Volume 64, Issue 2, 2020
Journal of Scientific Research
Institute of Science,
Banaras Hindu University, Varanasi, India.

Structure Prediction of 30s Ribosomal S3, RNA Polymerase II Proteins of *Phoenix Pusilla* using Bioinformatics Tools

S. Vijaya Bharathi[1,2] and V. Anuradha[*2]

[1]Department of Biotechnology, FSH, SRM IST, Kattankulathur, Chennai. shavisa04@gmail.com
[2]PG & Research Department of Biochemistry, Mohammed Sathak college of Arts & Science, Shollinganallur. vanuradha.2003@gmail.com*

Abstract: **Proteins have different functions like defense action, structural function, transport action, intracellular signalling etc. As proteins are important in many ways its composition analysis, structure and function determination is required. Protein structure prediction is necessary for drug designing, protein engineering and prediction of binding sites. *Phoenix pusilla* proteins primary structure prediction was done with Expasy protoparam, secondary structure with SOSUI, tertiary structure with swiss model and transmembrane region was identified with TMHMM v 2.0. Secondary structure prediction of 30SRibosomal protein S3 and RNA polymerase II, showed that α – helix, random coil, β – turn and extended strand predominates. Transmembrane region prediction showed that both the proteins were soluble form. As RNA polymerase II of *Phoenix pusilla* has only 27aminoacid residues, 3D structure has to be predicted with other logarithmic tool and so 30Sribosomal protein 3D structure was predicted with Swiss model server.**

***Index Terms*: *Phoenix pusilla*, SOPMA, TMHMM v 2.0.**

I. INTRODUCTION

Phoenix pusilla (PP) is a stemless plant found in India and Srilanka belongs to arecaceae family, commonly called as small date palm. It is seen in open area along with the other plants in clumps (Kinhal and Parthasarthy, 2010). All the parts of PP have traditional uses, leaves are used to make mats, brooms, baskets, fruits are used to treat fever, pith is used in treating gonorrhoea and eaten as food by starving people. Roots are used as disposable toothbrush (Roop, 2018). Structure prediction is an important aspect which helps in drug discovery protocol. Nowadays various online tools are available to predict the structure. But selection of the tool has to be done properly so that accuracy of prediction can be maintained. Tertiary structure prediction tools include comparative modelling, threading based and free modelling (Zhang, 2009). Comparative and threading methods are referred as template based method. In template based methods the structure is predicted by identifying the unfolded structure which are similar or homologous to the protein to which the structure has to be predicted whereas in free modelling the structure is constructed by fragment assembly (Deng et al., 2018). In NCBI database three different proteins entry were seen for *Phoenix pusilla*. 30SRibosomal protein S3 plays important role in synthesizing chloroplast genome proteins (Beri et al., 2017). RNA polymerase II second largest subunit seen in *Phoenix pusilla* like other RNA polymerase involves in transcription process (Naryshkina, 2003). For NBS- LRR resistance protein the structure was available. So in this present study, structure of two proteins (30s ribosomal protein S3, RNA polymeraseII) of dwarf date palm was predicted.

II. MATERIALS AND METHODS

A. Sequence Retrieval

Sequence of the *Phoenix pusilla* proteins were downloaded from NCBI database in FASTA format (https://www.ncbi.nlm.nih.gov/).

B. Primary Structure Prediction

For Physio-chemical characterization, theoretical Isoelectric Point (pI), molecular weight, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index and grand average of hydropathy (GRAVY) were computed using the Expasy Protparm server (Pramanik et al., 2017; Akilesh et al., 2012). Primary structure analysis is important in understanding the biochemical and cellular function of protein (Toomula et al., 2011).

* Corresponding Author

C. Secondary structure prediction

SOPMA (Self Optimized Prediction Method with Alignment) was used for the 2° structure prediction. Neural network method (PHD) is the basis of this tool. In this method, the algorithm first searches the database for the homologous proteins, then filters the close homologues and finally submit the sequence, alignment data for structure prediction (Toomula et al., 2011). Secondary structure prediction is important for identifying the conformational changes in the target protein (Roy et al., 2015).

D. Swiss model

3D structure prediction was done with swiss model (Schwede et al., 2003). Swiss model online tool was developed by Torsten Schwede's structural bioinformatics group. The process in this online tool includes template recognition, target-template alignment, model building and model evaluation (Deng et al., 2018). Predicted structure were validated with PROCHECK (https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/) and ProSA web service (https://prosa.services.came.sbg.ac.at/prosa.php).

E. Functional characterization

TMHMM v.2.0 tool was used to identify the transmembrane region in the protein (Venkatasamy, 2013). Transmembrane proteins are important in performing functions like signalling, transport etc and also important drug target (Yu and Zhang, 2013).

III. RESULTS & DISCUSSION

FASTA sequence of both the proteins for which the structure prediction has to be done was retrieved from NCBI database (Table I). pI of 30s Ribosomal protein showed that it is an basic protein whereas RNA polymerase is an acidic protein. Isoelectric point (pI) of protein is the point where protein exists in zwitter ion for and so it is important in protein purification (Filiz and Koc, 2013). Also pI information of a protein helps in developing a proper buffer system for its purification. Instability index of both the protein was less than 40 which revealed that both the proteins were stable. Protein stability in lab condition depends on instability index (Gosh et al., 2017). Aliphatic index of 30s Ribosomal protein was 107.36 and RNA polymerase aliphatic index was 75.93. Aliphatic index provides information about the aliphatic side chain amino acids occupied and temperature of protein at which the protein is stable (Valli and Mythili, 2018). Aliphatic index value by Expasy Protparm tool revealed that the value of both the proteins of *Phoenix pusilla* showed that these proteins may be stable at wide range of temperature. Lowest GRAVY value of the proteins indicates better interaction of protein with water (Akhilesh et al., 2012) (Table II). Among the two proteins, 30S ribosomal protein showed lowest Gravy value (-0.326) than RNA polymerase II. Secondary structure prediction revealed that in both the proteins α – helix, random coil, β – turn and extended strand predominates (Fig. 1, 2). Alpha helix percentage was high when compared with the other conformations seen in both the proteins. In 30SRibosomal protein S3, alpha helix showed highest percentage, followed by random coil, extended strand and finally beta turn. Similarly in RNA polymerase II second largest subunit, high percentage of alpha helix conformation was followed by random coil but the beta turn percentage was higher than extended strand which was found reverse in the ribosomal protein. Transmembrane region was not found in 30Sribosomal proteinS3, chlorolastic protein as well as in RNA polymerase II. The oligostate of the predicted model was monomer (Fig. 3). Q mean score showed -2.25. Tertiary structure of proteins will give information about their functions and also its application in life science research (Waterhouse et al., 2018). Ramachandran plot showed that the percentage of residues in the favoured region was 90.8 (Fig. 4). Normally validation of 3D structure showing atleast 90% residues are considered as preferred structure (Laskoeski et al., 1993). ProSA web service Z-score of the predicted structure was -5.9 (Fig.5). Positive value is the indicator of error in the model (Prajapat et al., 2014). So the predicted structure of 30S ribosomal protein, chloroplast may be an appropriate model for studying the importance of the protein.

Table I FASTA sequence of proteins

Protein/ Accession number	FASTA sequence
30S ribosomal protein S3, partial (chloroplast)/ AGC11927.1	QTKNYSVSLQEDEKIRNCIKNYVQKNRRI SSGFEGIARIGIQKRIDLIQVIIHIGFPNLLM EGQTRGIEELQMNVQKELHSVNRRLNIAI TRIEKPYGQPNILAEYIALQLKNRVSFRK SMKKAIELTEQTDTKGIQVQIAGRIDGKE IARVEWIREGRVPLQTIRAKIDHCSYTIRT IYGVLGIKIWIFVNEFGYL
RNA polymerase II second largest subunit, partial/ ADF45488.1	QLIECIMGKVAAHMGKEGDATPFTDVT

Mol. Wt – molecular weight(Daltons), pI-Isoelectric pH, -R - Number of negative residues, +R – Number of Positive residues, GRAVY – Grand Average Hydropathicity, * - No Trp, Tyr or Cys residue (should not be visible by UV spectrophotometry.

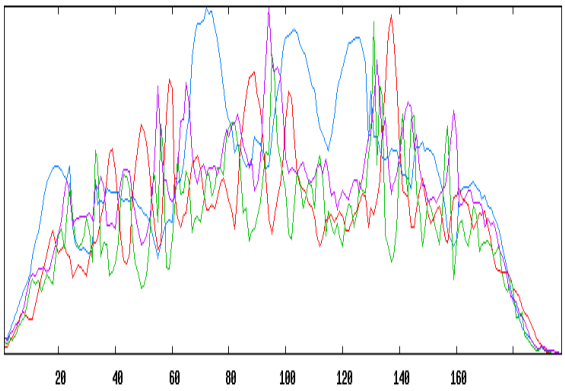


Fig. 1 Secondary structure of 30s Ribosomal protein S3

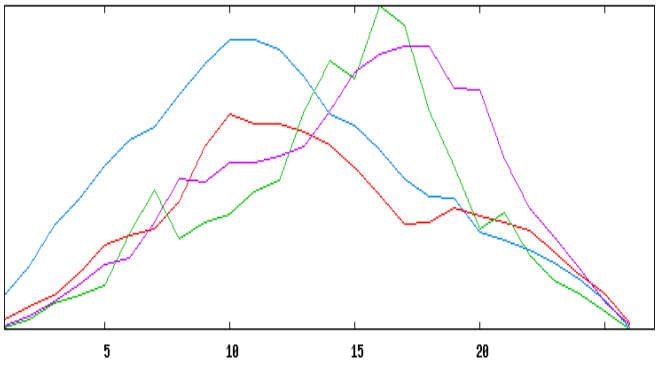


Fig. 2 Secondary structure of RNA polymerase II

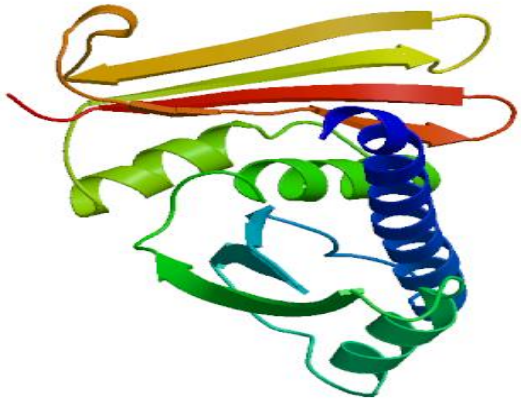


Fig. 3 3D Structure of 30S Ribosomal protein S3

Table II Protoparam parameters

Protein	Length	Mol.wt	pI	- R	+ R	Extinction Coefficient at 280 nm	Instability index	Aliphatic index	GRAVY
30SRibosomal protein S3, chloroplastic	197	22850.65	9.85	20	32	21555	39.53	107.36	-0.326
RNA polymerase II second largest subunit	27	2863.31	4.75	4	2	0	33.59	75.93	0.085

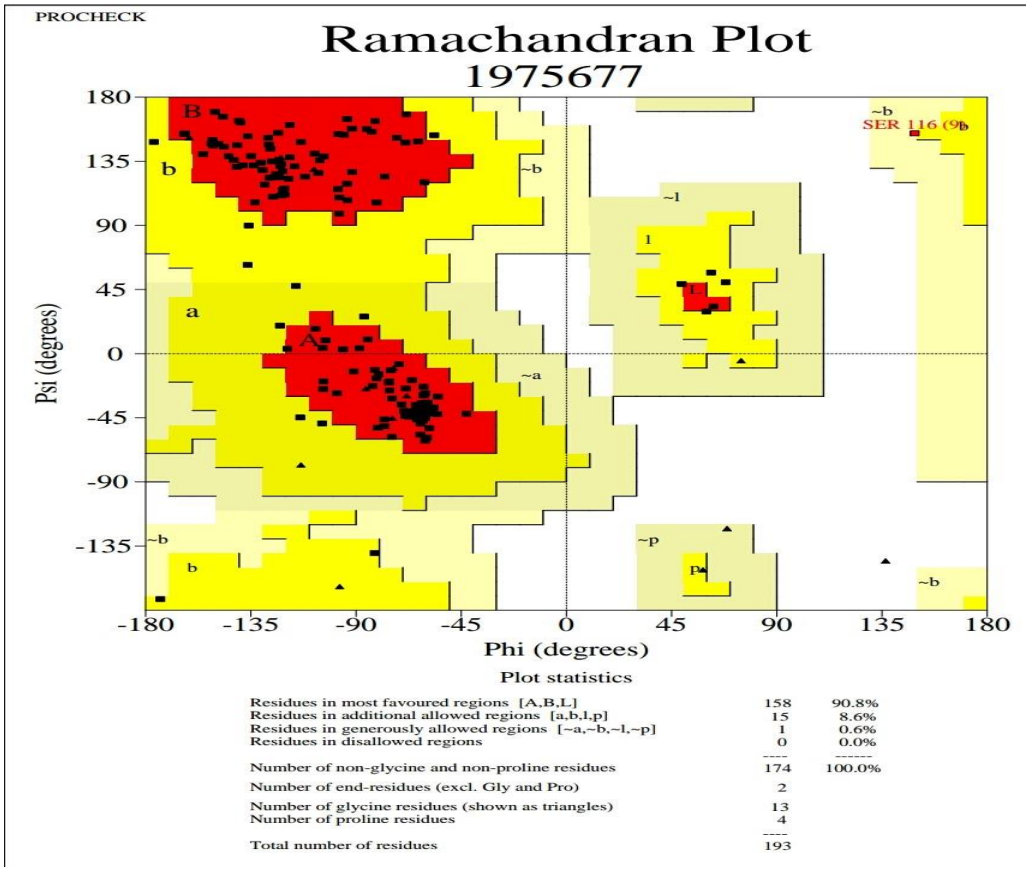
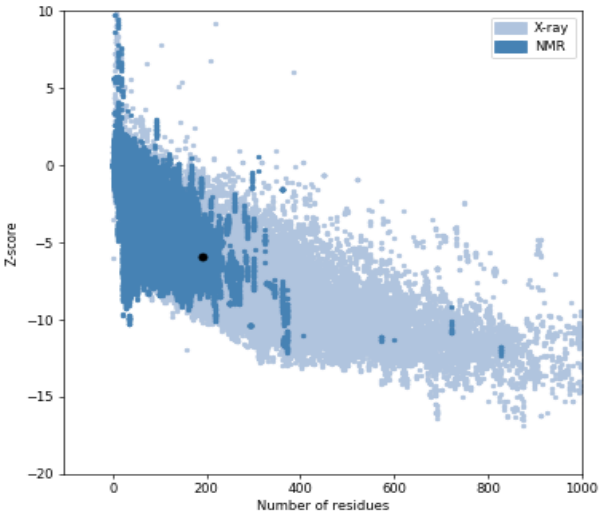


Fig. 4 Ramachandran plot



A. *B.* *C.*

Fig. 5 ProSA web plot

CONCLUSION

Protein structure plays important role in determining its function. *Phoenix pusilla* is a traditional medicinal plant, also proved to grow in extreme temperatures, prevents soil erosion. Structure predicted for the protein 30 s ribosomal protein in this study showed an acceptable Z-score. As 30S ribosomal protein, chloroplast is associated with photosynthesis the structure predicted for this protein may be useful in studying about the translation mechanism in chloroplast of *Phoenix pusilla*.

REFERENCES

Akhilesh, U., Bhaskar., Anusha., Sakshi, I., and Selvi, V. S. (2012). In Silico Characterization of Keratitis Causing Herpes Simplex Virus (HSV 1) Membrane Proteins using Computational Tools and Servers. Research Journal of Recent Sciences, 1(11), 27-31.

Bieri, P., Leibundgut, M., Saurer, M., Boehringer, D., and Ban, N. (2017). The complete structure of the chloroplast 70S ribosome in complex with translation factor pY. EMBO J, 36(4), 475-486.

Deng, H., Jia, Y., and Zhang, Y. (2018). Protein structure prediction. International Journal of Modern Physics, B 32, 1-17.

Filiz, E., and Koc, I. (2013). In silico analysis of dicer-like protein (DCLs) sequences from higher plant species. IUFS J Biol, 72, 53-63.

Ghosh, R., Upadhayay, A. D., and Roy, A. K. (2017). In silico analysis, structure modeling and phosphorylation site prediction of vitellogenin protein from Gibelion catla. JAppl Biotechnol Bioeng, 3(1), 265-270.

Kinhal, V and Parthasarathy, N. (2010). Nucleated succession by an endemic palm Phoenix pusilla enhances diversity of woody species in the arid Coromandel Coast of India. AoB PLANTS, 2010, plq007.

Laskowski, R. A , MacArthur., Moss, D. S , and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures", Journal of Applied crystallography, 26(2), 283-291.

Naryshkina, T., Bruning, A., Gadal, O., and Severinov, K. (2003). Role of Second-Largest RNA Polymerase I Subunit Zn-Binding Domain in Enzyme Assembly. Eukaryotic cell, 2(5), 1046–1052.

Prajapat, R., Marwal, A., and Gaur, R. K. (2014). Recognition of Errors in the Refinement and Validation of Three-Dimensional Structures of AC1 Proteins of Begomovirus Strains by Using ProSA-Web. Journal of Viruses, 2014, 6pages.

Pramanik, K., Ghosh, K.P, Ray, S., Sarkar, A., Mitra, S., and Maiti, K. T. (2017). An in silico structural, functional and phylogenetic analysis with three dimensional protein modeling of alkaline phosphatase enzyme of Pseudomonas aeruginosa. Journal of Genetic Engineering and Biotechnology, 15, 527–537.

Roop J.K. (2018). Phoenix pusillaGaertn. - A Review. IJSRR, 7(1), 496–503.

Roy, S., Banerjee, V., and Das, K. P. (2015). Understanding the Physical and Molecular Basis of Stability of *Arabidopsis* DNA Pol λ under UV-B and High NaCl Stress. PLoS ONE, 10(7).

Schwede, T., Kop, J., and Peitsch, M. C. (2003). SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Research, 31(13), 3381–3385.

Toomula, N., Kumar, D.S, Kumar, A.V.V.L.P. (2011). Computational Methods for Protein Structure Prediction and Its Application in Drug Design. J Proteomics Bioinform, 4, 12.

Valli, S. A., and Mythili. (2018). Bioinformatic study of an antitumor protein, Azurin. Asian J Pharm Clin Res, 11(6), 169-176.

Venkatasamy, K. (2013). In silico Analysis and Homology Modeling of Putative Hypothetical Protein Q4QH83 of Leishmania major. Advanced BioTech, 13 (3), 1-4.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, T. F, Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Research, 46, W296- W303.

Yu, B., and Zhang, Y. (2013). A simple method for predicting transmembrane protens based on Wavelet transfom. International journal of biological Science, 9(1), 22-33.

Zhang, Y. (2009). Protein Structure Prediction: Is It Useful?. Curr Opin Struct Biol, 19(2), 145–155.

***