molecular systems biology

NEWS AND VIEWS

Escalating model sizes and complexities call for standardized forms of representation

Michael Hucka[1] and Andrew Finney[2]

[1] Control and Dynamical Systems, California Institute of Technology, Pasadena, CA, USA and [2] Biocomputation Research Group, Science and Technology Research Institute, University of Hertfordshire, Hatfield, UK

The recent work of Kitano *et al* on a comprehensive EGFR Pathway Map (*Mol Systems Biol*, this issue) represents a tremendous amount of intellectual effort. The scale of the model is breathtaking. No doubt some readers will assail the effort on the grounds that models of this size and complexity are difficult to verify, but while this may be true for today's methods, it is an unhelpful criticism. The inescapable reality in systems biology is that models (that is to say, hypotheses cast in a computational form) will continue to grow in size, complexity, and scope. Rather than grouse, we should be thinking about how to develop ways of analyzing and verifying models of this scale. We also need to improve our methods of sharing and understanding each other's work in order to facilitate the iterative processes of review and refinement that are fundamental to modeling.

An important first step is to reach agreement on how to communicate models. Kitano *et al* have been developing a new visual notation for diagrams (Figures 1 and 3 in the article). It represents an attempt to add more rigor and consistency to the usually *ad hoc* diagrams that often accompany published research on biological networks. The expanded visual vocabulary of their iconography allows for greater expressiveness while maintaining compactness, a necessary feature as we work with increasingly larger networks. The real payoff will come when more people and software adopt such a common visual notation and it becomes as familiar to them as circuit schematics are to computer engineers. When researchers are saved the time and effort required to familiarize themselves with different notations, they can spend more time thinking about the underlying networks being depicted.

A complement to visual notations is a computational representation that allows software tools to process, analyze, store, and communicate the underlying model. The Systems Biology Markup Language (SBML)—which also owes its genesis to Kitano—addresses this need. SBML is an open format for representing computational models of biological networks (Finney and Hucka, 2003; Hucka *et al*, 2003, 2004). By supporting SBML as a format for reading and writing models, different software tools (including editing programs such as CellDesigner, simulation programs, databases, and other systems) can directly communicate and store the same computable representation of those models. Not only does this reduce errors due to human translation from one format to another, but it also permits models to be reused more effectively, built upon more directly, and published more precisely.

To date, we know of over 80 software tools worldwide supporting SBML, including several commercial packages. The wealth of software now available is a boon to researchers, who can mix and match tools to suit their research needs yet still be able to exchange their models easily between the tools. The surprisingly fast take-up of SBML has recently spread to publications, with *Molecular Systems Biology* spearheading the trend of accepting computational models in SBML format as supplementary information accompanying published articles. In addition, on the near-term horizon is the ongoing development of centralized, curated, public databases that accept and store published computational models in SBML format. JWS Online's (Olivier and Snoep, 2004) and SigPath's (Campagne *et al*, 2004) support for SBML, and the recent BioModels.net initiative and BioModels Database (Donizelli and Le Novère, 2005; Leslie, 2005), are already steps in this direction.

Standardizing on a common format such as SBML is essential for being able to move forward with large-scale modeling efforts such as Kitano's. It removes an impediment to sharing results and permits other researchers to start with an unambiguous representation of the network, examine it carefully, propose precise corrections and extensions, and apply new techniques and approaches—in short, to do better science. Moreover, as infrastructure for working with standardized formats becomes commoditized and widely available, the cost of experimenting with creative new tools decreases. As a consequence, developers are encouraged to differentiate their products on the basis of innovation and performance.

These resources take time to build, and it is about time things got started. The use of computational modeling is clearly increasing in all areas of biology, from analyzing and extracting understanding from the vast quantities of data saturating researchers today, to designing biological circuits (Church, 2005). It does not take prescience to see that infrastructure such as SBML, databases, and more powerful analysis tools are needed to support continued progress in systems biology.

References

Campagne F, Neves S, Chang CW, Skrabanek L, Ram PT, Iyengar R, Weinstein H (2004) Quantitative information management for the biochemical computation of cellular networks. *Science STKE* **248:** PL11

Church GM (2005) From systems biology to synthetic biology. *Mol Systems Biol* 29 March 2005; doi:10.1038/msb4100007

Donizelli M, Le Novère N (2005) BioModels Database. Available on the World Wide Web at http://www.ebi.ac.uk/biomodels/

Finney A, Hucka M (2003) Systems biology markup language: level 2 and beyond. *Biochem Soc Trans* **31:** 1472–1473

Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmeyr J-H, Hunter PJ, Juty NS, Kasberger JL, Kremling A, Kummer U, Le Novere N, Loew LM, Lucio D, Mendes P, Minch E, Mjolsness ED, Nakayama Y, Nelson MR, Nielsen PF, Sakurada T, Schaff JC, Shapiro BE, Shimizu TS, Spence HD, Stelling J, Takahashi K, Tomita M, Wagner J, Wang J (2003) The Systems Biology Markup Language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19:** 524–531

Hucka M, Finney A, Bornstein BJ, Keating SM, Shapiro BE, Matthews J, Kovitz BL, Schilstra MJ, Funahashi A, Doyle JC, Kitano H (2004) Evolving a lingua franca and associated software infrastructure for computational systems biology: The Systems Biology Markup Language (SBML) project. *Systems Biol* **1:** 41–53

Leslie M (2005) NetWatch. *Science* **308:** 475

Olivier BG, Snoep JJ (2004) Web-based kinetic modelling using JWS Online. *Bioinformatics* **20:** 2143–2144