ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Comparative modeling of Pathogenesis Related Pr5 Protein of *Musa acuminata* with the three-dimensional structure of 1zq3 of *Musa paradaisiaca.*

D Malakonda Reddy, GS Akash, and MS Ekklesia Sesham*.

Department of Biotechnology, KLEF (Deemed To Be University), Vaddeswaram, Guntur-522502, A.P, India

ABSTRACT

The objective of this work is to perform comparative modelling of pathogenesis related protein (PR5) of *Musa acuminata* with the three dimensional structure of chain A protein (1ZQ3). Pathogenesis related proteins are associated with the defensive mechanism, of plants, during pathogenic attack. Usage of antifungal agents, containing PR5 protein, against bacteria and pathogens, will help in increasing the crop yield. This work aims for modelling of PR5 structure using homology modelling, considering the available structure of 1ZQ3 of *Musa paradaisiaca* as a reference. Since PR5 is important in pathogenic activities in plants, against the fungi and bacteria, the three dimensional protein structures was modelled by using Modeller 9.17, a simulator that would help in perceiving the modelled structure better. The PR5 and 1ZQ3 sequences when aligned using BLAST P, showed significant similarity. Comparative modelling of PR5 helps in the study of defensive mechanism in the *Musa acuminata* and its structural integrities.

Keywords: *Musa acuminata*, PR5, 1ZQ3, *Musa paradaisiaca*, Modeller 9.17

Corresponding author

ISSN: 0975-8585

INTRODUCTION

Musa paradaisiaca (banana) known for its nutritional values, is tended to in regions with slight humidity and relatively higher temperatures. The fruits produced by this plant type are oblong and fleshy, while the physical characteristics show pseudostem and elongated leaves with a central midriff. The fruit is generally used for treating diarrhoea, hypertension, cardiac diseases, diabetes etc. [1]. *Musa acuminata* is another species of banana that is found majorly in Southeast Asia, and significant amount of edible variety is contributed from this species [2].

By nature, living things have a tendency to defend, by a built-in defensive mechanism. In plants too, there are defense mechanisms to fight off various biotic and abiotic stress factors. One such mechanism is the ability to produce PR proteins, standing for pathogenesis related proteins. These PR proteins are broadly classified into 16 families, and our present focus PR5 is a part of the PR protein family. These PR5 proteins show antifungal and antibacterial properties, which help the plant to ward off the attacking elements to an extent.

But due to the lack of a three dimensional structure of the PR5 protein, it was proving hard to understand the defense mechanism. Homology modeling is a novel way to get an idea of the 3-D structure of proteins by comparative analyses to form a new structure with reference to a similar already existing crystal structure. With the new advances in computational biology, it has become relatively easier to predict structures using amino acid sequences of proteins. Once the structure of protein is established it helps to understand the working mechanisms of the protein.

PROCEDURE OF WORK DONE:

SEQUENCE RETRIEVAL OF TARGET PROTEINS

Retrieval of the sequence of pathogenesis related protein 5 (PR5) of *Musa paradaisiaca* (Accession Number: I3RTU5; length: 251 amino acid residues) was done from Uniprot in FASTA format[3].

PROTEOMIC ANALYSIS

Protoparam tool of expasy proteomic server was used for the primary structural analysis (http://expasy.org/cgi-bin/protoparam). Aliphatic index, amino acid composition, atomic composition, estimated half-life, extinction coefficient, grand average of hydropathicity (GRAVY), instability index, molecular weight and theoretical pI were the parameters that were computed using Protoparam. Prediction of the secondary structure of PR5 protein was done with the help of PSIPRED server. For better understanding of the domain arrangement within the protein, protein family and super family of protein, the InterPro Scantool (http://www.ebi.ac.uk/Tools/pfa/iprscan/)[5] was used. Pfam was utilized to explore the conserved domains of PR5 protein (http://pfam.janelia.org/)[3][4].

SCREENING FOR THE BEST HOMOLOGOUS TEMPLATE FOR MODELLING OF BANANA PR5

Search for PR5 template was carried out using BLASTP program against PDB database (http://www.rcsb.org/pdb) server. On considering and analysing the alignment scores, query coverage, expected value, similarity and percentage of identity, the best template was chosen[3].

MULTIPLE SEQUENCE ALIGNMENT

An essential tool for prediction of protein structure is multiple sequence alignment (MSA). It also performs other tasks of sequence analysis such as function prediction and phylogeny inference. One such multiple sequence alignment program, T-COFFEE, tree based consistency objective function for alignment evolution, was used to detect the conserved regions of banana PR5[6].

ISSN: 0975-8585

HOMOLOGY MODELLING:

MODELLER 9.17, software, often used for comparative modelling or homology modelling, was used to build three dimensional structure of PR5 using 1ZQ3 of *Musa paradaisiaca* as a template. Sequence alignment with related structure, available on the database provided by the user and the model predicted by MODELLER displays all non-hydrogen atoms. Visualization of the model was later carried out using SPDBV[3][10].

MODEL VALIDATION:

Assessment of quality of model was done using validation tools, such as, PROCHECK[7] and PROSA1. The protein quality was validated, with the help of Ramachandran Plot, by PROCHECK. A program, PROSA1[8], a tool developed for structure prediction, refinement and experimental protein structure validation, was used for the same. The compatibility of a sequence and its structure is measured using Z-score, while keeping in mind that the Z-scores of the model and template should be comparable[3]. Anolea, Qmean and Gromos which are three different molecular dynamic force fields, were used for energy minimization of the three dimensional structures[9][3].

MODEL VISUALIZATION:

The final predicted structure of the model, I3RTU5 was visualized with the help of SPDBV[3].

RESULTS

PR5 (*Musa acuminata*) Sequence Retrieval and Screening for Best Homologous Template:

The Uniprot retrieved sequence of PR5 protein (ID: I3RTU5) was used throughout the study. Homologous template was selected using the search with BLAST against the PDB for target PR5 protein. From blast one sequence selected as a template having 74% identity and 83% positives with target PR5 sequence. The BLAST results present in the Table 1.

Table 1: Results of BLAST search against PDB using I3RTU5 as template.

Protein name	Max score	Total score	Query cover	E value	Ident	Accession
Chain A, Resolution Of The Structure Of The Allergenic And Antifungal Banana Fruit Thaumatin-like Protein At 1.7a	309	309	79%	1.00E-107	74%	1Z3Q A
Chain A, The Crystal Structure Of Zeamatin	261	261	79%	1.00E-88	63%	1DU5 A

PR5 Protein Structure Analysis:

Physiochemical properties of PR5 proteins were studied with the help of protoparam tool. Figures 1 and 2 represent the results of protoparam. The molecular weight of PR5 protein was estimated as 25484.73 daltons (Fig.2). The proportion of amino acids present in the PR5 proteins are Alanine 6.6%, Arginine 3.7% Aspartic acid 5.8%, Asparagine 4.5%, Cysteine 6.6%, Glutamine 2.1%, Glycine 12.8%, Glutamic acid 2.1%, Isoleucine 3.3%, Leucine 5.8%, Lysine 2.9%, Methionine 0.8%, Proline 7.4%, Phenylalanine 7.0%, Serine 9.5, Tryptophan 1.2%, Threonine 8.2%, Tyrosine 2.9%, Valine 7.4% and Histidine, Pyrrolysine, Selenocysteine are 0%. From the protoparam results we concluded that instability index as 27.37; aliphatic index as 63.37 and GRAVY index was 0.040.

Secondary structure of PR5 was studied using PSIPRED (Fig.3) and Interpro. The secondary structure of PR 5 protein contains 3 helix structures and 8 sheets and 1 domain. Domain starts from 28 and ends at 214. The superimposition showed that the structural alignment in the allowed root mean square value and its value is 0.8 A^0 indicates the perfect structural alignment and high quality.

ISSN: 0975-8585

```
Number of amino acids: 243

Molecular weight: 25484.73

Theoretical pI: 4.91
```

Figure 1: Primary structure analysis of I3RTU5 using protoparam.

```
Amino acid composition:
Ala (A)  16       6.6%
Arg (R)   9       3.7%
Asn (N)  11       4.5%
Asp (D)  14       5.8%
Cys (C)  16       6.6%
Gln (Q)   5       2.1%
Glu (E)   5       2.1%
Gly (G)  30      12.3%
His (H)   0       0.0%
Ile (I)   8       3.3%
Leu (L)  14       5.8%
Lys (K)   7       2.9%
Met (M)   2       0.8%
Phe (F)  17       7.0%
Pro (P)  18       7.4%
Ser (S)  23       9.5%
Thr (T)  20       8.2%
Trp (W)   3       1.2%
Tyr (Y)   7       2.9%
Val (V)  18       7.4%
Pyl (O)   0       0.0%
Sec (U)   0       0.0%

 (B)   0          0.0%
 (Z)   0          0.0%
 (X)   0          0.0%


Total number of negatively charged residues (Asp + Glu): 19
Total number of positively charged residues (Arg + Lys): 16
```

Figure 2: Amino acid composition in I3RTU5 (protoparam).

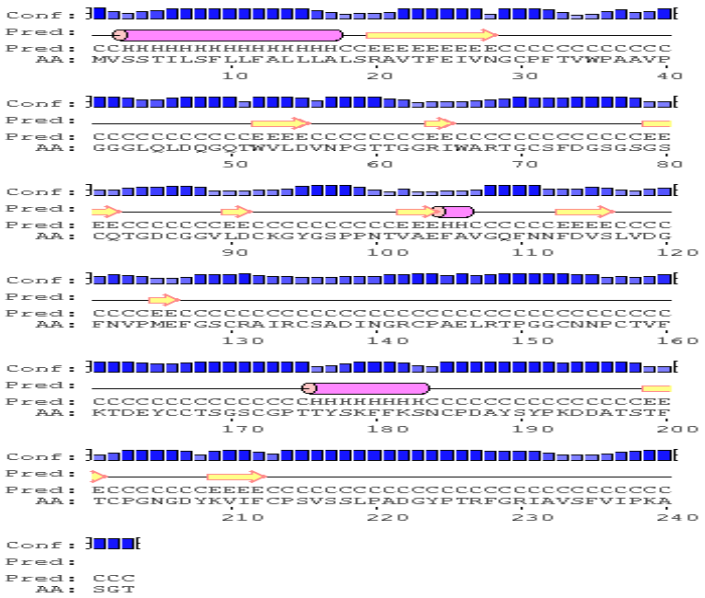

Figure 3: Secondary structure analysis (PSIPRED).

Multiple Sequence Alignment:

Multiple sequence alignment performed by the tool T-COFFEE states that the highly conserved amino acids with good identity are marked with asterisk symbol. Results of T-COFFEE (Fig.4) shows 99% consensus region between the target and template regions and these regions represents the functional domains of PR5 protein.

ISSN: 0975-8585

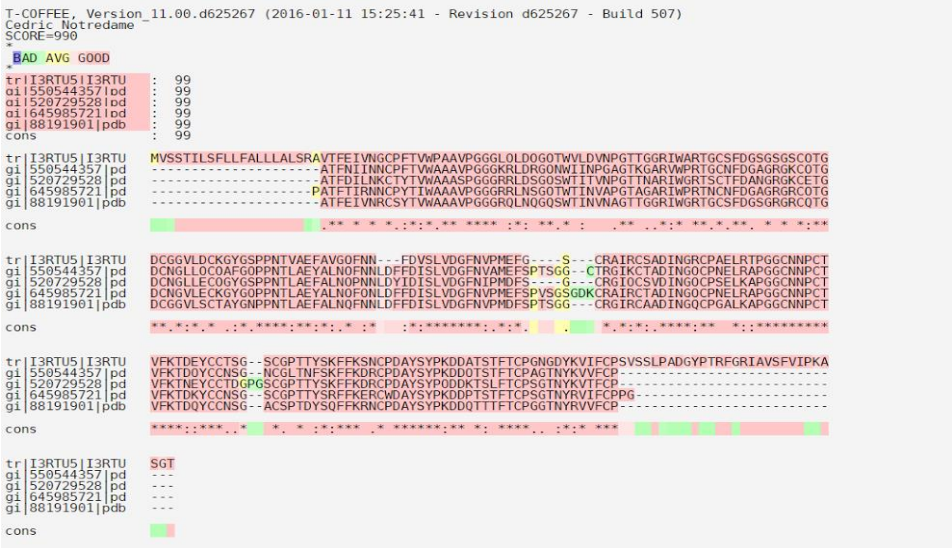

Figure 4: Multiple sequence alignment (T-COFFEE).

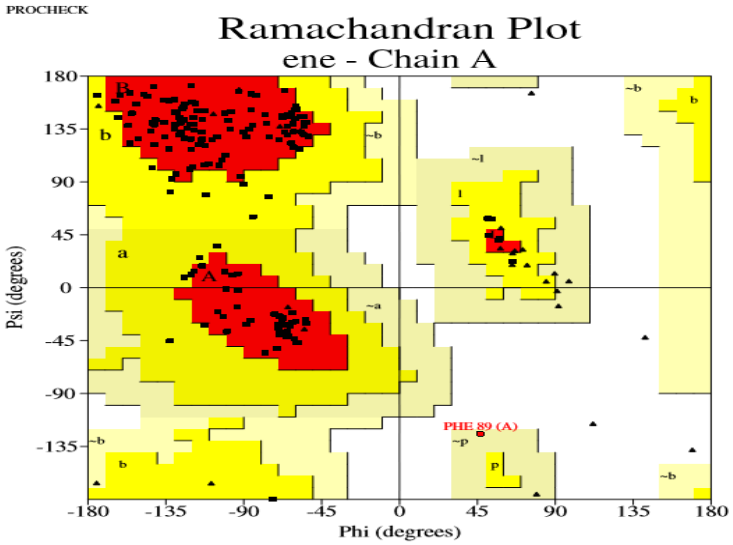

Figure 5: Ramchandran plot (PROCHECK).

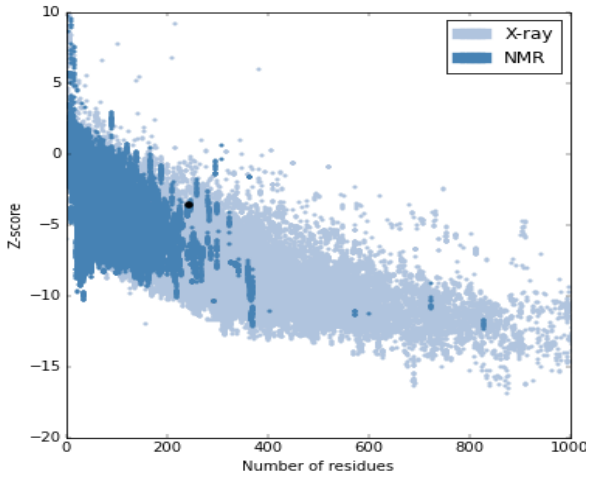

Figure 6: PROSA analysis (z-score).

ISSN: 0975-8585

Homology Modelling of PR5 Protein:

The assumed/proposed model of PR5 protein had been modelled using MODELLER 9v17 since the structure of PR5 proteins was unavailable. The 3D structure was generated by using the sequence of PR5 protein and sequence of chain A , allergenic and antifungal banana fruit thaumatin like protein of *Musa acuminata* (1ZQ3) as the target and templates respectively. The structure of PR5 protein which was generated using modeller 9v17visualized using SPDBV(fig.7).

Figure 7: Modelled structure of PR 5 protein using modeller 9v17.

Validation of modelled structure:

The evaluation of quality of generated structure was done by various tools to check the reliability of the model as well as the internal consistency. The best model selected based on the validity which was checked using PROCHECK analysis[7], PROSA1[8] and the molecular dynamic force fields which are ANOLEA[] and Qmean[]. Stereochemical quality of polypeptide backbone and the side chains were studied based on the Ramchandran plot (Fig.5) generated during the PROCHECK analysis. It revealed that the most favoured regions showed residues present as 87.4%, in additional allowed regions, the residues present were given as 11.9%, in generously allowed regions and in disallowed regions residues present as 0.6% and 0% respectively. The energy of the structure is minimized by using dynamic force fields like ANOLEA, Qmean and GROMOS. Model energy profile and the Z-score was detected using PROSA server and the interaction energy per residues distance based pair potential. The analysis of structure using PROSA shows (Fig.6) Z-score of -4.28 indicating the model is good in quality[3].

DISCUSSION

The homology model of PR5 protein of *Musa paradaisiaca* having 243 amino acids length is described. Protein structure which is determined by the experimental procedure is necessary for the comparative homology modelling. The template sequence for homology modelling of 1ZQ3 of banana (*Musa acuminata*) was retrieved from the Uniprot. This protein is closely related with the PR5 protein with 74% identity, 83% positives, 1.00E-107 e-score and 309 alignment score. The template was chosen by taking into considerations the percentage of identity, similarity, e-value, and alignment score. Maximum identity and the lesser e-values

are the means for the generation of good structure. 1ZQ3 selected as a template because it shows maximum homology with the targeted protein. Proteomic analysis provides vast information about the target protein. The information about instability index as 27.37, given by Protoparam explains the stable nature of the protein. High aliphatic index (63.37) predicts the stability of protein at wide range of temperature and low GRAVY index (0.040) explains the high affinity of the protein for water[3]. PSIPRED was used for the secondary structure of PR 5 and the information regarding the domains in the target protein was taken from Interpro. T-COFFEE was used for confirmation of the homology between the target PR5 and template 1ZQ3 with consensus regions of 99%. Accurate method for prediction of 3d structure for the target PR5 protein is comparative homology modelling. Comparative homology modelling utilizes the homology between the target and template sequences because the similar sequences share similar tertiary structures. PROCHECK is used to generate the Ramchandran plot which splits the protein into four areas, which are, additional allowed, disallowed, generously allowed and most favoured regions. Good models have most residues in the most favored regions and least residues in the disallowed regions. The model structure of PR5 proteins shows 86.7 % and 0% of residues in most favored regions and disallowed regions, respectively. The reliability and internal consistency of the protein model structure was estimated using the PROSA analysis. Z-score value was predicted as -4.28. The negative of the Z-value represents the reliability and quality of the modelled structure[3]. The above stated results reveal that the model generated using the homology modelling is most acceptable and useful in future studies. Validation of the generated model suggested the resulting three dimensional structure of PR5 protein as reliable.

CONCLUSION

In the carried out procedure, 3D structure of PR5 protein has been generated with the help of available protein structure which has the more identity with PR5 protein. 3D structure of PR5 protein generated using homology modelling which is useful in understanding the defensive mechanism during the fungal attack in banana and in other plants which have the similar mechanism and shares the homology. Based on the similarity between the I3RTU5 (PR5 protein) and 1ZQ3, we can include PR 5 protein in the protein family which containing the 1ZQ3 protein. Elucidation of structure of PR5 proteins enables us to study novel structure of PR5 protein, extraction and prediction of protein from its sources.

REFERENCES

[1] Mohammad Zafar Imam and Saleha Akter, Musa paradisiaca L. and Musa sapientum L.: A Phytochemical and Pharmacological Review, Journal of Applied Pharmaceutical Science 01 (05); 2011: 14-20.

[2] Carol Wong, Ruth Kiew, Jin Phang Loh, Leong Huat Gan, Ohn Set, Sing Kong Lee, Shawn Lum and Yik Yuen Gan, Genetic Diversity of the Wild Banana Musa acuminata Colla in Malaysia as Evidenced by AFLP, Annals of Botany 88; 2001: 1017-1025 .

[3] Showmy K.S1, Sinosh Skariyachan and Yusuf A. “Comparative modelling of pathogenesis related 4b protein (Q6T5J8) of oryza sativa subsp. Indica with the three-dimensional structure of barley 1bw3”. *International journal of plant, animal and environmental sciences*, 2014, 4(4).

[4] Finn RD, Mistry J, Tate J, et al. (2010). The Pfam protein families database. Nucleic Acids Res 38, D211–222.

[5] ZdobnovEM, and Apweiler R. (2001). InterProScan—An integration platform for the signature-recognition methods in InterPro. Bioinformatics 17, 847–848.

[6] Notredame, C., Higgins, D.G., Heringa, J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J MolBiol, 302, pp. 205-217.

[7] Laskowski, R.A., Macarthur, M.W., Thornton, J.M. 1993. PROCHECK: A program to check the stereochemical quality of protein. J ApplCryst 26, pp. 283-291.

[8] Markus, W., Manfred, J. S. 2007. ProSA-web: interactive web service for the recognition of errors in three dimensional structures of proteins. Nucleic Acids Research, 35, pp. 407-410.

[9] Pascal, B., Michael, K., Torsten, S. 2009. QMEAN server for protein model quality estimation. Nucleic Acids Research, Vol. 37, web server issue.

[10] Andrej, S., Tom, L. B. 1993. Comparative protein modeling by satisfaction of spatial restraints J. Mol. Biol., 5, pp. 779–815 .