

Author: Simeon Borko^{1,2}, supervisor: Jiří Hon^{1,2}

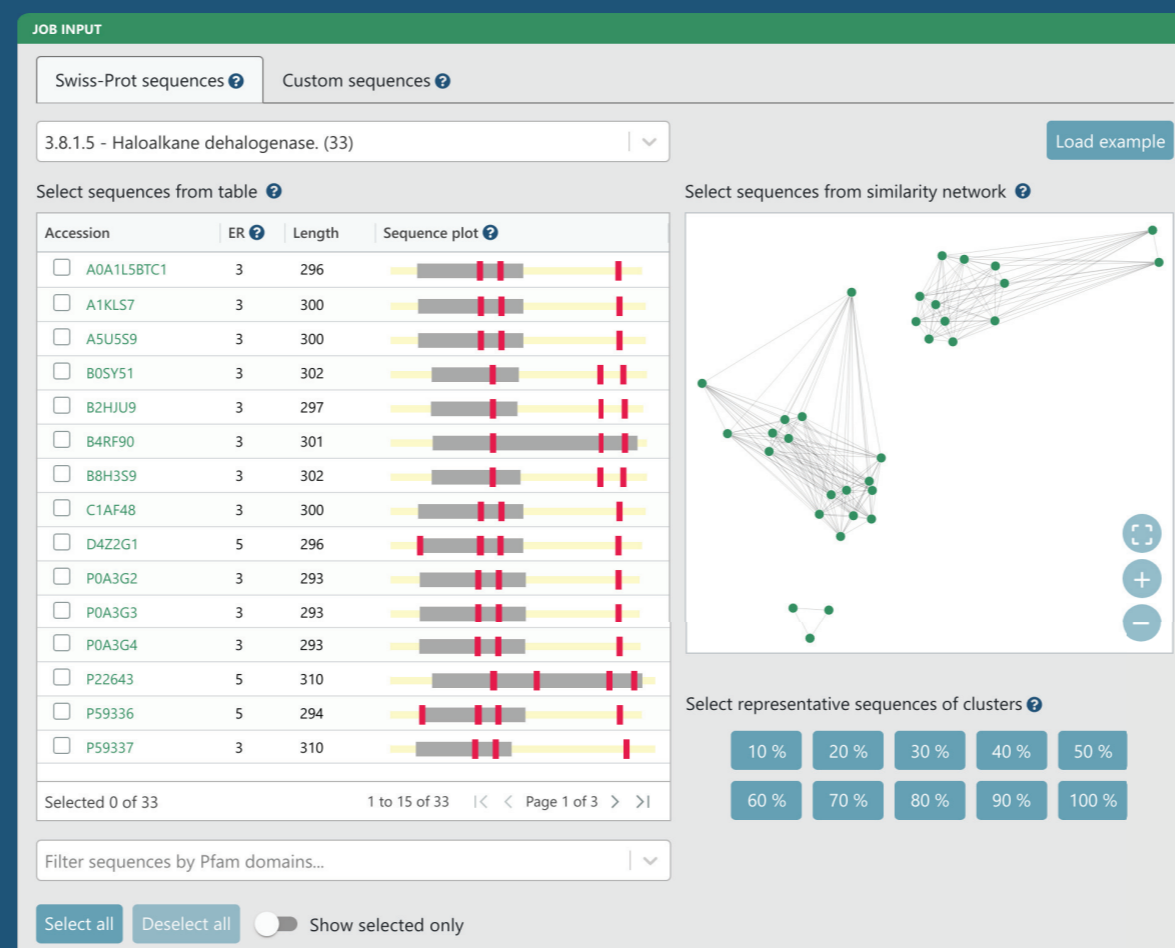
¹Department of Information Systems, Faculty of Information Technology, Brno University of Technology
²Loschmidt Laboratories, RECETOX, Department of Experimental Biology, Faculty of Science, Masaryk University

<https://loschmidt.chemi.muni.cz/enzymeminer/>



Motivation

- Enzymes are proteins catalysing chemical reactions with great applications in the ecological production of chemical compounds
- Millions of protein sequences are being discovered and uploaded to protein databases every year
- Impossible to test all new sequences in laboratory for specific enzymatic function
- EnzymeMiner finds novel enzymes from protein databases and facilitates selection of promising targets for laboratory tests



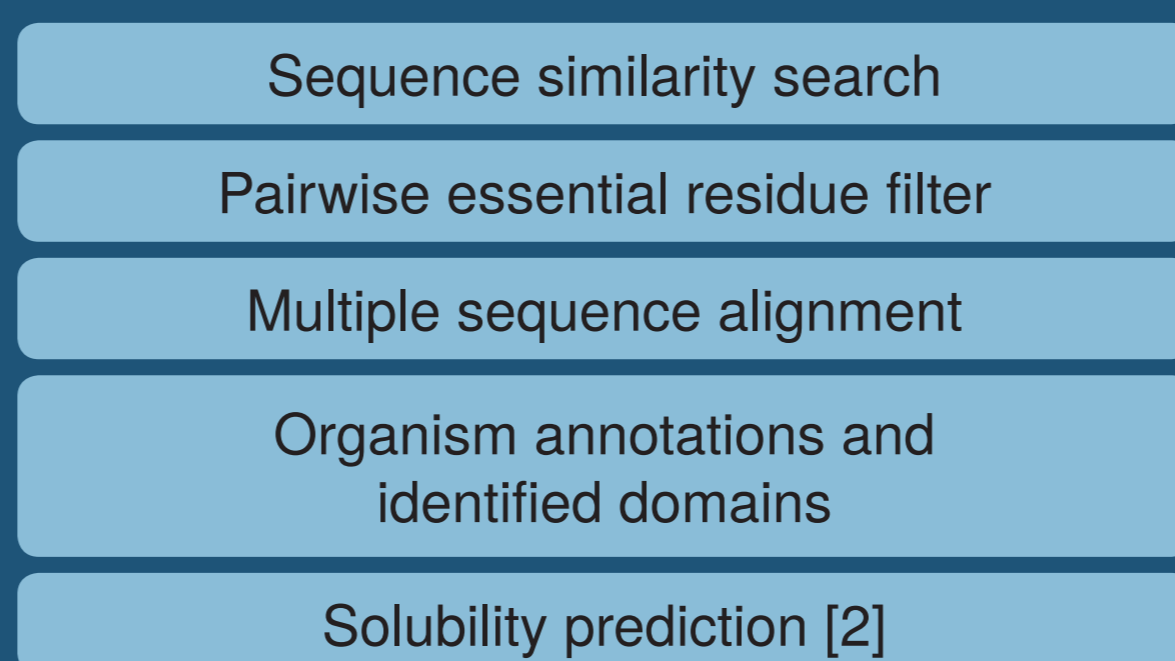
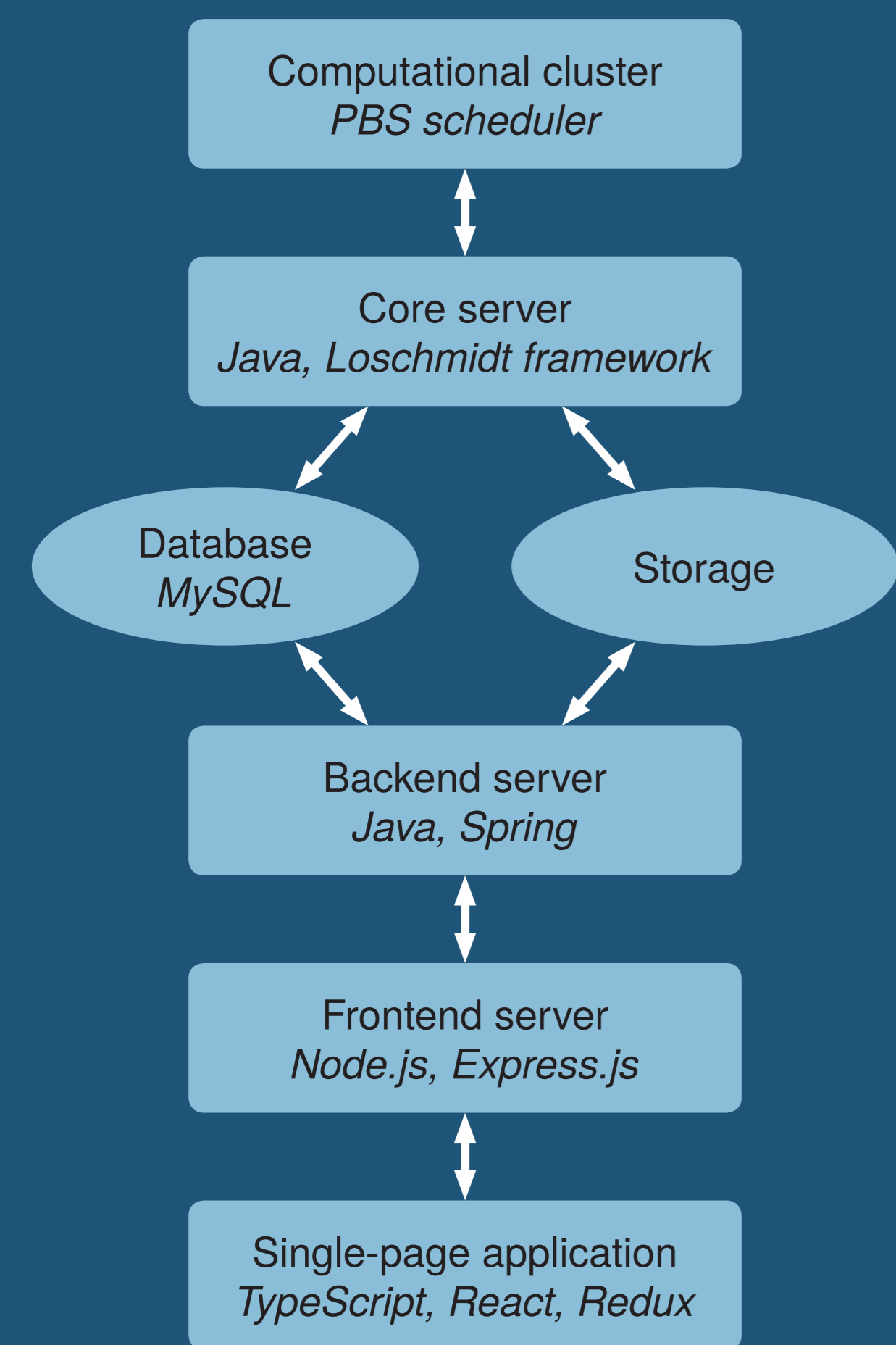
Interactive input form using EC number and SwissProt database



Conclusions

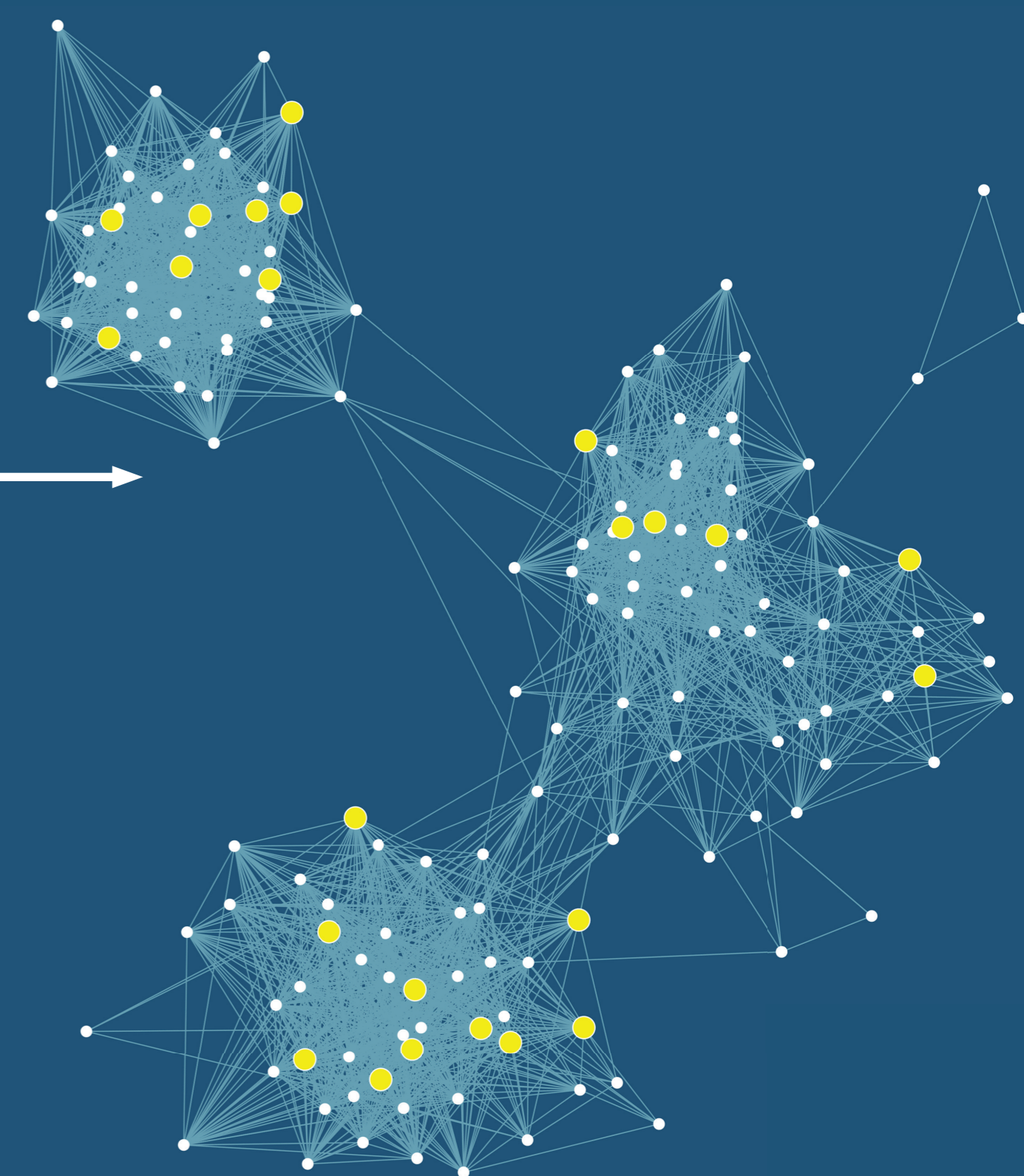
- EnzymeMiner identifies novel putative enzymes and facilitates their prioritization and selection for experimental characterization
- Previously published methodology [1] validated on 20 putative haloalkane dehalogenase genes
- Robust web server architecture using high-performance computational cluster of ELIXIR CZ
- Easy-to-use web interface with interactive components implemented using Typescript and React
- Accepted to Nucleic Acids Research journal (JIF 11.15)

Web server architecture



Selected	Full Dataset	Extra Domain	Organism	Temperature	Salinity	Biotic Relationship	Disease	Transmembrane	3D Structure	Network
<input type="checkbox"/>	3P5D.A	Chain A, Crystal Struc...	D422G1.S1	415	€	0.9715	318	Abhydri...		
<input type="checkbox"/>	3P5E.A	Chain A, Crystal Struc...	D422G1.S1	415	€	0.9615	310	Abhydri...		
<input type="checkbox"/>	3P5F.A	Chain A, Crystal Struc...	D422G1.S1	415	€	0.9614	319	Abhydri...		
<input type="checkbox"/>	3P5H.A	Chain A, Crystal Struc...	D422G1.S1	412	€	0.9586	311	Abhydri...		
<input type="checkbox"/>	NP_017515177.1	haloalkane dehalogen...	D422G1.S1	708	€	0.9399	270	Abhydri...		
<input type="checkbox"/>	3S8D.A	Chain A, structure of ...	D422G1.S1	462	€	0.9393	311	Abhydri...		
<input type="checkbox"/>	4BZ2.A	Chain A, Haloalkane D...	D422G1.S1	617	€	0.9357	290	Abhydri...		
<input type="checkbox"/>	AC54806.1	haloalkane dehalogen...	D422G1.S1	613	€	0.925	282	Abhydri...		
<input type="checkbox"/>	NP_003023101.1	haloalkane dehalogen...	D422G1.S1	693	€	0.9219	300	Abhydri...		
<input type="checkbox"/>	4C5H.A	Chain A, Haloalkane D...	D422G1.S1	615	€	0.9202	291	Abhydri...		
<input type="checkbox"/>	3Z9B.A	Chain A, Crystal Struc...	D422G1.S1	695	€	0.9187	297	Abhydri...		
<input type="checkbox"/>	3Z9H.A	MULTISPECIES haloa...	D422G1.S1	693	€	0.917	300	Abhydri...		
<input type="checkbox"/>	NP_017342388.1	haloalkane dehalogen...	D422G1.S1	689	€	0.9162	300	Abhydri...		
<input type="checkbox"/>	CM885140.1	haloalkane dehalogen...	D422G1.S1	679	€	0.9141	275	Abhydri...		
<input type="checkbox"/>	VCL.A	Chain A, PROTEIN (HA...	P23643.2	997	€	0.9136	310	Abhydri...		
<input type="checkbox"/>	NP_053370782.1	haloalkane dehalogen...	D422G1.S1	693	€	0.9134	300	Abhydri...		
<input type="checkbox"/>	348R.A	Chain A, Crystal Struc...	D422G1.S1	661	€	0.911	304	Abhydri...		

Interactive selection table with multiple views and filtering options



Interactive sequence similarity networks in input form and results page

2013
 First version of EnzymeMiner pipeline specialized for haloalkane dehalogenase (HLD) enzyme family.

2017
 First 20 putative HLDs selected for experimental characterization using conventional methods.

2018
 Publication of EnzymeMimer pipeline for HLDs and results of experimental characterization of first 20 selected HLDs.

2018
 Next 47 putative HLDs for experimental characterization using high-throughput methods.

2020
 Publication of EnzymeMiner web server and results of experimental characterization of next 47 selected HLDs.

2020

2021

2021
 Support for homology modelling and analysis of pockets and tunnels. Validation using other enzyme families.

References & acknowledgements

- Vanacek et al. 2018, Exploration of Enzyme Diversity by Integrating Bioinformatics with Expression Analysis and Biochemical Characterization. ACS Catalysis 8: 2402–2412
- Hon et al. 2019, SoluProt: Tool for Prediction of Protein Solubility. Bioinformatics (in preparation)



This work was supported by ELIXIR CZ research infrastructure project (MEYS Grant No: LM2015047) including access to computing and storage facilities. Computational resources were supplied by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures".