



Department of Biological Sciences Seminar Announcement

Predicting allergenic proteins *with* bioinformatics

With many transgenic proteins introduced today, the ability to predict their potential allergenicity has become an important issue. Previous studies are based on either sequence similarity or the protein motifs identified from known allergen database. The similarity based approaches, although being able to produce high recalls, usually have low prediction precisions. The motif-based approaches have improved precisions on cross-validation experiments but not on real protein sequences. We have developed a system that combines the advantages of similarity-based and motif-based prediction. The new prediction system uses a clustering algorithm that groups the known allergenic proteins into clusters. Proteins within each cluster are assumed to carry one or more common motifs. After a multiple sequence alignment, proteins in each cluster go through a wavelet analysis program whereby conserved motifs will be identified. An HMM profile will then be prepared for each identified motif. The allergens that don't appear to carry detectable allergen motifs will be saved in a small database. The allergenicity of an unknown protein may be predicted by comparing it against the HMM profiles, and, if no matching profiles can be found, against the small allergen database by BLASTP. Over 80% of recall and over 90% of precision were observed using cross-validation experiments. Using the entire Swiss-Prot as the query, we reached a precision about 40%.

Profile

Dr. Kuo-Bin Li is a Young Investigator of the Bioinformatics Institute (BII), Singapore. He joined the institute in December 2001. He received his Ph.D. degree in computational chemistry from McGill University, Montreal, Canada in 1996. The focus was to develop novel algorithms for protein structure determination using NMR. He went on to the Rowland Institute at Harvard, Cambridge, MA, USA for the postdoctoral studies. From 2000 to 2001, he joined a bioinformatics team in the Industrial Technology Research Institute in Taiwan. He was involved in the development and implementation of a fast pairwise alignment software on distributed and parallel computer platforms. At the Bioinformatics Institute, Dr. Li leads the computational genomics group. The ongoing projects include studies of the regulatory roles of non-coding RNA molecules, novel algorithms for protein multiple sequence alignments and prediction of protein functional domains.

By Dr Li Kuo-Bin
Bioinformatics Institute,
Singapore

Fri, 28 May 2004

LT 20, 4 pm

Hosted by

Dr Chew Fook Tim