Int.J.Curr.Microbiol.App.Sci (2019) *8*(6): 2289-2296

International Journal of Current Microbiology and Applied Sciences
***ISSN: 2319-7706* Volume 8 Number 06 (2019)**
Journal homepage: http://www.ijcmas.com

Original Research Article https://doi.org/10.20546/ijcmas.2019.806.271

Isolation and Characterization of a Late Embryogenesis Abundant (LEA) Protein Coding Gene (*VrLEA2*) from Mung Bean (*Vigna radiata* (L) Wilczek)

U. Nandhini, M. Raveendran and S. Rajesh*

Department of Plant Biotechnology, Centre for Plant Molecular Biology and Biotechnology, Tamil Nadu Agricultural University, Coimbatore – 641003, India

**Corresponding author*

Keywords

Mung bean, *Vr*LEA2 gene, Sequence analysis, Group 1 LEA proteins, structure validation

Article Info

Accepted:
17 May 2019
Available Online:
10 June 2019

A B S T R A C T

Mung bean is a widely consumed tropical food grain legume and is generally affected by different abiotic stresses. Late Embryogenic Abundant (LEA) proteins help the plants to withstand these stresses by its multi-functional role. In this study, a *VrLEA2* gene isolated from mung bean using gene specific primers was sequenced and the *in silico* analysis of the gene sequence revealed the presence of eight ORF regions. ORF6region showed high similarity to Em like protein of *Vigna radiata* (L) Wilczek. The presence of conserved signature motif 'GETVVPGGT' revealed this protein to be hydrophilic and the presence of 20-mer aminoacid repeat 'GGQTRKQQLGSEGYHEMGRK' confirmed the protein to be belonging to Group 1 class of LEA protein families. Secondary structure prediction revealed it as coiled and disordered protein which specifies it has significant function of protecting biomolecules. The 3D structure of the protein was predicted using ROBETTA tool and was validated using PROCHECK, which found as better model based on several parameters. Further this gene can be validated in any expression system to understand the functional role played by this protein.

Introduction

Plants growth and productivity are influenced by biotic and abiotic environmental stress factors. Abiotic stresses like salt, heat, drought and osmotic stress affects plants in several ways. Generally, plants face all kinds of stresses in combination, for example if drought condition occurs then it will increase the chance of salt or high temperature stress. That's why studies on physiological, metabolic and molecular level studies are not restricted to single stress conditions (Khraiwesh *et al.*, 2015). Against these abiotic stresses, plants responses are complex and dynamic; it may leads to reversible or irreversible changes (Cramer, 2011).Role played by Late embryogenic abundant (LEA) proteins are one of the mechanism evolved by plant to overcome these stresses. LEA proteins are a class of low molecular weight protein accumulated in high amount during stress conditions. There are several classification available, now they are classified into eight families based on Pfam analysis of conserved protein motif (Finn *et*

al., 2013). From more than 200 organisms, information on 1139 LEA proteins were stored in LEAPdb public database (Hunault and Jaspard, 2010).

Early studies reported a presence of LEA protein in the developing mungbean seeds named as EM proteins and was the first to be reported in the Fabaceae family (Manickam and Carlier, 1980). These were classified under group1 LEA protein family based on the presence of 20-mer repeat motif sequence and also with its rich glycine residues. Generally the function of any protein can be detected based on its structural arrangements. But the isolation and experimental prediction of protein structure is tedious and time consuming. Hence *in silico* analysis of nucleotide and protein sequence reveals details about the structure and properties of proteins. In the present study, group specific characteristics of *VrLEA2* gene isolated from *Vigna radiata* (L) Wilczek were studied. The structure model for *Vr*LEA2 was predicted and validated was using *in silico* analysis.

Materials and Methods

Gene isolation and Sequencing

The *Vr*LEA2 gene was amplified from the genomic DNA of mung bean (*Vigna radiata*) by PCR using gene specific primers designed from the draft genome sequence of mung bean. The amplified *Vr*LEA2 gene was subjected to DNA sequencing (Agri-Genomics Private Limited, Kerala) and extensively analyzed using bioinformatics tools.

Sequence analysis

The sequence was assembled using CAP3 contigs assembly tool (http://www.insilico.uni-duesseldorf.de/Cap3.html) and searched for sequence similarity against the NCBI non-redundant database using BLASTn tool (Altschul *et al.,* 1997). The domains conserved in this sequence were predicted using conserved domain database (CDD) search tool. The number of open reading frame (ORF) and translation of the sequence was revealed by ORF finder at NCBI (https://www.ncbi.nlm.nih.gov/orffinder/).

These ORFs are matched against non-redundant and protein database (PDB) using BLASTp (Altschul *et al.,* 1997). For the valid ORF sequence, protein structure was predicted using SWISSPROT automated homology modeling tool.

Sequence characterization

The physico-chemical properties like molecular weight (MW), grand average of hydropathicity (GRAVY) and isoelectric point (pI) were determined using PROTPARAM tool (https://web.expasy.org/protparam/). The hydropathy analysis was also confirmed through Kyte and Doolittle scale (https://web.expasy.org/protscale/) (Kyteand Doolittle, 1982). Transmembrane domain was predicted using PROTTER server (http://wlab.ethz.ch/protter/#). The subcellular localization and the secondary structure were predicted using WOLF_PSORT (https://www. genscript.com /wolf-psort.html) and PRISPRED tool (http://bioinf.cs.ucl.ac.uk/psipred/).

Structure prediction of *Vr*LEA2 protein

The three dimensional (3D) structure of *VrLEA2* was predicted using ROBETTA tool (http://robetta.bakerlab.org/submit.jsp). The validation of this predicted structure was done by PROCHECK (https://servicesn.mbi.ucla.edu/PROCHECK/) (Laskowski, 1993) based on parameters like Ramachandran plot analysis, peptide bond planarity, overall G factor value, main chain hydrogen bond energy and Bad non-bonded interaction (Morris *et al.*, 1992).

Results and Discussion

Isolation and sequencing of *Vr*LEA2 gene

The *Vr*LEA2 gene got amplified at ~900bp using gene specific primer (Fig. 1) and the amplification conditions for this gene were optimized using gradient PCR. The amplicon was sequenced and the contigs were assembled. The sequence was deposited in the NCBI-GenBank.

Classification of VrLEA2 protein

The *VrLEA2* sequence was subjected to similarity search using BLAST tool and it showed 97% and 71%, sequence similarity with *Vigna radiata* var. *radiata* SLE2 protein and *Vigna radiata* (L) Wilczek Em like protein, respectively. Generally Group 1 Class of LEA proteins are hydrophilic and are α-helical in structure. The CDD search predicted the presence of hydrophilic signature domain 'GETVVPGGT' and also the presence of 20-mer 'GGQTRKQQ LGSEGYHEMGRK' motif unique for Group1 Class of LEA proteins (Fig. 2). The presence of glycine rich hydrophilic repeat at N-terminus is related to small hydrophilic plant seed protein of Pfam00477 super family. The hydrophilic nature of this protein might help in biological activity of this LEA protein during water stress. Even though the role of hydrophilins remains unclear, some evidence supports their involvement in adaptive stress tolerance like heterologous expression of LEA proteins in some plants and yeast confers tolerance to water deficient conditions (Swire-Clark and Marcotte, 1999; Zhang *et al.,* 2000) and chilling stress conditions (Rinne *et al.,* 1999). Further the deletion of RMF hydrophilins reduced osmotic tolerance in *E.coli* (Garay-Arroyo *et al.,* 2000). The ORF finder revealed the presence of eight ORF regions in these sequence (Fig. 3). The ORF 6 was found to have high similarity with Em-like protein of mung bean and both conserved motif region were present in this ORF.

Function assignments based on sequence characterization

The PROTOPARAM tool predicted the physico-chemical properties of the particular ORF 6 sequence like molecular weight, isoelectric point and GRAVY value to be 6.9 KDa, 8.01 and -1.6 respectively. The sequence was filled with hydrophilic residues predicted by hydropathy scale (Fig. 4). This peptide was mostly found to be located in nucleus without any transmembrane domains (Fig. 5). Though most LEA proteins are found in cytoplasm (Soulages *et al*., 2002), there are some evidence that supports its distribution in other organelles (Duan and Cai, 2012; George *et al.,* 2009) as well. Predicted polypeptide showed majority ofthe VrLEA2 protein constitutes Glycine about 16.1%and which is unique to Group 1 class of LEA proteins. The flexible conformation like intrinsically disordered state helps in binding property of LEA proteins to biomolecules. The helix/coil transition induced during temperature changes was stated in soybean group 1 class proteins (Soulages *et al.,* 2002). Secondary structure analysis revealed it to form mostly helical/coiled structure and highly disordered protein (Fig. 6) which suggested that it may have significant biological role during stress conditions.

Validation of predicted structure

Prediction of 3D structure is important in understanding the function of any protein molecule. Here, the 3D structure was displayed in cartoon model and colored based on the secondary structure predicted using ROBETTA tool (Fig. 7). The overall G factor value predicted was 0.3 for this model and the stereo chemical parameters of side chain like chi standard deviation of Gauche minus, trans and Gauche plus were better (Table 1). For

the structural model to be good, most of the residues should be distributed in allowed region. Here96% was present in allowed region for the *VrLEA2* predicted model and this suggest it as good hypothetical protein model. The distribution of residues was displayed through Ramachandran graph (Fig. 8) and the plot statistics (Table 2). From the analysis of all parameters predicted using PRISPRED, this model was found to be good.

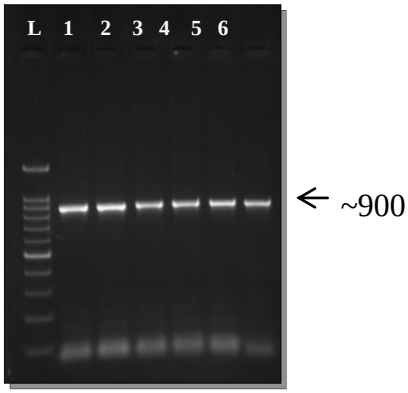
Table.1 Sterochemical data of emv2 protein using PROCHECK

Serial	Stereo chemical parameter	Data points	Parameter value	Comparison values		No. of Band widths from mean	Remarks
				Typical Band	band widths		
Main Chain							
A	%-tage residues in A, B, L	50	96.0	88.2	10.0	0.8	Inside
B	Omega angle St dev	61	16.5	6.0	3.0	3.5	Worse
C	Bad contacts / 100 residues	0	0.0	1.0	10.0	-0.1	Inside
D	Zeta angle St dev	52	1.1	3.1	1.6	-1.3	Better
E	H-bond energy St dev	45	0.7	0.7	0.2	-0.1	Inside
F	Overall G-factor	62	0.3	-0.2	0.3	1.8	Better
Side Chain							
A	Chi-1 gauche minus	6	4.2	13.6	6.5	-1.4	Better
B	Chi-1 trans	16	8.5	15.3	5.3	-1.3	Better
C	Chi-1 gauche plus	28	3.4	13.8	4.9	-2.1	Better
D	Chi-1 pooled	50	5.4	14.3	4.8	-1.8	Better
E	Chi-1 trans	25	7.1	17.7	5.0	-2.1	Better

Table.2 Ramachandran plot statistics of EMV-2 protein predicted using PRISPRED

Residues in quadrangles	Scattered residues	
	Number	Percentage
Most favoured regions [A,B,L]	48	96.0
Additional allowed regions [a,b,l,p]	2	4.0
Generously allowed regions [~a,~b,~l,~p]	0	0.0
Disallowed regions	0	0.0
Number of non-glycine and non-proline residues	50	100.0

Fig.1 PCR amplification of *VrLEA2*gene



Lane L – 100bp ladder; 1to 6 – PCR amplified product of *VrLEA2* gene

Fig.2 Conserved domain prediction using Conserved Domain Database (CDD) search tool; Green box denotes hydrophilic residues; Black box denotes 20-mer repeats

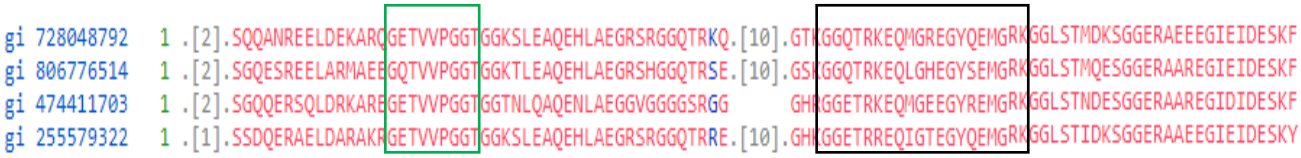


Fig.3 ORFs prediction of EMV-2 gene by ORF finder

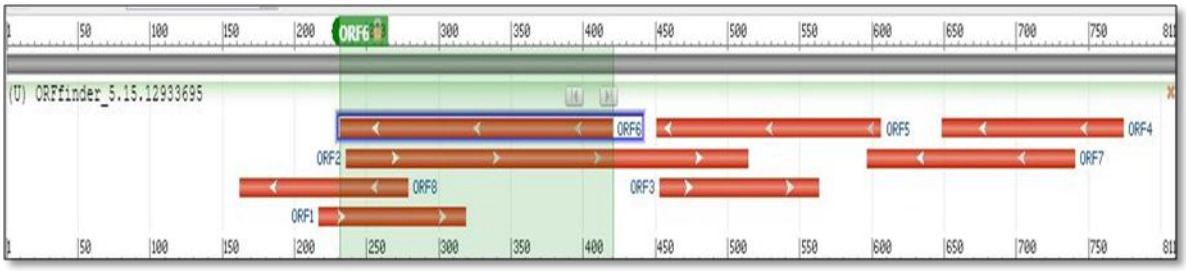


Fig.4 Kyte and Doolittle hydropathy analysis of VrLEA2. Region above zero denotes hydrophobicity and region below zero denotes hydrophilicity

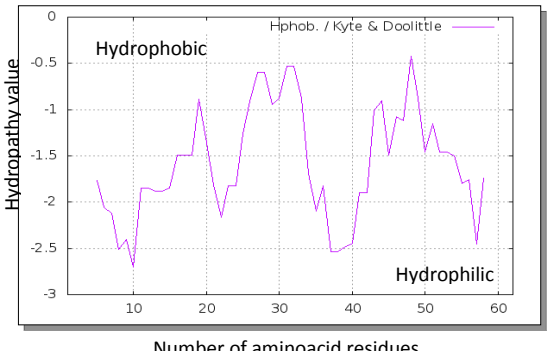


Fig.5 Protein localization of EMV-2 using PROTTER server

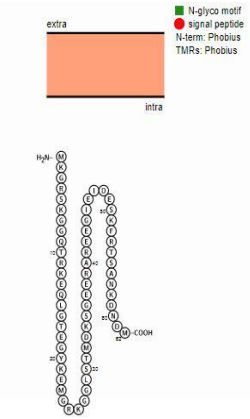


Fig.6 Secondary structure prediction of EMV-2 gene using PRISPRED;
A) Helix prediction; b) Disorder prediction

Fig.7 3D structure predicted using ROBETTA tool.
Ribbon view of VrLEA2 protein for amino acid residues 1-62

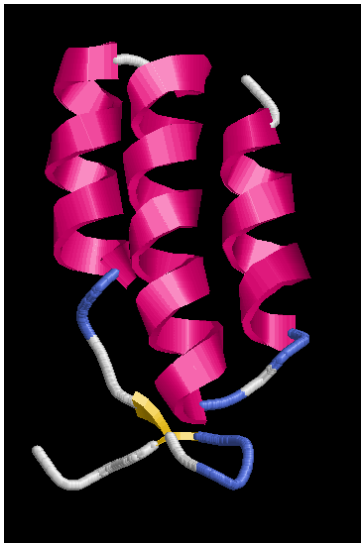
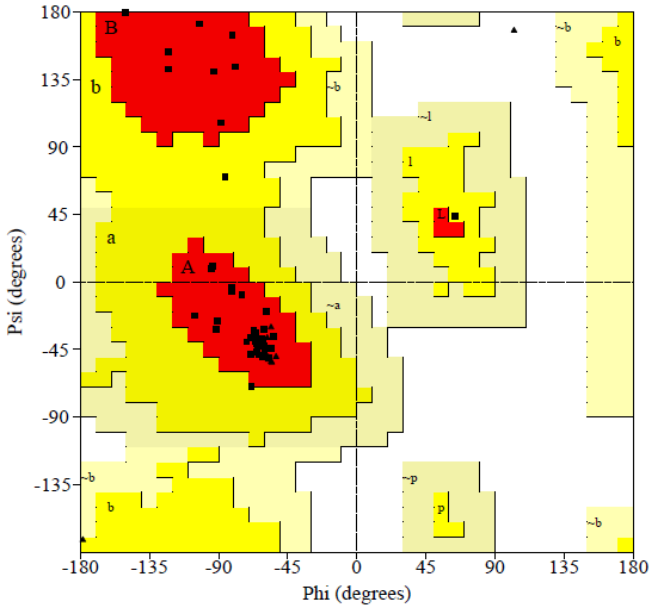


Fig.8 Ramachandran plot analysis of EMV-2 protein predicted using PRISPRED



In conclusion, LEA proteins are class of stress associated proteins with multi-functional role in both plants as well as in other organism. The bioinformatics analysis revealed *VrLEA2* gene belong to Group 1 class of LEA protein with predicted signature motif and also all unique feature like high Glycine content and hydrophilic nature added value to this prediction. This VrLEA2 protein is intrinsically disordered in nature and is presumed to play potential biological role during stress conditions like other LEA proteins. The analysis of predicted 3D structure will helps in understanding its functional role against various stress conditions..

References

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, *25*(17), 3389-3402.

Duan, J., and Cai, W. (2012). OsLEA3-2, an abiotic stress induced gene of rice plays a key role in salt and drought tolerance. *PLoS One*, *7*(9), e45117.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J., and Punta, M. (2013). Pfam: the protein families database. *Nucleic acids research*, *42*(D1), D222-D230.

Garay-Arroyo, A., Colmenero-Flores, J. M., Garciarrubio, A., and Covarrubias, A. A. (2000). Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. *Journal of Biological Chemistry*, *275*(8), 5668-5674.

George, S., Usha, B., and Parida, A. (2009). Isolation and characterization of an atypical LEA protein coding cDNA and its promoter from drought-tolerant plant *Prosopis juliflora*. *Applied biochemistry and biotechnology*, *157*(2), 244-253.

Hunault, G., and Jaspard, E. (2010). LEAPdb: a database for the late embryogenesis

abundant proteins. *BMC Genomics*, *11*(1), 221.

Khraiwesh, B., Qudeimat, E., Thimma, M., Chaiboonchoe, A., Jijakli, K., Alzahmi, A., Salehi-Ashtiani, K. (2015). Genome-wide expression analysis offers new insights into the origin and evolution of Physcomitrella patens stress response. *Scientific reports*, *5*, 17434.

Laskowski, R.A. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26, 283-291.

Manickam, A., and Carlier, A. R. (1980). Isolation and function of a low molecular weight protein of mung bean embryonic axes. *Planta, 149*(3), 234-240.

Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992). Stereochemical quality of protein structure coordinates. *Proteins: Structure, Function, and Bioinformatics*, *12*(4), 345-364.

Rinne, P. L., Kaikuranta, P. L., van der Plas, L. H., and van der Schoot, C. (1999). Dehydrins in cold-acclimated apices of birch (*Betula pubescens* Ehrh.): production, localization and potential role in rescuing enzyme function during dehydration. *Planta*, *209*(4), 377-388.

Soulages, J. L., Kim, K., Walters, C., and Cushman, J. C. (2002). Temperature-induced extended helix/random coil transitions in a group 1 late embryogenesis-abundant protein from soybean. *Plant Physiology*, *128*(3), 822-832.

Swire-Clark, G. A., and Marcotte, W. R. (1999). The wheat LEA protein Em functions as an osmoprotective molecule in *Saccharomyces cerevisiae*. *Plant molecular biology*, *39*(1), 117-128.

Zhang, L., Ohta, A., Takagi, M., and Imai, R. (2000). Expression of plant group 2 and group 3 lea genes in Saccharomyces cerevisiae revealed functional divergence among LEA proteins. *The Journal of Biochemistry*, *127*(4), 611-616.

How to cite this article:

Nandhini, U., M. Raveendran and Rajesh, S.S. 2019. Isolation and Characterization of a Late Embryogenesis Abundant (LEA) Protein Coding Gene (VrLEA2) from Mung Bean (*Vigna radiata* (L) Wilczek). *Int.J.Curr.Microbiol.App.Sci.* 8(06): 2289-2296.
doi: https://doi.org/10.20546/ijcmas.2019.806.271