Poster J-50

Monet Ontology as a tool to construct and explore integrated biological networks

Authors:
Ney Lemke (*IBB-Unesp*)
Renata Vieira (*Unisinos*)
João Paulo Müller da Silva (*Unisinos*)
José Guilherme Camargo de Souza (*UNISINOS*)
Marialva Sinigaglia (*IF-UFRGS*)
José Carlos Mombach (*IF-UFRGS*)

Short Abstract: We present MONET (Molecular Network) ontology, an integrated model for the integration of molecular cellular networks: metabolism, protein-protein interaction and regulation. Using the ontology we build a network model for E.coli molecular interactions. Applying machine learning techniques we predict gene essentiality with 85.7 recall.

Long Abstract:

One of the most important challenges for biology in the post-genomic era is to understand the structure and
behavior of the molecular interactions that controls cell behavior. Therefore is essential to integrate
biological data concerning these interactions, which are stored in different databases. The integration task is difficult because these data are distributed in public databases on the world wide web and each database has different management systems, formats and views of how to represent biological data. The two main problems involved here are the difficulty in parsing the data when dealing with heterogeneous flat file formats and the inconsistencies due to the absence of an unified vocabulary. As an alternative to facilitate these problems this work proposes MONET (the Molecular Network) ontology, an integrated model for the integration of different molecular networks that exist inside the cell. Such integrated view facilitates the understanding of the large-scale interactions responsible for the behavior of the cells, and the prediction of cellular behavior that can be tested experimentally.

The ontology integrates metabolic data and protein-protein interaction
for prokaryote and eukaryote organisms and also transcriptional-regulatory data only for prokaryote organisms.
As result, this work provides a standardization of the terms used in these areas of the ontology and the
population of the ontology with data referring to E. coli.

Using these data we build a network model for
E. coli molecular interactions. The nodes of this network are genes; genes g1 and g2 that code for proteins
p1 and p2 are linked if:
p1 and p2 interact physically, g1 regulates the transcription of gene g2, or a product generated by

a reaction catalyzed by p1 is consumed in a reaction catalyzed by p2 (we excluded from this analysis
the most frequently used compounds, such as ATP, NAD, H2O, etc.).
Following this procedure we obtained a network with 1508 genes and 37,636 interactions.

By measuring the degree connectivity and the clustering coefficient we
characterized the resulting graph as an hierarchical free-scale network. This is a clear indication
that the same kind of modularity found on metabolic networks occurs also when we consider a more complete
view of the cellular network.

An important biological question is the determination of gene essentiality, we can address this question
from both a experimental and theoretical perspective. Our network can be used as a tool to predict
gene essentiality using topological information. Using experimental data from PEC database
(http://www.shigen.nig.ac.jp/ecoli/pec/) and the J48, a decision tree artificial intelligence algorithm we
proposed a set of criteria to classify a gene as essential or non-essential. The algorithm uses as input
the following information: the number of network arcs originating from protein-protein, metabolic and
regulatory interactions. The proposed algorithm has a recall of 85.7% for essential genes.