Plant Omics Journal
1(1):17-25 (2008)
ISSN 1836-3644

Prediction of 3-dimensional structure of EMV1, a group 1 late embryogenesis abundant protein of *Vigna radiata* Wilczek.

[1,2*]Subramanian Rajesh, [1]Muthurajan Raveendran and [1]Ayyanar Manickam

[1]Centre for Plant Molecular Biology, Tamil Nadu Agricultural University, Coimbatore 641003, India*
[2]National Research Centre for Soybean, Indore 452001, Madhya Pradesh, India

[*] Corresponding authors: biotekrajesh@rediffmail.com

Abstract

Late embryogenesis abundant proteins (LEA) are associated with desiccation tolerance among photosynthetic organisms and have been reported in mono and dicot plants as well as in nematodes, yeast, bacteria and cyanobacteria. Although the functional role of LEA proteins remains speculative, there is evidence supporting their participation in acclimation and/or in the adaptive response to stress. EMV1 is a Group 1 LEA protein isolated from *Vigna radiata*, which is speculated to impart desiccation tolerance in plants. The homology model of this protein was generated by using the LOOPP software based on available structural homologues in protein databases. The final model obtained by molecular mechanics and dynamics method was refined and assessed by PROCHECK and shown to be reliable. The generated model could be helpful in understanding functional characteristics of this important class of desiccation tolerant protein.

***Keywords*:** LEA proteins, *Vigna radiata*, desiccation, homology modeling, validation

***Abbreviations*:** SCOP_Structural Classification Of Proteins; PDB_Protein Data Bank; LOOPP _Learning Observing and Outputting Protein Patterns.

Introduction

Late embryogenesis abundant protein genes are highly expressed during late stages of seed development at normal growth condition, but many of the LEA class genes are also frequently expressed in vegetative tissues when plants are exposed to environmental stress (Bray *et al*., 2000). Several groups of LEA protein genes have been demonstrated to confer water-deficit and salt stress tolerance.

On the basis of sequence similarities, LEA proteins have been classified in six groups (Dure, 1993; Bray, 1993). Group 2 LEA proteins or dehydrins are, by far the most frequently described LEA protein family and have been classified in distinct groups (Close, 1997) that differ in the arrangement and number of conserved motifs: the lysine-rich repeat (KIKEKL-PG) or K segment, the stretch of serine or S segment

and the V/T DEYGNP motif or Y segment. Some of these structural motifs are predicted to form amphipathic alpha helices, which may be important for their function in protecting plant cells against dehydration. Evidence of functional links between LEA protein accumulation and improved stress tolerance of transgenic yeast and plants support this hypothesis (Imai, 1996; Xu *et al*., 1996; Sivamani *et al*., 2000).

It was therefore proposed that most LEA and dehydrin proteins exist as largely unfolded structures in their native state, although a few members exist as dimers or tetramers (Ceccardi *et al*., 1994; Kazuoka and Oeda, 1994). Hydrophilicity is a common characteristic of LEA-type and other osmotic stress-responsive proteins. Our earlier work on the occurence of LEA proteins in the embryonic axes of *Vigna radiata* (L.) Wilczek referred as EMV proteins, the first ever report in the Fabaceae family (Manickam and Carlier, 1980). The cDNA encoding this proteins were isolated, characterized (Manickam *et al*., 1996). *In silico* analysis of the 20-mer motif of this EMV1 categorize this protein to Group1 LEA and hypothesize to function as DNA/RNA binding proteins in stabilizing membranes/macromolecules at the time of dehydration process (Rajesh and Manickam, 2006; Gillies *et al*., 2007).

In the present study, earnest effort was made to predict the three-dimensional (3D) structure of the EMV1 protein based on the available template structural homologues from Protein Data Bank and SCOP databases and then generated model was validated with standard protein check parameters. This study can be used useful in further characterization of this important group of functional proteins.

Materials and methods

Data sets

The peptide sequences of *Vigna radiata*, EMV1 (NCBI GenBank accession number U31210; UniProt acc. Nos. Q41684) and other sequences examined in this study were retrieved from the public databases, and http://www.ebi.ac.uk. Structurally homologous subsets of the experimentally determined 3D structures of the EMV proteins were retrieved from PDB and SCOP databases.

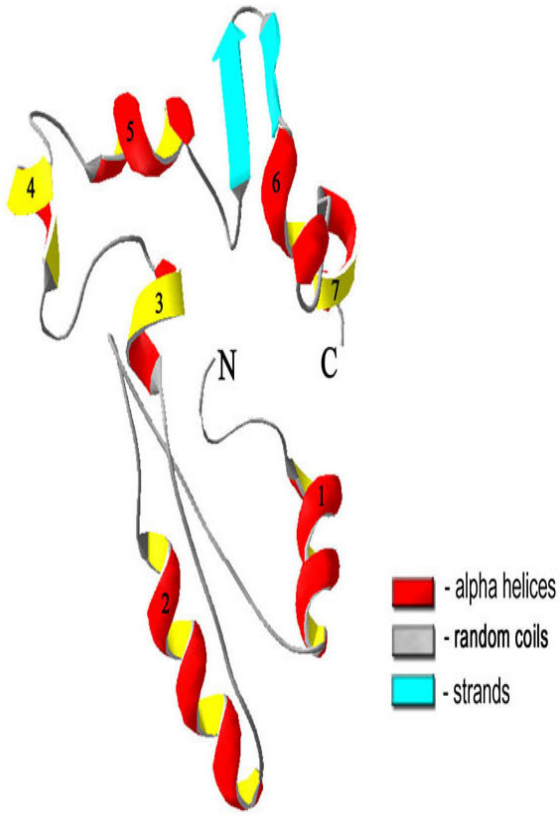

Fig 1. Predicted 3-Dimensional Structure of the EMV1 protein. Ribbon view of EMV1 structure for residues 1-112. Numbers represent the order of helices. The N and C termini of the protein are labeled.

Homology modeling of EMV1 protein

Tertiary structure of the *Vigna radiata* LEA protein, EMV1 was modeled by submitting the deduced amino acid sequences to the Computational Biology Service Unit, Cornell Theory Center, Cornell University. The Atomic coordinates for the protein models were generated by aligning to the structural homologues in the fold recognition program of LOOPP v3.0 server (Teoderescu *et al*., 2004).

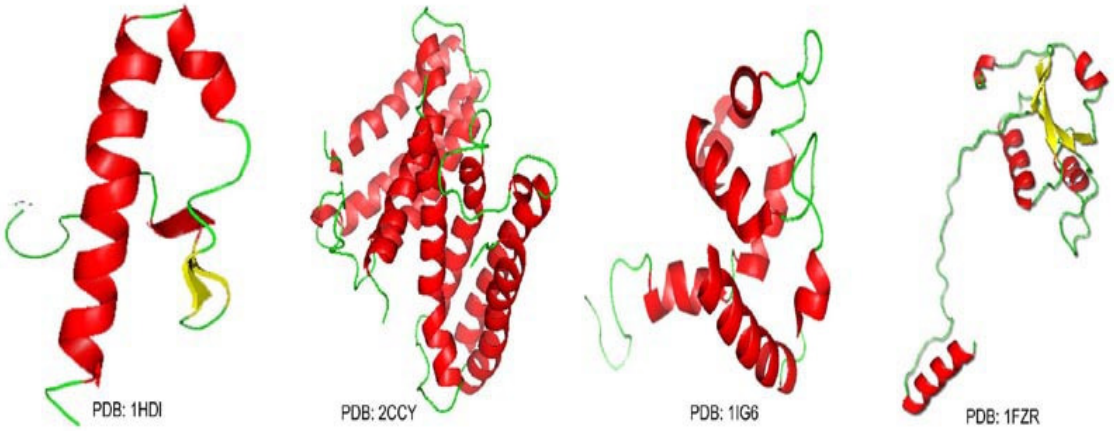

Fig 2. Closest structural homologues to EMV2 protein validation of protein structures of EMV1

PROCHECK validation of EMV1 model

PROCHECK, a versatile protein structure analysis program (Laskowski *et al*., 1993) (available at the Joint Centre for Structural Genomics, Bioinformatics core, University of California, San Diego) was used in validation of protein structure and models by verifying the parameters like Ramachandran plot quality, peptide bond planarity, Bad non-bonded interactions, main chain hydrogen bond energy, C-alpha chirality and over-all G factor and the side chain parameters like standard deviations of chi1 *gauche* minus, *trans* and plus, pooled standard deviations of chi1 with respect to refined structures (Morris *et al*., 1992).

Results

Comparative modeling of EMV1 protein

Tertiary structure of a protein is build by packing of its secondary structure elements to form discrete domains or autonomous folding units. Comparative modeling to build 3D structure of the EMV1 protein was made based on the experimentally solved structural homologous. The amino acid sequences of EMV1 were submitted to LOOPP server Cornel Bioinformatics Structural Unit (CBSU). The atomic coordinates for the proteins were generated based on Hidden Markov Model.

The hypothetical protein models created were stored as PDB output file. The hypothetical proteins were visualized and computed by Swiss PDB Viewer and Rastop. The 3D structure of the proteins were represented by cartoon display and colored based on the secondary structure (Fig 1). The structural homologues used as template in homology modeling are represented (Fig 2).

The hypothetical protein models generated were analyzed online by submitting to Joint Center for Structural Genomics (JCSG), Bioinformatics core, University of California, San Diego. Accuracy of the protein model generated was judged by validity report generated by PROCHECK. Parameter comparisons of these proteins were made with well-refined structures that have similar resolution. An overview of residue-by- residue analysis illustrated by graphics is represented for the EMV1 (Fig 3) protein models. The main chain parameters plotted are Ramachandran plot quality, peptide bond planarity, Bad non-bonded interactions, main chain hydrogen bond energy, C-

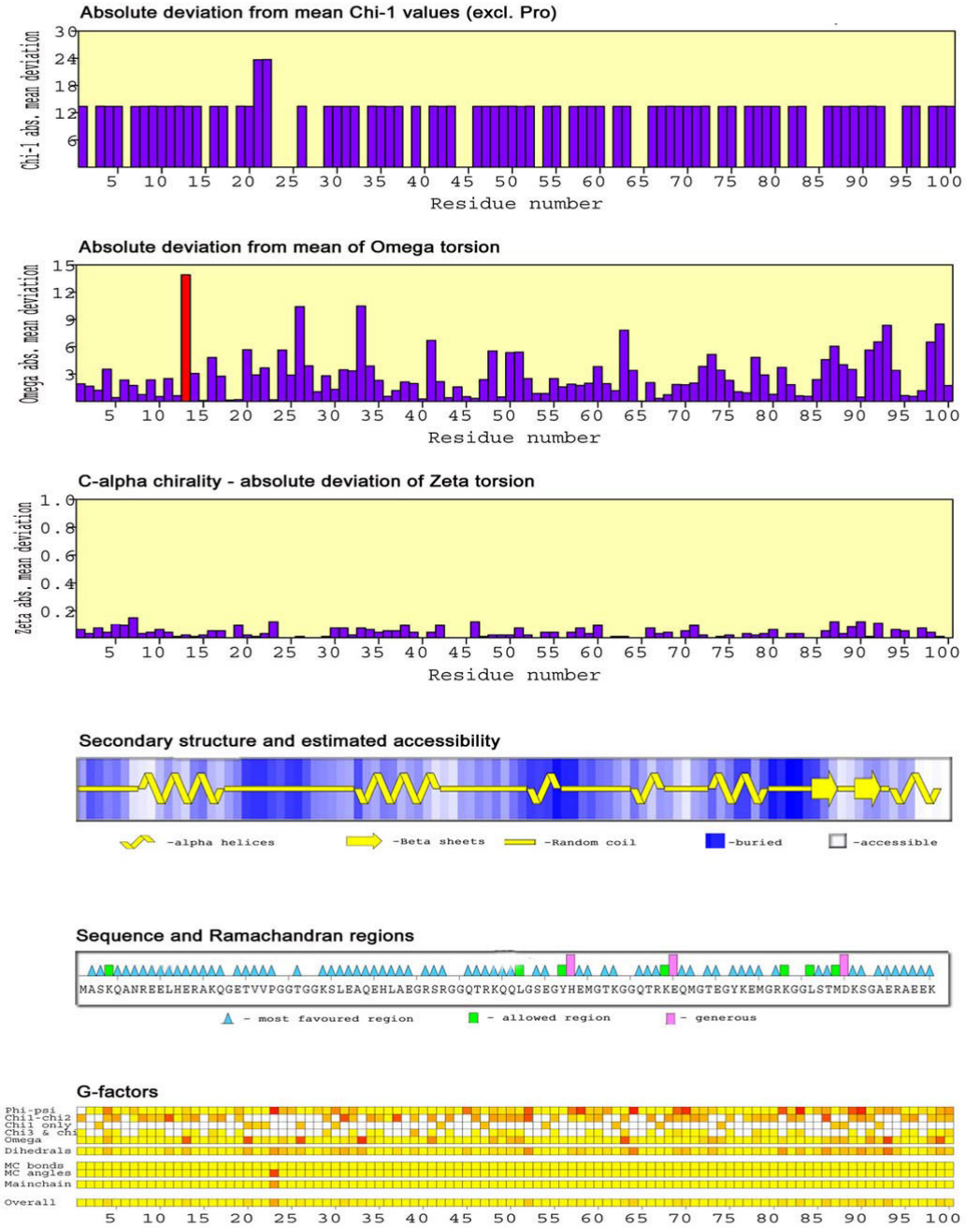

Fig 3. Residue by residue analysis of EMV2 protein Stereochemistry of main chain

alpha chirality and over-all G factor. The stereo-chemistry of main and side chains EMV1 are reported (Table 1).

Ramachandran plot of the protein models

The Phi/Psi angles of the amino acids that determine the secondary structural property of the hypothetical proteins were computed and represented as Ramachandran plot. The residues were classified according to its regions in the quadrangle. The Ramachandran map for EMV1 (Fig 4) and the plot statistics (Table 2) is represented.

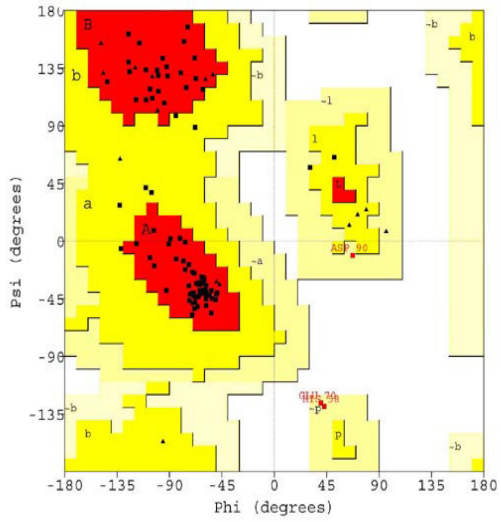

Fig 4. Ramachandran map of EMV1 protein. The plot calculation was done with PROCHECK program.

Non-bonded interactions check

Simple checks of non- bonded interactions are made for the hypothetical protein models and the values of bad contacts per 100 residues are reported as 28 for EMV1.

C-alpha chirality

The atoms C-alpha, N, C and C-beta define chirality. The chirality measure of the zeta torsion angle observed is 0.1 for EMV1.

Over-all G factor

The G-factor shows how normal or unusual the residue type in the main chain of the hypothetical proteins. The over-all G factor computed for the EMV1 is 0.1, which is higher than the ideal value (-0.4).

Stereochemistry of side chain

The protein models were analyzed for the side chain parameters like standard deviations of chi1 *gauche* minus, *trans* and plus, pooled standard deviations of chi1 with respect to refined structure. The standard deviations of Chi1 *gauche* minus, *trans* and plus are in better range and within limits for EMV1 hypothetical protein model.

Computations performed in the EMV1 model

The electrostatic potential of the residues in the hypothetical protein was computed based on Coulomb's method. The cloud of charged residues is represented in blue and red colors and the proteins visualized as density map of EMV1 (Fig 5).Force field energy was computed for EMV1 protein model. Positive values were observed ($E=19.2x10^9$). Model refinement was done by energy minimization. Energy minimization was carried out to reduce clashing amino acids, using GROMOS96 force field algorithm. Decrease in the force field energy was observed for both the protein models after successive energy minimization (E= -2251 for EMV1). The energy minimized models will however needs further refinement in order to reduce the non-bonded interactions for the model to be judged as a good homology model.

Discussion

3-D modeling of EMV1, LEA protein from Vigna radiata

Prediction of tertiary structure of a protein molecule signifies an important step towards understanding the structure –function relationships in the concerned protein family. Recently, the first solution structure of a LEA protein, LEA14 from *Arabidopsis thaliana* has been reported (Singh *et al*., 2005). In the present

Table 1. Stereochemistry of EMV1 protein validated using PROCHECK

	Stereochemical Parameter	Data points	Parameter value	Comparison values		No. of Band widths from mean	Remarks
				Typical value	Band width		
	Main Chain						
A	% Residues in A, B, L	89	87.6	83.8	10.0	0.4	Inside
B	Omega angle	111	3.7	6.0	3.0	-0.8	Inside
C	Bad contacts/100 residues	28	25.0	4.2	10.0	2.1	Worse
D	Zeta angle	92	0.1	3.1	1.6	-1.9	Better
E	H-Bond energy	64	1.2	0.8	0.2	1.8	Worse
F	Overall G-factor	112	0.1	-0.4	0.3	1.5	Better
	Side Chain						
A	Chi-1 gauche minus	0	0.0	18.1	6.5	-2.8	Better
B	Chi-1 trans	2	0.0	19.0	5.3	-3.6	Better
C	Chi-1 gauche plus	82	0.0	17.5	4.9	-3.6	Better
D	Chi-1 pooled	84	14.0	18.2	4.8	-0.9	Inside
E	Chi-2 trans	0	0.0	20.4	5.0	-4.1	Better

study, model of EMV1 LEA protein of *Vigna radiata* was generated from the LOOPP server, based on the structural homologues derived from the SCOP and protein data banks. There exists biological sequence-structure deficit with more than 3 lakhs protein sequences and millions of partial nucleotide sequences, available in the public non-redundant databases (Boguski *et al.*, 1994); and by contrast, the number of unique 3D structures in the protein data bank is still less than 1500 (Attwood and Parry-Smith, 2005). The difference of scale in sequence and structural information is an important factor to be considered when assigning functions to hypothetical proteins. Structure based functional implications of such proteins have always been speculative. Generally, under stress situations the plants may induce formation of coiled coil / folding of the natively unfolded proteins into more rigid structures upon binding to the partner molecules. Since all natively unfolded proteins have defined partner molecules that can be as small as nucleotide or cations or a macromolecule, LEA proteins being natively unfolded is believed to have such binding partners to attain a rigid structures. The LEA proteins from *Vigna radiata* are being classified as Group 1 LEA protein because of its extreme hydrophilicity and adoption of helical conformation as revealed by *ab initio* secondary structure predictions , in combination with the predominant random-coiled arrangement of the residues of *Vigna* LEA protein, is hypothesized to function as water replacement molecule.

Such a property may facilitate hydrogen bonding of this EMV proteins with essentially any macromolecular or membrane surface. However, additional experiments on physico-chemical analyses including examination of hydration properties of these proteins need to be done to determine if EMV proteins can adopt certain structures upon interaction with other macromolecules. Structure homologues identified for these EMV proteins show closest structural homology to proteins with helical bundles of small proteins and DNA/RNA binding proteins. Group 1 LEA proteins are reported to be very hydrophilic, loosely structured with predominantly random-coiled structures.

These proteins are reported to form regular α-helical structure when subjected to altered physiological conditions. Observations from *ab initio* predictions of these EM proteins of *Vigna* indicate 32.14 % of EMV1 proteins attain helical conformation as represented by helical blocks in the 3D hypothetical models (Manickam and Carlier, 1980). These observations are contradictory to the earlier findings from our group that the low molecular weight protein isolated from *Vigna* was believed to be located in the

cytoplasm The structural motifs of these proteins are predicted to form amphipathic α-helices which may be important for their function in protecting cells against dehydration. However, not all LEA proteins are folded and structured. Temperature-induced extended helix/random coil transition was reported for a Group 1 LEA protein from soybean. These proteins are by native, largely unstructured but attained 6-14% helical conformation under temperature stress or at high salt concentrations (Souglaes *et al*., 2002). Similar reports from Goyal *et al*. (2003) for AavLEA1, a Group 3 LEA protein from the nematode, *Aphelenchus avenae* indicate oligomerization of these proteins in immunoblotting and cross-linking experiments, however majority of these proteins was found to be monomeric in analytical ultracentrifugation and gel filtration studies. Also, formation of α-helical structures on drying was reported in partially characterized protein from *Typha latifolia,* probably a Group 3 LEA protein based on Fourier transform-Infra red (FT-IR) spectroscopy studies (Wolkers *et al*., 2001).

The LEA proteins from *Vigna radiata* are being classified as Group 1 LEA protein because of its extreme hydrophilicity and adoption of helical conformation as revealed by *ab initio* secondary structure predictions, in combination with the predominant random-coiled arrangement of the residues of *Vigna* LEA protein, is hypothesized to function as water- replacement molecule.

Such a property may facilitate hydrogen bonding of this EMV proteins with essentially any macro-molecular or membrane surface. However, additional experiments on physico-chemical analyses including examination of hydration properties of these proteins need to be done to determine if EMV proteins can adopt certain structures upon interaction with other macromolecules.

Validation of the model

The hypothetical protein model generated was subjected to structure validation, for testing the accuracy of the model. The quality of the final ensemble of conformers was assessed using PROCHECK, a protein structure validation program. The visual displays of the models were performed with either the Swiss PDB viewer (Guex and Peitsch, 1997) or RasTop (Sayler and Milner-white, 1995).

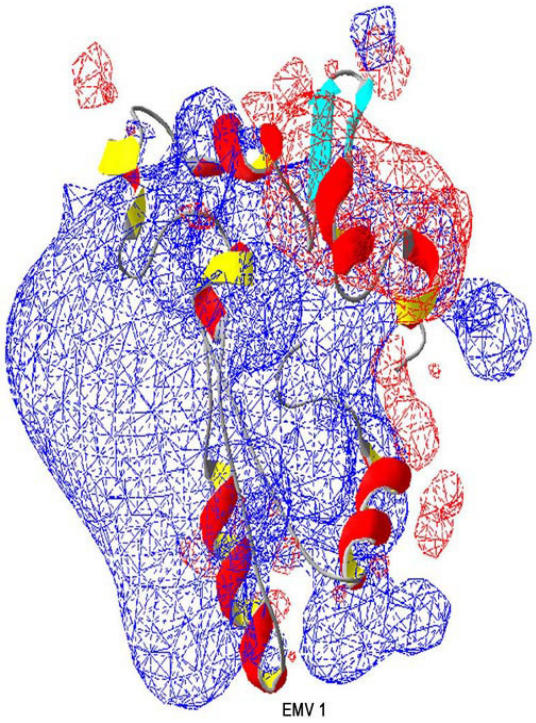

Fig 5. Electron density map of EMV1 protein structure. Positive potential cloud is drawn blue and negative in red.

Stereochemical parameters of the proteins like main- and side chains data of EMV1 was considered for determining the quality of the model. The main chain parameters like Ramachandran plot quality; peptide bond planarity, C-alpha chirality and over- all G factor are found to be within the limits for the model. However, the bad contacts per 100 residues are high. The side chain parameters are in better range and within the limits for EMV1. These parameters are compared to essentially satisfy the generated models with well-refined structures at similar resolution as described by Morris *et al*. (1992).

The validation reports for the protein models are analyzed, and energy minimization of the models was made after checking the force field energy of the models. For a model to be validated based on quality, a good quality protein model should have 90% or more residues in the most favored regions of quadrangle in the Ramachandran plot.

Table 2. Ramachandran plot statistics of EMV1 computed by PROCHECK

Residues in quadrangular regions	Scattered residues	
	Number	Percentage
Most favoured regions [A,B,L]	78	87.6
Additional allowed regions [a,b,l,p]	8	9.0
Generously allowed regions [~a,~b,~l,~p]	3	3.4
Disallowed regions [XX]	0	0.0
Non-glycine and non-proline residues	89	100.0

In the generated model of EMV1, the distribution of residues in the most favored regions is 87.6. This infers that EMV1 as fairly good hypothetical protein model. Also, a tertiary structure of a protein can be worth considering only from its solution structures, obtained from the experiment- tations using either NMR or crystallographic studies. The homology model of plant LEA proteins, generated in this study, could extend investigations at determining the mechanistic function of important class of proteins.

Acknowledgements

S.Rajesh is grateful to the Council of Scientific and Industrial Research, New Delhi for grant of Research Fellowship. We are grateful to the crews of NCBI, EBI, MRC Lab-UK and SIB for making computation-nal biology data/tools publicly available.

References

Attwood TK, Parry-Smith DJ (2005) In. Introduction to Bioinformatics. Ed Wood (ed.). Pearson Education Asia ltd, India. P: 3-10

Boguski MS, Tolstoshev CM, Bassett DE (1994) Gene discovery in DBEST. Science 265: 194-199

Bray EA (1993) Molecular responses to water deficit. Plant Physiol. 103: 1035-1040

Bray EAJ, Bailey-Serres E, Weretilnyk E (2000) Responses to abiotic stresses, In: Biochemistry and Molecular Biology of Plants. Buchanan, W, Gruissem R, Jones (Eds.), American Society of Plant Physiologists, 2000, pp. 1158–1176

Ceccardi TL, Meyer NC, Close TJ (1994) Purification of a maize dehydrin. Protein Express Purif. 5: 266–269

Close TJ (1997) Dehydrins: A commonality in the response of plants to dehydration and low temperature. Physiol. Plant. 100**:** 291-296

Dure L (1993) A repeating 11-mer amino acid motif and plant desiccation. Plant J. 3: 363-369

Gilles GJ, Hines KM, Manfre AJ, Marcotte WR (2007) A predicted N-terminal helical domain of a Group 1 LEA protein is required for protection of enzyme activity from drying. Plant Physiol. Biochem. 45(6-7): 389-399

Goyal K, Tisi L, Basran A, Browne J, Burnell A, Zurdo J, Tunnacliffe A (2003) Transition from natively unfolded to folded state induced by desiccation in an anhydrobiotic nematode protein. J. Biol. Chem. 278: 12977–12984

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

Imai R, Chang L, Ohta A, Bray EA, Takagi M (1996) A lea-class gene of tomato confers salt and freezing tolerance when expressed in *Saccharomyces cerevisiae*. Gene 170: 243–248

Kazuoka T, Oeda K (1994) Purification and characterization of COR85-oligomeric complex from cold-acclimated spinach. Plant Cell Physiol. 35: 601–611

Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst, 26: 283-291

Manickam A, Carlier AR (1980) Isolation and Function of a Low Molecular Weight Protein of Mung Bean Embryonic Axes. Planta 149: 234-240

Manickam A, Van Damme EJM, Kalaiselvi K, Verhaert PW, Peumans WJ (1996) Isolation and cDNA cloning of an Em-like protein from

mungbean (*Vigna radiata*) axes. Physiol. Plant. 97: 524-530 Morris AL, MacArthur MW, Hutchinson EG, Thornton JM (1992) Stereochemical quality of protein structure coordinates. Proteins 12: 345-364

Rajesh S, Manickam A (2006) Prediction of functions for two LEA proteins from mungbean. Bioinformation 1(4): 133-138

Sayle RA, Milner-White EJ (1995) RASMOL: Biomolecular graphics for all. Trends Biochem. Sci., 20: 374–376

Singh S, Cornilescu CC, Tyler RC, Cornilescu G, Tonelli M, Lee MS, Markley JI (2005) Solution structure of a late embryogenesis abundant protein (LEA14) from *Arabidopsis thaliana*, a cellular stress-related protein**.** Protein Science 14: 2601–2609.

Sivamani E, Bahieldin A, Wraith JM, Al-Niemi T, Dyer WE, Ho THD, Qu RD (2000) Improved biomass productivity and water use efficiency under water deficit conditions in transgenic wheat constitutively expressing the barley *HVA1* gene. Plant Sci. 155: 1-9

Soulages JL, Kim K, Walters C, Cushman JC (2002) Temperature induced extended helix/random coil transitions in a group 1 late embryogenesis-abundant protein from soybean. Plant Physiol., 128: 822–832

Teodorescu O, Galor T, Pillardy J, Elber R (2004) Enriching the sequence substitution matrix by structural information. Proteins Str. Fun. Genet. 54: 41-48

Wolkers WF, McCready S, Brandt S, Lindsey GG, Hoekstra FA (2001) Isolation and characterization of a D-7 LEA protein from pollen that stabilizes glasses *in vitro*. Biochim. Biophys. Acta 1544, 196–206 Xu D, Duan X, Wang B, Hong B, Ho TD,

Wu R (1996) Expression of a late embryogenesis abundant protein gene, *HVA1*, from barley confers tolerance to water deficit and salt stress in transgenic rice**.** Plant Physiol. 110: 249-257