METABOLIC FOOTPRINTING AND SYSTEMS BIOLOGY: THE MEDIUM IS THE MESSAGE

Douglas B. Kell, Marie Brown*, Hazel M. Davey‡, Warwick B. Dunn*, Irena Spasic* and Stephen G. Oliver§*

Abstract | One element of classical systems analysis treats a system as a black or grey box, the inner structure and behaviour of which can be analysed and modelled by varying an internal or external condition, probing it from outside and studying the effect of the variation on the external observables. The result is an understanding of the inner make-up and workings of the system. The equivalent of this in biology is to observe what a cell or system excretes under controlled conditions — the 'metabolic footprint' or exometabolome — as this is readily and accurately measurable. Here, we review the principles, experimental approaches and scientific outcomes that have been obtained with this useful and convenient strategy.

**School of Chemistry, University of Manchester, Faraday Building, PO Box 88, Sackville Street, Manchester M60 1QD, UK.*
‡Institute of Biological Sciences, University of Wales, Cledwyn Building, Aberystwyth SY23 3DD, UK.
§Faculty of Life Sciences, University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, UK.
Correspondence to D.B.K.
e-mail: dbk@manchester.ac.uk
doi:10.1038/nrmicro1177
Published online 10 June 2005

The METABOLOME is defined as the quantitative complement of low-molecular-weight metabolites present in a cell under a given set of physiological conditions[1–9]. It lends itself readily to functional genomic and other analyses, as changes in a cell's physiology as a result of gene deletion or overexpression are amplified through the hierarchy of the TRANSCRIPTOME and the PROTEOME, and are therefore more easily measurable through the metabolome, even when changes in metabolic fluxes are negligible[1,10–12]. In addition, the default assumption underlying transcriptome and proteome analysis — that an x-fold increase in a transcript or protein necessarily results in an x-fold increase in the effective activity — does not normally reflect reality. At any given time, the rate of an enzymatic reaction is a function of the available substrates, products and modifiers as well as gene expression[13], and the formalism of metabolic control analysis[14–16] tells us that, although changes in the expression level of individual proteins might have little influence on fluxes, they can and do have large effects on the concentrations of intermediary metabolites. Consequently, the metabolome is expected — and found[17] — to be more sensitive to perturbations than either the transcriptome or proteome. This is because the activities of metabolic pathways are reflected more accurately in the concentrations of pools of metabolites than in the concentrations of the relevant enzymes (or indeed the concentrations of the mRNAs encoding them).

Metabolomics is therefore considered to be, in many senses, more 'useful' (that is, 'discriminatory') than transcriptomics and proteomics, for the following reasons. Metabolomics is 'downstream' — changes in the metabolome are amplified relative to changes in the transcriptome and proteome, and are arguably numerically more tractable[1]. There is no need for a whole genome sequence or for large expressed-sequence-tag databases to be available for each species. Metabolic profiling is cheaper and more high-throughput than proteomics and transcriptomics, making it feasible to examine large numbers of samples from organisms that have been 'grown' under a wide range of conditions. Finally, the technology involved in metabolomics is generic, as a given metabolite — unlike a transcript or protein — is the same in every organism that contains it (other than for secondary metabolites, as originally defined[18]).

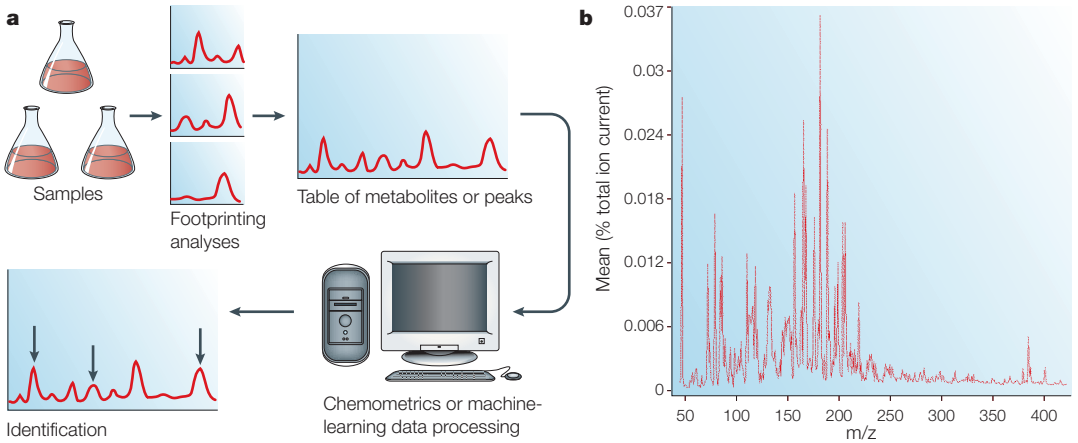


Figure 1 | **The basis of metabolic footprinting. a** | The basic metabolic footprinting strategy, starting with samples and proceeding through analysis to data processing until a classification can be effected or a scientific conclusion can be drawn. This can lead to the acquisition and analysis of further samples and subsequent loops around this cycle. **b** | A typical metabolic footprint (H.M.D. & D.B.K., unpublished data) using direct-injection positive-ionization low-resolution (to unit mass) electrospray mass spectrometry[54]. Shown is the average footprint from 250 single-gene yeast knockout strains. Note that there are peaks at every m/z (mass/charge ratio) value. Glucose can be identified at m/z 181, 203 (Na^+ adduct) and 219 (K^+ adduct). Other metabolites tentatively identified are glycerol-NH_4^+ (m/z 110), valine (m/z 118), histidine (m/z 156) and phenylalanine-Na^+ or methionine-K^+ (m/z 188).

METABOLOME
Nominally all of the small-molecular-weight metabolites in a sample. For practical reasons, this is rarely, if ever, achieved using a single extraction method, and subsets of metabolites ('metabolic profiles') are more typically obtained.

TRANSCRIPTOME
All the mRNA molecules (transcripts) in a sample.

PROTEOME
Nominally all the proteins in a sample. This measurement is not usually achieved because of poor solubilities, especially of membrane proteins, and sometimes protein digests are used, from which peptides representing the proteome are analysed. The concept of the proteome should also include all the post-translational modifications that might occur.

CE-MS
A technique in which metabolites are separated using capillary electrophoresis and then analysed using mass spectrometry. Two different runs are used for cations and anions, whereas neutral molecules can be separated using techniques that give them an effective charge.

GC-MS
A high-resolution analytical technique in which molecules, typically derivatized to enhance their volatility, are separated and then identified using mass spectrometry.

LC-MS
A technique in which metabolites are separated according to their polarity. In metabolomics, reverse-phase chromatography (in which the column is hydrophobic and an organic solvent such as methanol or acetonitrile is used for elution) is most popular, although polar chromatographies are also used.

METABOLIC FOOTPRINTING
A strategy for analysing the properties of cells or tissues by looking in a high-throughout manner at the metabolites that they excrete or fail to take up from their surroundings.

METABOLIC FINGERPRINTING
Classification of samples on the basis of their biological status or origin, using high-throughput methods, usually spectroscopic.

However, measurement of a complete set of intracellular microbial metabolites not only requires rapid quenching of metabolism[19,20] and a time-consuming and often inadequate extraction and separation procedure[20–22], but also the development of a method suitable for detecting and quantifying large numbers of metabolites that can be present at widely differing concentrations. Although several diverse techniques have previously been used for this purpose, including enzymatic metabolite quantification[12], nuclear magnetic resonance (NMR) spectroscopy[12,23], capillary electrophoresis coupled to mass spectrometry (CE-MS)[24], gas chromatography coupled to mass spectrometry (GC-MS)[25–28] and liquid chromatography coupled to mass spectrometry (LC-MS)[29–32], none of these completely fulfils the need for an accurate, simple and, in particular, rapid method with a broad dynamic range. We note especially the rapid timescale for the turnover of an intracellular metabolite (the turnover time is approximately equal to the concentration of the metabolite divided by the flux through the pathway of which it is a member[33]), which can be under 1 second in microbial systems, even for metabolites at mM concentrations[19].

Metabolite secretion, excretion and footprinting

It has long been known that microorganisms (and humans[7,23,34,35]) secrete a large number of metabolites into their external environment (especially under conditions of unbalanced growth). Fermented beverages provide an obvious example of what is now known to be microbial activity. Other examples include staling factors[36,37], uracil and xanthine[38], 1,5-anhydroglucitol[39], furanones[40,41], a vast array of intercellular signalling molecules[18,42–45], many charged but uncharacterized metabolites[46] and the many biologically active products of both 'primary'[47,48] and 'secondary' metabolism[18], which are often of applied interest[49].

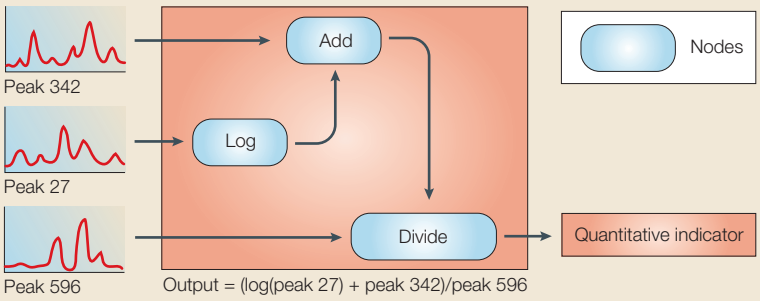
As such secretory activities clearly reflect cellular metabolic activity (and the level of transcription and translation of relevant genes), it occurred to us that the global measurement of secreted metabolites might be exploited to analytical advantage. In particular, the rapid quenching required to profile intracellular metabolites would be avoided, as the turnover time of metabolites diluted into the greater extracellular space is decreased in proportion to the relative ratios of extracellular and intracellular volumes. Although the characterization of metabolites in microbial culture broths has already been carried out using various detection methods[50,51] (and in culture off-gas using an electronic 'nose'[48,52,53]), such analyses were largely confined to specific metabolites and have not been done on a global scale.

Consequently, we devised METABOLIC FOOTPRINTING[54] as a novel method for the functional analysis and characterization of cells using the metabolome (FIG. 1). Metabolic footprinting as described relies not on the measurement of intracellular metabolites (a technique which is widely referred to as METABOLIC FINGERPRINTING[55]) but on the monitoring of metabolites consumed from, and secreted into, the growth medium by batch cultures of yeast using direct-injection MS, in which samples of culture media are injected directly into an electrospray ionization mass spectrometer. To maximize the excretion of metabolites, overflow metabolism[56] is stimulated by adding to the fully defined medium various carbon compounds that 'probe' metabolically active networks in the same way that an engineer might probe an electrical circuit[57]. Although such reasoning is not entirely watertight[58], it is reasonable to infer that those metabolites that do not appear in the medium in a particular gene knockout (or conversely those that increase in concentration when the encoding gene is cloned

Box 1 | **Genetic programming**

Genetic programming is an evolutionary computing or machine-learning technique. With such methods, we typically seek descriptive relationships between complex paired sets of highly multivariate inputs (in this case metabolite peak levels) and outputs representing a higher-order property of interest (for example, the presence or severity of, or susceptibility to, a disease, or a functional knockout, mode of action, physiological trait, and so on). Individual rules are encoded in the form of a tree (see figure, also known as a 'parse tree') in which the terminals represent the inputs, and nodes (curved rectangles) represent arithmetic operators such as +, –, *, /, log, √, or logical operators such as 'IF…THEN…', and so on. The tree is read from the terminals to the top and is assigned a fitness according to preference (usually this incorporates the idea that fitter individuals should have more examples classed correctly, and that simpler explanations are to be preferred over, and are fitter than, complex explanations). In each generation, typically thousands of such rules are evaluated and selection is used to favour (though not exclusively) those rules that are fitter. New rules are formed by manipulations that cause the mutation and recombination of these rules (trees), which occurs by changing or swapping parts of the tree below a node (this automatically preserves semantic correctness). This process continues until a desired stopping point (a certain number of generations or ideally a pre-determined degree of success on a 'hold-out' or 'validation' set not used in the formation of the rules). Finally, the rules are evaluated on an independent test set (one that was not used even in setting the stopping criterion).

Although there are some antecedents, genetic progamming was developed and popularized by John Koza[63,131]. For other surveys see REFS 61,62,132–135, and for evolutionary computing more generally REF. 136. In the example given in the text[60], a rule was evolved for a particular mode of action that decreased the number of variables involved in describing the effects of this mode of action on the exometabolome of yeast cells from 935 to just 2, therefore making the potential structural determination of the metabolites involved in these rules much easier.



downstream of a high-expression promoter) are produced through flux- or information-controlling pathways that more or less directly involve the gene product in question.

Metabolic footprinting has proved useful for detecting the patterns of metabolites in single-gene-knockout strains in functional genomic analyses[54,59] as well as in mode-of-action studies[60], in which the pattern of metabolites excreted when strains are challenged with sub-lethal concentrations of growth inhibitors makes it possible to discriminate the site or mode of action of those inhibitors. GENETIC PROGRAMMING[61–63] (BOX 1) allows the evolution of simple rules that can both discriminate the mode of action and explain how this discrimination works[60]. It is evident that this kind of metabolic footprinting strategy could be useful in ('forward') chemical genetics and/or chemical genomics[64,65] in which xenobiotic compounds affect cells at unknown sites of interaction.

GENETIC PROGRAMMING
A powerful but simple computational technique with which rules are evolved that can be used to solve classification or regression problems, for instance by using the metabolome or metabolic fingerprinting data as the inputs.

GC-TOF MS
A high-resolution analytical technique in which molecules, typically derivatized to enhance their volatility, are separated and then identified using time-of-flight mass spectrometry.

Global metabolite patterns

Because direct-injection MS is high-throughput, we could screen 250 mutants in triplicate over a period of 3 days. The large amounts of data (750 samples; mass/charge ratio (m/z values) from 45 to 1,000) reveal some interesting patterns when analysed in a 'global' sense using a data pipeline approach[66]. Some m/z values (relating to metabolites (M) present as $[M+H]^+$, $[M+NH_4]^+$, $[M+Na]^+$ or $[M+K]^+$ ions as minimal ion fragmentation is observed) are highly correlated with each other, indicating that they are metabolically linked or are different molecular ions of the same metabolite. Those m/z values that are not related to the same metabolite but are correlated with other metabolites are of greater interest[67], and exact mass measurements can assist in finding them. A similar phenomenon has been observed when studying the natural co-variation of metabolites in individual plants[68–70], but it is clearer when the range of concentrations is broadened (typically 3.5-fold) by looking at a series of gene knockouts (FIG. 2a). Other relationships between m/z values, such as a strong anti-correlation, are also commonly observed. More interesting perhaps is a frequent bimodal distribution in which individual m/z values show clear relationships, groups and clusters even if they are not well correlated *per se* (FIG. 2b). The analysis of the contribution of different m/z values to the total ion current (for example, FIG. 2c) also frequently shows bimodality. One interpretation of these observations is that metabolism is organized in 'modules'[71–73] (as is common in transcription networks[74–77]), so that a knockout in one module might have a large influence on flux within this module, whereas a knockout affecting a gene product not in this module will have correspondingly little effect (therefore bimodality: a gene product is either in a particular module or not). A second conclusion that can be drawn from plots such as that in FIG. 2c is that most gene knockouts have pleiotropic effects (see also REFS 78,79), causing changes in almost all m/z values. This emphasizes the power of metabolomics in providing a rich and highly discriminatory source of data for functional genomics[54]. Additionally, some gene knockouts will simply cause a slow-growth phenotype, for which the footprint is less precise in terms of the specific lesion; the same is true for transcriptome analysis[80].

From patterns to metabolites

The disadvantage of direct-injection MS is that we trade off knowledge of which metabolite(s) each m/z value represents for speed (the direct-injection method produces reliable spectra in under a minute). More recently, we have turned to methods that allow us to identify the metabolites in question chemically. Of these methods, gas chromatography coupled to time-of-flight mass spectrometry (GC-tof MS) is probably the most widely used[6,9,26]. By combining evolutionary algorithms[81,82] with an entirely automated closed-loop strategy[83], we have succeeded in doubling the number of peaks observed in yeast supernatants to >350 and trebling the number of peaks observed in human

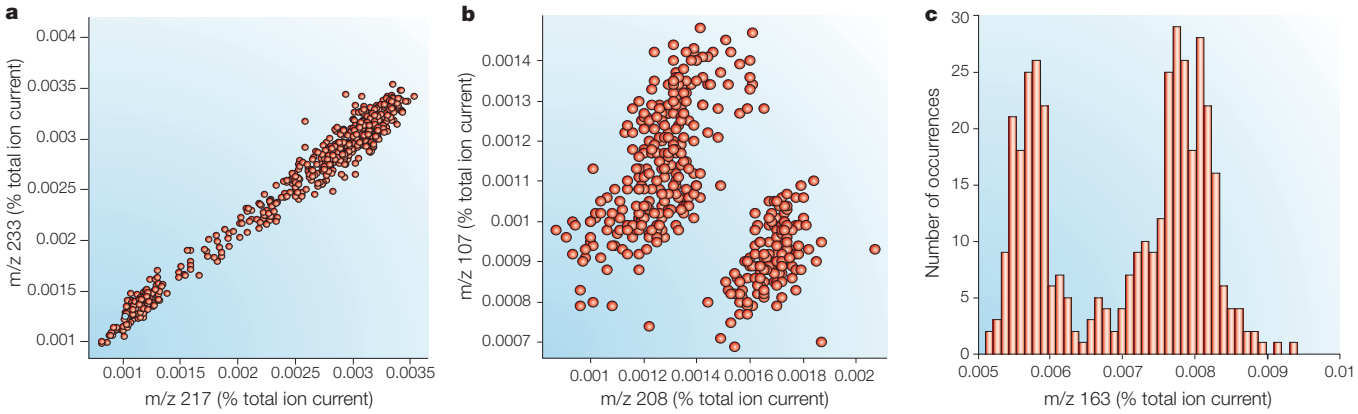


Figure 2 | **Patterns of m/z (mass/charge ratio) values in large-scale gene-knockout metabolic footprinting experiments.** The values plotted are derived from the m/z values normalized to total ion count from 250 single-gene knockout experiments carried out in triplicate. **a** | Strong correlation between m/z 233 and 217 (these could be Na^+ and K^+ adducts of the same metabolite). **b** | Bimodal clustering of pairs of m/z values (here 107 and 208), consistent with a modular view of metabolism. **c** | The variation of the contribution of m/z 163 to the total ion current.

serum to ~950 (REF. 28), while keeping the run time down to 13 and 20 minutes, respectively, and maintaining a good signal/noise (S/N) ratio. All the detected peaks are a combination of metabolite peaks, derivatization of which can produce more than one product per metabolite, and to a lesser extent blank related compounds such as phthalates. For yeast supernatants, >300 different metabolite peaks have been detected by one-dimensional GC-tof MS, and a typical result is 100–200 metabolite peaks per sample, corresponding to 50–150 metabolites. Current research strategies include building metabolomics-based retention index/mass spectral libraries to assist in metabolite identification[84]. In the authors' laboratories, ~70 metabolite peaks have been identified using chemical standards, and work is ongoing.

FIGURE 3a shows a typical GC-tof MS total ion chromatogram of yeast supernatant, with some of the metabolites identified. This approach is semi-quantitative to aid the high-throughput approach. The application of emerging orthogonal multidimensional separation techniques, including GC × GC-tof MS, will increase both the number of metabolites detected and the measured S/N ratio[85–88]. A typical GC × GC-tof MS 3-dimensional plot of yeast supernatant is shown in FIG. 3b; up to 500 peaks can be detected in yeast supernatant, not all of which are visible at a single magnification.

From yeast to humans

Most of our work has used yeast, for all the usual and obvious reasons (for example, easy and safe to work with and grow, the availability of 'bar-coded' gene-knockout mutants and the absence of ethical issues). However, yeast is also well recognized as a good model for mammalian studies[89] — 42% of yeast genes have human homologues[90] and trans-complementation has been shown for some genes[91]. One can also remark that, from the point of view of cells and tissues, biological fluids such as urine and even serum or plasma are the extracellular space into which molecules are secreted. Sampling these fluids is therefore, in some senses, equivalent to metabolic footprinting (and equivalently integrates the signals from all cells in a differentiated organism). A GC × GC-tof MS trace from human serum (FIG. 3d) shows both the dynamic range and the extent to which many overlapping peaks in the first dimension are successfully separated in the second dimension (>2,000 peaks are detectable with S/N>10 and relate to ~1,000 metabolite peaks; see also REF. 88).

Future directions: data standards and curation

An important need for large-scale studies of the type facilitated by metabolic footprinting is the ability to store metabolomics data in well-designed and curated databases that can store, handle and disseminate large amounts of metabolomics data efficiently and readily lend themselves to data mining and machine learning. Various omics data models (colloquially 'data standards') have been proposed, focusing on specific research areas and analytical methods. For example, MIAME (minimum information about a microarray experiment)[92] and MAGE-OM (microarray-gene-expression object model)[93], and PEDRo/PSI (proteomics experimental data repository/proteomics standards initiative)[94–97] are the emerging standards for transcriptomics and proteomics, respectively. Similar attempts are now coming through for metabolomics[98–100], and other MS data standards such as mzXML (in which MS data-acquisition files are converted into extensible markup language, XML[101]) will be useful for both proteomics and metabolomics.

These data models all enable the storage of both metadata and analytical data. Storing the metadata (that is, information about the specific conditions, protocols and parameters used in wet experiments) together with the experimental results is necessary to interpret the data and support comparison and reproducibility. The metadata types supported in the

GC x GC-tof MS
Similar to GC-tof MS, but involving two stages of gas-chromatographic separation in which a sample is taken (split) from the first dimension of the separation (typically carried out using a non-polar stationary-phase material) and effecting further 'orthogonal' separation (typically using a more polar stationary phase).

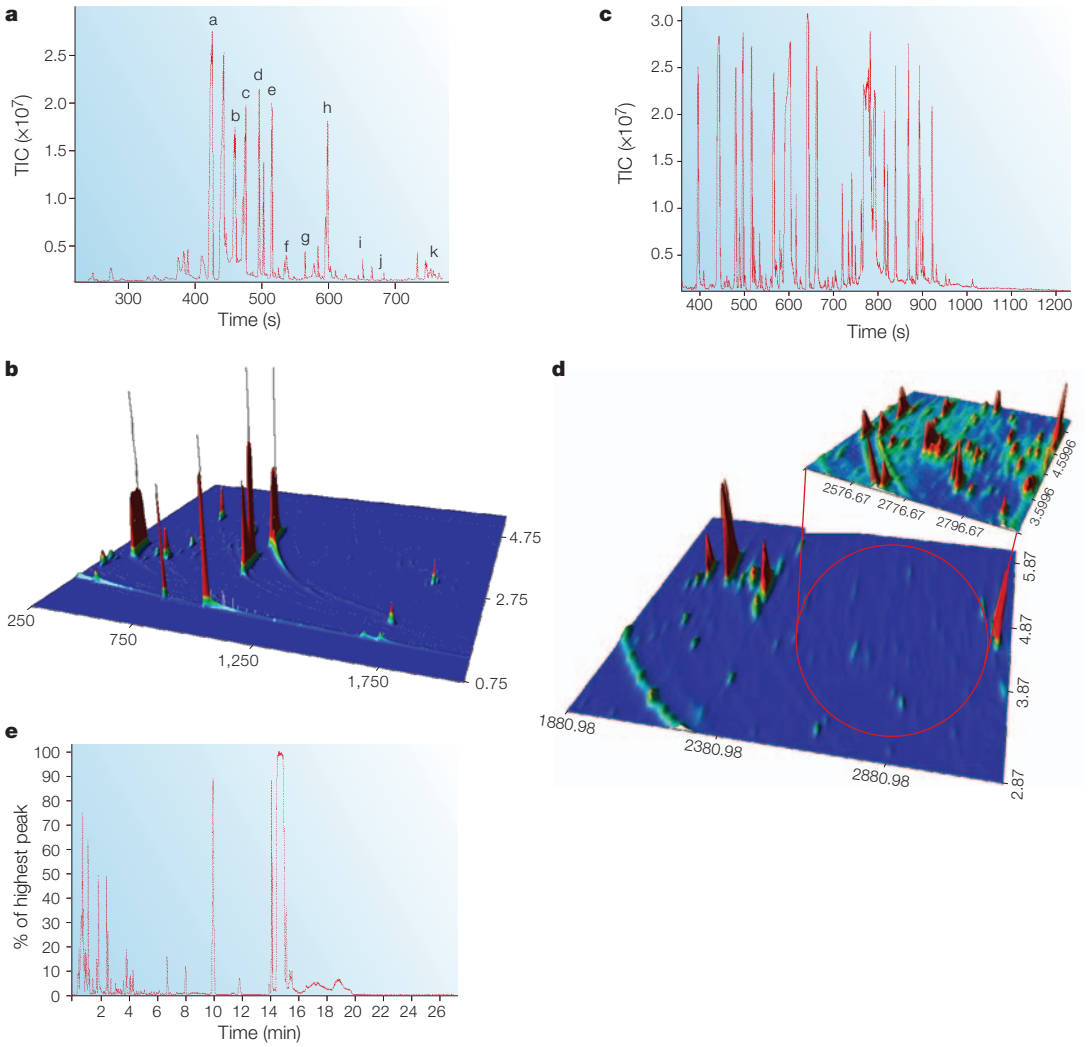


Figure 3 | **Modern technology applied to metabolic footprinting. a** | Gas chromatography coupled to time-of-flight mass spectrometry (GC-tof MS) total ion chromatogram for a typical metabolic footprint of *Saccharomyces cerevisiae*. a, glycerol; b, malonic acid; c, succinic acid; d, uracil; e, malic acid; f, 2-hydroxyglutaric acid; g, 2-oxoglutaric acid; h, citric acid; i, tryptophan; j, stearic acid; k, disaccharides. **b** | GC × GC-tof MS total ion chromatogram 3D plot for a typical metabolic footprint of *Saccharomyces cerevisiae*. **c** | GC-tof MS total ion chromatogram of human serum. **d** | GC × GC-tof MS traces of the human serum metabolome. **e** | Base peak ion chromatogram of human serum collected by UPLC (ultra-performance liquid chromatography)-tof MS. More than 10,000 peaks are typically detected with this technique, with the signal/noise ratio >5 after chromatogram deconvolution using the MarkerLynx software (Waters). The identification and number of metabolites and blank related compounds is not yet fully established, though it can be expected that 2,000–5,000 peaks are endogenous-metabolite-related, with others coming from xenobiotic metabolism in the gut and elsewhere. TIC, total ion current.

omics data models depend on the type of data that are acquired by analytical experiments, and the analytical techniques used to produce these data can therefore be seen as common ground in these models. For example, some experimental techniques and data features (such as peaks and fractions) are shared by proteomics and metabolomics. Therefore, there is room to standardize common parts across different omics models to facilitate their incorporation into an overall schema that reduces redundancies and reuses common components. For example, SysBio-OM (systems-biology object model)[102] has been developed as a single schema to document transcriptomics, proteomics and metabolomics experiments using a common platform to represent all three types of data. This has been done by integrating proteomics data (supported by PEDRo[94–96]) and metabolomics experiments (in particular MS and NMR methods) into the MAGE-OM model for transcriptomics[93].

Obviously, what must be supported in these data models are the analytical techniques. A good way to standardize models across different fields of research is to develop a universal separate module that can be fitted into each different data model (for example, by using name spaces in XML schemas) or reproduced if necessary. At present, there are plenty of 'dialects' describing analytical methods and data (BOX 2 lists some) but still no widely accepted standard. Some models (for example, AGML (annotated gel markup language)) target specific techniques, whereas others

Box 2 | **Some omics and related markup languages**

- **AGML (annotated gel markup language): specialized markup language for 2DE/MS experimental data[137].**
- **SpectroML[138]and mzXML[101]: markup languages for spectroscopic data.**
- **anIML (analytical information markup language): an effort to develop a single standard for analytical data[139].**
- **GAML (generalized analytical markup language): a compact, generalized format for analytical data.**
- **SBML (systems biology markup language)[105,106,140]: a still-developing markup language that is now widely accepted as the information standard for systems biology.**

(for example, anIML (analytical information markup language)) cover a range of different techniques.

Models have also been developed for some of the analytical aspects of metabolomics. Some parts (for example, MS and GC-MS experiments) already exist in ArMet[98] (architecture for metabolomics, a framework for the description of plant metabolomics experiments and their results), but we have also extended our model to support other analytical techniques, including NMR and Fourier-transform infrared (FT-IR) spectroscopy. Our approach is instrument-orientated (a separate schema for each machine) for three main reasons: the current lack of common standards; easy conversion of metadata into the schema format; and to prevent the loss of instrument-specific metadata that can be caused by using minimum information to describe a certain technique. We use an extensible modular approach in which different metadata modules (implemented as separate XML schemas) can be inputted into the overall metabolomics schema (both relational and XML versions). The XML representation (using the XML schema language) also facilitates the conversion from one format into another by using freely available tools (for example, Java Web Services Developer Pack) for automatic processing of the corresponding XML documents.

As predicted by Achard and colleagues[103], XML is becoming the predominant means of information modelling in the biological sciences (including metabolomics[104]), allowing the design of customized markup languages (BOX 2), web-enabled data exchange and full data management, including modelling (schemas defined by document type definition (DTD) or XML schema), storing (XML documents that can be stored as files or in databases) and querying (using XML query languages, for example, Xquery or XPath). Its scalability is particularly suitable for biological applications because new types of data and relationships are constantly emerging; data mining and machine learning also produce new data and complex models can be built using different data types and relationships.

It usually takes time for a research community to agree on, and fully embrace, new standards. However, the systems biology experts correctly recognized the potential of XML. Systems biology markup language (SBML)[105,106] is being developed as an XML format to represent models of biochemical reaction networks. Although far from complete, it has been widely accepted within the systems biology community as an information standard, allowing complex models to be shared, evaluated and developed cooperatively. Metabolomic models represented by SBML must still be integrated with proteome and metabolome data, and this integration represents a substantial part of the systems biology agenda. Specifically, the integrated understanding of cellular function requires large amounts of omics data, but the development of this understanding has been slowed because existing research programmes have focused on partial aspects (for example, transcriptomics, proteomics or metabolomics) of biological systems, neglecting the relationships between them.

Therefore systems biology, although not so new in its essential thinking[107,108], has emerged as an approach to the study of biological systems that aims to understand how the heterogeneous parts combine to form the whole[109,110] through systematic integration of technology, biology and computation[111]. As such, it shares the same agenda as metabolic control analysis[14–16,112] and biochemical systems theory[113]. Data-mining and machine-learning techniques applied to databases of omics data to extract hidden patterns, followed by simulation-based analyses to generate predictions that can be further tested *in vivo* and/or *in vitro,* can help achieve the systems-biology objective[114]. The first step, however, is to develop and standardize the different omics data models and fill them with curated data. The proteomics field has significantly advanced in such efforts: the Human Proteome Organization (HUPO) support the Protein Standards Initiative (PSI)[97,115,116] to develop an exchange standard for proteomics data (MIAPE — minimum information to describe a proteomics experiment) based on the PEDRo schema[94–96] (itself using the Pedro data-entry tool[95]). It is now essential for similar steps to be taken by the metabolomics community[99].

Concluding remarks

Metabolic footprinting was developed as a convenient, reproducible and high-throughput technique for the genome-wide, physiological-level characterization of microorganisms. The fluxes to extracellular products depend on the metabolic pathways operating and, in most cases, the metabolic footprint of even single-gene knockouts differs markedly from that of otherwise isogenic wild-type organisms[54]. As anticipated[117], one can 'calibrate' the system with knockouts of genes with known function, to infer or suggest the function of other genes with unknown function. Equivalently, large and reproducible changes in the exometabolome occur when cells are treated with sub-lethal concentrations of inhibitors with particular modes of action, thereby also allowing one to infer this[60]. Improved technical developments, such as CE-MS[24], high-throughput LC-MS[118,119] (FIG. 3e) and multi-dimensional chromatography, including GC × GC-tof MS, will enable wider coverage of the metabolome and make these measurements more straightforward and discriminatory.

2DE/MS
A proteomics technique in which proteins are separated according to their mass and charge, using two-dimensional gel electrophoresis, and subsequently identified, and sometimes quantified, using mass spectrometric methods.

Metabolic footprinting will also have applications for metabolic engineering. Metabolic engineering seeks to establish those gene products that are important in increasing the productivity of biological processes, and to manipulate their concentrations or activities accordingly[112,120,121]. As such, the detection in the extracellular space of all metabolites that are not the desired products of interest shows directly which non-productive pathways are operating and therefore should, if possible, be eliminated[122,123] (with concomitant savings in purification costs as well as in productivity enhancement).

Of course, the main programme to which metabolic footprinting data will contribute strongly is the generation and testing of mathematical models of cell behaviour, whose iterative interplay with 'wet' experiments is the hallmark of systems biology[109,114]. Most constraint-based metabolic models of microbial cells[124,125] can (but mainly do not) take large-scale metabolite secretion into account. However, in many cases cells are not in balanced growth and therefore, for stoichiometric reasons alone, must excrete large amounts of metabolites. Indeed, quantitative models with too few 'branches to exit'[126] will not even form a steady state[127], although one can be enforced by parameter optimization[128]. Therefore, measurement of the metabolic footprint (or exometabolome) before, or as well as, measurement of the intracellular (or endo-) metabolome[9,129] can provide a useful set of constraints for metabolic models. Searching parameter space to optimize a metabolic model that can reproduce experimental data is difficult[128]. However, once the model can reproduce the exometabolome accurately, many of the parameters will be sufficiently well determined to provide good starting points for the fine-tuning with which they can also reproduce the endometabolome, first as steady-state 'snapshots'[16] and then as time series.

As a large-scale integrator of cellular activity, the metabolic footprint includes all the metabolites that (enter and) exit the cell, and their measurement provides a potent set of constraints on any cellular metabolic model. Therefore, with apologies to Marshall McLuhan[130], in metabolic footprinting the medium is the message.

1. Oliver, S. G., Winson, M. K., Kell, D. B. & Baganz, F. Systematic functional analysis of the yeast genome. *Trends Biotechnol.* **16**, 373–378 (1998).
First use of the term 'metabolome' in the literature.
2. Fiehn, O., Kloska, S. & Altmann, T. Integrated studies on plant biology using multiparallel techniques. *Curr. Opin. Biotechnol.* **12**, 82–86 (2001).
3. Fiehn, O. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp. Funct. Genomics* **2**, 155–168 (2001).
4. Harrigan, G. G. & Goodacre, R. (eds) *Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function Analysis* (Kluwer Academic Publishers, Boston, 2003).
The first book on metabolomics.
5. Weckwerth, W. Metabolomics in systems biology. *Annu. Rev. Plant Biol.* **54**, 669–689 (2003).
6. Goodacre, R., Vaidyanathan, S., Dunn, W. B., Harrigan, G. G. & Kell, D. B. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol.* **22**, 245–252 (2004).
7. Nicholson, J. K., Holmes, E., Lindon, J. C. & Wilson, I. D. The challenges of modeling mammalian biocomplexity. *Nature Biotechnol.* **22**, 1268–1274 (2004).
Stresses the role of intestinal microorganisms in contributing to the human metabolome.
8. van der Greef, J., Stroobant, P. & van der Heijden, R. The role of analytical sciences in medical systems biology. *Curr. Opin. Chem. Biol.* **8**, 559–565 (2004).
The importance of analytical sciences to omics and systems biology.
9. Kell, D. B. Metabolomics and systems biology: making sense of the soup. *Curr. Opin. Microbiol.* **7**, 296–307 (2004).
Metabolomic data as an input to systems biology models.
10. Mendes, P., Kell, D. B. & Westerhoff, H. V. Why and when channeling can decrease pool size at constant net flux in a simple dynamic channel. *Biochim. Biophys. Acta* **1289**, 175–186 (1996).
11. Oliver, S. G. Yeast as a navigational aid in genome analysis. *Microbiology* **143**, 1483–1487 (1997).
12. Raamsdonk, L. M. *et al.* A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nature Biotechnol.* **19**, 45–50 (2001).
13. ter Kuile, B. H. & Westerhoff, H. V. Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. *FEBS Lett.* **500**, 169–171 (2001).
14. Fell, D. A. *Understanding the Control of Metabolism* (Portland Press, London, 1996).
Excellent introduction to metabolic control analysis.
15. Heinrich, R. & Schuster, S. *The Regulation of Cellular Systems* (Chapman & Hall, New York, 1996).
16. Kell, D. B. & Mendes, P. in *Technological and Medical Implications of Metabolic Control Analysis* (eds Cornish–Bowden, A. & Cárdenas, M. L.) 3–25 (Kluwer Academic Publishers, Dordrecht, 2000).
17. Urbanczyk-Wochniak, E. *et al.* Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. *EMBO Rep.* **4**, 989–993 (2003).
18. Kell, D. B., Kaprelyants, A. S. & Grafen, A. On pheromones, social behaviour and the functions of secondary metabolism in bacteria. *Trends Ecol. Evol.* **10**, 126–129 (1995).
19. De Koning, W. & van Dam, K. A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. *Anal. Biochem.* **204**, 118–123 (1992).
20. Teusink, B., Baganz, F., Westerhoff, H. V. & Oliver, S. G. in *Methods in Microbiology: Yeast Gene Analysis* (eds Tuite, M. F. & Brown, A. J. P.) 297–336 (Academic Press, London, 1998).
21. Gonzalez, B., Francois, J. & Renaud, M. A rapid and reliable method for metabolite extraction in yeast using boiling buffered ethanol. *Yeast* **13**, 1347–1355 (1997).
22. Castrillo, J. I., Hayes, A., Mohammed, S., Gaskell, S. J. & Oliver, S. G. An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry. *Phytochem.* **62**, 929–937 (2003).
23. Lindon, J. C., Nicholson, J. K. & Everett, J. R. NMR spectroscopy of biofluids. *Annu. Rep. NMR Spectrosc.* **38**, 1–88 (1999).
24. Soga, T. *et al.* Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *J. Proteome Res.* **2**, 488–494 (2003).
25. Greenaway, W., May, J., Scaysbrook, T. & Whatley, F. R. Identification by gas chromatography-mass spectrometry of 150 compounds in propolis. *Z. Naturforsch.* **46**, 111–121 (1991).
A pioneering paper in plant metabolomics.
26. Fiehn, O. *et al.* Metabolite profiling for plant functional genomics. *Nature Biotechnol.* **18**, 1157–1161 (2000).
An important paper that initiated the modern era of plant metabolomics.
27. Roessner, U. *et al.* Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* **13**, 11–29 (2001).
28. O'Hagan, S., Dunn, W. B., Brown, M., Knowles, J. D. & Kell, D. B. 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 (2005).
Explicit recognition of the need for optimization in metabolomics, and a new way of doing it.
29. van Eijk, H. M. H., Rooyakkers, D. R., Soeters, P. B. & Deutz, N. E. P. Determination of amino acid isotope enrichment using liquid chromatography-mass spectrometry. *Anal. Biochem.* **271**, 8–17 (1999).
30. Buchholz, A., Takors, R. & Wandrey, C. Quantification of intracellular metabolites in *Escherichia coli* K12 using liquid chromatographic–electrospray ionization tandem mass spectrometric techniques. *Anal. Biochem.* **295**, 129–137 (2001).
31. Lenz, E. M., Bright, J., Knight, R., Wilson, I. D. & Major, H. A metabonomic investigation of the biochemical effects of mercuric chloride in the rat using 1H NMR and HPLC-TOF/MS: time dependent changes in the urinary profile of endogenous metabolites as a result of nephrotoxicity. *Analyst* **129**, 535–541 (2004).
32. Colón, L. A., Cintron, J. M., Anspach, J. A., Fermier, A. M. & Swinney, K. A. Very high pressure HPLC with 1 mm id columns. *Analyst* **129**, 503–504 (2004).
33. Welch, G. R. & Easterby, J. S. Metabolic channeling versus free diffusion: transition-time analysis. *Trends Biochem. Sci.* **19**, 193–197 (1994).
34. Rashed, M. S. *et al.* Screening blood spots for inborn errors of metabolism by electrospray tandem mass spectrometry with a microplate batch process and a computer algorithm for automated flagging of abnormal profiles. *Clin. Chem.* **43**, 1129–1141 (1997).
35. Nicholson, J. K. & Wilson, I. D. Understanding 'global' systems biology: metabonomics and the continuum of metabolism. *Nature Rev. Drug Disc.* **2**, 668–676 (2003).
36. Barker, R. L., Gracey, D. E. F., Irwin, A. J., Pipasts, P. & Leiska, E. Liberation of staling aldehydes during storage of beer. *J. Inst. Brew.* **89**, 411–415 (1983).
37. Lynch, P. A. & Seo, C. W. Ethylene production in staling beer. *J. Food Sci.* **52**, 1270–1272 (1987).
38. Rinas, U., Hellmuth, K., Kang, R. J., Seeger, A. & Schlieker, H. Entry of *Escherichia coli* into stationary phase is indicated by endogenous and exogenous accumulation of nucleobases. *Appl. Env. Microbiol.* **61**, 4147–4151 (1995).
39. Shiga, Y., Mizuno, H. & Akanuma, H. Conditional synthesis and utilization of 1,5-anhydroglucitol in *Escherichia coli*. *J. Bacteriol.* **175**, 7138–7141 (1993).
40. Hayashida, Y., Kuriyama, H., Nishimura, K. & Slaughter, J. C. Production of 4-hydroxyfuranones in simple media by fermentation. *Biotechnol. Lett.* **21**, 505–509 (1999).
41. Slaughter, J. C. The naturally occurring furanones: formation and function from pheromone to food. *Biol. Rev. Camb. Philos. Soc.* **74**, 259–276 (1999).
42. Kaprelyants, A. S. & Kell, D. B. Do bacteria need to communicate with each other for growth? *Trends Microbiol.* **4**, 237–242 (1996).
43. Fuqua, C., Winans, S. C. & Greenberg, E. P. Census and consensus in bacterial ecosystems: the *LuxR–LuxI* family of quorum-sensing transcriptional regulators. *Annu. Rev. Microbiol.* **50**, 727–751 (1996).

44. Mukamolova, G. V., Kaprelyants, A. S., Young, D. I., Young, M. & Kell, D. B. A bacterial cytokine. *Proc. Natl Acad. Sci. USA* **95**, 8916–8921 (1998).
45. Weichart, D. H. & Kell, D. B. Characterization of an autostimulatory substance produced by *Escherichia coli*. *Microbiology* **147**, 1875–1885 (2001).
46. Firstenberg-Eden, R. & Eden, G. *Impedance Microbiology* (Research Studies Press, Letchworth, 1984).
47. Holms, H. Flux analysis and control of the central metabolic pathways in *Escherichia coli*. *FEMS Microbiol. Rev.* **19**, 85–116 (1996).
48. Wittmann, C. & Heinzle, E. MALDI-TOF MS for quantification of substrates and products in cultivations of *Corynebacterium glutamicum*. *Biotechnol. Bioeng.* **72**, 642–647 (2001).
49. Devlin, J. P. (ed.) *High Throughput Screening: The Discovery of Bioactive Substances* (Marcel Dekker, New York, 1997).
50. Abel, C. B. L. *et al.* Characterization of metabolites in intact *Streptomyces citricolor* culture supernatants using high-resolution nuclear magnetic resonance and directly coupled high-pressure liquid chromatography-nuclear magnetic resonance spectroscopy. *Anal. Biochem.* **270**, 220–230 (1999).
51. Bubb, W. A. *et al.* Heteronuclear NMR studies of metabolites produced by *Cryptococcus neoformans* in culture media: identification of possible virulence factors. *Magn. Res. Med.* **42**, 442–453 (1999).
52. Bachinger, T. & Mandenius, C. F. Searching for process information in the aroma of cell cultures. *Trends Biotechnol.* **18**, 494–500 (2000).
53. Gibson, T. D. *et al.* Detection and simultaneous identification of microorganisms from headspace samples using an electronic nose. *Sens. Actuators B* **44**, 413–422 (1997).
54. Allen, J. K. *et al.* High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nature Biotechnol.* **21**, 692–696 (2003).
The original paper describing metabolic footprinting.
55. Fiehn, O. Metabolomics: the link between genotypes and phenotypes. *Plant Mol. Biol.* **48**, 155–171 (2002).
56. Neijssel, O. M. & Tempest, D. W. The role of energy-splitting reactions in the growth of *Klebsiella aerogenes* NCTC 418 in aerobic chemostat culture. *Arch. Microbiol.* **110**, 305–311 (1976).
57. Lazebnik, Y. Can a biologist fix a radio? — Or, what I learned while studying apoptosis. *Cancer Cell* **2**, 179–182 (2002).
58. Westerhoff, H. V. & Kell, D. B. Matrix method for determining the steps most rate-limiting to metabolic fluxes in biotechnological processes. *Biotechnol. Bioeng.* **30**, 101–107 (1987).
59. Kaderbhai, N. N., Broadhurst, D. I., Ellis, D. I., Goodacre, R. & Kell, D. B. Functional genomics via metabolic footprinting: monitoring metabolite secretion by *Escherichia coli* tryptophan metabolism mutants using FT-IR and direct injection electrospray mass spectrometry. *Comp. Funct. Genomics* **4**, 376–391 (2003).
60. Allen, J. *et al.* Discrimination of modes of action of antifungal substances by use of metabolic footprinting. *Appl. Environ. Microbiol.* **70**, 6157–6165 (2004).
61. Langdon, W. B. *Genetic Programming and Data Structures: Genetic Programming + Data Structures = Automatic Programming!* (Kluwer Academic Publishers, Boston, 1998).
62. Kell, D. B. Genotype–phenotype mapping: genes as computer programs. *Trends Genet.* **18**, 555–559 (2002).
63. Koza, J. R. *et al. Genetic Programming: Routine Human–Competitive Machine Intelligence* (Kluwer Academic Publishers, New York, 2003).
64. Schreiber, S. L. Chemical genetics resulting from a passion for synthetic organic chemistry. *Bioorg. Med. Chem.* **6**, 1127–1152 (1998).
65. Huang, J. *et al.* Finding new components of the target of rapamycin (TOR) signaling network through chemical genetics and proteome chips. *Proc. Natl Acad. Sci. USA* **101**, 16594–16599 (2004).
66. Brown, M. *et al.* A metabolome pipeline: from concept to data to knowledge. *Metabolomics* **1**, 35–46 (2005).
67. Camacho, D., de la Fuente, A. & Mendes, P. The origins of correlations in metabolomics data. *Metabolomics* **1**, 53–63 (2005).
68. Kose, F., Weckwerth, W., Linke, T. & Fiehn, O. Visualizing plant metabolomic correlation networks using clique-metabolite matrices. *Bioinformatics* **17**, 1198–1208 (2001).
69. Fiehn, O. Metabolic networks of *Cucurbita maxima* phloem. *Phytochemistry* **62**, 875–886 (2003).
70. Steuer, R., Kurths, J., Fiehn, O. & Weckwerth, W. Observing and interpreting correlations in metabolomic networks. *Bioinformatics* **19**, 1019–1026 (2003).
71. Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabási, A. L. Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555 (2002).
72. Csete, M. E. & Doyle, J. C. Reverse engineering of biological complexity. *Science* **295**, 1664–1669 (2002).
Highlights the fact that most systems-biology problems are inverse problems, in which we must move from measured variables to the inference of parameters.
73. Barabási, A.-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nature Rev. Genet.* **5**, 101–113 (2004).
Useful review of network biology.
74. Bruggeman, F. J., Westerhoff, H. V., Hoek, J. B. & Kholodenko, B. N. Modular response analysis of cellular regulatory networks. *J. Theor. Biol.* **218**, 507–520 (2002).
75. Ihmels, J. *et al.* Revealing modular organization in the yeast transcriptional network. *Nature Genet.* **31**, 370–377 (2002).
76. Han, J. D. *et al.* Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* **430**, 88–93 (2004).
77. Tanay, A., Sharan, R., Kupiec, M. & Shamir, R. Revealing modularity and organization in the yeast molecular network by integrated analysis of highly heterogeneous genomewide data. *Proc. Natl Acad. Sci. USA* **101**, 2981–2986 (2004).
78. Hughes, T. R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).
79. Featherstone, D. E. & Broadie, K. Wrestling with pleiotropy: genomic and topological analysis of the yeast gene expression network. *Bioessays* **24**, 267–274 (2002).
80. Hayes, A. *et al.* Hybridization array technology coupled with chemostat culture: tools to interrogate gene expression in *Saccharomyces cerevisiae*. *Methods* **26**, 281–290 (2002).
81. Vaidyanathan, S., Broadhurst, D. I., Kell, D. B. & Goodacre, R. Explanatory optimisation of protein mass spectrometry via genetic search. *Anal. Chem.* **75**, 6679–6686 (2003).
82. Vaidyanathan, S., Kell, D. B. & Goodacre, R. Selective detection of proteins in mixtures using electrospray ionization mass spectrometry: influence of instrumental settings and implications for proteomics. *Anal. Chem.* **76**, 5024–5032 (2004).
83. King, R. D. *et al.* Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature* **427**, 247–252 (2004).
The 'robot scientist' strategy for principled and automated scientific hypothesis generation and testing.
84. Schauer, N. *et al.* GC-MS libraries for the rapid identification of metabolites in complex biological samples. *FEBS Lett.* **579**, 1332–1337 (2005).
85. Marriott, P. & Shellie, R. Principles and applications of comprehensive two-dimensional gas chromatography. *Trends Anal. Chem.* **21**, 573–583 (2002).
86. Wilson, I. D. & Brinkman, U. A. Hyphenation and hypernation: the practice and prospects of multiple hyphenation. *J. Chromatogr. A* **1000**, 325–356 (2003).
87. Guttman, A., Varoglu, M. & Khandurina, J. Multidimensional separations in the pharmaceutical arena. *Drug Discov. Today* **9**, 136–144 (2004).
88. Welthagen, W. *et al.* Comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (GC x GC-TOF) for high-resolution metabolomics: biomarker discovery on spleen tissue extracts of obese NZO compared to lean C57BL/6 mice. *Metabolomics* **1**, 65–73 (2005).
89. Castrillo, J. I. & Oliver, S. G. Yeast as a touchstone in post-genomic research: strategies for integrative analysis in functional genomics. *J. Biochem. Mol. Biol.* **37**, 93–106 (2004).
90. Foury, F. Human genetic diseases: a cross-talk between man and yeast. *Gene* **195**, 1–10 (1997).
91. Zhang, N. *et al.* Using yeast to place human genes in functional categories. *Gene* **303**, 121–129 (2003).
92. Brazma, A. *et al.* Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nature Genet.* **29**, 365–371 (2001).
93. Spellman, P. *et al.* Design and implementation of microarray gene expression markup language (MAGE–ML). *Genome Biol.* **3**, 9 (2002).
94. Taylor, C. F. *et al.* A systematic approach to modelling capturing and disseminating proteomics experimental data. *Nature Biotechnol.* **21**, 247–254 (2003).
95. Garwood, K. L. *et al.* Pedro: a configurable data entry tool for XML. *Bioinformatics* **20**, 2463–2465 (2004).
96. Garwood, K. L. *et al.* PEDRo: a database for storing, searching and disseminating experimental proteomics data. *BMC Genomics* **5** (2004).
97. Orchard, S., Hermjakob, H. & Apweiler, R. The proteomics standards initiative. *Proteomics* **3**, 1374–1376 (2003).
98. Jenkins, H. *et al.* A proposed framework for the description of plant metabolomics experiments and their results. *Nature Biotechnol.* **22**, 1601–1606 (2004).
Specific schemas for evolving metabolome data standards.
99. Bino, R. J. *et al.* Potential of metabolomics as a functional genomics tool. *Trends Plant Sci.* **9**, 418–425 (2004).
100. Kopka, J. *et al.* GMD@CSB. DB: the Golm metabolome database. *Bioinformatics* (2005).
101. Pedrioli, P. G. *et al.* A common open representation of mass spectrometry data and its application to proteomics research. *Nature Biotechnol.* **22**, 1459–1466 (2004).
102. Xirasagar, S. *et al.* CEBS object model for systems biology data, SysBio-OM. *Bioinformatics* **20**, 2004–2015 (2004).
103. Achard, F., Vaysseix, G. & Barillot, E. XML, bioinformatics and data integration. *Bioinformatics* **17**, 115–125 (2001).
104. Li, X. J. *et al.* in *Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function Analysis* (eds Harrigan, G. G. & Goodacre, R.) 293–309 (Kluwer Academic Publishers, Boston, 2003).
105. Hucka, M. *et al.* The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531 (2003).
A starting point for learning about SBML.
106. Finney, A. & Hucka, M. Systems biology markup language: level 2 and beyond. *Biochem. Soc. Trans.* **31**, 1472–1473 (2003).
107. von Bertalanffy, L. *General System Theory* (George Braziller, New York, 1969).
108. Iberall, A. S. *Toward A General Science of Viable Systems* (McGraw–Hill, New York, 1972).
109. Ideker, T., Galitski, T. & Hood, L. A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.* **2**, 343–372 (2001).
110. Henry, C. M. Systems biology. *Chem. Eng. News* **81**, 45–55 (2003).
111. Hood, L. Systems biology: integrating technology, biology, and computation. *Mech. Ageing Dev.* **124**, 9–16 (2003).
112. Kell, D. B. & Westerhoff, H. V. Metabolic control theory: its role in microbiology and biotechnology. *FEMS Microbiol. Rev.* **39**, 305–320 (1986).
Metabolic control analysis as the progenitor of metabolic engineering; why the metabolome is expected to be more discriminatory than the transcriptome or proteome.
113. Savageau, M. *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology* (Addison–Wesley, Reading, 1976).
114. Kitano, H. Systems biology: a brief overview. *Science* **295**, 1662–1664 (2002).
115. Orchard, S., Zhu, W., Julian, R. J., Hermjakob, H. & Apweiler, R. Further advances in the development of a data interchange standard for proteomics data. *Proteomics* **3**, 2065–2066 (2003).
116. Orchard, S. *et al.* Common interchange standards for proteomics data: public availability of tools and schema. *Proteomics* **4**, 490–491 (2004).
117. Kell, D. B. & King, R. D. 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 (2000).
Why supervised learning methods are more useful than unsupervised ones for functional genomics, and which constraints this entails.
118. Plumb, R. *et al.* 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 (2003).
119. Plumb, R. *et al.* Ultra-performance liquid chromatography coupled to quadrupole- orthogonal time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **18**, 2331–2337 (2004).
Very high resolution metabolomics using ultra-performance liquid chromatography.
120. Bailey, J. E. Toward a science of metabolic engineering. *Science* **252**, 1668–1675 (1991).

121. Stephanopoulos, G. & Sinskey, A. J. Metabolic engineering — methodologies and future prospects. *Trends Biotechnol.* **11**, 392–396 (1993).
122. Holms, W. H., Hamilton, I. D. & Mousdale, D. Improvements to microbial productivity by analysis of metabolic fluxes. *J. Chem. Technol. Biotechnol.* **50**, 139–141 (1991).
123. Askenazi, M. *et al.* Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing fungal strains. *Nature Biotechnol.* **21**, 150–156 (2003).
124. Reed, J. L. & Palsson, B. Ø. Thirteen years of building constraint-based *in silico* models of *Escherichia coli*. *J. Bacteriol.* **185**, 2692–2699 (2003).
125. Duarte, N. C., Herrgard, M. J. & Palsson, B. O. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.* **14**, 1298–1309 (2004).
A genome-scale metabolic model for *Saccharomyces cerevisiae*.
126. Kacser, H. The control of enzyme systems *in vivo*: elasticity analysis of the steady state. *Biochem. Soc. Trans.* **11**, 35–40 (1983).
127. Pritchard, L. & Kell, D. B. Schemes of flux control in a model of *Saccharomyces cerevisiae* glycolysis. *Eur. J. Biochem.* **269**, 3894–3904 (2002).
128. Mendes, P. & Kell, D. B. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics* **14**, 869–883 (1998).
129. van der Greef, J. *et al.* in *Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function Analysis* (eds Harrigan, G. G. & Goodacre, R.) 171–198 (Kluwer Academic Publishers, Boston, 2003).
130. McLuhan, M. & Fiore, Q. *The Medium is the Massage* (Penguin Books, London, 1971).
131. Koza, J. R. *Genetic Programming: on the Programming of Computers by Means of Natural Selection* (MIT Press, Cambridge, 1992).
132. Banzhaf, W., Nordin, P., Keller, R. E. & Francone, F. D. *Genetic Programming: An Introduction* (Morgan Kaufmann, San Francisco, 1998).
133. Kell, D. B., Darby, R. M. & Draper, J. Genomic computing: explanatory analysis of plant expression profiling data using machine learning. *Plant Physiol.* **126**, 943–951 (2001).
134. Kell, D. B. Defence against the flood: a solution to the data mining and predictive modelling challenges of today. *Bioinformatics World (part of Scientific Computing News)* **1**, 16-18 (2002).
135. Goodacre, R. & Kell, D. B. in *Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function Analysis* (eds Harrigan, G. G. & Goodacre, R.) 239–256 (Kluwer Academic Publishers, Boston, 2003).
136. Bäck, T., Fogel, D. B. & Michalewicz, Z. (eds) *Handbook of Evolutionary Computation* (IOP Publishing/Oxford University Press, Oxford, 1997).
137. Stanislaus, R., Jiang, L., Swartz, M., Arthur, J. & Almeida, J. An XML standard for the dissemination of annotated 2D gel electrophoresis data complemented with mass spectrometry results. *BMC Bioinformatics* **5** (2004).
138. Kramer, G., Ruehl, M. & Schafer, R. SpectroML — a markup language for molecular spectrometry data. *J. Assoc. Lab. Automat.* **6**, 76–82 (2001).
139. Davies, T., Lampen, P., Fiege, M., Richter, T. & Frohlich, T. AnIMLs in the spectroscopic laboratory? *Spectroscopy Europe* **15**, 25–28 (2003).
140. Shapiro, B. E., Hucka, M., Finney, A. & Doyle, J. MathSBML: a package for manipulating SBML-based biological models. *Bioinformatics* **20**, 2829–2831 (2004).

Acknowledgements

We thank the Biotechnology and Biological Sciences Research Council and the Natural Environment Research Council for financial support. D.B.K. is supported by the Royal Society of Chemistry and the Engineering and Physical Sciences Research Council. We thank N. Bukowski, J. Heald and H. Major for assistance with chromatographic and mass spectrometric analyses, and N. Burton for assistance with yeast microbiology and genetics.

Competing financial interests statement

The authors declare competing financial interests; see Web version for details.

Online links

FURTHER INFORMATION

Douglas Kell's laboratory: http://dbk.ch.umist.ac.uk/
AGML: http://s04.muh.musc.edu/%7Eromy/agml_central/index.php
AnIML: http://animl.sourceforge.net
ArMet: http://www.armet.org
GAML: http://www.gaml.org
Java Web Services Developer Pack: http://java.sun.com/webservices/jwsdp
MAGE-OM: http://www.mged.org/Workgroups/MAGE/mage.html
The Metabolic Control Analysis Web: http://dbk.ch.umist.ac.uk/mca_home.htm
MIAME: http://www.mged.org/index.html
mzXML: http://tools.proteomecenter.org/mzXMLschema.php
PEDRo/PSI: http://psidev.sourceforge.net
SBML: http://www.sbml.o rg
SysBioOM: http://cebs.niehs.nih.gov/content_sysbio_om.html
XML schemas: http://www.w3.org/XML/Schema
Access to this interactive links box is free online.