Simplified Approaches to Complex Biological Systems

N. V. Dokholyan

http://www.fz-juelich.de/nic-series/volume40

Simplified Approaches to Complex Biological Systems

Nikolay V. Dokholyan

Department of Biochemistry and Biophysics, The University of North Carolina at Chapel Hill,
School of Medicine, Chapel Hill, NC 27599, USA
E-mail: dokh@med.unc.edu

Some of the emerging goals in modern medicine are to uncover the molecular origins of human diseases, and ultimately contribute to the development of new therapeutic strategies to rationally abate disease. Of immediate interests are the roles of molecular structure and dynamics in certain cellular processes leading to human diseases and the ability to rationally manipulate these processes. Despite recent revolutionary advances in experimental methodologies, we are still limited in our ability to sample and decipher the structural and dynamic aspects of single molecules that are critical for their biological function. Thus, there is a crucial need for new and unorthodox techniques to uncover the fundamentals of molecular structure and interactions. We follow a hypothesis-driven approach, which is based on tailoring simplified protein models to the systems of interest. Such an approach allows significantly extending the length and time scales for studies of complex biological systems. Here we describe several recent studies that signify the predictive power of simplified protein models within the hypothesis-driven modeling approach utilizing rapid Discrete Molecular Dynamics (DMD) simulations.

1   Introduction

Despite the multiple innovations in the field of molecular simulations, the size of biological molecules and complexes and the time scales at which they function remain unreachable to traditional computational approaches[1]. This roadblock hallmarks the principal challenge in computational structural biology and is the subject of our current research.

To circumvent the problem with reaching biologically-relevant time and length scales, one must simplify a biological system to elements, essential to a regime of interest. For example, if we are interested in large-scale motions of proteins occurring at the milliseconds to seconds time scale (e.g. protein folding), it is often safe to eliminate atomic vibrations occurring at time scales of picoseconds. Such time scale decoupling is an important maneuver that has been utilized in molecular dynamics simulations. However, despite such innovations as world-wide distributed computing and hardware-customized molecular dynamics, the time scales reachable by these sophisticated techniques are still limited by microseconds time for relatively small systems.

Developed in 1959 approach Discrete Molecular Dynamics (DMD)[2] has a philosophically different event-handling scheme, which makes it extremely efficient. The algorithm is based on satisfying the same basic physical principles as traditional molecular dynamics. However, instead of integrating equations of motions (integrating time $\sim N^3$, where N is the number of particles), DMD engine searches for collision events (search time $\sim N \log N$). This difference makes DMD speed by far more superior to traditional algorithms. With developed all atom force field Medusa[3], DMD simulations allow to extend visible simulation range to microseconds simulations. Next, we will describe applications of DMD engine and Medusa force field to studies of protein folding and modeling, and demonstrate various biotechnological applications.

2   Protein Design and Evolution

Studies of known proteins have revealed the intriguing co-organization of their sequences and structures. Proteins with sequence similarity higher than 25-30% usually adopt a similar structure and are called homologs, while those with low sequence similarity (<20%) can share the same structure and are referred as analogs3. The origin of such co-organization has been the topic of extensive discussions in the protein folding, design and evolution communities, since an understanding of the emergence of homologs and analogs in the protein universe has broad implications on our ability to rationally manipulate proteins. We developed a molecular modeling and design method, Medusa, to computationally design diversified protein sequences that correspond to similar backbone structures. It is these backbone structures that determine protein fold family. Using protein design, we directly demonstrated the formation of distinct protein homologs within a specific fold family when the structure deviates only 1-2 Å away from the original structure[3]. The study suggests that subtle structural changes, which appear due to accumulating mutations in the families of homologs, lead to a distinct packing of the protein core and, thus, novel compositions of core residues. The latter process leads to the formation of distinct families of homologs. This work was important in two ways: Firstly, we demonstrated that using protein design we could mimic the formation of the extremely complex protein universe. Secondly, we developed a new force field that enables us to design proteins and predict protein structure when coupled with the rapid DMD simulations technique.

3   Protein Design and Biotechnology

Mutagenesis is a central tool of molecular biology, genetics, and biotechnology. Knowing to what extent mutations affect the thermodynamic stability and structures of proteins is often vital for designing experiments. Estimation of protein stabilities remains a paramount challenge in computational molecular biology. We extended Medusa to a novel methodology, Eris (http://eris.dokhlab.org), for accurately predicting the mutation-induced protein stability changes[4,5]. Due to the complex nature of the interactions involved in protein folding, existing stability prediction methods often use empirical parameters trained on experimental protein stability data, which makes these methods highly dependent upon the training databases. Limited by their capability to model the structural changes induced by mutations, the applications of the developed methods are often restricted to mutations from large residues to small ones. We addressed these deficiencies with a unique approach that combines a physical force field and a fast conformation-sampling algorithm in an atomic framework of proteins. We showed that Eris could effectively detect and resolve the atomic clashes and structure strains introduced by the mutations and yield reliable predictions of the stability change for these mutants. We expect Eris, which is freely accessible through our server, to have applications on a broad range of mutations in the course of protein engineering. In fact, we have already used Eris to design a peptide to a cysteine-rich intestinal protein one (CRIP1), which has been identified as a novel marker for early detection of cancers. The designed peptide binds to CRIP1 at $K_d \sim 3\mu M$ *in vivo*[6]. Our all atom DMD models now feature Medusa force field, which offers a new automated technology for rational protein engineering and conformational exploration.

4   Protein Modeling

The deletion of the single residue Phe-508 in CFTR present in ∼90% of cystic fibrosis patients prevents maturation of CFTR in endoplasmic reticulum. While this mutation does not significantly affects the structure and thermodynamics of the nucleotide binding domain's (NBD1) where it resides, clearly Phe-508 deletion disrupts tertiary interactions with other domains. The lack of CFTR's 3D structure hampers our understanding how mutation affects channel's both tertiary organization and function. Using homology modeling and DMD/Medusa simulations, we constructed a three-dimensional model of CFTR structure[7]. We have further validated the structure *in vitro* by testing the residues' proximities in our model via engineered disulfide bonds between tested residues. We found that Phe-508 mediates a tertiary interaction between the surface of NBD1 and a cytoplasmic loop (CL4) in the C-terminal membrane-spanning domain. This crucial cytoplasmic membrane interface, which is dynamically involved in regulation of channel gating, explains the known sensitivity of CFTR assembly to many disease-associated mutations in CL4 as well as NBD1 and provides a sharply focused target for small molecules to treat cystic fibrosis. In addition to identifying a key intramolecular site to be repaired therapeutically, our findings advance understanding of CFTR structure and function and provide a platform for focused biochemical studies of other features of this unique ATP-binding cassette (ABC) transporter family ion channel.

5   Concluding Remarks

All atom DMD simulations with Medusa force field allows us to perform redesign protein structures with flexible backbone algorithm, which in turn permits evaluation of the impact of mutations on stabilities of proteins. Furthermore, coupling DMD with Medusa allows one to improve the homology models and refine the structures.

References

1. F. Ding, and N. V. Dokholyan, *Simple but predictive protein models*, Trends Biotechnol. **23**, 450–455, 2005.
2. N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley, and E. I. Shakhnovich, *Discrete molecular dynamics studies of the folding of a protein-like model*, Folding and Design **3**, 577, 1998.
3. F. Ding, and N. V. Dokholyan, *Emergence of protein fold families through rational design*, PLoS computational biology **2**, e85, 2006.
4. S. Yin, F. Ding, and N. V. Dokholyan, *Eris: An automated estimator of protein stability*, Nature methods **4**, 466–467, 2007.
5. S. Yin, F. Ding, and N. V. Dokholyan, *Modeling backbone flexibility improves protein stability estimation*, Structure **15**, 1567–1576, 2007.
6. J. Hao, A. W. R. Serohijos, G. Newton, G. Tassone, Z. Wang, D. C. Sgroi, N. V. Dokholyan, and J. P. Basilion, *Identification and rational redesign of peptide ligands to CRIP1, a novel biomarker for cancers*, Public Library of Science Computational Biology, in press (2008).

7. A. W. R. Serohijos. T. Hegedus, A. A. Aleksandrov, L. He, L. Cui, N. V. Dokholyan, and J. R. Riordan, *Phenylalanine-508 mediates a cytoplasmic-membrane domain contact in the CFTR 3D structure crucial to assembly and channel function*, Proceedings of the National Academy of Sciences of the United States of America **105**, 3256–3261, 2008.