

NEWS AND VIEWS

Bringing metabolomics data into the forefront of systems biology

Steve Van Dien and Christophe H Schilling

Genomatica, San Diego, CA, USA

Molecular Systems Biology 20 June 2006; doi:10.1038/msb4100078

Systems Biology has been defined in various ways since the term was first used less than two decades ago, and the boundary between what does and does not constitute systems biology is not likely to be defined anytime soon. Two aspects that appear to be irrefutable are that systems biology involves the use of mathematical models and high-throughput ‘omics’ data. A model should use the experimental results to understand the complex relationships and interactions among the various parts of the system, not merely organize and catalog the data into arbitrary classifications. Ideally, one would develop a model starting from the kinetic equations governing each molecular step in all aspects of the cell’s existence: signaling, metabolism, growth, etc. However, owing to a lack of comprehensive knowledge, data, and computational power, it is not possible to formulate or solve this level of description on a genome scale.

One mathematical framework that has gained wide acceptance in the systems biology community, particularly for the study of metabolism, is the general approach of constraint-based modeling (Price *et al*, 2004). Instead of attempting to calculate an exact phenotypic ‘solution’, physico-chemical constraints are imposed on a metabolic network to determine a feasible solution space in which the cell must operate. In this way, models and experimental data can be more easily reconciled and studied on a whole-cell or genome-scale level. Experimental data sets can first be examined for their consistency against the underlying biology and chemistry represented in the models. The data sets can then be further analyzed in the context of models to improve our understanding of metabolism and become a driver of the biological discovery process. In an article currently published in *Molecular Systems Biology*, Heinemann and co-workers (Kümmel *et al*, 2006) illustrate both these concepts through their successful utilization of high-throughput metabolomics data within a constraint-based framework, in conjunction with the second law of thermodynamics, to examine consistency of the data with the models and also to provide an insight into cell physiology.

Two important measures of cell physiology are the fluxes through metabolic pathways and the intracellular metabolite concentrations. As high-throughput data sets move down the hierarchy from transcriptome and proteome to metabolome (cell metabolite concentrations) and fluxome (intracellular reaction rates), we are getting more direct measurements of the actual metabolic phenotype. Metabolomics data sets are currently being generated by a number of industrial and

academic research groups, and appear to be improving rapidly in both the number of metabolites that can be identified and in the quantitative accuracy of intracellular concentration measurements (Goodacre *et al*, 2004; van der Werf *et al*, 2005). Thus far, computational study on metabolomics data has been restricted to statistical techniques such as principal components analysis to look at trends between different data sets. Such work has proven useful in discovering biomarkers and identifying strains (Vaidyaraman and Goodacre 2003), but provides minimal insight into the underlying biology or the means to modulate it for therapeutic or industrial purposes. In order to incorporate the knowledge of the underlying metabolic system into the analysis of the data requires novel computational approaches and methods that go beyond statistical techniques.

Kümmel *et al* (2006) have made one of the first significant steps in bridging the gap between data and true biological understanding by developing a methodology for analyzing quantitative metabolomics data in the context of the entire metabolic network. Their paper introduces a method called network-embedded thermodynamic (NET) analysis for the model-based interpretation of quantitative metabolite data. NET analysis uses the Gibbs energies of formation, known reaction directions, and the second law of thermodynamics to calculate feasible concentration ranges of all metabolites. ΔG of each reaction is constrained by the thermodynamic interdependencies of all reactions in the network, and thus metabolite concentrations have to be feasible not only in view of one specific reaction but also with respect to the entire network (Kümmel *et al*, 2006). The first application is to examine the consistency of metabolic data sets; that is, do the concentrations all fall within the thermodynamically feasible ranges. Of seven published *Escherichia coli* data sets, only four were determined to be consistent, which underscores the need for quality control of omics data sets before use in modeling efforts. Next, the authors test the ability of NET analysis to predict unmeasured metabolite concentrations. Although the state-of-the-art for metabolomics is advancing, there are still a number of key central metabolites that can only be detected as pooled concentrations. NET analysis can be used to resolve these pools, and in some cases specify rather narrow concentration ranges, for example, 0.56–0.70 mM for DHAP, which is pooled with G3P in the experimental data set. Finally, results of the analysis are used to identify putative regulatory sites. Recent work in the area of metabolic control analysis (MCA) indicates that reactions operating far from equilibrium

are more likely to impose flux control, and thus more likely to be regulated by the cell (Crabtree *et al.*, 1997; Wang *et al.*, 2004). Not only did they identify reactions previously known to be regulated, pyruvate kinase and phosphofructokinase, but they also found the cytoplasmic transhydrogenase (*udh*) to be operating far from equilibrium. Regulation of the transhydrogenase makes sense for physiological reasons, because any shift in the NAD(H) pool ratio could be detrimental to the cell by shifting the direction of key enzymes.

In summary, the paper by Kümmel *et al.* represents the first compelling example of using quantitative metabolomics data to gain true biological insight into cell physiology and metabolism. The beauty of the NET approach lies in its simplicity and scalability. As it relies only on thermodynamic principles and network stoichiometry, rather than sophisticated model structures, it can be applied in the same manner regardless of how limited the metabolomics data set is. Furthermore, it can be integrated directly with existing constraint-based modeling approaches in a quantitative sense. Constraints derived from metabolomics data can sharpen model predictions and thus improve the value of *in silico* methods for applications such as metabolic engineering. Widespread use of this technique would significantly advance the field of systems biology by bridging the gap between models and metabolomics data. Expect this approach to usher in a series of new computational strategies that will allow us to extract rich information from metabolomics data that will

exceed that which has thus far been accomplished with transcriptomic and proteomic data sets.

References

- Crabtree B, Newsholme EA, Reppas NB (1997) Principles of regulation and control in biochemistry: a pragmatic, flux-oriented approach. In *Handbook of Physiology*, Hoffman JF, Jamieson JD (eds) Vol. Section 14: Cell Physiology, pp 117–180. New York, USA: Oxford University Press
- Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB (2004) Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol* **22**: 245–252
- Kümmel A, Panke S, Heinemann M (2006) Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Mol Syst Biol* doi:10.1038/msb4100074
- Price ND, Reed JL, Palsson BO (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol* **2**: 886–897
- Vaidyaraman S, Goodacre R (2003) Metabolome and proteome profiling for microbial characterization. In: *Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function Analysis*, Harrigan GG, Goodacre R (eds) pp 9–38. Boston: Kluwer
- Wang LQ, Biro I, Hatzimanikatis V (2004) Metabolic control analysis under uncertainty: framework development and case studies. *Biophys J* **87**: 3750–3763
- van der Werf MJ, Jellema RH, Hankemeier T (2005) Microbial metabolomics: replacing trial and error by the unbiased selection and ranking of targets. *J Ind Microbiol Biotechnol* **32**: 234–252