



SEMINAR

Protein Universe and its applications

Professor Stephen S.-T. Yau
Tsinghua University, Beijing

Abstract

Current methods cannot tell us what the nature of the protein universe is concretely. Here we develop a novel method to realize the nature of protein universe. We show the protein universe can be realized as a protein space in 60-dimensional Euclidean space by using the distribution information of each amino acid within the protein. Every protein is in one-to-one correspondence with a point in protein space, where proteins with similar properties stay close together. Thus the distance between two points in protein space represents the biological distance of the corresponding two proteins. We also propose a natural graphical representation for inferring phylogeny. The representation is natural and unique based on the biological distances of proteins in protein space. This will solve the fundamental question of how proteins are distributed in the protein universe.

As an application, we shall answer the following fundamental question. What kinds of amino acid sequences could possibly be protein sequences? From all existing databases that we can find, known proteins are only a small fraction of all possible combinations of amino acids. Here we show that when the collection of arbitrary amino acid sequences is viewed in an appropriate geometric context, the protein sequences cluster together. This leads to a new computational test, described here, that has proved to be remarkably accurate at determining whether an arbitrary amino acid sequence can be a protein. Even more, if the results of this test indicate that the sequence can be a protein, and it is indeed a protein sequence, then its identity as a protein sequence is uniquely defined.

Date: May 4, 2016 (Wednesday)

Time: 1:30 – 2:30pm

Place: Room 210, Run Run Shaw Bldg., HKU