GPU-accelerated analysis of linear trends in a protein motion*

D. Suplatov[1], K. Kopylov[2], V. Švedas[1]

Lomonosov Moscow State University, Faculty of bioengineering and bioinformatics and Belozersky Institute of Physicochemical Biology[1] and Faculty of chemistry[2]

Computational structure analysis reveals that proteins and enzymes along with previously annotated functional sites (active sites) possess a significant amount of previously unexplored potential binding pockets on their surfaces [1, 2]. This comes along with the recently made suggestion that allostery – regulation of protein function by binding of low-molecular weight compounds in topologically independent regulatory sites – may be an intrinsic property of virtually all proteins [3]. It is generally accepted that conformational changes in a protein structure induced by a binding of a ligand to an allosteric center eventually leads to a change of its functional properties. Nevertheless, little is known about the particular molecular mechanisms of this phenomenon. Structural changes that occur as a result of ligand binding can be studied by molecular dynamics (MD). Modern supercomputers provide an opportunity to run multiple long time scale simulations in order to improve sampling and collect more information for statistical comparison, however novel approaches are needed to analyze this vast amount of data.

A new algorithm has been developed to reveal induced conformational changes by comparing the MD trajectories of a free (unbound) protein versus protein in a complex with a regulatory ligand. To compare mobility of structural elements in two molecular systems we determine the basic linear component (the trend in protein deviation from to the initial conformation) in the root-mean-square deviation that demonstrates dependence of the molecular displacement of protein backbone atoms from the initial reference position in time. Representation of protein motion in terms of linear trends provides an opportunity to process in a reasonable time a vast amount of data extracted from different MD runs. The output of the algorithm is a list of residues whose structural behavior is induced by a binding of a ligand. These motions can be ranked by their statistical significance and atomic displacement. The proposed method can be applied to study the structure-function-regulation relationship and can help in designing new selective allosteric effectors/inhibitors.

We have shown that analysis of linear trends in protein dynamics is a very useful tool to identify particular structural regions and amino acid residues with different mobility and prepare enzyme variants with improved properties [4, 5]. On the next step an evaluation on a large data sample will be performed to quantitatively assess the efficiency of our approach in detecting ligand-induced conformational changes in protein structures. A non-redundant set of high-quality structures corresponding to protein complexes with known regulatory ligands has been collected from the Protein Data Bank. CUDA version of the AMBER14 molecular dynamics engine will be used to perform MD simulations which will be further analyzed using our method and the results assessed by the ROC and PR characteristics. Our tests indicate that Tesla K40s GPUs installed on the “Lomonosov-2” supercomputer provide a very significant acceleration to the MD, allowing dozens of nanoseconds to be simulated by one unit per day for a 70'000-atom molecular system. This opens an opportunity to collect a large amount of MD data for further analysis on a supercomputer [6, 7] in order to improve efficacy of the suggested approach.

References

1. Suplatov D., Švedas V. Study of Functional and Allosteric Sites in Protein Superfamilies // Acta Naturae. 2015. Vol. 7. No. 4. P. 34.
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.

* This work was supported by Russian Foundation for Basic Research grant No. 16-34-01157

3. Gunasekaran K., Ma B., Nussinov R. Is allostery an intrinsic property of all dynamic proteins? // Proteins. 2004. Vol. 57. No. 3. P. 433-443.
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.
5. 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.
6. Suplatov, D., Popova, N., Zhumatiy, S., Voevodin, V., Švedas, V. Parallel workflow manager for non-parallel bioinformatic applications to solve large-scale biological problems on a supercomputer // J. Bioinform. Comput. Biol. 2016. Vol. 14. No. 2. P. 1641008
7. Suplatov, D., Popova, N., Kopylov, K., Shegai, M., Voevodin, V., Švedas, V. High-performance computing in bioinformatic analysis of protein superfamilies to design enzymes with new properties. // Russian Supercomputing Days 2015, 2015. P. 537.