New Concepts and Software Tools for Rational Design of Enzymes

Jiri Damborsky, Jaroslav Bendl, Jan Stourac, Eva Sebestova, Jan Brezovsky

Loschmidt Laboratories, Department of Experimental Biology and Research Centre for Toxic Compounds in the Environment, Faculty of Science, Masaryk University, Czech Republic

Enzymes are biological catalysts accelerating chemical reactions in living organisms. Protein engineering focuses on construction of enzymes with improved activities, specificities, and stabilities for biotechnological applications. Protein engineering greatly benefits from computational tools that provide guidance for identification of structural elements important for protein function and design of specific mutations [1] or smart mutant libraries [2].

We have released a new version of **CAVER** (http://www.caver.cz/), the software tool for analysis of tunnels and channels in biomolecules [3-4]. The tunnels and channels are important structural features serving for transport of small molecules, ions and water solvent through the protein structures. The mutagenesis targeting protein tunnels provides enzymes with improved activity [5], enantioselectivity [6] and stability [7-8]. New Caver enables analysis of ensembles from molecular dynamic simulations (**Fig. 1**), which is important for integration of dynamics into protein design [9]. Mobile elements called molecular gates are important components of protein tunnels and channels [10], providing specificity and selectivity. Molecular gates add another level of sophistication to rational design of tunnels and channels and this task very challenging.

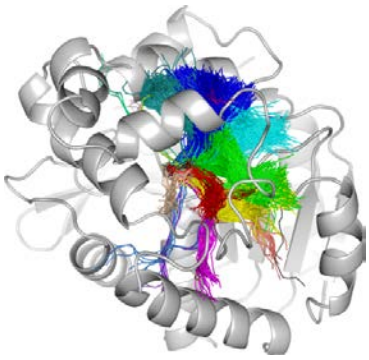

Fig.1. Analysis and visualization of tunnels in dynamic protein structure of haloalkane dehalogenase by Caver 3.0 [3] and Caver Analyst 1.0 [4].

HOTSPOT WIZARD (http://loschmidt.chemi.muni.cz/hotspotwizard/) predicts hot spots for site-directed mutagenesis or focused directed evolution [11]. Identification of suitable targets for mutagenesis experiments is based on combination of structural, functional and evolutionary information obtained from bioinformatics databases and computational tools (**Fig. 2**). Hotspot Wizard lists residues ordered by predicted suitability for mutagenesis, together with information about their conservation level and potential function. Although the current version of HotSpot Wizard has become quite popular with more than 4,000 analyzed structures since its launch in 2009, it currently has two major limitations: (i) the three-dimensional structure of analyzed protein must be known and (ii) the tool identifies positions, but does not predict substitutions. The new version of HotSpot Wizard will address these shortages by integration of structure prediction using homology modelling and estimation of the effect of substitutions on protein function. All databases and computational tools will also be updated. New interactive and intuitive user interface will offer an easy way to perform several structural and evolutionary analyses at once with minimal demands on users, making HotSpot Wizard potentially useful for experimentalists with no prior knowledge of computer modelling or rational protein design.

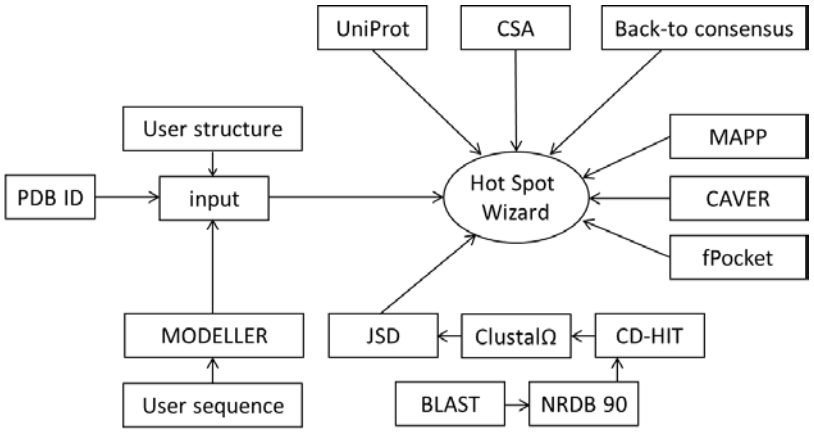

Fig. 2. Workflow of HotSpot Wizard 2.0 [11].

The software tool **PREDICTSNP** (http://loschmidt.chemi.muni.cz/predictsnp/) provides robust and accurate prediction of the effect of mutations on protein function [12]. We have used 43,000 mutations of known effect from scientific literature and patents for the unbiased evaluation of eight established prediction tools: MAPP, nsSNPAnalyzer, PANTHER, PhD-SNP, PolyPhen-1, PolyPhen-2, SIFT and SNAP. The six best performing tools were combined into a consensus classifier PredictSNP, resulting into significantly increased prediction performance and improved robustness (**Fig. 3**). A user-friendly web interface enables easy access to all eight prediction tools, the consensus classifier PredictSNP and annotations from the Protein Mutant Database and the UniProt knowledgebase. The web server and the datasets are freely available to the academic and medical community.

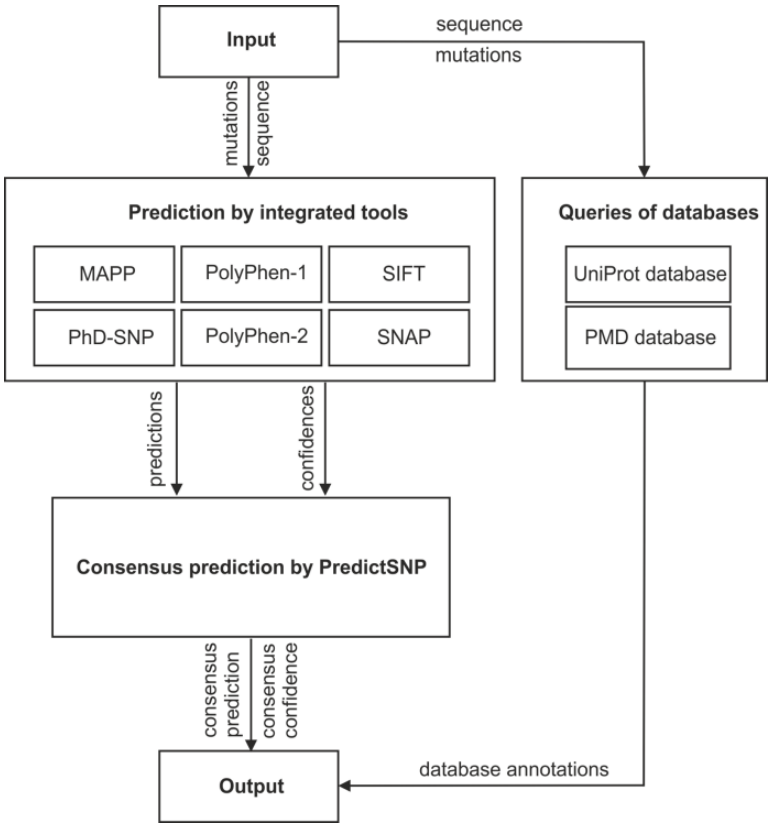

Fig 3. Workflow of PredictSNP 1.0 [12].

[1] Damborsky, J. and Brezovsky, J., 2014: Computational Tools for Designing and Engineering Enzymes. *Current Opinion in Chemical Biology* 19: 8-16.

[2] Sebestova, E., et al. 2014: Computational Tools for Designing Smart Libraries. In. Gillam, E.M.J., Copp, J.N., Ackerley, D.F., (Eds.), Directed Evolution Library Creation: Methods and Protocols, *Methods in Molecular Biology*, 2nd edition, Humana Press, New York, pp. 291-314.

[3] Chovancova, E., et al. 2012: CAVER 3.0: A Tool for Analysis of Transport Pathways in Dynamic Protein Structures. *PLOS Computational Biology* 8: e1002708.

[4] Kozlikova, B., et al. 2014: CAVER Analyst 1.0: Graphic Tool for Interactive Visualization and Analysis of Tunnels and Channels in Protein Structures. *Bioinformatics* 30: 2684-2685.

[5] Pavlova, M., et al. 2009: Redesigning Dehalogenase Access Tunnels as a Strategy for Degrading an Anthropogenic Substrate. *Nature Chemical Biology* 5: 727 - 733.

[6] Prokop, Z., et al. 2010: Enantioselectivity of Haloalkane Dehalogenases and its Modulation by Surface Loop Engineering. *Angewandte Chemie* 49: 6111-6115.

[7] Koudelakova, T., et al. 2013: Engineering Enzyme Stability and Resistance to an Organic Cosolvent by Modification of Residues in the Access Tunnel. *Angewandte Chemie* 52: 1959-1963.

[8] Damborsky, J., et al. 2013: Method of Thermostabilization of a Protein and/or Stabilization Towards Organic Solvents. Patent US 8,580,932.

[9] Sykora, J., et al. 2014: Dynamics and Hydration Explain Failed Functional Transformation in Dehalogenase Design. *Nature Chemical Biology* 10: 428-430.

[10] Gora, A., et al. 2013: Gates of Enzymes. *Chemical Reviews* 113: 5871–5923.

[11] Pavelka, A., et al. 2009: HotSpot Wizard: A Web Server for Identification of Hot Spots in Protein Engineering. *Nucleic Acids Research* 37: W376–W383.

[12] Bendl J., et al. 2014: PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations. *PLOS Computational Biology* 10: e1003440.