Structure-Function Characterization of Tail-Anchored Protein Translocation Pathway in Plants

Thesis Submitted to AcSIR
For the Award of the Degree of

DOCTOR OF PHILOSOPHY
In
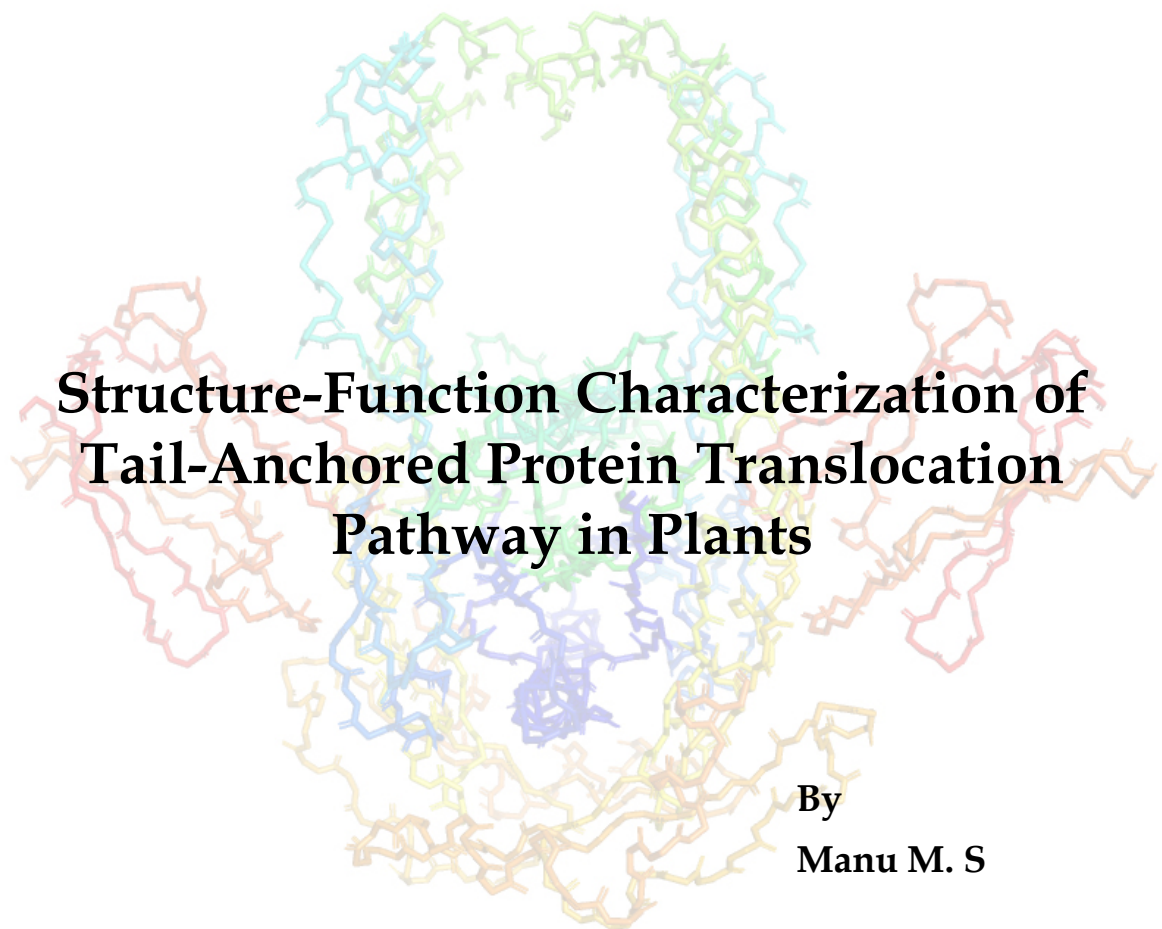## BIOLOGICAL SCIENCES

By

MANU M. S
10BB11A26055

Under the guidance of

Dr. Bhushan P. Chaudhari   **Dr. Sureshkumar Ramasamy**
(Supervisor)   (Co-Supervisor)

Biochemical Sciences Division
CSIR-National Chemical Laboratory
Pune – 411008, India

November 2018

Structure-Function Characterization of Tail-Anchored Protein Translocation Pathway in Plants

By

Manu M. S

सीएसआईआर – राष्ट्रीय रासायनिक प्रयोगशाला
(वैज्ञानिक तथा औद्योगिक अनुसंधान परिषद)
डॉ. होमी भाभा मार्ग, पुणे - 411 008, भारत

CSIR - NATIONAL CHEMICAL LABORATORY
(Council of Scientific & Industrial Research)
Dr. Homi Bhabha Road, Pune - 411 008, India

Certificate

This is to certify that the work incorporated in this Ph.D. thesis entitled **Structure-Function Characterization of Tail-Anchored Protein Translocation Pathway in Plants** submitted by Mr. **Manu M. S.** to Academy of Scientific and Innovative Research (AcSIR) in fulfillment of the requirements for the award of the Degree of **Doctor of Philosophy**, embodies original research work carried out under my supervision. I further certify that this work has not been submitted to any other University or Institution in part or full for the award of any degree or diploma. Research material obtained from other sources has been duly acknowledged in the thesis. Any text, illustration, table etc., used in the thesis from other sources, have been duly cited and acknowledged.

Dr. Bhushan P. Chaudhari
(Research Supervisor)

Manu M. S.
(Research Student)

Dr. Sureshkumar Ramasamy
(Research Co-supervisor)

Biochemical Sciences Division
CSIR-National Chemical Laboratory
Pune-411008

Place: Pune
Date: 16-11-2018

Communication Channels

☎
NCL Level DID : 2590
NCL Board No. : +91-20-25902000
EPABX : +91-20-25893300
: +91-20-25893400

FAX
Director's Office : +91-20-25902601
COA's Office : +91-20-25902660
SPO's Office : +91-20-25902664

WEBSITE
www.ncl-india.org

Declaration of Authorship

I hereby declare that this Ph.D. thesis entitled **Structure-Function Characterization of Tail-Anchored Protein Translocation Pathway in Plants** was completely carried out by me for the degree of **Doctor of Philosophy** in **Biological Sciences** under the guidance and supervision of **Dr. Bhushan P. Chaudhari (Research Supervisor)** and **Dr. Sureshkumar Ramasamy (Research Co-Supervisor)**, CSIR-National Chemical Laboratory, Pune, India.

I confirm that this thesis research is my own work while in candidature for a research degree at this institution and the contents of this thesis is original. I also affirm that no part of this thesis has previously been submitted for a degree or any other qualification at this institution or any other institution. Also, the interpretations put forth are based solely on my knowledge and understanding of the original research articles and all resources have been duly cited and acknowledged as and when appropriate.

Place: Pune
Date: 16 - 11 - 2018

Manu M. S
(Research Student)
Biochemical Sciences Division
CSIR-National Chemical Laboratory
Pune-411008

"If we knew what it was we were doing, it would not be called research, would it?"

- Albert Einstein

Dedicated to.....

My Family

Acknowledgements

It is my pleasure to take this opportunity to thank all the people who helped me throughout this journey of my PhD.

My utmost and sincere gratitude is to **Dr. Sureshkumar Ramasamy**, who supported me completely and helped me in solving all the queries related to this research topic. Although I am new to this field, I was lucky to have him as my supervisor because of which I could now learn structural biology. He has offered his help in several difficulties and I am thankful to him for being there for me in every moment of need. I am privileged to mention here that I feel very lucky to work with him for my PhD.

My profound gratitude is to **Dr. Bhushan P. Chaudari**, for allowing me to pursue my degree under his supervision. I would like to extend my sincere gratitude to **Dr. C. G. Suresh**, for initial supervision of my PhD. He has been a very good mentor for me in teaching the basics about structural biology. I remember and cherish all the thoughtful discussions with him and lectures of him. He has supported me in several situations both in profession and in personal aspects. I also thank **Dr. Bhaskar Saha**, NCCS, Pune, with whom I started my initial phase of PhD.

I am also grateful to my Doctoral Advisory Committee (DAC) - Drs. Dhanashekaran Shanmugam, Mahesh J Kulkarni and H V Thulasiram, for their valuable interactions and suggestions in every DAC meetings. I extend my thanks to Dr. M. V. Deshpande, Dr. Kiran Kulkarni, Dr. Subhash Chandra Bose, Dr. Narendra Kadoo and Dr. Anu Raghunathan for their help and interactions.

I am happy that I worked in Biochemical Division. At this moment, I would like to thank the **Heads of Division** during my tenure for making the work atmosphere peaceful and pleasant. External expert committee for my fellowship extension and screening, Dr. Saikrishnan Kayarat and Dr. Thomas J Pucadiyl is also remembered and acknowledged for their presence and evaluations. I also thank **Prof. Bil Clemons** and **Justin Chartron** at **Caltech, Pasadena, USA** for providing AtGet3 clone; **Dr. Ravi Maruthasalam** at IISER, TVM, India for knockout seeds.

My biggest acknowledgement and love to my dearest friend, **Dr. Deepak Chand** for his constant support throughout my PhD. I would say that this journey wouldn't be possible without him as a companion in the work place. He helped and encouraged me in several situations. He has also offered his help in data collection for XRD data and troubleshooting some experiments.

I would like acknowledge Dr. Nishant Varshney for his help in data collection at ELETTRA, Italy; Selva Rupa Immanuel for confocal microscopy; Shakuntala and Shanthakumari for her help in acquiring mass spectrometry data; Priyanka, Vaishnavi, Rahul and Thejas for their help in growing *Arabidopsis*; Dr. Anjan Banerjee and Ron Sunny (IISER, Pune) for microtomy; NCL venture center, Pune for instruments, NCCS Pune and NCL pune for crystallization facility ; The EMBL staff Dr. Hassan Belrhali or/and Dr. Babu A. Manjasetty for providing support on the beamline and EMBL-

Table of Contents

Chapter 3: Analysis of TA Protein Targeting Pathway in Plants   45

Chapter 4: Functional characterization of Get3 in plants   63

Chapter 5: Structural characterization of Get3 in plants   81

Chapter 6: Summary and Conclusion   105

Appendices 115

Bibliography..143

List of Publications 155

Author's curriculum vitae 157

*****~~*****

Abbreviations

aa	Amino acids
ADP	Adenosine 5′-diphosphate
AMP-PNP	Adenylyl-imidodiphosphate
AtGet3	*Arabidopsis thaliana* Guided Entry of Tail-anchored protein 3
ATP	Adenosine 5′-triphosphate
AU	Absorbance unit
BLAST	Basic Local Alignment Search Tool
BME	beta-mercaptoethanol
bp	Base pair
BSA	Bovine serum albumin
CCP4	Collaborative Computational Project No. 4
DAB	3,3′-Diaminobenzidine
DOPE	Discrete optimised protein energy
DTT	Dithiothreitol
ER	Endoplasmic reticulum
GET	Guided Entry of Tail-anchored protein
HRP	Peroxidase from horseradish
HSP	Heat shock protein
IP	Immunoprecipitation
IPTG	Isopropyl β-D-1-thiogalactopyranoside
iTOL	Interactive Tree Of Life
LB	Luria-Bertani media
Ns	nanoseconds
OD	Optical density
PBS	Phosphate buffered saline
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
rcf	relative centrifugal force
RMSD	Root mean square deviation
Rpm	Revolutions per minute
ScGet3	*Saccharomyces cerevisiae* Guided Entry of Tail-anchored protein 3
SDS	Sodium dodecyl sulphate
TA protein	Tail-anchored protein
TM	Transmembrane
TMD	Transmembrane domain

List of Figures

*****~~*****

List of Tables

*****~~*****

Thesis Abstract

Tail-anchored (TA) proteins are a special class of membrane proteins that carry out vital functions in all living cells. Targeting mechanisms of TA proteins are investigated as the best example for post-translational protein targeting in yeast. Of the several mechanisms, Guided Entry of Tail-anchored protein (GET) pathway plays a major role in TA protein targeting. This study majorly investigates GET pathway in plants by selecting *Arabidopsis thaliana*, *Oryza sativa* subsp. Indica and *Solanum tububerosum* as model systems. **In this thesis,** both experimental and *in silico* analyses have been performed to characterize the GET pathway. From the *in silico* analyses, 508 and 912 TA proteins are identified in *Oryza sativa* subsp. Indica and *Solanum tuberosum* respectively and their localization with respect to endoplasmic reticulum (ER), mitochondria, and chloroplast has been delineated. Similarly, the organelle associated GET proteins were also identified and analysed thoroughly. Our analysis revealed that the chloroplast specific Get3, which ferries the TA protein to chloroplast has multiple paralogs with different domain architectures. In *Arabidopsis thaliana*, one of the paralogs (AtGet3b) exhibits significant structural/sequence similarity to its cytoplasmic counterpart from yeasts. But the other paralog, AtGet3d distinctly possesses an HSP domain at C-terminal. Also, our data on pull down and mass spectrometry analysis suggests that AtGet3d might be involved in membrane protein homeostasis. It is very interesting to establish the relationship between the TA protein targeting and membrane protein quality control. It is fairly well recognized that Get3 itself operates as a molecular chaperon in stress conditions in other eukaryotes. In this context, results from the present study reveal that Get3d with its extra HSP domain probably functions as a dual headed chaperone and play a crucial role in chloroplast membrane protein homeostasis, as it is well conserved across the plant kingdom.

Here-in, this work reports the crystal structure of AtGet3d, for the first time and demonstrates that AtGet3d shows structural similarity to yeast Get3 and possesses a HSP domain at the C-terminal. It also demonstrates the domain fusion event in Get3 with HSP domain for the first time. AtGet3d also exhibits conserved ATPase activity and localized to the chloroplast. AtGet3d specifically binds and targets the TA proteins to chloroplast and has been found to be expressed during osmotic, salt and heat stress. Further, it interacts with many proteins associated with protein quality control through the HSP domain.

Thesis Overview

The detailed experimental and *in silico* analyses in this thesis is organised into six chapters.

Chapter 1: Introduction

Chapter 1 introduces the content of the thesis with the general introduction and literature review of TA protein targeting pathway, GET pathway and its components. This chapter also describes the physiological role of Get3 on the basis of previous works from various sources. Recent works from plant GET pathway are also described in this chapter.

Chapter 2: Materials and Methods

All materials and methods used in this thesis are depicted in chapter 2. This contains the details of protein purification, crystallization, structure determination, co-immuno precipitation, co-expression, pull down and other techniques used in this thesis.

Chapter 3: Analysis of TA Protein Targeting Pathway in Plants

Chapter 3 deals with the detailed *in silico* analysis of TA proteins and its targeting pathway in plant system. In this study, *Arabidopsis thaliana* along with two major crop plants *Oryza sativa* and *Solanum tuberosum* were analysed in-depth for TA proteins and its targeting mechanism. All analysed plant species have less than 2% TA proteins while comparing to their total proteome. Also, the number of Get3 differs across the plant species studied. These identified Get3 are highly organelle specific in TA protein targeting. The phylogenic analysis shows that Get3 in clade – d targets TA proteins to chloroplast that have a domain fusion event with α-crystalline domain.

Chapter 4: Functional characterization of Get3 in plants

Chapter 4 describes the functional characterization of *A. thaliana* Get3, with emphasis on AtGet3d. Immunofluorescence and co-expression studies show that AtGet3d localized to chloroplast surface and can bind with chloroplast TA proteins. Co-immunoprecipitation and LC-MS/MS identified the proteins that are interacting with AtGet3d through HSP domain. AtGet3d is non-essential for the growth of the plant and have slow ATPase activity.

Chapter 5: Structural characterization of Get3 in plants

This chapter details the structural analysis of chloroplast AtGet3d. The AtGet3Δd structure was solved at 2.5Å resolution. This chapter here-in reports the first crystal structure of Get3 from plant system. Also, Homology modelling was employed for the structural studies of

Get3 in *O. sativa* and *S. tuberosum*. All the refined structures were compared with the structure of ScGet3 and All4481 protein from Nostoc sp. PCC 7120. All analysed chloroplast Get3d have HSP domain at C-terminal region.

Chapter 6: Summary and Conclusion

This chapter concludes the thesis by emphasizing the important findings from this research work and by highlighting some of the future perspectives.

Chapter 1

Introduction

All biological cell membranes are occupied with several types of membrane proteins. These membrane proteins perform important functions of the cells as enzymes, transporters and receptors. The nascent proteins synthesized in the cell are often localized to their target locations for their specific functions. The biological question of interest here is: **How do these proteins reach their target location?** The mechanisms that facilitate these nascent proteins to reach their target locations are generally known as "protein targeting" or "protein sorting". Günter Blobel (awarded Nobel Prize in 1999) discovered that most of the newly-synthesized proteins have a targeting signal sequence in them that determine the final location of a protein in the cell (Nałecz n.d.). Each organelle in the cell also has its own proteins targeted to their location. The proteins belonging to different organelles can be targeted specifically either by co-translational targeting or by post-translational targeting mechanism that is quite essential for their functions. This necessitates the importance for unveiling the protein targeting mechanism that becomes crucial for each protein to function. With this motivation and background, the complete doctoral research work encompassed in **this thesis** "Structure-Function Characterization of Tail-Anchored Protein Translocation Pathway in Plants", deals with exploring a recently identified post-translational protein targeting pathway known as GET (Guided Entry of Tail-anchored translocation) pathway. Here-in, GET pathway is analyzed in plant system with a special attention to a chloroplast targeting protein, Get3d. This pathway is dedicated to target and localize a specific type of proteins called "Tail-anchored" proteins to their specific target location.

1.1. Proteins and protein synthesis

Proteins are the most abundant macromolecules present in every cell. They are the final product of central dogma through which genetic information is expressed (Crick 1970). Proteins are polypeptides that are made up of amino acids linked by special covalent bonds called "peptide bonds". Depending on the coding sequence of DNA, the amino acid composition of each protein will vary that ultimately results in functional diversity. In all biological systems, proteins are synthesized by a process called "translation". During translation, the ribosomal machinery converts the transcribed genetic information stored in the triplet codon of mRNA to proteins (Clancy Suzanne 2008).

1.2. Protein structures

In general, protein structures are formed by three-dimensional arrangement of the polypeptide

chain. These structures can be classified into four types as, primary, secondary, tertiary and quaternary structures. The primary structure is a simple linear arrangement of amino acids in the polypeptide chain. These linear polypeptide chains can also fold in a particular manner to form local sub-structures called secondary structures. Two main types of secondary structures are α-helix and β-sheet. These α-helixes and β-sheets further fold into a compact globular structure to form the tertiary structure. Multiple subunits of proteins fit together to form the quaternary structures. These structures are mainly stabilized by non-covalent interactions, hydrogen bonds and salt bridges (Lehninger Albert L, David L. Nelson 2000).

1.3. Functions of proteins

As discussed earlier, proteins perform a vast variety of functions (Table 1.1). They are classified according to their functions as enzymes, transporters, storage molecules, structural components and messengers (Anon 2010).

Table 1.1: General functions of proteins

Functions	Descriptions	Example
Enzyme	Proteins as enzymes accelerate or catalyze the chemical reactions.	Cellulases
Transport/storage	Helps in the transport of small molecules within the cell and between the cells	Ferritin
Structural components	Proteins provide structural integrity to the cell	Actin
Messenger	Interconnected protein networks help in signal transduction and signal amplification.	G-protein coupled receptors.

1.4. Membrane proteins

Other than functional classification, proteins can also be considered as soluble proteins and membrane proteins, based on cellular location. The hydrophobic regions in the soluble proteins are masked inside the structure and hydrophilic regions are exposed to the solvent. Proteins that are present in the membrane and associated with the functions of biological membranes are called "membrane proteins". As these are embedded in biological membranes, their hydrophobic regions are masked by the lipid bilayer. There are two kinds of membrane proteins: (i) integral membrane proteins and (ii) peripheral membrane proteins. Integral membrane proteins are also known as transmembrane proteins with at-least one transmembrane domain that spans across the membrane. Unlike integral membrane protein,

peripheral proteins are attached to the surface of the membrane through hydrophobic, electrostatic and other non-covalent interactions.

1.5. Types of membrane proteins

Membrane proteins can be classified into different types by their topology and biogenesis. According to Chou and Cai, 2005, there are six types of membrane proteins (Figure 1.1) that include (1) type I single-pass membrane proteins, (2) type II single-pass membrane proteins (3) multi-pass transmembrane proteins, (4) lipid chain-anchored membrane proteins (5) GPI-anchored membrane protein and (6) peripheral membrane proteins. Type I membrane proteins have an extracellular N- terminus and a cytoplasmic C- terminus. Type II membrane proteins are just reverse of Type I membrane proteins, having an extracellular C-terminus and a cytoplasmic N-terminus. Type III is a multi-pass membrane protein, where polypeptide chain passes the lipid bilayer multiple times. Both lipid chain-anchored membrane proteins and GPI-anchored membrane proteins are membrane-anchored proteins. Lipid chain-anchored membrane proteins establish an association with the membrane by covalent bond through prenyl group. While in GPI-anchored membrane proteins, the interaction with lipid bilayer is facilitated by glycosylphosphatidylinositol (GPI) anchor. Peripheral membrane proteins bond to the membrane indirectly through non-covalent interaction with other membrane proteins.
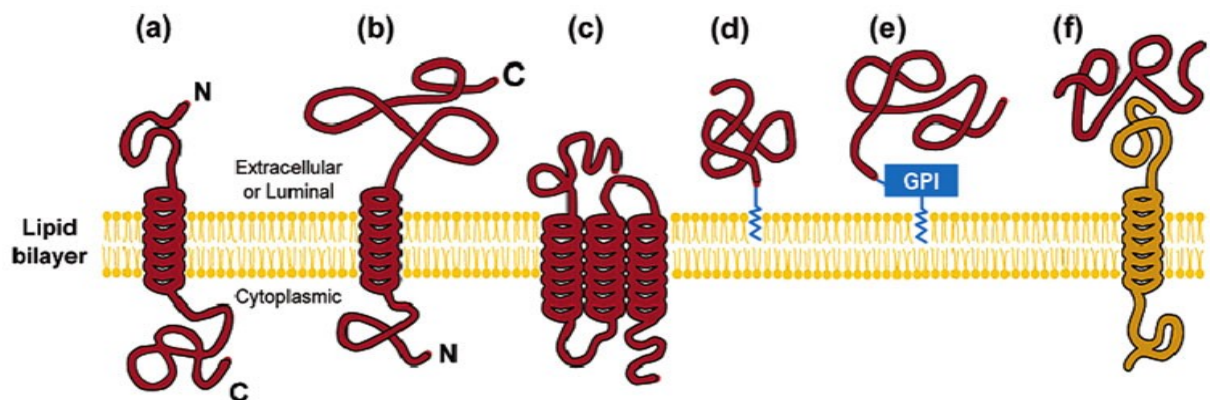
Figure 1.1: Types of membrane proteins (Schematic representation). (a) type I (b) type II (c) multipass trans-membrane protein (d) lipid chain-anchored membrane protein (e) GPI-anchored membrane protein (f) Peripheral membrane proteins (figure adapted from (Chou & Cai 2005)).

Besides these above-mentioned membrane proteins, some proteins also differ in their topology and biogenesis. Example of one such special type of membrane protein is Tail-anchored membrane proteins that do not belong to any of the above-said classifications

because of their unique features.

1.6. Tail-anchored membrane proteins

Tail-anchored (TA) proteins or C-terminal TA proteins, are a special class of single-pass membrane proteins, characterized by (i) an N-terminal functional cytosolic domain, (ii) a single transmembrane domain near to C-terminal and (iii) a short lumenal polar sequence (Figure 1.2) (Kutay et al. 1993). TA proteins are present in all the three forms of life (Archea, bacteria and eukaryotes) and involved in several vital functions such as vesicular fusion, protein translocation, enzyme catalysis and photosynthesis (Beilharz et al. 2003; Kalbfleisch et al. 2007; Pedrazzini 2009; Kriechbaumer et al. 2009; Borgese & Righi 2010). Examples of known TA proteins and their reported functions are given in the Table 1.2.
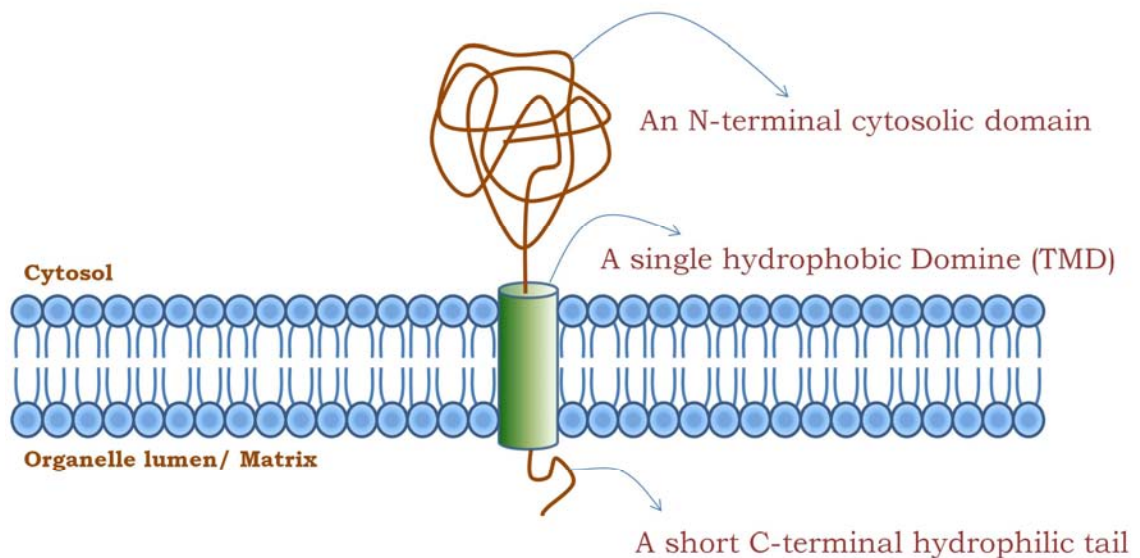
Figure 1.2: Cartoon representation of TA protein structure.

Table 1.2: Functions of TA proteins

Function	Examples	Localization
Enzymatic	Cytochrome b(5), Heme oxygenase I and II, UBC6	ER, MOM
Protein localization	Sec61γ, Sec61β,TOM5, TOM6, Pex15p, OMP25	ER,MOM, Peroxisomes
Vesicular traffic	Target SNAREs (Syntaxins), Vesicular SNAREs (e.g., Synaptobrevins), Giantin	Target membranes for vesicular fusion
Regulation of apoptosis	Bcl-2, Bcl-XL, Bax	MOM, ER
Constituent of the viral envelope	Us9 protein of α herpes viruses	Trans-Golgi network

TA proteins are present in almost all the cellular organelles inside the cell. Unlike other single-pass membrane proteins, TA proteins do not have an N-terminal signal sequence. The TMD located towards the C-terminal carries the signal for the TA protein to reach the target location (Hwang et al. 2004). This specific targeting is also determined by several factors such as overall hydrophobicity, length of TMD, physical and chemical properties of amino acid sequence etc.

1.7. Protein targeting pathways

The synthesis of proteins is mainly carried out in the cytoplasm of the cell. The free ribosomes and membrane-associated ribosomes continuously synthesize the proteins according to the genetic code from mRNA. These synthesized naïve proteins need to reach their target location in order to perform their functions. How are these naïve proteins transported to their target locations once it is synthesized?. It is an interesting research problem to explore.

Several protein targeting pathways are operational inside the cell that ensure the precise targeting of newly synthesized proteins to their target locations. Any defect in the targeting mechanism resulting in abnormalities in the living cells leading to disorders. Most of the proteins have a continuous stretch of amino acid sequence called "signal sequence" that carries the targeting information. This signal sequence can be found at the starting of proteins or in the internal portion of the proteins (Nałecz n.d.). Depending on the location of these signal sequences, the protein translocation can be a co-translational translocation or a post-translational translocation.

1.7.1. Co-translational translocation

Most of the secretory proteins and membrane-bound proteins follow co-translational targeting pathway (Figure 1.3). This is an evolutionarily conserved mechanism for secretory protein targeting. In this pathway, the nascent proteins with N-terminal signal sequences are identified by a signal recognition particle (SRP) at the ribosome. This SRP, then targets this specific nascent protein with ribosome to the translocon of Endoplasmic reticulum (ER) through the SRP receptor.

On translocon, ribosome continues the protein synthesis and the newly synthesized protein is transferred directly into the ER lumen. Signal peptidases that are associated with the translocon then cleave the N-terminal signal peptide of the nascent protein upon completion

of the translation process. Inside the lumen, proteins undergo post-translational modifications and properly fold into its final conformation (Nyathi et al. 2013).
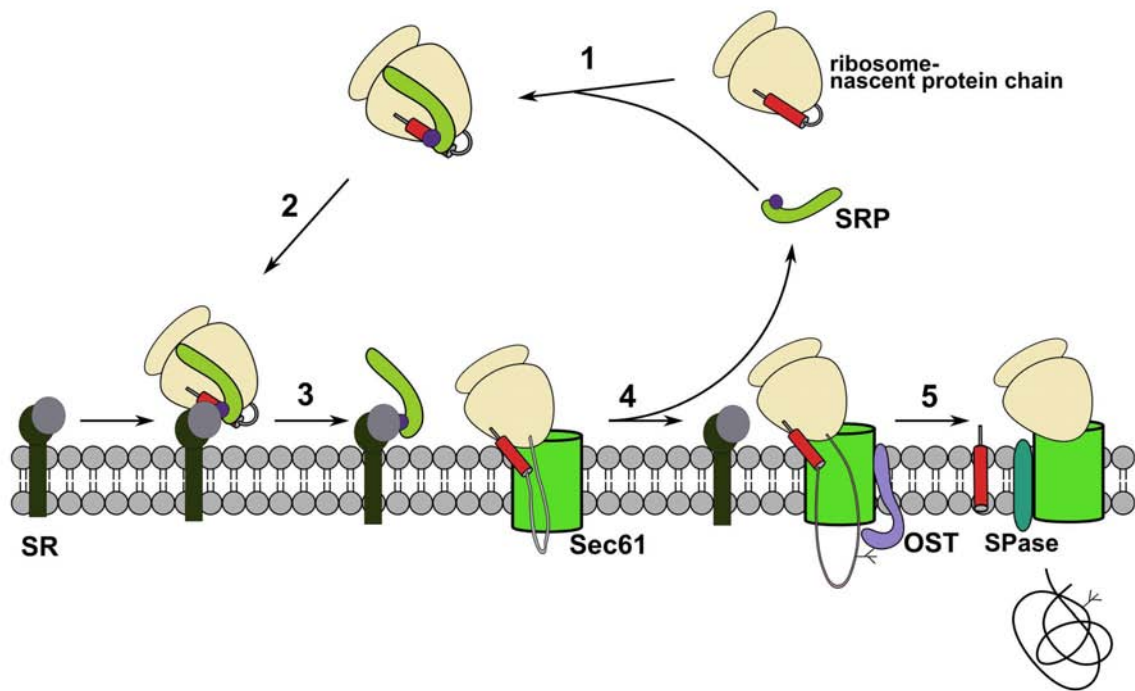
Figure 1.3: SRP mediated co-translation protein targeting pathway (based on Nyathi, Wilkinson and Pool, 2013). (1) SRP recogonise and binds with N-terminal signal sequence of nascent proteins. (2) SRP takes them to ER membrane and tether with SRP-receptor (SR). (3) Ribosome-nascent protein chain gets transferred to translocon (Sec61) of ER membrane. (4) SRP gets released for next cycle. Ribosome continues to synthesize the protein. Oligosaccharyltransferase (OST) adds N-linked glycans to the protein. (5) Signal peptidase(SPase) cleaves the N-terminal signal and the properly folded protein will release to ER lumen.

1.7.2. Post-translational translocation

In post-translational pathway, proteins are targeted to their location after the complete translation process. It is not necessary that the targeting signal should reside at the N-terminal location. The internal signal for the target location of nascent proteins is identified by cytosolic proteins like Get3/Asna1, HSP70, HSP40 and Pex19 (Figure 1.4). Then these nascent proteins will be taken to the target location by binding the membrane receptors present on the target membrane (Johnson et al. 2013). Since the targeting signal for the TA proteins is located at the C-terminal TMD, the co-translational signal recognition particle (SRP) mediated targeting pathway cannot function properly in this case. Hence, most of the TA proteins are targeted post-translationally. This targeting mechanism can be divided into

two, (i) unassisted and (ii) assisted. In unassistant mechanisms, TA proteins do not require any assistant protein to reach the target location. But in the assisted mechanism, TA proteins require several chaperones to reach the specific target. In general, the assisted mechanism can further be classified into three types (i) SRP mediated (ii) HSP70/90 mediated and (iii) Get3 mediated (Borgese & Fasana 2011).

Recently explored Guided entry of Tail-anchored Protein (GET) pathway is an example for post-translational translocation pathway. This pathway is responsible for the proper targeting of TA proteins to endoplasmic reticulum (Denic et al. 2013).
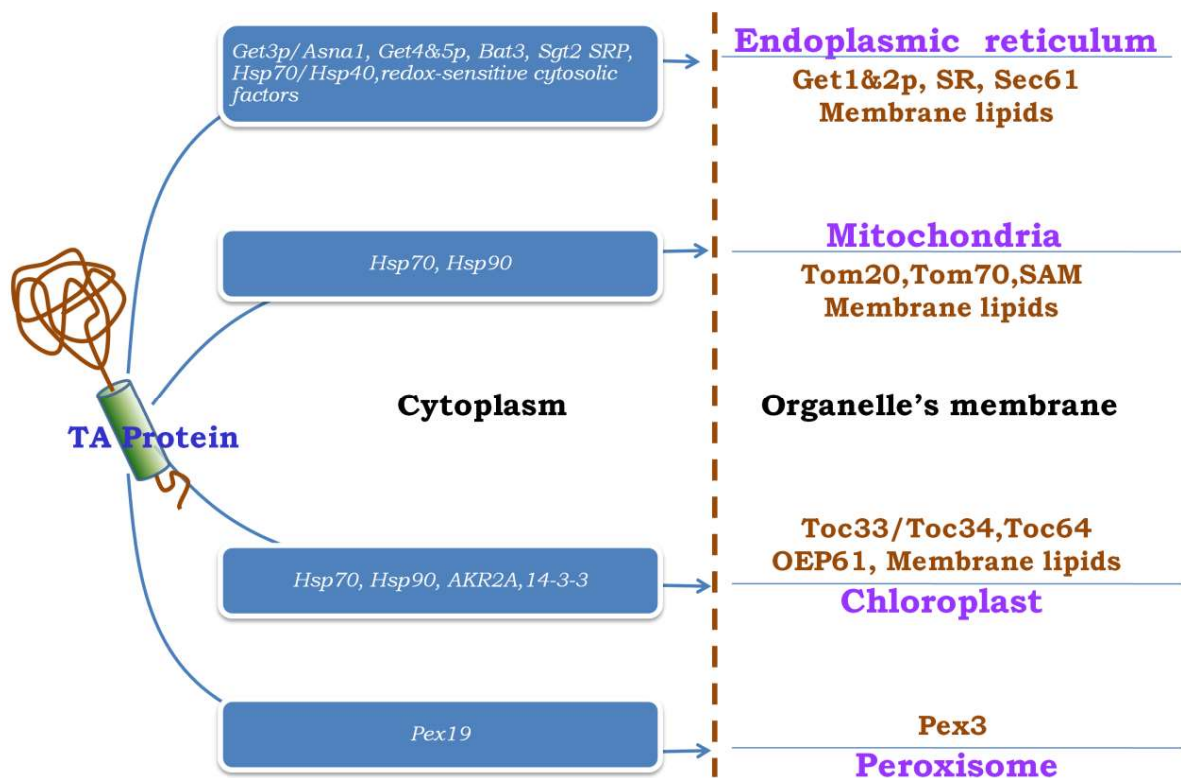
Figure 1.4: Post-translational TA protein targeting pathways. TA proteins are properly targeted to its location with the combined action of both cytoplasmic targeting factor and their membrane counterparts.

1.8. Guided entry of tail-anchored protein (GET) pathway

GET pathway is a most conserved pathway from archaea to higher plants for targeting tail-anchored proteins. This pathway is well explored in *Saccharomyces cerevisiae*. The timely and coordinated action of GET pathway component ensures the accurate insertion of TA proteins to their target location specifically (Table 1.3).

Table 1.3: Location of GET pathway proteins in Yeast

Members	Location
Get1 Get2	ER Membrane
Get3 Get4 Get5	Cytosol
Ydj1 Sgt2 Bag6	Cytosol (Accessory Proteins)

The main component in this pathway that connects pre-targeting and post-targeting complex is Get3. Get3 was initially annotated as Arr4 because of its sequence homology with ArsA (Leipe et al. 2002). Other major and accessory components of this pathway include Get5, Get4, Get2, Get1, Sgt2 and Ydj1 (Denic 2012).

1.8.1. Mechanism of GET pathway

GET pathway is initiated by the binding of Get4/Get5 to the trans-membrane domain (TMD) of nascent TA proteins (Figure 1.5). This interaction is assisted by Sgt2/Ydj1 complex. Get4/Get5 interaction with hydrophobic TMD prevents TA proteins from aggregation in the cytosol. This pre-targeting complex can be recognized by Get3. Get3 can change its conformation from open to close depending on the presence of nucleotides.

Get3 now binds with Get4 in the presence of ATP and loads the TA protein. ATP hydrolysis converts this open form to close form and ensure the tight binding with TMD of TA protein (Mariappan et al. 2011; Chang et al. 2010; Gristick et al. 2014; Wang et al. 2014). This Get3-TA complex is then directed to the target location specific to the signal sequence. Get3 then binds with its receptors, Get2 and Get1, on the surface of ER and insert the nascent TA protein into the ER Membrane.

In *S. cerevisiae,* Get3 interacts with TMD of TA protein specific for ER. Deletion or mutation in the ScGe3 leads to misinsertion of ER TA proteins into mitochondria or accumulation in the cytosol (Schuldiner et al. 2008).
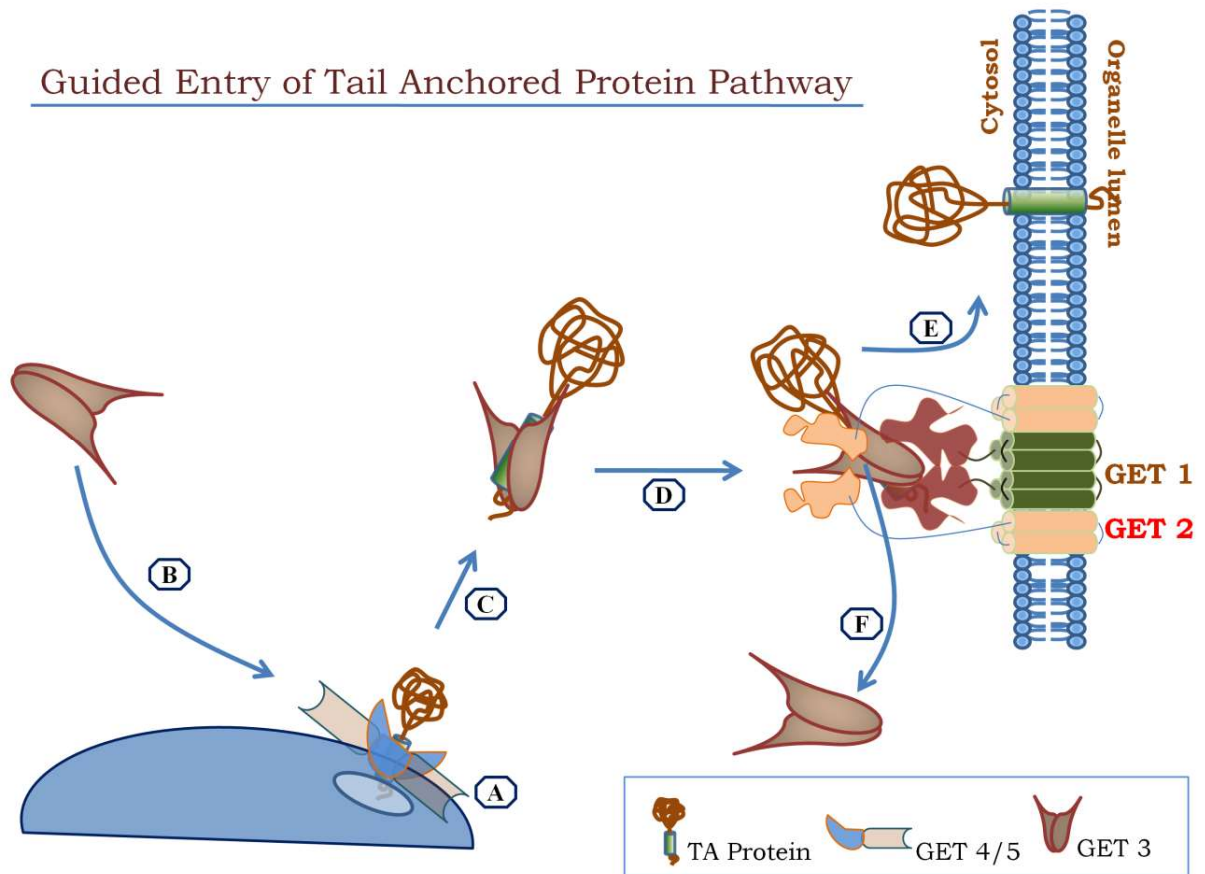
Figure 1.5: GET Pathway for TA protein targeting. (A) Binding of Get4/5 complex with nascent TA protein. This interaction stabilizes the nascent TA propein from aggregation. (B) Get3 binds with TA protein and release Get4/5 complex. (C) Get3 form a stable complex TA protein (Get3-TA complex). (D) Get3-TA complex will get targeted to ER membrane and coupled to Get1/2 complex (E) TA protein gets insert into target membrane (F) Regeneration of Get3 for next Cycle.

1.9. Components of GET pathway

1.9.1. Get5

Get5 helps to establish the interaction of Sgt2 and Get4. The amino domain of Get5 interacts with Get4 and form a heterodimer. Also, Get5 interacts with Sgt2 through its ubiquitin-like domain (Ubl). C-terminal homodimerization domain of Get5 interacts with Get4 to form a heterotetrameric Get4/Get5 complex (Chartron et al. 2011).

1.9.2. Get4

Get4 is a highly conserved protein localized in the cytoplasm. It interacts with Ge3 and Get5. Even though Get4 is nonessential, knocking out this in yeast leads to sensitivity in several growth conditions. Get4 forms dimeric complex with Get5 (Chang et al. 2010).

1.9.3. Get3

Get3 is a highly conserved ATPase, initially annotated as Asna1 because of its homology with bacterial arsenic export pump ArsA. In bacteria, ArsD transfers the As(III) to ArsA ATPase (catalytic subunit of ArsAB pump) (Lin et al. 2006; Ye et al. 2010). Both the mammalian homolog TRC40 and yeast Get3 selectively bind to TMD of TA proteins in the cytosol.

Get3 is the major player in GET pathway. Get3 connect pre-targeting (Get5/4 + TA protein) and post-targeting complexes (TA protein + Get1/2). Get3 specifically binds to the conserved surface of Get4 in a nucleotide-dependent manner. During this binding process, Get4/5 complex transfers TA protein to Get3. Hydrolysis of nucleotide converts open form of Get3 to closed form. This ensures the tight binding of Get3 to TM of TA protein (Mateja et al. 2009) as mentioned earlier in the mechanism.

1.9.4. Get1 and Get2

Get1 and Get2 are integral membrane proteins present on the surface of ER membrane. The mammalian counterpart of Get1/Get2 is WRB/CAML. Get3-TA protein complex dock with Get2/Get1 and insert TA protein into the ER membrane. Get2 initially binds with Get3-TA targeting complex, while Get1 facilitates the release of TA protein from Get3. TA protein interacts with TMD of Get2 and Get1. With the help of this interaction, TA protein gets inserted into ER membrane (Mariappan et al. 2011). Recent studies reveal that single Get1/2 heterodimer is sufficient for TA insertion to ER membrane (Zalisko et al. 2017).

1.9.5. Sgt2

Sgt2 is a heat shock protein (HSP) co-chaperone that make the first decision step in TA protein targeting. The internal tetratricopeptide repeat (TPR) domain of Sgt2 can bind with several families of HSP proteins (Chartron et al. 2011). C-terminal domain of Sgt2 can bind with hydrophobic amino acid sequences of length six or more (Liou & Wang 2005). N-terminal homodimerization domain interacts with a single molecule of Get5 (Chartron et al. 2011; Kohl et al. 2011; Chang et al. 2010).

In GET pathway, C-terminal of Sgt2 interacts with the TM domain of TA protein and the N-terminal domain interacts with Get5. This triggers the ER-destined TA protein targeting. Studies showed that the knockout of Sgt2 caused mislocalization of ER-targeted TA proteins to Mitochondria (Costanzo et al. 2010).

1.10. Structural features of GET pathway components

1.10.1. Structural features of yeast Get3

Structures of yeast Get3 have been delineated through several independent studies (Mateja et al. 2009; Suloway et al. 2009; Hu et al. 2009; Yamagata et al. 2010). Crystal structures show that Get3 exists as a symmetric homodimer. Each monomer of Get3 is composed of a core nucleotide-binding subdomain and an α-helical subdomain. Each nucleotide-binding domain can accommodate single molecule of ATP. Besides this, the ATPase subdomain of Get3 has structural similarity with SIMIBI protein family (Leipe et al. 2002). Structures of both nucleotide-free and nucleotide-bound forms have been determined (Figure 1.6A and Figure 1.6B). The binding of nucleotide led to conformational changes. In the nucleotide-free state, Get3 occurs as open form, while the nucleotide binding led to closed conformation. The dimer interface area of the closed form ($\sim$2400Å^2) was found to be much higher compared to open state ($\sim$900 Å^2).

| Open dimer | Closed dimer | Get3-TA complex |
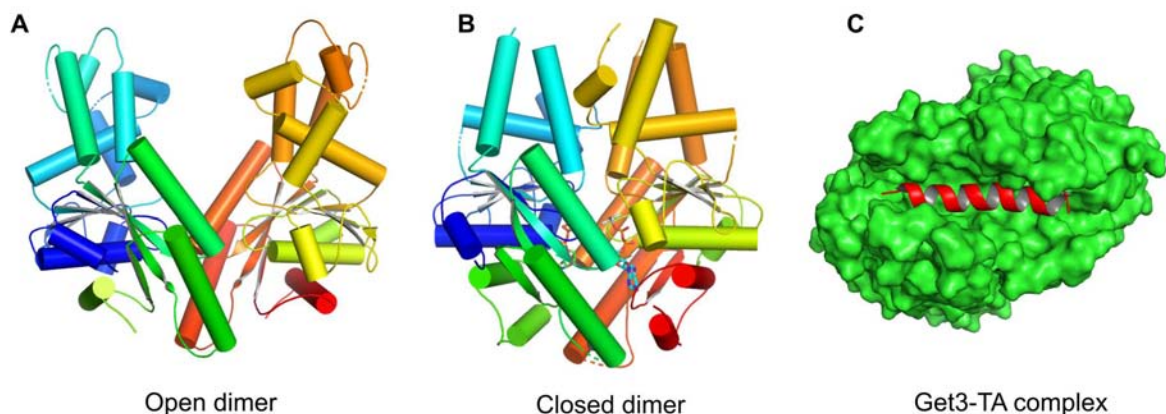
Figure 1.6: Yeast Get3 structures (A) Open dimer of Get3 (*S. pombe*) (B) Closed dimer of Get3 (*S. cerevisiae*) (C) Complex of Get3 with TMD of Pep12.

In Yeast, the structure of Get3 is a dynamic homodimer, stabilized by zinc ion at the CXXC motif. Any mutation in cysteine in CXXC motif was found to cause serious growth defects in yeast. The TA protein binding groove of Get3 is composed of ten helices of α-helical subdomain. In closed dimer state, α-helical subdomain comes closer to form a continuous hydrophobic groove. This groove has a length of $\sim$30Å and diameter of $\sim$15Å. These parameters of solvent-exposed groove are sufficient for accommodating $\sim$21 amino acid TM of TA proteins (Figure 1.6C).

Beside Yeast, the structure of Get3 from other organisms including archeal was solved by several independent research groups (Figure 1.7). *Aspergillus fumigates* (Suloway et al. 2009), *Chaetomium thermophilum* (Bozkurt et al. 2009) and *Methanothermobacter thermautotrophicus* (Sherrill et al. 2011) have structural features similar to yeast Get3 and exists as a dimer in solution. But an archeal Get3 from *Methanocaldococcus jannaschii* (Suloway et al. 2012) was found to be a tetramer.

The moonlight function of Get3 as an effective ATP-independent chaperone, when oxidized, has been previously reported, that plays a role in protecting eukaryotic cells against oxidative protein damage. This process is a fully reversible; involving disulfide bond formation and metal release; and it adopts into distinct, higher oligomeric structures. Mutagenesis studies demonstrated that the chaperone activity of Get3 is functionally distinct from and likely mutually exclusive with its TA protein binding and targeting function (Voth et al. 2014).
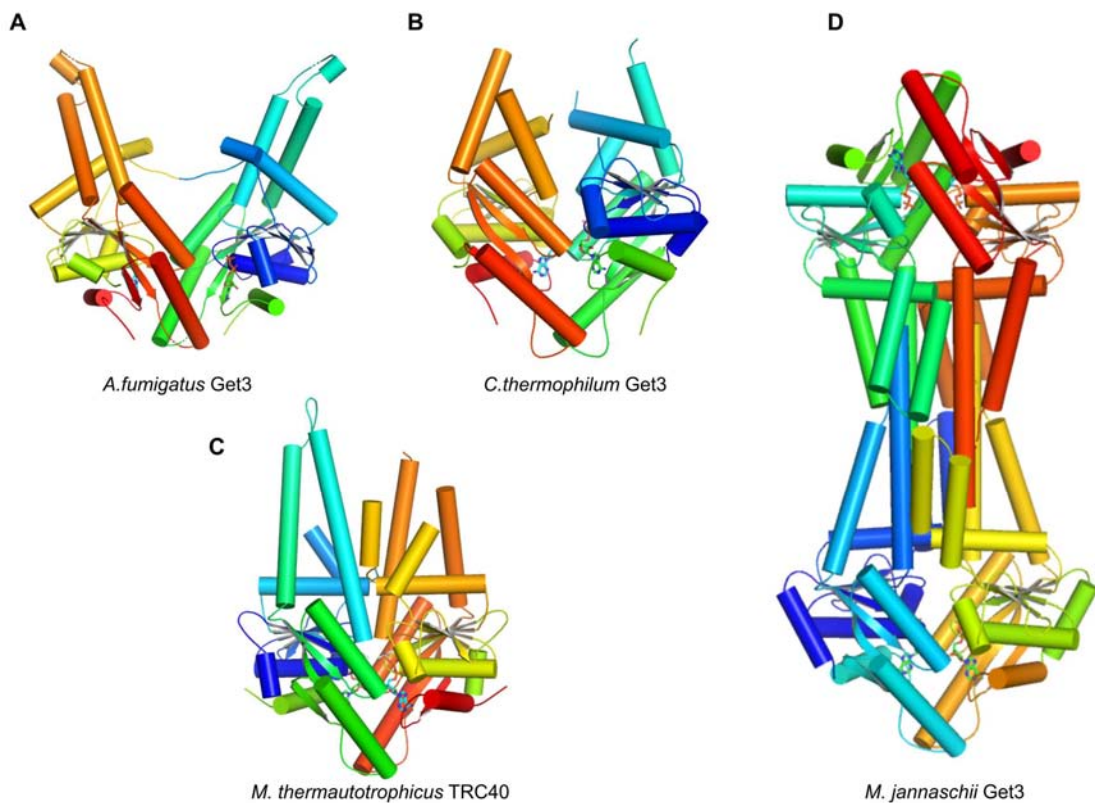
Figure 1.7: Structures of Get3 from (A) *Aspergillus fumigatus* (PDB:3IBG) (B) *Chaetomium thermophilum* (PDB:3IQX) (C) *Methanothermobacter thermautotrophicus* (PDB:3ZQ6) (D) *Methanocaldococcus jannaschii* (PDB:3UG7). Archeal Get3 (*M. jannaschii*) occurs as tetramer.

CXXC motifs are engaged in the high-affinity binding of Zn ion. Zn principally serves as to prime the cysteines for rapid oxidation by coordinating them in a highly reactive thiolate anion form and it also stabilizes them to prevent their premature oxidation (Voth et al. 2014). Under stress, cysteines engage in the formation of disulphide bonds, resulting in the release of Zn that ultimately leads to conformational change. This confirmation is not favourable for Get3 ATPase activity and it exposes the hydrophobic surface and forms the higher oligomeric species, which might generate interaction sites for the unfolded substrate. Several experimental studies have substantiated that Get3 serves as molecular chaperon in *Caenorhabditis elegans* and in other organisms (Hemmingsson et al. 2010).

1.10.2. Structural features of yeast Get4

The overall structure of yeast Get4 is composed of mainly 14 α-helices, an antiparallel β-sheet and several helix-turn-helix motifs. The N-terminal region of Get4 has similarity with Tetratrico Peptide Repeat (TPR)-like fold (Bozkurt et al. 2010). Get4 make interaction both with Get5 and Get3. C-terminal region of Get4 can accommodate N-terminal of Get5 in the groove between α12, α13 and β-tongue (Figure 1.8). N-terminal region of Get4 makes two interfaces with Get3, an anchoring interface and a regulatory interface. Anchoring interface is formed between Get4-α2, Get3-α10 and Get3-α11. Similarly, Get4-α4 and Get3-α3 form the regulatory interface. The anchoring interface mediate Get3-Get4 interaction, while regulatory interface has an influence on ATP hydrolysis by Get3 (Chartron et al. 2010; Gristick et al. 2014).
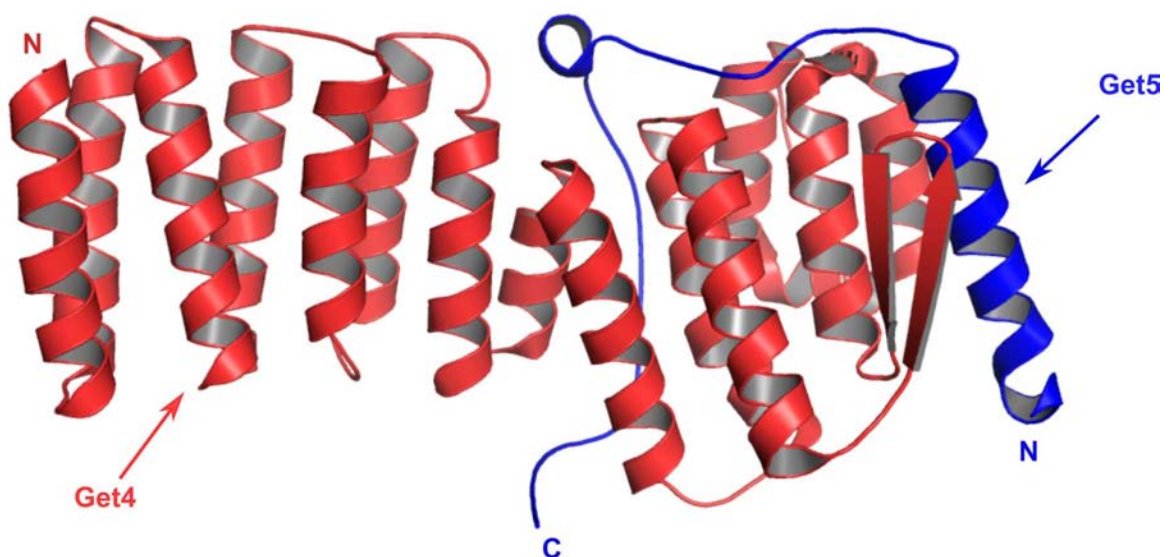
Figure 1.8: Structure of *S. cerevisiae* Get4 in complex with an N-terminal fragment of Get5.

1.10.3. Structural features of yeast Get5

Structure of N-terminal fragment of yeast Get5 (Get5N) was determined in complex with Get4 (Figure 1.8). Get5N has an α-helix at N-terminal followed by two long loops. The Get4-Get5N complex is mainly stabilized through hydrophobic interaction. Get5N is docked in a groove that is formed between β-sheets, α12 and α13 of Get4. Also, the backbone of Get5 interacts with α8, α9, α10, α11 of Get4 (Chartron et al. 2010; Gristick et al. 2014).

1.10.4. Structural features of yeast Get1/Get2

Get1 and Get2 did not interact with each other directly, but they both interact with Get3. Cytosolic domain of Get2 consists of two helices connected by a short linker. Get3 homodimer can bind with two Get2 molecules (Figure 1.9A). The cytosolic domain of Get1 has an antiparallel coiled-coil structure. Similar to Get2, two fragments of Get1 can interact with Get3 (Figure 1.9B) (Mariappan et al. 2011).
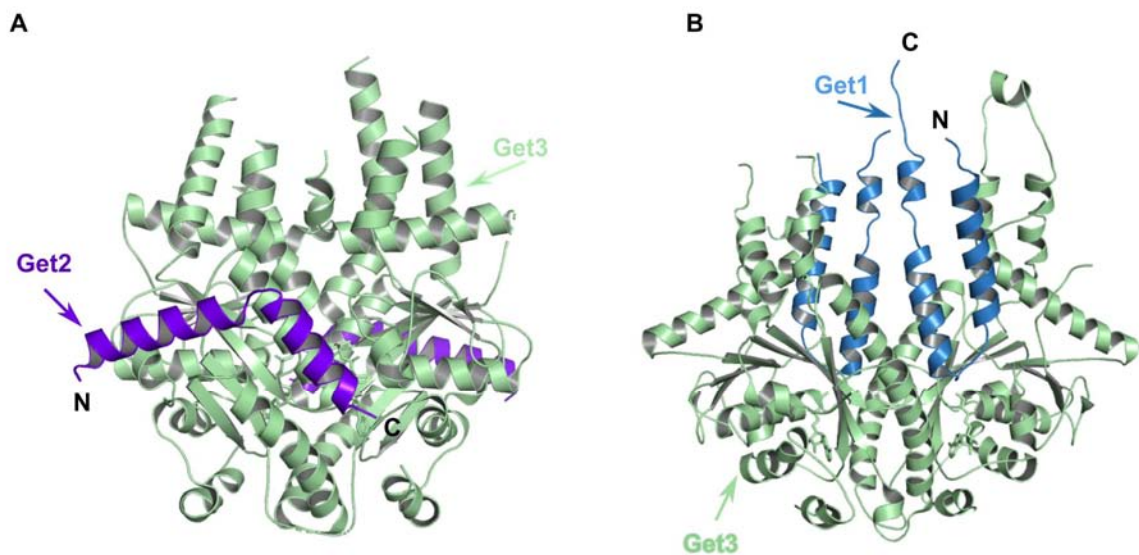
Figure 1.9: Structure of (A) Get2 and (B) Get1 in complex with Get3.

1.10.5. Structural features of yeast Sgt2

Structure of N-terminal homodimerization domain of both yeast Sgt2 and human SGTA have been delineated by Chartron et al., (Chartron et al. 2012). Sgt2-N monomer consists of three helices. The length of first two helices is similar and these mediate the dimerization. The third helix is shorter and orients away from dimer interface.

1.11. Physiological role of Get3

The cooperative action of all GET pathway components is necessary for the proper targeting of TA proteins to their target. Out of all, Get3 is an important component since it connects the pre-targeting (Get5/Get4/Sgt2/Ydj1/TA protein) and post-targeting (Get1/Get2/TA protein) complexes. Any defect in this targeting mechanism, especially functional disruption of Get3 leads to the mislocalization of ER TA proteins to mitochondria and accumulation in the cytosol (Schuldiner et al. 2008). There are other important implications reported for the GET components till date. Get3 is also shown to act as a holdase chaperone during stress conditions especially in redox stress (Voth et al. 2014; Shen et al. 2003). The loss of Get3 in yeast showed no clear growth defect on synthetic complete media at $30^{\circ}C$, but was unable to grow on media containing Cu^{2+} or hydroxyurea. Also, Get3 knockout yeasts was shown to have less potential to grow at elevated temperature. In higher organisms, loss of Get3 or Get3 ortholog TRC40 led to serious complications. The lack of mammalian Get3 ortholog TRC40 was found to be lethal to embryonic development in mice (Mukhopadhyay et al. 2006). Also in *C.elegan*, knockout of ASNA1 arrested growth at the L1 stage (Kao et al. 2007).

Several *in vivo* tissue-specific knockout studies showed that the GET pathway is involved in diverse physiological processes like photoreceptor functions, auditory perception and insulin secretion (Daniele et al. 2016; Vogl et al. 2016; Norlin et al. 2015). Recent studies on *Arabidopsis thaliana* have revealed that the loss of GET pathway orthologs led to serious root hair growth defects, and influences SNARE abundances and ER stress (Srivastava et al. 2017; Xing et al. 2017).

1.12. Quality control mechanism for TA protein insertion

In yeast, TA proteins that are designated for ER are targeted in a highly coordinated mechanism. Any defect in GET pathway leads to mis-insertion of ER TA proteins to mitochondria instead of targeting it to ER. The quality control for the proper insertion of TA protein commences even from Sgt2 recognition of the TM of TA proteins. Some of the deciding factors for the target location of nascent TA proteins are (i) the amino acid composition of the transmembrane domain, (ii) post-translational modification of TM, (iii) length of TM and (iv) the lipid composition of target membrane. In the case of cytochrome b5 proteins, the targeting signals that reside in the TM also play a crucial role in accurate targeting (Hwang et al. 2004).

Recent studies show the presence of a proof-reading protein for mis-localized ER TA proteins

that gets accumulated in the cytoplasm due to some defects in GET pathway (Matthew L Wohlever et al. 2017). For example, MSP1 is a dislocase protein that identifies the mis-localized ER TA proteins in mitochondria and removes them from mitochondrial membrane. This MSP1 is also an AAA ATPase that forms hexamer in solution in ATP-dependent manner. Central pore conserved residues of MSP1 are found to be essential for the removal of mis-localized TA proteins (Matthew L. Wohlever et al. 2017; Okreglak & Walter 2014).

1.13. GET pathway in plants

Several attempts were carried out to characterize TA protein targeting pathway in bacteria, yeast, archaea and mammalian systems (Beilharz et al. 2003; Kalbfleisch et al. 2007; Pedrazzini 2009; Kriechbaumer et al. 2009; Borgese & Righi 2010). Among eukaryotic systems, plant cells differ from animal cells structurally and functionally. Due to the presence of chloroplast, the protein targeting requires a highly specific mechanism for different membrane system within the plant cell. In view of this complexity, TA proteins are analyzed in the plant systems in this study. The sequence analysis in bacteria predicted several TA proteins that possibly suggest the presence of TA proteins and its targeting mechanisms in chloroplast and mitochondria (Borgese & Righi 2010).

TA proteins associated with the plant cell membrane were recently reviewed in *A.thaliana* (Kriechbaumer et al. 2009). Cytochrome b5 (Cb5) is a well-known example of TA proteins in plants. In *A. thaliana*, at least five Cb5 proteins are present, of which, some are localized to ER and some are localized to chloroplast or mitochondria. Ascorbate peroxidase (APX) and monodehydroascorbate peroxidase (MDAR) are other TA proteins found in *A. thaliana*, as isoforms, working cooperatively in NADH-dependent electron transport chain. Some of the functions of TA proteins include SNARE, disease resistance, transcription factor and protein translocation.

Even though some preliminary analysis of GET pathway was carried out in *A. thaliana* (Srivastava et al. 2017; Xing et al. 2017), the detailed investigation of GET pathway is missing in plant system. It was reported that *A. thalian*a has three Get3 for TA targetting to different organelle. But our analysis shows the presence of four Get3 in *A. thaliana*. To explore the GET pathway in plants, **this study focused on structural and functional characterization of a chloroplast Get3 from *A.thaliana* that is novel in indentification**.

Also, the crop plants that are growing in stringent conditions are not investigated for the presence of TA proteins. In order to understand GET pathway in crop plants, this doctoral

research work also highlights the analysis in selected two crop plants, *Oryza sativa* subsp. Indica (*O. sativa*) and *Solanum tuberosum* (*S. tuberosum*). These belong to monocot and dicot systems respectively. In this study, we have identified TA proteins in *O. sativa* subsp. Indica and *S. tuberosum* through *in silico* analysis. Predictions of functional and other physiological distribution of TA proteins and trans-membrane domain analyses are performed.

1.14. Statement of Problem

Eukaryotic systems are complex in their cellular architecture starting from their basic components to their higher molecular hierarchy. It is also known and reported that the evolution of eukaryotes (animal and plant cells) involved internalization of many unicellular organisms and this subsequently formed the specific organelles like mitochondria and chloroplast. During this evolution, the information for synthesis of proteins those are specific for these organelles are also conserved in the nuclear DNA. This facilitated the synthesis of these proteins in the cytoplasm and so the synthesized proteins are needed to be transported to their specific locations like chloroplast membrane or mitochondrial membrane for their function. Especially, the morphology and functionality of these organelle-specific membranes are also different. The inter-membrane space of these organelles is also varied depending on their membrane composition. This adds the complexity of characterizing membrane proteins and their targeting mechanisms.

Having known the importance and unique features of TA proteins, the targeting mechanism of these proteins is explored in this doctoral research. As mentioned earlier, TMD of TA proteins that are specific to different organelles vary in their amino acid composition and length; this could be the possible reason for the presence of multiple forms of Get3 to target TA proteins to different organelles in higher organism. Moving from the complexities in understanding the mechanisms of membrane protein targeting to different molecular hierarchies, the plant systems posses an extra organelle, chloroplast, that increase the importance of this research problem.

In this work, the mechanistic basis of GET pathway in plant system has been explored in three model systems that include *A. thaliana, O. sativa* subsp. Indica (*O. sativa*) and *S. tuberosum* (*S. tuberosum*) by detailed analysis. The complete doctoral research work in this thesis emphasize on the structure-function characterization of GET pathway in these plant systems to understand the crux in establishing the organelle-specific targeting of TA proteins. The first crystal structure of AtGet3Δd (*Arabidopsis thaliana* Get3), the chloroplast Get3 with

57 amino acid N' terminal truncation, has been determined along with *in silico* and biochemical characterization to validate the proposed mechanism.

Objectives

The specific aim of this study is to delineate the TA protein targeting mechanism in plant systems. This encompasses

- Analysis of GET pathway in selected plants.
- Functional characterization of Get3 in plants.
- Structural characterization of Get3 in plants.

Chapter 2

Materials and Methods

2.1. Materials

LB media and LB agar used for all the bacterial culture were purchased from Hi-media. Antibiotics used for selection of transformants such as kanamycin, Ampicillin, and chloramphenicol were obtained from Sigma-Aldrich, USA. Chemicals used for the purification of AtGet3a and AtGet3d such as Trizma, Sodium chloride, Imidazole, β-mercaptoethanol, tritonX 100, Magnesium chloride, glycerol, DTT, Nickel sulphate, Bromo-phenol-blue (BPB), Acrylamide, N,N'-methylene bisacrylamide, Sodium dodecyl sulfate (SDS), Acetic acid, Methanol, TEMED (N,N,N',N'-Tetramethylethylenediamine), Ammonium persulfate (APS), were purchased from Sigma-Aldrich, USA. Molecular weight marker for SDS-PAGE was purchased from Bio-Rad Laboratories, USA. Ni-NTA beads and Amylose resins used for affinity purification were purchased from Qiagen, Germany and New England BioLabs respectively. Size exclusion columns were obtained from GE, USA. Protein samples were concentrated using Amicon$^{®}$ ultra centrifugal filters procured from Merck-Millipore, USA.

Chemicals used in ATPase assay such as phosphoenolpyruvate, lactate dehydrogenase, pyruvate kinase, NADH and ATP were procured from Sigma-Aldrich, USA. Chemicals used in Plant cell organelle isolation and immunolocalization such as Sorbitol, Sodium Pyrophosphate, MgCl2, NaAscorbate, bovine serum albumin, OptiPrep^{TM} density gradient medium Sucrose, Na$_2$EDTA, MgCl$_2$, HEPES, DTT, BSA, paraformaldehyde, ethanol, butanol, Xylene, Paraplast X-TRA were brought from Sigma-Aldrich, USA. Antibody against 6XHis tag was obtained from Sigma-Aldrich, USA. Alexa Fluor 488 conjugated anti-rabbit IgG was obtained from Invitrogen, USA. Polyclonal antibodies against AtGet3d and AtGet3b were custom-made from ABGENEX Pvt. Ltd, India.

Commercial screens procured from Hampton research, USA and Qiagen, Germany were used for initial crystallization screening. Sodium cacodylate, lithium sulfate, PEG 4000, Ethylene glycol, glycerol, 2-methyl-2,4-pentanediol, propan-2-ol etc used in crystallization trials were obtained from Sigma-Aldrich, USA. Two-well sitting-drop plates were obtained from Hampton research, USA. 24 well plates and coverslips were obtained from Corning$^{®}$ (Sigma-Aldrich, USA) and Blue Star, India respectively. Other specialized chemicals and instruments used in the experiments are mentioned in the appropriate places.

Diffraction experiments were performed at INDUS-II (BL21 beamline, India), Elettra (Italy) and ESRF (France). High-resolution data were collected at ESRF, BM-14 beamline.

Crystallographic software XDS and MOSFLM were used to integrate the data. Data were scaled using AIMLESS or SCALA. Molecular Replacement was carried out with Phaser-MR of phenix suit using PDB 3IGF as a search model. Refinement was done using several cycles of Refmac5 and coot.

2.2 Constructs used in this study

The constructs used in this thesis are listed in Table 2.1.

Table 2.1. Constructs used in this study

SI	Gene	Vector	Forward primer	Reverse primer
1	AtGet3d	pEt33b	Gift from Professor Bil Clemons, Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA.	
2	AtGet3Δd	pET22b	GCTTCATATGACCAAATTCGTCAC CTTTCTCGG	CGGAAGCTTCCGCATTGTGACGATGAG AC
3	Atget3a	pET22b	GAGATACATATGGCGGCGGATTT GCCG	GATCTCGAGGCCACTCTTGACCCGTTC GAG
4	Atget3a	pET28 a+	GAGATACATATGGCGGCGGATTT GCCG	GATCTCGAGTTAGCCACTCTT GAC CCG TTC GAG
5	VA722	pMAL C2	CGGGATCCATGGCGCAACAATCG TTGATCTAC	ACGCGTCGACTTATTTACCGCAGTTGA ATCCCCC
6	U603	pMAL C2	TATGGATCCATGGAGACCCTTCTC TCCCCTC	ACGCGTCGACTTACTTCCTGGAGACAT AAGCAAAG
7	PMD2	pMAL C2	TATGGATCCATGGCGGAAGAGAG GAGCTTG	ACGCGTCGACTCAAACCCTCCTCGAGT GGTAAATG

2.3. Choice of plant systems

A.thaliana (UniProt Taxon identifiers: 3702), *Oryza sativa* subsp. Indica (UniProt Taxon identifiers: 39946) and *S. tuberosum* (UniProt Taxon identifiers: 4113) were selected for analyses.

2.4. Identifying the Tail-anchored proteins in selected plants

Complete proteome of *O. sativa* subsp. Indica and *S. tuberosum* were retrieved from UniProt (The UniProt Consortium 2017). TMHMM and Phobius server (Krogh et al. 2001; Kall et al. 2007) were used to identify Proteins with TM domains (TMDs) and proteins with single TM were selected (zero or more than 1 TM were rejected). Sequences were reanalyzed to find out the protein with single TM at C-terminal within last 50 amino acids. Proteins thus obtained were further analyzed using SignalP 4.1, Protein Prowler and TargetP 1.1 servers(Petersen et al. 2011; Emanuelsson et al. 2007; Boden & Hawkins 2005). Proteins with N-terminal signal peptides were identified using SignalP server and excluded from the analysis. Proteins

without N-terminal signal peptides were selected for further analyses. Protein Prowler program was used to identify the proteins with secretory signal sequence. Proteins with a probability of more than 0.5 for secretory signal sequence were rejected. TargetP was used to identify secretory pathway signals and mitochondrial or plastidial targetting sequences. All the results were compared and analyzed to select proteins that are not targeted by N-terminal signal and non-secretary.

2.5. Functional Annotation of TA proteins

Functional annotation of identified TA proteins of *O. sativa* and *S. tuberosum* was done by using Blast2GO, a powerful annotation tool (Götz et al. 2008). Blast, mapping and annotation of TA proteins were performed according to Blast2GO instructions. Proteins with similar functions were segregated based on their GO annotations.

2.6. Analysis of Predicted TA proteins

The length, molecular weight and amino acid sequence of the predicted TA proteins were retrieved from Uniprot. The Tm-region was predicted using Phobius and then the TM (Transmembrane) sequence and TM length were extracted from the protein sequence using R-script. For analyzing the Hydrophobicity of the total protein and the TM region of each TA protein the Kyte-Doolittle score was calculated using the Peptides package in R. Box plots for each parameters were plotted using R.

2.7. Identification of GET pathway component

GET pathway members of *O. sativa* and *S. tuberosum* were identified by analysing the whole proteome in comparison with Yeast and *A. thaliana* Get pathway proteins. Both crop plants have multiple forms of Get3 and localized to different organelle. Get2 and Get5 were not observed in *O. sativa* and *S. tuberosum*. Domains were confirmed by InterPro analysis (Finn et al. 2017).

2.8. Cloning and Over-expression of AtGet3

2.8.1 Cloning and Over-expression of AtGet3a

2.8.1.1. Cloning of AtGet3a

Total RNA of *Arabidopsis thaliana* was isolated using Spectrum RNA isolation kit (Sigma). cDNA was synthesized using Invitrogen superscript III (modified). The full-length AtGet3a was amplified from the cDNA and cloned between NdeI/XhoI site of pET 28a+ with N-terminal 6XHis tag. Primers used for the Amplification of AtGet3a are given in the table 2.1.

The clone was confirmed by double digestion and DNA sequencing. For better expression of AtGet3a, it was again re-cloned in pET22b(+) (Figure 2.1) with C-terminal 6XHis tag.

Following reaction conditions were used for PCR amplification of AtGet3a: 1 cycle of pre-denaturation at 95°C (2 min) followed by 28 cycles of denaturation at 95°C (30 s), primer annealing at 58°C (45 s) and extension at 72°C (1 min), a final extension at 72°C (10 min), and then cooled to 4°C.
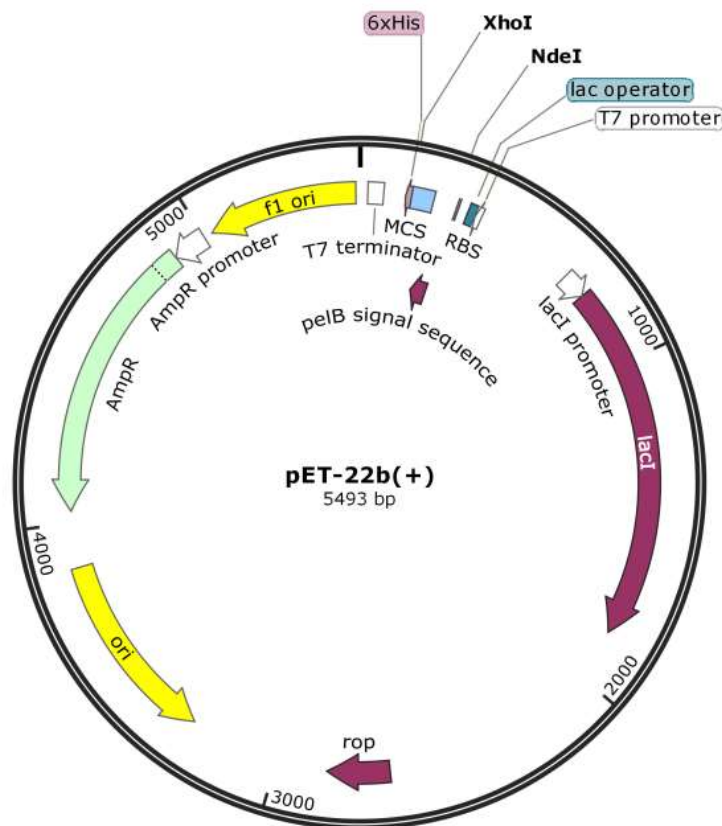
Figure 2.1: Map of pET 22b(+) expression vector.

2.8.1.2. Over-expression of AtGet3a

To obtain the recombinant AtGet3a, the expression host *E. coli* Rosetta DE3 was transformed with AtGet3a /pET22b(+) construct by calcium chloride treatment. Selection of transformed colonies was performed on LB agar plate containing ampicillin and chloramphenicol. 1ml of the overnight culture was used to inoculate 100 ml of fresh LB broth with ampicillin and chloramphenicol. When the absorbance at 600 nm reached 1, the culture was shifted to 16°C and the AtGet3a expression was induced with 0.1 mM IPTG and kept overnight. After overnight induction, cells were harvested by centrifugation at 4000 rpm for 15 minutes.

2.8.2 Cloning and Over-expression of AtGet3d

2.8.2.1. Cloning of AtGet3d

AtGet3d clone was a gift from Prof. Bil Clemons, California Institute of Technology, CA. AtGet3d was cloned in a modified pET33b vector (Figure 2.2) between SalI and PstI restriction sites. The expressed protein has a 6XHis and Tev protease site at the N terminal. Initial attempts to obtain the purified full-length AtGet3d were not successful as the protein precipitated readily and degradation in the N-terminal region was observed. In order to increase the stability, 57 amino acids were deleted from N-terminal and AtGet3Δd was generated.
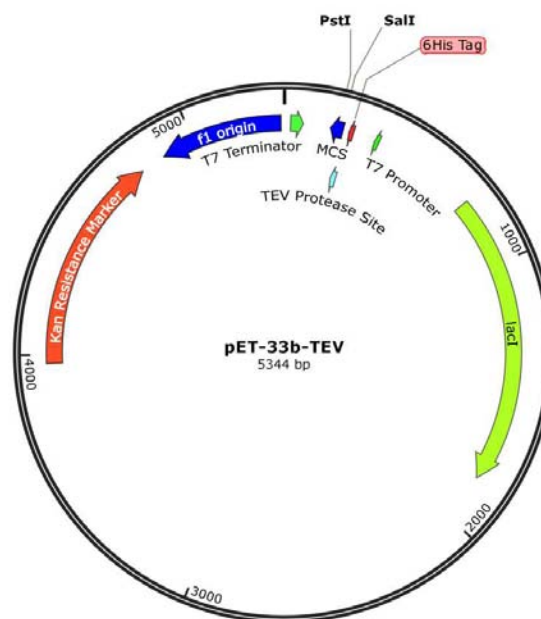
Figure 2.2: Map of pET33b (Modified) expression vector

2.8.2.2. Cloning of AtGet3Δd

As mentioned above, the initial trials for getting purified AtGet3d were failed. During purification, it was observed that two proteins came together in all purification steps (Ni-NTA affinity chromatography, Ion exchange and size exclusion) (Figure 2.3). LC-MS analysis of these two proteins showed that these belong to same proteins. The disorder prediction was carried out using online Protein DisOrder prediction System (http://prdos.hgc.jp) showed that there is a higher disorder probability at N-terminal of AtGet3d (Figure 2.4). Besides this, the protein precipitated readily and degradation in the N-terminal region was observed. To overcome these problems, 57 amino acids from N-terminal is truncated and recloned in pET 22b vector. AtGet3Δd was amplified from full-length AtGet3d. Primers used for the

amplification is given in the table 2.1 Following reaction conditions were used for PCR amplification of AtGet3Δd: 1 cycle of pre-denaturation at 95°C (2 min) followed by 28 cycles of denaturation at 95°C (30 s), primer annealing at 58°C (45 s) and extension at 72°C (1 min), a final extension at 72°C (10 min), and then cooled to 4°C. The PCR amplified products were analyzed on 1% agarose gel. Amplified AtGet3Δd was cloned between NdeI/XhoI site of pET 22b+ vector with C-terminal 6XHis tag.
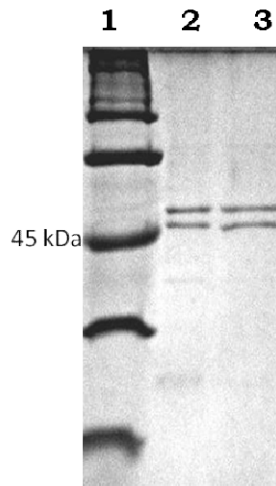
Figure 2.3: SDS PAGE analysis of Purified AtGet3d. Lane 1: Marker, lane 2-3: purified AtGet3d
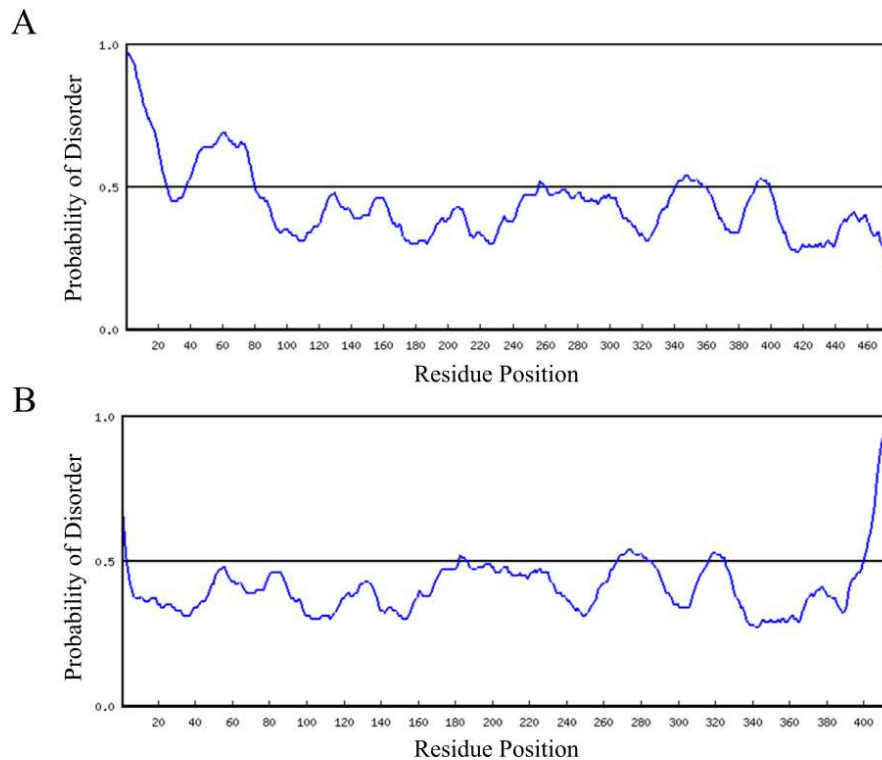
Figure 2.4: Disorder prediction of proteins (A) full length AtGet3d in pET33b expression vector and (B) AtGet3Δd in pET22b(+) expression vector.

2.8.2.3. Over-expression of AtGet3Δd

To obtain the recombinant AtGet3Δd, the expression host *E. coli* BL21 star DE3 was transformed with AtGet3Δd /pET22b construct by calcium chloride treatment. Selection of transformed colonies was performed on LB agar plate containing ampicillin. 1ml of the overnight culture was used to inoculate 100 ml of fresh LB broth with ampicillin. When the absorbance at 600 nm reached above 1, the culture was shifted to 16°C and the AtGet3Δd expression was induced with 0.1 mM IPTG and kept overnight. After overnight induction, cells were harvested by centrifugation at 4000 rpm for 15 minutes.

2.9. Purification of AtGet3

2.9.1. Purification of AtGet3a

AtGet3a overexpressed cells were resuspended in lysis buffer (25 mM Tris pH 7.5, 500 mM NaCl, 30 mM imidazole, 2 mM β-mercaptoethanol, 0.1% (v/v) triton X 100, 2mM $MgCl_2$ and 10% (v/v) glycerol) and sonicated for 5 minutes. After the cell disruption, the lysate was centrifuged for 30 minutes at 13000 rcf and the supernatant was collected. This supernatant was allowed to pass through the pre-equilibrated Ni-NTA column. The column was washed with 10 times the column volume of wash buffer (25 mM Tris pH 7.5, 500 mM NaCl, 30 mM imidazole, 2 mM β-mercaptoethanol (β-ME), 2mM $MgCl_2$ and 10% glycerol). The elution of protein was carried out with elution buffer containing 250 mM imidazole (25 mM Tris pH 7.5, 300 mM NaCl, 250 mM imidazole, 2 mM β-ME, 2mM $MgCl_2$ and 10% glycerol). Eluted protein fractions were pooled and concentrated using amicon 30K concentrator. The concentrated protein was further pass through pre-equilibrated (25 mM Tris pH 7.5, 150 mM NaCl, 2 mM DTT, 1mM $MgCl_2$ and 10% glycerol) S-200 size exclusion column. The eluted fractions were pooled and further concentrated before crystallization. Concentration of AtGet3a was measured using Bradford method and the purity was checked by SDS-PAGE.

2.9.2. Purification of AtGet3Δd

After overnight induction, cells were harvested by centrifugation and resuspended in lysis buffer (25 mM Tris pH 7.5, 500 mM NaCl, 30 mM imidazole, 2 mM β-ME, 0.1% triton X 100, 2mM $MgCl_2$ and 10% v/v glycerol) followed by sonication for 5 minutes. After the cell disruption, the lysate was centrifuged for 30 minutes at 13000 rcf and the supernatant was collected. This supernatant was allowed to pass through the pre-equilibrated Ni- NTA column. The column was washed with 10 times the column volume of wash buffer (25 mM Tris pH 7.5, 500 mM NaCl, 30 mM imidazole, 2 mM β-ME, 2mM $MgCl_2$ and 10% v/v glycerol). The

elution of protein was carried out with elution buffer containing 250 mM imidazole (25 mM Tris pH 7.5, 300 mM NaCl, 250 mM imidazole, 2 mM β-ME, 2mM $MgCl_2$ and 10% v/v glycerol). Eluted protein fractions were pooled and concentrated using amicon 30K concentrator. The concentrated protein was further pass through pre-equilibrated (25 mM Tris pH 7.5, 150 mM NaCl, 2 mM DTT, 1mM $MgCl_2$ and 10% v/v glycerol) S-200 size exclusion column. The eluted fractions were pooled and further concentrated before crystallization. Concentration of AtGet3Δd was measured using Bradford method and the purity was checked by SDS-PAGE.

2.10. Cloning of TA proteins

Tail-anchored proteins with average size 25 kDa were selected for co-expression studies. List of TA proteins selected is given in table 2.2. Several cloning attempts were carried out to incorporate these selected proteins into different vectors including pET22b DUET, PACYC-DUET and pMAL-C2. Out of these selected TA proteins, VA722_ARATH, U603_ARATH and PMD2_ARATH were successfully cloned between BamHI and SalI restriction sites of pMAL-C2 vector (Figure 2.5 and Figure 2.6). Primers used for the amplification of these three TA proteins are given in table 2.1. The expressing proteins had a MBP tag at N-terminal.
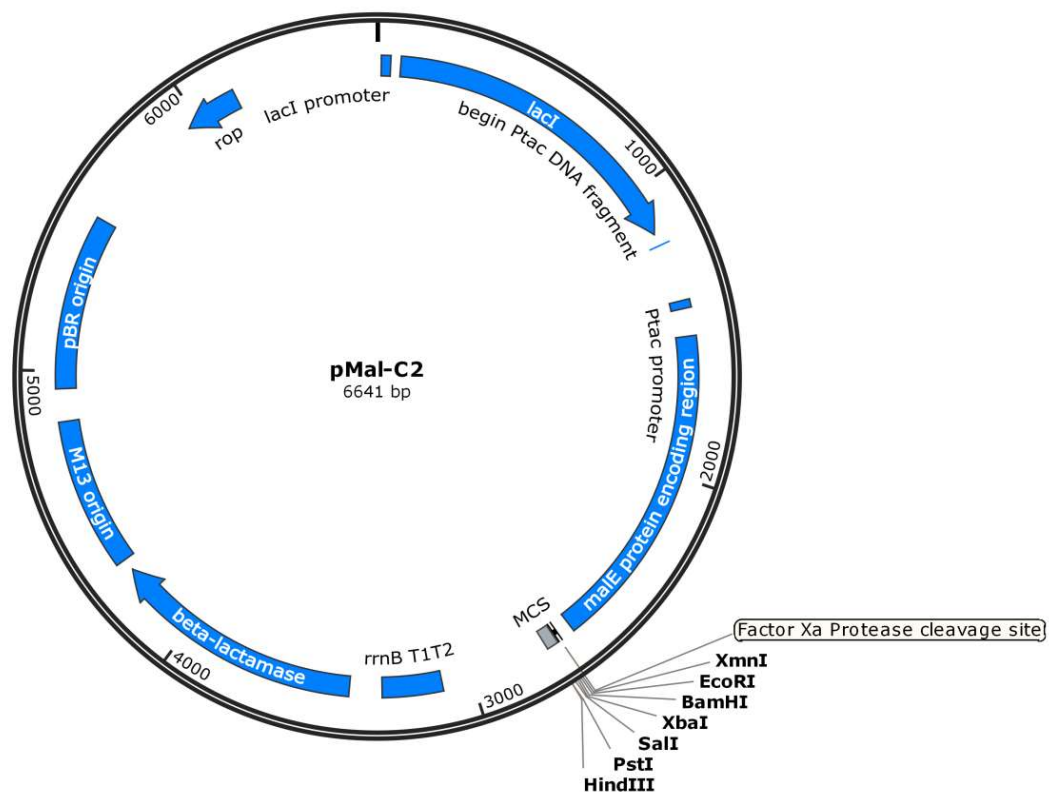
Figure 2.5: Map of pMal-C2 expression vector.

Table 2.2: Selected TA proteins for cloning and co-expression

SI No	Uniprot ID	Protein Name	Protein Size (in Da)	Localized
1	P93030	RMA2_ARATH	22,178.3	Cytoplasm/ER
2	O81045	P24D8_ARATH	24,565.4	Cytoplasm/ER
3	Q94AU2	SEC22_ARATH	25,332.2	Cytoplasm/ER
4	Q9SJL6	MEM11_ARATH	25,628.3	Cytoplasm/ER
5	Q9SHJ6	PMD2_ARATH	36,119.0	Mitocondrial OM
6	Q9SHC8	VAP12_ARATH	26,442.1	Cytoplasm/ER
7	P47192	VA722_ARATH	24,928.0	Cytoplasm/ER
8	Q9STT2	VPS29_ARATH	20,968.2	Cytoplasm/ER
9	Q9ZVL6	U603_ARATH	31,139.0	Chloroplast
10	F4KER9	TraB family protein	28,173.1	Cytoplasm/ER
11	Q9LM91	CCB25_ARATH	30,801.1	Cytoplasm/ER
12	F4I2U7	F4I2U7_ARATH	22,346.0	Cytoplasm/ER

Figure 2.6: Restriction digestion profile of successfully cloned TA proteins (**A**)VA722_ARATH (lane1: 1Kb plus marker, lane 2: double digestion of cloned plasmid) (**B**) PMD2_ARATH (lane1: 1Kb marker NEB, lane 2: double digestion of cloned plasmid) (**C**) U603_ARATH (lane1: 1Kb marker NEB, lane 2: double digestion of cloned plasmid)

2.11. Co-expression and Pull down of TA proteins and AtGet3

2.11.1. Co-transformation

AtGet3d/pET33b construct and TA protein constructs were co-transformed into *E. coli* C41 (DE3) expression strain by chemical method. Selection of co-transformed cells was carried out in LB agar in the presence of kanamycin and ampicillin antibiotics. Bacterial cells with both constructs only will survive in the selection plate. Positive colonies were selected and expression studies were carried out. The same approach was adopted for co-transformation of AtGet3a/pET28 construct and TA protein.

2.11.2. Co-expression

AtGet3d or AtGet3a and TA proteins were Co-expressed in *E. coli* C41 (DE3) expression strain. 1ml of the overnight culture was used to inoculate 100 ml of fresh LB broth with kanamycin and ampicillin. When the absorbance at 600 nm reached above 1, the culture was shifted to 16°C and the expression of both AtGet3 and TA proteins were induced with 0.1 mM IPTG and kept for overnight. After overnight induction, cells were harvested by centrifugation at 4000 rpm for 15 minutes.

2.11.3. Pull-down analysis

The interaction of AtGet3d/AtGet3a with TA proteins was confirmed by two-step affinity chromatography. Harvested bacterial cells were resuspended in lysis buffer (25 mM Tris pH 8, 150 mM NaCl, 10% v/v glycerol, 2mM DTT and 10mM imidazole), followed by sonication for 5 minutes with 5 sec on/off pulse and 45% amplitude. After the cell disruption, the lysate was centrifuged for 30 minutes at 12000 rpm. The supernatant was collected and subjected to Ni-NTA affinity chromatography. Fractions eluted from Ni-NTA were further pooled and passed through amylose column. The elution of protein from amylose column was carried out with elution buffer containing 20 mM Maltose (25 mM Tris pH 8, 200 mM NaCl, 10% glycerol, 1mM DTT and 20mM maltose). The eluted fractions from amylose column for all co-expressed samples were run on 12% SDS-PAGE.

2.12. ATPase activity assay

ATPase activity was determined at 30°C using a microplate photometric assay in which ATP hydrolysis is coupled to NADH oxidation (Figure 2.7) (Kiianitsa et al. 2003).
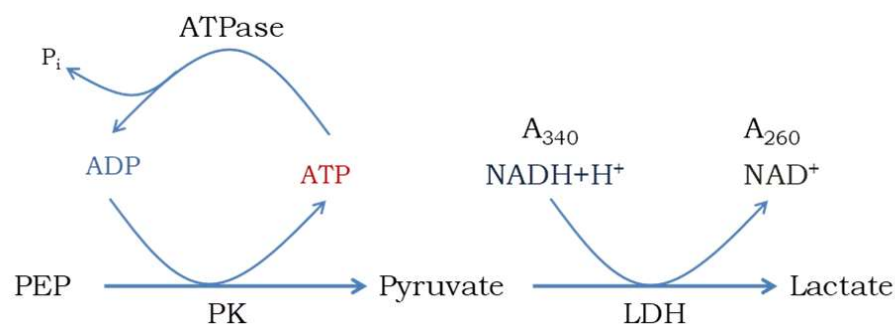
Figure 2.7: Determination of ATPase activity. The reaction is initiated with the conversion of ATP to ADP by ATPase. This triggers the conversion of phosphoenolpyruvate (PEP) to pyruvate by pyruvate kinase (PK). Once pyruvate is formed in the reaction, lactate dehydrogenase (LDH) acts on it and convert to lactate by utilizing NADH. Reduction in

NADH concentration can be measured spectrophotometrically at 340nm. This reduction of NADH reflects in ATPase activity.

The assay buffer contained 50mM Tris pH7.5, 20mM NaCl, 5mM $MgCl_2$, 1mM DTT, 5% v/v glycerol, 4.5mM phosphoenolpyruvate, 8.0U lactate dehydrogenase , 6.3U pyruvate kinase , 0.3mM NADH and 2µM AtGet3d/AtGet3a, and reactions were carried out in a final reaction volume of 200µl. The reaction was initiated by adding ATP and the decrease in NADH concentration was measured spectrophotometrically at 340nm. All chemicals were obtained from Sigma, USA.

2.13. Antibodies

Rabbit polyclonal antibodies against AtGet3a and AtGet3d were custom-made from ABGENEX Pvt. Ltd.

2.14. Organelle Isolation

Fresh leaf samples from *A. thaliana* plants were washed thoroughly with distilled water followed by ultrapure Milli-Q water. The midrib veins were removed and then the leaves were cut into small pieces. These leaves were then soaked in an equal volume of ice-cold Extraction Buffer (330 mM Sorbitol, 10 mM Sodium Pyrophosphate, 5 mM MgCl2, 2 mM NaAscorbate, 0.1% bovine serum albumin, pH 6.5) for few minutes and then ground to homogeneity in a sterile, ice-cold mortar and pestle. ~70 g of sample was used. The homogenate was separated into two batches. The first batch was used to separate the organelles by a modified method of Hartman et al., 2007 (Hartman et al. 2007). The first batch was spun twice, 5 minutes each, at 2200g to remove nuclei and cell wall components. The resultant supernatant was centrifuged in a Fixed angle 70 Ti rotor in a Beckman Coulter ultracentrifuge at 100000g onto a 10-mL 18% OptiPrep cushion (Axis-shield, Norway) (0.25 M Sucrose, 1 mM Na_2EDTA, 1 mM $MgCl_2$, 50 mM HEPES, 1mM DTT, 0.1% BSA, pH 7.5) for 2 h at 4 °C in 26.3 ml polycarbonate centrifuge bottles. The green pellet obtained were scooped and resuspended in Resuspension Buffer (2 mM Na_2EDTA, 5 mM $MgCl_2$, 1 mM $MnCl_2$, 50 mM HEPES, 330 mM Sorbitol, 0.1% BSA, pH 7.6 adjusted with KOH) containing chloroplasts. Concentrated membranes were carefully collected from the top of the cushion, adjusted to a 16% OptiPrep cushion of the same volume, and spun for 3 h at 4 °C in the same rotor at 350,000g to separate organelles in a self-generating iodixanol density gradient. Finally, 0.5 mL fractions were isolated from the top of the gradient which contained mitochondrial and cytosolic fractions. All isolated fractions of cytosol, chloroplast and

mitochondria were used for the further experiments. The batch was used to isolate intact chloroplast, according to Laganowsky et al., 2009 (Laganowsky et al. 2009). Homogenate was filtered through 8 layers of diaper liners (Gerber) and spun in a centrifuge at 200g for 10 min at 4 °C. The white pellet containing cell debris was removed and the supernatant containing chloroplast was centrifuged at 1200g. The pellet was resuspended in resuspension buffer. The resuspended pellet was layered onto a 36% OptiPrep pad (0.27 mM Sucrose, 2 mM Na$_2$EDTA, 1 mM MgCl$_2$, 50 mM HEPES, 0.2% BSA, pH 7.6) and spun at 1200g for 10 min at 4 °C. The pellets from the OptiPrep density gradient (containing intact chloroplast) were resuspended in resuspension buffer without BSA.

2.15. Western Blot of Isolated Organelles

Cytosolic, mitochondrial and chloroplast fractions were run on 12% SDS PAGE gel and transferred to PVDF membrane by semi-dry method using Bio-Rad Trans-Blot® Turbo™ Transfer System. The membrane was then blocked by 5% BSA in PBS for 1 hour and incubated overnight in the anti-AtGet3d polyclonal antibody. The membrane is then washed with PBST (10min X3) and incubated with Anti-rabbit IgG conjugated with HRP. After washing with PBST (10min X3), the blot was developed by DAB Peroxidase Substrate (Sigma, USA).

2.16. Immunolocalization Experiments

Immunolocalization experiments were performed according to modified protocol from Divol et al. (Divol et al. 2013). 2 to 3 week old leaves or 1-week old seedlings were fixed in 4% paraformaldehyde in 10mM PBS pH 7.2 by vacuum infiltration. The samples were dehydrated in successive baths of 75, 95 and 100 % ethanol, ethanol/butanol 1:1 (v/v) and 100 % butanol overnight. Samples were embedded in wax by soaking the sample in successive 1-hour baths of increasing concentration of xylene (Butanol/xylene 2:1, 1:1, 0:1) and then in 1-hour baths in increasing Paraplast X-TRA concentrations (Xylene/Paraplast X-TRA 3:1, 1:1, 1:3). Before taking sections, samples were kept in 100% Paraplast X-TRA for at least 1 night. Wax blocks with samples were prepared and sections were taken using Leica RM2255 microtome. 8 to 10 micrometre thick sections were deposited and dried onto silanized-slides. Samples were de-waxed and rehydrated by reversing the above steps. Sections on the slides were washed with PBS, blocked by BSA (1% BSA in PBS) at room temperature for 1 hour and incubated with anti-AtGet3Δd polyclonal antibody overnight. The sections were then washed with PBS (10 min x 3) and incubated for one hour in dark with anti-rabbit IgG

conjugated to the Alexa Fluor 488 fluorochrome (Invitrogen, USA). After washing in PBS (10 min x 3), the sections were mounted in mounting solution (Sigma, USA).

2.17. Confocal Laser Scanning Microscopy

The microscopy images were obtained using Leica TCS SP8 laser scanning microscope. A laser at 488 nm was used for exciting Alexa Fluor 488 and chloroplast autofluorescence. A long pass filter of 585 nm and a band pass filter of 502 to 550 nm were used for detecting the emission signals of chloroplast autofluorescence and green fluorescence respectively. Sections were observed with HC PL APO CS2 63x objective and images were processed by LAS X software.

2.18. Pull-down and Mass Spectrometric Analysis

In order to find the proteins interacting with AtGet3Δd, total protein fraction was extracted from wild-type, mutant Arabidopsis leaf tissues and wild-type chloroplasts using P-PER™ Plant Protein Extraction Reagent (Thermo Scientific, USA) according to the manufacturer's protocol. Proteins interacting with AtGet3d were pulled down using anti-AtGet3Δd and TrueBlot® Anti-Rabbit Ig IP Agarose Bead (Rockland antibodies and assays, USA) and identified by Mass spectrometry (Figure 2.8).

Total protein isolated from both wild-type or mutant leaf tissues were mixed with Protein-A beads. After incubation for 30 minutes, the samples were subjected to centrifugation and the supernatant was collected. These pre-cleared samples were further used for the pull-down experiments.
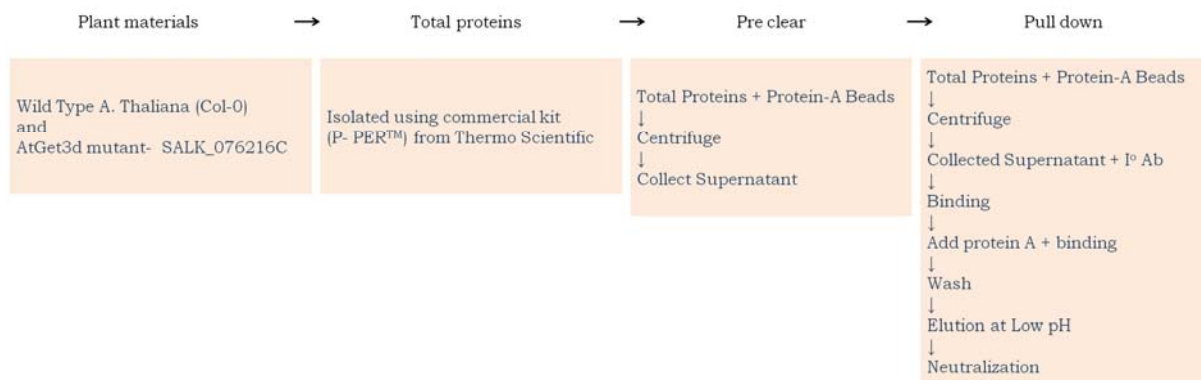
Figure 2.8: Procedure followed for pull-down experiment to identify the interacting proteins.

Collected supernatants were mixed with primary antibody (against AtGet3d) and incubated for 60 minutes. Protein A beads were added to the incubated samples and continued the

incubation for 60 minutes. After the incubation, the beads were washed with wash buffer (50mM Tris HCl pH 8.0; 150mM NaCl; 1% NP-40) and the proteins were eluted with low pH elution buffer (0.1 M glycine-HCl, pH 2.5). The eluted samples were neutralized with 1 M Tris, pH 9.5 before proceeding to LC-MS/MS analysis. Proteins were identified by the standard operating protocol for LC-MS/MS. The final list of interacting proteins was prepared by eliminating the proteins identified in mutant sample from wild-type sample.

LC-MS/MS was further employed to investigate the location of AtGet3d in Chloroplast. For finding the location of the AtGet3Δd in the chloroplast, intact chloroplasts were treated with thermolysin (Promega Corporation, USA) according to manufacturer protocol. Thermolysin cleaves membrane proteins on the surface of the membrane without disturbing the chloroplast membrane. After the treatment, the supernatant was collected and subjected to Mass spectrometry.

2.19. Expression Data and Heatmap

Microarray expression data for genes At1g26090 (AtGet3d), At5g60730 (AtGet3c), At1g01910 (AtGet3a) and At3g10350 (AtGet3b) were collected from NASCArrays expression database through e-Northerns Expression Browser of BAR (Craigon et al. 2004; Toufighi et al. 2005). Expression data were normalized to control value and Heatmap was generated using ClustVis web tool (Metsalu et al. 2015).

2.20. Crystallization

To find the crystallization condition of AtGet3Δd, the concentrated protein was subjected to screening with several commercially available and manually prepared screens. The solution with 50mM Sodium cacodylate (pH-5.47), 50 mM lithium sulfate and 30% PEG 4000 yield plate-like crystals with poor diffraction quality. In order to achieve proper packing and a good diffraction-quality crystal, purified AtGet3ΔL was incubated with 2 mM ADP and introduced for the crystallization in the above-mentioned buffer condition. Crystals of AtGet3Δd obtained were checked for diffraction quality and they diffracted up to 2 Å.

Similarly, crystallization trials for AtGet3a were also carried out using several screens and got crystals in the solution with 100mM Tris pH-8.5 and 3M Sodium chloride. Standardization of the condition for the crystal growth for AtGet3a is progressing.

2.20.1. Crystallization of AtGet3Δd

Purified AtGet3Δd was concentrated to 10 mg/ml and crystallization trials were carried out with sitting drop vapour diffusion method. Hits were obtained in the condition with 50mM Sodium cacodylate (pH-5.47), 50mM lithium sulfate and 30% PEG 4000. Crystals were reproduced in the same condition with hanging drop vapour diffusion method. But these crystals failed to diffract during diffraction experiments.

After several trials of crystallization and diffraction, AtGet3d crystals found to be diffracted in the presence of nucleotides. For that, crystals of 6XHis tagged AtGet3Δd were grown at room temperature by mixing an equal volume of protein solution containing 2mM ADP/AMPPNP with reservoir solution containing 50mM Sodium cacodylate (pH-5.47), 50mM lithium sulfate and 30% PEG 4000.

2.21. Cryo-protection and X-ray Diffraction

In general, crystals are exposed to high-intensity ionizing radiation during X-ray diffraction. In order to minimize the radiation damage during this exposure, crystals were diffracted in a cryogenic temperature at 100K. Glycerol, 2-methyl-2,4-pentanediol, sucrose, PEG 400, perfluoropolyether oil etc are commonly used cryoprotectants. Different cryo-protecting agents were tried for AtGet3Δd and Glycerol was found to give better diffraction pattern compared to others. Crystals of AtGet3Δd were cryoprotected in the mother liquid (50mM Sodium cacodylate (pH-5.47), 50mM lithium sulfate and 30% PEG 4000 and 2mM ADP/AMPPNP) supplemented with 30% glycerol before flash-freezing in liquid nitrogen.

Several rounds of X-ray diffraction experiments were conducted at different synchrotron facilities, INDUS-II(India), Elettra(Italy) and ESRF (France). High-resolution data were collected from ESRF BM-14 beamline.

2.22. Data collection, structure solution and refinement

Data collection was done at BM-14 beamline, ESRF, France at 100K. Data collection was carried using CCD MAR mosaic 225 detector at 100K temperature (Figure 2.9). The data were integrated with XDS and scaled with Aimless (CCP4 suite) in space group P 2_1 (Kabsch 2010; Evans & Murshudov 2013; Evans 2006). Output parameters of Aimless include R-merge, X^2, I/σI. A processed dataset with less than 10% R merge is considered as a good one. R merge alone cannot be a deciding factor in high resolution cut off, since it is dependent on the redundancy of data. A well-refined data have χ2 values closer to 1.0. Number of

molecules in the crystallographic asymmetric unit can be predicted by Matthew's coefficient. Based on the knowledge of unit cell parameters and space group, it is quite possible to determine the number of molecules in the unit cell and solvent content once we know the molecular weight of the protein. The monomeric molecular weight of 45kDa was used to calculate Mathew's coefficient, $2.15\text{Å}^3 \text{ Da}^{-1}$, for AtGet3Δd crystals.

AtGet3Δd crystals diffracted up to 2Å. Based on data quality statistics, for a better structural solution, the resolution was limited to 2.5 Å. Structure of AtGet3Δd was determined to 2.5Å by molecular replacement. Rotation and translation functions were used to properly place the model in new unit cell. MOLREP and PHASER are the most commonly used programmes for Molecular Replacement. PHASER makes use of maximum likelihood-based method for performing the rotation and translation searches.
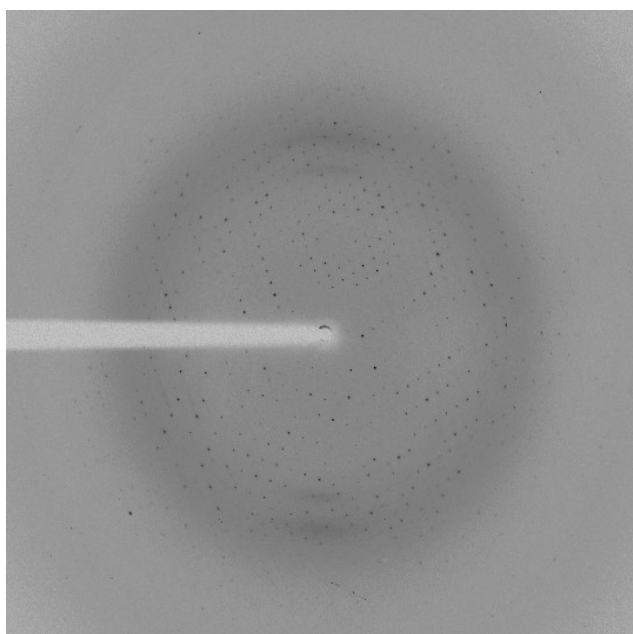
Figure 2.9: Diffraction image obtained from crystals of AtGet3Δd with ADP.

Structure of Atget3Δd was determined by molecular replacement with PHASER using All4481 protein from Nostoc sp. PCC 7120 (PDB: 3IGF) as a search model (McCoy et al. 2007; Forouhar et al. n.d.). Several rounds of model building and refinement were carried out with COOT 0.8.x and CCP4/Refmac5 (Emsley et al. 2010; Winn et al. 2011; Murshudov et al. 2011). A single molecule of AtGe3Δd dimer was found in the asymmetric unit. The density of side-chain is generally weak in the α-helical subdomains, and density was missing for residues 186-205, 320-328 and 399 in both subunits. Data collection and refinement statistics are listed in table 5.1. Structure figures were generated using PyMOL (http://www.pymol.org/).

2.23. Accession codes

Coordinates and structure factors have been deposited in the Protein Data Bank under accession code 5YQK.

2.24. AtGet3b Models

Three dimensional model of AtGet3b monomer was generated using X-ray coordinates of 3ZQ6 and 3UG7 (Sherrill et al. 2011; Suloway et al. 2012). 92 amino acids from N-terminal were deleted since there was no proper sequence alignment. Homology modelling was performed using MODELLER version 9.17 (Sali & Blundell 1993; Webb & Sali 2016). Model with lowest discrete optimised protein energy (DOPE) score was selected for further analysis. Dimer generation and refinement of AtGet3b model was performed in Galaxy WEB server (Ko et al. 2012; Shin et al. 2014).

2.25. Modelling of GET Pathway Proteins from *O. sativa* and *S. tuberosum*

Identified GET pathway components were modelled using MODELLER ver.9.17 (Webb & Sali 2016). Yeast Get3 (PDB ID: 2WOO) was used as template to model all cytoplasmic Get3s of both *O. sativa* and *S. tuberosum* (Mateja et al. 2009). The cytoplasmic domain of Get1 of both *O. sativa* and *S. tuberosum* were modelled using Yeast Get1 (PDB ID: 3ZS8) as template (Mariappan et al. 2011). *Chaetomium thermophilum* Get4 (PDB ID: 3LPZ) and human TRC 35 (PDB ID: 6AU8) were used as templates for modelling *O. sativa* Get4 and *S. tuberosum* Get4 respectivily (Bozkurt et al. 2010; Mock et al. 2017). Chloroplastic Get3 of both *O. sativa* and *S. tuberosum* have more than 40% identity with *A. thaliana'*s chloroplastic Get3 (AtGet3d). Structure of AtGet3d solved in this study (PDB: 5YQK) was used as template to model chloroplastic Get3 of both *O. sativa* and *S. tuberosum*. Models with lowest discrete optimized protein energy (DOEP) scores were selected and further refined using GalaxyWeb server (Ko et al. 2012).

2.26. Docking and simulation

Docking and molecular dynamics simulation studies of AtGet3Δd with ADP were done with Glide 5.8 and Desmond 3.1(Desmond Molecular Dynamics System, version 3.1, D. E. Shaw Research, New York, NY.) of maestro (Maestro, Schrödinger, LLC, New York,) with default settings (Friesner et al. 2006; Bowers et al. 2006). ADP molecule considered in the present study was downloaded from ZINC compound database in the mol2 format. The protein and ligand were prepared first, before proceeding with the docking studies. The protein structure

was prepared using protein preparation utility of Maestro. The minimization of the structure was done using impref utility of Maestro. ADP ligand was refined with the help of LigPrep 2.5 to define their charged state and enumerate their stereoisomers. The processed AtGet3Δd and ligand were further used for the docking studies using Glide 5.8.

2.26.1. Molecular dynamics simulations: The docked complexes were prepared first using protein preparation wizard and then subjected to molecular dynamics simulations for a time scale of 15 nanoseconds (ns) using Desmond 3.1 of maestro. OPLS2005 force field was applied on docked complexes placed in the centre of the orthorhombic box solvated in water. Protein was immersed in orthorhombic water box of SPC water model. Total negative charges on the docked structures were balanced by a suitable number of counter-ions to make the whole system neutral (10 Cl⁻ ions). The system was initially energy minimized for maximum 2000 iterations of the steepest descent (500 steps) and the limited memory Broyden–Fletcher– Goldfarb- Shanno (BFGS) algorithm with a convergence threshold of 1.0kcal/mol/Å. The short and long-range Coulomb interactions were handled by Cutoff and Smooth particle mesh Ewald method with a cut off radius of 9.0Å and Ewald tolerance of 1e -09. Periodic boundary conditions were applied in all three directions. The relaxed system was simulated for 10 ns with a time step of 2.0 femtosecond (fs), NPT ensemble using a Berendsen thermostat at 300K temperature and atmospheric pressure of 1bar. The energies and trajectories were recorded after every 2.0 picosecond (ps). The energies and RMSD of the complex in each trajectory were monitored with respect to simulation time. The C-alpha atom root mean square fluctuation (RMSF) of each residue was analyzed. The intermolecular interactions between the target and substrate were assessed to check the stability of the complexes.

Analysis of TA Protein Targeting Pathway in Plants

Note:

Some parts of this chapter have been published in '*Biochemistry and Biophysics Reports*' (Manu et al. 2018). Some figures are used as such from this publication wherever required.

3.1. Introduction

GET protein translocation pathway is well analyzed in yeast. Even though several reports are available about the TA protein translocation in plants, the information available is limited. Recent analyses of *A. thaliana* show the presence of higher number of TA proteins and Get3s compared to yeast (Kriechbaumer et al. 2009; Duncan et al. 2013; Xing et al. 2017). But the detailed analysis of GET pathway in plants is yet not complete. In this Chapter 3, *A. thaliana* along with two major crop plants *O. sativa* and *S. tuberosum* were analysed indepth for TA proteins and its targeting mechanism.

3.2. Results

3.2.1. TA Proteins in *A. thaliana*.

In silico analysis of TA proteins in *A. thaliana* by Kriechbaumer *et al.*,(Kriechbaumer et al. 2009) reported that *A. thaliana* has 502 TA proteins. Re-analysis of TA proteins in *A. thaliana* in this chapter explored that 14% of the identified TA proteins are localized to the endoplasmic reticulum. Mitochondria, plastids and peroxisomes are other locations where TA proteins are localized (Figure 3.1).
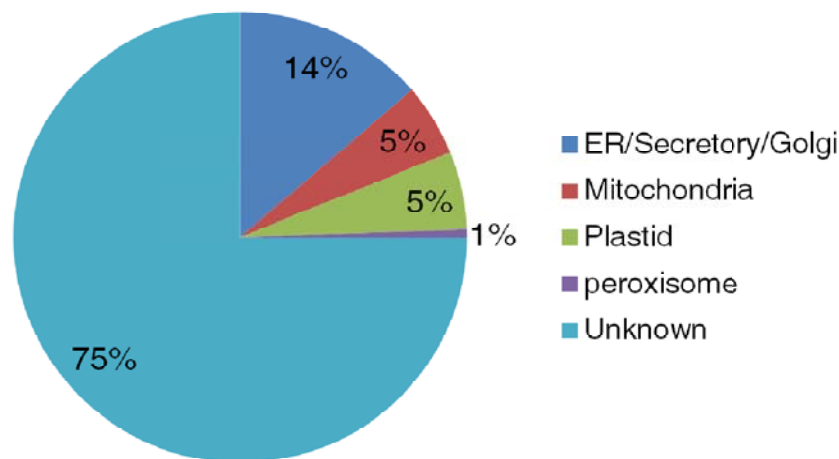
Figure 3.1: Organelle distribution of TA proteins in *A. thaliana.*

Inspite of being 67% TA proteins withunknown function, the remaining TA proteins are involved in several activities like SNARE, transcription factors, disease resistance, heat shock proteins etc. In functionally known category, around 10% proteins are involved in SNARE activity and 1% act as cytochromes (Figure 3.2).

Even though the TA protein and its targeting pathway is analyzed in *A. thaliana*, there is no detailed analysis of this pathway in other plant systems especially in crop plants. In this study

two crop plants, *O. sativa* and *S. tuberosum* are analyzed *in silico* for TA proteins and its targeting components.
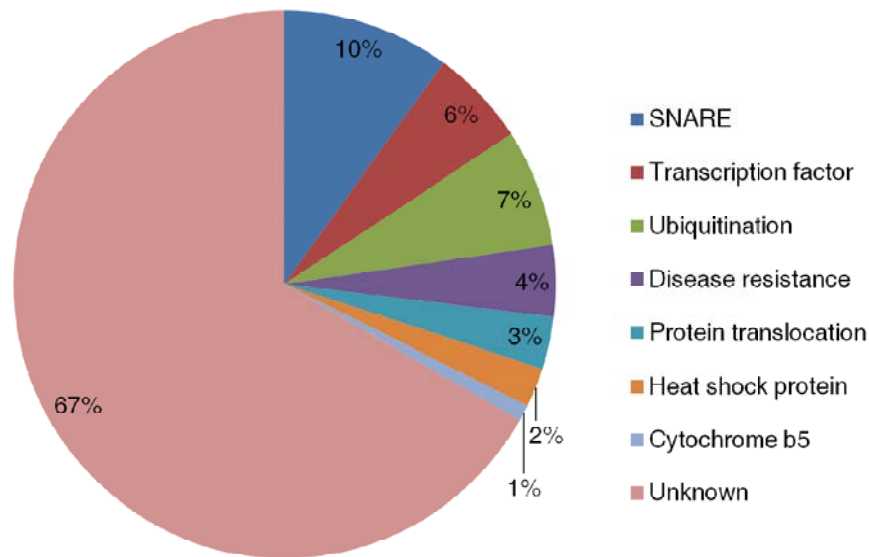
Figure 3.2: Functional distribution of TA proteins in *A. thaliana*

3.2.2. Identification of TA proteins in *O. sativa* and *S. tuberosum*

TA proteins were identified based on its definition. Proteome of *O. sativa* and *S. tuberosum* were downloaded from Uniprot. 37383 and 53105 total number of TA proteins were found in *O. sativa* and *S. tuberosum* respectively. Proteins with single TM were obtained after analysis through TMHMM and Phobius servers. 723 and 1287 membrane proteins were found in *O. sativa* and *S. tuberosum* respectively, after filtering the proteins with single TM at C- terminal, within 50 amino acids. From this list, proteins with N-terminal signal and secretary sequence were omitted. From 37383 of total proteins from *O. sativa*, 508 were found to be TA proteins. Similarly for *S. tuberosum* 912 found to be TA protein from Total 53105 proteins (Table 3.1, Appendix B, Table B1, Table B2 and (Manu et al. 2018)).

Table 3.1: Number of TA proteins in *O. sativa* and *S. tuberosum*

Plant Species	Total Number of proteins	Proteins with one TMD	Proteins with C-terminal TMD	Tail-anchored proteins
Oryza sativa subsp. *Indica*	37383	4379	723	508
Solanum tuberosum	53105	5022	1287	912

3.2.3. Organelle distribution of TA proteins in *O. sativa* and *S. tuberosum*

Organelle distribution of TA proteins was analyzed for both *O. sativa* and *S. tuberosum* (Figure 3.3 and Figure 3.4). In case of *O. sativa*, out of 508 TA proteins, 88 proteins are found to be localized to chloroplast, 107 belong to mitochondria, 16 proteins belong to ER/Golgi/Secretary and for 297, location was unknown. Similarly, for *S. tuberosum* out of 912 TA proteins, 107 were localized to chloroplast, 128 belong to mitochondria, 28 localized to ER/Golgi/Secretary and 649 have unknown location. In both *O. sativa* and *S. tuberosum*, 3% of total TA proteins were found localized to ER. In *O. sativa*, 21% TA proteins were found localized to mitochondria that are higher than *S. tuberosum* mitochondrial TA proteins (14%).
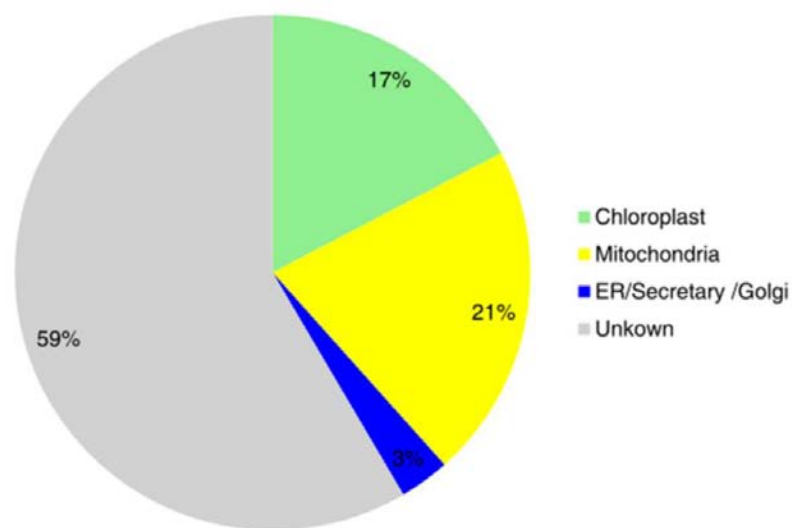
Figure 3.3: Organelle distributions of TA proteins in *O. sativa*
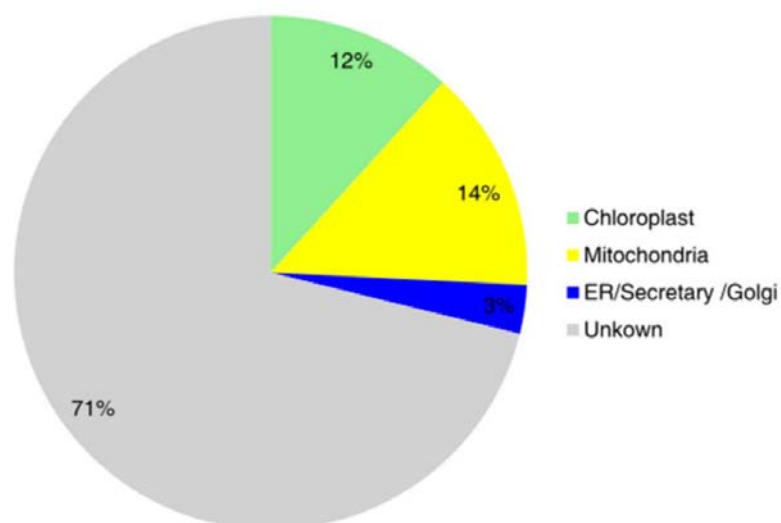
Figure 3.4: Organelle distributions of TA proteins in *S. tuberosum*

3.2.4. Functional distribution of TA proteins in *O. sativa* and *S. tuberosum*

The identified TA proteins were grouped by functional similarity and their biological process. The identified TA proteins are involved in several biological processes like vesicular transport, gene expression, cellular metabolism etc (Figure 3.5 and Figure 3.6). TA proteins have a vast variety of functional divergence. They function as SNARE, transcription factor, involved in disease resistance etc. The functional distribution of *O. sativa* and *S. tuberosum* TA proteins are shown in Figure 3.7 and Figure 3.8. In the functionally known category, SNARE binding proteins found to be higher for *O. sativa* and Protein binding found to be higher for *S. tuberosum*. In addition, TA proteins in *O. sativa* are majorly involved in vesicle fusion, protein transport, gene expression etc. Similarly, in the case of *S. tuberosum*, they are involved in vesicle-mediated transport, oxidation-reduction process, gene expression etc.
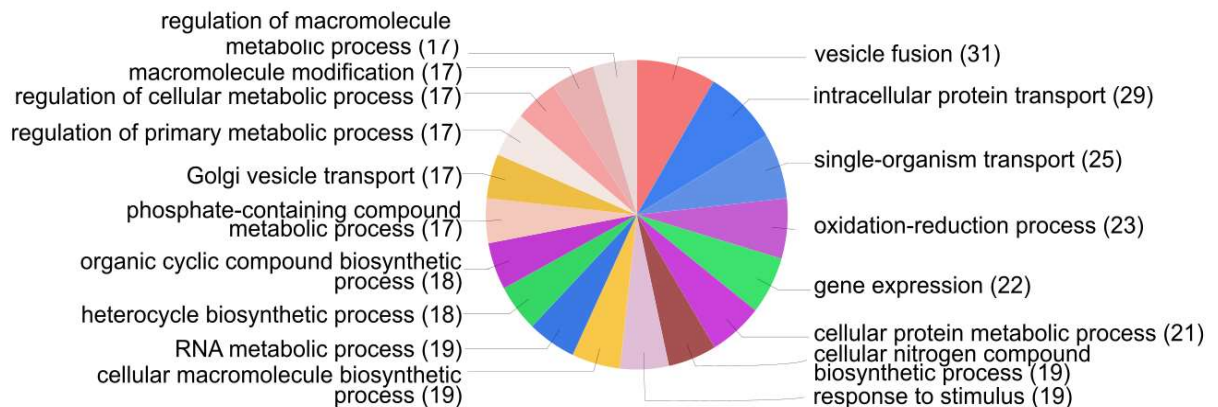
Figure 3.5: Biological process distributions of *O. sativa* TA proteins.
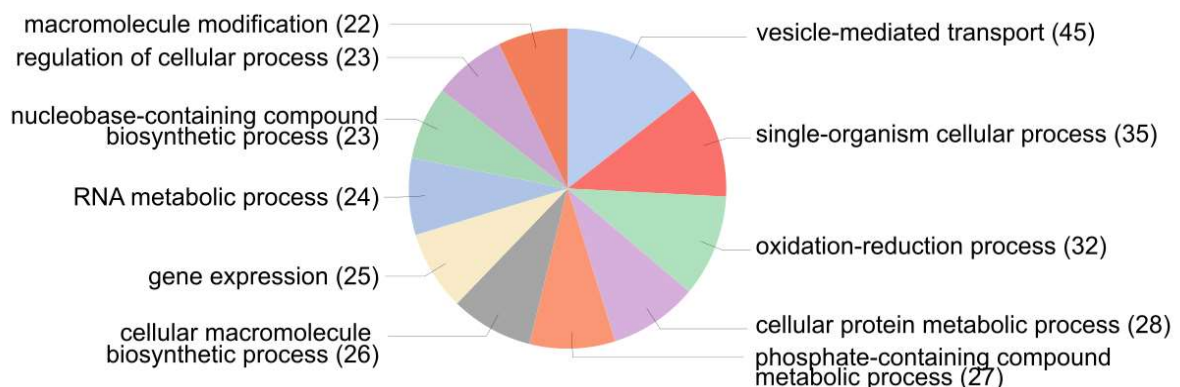
Figure 3.6: Biological process distributions of *S. tuberosum* TA proteins.
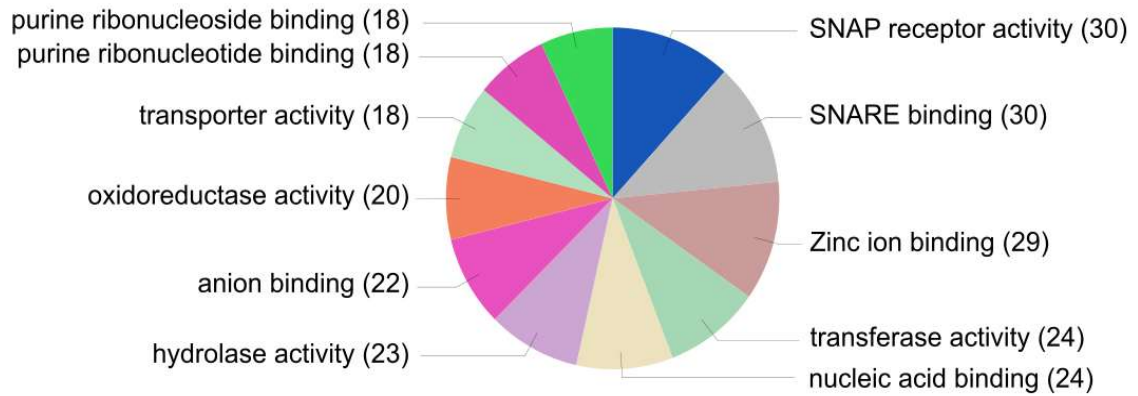
Figure 3.7: Functional distributions of *O. sativa* TA proteins.
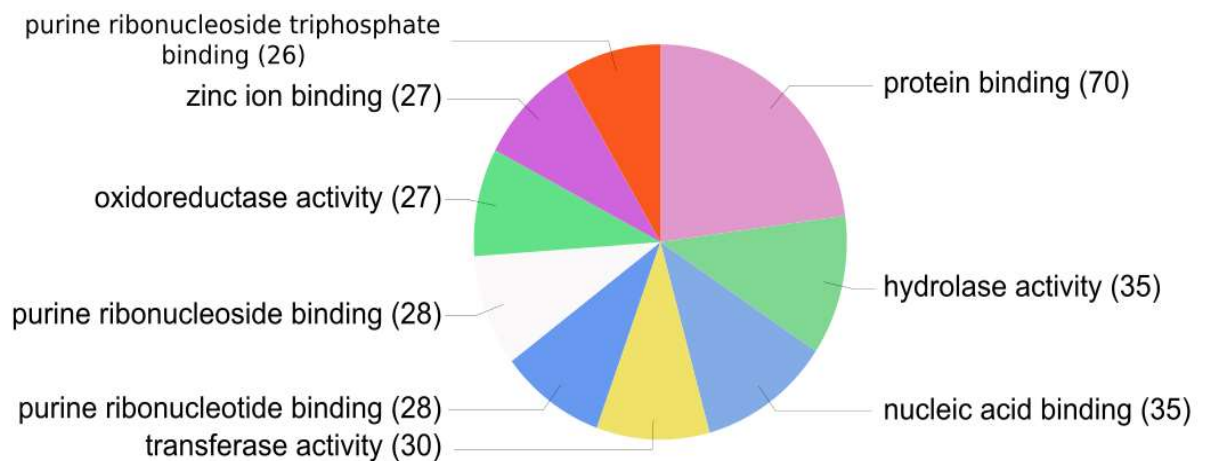
Figure 3.8: Functional distributions of *S. tuberosum* TA proteins.

3.2.5. Molecular weight distribution of TA proteins in *O. sativa* and *S. tuberosum*

Most of the TA proteins have molecular weight below 50kDa. If we compare the molecular weight of TA proteins targeted to choloroplast of both *O. sativa* and *S. tuberosum*, both of them have proteins falls into same molecular weight scale. But, in case of mitochondrial TA proteins, *S. tuberosum* has proteins with molecular weights less compared to *O. sativa* (Figure 3.9).

3.2.6. Transmembrane domain analysis of TA proteins

In case of TA proteins, the TMD is the major determinant factor for their location. These TMDs of TA proteins were analyzed for their length, hydrophobicity and amino acid frequencies.
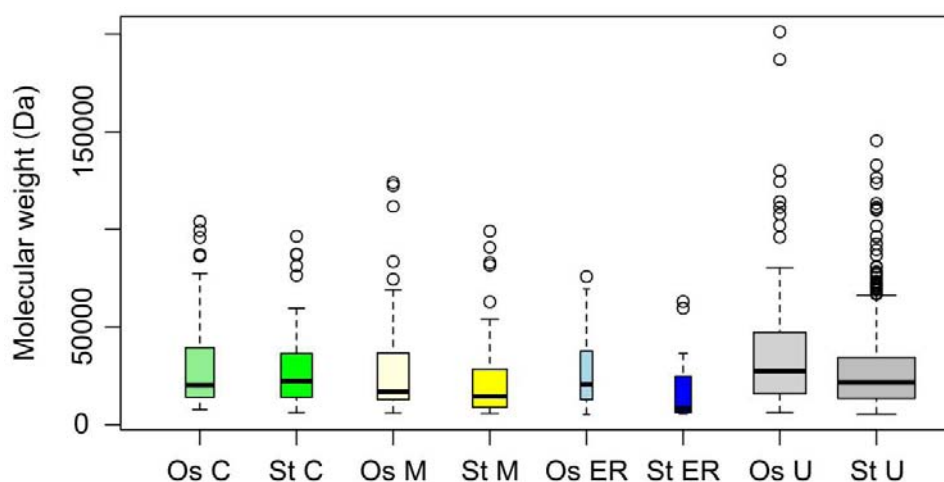
Figure 3.9: Organelle-wise molecular weight distribution of TA proteins in *O. sativa* and *S. tuberosum*. Os C - *O. sativa* chloroplast; Os M - *O. sativa* mitochondria; Os ER - *O. sativa* endoplasmic reticulum; Os U - *O. sativa* unkown location. St C - *S. tuberosum* chloroplast; St M - *S. tuberosum* mitochondria; St ER - *S. tuberosum* endoplasmic reticulum, St U - *S. tuberosum* unkown location.

3.2.6.1. Length distribution of TMD of TA proteins in *O. sativa* and *S. tuberosum*

The average length of the TMD was found to be 21 amino acids for most of the organelles in both *O. sativa* and *S. tuberosum*. In the case of *O. sativa*, maximum variation in length of TMD was observed in chloroplastic TA proteins. But in the case of *S. tuberosum*, TA proteins belonging to ER has the maximum variation in their length (Figure 3.10).
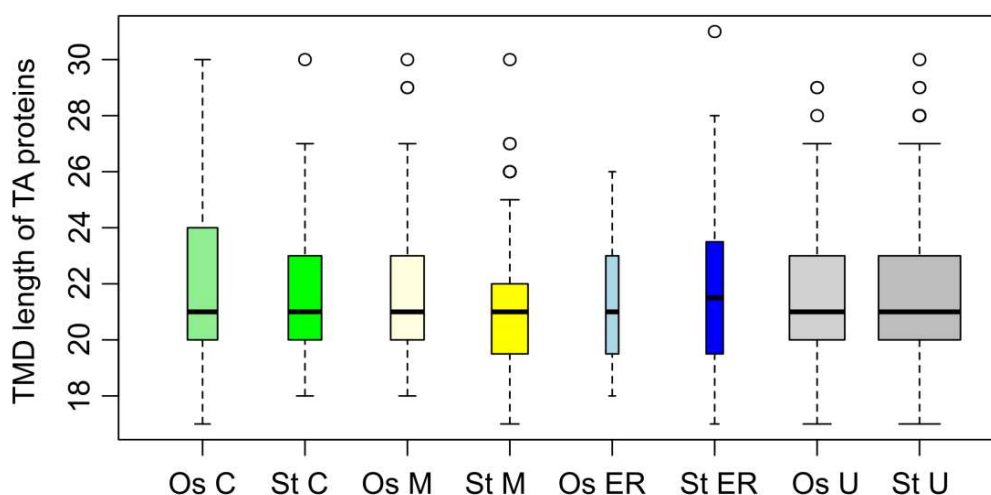
Figure 3.10: Organelle-wise trans-membrane length distribution analysis of identified TA proteins. Os C - *O. sativa* chloroplast; Os M - *O. sativa* mitochondria; Os ER - *O. sativa* endoplasmic reticulum; Os U - *O. sativa* unkown location. St C - *S. tuberosum* chloroplast; St M - *S. tuberosum* mitochondria; St ER - *S. tuberosum* endoplasmic reticulum, St U - *S. tuberosum* unkown location.

3.2.6.2. TMD Hydrophobicity analysis of TA Proteins in *O. sativa* and *S. tuberosum*

Hydrophobicity of TMD and TA proteins of both *O. sativa* and *S. tuberosum* were estimated based on Kyte and Doolittle scores. The overall hydrophobicity score for TA proteins are found less than zero. Average hydrophobicity of TMD was found higher for *O. sativa* TA proteins.
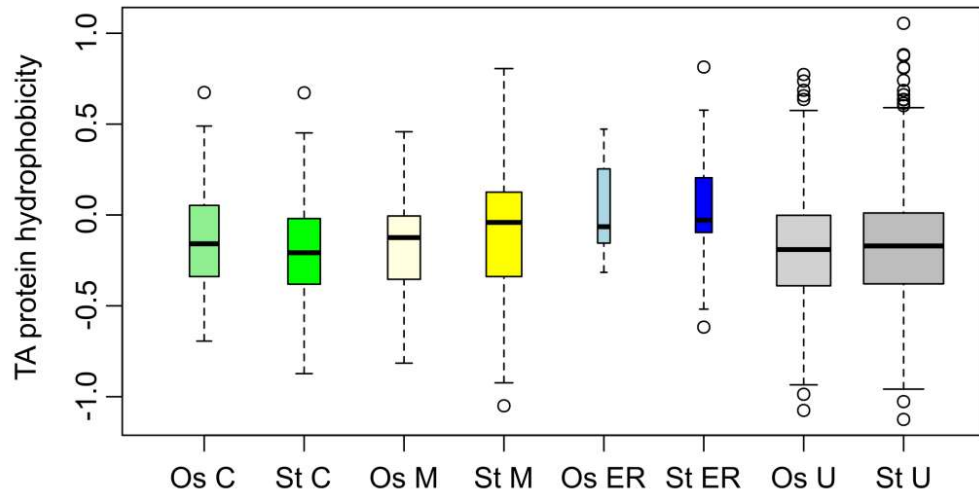
Figure 3.11: Organelle-wise hydrophobicity distribution of identified TA proteins. Os C - *O. sativa* chloroplast; Os M - *O. sativa* mitochondria; Os ER - *O. sativa* endoplasmic reticulum; Os U - *O. sativa* unkown location. St C - *S. tuberosum* chloroplast; St M - *S. tuberosum* mitochondria; St ER - *S. tuberosum* endoplasmic reticulum, St U - *S. tuberosum* unkown location.
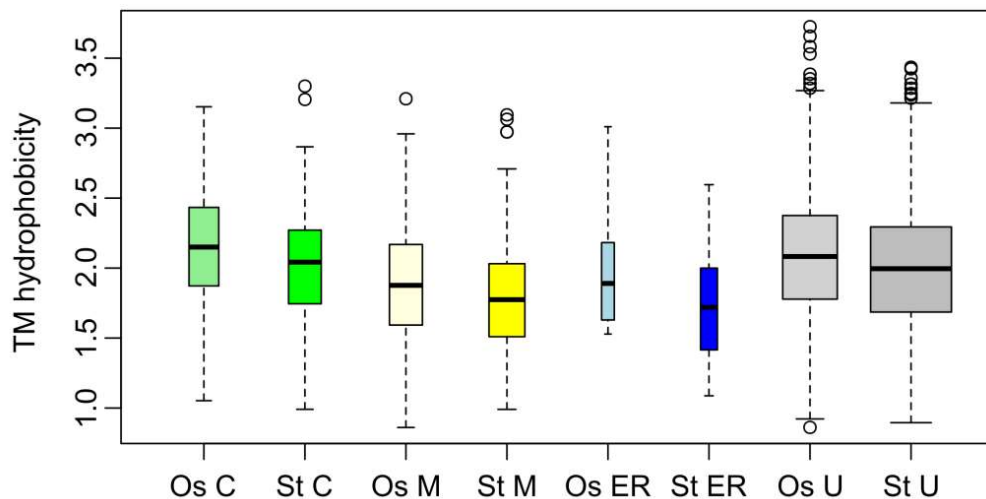
Figure 3.12: Organelle-wise trans-membrane hydrophobicity distributions of identified TA proteins. Os C - *O. sativa* chloroplast; Os M - *O. sativa* mitochondria; Os ER - *O. sativa* endoplasmic reticulum; Os U - *O. sativa* unkown location. St C - *S. tuberosum* chloroplast; St M - *S. tuberosum* mitochondria; St ER - *S. tuberosum* endoplasmic reticulum, St U - *S. tuberosum* unkown location.

Among all TA proteins, TMDs of chloroplast TA proteins in *O. sativa* were predicted to have higher hydrophobicity. The lowest hydrophobicity score for TMD was observed in mitochondrial TA proteins of *S. tuberosum*, while overall hydrophobicity of TA proteins was found to be higher for *S. tuberosum* mitochondria (Figure 3.11 and Figure 3.12). Highest average hydrophobicity score for TMD was observed in *O. sativa* chloroplastic TA proteins.

3.2.6.3. Amino acid frequency in TMD of TA proteins

The global amino acid frequency for TMD of TA proteins was also analyzed. The frequency of Leucine was higher compared to other amino acids. Valine, Isoleucine, Alanine and Phenylalanine were also identified in major proportion (Figure 3.13).
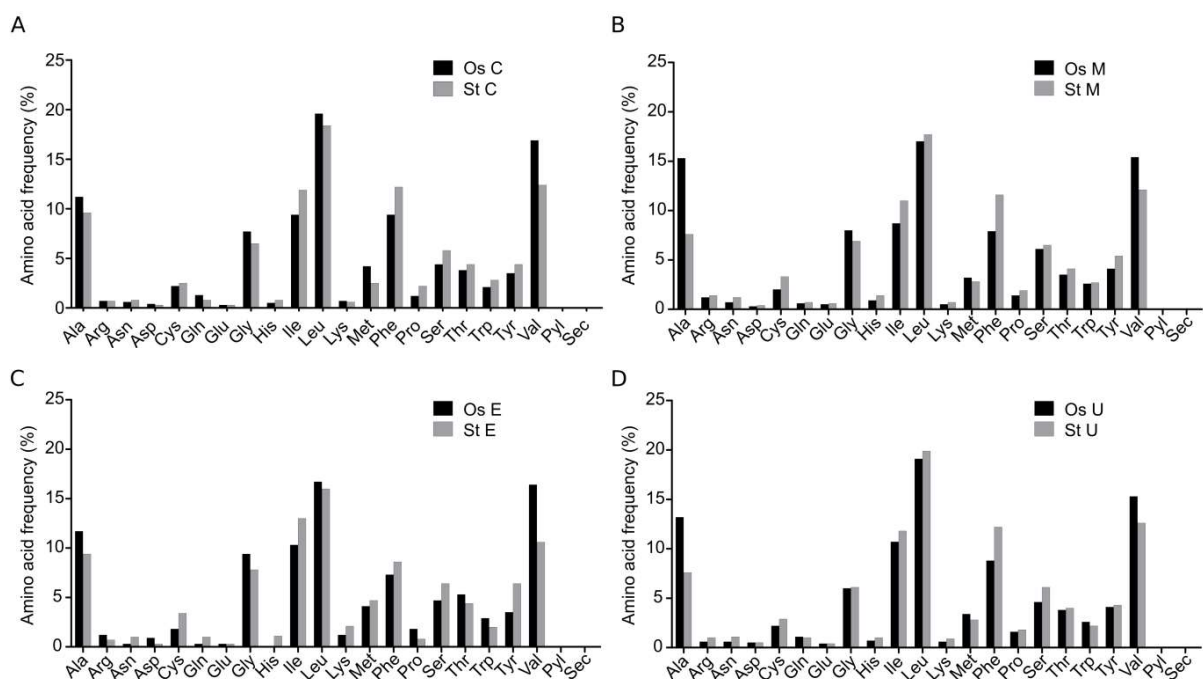
Figure 3.13: Global amino acid frequency analysis of transmembrane domain of *O. sativa* and *S. tuberosum* TA proteins. Amino acid frequency in transmembrane domain of TA proteins present in (A) chloroplast, (B) mitochondria, (C) endoplasmic reticulum and (D) unknown locations are shown. Os C - *O. sativa* chloroplast; Os M - *O. sativa* mitochondria; Os ER - *O. sativa* endoplasmic reticulum; Os U - *O. sativa* unkown location. St C - *S. tuberosum* chloroplast; St M - *S. tuberosum* mitochondria; St ER - *S. tuberosum* endoplasmic reticulum, St U - *S. tuberosum* unkown location.

3.2.7. GET Pathway components in selected plants

GET pathway is well studied in the yeast system. The homologous proteins for yeast GET pathway is also identified in the mammalian system (Table 3.2). Compared to other systems, plant system is more complicated because of (1) several GET pathway components like Get2 are missing (2) presence of more than one form of Get3.

Table 3.2: GET pathway components of yeast, mammalian and plant systems

Yeast	Mammalian	Plant
Get1	WRB	WRB
Get2	CAML	
Get3	TRC40	Get3*
Get4	TRC35	Get4
Get5	UBL4A	Get5
Ydj1		
Sgt2	SGTA	?
	Bag6	Bag6

Analysis using Pfam and orthoDB databases explored that *A. thaliana* has 4 unique Get3 sequences (out of 7 deposited) (Figure 3.14). Similarly, *O. sativa* and *S. tuberosum* have 3 and 5 Get3 respectively.

Figure 3.14: OrthoDB output for Get3. Plant species have more than one orthologs are shown in black colour.

Get2, a membrane protein to which the Get3 tethers during TA protein transport, was found missing in all analyzed plant species. Gene encoding Get5 was recently identified in *A. thaliana* that was not observed in *O. sativa* and *S. tuberosum*. All identified GET pathway components for analyzed plant species are listed in table 3.3.

Table 3.3: GET pathway components in selected plants

Yeast	Mammalian	*A. thaliana*	*O. sativa*	*S. tuberosum*
Get1	WRB	Q1H5D2	A2YAC8	M0ZME5
Get2	CAML	-	-	-
Get3	TRC40	Q949M9, A1L4Y1, Q5XF80, Q6DYE4	B8BDK7, A2X8V0, B8AIG1	M1AE77, M1A9X9, M0ZFY4, M1AND2, M0ZJQ4
Get4	TRC35	Q6GKV1	B8ADF2	M1ACL9
Get5	UBL4A	Q3E7K8	-	-

3.2.8. Phylogenetic analysis of Get3 in plants

All Get3 sequences of viridiplantae were analysed for their evolutionary relationship. Sequences were aligned and assembled in a phylogenetic tree. Interestingly, these sequences were clustered in three different clades: clade-a, clade-bc and clade-c (Figure 3.15). The further analysis reveals that the sequences are clustered based on their location pattern. Clade-a has the proteins which are localized to ER/cytoplasm. Clade-bc has proteins which are localized to chloroplast and mitochondria. Clade-d, a newly evolved clade in this study, mainly contains proteins localized to chloroplast. Based on this arrangement, the Get3 in plants are named as Get3a, Get3b, Get3c and Get3d based on their location. Even though Get3b and Get3d are localized to chloroplast, Get3d is different from Get3b. Out of all plant species analyzed, 51.7% plant species have both Get3b and Get3d. Also only 13.8% plant species have only Get3b. Interestingly 34.5% species have chloroplastic Get3d (Figure 3.16). Cytoplasmic Get3s, belong to clade-a, have much higher sequence identity with Get3 of other phyla. Get3d is evolutionarily very distinct form Get3a.

Figure 3.15: Phylogenetic tree of Get3 from viridiplantae.

Figure 3.16: Venn diagram of chloroplast Get3 in sequenced plants. 51.7 % of total sequenced plant species have two paralogs of chloroplast Get3.

3.2.9. Get3 in *O. sativa* and *S. tuberosum*

As discussed earlier, in comparison to other organisms, plant species have several orthologues for Get3. Also, several GET Pathway members are not yet characterized. Yet, these details were not investigated in crop plants. The Pfam analysis identified that *O. sativa* has three Get3 orthologues under three genes. Similarly, in *S. tuberosum*, five Get3 sequences were present under five genes. Target locations of these Get3s were predicted using the TargetP server (Table 3.4). Both the plant species have Get3 orthologues that belong to different organelles where ER, Mitochondria and Chloroplast are the main organelles. In *O. sativa*, ER TA proteins are targeted by a single Get3 (B8BDK7), but in *S. tuberosum*, three cytoplasmic Get3 (M1A9X9, M0ZFY4 and M1AND2) are present for targeting ER TA proteins. Both species have Get3 specifically for chloroplast and mitochondrial TA protein targeting.

Table 3.4: Locations of Get3 in selected plants

Organism	No of Get3	Name(Uniprot ID)	Location
A. thaliana	4	AtGet3a (Q949M9)	Cytoplasm/ER
		Atget3b (A1L4Y1)	Chloroplast
		AtGet3c (Q5XF80)	Mitochondria
		AtGet3d (Q6DYE4)	Chloroplast
O. sativa	3	B8BDK7	Cytoplasm/ER
		A2X8V0	Chloroplast
		B8AIG1	Mitochondria
S. tuberosum	5	M1AE77	Mitochondria
		M1A9X9	
		M0ZFY4	Cytoplasm/ER
		M1AND2	
		M0ZJQ4	Chloroplast

3.2.10. Get3 in *A. thaliana*

Preliminary analysis of GET pathway in plant system showed the presence of multiple Get3s. Since *A. thaliana* is a model system for several studies, this has been selected for structural and functional characterization of GET pathway in plants. Analysis of public databases for the sequences with ArsA/Get3/TRC40 motifs retrieves 10 entries for Get3 in *A.thaliana* (Figure 3.17). The detailed analysis of these 10 sequences (eliminating duplicate entries and incomplete sequence) gives four unique Get3 sequences. The Uniprot ids for the final sequences are Q949M9, A1L4Y1, Q5XF80 and Q6DYE4 (Figure 3.18). *In silico* localization predictions using TargetP and SignalP servers reveals that these four Get3 are localized to three different organelles. Also in the phylogenetic tree, these Get3 are clustered in three different clades. Based on the localization pattern and the position in the phylogenetic tree, these identified Get3s are further renamed. Q949M9 which is localized to ER/cytoplasm is named as AtGet3a. A1L4Y1 which is localized to chloroplast is named as AtGet3b. Similarly, Q5XF80 and Q6DYE4 were renamed as AtGet3c and AtGet3d. Q5XF80 is localized to mitochondria and Q6DYE4 is localized to Chloroplast.
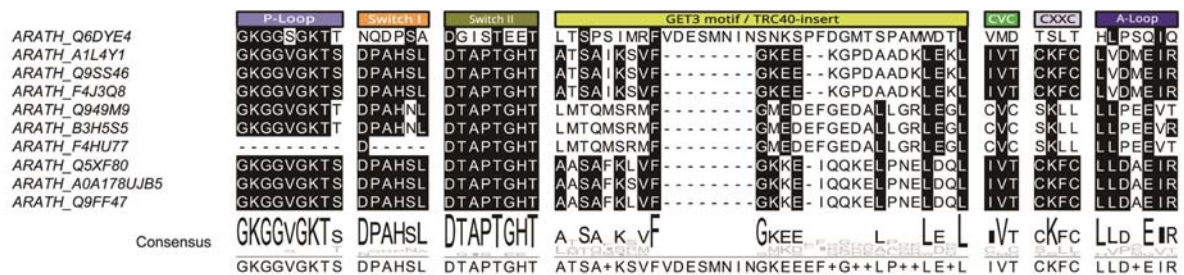
Figure 3.17: Sequence alignment of all *A. thaliana* Get3 annotated in UniProt (important domains marked).
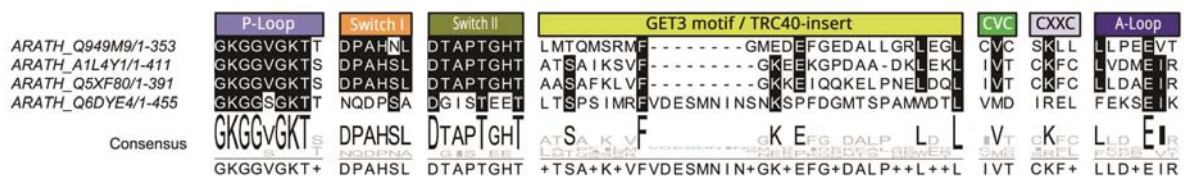
Figure 3.18: Sequence alignment of four *A. thaliana* Get3 (domains marked).

The phylogenetic analysis of AtGet3 with yeast Get3 displays that the cytoplasmic AtGet3a is more related to yeast Get3. Also, AtGet3b and AtGet3c are clustered in the same clade (Figure 3.19). AtGet3d is distinct from all other *A. thaliana* Get3 and found in a separate clade. The sequence features of AtGet3d are different from other known Get3. Like other Get3, AtGet3d

also has conserved P-loop and Get3 motif. In addition to these features, AtGet3d has an α-crystalline domain at C-terminal region. Presence of this domain may make AtGet3d distinct from other Get3 phylogenetically.
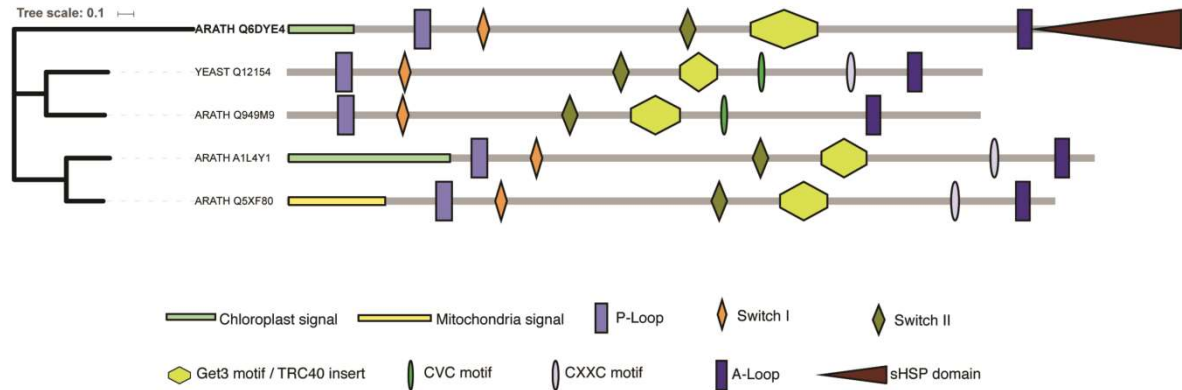
Figure 3.19: Phylogenetic tree of *A. thaliana* Get3 paralogs with yeast Get3. Q6DYE4 (AtGet3d) is found in separate clad. Cytoplasmic Get3 (Q949M9) clustered together with yeast Get3 (Q12154).

Sequence alignment of Arabidopsis Get3 with other Get3 has been carried out (Figure 3.20). The signature motif for P-loop was conserved in all the analyzed Get3s. AtGet3d has a single amino acid variation in the P-loop region. CXXC motif was absent in both cytoplasmic AtGet3a and chloroplastic AtGet3d.
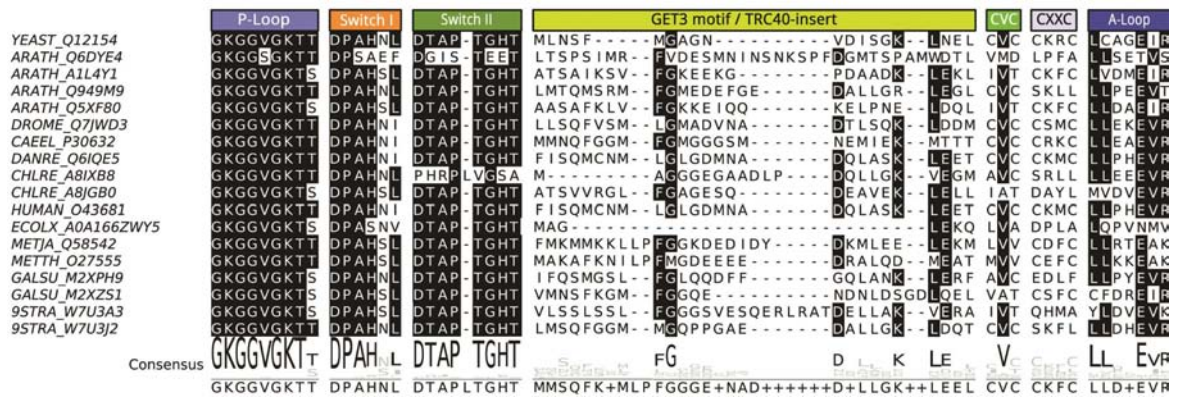
Figure 3.20: Sequence alignment of Get3 paralogs from different species.

Similarly, *A. thaliana* Get3 were phylogenetically analyzed with the representative Get3 members of other phyla (Figure 3.21). Cytoplasmic Get3 is found to cluster with yeast Get3 along with the Get3 of other organisms including human Get3 homologue TRC40. AtGet3a is more closely related to *Clamydomonas reinhardtii* Get3 (A8IXB8). AtGet3b and AtGet3c are evolved from a common ancestor. They are clustered together in a clade along with other Get3 orthologues of rhodophyta, chlorophyta and stramenopites. AtGet3d is different from

other *A.thaliana* Get3 and clustered with *E. coli* (proteobacteria) Ars ATPase and archeal Ars ATPase.

Figure 3.21: Phylogenetic tree of Get3 paralogs with domain structure from different species. Q6DYE4(AtGet3d) is found in separate clad. Species colour code: green, viridiplante; orange, metazoa; yellow, archaea; red, red algae; purple, yeast; grey, bacteria.

More than 80% of sequenced plant species have Get3 which is similar to AtGet3d. Sequence alignment of AtGet3d with other plant Get3d shows well conservation in P-loop region and C-terminal HSP region. Also, analyzed Get3d sequences have serine instead of valine in P-loop region. α-crystalline domains of these Get3d are well conserved among the plant species (Figure 3.22).

Figure 3.22: Sequence alignment of α-crystallin (HSP) domains in Get3d. HSP domain of Get3d is well conserved in analysed plant species.

3.3. Discussion

The previous study in *A. thaliana* reported the presence of 502 TA proteins in its proteome (Kriechbaumer et al. 2009). This study in this Chapter 3 explored *Oryza sativa* subsp. Indica and *Solanum tuberosum*, *in silico*, for the presence of TA protein targeting pathway (Manu et al. 2018). Also, detailed analysis was carried out in *A. thaliana* for GET pathway. All analyzed plant species have less than 2% TA proteins while comparing to their total proteome. Yet, these TA proteins are involved in several vital cellular processes like vesicular transport, vesicle fusion etc.

The number of Get3 also differs across the plant species studied. *O. sativa* has three Get3, *S. tuberosum* has five Get3 and *A. thaliana* has four Get3 as predicted by our analysis. These identified Get3 are highly organelle specific in TA protein targeting. Phylogenetic analysis reveals that the Get3s belong to clade-d are diverged from rest of the Get3s. Also, clade-d has Get3 which target TA proteins to chloroplast. Get3 clustered in clade-d have a domain fusion event with α-crystalline domain. Moreover, 80% of the sequenced plant species have chloroplast Get3s from clade-d compared to clade-b. This is the first study that highlights the detailed analysis of chloroplast Get3 with an HSP domain. Further structural and functional characterization of chloroplast Get3 is performed and mentioned in the following chapters in this thesis.

Functional characterization of Get3 in plants

4.1. Introduction

In silico analysis in Chapter 3 showed that *A. thalina*, *O. sativa* and *S. tuberosum* have more than 500 TA proteins in their proteome. These proteins are localized to different organelles like ER, mitochondria, chloroplast etc. Besides this, organelle-specific Get3 are also present in these organisms. In *A. thaliana*, two Get3 were identified for chloroplast, AtGet3b and AtGet3d. From such *in silico* analysis, it was observed that Get3 which is similar to AtGet3d, is found in more than 80% sequenced plants. Also, it is evident from Chapter 1 literature reviews and Chapter 3 *in silico* analysis, there is no detailed functional and structural analysis available for Get3d of plants. Henceforth for further characterization and analysis, we have selected AtGet3d (Chloroplast) and AtGet3a (Cytoplasm) from *A. thaliana*. In this Chapter 4, the functional characterization of AtGet3d (Chloroplast) has been performed to understand the organelle-specificity of this GET pathway in *A. thaliana*.

4.2. Results

4.2.1. Cloning of AtGet3a

The coding region of AtGet3a (Cytoplasmic Get3) was amplified from cDNA using primers listed in table 2.1. The amplified product was then cloned between the restriction sites of pET22b. The positive clones were selected by colony PCR. The isolated plasmids were subjected to restriction digestion for further confirmation. Restriction digestion shows the release of 1 kb insert (Figure 4.1). Positive clones were maintained and the plasmid was isolated for confirmation using sequencing. Sequencing results thus confirmed that AtGet3a has been cloned properly without any mutation.
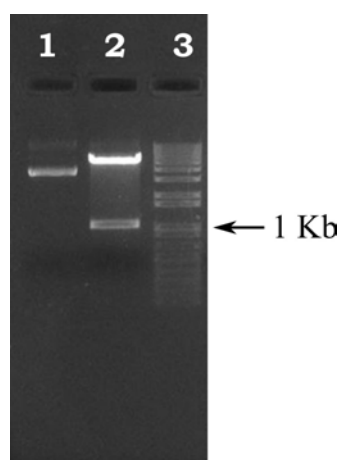
Figure 4.1: Double digestion profile of AtGet3a /pET22b: Lane 1: Undigested plasmid; Lane 2: Digested positive clone; and Lane3: Marker

4.2.2. Cloning of AtGet3Δd

Initial clone of AtGe3d was a gift from Prof. Bil Clemons. Full-length AtGet3d was cloned in in pET33b vector between SalI and PstI restriction sites. This clone was also confirmed by restriction digestion and sequencing (Figure 4.2A). During initial purification trials, SDS-PAGE showed that two protein bands were always coming together (Figure 4.2B). The LC-MS/MS analysis of these proteins further confirmed that both these bands belong to same proteins with N-terminal truncation (Peptides identified are listed in Table A4). In order to get a stable and homogeneous protein, AtGet3d was further recloned in pET22b.
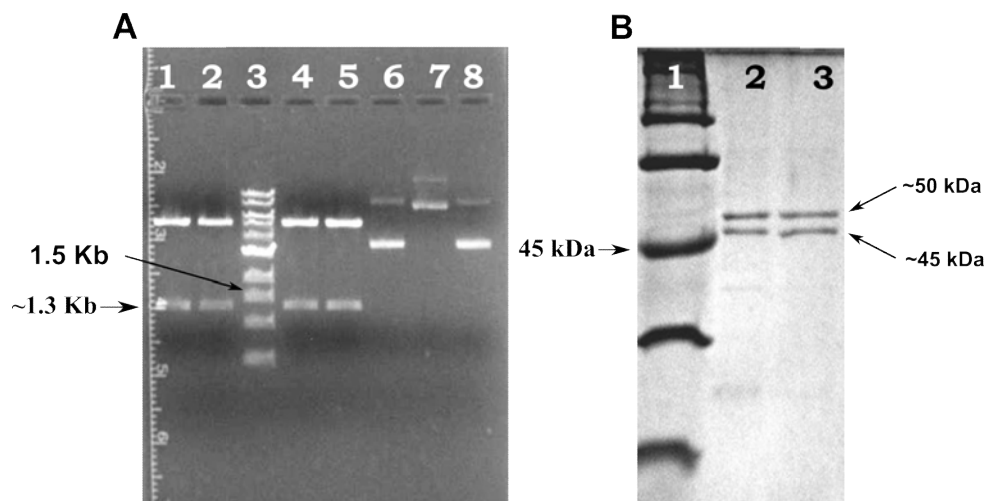
Figure 4.2: (A) Double digestion profile of AtGet3d /pET33b: lane:: 1-2: Digested; 3: Marker; 4-5: Digested; 6-8: Undigested **(B)** SDS PAGE profile of AtGet3d during purification. Lane:: 1:marker; 2-3: purified AtGet3d

The disorder prediction showed that the N-terminal region of AtGet3d has a disorder and this region encodes for the targeting signal of AtGet3d. These are the possible reasons for getting two proteins bands during purification.
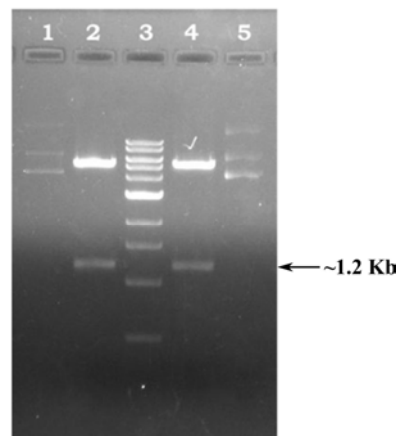
Figure 4.3: Double digestion profile of AtGet3Δd /pET22b Lane 3: Marker; Lane 1 and 5: Undigested plasmid; Lane 2 and 4: Digested positive clones

In order to get a stable protein during expression and purification, 57 amino acids from N-terminal were removed and the truncated AtGet3Δd was cloned between NdeI and HindIII restriction sites of pET22b. Positive clones were identified by PCR and restriction digestion (Figure 4.3). The sequencing results further confirmed that AtGet3Δd has been properly inserted in the pET22b(+) vector.

4.2.3. Purification of AtGet3a

Bacterial expression system was used to express AtGet3a as mentioned in the methods section in Chapter 2. AtGet3a was expressed in *E. coli* Rosetta DE3 cells with C-terminal 6X His tag. The stable and soluble protein was then over-expressed by inducing with 0.1M IPTG at an optimized temperature of 16° C.
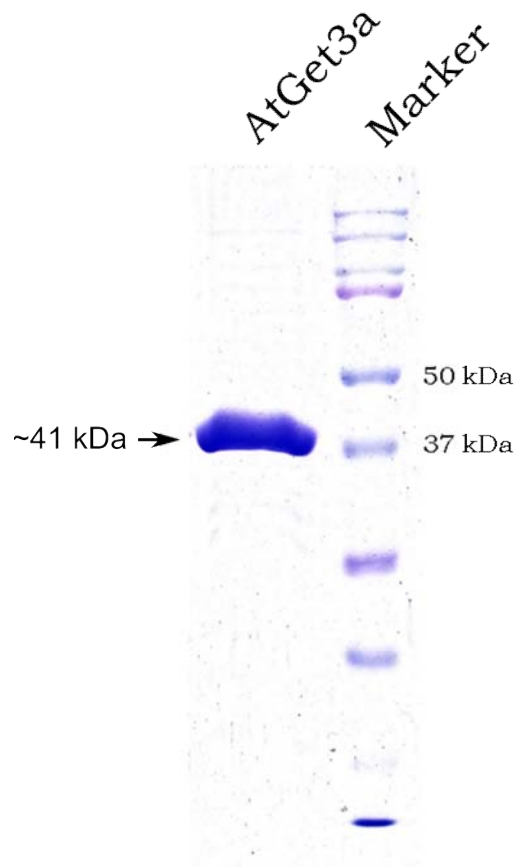
Figure 4.4: SDS PAGE of purified AtGet3a

Purification of over-expressed AtGet3a was achieved by two-step process involving Ni-NTA affinity chromatography and size exclusion chromatography (Superdex 200/GE) and the SDS-PAGE result is shown (Figure 4.4).

Some aggregation was observed during the purification process and the properly folded proteins were selected for further analysis. The size exclusion chromatography profile showed that AtGet3a occurs as a tetramer in solution (Figure 4.6).

4.2.4. Purification of AtGet3Δd

Truncated AtGet3d (AtGet3Δd) was cloned in pET22b+ expression vector and protein was expressed in *E. coli* BL21 star (DE3) cells with C-terminal 6XHis tag. 57 amino acids deleted from N-terminal of AtGet3d are highlighted in figure A1. Expression of AtGet3Δd was induced with 01.M IPTG at 16°C overnight.
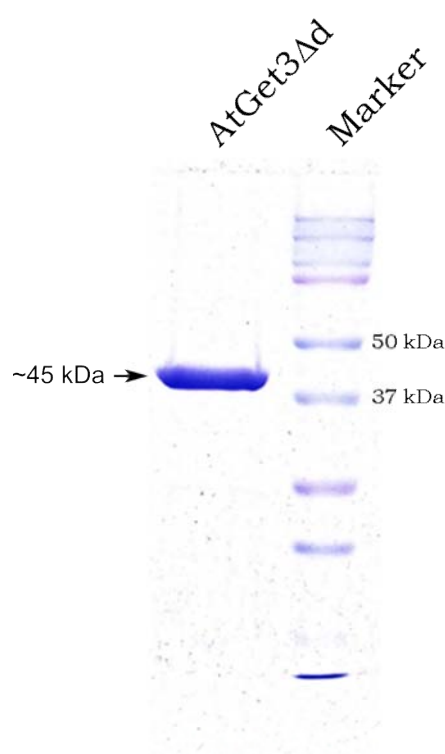
Figure 4.5: SDS PAGE of purified AtGet3Δd

The over-expressed soluble protein was purified to homogeneity by Ni-NTA affinity chromatography followed by size exclusion chromatography (Superdex 200/GE) and the SDS-PAGE result is shown (Figure 4.5). Size exclusion profile shows that AtGet3Δd occurs as a dimer in solution (Figure 4.6). It was also observed that AtGet3Δd tends to precipitate at 4°C upon storage. So the purified AtGet3Δd was immediately used for further experiments including crystallization.
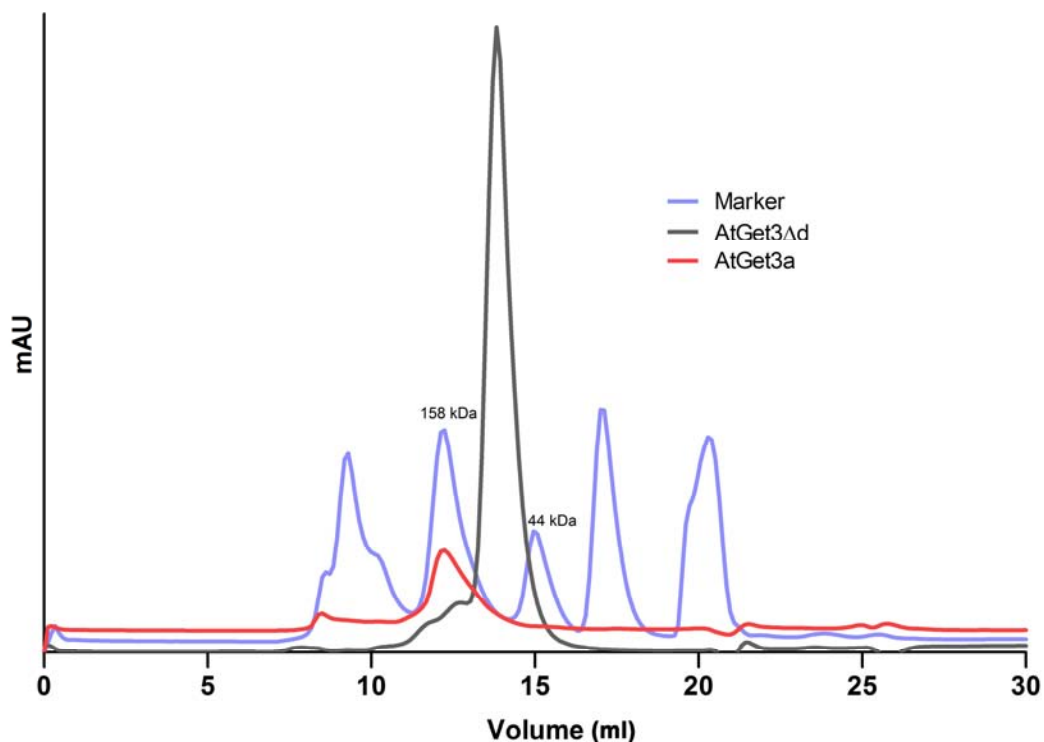
Figure 4.6: Size exclusion chromatography profile of purified AtGet3a and AtGet3Δd in superdex 200 10/300GL. AtGet3Δd occurs as dimer in solution while AtGet3a forms tetramer in solution.

4.2.5. ATPase activity of AtGet3d

ATPase activity of both AtGet3a and AtGet3d were checked according to the protocol mentioned in Chapter 2.14. Both AtGet3a and AtGet3d were found to be able to hydrolyse ATP.
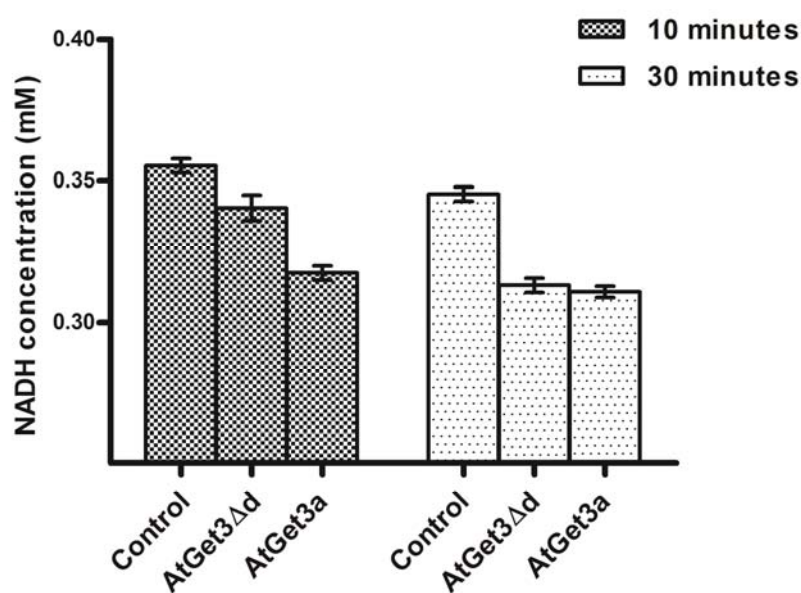
Figure 4.7: ATPase activity of both AtGet3a and AtGet3d.

Although AtGet3a hydrolysed ATP at a higher initial rate, both these Get3s were able to hydrolyse the same amount of ATP after 30 min (Figure 4.7). This difference in the hydrolysis rate may be associated with P-loop residues. In case of AtGet3d, conserved Val in the P-loop was replaced with Ser (Figure 3.20). Structural analysis of P-loop of both AtGet3a and AtGet3d would possibly correlate the reason for this difference in their activity (Chapter 5.2.7).

4.2.6. Gene expression analysis of AtGet3

Gene expression data for all *A. thaliana* Get3 under stress condition were collected from NASC Arrays expression database through BAR as mentioned in Chapter 2.19. All the expression values were normalized to control samples and plotted as a heat-map. Gene expression under stress conditions like drought, heat, osmotic pressure, salt and UV-B were analyzed (Figure 4.8). Expression patterns of AtGet3b and AtGet3a are highly correlated compared to other AtGet3. Expression of AtGet3c is found higher during UV-B stress, suggesting that UV-B has more effect on mitochondrial-associated proteins. Interestingly, expression patterns of AtGet3d during stress conditions are different from other AtGet3. Stress conditions like salt, osmotic, heat and drought are found to be triggering the expression of AtGet3d. Previous reports suggested that in yeast, during stress condition, Get3 acts as chaperon.
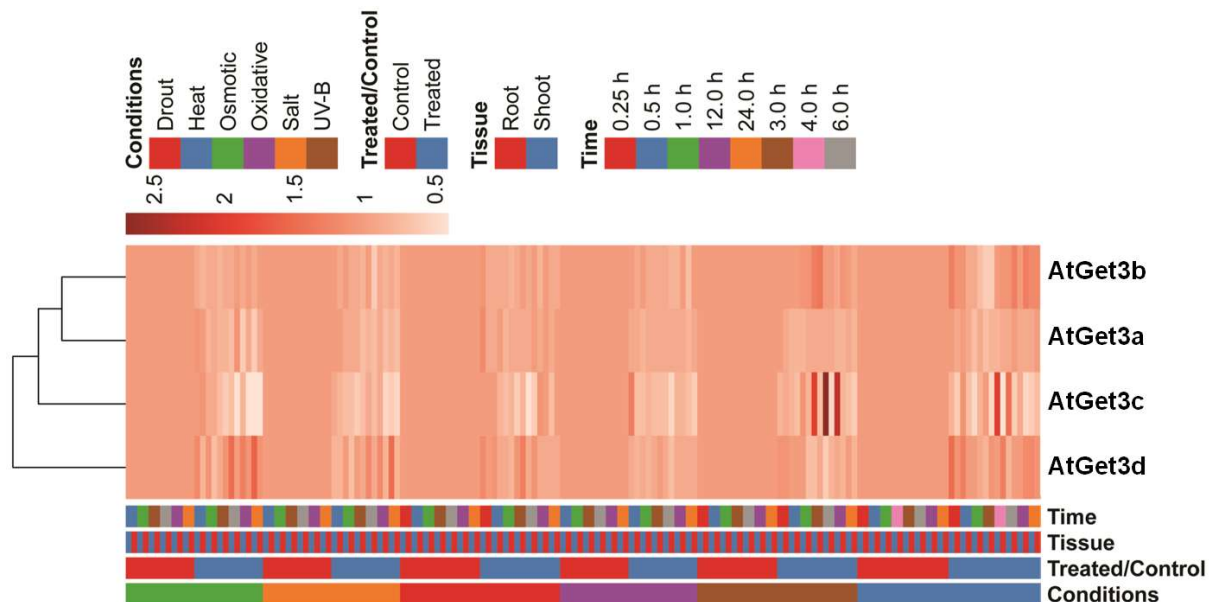
Figure 4.8: Expression analysis of AtGet3 genes (At1g26090, At5g60730, 56 At1g01910 and At3g10350) under different stress conditions. Expression data were collected from NASC Arrays expression database through e-Northerns Expression Browser of BAR and heat map was generated using ClustVis web tool.

Over-expression of AtGet3d during stress condition also supports and suggests that AtGet3d can act as chaperon. Expression of AtGet3d in different plant organs was also analyzed and found to be higher in aerial tissues especially in leaves. Besides leaf tissues, expression was also found in shot apex and flowers (Figure 4.9).
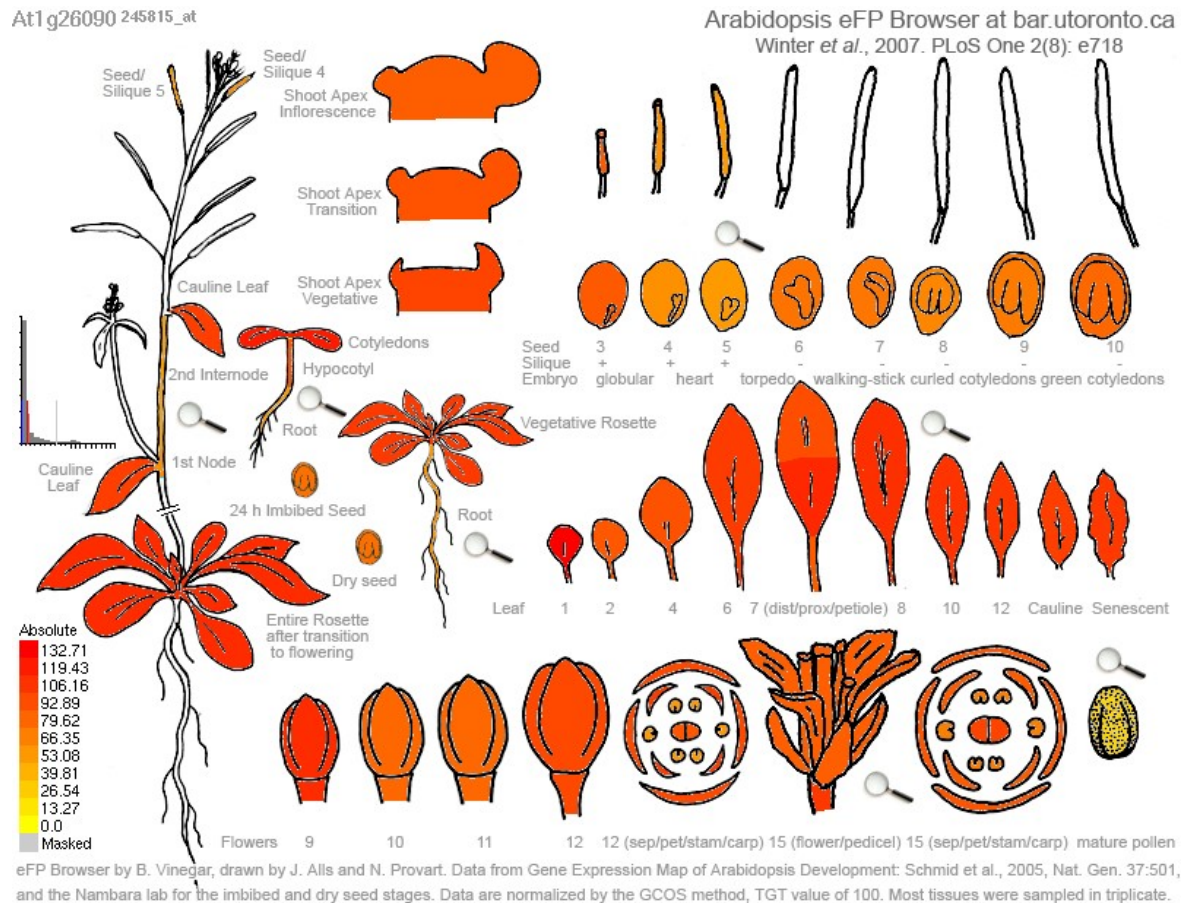
Figure 4.9: Organ wise expression pattern of AtGet3d in *A. thaliana*. Expression of AtGet3d (At1g26090) is found higher in aerial tissues.

4.2.7. Phenotypic characterization

In order to delineate the characteristics of the phenotype in the absence of AtGet3d, the *Arabidopsis* T-DNA insertion mutant of AtGet3d (SALK_076216C) was obtained from Arabidopsis Biological Resource Center (ABRC) (Alonso et al. 2003) and used. Both wild-type and AtGet3d insertion mutant seeds were germinated on MS agar plate. Wild-type seeds germinated much faster than the mutant strains (Figure 4.10). But in the later stage, no variation was observed with wild-type plants indicating that the mutational effect of AtGet3d is not lethal and possibly some alternate mechanism is present to nullify the effect of AtGet3d deletion.
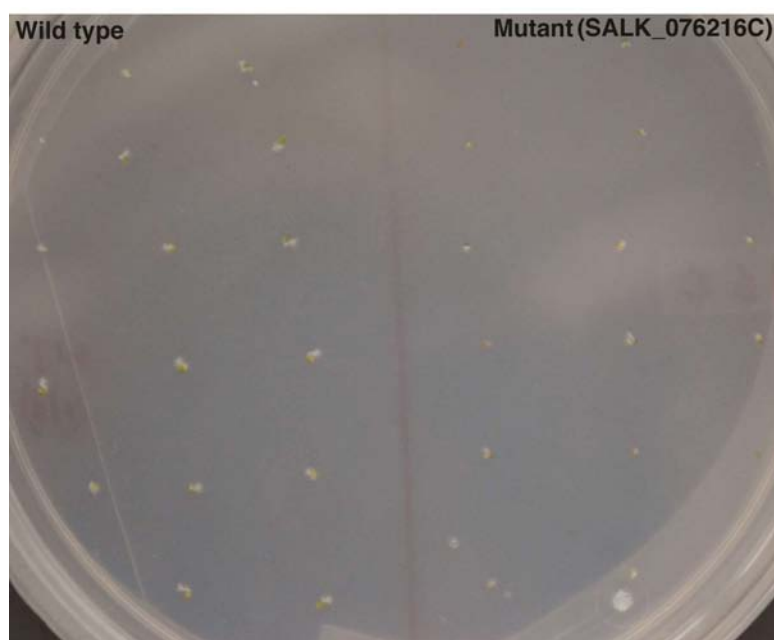
Figure 4.10: Phenotype analysis of AtGet3d mutant. One week old germinating seeds of wild-type (WT) and mutant (MT) *A. thaliana* in MS agar plate. WT germinated fast compared to mutant. But in later stage, no significant difference was observed.

4.2.8. Subcellular localization of AtGet3d

The previous reports in *Arabidopsis* have shown that AtGet3a, AtGet3b and AtGet3c were localized into cytoplasm/ER, chloroplast and mitochondria, respectively (Table 4.1). No detailed investigation is carried out in the case of AtGet3d. Sequence analysis showed that Atget3d is encoded by the nuclear DNA of *A. thaliana*, and signal prediction by TargetP 1.1 server suggested that it can localize in chloroplast (Emanuelsson et al. 2000).

Table 4.1: Localization pattern of AtGet3

AtGet3	Location	Reference
AtGet3a	Cytoplasm/ER	O. Duncan, M. J. van der Merwe, D. O. Daley, J. Whelan, *Trends Plant Sci.* 18, 207–217 (2013).
AtGet3b	Chloroplast	
AtGet3c	Mitochondria	
		S. Xing *et al.*, *Proc. Natl. Acad. Sci.* 114, E1544–E1553 (2017).
AtGet3d	???	

To confirm the localization and the specificity of organelle targeting of AtGet3d, we have

employed organelle separation and western blot, immunofluorescence microscopy and thermolysin treatment followed by mass spectrometry.

4.2.8.1. Immunoblot for identifying the subcellular localization of AtGet3d

Plant cell organelles were isolated from mature leaves of *A. thaliana* and separated by density gradient ultra-centrifugation as mentioned in Chapter 2. The isolated organelle fractions of cytosol, mitochondria and chloroplast were subjected to western blot analysis against custom-made anti-AtGet3Δd antibody. The signal was detected in the chloroplast fraction, suggesting that AtGet3d is localized in chloroplast (Figure 4.11).
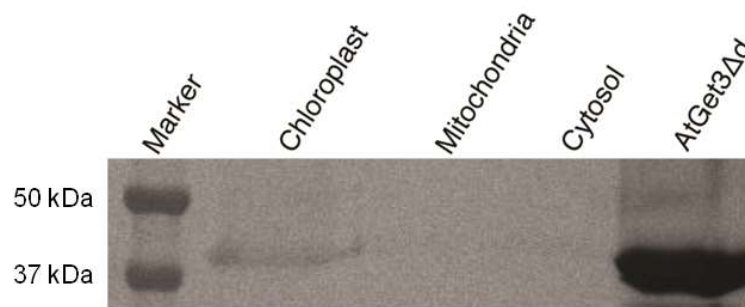
Figure 4.11: Immunoblot analysis to identify the subcellular localization of AtGet3d. Organelle isolated and subjected to western blot using polyclonal anti- AtGet3Δd antibody. The signal from chloroplast fraction indicates AtGet3d localized to chloroplast.

4.2.8.2. Immunofluorescence assay for identifying the subcellular localization of AtGet3d

The subcellular localization of AtGet3d was further confirmed by immunofluorescence followed by confocal microscopy. Wild-type and mutant leaves were fixed in 4% paraformaldehyde followed by microtomy. 8-10 micrometre thick sections were subjected to immune-staining with custom-made AtGet3Δd antibodies. Anti-rabbit IgG conjugated to the Alexa Fluor 488 fluorochrome was used as secondary antibody. A long pass filter of 585 nm and a band pass filter of 502 to 550 nm were used for detecting the emission signals of chloroplast autofluorescence and green fluorescence respectively.

A strong co-localization signal with chlorophyll autofluorescence (red) and Alexa Fluor 488 fluorochrome (green) was observed in immunofluorescence assay of wild-type mature leaves, but absent in AtGet3d mutant. Fluorescence signals from the surface of chloroplast suggest that anti-AtGet3Δd binds on the surface of chloroplast (Figure 4.12). This result further confirms that AtGet3d gets localized onto chloroplast membrane.
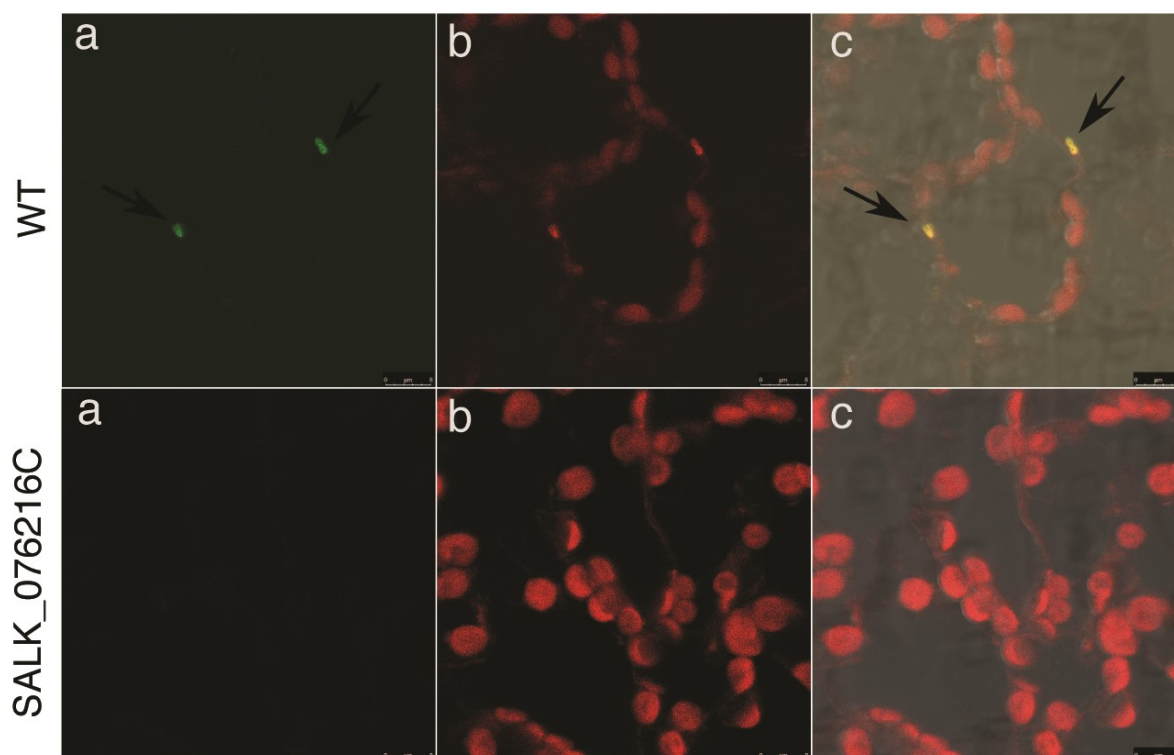
Figure 4.12: Confocal microscopic images of leaf sections of both wild-type and mutant (SALK_076216C) Arabidopsis thaliana (Immunofluorescence using anti- AtGet3Δd antibody as primary antibody and Alexa Fluor 488 as secondary antibody). Arrows indicated shows AtGet3Δd localized in chloroplast. (a, Alexa Fluor 488 green fluorescence; b, chloroplast autofluorescence; c, merged)

4.2.8.3. Digestion with thermolysin to identify the subcellular localization of AtGet3d

To further confirm whether AtGet3d is localized on the surface of the chloroplast or inside, intact chloroplasts were treated with thermolysin. Thermolysin is known to cleave extracellular part of membrane proteins or protein that are interacting with membrane proteins without disturbing the chloroplast membrane or chloroplast. Following thermolysin treatment, AtGet3d (Uniprot ID - Q6DYE4) was identified in the supernatant by mass spectrometry (highlighted in Table 4.2 and Figure A3), which further confirms that AtGet3d is localized on the surface of chloroplast.

Table 4.2: Identification of proteins localized on chloroplast surface

Uniprot ID	Localization	Uniprot Description
Q9SY97	Chloroplast	Photosystem I chlorophyll a/b-binding protein 3-1
P38418	Chloroplast	Lipoxygenase 2
P23321	Chloroplast	Oxygen-evolving enhancer protein 1-1

Uniprot ID	Localization	Uniprot Description
Q9S841	Chloroplast	Oxygen-evolving enhancer protein 1-2
Q9SYW8	Chloroplast	Photosystem I chlorophyll a/b-binding protein 2
Q9SUI7	Chloroplast	Photosystem I reaction center subunit VI-1
Q9SUI6	Chloroplast	Photosystem I reaction center subunit VI-2
Q9SE50	ER	Beta-D-glucopyranosyl abscisate beta-glucosidase
Q9SA56	Chloroplast	Photosystem I reaction center subunit II-2
Q9S7H1	Chloroplast	Photosystem I reaction center subunit II-1
Q9LPW0	Chloroplast	Glyceraldehyde-3-phosphate dehydrogenase GAPA2
P25856	Chloroplast	Glyceraldehyde-3-phosphate dehydrogenase GAPA1
Q84X02	Unknown	FBD-associated F-box protein
Q01667	Chloroplast	Chlorophyll a-b binding protein 6
P56779	Chloroplast	Cytochrome b559 subunit alpha
P10896	Chloroplast	Ribulose bisphosphate carboxylase/oxygenase activase
O03042	Chloroplast	Ribulose bisphosphate carboxylase large chain
Q6DYE4	**Chloroplast**	**Uncharacterized protein**
O65782	Membrane	Cytochrome P450 83B1
Q9FHG4	Membrane	Probable L-type lectin-domain containing receptor kinase S.7
P27521	Chloroplast	Chlorophyll a-b binding protein 4
O04153	ER	Calreticulin-3
Q0WKZ3	Mitochondria	Pentatricopeptide repeat-containing protein
Q6NKS1	Unknown	Probable protein phosphatase 2C 65

4.2.9. Co-expression and pull-down of AtGet3a and AtGet3d with selected TA proteins

To access the organelle-specific targeting ability of AtGet3Δd, TA proteins that belong to different organelles were selected for co-expression in *E. coli C41* (DE3) and pull-down studies (see methods section 2.11.2 and 2.11.3). In yeast, Get3 is reported to target TA proteins to ER. In plant system, more than one orthologues of Get3 are identified. In Arabidopsis, four Get3 are identified with different organelle specificity. This study has identified that AtGet3d is localized to the chloroplast. In order to find out whether AtGet3d is able to bind with TA proteins, 12 different TA proteins (Table 4.3) were selected, out of which 3 TA proteins (highlighted in Table 4.3) were successfully cloned and co-expressed with

AtGet3a and AtGe3d. Vesicle-associated membrane protein 722 (VA722), Thylakoid lumen 18.3 kDa protein (U603) and Peroxisomal and mitochondrial division factor 2 (PMD2) whose N-terminal signals were predicted to target to the cell membrane, chloroplast and mitochondria respectively were co-expressed with AtGet3a and AtGet3d.

Table 4.3: Selected TA proteins for co-expression with AtGet3a and AtGet3d. Successfully cloned TA proteins are highlighted.

SI No	Uniprot ID	Protein Name	Protein Size (in Da)	Localized
1	P93030	RMA2_ARATH	22,178.3	Cytoplasm/ER
2	O81045	P24D8_ARATH	24,565.4	Cytoplasm/ER
3	Q94AU2	SEC22_ARATH	25,332.2	Cytoplasm/ER
4	Q9SJL6	MEM11_ARATH	25,628.3	Cytoplasm/ER
5	Q9SHJ6	PMD2_ARATH	36,119.0	Mitocondrial OM
6	Q9SHC8	VAP12_ARATH	26,442.1	Cytoplasm/ER
7	P47192	VA722_ARATH	24,928.0	Cytoplasm/ER
8	Q9STT2	VPS29_ARATH	20,968.2	Cytoplasm/ER
9	Q9ZVL6	U603_ARATH	31,139.0	Chloroplast
10	F4KER9	TraB family protein	28,173.1	Cytoplasm/ER
11	Q9LM91	CCB25_ARATH	30,801.1	Cytoplasm/ER
12	F4I2U7	F4I2U7_ARATH	22,346.0	Cytoplasm/ER

Figure 4.13: Coomassie blue staining of co-expressed and pulldown samples. TA proteins specific to ER (VA722), chloroplast (U603), and mitochondria (PMD2) were co-expressed and pulled down with AtGet3d and AtGet3a. AtGet3a was found to bind effectively with VA722. Both AtGet3a and AtGet3d can bind with U603, but the binding capacity is observed higher for AtGet3d.

Co-expression and pull-down assays identified that AtGet3a has a higher affinity for VA722. While both AtGet3a and AtGet3d can bind with chloroplast TA protein, the pull-down efficiencies are different. AtGet3d can pull down chloroplastic TA protein more competently compared to AtGet3a with ~2.5 fold increased efficiency (Figure 4.13 and Figure 4.14). Both the Get3s could only pull down very feeble or negligible amount of mitochondrial TA protein.
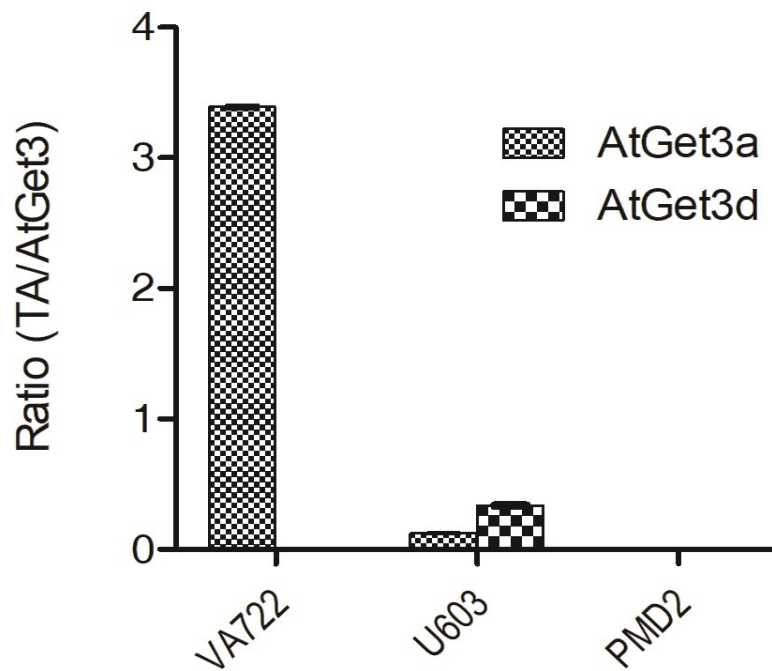
Figure 4.14: Graphical representation of pull-down efficiency. Graph indicates VA722 was pulldown by AtGet3a efficiently. Both AtGet3d and AtGet3a can bind with U603, but 2.5 times greater binding efficiency is seen in case of AtGet3d.

All three selected TA proteins have TMD of 21 amino acids. Hydrophobicity scale of TMD of these TA proteins shows that VA722 is more hydrophobic at the middle region compared to other TA proteins (Figure 4.15). Gravy score was also higher for TM of VA722 compared to other TA proteins. This difference in hydrophobicity distribution pattern of TMD can possibly have some influence in the binding preference of Get3 and TA proteins.

4.2.10. Co-immunoprecipitation and Mass spectrometry analysis to identify proteins interacting with AtGet3d

Sequence analysis of AtGet3d highlighted the presence of an HSP domain at the C-terminal region. Also, the expression analysis and co-expression studies indicated that AtGet3d is over-expressed during stress condition and is able to bind effectively with TA proteins specific to chloroplast. In comparison with yeast, several GET pathway components are missing in

plants. The functionality of these missing components may have been replaced by other proteins. Co-immunoprecipitation followed by Mass spectrometric (MS) analysis is employed here-in to uncover all the proteins that interact with AtGet3d.
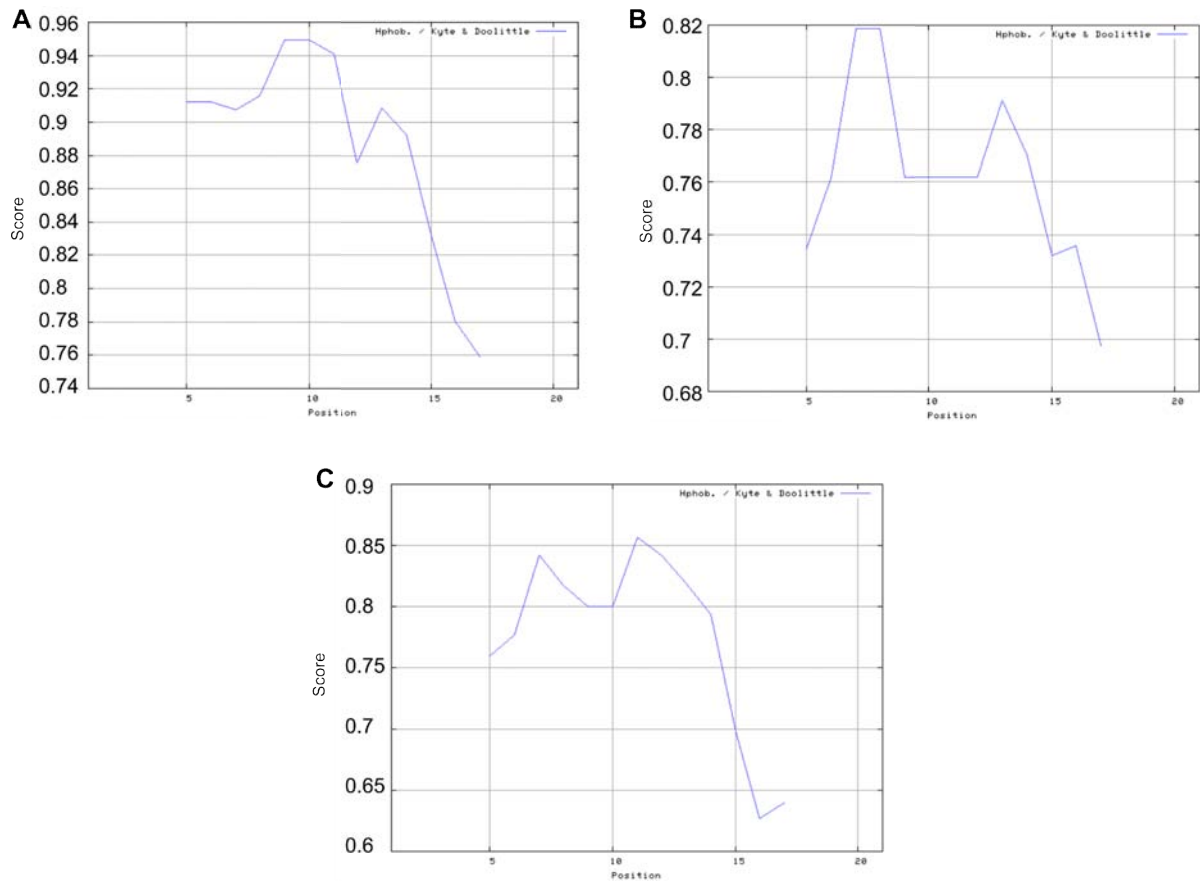
Figure 4.15: Hydrophobicity profile of TMDs (**A**) P47192 (VA722) (**B**) Q9SHJ6 (PMD2) and (**C**) Q9ZVL6 (U603). Hydrophobic profiles are calculated by ExPASy-Protscale (Kyte & Doolittle).

Total proteins were isolated from leaf samples of wild-type, mutant and wild-type chloroplast of *A. thaliana*. Proteins interacting with AtGet3d were pulled down by anti-AtGet3d antibody and protein A beads. The detailed experimental procedure is provided in Chapter 2.20. The specificity of antibody is tested with western blot (Figure: A5).

MS analysis of pull-down sample identified around 253 proteins that are interacting with AtGet3d (Table 4.4 and Appendix-A, Table A1 & Figure A4). These 253 proteins include both soluble and membrane proteins of which maximum number of proteins are localized to chloroplast. These soluble proteins that are interacting with AtGet3d can include different protein families ranging from heat shock proteins, actins, ubiquitins and ribosomal etc.

Similarly, Aquaporin, cytochromes, proteins in photosystems are the majority of those membrane proteins identified. Interestingly, all these identified proteins are reported to have

interactions with HSPs. This clearly implicates the presence of HSP domain could potentially coordinate these interactions of AtGet3d with these identified protein binding partners.

Out of 60 membrane proteins, 9 are TA proteins. Thylakoid lumen 18.3 kDa protein (U603), one of the TA proteins used in the co-expression studies, was detected in this pulled down sample analysis as well (Table 4.5). Besides U603, several other TA proteins that belong to chloroplast were also identified in this analysis as interacting with AtGet3d. This further confirms the specificity of AtGet3d that it could interact with chloroplast TA proteins for their targeting.

Table 4.4: Identified AtGet3d interacting proteins. Immunoprecipitation using anti-AtGet3Δd antibody identified 253 proteins belonging to different categories that can interact with AtGet3d. Out of 60 membrane proteins, 9 were TA proteins as shown in parenthesis.

Protein class	Total Proteins- 253	
	Soluble proteins 193	Membrane Proteins 60
HSP	4	---
Actin	4	---
ATPase	5	3
Ubiquitin	14	---
Ribosomal	12	---
Photosystem	9	13 (5- TA proteins)
ATP synthase	9	4
Peroxisomal	3	---
Chaperon	4	---
Cytochrome	---	6 (2- TA proteins)
Aquaporin	---	7
Other	129	27 (2- TA proteins)

Another interesting feature identified from this pull down analysis is the presence of polyubiquitin-12 (Q3E7K8 - encoded by At1g55060) among the identified AtGet3d interacting proteins. This At1g55060 gene is a conserved gene and gets over-expressed under abiotic stress conditions (Sanchita et al. 2014) compared to its basal expression levels. On the other hand, the interaction of ubiquitin and HSP is also well documented. Also, Arabidopsis gene At1g55060 is recently reported to encode the yeast homolog of Get5 (Srivastava et al. 2017). Although further detailed validations are needed, this result could aid in the hypothesis that the HSP domain of AtGet3d bypasses the interaction of Get4/Get5 and could potentially initiate the direct interaction with ubiquitin like domain of Get5.

Table 4.5: TA proteins interacting with AtGet3d pulled down by Co-IP

Uniprot ID	Subcellular location	Description
P60129	chloroplast thylakoid membrane	Photosystem II reaction center protein L (PSII-L)
P56779	chloroplast thylakoid membrane	Cytochrome b559 subunit alpha
P56780	chloroplast thylakoid membrane	Photosystem II reaction center protein H
Q6IDL4	Endoplasmic reticulum membrane	Transmembrane emp24 domain-containing protein p24delta3
Q9SHE8	chloroplast thylakoid membrane	Photosystem I reaction center subunit III, chloroplastic
Q9ZVL6	chloroplast thylakoid membrane	UPF0603 protein At1g54780, chloroplastic
Q9SUI6	chloroplast thylakoid membrane	Photosystem I reaction center subunit VI-2, chloroplastic
Q9SUI7	chloroplast thylakoid membrane	Photosystem I reaction center subunit VI-1, chloroplastic
P56771	chloroplast thylakoid membrane	Cytochrome f

4.3. Discussion

Out of the four Get3s present in *A. thaliana*, AtGet3d is functionally characterized in this chapter. Expression profile analysis shows that AtGet3d expresses in aerial tissues and the level of expression elevates during stress condition like salt, osmotic, heat and drought. Immunofluorescence and co-expression studies identified that AtGet3d gets localized to chloroplast surface and can bind with chloroplast TA proteins. Co-immunoprecipitation and LC-MS/MS analysis identified all the proteins that interact with AtGet3d through HSP domain. All the above observations form the basis for the hypothesis that AtGet3d can have dual functions as a TA translocating factor and as a chaperone.

Structural characterization of Get3 in plants

5.1. Introduction

In this Chapter 5, structural features of plant Get3 are explored using X-ray crystallographic techniques and homology modelling. There are already reported studies on Get3 from other organisms including yeast, archaea, etc that were characterized using X-ray crystallography. Since no plant Get3 structure is available, we have characterized the structure of AtGet3d from *A. thaliana*. Also, Homology modelling was employed for the structural studies of Get3 in *O. sativa* and *S. tuberosum*. The structure of AtGet3Δd was determined by molecular replacement method using All4481 protein from Nostoc sp. PCC 7120 (PDB: 3IGF) as a search model and refined at a resolution of 2.5Å. All the refined structures were compared with yeast Get3 structures.

5.2. Results

5.2.1. Protein purification, crystallization and data collection of AtGet3d

Initial attempts to obtain the purified full-length AtGet3d were not successful as the protein precipitated readily and degradation in the N-terminal region was observed. We were able to purify a truncated version AtGet3Δd, with 57 amino acids removed from the N-terminal. The construct was expressed in *E. coli* (BL21) DE3 followed by protein purification on Ni-affinity and size exclusion chromatography under reducing condition. AtGet3Δd was eluted as a homodimer. AtGet3d was purified to homogeneity as described in Chapter 2.11.

Crystallization trials of AtGet3Δd were carried out using commercial screens and custom-made screens with the sitting-drop method. Custom-made screens produced the AtGet3Δd crystals with different morphologies. Crystals of AtGet3d were reproduced by hanging-drop method with 50mM Sodium cacodylate (pH-5.47), 50mM lithium sulfate and 30% (w/v) PEG 4000. But these crystals failed to diffract (Figure 5.1a).

Since AtGet3d is an ATPase, further crystallization trials were performed in the presence of non-hydrolysable ATP analogue AMP-PNP and ADP. Crystals of 6xHis tagged AtGet3Δd were grown at room temperature by mixing an equal volume of protein solution containing 2 mM ADP/AMPPNP with reservoir solution containing 50 mM Sodium cacodylate (pH-5.47), 50 mM lithium sulfate and 30% (w/v) PEG 4000 (Figure 5.1b and c). In the presence of AMP-PNP, AtGet3Δd crystals diffracted up to 2Å. AtGet3d crystals belong to space group P2$_1$ with unit cell parameters of a= 59.25 Å, b=67.05 Å, c=99.40 Å, β=97.81°. Data collection

statistics for AtGet3Δd are given in Table 5.1.

a. Crystals of AtGet3Δd without nucleotide
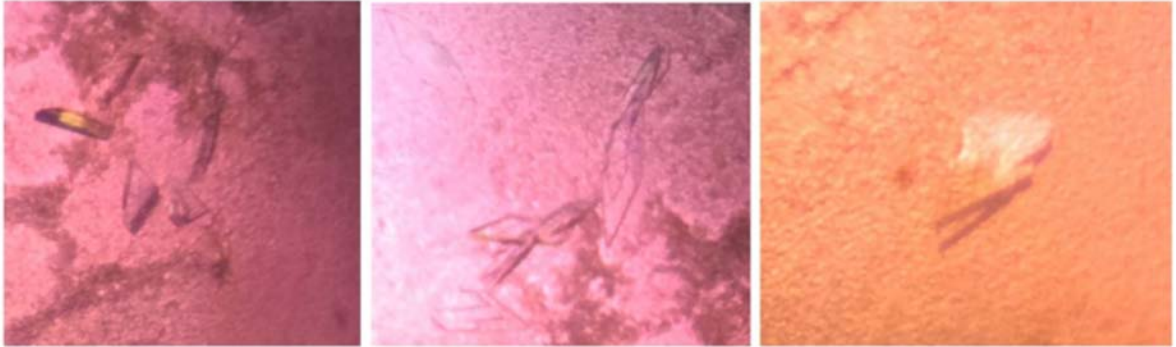
b. Crystals of AtGet3Δd with AMP-PNP

c. Crystals of AtGet3Δd with ADP

Figure 5.1: Crystals of AtGet3Δd grown at room temperature (a) without nucleotide (b) with AMP-PNP (c) with ADP.

5.2.2. Structure solution, refinement and validation

Data collection was done at ESRF, BM-14 beamline at 100K. The data sets were recorded with MARmosaic 225 CCD detector. The data were integrated with XDS and scaled with Aimless (CCP4 suite) in space group P 2_1(Kabsch 2010; Evans & Murshudov 2013; Evans 2006). 5% of the data was reserved for cross-validation.

Structure of Atget3Δd was determined to 2.5Å by molecular replacement with PHASER

using All4481 protein from Nostoc sp. PCC 7120 (PDB: 3IGF) as a search model (McCoy et al. 2007; Forouhar et al. n.d.). Iterative rounds of model building and refinement were carried out with COOT and CCP4/Refmac (Emsley et al. 2010; Winn et al. 2011; Murshudov et al. 2011).

Table 5.1: Data collection and refinement statistics of AtGet3Δd (PDB ID: 5YQK). Values of outer shell are shown in parenthesis.

Data Collection	
Beamline (ESRF)	BM-14
Wavelength (Å)	0.97625
Space group	P 2_1
Unit cell	a= 59.25 Å b=67.05 Å c=99.40 Å β=97.81°
Molecules per AU	Dimer
Solvent content (%)	42.75
Unique number of reflections	26818
Multiplicity	4.2 (4.2)
Completeness (%)	99.5 (100)
{I/σ(I)}	28.3 (6.0)
R_{sym} (%)	3 (20.5)
Refinement	
Resolution (Å)	58.70 - 2.50
Number of reflections	25502
R work / R free	22.24 / 25.84
Number of atoms	11524
Average B-factor (Å)	56.59
R.m.s. deviations	
Bond lengths (Å)	0.0141
Bond angles (°)	1.787
Ramachandran plot quality (%)	
Most favoured	90.5
Additionally allowed	7.4
Generously allowed	1.3

A single molecule of AtGet3Δd dimer was found in the asymmetric unit. The density of side-chain is generally weak in the α-helical subdomains, and density for residues 186-205, 320-328 and 399 were not observed in both subunit A and B. The refinement statistics are given in table 5.1. The crystal structure of AtGet3Δd was obtained in the presence of ADP, as crystals of AtGet3Δd failed to diffract in the absence of nucleotides. The structure reveals that AtGet3Δd is a symmetric homodimer. Even though nucleotides are missing, Mg^{2+} ions were observed in the P-loop regions. The final refined structure encompassed 368 residues out of 398 residues and refined to an R factor 0.22 and Free-R factor 0.26. Refined structure was deposited in PDB with PDB ID: 5YQK.

5.2.3. Overall Structure of Chloroplast AtGet3Δd

To comprehend the unique structural features of chloroplast Get3 of *A.thaliana,* the crystal structure of AtGet3Δd was determined through X-ray crystallography, using Nostoc Get3 (PDB ID: 3IGF) as the structural template. The structure of AtGet3Δd monomer contains 11 α-helices and 15 β-strands (Figure 5.2). Like the previously reported Get3 structures from *S. cerevisiae* and *C. thermophilum*, AtGet3Δd also comprises the core nucleotide binding domain (NBD) and an α-helical subdomain. In addition, the AtGet3Δd also has an extra α-crystallin domain at C-terminal. Analysis showed that this feature is unique to a paralog of chloroplast Get3.

The structure of AtGet3Δd was homologous to Get3 (All4481) from Nostoc sp. PCC 7120, with RMSD 1.97 Å. Individual subunits of AtGet3Δd are identical to those from 3IGF (Nostoc Get3) structures (Figure 5.5c and Appendix A-Table A2). The crystal structure is similar to 3IGF with conserved HSP domain at C-terminal. Length of the loop connecting α5 and β5 was found to be little longer with 12 extra amino acids in the case of AtGet3Δd compared to 3IGF.

The NBD was made up of Rossmann fold and like in G-type hydrolases, it is formed by eight β-strands and is parallel to each other except β3 which is anti-parallel to each other (Rao & Rossmann 1973; Sprang 1997). Even though ADP and AMP-PNP were added in the crystallization condition, density was not observed in the refined structure.

The crystal structure of ScGet3 is a dimer and contains Zn molecule. It has been established that Get3 is an obligate dimer and exhibits functionally distinct conformation of this dimer (Mateja et al. 2009). This Zn-Cystine motif serves as a hinge for the switch between multiple differentially open and closed conformation. Structural comparison of Atget3Δd with Get3

from other organisms indicated that it is comparable to the closed form of ScGet3 in binding groove orientation and α-helical subdomain (Appendix A- Table A3).
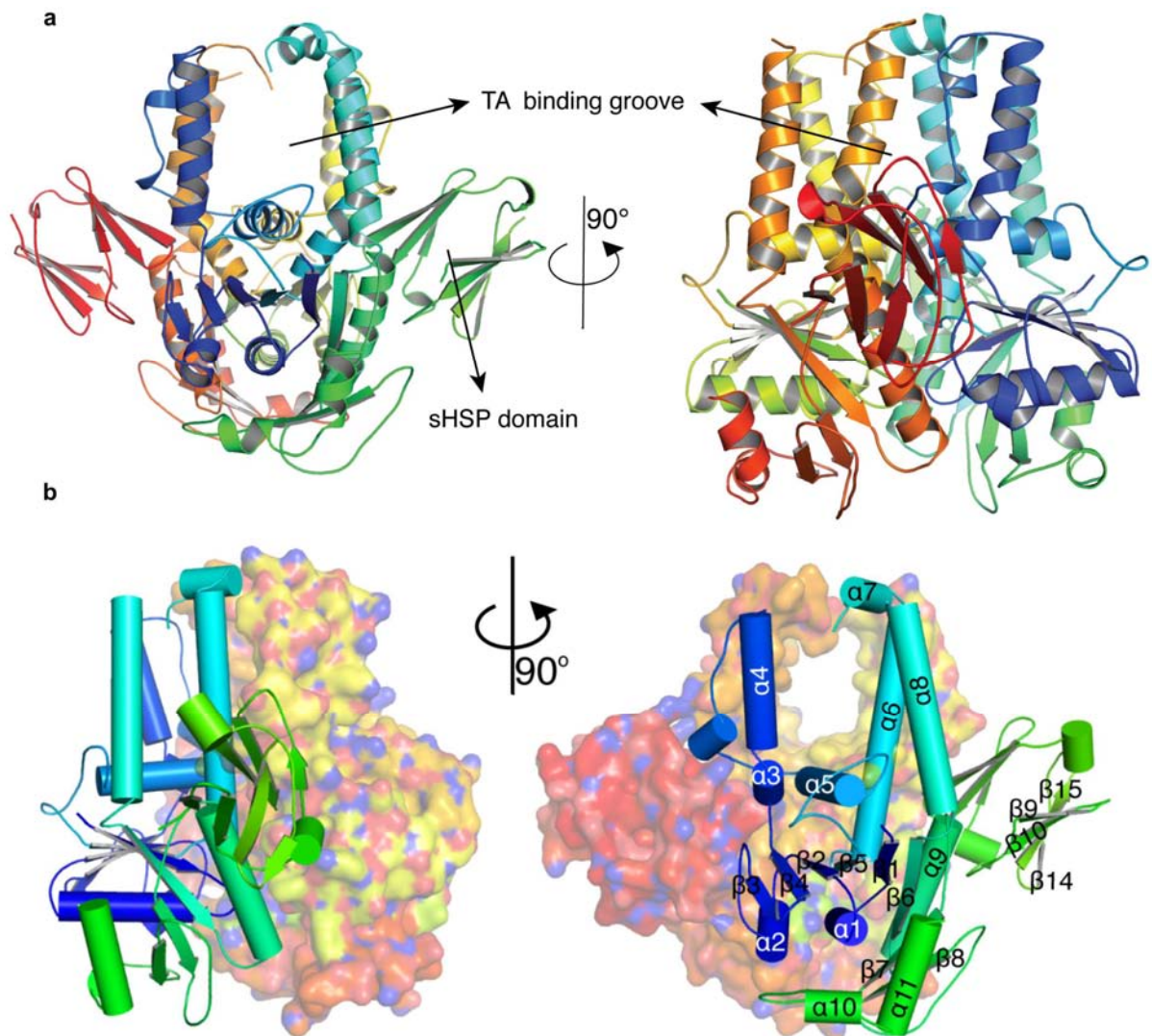
Figure 5.2: Overall structure of AtGet3Δd (a) cartoon representation (b) one monomer in molecular surface view.

5.2.4. Structural comparison of AtGet3d with AtGet3b

The AtGet3Δd structure exhibited differences in many structural features compared to a model of AtGet3b, with an RMSD of 4.7 Å. The P-loop and CXXC motifs are present in AtGet3b model as like in ScGet3, while the AtGet3d lacks CXXC motif and have variation in P-loop (Figure 5.3).
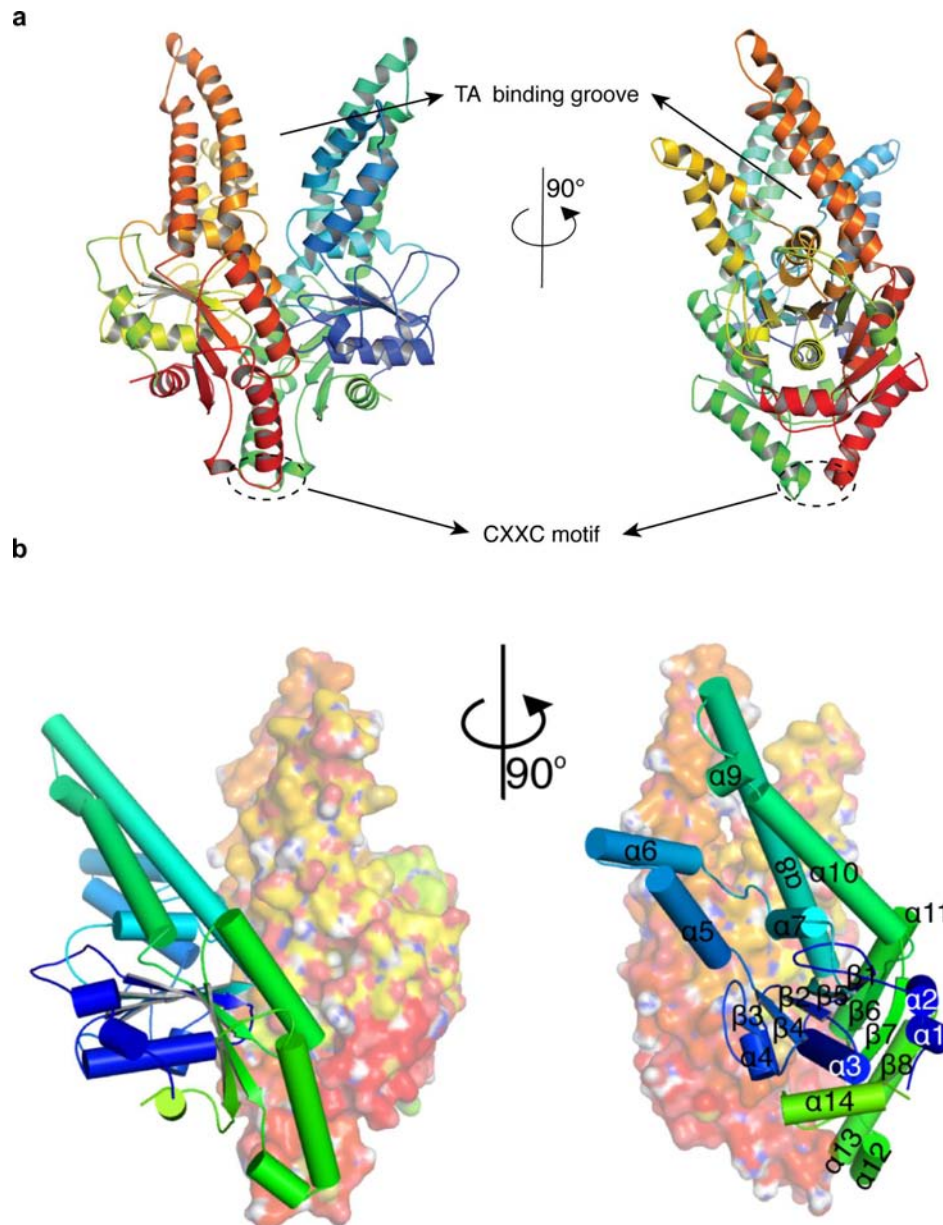
Figure 5.3: Model of AtGet3b. CXXC motif is absent in AtGet3Δd while present in AtGet3b.
(a) cartoon representation (b) one monomer in molecular surface view.

Major differences were observed at the TM binding groove. Helices that form the side of the groove are more deviated from the centre of the groove, compared to Atget3Δd. Also the R.M.S.D of main chain atoms of AtGet3b at the TM binding groove shows the highest variation with AtGet3d (Figure 5.5a and Figure A2). Similar to ScGet3, AtGet3b is also devoid of the HSP domain at C-terminal, making it likely that the substrate preference and targeting mechanism for AtGet3d and AtGet3b might be different. Surface conservation analysis reveals that HSP domain of AtGet3d is more conserved while AtGet3b found to have more conservation at dimer interface (Figure 5.4).
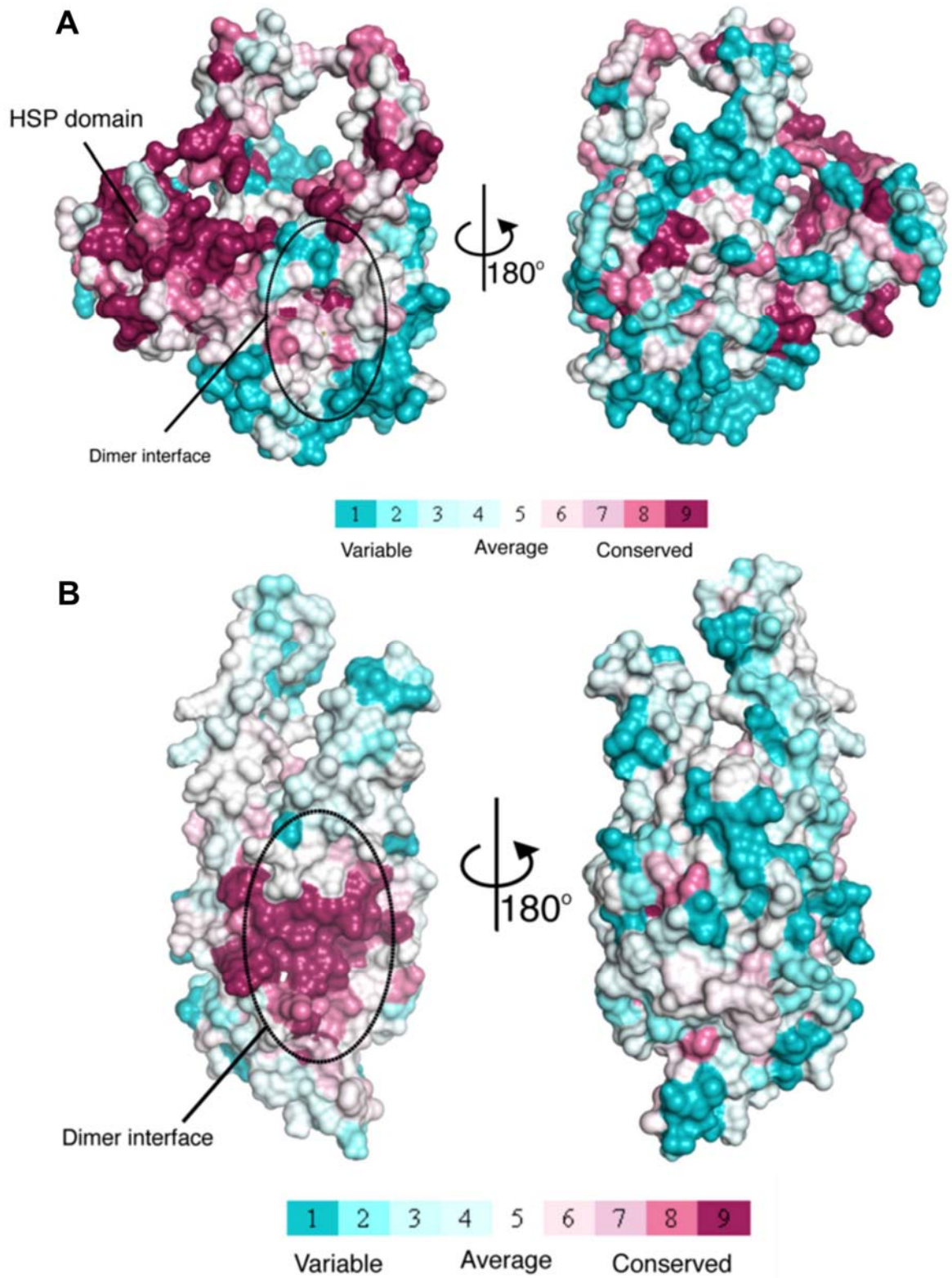
Figure 5.4: Surface conservation analysis of (**A**) AtGet3Δd and (**B**) AtGet3b monomers using consurf server
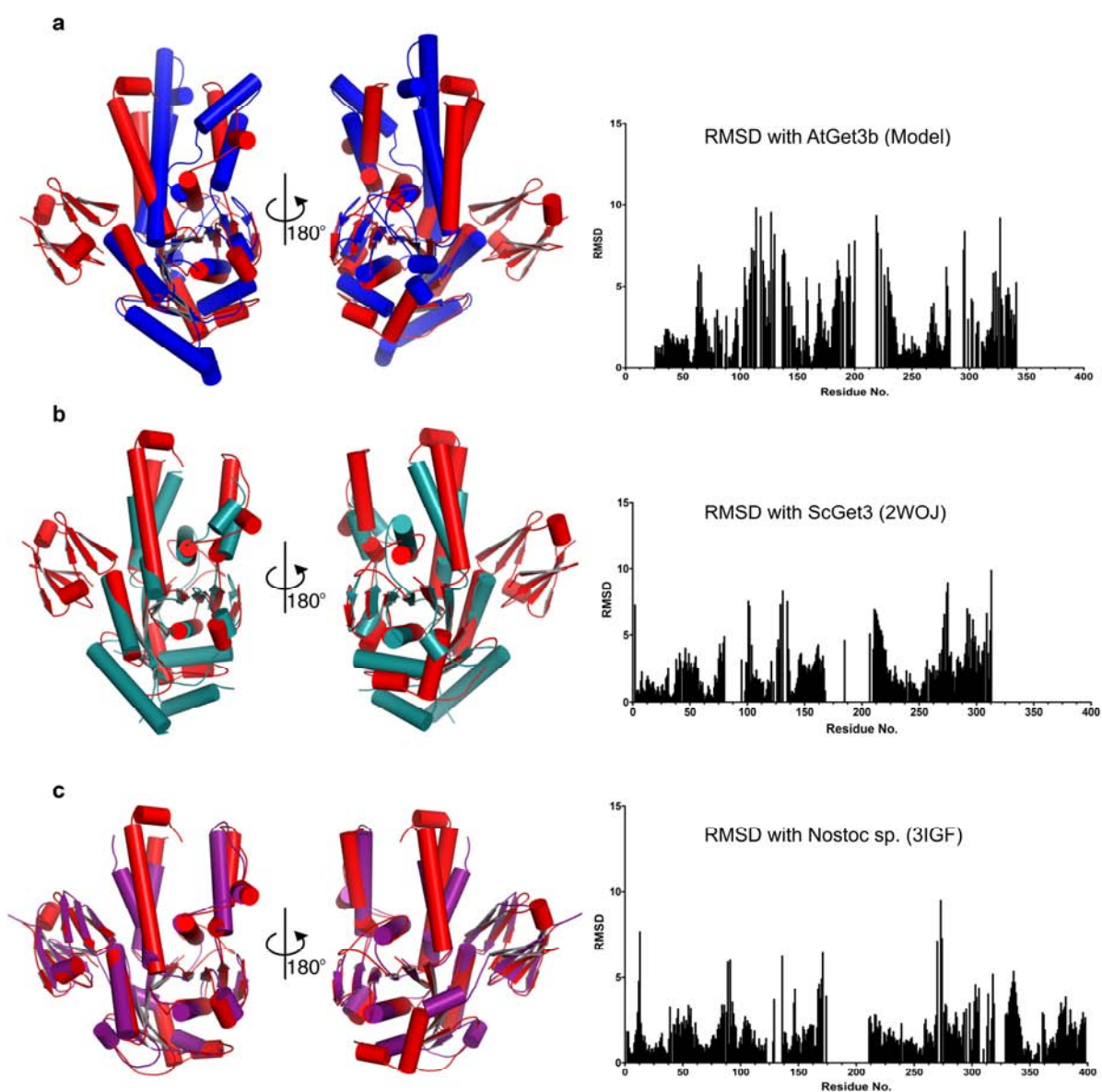
Figure 5.5: Structural comparison of AtGet3Δd (red) with (a) AtGet3b (blue) (b) ScGet3 (deep teal) and (c) Nostoc Get3 (purple). RMSD analysis shows that more variations in TA binding groove region. Gaps in the RMSD graph indicate missing residues. Nostoc Get3 has HSP domain matched with AtGet3Δd's HSP domain.

5.2.5. AtGe3d Crystallized in Closed State

The closed conformation of Get3 dimer has a larger interacting surface compared to its open conformation. The interface area in AtGet3Δd closed conformation (~3400 Å^2) is greatly larger than in ScGet3 (~2400 Å^2). The open form of ScGet3 has a loop at Switch-II region, which is converted to helix in closed form. AtGet3Δd also has a helix in the region that corresponds to switch-II (Figure 5.7). These parameters assert that the refined crystal structure corresponds to the closed form of AtGet3Δd, despite the missing electron density for nucleotides at the P-loop region. Another striking feature of AtGet3d is the CXXC motif,

which is absent also in AtGet3a but present in AtGet3b. The change in open to closed conformation in ScGet3 is coordinated by Zn ion found at the dimer interface. Hence it is apparent that the mechanistic basis of open to close conformation change in AtGet3Δd is still not well understood. The RMSD of AtGet3Δd with closed form and open form of ScGet3 were found to be 3.158 Å and 3.597 Å respectively. (Figure 5.6)
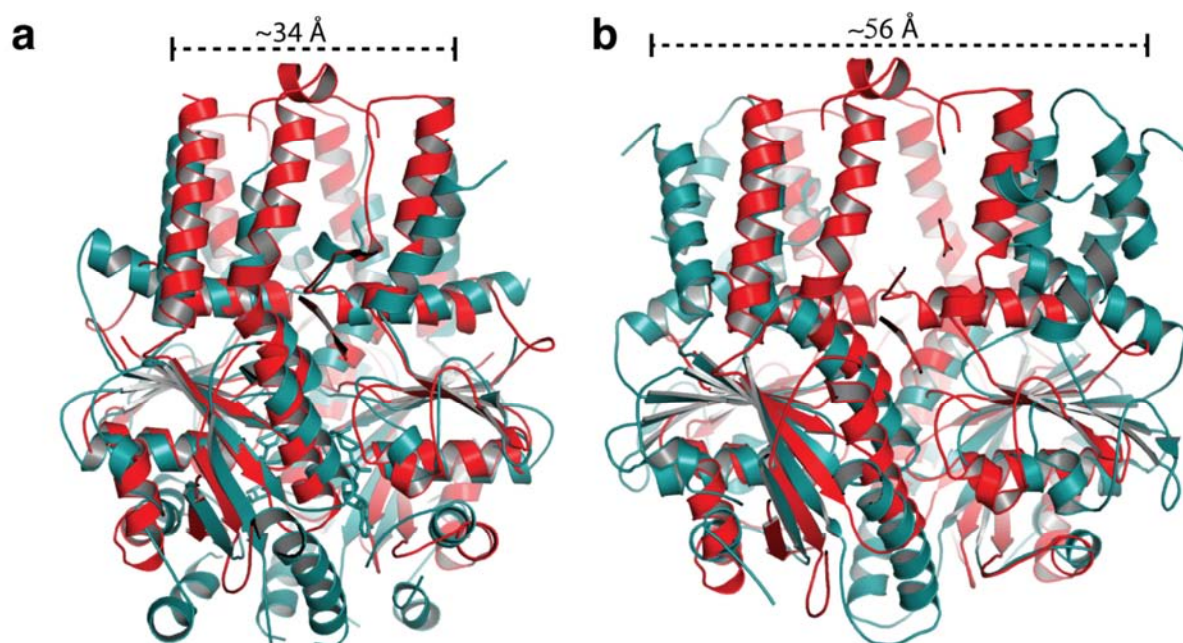
Figure 5.6: AtGet3Δd superposed with **a)** closed ScGet3 **b)** Open ScGet3. RMSD of AtGet3Δd with closed form and open form of ScGet3 were found to be 3.158 Å and 3.597 Å respectively.
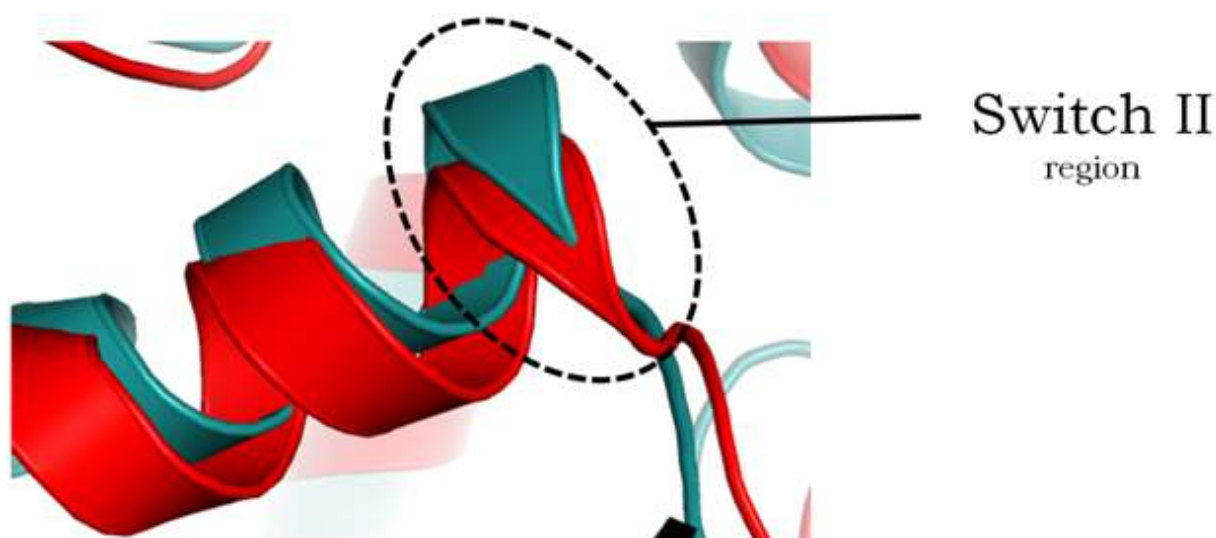
Figure 5.7: Switch II motif region of AtGet3Δd (red) compared with the closed form of ScGet3 (deep teal). In both cases, helix is observed in switch II motif region, while loop was present in the open form of ScGet3**.**

AtGet3d showed significant similarity to the closed form of ScGet3 with several differences. However, major differences were observed at the nucleotide binding region and Zn binding region. In the case of ScGet3, α-helix 11 containing CXXC motif is found to be present and the same is absent in AtGet3Δd. The change in open to closed conformation in ScGet3 is coordinated by Zn ion found at the dimer interface, whereas Zn doesn't appear in the AtGet3Δd structure. Hence it is apparent that the mechanistic basis of open to close conformation change in AtGet3Δd is still not well understood. Get3 functions as homodimer as seen in many cases, although tetramer has also been reported in an archaeal homolog, and tetramerisation upon target binding has been observed (Suloway et al. 2012).

5.2.6. TM Binding Groove

The TM binding groove of AtGet3Δd is comprised of six amphipathic helices from α-helical sub-domain or finger domain. Two crossing helices (α5) form the bottom of the groove, while the side is formed by three extended helices α4, α6 and α8 (Figure 5.8a and Figure 5.8b). Electron density at the TM binding groove and the surface were very weak implying that these regions are highly flexible and dynamic.

The dimensions of the hydrophobic groove (~32 Å x20 Åx~20 Å) appear sufficient to bind the hydrophobic TM region of ~21 amino acids. The overall alignment of these loops is highly complicated but the hydrophobic residues and glycines are usually conserved. The leucine-rich groove is arranged in a manner that the hydrophobic amino acids face towards the groove and hydrophilic amino acids like Arg, Lys and Gln are exposed to the surface (Figure 5.8c). Also the hydrophobicity distribution pattern in the binding groove of ScGet3, AtGet3d and AtGet3b were not similar suggesting the different substrate specificity of these Get3s (Figure 5.9).
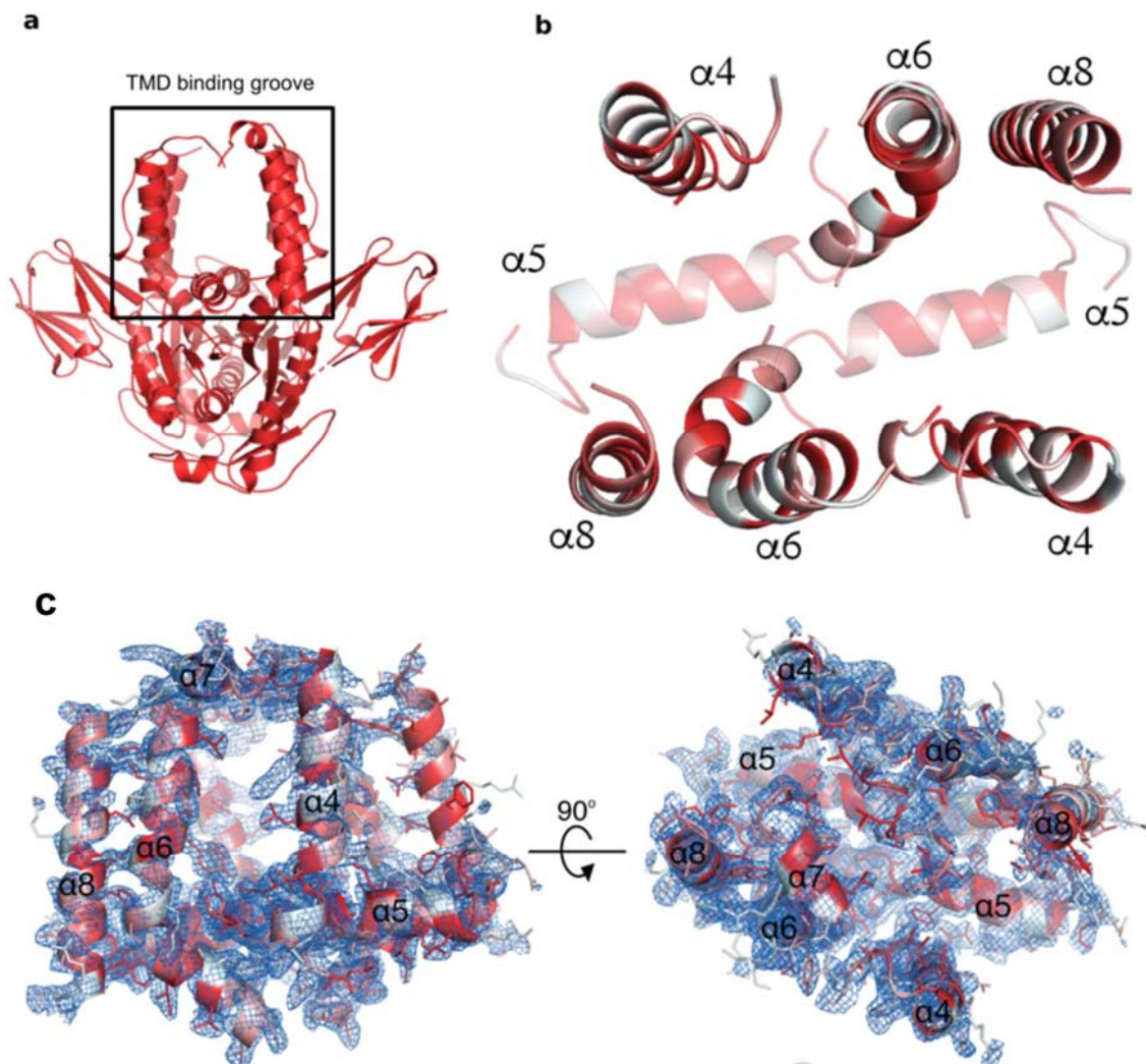
Figure 5.8: TM binding groove of AtGet3Δd. (a) overall structure of AtGet3d in which TM binding groove is highlighted (b) TM binding groove with helixes are labeled. (c) Electron density map (2Fo-Fc) of TMD binding groove at 1σ. Weak density observed in several regions because of its highly dynamic nature.
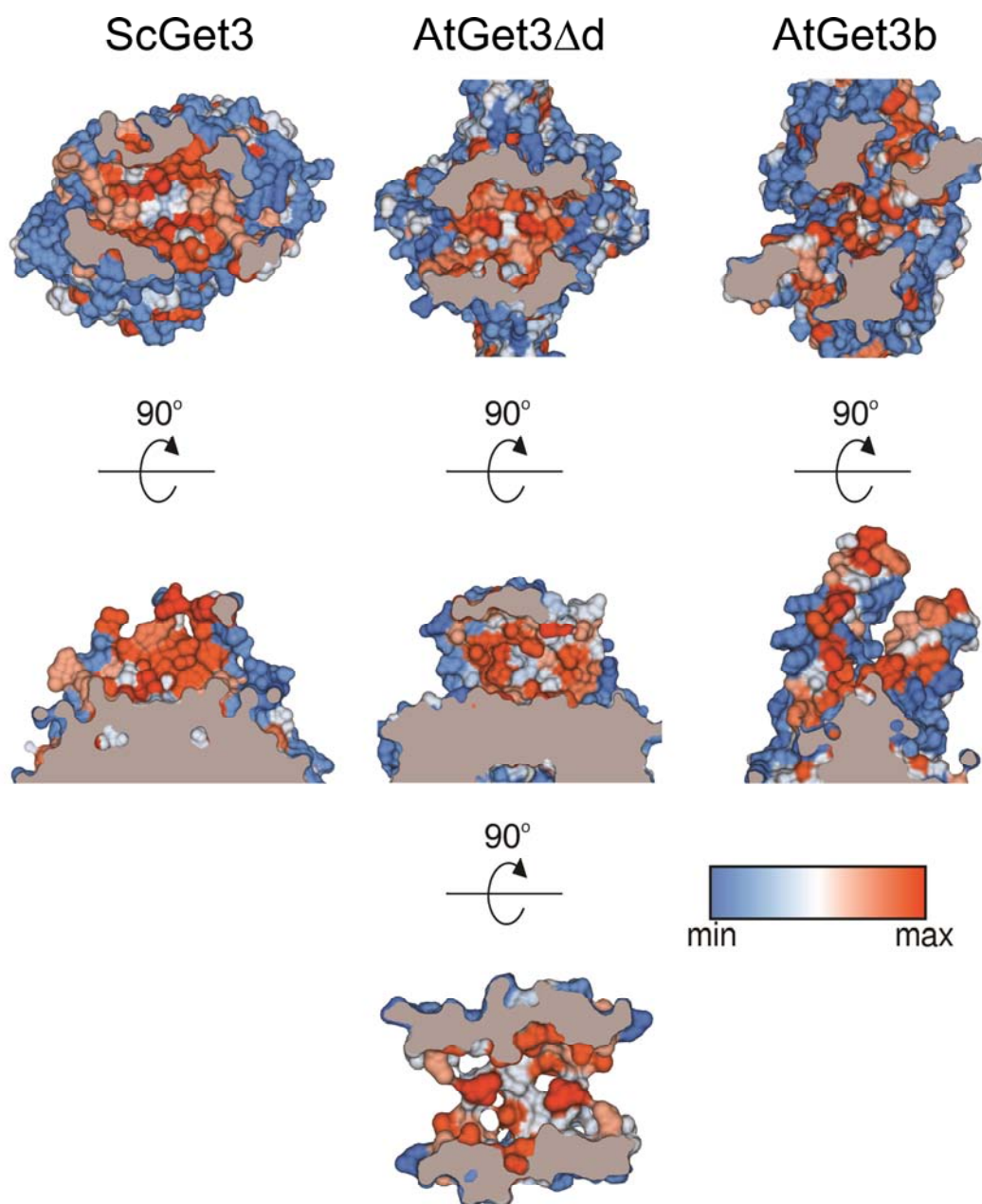
Figure 5.9: Distribution of hydrophobic residues in TMD binding groove of ScGet3 (PDB Id: 2WOJ), AtGet3Δd and AtGet3b model. Images were generated in chimera with Kd hydrophobicity scale.

5.2.7. P-Loop of AtGet3d

Phosphate-binding loop or Walker-A motif is a common motif present in ATP and GTP binding proteins. The sequence of P-loop has a signature pattern of G-X(4)-GK-[TS], where X is any amino acid. In the case of AtGet3d, P-loop consists of "GKGGSGKT". As shown in the previous Chapter 3, there is a single amino acid variation (val to ser) in the non-conserved amino acid region of P-loop. Even though nucleotides were added during crystallization, the

density for the same was not observed, while Mg^{2+} ions were present in the P-loop region (Figure 5.10a and Figure 5.10b).

As discussed in Chapter 4, the ATPase activity of AtGet3d was slow compared to AtGet3a. In ScGet3, Val in the P-loop forms two hydrogen bonds with Gln273 and Asn272. On the other hand, Ser in the P-loop in AtGet3Δd is connected by two hydrogen bonds with Lys 10 and one hydrogen bond with Val 236. This extra hydrogen bond may bring rigidity to P-loop movement and in turn results in slow ATPase activity (Figure 5.10c and Figure 5.10d).
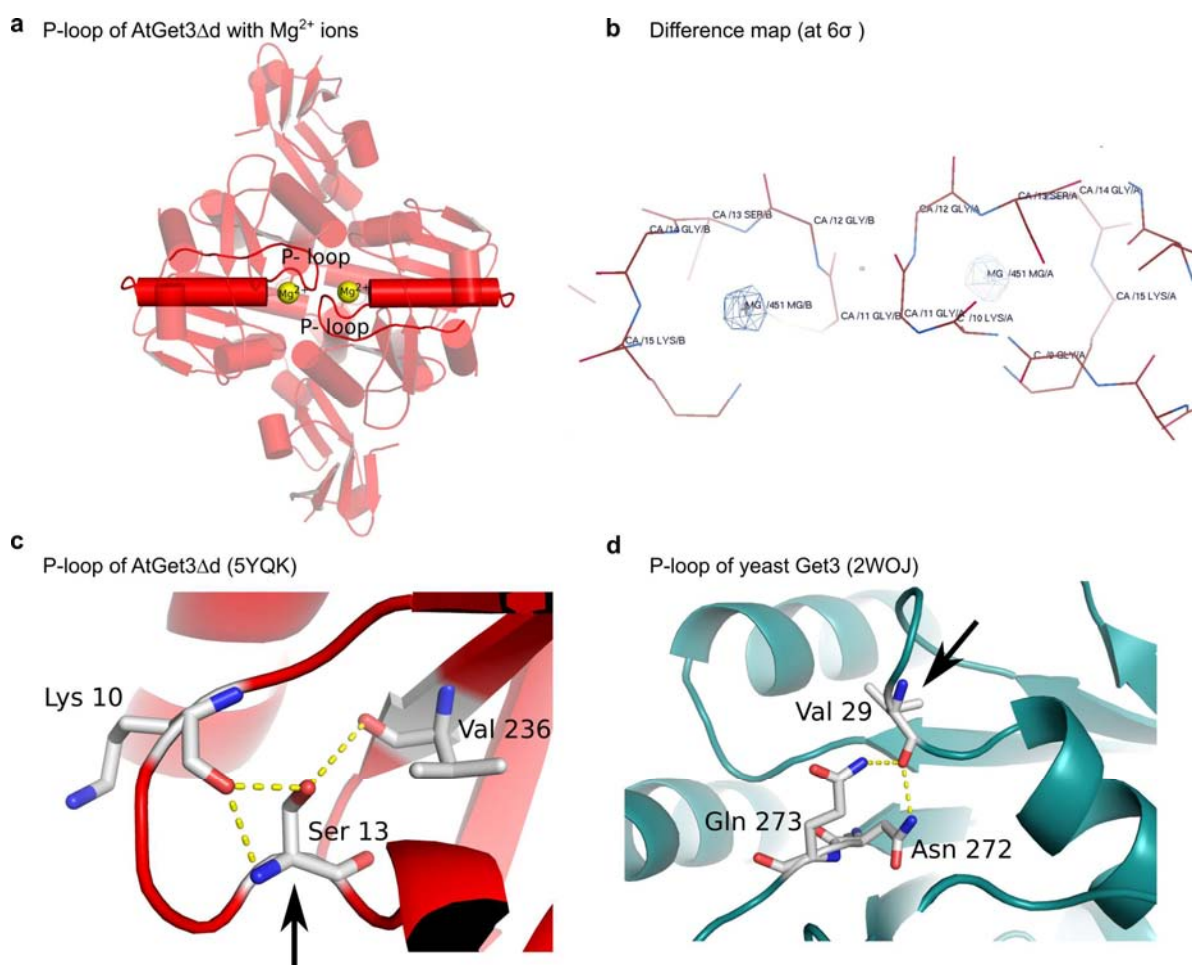
Figure 5.10: P-loop of AtGet3Δd. (a) Each monomer of AtGet3Δd has one Mg^{2+} ion at P-loop region. (b) Electron density map (2Fo-Fc) of P-loop at 6σ. Densities for Mg^{2+} ions are clearly visible. (c) P-loop Ser13 of AtGet3Δd can form three hydrogen bonds with Lys10 and Val 236. (d) P-loop Val 29 of ScGet3 can form two hydrogen bonds with Gln273 and Asn272. The difference in hydrogen bond pattern may be the reason for slow ATPase activity of AtGet3Δd.

ADP molecule was successfully docked into the nucleotide binding site by following the method section 2.26. The docking score and glide scores were -13.952 and -14.044

respectively. A 5 ns molecular dynamics simulation of Atget3Δd structure docked with ADP exhibited good ligand stability in the binding site and forms 9 hydrogen bonds with the amino acids. The bond formation with P-loop residues and Mg^{2+} ion are identical in case of Atget3Δd and ScGet3. Mg^{2+} ion also interacts with phosphate group of ADP and Asp41 and Thr16. P-loop residues Gly12, Gly14 and Lys15 make strong hydrogen bonds with phosphate residue of ADP (Figure 5.11).
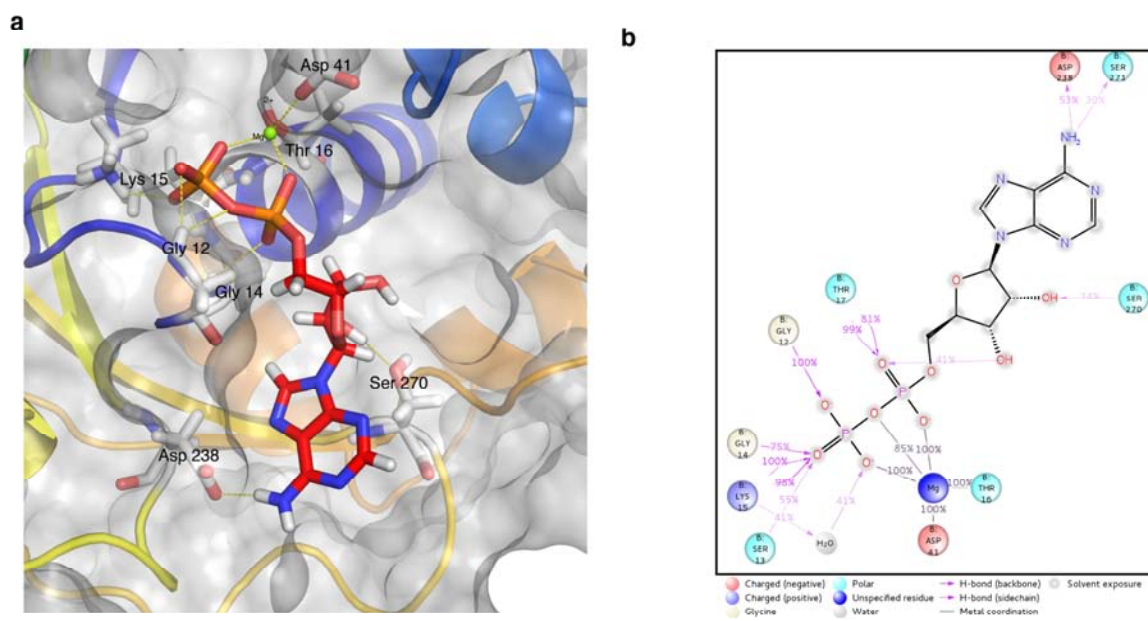
Figure 5.11: Docking of ADP at P-loop of AtGet3Δd (a) ADP molecule is docked into nucleotide binding region of AtGet3Δd monomer. Nucleotide interacting residues are marked. (b) Protein-ligand interaction diagram of ADP docked AtGet3Δd.

5.2.8. HSP domain of AtGet3Δd

A distinctive feature of AtGet3Δd is the presence of HSP domain at the C-terminal (Figure 5.2 and Figure 5.12). HSP domain of AtGet3Δd consists of anti-parallel 7 stranded β-sheets. DALI (Holm & Rosenström 2010) analysis showed that the C-terminal domain is an HSP20 α-crystallin domain (Figure 5.13 and Table 5.2). This unique domain combination is not observed in any other Get3 structure so far reported. The α-crystallin domain of HSP20 consists of several β-sheets and is located towards the C-terminal part. HSP domain of AtGet3Δd is located in the region where other GET pathway components like Get4-Get5 and Get1-Get2 interact with Get3 in the case of yeast (Figure 5.14). We speculate that HSP domain may bypass interactions meant for cytosolic Get3 or it may need significant conformational change to uncover the interacting surface of other components. Loop regions where HSP interact with other proteins are well conserved in HSP domain of AtGet3d (Figure

5.12b). These imply AtGet3d may have an alternate mechanism for targeting TA proteins to the chloroplast. Electron density for the loop that connects HSP domain of AtGet3Δd is missing.
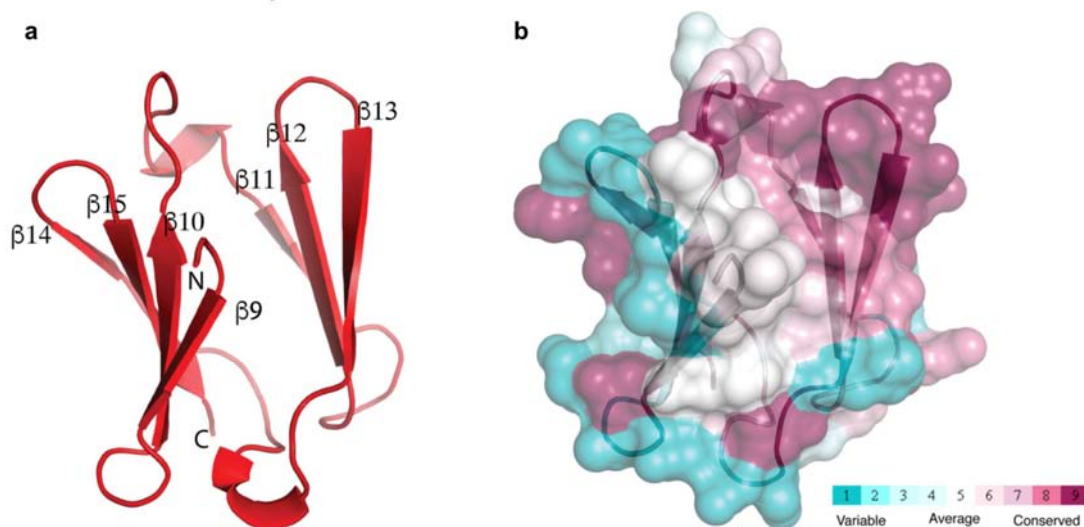
Figure 5.12: (a) HSP domain of AtGet3Δd consists of anti-parallel 7 stranded β-sheets. (b) Consurf analysis shows high conservation in loop regions.

Small heat shock proteins such as HSP20 prevent protein aggregation during stress conditions, especially in temperature stress. HSP20 can form both low molecular and high molecular multimeric complexes. HSP20 can be phosphorylated by several protein kinases, which influences their interactions with other proteins (Gusev et al. 2005). BAG3 is a co-chaperone which can link sHSPs to HSP70 (Rauch et al. 2017). HSP20 can interact with actin in a phosphorylation-dependent manner (Tessier et al. 2003). Several Pfam entries show that HSP20 is fused with other domains such as ARID/BRIGHT DNA binding domain, PCAF, protein kinases etc.

Table 5.2: Top 5 hits from DALI server for HSP domain of AtGet3Δd

PDB ID	Z score	RMSD	LALI score	% Identity	PDB Description
3AAC(A)	8.6	2.0	65	29	Small heat shock protein hsp14.0 with the mutations
4RZK(B)	8.4	2.1	68	22	Small heat shock protein hsp20 family
4YLB(D)	8.3	2.2	66	23	Heat shock protein Hsp20
5DS2(F)	8.7	2.3	68	19	18.1 kDa class I heat shock protein
1GME(B)	7.3	2.7	68	15	Heat shock protein 16.9b

*Parenthesis represents chain ID
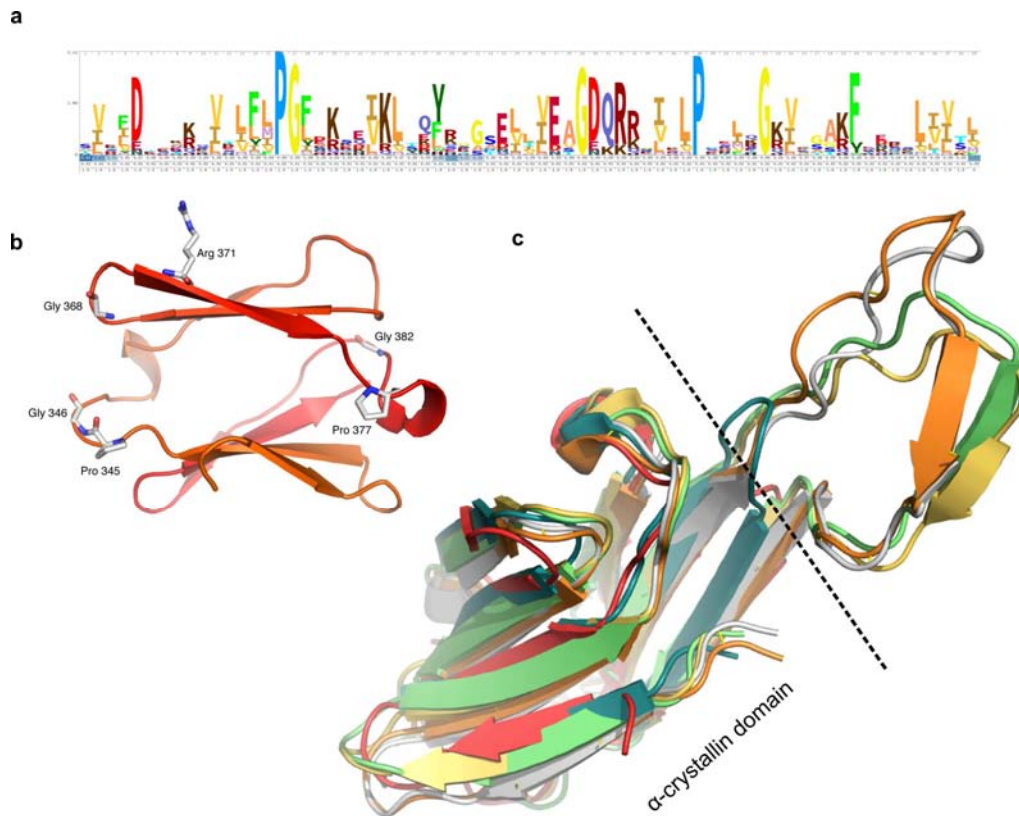
Figure 5.13: Conservation of HSP domain (a) Web logo sequence alignment from DALI server (b) HSP domain of AtGet3Δd with conserved residues are marked.(c) Structural alignment of HSP domain of AtGet3Δd (red) with other sHSPs found in DALI server. 1GME (gray), 4YLB (green), 4RZK (deep teal), 5DS2 (orange) and 3AAC (yellow).
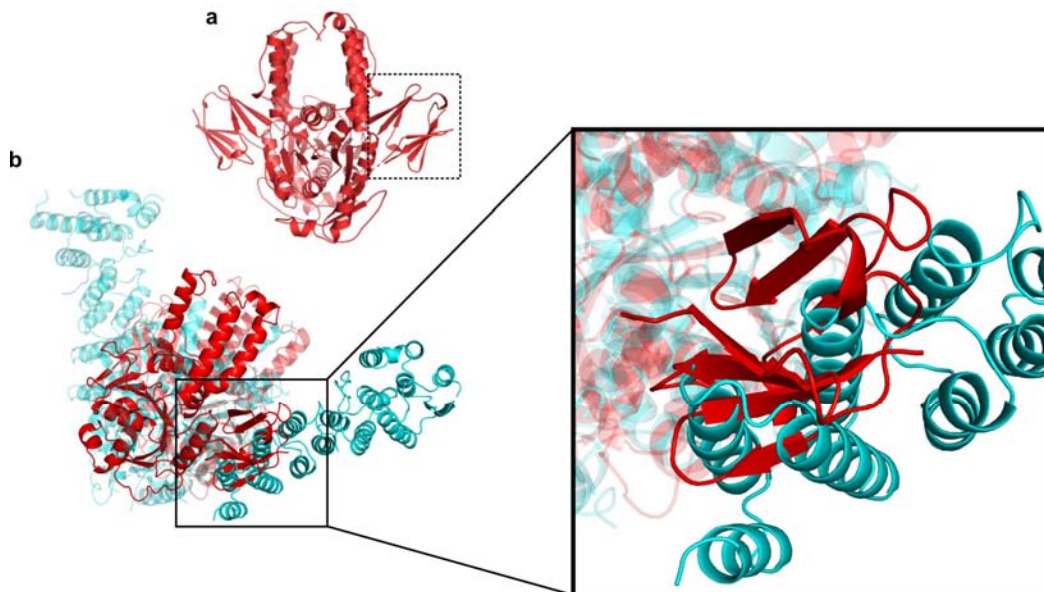
Figure 5.14: (a) overall structure of AtGet3Δd in which HSP domain is highlighted (b) AtGet3Δd (red) is superposed with ScGet3-Get4-Get5 complex (teal). HSP domain of AtGet3Δd is showed in zoom. HSP domain is positioned in the place where Get3-Get4 interaction occurs. Clash in this region indicates that AtGet3Δd may not interact with Get4 or HSP domain has to move/change confirmation to allow Get4 to bind with AtGet3Δd.

The modelled open form of AtGet3Δd (Figure 5.15a) indicates that HSP domain may come close together to function as the chaperon. Centroid distance between two HSP domain of AtGet3Δd showed ~12Å movement as compared to close form, while the centroid of loop present between β12and β13 aligned face to face showing movement of ~15Å (Figure 5.15b).
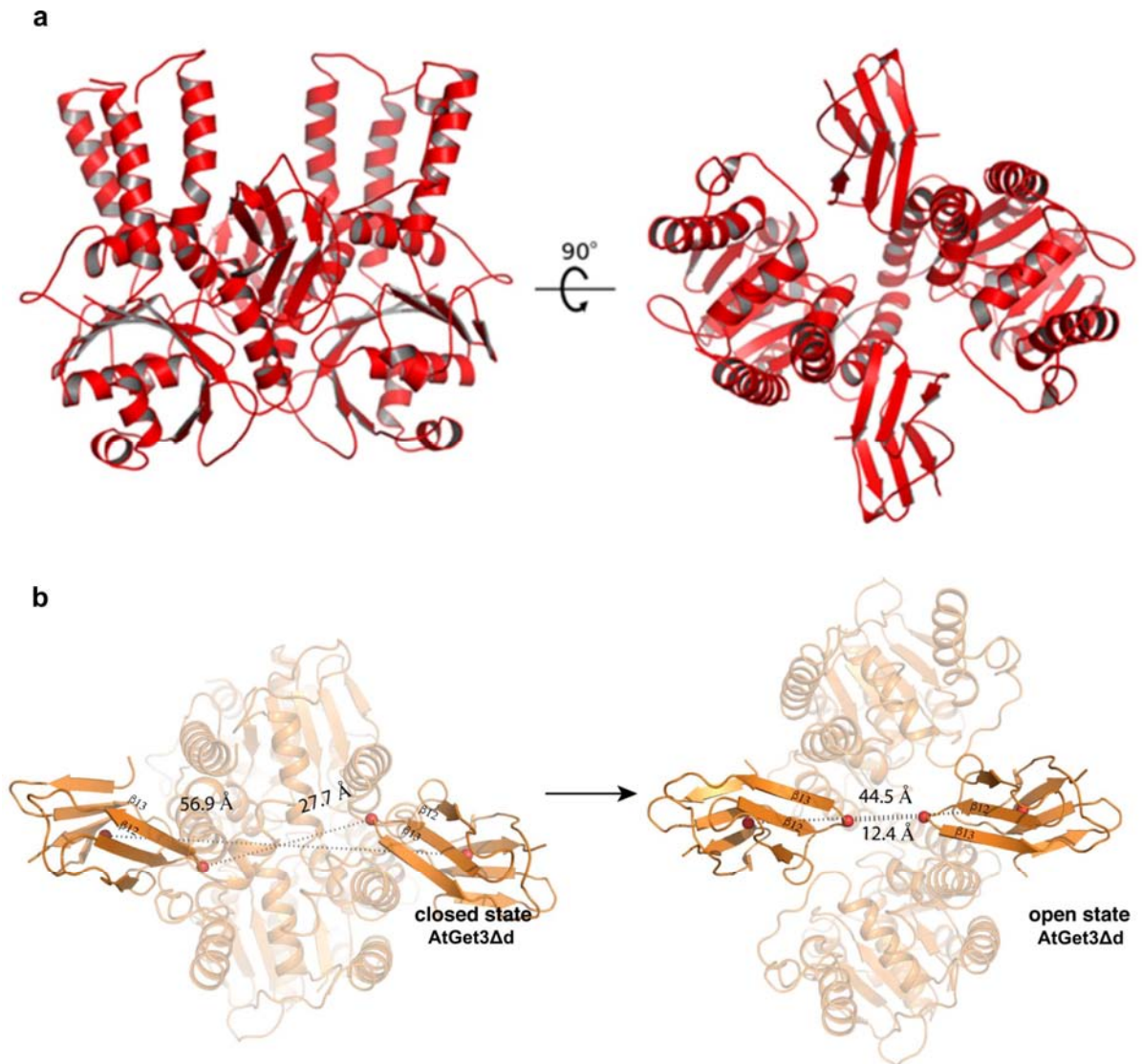
Figure 5.15: (a) Model of Open form of AtGet3d (b) Centroid distances between HSP and between loop (β12/ β13) of closed and open state of AtGet3Δd. In open state, HSP domains move closer.

5.2.9. Docking of TMD in the hydrophobic groove

Using previously reported ScGet3-Pep12 complex structure (PDB ID: 4XTR)(Mateja et al. 2015) was used as reference for docking of Pep12 into hydrophobic groove of AtGet3Δd. The TM domain of TA protein Pep12 was docked into the hydrophobic groove of AtGet3Δd without any steric hindrance (Figure 5.16). Pep12 has a 21 amino acid transmembrane

domain. The TMD was stabilized in the groove during 5 ns simulation. Hydrophobic interaction was observed to be stabilizing the TMD in the binding groove (Figure 5.17). This confirms that AtGet3d can accommodate TMD of TA proteins in the binding groove in the closed state.
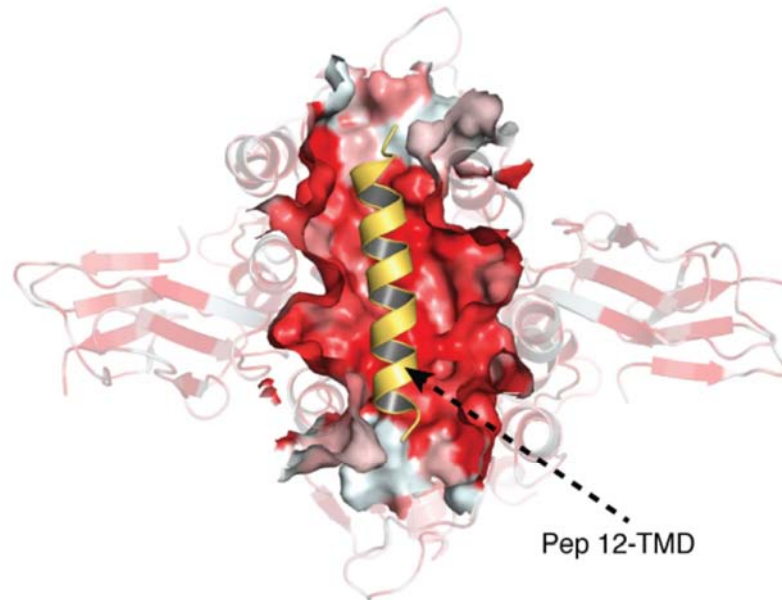
Figure 5.16: TMD of Pep12 docked in the groove of AtGet3Δd.
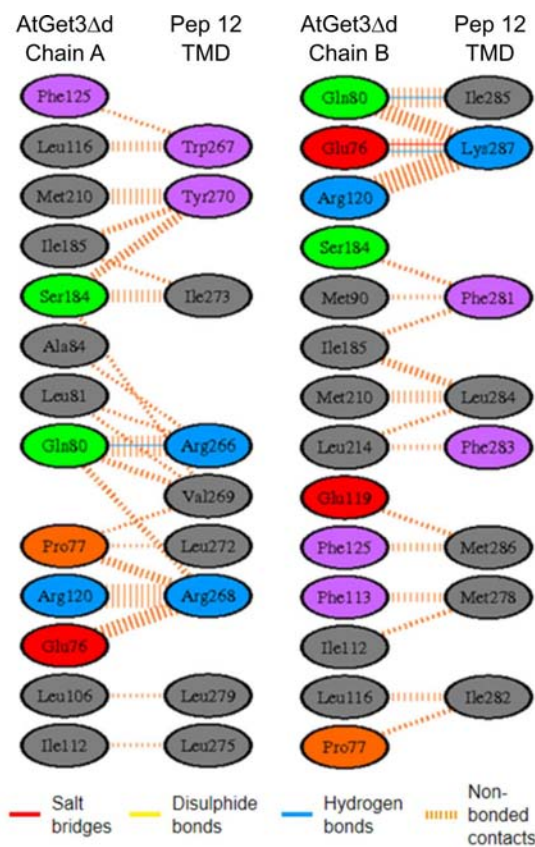
Figure 5.17: Interactions of Pep12 TMD with the AtGet3d groove residues.

5.2.10. Structural analysis of GET pathway members in *O. sativa* and *S. tuberosum*.

To extend the understanding of structural features of GET pathway in *O. sativa* and *S. tuberosum*, we have modelled the structure of cytoplasmic Get3, Get1 and Get4 from both *O. sativa* and *S. tuberosum* (Figure 5.18). *O. sativa* has single cytoplasmic Get3 while *S. tuberosum* has three cytoplasmic Get3s. Yeast Get3 (PDB ID: 2WOO) without any bound molecule (open form) was used as a template for all the cytoplasmic Get3s. The structure of *O. sativa* cytoplasmic Get3 is much similar to Yeast Get3 except at the C-terminal. In *S. tuberosum*, out of three cytoplasmic Get3s, M1AND2 and M0ZFY4 are similar to Yeast Get3. But M1A9X9 has a loop between β3 and α4 that is absent in Yeast and other Get3s of *S. tuberosum*. Also, M1A9X9 has more amino acids compared to M1AND2 and M0ZFY4. All *S. tuberosum* cytoplasmic Get3 models have a disordered loop at C-terminal due to lack of the corresponding residue in Yeast Get3. Cytosolic domain of Get1 was modelled for both *O. sativa* and *S. tuberosum*. In this analysis, Yeast Get1 structure (PDB ID: 3ZS8) was used as template for model generation. Both signal sequence and trans-membrane domain were omitted for generating models. RMSD is found much higher for *O. sativa* Get1 with Yeast Get1. *O. sativa* Get1 and *S. tuberosum* Get1 have RMSDs of 3 and 0.47 respectively with Yeast Get1. Four helices were observed in the overall model of *S. tuberosum*, in comparison with Yeast Get1 having two alpha helices with lowest RMSD. Also in *S. tuberosum* Get1, an additional α-turn is observed between α2 and α3. In *O. sativa* Get1 model, two α-helices were observed and a loop is present between two helices.

In addition, Get4 of both *O. sativa* and *S. tuberosum* were modelled (Figure 5.18). *Chaetomium thermophilum* Get4 (PDB ID: 3LPZ) and human TRC 35 (PDB ID: 6AU8) were used as templates for modelling *O. sativa* Get4 and *S. tuberosum* Get4 respectively. These modelled structures were compared to Yeast Get4. In Yeast Get4, two β sheets are present between α11 and α12 but these are not present in both *O. sativa* and *S. tuberosum* Get4. In place of β- sheets, a loop is observed in both where loop length is small in *S. tuberosum* compared to *O. sativa*. The loop connecting between α5 and α6 is also longer in *S. tuberosum*. RMSD (7.2 and 2.6 for *O. sativa* and *S. tuberosum* respectively) in comparison with Yeast Get4 is estimated in the analysis. By using such structure modelling analyses, this study provides insight into the deeper understanding of GET pathway mechanisms.
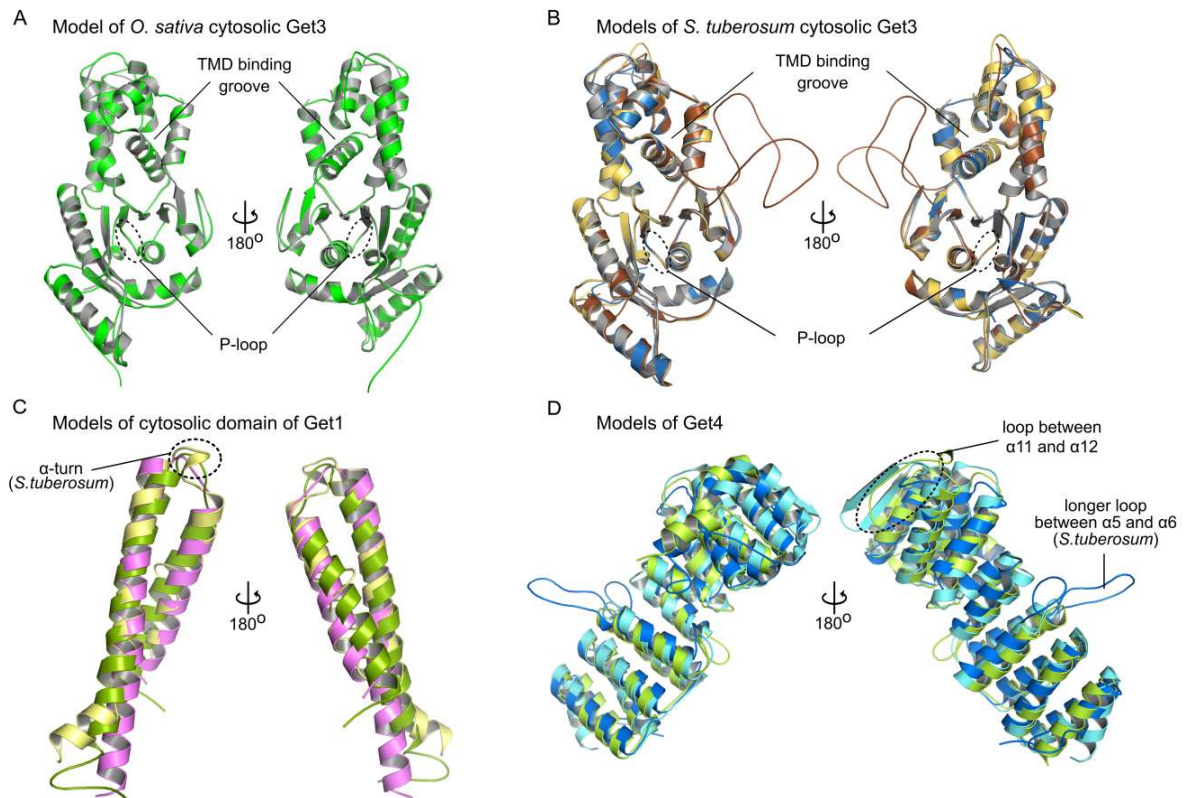
Figure 5.18: Structural models of identified GET pathway members of *O. sativa* and *S. tuberosum*. (A) model of *O. sativa* cytosolic Get3(green) superposed with Yeast Get3 (gray). (B)Models of *S. tuberosum* cytosolic Get3s M1AND2 (sky blue), M1A9X9 (brown) and M0ZFY4 (yellow orange) superposed with Yeast Get3 (gray). (C) models of *O. sativa* Get1(split pea) and *S. tuberosum* Get1 (pale yellow) superposed with Yeast Get1 (violet). (D) models of *O. sativa* Get4 (lemon) and *S. tuberosum* Get4 (marine) superposed with Yeast Get4 (aquamarine).

5.2.11 Models of chloroplast Get3 of *O. sativa* and *S. tuberosum*

Sequence analysis of both *O. sativa* and *S. tuberosum* in Chapter 3 showed that they have chloroplast Get3 that is similar to *A. thaliana* Get3d. Also, Get3 which is homologous to AtGet3b is absent. This motivates to model the structure of chloroplast Get3 from both *O. sativa* and *S. tuberosum*. Structural coordinates of AtGet3d (PDB ID: 5YQK, structure from this study) and All4481 protein from Nostoc sp. PCC 7120 (PDB ID: 3IGF) were used as templates for the model generation. AtGet3d has more than 40% sequence similarity with both *O. sativa* and *S. tuberosum* chloroplast Get3. Models of chloroplast Get3 of both *O. sativa* and *S. tuberosum* are shown in the figure 5.19. Structure of *O. sativa* chloroplast Get3 is identical to the AtGet3d structure with RMSD of 0.687. While comparing with All4481 of Nostoc, *O. sativa* chloroplast Get3 has more structural deviation at HSP region with overall

R.M.S.D. of 2.028. Like AtGet3d, P-loop is present in *O. sativa* chloroplast Get3. Also CXXC motif is absent.

Unlike *O. sativa* chloroplast Get3, *S. tuberosum* chloroplast Get3 doesn't have a P-loop at nucleotide binding region. *S. tuberosum* chloroplast Get3 has R.M.S.D. of 0.786 and 2.246 with AtGet3d and Nostoc respectively. Since P-loop is absent in *S. tuberosum* chloroplast Get3, the nucleotide-dependent conformational changes may be absent in this Get3 or may have any other mechanism for TA protein targeting. It is also possible that any one of the cytoplasmic Get3 (out of three) may function as targeting factor for chloroplast TA protein in *S. tuberosum*.
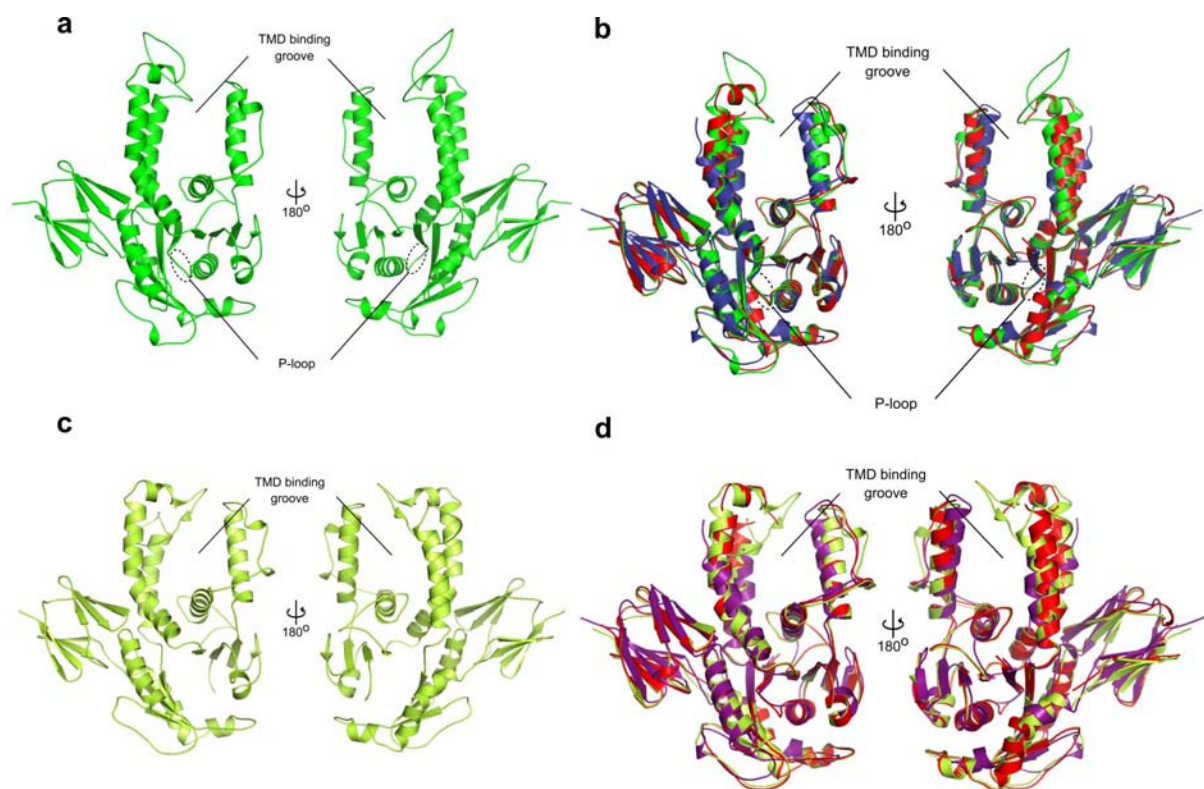
Figure 5.19: Models of Chloroplast Get3 of *O. sativa* and *S. tuberosum*. (a) Model of chloroplast Get3 of *O. sativa* (b) model of *O. sativa* cholorplast Get3 (green) superposed with AtGet3d (red) and Nostoc (deep purple). (c) Model of chloroplast Get3 of *S. tuberosum* (d) model of *S. tuberosum* cholorplast Get3 (lemon) superposed with AtGet3d (red) and Nostoc(deep purple).

5.3. Discussion

Structure of Get3 from different organisms including fungi and archaea were investigated by several independent studies (Mateja et al. 2009; Suloway et al. 2009; Hu et al. 2009; Yamagata et al. 2010; Bozkurt et al. 2009; Sherrill et al. 2011; Suloway et al. 2012). But

structural information of Get3 from higher organisms are yet lacking. In this study, Get3 from plant system were analyzed both structurally and functionally. Structure of chloroplast Get3 (AtGet3Δd) from *A. thaliana* was elucidated using X-ray crystallographic techniques. This is the first structural report of Get3 from plant system. The structural analysis indicates that AtGet3d has an α-crystalline domain (sHSP domain) at C-terminal that coordinates the interaction with other proteins. This study first time identified the domain fusion event in Get3 with HSP domain. The structural analysis, in this Chapter 5, has well-coordinated with the functional analysis in Chapter 4 including nucleotide binding, TA protein binding and interaction of other stress-related protein with AtGet3d. The predicted conformational changes of AtGet3d point outs that during stress condition AtGet3d can act as chaperone. Also, crop plants analyzed in this study, have Get3 that are structurally homologous to AtGet3d with HSP domain at C-terminal. The presence of Get3 that are similar to AtGet3d in more than 80% sequenced plants especially in crop plants highlights the importance of Get3d over Get3b.

Chapter 6

Summary and Conclusion

Out of the several known types of membrane proteins, TA proteins attracts a special attention because of their distinct topology and functional importance. Since the location of TA protein is determined by the C-terminal TM region, the co-translational targeting pathway is less effective for their targeting. TA protein targeting is mainly carried out by post-translational chaperonin-assisted mechanisms. GET pathway, one of thewell-studied pathways in yeast, is responsible for the targeting of TA proteins to ER membrane. Among the other components of GET pathway, Get3 plays an important role by connecting the pre-targeting and post-targeting complexes. Get3 homologues proteins are present in all the forms of life ranging from archaea to mammals. This doctoral research work dealt with exploring the GET pathway of TA protein targeting with special implications to Get3 characterization in plants.

6.1. Summary

Although several computational and experimental studies are available to identify TA protein targeting pathways in different organisms including the plant model of *A. thaliana*, there were no detailed studies on plants. In this study, we have identified and delineated the TA protein distribution and their GET components through *in silico* analyses in two crop plants, *Oryza sativa* subsp. Indica and *Solanum tuberosum* in Chapter 3. Beside these analyses, this study, from Chapter 4 to 6, also explored the functional and structural characteristics of chloroplastic Get3 from *Arabidopsis thaliana*.

These selected crop plant species (*O. sativa* subsp. Indica and *S. tuberosum*) have more than 35,000 total proteins but the TA protein content of these plants accounts for less than 2%. Form our analysis in Chapter 3, it is relevant that *O. sativa* and *S. tuberosum* have 508 and 912 TA proteins respectively. Also, the previous studies have shown that *A. thaliana* has 520 TA proteins. TA proteins are involved in several vital cellular processes like vesicular transport, vesicle fusion etc. Also, TA proteins are present in almost all organelles of the cell. The presence of extra organelles in plants compared to other phyla, like chloroplast, makes the TA protein targeting more complex than other systems.

The average length of TMD of TA proteins is found to be 21 amino acids. But the TMD length shows greater variations depending on the respective organelle membrane. Hydrophobicity pattern, amino acid composition and post-translational modification are predicted to be influenced by the organelle specificity of TA proteins. In plants, more than one

Get3s are present that are specific to different organelles due to their complex nature of different membranes of organelles and their respective TA proteins. The number of Get3s also differs across the plant species studied. *O. sativa* has three Get3s, while *A. thaliana* and *S. tuberosum* have four and five Get3s respectively as predicted by our analysis. Also in *S. tuberosum*, three out of five predicted Get3s are specific for ER TA protein targeting. In these three Get3, 85% identify is observed between M0ZFY4 and M1AND2 as specified in Chapter 3. Modelling studies reveled that M1A9X9 is different compared to both M0ZFY4 and M1AND2, having a loop between β3 and α4. Both *O. sativa* and *S. tuberosum* have chloroplast Get3 which have more than 40% similarity with *A. thaliana* AtGet3d.

The percentage existence of TA proteins in ER is more or less the same for both *O. sativa* and *S. tuberosum*, yet the number of Get3s differs across them. This needs further experimental validation especially in case of *S. tuberosum* where three Get3s are predicted for ER. Besides these Get3s, both selected crop plants have Get3s specific for mitochondria and chloroplast as well. Only one Get1 and Get4 are identified in both the plant species. Other components (Get2, Get5, Sgt2, Ydj1) that are known in Yeast, are not present in these crop plants. Further studies are essential to validate the functioning of GET pathway in the absence of these unpredicted components in *O. sativa* and *S. tuberosum*. This *in silico* analysis in Chapter 3 highlights the use of such predictive analyses in identifying the existence of TA proteins in plants and therefore forms the basis for further experimental characterization of GET mediated TA protein targeting mechanisms.

A. thaliana has four Get3s specific for three different organelles. Sequence and phylogenetic analysis showed that Get3 belongs to clade-d that is diverged from the rest of the Get3. Also, Get3s in clade-d have a domain fusion event with HSP domain at C-terminal. AtGet3a is responsible for targeting TA proteins designed to ER. Similarly, AtGet3d and AtGet3d belong to the chloroplast. AtGet3c target TA proteins to mitochondria. The sequence analysis shows that ~80% of sequenced plant species have AtGet3d. Also, AtGet3d has a unique domain combination of α-crystalline domain at C-terminal. Immuno-localization, immuno-blot and LC-MS/MS studies in Chapter 4 show that AtGet3d is localized to the surface of the chloroplast. Also, co-expression and pulldown studies indicated that AtGet3d can effectively bind with chloroplast TA proteins. But the ATP hydrolyzing power of AtGet3d was found relatively less compared to AtGet3a in the ATPase assay in Chapter 4. This can bedue to the extra hydrogen bonds that Ser13 makes between Lys10 and Val 236. The knockout seeds of AtGet3d germinated very slowly without undergoing lethality compared to wild-type

highlighting that the functionality of AtGet3d can be replaced by some other proteins.

In Chapter 5, AtGet3Δd was crystallized in closed conformation. Eventhough nucleotides were added during the crystallization trials, the density for the nucleotides was missing in the data. Only Mg^{2+} ions were observed in the P-loop region. Docking and simulation identified that AtGet3d can accommodate nucleotide at the P-loop region and TMD of TA proteins at TMD binding groove. One unique feature of AtGet3d is the presence of HSP domain at the C-terminal region. Expression of AtGet3d was found higher in aerial tissues during stress conditions. The presence of HSP domain with Get3 makes AtGet3d compatible for dual purpose. Immuno-precipitation followed by LC-MS/MS analysis showed that AtGet3d interacts with different protein families including actin, HSPs, ubiquitins, photosystem subunits, ATP synthase, aquaporins etc. These are the protein families reported to have interaction with HSP/α-crystalline domain. It is also evident from the model of the open form of AtGet3d that HSP domain can move closer to form a large hydrophobic cavity. This cavity/binding site can accommodate large miss-folded proteins during stress conditions with the help of other interacting proteins. In summary, the dual fuction of AtGet3d has been identified: (i) during normal physiological conditions, AtGet3d can bind with chloroplast TA protein for targeting and (ii) during stress condition, AtGet3d can act as chaperonin with the help of HSP domain.

6.2. Conclusions

In Eukaryotes, the efficient and precise insertion of membrane proteins is an imperative step for their accurate function in various organelles. Any targeting error may lead to mislocalization of these proteins with unfavourable cellular effects. GET pathway plays a critical role in TA protein targeting. In yeast and lower eukaryotes, the GET pathway includes five Get Proteins (Get1-5) and co-chaperone that contains tetratricopeptide repeat-containing protein 2 (Sgt2). However, it appears that plants exhibit considerable variation in the components of the pathway. The most striking difference is the presence of multiple paralogs for Get3 compared to other phyla of life, and they are with distinct domain combinations. The chloroplast-targeting assay with purified full-length AtGet3d, truncated AtGet3Δd and AtGet3d without HSP domain failed to localize, indicating that it might require some additional cytoplasmic factors such as SGTA or Get4 for chloroplast targeting. Yeast Get3 mutants displayed conditional lethality, with the disruption in Get3 function causing the mislocalization of TA proteins into the mitochondrial membrane (Schuldiner et al. 2008).

Moreover, in mutants, the TA proteins may use the alternate pathways for insertion into the ER. On the other hand, the Get3 mutant leads to embryonic lethality in mammals, indicating a divergent role of GET components in different organisms (Mukhopadhyay et al. 2006).

Plants are continuously challenged by a diverse variety of biotic and abiotic stress conditions on account of their immobile nature, to cope with which they have evolved countermeasures including sensors and response mechanisms (Suzuki et al. 2014). The foremost indispensable task during stress involves maintaining protein homeostasis by keeping them in functional native conformation. In plants, a wide range of HSPs function as molecular chaperones in the quality control of membrane proteins from different organelles. Generally, they form a stable complex with the unfolding client protein and protect the proteins from irreversible aggregation till stress conditions persist. AtGet3d is one such specialized molecular chaperone that is responsible for upholding protein homeostasis in addition to correct organelle targeting and membrane protein insertion. The C-terminal of Get3d encompasses the α-crystallin domain of small HSP (sHSP) specifically very similar to HSP20. HSP20 is generally known to preserve the denatured proteins in a folding state and proceed with consecutive ATP-dependent disaggregation through HSP70/90 chaperon system (Liberek et al. 2008). However, further investigation is needed to shed light on the significance of fusion of sHSP20 and TA-protein binding domains. It is worthy to mention that there are many sHSPs (P31170, P13853 and P19037) in Arabidopsis.

sHSPs are involved in a wide variety of cellular functions including but not limited to response to diverse types of stress, modulation of cytoskeleton, cell growth, differentiation and signal transduction. MS analysis in Chapter 4 revealed the interaction of Get3d HSP domain with actin, HSPs, ubiquitins, photosystem subunits, ATP synthase, aquaporins etc. Several sHSPs have been reported to interact with actin and intermediate filament to maintain the integrity of cells (Mounier & Arrigo 2002). Oligomerisation and phosphorylation are two imperative properties of sHSPs closely related with this function. The MAP kinase cascade is responsible for phosphorylation of sHSPs especially at several serine residues (Guay et al. 1997; Brophy et al. 1999). Actin-HSPs interaction needs further study to comprehend the overall relationship of this system. Many reports have demonstrated that α-crystallin binds and prevents the aggregation of multispan transmembrane proteins such as Aquaporin (Swamy-Mruthinti et al. 2013). So it may also be involved in preventing the aggregation of TA proteins that are targeted to indispensable chloroplast proteins such as photosynthetic components and cytochrome b.

The overall function of AtGet3d (Figure 6.1) explains its role in normal growth condition and under stress. It is also speculated that the chaperone function may get activated in response to a different stress condition when compared to Get3a. Due to the absence of the CXXC motif, AtGet3d might be insensitive to the oxidative stress and might activate during other form of
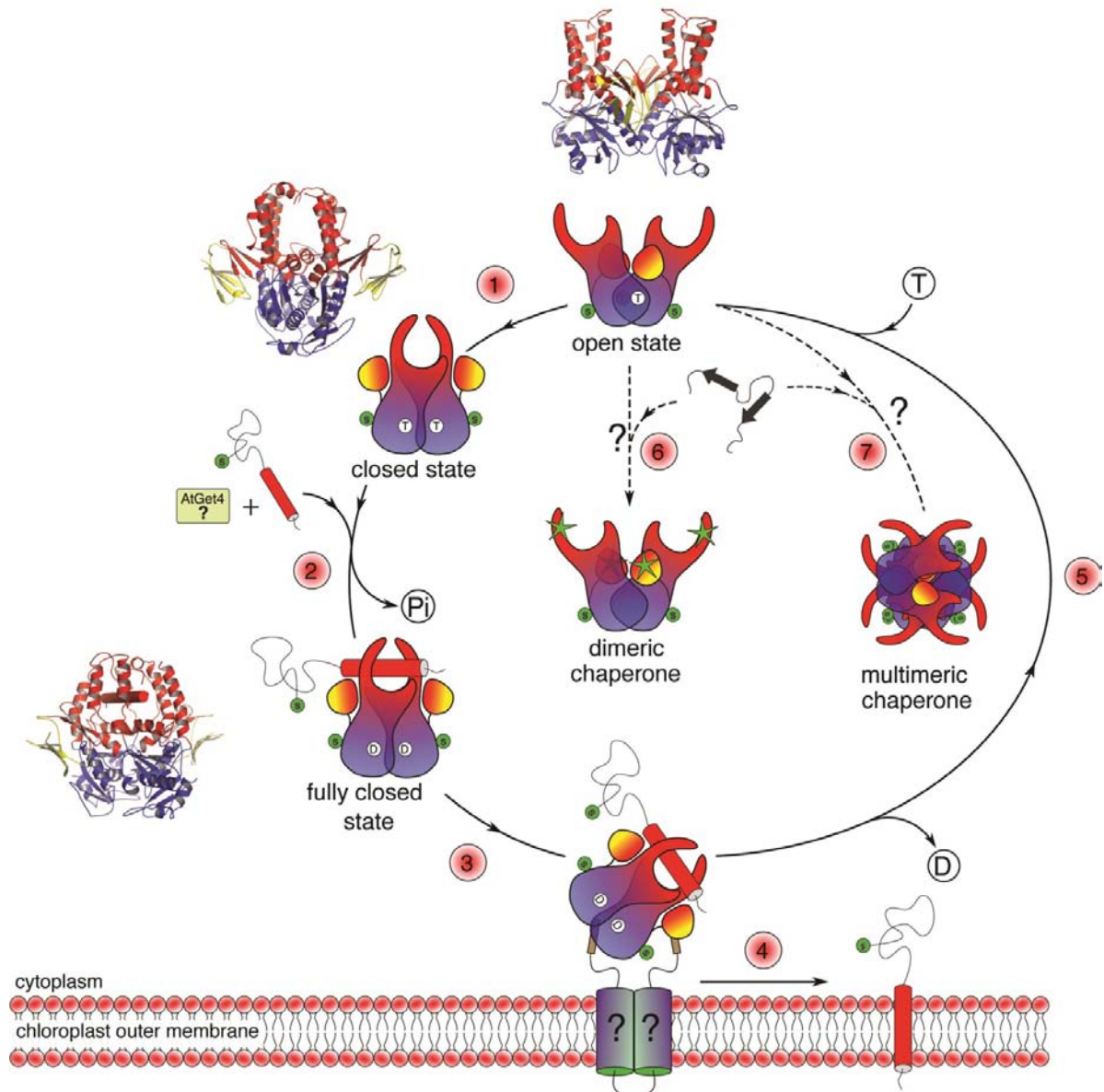
Figure 6.1. Schematic representation of probable AtGet3d functions. HSP domain orientation in open form of AtGet3d determines the faith of AtGet3d. ATP binding converts open form to closed state, trigger the TA targeting pathway along with some unknown components (**step 1**). ATP hydrolysis ensures tight binding of TA with AtGet3d (**step 2**). Signal sequence of both AtGet3d and TA proteins together determine the target location. This complex can tether with unknown membrane protein on the surface of chloroplast outer membrane (**step 3**) and transfer TA proteins into chloroplast outer membrane (**step 4**). After TA protein targeting, ADP will be release and AtGet3d will be recycled for next cycle (**step

5). During stress condition (dotted lines) the open form of AtGet3d can act as chaperone either in dimeric state or in multimeric state. In open state hydrophobic regions of both TA binding region and HSP (indicated in green stars) will get expose and can accept bigger substrates like aggregated/misfolded proteins (**step 6**). Similarly, AtGet3d can form multimeric chaperones through HSP domain interaction and can also bind with aggregated proteins (**step 7**).

stress (Figure 4.8). Further, the increased interface area in AtGet3d over ScGet3 indicates that comparatively larger substrates could be protected by Atget3d. In AtGet3b, the features of the signal sequence are considerably different. The extra sequences may be necessary for targeting it to different compartments of the chloroplast. The presence of CXXC motif in AtGet3b also implies that it can play a chaperone role during redox stress in the chloroplast. Similar to the ER which has a coordinated and regulated protein quality control along with its protein targeting and insertion machinery, other organelles like chloroplasts also require a certain degree of sophistication to maintain their protein homeostasis. This study sheds light on this type of machinery; further studies in this area would reveal more insights into the working of this system. The prominent question that remains to be addressed is whether for the Get3d, with dual function, are both these functions equally relevant *in vivo* or is one more important than the other? What features are common among the client proteins to bind to the Get3d chaperon?

6.3. Future Scope

The present study shows that AtGet3d is different from other reported Get3 both structurally and functionally. Yeast complementation assay can give more information about the function of AtGet3d at *in vivo* level. The ATPase activity of AtGet3d found relatively slow compared to AtGet3a in this study. From the structure, it is speculated that the slow activity is due to the extra hydrogen bonds that Ser13 make between Lys10 and Val 236. The mutational studies of P-loop residues and the assessment of kinetic parameters will give more insights about the ATPase hydrolyzing ability of AtGet3d. As shown in this study, HSP domain of AtGet3d may hinder the interaction of Get4 with AtGet3d. Hence it will be interesting to elucidate the mechanism of AtGet3d interaction with other GET pathway members in near future.

It may also possible that AtGet3d can act as an independent TA protein transporter without interacting with other known GET pathway members. Previous studies show that Msp1 act as membrane protein dislocase for TA proteins, which are wrongly integrated into the mitochondrial membranes in yeast. But in the plant system, the presence of Msp1or protein

similar to Msp1 is yet to be identified. The expression of AtGe3d is found higher during the stress conditions. Also, the pull down studies shows that AtGet3d interact with stress-related proteins. From the model of open from of AtGet3d, it is assumed that AtGet3d can accommodate bigger unfolded protein during stress conditions. This chaperonin activity of AtGet3d has to be validated by *in vivo* experiments. Initial phenotypic characterization of AtGet3d in this study showed that the mutant seeds were germinated slowly compared to wild-type. Even though the mutant plants didn't show any phenotypic difference with wild-type, the effects of various stress conditions like the presence of metal ions, osmotic stress, heat stress etc. on the mutant plants needs further investigation.

Appendices

Table A1: List of proteins interacting with AtGet3d identified by co-immuno precipitations and LC-MS/MS.

Type	Uniprot ID	Description	Soluble/Membrane Protein
HSP	O03986	Heat shock protein 90-4	Soluble
	P51818	Heat shock protein 90-3	
	P55737	Heat shock protein 90-2	
	Q9STW6	Heat shock 70 kDa protein 6	
Actin	P53492	Actin-7	Soluble
	P53494	Actin-4	
	P53496	Actin-11	
	P53497	Actin-12	
Ubiquitin	B9DHA6	Ubiquitin-60S ribosomal protein L40-1	Soluble
	P0CH32	Polyubiquitin 4	
	P0CH33	Polyubiquitin 11	
	P59232	Ubiquitin-40S ribosomal protein S27a-2	
	P59233	Ubiquitin-40S ribosomal protein S27a-3	
	P59271	Ubiquitin-40S ribosomal protein S27a-1	
	Q1EC66	Polyubiquitin 3	
	Q3E7K8	Polyubiquitin 12	
	Q3E7T8	Polyubiquitin 14	
	Q42202	Ubiquitin-60S ribosomal protein L40-2	
	Q8H159	Polyubiquitin 10	
	Q8RUC6	Ubiquitin-NEDD8-like protein RUB2	
	Q9FHQ6	Polyubiquitin 9	
	Q9SHE7	Ubiquitin-NEDD8-like protein RUB1	
Ribosomal	P17094	60S ribosomal protein L3-1	Soluble
	P36210	50S ribosomal protein L12-1, chloroplastic	
	P36212	50S ribosomal protein L12-3, chloroplastic	
	P42791	60S ribosomal protein L18-2	
	P51407	60S acidic ribosomal protein P2-1	
	Q08682	40S ribosomal protein Sa-1	
	Q08770	60S ribosomal protein L10-2	
	Q8H173	40S ribosomal protein Sa-2	
	Q93VT9	60S ribosomal protein L10-1	
	Q93W22	60S ribosomal protein L10-3	
	Q940B0	60S ribosomal protein L18-3	
	Q9SLF7	60S acidic ribosomal protein P2-2	
ATP Synthase	P56758	ATP synthase subunit a, chloroplastic	Membrane

Type	Uniprot ID	Description	Soluble/Membrane Protein
	P56759	ATP synthase subunit b, chloroplastic	
ATP Synthase	P56760	ATP synthase subunit c, chloroplastic	Membrane
	P60112	ATP synthase subunit 9, mitochondrial	
	P09468	ATP synthase epsilon chain, chloroplastic	Soluble
	P19366	ATP synthase subunit beta, chloroplastic	
	P56757	ATP synthase subunit alpha, chloroplastic	
	P83483	ATP synthase subunit beta-1, mitochondrial	
	P83484	ATP synthase subunit beta-2, mitochondrial	
	P92549	ATP synthase subunit alpha, mitochondrial	
	Q01908	ATP synthase gamma chain 1, chloroplastic	
	Q9C5A9	ATP synthase subunit beta-3, mitochondrial	
	Q9SSS9	ATP synthase subunit delta, chloroplastic	
ATPase	P0CW77	Putative inactive cadmium/zinc-transporting ATPase HMA3	Membrane
	P0CW78	Cadmium/zinc-transporting ATPase HMA3	
	Q43128	ATPase 10, plasma membrane-type	
	O23654	V-type proton ATPase catalytic subunit A	Soluble
	P11574	V-type proton ATPase subunit B1	
	Q39258	V-type proton ATPase subunit E1	
	Q8W4E2	V-type proton ATPase subunit B3	
	Q9SZN1	V-type proton ATPase subunit B2	
Photosystem	P56761	Photosystem II D2 protein	Membrane
	P56766	Photosystem I P700 chlorophyll a apoprotein A1	
	P56767	Photosystem I P700 chlorophyll a apoprotein A2	
	P56777	Photosystem II CP47 reaction center protein	
	P56778	Photosystem II CP43 reaction center protein	
	P83755	Photosystem II protein D1	
	Q949Q5	Photosystem I subunit O	
	Q9S7N7	Photosystem I reaction center subunit V, chloroplastic	
	Q9SUI4	Photosystem I reaction center subunit XI, chloroplastic	
	Q9SUI5	Photosystem I reaction center subunit psaK, chloroplastic	
	P56780	Photosystem II reaction center protein H (PSII-H) (Photosystem II 10 kDa phosphoprotein)	
	P60129	Photosystem II reaction center protein L (PSII-L)	
	Q9SHE8	Photosystem I reaction center subunit III, chloroplastic (Light-harvesting complex I 17 kDa protein) (PSI-F)	Membrane
Photosystem	P62090	Photosystem I iron-sulfur center	Soluble
	Q9S714	Photosystem I reaction center subunit IV B, chloroplastic	
	Q9S831	Photosystem I reaction center subunit IV A, chloroplastic	
	Q9SA56	Photosystem I reaction center subunit II-2, chloroplastic	
	Q9SUI6	Photosystem I reaction center subunit VI-2,	

Type	Uniprot ID	Description	Soluble/Membrane Protein
		chloroplastic	
	Q9SUI7	Photosystem I reaction center subunit VI-1, chloroplastic	
	Q9SY97	Photosystem I chlorophyll a/b-binding protein 3-1, chloroplastic	
	Q9SYW8	Photosystem I chlorophyll a/b-binding protein 2, chloroplastic	
	Q9XF91	Photosystem II 22 kDa protein, chloroplastic	
Aquaporin	P25818	Aquaporin TIP1-1	Membrane
	P43286	Aquaporin PIP2-1	
	P61837	Aquaporin PIP1-1	
	Q06611	Aquaporin PIP1-2	
	Q08733	Aquaporin PIP1-3	
	Q39196	Probable aquaporin PIP1-4	
	Q41963	Aquaporin TIP1-2	
Cytochrome	P56771	Cytochrome f	Membrane
	P56773	Cytochrome b6	
	P56779	Cytochrome b559 subunit alpha	
	Q6NKZ8	Cytochrome P450 714A2	
	Q9SZU1	Cytochrome P450 81F4	
	Q9ZR03	Cytochrome b6-f complex iron-sulfur subunit, chloroplastic	
Peroxisomal	Q9C8D4	Butyrate--CoA ligase AAE11, peroxisomal	Soluble
	Q9LRR9	Peroxisomal (S)-2-hydroxy-acid oxidase GLO1	
	Q9LRS0	Peroxisomal (S)-2-hydroxy-acid oxidase GLO2	
Chaperone	P21240	Chaperonin 60 subunit beta 1, chloroplastic	Soluble
Chaperone	Q9FI56	Chaperone protein ClpC1, chloroplastic	
	Q9LJE4	Chaperonin 60 subunit beta 2, chloroplastic	
	Q9SXJ7	Chaperone protein ClpC2, chloroplastic	
Other	O04616	Protein CURVATURE THYLAKOID 1A, chloroplastic	Membrane
	O65530	Proline-rich receptor-like protein kinase PERK14	
	O80632	Metal tolerance protein 11	
	P31167	ADP,ATP carrier protein 1, mitochondrial	
	P42699	Plastocyanin major isoform, chloroplastic	
	P93319	Uncharacterized mitochondrial protein AtMg00670	
	Q0WNW4	Myosin-binding protein 3	
	Q3EDL4	Probable serine/threonine-protein kinase At1g01540	
	Q42534	Pectinesterase 2	
	Q84WF0	Serine carboxypeptidase-like 37	
	Q84WP5	CASP-like protein 4A3	
	Q8GW38	RING-H2 finger protein ATL47	
	Q8RY65	Protein NSP-INTERACTING KINASE 2	
	Q9C8H1	ABC transporter C family member 11	
	Q9FL07	RING-H2 finger protein ATL46	

Type	Uniprot ID	Description	Soluble/Membrane Protein
	Q9LJG3	GDSL esterase/lipase ESM1	
	Q9LRL6	Putative cysteine-rich repeat secretory protein 23	
	Q9LUG9	Mediator of RNA polymerase II transcription subunit 33A	
	Q9LXT9	Callose synthase 3	
	Q9M129	WAT1-related protein At4g01450	
	Q9SE50	Beta-D-glucopyranosyl abscisate beta-glucosidase	
	Q9SIE7	PLAT domain-containing protein 2	
	Q9ZQR4	DUF21 domain-containing protein At2g14520	
	Q9ZSR7	Triose phosphate/phosphate translocator TPT, chloroplastic	
	Q9ZUT8	ABC transporter G family member 33	
	Q6IDL4	Transmembrane emp24 domain-containing protein p24delta3	
	Q9ZVL6	UPF0603 protein At1g54780, chloroplastic (Thylakoid lumen 18.3 kDa protein)	
Other	F4I1T7	Nuclear pore complex protein NUP214	Soluble
	F4JNB7	Disease resistance protein RPP5	
	O03042	Ribulose bisphosphate carboxylase large chain	
	O04019	26S proteasome regulatory subunit 6A homolog B	
	O04309	Jacalin-related lectin 35	
	O04496	Aspartyl protease AED3	
	O23140	AP-2 complex subunit mu	
	O23255	Adenosylhomocysteinase 1	
	O49298	Probable pectinesterase/pectinesterase inhibitor 6	
	O49647	Putative F-box protein At4g22660	
	O50008	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase 1	
	O64766	Pentatricopeptide repeat-containing protein At2g35030, mitochondrial	
	O64789	Probable disease resistance protein At1g61310	
	O65621	Probable cinnamyl alcohol dehydrogenase 6	
	O80852	Glutathione S-transferase F9	
	O82794	MADS-box protein AGL24	
	P04778	Chlorophyll a-b binding protein 1, chloroplastic	
	P0CJ48	Chlorophyll a-b binding protein 2, chloroplastic	
	P0DH91	ADP-ribosylation factor 2-B	
	P0DH99	Elongation factor 1-alpha 1	
	P10795	Ribulose bisphosphate carboxylase small chain 1A, chloroplastic	
	P10797	Ribulose bisphosphate carboxylase small chain 2B, chloroplastic	
	P10896	Ribulose bisphosphate carboxylase/oxygenase activase, chloroplastic	
	P17745	Elongation factor Tu, chloroplastic	

Type	Uniprot ID	Description	Soluble/Membrane Protein
	P23321	Oxygen-evolving enhancer protein 1-1, chloroplastic	
	P25856	Glyceraldehyde-3-phosphate dehydrogenase GAPA1, chloroplastic	
	P25857	Glyceraldehyde-3-phosphate dehydrogenase GAPB, chloroplastic	
	P27140	Beta carbonic anhydrase 1, chloroplastic	
	P27521	Chlorophyll a-b binding protein 4, chloroplastic	
Other	P32962	Nitrilase 2	Soluble
	P34791	Peptidyl-prolyl cis-trans isomerase CYP20-3, chloroplastic	
	P36397	ADP-ribosylation factor 1	
	P38418	Lipoxygenase 2, chloroplastic	
	P39207	Nucleoside diphosphate kinase 1	
	P42737	Beta carbonic anhydrase 2, chloroplastic	
	P42760	Glutathione S-transferase F6	
	P42763	Dehydrin ERD14	
	P43297	Cysteine proteinase RD21A	
	P46422	Glutathione S-transferase F2	
	P48785	Pathogenesis-related homeodomain protein	
	P52410	3-oxoacyl-[acyl-carrier-protein] synthase I, chloroplastic	
	P57106	Malate dehydrogenase 2, cytoplasmic	
	P93043	Vacuolar protein sorting-associated protein 41 homolog	
	P93819	Malate dehydrogenase 1, cytoplasmic	
	Q01667	Chlorophyll a-b binding protein 6, chloroplastic	
	Q07473	Chlorophyll a-b binding protein CP29.1, chloroplastic	
	Q0WL56	Elongation factor 1-alpha 3	
	Q0WQ57	Auxilin-related protein 2	
	Q1G3U6	Plant cysteine oxidase 3	
	Q39102	ATP-dependent zinc metalloprotease FTSH 1, chloroplastic	
	Q39160	Myosin-5	
	Q3EBY8	F-box protein At2g17690	
	Q42029	Oxygen-evolving enhancer protein 2-1, chloroplastic	
	Q42547	Catalase-3	
	Q43127	Glutamine synthetase, chloroplastic/mitochondrial	
	Q56YA5	Serine--glyoxylate aminotransferase	
	Q5PNS9	Probable protein phosphatase 2C 64	
	Q6DYE4	Uncharacterized protein At1g26090, chloroplastic	
	Q708Y0	EIN3-binding F-box protein 2	
	Q84J62	UPF0725 protein At2g19200	
Other	Q84TH4	Serine/arginine-rich splicing factor SR45a	Soluble
	Q8GTY0	Elongation factor 1-alpha 4	
	Q8GUI6	Probable lysine-specific demethylase JMJ14	
	Q8GYM1	Glutathione S-transferase U22	

Type	Uniprot ID	Description	Soluble/Membrane Protein
	Q8L7G4	Terpenoid synthase 9	
	Q8LFS6	RNA-binding protein BRN1	
	Q8RWV0	Transketolase-1, chloroplastic	
	Q8VZ87	Chlorophyll a-b binding protein 3, chloroplastic	
	Q8VZJ2	Probable glucan endo-1,3-beta-glucosidase At4g16260	
	Q8W4C8	ADP-ribosylation factor-like protein 8c	
	Q8W4H7	Elongation factor 1-alpha 2	
	Q8W4J9	Disease resistance protein RPP8	
	Q941D3	Probable plastid-lipid-associated protein 8, chloroplastic	
	Q94CJ5	Protein RETICULATA-RELATED 4, chloroplastic	
	Q94LA4	Probable delta-aminolevulinic acid dehydratase 2, chloroplastic	
	Q96262	Plasma membrane-associated cation-binding protein 1	
	Q9C681	Probable histone H2A.1	
	Q9C6I6	Electron transfer flavoprotein subunit alpha, mitochondrial	
	Q9C8Y5	FBD-associated F-box protein At1g66320	
	Q9C8Y6	FBD-associated F-box protein At1g66310	
	Q9C8Y8	Putative F-box/FBD/LRR-repeat protein At1g66290	
	Q9CAH0	Multiple organellar RNA editing factor 7, mitochondrial	
	Q9FH02	ATP-dependent zinc metalloprotease FTSH 5, chloroplastic	
	Q9FIL7	Calmodulin-binding receptor-like cytoplasmic kinase 1	
	Q9FIM2	ATP-dependent zinc metalloprotease FTSH 9, chloroplastic	
	Q9FX54	Glyceraldehyde-3-phosphate dehydrogenase GAPC2, cytosolic	
Other	Q9LD57	Phosphoglycerate kinase 1, chloroplastic	Soluble
	Q9LJD8	MAP3K epsilon protein kinase 1	
	Q9LK36	Adenosylhomocysteinase 2	
	Q9LNJ9	Bifunctional fucokinase/fucose pyrophosphorylase	
	Q9LPW0	Glyceraldehyde-3-phosphate dehydrogenase GAPA2, chloroplastic	
	Q9LQC8	ADP-ribosylation factor 2-A	
	Q9LRZ5	Phospholipase D zeta 1	
	Q9LU73	Protein SMAX1-LIKE 5	
	Q9LW57	Plastid-lipid-associated protein 6, chloroplastic	
	Q9LZ06	Glutathione S-transferase L3	
	Q9M063	Putative GEM-like protein 3	
	Q9M0M4	Putative MO25-like protein At4g17270	
	Q9M5K2	Dihydrolipoyl dehydrogenase 2, mitochondrial	
	Q9M5K3	Dihydrolipoyl dehydrogenase 1, mitochondrial	
	Q9MBA2	Putative septum site-determining protein minD homolog, chloroplastic	
	Q9S7E4	Formate dehydrogenase, chloroplastic/mitochondrial	

Type	Uniprot ID	Description	Soluble/Membrane Protein
	Q9S7J7	Chlorophyll a-b binding protein 2.2, chloroplastic	
	Q9S7M0	Chlorophyll a-b binding protein 3, chloroplastic	
	Q9S841	Oxygen-evolving enhancer protein 1-2, chloroplastic	
	Q9SA52	Chloroplast stem-loop binding protein of 41 kDa b, chloroplastic	
	Q9SCT6	WEB family protein At3g51720	
	Q9SEI2	26S proteasome regulatory subunit 6A homolog A	
	Q9SH42	Probable cysteine-rich repeat secretory protein 6	
	Q9SH43	Putative cysteine-rich repeat secretory protein 5	
	Q9SHR7	Chlorophyll a-b binding protein 2.1, chloroplastic	
	Q9SKK4	Probable 2-oxoacid dependent dioxygenase	
	Q9SLM6	Glutathione S-transferase F3	
	Q9SR66	DEMETER-like protein 2	
	Q9SRY5	Glutathione S-transferase F7	
	Q9SSK1	Asparagine--tRNA ligase, cytoplasmic 3	
	Q9STF2	Protein plastid transcriptionally active 16, chloroplastic	
Other	Q9SU08	Auxilin-related protein 1	Soluble
	Q9SW48	Probable alkaline/neutral invertase B	
	Q9SYG7	Aldehyde dehydrogenase family 7 member B4	
	Q9SYT0	Annexin D1	
	Q9SZH4	RNA-binding KH domain-containing protein PEPPER	
	Q9SZJ5	Serine hydroxymethyltransferase 1, mitochondrial	
	Q9XF87	Chlorophyll a-b binding protein 2.4, chloroplastic	
	Q9XF88	Chlorophyll a-b binding protein CP29.2, chloroplastic	
	Q9XF89	Chlorophyll a-b binding protein CP26, chloroplastic	
	Q9XFS9	1-deoxy-D-xylulose 5-phosphate reductoisomerase, chloroplastic	
	Q9XFT3	Oxygen-evolving enhancer protein 3-1, chloroplastic	
	Q9ZV94	Putative F-box/FBD/LRR-repeat protein At1g78760	

Table A2. Top 10 hits from DALI server for AtGet3Δd

PDB ID	Z-Score	RMSD	LALI score	% Identity	PDB Description
3igf(A)	33.3	2.6	322	31	ALL4481 PROTEIN
3zq6(A)	21.2	3.2	250	20	PUTATIVE ARSENICAL PUMP-DRIVING ATPASE
3ug6(D)	20.1	4.2	260	21	ARSENICAL PUMP-DRIVING ATPASE
4xtr(B)	19.5	3.2	253	18	ATPASE GET3
2woj(B)	19.5	3.2	246	18	ATPASE GET3

PDB ID	Z-Score	RMSD	LALI score	% Identity	PDB Description
3ug7(A)	19.3	4.2	260	21	ARSENICAL PUMP-DRIVING ATPASE
3iqw(B)	18.5	3.3	212	16	TAIL-ANCHORED PROTEIN TARGETING FACTOR GET3
5auq(B)	16.8	3.3	193	16	ATPASE INVOLVED IN CHROMOSOME PARTITIONING
4rz3(A)	16.6	3.1	194	15	SITE-DETERMINING PROTEIN
5auo(B)	16.5	3.2	190	16	PROBABLE HYDROGENASE NICKEL INCORPORATION PROTEIN

*Parenthesis represents chain ID

Proteins which are similar to the AtGet3Δd structure were identified using DALI server. ALL4481 protein from Nostoc has maximum identity and structural similarity with AtGet3Δd.

Table A3. Comparison of ScGet3 residues with AtGet3Δd and AtGet3b residues.

Properties	AtGet3Δd[±]	AtGet3b-Model[#]	Yeast Get3	Reference
Interface Residues	Gln40, Asp41, Thr70, Thr71, Gln92, Val99, Val105, Gly108, Tyr164, Tyr167, Thr175, Arg250, Gln370, Arg371, Glu149, Asp369, Glu150, Gly104, Gly97, Gly100, Leu103, His38, Lys51, Arg153, Arg358, Glu351, Asp110, Glu102	Ser67, Ser256, Arg260, Arg327, His66, Glu257, Asp70, Val252, Lys137, Arg176, Glu140, Glu99, Arg102, Arg176	Asn61, Ser132, Arg175, Thr182, Phe246, Arg254, Arg287, Arg291, Tyr298, Glu320, His 60, Asn61, Glu251, Leu275, Met 294, Cys317, Glu320, Tyr 348, His172, Lys293, Asp64, Asp137, Glu138	CCP4-PISA(Krissinel & Henrick 2007)
TA binding residues	Gln80, Phe125, Leu116, Met210, Ile185, Ser184, Ala84, Pro77, Arg120, Ile112,	Gly127, Asp150, Ile153, Leu184, Ser187,	Met97, Leu126, Met143, Met146, Leu183, Leu186, Phe190, Leu216, Leu219	4XTR(Mateja et al. 2015)

	Glu76, Phe113	Lys190, Leu222, Leu225		
Nucleotide Binding residues	Gly12, Ser13, Lys15, Gly14, Thr16, Thr17, Asp238, Ser270, Ser271, His272	Gly36, Lys39, Gly38, Thr40, Arg327, Val322, Asn278, Ser41	Gly28, Lys31, Gly30, Thr32, Ile321, Cys317, Asn272, Thr33	2WOJ(Mateja et al. 2009)
Get4 binding residues	Pro242, Lys246, Leu249, Gly253, Cys254, Pro284	Val252, Ser256, Ser259, Ala263, Ser264, Glu307, Ala313	Phe246, Tyr250, Glu253, Gln257, Glu258, Glu304, Asp308	4PWX(Gristick et al. 2014)
Get1 binding residues	Pro242, Val245. Lys246, Leu249, Asp281, Phe282	Val252, Val255, Ser256, Ser259, Ala301, Asp308, Ser312	Phe246, Leu249, Tyr250, Glu253, Tyr298, Leu305, Tyr306	3ZS8(Mariappan et al. 2011)
Get2 binding residues	His261, Val283, Pro284, Leu249	Pro271, Ala313, Ser259	Asp265, Glu307, Asp308, Glu253	3ZS9(Mariappan et al. 2011)

±57 amino acids from N-terminal of full-length AtGet3d was deleted and renumbered accordingly.

\# 92 amino acids from N-terminal of full-length AtGet3b was deleted and renumbered accordingly.

Corresponding residues were identified by structural alignment in chimera except for interface residues. Interface residues are identified by PISA (CCP4). Residues forming hydrogen bonds and salt bridges are considered.

AtGet3d Sequence:

MVSLVNSSLTCSSLTLNLLPILRTETPSLSRKRRAAYVAATSSRDVN
DTAADSSQKLTKFVTFLGKGGSGKTTAAVFAAQHYALAGLSTCLVI
HNQDPSAEFLLGSKIGTSPTLINDNLSVIRLETTKMLLEPLKQLKQA
DARLNMTQGVLEGVVGEELGVLPGMDSIFSMLELERLVGFFRQAT
RKNHKGKPFDVIIYDGISTEETLRMIGLSSKTRLYAKYLRSLAEKTD
LGRLTSPSIMRFVDESMNINSNKSPFDGMTSPAMWDTLERFLETGA
SAWRDPERFRSFLVMDPNNPMSVKAALRYWGCTVQAGSHVSGAF
AISSSHLTSQIPKADFVPLPFASASVPFTITGLDWDKILLDQANSSIRE
LLSETVSHGTSLTQTVMFDTAKKLVTLFMPGFEKSEIKLYQYRGGS
ELLIEAGDQRRVIHLPSQIQGKVGGAKFVDRSLIVTMR

Figure A1: Primary sequence of AtGet3d. 57 amino acids from N-terminal are truncated to generate AtGet3Δd (Highlighted).
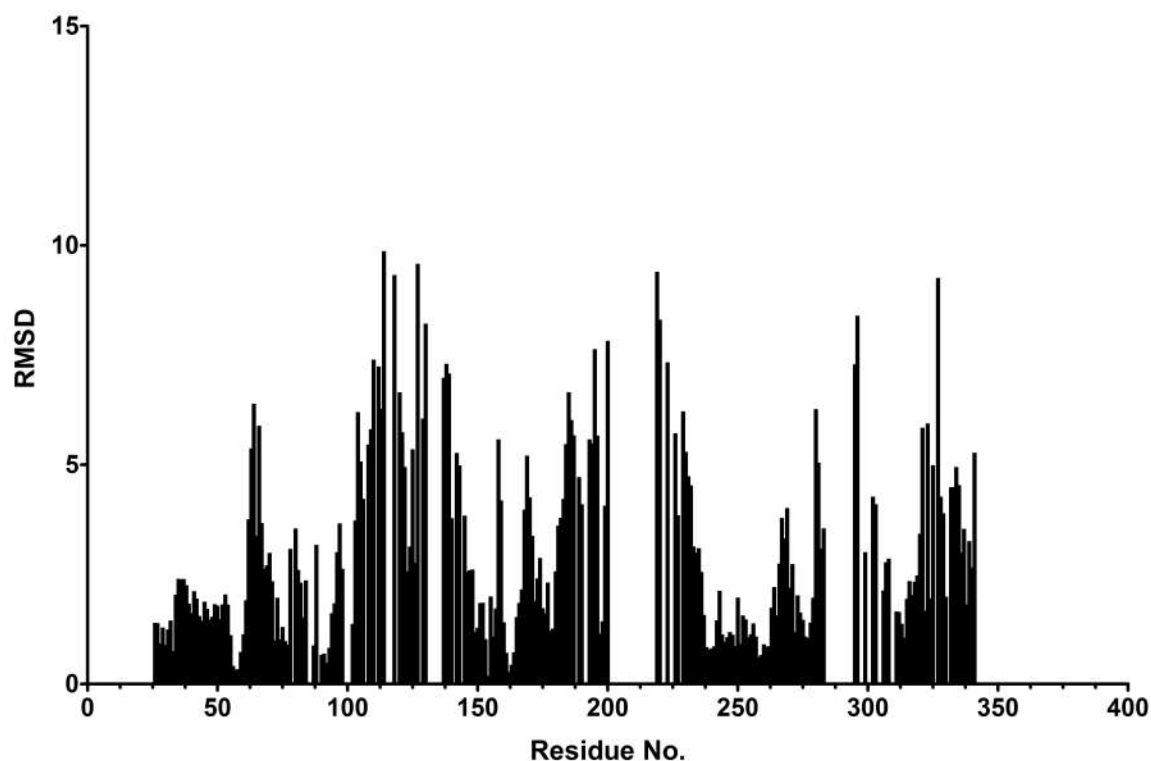
Figure A2: Residue wise RMSD comparison AtGet3b with AtGet3Δd. Gaps in the graph indicate the missing corresponding residue of AtGet3b compared to AtGet3Δd. Values 345 to 400 indicate the position of HSP domain in AtGet3d which is absent in AtGet3b (hence zero values).

Table A4: List of peptide dected during LC- MS analysis of purified full length AtGet3d. Two bands were observed in the SDS –PAGE. Both were Gel eluted and did the LC-MS analysis. Results show that both bands are belonging to same protein.

Peptide Sequence identified in LC-MS	
~50Kda band	**~45kDa band**
(R)VIHLPSQIQGK(V)	(R)GGSELLIEAGDQR(R)
(K)IGTSPTLINDNLSVIR(L)	(K)IGTSPTLINDNLSVIR(L)
(R)GGSELLIEAGDQR(R)	(R)VIHLPSQIQGK(V)
(R)DVNDTAADSSQKLTK(F)	(R)FLETGASAWRDPER(F)
(R)FVDESMNINSNK(S)	(R)AAYVAATSSR(D)
(R)FLETGASAWRDPER(F)	(R)FVDESMNINSNK(S)
(K)FVTFLGK(G)	(R)TETPSLSR(K)
(R)AAYVAATSSR(D)	(K)ILLDQANSSIR(E)
(R)TETPSLSR(K)	(R)DVNDTAADSSQKLTK(F)
(R)GGSELLIEAGDQRR(V)	(K)FVTFLGK(G)
(R)SLAEKTDLGR(L)	(K)MLLEPLKQLK(Q)
(K)LVTLFMPGFEK(S)	(R)RAAYVAATSSR(D)
(R)AAYVAATSSRDVNDTAADSSQK(L)	(R)GGSELLIEAGDQRR(V)
(K)SEIKLYQYR(G)	(R)LVGFFR(Q)
(K)MLLEPLKQLK(Q)	(R)DVNDTAADSSQK(L)
(K)ILLDQANSSIR(E)	(K)LYQYR(G)
(R)MIGLSSK(T)	(R)SLAEKTDLGR(L)
(R)RVIHLPSQIQGK(V)	(K)MLLEPLK(Q)
(R)FLETGASAWRDPERFR(S)	(K)LVTLFMPGFEK(S)
(R)LVGFFR(Q)	(R)FLETGASAWR(D)
(K)MLLEPLK(Q)	(R)MIGLSSK(T)
(K)SPFDGMTSPAMWDTLER(F)	(K)SPFDGMTSPAMWDTLER(F)
(K)IGTSPTLINDNLSVIRLETTK(M)	(R)SFLVMDPNNPMSVK(A)
(R)SLIVTMR(-)	(R)SLIVTMR(-)
(R)LTSPSIMR(F)	(R)FLETGASAWRDPERFR(S)
(R)SFLVMDPNNPMSVK(A)	(K)SEIKLYQYR(G)
(K)LYQYR(G)	(R)LTSPSIMR(F)
(R)RAAYVAATSSR(D)	(R)RVIHLPSQIQGK(V)
(R)AAYVAATSSRDVNDTAADSSQKLTK(F)	(R)AAYVAATSSRDVNDTAADSSQK(L)
(R)RAAYVAATSSRDVNDTAADSSQK(L)	(K)IGTSPTLINDNLSVIRLETTK(M)
(R)FLETGASAWR(D)	(R)FRSFLVMDPNNPMSVK(A)
(K)VGGAKFVDR(S)	(K)KLVTLFMPGFEK(S)
(K)FVTFLGKGGSGK(T)	(K)GKPFDVIIYDGISTEETLR(M)
(K)LVTLFMPGFEKSEIK(L)	(E)TPSLSR(K)
(K)SPFDGMTSPAMWDTLER(F)	(E)TPSLSR(K)
(K)KLVTLFMPGFEKSEIK(L)	(T)PSLSR(K)
(K)KLVTLFMPGFEK(S)	(P)SLSR(K)
(K)FVDRSLIVTMR(-)	(R)RAAYVAATSSR(D)
(K)TDLGRLTSPSIMR(F)	(A)YVAATSSR(D)

~50Kda band	~45kDa band
(K)GKPFDVIIYDGISTEETLR(M)	(V)AATSSR(D)
(E)TPSLSR(K)	(A)ATSSR(D)
(E)TPSLSR(K)	(R)DVNDTAADSSQKLTK(F)
(T)PSLSR(K)	(V)NDTAADSSQKLTK(F)
(R)RAAYVAATSSR(D)	(N)DTAADSSQKLTK(F)
(R)AAYVAATSSR(D)	(A)DSSQKLTK(F)
(R)AAYVAATSSRD(V)	(K)FVTFL(G)
(V)AATSSR(D)	(K)FVTFLGK(G)
(A)ATSSR(D)	(F)VTFLGK(G)
(R)DVNDTAADSSQKLTK(F)	(V)TFLGK(G)
(R)DVNDTAADSSQKLTK(F)	(T)FLGK(G)
(V)NDTAADSSQKLTK(F)	(M)LLEPLK(Q)
(N)DTAADSSQKLTK(F)	(M)LLEPLKQLK(Q)
(K)FVTFLGK(G)	(L)LEPLK(Q)
(F)VTFLGK(G)	(L)LEPLKQLK(Q)
(V)TFLGK(G)	(L)VGFFR(Q)
(V)TFLGKGGSGK(T)	(V)GFFR(Q)
(T)FLGK(G)	(G)FFR(Q)
(M)LLEPLK(Q)	(M)IGLSSK(T)
(M)LLEPLKQLK(Q)	(I)GLSSK(T)
(L)LEPLK(Q)	(S)LAEKTDLGR(L)
(L)LEPLKQLK(Q)	(L)AEKTDLGR(L)
(L)VGFFR(Q)	(A)EKTDLGR(L)
(V)GFFR(Q)	(A)EKTDLGR(L)
(G)FFR(Q)	(R)LTSPSIMR(F)
(M)IGLSSK(T)	(T)SPSIMR(F)
(I)GLSSK(T)	(S)PSIMR(F)
(R)SLAEKTDLGR(L)	(R)FLETGASAWRDPER(F)
(S)LAEKTDLGR(L)	(L)ETGASAWRDPER(F)
(A)EKTDLGR(L)	(L)ETGASAWRDPERFR(S)
(A)EKTDLGR(L)	(E)TGASAWRDPER(F)
(K)TDLGR(L)	(T)GASAWRDPER(F)
(R)LTSPSIMR(F)	(K)SEIKLYQYR(G)
(T)SPSIMR(F)	(E)IKLYQYR(G)
(S)PSIMR(F)	(I)KLYQYR(G)
(R)FLETGASAWRD(P)	(L)YQYR(G)
(R)FLETGASAWRD(P)	(Y)QYR(G)
(R)FLETGASAWRDPER(F)	(R)GGSELLIEAGDQRR(V)
(R)FLETGASAWRDPERFR(S)	(G)SELLIEAGDQRR(V)
(F)LETGASAWRDPERFR(S)	(G)SELLIEAGDQRR(V)
(L)ETGASAWRDPER(F)	(S)ELLIEAGDQRR(V)
(L)ETGASAWRDPERFR(S)	(E)LLIEAGDQRR(V)
(E)TGASAWRDPER(F)	(L)LIEAGDQRR(V)

~50Kda band	~45kDa band
(E)TGASAWRDPERFR(S)	(L)IEAGDQRR(V)
(T)GASAWRDPER(F)	(R)VIHL(P)
(D)PERFR(S)	(R)VIHLPS(Q)
(K)KLVTLF(M)	(R)VIHLPSQ(I)
(M)PGFEKSEIK(L)	(R)VIHLPSQIQGK(V)
(K)SEIK(L)	(V)IHLPSQIQGK(V)
(K)SEIKLYQYR(G)	(I)HLPSQIQGK(V)
(S)EIKLYQYR(G)	(H)LPSQIQGK(V)
(E)IKLYQYR(G)	(L)PSQIQGK(V)
(I)KLYQYR(G)	(P)SQIQGK(V)
(L)YQYR(G)	(S)QIQGK(V)
(Y)QYR(G)	(Q)IQGK(V)
(R)GGSELLIEAGDQRR(V)	(L)IVTMR(-)
(G)SELLIEAGDQRR(V)	(I)VTMR(-)
(G)SELLIEAGDQRR(V)	(V)TMR(-)
(S)ELLIEAGDQRR(V)	(R)TETPSLSR(K)
(E)LLIEAGDQRR(V)	
(L)LIEAGDQRR(V)	
(L)IEAGDQRR(V)	
(L)IEAGDQRR(V)	
(R)RVIHLPSQ(I)	
(R)VIHL(P)	
(R)VIHLPS(Q)	
(R)VIHLPSQ(I)	
(R)VIHLPSQIQGK(V)	
(R)VIHLPSQIQGK(V)	
(V)IHLPSQIQGK(V)	
(I)HLPSQIQGK(V)	
(H)LPSQIQGK(V)	
(L)PSQIQGK(V)	
(P)SQIQGK(V)	
(S)QIQGK(V)	
(V)GGAKFVDR(S)	
(L)IVTMR(-)	
(I)VTMR(-)	
(V)TMR(-)	
(R)TETPSLSR(K)	
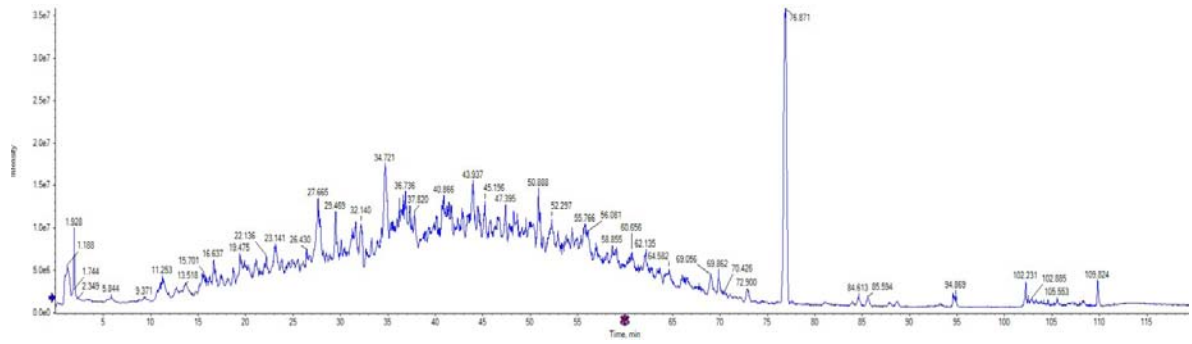
Figure A3: LC-MS/MS spectra of thermolysin treated sample
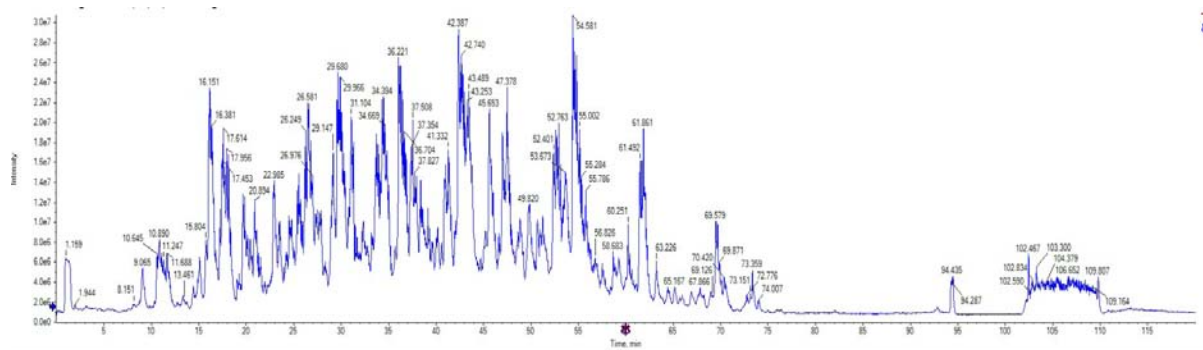
Figure A4: LC-MS/MS spectra of proteins interacting with AtGet3d
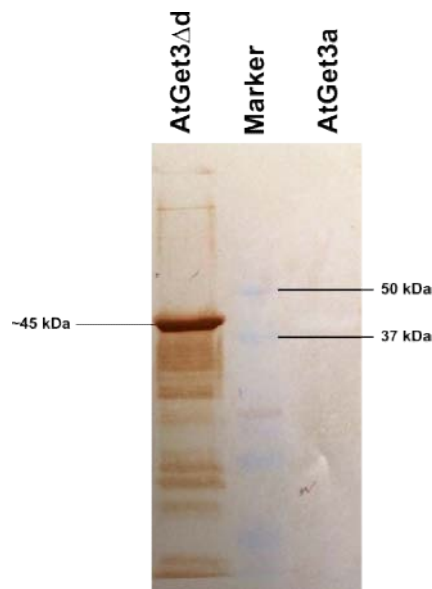
Figure A5: Custom made anti-AtGet3Δd antibody is tested for its specificity towards AtGet3d. Western blotting was carried for both recombinantly expressed AtGet3Δd (chloroplast Get3) and AtGet3a (cytosol Get3) with anti-AtGet3Δd antibody. Anti-AtGet3Δd antibody can bind specifically to AtGet3d.

(Additional soft data)

Some of the detailed analyses of this study are provided as additional soft data in the CD attached at the back of the hard copy of this thesis. These are also provided with the soft copy of the thesis in the CD.

Details of the soft copy files provided in the CD:

Table B1: List of identified TA proteins in *O. sativa* (.xlsx file).

Table B2: List of identified TA proteins in *S. tuberosum* (.xlsx file).

(Contributions to other projects)

C. Study of *Leishmania major* Mitogen Activated Protein Kinase 4 (Lmjmapk4) as a Drug Target against Leishmaniasis Disease.

C1. Introduction

Leishmania, a protozoan parasite, inflicts the disease leishmaniasis in its mammalian hosts. *Leishmania* exhibit dimorphic life cycle (Sand fly and mammalian host). *Leishmania* resides and replicates as amastigotes in myeloid cells, such as macrophages and dendritic cells (DCs), of the mammalian host (Figure C1).
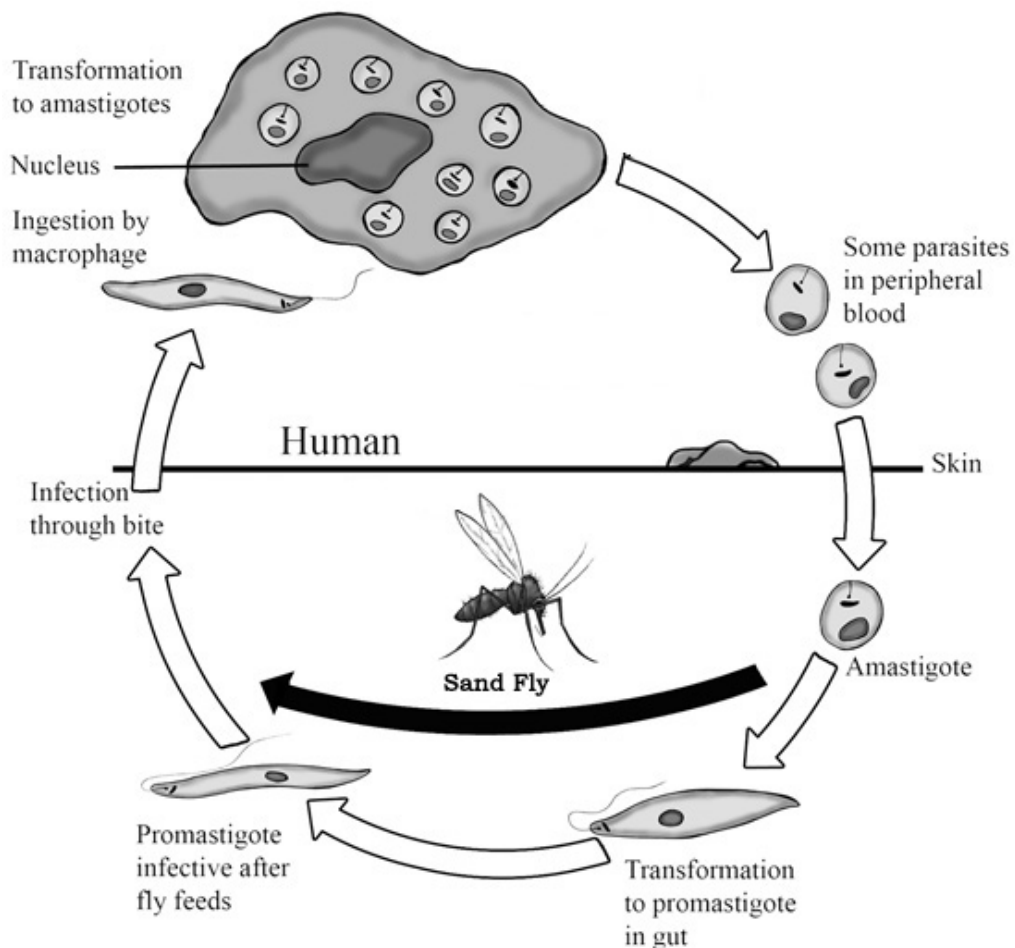
Figure C1. Life cycle of *Leishmania*

The previous studies proved that the host-protective effect againt leishmaniais due to CD40-induced IL-12 production. It also suggested that *L. major* infection selectively modulated

CD40 signaling in macrophages, resulting in reciprocal regulation of the CD40-induced p38MAPK and ERK-1/2 activation and differential modulation of IL-10 and IL-12 production (Figure C2). LmjMAPK4 is found homologous to human ERK1 and involved in pathway shifting from p38MAPK to ERK-1/2 to help parasite survival in macrophages.
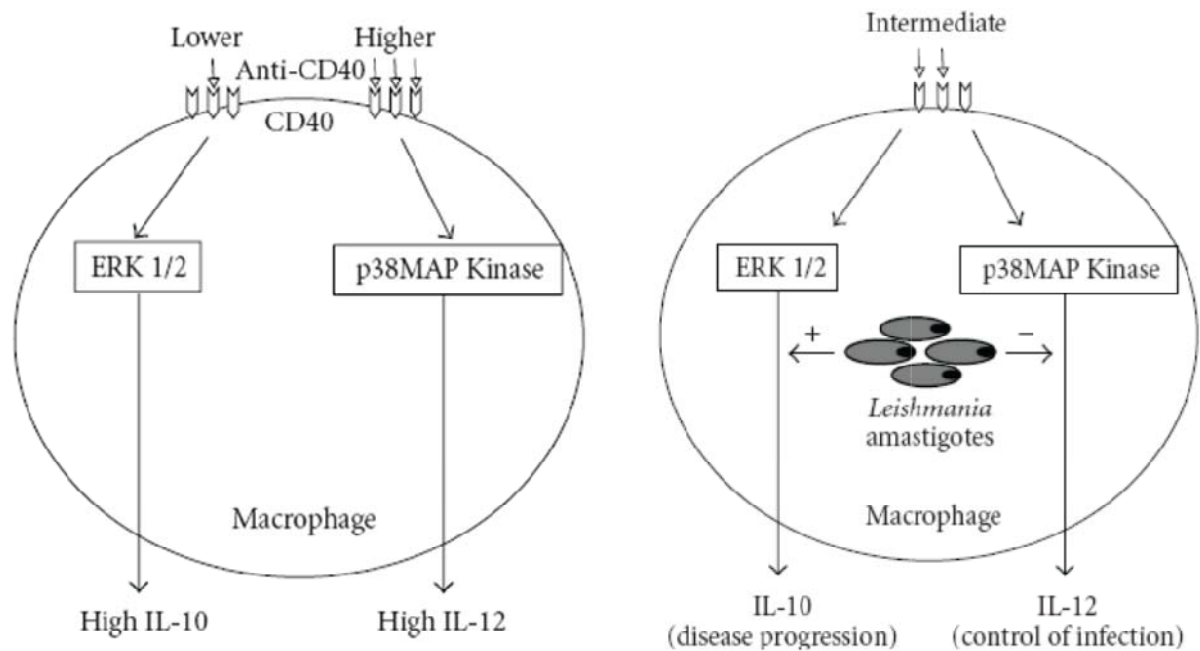
Figure C2. Modulation of CD40 signalling in macrophages by *L. major* infection.

In order to understand the mechanistic basis of LmjMAPK4 function, this study tried to over-express and purify LmjMAPK4 in prokaryotic expression systems. But, unfortunately, LmjMAPK4 is not stable during expression and found to be forming inclusion bodies.

C2. Materials and methods

C2.1. LmjMAPK4 clone

Cloned full length LmjMAPK4 (NCBI Accession Number: XM_001682680.1) in pET (28a+) vector was obtained from Lab#5 NCCS Pune. While expression of LmjMAPK4, it is found to be forming inclusion bodies.

C2.2. Cloning of Kinase domain of LmjMAPK4 with N terminal His tag.

To construct N terminal His tag LmjMAPK4Δ/pET28(a+), primers Pl (5'-GGGAATTCCATATGTATGATCTGGTCAAGGTTG -3') and P2 (5'-CCGCTCGAGTTAAAAATATGGATGTTCCATCAC -3') were used for PCR amplification of the Kinase domain of LmjMAPK4 from a plasmid encoding the full-length LmjMAPK4 (Clonned at NCCS Lab#5) and the resulting PCR fragment was digested with NdeI and

XhoI, and cloned between the Xho I and NdeI sites of the expression vector pET28(a+) (Figure C3).

1: Marker
2: Amplified LmjMAPK4 Δ

1&2: Insert
3&4: vector
5: Marker

1: Digested Clone
2: Marker
3: Undigested Clone
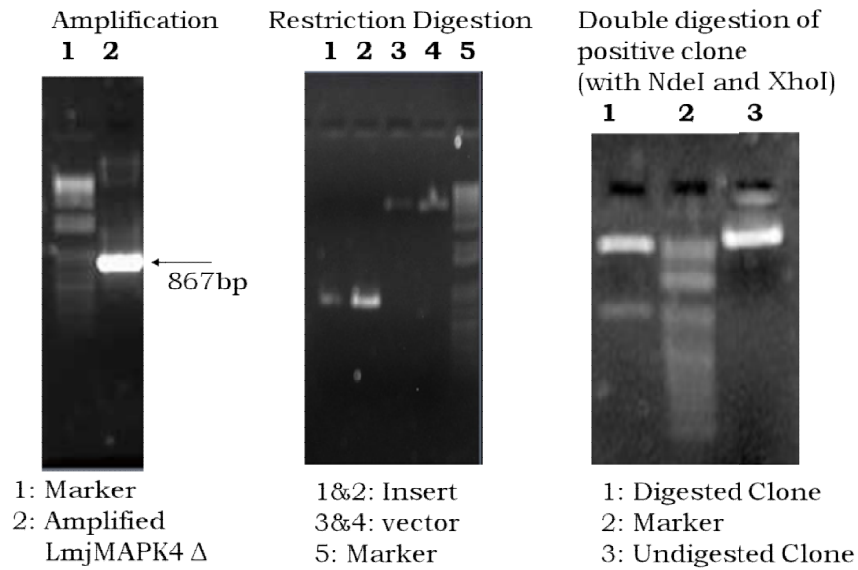
Figure C3. Cloning of LmjMAPK4 Kinase domain In pET 28 (a+)

C2.3 Cloning of LmjMAPK4 in pSSDS vector

In order to increase the solubility of LmjMAPK4, full-length LmjMAPK4 is cloned in pSSDS with sumo* tag. To construct N terminal His tag LmjMAPK4/pSSDS, USER mediated cloning strategy is used. Primers Pl (5'-CGGGGAUATGGCTCAACTCGTCCCTTTAGCTGAAC -3') and P2 (5'-CCCCGAUTTCGTTCAATTGTGAATGGGCTTCAACAACCC -3') were used for PCR amplification of LmjMAPK4 from a plasmid encoding the full-length LmjMAPK4 (LmjMAPK4 in pET 28a+). The resulting PCR fragment was introduced into pSSDS vector by USER mediated ligase independent cloning method. The positive clones were confirmed by digestion with Xba1 (Figure C4).
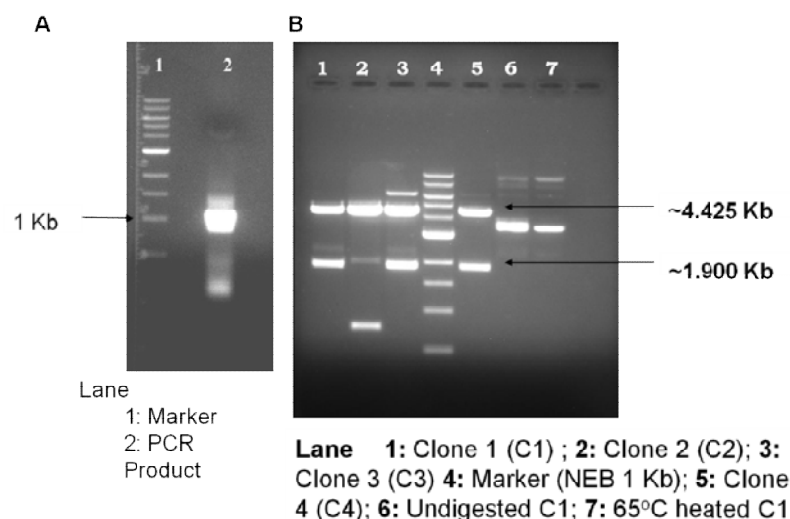
Figure C4. (A)Amplification of LmjMAPK4 and (B) restriction digestion with XbaI

C2.4 Cloning of LmjMAPK4 in P^{Opinss} vector

Cloning was carried out by USER mediated method as described above. To construct N terminal His tag LmjMAPK4/P^{Opinss}, USER mediated cloning strategy is used. Primers Pl (5'-CGGGGAUATGGCTCAACTCGTCCCTTTAGCTGAAC -3') and P2 (5'-CCCCGAUTTCGTTCAATTGTGAATGGGCTTCAACAACCC -3') were used for PCR amplification of LmjMAPK4 from a plasmid encoding the full-length LmjMAPK4 (LmjMAPK4 in pET 28a+). The resulting PCR fragment was introduced into pSSDS vector by USER mediated ligase independent cloning method. The positive clones were confirmed by digestion with Xba1.

C3. Results

C3.1. Expression of LmjMAPK4Δ in expression host BL21 DE3

To obtain the recombinant LmjMAPK4 kinase domain, the expression host E.coli BL21 DE3 was transformed with LmjMAPK4Δ/pET28(a+) construct by calcium chloride treatment. Selection of transformed colonies was performed on LB agar plate containing kanamycin. One millilitre of overnight culture was used to inoculate 100 ml of fresh LB broth with kanamycin. When the A600nm was 0.5, 1mM IPTG was added to induce LmjMAPK4 synthesis. At 4 hour after induction, cells were harvested by centrifugation at 4000 rpm for 15 minutes.

The bacterial cells were resuspended in lysis buffer(50mM Tris pH 8.5, 300mM NaCl, 10mM MgCl2, 1mM DTT, 10mM Imidazole 0.1% Triton X100 and 5% glycerol) and sonicate for 10 minute with 10 sec on/off pulse and 80% amplitude. After the cell disruption, the lysate was

centrifuged 15minute at 13000 rcf and collected the supernatant. Both supernatant and pellet were loaded in 12% SDS gel. LmjMAPK4Δ is found to be forming inclusion bodies during IPTG induction (Figure C5).
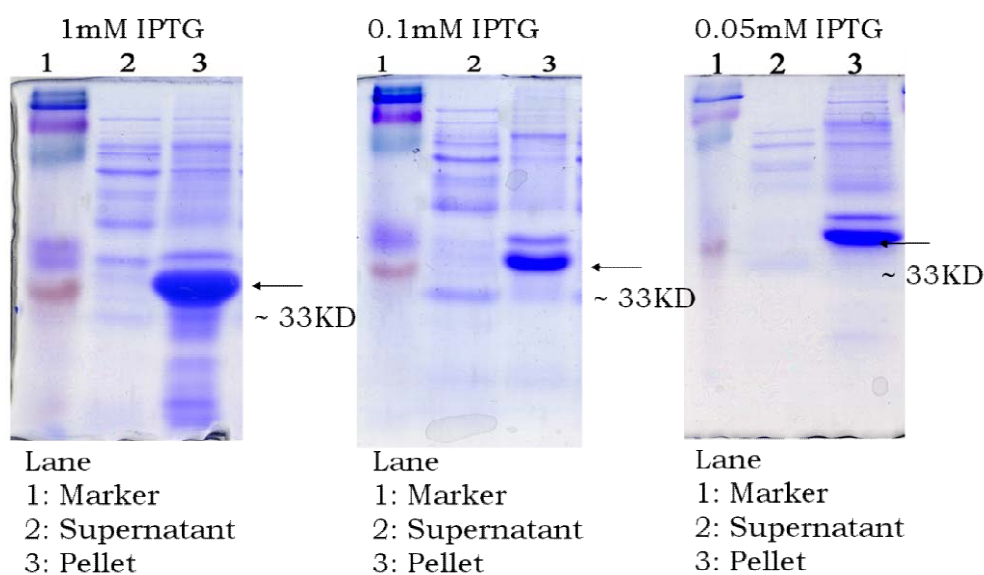
Figure C5. SDS PAGE analysis during expression and purification of LmjMAPK4 Δ in pET28(a+).

C3.2 Conditions tried to increase the solubility

LmjMAPK4 is expressed in different expression hosts such as JM109DE3, BL21DE3, Rosetta gami DE3 & Rosetta DE3. Different parameters such as IPTG concentration, Temperature of induction, Different growth media etc are tried to increase the solubility of LmjMAPK4. Parameters tried to increase the solubility are given in table C1.

Table C1. Parameters tried to increase the solubility of LmjMAPK4

Media Used	Organism	Inducer	Temperature of Incubation	Buffer for lysis	Observation
LB	JM109DE3	IPTG, Lactose	10, 16, 25& 37	Native and Denaturation conditions	In denaturation condition Lmj MAPK4 in soluble Fraction.
	BL21DE3				
	Rosetta gami DE3 &	IPTG			

Media Used	Organism	Inducer	Temperature of Incubation	Buffer for lysis	Observation
	Rosetta DE3				
LB with Glucose and Glycerol	BL21DE3	IPTG, Lactose	37,25		
ZYP 5052	BL21DE3	Auto induction	16,25 and 37		
	Rosetta gami DE3	Auto induction			

C3.2.1 Different concentration of IPTG

IPTG concentration varying from 0.05mM to 1mM is used for induction.

C3.2.2 Induction using Lactose

Instead of IPTG lactose is used for induction. 2mM and 5mM lactose are used to induce LmjMAPK4 at 37oC and 20oC.

C3.2.3 Used catabolic repressors

Catabolic repressors like glucose and glycerol are used in the media to increase the solubility during protein expression.

C3.2.4 Cold shock treatment

After IPTG induction the bacterial culture is alternatively exposed to lower temperature and $37^{o}C$. The aim was to induce the expression of the native heat shock proteins present in the bacteria and assist folding of expressed LmjMAPK4 bacterial cells.

C3.2.5 Continuous harvesting

After IPTG induction, at frequent intervals, growth media was replaced with fresh media.

C3.2.6 Added Arginine in LB media

During secondary culture, 0.2mM arginine is added to the LB media. But no bacterial growth observed during secondary culture.

C3.2.7 Different Lysis buffers

Buffers with ph ranging from 7.5 to 8.5 are used to lyse the bacterial cells. Different additives such as detergents and Glycerol are added in lysis buffer to stabilize the protein.

C3.2.8 Different Sonication parameters

Sonication parameters such as sonication time, pulse on/off time and amplitude are verified during sonication. Normally, 10 sec on and 10 sec off pulse for total 10 minutes at 80% amplitude is used for lysing the bacterial cells.

C3.3 Expression of LmjMAPK4/pSSDS in expression host B834

To obtain the recombinant sumo*tagged LmjMAPK4, the expression host E.coli B834 DE3 was transformed with LmjMAPK4/pSSDS construct by calcium chloride treatment. Selection of transformed colonies was performed on LB agar plate containing ampicillin and chloramphenicol. One millilitre of overnight culture was used to inoculate 100 ml of fresh LB broth with ampicillin and chloramphenicol. When the A600nm was 0.5, 0.5mM IPTG was added to induce LmjMAPK4 synthesis and shifted the culture at $16^{o}C$ for overnight. After induction, cells were harvested by centrifugation at 4000 rpm for 15 minutes.

The bacterial cells were resuspended in lysis buffer (50mM Tris pH 8, 500mM NaCl, 1mM 2 B mercaptoethanol, 2mM Imidazole and 10% glycerol) and sonicated for 5 minutes with 5 sec on/off pulse and 45% amplitude. After the cell disruption, the lysate was centrifuged for 45 minutes at 14000 rpm and collected the supernatant. Supernatants was loaded in 12% SDS gel (Figure C6A).

The clarified supernatant after cell lysis is passed through the pre-equilibrated Ni NTA beads with lysis buffer. The beads are washed with wash buffer (50mM Tris pH 8, 500mM NaCl, 2mM β mercaptoethanol, 2mM Imidazole and 10% glycerol) in order to remove the unbounded proteins. The bounded proteins were eluted with imidazole gradient from 2mM to 250mM (Figure C6A). Induced proteins were checked by SDS PAGE and western blot.

The eluted fractions were pooled together and treated with tev protease in order to remove the sumo*tag along with dialysis (dialysis buffer: 10mM Tris pH8, 150mM NaCl, 2mM DTT and 10% glycerol). The dialyzed sample was then concentrated and desalted. The desalted sample was passed through second Ni NTA column and collected the unbounded fractions. LmjMAPK4 without N- terminal His tag and Sumo* was expected to find in the unbounded fractions (Figure C6).
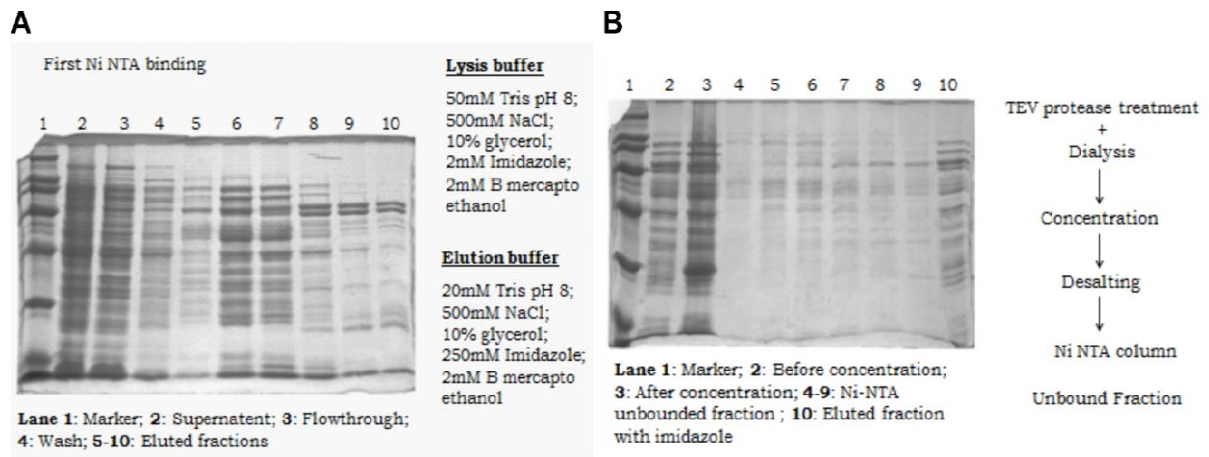
Figure C6. SDS PAGE analysis during expression and purification of LmjMAPK4 in pSSDS vector. (A)Ni-NTA affinity step before TEV protease treatment (B) Ni-NTA affinity step after TEV protease treatment.

C3.4 Expression of LmjMAPK4/p^{Opinss} in expression host B834

To obtain the recombinant sumo*tagged LmjMAPK4, the expression host E.coli B834 DE3 was transformed with LmjMAPK4/p^{Opinss} construct by calcium chloride treatment. Selection of transformed colonies was performed on LB agar plate containing ampicillin and chloramphenicol. One millilitre of overnight culture was used to inoculate 100 ml of fresh LB broth with ampicillin and chloramphenicol. When the A600nm was 0.5, 0.5mM IPTG was added to induce LmjMAPK4 synthesis and shifted the culture at 16°C for overnight. After induction, cells were harvested by centrifugation at 4000 rpm for 15 minutes.

The bacterial cells were resuspended in lysis buffer (50mM Tris pH 8, 500mM NaCl, 1mM 2 B mercaptoethanol, 2mM Imidazole and 10% glycerol) and sonicated for 5 minutes with 5 sec on/off pulse and 45% amplitude. After the cell disruption, the lysate was centrifuged for 45 minutes at 14000 rpm and collected the supernatant. Both supernatant and pellet were loaded in 12% SDS gel. An induced protein band is found to be expressed around 66 Kd.

C3.5. Refolding of LmjMAPK4 by On-Column methods

C3.5.1 Inclusion Body Expression and purification

To obtain the recombinant LmjMAPK4, the expression host E.coli Rosetta DE3 was transformed with pET 289(a+)-LmjMAPK4 construct by calcium chloride treatment. Selection of transformed colonies was performed on LB agar plate containing kanamycin and chloramphenicol. One millilitre of overnight culture was used to inoculate 100 ml of fresh LB broth with kanamycin and chloramphenicol. When the A600nm was 0.8, 0.25mM IPTG was added to induce LmjMAPK4 synthesis. At 4 hour after induction, cells were harvested by centrifugation at 4000 rpm for 15 minutes.

The bacterial cells were resuspended in lysis buffer and sonicated for 10 minutes with 10 sec on/off pulse and 80% amplitude. After the cell disruption, the inclusion bodies were isolated by centrifugation at 13000 rcf for 20 minutes. Inclusion bodies that were sedimented as pellet was washed twice with wash buffer1and reisolated by centrifugation described above. A final wash was given with wash buffer 2 and collected the purified inclusion bodies by centrifugation and stored at -80oC before further processing.

C3.5.2 Refolding trials using Ni NTA Superflow beads

After inclusion bodies solubilisation in 8M urea, 6M Gd HCl and 0.3% Sarcosine, Ni Nta binding trials are carried out for LmjMAPK4. It is observed that LmjMAPK4 does not bind to Ni NTA beads after solubilized in 8M urea and 6M Gd HCl. But it is found that LmjMAPK4 binds to Ni NTA Superflow beads after solubulise in CAPS buffer with 0.3% sarcosine

C3.5.3 Ni-Nta binding of LmjMAPK4 for refolding

Ni NTA binding was carried out by batch- absorption method. Equilibrate the Ni-NTA beads in equilibration buffer (50mM CAPS, pH 11, 100mM NaCl and 10mM Imidazole). Dilute the solubilized LmjMAPK4 in equilibration buffer in 1:1 ratio. Mix the diluted protein with equilibrated Ni-NTA beads (1ml resin for 5mg of protein) and keep in IP rotor for overnight.

C3.5.4 On Column refolding

The resin was packed into a column. All chromatographic steps were performed under gravity. First, the column was washed with 10 times column volume of wash buffer 1 (50mM CAPS pH 11, 100mM NaCl, 5mM β- cyclodextrin and 10mM B- mercaptoethanol). In the next step column was washed with 10 times column volume of wash buffer 2 (50mM CAPS pH 11, 500mM NaCl and 10mM B- mercaptoethanol) in order to remove the non-specific

binding and b- cyclodextrin. In the last step, the protein was eluted with the buffer containing 50mM CAPS pH 11, 100mM NaCl, 10mM B- mercaptoethanol and 300mM imidazole. Eluted protein was dialyzed against the dialysis buffer (20mM tris ph 8.5, 100mM NaCl and 1% glycerol) in the ratio of 1:500 (protein: dialysis buffer) overnight. But it was observed that during dialysis LmjMAPK4 tends to aggregate (Figure C7).
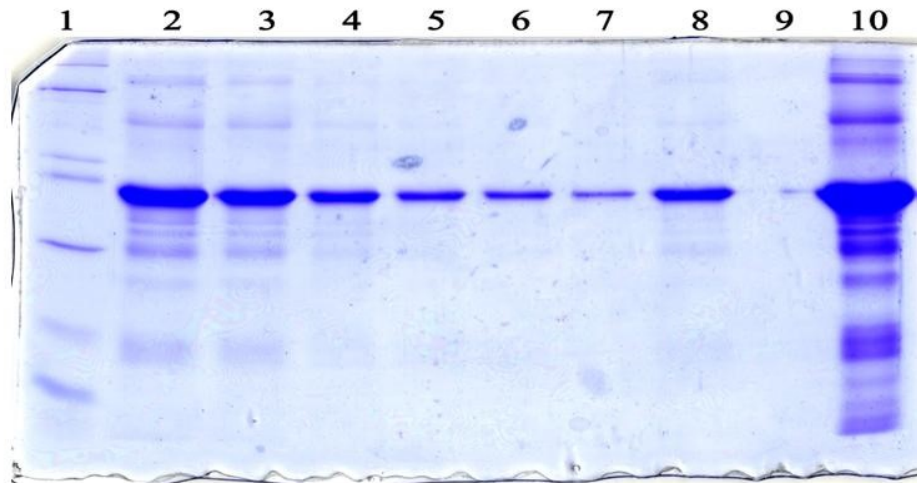
Lane 1: Marker, **2:** Elute 1, **3:** Elute 2, **4:** Elute 3, **5:** Elute4, **6:** Elute 5, **7:** Elute 7, **8:** All eluted fraction mixed , **9:** pH 8 dialyzed supernatant, **10:** pH 8 dialyzed aggregated M4

Figure C7. Modified column refolding method for LmjMAPK4. LmjMAPK4 found to be stable at pH above 10. While trying to bring the pH to 8, LmjMAPK4 tends to aggregate.

****

Bibliography

Alonso, J. M. et al. (2003) 'Genome-Wide Insertional Mutagenesis of Arabidopsis thaliana', Science, 301(5633), pp. 653–657. doi: 10.1126/science.1086391.

Beilharz, T. et al. (2003) 'Bipartite signals mediate subcellular targeting of tail-anchored membrane proteins in Saccharomyces cerevisiae.', The Journal of biological chemistry, 278(10), pp. 8219–23. doi: 10.1074/jbc.M212725200.

Boden, M. and Hawkins, J. (2005) 'Prediction of subcellular localization using sequence-biased recurrent networks', Bioinformatics. Oxford University Press, 21(10), pp. 2279–2286. doi: 10.1093/bioinformatics/bti372.

Borgese, N. and Fasana, E. (2011) 'Targeting pathways of C-tail-anchored proteins', Biochimica et Biophysica Acta - Biomembranes. Elsevier B.V., 1808(3), pp. 937–946. doi: 10.1016/j.bbamem.2010.07.010.

Borgese, N. and Righi, M. (2010) 'Remote origins of tail-anchored proteins', Traffic, 11(7), pp. 877–885. doi: 10.1111/j.1600-0854.2010.01068.x.

Bowers, K. et al. (2006) 'Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters', in ACM/IEEE SC 2006 Conference (SC'06). IEEE, pp. 43–43. doi: 10.1109/SC.2006.54.

Bozkurt, G. et al. (2009) 'Structural insights into tail-anchored protein binding and membrane insertion by Get3.', Proceedings of the National Academy of Sciences of the United States of America, 106(50), pp. 21131–21136. doi: 10.1073/pnas.0910223106.

Bozkurt, G. et al. (2010) 'The structure of Get4 reveals an alpha-solenoid fold adapted for multiple interactions in tail-anchored protein biogenesis.', FEBS letters, 584(8), pp. 1509–14. doi: 10.1016/j.febslet.2010.02.070.

Brophy, C. M., Lamb, S. and Graham, A. (1999) 'The small heat shock-related protein-20 is an actin-associated protein.', Journal of vascular surgery, 29(2), pp. 326–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9950990 (Accessed: 28 October 2017).

Chang, Y. W. et al. (2010) 'Crystal structure of Get4-Get5 complex and its interactions with

Sgt2, Get3, and Ydj1', Journal of Biological Chemistry, 285(13), pp. 9962–9970. doi: 10.1074/jbc.M109.087098.

Chartron, J. W. et al. (2010) 'Structural characterization of the Get4/Get5 complex and its interaction with Get3.', Proceedings of the National Academy of Sciences of the United States of America, 107(27), pp. 12127–12132. doi: 10.1073/pnas.1006036107.

Chartron, J. W., Gonzalez, G. M. and Clemons, W. M. (2011) 'A structural model of the Sgt2 protein and its interactions with chaperones and the Get4/Get5 complex.', The Journal of biological chemistry, 286(39), pp. 34325–34. doi: 10.1074/jbc.M111.277798.

Chartron, J. W., VanderVelde, D. G. and Clemons, W. M. (2012) 'Structures of the Sgt2/SGTA Dimerization Domain with the Get5/UBL4A UBL Domain Reveal an Interaction that Forms a Conserved Dynamic Interface', Cell Reports. The Authors, 2(6), pp. 1620–1632. doi: 10.1016/j.celrep.2012.10.010.

Chou, K.-C. and Cai, Y.-D. (2005) 'Prediction of membrane protein types by incorporating amphipathic effects.', Journal of chemical information and modeling, 45(2), pp. 407–13. doi: 10.1021/ci049686v.

Clancy Suzanne, W. B. (2008) 'Translation: DNA to mRNA to Protein', Nature Education, 1(1), p. 101.

Costanzo, M. et al. (2010) 'The genetic landscape of a cell.', Science (New York, N.Y.), 327(5964), pp. 425–31. doi: 10.1126/science.1180823.

Craigon, D. J. et al. (2004) 'NASCArrays: a repository for microarray data generated by NASC's transcriptomics service.', Nucleic acids research. Oxford University Press, 32(Database issue), pp. D575-7. doi: 10.1093/nar/gkh133.

Crick, F. (1970) 'Central dogma of molecular biology.', Nature, 227(5258), pp. 561–3. Available at: http://www.ncbi.nlm.nih.gov/pubmed/4913914 (Accessed: 26 June 2018).

Daniele, L. L. et al. (2016) 'Mutation of wrb, a Component of the Guided Entry of Tail-Anchored Protein Pathway, Disrupts Photoreceptor Synapse Structure and Function.', Investigative ophthalmology & visual science. Association for Research in Vision and

Ophthalmology, 57(7), pp. 2942–54. doi: 10.1167/iovs.15-18996.

Denic, V. (2012) 'A portrait of the GET pathway as a surprisingly complicated young man', Trends in Biochemical Sciences, 37(10), pp. 411–417. doi: 10.1016/j.tibs.2012.07.004.

Denic, V., Dotsch, V. and Sinning, I. (2013) 'Endoplasmic Reticulum Targeting and Insertion of Tail-Anchored Membrane Proteins by the GET Pathway', Cold Spring Harbor Perspectives in Biology. Cold Spring Harbor Laboratory Press, 5(8), pp. a013334–a013334. doi: 10.1101/cshperspect.a013334.

Divol, F. et al. (2013) 'The Arabidopsis YELLOW STRIPE LIKE4 and 6 transporters control iron release from the chloroplast.', The Plant cell, 25(3), pp. 1040–55. doi: 10.1105/tpc.112.107672.

Duncan, O. et al. (2013) 'The outer mitochondrial membrane in higher plants', Trends in Plant Science. Elsevier Ltd, 18(4), pp. 207–217. doi: 10.1016/j.tplants.2012.12.004.

Emanuelsson, O. et al. (2000) 'Predicting subcellular localization of proteins based on their N-terminal amino acid sequence.', Journal of molecular biology, 300(4), pp. 1005–16. doi: 10.1006/jmbi.2000.3903.

Emanuelsson, O. et al. (2007) 'Locating proteins in the cell using TargetP, SignalP and related tools.', Nature protocols, 2(4), pp. 953–71. doi: 10.1038/nprot.2007.131.

Emsley, P. et al. (2010) 'Features and development of Coot', Acta Crystallographica Section D Biological Crystallography, 66(4), pp. 486–501. doi: 10.1107/S0907444910007493.

Evans, P. (2006) 'Scaling and assessment of data quality.', Acta crystallographica. Section D, Biological crystallography, 62(Pt 1), pp. 72–82. doi: 10.1107/S0907444905036693.

Evans, P. R. and Murshudov, G. N. (2013) 'How good are my data and what is the resolution?', Acta crystallographica. Section D, Biological crystallography, 69(Pt 7), pp. 1204–14. doi: 10.1107/S0907444913000061.

Finn, R. D., Attwood, T. K. and Babbitt (2017) 'InterPro in 2017-beyond protein family and domain annotations.', Nucleic acids research, 45(D1), pp. D190–D199. doi: 10.1093/nar/gkw1107.

Forouhar, F. et al. (no date) 'Northeast Structural Genomics Consortium Target NsR300', To be Published. doi: 10.2210/PDB3IGF/PDB.

Friesner, R. A. et al. (2006) 'Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes.', Journal of medicinal chemistry, 49(21), pp. 6177–96. doi: 10.1021/jm051256o.

Götz, S. et al. (2008) 'High-throughput functional annotation and data mining with the Blast2GO suite.', Nucleic acids research, 36(10), pp. 3420–35. doi: 10.1093/nar/gkn176.

Gristick, H. B. et al. (2014) 'Crystal structure of ATP-bound Get3-Get4-Get5 complex reveals regulation of Get3 by Get4.', Nature structural & molecular biology. Nature Publishing Group, 21(5), pp. 437–42. doi: 10.1038/nsmb.2813.

Guay, J. et al. (1997) 'Regulation of actin filament dynamics by p38 map kinase-mediated phosphorylation of heat shock protein 27.', Journal of cell science, 110 (Pt 3), pp. 357–68. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9057088 (Accessed: 28 October 2017).

Gusev, N. B., Bukach, O. V. and Marston, S. B. (2005) 'Structure, Properties, and Probable Physiological Role of Small Heat Shock Protein with Molecular Mass 20 kD (Hsp20, HspB6)', Biochemistry (Moscow). Nauka/Interperiodica, 70(6), pp. 629–637. doi: 10.1007/s10541-005-0162-8.

Hartman, N. T. et al. (2007) 'Proteomic Complex Detection Using Sedimentation', Analytical Chemistry, 79(5), pp. 2078–2083. doi: 10.1021/ac061959t.

Hemmingsson, O. et al. (2010) 'ASNA-1 Activity Modulates Sensitivity to Cisplatin', Cancer Research, 70(24), pp. 10321–10328. doi: 10.1158/0008-5472.CAN-10-1548.

Holm, L. and Rosenström, P. (2010) 'Dali server: conservation mapping in 3D.', Nucleic acids research, 38(Web Server issue), pp. W545-9. doi: 10.1093/nar/gkq366.

Hu, J. et al. (2009) 'The Crystal Structures of Yeast Get3 Suggest a Mechanism for Tail-Anchored Protein Membrane Insertion', PLoS ONE, 4(11). doi: 10.1371/journal.pone.0008061.

Hwang, Y. T. et al. (2004) 'Novel targeting signals mediate the sorting of different isoforms of the tail-anchored membrane protein cytochrome b5 to either endoplasmic reticulum or mitochondria.', The Plant cell. American Society of Plant Biologists, 16(11), pp. 3002–19. doi: 10.1105/tpc.104.026039.

Johnson, N., Powis, K. and High, S. (2013) 'Post-translational translocation into the endoplasmic reticulum', Biochimica et Biophysica Acta - Molecular Cell Research. Elsevier B.V., 1833(11), pp. 2403–2409. doi: 10.1016/j.bbamcr.2012.12.008.

Kabsch, W. (2010) 'XDS', Acta Crystallographica Section D Biological Crystallography. International Union of Crystallography, 66(2), pp. 125–132. doi: 10.1107/S0907444909047337.

Kalbfleisch, T., Cambon, A. and Wattenberg, B. W. (2007) 'A bioinformatics approach to identifying tail-anchored proteins in the human genome', Traffic, 8(12), pp. 1687–1694. doi: 10.1111/j.1600-0854.2007.00661.x.

Kall, L., Krogh, A. and Sonnhammer, E. L. L. (2007) 'Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server', Nucleic Acids Research, 35(Web Server), pp. W429–W432. doi: 10.1093/nar/gkm256.

Kao, G. et al. (2007) 'ASNA-1 Positively Regulates Insulin Secretion in C. elegans and Mammalian Cells', Cell, 128(3), pp. 577–587. doi: 10.1016/j.cell.2006.12.031.

Kiianitsa, K., Solinger, J. A. and Heyer, W.-D. (2003) 'NADH-coupled microplate photometric assay for kinetic studies of ATP-hydrolyzing enzymes with low and high specific activities.', Analytical biochemistry, 321(2), pp. 266–71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14511695 (Accessed: 6 November 2017).

Ko, J. et al. (2012) 'GalaxyWEB server for protein structure prediction and refinement', Nucleic Acids Research. Garland Publishing Inc., New York, 40(W1), pp. W294–W297. doi: 10.1093/nar/gks493.

Kohl, C. et al. (2011) 'Cooperative and independent activities of Sgt2 and Get5 in the targeting of tail-anchored proteins', Biological Chemistry, 392(7), pp. 601–608. doi: 10.1515/BC.2011.066.

Kriechbaumer, V. et al. (2009) 'Subcellular distribution of tail-anchored proteins in arabidopsis', Traffic, 10(12), pp. 1753–1764. doi: 10.1111/j.1600-0854.2009.00991.x.

Krissinel, E. and Henrick, K. (2007) 'Inference of macromolecular assemblies from crystalline state.', Journal of molecular biology, 372(3), pp. 774–97. doi: 10.1016/j.jmb.2007.05.022.

Krogh, A. et al. (2001) 'Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.', Journal of molecular biology, 305(3), pp. 567–80. doi: 10.1006/jmbi.2000.4315.

Kutay, U., Hartmann, E. and Rapoport, T. A. (1993) 'A class of membrane proteins with a C-terminal anchor.', Trends in cell biology, 3(3), pp. 72–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14731773 (Accessed: 25 January 2018).

Laganowsky, A. et al. (2009) 'Hydroponics on a chip: analysis of the Fe deficient Arabidopsis thylakoid membrane proteome.', Journal of proteomics, 72(3), pp. 397–415. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19367733 (Accessed: 6 November 2017).

Lehninger Albert L, David L. Nelson, M. M. C. (2000) Lehninger principles of biochemistry. 6th edn. New York: Worth Publishers.

Leipe, D. D. et al. (2002) 'Classification and evolution of P-loop GTPases and related ATPases', Journal of Molecular Biology, 317(1), pp. 41–72. doi: 10.1006/jmbi.2001.5378.

Liberek, K., Lewandowska, A. and Ziętkiewicz, S. (2008) 'Chaperones in control of protein disaggregation', The EMBO Journal, 27(2), pp. 328–335. doi: 10.1038/sj.emboj.7601970.

Lin, Y.-F., Walmsley, A. R. and Rosen, B. P. (2006) 'An arsenic metallochaperone for an arsenic detoxification pump', Proceedings of the National Academy of Sciences, 103(42), pp. 15617–15622. doi: 10.1073/pnas.0603974103.

Liou, S.-T. and Wang, C. (2005) 'Small glutamine-rich tetratricopeptide repeat-containing protein is composed of three structural units with distinct functions.', Archives of biochemistry and biophysics, 435(2), pp. 253–63. doi: 10.1016/j.abb.2004.12.020.

Manu, M. S. et al. (2018) 'Analysis of tail-anchored protein translocation pathway in plants', Biochemistry and Biophysics Reports, 14, pp. 161–167. doi: 10.1016/j.bbrep.2018.05.001.

Mariappan, M. et al. (2011) 'The mechanism of membrane-associated steps in tail-anchored protein insertion.', Nature. Nature Publishing Group, 477(7362), pp. 61–66. doi: 10.1038/nature10362.

Mateja, A. et al. (2009) 'The structural basis of tail-anchored membrane protein recognition by Get3.', Nature. Nature Publishing Group, 461(7262), pp. 361–366. doi: 10.1038/nature08319.

Mateja, A. et al. (2015) 'Structure of the Get3 targeting factor in complex with its membrane protein cargo', Science, 347(6226), pp. 1152–1155. doi: 10.1126/science.1261671.

McCoy, A. J. et al. (2007) 'Phaser crystallographic software', Journal of Applied Crystallography. International Union of Crystallography, 40(4), pp. 658–674. doi: 10.1107/S0021889807021206.

Metsalu, T. et al. (2015) 'ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap', Nucleic Acids Research. International Society for Optics and Photonics, San Jose, CA, 43(W1), pp. W566–W570. doi: 10.1093/nar/gkv468.

Mock, J.-Y. et al. (2017) 'Structural basis for regulation of the nucleo-cytoplasmic distribution of Bag6 by TRC35.', Proceedings of the National Academy of Sciences of the United States of America, 114(44), pp. 11679–11684. doi: 10.1073/pnas.1702940114.

Mounier, N. and Arrigo, A.-P. (2002) 'Actin cytoskeleton and small heat shock proteins: how do they interact?', Cell stress & chaperones, 7(2), pp. 167–76. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12380684 (Accessed: 3 November 2017).

Mukhopadhyay, R. et al. (2006) 'Targeted disruption of the mouse Asna1 gene results in embryonic lethality', FEBS Letters, 580(16), pp. 3889–3894. doi: 10.1016/j.febslet.2006.06.017.

Murshudov, G. N. et al. (2011) 'REFMAC5 for the refinement of macromolecular crystal structures', Acta Crystallographica Section D Biological Crystallography, 67(4), pp. 355–367. doi: 10.1107/S0907444911001314.

Nałecz, K. A. (no date) '[The 1999 Nobel Prize for physiology or medicine].', Neurologia i neurochirurgia polska, 34(2), pp. 233–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10962717 (Accessed: 4 July 2018).

Norlin, S. et al. (2015) 'Asna1/TRC40 controls beta cell function and ER homeostasis by ensuring retrograde transport', Diabetes, p. db150699. doi: 10.2337/db15-0699.

Nyathi, Y., Wilkinson, B. M. and Pool, M. R. (2013) 'Co-translational targeting and translocation of proteins to the endoplasmic reticulum', Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. Elsevier, 1833(11), pp. 2392–2402. doi: 10.1016/J.BBAMCR.2013.02.021.

Okreglak, V. and Walter, P. (2014) 'The conserved AAA-ATPase Msp1 confers organelle specificity to tail-anchored proteins', Proceedings of the National Academy of Sciences. National Academy of Sciences, 111(22), pp. 8019–8024. doi: 10.1073/pnas.1405755111.

Pedrazzini, E. (2009) 'Tail-Anchored proteins in plants', Journal of Plant Biology, 52(2), pp. 88–101. doi: 10.1007/s12374-009-9014-1.

Petersen, T. N. et al. (2011) 'SignalP 4.0: discriminating signal peptides from transmembrane regions', Nature Methods, 8(10), pp. 785–786. doi: 10.1038/nmeth.1701.

'Protein Function' (2010) in Essentials of Cell Biology. Available at: https://www.nature.com/scitable/topicpage/protein-function-14123348 (Accessed: 26 June 2018).

Rao, S. T. and Rossmann, M. G. (1973) 'Comparison of super-secondary structures in proteins.', Journal of molecular biology, 76(2), pp. 241–56. Available at: http://www.ncbi.nlm.nih.gov/pubmed/4737475 (Accessed: 27 October 2017).

Rauch, J. N. et al. (2017) 'BAG3 Is a Modular, Scaffolding Protein that physically Links Heat Shock Protein 70 (Hsp70) to the Small Heat Shock Proteins', Journal of Molecular

Biology, 429(1), pp. 128–141. doi: 10.1016/j.jmb.2016.11.013.

Sali, A. and Blundell, T. L. (1993) 'Comparative protein modelling by satisfaction of spatial restraints.', Journal of molecular biology, 234(3), pp. 779–815. doi: 10.1006/jmbi.1993.1626.

Sanchita, Dhawan, S. S. and Sharma, A. (2014) 'Analysis of differentially expressed genes in abiotic stress response and their role in signal transduction pathways', Protoplasma. Springer Vienna, 251(1), pp. 81–91. doi: 10.1007/s00709-013-0528-5.

Schuldiner, M. et al. (2008) 'The GET Complex Mediates Insertion of Tail-Anchored Proteins into the ER Membrane', Cell. Elsevier Inc., 134(4), pp. 634–645. doi: 10.1016/j.cell.2008.06.025.

Shen, J. et al. (2003) 'The Saccharomyces cerevisiae Arr4p is involved in metal and heat tolerance.', Biometals : an international journal on the role of metal ions in biology, biochemistry, and medicine, 16(3), pp. 369–78. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12680698 (Accessed: 28 October 2017).

Sherrill, J. et al. (2011) 'A Conserved Archaeal Pathway for Tail-Anchored Membrane Protein Insertion', Traffic. Blackwell Publishing Ltd, 12(9), pp. 1119–1123. doi: 10.1111/j.1600-0854.2011.01229.x.

Shin, W.-H. et al. (2014) 'Prediction of Protein Structure and Interaction by GALAXY Protein Modeling Programs', Bio Design l bdjn.org Bio Design l, 22(11). Available at: http://www.bdjn.org/APP_PDF/BDJN002-01-01.pdf (Accessed: 6 November 2017).

Sprang, S. R. (1997) 'G protein mechanisms: insights from structural analysis.', Annual review of biochemistry, 66(1), pp. 639–78. doi: 10.1146/annurev.biochem.66.1.639.

Srivastava, R. et al. (2017) 'The GET System Inserts the Tail-Anchored Protein, SYP72, into Endoplasmic Reticulum Membranes', Plant Physiology, 173(2), pp. 1137–1145. doi: 10.1104/pp.16.00928.

Suloway, C. J. M. et al. (2009) 'Model for eukaryotic tail-anchored protein binding based on the structure of Get3.', Proceedings of the National Academy of Sciences of the United States of America, 106(35), pp. 14849–14854. doi: 10.1073/pnas.0907522106.

Suloway, C. J., Rome, M. E. and Clemons, W. M. (2012) 'Tail-anchor targeting by a Get3 tetramer: the structure of an archaeal homologue', The EMBO Journal. Nature Publishing Group, 31(3), pp. 707–719. doi: 10.1038/emboj.2011.433.

Suzuki, N. et al. (2014) 'Abiotic and biotic stress combinations', New Phytologist, 203(1), pp. 32–43. doi: 10.1111/nph.12797.

Swamy-Mruthinti, S. et al. (2013) 'Thermal stress induced aggregation of aquaporin 0 (AQP0) and protection by α-crystallin via its chaperone function.', PloS one. Edited by J. Saad, 8(11), p. e80404. doi: 10.1371/journal.pone.0080404.

Tessier, D. J. et al. (2003) 'The small heat shock protein (HSP) 20 is dynamically associated with the actin cross-linking protein actinin.', The Journal of surgical research, 111(1), pp. 152–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12842460 (Accessed: 3 November 2017).

The UniProt Consortium (2017) 'UniProt: the universal protein knowledgebase', Nucleic Acids Research, 45(D1), pp. D158–D169. doi: 10.1093/nar/gkw1099.

Toufighi, K. et al. (2005) 'The Botany Array Resource: e-Northerns, Expression Angling, and promoter analyses.', The Plant journal : for cell and molecular biology, 43(1), pp. 153–63. doi: 10.1111/j.1365-313X.2005.02437.x.

Vogl, C. et al. (2016) 'Tryptophan-rich basic protein (WRB) mediates insertion of the tail-anchored protein otoferlin and is required for hair cell exocytosis and hearing', The EMBO Journal, 35(23), pp. 2536–2552. doi: 10.15252/embj.201593565.

Voth, W. et al. (2014) 'The protein targeting factor Get3 functions as ATP-Independent chaperone under oxidative stress conditions', Molecular Cell. Elsevier Inc., 56(1), pp. 116–127. doi: 10.1016/j.molcel.2014.08.017.

Wang, F. et al. (2014) 'The Get1/2 transmembrane complex is an endoplasmic-reticulum membrane protein insertase.', Nature. Nature Publishing Group, 512(7515), pp. 441–4. doi: 10.1038/nature13471.

Webb, B. and Sali, A. (2016) 'Comparative Protein Structure Modeling Using MODELLER', in Current Protocols in Bioinformatics. Hoboken, NJ, USA: John Wiley & Sons, Inc.,

p. 5.6.1-5.6.37. doi: 10.1002/cpbi.3.

Winn, M. D. et al. (2011) 'Overview of the CCP4 suite and current developments.', Acta crystallographica. Section D, Biological crystallography, 67(Pt 4), pp. 235–42. doi: 10.1107/S0907444910045749.

Wohlever, M. L. et al. (2017) 'Msp1 Is a Membrane Protein Dislocase for Tail-Anchored Proteins', Molecular Cell. Elsevier Inc., 67(2), p. 194–202.e6. doi: 10.1016/j.molcel.2017.06.019.

Wohlever, M. L. et al. (2017) 'Msp1 Is a Membrane Protein Dislocase for Tail-Anchored Proteins', Molecular Cell, 67(2), p. 194–202.e6. doi: 10.1016/j.molcel.2017.06.019.

Xing, S. et al. (2017) 'Loss of GET pathway orthologs in Arabidopsis thaliana causes root hair growth defects and affects SNARE abundance', Proceedings of the National Academy of Sciences, 114(8), pp. E1544–E1553. doi: 10.1073/pnas.1619525114.

Yamagata, A. et al. (2010) 'Structural insight into the membrane insertion of tail-anchored proteins by Get3', Genes to Cells, 15(1), pp. 29–41. doi: 10.1111/j.1365-2443.2009.01362.x.

Ye, J. et al. (2010) 'The 1.4 Å Crystal Structure of the ArsD Arsenic Metallochaperone Provides Insights into Its Interaction with the ArsA ATPase', Biochemistry. American Chemical Society, 49(25), pp. 5206–5212. doi: 10.1021/bi100571r.

Zalisko, B. E. et al. (2017) 'Tail-Anchored Protein Insertion by a Single Get1/2 Heterodimer', Cell Reports, 20(10), pp. 2287–2293. doi: 10.1016/j.celrep.2017.08.035.

List of Publications

- **Manu, M. S.** *et al.* (2018) 'Analysis of tail-anchored protein translocation pathway in plants', *Biochemistry and Biophysics Reports*. Elsevier, 14, pp. 161–167. **doi: 10.1016/j.bbrep.2018.05.001**.

- **Manu, M. S.** and Ramasamy, S. (2017) 'Crystal structure of AtGet3ΔL, a chloroplast Get3 from *Arabidopsis thaliana*', *Acta Crystallographica Section A Foundations and Advances*, 73(a2), pp. C1179–C1179. **doi: https://doi.org/10.1107/s2053273317083954**.

- **Manu, M. S.** *et al.* 'Dual function of plant Get3 ortholog with HSP domain', Manuscript under communication.

- **Manu, M. S.** *et al.* 'Tail-anchored protein targeting pathways in Plants', Review in preparation.

Analysis of tail-anchored protein translocation pathway in plants

M.S. Manu[a,b], Deepanjan Ghosh[a,b], Bhushan P. Chaudhari[a,b], Sureshkumar Ramasamy[a,b,*]

[a] Biochemical Sciences Division, CSIR-National Chemical Laboratory, Pune 411008 India
[b] Academy of Scientific and Innovative Research (AcSIR), CSIR-National Chemical Laboratory, Pune 411008, India

ARTICLE INFO

ABSTRACT

Tail-anchored (TA) proteins are a special class of membrane proteins that carry out vital functions in all living cells. Targeting mechanisms of TA proteins are investigated as the best example for post-translational protein targeting in yeast. Of the several mechanisms, Guided Entry of Tail-anchored protein (GET) pathway plays a major role in TA protein targeting. Many in silico and in vivo analyses are geared to identify TA proteins and their targeting mechanisms in different systems including *Arabidopsis thaliana*. Yet, crop plants that grow in specific and/or different conditions are not investigated for the presence of TA proteins and GET pathway. This study majorly investigates GET pathway in two crop plants, *Oryza sativa* subsp. Indica and *Solanum tuberosum*, through detailed *in silico* analysis. 508 and 912 TA proteins are identified in *Oryza sativa* subsp. Indica and *Solanum tuberosum* respectively and their localization with respect to endoplasmic reticulum (ER), mitochondria, and chloroplast has been delineated. Similarly, the associated GET proteins are identified (Get1, Get3 and Get4) and their structural inferences are elucidated using homology modelling. Get3 models are based on yeast Get3. The cytoplasmic Get3 from *O. sativa* is identified to be very similar to yeast Get3 with conserved P-loop and TA binding groove. Three cytoplasmic Get3s are identified for *S. tuberosum*. Taken together, this is the first study to identify TA proteins and GET components in *Oryza sativa* subsp. Indica and *Solanum tuberosum*, forming the basis for any further experimental characterization of TA targeting and GET pathway mechanisms in crop plants.

1. Introduction

Many integral membrane proteins with several vital functions are present in biological membranes. Among these membrane proteins, tail-anchored (TA) proteins gain importance because of their topology, biogenesis and functionality [1]. Around 5% of the total membrane proteins in eukaryotes are TA proteins. TA proteins are a special class of proteins with a single pass C-terminal trans-membrane domain (TMD) and whole functional N-terminal domain facing towards the cytoplasm [2]. TA proteins are found on several organelle membranes, involving in vesicular trafficking, redox reaction, apoptosis etc [1]. Signal for the TA protein to reach the target is located in the TMD [3]. Specific targeting is also determined by several factors such as overall hydrophobicity, length of TMD, physical and chemical properties of amino acid sequence etc. Since the targeting signal for the TA proteins is located at the C-terminal TMD, the co-translational signal recognition particle (SRP) mediated targeting pathway cannot function properly in this case. Hence, most of the TA proteins are targeted post-translationally. This targeting mechanism can be divided into two, (i) unassisted and (ii) assisted. In unassistant mechanisms, TA proteins do not require any assistant protein to reach the target location. But in the

assisted mechanism, TA proteins require several chaperones to reach the specific target. In general, the assisted mechanism can further be classified into three types (i) SRP mediated (ii) HSP70/90 mediated and (iii) Get3 mediated [4]. The involvement of Get3 in TA protein targeting was identified in recent years by several independent investigations. Most of the TA proteins follow GET pathway to reach their location. One of the well-studied pathway in yeast for efficient TA protein targeting is GET pathway. Get3 in yeast (mammalian homologue TRC40) has sequence similarity with *E. coli* ArsA. ArsA is included in nucleotide binding protein class, SIMIBI (SRP, MinD, BioD) [5]. The GET pathway from yeast composed of several components that include Get1, Get2, Get3, Get4 and Get5. GET pathway gets initiated by the recruitment of sorting complex (sgt2/Get4/Get5) to the TMD of nascent TA proteins. This sorting complex transfers the appropriate TA proteins to Get3 ATPase. Get3 now targets the protein to endoplasmic reticulum (ER) membrane through Get1/Get2 complex [6–9]. Get3 is the major component that connects pre- and post-targeting of TA protein complex.

Attempts to identify TA proteins computationally are done in eukaryotes and prokaryotes [10–14]. Among eukaryotes, plant and animal systems differ mostly in the presence of differential number of

MANU M. S

SENIOR RESEARCH FELLOW

PROFILE

Passionate Researcher with vision of social improvement. Extremely motivated to constantly develop my skills and grow professionally.

CONTACT

📍 Lab 1875,Biochemical Sciences
CSIR-National Chemical Laboratory
Dr.Homibhabha Road,Pune
Maharashtra,India – 411008

✉ manuxtal@gmail.com

☏ +91-9561342439

in /in/manu-m-s/

EDUCATION

Ph.D. in Biological Science
AcSIR,India (Thesis Submitted)
> **Thesis focus :** Structure-Function Characterization of Tail-Anchored Protein Translocation Pathway in Plants

M.Sc. in Biotechnology | 2010
Bharathiar University, India
> **Thesis focus :** Evaluation of the free radical scavenging potential and phytochemical profiling of *Terminalia chebula* Retz.

B.Sc. in Botany and Biotechnology | 2008
Kerala University, India

AWARDS

- CSIR-JRF/NET,India | 2011
- GATE (Biotechnology),India | 2011

REFERENCES

3 to 5 references to support my candidature will be provided on request.

EXPERIENCE

SENIOR RESEARCH FELLOW | AUG 2013- PRESENT
CSIR-National Chemical Laboratory | Pune, India
- Involved in the cloning, purification, crystallization and structural determination of proteins in GET pathway.

JUNIOR RESEARCH FELLOW | AUG 2011- JULY 2013
CSIR-National Chemical Laboratory | Pune, India
- Involved in the cloning, purification and characterization of LmjMAPK4 from *Leishmania major*.

BCIL TRAINEE (R&D) | JAN 2011-JUN 2011
PRAJ MATRIX-THE INNOVATION CENTER | Pune, India
- Involved in the downstream process development of lactic acid production.

SKILLS

Protein X-ray crystallography
> Protein Crystallization | Diffraction and Data collection | Refinement | Structure determination | protein structural analysis

Molecular biology techniques
> Molecular cloning |Site Directed Mutagenesis | Total RNA isolation from plants| cDNA synthesis

Protein purification techniques
> Affinity chromatography | Size Exclussion chromatography | Ion-Exchange Chromatography | Protein refolding | Protein co-expression and pulldown

Protein functional analysis
> ATPase assay | Kinase assay | Co-immuno precipitation | Immuno staining and confocal analysis

Fermentation techniques

Softwares
> CCP4 | Phenix | Coot | Pymol |Modeller | Prime & Desmond | GraphPad Prism | Adobe illustator and Inkscape

CONFERENCES & WORKSHOPS ATTENDED

- 24th congress and General Assembly of the International Union of Crystallography | HICC, Hyderabad, India (2017)
- Computational Biotechnology at the Nanoscale: CCP4 Workshop 2016 | Regional Center For Biotechnology, Faridabad, India (2016)
- 44th National Seminar on Crystallography | IISER, Pune,India (2016)
- National Colloquium on 100 years of Crystallography | SRIBS-KSCSTE, Thiruvananthapuram, India (2013)

PUBLICATIONS

- **M.S. Manu**, D. Ghosh, B.P. Chaudhari, S. Ramasamy, Biochem. Biophys. Reports. 14 (2018) 161–167. doi:10.1016/j.bbrep.2018.05.001.
- **M.S. Manu**, S. Ramasamy, Acta Crystallogr. Sect. A Found. Adv. 73 (2017) C1179–C1179. (Conference paper)
- J. Elias, M.G. Rajesh, N.P. Anish, **M.S. Manu**, I.C. Varkey, Asian J. Res. Chem. 4 (2011) 445–449.