

Hydroxyl radical scavenging potential of the late embryogenesis abundant proteins (LEAPs) from *Ramonda serbica* – *in silico* approach

Sonja Milić Komić¹, Strahinja Stevanović², Marija Vidović ²

¹ *Institute for Multidisciplinary Research, Department of Life Sciences, University of Belgrade, 11030 Belgrade, Serbia;*

² *Institute of Molecular Genetics and Genetic Engineering, Laboratory for Plant Molecular Biology, University of Belgrade, Vojvode Stepe 444a, Belgrade, Serbia*



e-mail: sonjamilic@imsi.rs

Ramonda serbica Panc. is a resurrection plant that can survive long desiccation periods (extreme loss of cellular water). The accumulation of late embryogenesis abundant proteins (LEAPs) is a crucial step in desiccation tolerance mechanism. Based on *in vitro* studies, LEAPs can be involved in antioxidative defense, ion sequestration, structural stabilization of both membranes and enzymes during freezing or drying, while by forming intracellular proteinaceous condensates they increase structural integrity and intracellular viscosity of cells during desiccation. Here we investigated the antioxidative potential of LEAPs identified by *de novo* transcriptomics of *R. serbica*, based on their primary and secondary confirmation.

In our previous work [1], we displayed the antioxidative capacity of 20 free proteogenic amino acids (FAA) through determining their hydroxyl radical ($\cdot\text{OH}$, generated in Fenton reaction) scavenging rate by using electron paramagnetic resonance. These results served as a basis for generating a model for prediction of $\cdot\text{OH}$ scavenging activity for selected proteins. In addition, the model was built based on protein primary sequences, hydrophobicity, 3D structure and predicted solvent accessible area. Manually curated data for peptides and proteins with experimentally determined $\cdot\text{OH}$ scavenging rate were used for training and testing. The model was fed into machine learning algorithm and $\cdot\text{OH}$ scavenging potential scale was created using IC_{50} values. By applying our model, we classified 164 LEAPs according to their potential for $\cdot\text{OH}$ scavenging. Further work will focus on the experimental evaluation of the obtained model by measuring of the rate of $\cdot\text{OH}$ scavenging in the presence of recombinantly produced LEAPs.

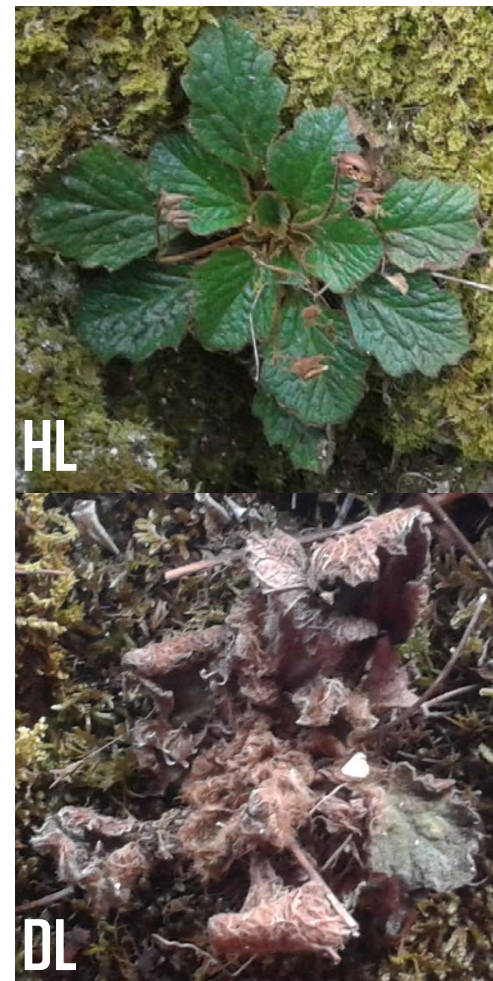
Keywords: desiccation tolerance, electron paramagnetic resonance (EPR), intrinsically disordered proteins, machine learning, resurrection plants, secondary structure.

Acknowledgements: This work was funded 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, the Republic of Serbia (Contract No. 451-03-9/2021-14/200042 and 451-03-9/2021-14/200053).

References:

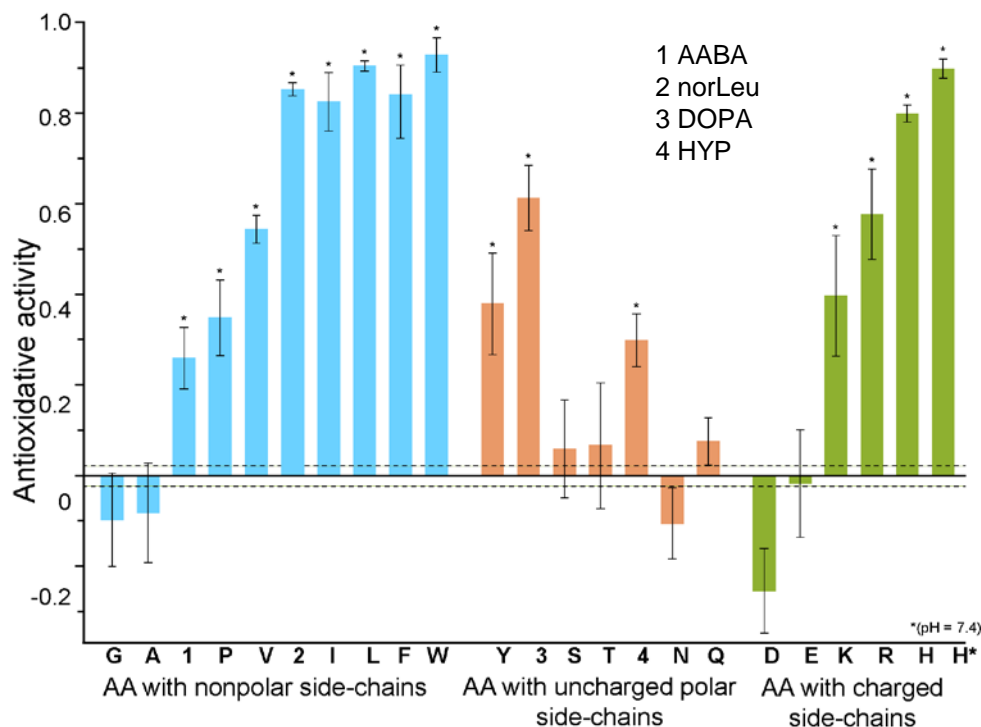
[1] Milic S, Bogdanović Pristov J, Mutavdzic D, Savic A, Spasić, M and Spasojevic I, 2015. *Environ. Sci. Technol.*, 49, 4245-54.

- Resurrection plant *Ramonda serbica* Panc. survives desiccation for a long period and fully recovers metabolic functions already within one day upon watering.
- Desiccation (extreme dehydration) induces protein unfolding and aggregation, destabilization or loss of cellular membrane integrity. Besides, desiccation provokes the accelerated generation of reactive oxygen species.
- Desiccation tolerance mechanism is related to the accumulation of late embryogenesis abundant proteins (**LEAPs**), intrinsically disordered group of proteins (IDP).
- Proposed protective functions of **LEAPs** are associated with antioxidative defense, ion sequestration, structural stabilization of both membranes and enzymes during freezing or drying, while by forming intracellular proteinaceous condensates they increase structural integrity and intracellular viscosity of cells during desiccation.



Vidović M, Franchin C, Morina F, Veljović-Jovanović S, Masi A and Arrigoni G. 2020. *Anal Bioanal Chem*, 412:8299-312.

- By integrating previously done *de novo* transcriptome and homologues protein motifs we have identified 164 members of LEA gene family in *R. serbica*.
- Previously we have determined the antioxidative capacity of 20 free proteogenic amino acids (FAA) was through measuring their hydroxyl radical ($\bullet\text{OH}$, generated in Fenton reaction) scavenging rate by using electron paramagnetic resonance spectroscopy (EPR).



Aim:

To predict the hydroxyl radical scavenging activity (**HRSA**) of identified LEAPs according to their amino acid sequence, taking into account predicted secondary structure, surface accessible area.

Milic et al. 2015. *Environ. Sci. Technol.*, 49: 4245-54.

- A set of peptides and proteins with experimentally determined HRSA (IC₅₀ values)* and their primary sequences studied in Fenton reaction system was used for obtaining manually curated data.
- **Labels** represent values of probability for higher/lower HRSA potential.
 - Binary type, with values 0 and 1.
 - (With 1 meaning higher HRSA potential, and 0 meaning less likely to be HRSA potent)
 - The binary IC₅₀ classification was set with a threshold value of 100 μM.
 - HRSA < 100 uM are considered to have higher HRSA potential.

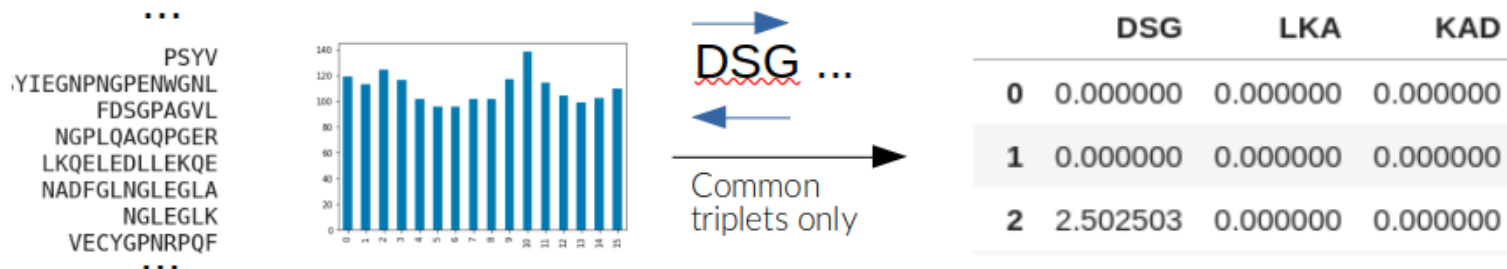
*You et al. 2010 *Food Res Int*, 43, 1167-73
Hou et al. 2001 *J Agr Food Chem*, 49, 4956-60
Mendis et al. 2005 *Life Sci* 77, 2166-78
Qian et al. 2008 *Bioresource Technol*, 99, 3365-71
Rajapakse et al. 2005 *J Nutr Biochem*, 16, 562-569
Sheih et al. 2009 *Bioresource Technol*, 100, 3419-25

Liu et al. 2003 *J Agr Food Chem*, 51, 4389-93
Je et al. 2007 *J Med food*, 10, 401-07
Suetsuna et al. 2000 *J Nutr Biochem* 11, 128-131
Ranathunga et al. 2006 *Eur Food Res Technol* 222, 310-15
Ueno et al. 2013 *J Clin biochem Nutr*, 12-75



Data preparation

- Features:** distinct traits (individual independent variables) of analyzed peptides/proteins which are used as input. They are divided into 5 groups and are `min/max` normalized.
 - F1:** 1-dimensional (percent of each amino acid in the given sequence).
 - F2:** 2-dimensional (triplets), percentage of each triplet present in the sequence, surrounding is taken into account.



Triplet parsing and calculation workflow, e.g. in 2, triplet “DSG” takes 0.25% of the sequence.

- F3:** Amino acid HRSA values determined by EPR (Milic et al. 2015)
- F4:** High-Resolution Hydroxyl Radical Protein Footprinting, (Kiselar et al. 2018)
- F5:** Surface Accessible Area, calculated using Acconpred software.

Equation for F3-F5:

$$F_n = \text{Sum}(F) / \text{sequence_length}$$

F: return value of a function which multiplies amino acid percent with the given parameter

We designed a **deep neural network model** using tensorflow [www.tensorflow.org].

■ **Model definition**

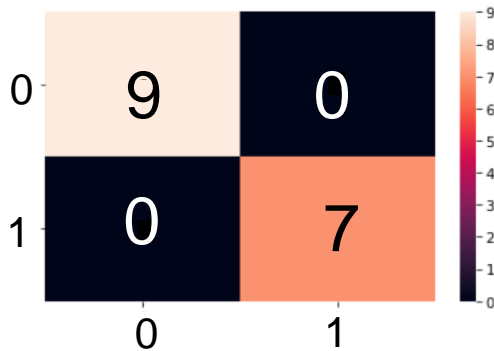
- Sequential model
- 18 dense layers with Rectified Linear Unit (ReLU) activation function
- 12 dense layers with hyperbolic tangent function (tanh) activation
- Output layer with sigmoid activation function

Abadi M. et al. 2015 TensorFlow: Large-scale machine learning on heterogeneous systems.

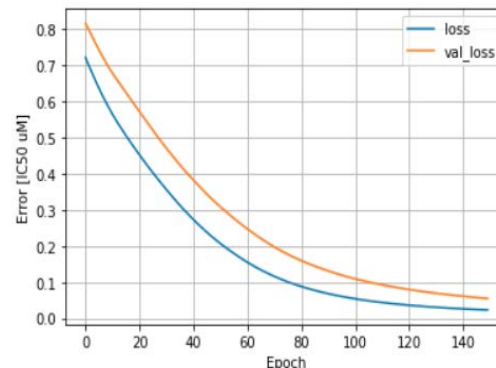
■ **Model compilation and fitting**

- Compiled with a binary loss function, using Adam optimizer.
- Dataset is separated on training & testing based on random seed with 0.25 split parameter
- Fitted using 150 epochs and 256 batch size

99.99 % accuracy is obtained
However, further application tests are required.



Confusion matrix:(9 true positives, 7 true negatives)

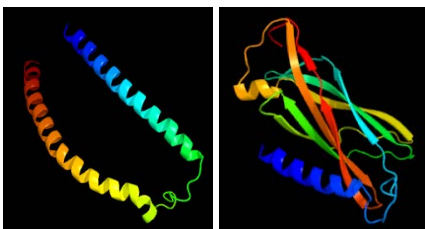


Error of prediction for training (loss) and testing (val_loss) data.

Characteristics of 7.9% of all analyzed LEAPs predicted to have high HRSA.

lea_code	HRSA	HRSA_Hit	Gravy	Helix MEAN	Sheet MEAN	Coil MEAN	Disorder MEAN
RsLEA15	86	TRUE	-0.1131	15.39	38.56	42.37	10
RsLEA136	85	TRUE	-1.3775	72.53	2.12	32.39	81.83
RsLEA1	84	TRUE	-0.0293	17.21	33.05	36.98	12.38
RsLEA98	83	TRUE	-0.1262	20.19	27.12	34.34	21.25
RsLEA19	83	TRUE	-0.1108	16.79	27.46	46.25	15.87
RsLEA53	81	TRUE	-0.5842	21.99	8.64	57.89	55.88
RsLEA154	81	TRUE	-0.0006	25.62	23.63	35.58	11.71
RsLEA56	81	TRUE	-0.5918	17.35	13.95	53.74	55.93
RsLEA85	80	TRUE	-0.582	24.06	8.64	53.76	55.22
RsLEA8	80	TRUE	-0.0247	17.26	29.02	41.95	8.27
RsLEA72	80	TRUE	-0.4154	18.74	5.77	54.16	52.15
RsLEA10	80	TRUE	0.0078	19.15	28.21	43.34	16.36
RsLEA22	80	TRUE	0.1043	16.63	39.48	38.52	6.2

- Using our established deep neural network model, we tested 164 LEAPs identified in *R. serbica* for their HRSA.
- The results showed 72% of the identified LEAPs has a high probability to have HRSA values below the preset criteria and 7.9%, which presents good candidates for the experimental validation by measuring IC_{50} values via EPR.
- Model needs to be updated and tested further by applying various experimental techniques which could validate the method, thus giving more reliable predictions.



Acknowledgements:

This work was funded 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, the Republic of Serbia (Contract No. 451-03-9/2021-14/200042 and 451-03-9/2021-14/200053).