

Integrated analysis of regulatory and metabolic intra- and intercellular networks

Edgar Wingender

Dept. of Bioinformatics, University Medical Center Göttingen, Georg August University, Göttingen, Germany
BIOBASE GmbH, Wolfenbüttel, Germany

Aim

Systems biology aims at analyzing and predicting the behavior of biological systems, usually: of cells. This requires knowledge about the underlying network: the wiring diagram has to be known (or reconstructed from the experimental observations), the parts that are most relevant for the observed process have to be identified, and have to be made amenable to dynamic simulation. In this presentation, the required knowledge bases and approaches will be exemplified and discussed.

The data & network construction

Most of the resources used for constructing the network to be analyzed were different modules of the BIOBASE Knowledge Library™, BKL: The signaling network was extracted from the TRANSPATH® database, the mammalian transcription network from TRANSFAC®, the yeast protein-protein and a yeast genetic network interaction network from YPD™. In addition, public domain datasets about metabolic and regulatory pathways and networks were used for comparison.

For the integration of the different kinds of networks, a common semantic had to be developed for the network vertices and/or edges. In a first approach, we projected all encoded molecules on the corresponding genes, the edges between them representing any kind of functional interaction, or exchange, between them. This means that the metabolic network vertices represent genes of enzymes (or their subunits), whereas the edges stand for the exchange of metabolites. In transcription networks, the vertices represent transcription factor genes, the edges summarize their expression plus the regulation of another TF gene by the gene product, and so on.

Extension of regulatory networks by predicted TF and miRNA targets

In particular, the eukaryotic (mainly: mammalian) transcription network suffers from lack of sufficient knowledge; conservative estimates tell us that the experimental knowledge presently available may cover not more than 1% of all in vivo relevant transcription factor binding sites. Thus, dealing with transcription networks requires reliable prediction approaches for adding further TF-target gene relations to the regulatory network of a mammalian cell. However, to do these predictions in a reliable way is still a major challenge. The same is true for other important players in gene regulation control, micro-RNAs and their target sites. Some recent contributions to the state of the art will be discussed.

Topological features of different networks

The networks that we were able to construct so far, along with a couple of freely available networks, were analyzed for their global and local topological characteristics. The different kinds of networks show some significant differences with regard to some of the well-known topological parameters. However, they hardly help identify the most relevant components in a network. For this, *centrality* (in particular: *betweenness centrality*) may be a useful criterion. To overcome some of the shortcomings of this parameter, we have established and computed the *pairwise disconnectivity index* of vertices, edges and motifs of a network. The particular advantages of this new characteristic will be demonstrated.

Mapping expression data onto knowledge-derived networks

Based on the aforementioned knowledge bases and methodologies, we have developed a platform (ExPlain™) which analyzes transcriptomics or proteomics data in an integrated way, including functional categorization, pathway and promoter analysis. It allows to identify potential master regulators for the observed process, which may then be subjected to targeted experimental validation. Proof-of-principle for this novel concept will be presented.

Simulation of network dynamics

Most attempts to simulate complex networks require huge computational facilities, in particular if they try to apply systems of ordinary differential equations (ODEs). What is needed are approaches to identify the appropriate granularity on which simulation has to be executed, and to isolate that core subgraph which has to, and can, be simulated in all details. How to transfer graphs that were identified by – omics analyses into a simulation platform such as Cell Illustrator™ (CIO™) will be demonstrated.