18. A semi-automated method to study the intersection of metabolic network and interactome

Joan Planas Iglesias

Structural Bioinformatics Laboratory, Research Group on Biomedical Informatics (GRIB), IMIM/UPF. c/ Dr. Aiguader 80. 08003 Barcelona, Spain

We propose a semi-automated method to study the intersection of metabolic network and interactome that aids inferring similar protein-protein relationships in other organisms than the few from which the actual experiments were performed. Inferred relationships are scored and compared against metabolic annotations (KEGG pathways) to assess their reliability.

The understanding and characterization of data that comes from proteomic experiments is one of the major challenges of the post-genomic era. Recent studies on the interactome (set of protein interactions) and metabolic network (set of metabolic pathways) have yielded a large corpus of data, but the biological significance of their complexity is still poorly understood and the validity of some interactions and protein relationships is questioned. A graph-based analysis of the intersection of both, the interactome and the metabolic network, may help to improve the knowledge of this complexity and to propose similar relationships in other organisms than the few from which the actual experiments were performed. Numerically describing protein-protein metabolic relationships as a function depending on reactors and reagents allows to compare proposed metabolic relationships against a golden standard (such as KEGG)

The interactome was obtained by XML parsing of the DIP database (January 5th 2004; Salwinski et al. 2004) containing 22872 proteins and 50797 protein-protein interactions. On the other hand, we reconstructed a metabolic network by text parsing the KEGG-integrated LIGAND database (Release29.0; Goto et al. 2002), linking two enzymes A and B whenever both were present at the same organism and a chemical compound different than water was acting simultaneously as product of A and substrate of B (hereafter named functional link). We scored those functional links depending on the relative frequency of the shared reagents (compounds acting simultaneously as product of A and substrate of B) and the frequency of non-shared reagents (compounds acting exclusively in A or in B), obtaining a metabolic score for each pair. Organism specific information of enzymes was obtained by scanning the SwissProt database (release 46.5, April 12th 2005; Boeckmann et al. 2003). We obtained a total of 118082 functional links involving 17397 distinct proteins. Remarkably, graphs formed with these links are not classical metabolic pathways because there is no orientation in the direction of their links, while for a metabolic pathway that orientation is needed.

The analysis of the intersection between interactome and our putative metabolome was performed at three different graph levels: G1 (intersection graph) for the raw intersection, representing proteins as nodes and interactions as edges; G2 (abstracted graph) where nodes from G1 were substituted by their codification in the Cluster of Orthologous Groups (COG) and the graph was rearranged to remove redundant edges; and G3 where nodes from G2 were substituted by all proteins encoded in the COG and nodes and edges for each specific organism were rearranged to obtain a particular graph for each organism. The construction for the G3 graphs (reconstructed graphs) implies the assumption of interologs: an interaction between a pair of proteins in an organism will also exist for the orthologous pair in other organisms and is named interolog (Park et al, 2001; Matthews et al, 2001). Maximal connected subgraphs of the studied graphs or components (Gross & Yellen 1999) were considered

during the analysis as well as the presence of one or more common KEGG-annotated metabolic pathways for all enzymes represented in a component.

Three main effects were observed in G1 to G3 transition: increment of total number of links, reconstruction of components in G1 poorly represented organisms (eg. human), and joining of distinct components in organisms which were highly represented in G1, as yeast. This last effect is driven by the joining of some G1 components in G2 through a shared COG-code. Complexity of functional interaction networks was compared for different selected organisms in G3, observing an increase of complexity in higher eukaryotes.

From this work we conclude that the intersection of metabolic network and interactome aids inferring similar protein-protein relationships in other organisms than the few from which the actual experiments were performed. Validity of these newly inferred protein-protein metabolic relationships is assessed by their metabolic score: only pairs with metabolic score values similar to those in KEGG are allowed.

References

Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M. (2003)
The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res, 31, 365-70

Goto S, Nishioka T, Kanehisa M.(1998)
LIGAND: chemical database for enzyme reactions.
Bioinformatics, 14, 591-9

Gross J, Yellen J.(1999) Graph Theory and its Applications. RCC Press, Boca Raton, FL, U.S.A.

Matthews LR, Vaglio P, Reboul J, Ge H, Davis BP, Garrels J, Vincent S, and Vidal M. (2001)
Identification of Potential Interaction Networks Using Sequence-Based Searches for Conserved Protein-Protein Interactions or "Interologs"
Genome Res, 11, 2120-2126

Park J, Lappe M, Teichmann SA. (2001)
Mapping protein family interactions: intramolecular and intermolecular protein family interaction repertoires in the PDB and yeast1
J Mol Biol, 307, 929-38

Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D. (2004)
The Database of Interacting Proteins: 2004 update
NAR, 32, 449-51