence that the change in curvature by the concerted movement of secondary structural elements may be necessary for ligand binding. This observation supports a recent study on a TPR containing protein, MamA, where 3 Å radial movement by two N-terminal TPR motifs
was observed in different crystal structures upon ligand imitator binding (Zeytuni et al., 2011). Unlike the case of Hop, where there are two TPR domains with specificity for Hsp binding, proteins such as CHIP and CNS1 (Qbadou et al., 2006, Hainzl et al., 2004) recognize both Hsp70 and Hsp90 with single TPR domain. In CHIP, I^{120} (Hsp70)/M^{120} (Hsp90) are accommodated in a hydrophobic pocket, which is absent in Hop (Scheufler et al., 2000). In our case, Ile^{120} and Val^{123} sandwich the Phe^{12} of AtToc64_TPR, thus enhancing the interaction through van der Waals contacts.

In the current work, the two peptides used for this study exhibit a conserved EEVD motif, however residues N-terminal to this conserved sequence differ between Hsp70 (sequence: G^{117}I^{118}T^{119}I^{120}) and Hsp90 (sequence: T^{117}S^{118}R^{119}M^{120}). Despite these differences, the two peptides bind to the AtToc64_TPR with similar affinity. The differences in the observed binding mechanism (entropy vs enthalpy driven reactions) is likely to be a consequence of the contributions of the different residues located at the N-terminus, where peptides with different sequences bind with similar affinity to the same protein. Interestingly, the pentapeptide versions of Hsp70 (sequence: I^{120}E^{121}E^{122}V^{123}D^{124}) and Hsp90 (sequence: M^{120}E^{121}E^{122}V^{123}D^{124}) demonstrate entropically driven interactions as observed by ITC studies, although with weak micromolar affinity than the octapeptides. Additionally the residues glycine and proline do not interact with the protein in the case of the TPR-Hsp70 system. It is tempting to conclude from the above observed results that the interactions involved with the N-terminal residues of the Hsp90 octapeptide, namely threonine, serine and arginine, contribute towards an enthalpically driven interaction. Additionally, one might hypothesize that the interaction of hydrophobic residues such as I^{120} and M^{120} might assist in the release of water molecules from the interaction interface. Furthermore, although mutation of I^{120} to alanine caused improved enthalpy, this improvement is lost by entropy.

Furthermore, computational studies using essential dynamics suggests that a higher conformational spread is observed in the TPR-Hsp70 system than in the TPR-Hsp90 system resulting in an entropy driven process in the former system. The observed flexibility in the terminal helices, H1 and H7, upon ligand binding may further contribute to the entropy of the system. However, in the TPR-Hsp90 system, the
interaction with the N-terminal residues (T\textsuperscript{117}, S\textsuperscript{118}, R\textsuperscript{119}) appears to dominate over the entropy due to the increased flexibility thus resulting in an enthalpy driven process.

This is the first study reported till date, where we throw light on the \textit{Ar}Toc64\_TPR and C-term Hsp70/Hsp90 interaction at a molecular level using \textit{in vitro} studies with ITC and \textit{in silico} studies with molecular dynamics. Our \textit{in vitro} and \textit{in silico} studies suggest that \textit{Ar}Toc64\_TPR can recognize the C-terminal octapeptide of both Hsp70 and Hsp90 with similar affinity and mutation of residues within this peptide to alanine destabilizes the interactions.
3.6 SUPPORTING MATERIALS

Additional table to show the thermodynamic parameters for interactions of alanine mutants of Hsp70 and Hsp90 peptides with AtToc64_TPR-H6.

Table 3-S1 Alanine scanning of both ligands (Hsp70/90) using ITC.

<table>
<thead>
<tr>
<th>Hsp70 mutants</th>
<th>N</th>
<th>ΔG (kcal/mol)</th>
<th>ΔH (kcal/mol)</th>
<th>TΔS (kcal/mol)</th>
<th>Kd (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>APTIEEV</td>
<td>1</td>
<td>-4.2</td>
<td>-1.0</td>
<td>3.2</td>
<td>740</td>
</tr>
<tr>
<td>GATIEEV</td>
<td>1</td>
<td>-4.2</td>
<td>-2.4</td>
<td>1.8</td>
<td>657</td>
</tr>
<tr>
<td>GPAIEEV</td>
<td>1</td>
<td>-4.1</td>
<td>-1.6</td>
<td>2.5</td>
<td>769</td>
</tr>
<tr>
<td>GPTAEV</td>
<td>1</td>
<td>-3.1</td>
<td>-5.5</td>
<td>-2.4</td>
<td>4651</td>
</tr>
<tr>
<td>GPTIAEV</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>No binding</td>
</tr>
<tr>
<td>GPTIEAV</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>No binding</td>
</tr>
<tr>
<td>GPTIEAD</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>No binding</td>
</tr>
<tr>
<td>GPTIEVA</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>No binding</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hsp90 mutants</th>
<th>N</th>
<th>ΔG (kcal/mol)</th>
<th>ΔH (kcal/mol)</th>
<th>TΔS (kcal/mol)</th>
<th>Kd (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASRMEEV</td>
<td>1</td>
<td>-4.2</td>
<td>-5.5</td>
<td>1.3</td>
<td>854</td>
</tr>
<tr>
<td>TARMEV</td>
<td>1</td>
<td>-4.7</td>
<td>-1.7</td>
<td>3.0</td>
<td>344</td>
</tr>
<tr>
<td>TSMEEV</td>
<td>1</td>
<td>-5.4</td>
<td>-2.2</td>
<td>3.2</td>
<td>100</td>
</tr>
<tr>
<td>TSRAEV</td>
<td>1</td>
<td>-4.0</td>
<td>-2.3</td>
<td>1.7</td>
<td>1080</td>
</tr>
<tr>
<td>TSRAEVD</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>No binding*</td>
</tr>
<tr>
<td>TSRMEEV</td>
<td>1</td>
<td>-3.7</td>
<td>-2.3</td>
<td>1.4</td>
<td>1908</td>
</tr>
<tr>
<td>TSRMEEAD</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>No binding*</td>
</tr>
</tbody>
</table>

Synthetic octapeptides were used, where point mutation of individual residue into alanine, helped in understanding the contribution of the same towards the thermodynamics of binding. Asterik (*) suggests that the heat change during binding event is quite low and the signal to noise ratio is high. Hence the Kd was considered as no binding. The reported values have an error range of ± (5-10)%
Additional figure to show the biophysical characterization of AtToc64_TPR_H6.

Figure 3-S1 Characterization of the AtToc64_TPR-H6.

A. Cradle structure of the 3-TPR domain. Helices are numbered as H1 to H7. Each TPR forms a helix turn helix structure which repeats itself three times followed by a solvation or capping helix. B. 16% denaturing SDS PAGE gel showing the purification of AtToc64_TPR-H6 after Ni-NTA chromatography. The band observed in the red box indicates the presence of AtToc64_TPR-H6. C. Thermal denaturation curve of AtToc64_TPR-H6. \( T_m \) was calculated to be \(-35 ^\circ C\).
Additional figure to show the reliability of the I-TASSER generated model for \textit{AtToc64\_TPR}.

Figure 3-S2 Tube depiction of the superposition of the TPR model obtained from I-TASSER on the high-resolution crystal structure of the TPR domain from Hop complexes with the Hsp70 octapeptide (1ELW).

\textit{The modelled structure of the TPR domain is shown in blue color and that of the crystal structure is shown in yellow. The RMSD was found to be 1.36 Å, suggesting that the modelled structure was quite reliable for use in MD studies.}
Additional figure to show the percentage of residues in the allowed region in the I-TASSER generated model for *ArToc64_TPR*.

**Figure 3-S3** Ramachandran plot of modelled *ArToc64_TPR* (Apo).
Additional figure to show the percentage of residues in the allowed region in the I-TASSER generated model for AtToc64_TPR in complex with octapeptide mimicking the C-terminal of Hsp70.

Figure 3-S4 Ramachandran plot of C-Hsp70 (octapeptide) bound form (T_C70).
Additional figure to show the percentage of residues in the allowed region in the I-TASSER generated model for AtToc64_TPR in complex with octapeptide mimicking the C-terminal of Hsp90.

**Figure 3-S5** Ramachandran plot of C-Hsp90 (octapeptide) bound form (T_C90).
Additional figure to show the secondary structure change during the whole simulation (50 ns) in all the systems.

**Figure 3-S6** Secondary structure maps of all the trajectories created by DSSP.

*Color codes: green represents alpha helices, black represents turns, grey represents 3-10 helices and white represents coils.*
Additional figure to show the dynamic cross-correlation in the three different systems.

**Figure 3-S7** Dynamic cross-correlation maps.
A. Apo, B. C-Hsp70 bound form of TPR and C. C-Hsp90 bound for of TPR using Ca atoms.
Additional figure to show the molecular surfaces of the receptors in the two complexes.

**Figure 3-S8** Molecular surface representations.

A. C-Hsp70 octapeptide is bound to AtToc64_TPR; B. C-Hsp90 octapeptide is bound to AtToc64_TPR. The averaged pdb of the final 5ns of the simulation from each trajectory was used to create these maps. Electronegative and electropositive charges are colored in red and blue respectively.
Chapter 4

Exploring ligand recognition, selectivity and dynamics of TPR domains of chloroplast Toc64 and mitochondria Om64 from *Arabidopsis thaliana*
4.1 ABSTRACT

The study aims to gain insight into the mode of ligand recognition by TPR domains of chloroplast Toc64 and mitochondrial Om64, two paralogous proteins that mediate import of proteins into chloroplast and mitochondria respectively. Chaperone proteins associate with precursor proteins in the cytosol to maintain them in a translocation competent conformation, and are recognized by Toc64 and Om64 that are located on the outer membrane of the target organelle. Hsp70 and Hsp90 are two chaperones, which are known to play import roles in protein import. The C-termini of these chaperones are known to interact with the Tetratricopeptide Repeat (TPR) domain of chloroplast Toc64, and mitochondrial Om64 in Arabidopsis thaliana (At). Using a molecular dynamics approach and binding energy calculations, we identify important residues involved in the interactions. Our findings suggest that the TPR domain from AtToc64 has higher affinity towards C-terminal residues of Hsp70 (Figure 4-1). The interaction occurs as the terminal helices move towards each other enclosing the cradle on interaction of AtHsp70 with the TPR domain. In contrast, the TPR domain from AtOm64 does not discriminate between the C-termini of Hsp70 and Hsp90. These binding affinities are discussed with respect to our knowledge of protein targeting and specificity of protein import into endosymbiotic organelles in plant cells.
Figure 4-1 Graphical abstract showing a model for the interactions between C-termini of chaperones Hsp70 and Hsp90 with the TPR domains of two homologues, Toc64 and Om64 from *Arabidopsis thaliana*.

*Interactions have been investigated using simulation studies. The C-terminus of Hsp90 exhibited higher affinity towards the TPR domain from Toc64 whereas that of Om64 does not show any preference for Hsp90 or Hsp70. Additionally, the mode of interactions and the contribution of residues towards interactions are studied in detail.*

4.2 INTRODUCTION

Mitochondria and chloroplasts are two organelles in the plant cell, which originated from two distinct endosymbiotic events. The majority of proteins in these organelles are encoded in the nucleus, synthesized in the cytosol and delivered to the destined locations with the help of different translocation machineries, located on the outer and inner membranes of each organelle (Li and Chiu, 2010, Lithgow and Schneider, 2010). These proteins are synthesized in precursor forms, containing targeting signals, which are necessary and sufficient to direct them to the specific organelles. Heat shock proteins (Hsp) or chaperones play an important role in the import process; they maintain the precursor proteins in a translocation competent state while preventing them from aggregating in the cytosol (Flores-Pérez and Jarvis, 2013). The coexistence of these two organelles in the plant cell, requires the functioning of highly specific protein targeting
systems in order to avoid mis-targeting between organelles (Peeters and Small, 2001). However, as many as 100 proteins, including pea glutathione reductase and amino-acyl-tRNA synthetases, have been reported to be dual targeted to both of these organelles (Xu et al., 2013, Carrie and Whelan, 2013). The signal sequences for proteins destined to both these organelles demonstrate both high sequence similarity and a high content of positively charged residues. Mitochondrial targeting signals have the potential to form amphiphilic α-helices and those for chloroplast are thought to form random coils, although recently the latter are known to form secondary structures similar to the mitochondrial counterpart (Bruce, 2000). Thus mis-targeting could be a consequence of unusual properties of targeting sequences. In vivo mis-targeting was first demonstrated in transgenic tobacco plants, where the presequence of yeast mitochondrial cytochrome oxidase subunit Va fused to the protein chloramphenicol acetyltransferase could act as both a chloroplast and a mitochondrial targeting signal (Huang et al., 1990). Mis-targeting of chloroplast proteins into the mitochondria has been reported; for example, dihydrofolate reductase containing a chloroplast targeting signal from the small subunit of ribulose-1,5-bisphosphate carboxylase from green algae is targeted to yeast mitochondria (Hurt et al., 1986). Similarly triosephosphate-3-phosphoglycerate phosphate translocator from spinach chloroplast can be imported to yeast mitochondria (Brink et al., 1994). The above demonstrated mis-targeting in heterologous in vitro import systems. In vivo the study of targeting involves the use of intact cellular systems. These systems do not however provide an understanding of the kinetics and the underlying mechanism involved in protein import (Rudhe et al., 2002). However, in vitro dual import systems rule out the above limitations and additionally help in understanding sorting between the two organelles. The chloroplast acts as a competing organelle in mitochondrial import assays in the dual import system as well as abolishing the mis-targeting of chloroplast proteins. Interestingly mis-targeting of mitochondrial proteins into the chloroplast has not been reported. This suggests that other factors might be involved to prevent mis-targeting in vivo.

In the chloroplast, the recognition and translocation is initiated by subunits of a multimeric complex called the Translocon at the outer envelope of chloroplast (Toc). Toc64 is one of the subunits of the Toc core, and contains a C-terminal cytosolic Tetra tricopeptide repeat domain (TPR), which recognizes the C-terminal region of Hsp90 bound precursor protein (Li and Chiu, 2010, Qbadou et al., 2007, Sohrt and Soll,
Similarly, in the case of mitochondria, the precursors are recognized and imported by the subunits of multimeric Translocon of the outer membrane (Tom) complex. Om64 is an integral membrane protein found on the outer membrane of some plant mitochondria, which has 67% sequence identity and a similar domain architecture to that of Toc64 (Chew et al., 2004), but is not a core subunit of the Tom complex (Lister et al., 2007). *In silico* analyses suggest that Om64 appears to have arisen from a recent evolutionary event as it is only found in a subset of vascular plants (Carrie et al., 2010). *In vivo* studies have suggested that depletion of Om64 can affect import of certain mitochondrial proteins (Lister et al., 2007). While a number of possible roles for Om64 have been proposed, such as a replacement for the absence of Tom70 in plant mitochondria and as a receptor for dual-targeted proteins (Duncan et al., 2013, Chew et al., 2004, Peeters and Small, 2001), none of these roles have been experimentally verified. It has been observed that *Arabidopsis* mutants lacking the functional genes encoding three Tom20 isoforms and Om64 produced early embryo lethal phenotype (Duncan et al., 2013). This may be due to the fact that the direct presequence recognition pathway by Tom20 and the chaperone assisted recognition of preproteins by Om64 are absent for delivering preproteins to the import pore.

The domain architecture of both Toc64 and Om64 suggests an N-terminal membrane spanning region and a C-terminal 3-TPR domain. TPR motifs are degenerate 34 amino acids repeats, which are composed of a canonical helix-turn-helix fold. The helices are stacked in an anti-parallel fashion forming a grooved surface with a concave and a convex face and thus forming a solenoid structure. The concave face forms a protein-protein interaction platform (D’Andrea and Regan, 2003, Scheufler et al., 2000). TPR-ligand interaction is usually specific and depends on ligand sequence, secondary structure, different surface residues and the presence of a hydrophobic pocket on the TPR motif. Molecular mechanisms underlying ligand binding have been previously studied including the TPR domains from Hsp70 and Hsp90 organizing protein (Hop) (Scheufler et al., 2000) and p67phox (Lapouge et al., 2000). These structural domains form multiprotein complex mediators and are found in all kingdoms of life (Allan and Ratajczak, 2011, Lapouge et al., 2000). Receptors containing a TPR domain play an important roles in both mitochondria (Baker et al., 2007) and chloroplast (Mirus et al., 2009) protein import.
In this study, we focus on the mode of ligand recognition by the TPR domains of Toc64 and Om64. Previous *in vivo* studies have shown that the TPR domain from Toc64 interacts preferentially with the C-terminus of Hsp90 chaperone (Qbadou et al., 2007). The C-termini of cytosolic chaperones as Hsp70 and Hsp90 contain an EEVD motif, which is highly conserved across all eukaryotic lineages (*Figure 4-S1A*). In the case of Hop from humans, the aspartate of this motif has been shown to anchor to the TPR domains by conserved dicarboxylate clamp (Scheufler et al., 2000), which comprise of mostly hydrogen bonding and electrostatic interactions. Further the C-terminal of the peptide faces the N-terminal of the TPR domain (*Figure 4-S1B*). Although the TPR residues forming the carboxylate clamp are conserved (Scheufler et al., 2000), analysis of the solution structure of Tah1 bound to MEEVD motif of Hsp90 (PDB: 2L6J) (Jiménez et al., 2012) revealed that not all of these residues interact with the terminal aspartate of the peptide. Furthermore, the arrangement of the peptide on the TPR domain of Tah1 is opposite to that observed in the TPR domain from Hop (*Figure 4-S1C*). Deletion or alanine mutation of residues N-terminal to this motif have been reported to reduce binding affinity (Brinker et al., 2002) suggesting that these residues play a role in recognition. Differences in residues N-terminal to the conserved EEVD are evident between plants and humans. For example, in *Arabidopsis thaliana* (*At*), a conserved lysine is present prior to the EEVD in both Hsp70 and Hsp90, however in humans the residue at the same position is threonine in Hsp70 and arginine in Hsp90. Understanding the role of these residues in interaction to the TPR domain is important in detailing selectivity and specificity in preprotein import. Furthermore, the mode of interactions and preferences of Hsp70 and Hsp90 with respect to the TPR domain from Om64 has not been reported to date. In the current study, we explore the mode of binding, selectivity and dynamics of the TPR domains of Toc64 and Om64 to C-terminal regions of Hsp70 and Hsp90 from *Arabidopsis thaliana*. In the absence of experimental structural data, molecular modelling, using multiple threading approaches, have been undertaken to obtain the structures of the domains, followed by docking of C-terminal octapeptides of Hsp70 and Hsp90 to obtain the structures of the protein-peptide complexes. Computational simulation studies were also undertaken to characterize the role of protein dynamics in protein-peptide interactions. Noticeably, glutamate residues of the EEVD motif play important roles in the interactions. While the TPR domain of Toc64 exhibits a higher affinity towards the C terminal residues of
Hsp70 than Hsp90, that of Om64 interacts with both C-terminal peptides with similar affinity.

4.3 MATERIALS AND METHODS

4.3.1 Starting structures

The sequences of AtToc64 (accession number NP_188424, At3g17970), AtOm64 (accession number NP_196504, At5g09420), C-terminus of AtHsp70 (GPKIEEVD) and C-terminus of AtHsp90 (GSKMEEVD) were used for the study. In order to obtain a reliable model for in silico studies, the protein sequences which constitute the TPR domain in AtToc64 (residues 474-589) and AtOm64 (residues 488-603) were submitted for automated protein structure modelling using the I-TASSER pipeline. This server builds 3D models based on multiple threading alignments and iterative structure assembly simulations (Roy et al., 2011, Zhang, 2008, Roy, 2010, Roy, 2012) and has outperformed other servers in specificity of predictions in the Critical Assessment of Techniques for Protein Structure Prediction (CASP) [7,8,9,10] experiments. Since the above method used multiple threading alignments, there is minimal bias towards a particular structural model as in the case of homology modeling (Chung, 1996, Venclovas, 2005). The model systems for protein-ligand complexes were built using the coordinates of the peptide obtained from the high-resolution crystal structure of the TPR domain of Hop in complex with the C-terminal octapeptide of Hsp70 (Scheufler et al., 2000) (1ELW obtained from the Protein Data Bank) and models of TPR domains obtained from I-TASSER server as templates. For further validation of the above model, the co-ordinates of the I-TASSER model (apo form) and that of the ligand were docked using the molecular docking algorithm ZDock (Pierce et al., 2011). This docking algorithm uses an energy scoring function to establish the best binding interaction between the two partners after searching all possible binding modes in translational and rotational space. The docked structure obtained from ZDock was identical to that of the modeled complex, thus validating our docking approach. In addition, a high-resolution peptide docking and refinement protocol using Rosetta FlexPepDock (London et al., 2011, Raveh et al., 2010) was used to further refine the above docked systems. The starting structures were refined in 200 independent FlexPepDock simulations, which consisted of a low-resolution pre-optimization step followed by high-resolution
refinement, and high-resolution mode simulations. This procedure produced refined energy minimized protein-peptide complexes, which were used as the reference structures for Molecular Dynamics (MD) Simulations. The apo and liganded starting structures were further validated by a Ramachandran plot (Lovell et al., 2003). Thus six systems (three for each TPR domain) were prepared: AToc64 receptor only (T_Apo), AToc64_TPR bound to the C-terminus of Hsp70 (T_C70) and AToc64_TPR bound to the C-terminus of Hsp90 (T_C90), AtOm64_TPR receptor only (Om_Apo), AtOm64_TPR bound to the C-terminus of Hsp70 (Om_C70) and AtOm64_TPR bound to the C-terminus of Hsp90 (Om_C90).

4.3.2 Preparation for simulation

The simulation studies were carried out using the Graphics Processing Unit (GPU) accelerated Assisted Model Building with Energy Refinement (AMBER) suite version 12 (Goetz et al., 2012) associated with the latest all-atom ff12SB force field (Case et al., 2012, Case et al., 2005). The starting structures for the simulations were neutralized using Na\(^+\) and Cl\(^-\) ions. The tleap module from AMBERTOOLS12 was used to position hydrogen atoms. The protein was centered in a solvent truncated octahedron box, which was made of TIP3P (3-point charged) triangulated water molecules with a 12 Å cut off in all directions (Jorgensen et al., 1983). The total number of atoms including water molecules was approximately 27,000 across each of the various systems. The systems were minimized using a two-phase energy minimization procedure, which included 2500 cycles of steepest descent and 2500 cycles of conjugate gradient with solute atoms restrained by a harmonic potential with a force constant of 50 kcal mol\(^{-1}\) Å\(^2\). This was followed by 5000 steps of unrestrained whole system minimization. 50 ps of density equilibration with weak harmonic restraints (2 kcal mol\(^{-1}\) Å\(^2\)) on the solute molecule was performed followed by unrestrained equilibration for 500 ps was then carried out using constant-pressure and constant-temperature conditions. All simulations ran with constraints using the SHAKE algorithm (Ryckaert et al., 1977) on hydrogen-linked bonds with a tolerance of 0.0001. To evaluate long-range electrostatic interactions the Particle Mesh Ewald (PME) method (Salomon-Ferrer et al., 2013) was used employing a 9 Å cut-off. A 2 fs integration time step was used to numerically solve Newton’s equations of motion. Langevin dynamics was used to maintain a constant temperature of 300 K through out the simulations. All the simulations were carried out using the PMEMD module in AMBER12. Production runs were performed for 50 ns in an explicit solvent environment and isothermal-isobaric (NPT) ensemble.
4.3.3 Analysis of trajectory

After a 50 ns of production run, each of the trajectories was analyzed based on the variation in kinetic and potential energies under the NPT ensemble using the ptraj program. The root-mean-square deviation (RMSD) and atomic positional fluctuation per residue (RMSF) were analyzed to understand the overall conformational change through out the trajectory. Hydrogen bond and hydrophobic interactions and their occupancies were calculated using HBPLUS using the criteria described by McDonald and Thornton for the definition of a hydrogen bond (McDonald and Thornton, 1994). Only those hydrogen bonds with occupancies above 50% were considered for analysis. Interactions of arginine or lysine side chains with oxygen atoms of aspartate or glutamate residues that were ≤ 4 Å in distance were considered as electrostatic interactions. In addition, the solvent accessible surface area and binding energies for the protein-peptide interactions were calculated using NACCESS (Hubbard and Thornton, 1993) and MMPBSA tools (Kollman et al., 2000) respectively. Computational alanine scanning was performed with each peptide, in order to understand the contribution of individual amino acid residues towards binding. When calculating the difference in free energies (\(\Delta\Delta G\)) between the wild type and the mutants, the results of using Generalized Born calculations (GB) were taken into consideration, as they are well suited for protein-ligand and protein-protein interaction calculations. These values were calculated according to the following equation:

\[
\Delta\Delta G = \Delta G^{\text{wild}} - \Delta G^{\text{mutant}}
\]

For all of the above analyses, the last 5ns of the trajectory were analysed. The APBS software was used to compute the electrostatic potential surface (Baker et al., 2001). LIGPLOT was used to map the hydrogen and hydrophobic bonding patterns between the peptide residues and the residues of TPR involved in interaction (Wallace et al., 1995). The last 5ns averaged structures from each simulation were used to calculate the electrostatic potential and for schematic representation of protein-peptide interactions using ligplots. Visual Molecular Dynamics (VMD) was used for visualizing the trajectories of the simulations of the three systems (Humphrey et al., 1996). Dictionary of protein secondary structure (DSSP) was used to assign the protein secondary structure per amino acid along the different trajectories (Kabsch and Sander, 1983). Further performing Principal Component Analysis (PCA) substantiated the above results (Amadei et al., 1993). Collective coordinates for protein motions were extracted using Covariance matrices. These matrices were constructed using the backbone atoms.
of the protein (N, Cα, C) and were used to calculate the eigen values of maximum magnitude using GROMACS (Pronk et al., 2013). In this way all the linearly correlated motions were analyzed. Correlated atomic motion in the apo and ligand bound forms were obtained by analyzing the dynamical cross correlation map (DCCM) of Ca atoms using Bio3D (Ichiye and Karplus, 1991, Grant et al., 2006). This provided a means to understand the correlation of motions of neighboring and/or distant residues. PyMOL was used to generate all the graphical representations (DeLano, 2002).

4.4 RESULTS AND DISCUSSION

A molecular dynamics approach to gain an molecular insight into the structures of Toc64_TPR and Om64_TPR in complex with the C-terminal peptides of two chaperone proteins (Hsp70 and Hsp90) from Arabidopsis thaliana has been undertaken. The results provide important insights into the role of the TPR domain structure and dynamics in complex formation. Models of TPR domains of Toc64 and Om64 from Arabidopsis, generated using I-TASSER, showed that the structure is composed of seven alpha helices arranged in a head to tail fashion similar to the peptide free TPR domain from PP5 (Das et al., 1998) and peptide bound TPR1 and TPR2A domains from Hop (Scheufler et al., 2000). The RMSD between the I-TASSER modelled TPR domain from Toc64 and the crystal structure of TPR1 domain of Hop (PDB:1ELW) was 1.36 Å, while that of TPR domain from Om64 with the above crystal structure was 1.34 Å. The sequence similarity and sequence identity between the TPR domains of Toc64/Om64 and that from the crystal structure (PDB: 1ELW) was 51% and 39% respectively. The Ramachandran plot analyses showed that in T-Apo/Om_Apo, T_C70/Om_C70 and T_C90/OmC90 systems, 93% (94%), 94% (95%) and 97% (95%) residues respectively were present in favoured regions of the plot (Figure 4-S2 – S3).

The RMSDs, of the Cα atoms from the appropriate starting structure for each simulation were used to assess the overall stability of the simulations. The average Cα-RMSD were found to be 2.10 Å, 1.42 Å, and 1.49 Å for the T_Apo, T_C70 and T_C90 systems respectively and 2.14 Å, 2.13 Å, and 1.89 Å for the Om_Apo, Om_C70 and Om_C90 systems respectively (Figure 4-2A-B). This suggested that the complexed structures were stable and retained their overall structure during the simulation. The observed RMSD values for Om_Hsp systems were in the same range as
reported previously (Mirus et al., 2009). Further analyses of the six independent simulations indicated that, in all cases, the cradle topology of the TPR was maintained. An analysis of Ca RMSF for each of the systems showed that the N and C termini exhibited higher temperature factors (B factor). Additionally, higher B factors were also observed for the residues in the loop connecting the helices H1 and H2 and the loop connecting the helices H2 and H3 (Figure 4-2C) in T_C90 system. Similar results were observed for the residues in the loops connecting helices H1 and H2, helices H2 and H3 and helices H6 and H7 in the Om_Apo system. In addition, residues in the loops connecting helices H5 and H6 and helices H6 and H7 in Om_C90 system also exhibited higher B factors (Figure 4-2D). Finally, the RMSF of the terminal residues of AtToc64_TPR decreased when bound to the Hsp90 octapeptide as compared to the apo form. RMSF plots of Om_Hsp systems demonstrated a similar profile as reported previously (Mirus et al., 2009). Electrostatic surface potential representation is a continuum model, which describes the electrostatic interactions between the protein and the peptide component in each system (Figure 4-S4). Molecular surfaces of the TPR domain in the vicinity of negatively charged residues of the peptide such as Asp124 showed blue coloration confirming the interaction with basic residues of the protein. Similarly, uncharged surfaces on the TPR domain are colored white. Peptide residues that do not interact with the TPR domain such as G117 and P118 in T_C70 system are exposed out of the cradle.
Ligand selectivity of Toc64 and Om64

Figure 4-2 Analysis of the molecular dynamics trajectories obtained after 50 ns of simulation in Toc64_TPR systems and Om64_TPR systems.

A. RMSD plot for the apo form of Toc64_TPR receptor (black trace), the Hsp70 C-terminal octapeptide bound form of the receptor (T_C70: light green trace) and the Hsp90 C-terminal octapeptide bound form of the receptor (T_C90: magenta trace).

B. RMSD plot for the apo form of Om64_TPR receptor (black trace), the Hsp70 C-terminal octapeptide bound form of the receptor (Om_C70: dark green trace) and the Hsp90 C-terminal octapeptide bound form of the receptor (Om_C90: purple trace). Secondary structure of Toc64_TPR is shown in red cartoon.

C. The atomic positional fluctuation (RMSF) plot obtained for Toc64_TPR apo and liganded systems.

D. RMSF plot obtained for Om64_TPR systems. Secondary structure of Om64_TPR is shown in purple cartoon. The regions of the TPR domains with high B factor are shown in red.

4.4.1 Comparative analysis of protein–peptide interaction

As reported in case of other TPR domains, the concave surface of the cradle forms the interaction platform for the octapeptides through intermolecular hydrogen bonding and nonbonding interactions. On an average, approximately 10 hydrogen bonded and nonbonded interactions each were seen in the T_C70 system, whereas approximately 6 - 8 of each of the above kind of interactions were observed in all other peptide bound conformations (Figure 4-S5). The homology model of AtOm64 with octapeptide from C-terminal of Hsp90 reported previously had revealed that approximately 10 of these interactions are observed in more than 20% of the trajectory (Mirus et al., 2009). This is similar to our finding approximately 8 hydrogen bonding interactions were observed in
50% of the trajectory in the above system. Sequence specific and sequence independent interactions assist in stabilizing the ligand bound states of the TPR domains. Additionally, intrapeptide bonds help in providing the necessary conformation to the peptides for interaction with the TPR domains. A detailed analysis of these interactions provides an in-depth understanding of the process of molecular recognition. An example of such analysis representing interaction of Ar/Toc64_TPR with C-terminus of Hsp70 is shown in Movie 1. For the Hsp70 C-terminal peptide, the N-terminal Gly residue is termed Gly117, and the following residues are arranged in ascending order, e.g. Pro118, Thr119, Ile120, Glu121, Glu122, Val123, Asp124. A similar nomenclature for Hsp90 C-terminal peptide is used. The numbering of residues of the TPR domain are shown as superscript, e.g. Lys$^{70}$.

### 4.4.2 An anchor using a carboxylate clamp

The C-terminal conserved residues of the peptide “EEVD” have been known to play a key role in TPR and Hsp interaction. The crystal structure of the TPR domain from Hop in complex with C-terminal peptides of human Hsp70 and Hsp90, show that the carboxylate moieties of the terminal Asp residue (Asp$^{124}$) are held in place by electrostatic interactions with the conserved residues of the TPR domains. This interaction has been referred to as the carboxylate clamp (Scheufler et al., 2000). For the T_C70 system, the carboxylate groups of the conserved Asp124 of the peptide were positioned near to Lys$^{5}$ and Asn$^{9}$ of H1 helix, Asn$^{40}$ of H3 helix and Asn$^{68}$ located in the loop connecting the H4 and H5 helices and Lys$^{70}$ of H5 helix (Figure 4-3A). In the T_C90 system, the carboxylate groups are positioned by Lys$^{70}$ of helix H5, Asn$^{102}$ located in the loop connecting H6 and H7 helices and Lys$^{103}$ and Arg$^{104}$ of helix H7 (Figure 4-3B). Involvement of Asn$^{102}$ and Lys$^{103}$ in carboxylate clamp formation have not been observed in TPR-Hsp70/90 complexes from Hop.
Ligand selectivity of Toc64 and Om64

Figure 4-3 The carboxylate clamp.  
A. Key interacting residues of the Toc64_TPR domain with the carboxylate of Asp$^{124}$ of Hsp70;  B. Key interacting residues of the Toc64_TPR domain with the carboxylate of Asp$^{124}$ of Hsp90.  C. Key interacting residues of the Om64_TPR domain with the carboxylate of Asp$^{124}$ of Hsp90.  D. Key interacting residues of the Om64_TPR domain with the carboxylate of Asp$^{124}$ of Hsp90. The oxygen atoms of the carboxylate moieties are labeled. The peptide residues are coloured with green bonds for Hsp70 and with magenta bonds for Hsp90. The ribbon representation depicts the structures of the Toc64_TPR domain (magenta) and the OM64_TPR domain (teal).

In the Om_C90 system, as with the T_C70 system, the carboxylate anchor is formed by interaction of carboxylate groups of Asp124 with Asn$^9$ and Lys$^{13}$ (H1 helix), Asn$^{40}$ (H3 helix) and Lys$^{70}$ (H5 helix) (Figure 4-3D). The study on the similar system carried out previously using computational methods did not report the involvement of H3 helix in interaction (Mirus et al., 2009). Interestingly in the Om_C70 system, no hydrogen bonding interaction is observed between the Asp124 carboxylate group and residues of Om64_TPR. Van der Waals contacts and electrostatic interactions are observed between the side chains of Lys$^{103}$ of helix H7 and one of the carboxylate moieties; these appear to aid in stabilization of the system (Figure 4-3C). In summary, the involvement of
conserved clamp residues in interactions with the terminal aspartate is observed in T_C70 and Om_C90 systems, however, in T_C90 and Om_C70 systems, the electrostatic interactions are maintained by additional residues of TPR domains.

4.4.3 Key protein-peptide interactions

Hydrogen bonds play a pivotal role in protein-ligand interaction and thus contribute to the binding affinity. These interactions with the peptide involve both the side chains (sequence specific) and main chain (sequence independent) atoms of the TPR domain. In the T_C70 system residues N-terminal to Lys119 do not interact with the TPR domain (Figure 4-4A) whereas in the T_C90 system, Gly117 is held in place by hydrogen bonding interactions with Glu14 and Lys15 (Figure 4-4B). Van der Waals and hydrophobic interactions are observed between Ser118 of the peptide and Phe12 and the aliphatic portion of Lys13 and Glu47 of the TPR domain. Lys119 interacts with Glu47/Leu46 and Glu47 in the T_C70 and T_C90 system respectively. While the side chain of Ile120 forms hydrophobic contacts with side chain of Phe12 in the T_C70 system, van der Waals and hydrophobic contacts with Phe12, Leu46 and Met81 positions the Met120 of the peptide in the T_C90 system. Side chain interactions exist between Glu121 and Phe12, Ala43, Arg74, and Arg104 in the case of T_C70, and between, Lys13 and Arg104 in the T_C90 system. Glu122 interacts with Lys70/Asn9, Leu73/Lys13, Arg74/Asn40, Arg104/Asn74 and Asn104 in T_C70/T_C90 systems respectively.
Figure 4-4 Interactions occurring at the protein-peptide interface in AtToc64_TPR-Hsp systems generated by Ligplot.

A. Interactions between the TPR domain from Toc64 and the C-terminal octapeptide of Hsp70;

B. Interactions between the TPR domain from Toc64 and C-terminal octapeptide of Hsp90.

Similarly, Gly117 and Ser118 do not interact with the TPR domain in either of the Om–Hsp systems. Although there are no hydrogen bonding interactions observed between Om64_TPR and Pro118, other non-bonded interactions exist with Leu$^{46}$, Phe$^{51}$ and Ser$^{81}$ (Figure 4-5A). While in the Om_C90, Lys119 interacts only with Glu$^{47}$ (Figure 4-5B), in Om_C70, interactions are observed between Lys119 and Thr$^{77}$, Glu$^{80}$, Ser$^{81}$ and Arg$^{111}$ from the TPR domain. Sequence independent main chain interactions exist between Ile120 and Thr$^{77}$, whereas Met120 is held in place by nonbonded interactions with Tyr$^{12}$ and Lys$^{15}$. Glu121 interacts with Leu$^{46}$/Tyr$^{12}$, Lys$^{70}$/Lys$^{70}$, Arg$^{74}$/Arg$^{74}$ and Thr$^{77}$ in case of Om_C70 and Om_C90 systems respectively. Similarly, Glu122 interacts with Thr$^{104}$/Lys$^{13}$, Ala$^{108}$/Lys$^{70}$ and Arg$^{111}$/Arg$^{74}$ in Om_C70/Om_C90 respectively.
Figure 4-5 Interactions occurring at the protein-peptide interface in AtOm64_TPR-Hsp systems generated by Ligplot.
A. Interactions between the TPR domain from Om64 and the C-terminal octapeptide of Hsp70; B. Interactions between the TPR domain from Om64 and C-terminal octapeptide of Hsp90.

4.4.4 A cavity for Valine

Valine is held in position by a number of hydrophobic and van der Waals interactions involving Asn\(^9\), Phe\(^{12}\), Asn\(^{40}\), Lys\(^{70}\) and Arg\(^{74}\) in the case of the T_C70 system and Lys\(^{70}\), Leu\(^{73}\), Arg\(^{74}\) and Arg\(^{104}\) in the case of the T_C90 system. Similarly in the Om_C90 system, the side chains of Asn\(^9\), Tyr\(^{12}\), Asn\(^{40}\), Ala\(^{43}\) and Arg\(^{74}\) make a hydrophobic groove for Val123. In contrast, Val123 in Om_C70 does not interact with the TPR domain.

4.4.5 Intrapeptide interactions

Intrapeptide interactions allow the peptide to attain a favorable conformation so as to interact with the protein. This analysis is crucial for understanding the additional factors that might aid the protein-peptide interaction. The following key intrapeptide hydrogen bonding interactions are observed in the T_C70 system: the main chain nitrogen of
Val123 with main chain oxygen of Glu121; main chain nitrogen of Glu\textsuperscript{122} with side chain oxygen of Glu\textsuperscript{121} (Figure 4-6A). In the T\_C90 system, the side chain oxygen of Ser118 hydrogen bonds with the main chain nitrogen and oxygen of Met120 and main chain nitrogen of Glu122 forms a hydrogen bond with side chain oxygen of Glu\textsuperscript{121} (Figure 4-6B).

**Figure 4-6** Key intrapeptide interactions.

A. T\_C70 system B. T\_C90 system. C. Om\_C70 system D. Om\_C90 system. The TPR domains are shown as ribbon representations in magenta (AtToc64\_TPR) and teal (AtOm64\_TPR). The interacting peptide residues are shown in stick representation. Green coloring denotes AtHsp70 and magenta coloring denotes AtHsp90. Key intrapeptide hydrogen bonds are included in dash lines along with their distances shown.

Similarly in Om\_C70 the side chain nitrogen of Lys119 forms hydrogen bond interactions with main chain and side chain oxygen atoms of Ile120 and Glu122 respectively (Figure 4-6C) whereas the side chain nitrogen of Lys119 forms a hydrogen bond with its own main chain oxygen in the Om\_C90 system. Additionally,
identical intrapeptide bonds to that observed in the T_C70 system involving Glu121, Glu122 and Val123 are also present (Figure 4-6D).

### 4.4.6 Binding energy and computational alanine scanning

An important aspect of the interaction between the C-terminal octapeptide moieties of AtHsp70 and AtHsp90 and the TPR domains of Toc64 and Om64 include characterizing the differences between their binding energy. The $\Delta G$ of interaction between the C-term octapeptide of AtHsp70 and Toc64_TPR was calculated to be $-49.2$ kcal mole$^{-1}$, whereas for the T_C90 system, the binding energy was found to be $-25.5$ kcal mole$^{-1}$. For the Om_C70 and Om_C90 systems, the binding free energies were calculated to be $-22.9$ kcal mole$^{-1}$ and $-24.4$ kcal mole$^{-1}$ respectively.

**Table 4-1** Computational alanine scanning for the four ligand bound systems.

<table>
<thead>
<tr>
<th>A. Toc_TPR System</th>
<th>$\Delta \Delta G$ (kcal/mol)</th>
<th>B. Om_TPR System</th>
<th>$\Delta \Delta G$ (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hsp70</td>
<td></td>
<td>Hsp70</td>
<td></td>
</tr>
<tr>
<td>GPKIEEVD</td>
<td>-0.4</td>
<td>GPKIEEVD</td>
<td>-1.2</td>
</tr>
<tr>
<td>GAIEEVD</td>
<td>-0.3</td>
<td>GPAIEEVD</td>
<td>2.1</td>
</tr>
<tr>
<td>GPKAEEVD</td>
<td>-1.5</td>
<td>GPKAEEVD</td>
<td>0.2</td>
</tr>
<tr>
<td>GPKIAEVD</td>
<td>-22.4</td>
<td>GPKIAEVD</td>
<td>-5.8</td>
</tr>
<tr>
<td>GPKIEAED</td>
<td>-3.1</td>
<td>GPKIEAED</td>
<td>-16.9</td>
</tr>
<tr>
<td>GPKIEEAD</td>
<td>-4.0</td>
<td>GPKIEEAD</td>
<td>0.1</td>
</tr>
<tr>
<td>Hsp90</td>
<td></td>
<td>Hsp90</td>
<td></td>
</tr>
<tr>
<td>GSKMEEVD</td>
<td>-1.3</td>
<td>GSKMEEVD</td>
<td>0.2</td>
</tr>
<tr>
<td>GAAMEEVD</td>
<td>-1.5</td>
<td>GAAMEEVD</td>
<td>1.2</td>
</tr>
<tr>
<td>GSAMEEVD</td>
<td>-2.7</td>
<td>GSAMEEVD</td>
<td>-2.2</td>
</tr>
<tr>
<td>GSKMAEVD</td>
<td>-2.3</td>
<td>GSKMAEVD</td>
<td>-10.5</td>
</tr>
<tr>
<td>GSKMAD</td>
<td>-16.7</td>
<td>GSKMAD</td>
<td>-3.2</td>
</tr>
<tr>
<td>GSKMEEAD</td>
<td>-2.0</td>
<td>GSKMEEAD</td>
<td>-4.4</td>
</tr>
</tbody>
</table>

Point mutations to alanine were performed computationally for each residue of the Hsp70 and Hsp90 octapeptides. Changes in binding free energy ($\Delta \Delta G$) in kcal/mol were calculated using MMGBSA. The mutations sites are shown in green for the Hsp70 mutants and in magenta for the Hsp90 mutants.
To further investigate the contribution of each residue of the peptides towards interaction, computational alanine scanning calculations was performed and $\Delta\Delta G$ was determined. The terminal amino acids could not be mutated by MMGBSA due to limitations in the program. In the Toc64_TPR-Hsp systems, mutations of Val123, Glu122, Glu121, Met120 and Ile120 individually to alanine caused a decrease in binding affinity, with the largest effect observed in the case of Glu121 in T_C70 system (Table 4-1A). This residue is held by a number of sequence dependent hydrogen bonding interactions, which would be abolished upon mutation to alanine. In the T_C90 system there are only 4 interactions involving Glu122 however in the T_C70 system there are 8 interactions (Figure 4-4). Similarly, mutation of Glu121 in the Om_C90 system and Glu122 in the Om_C70 system caused a significant reduction in binding energy making the interaction unfavorable (Table 4-1B). Mutation of Glu121 in the Om_C70 system and Glu122 in Om_C90 system did reduce the binding, but to a lesser extent than observed ArToc64_TPR-Hsp systems. Mutations of Val123 in the Om_C70 system and Ser118 in the Om_C90 system to alanine, did not affect binding, as these residues do not interact with the TPR domain. Mutation Ile120 in the Om_C70 system to alanine did not affect binding possibly due to the fact that this residue only uses main chain atoms for interaction (Figure 4-5). Although Lys119 is conserved between ArHsp70 and ArHsp90, mutation of this residue to alanine improved binding as evidenced by a positive increase in $\Delta\Delta G$. Hence the role of this conserved lysine in plants is not fully understood.

In summary, alanine scanning experiments in the T_C70 system suggest that alanine mutants did decrease the binding affinity although to various extent due to loss of stabilizing interactions with their partners. Similar observation was recorded in T_C90 system although the decrease in affinity was lesser compared to the T_C70 system. Further, in the case of Om_C70 system, alanine mutations of K119, I120 and V123 favoured binding, suggesting that these residues in the wild type Hsp70 did not significantly favour binding. Similarly favourable binding event was observed on mutation of S118 and K119 to alanine in Om_C90 system. This analysis explained why T_C70 system exhibits the highest binding affinity among the four systems.
4.4.7 Dynamics in the protein-peptide interaction

The dictionary of protein secondary structure (DSSP) map showed that the overall secondary structures of the TPR domains in the apo and various peptide bound forms are similar throughout the simulations. A number of residues near the C-terminus of the \textit{At}\textit{Toc64}\_TPR domain in the T\_apo and T\_C70 systems appear to adopt coil and turn conformations respectively along the trajectory (Figure 4-S6). This correlates with the higher RMS fluctuation as discussed previously. Similarly, in the case of the Om\_Apo system, residues 30 - 34 undergo a secondary structure change between a 3-10 helix and a coil; where the coil conformation dominates the trajectory. This is stabilized in the peptide bound systems, where predominantly 3-10 helices and turns are observed (Figure 4-S7).

4.4.8 Dynamic cross-correlation map (DCCM)

In order to understand the motion between pair of atoms, DCCMs were generated for different systems. The cross-correlation map for the T\_Apo system suggests a positive correlation between helices: H1 and H2, H3 and H4, H4 and H5 and to a small extent between H5 and H6 (Figure 4-S8A). Similar correlated motion is observed in the Om\_Apo system, where stronger correlation exists between the H5 and H6 helices in addition to a positively correlated movement between the H2 and H7 helices. H1 helix, in both the ligand unbound systems, exhibits negatively correlated motion with H4 and H6. In addition in case of Om\_Apo system, H2 helix shows negatively correlated motion with H3, H4, H5 and H6 helices.

In both the cases, peptide binding causes subtle variation in different kinds of correlated motions. Among the peptide bound systems, the T\_C70 system, where comparatively stronger affinity for peptide was observed, a higher number of residues exhibited correlated motion, such as positive correlation among residues 40-80 comprising a part of helix H3 and all of helices H4 and H5. Interestingly helix H2 showed a negative correlation with the above-mentioned regions (Figure 4-S8B). In contrast the majority of residues in the Om\_C90 system exhibited very little or no correlated motion compared to other liganded forms (Figure 4-S8F).
4.4.9 Principal component analysis

In order to understand the harmonic and large-scale motions in the protein occurring in the essential subspace, the first 2 vectors were considered significant. These two vectors accounted for ~50% of the total fluctuations in all the systems. Hence the first two principal components, namely PC1 and PC2, were plotted against each other in the form of 2D plots for all trajectories (Figure 4-7A-C and Figure 4-8A-C). The conformational space sampled was found to be similar in all the systems, with multiple local minima.

![Conformational space samples](image)

**Figure 4-7** Principal component analyses of Ar/Toc64_TPR-Hsp systems.

*Essential dynamics - 2D projection of individual trajectories with their first two eigen vectors: vector 1 and vector 2 (A - C) and their corresponding porcupine plots (D - F) for T_Apo (black), T_C70 (green) and T_C90 (magenta). Porcupine plots of the three systems displayed with a cone model. The length and orientation of the cone (red) is positively correlated with the magnitude and direction of motion.*

In the T_C70, Om_C70 and Om_C90 systems, the sampled space is slightly more restricted, whereas in the other systems a diffused pattern is observed. As seen in the apo form, the terminal loop regions of the H7 and H1 helices in the T_C70 and T_C90 systems respectively, have higher motion; this correlates well with the observed rms fluctuations in these systems. The H1 and H7 helices move towards each other as if to enclose the peptide in the T_C70 system (Figure 4-7E), which is contrary to the motion observed in T_C90 system. Additionally, the helices H3, H4 and H5, which
form the interaction platform of the cradle in T_C70 system, seem to move in an outward direction. Similarly, in the Om_C70 and Om_C90 systems, the terminal loop regions of the H1 helices have higher motion as seen in the rms fluctuation plot. Additionally the H1 helix in case of the Om_C70 system and the H7 helix in the Om_C90 system appear to move towards the TPR groove. Notably the magnitude of motion of the helices in Om-Hsp systems is smaller than that of T_C70 system. In the Om_Apo system, the terminal helices display an outwardly directed motion away from the cradle and the helices H3, H4 and H5 move towards the inner surface of the cradle as shown by eigen vector 2 (Figure 4-8D). This finding correlates with that reported previously (Mirus et al., 2009). However in the Om64_TPR-Hsp90 complexed system, the porcupine plot shows higher restriction of movement of the two terminal helices except for the N- and C-terminal ends compared to the previously reported findings. This observed difference in dynamics following peptide binding can be attributed to the conformation adopted by the peptide on the TPR cradle.

Figure 4-8 Principal component analyses of AtOm64_TPR-Hsp systems.

Essential dynamics -2D projection of individual trajectories with their first two eigen vectors: vector 1 and vector 2 (A - C) and their corresponding porcupine plots (D - F) for Om_Apo (black), Om_C70 (green) and Om_C90 (magenta). Porcupine plots of the three systems displayed with a cone model. The length and orientation of the cone (red) is positively correlated with the magnitude and direction of motion.
4.5 CONCLUSION

Protein-ligand interaction can be investigated with the help of \textit{in vitro} methods such as crystallography and/or NMR, which give insight into the interactions at the atomic level. However these techniques necessitate the availability of non-aggregated protein sample for NMR or diffracting crystals for crystallography. In the absence of suitable samples for such studies, and in the presence of reliable structural models, molecular dynamics simulations studies is a widely used approach in structural biology wherein, along with structural details, thermodynamic parameters for interactions can be calculated. Although the theoretical values are not absolute they can follow a similar trend to that of experimental findings. Here we have used simulation studies to establish a quantitative and predictive based understanding of the interactions between the TPR domains of two receptor proteins, Toc64 and Om64, with the C-terminal octapeptides of the chaperones Hsp70 and Hsp90. In our study, the peptides were docked onto the proteins using FlexPep dock, so as to obtain the most energetically favourable conformation, which is not biased by the orientation of the peptide observed in the previous crystal structures. \textit{Ab initio} modelling reveals the TPR domain as a seven helical cradle like structure. No substantial change in conformation has been observed in TPR domains on interaction with the peptides (Scheufler et al., 2000, Xu et al., 2013). Electrostatic potential surface representations of the ligand bound forms of ArToc-TPR and ArOm64_TPR show that the peptides are bound to a predominantly positively charged surface within the cradle of the TPR domains. Additionally the peptides in the T_C90 and OmC70 systems were not oriented in the similar manner to that observed in T_C70 and Om_C90 and Hop_TPR-Hsp complexes (Scheufler et al., 2000). In the former cases, electrostatic interactions with other lysine and arginine residues substitute for the conserved clamp residues. Thus the conserved residues may not always be involved in the dicarboxylate clamp type interaction and the orientation of the peptide on the TPR cradle may differ, which is in line with the observation from Tah1-MEEVD complex (Figure 4-S1C).

Although the absolute free energies of interactions are not achievable (Deng and Roux, 2009), MMGBSA calculations, provided insights into the underlying changes in the binding free energies of interactions due to alanine mutations. Key residues involved in the interactions are identified using MD simulations (an example shown in Movie1) and a computational alanine scanning approach. Despite high sequence conservations
between the C-terminal octapeptide regions of Hsp70 and Hsp90, the interactions with Toc64_TPR differed between the two systems. A similar finding was observed for the Om64_TPR systems. This observation leads us to conclude that, not only the conserved residues within the Hsp proteins are important for interaction; the N-terminal portions of the peptides also play an important role in the interaction. Notably the intrapeptide bonds are important as they confer particular shape to the peptide, which aids in proper presentation to the protein interface for interaction. Mutation of Glu\textsuperscript{122} in T_C70 and Om_C70 and Glu\textsuperscript{121} in Om_C90 to alanine caused dramatic reduction in binding, highlighting the importance of these residues. Interestingly, the lack of interaction between Val\textsuperscript{123} of Hsp70 and the TPR domain of Om64 may partly explain the lower observed binding affinity. The importance of the conserved Lys\textsuperscript{119} in plants is not fully understood from this study. Mutation of this residue in the T_C70 system decreases binding to a small extent, whereas it improves binding in the Om_C70 and Om_C90 systems. It has been reported that, in the TPR domain from Toc64 (\textit{Pisum}), a point mutation of Arg\textsuperscript{550} to Ala reduced its interaction with Hsp70/90 by 70 - 80\% and a point mutation of Asn\textsuperscript{516} to Ala preferentially recognized Hsp70 (Qbadou et al., 2007). The residue Arg\textsuperscript{550} corresponds to Arg\textsuperscript{74} in our models, which is involved in van der Waals interaction with Glu121, Glu122 and Val123 in the T_C70 system and with Glu122 and Val123 in the T_C90 system. This interaction analysis explains the above decrease in affinity towards Hsp70/90. Further this residue also interacts with Glu121 in the Om_C70 system whereas interaction is observed with Glu122 and Val123 in the Om_C90 system. Similarly, the residue Asn\textsuperscript{516} corresponds to Asn\textsuperscript{40} in our model. Our model demonstrates interaction of this residue with Val123 and Asp124 in the T_C70 system and with Glu122 in the T_C90 system. Additionally, the computational alanine scanning experiments have shown that mutation of Glu122 to Ala in the T_C90 causes a significant decrease in binding compared to mutations of Val123 or Asp124 to Ala indicating the importance of interactions associated with Glu122. Thus it is possible that in Toc64_TPR the mutation Asn40 to Ala can lead to preferential recognition of Hsp70 over Hsp90.

The overall significant decrease in binding on mutation of residues of the peptide in the alanine scanning experiments compared to other systems explain the highest binding affinity observed in the T_C70 system. Interaction studies of \textit{ArToc64} with Hsp70/90 isoforms using \textit{surface plasmon resonance spectroscopy} (SPR) revealed similar trend in
binding affinity as observed in our studies (Schweiger et al., 2013). Similar studies with AtOm64 suggested that the protein interacts with Hsp70 with higher affinity than the Hsp90 isoforms (Schweiger et al., 2013), which contradicts our findings. In our studies, complexes of octapeptides of Hsp70 and Hsp90 with the TPR domains were obtained using in silico methods. However, in the SPR studies longer stretches of Hsp70/90 isoforms were used, suggesting that probably the peptides attain a more favourable conformation for interaction with the TPR, with increase in the number of amino acids. Essential dynamics studies demonstrate that the H1 and H7 helices move inwards along with H3, H4 and H5 helices moving backwards in order to enclose the Hsp70 octapeptide within the Toc64_TPR cradle. In contrast, in the other systems the magnitude of motion observed is comparatively lower. Though the TPRs are known to have rigid structures (Cheng et al., 2006), this study suggests that changes in curvature of the cradle structure by the concerted movement of secondary structural elements may be necessary for ligand binding. Similar kinds of motion were observed in a TPR containing protein MamA, where a 3Å radial movement by two N-terminal TPR motifs was observed in different crystal structures upon ligand imitator binding (Zeytuni et al., 2011). Our study proposes a model for the recognition of AtHsp70 or AtHsp90 by the TPR domains of AtToc64 and AtOm64 based on computational methodologies (Figure 4-9). The ArToc64_TPR has higher propensity to recognize and interact with the C-terminus of ArHsp70, whereas ArOm64_TPR has equal propensity to interact with C-terminus of ArHsp70 and ArHsp90. In terms of significance in plants, it is notable that while mis-targeting of chloroplast precursor proteins to mitochondria has been reported, there are no reports of mis-targeting of mitochondrial precursor proteins to chloroplast.
Figure 4-9 A proposed model for interaction of Toc64 and Om64 with Hsp70 and Hsp90 in Arabidopsis based on computational studies.

The preproteins after being translated by the ribosome is associated with chaperones like Hsp70 and Hsp90 in the cytosol. These chaperones help in keeping the preproteins in an unfolded conformation thus preventing aggregation. Our studies suggest that there is higher propensity of Hsp70 chaperone to be recognized by the TPR domain from Toc64 thus aiding in the import of preproteins to the Toc core. TPR domain from Om64 has no specific preference towards either of the chaperones suggesting that both Hsp70 and Hsp90 might be aiding in the import of preproteins into the mitochondria on involvement of Om64. Color codes: TPR domains in pale green; non-TPR cytosolic part of the protein in white; region of protein embedded in the membrane in dark grey; region in the inter-membrane space in light blue. AtHsp70 and AtHsp90 shown in green and pink respectively. Green and pink arrows denote interaction of TPR domains with Hsp70 and Hsp90 respectively. High affinity interaction is represented by solid line and weak ones by dotted line. OM stands for Outer membrane; IMS stands for Inter-membrane space; TPR stands for Tetratricopeptide repeat.

This may be a result of the fact that mitochondrial precursor proteins associate with Hsp90 and that Toc64 on the chloroplast surface does not bind Hsp90 with high affinity. In contrast the ability of Om64 to bind with equal affinity to Hsp70 and Hsp90, may contribute to mis-targeting of chloroplast proteins to mitochondria. Notably, this mis-
targeting of chloroplast precursors to mitochondria is abolished if chloroplasts and mitochondria are present in the uptake assays (Rudhe et al., 2002). Thus our finding explains the above-mentioned biological phenomenon of mis-targeting of chloroplast precursor to the mitochondria at a molecular level.
4.6 SUPPORTING MATERIALS

Additional figure to show the interaction of C-terminal EEVD motif of Hsp70 and Hsp90 with TPR domains.

**Figure 4-S1** Analysis of C-terminal Hsp70/90 sequences and TPR-Hsp complexes.

A. Sequence alignment of C-terminal of cytosolic Hsp70 and Hsp90 representing diverse eukaryotes. B. TPR1 domain of Hop complexed with Hsp70 octapeptide (light pink) (PDB: 1ELW). C. TPR domain from Tah1 complexed with MEEVD motif of Hsp90 (magenta) (PDB: 2L6J). The terminal aspartates of the peptides are shown in stick model. The conserved clamp residues of the TPR domains (grey) are shown as green sticks. The N- and C-terminals of the domains are labelled to highlight the difference in directionality of alignment with respect to the former.
Additional figure to show the percentage of residues in the allowed region in the I-TASSER generated model for AtToc64_TPR apo and ligand bound structures.

**Figure 4-S2** Ramachandran plots of AtToc64_TPR system bound and unbound to octapeptides from C-terminals of Hsp70/90.

A. T_Apo system; B. T_C70 system; C. T_C90 system.
Additional figure to show the percentage of residues in the allowed region in the I-TASSER generated model for AtOm64_TPR apo and ligand bound structures.

Figure 4-S3 Ramachandran plots of AtOm64_TPR systems bound and unbound to octapeptides from C-terminals of Hsp70/90.

A. Om_Apo system; B. Om_C70 system; C. Om_C90 system.
Additional figure to show the molecular surfaces of the receptors in the four complexes.

**Figure 4-S4** Electrostatic potential surface representations of TPR domains from *At*Toc64 and *At*Om64 on interaction with the C-terminal residues of *At*Hsp70 and *At*Hsp90 obtained from the last 5ns averaged structures.

*A.* *T_C70* system; *B.* *T_C90* system; *C.* *Om_C70* system; *D.* *Om_C90* system. Electronegative and electropositive charges are colored in red and blue respectively.
Additional figure to show the number of hydrogen bonds between the protein-peptide complexes through out the 50 ns trajectories.

Figure 4-S5 Number of hydrogen-bonding (black trace) and non-bonding (coloured trace) interactions observed in the trajectories after 50 ns of simulation.  
A. T_C70 system  B. T_C90 system  C. Om_C70 system; D. T_C90 system.
Additional figure to show the secondary structure change during the whole simulation (50 ns) in all the Toc64_TPR systems.

**Figure 4-S6** Secondary structure map of the trajectories of different systems created by DSSP.  
* A. T_Apo, B. T_C70 and C. T_C90 systems. Color codes: green represents alpha helices, black represents turns, grey represents for 3-10 helices and white represents coils.
Additional figure to show the secondary structure change during the whole simulation (50 ns) in all the Om64_TPR systems.

Figure 4-S7 Secondary structure map of the trajectories created by DSSP.  
A. Om_Apo, B. Om_C70 and C. Om_C90. Color codes: green represents alpha helices, black represents turns, grey represents 3-10 helices and white represents coils.
Additional figure to show the dynamic cross-correlation in the three different systems.

**Figure 4-S8** Dynamic cross-correlation maps.  
A. Toc64_TPR Apo form, B. C-Hsp70 bound form of Toc64_TPR and C. C-Hsp90 bound form of Toc64_TPR. D. Om64_TPR Apo form, E. C-Hsp70 bound form of Om64_TPR and F. C-Hsp90 bound form of Om64_TPR. For all of the maps only the Ca atoms were used. Secondary structures of Toc64_TPR and Om64_TPR are shown in red and purple cartoons respectively and the helices are numbered from H1 – H7.
Movie 1:
A molecular model of the mode of interaction of C terminal octapeptide of AtHsp70 with TPR domain from AtToc64 obtained from simulation studies. The peptide is shown in green cartoon. Eventually each residue is highlighted as a stick model in green and the corresponding interacting residues of the TPR domain and/or peptides are represented as stick models in pink. All the highlighted residues are labelled. The TPR domain is shown in grey.
Chapter 5

Discussion
5.1 DISCUSSION

Mitochondria and chloroplasts are two organelles that coexist in the eukaryotes and appear to have arisen from independent endosymbiotic events. The protein import apparatus in mitochondria has been modified in order to meet the specific needs in plant, fungal and animal lineages. Further the existence of plastids in plant cell complicates the process of protein sorting and targeting. This has led to the evolution of components, which confer specificity for targeting to destined organelles.

Metaxin is a component of the mitochondrial import apparatus located on the outer membrane of mitochondria. The plant Metaxin has limited sequence similarity with the mammalian counterparts. Metaxin is known to be a part of the SAM complex and recognizes a variety of precursors in the plant mitochondria as shown previously by yeast two hybrid and pull-down assays (Lister et al., 2007). It specifically aids in the import and assembly of β-barrel proteins on the outer membrane. It has an N-terminal domain, which is cytosolically exposed and aids in precursor recognition. Our initial aims in this project were to biophysically characterize this cytosolic domain and understand its structural details. Bioinformatics analysis suggested the presence of disordered regions in the sequence and that the protein majorly comprised of helices with random coils. Several deletion constructs were designed with hexa-histidine tag and were over-expressed in E. coli followed by purification using Ni-NTA affinity chromatography. A deletion construct comprising the N-terminal 131 residues behaved as a dimer as assessed by size exclusion chromatography. However the protein aggregated as a function of time and increasing concentration as observed by elution of the protein in the void volume by gel filtration. Further thermal melt studies using the fluorescent dye, Sypro Orange, and circular dichroism showed that the protein lacked a proper folded tertiary structure as deduced from the presence of non-sigmoidal melt curves. This phenomenon was not circumvented by the addition of buffers or detergents. As the N-terminal domain of Metaxin contains a predicted GST-fold, it was hypothesized that the recombinant protein might bind to glutathione. Binding studies with both glutathione and the presequence of alternative oxidase by ITC failed to show interaction, however an in vivo competition import inhibition assay showed inhibition of import of AOX preproteins by recombinantly purified Metaxin constructs suggesting the expressed deletion constructs were active. Based on these findings it can be concluded that a likely cause for the inability to visualize binding in vitro by ITC was
the aggregation prone nature of the construct, which prevented optimization of concentration parameters necessary for the biophysical experiments. Attempts to obtain diffracting crystals were also unsuccessful, likely due to the same reasons as above. An in silico model for metaxin and simulation studies with ligand were not possible as the amino acid sequence of metaxin shared very low identity with the sequences of structures available in the Protein Data Bank thus leading one to question the reliability of any computationally obtained homology model.

Based on our findings, the following directions might help in further studies of the cytosolic domain of metaxin.

1. As the in vivo studies have shown that recombinant metaxin is active and the protein consisting of N-terminal 131 residues is a dimer by gel filtration analysis, constructs can be designed with different truncated versions of the AOX presequence (peptide library) linked to ΔMtx131 with the help of linkers/spacers of differing lengths. Rigorous screening of the peptide library and the length of the spacer in such bicistronic constructs can aid in obtaining stable complex(es) which might be ideal for obtaining diffraction quality crystals or aid in obtaining a solution structure using NMR. This approach has been successfully applied in structural studies of the mitochondrial outer membrane receptor, Tom20, in its liganded form with aldehyde dehydrogenase presequence (Saitoh et al., 2011, Saitoh et al., 2007).

2. Based on our observations, although the recombinantly expressed ΔMtx131 is a dimer, it aggregates nonspecifically as a function of time and concentration. This could be due to the presence of exposed hydrophobic surfaces, which cause instability in the protein. Metaxin is proposed to be a part of the SAM complex. The exposed hydrophobic surface(es) might act as interaction sites for other partners of this complex. It would be interesting to investigate its interacting partners/domains from the complex using methods such as a yeast two-hybrid assay. Recombinant co-purification of these interacting domains may aid in purification of a stable complex containing the cytosolic domain of metaxin domain thereby aiding further structural studies.

Another mitochondrial outer membrane protein, which is thought to act as receptor is Om64. A paralogue of Om64 in plastids is Toc64. These proteins are integral membrane
Discussion

proteins, which interact via their C-terminal cytosolic TPR domains with the C-terminal peptides of Hsp70/90 when the latter approaches the import apparatus bound to unfolded preproteins. The TPR domains of Om64 and Toc64 share 65.8% sequence identity. Previously the C-terminal regions of Hsp70/90 from human have been used for in vivo interaction analysis with Pisum Toc64, and these studies showed that the TPR domain interacts with Hsp90 with higher affinity (Qbadou et al., 2006). In our studies, we tried to further validate and investigate the details of the above interactions in vitro using ITC and molecular dynamics simulation. ITC studies suggest that the recombinantly expressed TPR domain interacts with the C-terminal octapeptides from the two heat shock proteins with similar micromolar affinity consistent with a 1:1 stoichiometric model. These findings were further confirmed by theoretical binding energy calculations using MMPBSA. Although the binding free energy (ΔG) was similar in both the cases, the TPR-Hsp90 octapeptide interaction was enthalpically driven suggesting a conformational change in one or both of the interacting components. In contrast, the TPR-Hsp70 octapeptide interaction was entropically driven. Interactions involving the conserved pentapeptide counterpart composed of the EEVD motif demonstrated a decrease in binding affinity and showed an entropically driven interaction suggesting elimination of water molecule from the interaction surface upon ligand binding. Alanine scanning using ITC revealed a complete loss or substantial decrease in the binding affinity on mutating I/M\textsuperscript{120}, E\textsuperscript{121}, E\textsuperscript{122}, V\textsuperscript{123} and D\textsuperscript{124} individually to alanine in Hsp70/90 octapeptides respectively. Additionally, mutations of residues N-terminal to the above pentapeptide motif to alanine also reduced the binding affinity suggesting that these upstream residues may also be involved in interaction with the TPR domain. Similar finding have been observed in the TPR domains from Hop protein where TPR1 interacts with Hsp70 and TPR2A interacts with Hsp90 (Brinker et al., 2002). Indeed, it had been shown that isoleucine and methionine in the C-terminal peptide of Hsp70 and Hsp90 respectively are the primary determinants of specificity for interaction with the Hop-TPRs. Thus contrary to the reports from in vivo studies (Qbadou et al., 2006), our studies suggest that the TPR domain of Toc64 interacts with the C-terminal octapeptide of Hsp70 and Hsp90 with similar affinity.

In order to understand the basis of interaction at a molecular level, a structural model of the ArToc64TPR domain was constructed using I-TASSER. The quality of the model was assessed using a Ramachandran plot, which indicated that 93.9% of the residues
were present in favoured regions of the plot. Finally, we compared the modeled structure of \( At\text{Toc64TPR} \) with a 1.6 Å resolution crystal structure of the TPR domain from Hop (1ELW). These two TPR domains exhibit 51% sequence similarity and 39% sequence identity. Interestingly, the RMSD between the Hop TPR domain and the modeled \( At\text{Toc64} \) TPR domain was found to be quite low (1.4 Å) providing further support that the domains exhibit close structural similarity and thus supporting the validity of our model. Interaction analysis suggested a dicarboxylate clamp formation of the terminal aspartate of the ligand on interaction with the TPR domain which has been previously observed in the case of Hop-TPR-Hsp interaction (Scheufler et al., 2000). Van der Waals interactions form the key determining factor in the interactions although electrostatic and hydrophobic contacts also exist. Interestingly, although G\(^{117}\) and P\(^{118}\), of Hsp70 do not interact with the TPR domain, they are thought to play a role in conferring a conformation to the peptide necessary for protein-peptide interaction. We performed principal component analysis, which showed that the unliganded TPR sampled more conformational space than the liganded forms. Furthermore, the two terminal helices, H1 and H7, demonstrated prominent backward motions in order to expose the cradle surface upon ligand binding contrary to the apo form, thus suggesting that changes in curvature is essential for ligand binding to \( At\text{Toc64_TPR} \).

Simulation studies analysing the interaction of \( At\text{Om64} \) with \( At\text{Hsp90} \) have been previously reported (Mirus et al., 2009), however similar information characterizing the interaction with Hsp70 with \( At\text{Om64} \) was not available. Identifying whether any difference in specificity exist for the C terminal octapeptide of Hsp70/90 with respect to their interaction with TPR domains from chloroplast Toc64 and mitochondrial Om64 would provide important insights into the biological preference of Hsp70/90 towards recognition by the TPR domain from Toc64 and Om64.

The sequence of the C-terminal octapeptides of Hsp70/90 in plant differ from that of human by the presence of a lysine residue instead of T/R\(^{119}\) in Hsp70/90 respectively and a glycine residue in the place of T\(^{117}\) in Hsp90. We carried out a comparative analysis of the interactions of C-terminal octapeptide of Hsp70/90 with TPR domains from \( At\text{Toc64} \) and \( At\text{Om64} \) using computational methods. A model for the TPR domain of \( At\text{Om64} \) was generated using I-TASSER. Simulation studies were carried out to analyze the interactions. Interestingly, after FlexPep dock, these peptides were oriented
in different conformations compared to that observed previously, demonstrating the most energetically stable conformation. Furthermore, these conformations did not change throughout the simulation run. This kind of flexibility in positioning of the Hsp peptides on the TPR domain has previously been observed in the NMR ensemble structure of the TPR domain of Tah1 with the MEEVD peptide derived from Hsp90 (PDB:2L6J) (Jiménez et al. 2012). However the importance of the conserved lysine was not completely understood. Mutation of this residue to an alanine in both the Hsps improved binding to AtOm64-TPR. Essential dynamics showed the motion in the terminal helices, H1 and H7, once again supporting the role of a change of the curvature of the TPR cradle in ligand binding. Further MMPBSA energy calculations suggested that the TPR domain from AtToc64 displays higher affinity towards the octapeptide from AtHsp70 whereas the TPR domain from AtOm64 displays equal propensity to interact with AtHsp70 and AtHsp90. Schweiger et al. (Schweiger et al., 2013) have recently shown a similar interaction analysis by surface plasmon resonance spectroscopy using 45mer isoforms of Hsp70 (Hsp70.1) and Hsp90 (Hsp90.1, Hsp90.2, Hsp90.3 and Hsp90.4). Their study showed that AtToc64-TPR displayed higher association constant for Hsp70.1 than Hsp90.1, Hsp90.2 and Hsp90.3 isoforms, which supports our findings. However Hsp90.4, which also displayed a higher binding constant, cannot be explained using our model due to the smaller size of the ligand and the fact that the seven C-terminal residues are conserved in the Hsp90 isoforms. The crystal structure of Hop-TPR and our I-TASSER model suggest that the TPR domain can accommodate a maximum of eight residues. Hence the role of upstream residues beyond this region in the interaction remains to be answered.

Mis-targeting of mitochondrial preproteins to the chloroplast have not been reported to date. This may be explained by the differences in affinity between the TPR domains of mitochondria versus chloroplasts and the chaperones Hsp70 and 90. Specifically, mitochondrial preproteins may associate tightly with Hsp90 however Toc64-TPR does not exhibit high affinity towards Hsp90 (Figure 4.9). In contrast, Om64-TPR interacts with equal affinity to Hsp70 and Hsp90. This may be the cause of mis-targeting of chloroplast proteins to the mitochondria. Interestingly, this mis-targeting can be avoided if both organelles are present in the uptake assays (Rudhe et al., 2002). Hence our results provide a molecular level insight into mis-targeting of proteins in the above organelles.
REFERENCES


References


ZHANG, J. W. & LAZAROW, P. B. 1996. Peb1p (Pas7p) is an intraperoxisomal receptor for the NH2-terminal, type 2, peroxisomal targeting sequence of thiolase: Peb1p itself is targeted to peroxisomes by an NH2-terminal peptide. *Journal of Cell Biology*, 132, 325-334.

## APPENDIX

<table>
<thead>
<tr>
<th>Buffers</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD Buffer</td>
<td>10 mM K$_2$HPO$_4$/ KH$_2$PO$_4$ pH 7.5.</td>
</tr>
<tr>
<td>Coomassie stain</td>
<td>40% Methanol (v/v), 10% Acetic acid (v/v) and 5 g/L Coomassie brilliant blue.</td>
</tr>
<tr>
<td>Coomassie destaining solution</td>
<td>20% Ethanol (v/v) and 10% Acetic acid (v/v).</td>
</tr>
<tr>
<td>DNA Loading Dye (6x)</td>
<td>0.25% Bromophenol blue and 30% Glycerol.</td>
</tr>
<tr>
<td>ITC buffer for Toc_TPR</td>
<td>25 mM Hepes pH 7.3 and 100 mM KCl.</td>
</tr>
<tr>
<td>Lysis buffer for Metaxin deletion constructs</td>
<td>50 mM Hepes pH 8.0, 700 mM KCl, 10% Glycerol and 1% Sarkosyl.</td>
</tr>
<tr>
<td>Lysis buffer for Toc64_TPR construct</td>
<td>50 mM Hepes pH 7.3, 500 mM KCl, 10% Glycerol and 0.2% Sarkosyl.</td>
</tr>
<tr>
<td>Ni-NTA Binding buffer for Metaxin deletion constructs</td>
<td>50 mM Hepes pH 8.0, 700 mM KCl, 10% Glycerol and 30 mM Imidazole.</td>
</tr>
<tr>
<td>Ni-NTA Binding buffer for Toc64_TPR construct</td>
<td>50 mM Hepes pH 7.3, 500 mM KCl, 10% Glycerol and 30 mM Imidazole.</td>
</tr>
<tr>
<td>Ni-NTA Elution buffer for Metaxin deletion constructs</td>
<td>50 mM Hepes pH 8.0, 700 mM KCl, 10% Glycerol and 1 M Imidazole.</td>
</tr>
<tr>
<td>Ni-NTA Elution Buffer for Toc64_TPR construct</td>
<td>50 mM Hepes pH 7.3, 500 mM KCl, 10% Glycerol and 1 M Imidazole.</td>
</tr>
<tr>
<td>SDS-PAGE Running Buffer (1x)</td>
<td>25 mM Tris-HCl pH 8.3, 200 mM Glycine, 0.1% (w/v) SDS.</td>
</tr>
<tr>
<td>SDS-PAGE Resolving gel</td>
<td>400 mM Tris-HCl pH 8.8, 0.1% SDS (v/v), 0.1% ammonium persulfate (w/v) and 15% bisacrylamide-acrylamide (v/v).</td>
</tr>
<tr>
<td>SDS-PAGE Stacking gel</td>
<td>200 mM Tris-HCl pH 6.8, 0.1% SDS (v/v), 0.1% ammonium persulfate (w/v) and 5% bisacrylamide-acrylamide (v/v).</td>
</tr>
<tr>
<td>SDS-PAGE sample buffer (5x)</td>
<td>0.2 M Tris-HCl, pH 6.8, 20% Glycerol (w/v), 10% SDS (w/v), 10 mM βME and 0.05% Bromophenol blue.</td>
</tr>
<tr>
<td>Superdex buffer for Metaxin deletion constructs</td>
<td>20 mM Hepes pH 8.0 and 50 mM KCl.</td>
</tr>
<tr>
<td>Superdex buffer for Toc64_TPR constructs</td>
<td>25 mM Hepes pH 7.3 and 100 mM KCl.</td>
</tr>
<tr>
<td>TBE (5x)</td>
<td>54 g/L Tris base, 27.5 g/L boric acid, 20 ml/L 0.5M ethylenediaminetetraacetic acid.</td>
</tr>
</tbody>
</table>