Available online at www.sciencedirect.com
ScienceDirect
Journal of Theoretical Biology 252 (2008) 538–543
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

The markup is the model: Reasoning about systems biology models in the Semantic Web era

Douglas B. Kell[a,c,*], Pedro Mendes[b,c,d]

[a]School of Chemistry, The University of Manchester, 131 Princess Street, Manchester M1 7DN, UK
[b]School of Computer Science, The University of Manchester, 131 Princess Street, Manchester M1 7DN, UK
[c]The Manchester Centre for Integrative Systems Biology, Manchester Interdisciplinary Biocentre, The University of Manchester, 131 Princess Street, Manchester M1 7DN, UK
[d]Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Washington Street 0477, VA 24061, USA

Available online 3 December 2007

Dedicated to the memory of Reinhart Heinrich (1946–2006).

Abstract

Metabolic control analysis, co-invented by Reinhart Heinrich, is a formalism for the analysis of biochemical networks, and is a highly important intellectual forerunner of modern systems biology. Exchanging ideas and exchanging models are part of the international activities of science and scientists, and the Systems Biology Markup Language (SBML) allows one to perform the latter with great facility. Encoding such models in SBML allows their distributed analysis using loosely coupled workflows, and with the advent of the Internet the various software modules that one might use to analyze biochemical models can reside on entirely different computers and even on different continents. Optimization is at the core of many scientific and biotechnological activities, and Reinhart made many major contributions in this area, stimulating our own activities in the use of the methods of evolutionary computing for optimization.

Keywords: Systems biology; Metabolic control analysis; Reinhart Heinrich

1. Introduction

For most of us, including the present authors, Reinhart Heinrich's great contribution was the independent co-invention of metabolic control analysis (MCA) (Heinrich and Rapoport, 1973, 1974), a crucial forerunner of the modern systems biology in which we recognize in particular that systems are logical graphs or networks, with emergent properties that depend in a highly non-linear manner on the 'local' properties of each of the elements of the network. Indeed, MCA is a highly principled version of sensitivity analysis, that has subsequently developed quite independently in other fields (e.g. Lüdtke et al., 2007; Rabitz and Aliş, 1999; Saltelli et al., 2004), a recent development being the fusing of elasticity analysis with flux balance analysis (Smallbone et al., 2007). Although D.B.K. had been traveling to East Berlin—and indeed the Humboldt University—for a couple of years previously (visiting the person who is now his wife), it was at the Il Ciocco meeting in 1989 (Cornish-Bowden and Cárdenas, 1990) that he first met Reinhart, and enjoyed many profound conversations with him over 17 years, including at the Yokohama International Conference on Systems Biology (ICSB) meeting when they discussed science together for the last time. In 1989, of course, there was no Web, and few or no common metabolic modeling packages. At Il Ciocco, P.M., then a fresh graduate, also met Reinhart for the first time, and demonstrated an early

Abbreviations: ICSB, International Conference on Systems Biology; MCA, metabolic control analysis; OWL, Web Ontology Language; RDF, Resource Description Framework; SBML, Systems Biology Markup Language; WSDL, Web Services Description Language; XML, eXtensible Markup Language.

*Corresponding author. School of Chemistry, The University of Manchester, 131 Princess Street, Manchester M1 7DN, UK.
Tel.: +44 161 306 4492.
E-mail address: dbk@manchester.ac.uk (D.B. Kell).
URL: http://www.dbkgroup.org (D.B. Kell).

0022-5193/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2007.10.023

version of Gepasi (Mendes, 1993, 1997, 2001; Mendes and Kell, 1998) on an Amstrad computer with no hard disk (Mendes, 1990). P.M. still remembers vividly the excitement that Reinhart displayed for being West of the Iron Curtain; little did we know that this curtain was about to fall within a year, by the time Reinhart organized a scientific meeting in Holzhau. It was also at the ICSB meeting in Yokohama that P.M. last talked with Reinhart. What meetings like Il Ciocco represent are opportunities for the global network of scientists to interact efficiently, and it is Reinhart's huge contribution to the world of metabolic network modeling and optimization, within the concept of the global family of science and scientists, that we would wish to acknowledge with pleasure.

Interactions between scientists concerned with modeling and systems biology involve the exchange and analysis of such models, but models could and can be created and represented in many ways that would not be comprehensible without access to all the software (and often the very same computer) that was used to create them. In other words, there were certainly no standards of interoperability between such programs, and it was not until 10 years later in Visegrad (Cornish-Bowden and Cárdenas, 2000; Kell and Mendes, 2000) (http://bip.cnrs-mrs.fr/bip10/meet99.htm), also attended by Reinhart, that such community discussion began in earnest, culminating so far in the massively important Systems Biology Markup Language (SBML) (Hucka et al., 2003) (www.sbml.org). More specifically, a mailing list entitled MMFF (metabolic model file format) was set up by one of us (P.M.) in April 1999, just after Visegrad. The members of this list created a draft specification of a common metabolic file format (PMB; portable metabolic binary format). Most of the subscribers to this list joined the SBML effort in April 2000. The MDL (model definition language) specification draft, a predecessor of SBML, was created in August 2000, as was the sysbio mailing list. By September 2000 (in the first revision of this MDL document), SBML is referred to as such among this community. At all events, it is this last aspect, the importance of a lingua franca (Kell, 2006b), on which we wish first to focus here.

1.1. The medium is the message

Marshall McLuhan's famous slogan (http://en.wikipedia.org/wiki/The_Medium_is_the_Massage) (e.g. (McLuhan and Fiore, 1971)) (that predates the original ARPANET by 2 years and the widespread civilian use of the Internet by 20) not only made widespread the concept of the global village but is intended to convey the idea that the generic form of a medium is more important than its 'meaning' or 'content'. Consistent with this, it appears (Kell, 2006a) that the initial difficulties of interoperability between modern software systems are much more about data *structures* (syntax) than about their *meaning* (semantics) (Siepel et al., 2001; Wilkinson et al., 2005), although these can and will be dealt with using separate but integrable technologies. It is also important to recognize that a model is 'just that'; it is a representation of reality (Kell and Knowles, 2006), just as is Magritte's famous painting of a pipe (http://en.wikipedia.org/wiki/The_Treachery_Of_Images).

SBML, now at level 2 version 3, allows one to describe accurately the structure of most models in which there is not a highly complex spatial segregation within a compartment nor combinatorial variants of, e.g., protein phosphorylation states. SBML is an XML (eXtensible Markup Language) that allows one to describe facts in a principled way by marking them up in a standardized 'language'. As well as allowing the exchange of models, SBML provides a very convenient means for storing them in model databases such as Biomodels (www.biomodels.net) (Le Novère et al., 2006). The present version of SBML also now allows one to mark up the models with rudimentary semantic information. In a similar way, the Semantic Web (e.g. Alesso and Smith, 2006; Baker and Cheung, 2007; Berners-Lee and Hendler, 2001; Berners-Lee et al., 2006; Fensel et al., 2003; Stevens et al., 2006; Taylor et al., 2006), will allow the semantic annotation of such XML documents using the Resource Description Framework (RDF) and Web Ontology Language (OWL), while the Web 2.0 concept (http://en.wikipedia.org/wiki/Web_2) relates to the general increase in social or community involvement in developing a particular domain.

Although it is plausible that RDF will add significantly to XML (Wang et al., 2005) (and SBML L2V3 integrates it now), the great power of SBML (Kell, 2006a, b) is its ability to lie at the focus of a distributed computing model in which loosely coupled software programs allow its integration in pipelines or workflows constructed via a distributed, service-oriented architecture (Foster, 2005; Hey and Trefethen, 2005). This loose integration reflects the thinking behind the Systems Biology Workbench (Sauro et al., 2003), but we consider (Kell, 2006a, b, 2007; Kell and Paton, 2007) that more global distributed workflow environments such as Taverna (Hull et al., 2006; Oinn et al., 2004, 2006, 2007) (www.taverna.sourceforge.net) are likely to provide the environment of choice. An obvious advantage of Taverna is that by using widely adopted interoperability standards (WSDL and Web Services), it can readily make available a large number of diverse services that can be incorporated into systems biology workflows. As usual, the cost of adoption of standards is largely paid for by the benefits that come from compatibility with other software. There are a great many modules that perform useful activities and that can beneficially write to SBML (see, e.g. Kell, 2006b, 2007 and Fig. 1). One example might be Web Services to retrieve bibliographic references that could be useful to annotate models (Ananiadou and McNaught, 2006; Ananiadou et al., 2006), while another whole area covers the optimization, in some sense or senses, of the models themselves (e.g. Mendes and Kell, 1998; Moles et al., 2003; Rodriguez-Fernandez et al., 2006).

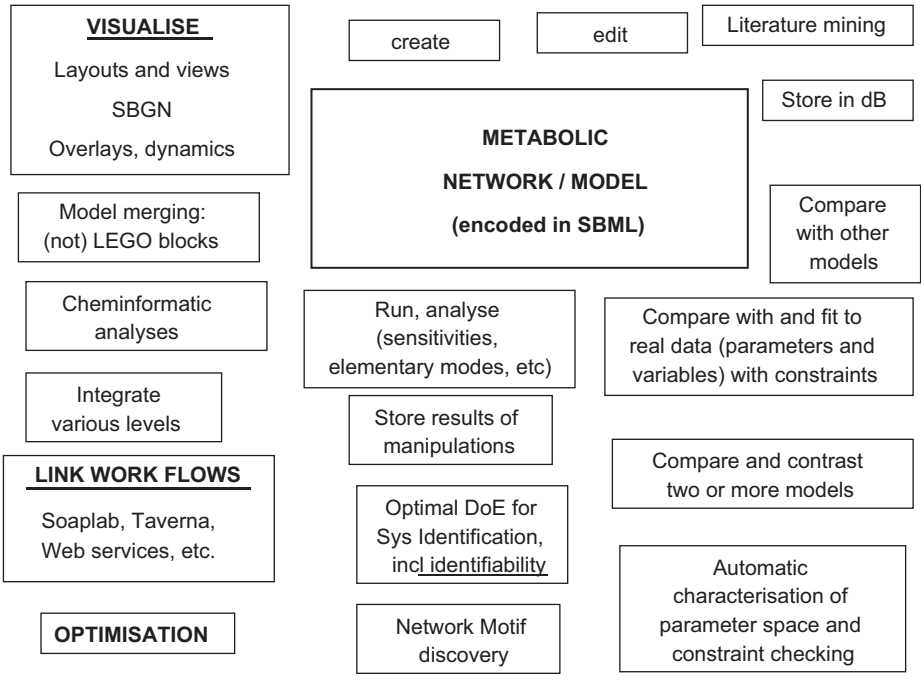

Fig. 1. Some of the modules or activities that one might wish to perform on a systems biology model encoded in SBML, and that may be stitched together to form workflows.

1.2. *Optimization of biochemical networks*

Another great contribution of Reinhart was his pioneering work on the application of optimization to metabolic networks (Heinrich and Schuster, 1998). Initially this focused more on optimality principles of metabolism (Heinrich et al., 1991) in terms of parameters such as time scales (Schuster and Heinrich, 1987) metabolite (Schuster and Heinrich, 1991; Schuster et al., 1991) or enzyme (Heinrich and Klipp, 1996; Klipp and Heinrich, 1999) concentrations. This was essentially analytical optimization using devices such as Laplace transforms. Later, they used optimization algorithms to explain the structure of metabolic pathways (Ebenhoh and Heinrich, 2001, 2003; Heinrich et al., 1997; Meléndez-Hevia et al., 1997; Stephani et al., 1999; Waddell et al., 1997). In this case, they mostly used evolutionary optimization algorithms, such as genetic algorithms, to perform combinatorial optimization.

Our own work has also relied on aspects of optimization, though mostly on numerical aspects. We proposed that metabolic engineering objectives (Kell and Westerhoff, 1986; Kell et al., 1989), usually maximization of a metabolic flux or an end-product concentration, could be formalized as an objective function whose maximum is the solution of the problem (Mendes and Kell, 1998). As these objective functions are non-linear in the parameters (such as kinetic constants) that are allowed to vary, the algorithms used must be appropriate for non-linear functions. In addition, we recognized that since the objective functions are specified as a system of non-linear differential equations for which there is no known analytical solution, the equations must be integrated numerically and the optimization must also be carried out numerically. Implementation of this approach in the Gepasi software (Mendes and Kell, 1998) provided one of the first applications of numerical optimization of biochemical networks, which has since been carried on to COPASI (Hoops et al., 2006). We also recognized that the same numerical optimization algorithms could be used to minimize the distance between a model's result and a set of data, such that they could be used to fit the model to the data. This is a powerful application which is at the core of most biochemical network modeling activities. More recently, we have identified that (notwithstanding the recognition that such statements cannot be universal for all domains (Wolpert and Macready, 1997)) evolutionary algorithms are often most efficient in solving these problems (Mendes, 2001; Moles et al., 2003; Patil et al., 2005; Rodriguez-Fernandez et al., 2006), although other strategies are also emerging (e.g. Wilkinson, 2007; Wilkinson et al., 2007).

Another application of evolutionary optimization algorithms has been the use of genetic programming methods (Koza, 1992; Koza et al., 2003; Langdon, 1998) to the reverse engineering or 'system identification' of biochemical pathway *structure* (Koza et al., 2001a, b, 2003), following the earlier work of Reinhart's group with genetic algorithms (Ebenhoh and Heinrich, 2001; Stephani et al., 1999). We too have found GP to be a very effective means for optimization and data analysis, independent of prejudicial hypotheses (Kell and Oliver, 2004), as applied to metabolism, spectroscopy and metabolomics (e.g. Gilbert et al., 1997; Johnson et al., 2000; Kell, 2002a, b; Kell et al., 2001; O'Hagan et al., 2005, 2007).

1.3. Whither biochemical network modeling?

> It has been said that we always overestimate what we can do in two years and underestimate what we can do in twenty (Ball and Garwin, 1992).

As well as looking back, it is appropriate, in an article of this type, to look forward. Obvious trends include the increasing scale and accuracy of biochemical network models, including improved knowledge of the human metabolic network (Duarte et al., 2007; Ma et al., 2007), improved abilities to effect measurements of the many uncharted metabolites that still exist (Harrigan and Goodacre, 2003; O'Hagan et al., 2007), much improved methods of system identification for estimating system parameters from measured variables, the improved recognition through metabolomics of metabolic pathway changes accompanying disease progression (e.g. Dunn et al., 2007; Kenny et al., 2005; van der Greef et al., 2006), and the bringing together of metabolomics measurements and systems biology models (Kell, 2004, 2006a, b, 2007). Some areas of biochemistry, especially human metabolite and drug transporters, are woefully underrepresented in the literature. To allow computational reasoning to assist us, it is absolutely vital that we use controlled vocabularies (Spasic et al., 2007) and traceable identifiers to describe the molecules in these models (see also, e.g. Le Novère et al., 2005). Calling a molecule 'glucose' or even 'glu' conveys nothing to a computer save those strings of letters, since these terms are of themselves devoid of semantic content and another user may easily use another term for the same chemical entity. Referring to it with a ChEBI, KEGG or PubChem identifier is a start, although of existing string-based representations only InChI strings (Coles et al., 2005; Stein et al., 2003) are likely to be genuinely unambiguous and will likely supplant the very useful but proprietary SMILES strings (Weininger, 1988). Web-accessible databases, including expression profiling information (e.g. Uhlen et al. (2005) and www.proteinatlas.org/), accessed by Taverna-type workflows, will allow the incorporation of such universal and traceable identifiers into SBML models, and such models will be both distributed and available to all. Then the clear goal of the 'digital human' (e.g. Hunter, 2004; Kell, 2007), an *in silico* representation of human biochemistry and its dynamic response to drugs, will begin to allow us to understand fully human physiology and medicine, and this systems approach will also help enormously to decrease the still-terrible attrition rates seen in drug development (Kola and Landis, 2004).

It is this legacy in particular—the systems approach to biology—for which we thank and remember Reinhart Heinrich with fondness and appreciation.

Acknowledgments

D.B.K. thanks the BBSRC, EPSRC and BHF for financial support, and the Royal Society and the Wolfson Foundation for a Research Merit award. P.M.'s Chair in Manchester is supported by the UK EPSRC. This is a contribution from the Manchester Centre for Integrative Systems Biology (www.mcisb.org/).

References

Alesso, H.P., Smith, C.F., 2006. Thinking on the Web: Berners-Lee, Gödel and Turing. Wiley, New York.

Ananiadou, S., McNaught, J. (Eds.), 2006. Text Mining in Biology and Biomedicine. Artech House, London.

Ananiadou, S., Kell, D.B., Tsujii, J.-I., 2006. Text Mining and its potential applications in Systems Biology. Trends Biotechnol. 24, 571–579.

Baker, C.J.O., Cheung, K.-H. (Eds.), 2007. Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences. Springer, New York.

Ball, P., Garwin, L., 1992. Science at the atomic scale. Nature 355, 761–766.

Berners-Lee, T., Hendler, J., 2001. Publishing on the semantic web. Nature 410, 1023–1024.

Berners-Lee, T., Hall, W., Hendler, J., Shadbolt, N., Weitzner, D.J., 2006. Computer science. Creating a science of the Web. Science 313, 769–771.

Coles, S.J., Day, N.E., Murray-Rust, P., Rzepa, H.S., Zhang, Y., 2005. Enhancement of the chemical semantic web through the use of InChI identifiers. Org. Biomol. Chem. 3, 1832–1834.

Cornish-Bowden, A., Cárdenas, M.L. (Eds.), 1990. Control of Metabolic Processes. Plenum Press, New York.

Cornish-Bowden, A., Cárdenas, M.L. (Eds.), 2000. Technological and Medical Implications of Metabolic Control Analysis. Kluwer Academic Publishers, Dordrecht.

Duarte, N.C., Becker, S.A., Jamshidi, N., Thiele, I., Mo, M.L., Vo, T.D., Srvivas, R., Palsson, B.Ø., 2007. Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc. Natl. Acad. Sci. 104, 1777–1782.

Dunn, W.B., Broadhurst, D.I., Sasalu, D., Buch, M., McDowell, G., Spasic, I., Ellis, D.I., Brooks, N., Kell, D.B., Neyses, L., 2007. Serum metabolomics reveals many novel metabolic markers of heart failure, including pseudouridine and 2-oxoglutarate. Metabolomics, in press.

Ebenhoh, O., Heinrich, R., 2001. Evolutionary optimization of metabolic pathways. Theoretical reconstruction of the stoichiometry of ATP and NADH producing systems. Bull. Math. Biol. 63, 21–55.

Ebenhoh, O., Heinrich, R., 2003. Stoichiometric design of metabolic networks: multifunctionality, clusters, optimization, weak and strong robustness. Bull. Math. Biol. 65, 323–357.

Fensel, D., Hendler, J., Lieberman, H., Wahlster, W. (Eds.), 2003. Spinning the Semantic Web. MIT Press, Cambridge, MA.

Foster, I., 2005. Service-oriented science. Science 308, 814–817.

Gilbert, R.J., Goodacre, R., Woodward, A.M., Kell, D.B., 1997. Genetic programming: a novel method for the quantitative analysis of pyrolysis mass spectral data. Anal. Chem. 69, 4381–4389.

Harrigan, G.G., Goodacre, R. (Eds.), 2003. Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function Analysis. Kluwer Academic Publishers, Boston.

Heinrich, R., Klipp, E., 1996. Control analysis of unbranched enzymatic chains in states of maximal activity. J. Theor. Biol. 182, 243–252.

Heinrich, R., Rapoport, T.A., 1973. Linear theory of enzymatic chains: its application for the analysis of the crossover theorem and of the glycolysis of human erythrocytes. Acta Biologica et Medica Germanica 31, 479–494.

Heinrich, R., Rapoport, T.A., 1974. A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. Eur. J. Biochem. 42, 89–95.

Heinrich, R., Schuster, S., 1998. The modelling of metabolic systems. Structure, control and optimality. Biosystems 47, 61–77.

Heinrich, R., Schuster, S., Holzhütter, H.G., 1991. Mathematical analysis of enzymatic reaction systems using optimization principles. Eur. J. Biochem. 201, 1–21.

Heinrich, R., Montero, F., Klipp, E., Waddell, T.G., Meléndez-Hevia, E., 1997. Theoretical approaches to the evolutionary optimization of glycolysis: thermodynamic and kinetic constraints. Eur. J. Biochem. 243, 191–201.

Hey, T., Trefethen, A.E., 2005. Cyberinfrastructure for e-Science. Science 308, 817–821.

Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., Kummer, U., 2006. COPASI: a COmplex PAthway SImulator. Bioinformatics 22, 3067–3074.

Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., Arkin, A.P., Bornstein, B.J., Bray, D., Cornish-Bowden, A., Cuellar, A.A., Dronov, S., Gilles, E.D., Ginkel, M., Gor, V., Goryanin, I.I., Hedley, W.J., Hodgman, T.C., Hofmeyr, J.H., Hunter, P.J., Juty, N.S., Kasberger, J.L., Kremling, A., Kummer, U., Le Novere, N., Loew, L.M., Lucio, D., Mendes, P., Minch, E., Mjolsness, E.D., Nakayama, Y., Nelson, M.R., Nielsen, P.F., Sakurada, T., Schaff, J.C., Shapiro, B.E., Shimizu, T.S., Spence, H.D., 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.

Hull, D., Wolstencroft, K., Stevens, R., Goble, C., Pocock, M.R., Li, P., Oinn, T., 2006. Taverna: a tool for building and running workflows of services. Nucleic Acids Res. 34, W729–W732.

Hunter, P.J., 2004. The IUPS Physiome Project: a framework for computational physiology. Prog. Biophys. Mol. Biol. 85, 551–569.

Johnson, H.E., Gilbert, R.J., Winson, M.K., Goodacre, R., Smith, A.R., Rowland, J.J., Hall, M.A., Kell, D.B., 2000. Explanatory analysis of the metabolome using genetic programming of simple, interpretable rules. Genetic Progr. Evolvable Machines 1, 243–258.

Kell, D.B., 2002a. Metabolomics and machine learning: explanatory analysis of complex metabolome data using genetic programming to produce simple, robust rules. Mol. Biol. Rep. 29, 237–241.

Kell, D.B., 2002b. Genotype:phenotype mapping: genes as computer programs. Trends Genet. 18, 555–559.

Kell, D.B., 2004. Metabolomics and systems biology: making sense of the soup. Curr. Opin. Microbiol. 7, 296–307.

Kell, D.B., 2006a. Metabolomics, modelling and machine learning in systems biology: towards an understanding of the languages of cells. The 2005 Theodor Bücher lecture. FEBS J. 273, 873–894.

Kell, D.B., 2006b. Systems biology, metabolic modelling and metabolomics in drug discovery and development. Drug Disc Today 11, 1085–1092.

Kell, D.B., 2007. The virtual human: towards a global systems biology of multiscale, distributed biochemical network models. IUBMB Life 59, 689–695.

Kell, D.B., Knowles, J.D., 2006. The role of modeling in systems biology. In: Szallasi, Z., et al. (Eds.), System Modeling in Cellular Biology: From Concepts to Nuts and Bolts. MIT Press, Cambridge, MA, pp. 3–18.

Kell, D.B., Mendes, P., 2000. Snapshots of systems: metabolic control analysis and biotechnology in the post-genomic era. In: Cornish-Bowden, A., Cárdenas, M.L. (Eds.), Technological and Medical Implications of Metabolic Control Analysis. Kluwer Academic Publishers, Dordrecht, pp. 3–25 (See http://dbkgroup.org/WhitePapers/mcabio.htm).

Kell, D.B., Oliver, S.G., 2004. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. Bioessays 26, 99–105.

Kell, D.B., Paton, N.W., 2007. Loosely coupled bioinformatic workflows for systems biology. In: Cassman, M., Brunak, S. (Eds.), US-EU Workshop on infrastructure needs for systems biology, in press.

Kell, D.B., Westerhoff, H.V., 1986. Metabolic control theory: its role in microbiology and biotechnology. FEMS Microbiol. Rev. 39, 305–320.

Kell, D.B., van Dam, K., Westerhoff, H.V., 1989. Control analysis of microbial growth and productivity. Symp. Soc. Gen. Microbiol. 44, 61–93.

Kell, D.B., Darby, R.M., Draper, J., 2001. Genomic computing: explanatory analysis of plant expression profiling data using machine learning. Plant Physiol. 126, 943–951.

Kenny, L.C., Dunn, W.B., Ellis, D.I., Myers, J., Baker, P.N., Kell, D.B., The GOPEC Consortium, 2005. Novel biomarkers for pre-eclampsia detected using metabolomics and machine learning. Metabolomics 1, 227–234 (online doi:10.1007/s11306-005-0003-1).

Klipp, E., Heinrich, R., 1999. Competition for enzymes in metabolic pathways: implications for optimal distributions of enzyme concentrations and for the distribution of flux control. Biosystems 54, 1–14.

Kola, I., Landis, J., 2004. Can the pharmaceutical industry reduce attrition rates? Nat. Rev. Drug Discov. 3, 711–715.

Koza, J.R., 1992. Genetic Programming: On the Programming of Computers by Means of Natural Selection. MIT Press, Cambridge, MA.

Koza, J.R., Mydlowec, W., Lanza, G., Yu, J., Keane, M.A., 2001a. Reverse engineering of metabolic pathways from observed data using genetic programming. Pac. Symp. Biocomput., 434–445.

Koza, J.R., Mydlowec, W., Lanza, G., Yu, J., Keane, M.A., 2001b. Automatic synthesis of both the topology and sizing of metabolic pathways using genetic programming. In: Spector, L., et al. (Eds.), Proc. GECCO-2001. Morgan Kaufmann, San Francisco, pp. 57–65.

Koza, J.R., Keane, M.A., Streeter, M.J., Mydlowec, W., Yu, J., Lanza, G., 2003. Genetic Programming: Routine Human-Competitive Machine Intelligence. Kluwer, New York.

Langdon, W.B., 1998. Genetic Programming and Data Structures: Genetic Programming + Data Structures = Automatic Programming!. Kluwer, Boston.

Le Novère, N., Finney, A., Hucka, M., Bhalla, U.S., Campagne, F., Collado-Vides, J., Crampin, E.J., Halstead, M., Klipp, E., Mendes, P., Nielsen, P., Sauro, H., Shapiro, B., Snoep, J.L., Spence, H.D., Wanner, B.L., 2005. Minimum information requested in the annotation of biochemical models (MIRIAM). Nat. Biotechnol. 23, 1509–1515.

Le Novère, N., Bornstein, B., Broicher, A., Courtot, M., Donizelli, M., Dharuri, H., Li, L., Sauro, H., Schilstra, M., Shapiro, B., Snoep, J.L., Hucka, M., 2006. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. Nucleic Acids Res. 34, D689–D691.

Lüdtke, N., Panzeri, S., Brown, M., Broomhead, D.S., Knowles, J., Montemurro, M.A., Kell, D.B., 2007. Information-theoretic Sensitivity Analysis: a general method for credit assignment in complex networks. J. R. Soc. Interface, in press, doi:10.1098/rsif.2007.1079.

Ma, H., Sorokin, A., Mazein, A., Selkov, A., Selkov, E., Demin, O., Goryanin, I., 2007. The Edinburgh human metabolic network reconstruction and its functional analysis. Mol. Syst. Biol. 3, 135.

McLuhan, M., Fiore, Q., 1971. The Medium is the Massage. Penguin Books, London.

Meléndez-Hevia, E., Waddell, T.G., Heinrich, R., Montero, F., 1997. Theoretical approaches to the evolutionary optimization of glycolysis. Chemical analysis. Eur. J. Biochem. 244, 527–543.

Mendes, P., 1990. GEPASI. In: Cornish-Bowden, A., Cárdenas, M.L. (Eds.), Control of Metabolic Processes. Plenum Press, New York/London, pp. 434–435.

Mendes, P., 1993. GEPASI: a software package for modelling the dynamics, steady states and control of biochemical and other systems. Comput. Appl. Biosci. 9, 563–571.

Mendes, P., 1997. Biochemistry by numbers: simulation of biochemical pathways with Gepasi 3. Trends Biochem. Sci. 22, 361–363.

Mendes, P., 2001. Modeling large scale biological systems from functional genomic data: parameter estimation. In: Kitano, H. (Ed.), Foundations of Systems Biology. MIT Press, Cambridge, MA, pp. 163–186.

Mendes, P., Kell, D.B., 1998. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. Bioinformatics 14, 869–883.

Moles, C.G., Mendes, P., Banga, J.R., 2003. Parameter estimation in biochemical pathways: a comparison of global optimization methods. Genome Res. 13, 2467–2474.

O'Hagan, S., Dunn, W.B., Brown, M., Knowles, J.D., Kell, D.B., 2005. Closed-loop, multiobjective optimisation of analytical instrumentation: gas-chromatography-time-of-flight mass spectrometry of the

metabolomes of human serum and of yeast fermentations. Anal. Chem. 77, 290–303.

O'Hagan, S., Dunn, W.B., Broadhurst, D., Williams, R., Ashworth, J.A., Cameron, M., Knowles, J., Kell, D.B., 2007. Closed-loop, multi-objective optimisation of two-dimensional gas chromatography (GC × GC-tof-MS) for serum metabolomics. Anal. Chem. 79, 464–476.

Oinn, T., Addis, M., Ferris, J., Marvin, D., Senger, M., Greenwood, M., Carver, T., Glover, K., Pocock, M.R., Wipat, A., Li, P., 2004. Taverna: a tool for the composition and enactment of bioinformatics workflows. Bioinformatics 20, 3045–3054.

Oinn, T., Greenwood, M., Addis, M., Alpdemir, M.N., Ferris, J., Glover, K., Goble, C., Goderis, A., Hull, D., Marvin, D., Li, P., Lord, P., Pocock, M.R., Senger, M., Stevens, R., Wipat, A., Wroe, C., 2006. Taverna: lessons in creating a workflow environment for the life sciences. Concurr. Comput.: Pract. Experience 18, 1067–1100.

Oinn, T., Li, P., Kell, D.B., Goble, C., Goderis, A., Greenwood, M., Hull, D., Stevens, R., Turi, D., Zhao, J., 2007. Taverna/myGrid: aligning a workflow system with the life sciences community. In: Taylor, I.J., et al. (Eds.), Workflows for e-Science: Scientific Workflows for Grids. Springer, Guildford, pp. 300–319.

Patil, K.R., Rocha, I., Förster, J., Nielsen, J., 2005. Evolutionary programming as a platform for *in silico* metabolic engineering. BMC Bioinform. 6, 308.

Rabitz, H., Aliş, Ö.F., 1999. General foundations of high-dimensional model representations. J. Math. Chem. 25, 197–233.

Rodriguez-Fernandez, M., Mendes, P., Banga, J.R., 2006. A hybrid approach for efficient and robust parameter estimation in biochemical pathways. Biosystems 83, 248–265.

Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., 2004. Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models. Wiley, New York.

Sauro, H.M., Hucka, M., Finney, A., Wellock, C., Bolouri, H., Doyle, J., Kitano, H., 2003. Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration. Omics 7, 355–372.

Schuster, S., Heinrich, R., 1987. Time hierarchy in enzymatic reaction chains resulting from optimality principles. J. Theor. Biol. 129, 189–209.

Schuster, S., Heinrich, R., 1991. Minimization of intermediate concentrations as a suggested optimality principle for biochemical networks. I. Theoretical analysis. J. Math. Biol. 29, 425–442.

Schuster, S., Schuster, R., Heinrich, R., 1991. Minimization of intermediate concentrations as a suggested optimality principle for biochemical networks. II. Time hierarchy, enzymatic rate laws, and erythrocyte metabolism. J. Math. Biol. 29, 443–455.

Siepel, A., Farmer, A., Tolopko, A., Zhuang, M., Mendes, P., Beavis, W., Sobral, B., 2001. ISYS: a decentralized, component-based approach to the integration of heterogeneous bioinformatics resources. Bioinformatics 17, 83–94.

Smallbone, K., Simeonidis, E., Broomhead, D.S., Kell, D.B., 2007. Something from nothing: bridging the gap between constraint-based and kinetic modelling. FEBS J 274, 5576–5585.

Spasic, I., Schober, D., Sansone, S.-A., Rebholz-Schuhmann, D., Kell, D.B., Paton, N., 2007. Facilitating the development of controlled vocabularies for metabolomics with text mining. BMC Bioinformatics, in press.

Stein, S.E., Heller, S.R., Tchekhovski, D., 2003. An open standard for chemical structure representation—the IUPAC chemical identifier. Proc. 2003 Nimes Int. Chem. Info. Conf., pp. 131–143.

Stephani, A., Nuno, J.C., Heinrich, R., 1999. Optimal stoichiometric designs of ATP-producing systems as determined by an evolutionary algorithm. J. Theor. Biol. 199, 45–61.

Stevens, R., Bodenreider, O., Lussier, Y.A., 2006. Semantic webs for life sciences. Pac. Symp. Biocomput., 112–115.

Taylor, K.R., Gledhill, R.J., Essex, J.W., Frey, J.G., Harris, S.W., De Roure, D.C., 2006. Bringing chemical data onto the Semantic Web. J. Chem. Inf. Model 46, 939–952.

Uhlen, M., Bjorling, E., Agaton, C., Szigyarto, C.A., Amini, B., Andersen, E., Andersson, A.C., Angelidou, P., Asplund, A., Asplund, C., Berglund, L., Bergstrom, K., Brumer, H., Cerjan, D., Ekstrom, M., Elobeid, A., Eriksson, C., Fagerberg, L., Falk, R., Fall, J., Forsberg, M., Bjorklund, M.G., Gumbel, K., Halimi, A., Hallin, I., Hamsten, C., Hansson, M., Hedhammar, M., Hercules, G., Kampf, C., Larsson, K., Lindskog, M., Lodewyckx, W., Lund, J., Lundeberg, J., Magnusson, K., Malm, E., Nilsson, P., Odling, J., Oksvold, P., Olsson, I., Oster, E., Ottosson, J., Paavilainen, L., Persson, A., Rimini, R., Rockberg, J., Runeson, M., Sivertsson, A., Skollermo, A., Steen, J., Stenvall, M., Sterky, F., Stromberg, S., Sundberg, M., Tegel, H., Tourle, S., Wahlund, E., Walden, A., Wan, J., Wernerus, H., Westberg, J., Wester, K., Wrethagen, U., Xu, L.L., Hober, S., Ponten, F., 2005. A human protein atlas for normal and cancer tissues based on antibody proteomics. Mol. Cell Proteomics 4, 1920–1932.

van der Greef, J., Hankemeier, T., McBurney, R.N., 2006. Metabolomics-based systems biology and personalized medicine: moving towards $n = 1$ clinical trials? Pharmacogenomics 7, 1087–1094.

Waddell, T.G., Repovic, P., MelendezHevia, E., Heinrich, R., Montero, F., 1997. Optimization of glycolysis: a new look at the efficiency of energy coupling. Biochem. Edu. 25, 204–205.

Wang, X., Gorlitsky, R., Almeida, J.S., 2005. From XML to RDF: how semantic web technologies will change the design of 'omic' standards. Nat. Biotechnol. 23, 1099–1103.

Weininger, D., 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci. 28, 31–36.

Wilkinson, D.J., 2007. Bayesian methods in bioinformatics and computational systems biology. Brief Bioinform. 8, 109–116.

Wilkinson, M., Schoof, H., Ernst, R., Haase, D., 2005. BioMOBY successfully integrates distributed heterogeneous bioinformatics Web Services. The PlaNet exemplar case. Plant Physiol. 138, 5–17.

Wilkinson, S.J., Benson, N., Kell, D.B., 2007. Proximate parameter tuning for biochemical networks with uncertain kinetic parameters. Mol. Biosyst., in press, doi:10.1039/b707506e.

Wolpert, D.H., Macready, W.G., 1997. No Free Lunch theorems for optimization. IEEE Trans. Evol. Comput. 1, 67–82.