International Conference in Systems Biology 2003
Extended Abstract
September 2003

Computational analysis of metabolic modules and pathways in the E.coli metabolic network

Victor Spirin
Harvard-MIT Division of Health Sciences and Technology
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139, USA
spirin@mit.edu

M. S. Gelfand
Integrated Genomics Moscow
Moscow, Russia

Leonid Mirny
Harvard-MIT Division of Health Sciences and Technology
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139, USA
leonid@mit.edu

Construction of genome-scale models of metabolic networks is an important problem in systems biology. Most of the studies of metabolic networks have focused of steady state flows in the native and mutated networks. However, to understand cellular adaptation and development, regulation of metabolism have to be taken into account. The approach of system biology would be to consider metabolism in a broad genomic context.

Here we undertake and in-depth exploration of *E.coli* metabolism by considering the metabolic network together with the network of gene regulation and co-regulation. *E.coli* has been selected for this study because, (1) it has a very well mapped metabolic network [2, 4], and (2) numerous close organisms have been sequenced allowing accurate prediction of co-regulation from chromosomal location and co-inheritance [6, 3]. We integrated metabolic network with the genomically predicted gene co-regulation. Co-regulation can also be deduced from expression profiling, such data are not available for *E.coli* in large amounts and may not be as accurate as genomic prediction that are based of dozens of complete bacterial genomes.

We first map the network on a graph in which vertices represent reactions and two types of edges: one that connect reactions sharing a metabolite (neighbors on the metabolic network), and the another that connect pairs of reactions that are catalyzed by co-regulated enzymes [6, 8] (co-regulated reactions). Our goals were to explore regulation of known pathways, reveal co-regulated pathways and modules, and understand principles of regulation of metabolism.

To achieve these goals we analyzed the metabolic network on three scales. *Macroscale* analysis focuses on general properties of the network such as the correlation between pathway distance and enzyme co-regulation. *Mesoscale* analysis focuses on regulation within and between known metabolic pathways, as well as search for clusters of coregulated reactions with short metabolic distance, novel coregulated modules. The key idea is that a metabolic module or pathway has many edges of both types (see above) between its enzymes and hence constitutes a highly-connected subgraph of interrelated and co-regulated metabolic reactions. Finally, *microscale* analysis focuses on co-regulated motifs of few enzymes that have a particular architecture of chemical reactions.

We use novel algorithms to search for highly-connected subgraphs in biological networks. These techniques combine exact enumeration of cliques, similar to that applied in [7], Super-paramagnetic Clustering — an algorithm developed by Domany and coworkers to cluster objects in a non-metric space of an arbitrary dimension [1], and novel Monte-Carlo optimization technique similar to that described in [7] to identify a highly-connected subgraphs in an arbitrary network.

We obtained the following main results.

(i) On *macroscale*, our analysis found that reactions short pathway distance apart are more likely to be catalyzed by the same enzymes or by co-regulated enzymes. In fact, neighboring reactions are 13 time more likely to be co-regulated than expected at random. Tendency to be co-regulated decays rapidly with no significant abundance of co-regulated reactions separated by more than 3 intermediate reactions (Figure 1).

Another important result concerns enzymes that catalyze the same reaction (as either isoenzymes, or as co-enzymes). 86% of pairs of enzymes that catalyze the same reaction are co-regulated, as many of them are subunits of large enzymatic complexes. This result strongly supports "the balance hypothesis" which suggests that imbalance in the concentration of proteins that constitute a single complex is deleterious [5]

(ii) On *mesoscale* level, we found that most of co-regulation occurs between the enzymes that are involved in the same known metabolic pathway. This finding is very much in line with biochemical intuition about metabolism regulation. However, we also find that several cases when a single metabolic pathways splits into two or more sub-pathways with strong co-regulation within, but not between them.

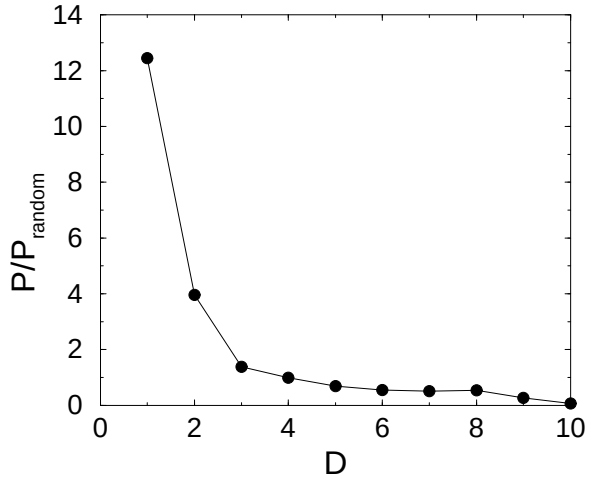

Figure 1: Probability that two reactions are catalyzed by coregulated enzymes as a function metabolic distance.

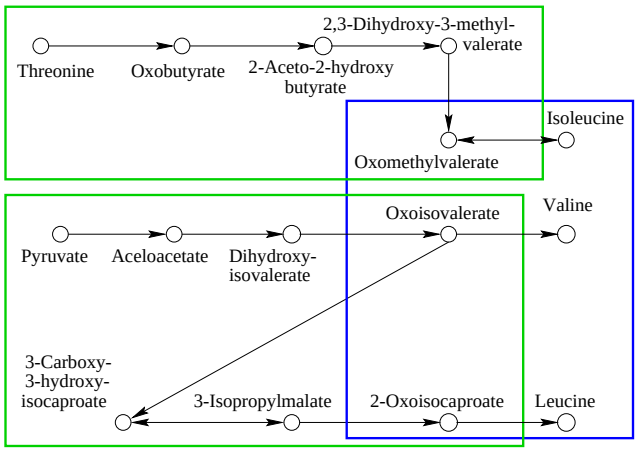

Figure 2: Coregulated clusters in Branched Chain Amino Acid biosynthesis.

This result shows that in many cases a cell does not regulate the whole metabolic pathway as a single entity, rather as a set of smaller sub-pathways. The one interesting example of this sort is biosynthesis of branched chain amino acids: Ile, Leu, and Val (Figure 2). Each of the three pathways that syntheze a precursor is stringly co-regulated. Surprisingly, the last step in the amino acid synthesis, is not co-regulated with the rest of the pathway. Instead, last steps for the Ile, Leu and Val biosynthesis are co-regulated with each other, suggesting somewhat different mode of regulation.

We also found several cases when two distinct biochemical pathways are strongly co-regulated. For example, Branched Chain Amino Acid Biosynthesis, Aromatic Amino Acids, Histidine Biosynthesis, and Threonine and Lysine Biosynthesis form a hypercluster of coregulated pathways. The number of links (pairs of coregulated enzymes) between any pair of these pathways is at least 6 to 10 times larger than would be expected if enzymes were coregulated at random.

Using our search algorithms we identified several non-overlapping metabolic modules. Some of them are parts of known pathways, while other consist of several groups of reactions from various pathways. The reactions within these modules are highly interconnected on the graph in terms of production of identical metabolites and high coregulation between their enzymes. So our definition of a pathways and cell's regulatory view of pathways may be quite different.

(iii) Finally, the *microscale* analysis focuses on motifs of 2-3 reactions in an attempt to find "architectures" of such motifs that are frequently co-regulated.

One interesting example concerns two irreversible reactions producing or utilizing a common metabolite. Such two reactions require the existence of a third reaction to serve as a sink or a source for this metabolite. Our results indicate that if the two irreversible reactions are co-regulated, there is a very high probability that either of them is co-regulated with all source or sink reactions which respectively produce or utilize the common metabolite.

In summary, we showed that regulation of cellular metabolism have several very distinct features. While some of these features are very intuitive, others are less intuintive. These features suggest new problems for biochemists aiming to understand regulation of metabolism.

1. REFERENCES

[1] M. Blatt, S. Wiseman, and E. Domany. Superparamagnetic clustering of data. *Physical Review Letters*, 76:3251–3254, 1996.

[2] P. Karp, M. Riley, S. M. Paley, and A. Pellegrini-Toole. The metacyc database. *Nucleic Acids Res.*, 30:59, 2002.

[3] A. Osterman and R. Overbeek. Missing genes in metabolic pathways: a comparative genomic approach. *Current Opinion in Chemical Biology*, 7:238–251, 2003.

[4] *E.coli* metabolic reactions database version 1.01. http://gcrg.ucsd.edu. UCSD Genomics Circuits Research Group.

[5] B. Papp, C. Pal, and L. D. Hurst. Dosage sensitivity and the evolution of gene families in yeast. *Nature*, 424:194–197, 2003.

[6] B. Snel, G. Lehmann, P. Bork, and M. A. Huynen. String: a web server to retrieve and display the repeatedly occurring neighborhood of a gene. *Nucleic Acids Res.*, 28:3442–3444, 2000.

[7] V. Spirin and L. A. Mirny. Protein complexes and functional modules in molecular networks. *To appear in Proc. Nat. Acad. Sci*, 2003.

[8] C. von Mering, M. A. Huynen, D. Jaeggi, S. Schmidt, P. Bork, and B. Snel. String: a database of predicted functional associations between proteins. *Nucleic Acids Res.*, 31:258–261, 2003.