HIGHLIGHT: Modeling of Protein Inner-Conformation Dynamics by Solving Mass Transportation Problem

Sergey Knyazev[1], Gaik Tamazian[2], Eugene Stepanov[3] and Yuri Porozov[1]

1 ITMO University, St. Petersburg, Russia
2 St. Petersburg State University, St. Petersburg, Russia
3 St. Petersburg Branch of the Steklov Mathematical Institute of the Russian Academy of Sciences, St. Petersburg, Russia

Abstract The understanding of a process of how a protein changes its own conformation from one known state to another is one of the most important problems in structural biology. It is because, a conformational flexibility allows proteins to implement its functionality, to regulate its interaction ability and chemical activity. To address this issue, there are a number of methods proposed to reveal protein dynamics, but there is still no universal method to investigate the protein dynamics with decent accuracy for all possible movements. For example, equipments which can observe the motion during the physical experiment, unfortunately, have small time resolution, and therefore, they can't represent the full picture of protein conformational movements. To overcome this problem, several mathematical methods were proposed. Some of them are based on an estimation of inner energy of proteins, and they can model with excellence but only short-scaled conformational changes. Another math methods are based on so called coarse-grain paradigm and capable to predict long-scale motions but with mediocre accuracy. So there is still a lack in new methods to model long-lasted conformational changes. There is a try to overcome the problem, and in this work, we have proposed one of our methods, which is intended to model big protein motions. Our approach is based on a solution of the mass transportation problem, which is well known as the Monge-Kantorovich problem. The simple idea of the method is to find a conformational motion of a protein which will transport masses of its atoms by as the shortest paths as possible. We have developed a special model and a functional by which we can estimates cost of a conformational motion. Our model is specially designed to prevent a violation of chemical properties of molecules during conformational changes. This allowed us to implement algorithms for finding an physically possible transformation with minimal transportation cost. We tested our methods on several examples, and verified adequacy of modeling results. To say more, several possible motions of proteins of interest were discovered, and provided a cue for biologists about internal chemical processes of the proteins. Based on our experiments, we believe that our model can be applied in practice and can have a good potential for predicting of long-time motions of protein conformations.

Keywords: protein dynamics, protein morphing, coarse-grained model, conformational motion

1 Introduction

Proteins play an essential role in biochemical processes of living organisms. They are macromolecules which are characterized by an amino-acid sequence as well as by a three-dimensional structure. And there is a generally recognized fact, that the structure and its flexibility influence proteins functional activity. Furthermore, proteins functional properties can be drastically changed during a transformation of a protein from one conformation to another [3].

So there is an solid interest in an investigating and a modeling of molecular dynamics, and, as consequences, there have appeared several tools for this purpose recently (for example see [2, 9, 12]). All of the methods are trying to extend knowledge about conformational mobility based on data from physical experiments. While methods of physical experiments, such as X-ray and NMR, can reveal only a limited set of possible protein conformations, the computational methods try to reveal a connection between these states. Most of the classical modeling methods are based on analysing of inner energy of conformations (for instance, Molecular Dynamics [2] or Elastic Network Models [9]) which are good methods for short-time movements (in range about 10^{-15} – $10^{-12}s$) with a small amplitude (about 0.2Å). Meanwhile, when we are interested in modeling of conformational functional mobility or protein domain motion, which can last for milliseconds, the classical methods reveal huge limitations and new methods are desired.

New approach for modeling of long scale motions, which is based on a rough approximation of a protein conformation and its movement, a new one of so called coarse-grained methods, is proposed here. The method was initially proposed in our laboratory [12] and series of extensions were implemented in [4,10,11,16] and some results are still unpublished yet.

2 Methods

Protein conformational model

Proteins consist of amino-acids, which are connected to each-other and form by this one whole and flexible chain. Such a chain consists of two main elements: a backbone and side chains. While the backbone forms a protein conformation or shape, the side chains are responsible for protein surface properties. The main chain consists of consecutively connected atoms $N^1 - C^1_\alpha - C^1 - \cdots - N^i - C^i_\alpha - C^i - \cdots - N^m - C^m_\alpha - C^m$, and the side chains are connected to C^i_α as essential part of i-th amino-acid.

Let us have a vector of Cartesian coordinates for atoms in main chain $\mathbf{x} = (x_1, \ldots, x_n)$, where x_1 represents N^1 atom, x_2 for C^1_α, …, x_n for C^m, and $n = 3m$.

So a backbone model looks like tuple of vectors:

$$C := \left\{ (r_i)_{i=1}^{n-1}, (\alpha_i)_{i=1}^{n-2}, (\chi)_{i=1}^{n-3} \right\} \tag{1}$$

where r_i is a covalent bonds length, α_i is a values of valence angels, χ_i is a values of torsion angles ϕ, ψ and ω.

$$\begin{aligned} r_i &= |\Delta x_i| \\ \alpha_i &= \arccos\left(\frac{\Delta x_i \cdot \Delta x_{i+1}}{r_i r_{i+1}}\right) \\ \chi_i &= \operatorname{atan2}\left(|\Delta x_{i+1}|\, \Delta x_i \cdot N_{i+1}, N_i \cdot N_{i+1}\right) \end{aligned} \tag{2}$$

where $|\cdot|$ denotes an Euclidean norm, $\Delta x_i = x_{i+1} - x_i$ and $N_i = \Delta x_i \times \Delta x_{i+1}$.

Protein transformation model

A transformation is a process of a protein motion from one configuration to another. Let us consider a backbone of a protein molecule which is consisting of n residues. A Lipschitz continuous function $\gamma : [0, 1] \to R^{3n}$, such that $\gamma(t)$ is an admissible conformation of the molecule for each $t \in [0, 1]$, is called as an admissible movement of the molecule. Note that an admissible movement changes neither covalent bonds lengths nor valence angles.

Let us have two conformations x_0 x_1 of the same protein. Our goal is to find an acceptable movement between these two conformations. We propose to do it by finding a minimum of a special functional among all admissible movements γ within a set of all movement T between given ones. So we introduce two functionals of an admissible movement γ: F_p and G_p , where p is the parameter so that $p \geq 1$.

$$F_p(T, \gamma) = \sum_{j=1}^{n} m_j l_j^p \tag{3}$$

where l_j is a total length of a path covered by the jth and m_j is an atomic mass of the jth atom in a backbone. For an alpha carbon atom, a total mass of atoms in the side-chain, which is connected to it, is also added to the corresponding m_j value.

$$G_p(T, \gamma) = \int_0^1 m_j |\dot{\gamma}|^p dt \tag{4}$$

Both functions have a quite similar meaning of a cost of a protein atom mass transportation between given conformations, so that minimizers of these functions tend to move less the “heavier” parts of a molecule. Both cost functions are quite similar to the Kantorovich-Wasserstein distances between two measures corresponding to the Monge-Kantorovich optimal mass transportation problem [6]. Roughly speaking, the difference between G_p and F_p is the difference between a

dynamic and a static optimal transportation problem formulations, respectively [1, 5].

In order to implement the above model numerically, we represent a motion of a protein as set of s intermediate conformations with given initial and final conformations.

The cost functional can be simplified to:

$$G_1(\tilde{\gamma}, T) = \sum_{i=1}^{n} m_j \sum_{i=0}^{M+1} |(x_{i+1})_j - (x_i)_j|^p \tag{5}$$

where $|\cdot|$ stands for the Euclidean norm in R^3 and $(x_i)_j$ denotes the jth atom of the conformation x_i.

3 Results

In accordance with our experiments [4,10–12,16], described principle often gives similar to other coarse-grained methods results. Nevertheless, our methods gives a slightly more accurate results than any other tested methods [7, 8, 14, 17], because of more strict constrains on protein backbone in our model [12, 16]. Furthermore, if we involve improving steps in our algorithm by analogy to [7], such as reconstruction of side chains [13] and energy minimization [15], the results became even more accurate [16] with firm inner energy levels of molecules.

We also have verified results of our modeling by using small-angle X-ray scattering [11] and have showed that such scattering is changing as expected in emulated motions of proteins.

References

1. Ambrosio, L., Gigli, N.: Gradient flows in metric spaces and in the space of probability measures. In: Lectures in Mathematics ETH Zürich, Birkhäuser Verlag (2005)
2. Atilgan, A.R., Durell, S.R., Jernigan, R.L., Demirel, M.C., Keskin, O., Bahar, I.: Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys. J. 80(1), 505–515 (Jan 2001)
3. Bahar, I., Lezon, T.R., Yang, L.W., Eyal, E.: Global dynamics of proteins: bridging between structure and function. Annu Rev Biophys 39, 23–42 (2010)
4. Buzdalov, M., Knyazev, S., Porozov, Y.: Protein conformation motion modeling using sep-cma-es. In: Machine Learning and Applications (ICMLA), 2014 13th International Conference on. pp. 35–40 (Dec 2014)
5. Elber, R., Ghosh, A., Cárdenas, A.: Long time dynamics of complex systems. Accounts of Chemical Research 35(6), 396–403 (2002)
6. Evans, L.C., Gangbo, W.: Differential equations methods for the monge-kantorovich mass transfer problem. Mem. Amer. Math. Soc 137, 653 (1999)
7. Flores, S., Echols, N., Milburn, D., Hespenheide, B., Keating, K., Lu, J., Wells, S., Yu, E.Z., Thorpe, M., Gerstein, M.: The Database of Macromolecular Motions: new features added at the decade mark. Nucleic Acids Res. 34(Database issue), 296–301 (Jan 2006)

8. Franklin, J., Koehl, P., Doniach, S., Delarue, M.: MinActionPath: maximum likelihood trajectory for large-scale structural transitions in a coarse-grained locally harmonic energy landscape. Nucleic Acids Res. 35(Web Server issue), W477–482 (Jul 2007)
9. Karplus, M., Petsko, G.A.: Molecular dynamics simulations in biology. Nature 347(6294), 631–639 (Oct 1990)
10. Knyazev, S., Tarakanov, S., Kuznetsov, V., Porozov, Y., Koucheryavy, Y., Stepanov, E.: Coarse-grained model of protein interaction for bio-inspired nano-communication. In: 2014 6th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). pp. 260–262 (Oct 2014)
11. Knyazev, S.N., Kalyakin, V.Y., Deryabin, I.N., Fedorov, B.A., Smirnov, A.V., Stepanov, E.O., Porozov, Y.B.: Prediction of Protein Conformational Mobility and Evaluation of Its Reliability Using Small-Angle X-ray Scattering. Biofizika 60(6), 1069–1076 (2015)
12. Koshevoy, A.A., Stepanov, E.O., Porozov, Y.B.: Method of prediction and optimization of conformational motion of proteins based on mass transportation principle. Biophysics 59(1), 28–34 (2014)
13. Krivov, G.G., Shapovalov, M.V., Dunbrack, R.L.: Improved prediction of protein side-chain conformations with SCWRL4. Proteins 77(4), 778–795 (Dec 2009)
14. Maiti, R., Van Domselaar, G.H., Wishart, D.S.: MovieMaker: a web server for rapid rendering of protein motions and interactions. Nucleic Acids Res. 33(Web Server issue), W358–362 (Jul 2005)
15. Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M.R., Smith, J.C., Kasson, P.M., van der Spoel, D., Hess, B., Lindahl, E.: GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29(7), 845–854 (Apr 2013)
16. Tamazian, G., Ho Chang, J., Knyazev, S., Stepanov, E., Kim, K.J., Porozov, Y.: Modeling conformational redox-switch modulation of human succinic semialdehyde dehydrogenase. Proteins 83(12), 2217–2229 (Dec 2015)
17. Zheng, W., Doniach, S.: A comparative study of motor-protein motions by using a simple elastic-network model. Proc. Natl. Acad. Sci. U.S.A. 100(23), 13253–13258 (Nov 2003)