High-performance computing in bioinformatic analysis of protein superfamilies to design enzymes with new properties*

D. Suplatov[1], N. Popova[2], K. Kopylov[3], M. Shegai[2], V. Voevodin[4], V. Švedas[1]

Lomonosov Moscow State University, Faculty of Bioengineering and Bioinformatics and Belozersky Institute of Physicochemical Biology[1], Faculty of Computational Mathematics and Cybernetics [2], Faculty of Chemistry [3], Research Computing Center [4]

Growing capacity of bioinformatic databases provides new opportunities to study structure-function relationship in large protein superfamilies and greatly increases the demand for high performance computing. However, the general-purpose parallel computing clusters do not provide optimal accommodation to bioinformatic applications which are usually written using openMP rather than MPI, Java or even Perl and Python. Distributed computing platforms are available as an inexpensive alternative but they lack the power of a dedicated computing cluster. Based on the Lomonosov Moscow State University supercomputer complex we are developing a platform which implements computational methods of bioinformatic analysis, molecular modeling and computational chemistry to study the structure-function relationship in large enzyme superfamilies and produce novel biocatalysts with improved properties. Co-design of cluster's hardware and software according to demands of computational biology will provide a solution for large-scale tasks of biocatalysis.

Homologous enzymes have evolved from a common ancestor to retain a general function but diverged in more specific features and can be divided into subfamilies with different functional properties such as catalytic activity, substrate specificity, enantioselectivity, stability, etc. Analysis of sequence and structural information in protein superfamilies is a promising trend in order to rationalize enzyme engineering and move away from unguided evolutionary stochastic approaches and empirical design [1]. We have recently developed a new method of bioinformatic analysis to identify function-related variable residues in protein structures that are responsible for functional divergence within superfamilies of homologous enzymes [2]. The developed methodology has been applied to study structure-functional relationship in various enzyme superfamilies: α/β-hydrolases, Ntn-hydrolases, penicillin-binding proteins, etc. Systematic bioinformatic analysis of genomic and structural information corresponding to each selected superfamily of enzymes has been carried out to identify functionally important amino acid residues as hotspots for enzyme engineering. It has been shown that bioinformatic analysis can be effectively used to design enzyme mutants with improved catalytic properties and to predict functional properties of enzymes [3, 4]. There is a need to implement these computationally demanding algorithms into the common laboratory practice to study the structure-function relationship in proteins and develop novel protein engineering strategies.

References

1. Suplatov D., Voevodin V., Švedas V. Robust enzyme design: Bioinformatic tools for improved protein stability //Biotechnol. J. – 2015. – Vol. 10. – No. 3. – P. 344-355.
2. Suplatov D., Kirilin E., Arbatsky M., Takhaveev V., Švedas V. pocketZebra: a web-server for automated selection and classification of subfamily-specific binding sites by bioinformatic analysis of diverse protein families //Nucl. Acids Res. – 2014. – Vol. 42. – No. W1 – P. W344-W349.
3. Suplatov D., Besenmatter W., Švedas V., Svendsen A. Bioinformatic analysis of alpha/beta-hydrolase fold enzymes reveals subfamily-specific positions responsible for discrimination of amidase and lipase activities //Protein Eng. Des. Sel. – 2012. – Vol. 25. – No. 11 – P. 689-697.
4. Suplatov D., Panin N., Kirilin E., Shcherbakova T., Kudryavtsev P., Švedas V. Computational Design of a pH Stable Enzyme: Understanding Molecular Mechanism of Penicillin Acylase's Adaptation to Alkaline Conditions //PloS one. – 2014. – Vol. 9. – No. 6. – P. e100643.

* This work was supported by Russian Foundation for Basic Research grant No. 14-07-00437