

Mapping of Protein Surfaces for Identification of Functional Similarity between Protein Motifs

Flávio A. C. Anjos, Health Informatics Department, Federal University of São Paulo - flavio-pg@dis.epm.br
Victor M. S. B. B. Oliveira, Biophysics Department, Federal University of São Paulo - vitor@biofis.epm.br
Paulo B. Paiva, Health Informatics Department, Federal University of São Paulo - paiva@unifesp.br

Abstract

With the diversity of data originated from projects for sequencing genomic DNA a great number of sequences is known. Intending to have a better knowledge of the molecular dynamics of organisms, proteins are extremely relevant as targets of study.

Certain conserved regions of protein structures are called motifs. This term is used in two manners in structural biology. As a particular amino acid sequence with a specific biochemical function, and as a continuous set of elements of the secondary structure that also has a specific function or defines an independent portion of the protein folding of the domain (1). Identification of protein motifs can be difficult since there may be stretches of amino-acid sequences pertaining to a motif which are not continuous in the primary sequence.

Determining protein function is an important objective in biology. Experimental techniques are expensive and time-consuming (2). In order to improve the process, computational techniques which analyze the structure of a target protein considering geometric and chemical similarities to known functional regions (motifs) (2). Using structural alignments to known motifs may help elucidate protein structure.

The proposed approach to this problem is the development of a software named MSProt (Mapping of Surface Protein) which allows users to visualize the three-dimensional structure of proteins and will evaluate geometrical and chemical similarity between protein motifs. The proposed software will not determine protein fold *ab initio*, it will use known tertiary structures aligned to the target protein. The resulting code will be placed in the public domain using an open-source license, allowing users to freely use, access the code, modify and redistribute it.

MSProt is being developed in C++ (3) using the OpenGL library for 3D programming and an open-source version of the Qt library version 4 (5). Binary distributions for Windows and Linux-based systems will be available since the development tools used allow porting for multiple platforms.

The development of the software will include an API (Application Programming Interface) based on

C++ objects, allowing other researchers to develop structural biology projects using the 3D environment and the functionalities implemented by MSProt.

The protein surface mapping will be accomplished by means of a geometric modeling, using differential geometry (8) to produce a level surface.

In the proposed mapping method, the accessibility of the solvent in the protein will be considered (6) in order to supply the necessary coordinates for the estimation of the level surface. The biochemical characteristics of the surface of the protein will be used (7).

The target protein is aligned to a known protein domain motif used as scaffold for mapping similar motifs on the target protein. A comparative analysis will produce a similarity score. The proposed similarity score is a scalar value which indicates the functional similarity between protein motifs. A larger score represents a greater functional similarity between the target protein and the motif.

A beta version of the software is available in the address: <http://compbio.epm.br/msprot/>

References

- [1] Petsko, G.A.; Ringe, D. (2004). Protein Structure and Function. New Science Press Ltd. pp. 34-35.
- [2] Chen, B. Y.; Fofanov, V. Y.; Bryant, D. H.; Dodson, B. D.; Kristensen, D. M.; Lisewski, A. M.; Kimmel, M.; Lichtarge, O.; Kavraki, L. E. (2006). Geometric Sieving: Automated Distributed Optimization of 3D Motifs for Protein Function Prediction. Research in Computational Biology: 10th Annual International Conference (RECOMB), Venice, Italy.
- [3] Stroustrup, B. (1999) Learning Standard C++ as a New Language. C/C++. Users Journal. pp 43-54.
- [4] The Industry's Foundation for High Performance Graphics OpenGL (2006) - (<http://www.opengl.org/>)
- [5] Trolltech (2006) - (<http://www.trolltech.com/>)
- [6] Connolly, M.L. Solvent-accessible surfaces of proteins and nucleic acids. Science, 221:709{713, 1983}.
- [7] Kinoshita, K.; Nakamura, H. (2003). Identification of protein biochemical functions by similarity search using the molecular surface database eF-site. Protein Science, 12: pp. 1589-1595.
- [8] Chen, B. Y.; Fofanov, V. Y.; Kristensen, D. M.; Kimmel, M.; Lichtarge, O.; Kavraki, L. E. (2005). Algorithms for Structural Comparison and Statistical Analysis of 3D Protein Motifs. In Pacific Symposium on Biocomputing, pp. 334-345, World Scientific, Hawaii, USA.