Journal of Advanced Bioinformatics Applications and Research.
Vol 1, Issue 1, June 2010, pp 7-10 http://www.bipublication.com

COMPARATIVE MODELING AND FUNCTIONAL CHARACTERIZATION OF 3-D STRUCTURE OF A LATE EMBRYOGENESIS ABUNDANT PROTEIN OF *ARABIDOPSIS THALIANA*

BOOBALAN. C, BHARATHI N.
Department of Plant Molecular Biology &Biotechnology, TamilNadu Agricultural University.

ABSTRACT:

LEA proteins have been widely analyzed in plants. They play an important role in seed maturation and osmotic stress. LEA proteins in seeds are related to desiccation tolerance. Among the 5 groups of Lea protein, Group-4 LEA protein of Arabidopsis *thaliana* is identified for the computational analysis. The homology modeling of the putative protein was generated by using the Modeller software. From 5 models generated from modeler, the final model is selected after analyzing and validating the structure by using the validation server WHAT IF interface. The model is further optimized by performing Energy minimization technique (Steepest Descent). The functional characterization of the modeled protein involves prediction of function, active site region and other structural analysis using various proteomics tools in Bioinformatics. The model used in further functional characterization of this LEA protein.

Keywords: LEA proteins; *Arabidopsis thaliana*; Homology modeling; validation

Abbreviations: PDB-Protein Data Bank, BLAST-Basic Local Alignment Search Tool, SAVS-Stuctural Analysis and Verification Server.

Introduction

Late embryogenesis abundant (LEA) 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 [1].

Several groups of LEA protein genes have been demonstrated to confer water-deficit and salt-stress tolerance.

It has been shown that protein structures are more conserved than protein sequences amongst homologues, but sequences falling below a 20% sequence identity can have very different structure, and reach what is called in homology modelling, the twilight zone. [2].

Evolutionarily related proteins have similar sequences and naturally occurring homologous proteins have similar protein structure. It has been shown that three-dimensional protein structure is evolutionarily more conserved than expected due to sequence conservation [3].

The sequence alignment and template structure are then used to produce a structural model of the target. Because protein structures are more conserved than DNA sequences, detectable levels of sequence similarity usually imply significant structural similarity [4]. On the basis of sequence similarities, LEA proteins have been classified in six groups [1]. In the present study, effort was made to generate the three-dimensional (3D) structure of the putative protein (*Arabidopsis thaliana*) based on the available template structural homologues from Protein Data Bank and the model validated with standard parameters. This study could prove useful in further functional

characterization of this important group of proteins.

Materials and Methods

The putative sequences of *Arabidopsis* (NCBI GenBank accession number NP_974009.1) and other sequences examined in this study were retrieved from the database http://www.ncbi.nlm.nih.gov.

Structurally homologous subsets of the experimentally determined 3D structures of the putative proteins (*Arabidopsis*) were retrieved from PDB. The template used for comparative modeling of putative protein (*Arabidopsis*) Chain A, Crystal Structure Of The Sodium-Potassium Pump from pig (PDB ID -3B8E) with sequence similarity 30%

Comparative Modeling of putative Protein (*Arabidopsis thaliana*)

Tertiary structure of the *Arabidopsis thaliana* LEA putative protein was modeled by submitting the deduced amino acid sequences to the MODELLER software. The Atomic coordinates for the protein models were generated by aligning to the structural homologues in the comparative modeling program of MODELLER.

Figure1: Predicted 3-Dimensional Structure of the **putative Protein (*Arabidopsis thaliana*)**

Validation of putative Protein (*Arabidopsis thaliana*) Model

The hypothetical protein models generated were analyzed by using energy minimization (steepest Descent) algorithms.

Functional characterization of putative Protein (*Arabidopsis thaliana*)

The putative protein function generated by using Pfam database. Active sites generated were analyzing online submitting to Qsiter.B factor was predicted by using visualization Tool pymol.

Results and Discussion

Comparative Modeling of putative Protein (*Arabidopsis thaliana*)

Tertiary structure of a protein is build by using homology modeling. Comparative modeling to build 3D structure of the putative Protein (*Arabidopsis thaliana*) was made based on the experimentally solved structural homologues. The amino acid sequences of putative protein (*Arabidopsis thaliana*) were submitted to MODELLER server and atomic coordinates for the proteins were generated.

The hypothetical protein models created were stored as PDB output file. The hypothetical proteins were visualized and computed by Swiss PDB Viewer and Rasmol. The 3D structure of the proteins were represented by cartoon display and colored based on the secondary structure (Fig. 1).

Validation of putative Protein (*Arabidopsis thaliana*) Model

The hypothetical protein models generated by using MODELLER. Among the 5 models one model was selected .Because the final energy minimization value of model-3 is NP_974009.1.B99990003-6090.705 KJ/mol.

Model No	Energy value before energy inimization	Energy value after energy minimization
NP_974009.1.B99990001	**594.324**	**-5529.510**
NP_974009.1.B99990002	**281.407**	**-6090.705**
NP_974009.1.B99990003	**-238.714**	**-5664.901**
NP_974009.1.B99990004	**141395.453**	**-5664.901**
NP_974009.1.B99990005	**47.74**	**-5686.493**

Table-1:energy minimization statistics of putative Protein (*Arabidopsis thaliana*) models computed by steepest Descent algorithms.

In the Ramamchandran plot analysis,the residues were classified according to its regions in the quadrangle. The Ramachandran map for putative Protein (Fig. 2). Among the 5 models ,NP_974009.1.B99990003 model only have low amino acids disallowed regions three:GLN5,THR120,GLN104).after performing loop modeling these amino acids are comes under allowed regions.

NP_974009.1.B99990003 model generated and were analyzed online submitting WHAT IF interface. WHAT IF interface was used for adding missing side chain in our model. Finally model will be generated. The final refined model is reliable.

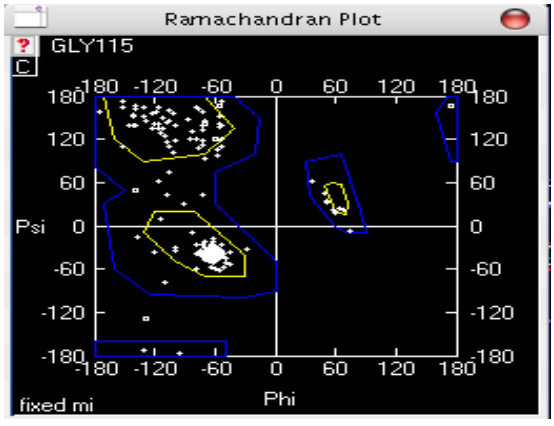

Figure2: Ramachandran map of putative Protein (*Arabidopsis thaliana*).

The Plot calculation was done with Steepest Descent algorithm.

Functional characterization of putative Protein (*Arabidopsis thaliana*) Model

Putative protein model generated and were function predicted by online submitting Pfam database.putative protein have a domain for LEA 4 group.they are confirmed that putative sequence is LEA 4 protein family.

.Pfam results shows that putative protein have a domain for LEA 4 group stress tolerance .other domain is present that domain doesn't have a precise function.

Active site for putative **models** is amino acids MASHGGGSTLAGTAGL and position is 1-16.this active site is predicted by using Active site prediction web server based on characteristics of spatial distribution of hydrophobicity in a protein molecule.B factor for putative models amino acids QILT **(Fig. 3)**

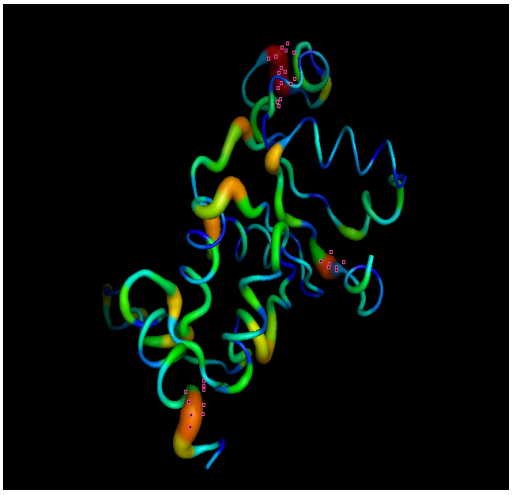

Figure 3: Dot lines indicate the B factor amino acids QILT respectively.

References

1. Bray EA (1993) Molecular responses to water deficit.Plant Physiol 103: 1035-1040.
2. Chothia C and Lesk AM (1986). The relation between the divergence of sequence and structure in proteins. *EMBO J*5:823-6.
3. Kaczanowski S and Zielenkiewicz P (2010). Why similar protein sequences encode similar three-dimensional structures? *Theoretical Chemistry Accounts* 125:543-50
4. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. (2000). Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 29: 291-325.