中国科学院数学与系统科学研究院
Academy of Mathematics and Systems Science, CAS

生物信息学研究中心

Center of Bioinformatics

学术报告

题　　目： A Knowledge-Based, Statistical Informatics Approach for Protein Structure Refinement

报告人： Prof. Zhijun Wu
Department of Mathematics, Program on Bioinformatics and Computational Biology
Iowa State University, USA

时　　间： 12 月 21 日（星期四）下午 3: 00

地　　点： 思源楼 712 室

摘　　要： The protein structures determined by conventional techniques usually are not as accurate as desired. Further refinement including human intervention is always required and sometimes critical. Therefore, the development of an efficient refinement technique is important, and as more and more structures are determined, the need is even more urgent, as the CASP prediction center explained for the call for a structure refinement competition in spring 2006.

Here, we describe a computational approach of deriving distance constraints from databases of known protein structures for structure refinement. We calculate the distributions of the distances of various types in known protein structures, and use them to obtain the most probable ranges or the mean-force potentials for the distances. We then impose the constraints on the structures to be refined or include the mean-force potentials in energy minimization so that more plausible structural models may be built. We show that many inter-atomic distances in low-resolution structures deviate significantly from their average distributions in known protein structures, and the structures can be refined when a selected set of distances are constrained to their most probable ranges or optimized with corresponding mean-force potentials.

We present the results from refining a set of NMR-determined protein structures by using database derived distance constraints and mean-force potentials, and show the improvements on the structures in terms of several standard measures. We also discuss our results from participating in the CASPR 2006 structural refinement experiments for comparative model refinement, using energy minimization, database derived distance constraints, and massively parallel computing. We describe the development of a database of protein inter-atomic distances that supports computing the distributions of the distances of various types in known protein structures and generating the constraints or potentials for the distances automatically. We discuss the possibilities of extending the system to a broader sense of protein geometry database and using it for structure analysis, classification, as well as refinement.