American-Eurasian J. Agric. & Environ. Sci., 14 (10): 1044-1048, 2014
ISSN 1818-6769

DOI: 10.5829/idosi.aejaes.2014.14.10.12424

Structural Validation and Homology Modeling of Lea 2 Protein in Bread Wheat

[1]Hassan Rasouli and [2]Bahman Fazeli-Nasab

[1]Agricultural Biotechnology, Graduated from Kermanshah University, Iran
[2]Center of Agricultural Biotechnology and Center of Agricultural Research,
University of Zabol, Zabol, Iran

Abstract: *LEA proteins* are late embryonic proteins abundant in higher plant embryos. It has been found that *LEA* genes are a gene family and play an important roles in the protection of water stress. *LEA genes* have been identified in many plant species. These genes are transcriptionally activated, produce the accumulated proteins and metabolites and protect cell structure from stress damage. In the current study, a complete *LEA* protein structure model was generated so as to understand the mechanistic detail of its functioning. The stereo chemical quality of the model was checked by PROCHECK-NMR, Wincoot Software, WHATIF, ProSA and QMEAN servers. The model was selected, the energy was minimized and simulated for 1.5ns. The result of the study may be a guiding point for further investigations on LEA 2 protein and its role in molecular mechanism of plant cells. Finally, the QMEAN Z-score of -1.06 indicates the overall model quality of *LEA 2 protein*.

Key words: Homology Modeling · Bioinformatics · *LEA Protein*

INTRODUCTION

Late embryogenesis abundant (*LEA*) proteins were initially identified in the late stage of seed maturation in cotton and wheat [1]. *LEA proteins* are associated with desiccation tolerance in orthodox seeds and resurrection plants and have been proposed to act as ion scavengers, molecular chaperones, or as cytoskeletal components [1, 2], although their molecular mechanisms are not fully elucidated. Recent studies reported that *LEA*-like proteins are also present in nematodes [3, 4] and rotifers [2]. 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 [5]. These biophysical features suggest that *LEA proteins* may carry out a bipartite function under different water states [1, 3, 5]. Bioinformatics is an interdisciplinary research area, which may be defined as the interface between biological and computational sciences. It greatly helps in management of complex and scattered biological data, sequence analysis and algorithmic designing. However, by using the *in silico* analysis we can analyze the protein sequences [6]. Hence the present paper enlists some of the physiochemical and functional properties of *LEA 2 protein* and provides information into its three-dimensional structure.

MATERIALS AND METHODS

Operating System: The study was conducted using Intel(R) Core (TM) i5-2410 M CPU @2.30Ghz 4 core processor and 64 bit operating system (HP Pavilion g6).

Multiple Sequence Alignment and Homology Modeling: PDB file of *LEA 2 protein* was generated by Phyre 2 servers. In order to build a model of protein domain, Multiple Sequence Alignment was performed between full length *LEA protein* sequence and another protein domain sequence in this database. To build the model of the *LEA* protein with more homology, high resolution (1.80 A) structure of *LEA* protein model in phyre 2 server was selected as template. Model construction and regularization (including geometry optimization) of model were done by optimization protocol in YASARA. The energy of the model was minimized using the standard protocols of combined application of simulated annealing, conjugate gradient and steepest descent.

Corresponding Author: Hassan Rasouli, Agricultural Biotechnology, Graduated from Kermanshah University, Iran.

The model of *LEA protein* was joined in a single coordinate file using in-house perl script. Output file was then utilized for the loop construction and the sequence given for loop construction was "KIKDKLPG".

Model Purgation: The newly constructed model was solvated and subjected to energy minimization using the steepest descent and conjugate gradient technique to eliminate unwanted contacts between structural water molecules and protein atoms. In this study, MD simulation study was undertaken by using YAMBER 3 [7] package for the model refinement, which was used to reduce the steric clashes between the residues. The constructed *LEA 2 protein* model had to be refined in order to stabilize the backbone. The data obtained after simulation was analyzed for trajectory projection.

Model Reputation: The accuracy of predicted model and its stereo chemical properties were evaluated by PROCHECK-NMR [8]. The model was selected on the basis of various factors such as overall G-factor, no. of residues in core that fall in generously allowed and disallowed regions in Ramachandran plot (Fig. 2) The model was further analyzed by WHATIF [9], QMEAN [10, 11] and ProSA [12]. ProSA was used for the display of Z-score and energy plots.

RESULTS AND DISCUSSION

Model Building: Sequence alignment of *LEA 2 protein* with sequences of troponin I domain and elongation factor 1-beta domain by using the phyre 2 server, revealed more sequence homology with troponin I domain (ID= 83%) which was selected as template for the model building of *LEA 2 protein*. To build the model, 6 times PSI-BLAST was done with the maximum E-value allowed for template being 0.005. The maximum number of templates considered for model building was 6 along with maximum of 5 ambiguous alignment, 4 oligomerization state and number of unaligned loop residue to add to termini being 10. Using troponin I domain sequence modeling of *LEA 2 protein* domains was done with the help of YASARA (Fig. 1).

Model Reputation: The model showed good stereo chemical property in terms of overall G-factor value of -0.68 indicating that geometry of the model corresponds to the probability conformation with 97.4% residues in the

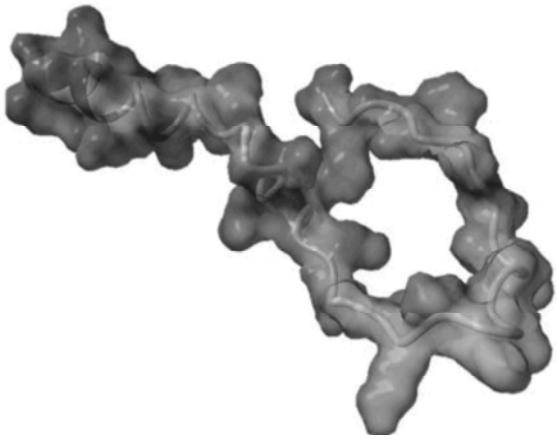

Fig. 1: LEA 2 protein model generated using YASARA.

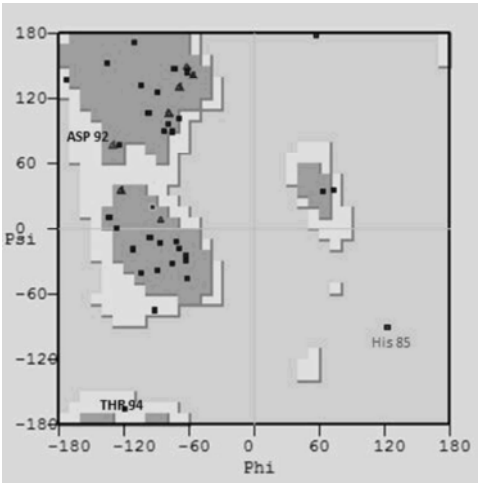

Fig. 2: Ramachandran Plot analysis of *LEA 2 protein*.

core region of Ramachandran plot showing high accuracy of model prediction. The number of residues in allowed and generously allowed region was 86% and 2.6%, respectively and none of the residues were present in the disallowed region of the plot (Figure 2).

The Plot Statistics Are: Total number of residues-124 with 86% in most favored regions [A, B, L], 21.4% in additional allowed regions [a,b,l,p], 2.6% in generously allowed regions and 0% in disallowed regions.

Model Refinement: Model refinement was carried out to improve the accuracy of *LEA 2 protein* model. The newly constructed model was solvated in a box with the dimension of 104.420 x 79.071 x 83.466 Å3 with 3324 number of water molecules and was subjected to

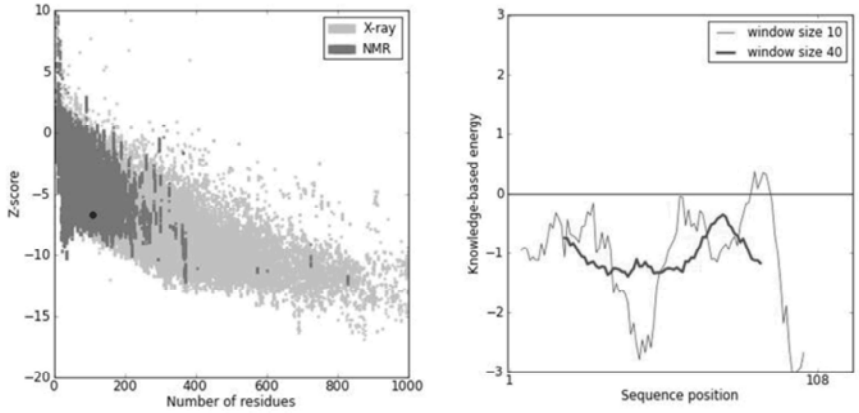

Fig. 3: ProSA web service analysis of LEA 2 protein model.

sequential application of energy minimization techniques. In the initial phase, the energy was minimized using Steepest Descent followed by Conjugate Gradient.

Finally, the global minimum of *LEA 2 protein* model was obtained by Simulated Annealing. This was performed to minimize strain energies and to eliminate unfavorable contacts between water molecules and protein atoms. YAMBER3 force field in YASARA dynamics was used for the model refinement, which was used to reduce the steric clashes between residues. The constructed *LEA 2 protein* model had to be refined in order to stabilize the backbone. After backbone refinement the energy was again minimized by the application of the above mentioned protocol.

The structure was then subjected to *nvt* ensemble (constant number of entities, isochoric and isothermal) based on dynamic simulation for 1.5 ns. The temperature was 298K, the density was 0.997 and the pH was 7.4 while carrying out refinement under physiological salt concentration of 150mM NaCl.

The trajectory was obtained for overall energy simulation of the modeled *LEA 2 protein* for 1500 picoseconds (ps) and it revealed that overall energy stabilized after a peak of -2589436.038 kJ/mol at 25 ps and tended to remain in plateau phase further for the rest of the period.

This reflected that simulation was achieved with stable energy for the rest of the period (40-1475ps) for the *LEA 2 protein*. Almost the similar trajectory was obtained for the plots of different energy contributions against simulation run time.

Due to steric parameters like bond strain, dihedral angle, bond coloumb and Van der Waal, the contribution was found maximum at 25 ps with the values of 329479.911 kJ/mol, 50047.973 kJ/mol, -3610916.379 kJ/mol and 512036.939 kJ/mol respectively, which stabilize further to a stationary phase for the rest of the period (50-1475ps), except for dihedral angle which shows variations in the value.

Model Validation: ProSA was used to check the three-dimensional model of *LEA 2 proteins* for potential errors. The program displays 2 characteristics of the input structure: its Z-score and a plot of its residue energies. The ProSA Z-score of -6.07 indicates the overall model quality of *LEA 2 protein* (Fig.3). Z-score also measures the deviation of total energy of the structure with respect to an energy distribution derived from random conformations. The scores indicate a highly reliable structure and are well within the range of scores typically found for proteins of similar size. The energy plot shows the local model quality by plotting knowledge-based energies as a function of amino acid sequence position. QMEAN analysis was also used to evaluate and validate the model.

The QMEAN score of the model was 0.40 and the Z-score was -1.06 which was very close to the value of 0 and this shows the good quality of the model because the estimated reliability of the model was expected to be in between 0 and 1 and this could be inferred from the density plot for QMEAN scores of the reference set (Fig. 4A). A comparison between normalized QMEAN score (0.40) and protein size in non-redundant set of PDB structures in the plot revealed different set of Z-values for different parameters such as C-beta interactions (-0.76), interactions between all atoms (-1.32), solvation (0.60), torsion (-4.34), SSE agreement (-0.18) and ACC agreement (1.25) (Fig. 4B).

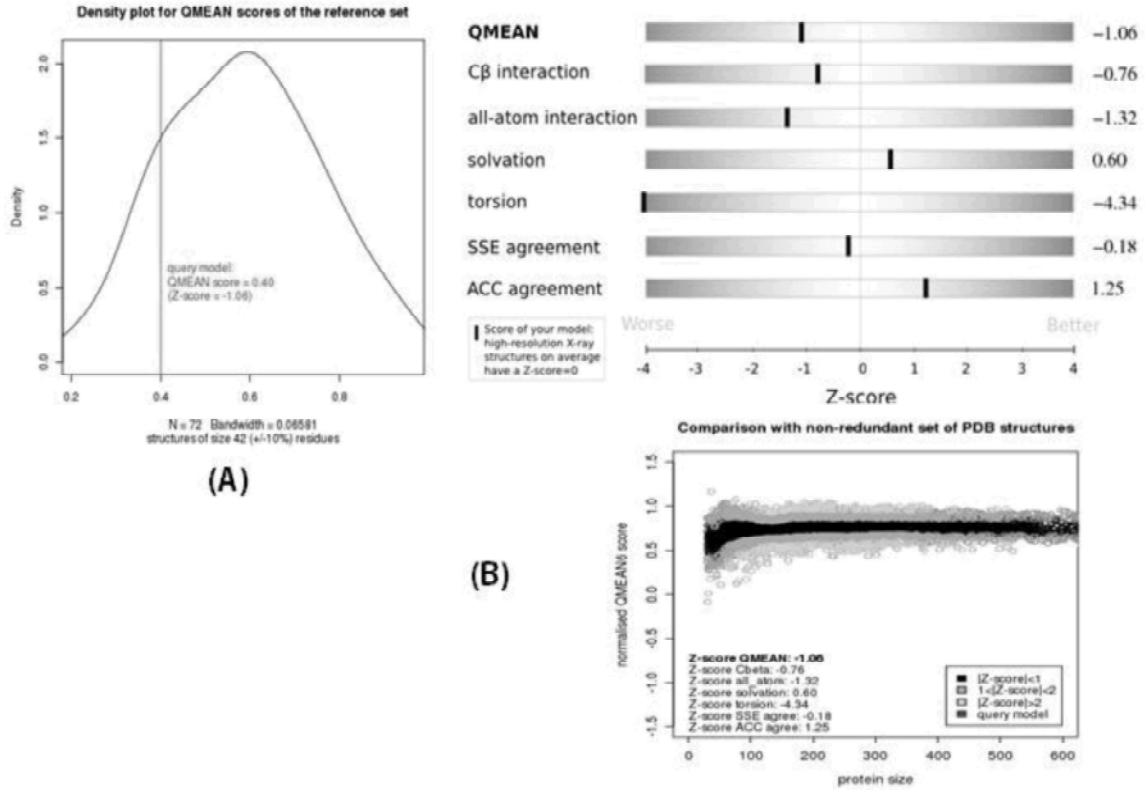

Fig. 4: A) Density plot for QMEAN showing the value of Z-score and QMEAN score. B) Plot showing the QMEAN value as well as Z-score.

CONCLUSION

The generated model could be helpful in understanding the functional characteristics of this important class of desiccation tolerant protein. The homology model of plant *LEA proteins*, generated in this study, could extend investigations in determining the mechanistic function of important class of proteins. *In silico* studies in general and molecular modeling with molecular dynamics simulation studies in particular have been of great help in understanding the structure, function and mechanism of the action of proteins, particularly the membrane proteins. The generated model was also subjected to structural validation. Structure validation by WHATIF, PROCHECK-NMR, ProSA and QMEAN confirmed the reliability of the model.

REFERENCES

1. Cuming, A.C., 1999. LEA Proteins, in: P.R. Shewry, R. Casey (Eds.), Seed Proteins. Kluwer Academic Publishers, the Netherlands, pp: 753-780.
2. Tunnacliffe, A., J. Lapinski and B. McGee, 2005. A putative LEA protein, but no trehalose, is present in anhydrobiotic bdelloid rotifers. Hydrobiologia 546: 315-321.
3. Gal, T.Z., I. Glazer and H. Koltai, 2004. An LEA group 3 family member is involved in survival of C. elegans during exposure to stress. FEBS Lett., 577: 21-26.
4. Goyal, K., C. Pinelli, S.L. Maslen, R.K. Rastogi, E. Stephens and A. Tunnacliffe, 2005. Dehydration-regulated processing of late embryogenesis abundant protein in a desiccation-tolerant nematode. FEBS Lett 579: 4093-4098.
5. Wu, X., F. Gong and W. Wang, 2013. Functional assignment to maize group 1 LEA protein EMB564 within the cell nucleus using computational analysis. Bioinformation, 9(6): 276-280.
6. Pevzner, P. and R. Shamir, 2011. Bioinformatics for Biologists: Cambridge University Press.
7. Krieger, E., T. Darden, S.B. Nabuurs, A. Finkelstein and G. Vriend, 2004. Making optimal use of empirical energy functions: force-field parameterization in crystal space. Proteins, 57(4): 678-683.

8. Vriend, G., 1990. WHAT IF: a molecular modeling and drug design program. Journal of Molecular Graph, 8(1): 52-56.
9. Benkert, P., S.C. Tosatto and D. Schomburg, 2008. QMEAN: A comprehensive scoring function for model quality assessment. Proteins, 71(1): 261-267.
10. Benkert, P., M. Künzli and T. Schwede, 2009. QMEAN server for protein model quality estimation. *Nucleic Acids Res*., 37: 4-10.
11. Novotny, W.F., T.J. Girard, J.P. Miletich and G.J.J. Broze, 1988. Platelets secrete a coagulation inhibitor functionally and antigenically similar to the lipoprotein associated coagulation inhibitor. Blood, 72(6): 2020-2025.
12. Wiederstein, M. and M.J. Sippl, 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research, 35: 407-410.