Arabidopsis Transmembrane Proteins Regulate Protein Turnover as Signaling Molecules During Internalization by Means of Ubiquitination

Lin Wang

University of Leeds, Leeds LS2 9JT, UK.

Abstract: This study attempts to characterize the structure, function, and other features of Arabidopsis thaliana based on its transmembrane protein amino acid sequence and assess its potential as a drug target and ubiquitination site. The transmembrane protein belongs to the Arabidopsis plant, but little is known about the structure-function of the protein, among others. We predicted the structure and transmembrane domains of the transmembrane protein by different bioinformatic tools, and the prediction results showed that its three poorly conserved transmembrane domains could be used as potential drug targeting sites. Among them, lysine ubiquitination sites play an important role in the recycling of plant membrane proteins, and we found the existence of ubiquitination sites by predicting the ubiquitination sites of transmembrane proteins, and this could be a new research direction in the future.

Keywords: Domain; Multiple Alignments; Structure Prediction; Protein Ubiquitination Sites

Introduction

As a preferred model for studying plants, Arabidopsis has played an important role in helping to understand plant growth and development and molecular genetics, etc. Thus Arabidopsis became the first plant to have its genome sequenced [1]. Membrane proteins have multiple functions, post-translational modifications are critical for the regulation of protein functions. These modifications include ubiquitination, acetylation and palmitoylation. Among them, ubiquitin has 7 lysine sites, which can produce different kinds of ubiquitination through different linkage and conformation, usually divided into mono-ubiquitination, multi-locus mono-ubiquitination and poly-ubiquitination [2].

In their study of ubiquitination modifications in yeast, Kölling and Losko identified a cytosolic cycle mediated by ubiquitination [3]. In a subsequent study, ubiquitination was found to play a role in vesicular transport [4]. Thus, a new chapter in the study of plant ubiquitination was opened, and it was also gradually found that ubiquitination has an important influence in the dynamic cycle of plant membrane proteins. Cui et al. showed that the ubiquitin ligase UBC32 can respond to BR signaling through the ERAD pathway in Arabidopsis thaliana.

The aim of this study was to explore the transmembrane protein of Arabidopsis thaliana regulated by ubiquitinated molecular signaling, leading to internalization and turnover of membrane proteins. The annotation score of this transmembrane protein is low because very little is known about its structure and function. Therefore, we aimed to determine the relationship between this protein and ubiquitination by predicting its structure, functional domains, ubiquitination site prediction, and multiple alignments.

Methods

Sequence selected: First search in Blastp based on the amino acid sequence provided by the module manager， the sequence was confirmed as a transmembrane protein of Arabidopsis thaliana based on E value, Percent Identity and Query Cover value. After that, the Gene sample AT2G18690 was searched directly in Uniprot and the Gene was identified as MSF 3.7. The sequence of this protein is shown in Figure 1 below, which is a sequence of 322 amino acids.

```
        10          20          30          40          50
MARFSFLNVV KDVVAILNES RKLFLKNKKL MFSVLVFPLL LNCLVYFLNI
        60          70          80          90          100
FVIVPEITNL ILEASLLPST DPTSPEYAAR LMRVFTDFRQ FVGSSYIFAA
       110         120         130         140         150
VSSIINLFSV LVIVHASAIT LKDENFNIKD FPVLSLKSWK GPLVTYFYIA
       160         170         180         190         200
LFSLGFGFLF FIILCPILLF SIKSGSVENI GFLAVEAGVL LIIFTVSQSY
       210         220         230         240         250
FAIYWNLSMV ISILEESYGF QALGKAAKIV KGMKTKLFLL NLFFGLLASG
       260         270         280         290         300
LAQILQLINM GRSLAVTLTT GFVLVCLVFA VRMFQLVTYT VAYFQCKSLQ
       310         320
GRDVESLRDV EYMALSSTTL TE
```

Figure 1: The FASTA format of the amino acid sequence of Transmembrane protein of Arabidopsis thaliana obtained from Uniprot. It is a 322 amino acid sequence

Domain prediction: Using pHmmer、Interpro and Phyre2 to predict the structural domain of the protein. The transmembrane topology of the protein can be detected by Phyre2. Interpro website can determine the conserved domain of this transmembrane protein. The function of this protein has not been predicted in Uniprot, and its similar protein in (thale cress) hypothetical protein Gene AXX17_AT2G18690 is the only protein for which a predicted function exists, i.e. the presence of an ATP binding site. Therefore, Websites such as NCBI and Pubmed will be used to search the literature and predict the function of this transmembrane protein.

Multiple alignments: Sequences were analyzed for comparison using Clustal Omega, which was done with yeast cells, animal cells, Arabidopsis thaliana (Mouse-ear cress) and human transmembrane proteins. Finally a phylogram was generated using the European Bioinformatics Institute (EBI) website.

Structure prediction: This transmembrane protein has no predicted structure in Uniprot, so we use Phyre2 to predict its structure. Confidence and Coverage values were considered when using Phyre2.

Prediction of protein ubiquitination sites in Arabidopsis thaliana: Using AraUbiSite's online predictive variables to predict the potential sites for ubiquitination of this transmembrane protein.

Results and Discussion

A search for similar protein using Uniprot showed that this protein has 50% identity with the Gene MSF 3.6 Arabidopsis thaliana (Mouse-ear cress) Transmembrane protein with the Gene AXX17_ AT2G18690 sequence for sequence comparison to analyze whether the two proteins have the same function.

Sequence Domains: The domains of this transmembrane protein were predicted using pHmmer, Phyre2 and Interpro tools, then the accuracy of the predictions between the different software was analyzed. pHmmer revealed the conserved domain of the Pfam family, which spans most of the protein's sequence (amino acids 30-284) shown in Figure 2, while the results showed that the protein protein has six transmembrane domains.

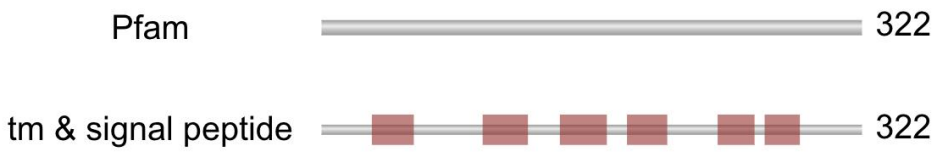


Figure 2: The domains of this sequence were obtained in the pHmmer tool. The sequence is a Pfam family domain that spans amino acids 30-284. According to pHmmer's predictions, the sequence possesses 6 transmembrane and signal peptides, indicated by the red-brown squares.

The transmembrane domain of this sequence was predicted to have six transmembrane domains using Phyre2 (shown in Figure 3) in agreement with the pHmmer results, with the protein's N and C terminals in the cytoplasmic region.

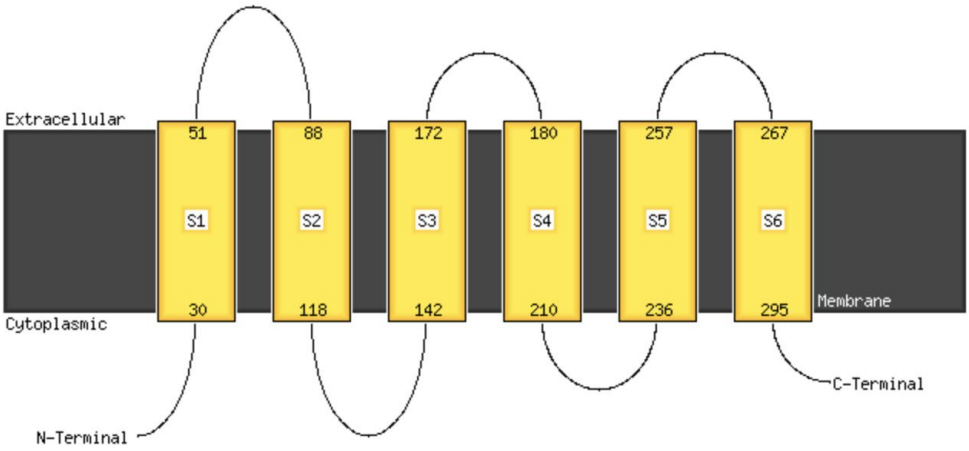


Figure 3: The transmembrane domains of protein were predicted using Phyre2. The protein spans 6 transmembrane domains, with the N and C terminals extending into the cytoplasmic space, as can be observed.

The conservation of the transmembrane domain of this sequence was analyzed by Interprot. Like the results of pHmmer, Phyre2, Interprot also predicted 6 transmembrane domains (amino acids 30-284)

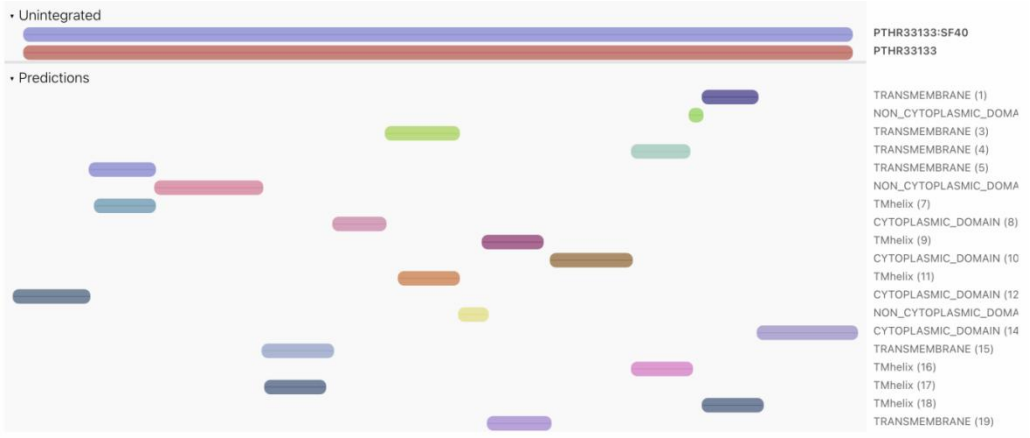


Figure 4: Transmembrane domains predicted by Interpro. The figure contains a total of 6 transmembrane domains

Also, the sequence predicts 3 non-cytoplasmic domains、 4 Cytoplasmic domains and 6 TM helix. Three non-cytoplasmic domains are present in amino acid residues 55-95, 171-181, and 259-263, respectively. The 6 TM helixes correspond to the

transmembrane domains, respectively. These non-conserved domains may be able to be targeted as drugs because there is less cross-reactivity with other species

To confirm the low conservation of the domains in Interpro, we performed a multiple sequence alignment with the (thale cress) hypothetical protein Gene AXX17_AT2G18690. The sequence comparison revealed that 3 of the structural domains present in the transmembrane protein have low conservation, residues 32-54, 148-170, 180-202 (red line boxed part), which further raises the possibility that these structural domains may be potential drug targets, and further studies are needed.

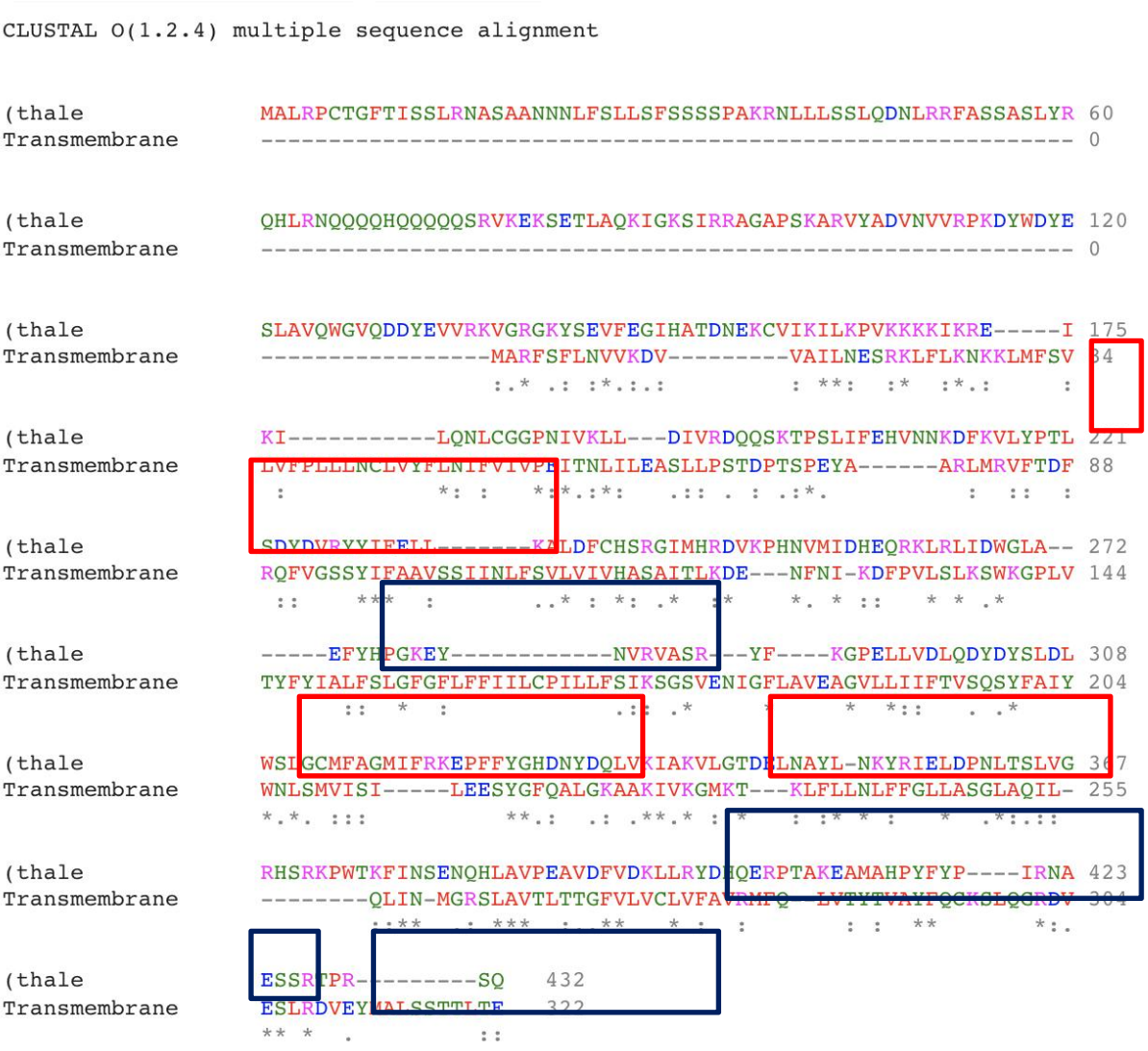
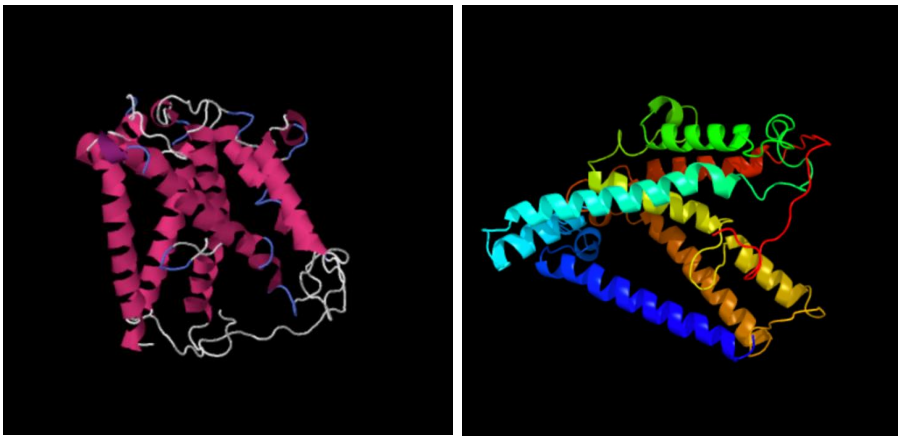


Figure 5: (thale cress) Multiplex sequence comparison of hypothetical protein and transmembrane protein. In this comparison, the transmembrane domains shown in Interpro are marked with boxes. Red boxes represent poorly conserved and black boxes represent well conserved.

Protein Structure

In our prediction of the structure of this protein, we used Phyre2 for structure prediction, which produced results with a higher quality, with a Coverage of 79%, and also compared with the homologous protein whose structure had been predicted and found to be similar, both being AlphaFold.



A1 **B:**

Figure 6: Structural model of this transmembrane protein predicted using Phyre2. The dimensions of the model (A) are X:54.476 Y:57.265 Z:66.656

The alpha helix and the loop region make up the secondary structure of the protein, and in Figure 6 above we can clearly see that the alpha helix also makes up the six transmembrane domains on the protein.

Prediction of protein ubiquitination sites in Arabidopsis thaliana

When studying the molecular mechanism of ubiquitination, it is important to identify the ubiquitination sites. Therefore, Chen et al. collected a large number of protein ubiquitination sites in Arabidopsis, adopted AAC, Binary and CKSAAP models to build a prediction model, and generated an online predictor named AraUbiSite (http://systbio.cau.edu.cn/araubisite)[5]. In Figure 7, the possible ubiquitination sites in this transmembrane protein sequence are listed, namely lysine residues 11, 22, 26, 28 and 29; however, we can also see that the SVM scores and Confidence of these ubiquitination sites are generally low. This protein is an unknown protein in Arabidopsis plants, although the SVM score and Confidence values are low, it still has the value for further study.

Protein	Sequence	Ubiquitination sites(SVM score)	Confidence: High Low Non
sequence	1 MARFSFLNVV KDVVAILNES RKLFLKNKKL MFSVLVFPLL 40 41 LNCLVYFL	11(0.0756893) 22(0.056168) 26(0.00372362) 28(9.42635e-06) 29(0.00408013)	

Figure 7: Transmembrane protein ubiquitination sites predicted by AraUbiSite. K marked in green font in the figure is the possible lysine ubiquitination site, SVM score represents the probability of the site, and the three colors represent Confidence values.

Conclusion

The main objective of our research project was to analyze the sequence of an uncharacterised protein to predict its possible structure and function, as well as its possible future application areas through different bioinformatic approaches. First, based on Uniprot's analysis of the protein sequence, we first learned that this protein belongs to the transmembrane protein of Arabidopsis thaliana, with a 322 amino acid sequence. We then predicted the transmembrane domain of this protein by pHmmer, Interpro and Phyre2 tools, and compared the differences between the predictions of different methods, we found that the results were consistent, the protein has 6 transmembrane domains, spanning amino acids 30-284, with the N and C terminals extending into the transmembrane domain. In addition, when predicted in the Interpro website, we found that the protein also has 6 TM helix, 3 non-cytoplasmic domains. The TM helix and transmembrane domains showed a one-to-one correspondence. To verify whether the transmembrane domains of this protein have low conservation, we selected a hypothetical protein in Arabidopsis with high sequence homology and found that three of the six transmembrane domains have low conservation.

Since the structure of the protein was not yet shown in Uniprot, we used Phyre2 tool to predict the structure and found that the structure of this protein is AlphaFold, which is similar to its homologous protein structure. After predicting the ubiquitination sites of this protein by online AraUbiSite predictor, we found that there are five possible lysine ubiquitination sites, which need to be verified by further studies because there are not many studies related to ubiquitination in Arabidopsis.

References

[1] The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408, 796–815 (2000).

[2] Ikeda, F. & Dikic, I. Atypical ubiquitin chains: new molecular signals. 'Protein Modifications: Beyond the Usual Suspects' review series. EMBO Rep 9, 536–542 (2008).

[3] Kölling, R. & Losko, S. The linker region of the ABC-transporter Ste6 mediates ubiquitination and fast turnover of the protein. EMBO J 16, 2251–2261 (1997).

[4] Barberon, M. et al. Monoubiquitin-dependent endocytosis of the IRON-REGULATED TRANSPORTER 1 (IRT1) transporter controls iron uptake in plants. PNAS 108, E450–E458 (2011).

[5] Chen, J., Zhao, J., Yang, S., Chen, Z. & Zhang, Z. Prediction of Protein Ubiquitination Sites in Arabidopsis thaliana. Current Bioinformatics 14, 614–620.