Structure prediction and modeling of the *lvaD* gene expression product contained in *lev* operon of *Pseudomonas putida*

Mikołaj Iwan[1], Jędrzej Kubica[2,3]

[1]Department of Environmental Microbiology and Biotechnology, Institute of Microbiology, Department of Biology, University of Warsaw

[2]Laboratory of Structural Bioinformatics, Institute of Evolutionary Biology, University of Warsaw, Poland

[3]Laboratory of Theory of Biopolymers, Faculty of Chemistry, University of Warsaw, Poland

Background

Pseudomonas putida KT2440 has been found out by **Rand et al.[1]** to contain *lev* operon, responsible for the catabolism of levulinic acid (LA). For biotechnological applications, the most interesting component of this cluster is *lvaD* gene which encodes a protein responsible for reduction of LA to 4-hydroxyvalerate (4-HV) which is important due to its use in polymer synthesis.

Predicting the structure of this protein (lvaD) could prove to be useful in furthering the understanding of the mechanism of LA reduction via this pathway. This is imporant from metabolic engineering point of view, since enzymes similar to the one encoded by *lvaD* are a promising candidate for efficient 4-HV production.

[1] Rand, J. M. *et al.* A metabolic pathway for catabolizing levulinic acid in bacteria. *Nat Microbiol* **2**, 1624–1634 (2017).

Methods

In the first step aminoacid sequence has been aligned using HHpred. Then the structure has been predicted via AlphaFold 2 followed by structure relaxation with FastRelax from PyRosetta.The structure has been predicted by RosettaFold as well.

Additionally, the electrostatic potential map of the protein surface has been prepared with SURFMAP package. Additional analyses comprised DeepTMHMM to predict where the protein is located relative to the cell membrane.

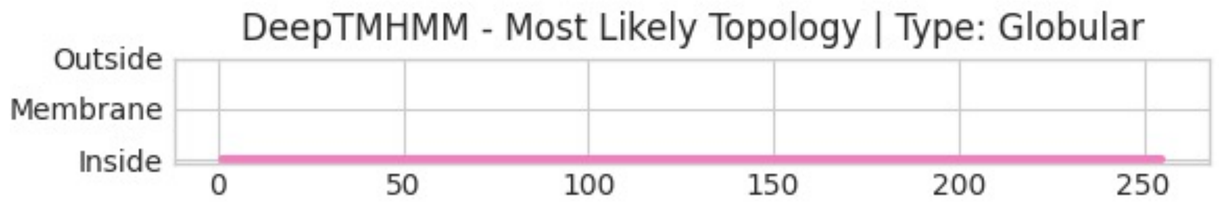

Fig. 1 DeepTMHMM prediction on topology

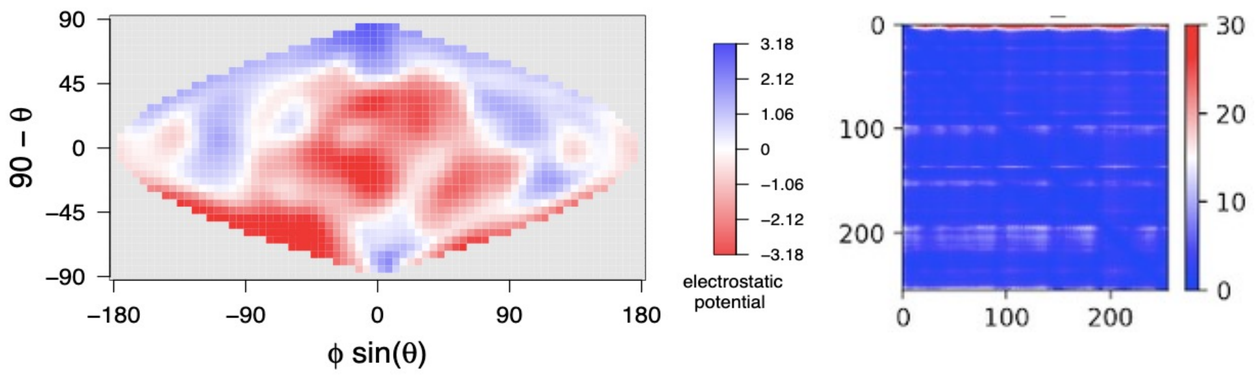

Fig. 2 Electrical potential distribution

Fig 3. PAE for AF2 model

Results

HHpred alignment showed high resemblance of query sequence to **acetoacetyl – ACP synthase found in *Mycobacterium smegmatis* [2]**. Both proteins have been shown to contain **SDR (short-chain dehydrogenase/ reductase) domain**. Especially interesting is the presence of Rossman fold charactarestic for enzymes binding $NADP^+$ and NAD^+ cofactors. **RosettaBox predicted that, with 99% certainty, that lvaD binds NADP [3].**

AF2 prediction relaxed by FastRelax showed a Rossman fold **(fig. 4.),** with high certainty **(fig. 3.)**. **Figure 2** presents results from electrical distribution on the Surface of protein.

DeepTMHMM shows that levD has globular structure and is located inside the cel with no transmembrane domains **(fig. 1)**.

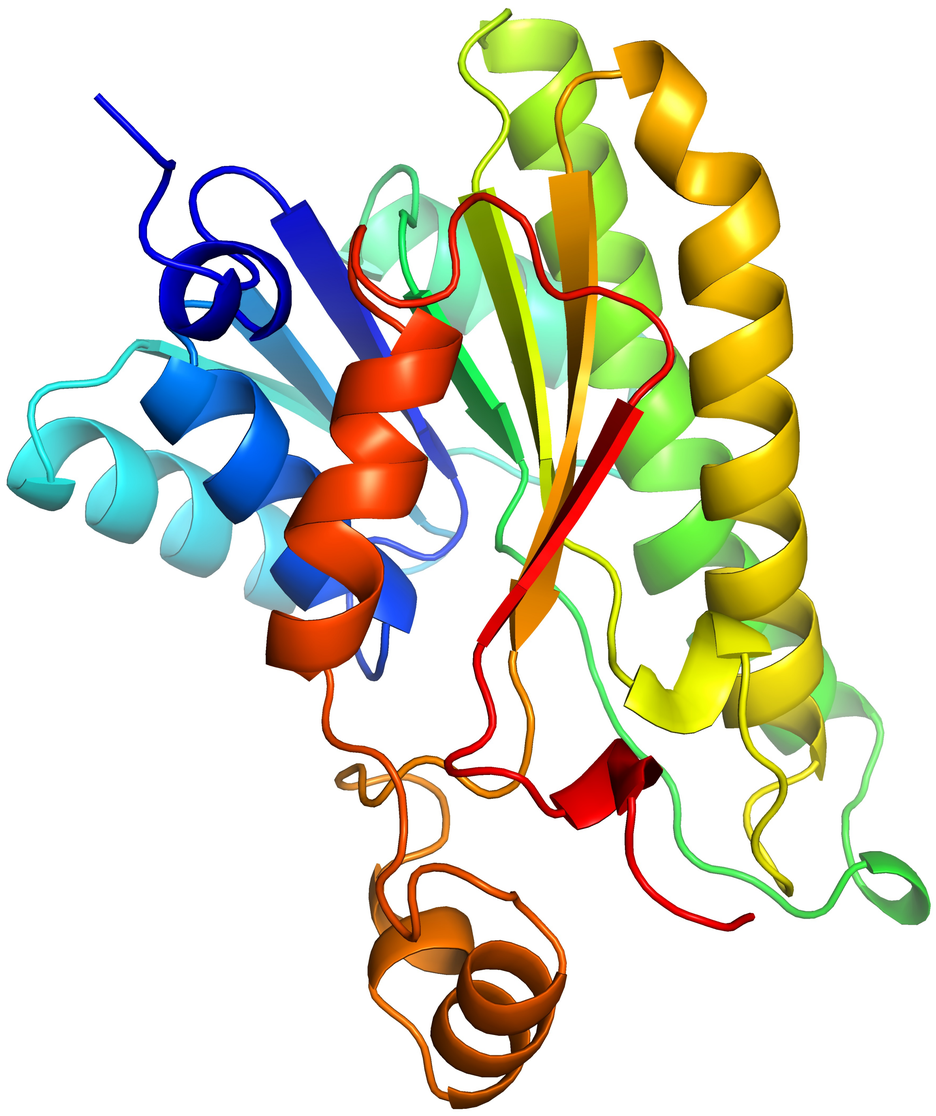

Fig. 4 Relaxed model of lvaD protein

Aminoacid sequence of levD protein

mqpnlarlfa ldgrralvtg assglgrhfa mtlaaagaev vvtarrqapl qalveaieva ggraqafald vtsredicrv ldaagpldvl vnnagvsdsq pllacddqtw dhvldtnlkg awavaqesar rmvvagkggs linvtsilas rvagavgpyl aakaglahlt ramalelarh girvnalapg yvmtdlneaf laseagdklr sripsrrfsv psdldgalll lasdagrams

[2] Lai CY, Cronan JE. Beta-ketoacyl-acyl carrier protein synthase III (FabH) is essential for bacterial fatty acid synthesis. J Biol Chem. 2003 Dec 19;278(51):51494-503. doi: 10.1074/jbc.M308638200. Epub 2003 Sep 30. PMID: 14523010.

[3] "*Rossmann-toolbox: a deep learning-based protocol for the prediction and design of cofactor specificity in Rossmann-fold proteins*" by Kamil Kaminski, Jan Ludwiczak, Maciej Jasinski, Adriana Bukala, Rafal Madaj, Krzysztof Szczepaniak, and Stanislaw Dunin-Horkawicz bioRxiv 2021.05.05.440912; doi: https://doi.org/10.1101/2021.05.05.440912