PLANT BREEDING AND SEED SCIENCE
Volume 70 2014

DOI: 10.1515/plass-2015-0010

Hassan Rasouli, Danial Kahrizi, Ezatollah Farshadfar,
Khirollach Yari, Mohammad Taher Moradi

Agronomy and Plant Breeding Department (APBD), College of Agriculture
and Natural Resources, Razi University, Kermanshah, Iran. P.B.671568543

COMPUTATIONAL ANALYSIS OF GROUP 2 LATE EMBRYOGENESIS PROTEIN (LEA) IN DIFFERENT CULTIVAR OF BREAD WHEAT (*TRITICUM AESTIVUM*)

ABSTRACT

Late Embryogenesis abundant protein has a crucial role as the cold-acclimation process in the wheat. These proteins encoded by *TaWdhn13* gene. This gene is transcriptionally activated and produces the accumulated proteins and metabolites and protection cell structure from freezing damage. The objectives of this study were to isolate the genomic DNA (g-DNA) sequence of *TaWdhn13*, to analyze structure, conserved domains of the gene, and to found a basis for association analysis of the functional sites associated with computational analysis. We here report on the functional assignment to *TaWdhn13* gene by computational analysis. The Three-Dimensional (3D) model of LEA protein drawing by using the phyre 2 server. For identify the conserved domain and motif of these gene sequence we used the Conserved Domain Database and DNA Motif Searching Database, however, the conserved domains and motif has been recognized. The results showed *TaWdhn13* conserved domain incudes: Dehydrin superfamily. Also, motifs structure for this gene includes: 2FE2S_FER_1 Motif, INTEGRIN_BETA Motif, VWFC_1 Motif, EGF_1 Motif and DEFENSIN Motif. Our results reveal that group 2 LEA proteins are most likely to function within the cell nucleus. The analysis of protein property showed that the protein had no trans-membrane domains. The isoelectric point of the protein was 3.41, which was charged with 5.34 negative electrons when pH value of the buffer was 7.0.

Key words: Computational analysis, Group 2 LEA protein, Functional assignment, *Triticum aestivum*

INTRODUCTION

During the late maturation stage of seed development, water content decreases greatly (Kobayashi *et al*. 2007 ; Alsheikh *et al*. 2003; Babu *et al*. 2004; Baker *et al*. (1988) ; Cheng *et al*. 2002). One of the most striking characteristics

Communicated by Grzegorz Żurek

of mature orthodox seeds is their ability to withstand severe desiccation (Alsheikh *et al*. 2003). Mechanisms of plant drought/desiccation tolerance have been studied by numerous groups, and a broad range of molecules have been identified to play some roles (Eom *et al*. 1996.; Hu, 2008; Wu *et al*. 2013; Battaglia *et al*. 2008; Bray, 2002). Examples are proline, oligosaccharide, and late embryogenesis abundant (LEA) proteins, and so on. LEA proteins were first described from mature cotton seeds decades ago. Since then, many LEA proteins were identified from vascular (Battaglia *et al*. 2008) and nonvascular plants, fungi, algae, and microbes, as well as anhydrobiotic animals such as protozoa, nematodes, insects, and crustaceans, and so on (Babu *et al*. 2004, Koag *et al*. 2003). The extensive distribution of LEA genes among diverse taxa implies that these genes might be primitive yet important and therefore maintained by these species. As a result of evolution, they may have a certain universal function—osmoprotection (Jeffrey *et al*. 1993). Hydrophilic LEA proteins are members of natively unfolded proteins in solution. After the removal of bulk cytoplasmic water, the structures of LEA proteins undergo desiccation induced folding (Kalemba and Pukacka, 2007). These biophysical features suggest that LEA proteins may carry out a bipartite function under different water states. During drought (Kobayashi *et al*. 2007 ; Alsheikh *et al*. 2003; Babu *et al*. 2004), LEA proteins may establish a water shell and decrease ion strength. After desiccation, they may enhance the bioglass strength and act as a water replacement to stabilize cellular components (Liu *et al*. 1998 ; Rajesh *et al*. 2008; Rasouli *et al*. 2013; Simon *et al*. 2009; Zhu *et al*. 2007).

MATERIAL AND METHODS

Plants Materials

A total of 3 cultivar of bread wheat were used to represent some of the diversity currently available. These cultivars were subdivided into 3 clusters, which consisted of Marvdasht, Azar 2 and Alborz cross.

DNA purification

Plants were grown for DNA isolation in the greenhouse and leaves were harvested 3 weeks after germination. DNA was extracted from different cultivar of bread wheat includes Marvdasht, Azar 2 and Alborz cross by using the CTAB protocol. Quantification of DNA was done using a Beckman Coulter DU-800 spectrophotometer to measure absorbance at 260 nm and 280 nm to determine the concentration.

Custom Primers

Custom primers with similar melting points were designed to amplify and sequence the *TaWdhn13* gene using Primer 3 program, http://biotools.unmassmed.edu/bioapps/primer3_www.cgi. Delta G values for homodimer, heterodimer, and hairpin loops were determined using OligoAnalyzer 4.0, http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/, by Integrated DNA Technologies.Custom primers (Table 1) were designed to amplify segment of the *TaWdhn13* CDs 385 bp.

Table 1

Sequence of PCR primers

Forward	5'-TAGGGACAAGTTGAGGGCAAG-3'
Reverse	5'-CTGGGCTTAGTGCTGTCCAG-3'

PCR amplification with Pwo DNA polymerases

PCR reagents and protocol were used at listed concentrations in 25 µL reactions. A magnesium curve from 2 mM to 4 mM was performed to identify optimal reaction conditions for High Fidelity *Pwo* DNA Polymerase. 1 µL of 575 ng/µL genomic template DNA was used in each reaction. A negative control without template DNA was performed in order to identify any potential errors. PCR was performed under varying conditions until successful. PCR amplification of the *TaWdhn13* was accomplished using the following program:

Initiation Denaturing Temperature at 94^{o}C for 4 minutes.

- For 35 cycle:
- Denaturing Temperature at 94^{o}C for 30 seconds.
- Annealing Temperature at 60^{o}C for 60 seconds.
- Extension Temperature at 72^{o}C for 90 seconds.
- Final Extension Temperature at 72^{o}C for 10 minutes.
- Sample hold at 4^{o}C.

Gel Electrophoresis

Gel electrophoresis was used to visually confirm the presence of PCR products. 1% agarose gel (Sigma-Aldrich) was cast using 0.5X TBE buffer and Ethidium Bromide (EB) (Sigma-Aldrich, CAS Number 1239-45-8) 1.5mL/45mL gel was added to stain the gel. 6X loading dye were added to each sample at appropriate dilutions. Banding lengths were determined by comparison to the 1 Kb plus Ladder by Invitrogen® (Carlsbad, California, USA). Gels were typically run at 110 V, until banding patterns were evident.

Sequence Analysis of TaWdhn13 gene

Isolated DNA from different cultivar of bread wheat includes Marvdasht, Azar 2 and Alborz cross was sequenced at BIONEER (in Korean) CO, using M13 Forward and Reverse primers, as well as designed custom sequencing primers. The PCR product with the fewest number of mutations furthest away from the start codon was selected for further experimentation. After the direction of the sequence to analysis the *TaWdhn13* gene we used the NCBI database.

Data sets

The peptide sequence from *TaWdhn13* gene generated by ExPasy database, http//:web.expasy.org/translate. Structurally homologous subsets of the experimentally determined 3D structures of the LEA proteins were retrieved from PDB and Phyre 2 and HMMER databases. To identify conserved domain, we used the Conserved domain database (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). To analyze the nucleotide sequence we searched for the motifs and the motif search software (http://www.genome.jp/tools/motif) was used to identify the motifs in nucleotide sequence. Sequence similarities were examined with the GenBank/EMBL database using the BLAST program (http://blast.ncbi.nlm.nih.gov/Blast.cgi). For nucleic acid and amino acid sequence alignment ClustalW1.8 software (http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html) was used. Analysis of hydropathy of deduced group 2 LEA proteins was done as described by (Kyte and Doolittle, 1982). For protein localization analysis PSORT (http://psort.ims.u-tokyo.ac.jp/) program was used. Phylogenetic relationship was analyzed by multiple alignments of plant LEA 2 g-DNA using ClustalW program (http://www.ebi.ac.uk/clustalw/).

RESULTS

Full length sequence of TaWdhn13 from genomic DNA

Using PCR-based walking method, the core region and the upstream and downstream flanking regions of *TaWdhn13* gene were amplified. The full length of *TaWdhn13* gene was 2408 bp. The length of the sequence from the start codon to the stop codon was 375 bp. The genomic sequence of this gene has been submitted to GenBank (accession number: in submitting). The single nucleotide polymorphisms (SNPs) and the indels of the nonallelic specific sequence were aligned between the *TaWdhn13* gene and the sequences of other Dehydrin gene, in NCBI database. The consensus sequences proved to be the genomic sequence of *TaWdhn13* gene. multiple sequence alignment showed (Fig. 1A) that the deduced genomic sequence of Marvdasht cultivar and Azar 2

cultivar sequence has 98% similarity between each other and shares about 90% similarity with *Wdhn13* gene sequence in NCBI data base (Fig. 1B).

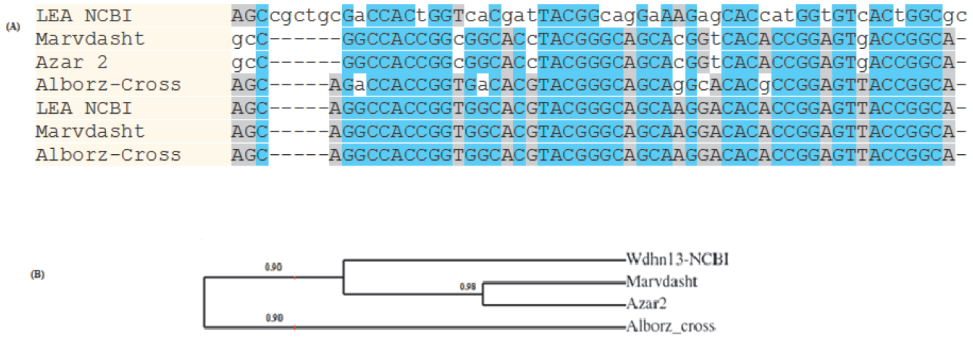

Fig 1. *In silico* analysis of *TaWdhn13* gene in different cultivar of wheat. (A) Comparison of deduced genomic sequence of *TaWdhn13* gene in different cultivar of wheat. (B) Phylogenetic analysis of *TaWdhn13* gene from different cultivar of bread wheat.

Gene structure

Sequence analysis of *TaWdhn13* gene revealed that the full length of the open reading frame, from the start codon ATG to the stop codon TAA, was 405 bp, which contained 1 exon (Fig. 2)

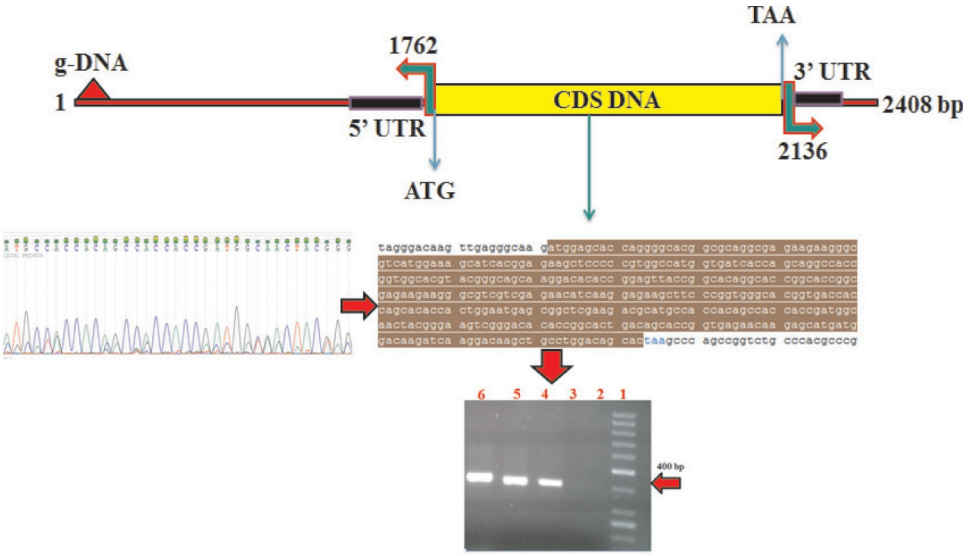

Fig 2. Structure of *TaWdhn13* gene in bread wheat.

Protein property and conserved domains

LAE protein encoded by *TaWdhn13* consisted of 124 amino acids with the molecular weight of 8.2 kD. The analysis of protein property showed that the protein had no trans-membrane domains. The isoelectric point of the protein

was 3.41, which was charged with 5.34 negative electrons when pH value of the buffer was 7.0. Conserved domain search analysis showed that pfam domain pfam00257 (E-value, 3e-04) was found from 1 to 73 amino acid of the predicated LEA protein in different cultivar of bread wheat in the present study. Predicted proteins of group 2 LEA protein in different cultivar of wheat showed preponderance of Gly, Thr, Ala and Glu that constitute 18, 17, 10 and 7.5%, respectively, but lack Trp and Cys. Hydropathy analysis showed that the predicted group 2 LEA protein in the present study is a hydrophilic protein (Fig. 3A).

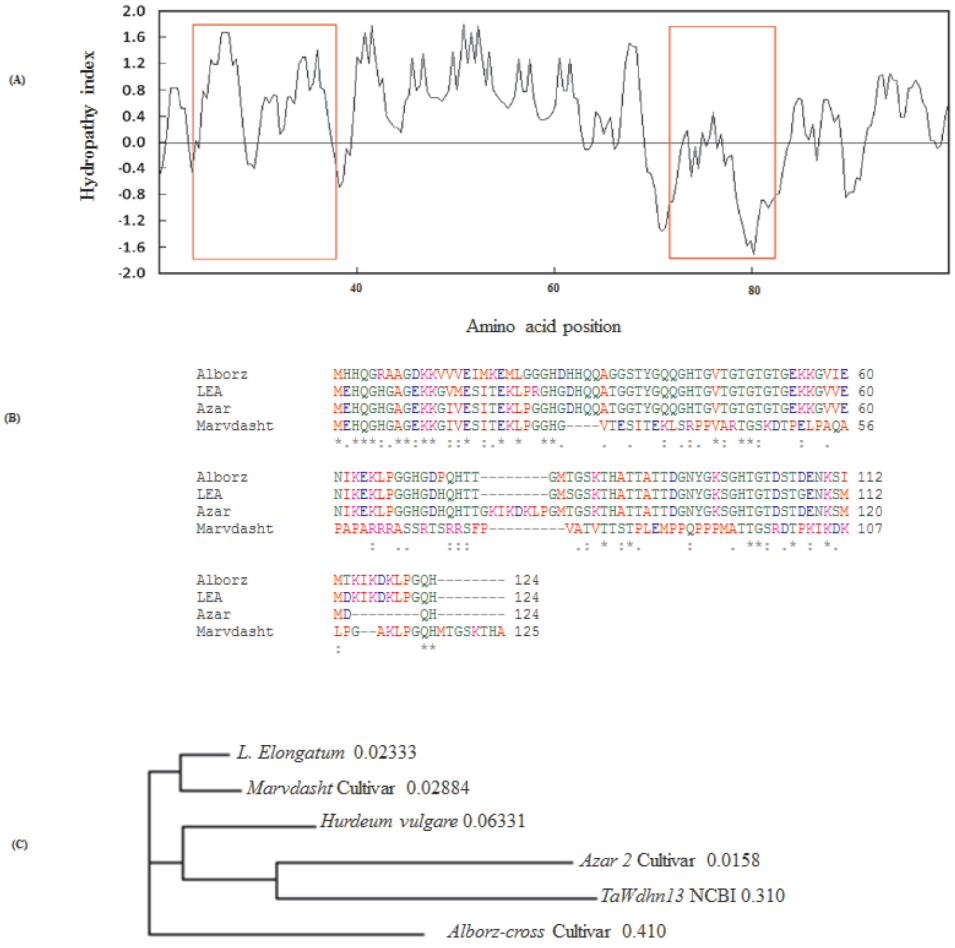

Fig. 3. Hydropathy and motif and phylogenetic analysis of LEA protein sequence. (A) Kyte and Doolittle hydropathy analysis of LEA 2 protein in present study. The first box contains the missing amino acids of LEA protein in the present study, the second box is the retained amino acids of LEA protein in NCBI. (B) Similarity alignment search of LEA protein sequence. (C) Phylogenetic analysis of group 2 LEA proteins from different plant species.

Phylogenetic analysis of pfam00257 containing plant proteins revealed those LEA 2 proteins from *Lophopyrum eleganthum* (AAC05923.1), *Hordeum vulgare* (AGT62697.1) and other species in the present study. Group 2 LEA protein sequence in Azar 2 cultivar is closer to LEA 2 in NCBI database Fig. 3C).

Prediction 3-dimensional structure of LEA

Comparative modeling to build 3D structure of the LEA protein was made based on the experimentally solved structural homologous (Fig. 4B). The amino acid sequences of LEA protein in from different cultivar of bread wheat includes Marvdasht, Azar 2 and Alborz cross were submitted to Phyre 2 server. The atomic coordinates for the proteins were generated based on Hidden Markov Model in the HMMER database. Secondary structure analysis showed that about 19% of amino acid residues in LEA protein forms helix (Fig. 4A). PSORT analysis revealed that wheat LEA protein has a bipartite nuclear localization signal (Kyte and Doolittle, 1982), KIKDKLPG (Fig. 4C).

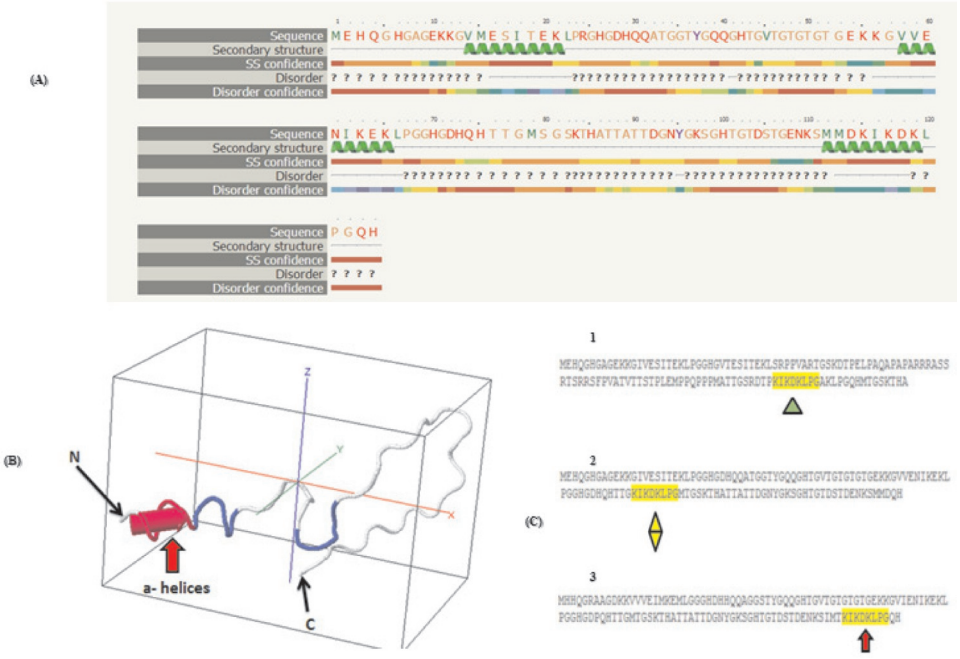

Fig. 4. *In silico* analysis sequence of the LEA protein in different cultivar of bread wheat. (A) Secondary structure of LEA protein. (B) 3-Dimensional structure of the LEA protein Ribbon view of LEA structure for residues 1-96. The N and C termini of the protein are labeled. (C) Position of conserved motif in LEA protein sequence in different cultivar of wheat includes Marvdasht (1), Azar 2 (2) and Alborz cross (3).

Ramachandran plot of the protein models

A Ramachandran plot is a way to visualize backbone dihedral angles ψ against φ of amino acid residues in protein structure (Kendrew *et al.* 1960). 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 (Pauling *et al*. 1951). The residues were classified according to its regions in the quadrangle (Fig. 5). Both right- and left-handed helices lie in regions of allowed conformations in the Ramachandran diagram. However, essentially all a helices in LEA protein is right-handed.

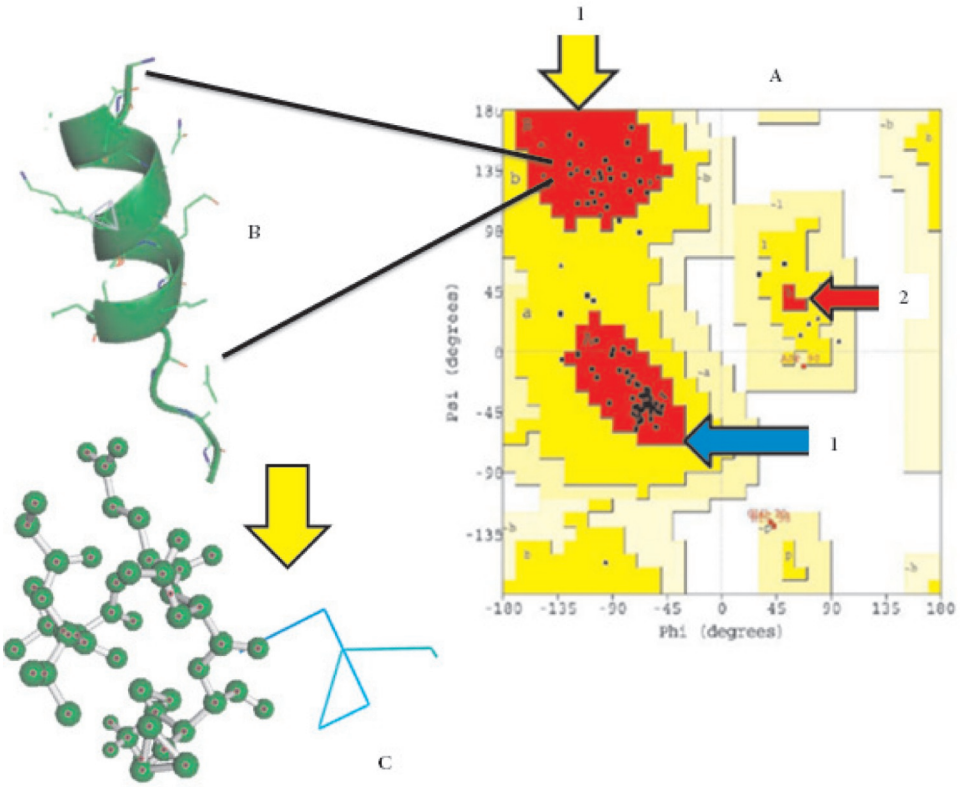

Fig. 5. Ramachandran plot of LEA protein in different cultivar of bread wheat. The plot calculation was done with VMD software. A) Ramachandran plot B) a-helix C) amino acids residue.

Conserved domain

Our results showed *TaWdhn13* gene conserved domain incudes: Dehydrin superfamily (pfam00257) (Fig. 6).

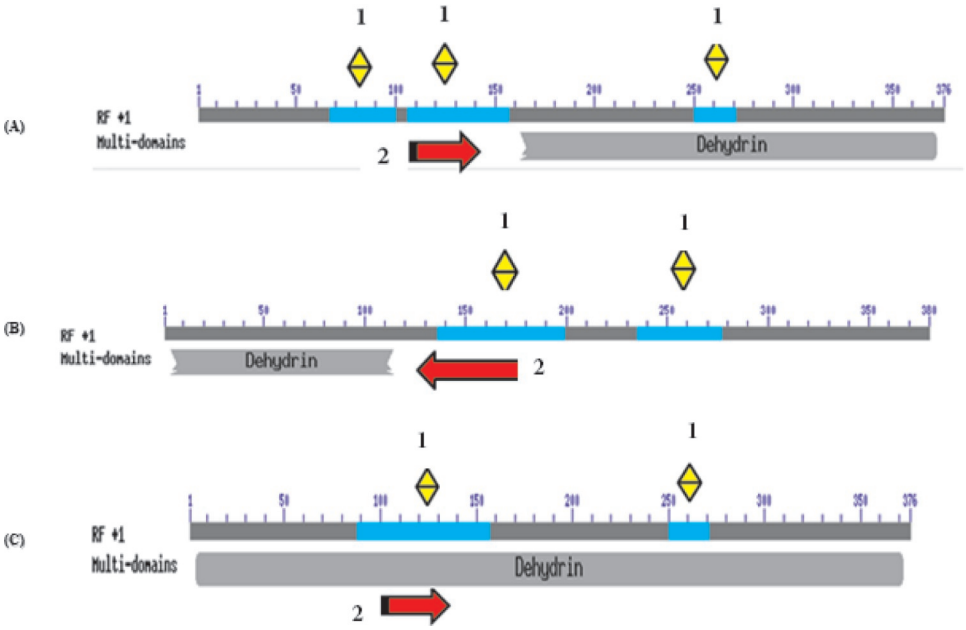

Fig. 6. Conserved domain and sequence for *TaWdhn13* gene in different cultivar of wheat. (A) Marvdasht cultivar. (B) Azar 2 cultivar. (C) Alborz-cross cultivar. In this picture No. 1 reveal the conserved sequences and No. 2 reveal the conserved domain.

Motif structure

In silico analysis showed the motif structure for this gene includes (Fig. 7 and Table 2):

a. 2FE-2S_FER_1 Motif, or ferredoxins, iron-sulfur binding region signature. Motif 2Fe-2S is a structural motif, from the comparison of the coding proteins between rice and spinach chlorine monooxygenase (CMOs), rice CMO potentially shares two conservative motifs including a Rieske-type [2Fe-2S] (Marwa *et al*. 2011) cluster and a mononuclear non-heme Fe binding sequence. These motifs are considered to be essential for the function of CMO (Marwa *et al*. 2011).

b. INTEGRIN_BETA Motif. Integrins areexpressed on the cell surface as a noncovalently linked heterodimer consisting of α and β subunit, which conveys specificity in cell-cell adhesion, cell–extracellular matrix (ECM) adhesion, immune cell recruitment, extravasation, and signaling (Adam *et al*. 2004).

c. VWFC_1 Motif. VWFC motif has conserved cysteine which was found in many cereal crops like rice (Fan *et al*. 2006).

d. EGF_1 Motif. The EGF_1 motif or EGF-like domain signature 1 is an evolutionary conserved protein domain, which derives its name from the epidermal growth factor where it was first described. It comprises

about 30 to 40 amino-acid residues and has been found in a large number of mostly animal proteins (Downing *et al*. 1996).

e. DEFENSIN Motif. Defensins motifs are small Cysteine-rich cationic proteins found in both vertebrates and invertebrates. They have also been reported in plants. They are, and function as, host defense peptides. They are active against bacteria, fungi and many enveloped and nonenveloped viruses (Selsted *et al*. 1985).

Table 2

Location and sequence to identified motifs in the *TaWdhn13* gene in different cultivar of bread wheat

Symbol	Location	Motif
	95-106	EGF_1
	20-33	INTEGRIN_BETA
	61-109	
	229-283	
	230-283	VWFC_1
	235-283	
	377-402	
	372-380	2FE2S_FER_1
	380-408	DEFENSIN

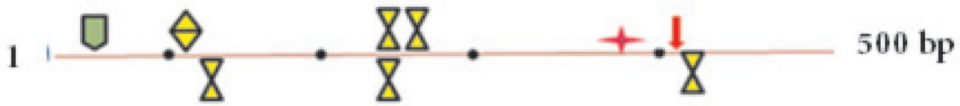

Fig. 7. Map of the motif structure sequence in the *TaWdhn13* gene in different cultivar of bread wheat.

Conserved motifs in TaWdhn13 gene

Finally, four conserved motifs, accgGTTGGCACGTAcgg, ggaCACACcgg and gGTGATCAc and aGCACca were found in the sequence of this gene (Fig. 8).

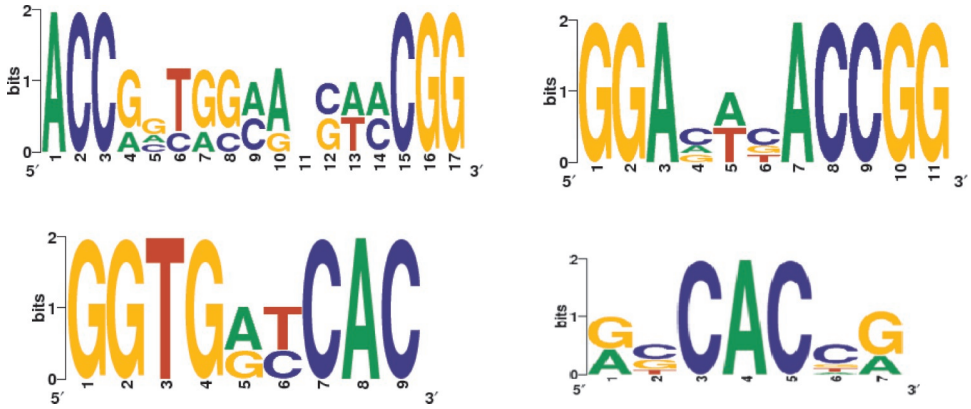

Fig. 8. WebLogo plot of consensus motif in this sequence

DISCUSSION

The fast and reliable method to obtain the full-length genomic sequence of *TaWdhn13* gene might be from genome sequencing. Prediction of 3D structure of a protein molecule signifies an important step towards understanding the structure–function relationships in the concerned protein family (Attwood and Parry-Smith, 2005). After 32 years since the first LEA protein discovered (Babu *et al*. 2004), the molecular functions and mechanisms of most LEA. proteins remain largely unknown (Baker *et al*. (1988)) LEA proteins are widely assumed to play a role in cellular dehydration tolerance and in controlling water uptake during imbibition (Babu *et al*. 2004). In the present study, model of LEA protein in the different cultivar of bread wheat was generated from the Phyre 2 server, based on the Structural homologues derived from the HMMER and protein databanks. The generated model could be helpful in understanding functional characteristics of this important class of desiccation tolerant protein. The homology model of plant LEA proteins, generated in this study, could extend investigations at determining the mechanistic function of important class of proteins. LEA proteins have been found in phylogenetically distant organisms and have always been related to abiotic stress tolerance, especially desiccation tolerance. However, no unifying concept for their physiological role(s) and modes of action has been attained so far. In this study a total of 5 motifs structure and one conserved domain identified in *TaWdhn13* gene sequence in different of bread wheat cultivar. In conclusion, these results confirm the importance of the 7-amino acid motif repeat units and hydrophilicity of the group 2 LEA protein.

REFERENCES

Adam L.F., Heidi K., Tresa C. 2004. Cellular integrins function as entry receptors for human cytomegalovirus via a highly conserved disintegrin-like domain. PANS, vol. 101. 15470-15475.

Alsheikh M.K., Heyen B.J., Randall S.K. 2003. Ion binding properties of the Dehydrin ERD14 are dependent upon phosphorylation. J Biol Chem 278, 40882–40889. DOI: 10.1074/jbc.M307151200

Attwood T.K., Parry-Smith D.J. 2005. In Introduction to Bioinformatics. Ed Wood (ed.). Pearson Education Asia ltd, India.

Babu R.C., Zhang J., Blum A., Ho T.H.D., Wu R., Nguyen H.T. 2004. HVA1, a LEA gene from barley confers dehydration tolerance in transgenic rice (Oryza sativa L.) via cell membrane protection. Protein Science 166, 855–862.

Baker J, Steel C, Dure L. (1988) Sequence and characterization of 6 LEA proteins and their genes from cotton. Plant Molecular Biology 11, 277–291.

Battaglia M., Olvera-Carrillo Y., Garciarrubio A., Campos F., Covarrubias A. 2008. The Enigmatic LEA Proteins and Other Hydrophilins. Plant Physiology. 148: 6-24. DOI: 10.1186/1471-2164-9-118

Bray E.A. 2002. Classification of the genes differentially expressed during water-deficit stress in Arabidopsis thaliana: An analysis using micro array and differential expression data. Ann. Bot. 89: 803-811. DOI: 10.1093/aob/mcf104

Cheng Z.Q., Targolli J., Huang X.Q., Wu R. 2002. Wheat LEA genes, PMA80 and PMA1959, enhance dehydration tolerance of transgenic rice (*Oryza sativa* L.). Molecular Breeding 10, 71–82. DOI: 10.1007/BF00027385

Downing A.K., Knott V., Werner J.M., Cardy C.M., Campbell I.D., Handford P.A. 1996. Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders. Cell 85 (4): 597–605.

Eom J., Baker W.R., Kintanar A., Wurtele E.S. 1996. The embryo-specific EMB-1 protein of Daucus carota is flexible and unstructuredin solution. Plant Sci. 115, 17-24.

Fan C.H., Xing Y.Z., Mao H.L., Lu T.T., Han B. 2006. GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor. Appl. Genet., 112: 1164-1171.

Gray J., Close P.S., Briggs S.P., Johal G.S. 1997. A novel suppressor of cell death in plants encoded by the Lls1 gene of maize. Cell, 89:25-31.

Hu T.Z. 2008. OsLEA3, A Late Embryogenesis Abundant Protein Gene from Rice, Confers Tolerance to Water Deficit and Salt Stress to Transgenic Rice. Russian Journal of Plant Physiology. 4, 530-537. DOI: 10.1134/S1021443708040158

Jeffrey L., Ried M.K., Walker-Simmons A. 1993. Group 3 Late Embryogenesis Abundant Proteins in Desiccation-Tolerant Seedlings of Wheat (Triticum aestivum). Plant Physiol. 102: 125-131.

Kalemba E.M., Pukacka S. 2007. Possible roles of LEA proteins and sHSPs in seed protection: a short review. Biological letter. 44 (1): 3-16.

Kendrew J.C., Dickerson R.E., Strandberg B.E., Hart R.G., Davies D.R, Phillips D.C., Shore VC 1960. Structure of myoglobin: a three-dimensional Fourier synthesis at 2Å resolution. Nature 185 (4711): 422–427

Koag M.C., Fenton R.D., Wilkens S., Close T.J. 2003. The binding of maize DHN1 to lipid vesicles. Gain of structure and lipid specificity. Plant Physiology. 131: 309–316. DOI: 10.1104/pp.900061

Kobayashi F., Ishibashi M., Takumi S. 2007. Transcriptional activation Cor/Lea genes and increase stress tolerance through expression of a wheat DREB2 homolog in transgenic tobacco. Transgenic Research, 17 (5): 755 – 767. DOI: 10.1007/s11248-007-9158-z

Kyte J., Doolittle R.F. 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105–132.

Liu Q., Kasuga M., Sakuma Y., Abe H., Miura S., Yamaguchi-Shinozaki K., Shinozaki K. 1998. Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways indrought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*. Plant Cell. 10: 1391–1406. DOI: 10.1105/tpc.10.8.1391

Marwa N.M.E., E-seuodi A.A., Abd-Elhady A.A., Edris S.E.A., Abdelhamid N.R., Hassan A.H.M. 2011. Utilization of Some EST Derived SSR Markers to Detect Glutenin Gene Loci in Wheat. Journal of Applied Sciences Research, 7(12): 2147-2155.

Pauling L., Corey H.R., Branson H.R. 1951. The Structure of Proteins: Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain. Proceedings National Academy of Sciences USA 37 (4): 205–211.

Rajesh R., Raveendran M., Manickam V. 2008. Prediction of 3-dimensional structure of EMV1, a group 1 late embryogenesis abundant protein of *Vigna radiata* Wilczek. Plant Omics Journal 1 (1): 17-25.

Rasouli H., Kahrizi D., Ghaderinia P. 2013. Identification of conserved domains and motifs for TaWdhn13 gene in *Triticum aestivum* by *in silico* analysis. Advances in Environmental Biology. 7 (4): 586-590

Selsted M.E., Harwing .SS., Ganz T., Schiling J.W., Lehrer R.I. 1985. Primary structure of three human neutrophil defensins. J.Clin. Invest. 76, 1436-1439.

Simon B., Alan T., Michael J., Brian M., Martin H., Volker L. 2009. Bioinformatics and protein expression analyses implicate LEA proteins in the drought response of *Collembola*, Journal of Insect Physiology. 55: 210–217

Wu X., Gong F., Wang W. 2013. Functional assignment to maize group 1 LEA protein EMB564 within the cell nucleus using computational analysis. Bioinformation 9(6): 276-280. DOI: 10.6026/97320630009276

Zhu W., Li Y., Hu M. 2007. Genome scale identification and analysis of LEA genes in rice (*Oryza sativa* L.). Plant Science. 172: 414-420.