Annotation of EST sequences by a structural bioinformatics approach - (session: Other)

L. Pugliese

via Don Grioli, 4 -Torino

In the past few years the complete sequences of more than 55 genomes have been published and at least 100 more are known to be near completion. One challenge of the genome era is to predict molecular functions and biological roles for the predicted gene product.

Most approaches for the tentative assignment of functions to predict proteins are based on pairwise sequence similarity searches against known proteins using sequence comparison programs such as FASTA and BLAST. However many proteins are multifunctional multidomain proteins for which the assignment of a single function results in loss of information. Also with more predicted proteins from genome projects being added to the databases, the best hit in pairwise sequence similarity searches is frequently a poorly annotated protein.

To overcome limitations of functional annotation based on pairwise sequence similarity searches, the addition of knowledge of the three dimensional structure of domains gains more and more importance. In this view the application of fold recognition methods coupled to homology model building and theoretical structure verification methods represents a way to get a lot of information in a short time.

The protocol applied in order to assign a function to an EST sequence involves the following steps:
1. Submission of the sequence to a fold recognition/structure prediction metaserver.
2. 3D alignment of sequences relative to templates receiving good scores from the metaserver.
3. Homology model building using as template the pdb files having the best scores within the metaserver output.
4. Evaluation of the models obtained by the program ProsaII and the server http://atlas.physbio.mssm.edu:8084/servers/pg/
5. Analysis of the literature concerning the template structures in order to extract information on the function of the new sequence.