



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

Keywords:

TA proteins
GET pathway
O. sativa
S. tuberosum
Crop plants

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

* Corresponding author at: Biochemical Sciences Division, CSIR-National Chemical Laboratory, Pune 411008, India.
E-mail address: s.ramasamy@ncl.res.in (S. Ramasamy).

compartmentalization due to the organelle variations including chloroplasts. Owing to this difference in the compartmentalization, plant cells are distinct from animal cells. In view of this, TA proteins are analysed in the plant systems in this study. The sequence analysis of bacteria predicted several TA proteins that possibly suggest the presence of TA proteins and its targeting mechanisms in chloroplast and mitochondria. TA proteins associated with the plant cell membrane were recently reviewed in *Arabidopsis thaliana* [13]. Cytochrome b5 (Cb5) is a well-known example of TA proteins. 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 reductase (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.

Recent studies show the function of GET pathway and TA proteins in *A. thaliana*. Yet, 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 study highlights the analysis on 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 *Oryza sativa* subsp. Indica and *Solanum tuberosum* through *in silico* analysis. Predictions of functional and other physiological distribution of TA proteins and transmembrane domain analyses are performed. Also, the identified GET pathway components (cytosolic Get3, Get1 and Get4) have been modelled to explore the TA protein targeting pathway in these crop plants. This is the first study to predict the existence of TA proteins and its targeting pathway in *Oryza sativa* subsp. Indica and *Solanum tuberosum* via detailed *in silico* analysis. Hence, it forms the basis for further experimental characterization and elucidation of TA targeting mechanisms in plant systems.

2. Material and methods

2.1. Choice of plant systems

Major crop plants, *Oryza sativa* subsp. Indica (UniProt Taxon identifiers: 39946) and *Solanum tuberosum* (UniProt Taxon identifiers: 4113) were selected for the *in silico* analyses.

2.2. Identification of TA proteins in selected plants

Complete proteome of *Oryza sativa* subsp. Indica and *Solanum tuberosum* were retrieved from UniProt [15]. TMHMM and Phobius server [16,17] were used to identify proteins with transmembrane domains (TMs) and proteins with single TM were selected (zero or more than 1 TM were rejected). Sequences were reanalysed to find out the protein with single TM at C-terminal within last 50 amino acids. Proteins thus obtained were further analysed using SignalP 4.1, Protein Prowler and TargetP 1.1 servers [18–20]. 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 targeting sequences. All the results were compared and analysed to select proteins that are not targeted by N-terminal signal and non-secretory.

2.3. Functional annotation of TA proteins

Functional annotation of identified TA proteins of *O. sativa* and *S. tuberosum* was done using Blast2GO, a powerful annotation tool [21]. Blast, mapping and annotation of TA proteins were performed

according to Blast2GO instructions. Proteins with similar functions were segregated based on their GO annotations.

2.4. Analysis of predicted TA proteins

The length, molecular weight and amino acid sequence of the predicted TA proteins were retrieved from Uniprot. The TM (Transmembrane) -region was predicted using Phobius and then the TM sequence and TM length was extracted from the protein sequence using R-script. For analysing 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 of the parameters were plotted using R.

2.5. Identification of GET pathway component

GET pathway members of *O. sativa* and *S. tuberosum* were identified by analysing the whole proteome with yeast and *A. thaliana* GET pathway proteins. Both the crop plants have multiple forms of Get3 and localized to different organelles. Get2 and Get5 were not observed in *O. sativa* and *S. tuberosum*. Domains were confirmed by InterPro analysis [22].

2.6. Modelling of GET pathway proteins

Identified GET pathway components were modelled using MODELLER ver.9.17 [23]. Yeast Get3 (PDB ID: 2WOO) was used as template to model all cytoplasmic Get3s of both *O. sativa* and *S. tuberosum* [24]. The cytoplasmic domain of Get1 of both *O. sativa* and *S. tuberosum* were modelled using yeast Get1 (PDB ID: 3ZS8) as template [6]. *Chaetomium thermophilum* Get4 (PDB ID: 3LPZ) and human TRC35 (PDB ID: 6AU8) were used as templates for modelling *O. sativa* Get4 and *S. tuberosum* Get4 respectively [25,26]. Models with lowest discrete optimised protein energy (DOEP) scores were selected and further refined using GalaxyWeb server [27]. Visualization and image processing of all models were performed using Pymol (<http://www.pymol.org/>).

3. Results

3.1. Identification of TA proteins

TA proteins were identified based on its definition. Proteome of *O. sativa* and *S. tuberosum* were downloaded from Uniprot. 37,383 and 53105 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 are 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 secretion sequence were excluded. From 37,383 of total proteins from *O. sativa*, 508 are found to be TA proteins. Similarly for *S. tuberosum*, 912 proteins are found to be TA proteins from a total of 53105 proteins (Table 1, Supplementary File 1 and 2).

Table 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
<i>Oryza sativa</i>	37,383	4379	723	508
subsp. <i>Indica</i>				
<i>Solanum tuberosum</i>	53105	5022	1287	912

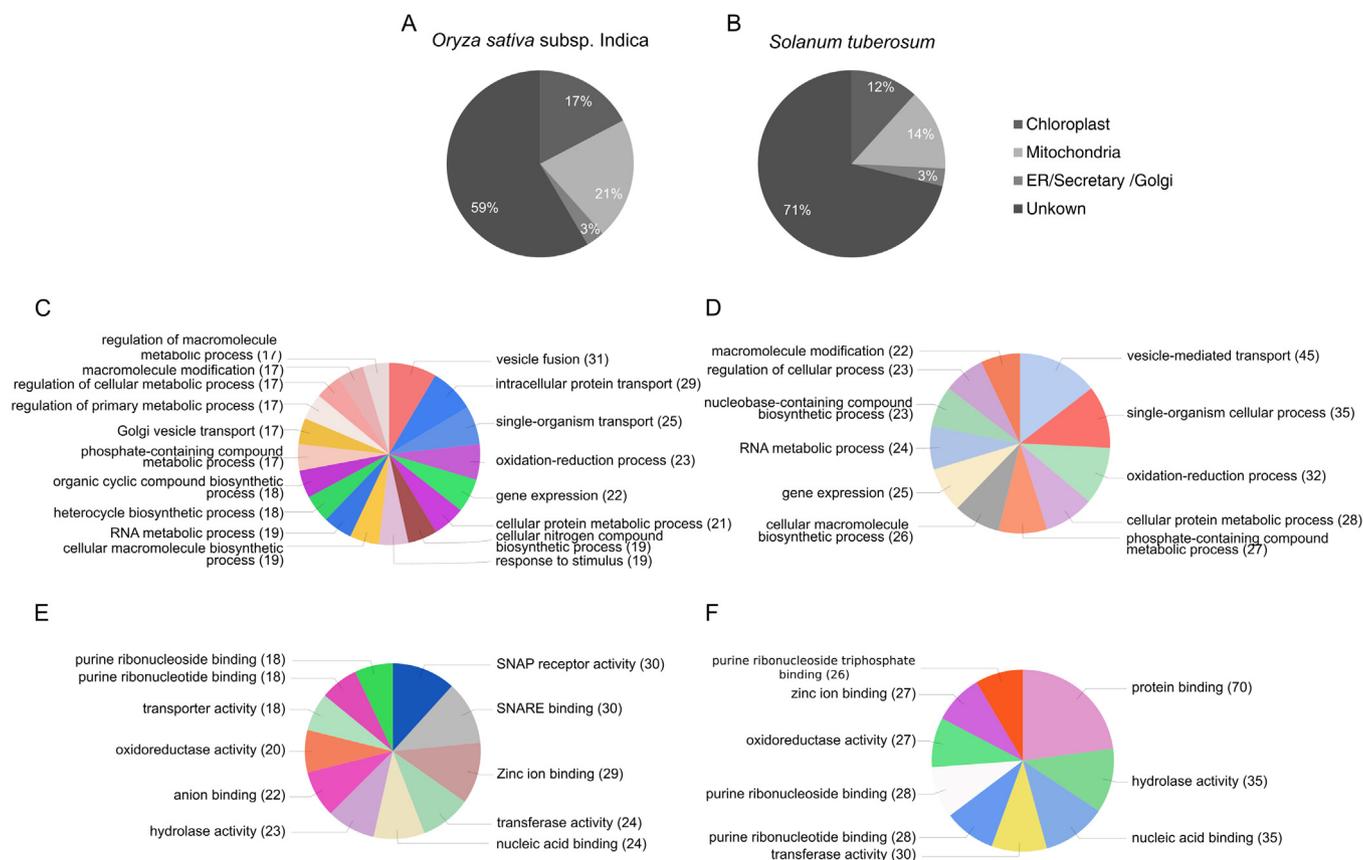


Fig. 1. Organelle and functional distributions of TA proteins. Organelle distribution of (A) *O. sativa* and (B) *S. tuberosum* TA proteins. Biological process distributions of (C) *O. sativa* and (D) *S. tuberosum* TA proteins. Functional distributions of (E) *O. sativa* and (F) *S. tuberosum* TA proteins.

3.2. Organelle distribution of TA proteins

Organelle distribution of TA proteins was analysed for both *O. sativa* and *S. tuberosum* (Fig. 1A and B). In the 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 proteins, the location was unknown. Similarly, in *S. tuberosum*, out of 912 TA proteins, 107 are 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 to be localized to ER. In *O. sativa*, 21% of total TA proteins are localized to mitochondria, that was higher than *S. tuberosum* mitochondrial TA proteins (14%).

3.3. Functional distribution of TA proteins

The identified TA proteins were grouped by functional similarity and their biological process. 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 *O. sativa* TA proteins are shown in Fig. 1C-F. In the functionally known category, SNARE binding proteins were found to be higher for *O. sativa* and protein binding was 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.

3.4. Molecular weight distribution of TA proteins

Most of the TA proteins have molecular weight below 50 kDa. On

comparison, the molecular weight of TA proteins targeted to chloroplast of both *O. sativa* and *S. tuberosum*, have proteins that fall into same molecular weight scale. But in the case of mitochondrial TA proteins, *S. tuberosum* has proteins that have less molecular weights compared to *O. sativa* (Fig. 2A).

3.5. Transmembrane domain analysis of TA proteins

In case of TA proteins, the transmembrane domain is the major determinant factor for their location. These transmembrane domains of TA proteins were analysed for their length, hydrophobicity and amino acid frequencies (Fig. 2 B-D, Fig. 3).

3.6. Length distribution of TMD in TA protein

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 cytoplasmic TA proteins. But in the case of *S. tuberosum*, TA proteins belonging to ER has the maximum variation in their length (Fig. 2B).

3.7. Hydrophobicity of TMD in TA proteins

Hydrophobicity of TMD and TA proteins of both *O. sativa* and *S. tuberosum* were estimated based on Kyte-Doolittle scores. Average hydrophobicity of TMD was found higher for *O. sativa* TA proteins. 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 (Fig. 2C and D).

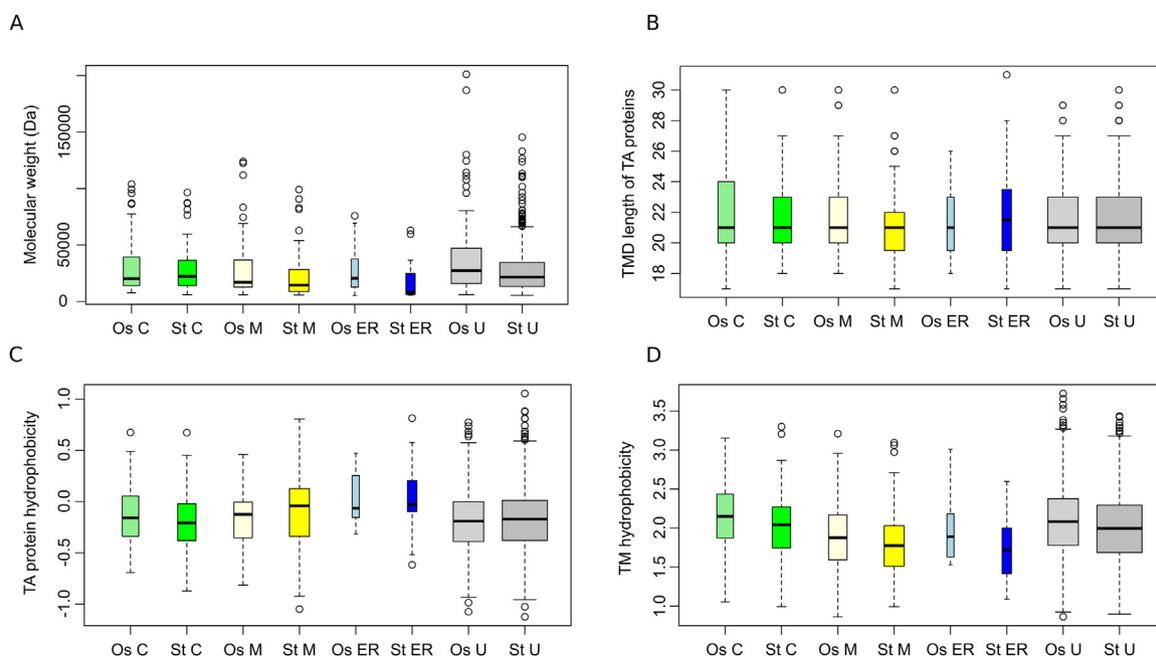


Fig. 2. Organelle wise analysis of identified TA proteins. TA protein's (A) molecular weight distribution, (B) trans-membrane length distribution, (C) hydrophobicity distribution and (D) trans-membrane hydrophobicity distributions were analysed and represented in box plot. Analyses of TA proteins were carried out in organelle wise. Os C - *O. sativa* chloroplast; Os M - *O. sativa* mitochondria; Os ER - *O. sativa* endoplasmic reticulum; Os U - *O. sativa* unknown location. St C - *S. tuberosum* chloroplast; St M - *S. tuberosum* mitochondria; St ER - *S. tuberosum* endoplasmic reticulum, St U - *S. tuberosum* unknown location.

3.8. Amino acid frequency in TMD of TA protein

The global amino acid frequency for TMD of TA proteins was analysed. The frequency of leucine was found higher compared to other amino acids. Valine, isoleucine, alanine and phenylalanine were also found in major proportion (Fig. 3).

3.9. GET components in *O. sativa* and *S. tuberosum*

GET pathway is well explored in yeast system. In case of plants, some recent studies were conducted in *Arabidopsis thaliana*. Compared to other organisms, plant species have several orthologues for Get3. Also, several GET Pathway members are not yet characterized. These details were not investigated in crop plants. The Pfam analysis shows that *O. sativa* has three Get3 orthologues under three genes. Similarly,

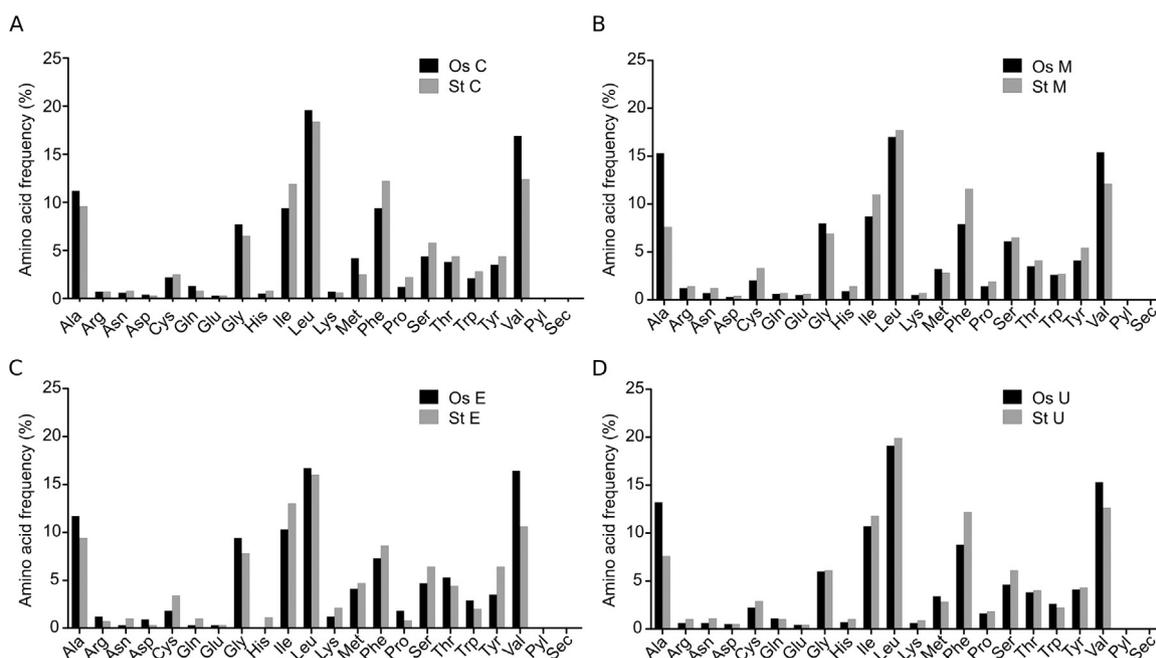


Fig. 3. Global amino acid frequency analysis of transmembrane domain of *O. sativa* and *S. tuberosum* TA proteins. Amino acid frequency in transmembrane domain of TA proteins present in (A) chloroplast, (B) mitochondria, (C) endoplasmic reticulum and (D) unknown locations are shown. Os C - *O. sativa* chloroplast; Os M - *O. sativa* mitochondria; Os ER - *O. sativa* endoplasmic reticulum; Os U - *O. sativa* unknown location. St C - *S. tuberosum* chloroplast; St M - *S. tuberosum* mitochondria; St ER - *S. tuberosum* endoplasmic reticulum, St U - *S. tuberosum* unknown location.

Table 2
Get3 orthologues of crop plants, *O. sativa* and *S. tuberosum*.

Organism	Gene	Get3	Location
<i>Oryza sativa</i>	OSI_32081	B8BDK7	Cytoplasm/ER
	OSI_08661	A2 × 8V0	Chloroplast
	OSI_08919	B8AIG1	Mitochondria
<i>Solanum tuberosum</i>	102593399	M1AE77	Mitochondria
	102598603	M1A9 × 9	Cytoplasm/ER
	102598324	M0ZFY4	Cytoplasm/ER
	102580025	M1AND2	Cytoplasm/ER
	102581371	M0ZJQ4	Chloroplast

in *S. tuberosum*, five Get3 sequences were present under five genes. Target locations of these Get3s were predicted using the TargetP server (Table 2). 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 (M1A9 × 9, M0ZFY4 and M1AND2) are present for targeting ER TA proteins. Both species have Get3 specifically for chloroplast and mitochondrial TA protein targeting. Besides Get3, Get1 and Get4 are also present in both species. But Get2 and Get5 were not found. The complete details of GET pathway members identified for both the plant species are given in Table 3.

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

To extend the understanding 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* (Fig. 4). *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 template for all the cytoplasmic Get3s. The structure of *O. sativa* cytoplasmic Get3 is much similar to yeast Get3 except at the C-terminal (Fig. 4A). In *S. tuberosum*, out of three cytoplasmic Get3s, M1AND2 and M0ZFY4 are similar to yeast Get3 (Fig. 4B). But M1A9 × 9 has a loop between β3 and α4 that is absent in yeast and other Get3s of *S. tuberosum*. Also, M1A9 × 9 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* (Fig. 4C). In this analysis, yeast Get1 structure (PDB ID: 3ZS8) was used as template for model generation. Both signal sequence and trans-membrane domain were excluded 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

Table 3
GET pathway members of *O. sativa* and *S. tuberosum* are compared with yeast. All proteins are represented by their Uniprot ID.

Component	Yeast	<i>Oryza sativa</i>	<i>Solanum tuberosum</i>
Get1	YGL020C	A2YAC8	M0ZME5
Get2	YER083C	–	–
Get3	YDL100C	–	M1AE77
		B8BDK7	M1A9 × 9
		A2 × 8V0	M0ZFY4
		B8AIG1	M1AND2
Get4	YOR164C	–	M0ZJQ4
		–	M1ACL9
Get5	Q12285	–	–

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 (Fig. 4D). *Chaetomium thermophilum* Get4 (PDB ID: 3LPZ) and human TRC35 (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 that connects α5 and α6 is 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.

4. Discussion

Though several computational and experimental studies are available to identify TA protein targeting pathways in different organisms including plant model of *A. thaliana*, there are no or very less report of detailed studies on crop plant available till date. TA proteins are distributed to almost all organelles of the cell. The presence of extra organelles in plants compared to other phyla makes the TA protein targeting more complex. In this study, we have identified and delineated the TA protein distribution and their GET components in two crop plants, *Oryza sativa* subsp. Indica and *Solanum tuberosum*. Though these selected crop plant species have more than 35,000 total proteins, the TA protein content of these plants accounts for less than 2%. The average length of TMD of TA proteins is predicted as 21 amino acids. Yet, the TMD length shows greater variations depending on the respective organelle membrane highlighting its organelle specificity.

Hydrophobicity pattern, amino acid composition and post-translational modification are also predicted to be influenced by the organelle specificity of TA proteins. In general, plants have more than one Get3s that are specific to different organelles. This is prevalent in our study and the number of Get3s also differs across the two plant species studied. *O. sativa* has three Get3s, while *S. tuberosum* has five Get3s as predicted by our analysis. Three out of five predicted Get3s are specific for ER TA protein targeting in *S. tuberosum*. 85% identify is observed between M0ZFY4 and M1AND2. But from modelling studies, M1A9 × 9 is different compared to both M0ZFY4 and M1AND2, having a loop between β3 and α4. The percentage existence of TA proteins in ER is more or less the same for both the plant species, yet the number of Get3s differ across them. This needs further experimental validation especially in case of *S. tuberosum* where three Get3s are predicted for ER. Besides these cytoplasmic 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 study thus highlights the use of such predictive analyses in identifying the existence of TA proteins and therefore necessitates further experimental characterization of GET mediated TA protein targeting mechanisms.

Acknowledgements

MSM and DG are thankful to Council of Scientific and Industrial Research CSIR, New Delhi India for providing the Senior Research Fellowship. SKR acknowledges Department of Science and Technology (DST), India for Ramanujan Fellowship. This Project GAP315226 is supported by Department of Biotechnology (DBT) India.

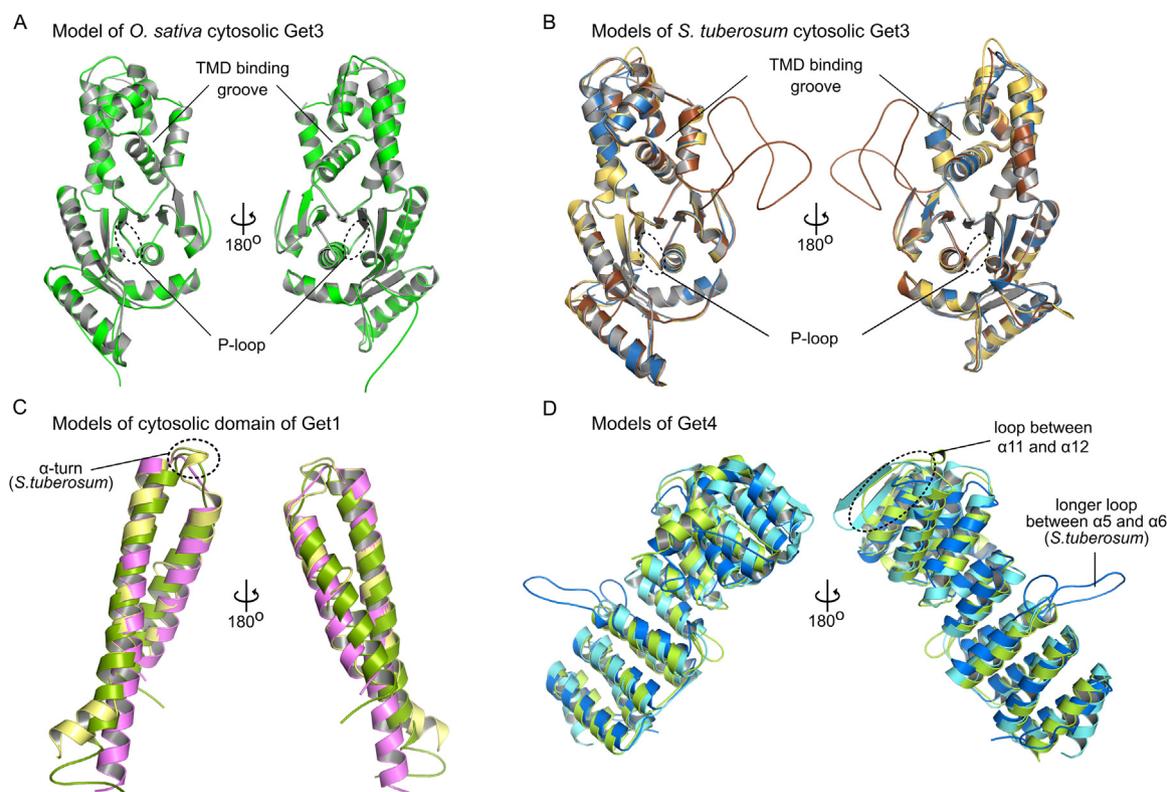


Fig. 4. Structural models of identified GET pathway members of *O. sativa* and *S. tuberosum*. (A) Model of *O. sativa* cytosolic Get3 (green) superposed with yeast Get3 (gray). (B) Models of *S. tuberosum* cytosolic Get3s M1AND2 (sky blue), M1A9x9 (brown) and MOZFY4 (yellow orange) superposed with yeast 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).

Author contributions

DG carried out the analysis of predicted TA proteins. MSM performed the GET pathway component analysis and modelling. SKR conceived the idea. MSM and SKR wrote the manuscript with input from all authors.

Appendix A. Transparency document

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.bbrep.2018.05.001>.

Appendix B. Supplementary material

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.bbrep.2018.05.001](https://doi.org/10.1016/j.bbrep.2018.05.001).

References

- [1] N. Borgese, S. Colombo, E. Pedrazzini, The tale of tail-anchored proteins: coming from the cytosol and looking for a membrane, *J. Cell Biol.* 161 (2003) 1013–1019, <http://dx.doi.org/10.1083/jcb.200303069>.
- [2] U. Kutay, E. Hartmann, T.A. Rapoport, A class of membrane proteins with a C-terminal anchor, *Trends Cell Biol.* 3 (1993) 72–75 <http://www.ncbi.nlm.nih.gov/pubmed/14731773> (accessed 25 January 2018).
- [3] Y.T. Hwang, S.M. Pelitire, M.P.A. Henderson, D.W. Andrews, J.M. Dyer, R.T. Mullen, Novel targeting signals mediate the sorting of different isoforms of the tail-anchored membrane protein cytochrome b5 to either endoplasmic reticulum or mitochondria, *Plant Cell.* 16 (2004) 3002–3019, <http://dx.doi.org/10.1105/tpc.104.026039>.
- [4] N. Borgese, E. Fasana, Targeting pathways of C-tail-anchored proteins, *Biochim. Biophys. Acta - Biomembr.* 1808 (2011) 937–946, <http://dx.doi.org/10.1016/j.bbmem.2010.07.010>.
- [5] D.D. Leippe, Y.I. Wolf, E.V. Koonin, L. Aravind, Classification and evolution of P-loop GTPases and related ATPases, *J. Mol. Biol.* 317 (2002) 41–72, <http://dx.doi.org/10.1006/jmbi.2001.5378>.
- [6] M. Mariappan, A. Mateja, M. Dobosz, E. Bove, R.S. Hegde, R.J. Keenan, The mechanism of membrane-associated steps in tail-anchored protein insertion, *Nature* 477 (2011) 61–66, <http://dx.doi.org/10.1038/nature10362>.
- [7] Y.W. Chang, Y.C. Chuang, Y.C. Ho, M.Y. Cheng, Y.J. Sun, C.D. Hsiao, C. Wang, Crystal structure of Get4-Get5 complex and its interactions with Sgt2, Get3, and Ydj1, *J. Biol. Chem.* 285 (2010) 9962–9970, <http://dx.doi.org/10.1074/jbc.M109.087098>.
- [8] H.B. Gristick, M. Rao, J.W. Chartron, M.E. Rome, S.-O. Shan, W.M. Clemons, Crystal structure of ATP-bound Get3-Get4-Get5 complex reveals regulation of Get3 by Get4, *Nat. Struct. Mol. Biol.* 21 (2014) 437–442, <http://dx.doi.org/10.1038/nsmb.2813>.
- [9] F. Wang, C. Chan, N.R. Weir, V. Denic, The Get1/2 transmembrane complex is an endoplasmic-reticulum membrane protein insertase, *Nature* 512 (2014) 441–444, <http://dx.doi.org/10.1038/nature13471>.
- [10] T. Beilharz, B. Egan, P.A. Silver, K. Hofmann, T. Lithgow, Bipartite signals mediate subcellular targeting of tail-anchored membrane proteins in *Saccharomyces cerevisiae*, *J. Biol. Chem.* 278 (2003) 8219–8223, <http://dx.doi.org/10.1074/jbc.M212725200>.
- [11] T. Kalbfleisch, A. Cambon, B.W. Wattenberg, A bioinformatics approach to identifying tail-anchored proteins in the human genome, *Traffic* 8 (2007) 1687–1694, <http://dx.doi.org/10.1111/j.1600-0854.2007.00661.x>.
- [12] E. Pedrazzini, Tail-anchored proteins in plants, *J. Plant Biol.* 52 (2009) 88–101, <http://dx.doi.org/10.1007/s12374-009-9014-1>.
- [13] V. Kriebchaumer, R. Shaw, J. Mukherjee, C.G. Bowsher, A.M. Harrison, B.M. Abell, Subcellular distribution of tail-anchored proteins in *Arabidopsis*, *Traffic* 10 (2009) 1753–1764, <http://dx.doi.org/10.1111/j.1600-0854.2009.00991.x>.
- [14] N. Borgese, M. Righi, Remote origins of tail-anchored proteins, *Traffic* 11 (2010) 877–885, <http://dx.doi.org/10.1111/j.1600-0854.2010.01068.x>.
- [15] The UniProt Consortium, UniProt: the universal protein knowledgebase, *Nucleic Acids Res.* 45 D158–D169. doi:10.1093/nar/gkw1099, 2017.
- [16] A. Krogh, B. Larsson, G. von Heijne, E.L. Sonnhammer, Predicting transmembrane protein topology with a hidden markov model: application to complete genomes, *J. Mol. Biol.* 305 (2001) 567–580, <http://dx.doi.org/10.1006/jmbi.2000.4315>.
- [17] L. Kall, A. Krogh, E.L.L. Sonnhammer, Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server, *Nucleic Acids Res.* 35 (2007) W429–W432, <http://dx.doi.org/10.1093/nar/gkm256>.
- [18] T.N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions, *Nat. Methods* 8 (2011) 785–786, <http://dx.doi.org/10.1038/nmeth.1701>.
- [19] O. Emanuelsson, S. Brunak, G. von Heijne, H. Nielsen, Locating proteins in the cell using TargetP, SignalP and related tools, *Nat. Protoc.* 2 (2007) 953–971, <http://dx.doi.org/10.1038/nprot.2007.131>.

- [20] M. Boden, J. Hawkins, Prediction of subcellular localization using sequence-biased recurrent networks, *Bioinformatics* 21 (2005) 2279–2286, <http://dx.doi.org/10.1093/bioinformatics/bti372>.
- [21] S. Götz, J.M. García-Gómez, J. Terol, T.D. Williams, S.H. Nagaraj, M.J. Nueda, M. Robles, M. Talón, J. Dopazo, A. Conesa, High-throughput functional annotation and data mining with the Blast2GO suite, *Nucleic Acids Res.* 36 (2008) 3420–3435, <http://dx.doi.org/10.1093/nar/gkn176>.
- [22] R.D. Finn, T.K. Attwood, Babbitt, InterPro in 2017-beyond protein family and domain annotations, *Nucleic Acids Res.* 45 (2017) D190–D199, <http://dx.doi.org/10.1093/nar/gkw1107>.
- [23] B. Webb, A. Sali, Comparative protein structure modeling using MODELLER, *Current Protocols in Bioinformatics*, John Wiley & Sons, Inc, Hoboken, NJ, USA, 2016, , <http://dx.doi.org/10.1002/cpbi.3>.
- [24] A. Mateja, A. Szlachcic, M.E. Downing, M. Dobosz, M. Mariappan, R.S. Hegde, R.J. Keenan, The structural basis of tail-anchored membrane protein recognition by Get3, *Nature* 461 (2009) 361–366, <http://dx.doi.org/10.1038/nature08319>.
- [25] G. Bozkurt, K. Wild, S. Amlacher, E. Hurt, B. Dobberstein, I. Sinning, The structure of Get4 reveals an alpha-solenoid fold adapted for multiple interactions in tail-anchored protein biogenesis, *FEBS Lett.* 584 (2010) 1509–1514, <http://dx.doi.org/10.1016/j.febslet.2010.02.070>.
- [26] J.-Y. Mock, Y. Xu, Y. Ye, W.M. Clemons, Structural basis for regulation of the nucleo-cytoplasmic distribution of Bag6 by TRC35, *Proc. Natl. Acad. Sci. USA.* 114 (2017) 11679–11684, <http://dx.doi.org/10.1073/pnas.1702940114>.
- [27] J. Ko, H. Park, L. Heo, C. Seok, GalaxyWEB server for protein structure prediction and refinement, *Nucleic Acids Res.* 40 (2012) W294–W297, <http://dx.doi.org/10.1093/nar/gks493>.