PDB50: A special symposium celebrating the 50th anniversary of Protein Data Bank, 4-5 May, 2021.

In silico structural survey of newly identified late embryogenesis abundant proteins (LEAPs) from *Ramonda serbica* and their structure - function relationship

Strahinja Stevanović and Marija Vidović
Institute of Molecular Genetics and Genetic Engineering, Laboratory for Plant Molecular Biology, University of Belgrade, Belgrade, Serbia

Introduction

Desiccation or extreme water loss leads to protein denaturation, aggregation, and degradation and impairs membrane lipid fluidity, resulting in loss of membrane integrity at the cellular level. The induction of late embryogenesis abundant proteins (LEAPs) is considered an essential component of desiccation tolerance strategy in so-called resurrection plants. This heterogeneous group of hydrophilic, non-globular proteins is characterized by a high structural plasticity that allows them to adopt a random conformation in aqueous solutions that transforms into α-helices during dehydration [1]. Therefore, LEAPs can interact with various ligands and partners, including ion sequestration and stabilization of membranes and enzymes during freezing or drying [2].

Methodology

We identified **164** members of the LEA gene family from endemic species *Ramonda serbica*. Then, we retrieved information from PFAM [3], InterPro [4], and Protein Data Bank [5] resources based on sequence homology. Additionally, multiple sequence alignment (MSA) in MAFFT (G-INS-i and L-INS-i) and Clustal OMEGA allowed us to produce the global alignment in agreement with a classification of *A. thaliana* LEA proteins [6, 7].

After classification, we built homology models using Phyre2, Modeller and trRosetta [8, 9, 10]. The resulting models were clustered and superimposed in order to obtain a consensus 3D structure for each subfamily.

Helix analysis were conducted using HelixQuest [11]. For data curation and plotting we used pandas and seaborn. For structure manipulation and visualization we used pymol.

Conclusion

We have built homology models using Protein Data Bank structure information. Homology models show variation in LEAPs secondary and tertiary structures with 30% on average primary sequence similarity between subfamilies. We found that LEAPs adopt diverse helix-bundle topologies, with an exception of LEA_2 which adopts alphabeta-fold type and is related to membrane, lipid interaction. LEA_4 and SMP may adopt structure which resembles α-synuclein.

The information obtained from the representative structural models is key to understanding the function of LEAPs and the regulation of their intrinsic structural disorder-to-order transition during desiccation. This will pave the way for the identification of LEAPs endogenous partners and their targets in the cell and provide further insights into the protective mechanisms of desiccation tolerance.

Ref.
[1] Hundertmark M, Hincha DK (2008). BMC Genom 9:1. [2] Chakrabortee S et al. (2012). Mol. BioSyst. 8:210-19. [3] Mistry S et al. (2021) Nucleic Acids Res. 49(D1):D412-D419. [4] Blum M et al. (2020) Nucleic Acids Res 49(D1):D344-D354. [5] Berman HM et al. (2000) Nucleic Acids Res., 28: 235-242. rcsb.org [6] Katoh K and Standley DM (2013) Mol Biol Evol 30(4):772-780. [7] Maderia F et al. (2019) Nucleic Acids Res. 47(W1):W636-W641. [8] Kelley LA et al. (2015) Nat. Protoc 10, 845-858. [9] Sali A and Blundell TL (1993) J. Mol. Biol. 234, 779-815. [10] Yang J et al. (2020) PNAS, 117: 1496-1503. [11] Gautier R et al. (2008) Bioinformatics 24(18):2101-2.

Acknowledgements

This work was supported by the Science Fund of the Republic of Serbia (PROMIS project LEAPSyn-SCI, grant number 6039663) and by the Ministry of Education, Science and Technological Development, Republic of Serbia (contract no. 451-03-9/2021-14/200042). On the cover: *Ramonda serbica* flower, photo credit to Dr. Lj. Jovanović.

Science Fund of the Republic of Serbia
Program for Excellent Projects of Young Researchers
LEAPSyn-SCI #6039663
https://leapsynsci.com/

Results

All identified LEAPs (164) were clustered into six subfamilies based on difference in sequence domains. The classification summary can be found on *Figure 1*. Further protein analysis is shown on *Figures 2-6*.

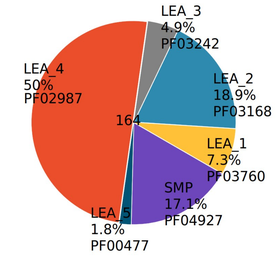

Figure 1. *Ramonda serbica* LEAPs
Classification according to protein family domains (PFAM) and MSA.

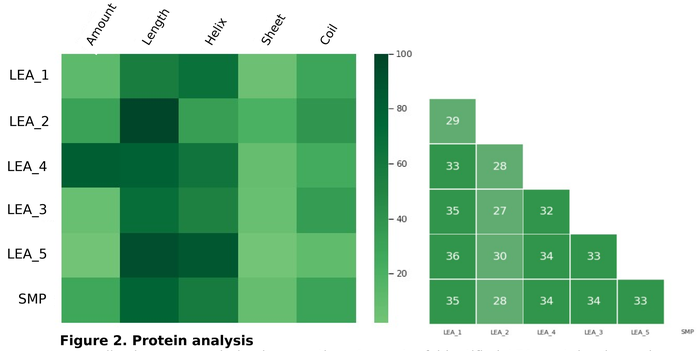

Figure 2. Protein analysis
Normalized sequence-derived properties: Amount of identified LEA proteins in each subfamily; Primary sequence length; Percent of secondary structure type on average (Helix, Sheet, Coil). Values show sequence similarity between subfamilies.

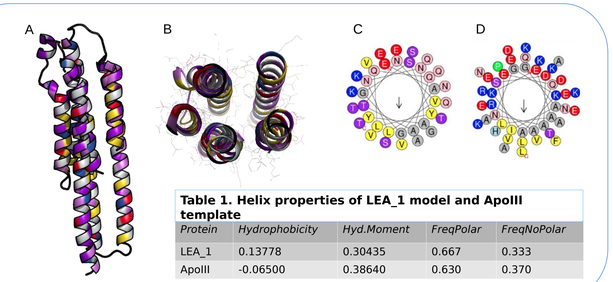

Table 1. Helix properties of LEA_1 model and ApoIII template

Protein	*Hydrophobicity*	*Hyd.Moment*	*FreqPolar*	*FreqNoPolar*
LEA_1	0.13778	0.30435	0.667	0.333
ApoIII	-0.06500	0.38640	0.630	0.370

Figure 3. A model for LEA_1 from *R.serbica*
Side view (A). Top view (B). Model's single helix properties (C) compared to a single helix of a template in (D) which is apolipophorin III (PDB code: 1eq1:A). Colored coded by hydrophobicity scale (C and D) – hydrophobic (yellow), polar (purple and pink), neutral (gray), positively charged (blue), negatively charged (red),

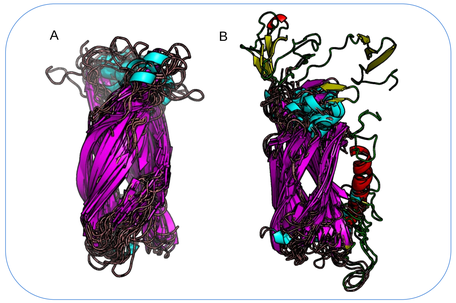

Figure 4. Models for LEA_2 from *R. serbica.*
Ensemble of superimposed 3D models with an average RMSD of 2.2 (A). Additional features unique to certain LEA_2 proteins (B). The structure template is putative LEAP from A. thaliana (pdb codes: 1xo8, 1yyc). The identified conformation is of alphabeta-fold type. It consist of 2 antiparallel β-sheets made of 7 β-strands. The structure resemblance can be found in viruses (PDB code: 3ub0) and animal cells (PDB code: 1x3d). Color coded by secondary structure: sheet (pink/yellow), helix (cyan/red)

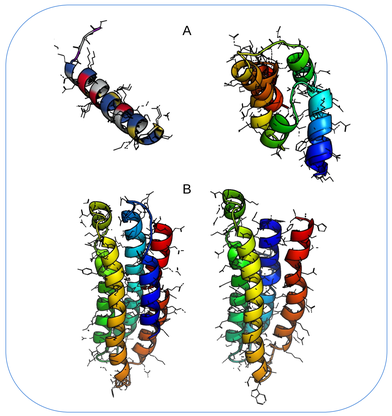

Figure 5. Models for *LEA_3 (A) and LEA_5 (B).* *This group is diverse in helical bundles architecture. Two LEA_5 models are shown, with a 5-helix and 4-helix bundle which dependence on a sequence length. Color coded by hydrophobicity scale (see Figure 3.) and N-to-C termini (blue-to-red).*

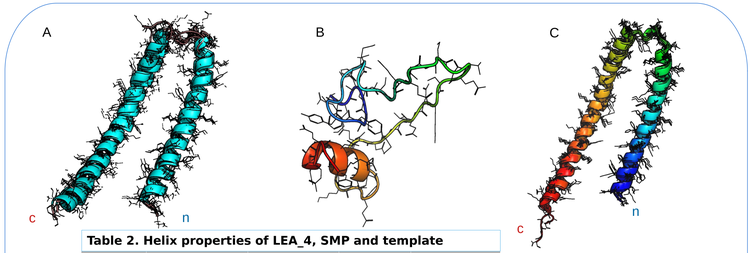

Table 2. Helix properties of LEA_4, SMP and template

*Protein**	*Hydrophobicity*	*Hyd.Moment*	*FreqPolar*	*FreqNoPolar*
α-syn_c	0.15579	0.31889	0.553	0.447
α-syn_n	0.28697	0.25412	0.545	0.455
LEA_4_c	-0.054267	0.303747	0.665000	0.335000
LEA_4_n	0.41565	0.257008	0.650167	0.349833
SMP_c	-0.096278	0.300718	0.6874	0.3126
SMP_n	-0.007516	0.257586	0.6712	0.3288

** calculated for both helices in each 3D structure, annotated with _n or_c*

Figure 6. Models for LEA_4 and SMPs from *R.serbica*
The 3D structure resemble α-synuclein (pdb code: 1xq8:A). Ensemble of superimposed homology models, with RMSD of 1.69 and 1.29 on average for LEA_4 (A) and SMP (B, C) are presented. The model (B) represents disordered structure and it may or may not be similar to any of the disorder-to-order conformation states. Color coded by secondary structure (see Figure 4.) and N-to-C termini (blue-to-red)