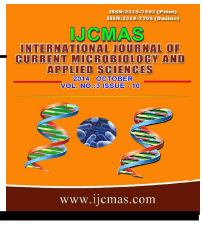
International Journal of Current Microbiology and Applied Sciences
ISSN: 2319-7706 **Volume 3 Number 10 (2014) pp. 177-187**
http://www.ijcmas.com



Original Research Article

Structure determination of Leghemoglobin using Homology Modeling

N.Dhurigai*, Rachel Regi Daniel and L.Rufus Auxilia

Department of Botany and Microbiology, Lady Doak College, Madurai, Tamilnadu, India
**Corresponding author*

K e y w o r d s

Leghemoglobin, Leguminous plants, Homology modeling, SWISS-MODEL, SAVeS

A B S T R A C T

Leghemoglobins are heme-proteins, first identified in root nodules of leguminous plants, where they are crucial for supplying sufficient oxygen to root nodule bacteria for nitrogen fixation to occur (Vainshtein *et al*., 1975 and Downie, 2005). The protein sequence of Leghemoglobin of plants *Psophocarpus tetragonolobus* and *Vigna unguiculata* were retrieved from the NCBI Protein Database. Tertiary structures were determined by homology modeling software 'SWISS-MODEL' an automated system for modeling the 3D structure of a protein from amino acid sequence using homology modeling technique (Biasini *et al*., 2014). The modeled structures were evaluated using SAVeS. The results showed that the structures obtained were reliable and could be used in the future studies.

Introduction

Protein sequencing is increasing at exponential rate, till date more than one million proteins are sequenced and are available in the databases (SwissProt, Protein Research Foundation and Protein Information Resource). The only protein structure database is PDB (Protein Data Bank). The gap between the number of protein sequences and protein structures has been increasing (Zhang, 2009). Determining the 3D structure of a protein sequence is a difficult and challenging problem. Basically X-ray crystallography or Nuclear Magnetic Resonance (NMR) techniques are used, which are expensive, time consuming and complex process. Therefore, computational methods/ algorithms including homology modeling, threading, and *ab initio* are developed. Homology modeling is the most accurate one.

The ultimate goal of protein modeling is to predict a structure from its amino acid sequence with an accuracy that is comparable to the best results achieved experimentally. This would allow users to safely use rapidly generated *in silico* protein models in all the contexts where today only experimental structures provide a solid basis (Krieger *et al.*, 2003).

The structure of a protein is uniquely determined by its amino acid sequence. Knowing the sequence should, at least in theory, suffice to obtain the structure. During evolution, the structure is more

stable and changes much slower than the associated sequence, so that similar sequences adopt practically identical structures and distantly related sequences still fold into similar structures. This relationship was first identified by Chothia and Lesk (1986) and later quantified by Sander and Schneider (1991). As long as the length of two sequences and the percentage of identical residues fall in the region marked as "safe," the two sequences are practically guaranteed to adopt a similar structure (Krieger *et al.,* 2003). If the similarity of the sequence is greater than 30% then it can act as template (Brindha *et al*., 2011).

SWISS-Model

(http://swissmodel.expasy.org/) is an automated system for modeling the 3D structure of a protein from its amino acid sequence using homology modeling techniques. SWISS-MODEL has been established 20 years ago as the first fully automated server for protein structure homology modeling and has been continuously developed and improved since then. The server features a user-friendly web interface, which allows also non-specialists to generate 3D models for their protein of interests from a simple web-browser without the need to install and learn complex molecular modeling software or to download large databases (Biasini *et al*., 2014).

Legumes are unique among crop plants (Becana *et al.,* 1995). A hallmark trait of legumes is their ability to develop root nodules and fix N_2 in symbiosis with compatible rhizobia (Graham and Vance, 2003). Nitrogen-fixing legume nodules exclusively synthesize a large amount of an oxygen-binding monomeric hemoprotein, leghemoglobin which belongs to globin family of protein (Brisson and Verma, 1982; Kawashima *et al.,* 2001). Owing to its extremely fast O_2 association rate and rather slow O_2 dissociation rate, leghemoglobin regulates O_2 tension in the nodule. Thereby it protects the O_2 sensitive nitrogenase from inactivation while still supporting the oxidative respiration of bacteroids (Kawashima *et al.,* 2001; Leong *et al.,* 1982). From 1983 to 1997 a total of 19 structures of leghemoglobin have been released which belongs to two plants namely *Glycine max* and *Lupinus luteus.* Since then no structure has been released yet for the rest of the plants in the NCBI protein database.

Validation of the predicted structures is essential because every homology model contains errors. SAVeS possesses the tools for validation of homology models namely PROCHECK, WHAT_CHECK, ERRAT, VERIFY 3D and PROVE. PROCHECK checks the stereo chemical quality of a protein structure by analyzing residue-by-residue geometry and overall structure geometry (Laskoswki *et al*., 1993). WHAT_CHECK derived from a subset of protein verification tools from the WHATIF program; this does extensive checking of many stereo chemical parameters of the residues in the model (Hooft *et al*., 1996). ERRAT analyzes the statistics of non-bonded interactions between different atom types and plots the value of the error function versus position of a 9-residue sliding window, calculated by a comparison with statistics from highly refined structures (Colovos *et al*., 1993). VERIFY 3D determines the compatibility of an atomic model (3D) with its own amino acid sequence (1D) by assigned a structural class based on its location and environment (alpha, beta, loop, polar, nonpolar etc) and comparing the results to good structures (Luthy *et al*., 1996). PROVE calculates the

volumes of atoms in macromolecules using an algorithm which treats the atoms like hard spheres and calculates a statistical Z-score deviation for the model from highly resolved (2.0 Å or better) and refined (R-factor of 0.2 or better) PDB-deposited structures (Pontius *et al*., 1996).

Thus an attempt was made to predict and validate the 3D structure of leghemoglobin for amino acid sequence of the plants namely *Psophocarpus tetragonolobus* and *Vigna unguiculata* available in NCBI protein sequence database. This could be used in further studies of leghemoglobin.

Materials and Methods

1. Selection of Target Sequence:

The leghemoglobin protein sequence of the leguminous plants namely *Psophocarpus tetragonolobus* and *Vigna unguiculata* was retrieved from NCBI Protein sequence database, in FASTA format.

2. Structure Prediction using Homology Modeling:

Homology modeling (or comparative modeling) relies on evolutionarily related structures (templates) to generate a structural model of a protein of interest (target). The process typically comprises the following steps: (i) Template identification, (ii) Template selection, (iii) Model building and (iv) Model quality estimation. (Biasini *et al*., 2014)

- Swiss model "Automatic modeling mode" was selected
- Protein sequence was entered in FASTA format in the space provided
- Modeling request was submitted.
- The constructed model was saved as a PDB file and was subjected to assessment.

3. Model Validation:

- Model was validated using the tools of Structural Analysis and Verification Server (SAVeS) available in the link http://nihserver.mbi.ucla.edu/SAVES/.
- The PDB file was selected and uploaded
- All the programs namely PROCHECK, WHAT_CHECK, ERRAT, VERIFY 3D and PROVE were selected and the command was given to Run all the programs.
- The results obtained were saved for further analysis.

4. Model Visualization:

The Model thus obtained was visualized by using RASMOL. PDB file was opened with RASMOL to visualize the protein.

Results and Discussion

1. Selection of Target sequence:

The FASTA format sequences of Leghemoglobin protein for the two plants are as follows:

>gi|20953|emb|CAA46704.1| leghemoglobin [*Psophocarpus tetragonolobus*]

MGGFTEKQEALVNSSYEAFKANVP
QYSVVFYTSILEKAPAAKDLFPFLA
NGVDPTNPKLIGHAEKLFGLVHDS
AAQLRAKGAVVADAALGSLHAQK
GVTDPQFVVVKEALLKTVKEAVG
DKWSDELSNAWEVAYNELAAALK
KAF

>gi|1177057|gb|AAA86756.1|
leghemoglobin I [*Vigna unguiculata*]
MVAFSDKQEALVNGAYEAFKANIP
KYSVVFYTTILEKAPAAKNLFSFLA
NGVDATNPKLTGHAEKLFGLVRDS
AAQLRASGGVVADAALGAVHSQK
AVNDAQFVVVKEALVKTLKEAVG
DKWSDELGTAVELAYDELAAAIK
KAY

2. Structure Prediction using Homology Modeling

For Both the plants the template was 1binB which belongs to Soybean plant. And the percentage identity of sequence was above 80% which is more than enough for Homology modeling.

3. Model Validation

The Ramachandran plot revealed that the modeled structure of Leghemoglobin of both the plants as a good and reliable model (Table 2, Fig I& II).

The ERRAT analysis for the modeled proteins revealed that the overall quality factor of the Leghemoglobin of *Vigna unguiculata* as 100.000 and that of *Psophocarpus tetragonolobu*s as 100.000 making them as good quality models. (Fig III, IV)

PROVE plot was used to calculate the atoms in the modeled structure of Leghemoglobin (Table 3, Fig V and VI)

4. Model Visualization

5.

The modeled leghemoglobin protein structures of *Psophocarpus tetragonolobus* and *Vigna unguiculata* were visualized using RASMOL. (Fig. VII and Fig VIII)

Being a hemo-protein Leghemoglobin possesses the protein and heme cofactor (an iron atom bound in a porphyrin ring). In both of the plants the modelled 3 D structure contains both the protein part and the heme part containing iron atom bound in a porphyrin ring.

Significance of this structure prediction is as follows:

- Both the plants belong to the Tribe Phaseoleae, and the leghemoglobin structure has been determined by automated mode of SWISS MODEL having 1binB as template. This template is of the plant *Glycine max* which again belongs to the same Tribe Phaseoleae. By this the ancient method of classifying the plants under the same tribe is supported by modern evidence given by this bioinformatics study. (Chemo taxonomy linked with bioinformatics work of studying 3D model of Leghemoglobin protein)
- Three dimensional structure of Leghemoglobin protein of both of these plants namely *Psophocarpus tetragonolobus* and *Vigna unguiculata* have not been experimentally determined so far and are not available in RCSB Protein Data Bank. Now by doing this Homology modeling, 3D structure has been predicted for the Leghemoglobin protein of both the plants.
- Through this study validity of proposed model was confirmed.

Table.1 Information of Modeled Structure

Information	Plants	
	Psophocarpus tetragonolobus	***Vigna unguiculata***
Modeled residue range	3 to 144	2 to 144
Based on template	1 bin B (2.20A)	1 bin B (2.20A)
Sequence identity %	85.915	81.818
E value	545252e-57	325602e-51
Q Mean Z Square	0.049	-0.213

Table.2 Ramachandran Plot

Plot Statistics	Plants	
	Psophocarpus tetragonolobus	***Vigna unguiculata***
Residues in most favoured region	92.9%	92.9%
Residues in additional allowed region	6.3%	7.0%
Residues in generously allowed region	0.8%	0.8%
Residues in dis-allowed region	0%	0%

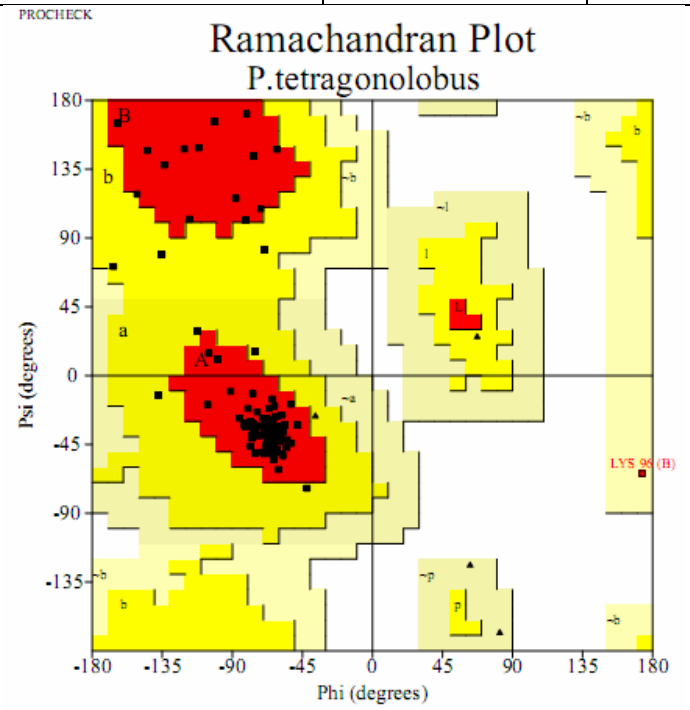


Fig.I Ramachandran plot of *Psophocarpus tetragonolobus*

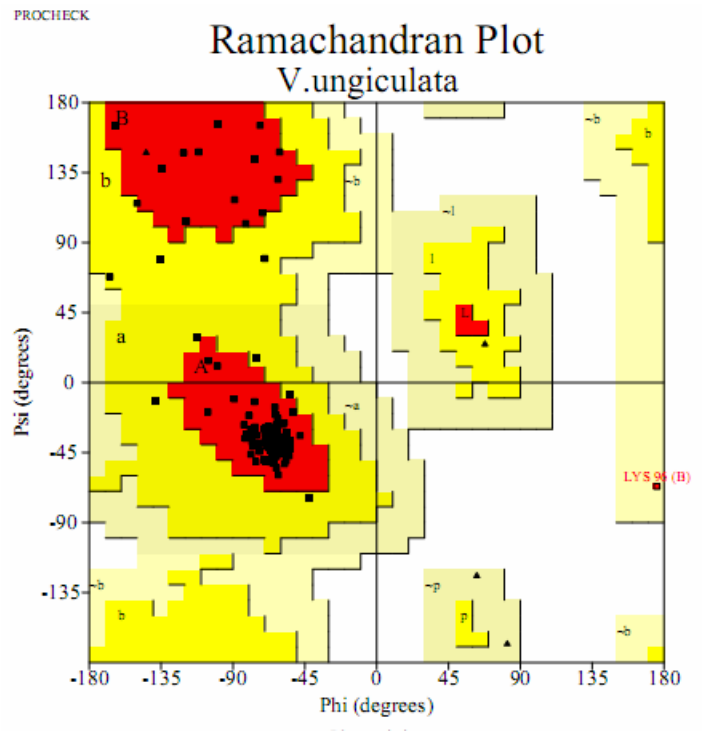


Fig.II Ramachandran plot of *Vigna unguiculata*

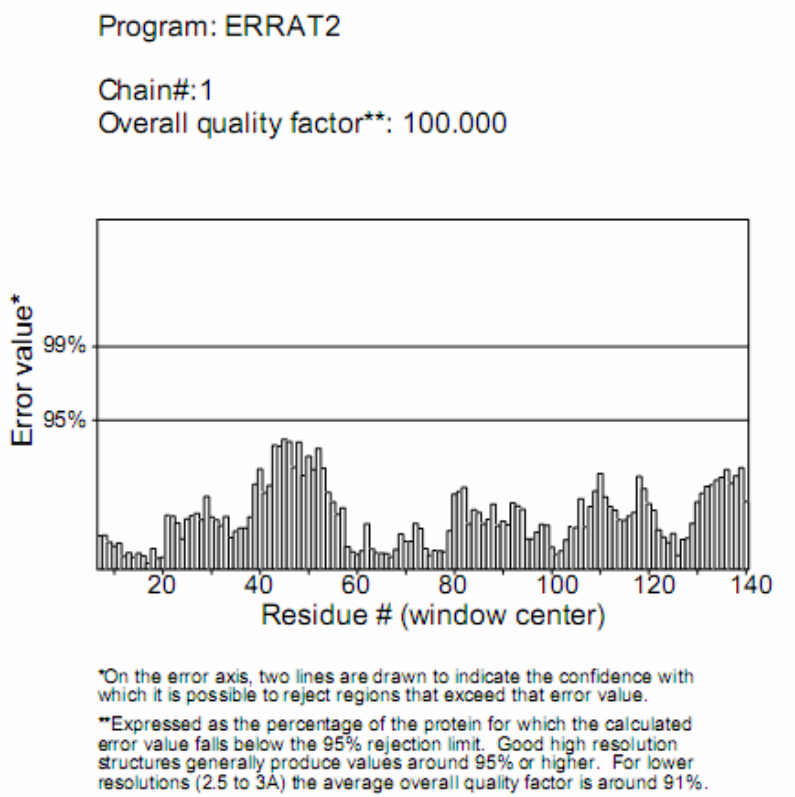


Fig.III ERRAT plot of *Psophocarpus tetragonolobus*

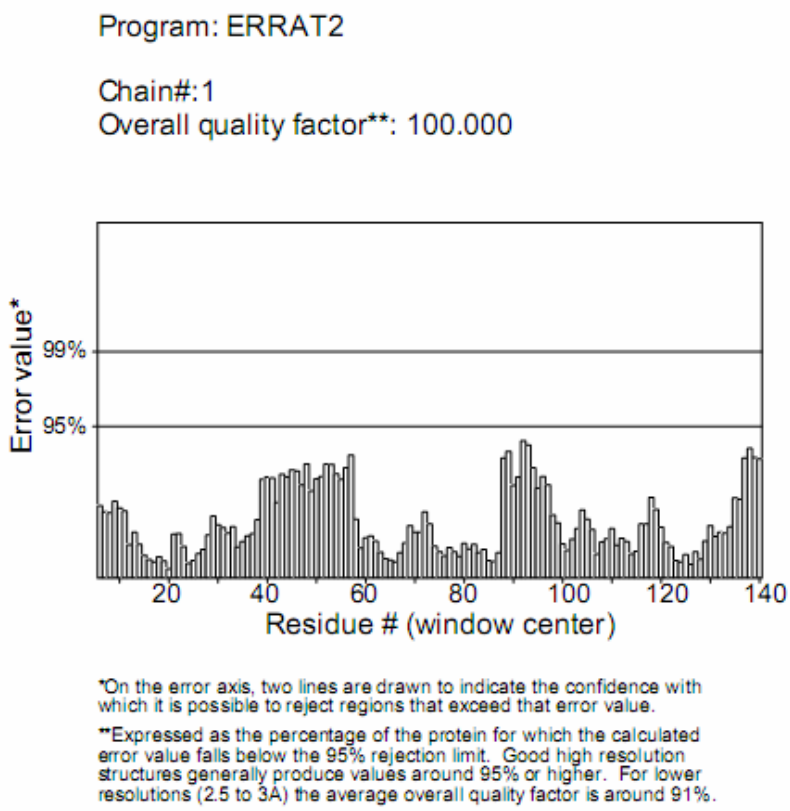


Fig.IV ERRAT plot of *Vigna unguiculata*

Table.3 PROVE PLOT values

Information	**Plants**	
	Psophocarpus tetragonolobus	***Vigna unguiculata***
Z Score mean	0.265	0.215
Z Score Std dev	1.277	1.265
Z Score RMS	1.303	1.282

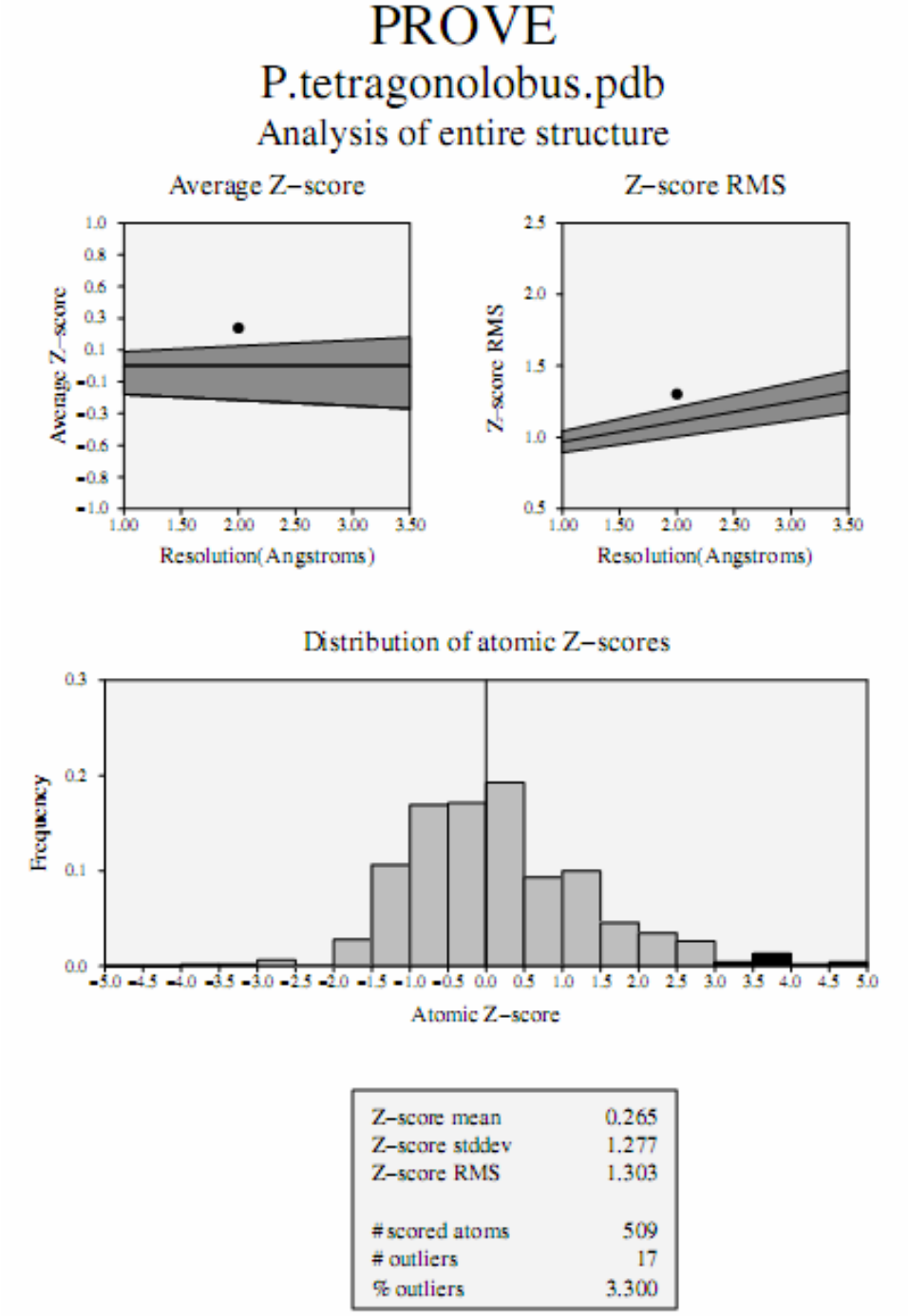


Fig.V PROVE plot of *Psophocarpus tetragonolobus*

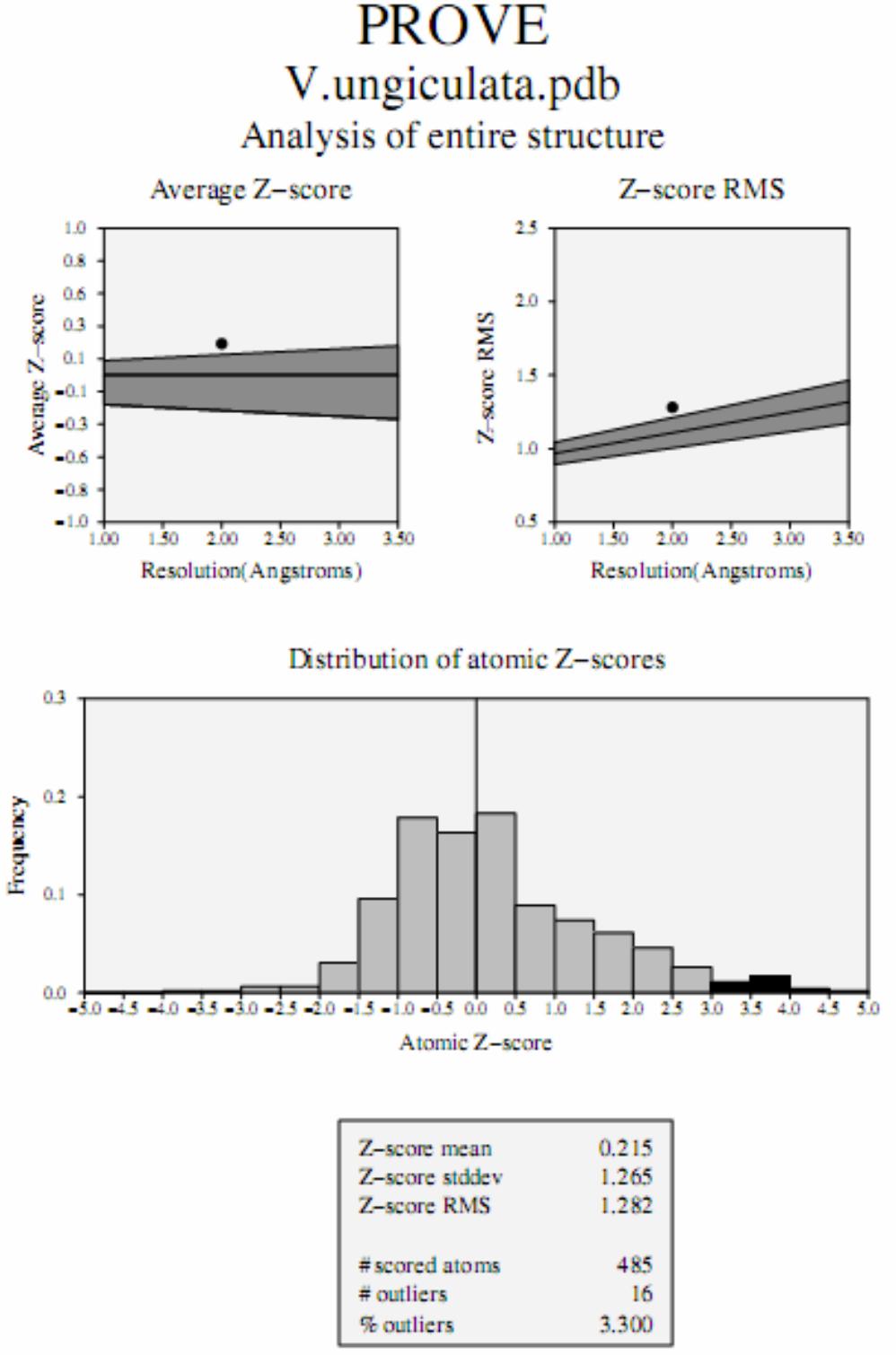


Fig.VI PROVE plot of *Vigna unguiculata*

Fig.VII Predicted 3D structure of Leghemoglobin of *Psophocarpus tetragonolobus*

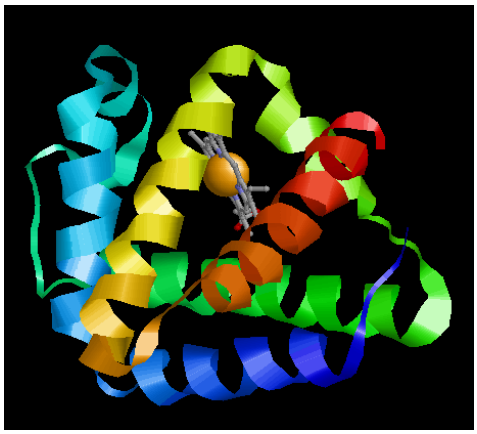


Fig.VIII Predicted 3D structure of Leghemoglobin of *Vigna unguiculata*

References

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling. Bioinformatics, 22,195-201.

Becana, M., Moran J. F., Iturbe-Ormaetxe, I., Gogorcena, Y. and Escuredo, P. R. (1995) Structure and function of leghemoglobins. W Estacion Experimental de Aula Dei (Zaragoza) 21 (3): 203-208.

Benkert, P., Biasini, M., and Schwede, T. (2011). "Toward the estimation of the absolute quality of individual protein structure models." Bioinformatics, 27(3):343-50.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T. and Schwede, T. (2014). SWISS-MODEL: modeling protein tertiary and quaternary structure using evolutionary information. Nucleic acids research, Vol 12 gku340.

Brindha, S., Sailo, S., Chhakchhuak, L., Kalita, P., Gurusubramanian, G., and Kumar, J. N. S. (2011) Protein 3D structure determination using homology modeling and structure analysis. Colloquium www.science.vision.in., 11(3):125–133.

Brisson, N. and Verma, D. P. (1982) Soybean leghemoglobin gene family: normal, pseudo and truncated genes. Proceedings of the National Academy of Sciences U.S.A. 79: 4055–4059.

Colovos, C., and Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 2, 1511-1519

Downie, J. A. (2005) Legume haemoglobins: symbiotic nitrogen fixation needs bloody nodules. Curr. Biol. 15, R196–R198

Graham, P. H. and Vance, C. P. (2003) Legumes: importance and constraints to greater use. Plant Physiology, 131: 872–877.

Guex, N. and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PDB Viewer: An environment for comparative protein modeling. Electrophoresis 18: 2714-2723

Hooft, R.W.W., Vriend, G., Sander, C., and Abola, E. E. (1996). Errors in protein structures. 381, 272-272.

Kawashima, K., Suganuma, N., Tamaoki, M. and Kouchi, H. (2001) Two types of pea leghemoglobin genes showing different O_2-binding affinities and distinct patterns of spatial expression in nodules. Plant Physiology, 125: 641-651.

Laskoswki, R. A., MacArthur, M. W., Moss, D. S. & Thorton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26, 283-291.

Leong, S. A., Ditta, G. S. and Helinski, D. R. (1982). Heme biosynthesis in Rhizobium: identification of a cloned gene coding for d-aminolevulinic acid synthetase from Rhizobium meliloti. Journal of Biological Chemistry, 257: 8724–8730.

Luthy, R., Bowie, J. U., and Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. Nature 356, 83-85.

Pontius, J., Richelle, J., and Wodak, S. J. (1996). Deviations from standard atomic volumes as a quality measure for protein crystal structures. J. Mol. Biol. 264, 121-136.

Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. (1963). "Stereochemistry of polypeptide chain configurations". Journal of Molecular Biology 7: 95–9.

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

Vainshtein, B. K., Harutyunyan, E. H., Kuranova, I. P., Borisov, V. V., Sosfenov, N. I., Pavlovsky, A. G., Grebenko, A. I. and Konareva, N. V. (1975). Structure of leghaemoglobin from lupin root nodules at 5 angstrom resolution. Nature 254, 163–164.

Zhang, Y. (2009). Protein structure prediction: when is it useful? Current opinion in structural biology, 19(2), 145-155.