Grenze International Journal of Engineering and Technology, Jan Issue

In-Silico Characterization and Homology Modelling of Membrane Bound Proteins in Different Rice Cultivars

Shrestha Kumar[1], Arup Kumar Mukherjee[2] and Satya Ranjan Dash[3]
[1]School of Biotechology, KIIT University, Odisha, India
Email: shreskum@gmail.com
[2]ICAR-National Rice Research Institute, Odisha, India
Email: titirtua@gmail.com
[3]School of Computer Applications, KIIT University, Odisha, India
Email: sdashfca@kiit.ac.in

***Abstract*—Various stresses which include majorly drought, salt and metal ions present in the soil adversely affects the growth and survival of the rice cultivars produced by various methods. To combat these researchers and scientist with the help of computational and systems biology are characterizing proteins involved for the resistance of these stresses to have the availability and reach of the rice cultivar across the globe with cost effectiveness keeping in mind. There are many of the proteins which are yet to be discovered, and many of them which have been discovered do not have their 3-D structure determined. So, there comes the application of homology modelling. Another reason for choosing homology modelling is, since proteins which are targeted here are membrane bound, they are hard to crystallize hence homology modelling becomes the most suitable way to determine and study their functionality. Proteins such as remorin 1.4, Phospholipase D, HIR protein LT1A, LT1B are some of the proteins which came into light after some research from already uploaded data from various journals and protein databases such as the UniprotKB. Using the database proteins were characterized based on biotic and abiotic stresses and proceeded accordingly. For modelling, an online software named SWISS Model was used to select the template and even helped in making the whole model with required scores to assess the quality of the model, respectively. To further refine the model modrefiner online software was used to refine the model to the extent where its score is better with bad angles and bonds removed respectively hence giving us a refined version of the model compared to the ones made by SWISS model directly. Templates or reference sequences were chosen based on ideal characteristics seen in Arabidopsis thaliana also seeing their sequence similarity which were automatically shown in the SWISS Model. Managed to model three of the total proteins mentioned which showed great scores after refining models. ERRAT online software was also used to give us the verification of the structure obtained.**

***Index Terms*— UniprotKB, SWISS model, homology modeling, Phospholipase D.**

I. INTRODUCTION

Responses generated by signal cascade due to various abiotic and biotic factors are proven to be in the plasma membrane (PM). Proteome analysis and their response to environmental stresses might lead to a better understanding towards creating cultivars and GMO crops. To study these factors and signals associated with

Grenze ID: 01.GIJET.9.1.743

ideal crop has been seen in Arabidopsis thaliana. Sub-cellular proteomics approach was applied to identify the proteins responsible for the abiotic and biotic stress responses.

There are many proteins expressed differentially, many of them were membrane transporters for example kinases and ATPases. Only few of them were of our natural interest some of which are remorins 1.4, phospholipases and HIR (Hypersensitive-induced response protein). Their experimental and predicted structural information are stored in public database known as UniProt Knowledge base.

Other proteins like hydrophobic proteins (LTI6A and LTI6B) have a functional role in response to chilling temperatures and salt stress, they were found in japonica rice (Oryza sativa). These responses are due to various signal cascades in which one of them is signaling through production of ABA (Abscisic Acid). The genes responsible for preventing dehydration and salt stress proteins are mainly responsible for cellular defenses against cell membrane and the plasma membrane. Conjunction of both proteins and various chemical signaling gives rise to defense mechanism against stress.

Protein 3D-structure collapsing from a straightforward grouping of amino acids were viewed as an extremely troublesome issue before. Nonetheless, it has advanced during that time into an operable test with agreeable and sensible exact forecasts of the time. To study their structural functionalities homology modelling is done when the model protein (with a known sequence and an unknown structure) is related to at least one other protein with both a known sequence and a known structure. SWISS-MODEL makes it quick and easy to submit a target sequence and get back an automatically generated homology model, provided an empirical structure with >30% sequence identity exists to use as a template.

Without tentatively decided protein 3D constructions, homology modelling displays and assumes a practical part in structure-based applications and the portrayal of protein properties and capacities. SWISS model's results can be verified and checked using ERRAT Save6.0 which checks and verifies the given model and provides with appropriate conclusion.

The purpose of targeting membrane bound proteins in different rice cultivars and modelling for the same is that they are the first and foremost barrier of the cell and are mostly associated with sensing and producing a response for given stress. The plasma layer is the principal site of extracellular biotic or abiotic detecting, so a comprehension of proteome elements may work with the improvement of new techniques for stress obstruction in crops. As the essential natural obstruction, the plasma film of a plant cell controls numerous organic cycles, for example, particle transport, endocytosis, cell separation and expansion, and sign transduction. Nonetheless, useful protein confinement is muddled because some layer proteins are firmly connected with the double lipid centre, while others are freely and reversibly related. Innovative advances in the extraction of sanitized plant plasma film proteins have made it conceivable to profile the proteome of the whole plant layer. Proteome profiling may explain which fundamental defensive reactions are started in plant plasma films presented to ecological pressure. Therefore, this profiling would then help in making various 3-D models using homology modelling through various software such as SWISS model or Modeller which is considered as the best for as it produces better and accurate results without errors compared to SWISS Model. Homology modelling intends to construct three-dimensional protein structure models utilizing tentatively decided designs of related relatives as formats. SWISS-MODEL workspace is a coordinated Web-based displaying master framework. For a given objective protein, a library of trial protein structures is looked to recognize appropriate layouts. Based on a grouping arrangement between the objective protein and the layout structure, a three-dimensional model for the objective protein is created. Model quality appraisal instruments are utilized to assess the dependability of the subsequent models. Homology modelling is as of now the most exact computational technique to produce solid underlying models and is regularly utilized in numerous organic applications.

II. Related Work

Genome sequencing projects are submitted as crude information, from which researcher endeavour to clarify the capacity of the anticipated quality items. The protein arrangements are put away in open data sets, for example, the UniProt Knowledgebase (UniProtKB) [1], where caretakers attempt to add anticipated and test useful data. Protein work forecast should be possible utilizing arrangement similitude look, yet an elective methodology is to utilize protein marks, which characterize proteins into families and areas. The significant protein signature information bases are accessible through the coordinated InterPro data set, which gives an order of UniProtKB arrangements. Just as portrayal of proteins through protein families, numerous scientists are keen on investigating the total arrangement of proteins from a genome (for example the proteome), and there are information bases and assets that give non-repetitive proteome sets and investigations of proteins from organic

entities with totally sequenced genomes [2]. The prediction of the 3D structure of a protein from its amino acid sequence remains a basic scientific problem.

Proteomics and various other methods compiled the study of good number of proteins. But amongst them only some of them played a vital role in response to the given stress majorly salt and metal ion stresses. Two-dimensional electrophoresis uncovered those 24 proteins were differentially separated accordingly to salt pressure. From these [3], eight proteins were distinguished by mass spectrometry examination. Most of the proteins distinguished are probably going to be PM-related and are known to be engaged with a few significant components of plant variation to salt pressure. These incorporate guideline of PM siphons and channels, film structure, oxidative pressure guard, signal transduction, protein collapsing, and the methyl cycle.

Various virus stress-incited qualities have been found as of late utilizing sub-atomic and hereditary methodologies as microarray tests [4-6]. The results of these unique virus stress-actuated qualities are by and large considered to improve cold pressure resistance and regulate quality articulation by means of flagging pathways. Plants react to cold pressure difficulties partially through changing their quality articulation levels, which at long last prompts different versatile reactions at cell and entire plant level [7].

A Material and Methods

All-atom segments that match the guiding positions can be obtained either by scanning all the known protein structures. In addition to that it includes those protein structures that are not related to the sequence being modelled, or by a conformational search restrained by an energy function. Modelling by satisfaction of spatial restraints based on the generation of many constraints or restraints on the structure of target sequence, using its alignment to related protein structures as a guide. Generation of restraints is based upon the assumption the corresponding distances between aligned residues in the template and the target structures are similar. SWISS Model was used to search the appropriate template with optimum alignment with the target sequence as mentioned in the above step [8]. SWISS Model uses HMM (Hidden Markov Model) to search for suitable templates against the built-in template library.

B. Model refinement.

Model refinement is a vital undertaking that requires effective examining for conformational space and a way to precisely distinguish close local constructions. Homology model structure measure develops through a progression of amino corrosive build-up replacements, inclusions, and cancellations. Model refinement depends on tuning arrangement, demonstrating circles and side chains. The model refinement cycle will typically start with an energy minimization step utilizing one of the atomic mechanics power fields and for additional refinement, strategies like sub-atomic elements, Monte Carlo and hereditary calculation-based inspecting can be -applied [9-10] An online software named Modrefiner can be used to refine the model with respect to the reference model given on Uniprot KB which is totally optional. The accuracy of alignment by modelling strongly depends on the degree of sequence similarity. Misalignment of the models some time results into the errors which may be hard to remove at the later stages of refinement.

C. Model validation.

Each progression in homology demonstrating is dependent on the previous cycles. Hence, mistakes might be coincidentally presented and engendered, in this way the model approval and evaluation of protein is vital for deciphering them. Errors in model are usually estimated by (1) superposition of model onto native structure with the structure alignment program Structural and calculation of RMSD of Cα atoms; (2) generation of Z-score, a measure of statistical significance between matched structures for the model, using the structure alignment program CE, scores four indicate good structural similarity and (3) development of a scoring function that is capable of discriminating good and bad models basic necessity for a constructed model is to have good stereochemistry. The Ramachandran plot is probably the most powerful determinant of the quality of protein, when Ramachandran plot quality of the model is comparatively worse than that of the template, then it is likely that error took place in backbone modelling. WHAT_CHECK determines Asn, His or Gln side chains need to be rotated by 180 degrees about their C2, C2 or C3 angle, respectively. Side chain torsion angles are essential for hydrogen bonding, sometimes altered during the modelling process. Conformational free energy distinguishes the native structure of a protein from an incorrectly folded decoy. A distinct advantage of such physically derived functions is that they are based on well-defined physical interactions, thus making it easier to learn and to gain insight from their performance. The validation programs are generally of two types: (1) first category (e.g., PROCHECK and WHATIF) checks for proper protein stereochemistry, such as symmetry checks, geometry checks (chirality, bond lengths, bond angles, torsion angles models. Solvation and structural packing

quality and (2) the second category (e.g., VERIFY3D and PROSAII) checks the fitness of sequence to structure and assigns a score for each residue fitting its current environment.

III. Result

The result has been shown in Fig. 1

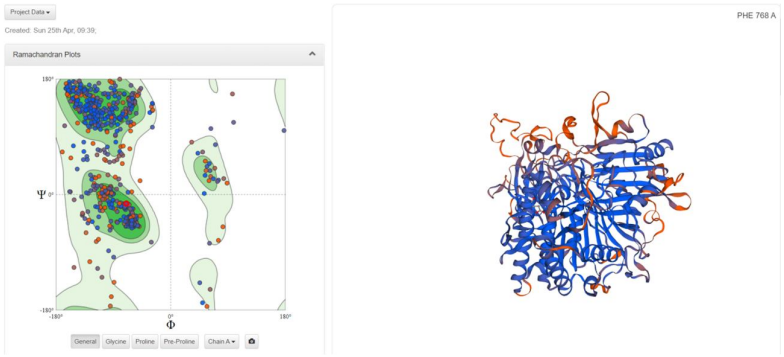

Figure 1 Phospholipase D model with Ramachandran plot

Using an appropriate template from the list provided by the SWISS Model site modelling was successfully completed. The mol probity score is 1.64 and 87.21 percent. The quality estimate is given in Fig 2.

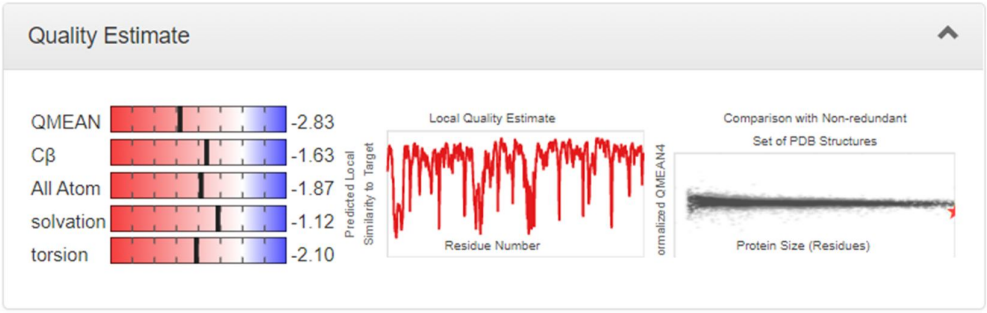

Figure 2Quality estimate scores of the model

The validation report was also generated using the SAVE6.0 server which provided us with ERRAT, verify 3D, PROVE, WHATCHECK, PROCHECK scores as shown in Fig 3.

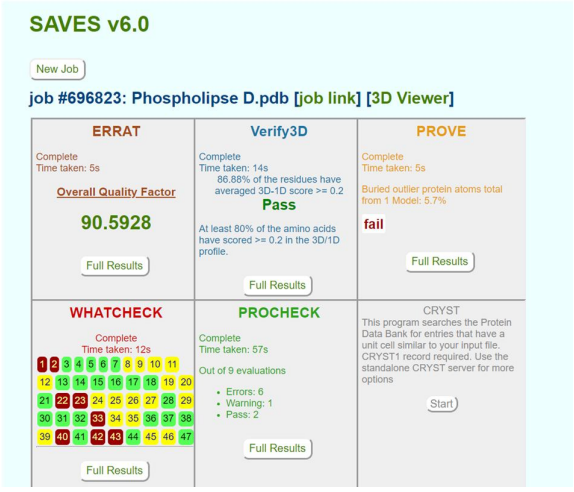

Figure 3 Validation report for Phospholipase D model

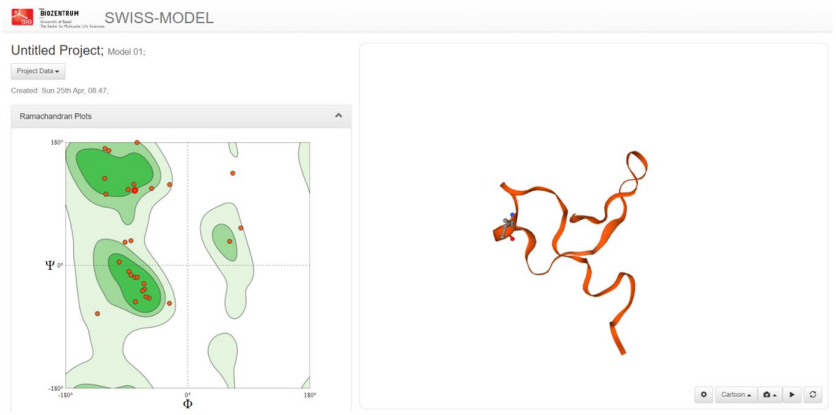

Figure 4 Remorin 1.4 model with Ramachandran Plot

As appropriate template was not available in templates tab in the SWISS model server, we had to move on to proteins present in Arabidopsis thaliana which has close similarity to remorin 1.4 with respect to characteristics and sequence both. Successfully a model was made as shown in the Fig4.

Figure 5Quality estimate with respect to remorin model

With quality estimate given as shown in Fig5, this model has a mol probity score of 1.95 and is 71.43 percent Ramachandran favoured.

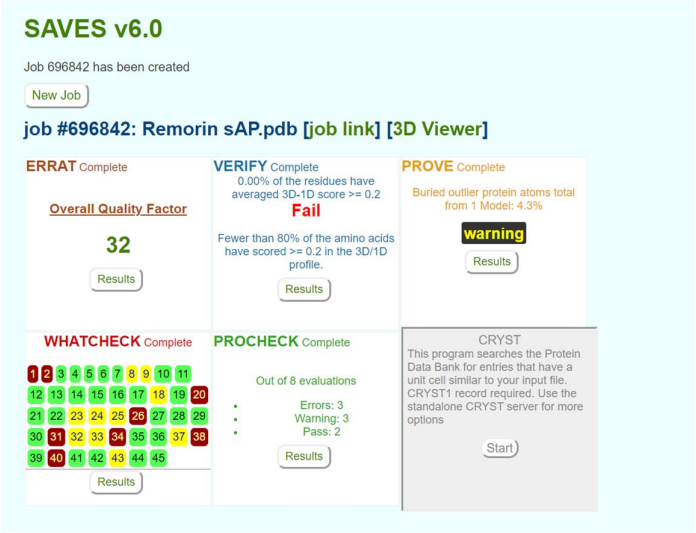

Figure 6Validation report of Remorin 1.4

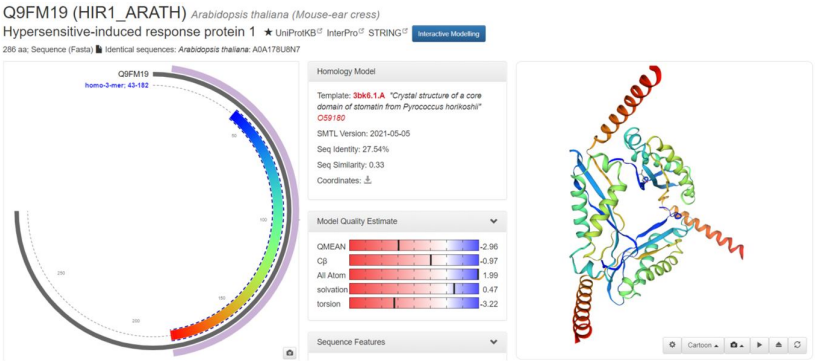

Figure 7HIR homology model with quality estimate

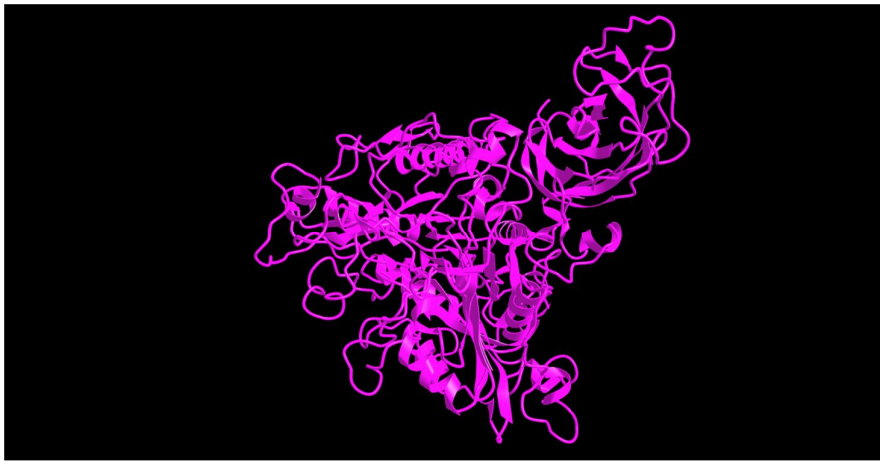

Figure 8 Refined model of Phospholipase D using Modrefiner

IV. Discussion

The model prepared for Phospholipase D has been made with a template named phospholipase D alpha 1 which is a crystal complex with phosphatidic acid. This choice of template comes from the fact that Cd-induced PLD activity and accumulation of phosphatidic acid (PA) is produced in ample amount with respect to PLD activity which is the response to the abiotic stress.

The template used for modelling Remorin 1.4 is Zinc finger AN1 and C2H2domain containing stress associated protein 13 which is responsible for stress associated with increase of abscisic acid, salt stress and metal ions. This is present in Arabidopsis thaliana which is referred as an ideal plant for most of the rice experiments. The validation scores are not up to the mark as they were expected to be but to ensure the hit and trial method for selection of appropriate template for the same would require refinement to some degree which would then tell us to move forward with other proteins related to Arabidopsis thaliana.

HIR protein already had a homology model prepared in SMR database. The details of the template which has been used as shown in the Fig7.

V. Conclusion and Future Work

The models achieved using the steps provided by the SWISS Model itself have relative errors and many of them are inaccurate, to avoid these and to get best fit model, Modeller can be used comparatively with respect to SWISS model. Phospholipase D homology model after refinement had the best results compared to other models (Remorin 1.4) which have their q mean statistics lower than the optimum range which must be considered for further re-modelling and refinement procedure. HIR protein already had its model submitted in SMR database which looks promising and does not need further change of angles or refinement. Template selection is critical in this process, to have a good template we must have amalgamation of both wet and dry computational lab

experiments. There is a need for reference models too which can be used to produce homology models without much haste.

REFERENCES

[1] Worawat Yuenyong Downstream components of the calmodulin signalling pathway in the rice salt stress response revealed by transcriptome profiling and target identification BMC Plant Biol. 2018; 18: 335.

[2] Hong□Yu Zhang Quantitative iTRAQ□based proteomic analysis of rice grains to assess high night temperature stress Proteomics. 2017 Mar; 17(5): 1600365.

[3] Arabinda Mahanty In-silico docking of natural compounds from plants against Rhizoctonia solani pectate lyase 6 November 2020.

[4] Mustafa R. Morsy the OsLti6 genes encoding low-molecular-weight membrane proteins are differentially expressed in rice cultivars with contrasting sensitivity to low temperature September 2004.

[5] Yongrong Liao Overexpression of a thylakoid membrane protein gene OsTMP14 improves indica rice cold tolerance 30 May 2017.

[6] Sahar Nohzadeh Malakshah Proteomics Reveals New Salt Responsive Proteins Associated with Rice Plasma Membrane September 7, 2007 [doi:10.1271/bbb.70027].

[7] Liming Yang the dynamic changes of the plasma membrane proteins and the protective roles of nitric oxide in rice subjected to heavy metal cadmium stress 26february2016.

[8] Ju Hee Kim Oryza sativa heat-induced ring finger protein 1 (oshirp1) positively regulates plant response to heat stress 07 February 2019.

[9] V. K. Vyas homology modelling a fast tool for drug discovery: current perspectives Indian J Pharm Sci. 2012 Jan-Feb; 74(1): 1–17.

[10] https://tess.elixir-europe.org/workflows/homology-modelling-with-swiss-model Elixir workflow of homology