Bioinformatic analysis of subfamily-specific positions reveals a previously unknown regulatory site in Glyceraldehyde 3-phosphate dehydrogenase family of proteins

Dmitry Suplatov, Elena Schmalhausen, Vladimir Muronets, Vytas Švedas

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

At the MCCMB'15 we presented the complex approach that combines methods of bioinformatics, molecular modeling and theoretical chemistry to study protein families, design enzymes with improved properties and search for new binding sites in enzyme structures [1]. This strategy is based on identification of subfamily-specific positions in protein structures which can be then used at studying functional diversity within large superfamilies, as hotspots at rational engineering of protein properties as well as a key criterion at selection of functionally important regulatory centers [2]. In the last two years the original approach was further developed and evaluated. Its adaptation to the supercomputer-based calculations provided an opportunity to boost the performance and make the procedures more accurate [3]. The upgraded approach has been applied to search for new mechanisms to regulate functional properties of Glyceraldehyde 3-phosphate dehydrogenases (GAPHDs). The bioinformatic analysis algorithm pocketZebra [4] has been used to identify a novel binding site that is topologically independent from the catalytic site in GAPHDs' structures and characterized by a high content of subfamily-specific positions. Screening of GAPHD from *Mycobacterium smegmatis* by a large *in silico* library of low molecular weight compounds has been carried out to identify selective modulators of the bacterial GAPHD catalytic activity. To improve the selection of the most promising specific modulators we have taken into account the interaction of candidate compounds with subfamily-specific amino acid residues located in the target binding site. Thus the presence of subfamily-specific positions previously considered as a key criterion at identifying and ranking functional and regulatory centers in protein structures [4], has been extended to the stage of selecting the most promising selective modulators. Inhibitory properties of in silico selected low molecular

weight compounds were experimentally evaluated with human GAPHD and bacterial enzyme from *Mycobacterium smegmatis*. Three inhibitors were shown to selectively suppress activity of the bacterial GAPHD but not of the human homolog. Molecular modeling has been implemented to study the detailed mechanism of selective ligand binding to the discovered site in GAPHDs. This study has shown that bioinformatic analysis can be used to disclose novel ways to regulate protein function by identifying and studying previously unknown binding sites.

This work was supported by the Russian Science Foundation grant #15-14-00069. The supercomputer complex of the Lomonosov Moscow State University was used for resource-intensive computations. We thank the Vitas-M Laboratory for providing the requested ligand samples from their collection for the experimental evaluation.

1. D. Suplatov, N. Panin, T. Shcherbakova, I. Khaliullin, V. Švedas (2015) Bioinformatic analysis of diverse protein superfamilies to design improved enzymes. *MCCMB'15*, July 16-19, 2015, Moscow, Russia

2. D. Suplatov, E. Kirilin, V. Švedas (2016) Bioinformatic analysis of protein families to select function-related variable positions, In: *Understanding Enzymes: Function, Design, Engineering and Analysis,* Allan Svendsen et al., 351-385 (Pan Stanford Publishing).

3. D. Suplatov, N. Popova, S. Zhumatiy, V. Voevodin, V. Švedas (2016). Parallel workflow manager for non-parallel bioinformatic applications to solve large-scale biological problems on a supercomputer. *J. Bioinform. Comput. Biol.*, **14(02)**: 1641008.

4. D. Suplatov, E. Kirilin, M. Arbatsky, V. Takhaveev, V. Švedas (2014). pocketZebra: a web-server for automated selection and classification of subfamily-specific binding sites by bioinformatic analysis of diverse protein families. *Nucl. Acids Res.*, **42(W1)**: W344-W349.