

A Bioinformatics Approach for Identification of Type III Signal Anchored Proteins in Rice

Amit Katiyar^{1,2}, Shuchi Smita^{1,2}, Dev Mani Pandey², Viswanathan Chinnusamy³ and Kailash Chander Bansal^{1*}

¹National Research Center on Plant Biotechnology, Indian Agricultural Research Institute Campus, New Delhi-110012, India

²Department of Biotechnology, Birla Institute of Technology, Mesra, Ranchi 835 215, Jharkhand, India

³Division of Plant Physiology, Indian Agricultural Research Institute, New Delhi-110012, India

* Corresponding author

Email: kailashbansal@hotmail.com

Phone: 91-11-25843554; Tele Fax: 91-11-25843554

Abstract - Single-pass transmembrane protein (type II, III, and IV) possessing a membrane-spanning domain which targets the protein to the endoplasmic reticulum (ER) membrane. In both type II and III membrane proteins, a single membrane-spanning domain serves both as a signal to initiate insertion and as a membrane anchor. These signal anchor sequences may direct membrane insertion with either an $N_{\text{cyt}}/C_{\text{exo}}$ or $N_{\text{exo}}/C_{\text{cyt}}$ orientation. This study focused on type III proteins, which possess single-anchor sequence but not having N-terminal signal peptide by definition. Type III proteins have the cluster of positively charged residues on the C-terminal side of the signal anchor. The distribution of charged residues flanking the hydrophobic core of the signal sequences play important role in the orientation of signal anchor proteins in membrane. However, the mechanism by which a signal-anchor sequence adopts a particular orientation is still unknown. Here, we performed genome wide screening to identify number of signal anchor proteins in rice genome, which will help to understand the general mechanism of protein orientation in type III membrane proteins.

Keywords: Single-Pas Membrane Protein, Type III Protein, Signal Anchor, Rice

1. Introduction

Subcellular protein sorting, in which proteins travel to their functional organelle within a cell, is an essential feature of cellular life. Typically, protein sorting depends on 'signal' content encoded in their primary structure of the transmembrane proteins. It contains number of hydrophobic and hydrophilic region or domain, which are exposed on one or both sides of the membrane. Single and multiple

membrane-spanning domains containing protein are known as single-pas and multi-pas protein, respectively. Two orientations of signal sequences (NH₂-terminal cleaved or uncleaved signal sequences), have been recognized, which can direct single-spanning membrane proteins to the endoplasmic reticulum (ER) [1]. The NH₂-terminal signal sequences are found on both secreted and membrane proteins [2] and cleaved from the protein by signal peptidase during its translocation across the ER membrane. Second classes of proteins which possess an uncleaved signal sequence target the protein to the ER and stably anchor the protein into the membrane [3, 4]. These proteins are known as signal-anchor (SA) protein and are differ from proteins with a cleaved signal sequence [5]. Two different kinds of orientations ($N_{\text{cyt}}/C_{\text{exo}}$ and $N_{\text{exo}}/C_{\text{cyt}}$), can be possible for single-spanning membrane proteins (Type II, III and IV) during protein targeting towards the ER membrane. In type II proteins the orientation is a luminal C terminus and a cytosolic N terminus, whereas in type III proteins, the orientation is a luminal N terminus and a cytosolic C terminus, which is just the opposite [5, 6]. The type III proteins contain single-anchor sequence but they have lack of N-terminal signal peptide, just like type II proteins. The cluster of positively charged amino acids in type II and III proteins, generally found adjacent to the N-terminal side and on the C-terminal side of the signal anchor sequence, respectively. These positively charged residues change the orientation of proteins in membrane by an uncertain mechanism. Therefore, the identification of new signal anchor proteins helps to determine the biological function and the mechanism of protein orientation. In this study, we did genome wide screening to produce a catalogue of putative signal anchor proteins encoded by the rice genome.

2. METHOD

2.1 Identification of Signal Anchored Proteins

Rice protein sequences for all 12 chromosomes were obtained from “The Institute of Genomic Research (TIGR)” database-version 6.1 (<http://rice.plantbiology.msu.edu/>) [7]. The server TMHMM, Version 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) was used to predict transmembrane domain in rice protein sequences. The proteins containing single transmembrane domain within 50 amino acid C-terminal were collected from manual eye inspection. Further, proteins not having N-terminal signal peptides were collected by using SignalP version 3.0 (<http://www.cbs.dtu.dk/services/SignalP/>). The parameters for SignalP v3.0 [8] tool were set as follows: eukaryotes, neural networks and Hidden Markov Model; truncated to first 70 residues. The overall strategy to predict signal anchor proteins were based on *In-silico* approach describe in Figure 1.

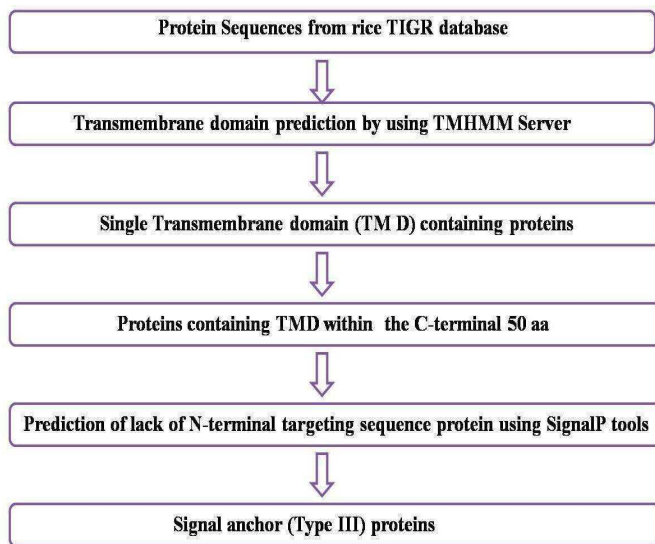


Figure 1: Shows the flow chart of obtaining signal anchor type III proteins.

2.2 Protein Localization and Functional Analysis

The molecular function of predicted signal anchor proteins were analyzed by using TIGR GOSlim Assignments (http://rice.plantbiology.msu.edu/downloads_gad.shtml) and AmiGO blast search (<http://amigo.geneontology.org/cgi-bin/>). The functional annotation was based on TIGR annotation release 6.1 and blast search against *Arabidopsis*, while subcellular localization was based on validated experimental data available on TIGR.

3. Results and Discussion

3.1 Genome-wide identification of Signal Anchored Proteins in rice

Single-spanning transmembrane proteins which contain signal anchor sequence but lack N-terminal peptide sequence are known as type III signal anchor protein (also known as reverse signal anchor) [9]. The function of a signal sequence was reported as targeting to the membrane, membrane insertion and translocation (secreted proteins) or retention (SA proteins). A bioinformatics approach has been previously applied to identify various transmembrane, tail anchored proteins in human, yeast and *Arabidopsis*. Here, we performed genome wide identification of single-pass signal anchored proteins (Type III) in rice using various computation tools. Signal anchor proteins in plants helps to understand general mechanisms about the changes occurred in the protein orientation in membrane. Thus, catalogue preparation of signal anchor proteins encoded by the rice genome is an important step to unravel the biological function. As a first step, we obtained 56,797 protein sequences from TIGR release 6.1 databases. Further, we identified proteins contained single transmembrane domain (TMD) by using transmembrane helix prediction server and we found 5,317 protein members. The next step descends the protein sequence up to 9, 36 as we extracted only those protein members having transmembrane domain within the C-terminal 50 residues. This step identified number of proteins with single TMD near to C-terminus which further required to process for knowing N-terminal signal peptide and if found discarded from list. Remaining 54 protein members contain single transmembrane helix (TMhelix) and lack N-terminal signal peptides are known as signal anchor proteins (Type III) and collected as signal anchor protein catalogue for rice. The gene ontology provide the molecular function of signal anchor proteins and we observed that majority of the protein members were involved in membrane protein transporter activities (Table 1). This exercise will help to understand biological functions of SA proteins in more detail.

Table 1: Shows the molecular function and localization for signal anchored proteins in rice

Rice TIGR Locus	Protein Description	GO-Molecular Function	GO- ID	GO-Cellular Component
LOC_Os01g03850	Hypothetical protein	Unknown	Unknown	Unknown
LOC_Os01g06750	verticillium wilt disease resistance protein precursor,	Nucleotide binding	GO:0000166	Plasma membrane
LOC_Os01g09270	Hypothetical protein	Unknown	Unknown	Unknown
LOC_Os01g15029	Expressed protein	Peptidyl-prolyl cis-trans isomerase activity	GO:0003755	Unknown
LOC_Os01g38070	Retrotransposon protein	Unknown	Unknown	Unknown
LOC_Os01g38510	protein transport protein SEC61 subunit beta	Transporter activity	GO:0005215	Unknown
LOC_Os01g41860	Hypothetical protein	Unknown	Unknown	Unknown
LOC_Os01g51240	Hypothetical protein	Unknown	Unknown	Unknown
LOC_Os01g54424	Expressed protein	Unknown	Unknown	Unknown
LOC_Os01g55860	Expressed protein	Unknown	Unknown	Mitochondrial
LOC_Os01g67960	GPI transamidase component	Unknown	Unknown	Endoplasmic reticulum
LOC_Os02g03880	Hypothetical protein	Protein transmembrane transporter activity	GO:0015450	Mitochondrial
LOC_Os02g08180	protein transport protein SEC61 subunit gamma	Protein transmembrane transport activity	GO:0006810	Unknown
LOC_Os02g10370	hrpN-interacting protein from Malus	Unknown	Unknown	Unknown
LOC_Os02g11705	Expressed protein	Unknown	Unknown	Unknown
LOC_Os02g17590	Hypothetical protein	Unknown	Unknown	Unknown
LOC_Os02g29400	Expressed protein	Protein transmembrane transporter activity	GO:0015450	Mitochondrial
LOC_Os02g32009	Expressed protein	Unknown	Unknown	Unknown
LOC_Os02g35610	Expressed protein	NADH dehydrogenase (ubiquinone) activity	GO:0005739	Mitochondrial
LOC_Os02g46830	Expressed protein	Unknown	Unknown	Unknown
LOC_Os03g02620	Hypothetical protein	Unknown	Unknown	Unknown
LOC_Os03g10160	Pentatricopeptide repeat-containing protein	Structural constituent of ribosome	GO:0003735	Chloroplast
LOC_Os03g12970	Hypothetical protein	Unknown	Unknown	Unknown
LOC_Os03g14334	Expressed protein	Unknown	Unknown	Unknown
LOC_Os03g38359	Expressed protein	Unknown	Unknown	Unknown
LOC_Os03g56784	Hypothetical protein	Unknown	Unknown	Unknown
LOC_Os04g44760	Hypothetical protein	Carbohydrate transmembrane transporter activit	GO:0015144	Membrane
LOC_Os04g50780	Expressed protein	Unknown	Unknown	Unknown
LOC_Os05g28020	Hypothetical protein	Unknown	Unknown	Unknown
LOC_Os05g30830	Retrotransposon protein	Unknown	Unknown	Chloroplast
LOC_Os05g42010	ubiquinol-cytochrome c reductase complex	Unknown	GO:0005739	Mitochondrial
LOC_Os05g50654	Mitochondrial import receptor subunit TOM7-1	Transporter activity	GO:0005215	Unknown
LOC_Os06g14340	Hypothetical protein	Unknown	Unknown	Unknown
LOC_Os06g15450	Retrotransposon protein	Iron ion binding	GO:0005506	Membrane
LOC_Os06g16620	Hypothetical protein	Unknown	Unknown	Unknown
LOC_Os06g21420	Retrotransposon protein	Unknown	Unknown	Unknown
LOC_Os06g23330	Conserved hypothetical protein	Unknown	Unknown	Unknown

LOC_Os06g44374	protein transport protein SEC61 subunit gamma	Protein transmembrane transport activity	GO:0006810	Unknown
LOC_Os07g04270	Hypothetical protein	Unknown	Unknown	Unknown
LOC_Os07g05084	Expressed protein	Transferase activity	GO:0016740	Unknown
LOC_Os07g31194	Expressed protein	Unknown	Unknown	Unknown
LOC_Os07g48244	Ubiquinol-cytochrome c reductase protein	Unknown	Unknown	Mitochondrial
LOC_Os08g03980	Retrotransposon protein	Unknown	Unknown	Unknown
LOC_Os08g06260	Hypothetical protein	Unknown	Unknown	Unknown
LOC_Os08g06270	Hypothetical protein	Unknown	Unknown	Unknown
LOC_Os08g13120	Retrotransposon protein	Unknown	Unknown	Unknown
LOC_Os08g41680	Expressed protein	Unknown	Unknown	Unknown
LOC_Os09g06930	Hypothetical protein	Unknown	Unknown	Unknown
LOC_Os09g25130	Hypothetical protein	Unknown	Unknown	Unknown
LOC_Os09g29640	Hypothetical protein	Unknown	Unknown	Unknown
LOC_Os10g40010	Expressed protein	Binding activity	GO:0005488	Chloroplast
LOC_Os11g41875	Membrane protein	Unknown	Unknown	Unknown
LOC_Os12g15390	Hypothetical protein	Unknown	Unknown	Unknown
LOC_Os12g32950	Membrane protein	Unknown	Unknown	Unknown

4. Conclusions

Our work provides the list of type III signal anchor proteins from rice genome that not having N-terminal targeting peptide but contain signal sequence and involved in protein targeting to the endoplasmic reticulum (ER). Gene ontology of SA proteins facilitates to understand molecular function of SA proteins. This work forms the foundation for molecular genetic and biochemical analysis that will help understand of the biological function of Type III signal anchor proteins in rice.

5. Acknowledgements

AK thanks to Indian Council of Agricultural Research (ICAR) for supporting this work by the ICAR-sponsored Network Project on Transgenic in Crop (NPTC).

6. References

[1] Wickner, W. T., and H. F. Lodish (1985) Multiple mechanisms of protein insertion into and across membranes. *Science (Wash. DC)* **230**: 400-407.

[2] Walter, P., and V. R. Lingappa (1986) Mechanism of protein translocation across the endoplasmic reticulum membrane. *Annu. Rev. Cell Biol.* **2**: 499-516.

[3] Spiess, M. and Lodish, H. F. (1986) An internal signal sequence: the asialoglycoprotein receptor membrane anchor. *Cell*, **44**: 177-185.

[4] Zerial, M., P. Melancon, C. Schneider, and H. Garoff (1986) The transmembrane segment of the human transferrin receptor functions as a signal peptide. *EMBO (Eur. Mol. Biol. Organ.) J.* **5**: 1543-1550.

[5] Lipp, J., and B. Dobberstein (1988) Signal and membrane anchor function overlap in the type II membrane protein 13, CAT. *J. Cell Biol.* **106**: 1813-1820.

[6] Holland, E. C., J. O. Leung, and K. Drickamer: Rat liver asialoglycoprotein receptor lacks a cleavable N-terminal signal sequence (1984) *Proc. Natl. Acad. Sci. USA*, **81**: 7338-7342.

[7] Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek R L, Lee Y, Zheng L, Orvis J, Haas B, Wortman J and Buell C R (2007) The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Research* **35**: D883-D887

[8] Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol*, **340**: 783-795

[9] Von Heijne, G. and Manoel, C (1990) Membrane proteins—from sequence to structure. *Protein Eng.*, **4**: 109-112.