A workflow for an orthology-based prediction of protein-protein interaction

Angeletti M (2), Bartocci E (1), Merelli E (1)

(1) Dipartimento di Informatica e Matematica, Università di Camerino, Camerino (MC)
(2) Dipartimento di Biologia Molecolare, Università di Camerino, Camerino (MC)

Motivation

Many genomes have been completely sequenced. However, detecting and analyzing their protein–protein interactions by experimental methods is not as fast as genome sequencing.These activities are very important to understand properties and mechanisms of cellulars system. The knowledge of the complete picture of the interaction of proteins with their ligands of an organism is defined as the interactome, which can be expressed by a map of (scalefree) interconnecting nodes where the edges represent a value of the interaction strength under particular conditions. Parallel to the application of experimental techniques to the determination of protein interaction networks and protein complexes, the first computational methods, based on sequence and genomic information, have emerged. In our approach we have considered the orthology-based method in which two proteins may interact if each one have at least an ancestor in its phylogenetic tree interacting with at least an ancestor of the other.

Methods

We have designed a workflow taking in input the aminoacid sequences of two proteins A and B. In the first step we search for A and B orthologues separately, using COG database (http://www.ncbi.nlm.nih.gov/COG/new), which contains a classification of proteins, from seven complete genomes and five major phylogenetic lineages, according to their orthologous relationships. Orthologous proteins share a common ancestor and they have been separated by a speciation event. They usually have the same function. We obtain two sets of A- and B-orthologues (AO and BO sets). Then we search for elements of the AO set and BO set which are known to interact.

In our approach we use BIND (http://www.bind.ca), BRITE (http://www.jaist.genome.ad.jp/brite/) and DIP (http://dip.doe-mbi.ucla.edu/) as protein-protein interactions datasets. From the search result in AO and BO subsets whose members are known to interact we obtain the XI and YI subsets. If they are not null, we can conclude that the query A and B proteins should interact. We have implemented this workflow using BioWMS[1].

Results

We are testing experimentally some produced results using this approach. This workflow will be available soon at http://litbio.unicam.it:8080/biowms.

Availability: http://litbio.unicam.it:8080/biowms

Contact email: ezio.bartocci@unicam.it

References

1. Bartocci E., Corradini F., Merelli E., Scortichini L., BioWMS: a web based Workflow Management System for Bioinformatics. Submitted to BITS 2006.