Zebra: web-server for bioinformatic analysis of large protein superfamilies to identify variable amino acid residues responsible for functional diversity

Dmitry Suplatov, Evgeny Kirilin, Vakil Takhaveev, Vytas Švedas

Lomonosov Moscow State University, Belozersky Institute of Physicochemical Biology and Faculty of Bioengineering and Bioinformatics, Vorobjev hills, 1-40, 119991, Moscow, Russia
vytas@belozersky.msu.ru

A new algorithm Zebra and a corresponding web-server have been developed to systematically study diverse protein superfamilies and identify the subfamily-specific positions (SSPs) – conserved only within functional subfamilies but different between them – that seem to be responsible for different substrate specificity, catalytic activity, stability, etc. [1]. It is known from experimental enzymology that mutations in the active site can change enantioselectivity, substrate specificity and catalytic promiscuity more effectively than distant ones. However, both close and distant mutations can be important for activity and stability thus highlighting complexity of evolutionary adaptation. Therefore, to identify functionally important SSPs a novel scoring function is suggested that incorporates structural information as well as physicochemical and residue conservation in protein subfamilies. The algorithm does not require pre-defined subfamilies and can propose multiple classifications automatically by graph based clustering at different fragmentation levels. Consequently, Zebra is the first application that provides specificity determinants at different levels of functional classification therefore addressing complex functional diversity of large superfamilies (Fig. 1A). Random shuffling and Bernoulli statistics are applied to rank hits by decreased significance and select highly valuable SSPs for further evaluation. Zebra results are provided in two ways – as a single all-in-one parsable text file and PyMol sessions with structural representation of SSPs (Fig. 1B).

Zebra has been applied to study how lipase and amidase catalytic activities are implemented into the alpha-beta hydrolase fold. Subfamily-specific positions of α/β-hydrolases with lipase and protease activities were identified and used as hotspots to introduce amidase activity into *Candida antarctica* lipase B (CALB). Molecular modeling was used to evaluate influence of selected residues on binding and catalytic conversion of amide substrate by corresponding *in silico* library of mutants and to select reactive enzyme-

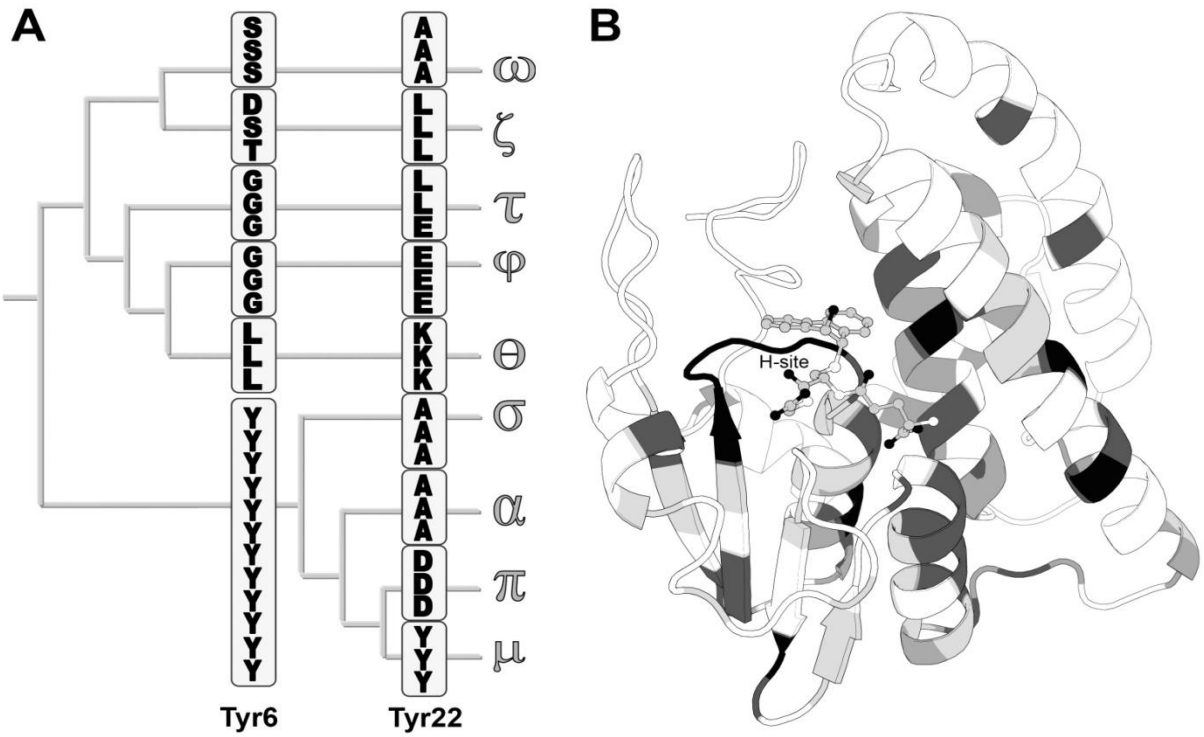

Fig. 1. A. Subfamily-specific positions at different levels of functional classification predicted by Zebra web-server in the glutathione S-transferase family. Greek letters represent different functional classes within the family. Position numbering as in the μ class enzyme from rat (PDB 2GST). **B.** Structural representation of subfamily-specific positions in the glutathione S-transferase family (automatically produced by Zebra web-server). Gradient paint corresponds to estimated specificity: black stands for highly significant hits, white – for non-specific positions. The product of enzymatic reaction is shown in ball-and-sticks to highlight the active site. The most significant hits are located in the H-site loop which contains catalytically important residues for binding structurally diverse xenobiotic substances, and also in surrounding regions – domain-domain, subunit-subunit and possibly dimer-dimer interfaces that provide shape and flexibility to the active site for complementarity with the substrate.

substrate complexes that satisfy knowledge-based criteria of amidase catalytic activity. Selected CALB variants were produced and showed significant improvement of experimentally measured amidase activity [2].

Zebra can be used to explore structure-functional relationship in enzymes and to select highly significant SSPs as hotspots for directed evolution or rational design experiments. Web-server, tutorials and examples are available at http://biokinet.belozersky.msu.ru/zebra.

Acknowledgement: this work was supported by the Russian Foundation for Basic Research (RFBR) grant 12-04-91335-ННИО_а and German Research Foundation (DFG) as a joint Russian-German project "RECESS".

[1] D. Suplatov *et al.* (2013). *J Biomol Struct Dyn*, doi: 10.1080/07391102.2012.750249.
[2] D. Suplatov *et al.* (2012). *Protein Eng Des Sel*, 25(11), 689-697.