Dang Huynh
Genomics and Bioinformatics

The Union of Conformational Flexibility Alignment and Structural Comparison

Introduction

Traditionally, the study of conformational flexibility has been the focus of protein docking and drug design. Recently, Keskin et al. (2000) shows that proteins with similar folds exhibit similar conformational flexibly characteristics, suggesting that it can be used for structural alignment. In addition, an efficient computational method that can predict conformational flexibility within a few seconds of CPU time has also been developed (Jacobs et al., 2001). These two combinations imply that conformational flexibility analysis can be automated and be used to assist structural alignment. This paper provides an overview of the field and proposes an automated procedure to employ conformational flexibility alignment in structural comparison.

Conformational Flexibility: In a Nutshell

Definition of conformational flexibility

Conformational flexibility is defined as the ability of a protein to undergo conformational changes within a "window" of native state. The window of native state is a range of free energy where most conformations populate (Price, 2000). It is the state at which the protein is functional (Price, 2000). On general, it prefers to the global minimum on the free energy profile or the most stable state explores by a protein (Carlson & McCammon, 2000). Within this population of stable conformations, one conformation can be converted to another conformation through bond-rotations. These rotational events lead to the local unfolding of a particular region within the protein. They involve the breaking and making of non-covalent interactions such as hydrogen bonds, salt bridges, van der Waals forces and hydrophobic interactions (Price, 2000). In a simplified case, the interconversion between the closed-liganded and the open-unliganded conformations is an example of conformational flexibility. Conformational flexibility differs from protein denature cases because it does not involve global unfolding of a protein. During the process of converting from a closed-liganded conformation to the open-unliganded form, the protein does not denature.

The significance of conformational flexibility in protein function

The essential role of conformational flexibility in protein function is first proposed by Koshland (1959) through his famous "induced-fit model". In the model, he postulated that upon binding of a substrate, the protein underwent conformational changes to accumulate the substrate. Today, conformational flexibility is known to control numerous biological processes. For example, the rotations of the ribosome during protein synthesis accompany numerous conformational changes within the protein complex (Agrawal & Frank, 2000). These conformational changes facilitate protein production. The high affinity of antibodies for antigens that allows the immune system to be carried out efficiently is governed by conformational flexibility. Conformational flexibility stimulates the production of ATP through the rotations of ATP synthase. These are only a few examples of how conformational flexibility influence biochemical processes.

Experimental methods to determine conformational flexibility

Although understanding conformational flexibility is important to comprehend biochemical reactions, few works have been done to characterize its properties in proteins on a large scale. The lack of efficient and high throughput methods to measure conformational flexibility has prevented such analysis. Traditionally, conformational flexibility characteristics of a protein can be inferred by looking at the B factors, obtained from X-ray crystallography, of each atom in the protein (Price, 2000). However, this method does not reveal the cooperative interactions occur during structural changes. Information on conformational flexibility can also be acquired by comparing multiple structures of the same protein (Price, 2000). Unfortunately, the laborious process of structural determination by NMR and X-ray crystallography techniques limits the application of the method. The most quantitatively approach to detect conformational flexibility is hydrogen exchange experiment (Price, 2000). When a protein is incubated in deuterated water, the amide hydrogen atoms of the protein can exchange with deuterium atoms from the solvent (Price, 2000). The rate of hydrogen exchange depends on the location of a particular residue in a protein. Surface residues have higher proton exchange rate than buried residues (Fields, 2001). As a protein undergoes conformational changes, buried residues become exposed, and their hydrogen atoms can undergo proton exchange with the solvent. The rate of proton exchange of buried residues is related to the rate of conformational fluctuations in a protein (Fields, 2001). Nonetheless, this method is difficult to be automated.

Computational approaches to predict conformational flexibility

Conformational flexibility of proteins can also be predicted using computational tools. These methods can be divided into two groups: Simulating protein motions and identifying rigid and flexible regions with a protein using a single conformation (Jacobs et al., 2001). The methods in the first group are popular, but they are computationally expensive. An example of such methods is GNM (Gaussian Network Model) (Haliloglu et al., 1997). In GNM, a protein is modeled as a network of strings. The strings correspond to the bonded and non-bonded interactions within a protein. The strings are connected through Cα atoms. A constant force of harmonic potential is applied to all residues, and the residues are allowed to vibrate around their mean coordinates. The calculated vibrational frequency depicts protein motions. In the second group, the methods are more applicable for automated purposes because they are relatively simple and computationally fast. Henceforth, the paper will only focus on the second group with a special emphasis on FIRST (Floppy Inclusion and Rigid Substructure Topography) method (Jacobs et al., 2001). In FIRST, the entire protein is modeled as a framework of constraints. The constraints are the bonds, which are covalent bonds and hydrogen bonds. Covalent bonds are always modeled as constraints while hydrogen bonds must meet a certain geometric and energy criteria to be modeled as constraints. By counting the number of constraints within a particular region, flexible and rigid clusters are identified. The flexibility features are then plotted against the residue number. The greatest advantage of FIRST is that it can capture conformational flexibility characteristics using only one conformation.

Conformational Flexibility Alignment and Structural Comparison

As mentioned previously, Keskin et al. (2000) have suggested that conformational flexibility properties might be used for structural alignment. However, they did not propose how such approach could be achieved. This paper will propose a procedure to align conformational

flexibility characteristics and use the alignment to quickly diagnose proteins with similar structure.

The analysis will involve the characterization of all known protein folds (605 folds)(Conte et al., 2002). Target proteins will be identified from SCOP database. Whenever possible, at least five proteins from each fold will be studied. The corresponding atomic coordinates will be downloaded from Protein Data Bank (Bernstein et al., 1977). After that, the following steps will be carried out a) aligning the conformational flexibility characteristics b) establishing the statistical significant of the alignments and c) identifying conserved conformational flexibility motif among proteins with the same fold. In the first step, FIRST will be used to characterize the conformational flexibility profiles of multiple conformations of the same protein to establish a standard similarity score. The conformational flexibility similarity matrix will be constructed similar to sequence alignment. The conformational flexibility characteristics of different structures will be aligned using dynamic programming. In the second step, the statistical significant of the alignment in term of P value will be calculated using method similar to Livitt and Gerstein approach (Levitt & Gerstein, 1998). The conformational flexibility characteristics of proteins from the same fold will be used as true positive samples whereas proteins from different folds will serve as true negative cases. Subsequently, conserved conformational flexibility motifs will be identified. The identified motifs will be stored in a database and will then be used to quickly diagnose proteins with similar structure.

Conclusions

The ultimate goal of the method is to translate structural similarity to conformational flexibility similarity and use conformational flexibility motifs to identify proteins with similar folds. The advantage of the method is that conformational flexibility alignment is expected to be quicker and simpler than structural alignment. In structural alignment when a new alignment occurs it affects previous optimal alignment, which will not be the case in conformational flexibility alignment. One could think of conformational flexibility alignment as an intermediate level between sequence alignment and structural alignment. However, the author does not intend to use conformational flexibility alignment to replace method such as root mean square (RMS) superposition but rather complement such technique. The objective of the method is to provide a simple and quick estimate of the similarity between two structures. It could be served as an initial step for RMS superposition. It should be suitable for non-expert usage. Conformational flexibility alignment would add a new dimension on how one could view relationship among proteins. Not only could one categorize proteins according to sequence, structure, and function, but one could also classify them according to their dynamics. The beauty of bioinformatics is to develop new approaches that allow researchers to interpret their data at different angles. Conformational flexibility alignment will be part of that endeavor.

References

Agrawal, R. K. & Frank, J. (2000). A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature* **406**, 318-322.

Bernstein, F. C., Koetzle, T. F., B., W. G. J., Meyer, E. F. J., D., B. M., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The Protein Data Bank: a computer based archival file for macromolecular structures. *Journal of Molecular Biology* **112**, 535-542.

Carlson, H. A. & McCammon, A. J. (2000). Accommodating protein flexiblity in computational drug design. *Molecular Pharmacology* **57**, 213-218.

Conte, L. L., Brener, S. E., Hubbard, T. J. P., Chothia, C. & Murzin, A. G. (2002). SCOP database in 2002: refinements accommodate structural genomics. *Nucleic Acids Research* **30**, 264-267.

Fields, P. (2001). Review: Protein function at thermal extremes: balancing stability and flexibility. *Comparative Biochemistry and Physiology* **129**, 417-431.

Haliloglu, T., Bahar, I. & Erman, B. (1997). Guassian dynamics of folded proteins. *Physical Review letters* **79**(16), 3090-3093.

Jacobs, D. J., Rader, A. J., Kuhn, L. A. & Thorpe, M. F. (2001). Protein flexibility predictions using graph theory. *Proteins* **44**, 150-165.

Levitt, M. & Gerstein, M. (1998). A unified statistical framework for sequence comparison and structural comparison. *Proceedings of the National Academy of Sciences of the USA* **95**, 5913 5920.

Price, N. C. (2000). Conformational issues in the characterization of proteins. *Biotechnology and Applied Biochemistry* **31**, 29-40.