ELSEVIER

Systems biology, metabolic modelling and metabolomics in drug discovery and development

Douglas B. Kell[1,2]

[1] School of Chemistry, Faraday Building, The University of Manchester. PO Box 88, Manchester, M60 1QD, UK
[2] The Manchester Centre for Integrative Systems Biology, The Manchester Interdisciplinary Biocentre, 131, Princess St, Manchester, M1 7DN, UK

Unlike signalling pathways, metabolic networks are subject to strict stoichiometric constraints. Metabolomics amplifies changes in the proteome, and represents more closely the phenotype of an organism. Recent advances enable the production (and computer-readable encoding as SBML) of metabolic network models reconstructed from genome sequences, as well as experimental measurements of much of the metabolome. There is increasing convergence between the number of human metabolites estimated via genomics (~3000) and the number measured experimentally. It is thus both timely, and now possible, to bring these two approaches together as an integrated (if distributed) whole to help understand the genesis of metabolic biomarkers, the progress of disease, and the modes of action, efficacy, off-target effects and toxicity of pharmaceutical drugs.

Systems biology and metabolic modelling in the 21st Century

Although there are many individual definitions, most commentators (including this one [1,2]) take it that systems biology involves an iterative interplay between more or less high-throughput and high-content 'wet' experiments, technology development, theory and computational modelling, and that it is the involvement of computational modelling, in particular, in the process that sets systems biology apart from the more traditional and more reductionist molecular biology. Metabolomics illustrates this amply. There is also the view that the perceived decrease in the effectiveness of the target-based drug discovery process [3–5], including the still-high levels of attrition [6], means that we must move towards understanding organisms at something more akin to a whole-system level [7–11].

The question then arises as to what part of a system one might first beneficially model? Although there has, unsurprisingly, been considerable interest in modelling the major signalling pathways (e.g. Refs [12,13]), there are several reasons why it is timely to turn our attention to the level of small molecule metabolism, which is the focus of this review.

Corresponding author: Kell, D.B. (dbk@manchester.ac.uk)

Metabolism is more discriminating

It has long been known, and proven through the formalism of metabolic control analysis [7,10,14–16], that whereas small changes in the concentrations of enzymes (and the transcripts that encode them) have only small effects on the fluxes through metabolic pathways, they have substantial effects on the concentrations of metabolic intermediates. Because the metabolome (nominally the concentrations of 'all' the metabolites measured in a system of interest [17]) is downstream of the proteome, it is thereby 'amplified' both in theory [18] and in practice [19,20] and represents a more sensitive level of organisation than do the macromolecular 'omes for understanding a complex biological system, and the changes in it that might be occasioned by disease or pharmaceutical intervention [21,22].

Metabolic reconstruction is now mature and timely

An attractive feature for the purposes of modelling is that metabolism, in contrast to signalling pathways, is subject to direct thermodynamic and in particular stoichiometric constraints [23,24]. As the product of one reaction is usually the substrate of another, and we know a considerable amount at a baseline level [25], the starting point for metabolic reconstruction is thus the genome itself. A combination of automated and manual procedures can help turn a genome sequence into a metabolic model,

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2006.10.004

TABLE 1

Useful Internet resources for metabolic modelling, metabolomics and systems biology

Resource	Website	Comment
Portals		
	http://dbkgroup.org/metabol.htm#links	Metabolomics links
	http://dbkgroup.org/sysbio.htm#links	Systems biology links
Metabolic pathways		
BRENDA	http://www.brenda.uni-koeln.de/index.php4	Enzyme database
Expasy	http://www.expasy.ch/cgi-bin/search-biochem-index	Classic maps
KEGG	http://www.genome.ad.jp/kegg/kegg2.html	A widely used site
Metacyc	http://www.metacyc.org/	Many pathways included
Metabolomics		
	http://www.metabolomics.ca/	Human metabolome database
	http://www.husermet.org/	Human serum metabolome project
Modelling		
SBML	http://www.sbml.org/	Links to most bio-modelling websites
SBGN	http://www.sbgn.org/	An emerging standard for visualising SBML models
Metabolic models		
	http://jjj.biochem.sun.ac.za/	'Triple-J site'
	http://www.ebi.ac.uk/biomodels/	Biomodels
Other		
	http://dbkgroup.org/memo/	MeMo data model
	http://www.metabolomicssociety.org/	Metabolomics Society

and a variety are available (e.g. Refs [26–30]). The qualitative metabolic network or logical graph, popularised in the biochemical wall charts [25] and resources such as Kyoto Encyclopedia of Genes and Genomes (KEGG) [31], is then the starting point for metabolic modelling. Given a homogenous compartment, the normal strategy is to develop the qualitative model into a quantitative model, in which each step is represented locally as an ordinary differential equation (ODE) that obeys a typical biochemical equation (such as that of Michaelis and Menten) and that can then be parametrised (in terms of kinetic and related constants, such as K_m and V_{max}). When the number of molecules is small, so-called stochastic simulations are required [32]. However, such detail can be more or less hidden from the modeller, as we now have available a representation that enables interoperability among many pieces of software, namely the Systems Biology Markup Language (SBML) model [33].

Generalised representation of metabolic and other biochemical models

SBML ([33]; http://www.sbml.org/; see also Table 1 for other Internet resources) is an eXtensible Markup Language (XML) that, in its present version, enables one straightforwardly to describe a biological network and its local equations in a manner that can be exchanged between any number of modelling systems, including what is probably [34] the most popular modelling software, viz. *Gepasi* [35]. A simple example of a model of glycolysis [36] is given in Figure 1. SBML therefore encodes the model in a manner that enables one to use it for any number of other desirable manipulations (Figure 2) and, as such, the importance of the SBML cannot be overestimated. As discussed later, SBML is one of the main ways of integrating metabolism, metabolomics and systems biology models.

Systems parameters and systems variables

Systems biology models make explicit the relationship between the elements of a system, namely the parameters (here the fixed or starting concentrations of proteins and controlled metabolites, and all the kinetic constants of the proteins for their substrates, products and effectors) and the variables (the time-dependent metabolite concentrations and fluxes). Therefore, a major requirement is the measurement of parameters [2], but much of our energy is expended on the measurements of the variables (i.e. the concentrations and fluxes, metabolomics and fluxomics). Systems biology needs to integrate all of these.

Metabolomics and technologies for its measurement

Metabolomics seeks to measure the concentrations of nominally all of the [small molecular weight (MW)] metabolites in a particular system, for example, a body fluid such as serum or an ensemble of cells [17,37,38], although normally a more restricted subset, the 'metabolic profile' is measured in practice. This is because of the huge chemical diversity, especially in terms of polarity, among different metabolites. As part of the emphasis on technology development above (and see Ref. [39]), the technologies that enable us to do metabolomics well have recently increased in power [17,40–43]. They normally include a separation step (gas or liquid chromatographies or electrophoresis) coupled to an identification step (typically mass spectrometry). Sophisticated optimisation methods are improving these still further. Thus, following the development of the 'Robot Scientist' approach

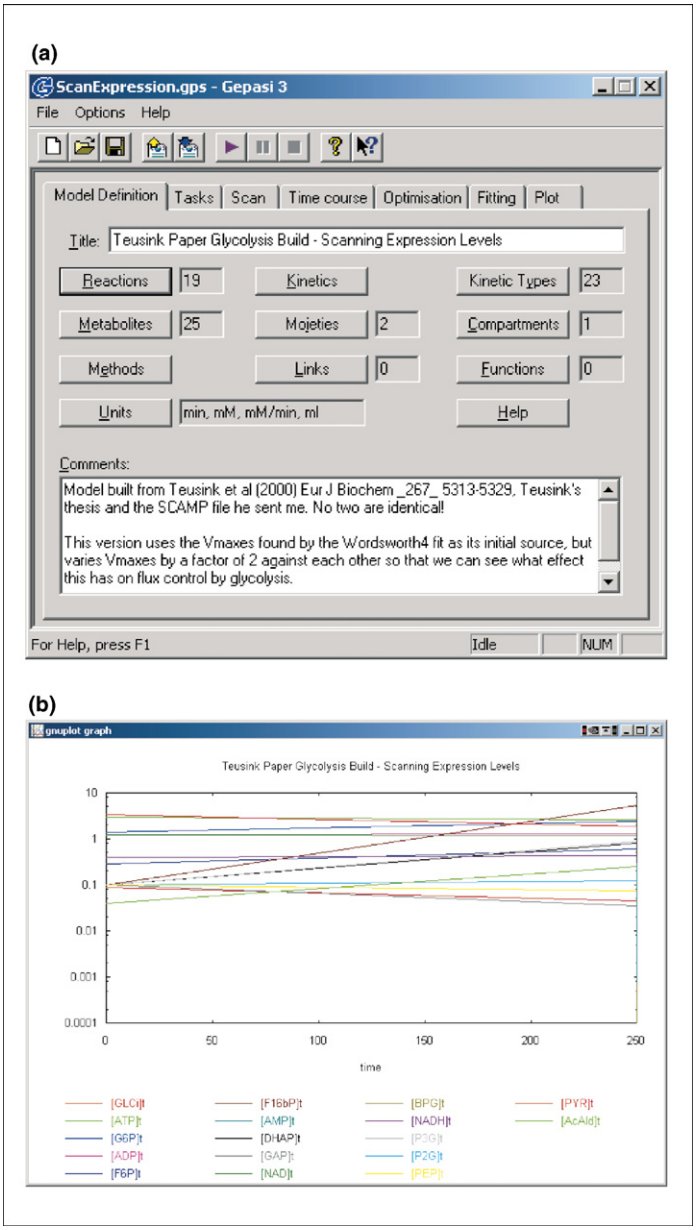

FIGURE 1

A typical metabolic model in *Gepasi*. (a) A screenshot of the model set-up and **(b)** the time series (in s) of variables that can then be compared with experiment. Note that not all the metabolites have reached a steady state.

[44], in which closed-loop methods of active learning are exploited to design and perform serial experiments in an intelligent manner, a comparable strategy has been exploited to increase hugely the number of metabolites that can be detected chromatographically. Using one-dimensional GC-tof-MS, the number of peaks observable in human serum was approximately trebled (compared with another GC-tof-MS method that was the starting point and that we thereby improved) to >1200 [45], which has already enabled the discovery of several novel biomarkers for pre-eclampsia [46] (and see later). Two-dimensional GCxGC-tof-MS has been exploited further to treble this number to >4000 raw peaks, which equates to ~1800 metabolite peaks [47,48]. Comparable liquid-phase chromatographies based on reverse-phase UPLC-MS [49,50] can measure thousands of peaks, and a new version of normal-phase UPLC-MS using HILIC [51,52] columns provides complementary measurements. Although what one measures can be influenced strongly by the means of extraction [53], how do these numbers compare with the known or inferred size of the human metabolome?

Sizes of the human and other metabolomes

An attraction of the metabolome has always been that it is numerically smaller, and thus more tractable, than the transcriptome or proteome [37]. In the case of baker's yeast (*Saccharomyces cerevisiae*), the latest models (e.g. Ref. [26]) give some 1200 reactions and 650 metabolites, with slightly smaller but broadly similar numbers for bacteria such as *Escherichia coli* [54–56] and *Streptomyces coelicolor* [28], most with a MW <500 [27,55]. The curated human metabolome [as reconstructed semi-manually from the consensus human genome sequence, build 31 [29] or 35, (Bernhard Palsson, pers. commun.)], presently contains respectively some 1100/3300 reactions and 700/2700 metabolites. (The number of enzymes inferred to be gene products is more than the number of reactions owing to the common existence of isozymes.) The availability of an accurate human metabolic network will revolutionise metabolomics, although the number of reactions and metabolites will be an underestimate for several reasons. First, it is recognised that some areas of metabolism are more 'represented' than others; transporters especially are highly underrepresented in terms of the literature. (This is particularly true for their activities in transporting xenobiotics and pharmaceuticals [57].) Second, many enzymes will have currently unknown substrates. Third, without the use of 'untargeted' metabolomics strategies (see below), it is hard to discover molecules whose existence one does not suspect, and so some molecules might be reasonably prevalent but of unknown chemical identity. (In plants and yeast, most metabolites measured by gas chromatography-mass spectrometry are presently of uncertain identity.) The measured metabolome is greater than that encoded by the genome, as it will include molecules acquired exogenously as drugs, foods or food additives, and will also include molecules derived from the microflora of the host [58].

Coming from the experimental end, Siuzdak and colleagues have measured some 3000 metabolite peaks in human serum [59], and nearly 2000 can be seen by GCxGC in serum, a similar number to that seen by Soga and colleagues using capillary electrophoresis-mass spectrometry in liver extracts [60]. An exciting initiative led by David Wishart and Lori Querengesser is the Human Metabolome Database (http://www.hmdb.ca), which is seeking to catalogue all the human metabolites with a concentration >1 μm in serum, and to confirm their identity using authentic standards. The current number of metabolites identified in this way is >800, with the expectation of reaching 1400 by the end of 2006. A broadly related goal is part of 'HUSERMET' ('the human serum metabolome in health and disease') project (http://www.husermet.org), where the database is based on the open-standard MeMo data model [61]. It is likely that the number of endogenous metabolites at significant concentrations in the human serum metabolome in the range of 1–10,000, and is likely to asymptote at ~3–3500. This statement somewhat ignores the large number of lipid combinations that are being picked up by the LIPIDMAPS consortium (http://www.lipidmaps.org).

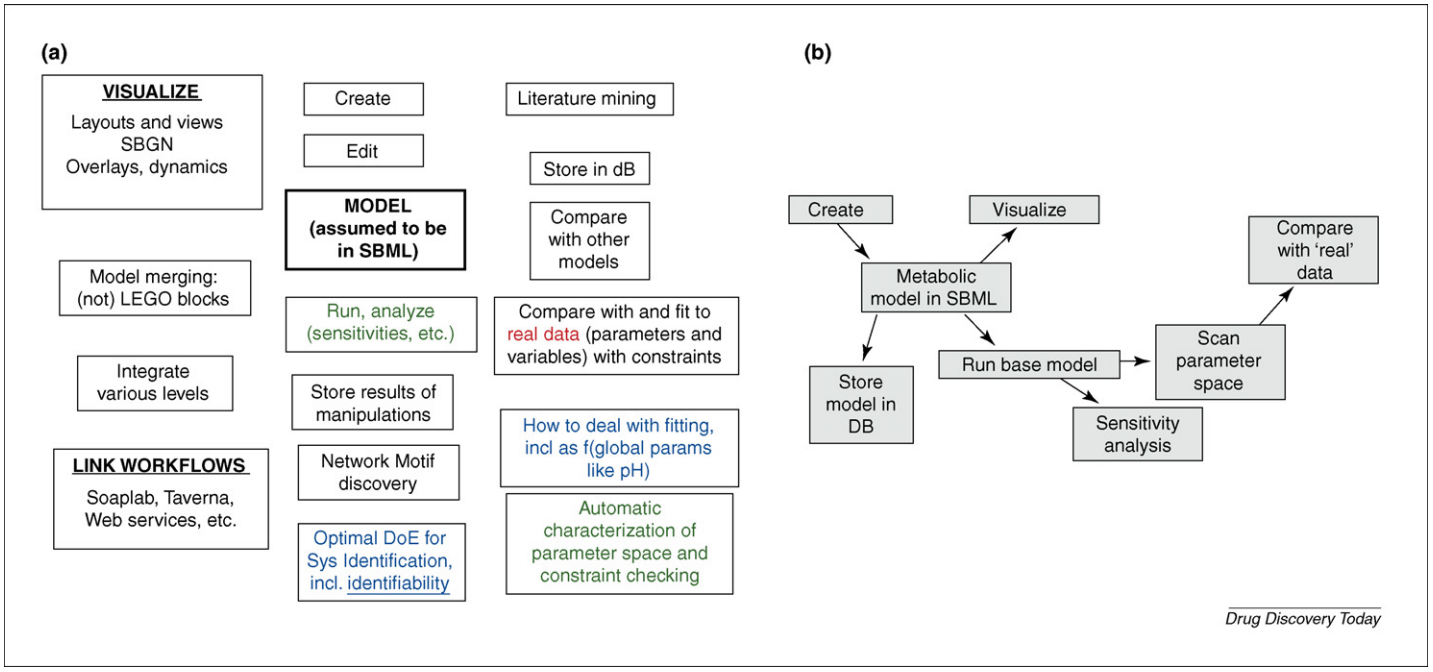

FIGURE 2

Integrating metabolic models into a systems biology workflow. (a) The representation of a metabolic model in SBML has many possibilities, including its creation, editing, visualisation, running, sensitivity analysis, comparison with experimental data, model merging, nonlinear dynamics analysis and so on. It is likely that the pieces of software that do any of these well will differ from each other, and so it is necessary to integrate them with each other in a distributed manner. Environments such as Taverna [100,102,105] enable the specification of the necessary bioinformatics workflows. **(b)** A workflow based on some of the components in (a).

What and where to measure?

One question that arises for those contemplating a metabolome project using measurements on biofluids, is whether to study urine or plasma and/or serum (there seems little difference between the latter; Dunn, W.B. *et al.*, unpublished). Overall, the general feeling is that urine reflects a more short-term state of the organism, whereas fasting serum and/or plasma changes represent more chronic or long-term snapshots of the system. There is also the influence of ethnicity, diet, diurnal rhythm, and so on, on the experimental metabolome to consider [62,63].

What is the metabolome potentially useful for?

Biomarker detection

In a certain sense, the metabolome is chemical pathology writ large, and just as many disease conditions are now assessed by measuring small molecule concentrations in biofluids, we can expect the metabolome to be of significant utility in various kinds of diagnosis [21,64]. Of course, the entire field of 'inborn errors of metabolism' is based on seeking diagnostic changes in the metabolome, and these are now measured routinely [65,66]. Such diagnostic biomarkers can of course be surrogates [67], as well as being genuinely diagnostic. However, studies that seek novel biomarkers are not without their difficulties, and bias is an ever-present danger [68,69]. In addition, because we measure multiple metabolites, their statistical analysis involves multiple hypothesis testing and there is a profound danger of false discoveries, especially when sample sizes are low [70–72]. This said, and although these are early days, there have been some interesting and unexpected findings, often from the perhaps more obscure parts of intermediary metabolism. Indeed, the great attraction of metabolomics (and other omics) is that specific hypotheses are not tested; instead, the data are left to tell us the answer [39,73]. Thus, Ringeisser and colleagues proposed that *N*-methylnicotinamide and *N*-methyl-4-pyridone-3-carboxamide were potential urinary and plasma biomarkers of peroxisome proliferation in the rat [74]; Soga *et al.* [60] discovered that the metabolite ophthalmic acid, γ-glutamyl-2aminobutyryl-glycine, was a novel biomarker for oxidative stress occurring as a result of glutathione depletion caused by the administration of acetaminophen (paracetamol).

He and colleagues were interested in finding ligands for the orphan GPCR GPR91, and discovered that these were in fact the citric acid cycle intermediates succinate and 2-oxoglutarate [75], thus opening up an interesting and novel area, and drawing attention further to the poorly recognised signalling roles of small molecules normally considered to be intermediary metabolites.

Using liquid chromatography-mass spectrometry, Sabatine, Gerszten and colleagues identified several biomarkers for exercise-induced myocardial ischaemia [76], including six members of the citric acid pathway that were among the 23 most changed metabolites. Furthermore, changes in six metabolites (citric acid, uric acid and GABA plus three unidentified metabolites) differentiated cases from controls with a high degree of accuracy. Using gas-chromatography-mass spectrometry [46] (Kenny, L.C., Dunn, W.B., Broadhurst, D.I., Brown, M.C., Ellis, D.I., Myers, J., Baker, P.N., The GOPEC Consortium and D.B. Kell, unpublished), a series of novel small MW metabolites have been discovered that serve as diagnostic biomarkers for pre-eclampsia (Figure 3). A separate study compared humans suffering from Huntington's disease with a mouse model of the disease, finding that the same metabolites were observed to have changed in each, thus confirming the utility

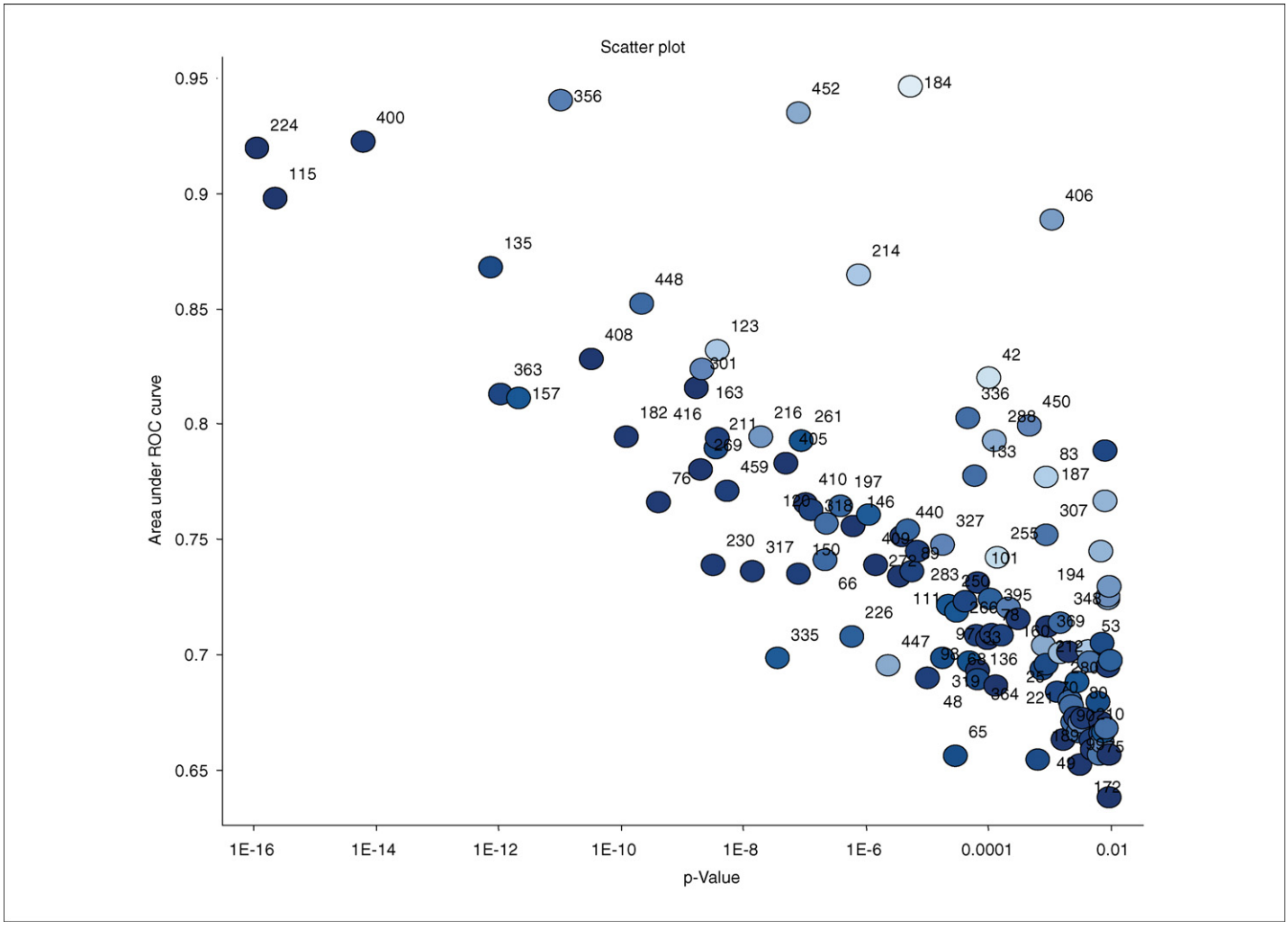

FIGURE 3

An example of modern metabolomic data. The importance of each of some 286 metabolites in discriminating pre-eclamptic from normal sera [46] is displayed as a plot of the area under the univariate ROC curve [127] versus the statistical *p* value in a non-parametric test [72].

of the mouse model [77]. Other examples include nucleosides for liver cancer [78], various lipids for type II diabetes [79], and a series of biomarkers for caloric restriction [80]. An attraction of many of these methods is that they can begin to give mechanistic insight into the relevant aetiologies of often progressive and complex physiologies and pathologies.

Mode of action studies

The notion that one might infer the site of action of an inhibitor in biology via the measurement of observable variables goes back to at least the crossover theorem (If you have a pathway A→B→C→D, the crossover theorem says that if you inhibit the step B→C then B will increase and C will decrease in concentration, thereby allowing you to assess the site of inhibition by looking at A, B, C, D. Except it doesn't work because of feeedback loops.) [81], albeit now discredited in general [82]; a more common view is that the site of a pharmaceutical intervention or genetic lesion can be inferred by using modern machine learning or pattern recognition techniques to look at the pattern of metabolic changes that ensue [42,83–86], calibrating as appropriate with molecules for which the answer is known [87] and validating using samples not involved in the formation of the predictive model.

In a similar vein, measurements of the metabolome can be coupled to appropriate breeding experiments so as to infer gene/QTL-metabolite linkages [88,89] Another related, useful and powerful strategy, at a genome-wide systems level, but which does not involve metabolomics directly, is based on the analysis of the differential effects of drugs on cells in response to specific gene dosages [90,91]. In principle, metabolomics measurements could add considerably to the value of such analyses in terms of understanding the effect of intervening with metabolic networks at known locations.

Toxicity analyses

Inferring the mode or target of action of a compound by measuring its effects on the metabolome is conceptually equivalent, in principle, to inferring mechanisms of toxicity by looking for 'tell-tale' metabolomic patterns, signatures or biomarker molecules, although one might hope that toxicity markers would be both more obvious and more greatly altered. (Indeed, I reviewed above two studies [60,74] under 'biomarkers' that were really looking at toxicological endpoints, i.e., some markers referred to as BIO markers are really TOX markers) As with the 'calibration' method for functional genomics [87], one obvious strategy is to build up a

database of the time- and dose-dependent metabolite patterns that occur when organisms are challenged with substances of 'known' toxicity (recognising that few, if any, are 'clean'), with a view to finding equivalent markers when 'unknown' drugs are tested; this is beginning to be undertaken [92]. Although it is early days, it is gratifying that the kinds of molecule being observed do correlate with what is to be expected consequently, e.g. upon nephrotoxicity or hepatotoxicity [93].

Efficacy

If one can infer a pattern of disease progression by measuring the metabolome, it is possible that the efficacy of a drug in more or less reversing that process might similarly be determined from the metabolome, and this is the motivation for some studies [22]. Although I am not aware of relevant data that show this clearly, it is likely to be part of the general strategy of theranostics [94], where modern diagnostics and therapeutics are integrated intelligently to assess the success or otherwise of therapeutic interventions. The personalised medicine agenda (e.g. Ref. [95]) will rely heavily on the availability of suitable biomarkers for all of the above.

Integrating metabolomics and metabolic modelling for systems biology

The methods for carrying out metabolic modelling, and the means for collecting, storing and analysing metabolomic data are considerably different, will normally be performed by individuals or in laboratories with different skill sets, and yet necessarily will deal with the same molecules. It is therefore extremely timely to bring together the known or inferred metabolic maps of suitable organisms with measurements of their metabolomes to provide a systems-level understanding of the metabolic fluxes and metabolite concentrations in these organisms, and how they might change under different conditions. The means by which this is to be accomplished is presaged in Figure 2, where it is recognised that, notwithstanding the core importance of the SBML representation, it needs to be exploited in an integrated environment. This, however, does not mean that we have to try and make an integrated environment by starting from scratch (and the legacy of excellent software and data would make this an act of folly), or that the integration has to be undertaken locally. What is needed is to link together the various elements and modules in which we are interested.

As was implicit almost from the beginning of such bioinformatic studies [96], the concept of the pipeline [97,98] or workflow [99–102] is now common in bioinformatics for the analysis of data. Here, the tools involved in the data analysis are stitched together using standardised environments or interfaces to form a workflow, after which they can then be enacted in a more or less automated manner. A convergence of various technologies now makes this easier [103]. Specifically, Web Services (WS) (http://w3.org/2002/ws/) is a distributed computing framework that enables computational resources such as databases and analysis tools to communicate with each other by the exchange of messages in XML based on the Simple Object Architecture Protocol (SOAP). All of this is well within the spirit of the Systems Biology Workbench [104] and of software Application Programming Interfaces more generally. Thus, distributed environments using systems such as Taverna [100,102,105] or others [106,107] to enact the necessary bioinformatic workflows provides an attractive way forward [2,108].

Future directions: bringing cheminformatics to metabolic systems biology

It is often the case that what are intellectually reasonably closely related subjects or disciplines can develop with little overlap, and two subjects that pertain closely to metabolic systems biology are cheminformatics and chemical genetics. Cheminformatics [109–111] is the application of informatics methods to solve chemical problems. Although it has largely been driven by the interests of the pharmaceutical industry whose concerns lie with xenobiotics, it is obvious that the same methods can be applied to the computational systems biology of natural metabolic systems, and we need to integrate the ideas and knowledge of cheminformatics into metabolomics, just as is happening with chemometrics [112]. Recent developments are increasing the richness of the representations that we can exploit [113], and bring the hope of adding chemical structure mining [114] to the emerging possibilities in literature and text mining (e.g. Refs [115–117]).

The modulation by small molecules of biological activities has proven of immense value historically in the dissection of biological pathways (e.g. in oxidative phosphorylation [118,119]). Chemical genetics or chemical genomics (e.g. Refs [120–123]) describes an integrated strategy for manipulating biological function using small molecules (the integration aspect specifically including cell biology-based assays and the databases necessary to systematise the knowledge and from that quantitative structure-activity relationships may be discerned [124]). Again, it is clear that the tools, including informatics tools, that are valuable for chemical genetics bear closely on those of value to the metabolic modeller, and having all of the data in a sensible, computer-readable form will enable the emergence of comparative metabolomics, which I predict will be as useful and powerful as comparative genomics is proving to be [125,126].

Acknowledgements

I thank the BBSRC, EPSRC and BHF for financial support of several projects, and AstraZeneca and GlaxoSmithKline for support of the HUSERMET project. Many individuals, especially David Broadhurst, Rick Dunn, Carole Goble, Roy Goodacre, Peter Li, Steve Oliver, Bernhard Palsson and Norman Paton are thanked for useful discussions. I thank Phil Baker, David Broadhurst, Rick Dunn and Louise Kenny for Figure 3. This is a contribution from the Manchester Centre for Integrative Systems Biology (http://www.mcisb.org).

References

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

2 Kell, D.B. (2006) 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

3 Lindsay, M.A. (2005) Finding new drug targets in the 21st century. *Drug Discov. Today* 10, 1683–1687
4 Sams-Dodd, F. (2005) Target-based drug discovery: is something wrong? *Drug Discov. Today* 10, 139–147
5 Schmid, E.F. and Smith, D.A. (2005) Is declining innovation in the pharmaceutical industry a myth? *Drug Discov. Today* 10, 1031–1039
6 Kola, I. and Landis, J. (2004) Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* 3, 711–715
7 Cascante, M. *et al.* (2002) Metabolic control analysis in drug discovery and disease. *Nat. Biotechnol.* 20, 243–249
8 Davidov, E. *et al.* (2003) Advancing drug discovery through systems biology. *Drug Discov. Today* 8, 175–183
9 Hornberg, J.J. *et al.* (2006) Cancer: a systems biology disease. *Biosystems* 83, 81–90
10 Klipp, E. *et al.* (2005) *Systems Biology in Practice: Concepts, Implementation and Clinical Application*. Wiley/VCH
11 Cho, C.R. *et al.* (2006) The application of systems biology to drug discovery. *Curr. Opin. Chem. Biol.* 10, 294–302
12 Schoeberl, B. *et al.* (2002) Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat. Biotechnol.* 20, 370–375
13 Nelson, D.E. *et al.* (2004) Oscillations in NF-κB signalling control the dynamics of target gene expression. *Science* 306, 704–708
14 Kell, D.B. and Westerhoff, H.V. (1986) Metabolic control theory: its role in microbiology and biotechnology. *FEMS Microbiol. Rev.* 39, 305–320
15 Fell, D.A. (1996) *Understanding the Control of Metabolism*. Portland Press
16 Heinrich, R. and Schuster, S. (1996) *The Regulation of Cellular Systems*. Chapman & Hall
17 Goodacre, R. *et al.* (2004) Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol.* 22, 245–252
18 Mendes, P. *et al.* (1996) Why and when channeling can decrease pool size at constant net flux in a simple dynamic channel. *Biochim. Biophys. Acta* 1289, 175–186
19 Raamsdonk, L.M. *et al.* (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat. Biotechnol.* 19, 45–50
20 Urbanczyk-Wochniak, E. *et al.* (2003) Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. *EMBO Rep.* 4, 989–993
21 Harrigan, G.G. and Goodacre, R., eds (2003) *Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function Analysis*, Kluwer Academic Publishers
22 van der Greef, J. *et al.* (2004) The role of analytical sciences in medical systems biology. *Curr. Opin. Chem. Biol.* 8, 559–565
23 Palsson, B.Ø. (2006) *Systems biology: properties of reconstructed networks*. Cambridge University Press
24 Alon, U. (2006) *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman and Hall/CRC
25 Michal, G., ed. (1999) *Biochemical Pathways: an Atlas of Biochemistry and Molecular Biology*, John Wiley & Sons
26 Duarte, N.C. *et al.* (2004) Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.* 14, 1298–1309
27 Kell, D.B. (2004) Metabolomics and systems biology: making sense of the soup. *Curr. Opin. Microbiol.* 7, 296–307
28 Borodina, I. *et al.* (2005) Genome-scale analysis of *Streptomyces coelicolor* A3(2) metabolism. *Genome Res.* 15, 820–829
29 Romero, P. *et al.* (2005) Computational prediction of human metabolic pathways from the complete human genome. *Genome Biol.* 6, R2
30 Arakawa, K. *et al.* (2006) GEM System: automatic prototyping of cell-wide metabolic pathway models from genomes. *BMC Bioinformatics* 7, 168
31 Kanehisa, M. *et al.* (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* 34, D354–D357
32 Andrews, S.S. and Bray, D. (2004) Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Phys. Biol.* 1, 137–151
33 Hucka, M. *et al.* (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19, 524–531
34 Alves, R. *et al.* (2006) Tools for kinetic modeling of biochemical networks. *Nat. Biotechnol.* 24, 667–672
35 Mendes, P. and Kell, D.B. (1998) Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics* 14, 869–883
36 Pritchard, L. and Kell, D.B. (2002) Schemes of flux control in a model of *Saccharomyces cerevisiae* glycolysis. *Eur. J. Biochem.* 269, 3894–3904
37 Oliver, S.G. *et al.* (1998) Systematic functional analysis of the yeast genome. *Trends Biotechnol.* 16, 373–378
38 Fiehn, O. (2001) Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp. Func. Genomics* 2, 155–168
39 Kell, D.B. and 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
40 Dunn, W.B. and Ellis, D.I. (2005) Metabolomics: current analytical platforms and methodologies. *Trends Anal. Chem.* 24, 285–294
41 Dunn, W.B. *et al.* (2005) Measuring the metabolome: current analytical technologies. *Analyst* 130, 606–625
42 Kell, D.B. *et al.* (2005) Metabolic footprinting and Systems Biology: the medium is the message. *Nat. Rev. Microbiol.* 3, 557–565
43 Weckwerth, W. and Morgenthal, K. (2005) Metabolomics: from pattern recognition to biological interpretation. *Drug Discov. Today* 10, 1551–1558
44 King, R.D. *et al.* (2004) Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature* 427, 247–252
45 O'Hagan, S. *et al.* (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
46 Kenny, L.C. *et al.* (2005) Novel biomarkers for pre-eclampsia detected using metabolomics and machine learning. *Metabolomics* 1, 227–234
47 Ong, R.C. and Marriott, P.J. (2002) A review of basic concepts in comprehensive two-dimensional gas chromatography. *J. Chromatogr. Sci.* 40, 276–291
48 Blumberg, L.M. (2003) Comprehensive two-dimensional gas chromatography: metrics, potentials, limits. *J. Chromatogr. A.* 985, 29–38
49 Plumb, R. *et al.* (2004) Ultra-performance liquid chromatography coupled to quadrupole- orthogonal time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 18, 2331–2337
50 Wilson, I.D. *et al.* (2005) High resolution ''ultra performance'' liquid chromatography coupled to oa-TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies. *J. Proteome Res.* 4, 591–598
51 Tolstikov, V.V. and Fiehn, O. (2002) Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. *Anal. Biochem.* 301, 298–307
52 Zywicki, B. *et al.* (2005) Comparison of rapid liquid chromatography-electrospray ionization-tandem mass spectrometry methods for determination of glycoalkaloids in transgenic field-grown potatoes. *Anal. Biochem.* 336, 178–186
53 Want, E.J. *et al.* (2006) Solvent-dependent metabolite distribution, clustering, and protein extraction for serum profiling with mass spectrometry. *Anal. Chem.* 78, 743–752
54 Reed, J.L. *et al.* (2003) An expanded genome-scale model of *Escherichia coli* K12 (iJR904 GSM/GPR). *Genome Biol.* 4, R54 (http://genomebiology.com/2003/2004/2009/R2054)
55 Nobeli, I. *et al.* (2003) A structure-based anatomy of the *E. coli* metabolome. *J. Mol. Biol.* 334, 697–719
56 Nobeli, I. and Thornton, J.M. (2006) A bioinformatician's view of the metabolome. *Bioessays* 28, 534–545
57 Sai, Y. and Tsuji, A. (2004) Transporter-mediated drug delivery: recent progress and experimental approaches. *Drug Discov. Today* 9, 712–720
58 Nicholson, J.K. *et al.* (2005) Gut microorganisms, mammalian metabolism and personalized health care. *Nat. Rev. Microbiol.* 3, 431–438
59 Smith, C.A. *et al.* (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* 78, 779–787
60 Soga, T. *et al.* (2006) Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption. *J. Biol. Chem.* 281, 16768–16776
61 Spasic, I. *et al.* (2006) MeMo: a hybrid SQL/XML approach to metabolomic data management for functional genomics. *BMC Bioinformatics* 7, 281
62 Plumb, R. *et al.* (2003) Metabonomic analysis of mouse urine by liquid-chromatography-time of flight mass spectrometry (LC-TOFMS): detection of strain, diurnal and gender differences. *Analyst* 128, 819–823
63 Lenz, E.M. *et al.* (2004) Metabonomics, dietary influences and cultural differences: a 1H NMR-based study of urine samples obtained from healthy British and Swedish subjects. *J. Pharm. Biomed. Anal.* 36, 841–849
64 Vaidyanathan, S. and et, al., eds (2005) *Metabolome Analyses: Strategies for Systems Biology*, Springer
65 Rashed, M.S. (2001) Clinical applications of tandem mass spectrometry: ten years of diagnosis and screening for inherited metabolic diseases. *J. Chromatogr. B. Biomed. Sci. Appl.* 758, 27–48

66 Pitt, J.J. *et al.* (2002) Comprehensive screening of urine samples for inborn errors of metabolism by electrospray tandem mass spectrometry. *Clin. Chem.* 48, 1970–1980

67 Lesko, L.J. and Atkinson, A.J., Jr (2001) Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: criteria, validation, strategies. *Annu. Rev. Pharmacol. Toxicol.* 41, 347–366

68 Ransohoff, D.F. (2005) Bias as a threat to the validity of cancer molecular-marker research. *Nat. Rev. Cancer* 5, 142–149

69 Kirschenlohr, H.L. *et al.* (2006) Proton NMR analysis of plasma is a weak predictor of coronary artery disease. *Nat. Med.* 12, 705–710

70 Wacholder, S. *et al.* (2004) Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J. Natl. Cancer Inst.* 96, 434–442

71 Ioannidis, J.P. (2005) Why most published research findings are false. *PLoS Med.* 2, e124

72 Broadhurst, D. and Kell, D.B. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics* (in press)

73 Breiman, L. (2001) Statistical modeling: The two cultures. *Stat. Sci.* 16, 199–215

74 Ringeissen, S. *et al.* (2003) Potential urinary and plasma biomarkers of peroxisome proliferation in the rat: identification of *N*-methylnicotinamide and *N*-methyl-4-pyridone-3-carboxamide by 1H nuclear magnetic resonance and high performance liquid chromatography. *Biomarkers* 8, 240–271

75 He, W. *et al.* (2004) Citric acid cycle intermediates as ligands for orphan G-protein-coupled receptors. *Nature* 429, 188–193

76 Sabatine, M.S. *et al.* (2005) Metabolomic identification of novel biomarkers of myocardial ischemia. *Circulation* 112, 3868–3875

77 Underwood, B.R. *et al.* (2006) Huntington's disease patients and transgenic mice have similar pro-catabolic serum metabolite profiles. *Brain* 129, 877–886

78 Yang, J. *et al.* (2004) Diagnosis of liver cancer using HPLC-based metabonomics avoiding false-positive result from hepatitis and hepatocirrhosis diseases. *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.* 813, 59–65

79 Yang, J. *et al.* (2004) Discrimination of Type 2 diabetic patients from healthy controls by using metabonomics method based on their serum fatty acid profiles. *J. Chromatogr. B.* 813, 53–58

80 Shi, H. *et al.* (2002) Characterization of diet-dependent metabolic serotypes: primary validation of male and female serotypes in independent cohorts of rats. *J. Nutr.* 132, 1039–1046

81 Chance, B. and Williams, G.R. (1955) Respiratory enzymes in oxidative phosphorylation. III The steady state. *J. Biol. Chem.* 217, 409–427

82 Westerhoff, H.V. and Kell, D.B. (1987) Matrix method for determining the steps most rate-limiting to metabolic fluxes in biotechnological processes. *Biotechnol. Bioeng.* 30, 101–107

83 Aranibar, N. *et al.* (2001) Automated mode-of-action detection by metabolic profiling. *Biochem. Biophys. Res. Commun.* 286, 150–155

84 Allen, J. *et al.* (2003) High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nat. Biotechnol.* 21, 692–696

85 Allen, J. *et al.* (2004) Discrimination of the modes of action of antifungal substances by use of metabolic footprinting. *Appl. Env. Micr.* 70, 6157–6165

86 Ott, K.-H. *et al.* (2003) Metabonomics classifies pathways affected by bioactive compounds. artificial neural network classification of NMR spectra of plant extracts. *Phytochemistry* 62, 971–985

87 Kell, D.B. and King, R.D. (2000) On the optimization of classes for the assignment of unidentified reading frames in functional genomics programmes: the need for machine learning. *Trends Biotechnol.* 18, 93–98

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

89 Schauer, N. *et al.* (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat. Biotechnol*

90 Giaever, G. *et al.* (1999) Genomic profiling of drug sensitivities via induced haploinsufficiency. *Nat. Genet.* 21, 278–283

91 Giaever, G. *et al.* (2004) Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 101, 793–798

92 Lindon, J.C. *et al.* (2005) The Consortium for Metabonomic Toxicology (COMET): aims, activities and achievements. *Pharmacogenomics* 6, 691–699

93 Robertson, D.G. (2005) Metabonomics in toxicology: a review. *Toxicol. Sci.* 85, 809–822

94 Picard, F.J. and Bergeron, M.G. (2002) Rapid molecular theranostics in infectious diseases. *Drug Discov. Today* 7, 1092–1101

95 Evans, W.E. and Relling, M.V. (2004) Moving towards individualized medicine with pharmacogenomics. *Nature* 429, 464–468

96 Goble, C.A. *et al.* (2001) Transparent access to multiple bioinformatics information sources. *IBM Syst. J.* 40, 532–551

97 Potter, S.C. *et al.* (2004) The Ensembl analysis pipeline. *Genome Res.* 14, 934–941

98 Brown, M. *et al.* (2005) A metabolome pipeline: from concept to data to knowledge. *Metabolomics* 1, 35–46

99 Peleg, M. *et al.* (2002) Modelling biological processes using workflow and Petri Net models. *Bioinformatics* 18, 825–837

100 Stevens, R. *et al.* (2004) ^{my}Grid and the drug discovery process. *DDT Biosilico* 4, 140–148

101 Romano, P. *et al.* (2005) Web services and workflow management for biological resources. *BMC Bioinformatics* 6 (Suppl 4), S24

102 Oinn, T. *et al.* (2006) Taverna/myGrid: aligning a workflow system with the life sciences community. In *Workflows for eScience*, M. Shields (Taylor, I., Deelman, E., Gannon, D., eds.) pp. 299–318, Springer

103 Stein, L. (2002) Creating a bioinformatics nation. *Nature* 417, 119–120

104 Sauro, H.M. *et al.* (2003) Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration. *OMICS* 7, 355–372

105 Oinn, T. *et al.* (2004) Taverna: a tool for the composition and enactment of bioinformatics workflows. *Bioinformatics* 20, 3045–3054

106 Lu, Q. *et al.* (2005) KDE Bioscience: Platform for bioinformatics analysis workflows. *J. Biomed. Inform.* 39, 440–450

107 Curcin, V. *et al.* (2005) Web services in the life sciences. *Drug Discov. Today* 10, 865–871

108 Chen, M. and Hofestädt, R. (2006) A medical bioinformatics approach for metabolic disorders: biomedical data prediction, modeling, and systematic analysis. *J. Biomed. Inform.* 39, 147–159

109 Gasteiger, J., ed. (2003) *Handbook of Chemoinformatics: From Data to Knowledge*, Wiley/VCH

110 Gasteiger, J. (2003) *Basic Chemoinformatics: A Textbook*, Wiley/VCH

111 Bajorath, J., ed. (2004) *Chemoinformatics: Concepts, Methods and Tools for Drug Discovery*, Humana Press

112 van der Greef, J. and Smilde, A.K. (2005) Symbiosis of chemometrics and metabolomics: past, present, and future. *J Chemometrics* 19 (5–7), 376–386

113 Taylor, K.R. *et al.* (2006) Bringing chemical data onto the Semantic Web. *J. Chem. Inf. Model.* 46, 939–952

114 Banville, D.L. (2006) Mining chemical structural information from the drug literature. *Drug Discov. Today* 11, 35–42

115 Ananiadou, S. and McNaught, J., eds (2006) *Text Mining in Biology and Biomedicine*, Artech House

116 Jensen, L.J. *et al.* (2006) Literature mining for the biologist: from information retrieval to biological discovery. *Nat. Rev. Genet.* 7, 119–129

117 Ananiadou, S. *et al.* Text Mining and its potential applications in Systems Biology. *Trends Biotechnol.* (in press)

118 Hitchens, G.D. and Kell, D.B. (1983) Uncouplers can shuttle rapidly between localised energy coupling sites during photophosphorylation by chromatophores of *Rhodopseudomonas capsulata* N22. *Biochem. J.* 212, 25–30

119 Westerhoff, H.V. and Kell, D.B. (1988) A control theoretical analysis of inhibitor titrations of metabolic channeling. *Comments Mol. Cell. Biophys.* 5, 57–107

120 Schreiber, S.L. (1998) Chemical genetics resulting from a passion for synthetic organic chemistry. *Bioorg. Med. Chem.* 6, 1127–1152

121 Brenner, C. (2004) Chemical genomics in yeast. *Genome Biol.* 5, 240

122 Smukste, I. and Stockwell, B.R. (2005) Advances in chemical genetics. *Annu. Rev. Genomics Hum. Genet.* 6, 261–286

123 Walsh, D.P. and Chang, Y.T. (2006) Chemical genetics. *Chem. Rev.* 106, 2476–2530

124 Haggarty, S.J. *et al.* (2004) Mapping chemical space using molecular descriptors and chemical genetics: deacetylase inhibitors. *Comb. Chem. High Throughput Screen.* 7, 669–676

125 Rubin, G.M. *et al.* (2000) Comparative genomics of the eukaryotes. *Science* 287, 2204–2215

126 Chain, P. *et al.* (2003) An applications-focused review of comparative genomics tools: capabilities, limitations and future challenges. *Brief. Bioinform.* 4, 105–123

127 Baker, S.G. (2003) The central role of receiver operating characteristic (ROC) curves in evaluating tests for the early detection of cancer. *J. Natl. Cancer Inst.* 95, 511–515