S6-1

PreADME : Virtual Screening for Early ADME Prediction

Sung Kwang Lee, In Hye Lee, Hyun Ji Kim, Gue Su Chang, Jae Eun Chung and Kyoung Tai No
Bioinformatics & Molecular Design Technology Innovation Center

A significant bottleneck in the drug discovery is analysis of the ADME and toxicity properties of drug candidates. Over 50% of the candidates failed due to poor ADME/Tox properties during development. To avoid this failure at the development, a set of in vitro ADME screens has been implemented in most pharmaceutical companies with the aim of discarding compounds in the discovery phase that are likely to fail further down the line. For this reason it is necessary to develop computational methods for predicting drug-likeness. Pre-ADME is a program designed to predict physico-chemical, drug absorption and drug-like properties. It calculates many physical properties and some mathematical descriptions, descriptors such as polar surface area, water solubility and logP. The absorption properties (Caco-2, MDCK, blood brain barrier and human intestinal absorption) of compounds can be predicted by PreADME. The prediction system is composed of Artificial Neural Network and calculation of descriptors which were selected with Genetic Algorithm. PreADME is useful for chemical library design considering drug-likeness, drug absorption and water solubility. You can evaluate this program at http://camd.ssu.ac.kr/adme, free of charge.

S6-2

Top-down EST Clustering Using the Draft Human Genome Map

Namshin Kim[1], Seokmin Shin[1] and Sanghyuk Lee
[1]Dept. of Chemistry, Seoul National University and Div. of Molecular Life Sciences, Ewha Womans University

EST clustering plays a central role in finding unknown genes and provides valuable information on gene expression, as can be seen the UniGene at NCBI and STACKdb in SANBI. We developed a new EST clustering algorithm utilizing the draft genome map. Human ESTs in dbEST database are mapped onto the UCSC assembly of human genome (so-called the goldenpath) using the BLAT program. Alignments at multiple loci are allowed. Alignments with over 95% match and with significant portion of EST (over 80%) are included in the clustering. The significant match excluding the repeat sequences (obtained by the RepeatMasker) should be at least 100 base pairs. The resulting alignments are clustered in top-down fashion. mRNAs and spliced ESTs are clustered initially, and singleton ESTs (ESTs without intron) are added in the subsequent step. Exon-intron mismatches are not allowed to find alternative splicing variants, thereby creating only one consensus sequence for each cluster. Comparison with the UniGene and TIGR Gene Indices shows that our EST clusters are significantly different from those clusters partly because alignments at multiple loci are allowed.