Semana de Engenharia 2010
Guimarães, 11 a 15 de Outubro

COMPUTATIONAL TOOLS FOR DATA INTEGRATION AND REGULATORY NETWORK INFERENCE IN SYSTEMS BIOLOGY

José P. Pinto
CCTC and CEB/IBB
Campus Gualtar, Univ. Minho
Braga, Portugal
tel: +351253604421
josepedr@di.uminho.pt

Isabel Rocha
CEB/IBB
Campus Gualtar, Univ. Minho
Braga, Portugal
tel: +351253604408
irocha@deb.uminho.pt

Miguel Rocha
CCTC
Campus Gualtar, Univ. Minho
Braga, Portugal
tel: +351253604435
mrocha@di.uminho.pt

Over the last few years, the field of Metabolic Engineering (ME), dealing with the design of strains with enhanced production capabilities of desired compounds, has grown considerably given the growing demands of the white Biotechnology industry (Stephanopoulos et al. 1998). In this area, there have been recent important advances in the development of genome-scale mathematical models of metabolism for a growing number of microorganisms. These have been used for phenotype simulation (using methods such as Flux Balance Analysis) (Edwards et al. 2002)) and also to address strain optimization tasks (Rocha et al. 2008).

To make these methods usable by a larger audience, recently the authors' research group has proposed OptFlux (http://www.optflux.org) (Rocha et al. 2010), an open-source software platform that uses mathematical models to simulate and optimize the behavior of microorganisms. Its main objective is to be a reference platform for research in ME. It incorporates strain optimization methods, allowing to find in silico mutant strains of microorganisms with enhanced capabilities, regarding an user defined objective function, involving the production of a compound of industrial interest. It also enables other functionalities such as several phenotype simulation methods, both for wild type and mutant strains (e.g. Flux Balance Analysis), Metabolic Flux Analysis, flux variability analysis and Elementary Modes analysis. All these features are freely provided in a user-friendly environment that can also be easily extended, given its modular plug-in based architecture.

A distinct approach to the analysis of metabolism takes stoichiometric models, which are composed by the metabolites, reactions and their stoichiometry and reversibility, and represents these biological systems as networks or graphs. This path has led to the discovery that metabolic networks share a similar architecture with other complex networks, indicating that similar laws may govern most complex networks in nature. This also allows the expertise from large and well-mapped non-biological systems to be used to characterize the intricate interwoven relationships that govern cellular functions (Barabási and Oltvai 2004).

As a result from this effort, and among many other results, metabolic networks have been characterized as highly connected networks that exhibit scale-free degree distributions (Jeong et al 2000), display the small-world property (Fell and Wagner) and exhibit a high degree of modularity (Ravasz et al 2002).

Although some interesting results have been obtained both by the model-based and network-based methods aforementioned, these strategies have largely remained independent. The area of pathway analysis, including the calculation and analysis of Elementary Flux Modes (Schuster et al 2000) provides a bridge between network topology and metabolic models, but given its complexity, it cannot be applied to genome-scale models.

In this work, we aim to provide software tools to bridge this gap by providing within the OptFlux ME platform a number of tools that allow the analysis of the topological features of networks created from the same metabolic models that can be used to perform phenotype simulation and strain optimization. Given OptFlux's plug-in based architecture, which allows the extension of its functionality through the inclusion of plug-ins, the logical way to add network creation and analysis capabilities was though the creation of a plug-in, named TopologyNetworkAnalyser for OptFlux, or briefly TNA4OptFlux.

Currently three variants of metabolic networks are supported by TNA4OptFlux: (i) a directed graph composed by two types of nodes, metabolites and reactions, where edges represent the consumption or production of metabolites by reactions; (ii) a directed

graph where nodes represent the metabolites and the edges represent the reactions; (iii) a directed graph where nodes represent reactions and edges stand for metabolites that are shared by the reactions.

TNA4OptFlux's tools for topologic analysis can be divided into three types:

1-Node degrees analysis: probably the most basic of all the network metrics, the degree of a vertex is simply the number of edges that connect with it. This metric is used in many different operations of graph analysis and in the calculation of more complex metrics. TNA4OptFlux automatically calculates the degree of all nodes in a network (discriminating between in and out degree) and can be used to calculate the degree distribution.

2-Shortest paths analysis: Besides calculating the shortest path between a selected pair of nodes TNA4OptFlux also supports some shortest paths related metrics: (i) the average path distance from a node to all the nodes it is connected to, (ii) the average path distance to get to a node, (iii) it can identify all the nodes that are connected to a selected node (iv) it can be used to calculate the shortest path spectrum and the mean path length of the network.

3- Clustering: The plug-in also allows the user to calculate the clustering coefficient for all nodes in the network and ranks the nodes according to this criterion. Also, considering the whole network, the user can calculate the average clustering coefficient and also use the function ACC(k) that gives the average clustering coefficient for all nodes with degree equal to k.

Beside the topology analysis TNA4OptFlux also contains some ranking algorithms these are metrics that classify all the nodes in the network according with their properties, the current version of the plug-in allows the nodes to be ranked according to three different algorithms: betweenness centrality, closeness centrality and HITS.

There are some situations when it may be useful to separate part of the network and analyze that sub-network as an independent graph. With that in mind, the possibility of creating sub-networks was included in TNA4OptFlux.

TNA4OptFlux's filters can be based in:

1. Degree.
2. Manually selected nodes.
3. Ranker values.
4. Removal of external metabolites and transport reactions.
5. OptFlux simulation results.

At the moment the filters are the only functionality that uses the results of the topological analysis, since the user can define a filter that removes reaction nodes based in their flux value in a selected OptFlux simulation. This allows the comparison among several different results obtained for different environmental conditions (e.g. different carbon sources, aerobic/ anaerobic conditions) or genetic modifications (e.g. gene deletions). This option allows to filter the original network, creating networks where only the reactions active in these simulations are active (flux higher than a user-defined value).

References

Barabási, AL and Oltvai, Z. N. 2004. Network biology: understanding the cell's functional organization. *Nature Reviews*, 5:101-114, Feb 2004.

Edwards JS, Covert M, Palsson B. Metabolic modelling of microbes: the flux-balance approach. *Environmental Microbiology*, 4:133-140, 2002.

Fell DA, Wagner A. The small world of metabolites. *Nature* Biotech 18(11):1121-1122.

Jeong H., Tombor B., Albert R., Oltvai Z., Barabasi AL, The large-scale organization of metabolic networks. *Nature* 407:651-654, 2000.

Ravasz E., Somera AL, Mongru DA, Oltvai ZN, Barabasi AL. Hierarchical organization of modularity in metabolic networks. *Science* 297:1551-1555, 2002.

Rocha I, Maia P, Evangelista P, Vilaça P, Soares S, Pinto JP, Nielsen J, Patil KR, Ferreira EC, Rocha M. OptFlux: an open-source software platform for in silico metabolic engineering. BMC Systems Biology, in press, 2010.

Rocha M, Maia P, Mendes R, Pinto JP, Ferreira EC, Nielsen J, Patil KR, Rocha I: Natural computation meta-heuristics for the *in silico* optimization of microbial strains. *BMC Bioinformatics*, 9:499, 2008.

Schuster S, Fell DA, Dandekar T: A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nature Biotechnology 2000. 18:326-332.

Stephanopoulos G, Aristidou A, Nielsen J. Metabolic engineering. San Diego: Academic Press; 1998.

AUTHOR BIOGRAPHY

JOSÉ PEDRO PINTO, graduated from LESI in the University of Minho, later he took a Post graduation in Data Analysis Systems, currently he is a PhD student in the same university.