Biotechnol Res 2017; Vol 3(4):94-98
eISSN 2395-6763

ORIGINAL RESEARCH

In Silico Analysis of Type III Signal Anchor Protein in *Citrus sinensis*

Archana KUMARI, Robina YEASMIN, Jatin SARMAH*
DBT sponsored Bioinformatics Infrastructure Facility, Department of Biotechnology, Bodoland University, Kokrajhar, Assam, India
*Corresponding Author email: jatinsarmahindia@gmail.com

• *Received: 29 June 2017* • *Revised: 18 August 2017* • *Accepted: 01 September 2017* • *Published: 25 September 2017* •

ABSTRACT
Single-pass transmembrane protein having a membrane-spanning domain which helps in movement of protein towards the endoplasmic reticulum (ER) membrane. Different types of trans-membrane protein (type I, II, III, and IV) is responsible for different function where both type II and III membrane proteins contain a special single membrane-spanning domain. In both proteins due to the presence of single membrane-spanning domain it acts as a signal to initiate insertion and as a membrane anchor. Further, this signal anchor may help in membrane insertion with either an Ncyt/Cexo orientation or Nexo/Ccyt orientation. Our study focused only on type III proteins. As this protein only possess single-anchor sequence and there is no N-terminal signal peptide. Usually, Type III proteins are positively charged due to the presence of positive cluster residues on the C-terminal side of the signal anchor. The dispersal of charged flanking the hydrophobic core of the signal sequences plays important role in coordination of signal anchor proteins in membrane. Although, the mechanism by which a signal-anchor sequence implements a specific orientation is still unidentified. Here, we performed genome wide screening to identify number of signal anchor proteins in *Citrus sinesis* genome, which will help to understand the general mechanism of protein orientation in type III membrane proteins.

KEY WORDS: *Single-Pas Membrane Protein, Type III Protein, Signal Anchor, Citrus sinensis*

INTRODUCTION

Importance of subcellular proteins can be understood as it helps other proteins to reach their target functional organelle within a cell. It acts as an essential feature for cell survival. Generally, protein sorting influenced by the 'signal' encoded in their primary structure of the transmembrane proteins. It consist a large number of hydrophobic and hydrophilic regions which usually exposed both side of membrane. Membrane-spanning domain which have multiple or single domain known as multi-pass or single-pass respectively. Normally, there are two orientations of signal sequences one is NH_2-terminal cleaved and other is uncleaved signal sequences. Signal sequences are responsible to direct single-spanning membrane proteins up to the endoplasmic reticulum (David *et al*., 1992; Wickner *et al.,* 1985). The NH_2-terminal signal sequences are found on both secreted and membrane proteins (Walter *et al.,* 1986) and cleaved from the protein by signal peptidase during its translocation across the endoplasmic reticulum membrane. On the other hand, uncleaved signal sequences targeted by the second class of proteins and also help in the transportation as well as anchor of protein into the membrane (Spiess *et al.,* 1986; Zerial *et al.,* 1986). These proteins are recognized as signal-anchor (SA) protein and are differ from proteins with a cleaved signal sequence (Lipp *et al.,* 1988; Kumari *et al.,* 2016, Kumari *et al.,* 2017, Ranjan *et al.,* 2014). Typically, there are two types of orientations like Ncyt/Cexo and Nexo/Ccyt, of single-spanning membrane proteins (Type II, III and IV) during protein targeting towards the ER membrane. Generally, a cytosolic N terminus and a luminal C terminus orientation found in type II proteins whereas vice

versa in type III proteins (Holland *et al.,* 1984). Remarkable part is the type III proteins contain only single-anchor sequence and lacking N-terminal signal peptide, similar 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. Due to the presence of positively charged residues and unknown mechanism the orientation of protein changes in membrane. Therefore, the identification of new signal anchor proteins helps to determine the biological function and the mechanism of protein orientation. In our study, we did genome wide screening to identify a putative signal anchor proteins encoded by the *Citrus sinensis* genome.

RESULTS AND DISCUSSION

Genome-wide identification of signal anchor proteins in *Citrus sinensis*

Single-spanning transmembrane protein contain signal anchor sequence but they don't have N-terminal peptide sequence are known as Type III singal anchor protein. The signal anchor protein also known as reverse signal anchor (Heijne *et al.,* 1990; Blumenthal *et al.,* 2000). Proteins which contain KFERO amino acid are known as Signal peptide and they help to reach lysosomes for degradation. KFERO means lysine, phenylalanine, glutamic acid, arginine and glutamine amino acid (Lawless *et al.,* 1995; Ajoy *et al*., 2015; Raji *et al*., 2017). According to various literatures it has been confirmed that the function of Type III anchor proteins are target to membrane (like secreted proteins), membrane insertion and translocation or retention (SA proteins). A computational biology approach has been previously applied to identify a number of transmembrane domain, tail anchored proteins in yeast, human and Arabidopsis. Here, in our research work we performed genome wide identification of single-pass signal anchored proteins (Type III) in *Citrus sinensis* using various bioinformatics tools (Jennifer et al., 2003; Ito *et al.,* 2014; Luigi *et al.,* 2017).

In previous study, bioinformatics was applied for identification of transmembrane protein and tail anchor protein in Human, Yeast and Arabidopsis (Mathias *et al.,* 2007; Ouyang *et al.,* 2007). In this study we used various bioinformatics tools to perform genome wide identification of signal-pass type III signal anchored proteins of *Citrus sinensis*. Signal anchor protein of plants helps to understand the general mechanism of transmembrane protein and change occurred in membrane due to orientation of protein (Valera *et al.,* 2004; Uwe *et al.,* 2005).

At first step we used NCBI release database for protein sequence and we obtained 35,648 protein sequence (Figure-1). From NCBI database we predicted the protein sequence database and we observe that unplaced sequence contain highest protein sequence (5,825) and pltd contain lowest protein sequence (87). Then we identify 2,483 number of protein sequence which contain single transmembrane domain by using transmembrane helix prediction server (Figure-2). From this graph we concluded that unplaced protein contains the highest number of single transmembrane domain (404) and pltd contains the lowest (32).

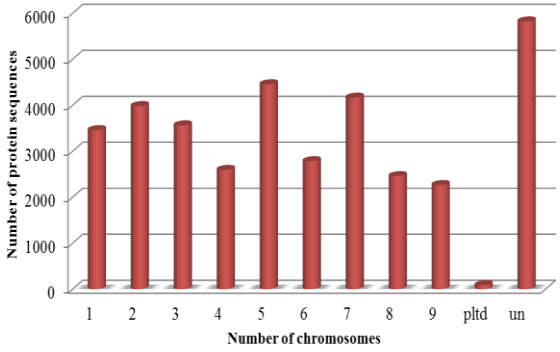

Figure 1: Showing whole protein sequence of *Citrus sinensis* from NCBI database

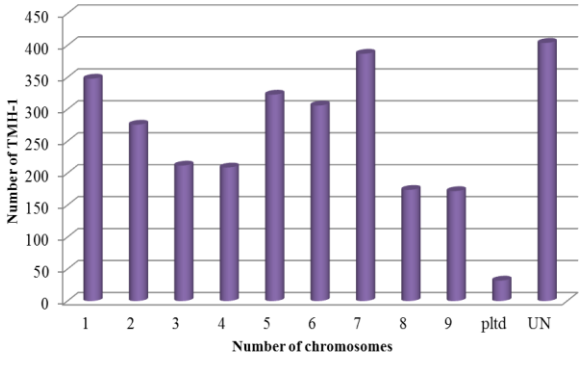

Figure 2: Showing transmembrane domain protein by using TMHMM

The next step we extracted protein sequence which contain only those transmembrane domain protein within the C-terminal 50 residue. The number of protein sequence which having transmembrane domain were 2,161 (Figure-3).

Table 1: Shows gene ontology of *Citrus sinensis*

Gene Name	GO ID	GO Molecular Function	GO Cellular Component	Protein Description
CISIN_1g010129mg	GO:0004190	Aspartic-type endonuclease activity	unknown	Unknown
CICLE_v10000393mg	GO:0004565	Beta-galactosidase activity	unknown	Beta -galactosidase
CISIN_1g033263mg	unknown	unknown	unknown	Unknown
CISIN_1g034691mg	GO:0016020	unknown	membrane	Unknown
CISIN_1g043104mg	GO:0003824	Catalytic activity	unknown	Unknown
CISIN_1g10015924mg	GO:0008233	Peptidase activity	unknown	Unknown
CICLE_v10000393mg	GO:0046872	Metal ion binding	Extracellular region	Unknown
CISIN_1g033170mg	unknown	unknown	unknown	Unknown
CISIN_1g034455mg	GO:0016020	unknown	Membrane	Unknown
CISIN_1g010230mg	GO:0008233	Peptidase activity	unknown	Unknown
CICLE_v10000393mg	GO:0008236	Serine type endopeptidase activity	Cell wall	Unknown
CISIN_1g021206mg	GO:0045492	unknown	unknown	Unknown
CICLE_v10020393mg	unknown	unknown	unknown	Unknown
CISIN_1g009054mg	GO:0005507	Copper ion binding	unknown	Unknown
CISIN_1g018968mg	GO:0008233	Peptidase activity	unknown	Unknown
CISIN_1g034826mg	GO:0006952	unknown	unknown	Unknown
CICLE_v10011553mg	GO:0004190	Aspartic-type endonuclease activity	unknown	Unknown
CICLE_v10031232mg	GO:0003824	Catalytic activity	unknown	Unknown
CICLE_v10015924mg	GO:0006979	Heme binding, peroxidase activity	membrane	peroxidase
CISIN_1g034691mg	GO:0016021	integral component of membrane	Membrane	Unknown
CISIN_1g043104mg	GO:0055114	Flavin adenine dinucleotide binding ,	unknown	Unknown
CISIN_1g019063mg	GO:0006508	Cyestein type peptidase	unknown	Unknown
CISIN_1g033170mg	unknown	unknown	unknown	Unknown
CISIN_1g034455mg	GO:0016020	unknown	Membrane	Unknown
CISIN_1g010230mg	unknown	Aspartic-type endopeptidase	unknown	Unknown
CISIN_1g021206mg	GO:0045492	Xylan biosynthetic process	unknown	Unknown
CISIN_1g009054mg	unknown	unknown	unknown	Unknown
CISIN_1g018968mg	GO:0006509	Cystein type peptidase activity	unknown	Unknown
CISIN_1g034826mg	GO:0006952	Defense response	unknown	Unknown

This transmembrane domain protein identified the single transmembrane domain near the C-terminus and which further help to know N-terminal single peptide and if we found N-terminal single peptide domain, we discarded them from the list (Figure-4). Finally 82 proteins contain single transmembrane helix (TMhelix) were collected as type III single anchor proteins and they don't have N-terminal single peptides. To know the molecular function and cellular component of anchor protein we retrieve the data in UniProtKB for gene ontology which gives the result. We observed out of 82 we found the function of only 29 proteins and for the rest no information was available on UniProtKB. According to our analysis most of the anchor proteins are involved in peptidase activity, some are involved in endonuclease activity, some are involved in hydrolase activity and some are involved in molecule binding protein. Four proteins do not have any function. Below the table showing gene ontology of *Citrus sinensis* of signal anchored protein to understand the molecular function and cellular component in details (Table-1).

CONCLUSION

Our work suggest list of anchored protein of *Citrus sinensis* which does not have N-terminal signal peptide but they contain signal peptide that are involved in protein targeting through endoplasmic reticulum. Different types of graph we draw showing the highest and lowest sequence through graph in different step. Gene ontology of signal anchored

protein helps to understand the molecular functions of signal anchored protein.

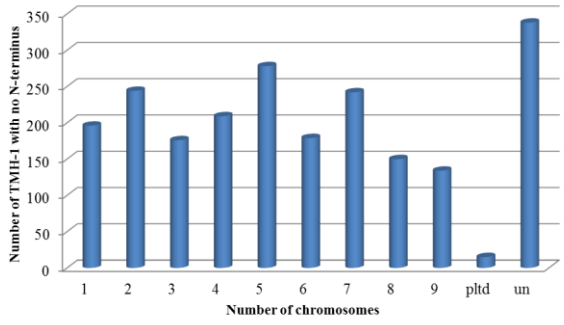

Figure 3: It is representing the number of TMH1 with no N-terminus

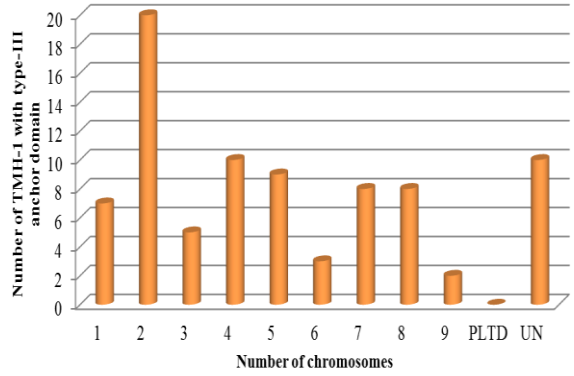

Figure 4: Final selected proteins having transmembrane domain and signal anchored type III protein of *Citrus sinensis*

From the work we found the molecular function and biochemical analysis that will help us to understand the biological function of type III anchored protein in *Citrus sinensis*.

METHODS

Identification of Signal Anchored Proteins

Citrus sinensis protein sequences for all 11 chromosomes were obtained from NCBI database. The server TMHMM, Version 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) was used to predict transmembrane domain in *Citrus sinensis* 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 4.1 (http://www.cbs.dtu.dk/services/SignalP/). The parameters for SignalP v4.1 (Bendtsen *et al.,* 2004) tool were set as follows: eukaryotes, neural networks (NN) and Hidden Markov Model (HMM); truncated to first 70 residues. The overall strategy to predict signal anchor proteins were based on *In-silico* approach describe in Figure 5

Protein Localization and Functional Analysis

The biological function of screened signal anchor proteins were analyzed by using GO Slim (http://www.geneontology.org/page/go-slim-and-subset-guide) and AmiGO blast search (http://amigo.geneontology.org/cgi-bin/). The functional annotation was based on GO Slim and blast search against Arabidopsis, while subcellular localization was based on validated experimental data available on TAIR.

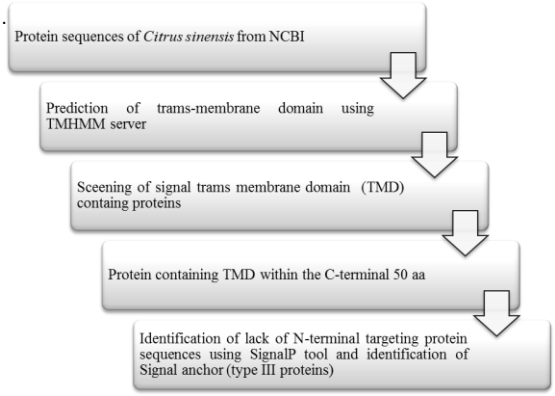

Figure 5: Flowchart showing steps to obtain signal anchor type III proteins.

ACKNOWLEDGEMENTS

DBT-BIF, Department of Biotechnology, Bodoland University is greatly acknowledged for providing necessary funds and facilities. Technical support provided by Dr. Prasun Kumar is fully acknowledged for the execution of above work.

CONFLICT OF INTERESTS

No conflict of interest.

REFERENCES

Ajoy, P., Sudhangsu, D. (2015). Text book of Cell and molecular biology; books and allied; 3(1): 143-158

Bendtsen, J.D., Nielsen, H., Heijine, G. and Brunak, S. (2004). Improved prediction of signal peptides: SignalP4.1. Journal of Molecular Biology, 340, 783-795.

Blumenthal, M., Stuart, M. and Lawless, J. (2000). The Illustrated encyclopedia of essential oils ;Herbal Medicine; Expand commission E Monographs, Austin, Texas, American Botanical Council; 12(1); 104-109

David, A.W., Jason, Y.C., Laurence, M.F. and Gregory, C.J. (1992). The role of the N region in signal sequence signal anchor function. *The journal of biological chemistry*, 267(11), 7761-7769.

Heijne, V., Manoil, G.C. (1990). Membrane proteins from sequence to structure. *Protein Eng.,* 4, 109-112.

Holland, E.C., Leung, J.O. and Drickamer, K.(1984). Rat liver asialoglycoprotein receptor lacks a cleavable N-terminal signal

sequence. *Proc .Natl. Acad. Sci. USA*, 81, 7338-7342.

Ito, T.M., Polido, P.B., Rampim, M.C., Kaschuk, G. and Souza, S.G.H. (2014). Genome-wide identification and phylogenetic analysis of the AP2/ERF gene superfamily in sweet orange (Citrus *sinensis*). *Genetic and Molecular Research*, 13(3), 7839-7851.

Jennifer, A., Corcoran and Roy, D. (2003). Reptilian reovirus utilizes a small type III protein with external myristylated amino terminus to mediate cell-cell fusion. *American society for microbiology*, 78(8), 4342-4351.

Kumari, A., Kanchan, S., Sinha, R., P., Kesheri, M. (2016). Applications of Bio-molecular Databases in Bioinformatics. In book: Medical Imaging in Clinical Applications, 329-351.

Kumari, A., Sumer, S., Jalan B., Nongbri, P., L., Laskar, M., A. (2017). Impact of next-generation sequencing technology in plant–microbe interaction study; Microbial Applications Vol.1, 269-294.

Lawless, J., Bown, D., Michael, W. and Kalish, B.C. (1995). The Royal Horticultural society New Encyclopedia of Herbs and their uses; London; England; Dorling Kindersley; 3(1); 182-191

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

Luigi, M.D., Domenico, C., Domenico, P. and Luigi, S. (2017). Experimental evidence and in silico identification of tryptophan decarboxylase in *Citrus* genus; MDPI; *Molecules*, 22, 272.

Mathias, R.L. and Arme, PL. (2007). Comprehensive analysis of glycosylphosphatidylinositol-anchor proteins in candidate albicans; *American society for Microbiology*, 6(2), 119-133.

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 Bull, C.R. (2007). The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acid Research*, 35, D883-D887.

Raji, P., Antony, S.V., Divya, D. and Boniface, A. (2017). In vitro and in silico approaches to study the bioactivity of citrus limon leaf extracts. *J Young Pharm*, 9(2), 290-295.

Ranjan, A., Kumari, A., Pandey, D., M. (2014). Annotation of stress responsive candidate genes in peanut ESTs. Interdiscip Sci. doi 10.1007/s12539-013-0054-3.

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

Uwe, A., Thomas, W., Walter, N. and Doron, N. (2005). Signal-anchored proteins follow a unique insertion pathway into the outer membrane of mitochondria. *Journal of biological chemistry*, 280(1), 48-53.

Valera, V., Peremy, S., Yung, P.W. and Valerian, D.V. (2004). Movement protein of a clsterovirus is a type III integral transmembrane protein localized to the endoplasmic reticulum. *American society for microbiology*, 78(7), 3704-3709.

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

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

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