Conformational Alphabets in the Study of Protein Structure

Wei-Mou Zheng (zheng@itp.ac.cn, zhengwm@genomics.org.cn) Institute of Theoretical Physics, Academia Sinica, Beijing, China

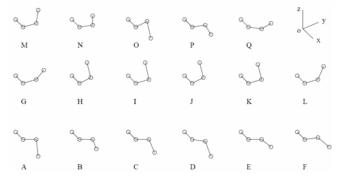
Introduction

Life science, bringing us ever-growing unexplained data, is attracting increasing interest from physicists. Protein structure is a good subject for physicists to study, being simple enough yet, at the same time, complex enough.

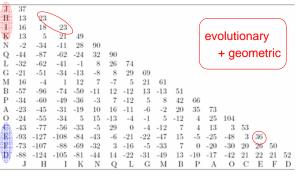
Conformational alphabets

Conformational alphabets (CAs) provide discrete representation for the protein local structure. We have proposed an alphabet:

a letter = a *cluster of combinations of three angles* formed by C_{α} pseudobonds of four contiguous residues (obtained by *clustering* according to the probability distribution) = a discretized state of 3D segmental conformations, and constructed its substitution matrix CLESUM (Conformational LEtter SUbstitution Matrix), which measures similarity between fragmental states.



Centers of 17 conformational letters



Similarity between conformational letters

Conformational alphabets in the study of protein structure

We have succeeded in developing a fast alignment tool for multiple protein structures by means of our CA called BLOMAPS, which is faster than other tools by 2 to 3 orders. Multiple alignment carries significantly more information than pairwise alignment, and hence is a much more powerful tool for classifying proteins, detecting evolutionary relationship and common structural motifs, and assisting structure/function prediction.

Such conformational alphabets, bridging the secondary and 3D structures, facilitate computations for protein structures. Our focus is on the development of a reliable statistical potential function using CAs, and improving structure prediction.

This is a rich territory for collaboration amongst physicists, computer scientists, mathematicians, and biologists.

References

1, WM ZHENG and X LIU, A protein structural alphabet and its substitution matrix CLESUM, in Lecture Notes in Bioinformatics **3680**, pp. 59-67, Springer, Berlin, 2005.

2, S Wang and WM Zheng, *Fast multiple alignment of protein structures using conformational letter blocks*, The Open Bioinformatics J., **3** (2009) 69-83.

On-going Project between Europe and China

of Metagenomic sequencing of gut microbiomes (conducted in Beijing Genomics Institute, Shenzhen) within the frame of the European Commission funded program MetaHIT (Metagenomics of the Human Intestinal Tract).