

THE THEODOR BÜCHER LECTURE

Metabolomics, modelling and machine learning in systems biology – towards an understanding of the languages of cells

Delivered on 3 July 2005 at the 30th FEBS Congress and 9th IUBMB conference in Budapest

Douglas B. Kell^{1,2}

1 School of Chemistry, Faraday Building, The University of Manchester, UK

2 Manchester Centre for Integrative Systems Biology, Manchester Interdisciplinary Biocentre, UK

Keywords

hypothesis generation; genetic programming; evolutionary computing; signal processing elements; technology development; systems biology

Correspondence

D.B. Kell, School of Chemistry, University of Manchester, Faraday Building, Sackville Street, Manchester M60 1QD, UK

Tel: +44 161 3064492

E-mail: dbk@manchester.ac.uk

Website: <http://dbk.ch.umist.ac.uk>, <http://www.mib.ac.uk/>, <http://www.mcisb.org/>

(Received 15 November 2005, revised 7 January 2006, accepted 16 January 2006)

doi:10.1111/j.1742-4658.2006.05136.x

The newly emerging field of systems biology involves a judicious interplay between high-throughput 'wet' experimentation, computational modelling and technology development, coupled to the world of ideas and theory. This interplay involves iterative cycles, such that systems biology is not at all confined to hypothesis-dependent studies, with intelligent, principled, hypothesis-generating studies being of high importance and consequently very far from aimless fishing expeditions. I seek to illustrate each of these facets. Novel technology development in metabolomics can increase substantially the dynamic range and number of metabolites that one can detect, and these can be exploited as disease markers and in the consequent and principled generation of hypotheses that are consistent with the data and achieve this in a value-free manner. Much of classical biochemistry and signalling pathway analysis has concentrated on the analyses of changes in the concentrations of intermediates, with 'local' equations – such as that of Michaelis and Menten $v = (V_{\max} \cdot S)/(S + K_m)$ – that describe individual steps being based solely on the instantaneous values of these concentrations. Recent work using single cells (that are not subject to the intellectually unsupportable averaging of the variable displayed by heterogeneous cells possessing nonlinear kinetics) has led to the recognition that some protein signalling pathways may encode their signals not (just) as concentrations (AM or amplitude-modulated in a radio analogy) but via changes in the dynamics of those concentrations (the signals are FM or frequency-modulated). This contributes in principle to a straightforward solution of the crosstalk problem, leads to a profound reassessment of how to understand the downstream effects of dynamic changes in the concentrations of elements in these pathways, and stresses the role of signal processing (and not merely the intermediates) in biological signalling. It is this signal processing that lies at the heart of understanding the languages of cells. The resolution of many of the modern and postgenomic problems of biochemistry requires the development of a myriad of new technologies (and maybe a new culture), and thus regular input from the physical

Abbreviations

MCA, metabolic control analysis; ODE, ordinary differential equations.

sciences, engineering, mathematics and computer science. One solution, that we are adopting in the Manchester Interdisciplinary Biocentre (<http://www.mib.ac.uk/>) and the Manchester Centre for Integrative Systems Biology (<http://www.mcisb.org/>), is thus to colocate individuals with the necessary combinations of skills. Novel disciplines that require such an integrative approach continue to emerge. These include fields such as chemical genomics, synthetic biology, distributed computational environments for biological data and modelling, single cell diagnostics/bionanotechnology, and computational linguistics/text mining.

The belief that an organism is ‘nothing more’ than a collection of substances, albeit a collection of very complex substances, is as widespread as it is difficult to substantiate...The problem is therefore the investigation of *systems*, i.e. components related or organized in a specific way. The properties of a system are, in fact, ‘more’ than (or different from) the properties of its components, a fact often overlooked in zealous attempts to demonstrate ‘additivity’ of certain phenomena. It is with the ‘systemic properties’ that we shall be mainly concerned.

H. Kacser (1957) in *The Strategy of the Genes* (ed. CH Waddington), pp. 191–249. Allen & Unwin, London

Progress in science depends on new techniques, new discoveries, and new ideas, probably in that order.

Sydney Brenner, *Nature*, June 5, 1980

Systems biology as such is not especially new [1–3], but while it is not hard to find prescient comments from Henrik Kacser and from Sydney Brenner [4], those given above might be seen as epitomizing the key features of the more recent move towards, and interest in, Systems Biology [5–14] (Fig. 1).

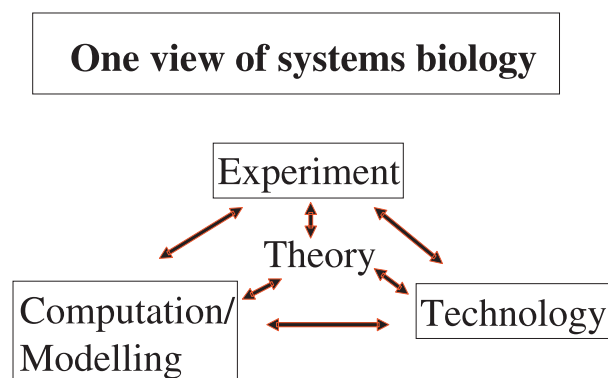


Fig. 1. Systems biology is usually seen as an iterative activity integrating computational work, high-throughput ‘wet’ experimentation and technology development with the world of theory and novel ideas.

Parallelling the Brenner quote, my lecture also chose to highlight three aspects of our current work with collaborators. The first involves the philosophical underpinnings of our scientific strategy and of the systems biology agenda, which can each be considered to involve an iterative interplay [15–17] between a series of linked activities. These activities include data (observations) and ideas (hypotheses); theory, computation and experiment; and the iterative assessment of the parameters and variables in such computational models and experiments. The second area relates to the actual development of technology for systems biology, specifically analytical and computational technology – especially in metabolomics – to help provide both high quality data and the concomitant modelling that relies on it. The third strand develops various ideas that emerged following our recent findings [18–20] that protein signalling pathways – specifically those involving the nuclear transcription factor NF- κ B – may encode their signals not so much in terms of changes in the concentrations of the observable signalling intermediates but in terms of their frequency or dynamics. Such signals must be perceived by downstream signal processing elements that respond to their dynamics, and so to understand such pathways properly one needs to understand and focus on not only the intermediates (the medium) but also the ‘downstream’ means (‘network motifs’ – see, e.g. [21–23], or ‘design elements’ [24]) by which such signals are perceived (to make the message). This leads to a profoundly different view of the significance of networks in systems biology, and one that allows one a much better understanding of signalling as signal processing. Put another way, and again quoting Henrik Kacser [25,26], ‘But one thing is certain: to understand the whole one must study the whole’.

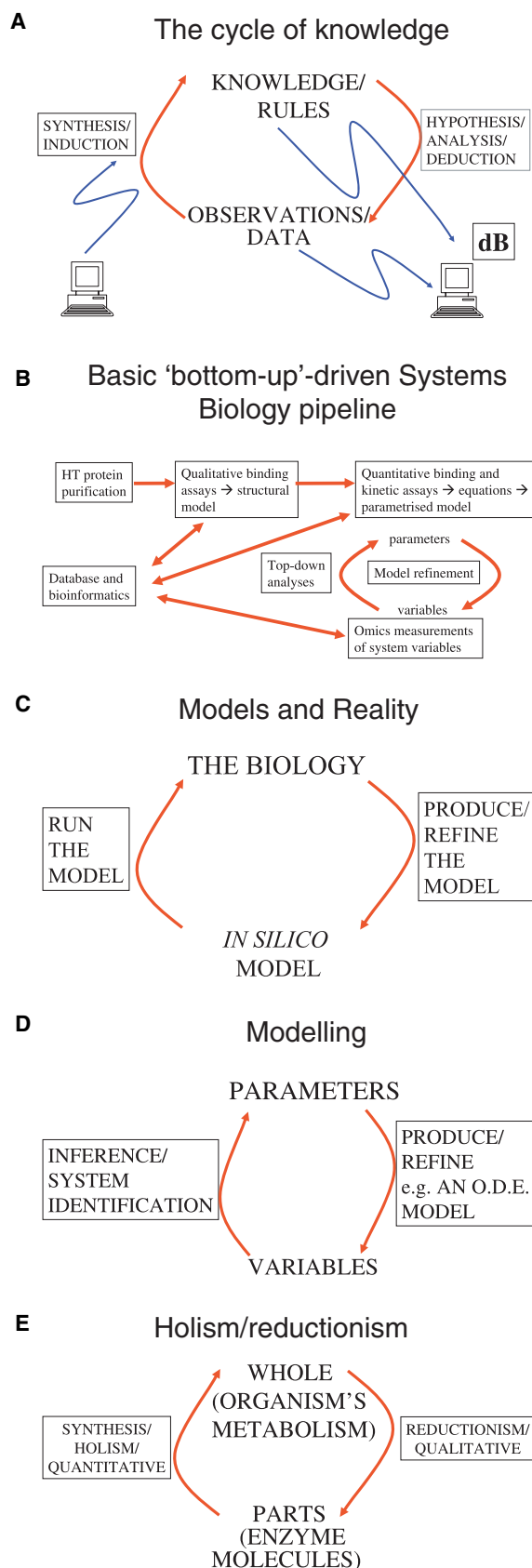
Philosophical elements of systems biology

As in Fig. 1, most commentators (summarized, e.g. in [12]), as I do [17,27], take the systems biology agenda to include pertinent technology development, theory,

computational modelling and high-throughput experimentation. Hypothesis-driven science is only a partial component of this, and not the major one [16]. More specifically, in systems biology, studies are performed purposively in an iterative manner, in a way that contrasts with previous strategies. This iteration is multi-dimensional, and can be described or seen in various ways, including both wet (experimental) and dry (computational and theoretical), reductionist and synthetic, qualitative and quantitative, and a systems biologist would lay more stress than is conventional on the right-hand arcs of the diagrams in Fig. 2. A particular feature is the ‘vertical’ focus of systems biology in seeking to relate ‘lower’ levels of biological organization such as enzymatic properties to higher levels of biological organization, and in this sense systems biology shares the same agenda as the long-established approaches of Metabolic Control Analysis [11,26,28–32]) and Biochemical Systems Theory [33,34].

It is a curious fact that in physics and chemistry (and indeed in economics) ‘theory’ has a status almost equal with that of experiment, and has claimed many Nobel Prizes, but in modern biology this is not the

Fig. 2. Some of the iterative elements of systems biology. (A) Science can be said to advance via an iterative interplay between the worlds of ideas and of experimental data. The world of ideas includes theories, hypotheses, human knowledge and any other mental constructs, while the world of data consists of experimental observations and other facts, sometimes referred to as ‘sense data’ in the philosophical literature as an iterative process, movement between these two worlds is not simply a reversible action: analysis is not the reverse of synthesis [339]. (B) One view of systems biology, reflecting a largely bottom-up approach, as in the ‘silicon cell’ [340]. First we need what we term a ‘structural model’ (this describes the network’s structure, and has nothing to do with structural biology) that defines the participants in the process of interest and the (qualitative) nature of the interactions between them; then we try to develop equations, preferably mechanistic rather than empirical, that best describe the relationships, then finally we seek to parameterize those equations (recognizing that if errors occur in the earlier phases we may need to return and correct them in the light of further knowledge). (C) The hallmark of modelling as a comparison between the mathematical models and the ‘reality’ (i.e. observed experimental data plus noise), again as an iterative process. (D) Producing and refining a model: data on kinetic parameters allow one to run a forward model. However, invoking such parameters from measured omics data (fluxes and concentrations) is referred to as an inverse or system identification problem (e.g. [86–88,90,91,341–347]) and is much harder. One strategy is to make estimates of the parameters and on the basis of the consequent forward model refine those estimates iteratively until some level of convergence (with statistical confidence levels) is achieved. (E) The iteration in models/mapping between levels of biological organization, e.g. in the case illustrated between the overall metabolism of an organism and its enzymatic parts.



case. 'Pure' theoreticians do not easily make a living (and only partly for sociological reasons connected with their perceived grant-winning abilities). Equivalently, it would be laughable for an engineer not to make a mathematical model of a candidate design for a bridge or an aeroplane before trying to build one, since the chance of it 'working' would be remote (because it is 'complex', and this is because its components are many and they act in nonlinear ways). By contrast, making mathematical models of the biological systems one is investigating (and seeing how they perform *in silico*) is generally considered a minority sport, and one not to be indulged in by those who prefer (or who prefer their postdocs and students) to spend more time with their pipettes.

Fairly obviously, it is easy to recognize that molecular biology concentrated perhaps too heavily on parts rather than wholes in its development, or at least that it is time, now that we have the postgenomic parts list of the genes and proteins (though not yet the metabolites) of most organisms of immediate interest, for working biologists to incorporate the skills of the numerical modeller (or indeed the radio engineer [35]), just as the more successful ones needed to become acquainted with the techniques of molecular biology when they began to be developed 30 years ago. In 10 years' time the referees of grant proposals and papers will normally ask only why one did not model one's system before studying it experimentally, not why one might wish to.

This said, it is useful to rehearse the variety of reasons why one might wish to model a biological systems that one is seeking to understand and study experimentally [36] (and see also [12,13,37]):

- testing whether the model is accurate, in the sense that it reflects – or can be made to reflect – known experimental facts. This amounts to 'simulation';
- analysing the model to understand which parts of the system contribute most to some desired properties of interest;
- hypothesis generation and testing, allowing one to analyse rapidly the effects of manipulating experimental conditions in the model without having to perform complex and costly experiments (or to restrict the number that are performed);
- testing what changes in the model would improve the consistency of its behaviour with experimental observations.

The last two points amount to 'prediction'.

The techniques of modelling

Most strategies for creating mathematical models of biological systems recognize that the nonoptical, high-

resolution experimental analysis of spatial distributions beyond macro-compartments is not yet available and thus it is appropriate to use ordinary differential equations (ODEs) that assume such compartments both to be well-stirred and with their components in high enough concentrations that they are 'homogeneous'. If the former assumption breaks down one can create subcompartments [38], while the latter requires one to resort to so-called 'stochastic' methods [39,40].

Modern ODE solvers can deal with essentially any system, even when its 'local' kinetics are on very different timescales (so-called 'stiff' systems), and many have been devised by and for biologists, thus making them particularly easy to use. A particular trend is towards making models that are interoperable between laboratories, and the website of the Systems Biology Markup Language <http://www.sbml.org> [41,42] lists many, including *Gepasi* [38,43,44].

Figure 2 shows various views of the systems biology agenda. Figure 2A stresses the importance of inductive methods of hypothesis *generation*; these have unaccountably had far less emphasis than they should have done because of the traditional obsession in twentieth century biology with hypothesis *testing* [16]. However, the *search* for good hypotheses can be seen as a heuristic search over a huge landscape of 'possible' hypotheses, of the form familiar in heuristic and combinatorial optimization problems [45–47], and the choice of where to look next – this is the 'principled' part – is known as 'active learning' [48–54]. It can be and has been automated in areas such as functional genomics [55,56], in clinical [57,58] and analytical chemistry [59], and in the coherent control of chemical reactions [60]. Principled hypothesis generation is clearly at least as important as hypothesis testing, and appropriate experimental designs, such as those used in active learning (and these go far beyond those usually described in textbooks of experimental design [61–65]), ensure that the search for good candidate data is not an aimless fishing expedition but one which is likely to find novel answers in unexpected places (e.g. [15,16,66–69]).

Figure 2B sets down the overall strategy, usually known as a 'bottom up' strategy, that we consider to be appropriate for most systems biology problems of interest to readers of the FEBS Journal. As whole-genome models of metabolism have become available (e.g. [70–72]), it has become evident that one can learn much merely from the structure plus constraints of a qualitative but stoichiometric model of the network (e.g. [14,73–80]). This leads one to stress the importance of first getting the structural model (the fundamental building blocks that determine and constrain

the 'language' of cells). From the qualitative model, we then require suitable equations that can represent the quantitative nature of the interactions set down in the structural model. Such equations are preferably mechanistic, as is common in molecular enzymology [81–84], but may also be empirical if they serve to fit the data over a suitably wide range [33,34,85]. After this, one must parametrize the kinetic data, as the parametrized equations (recast into the form of coupled ordinary differential equation) can then be used directly in forward models (e.g. [38,44]). Figure 2C, D and E highlight the basic and iterative relations between computational models and reality on one hand and between changes in the model that are invoked and its subsequent dynamic behaviour, leading to an understanding of how events at one level (e.g. the enzymatic) can be used to gain an understanding of events at a higher level (e.g. physiology or whole-cell metabolism). As mentioned above, the goal of systems biology in integrating these different levels of organization thus shares many similarities with those of metabolic control analysis and biochemical systems theory.

A particular issue with systems biology, which is why we stress the need to measure parameters, is that it is the parameters that control the variables and not the other way round, while omics measurements usually determine only the variables (e.g. in metabolism/metabolomics the metabolic fluxes and concentrations). Going from the variables to the parameters involves solving an inverse or 'system identification' problem [86], and this is typically very hard [87–91] as these problems are often heavily underdetermined (many parameter combinations can give the same variables), even if the structural model is correct.

Metabolomics and metabolomics technology development

As enshrined in the formalism of Metabolic Control Analysis (MCA) [11,26,28–32], it has been known for over 30 years that small changes in the activities of individual enzymes lead only to small changes in metabolic fluxes but can lead to large changes in concentrations. These facts are causally related, expected and mathematically proven. Metabolomics, being downstream of transcriptomics and proteomics, thus represents a more suitable level of biological organization for analysis [92] since metabolites are both more tractable in number and are amplified relative to changes in the transcriptome, proteome or gross phenotype [93]. Although we must in due time seek to integrate all the omes, metabolomics is thus the strategy of choice for

the purposes of functional genomics, biomarker development and systems biology (e.g. [94–104]).

If we consider metabolic systems, most analysts take discrete samples and provide what we have referred to as 'metabolic snapshots' [26]. Typical model microbes such as baker's yeast [70] contain upwards of 1000 known metabolites, and most of these have a relative molecular mass of less than 1000 [27]. Indeed, metabolomics is usually considered to mean 'small molecule metabolomics', even if cell wall polymers and the like are necessarily produced by metabolism.

The actual number of measurable metabolites in a given biological system is unknown, but numbers such as 10–13 000 have already been observed in mouse urine [105], albeit that some or many are of gut microbial origin [101]. Most of these have yet to be identified chemically.

The history of biomedicine as perceived via the awards of the Nobel Committee indicates the importance to our understanding of the subject of both small molecules (examples: ascorbic acid, coenzyme A, penicillin, streptomycin, cAMP, prostaglandins, dopamine, NO) and novel analytical methods (examples: paper chromatography, X-ray crystallography, the sequencing of proteins and of nucleic acids, radioimmunoassay, PCR, soft ionization MS, biological NMR). An important area of metabolomics thus consists of maximizing the number of metabolites that may be measured reliably [106–109], as a prelude to exploiting such data via a chemometric and computational pipeline [27,107,110]. As above, it transpires that optimizing scientific instrumentation is a combinatorial problem that scales exponentially with the number of experimental parameters. Thus, if there are 14 adjustable settings on an electrospray mass spectrometer, each of which can take 10 values, the number of combinations to be tested via exhaustive search is 10^{14} [111]. Since the lifetime of the Universe is about 10^{17} s [112], it is obvious that trying all of these ('exhaustive search') is impossible. So-called heuristic methods [113–117] are thus designed to find good but not provably optimal solutions, and methods [111,118] based on evolutionary algorithms [119] have proved successful. However, they are still slow because the run times are inconvenient and there is a human being in the loop, and the number of experiments that can be evaluated is correspondingly small.

As indicated above, active learning methods are attractive, and, in a manner related to the computationally driven supervised [120] and inductive [16] discovery of new biological knowledge [121], we have contributed to the Robot Scientist project [55]. This was concerned with automating principled hypothesis

generation in the area of experimental design for functional genomics. In this arrangement, one seeks to optimize the order in which one does a series of experiments, given that the number of possible experiments n can be done serially in $n!$ (n factorial) possible orders. For $n = 15$, $n! \approx 1.3 \cdot 10^{12}$. In the Robot Scientist paper [55] a computational system was used: (a) to hold background knowledge about a biological domain (amino acid biosynthesis, modelled as a logical graph); (b) to use that knowledge to design the ‘best’ (most discriminatory) experiment in order to find the biochemical location in that graph of a specific genetic lesion; (c) to perform that experiment using microbial growth tests, and to analyse the results; and (d) on the basis of these to design, perform and evaluate the next experiment, the whole continuing in an iterative manner (i.e. in a closed loop, without human intervention) until only one ‘possible’ hypothesis remains.

We have now combined these ideas to use heuristic search methods in an automated closed loop (the ‘Robot Chromatographer’) to maximize simultaneously the number of peaks observed while also minimizing the run time [59], and in addition maximizing a metric based on the signal : noise ratio. Depending on the sample (serum [107] or yeast supernatant [122–124]), this has more than trebled the number of metabolite peaks that we can reliably observe using GC TOF MS [59] (Fig. 3), thereby allowing us to discover important new biomarkers for metabolic and other diseases

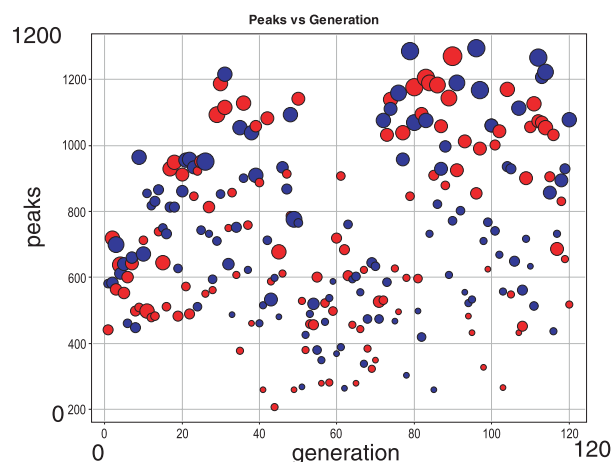


Fig. 3. Closed loop evolution of improved peak number in GC-MS experiments. Run time is encoded in the size of the symbols. It may be observed in the figure that this PESA-II algorithm [348] serially explores areas of space that can improve both the number of peaks and the run time. The size of the search space exceeded 200 000 000. Each generation contains two experiments, encoded via the two colours. Data are from the experiments described in [59].

including pre-eclampsia [125], peaks that were not observed in the original, previously optimized run conditions. The new technology thus led directly to the discovery of new biology, as in previous work in metabolomics (e.g. [67,68]). Sometimes it is a lack of unexpected differences that is the result of interest [126]. An especially useful strategy in microbiology is to study the exometabolome or ‘metabolic footprint’ [122–124,127] of metabolites excreted by cells, as this gives important clues as to their intracellular metabolism but is much easier to measure. Current work is concentrating on the optimization of 2D GC technology (GC×GC-TOF) [128–130] and ultra-performance liquid chromatography [105,124,131,132].

Creating and analysing systems biology models: network motifs, sensitivity analysis, functional linkage and signal processing

As postgenomic, high-throughput methods develop, it is increasingly commonplace to have access to large datasets of variables (‘omics data’) against which to test a mathematical model of the system that might generate such data. In these cases, the model will usually be an ODE model, and finding a good model is a system identification problem [44,86].

Much less frequently [133], the kinetic and binding constants are available, and a reliable ‘forward’ model can be generated directly. One such case [134] is the NF- κ B signalling pathway [135–138]. NF- κ B is a nuclear transcription factor that is normally held inactive in the cytoplasm by being bound to one or more isoforms of an inhibitor (I κ B). When I κ B is phosphorylated by a kinase (IKK) it is degraded and free NF- κ B can translocate to the nucleus, where it induces the expression of genes (including those such as I κ B that are involved in its own dynamics). The NF- κ B system is considered to be ‘involved’ in both cell proliferation and in apoptosis, as well as diseases such as arthritis, although how a cell ‘chooses’ which of these orthogonal processes will happen simply from the changes in the concentration of NF κ B in a particular location or compartment is neither known nor obvious. (In a sense this is the same problem as that of ‘commitment’ in developmental biology generally.) Earlier experimental measurements showed oscillations in nuclear NF- κ B in single cells, though these were damped when assessed as an ensemble since individual cells were necessarily out of phase ([139], and see also [140] for a different example and [141,142] for a similar philosophy underpinning the use of single-cell measurements in flow cytometry). More recently, with improved constructs and detector technology, the oscillations could

clearly be measured accurately in individual cells alone [19]. This ability to effect accurate measurements in individual cells is absolutely crucial for the analysis of nonlinear dynamic systems.

Based on the model of Hoffmann and colleagues [134] (see also [143,144]), and using *Gepasi* [43,44] we have modelled the ‘downstream’ parts of this pathway (there are 64 reactions and 23 variables), successfully reproducing the main features of the oscillations

observed experimentally in single cells (Fig. 4A and B) and performed sensitivity analysis on the model [18]. The model itself is/will soon be available via the ‘triple-J’ website <http://ijj.biochem.sun.ac.za/>. Sensitivity analysis is a generalized form of MCA [30] that is arguably the starting point for the analysis of any model [36], and that is useful in many other domains (e.g. [145]). This sensitivity analysis showed that only about eight of the 64 reactions exerted any serious

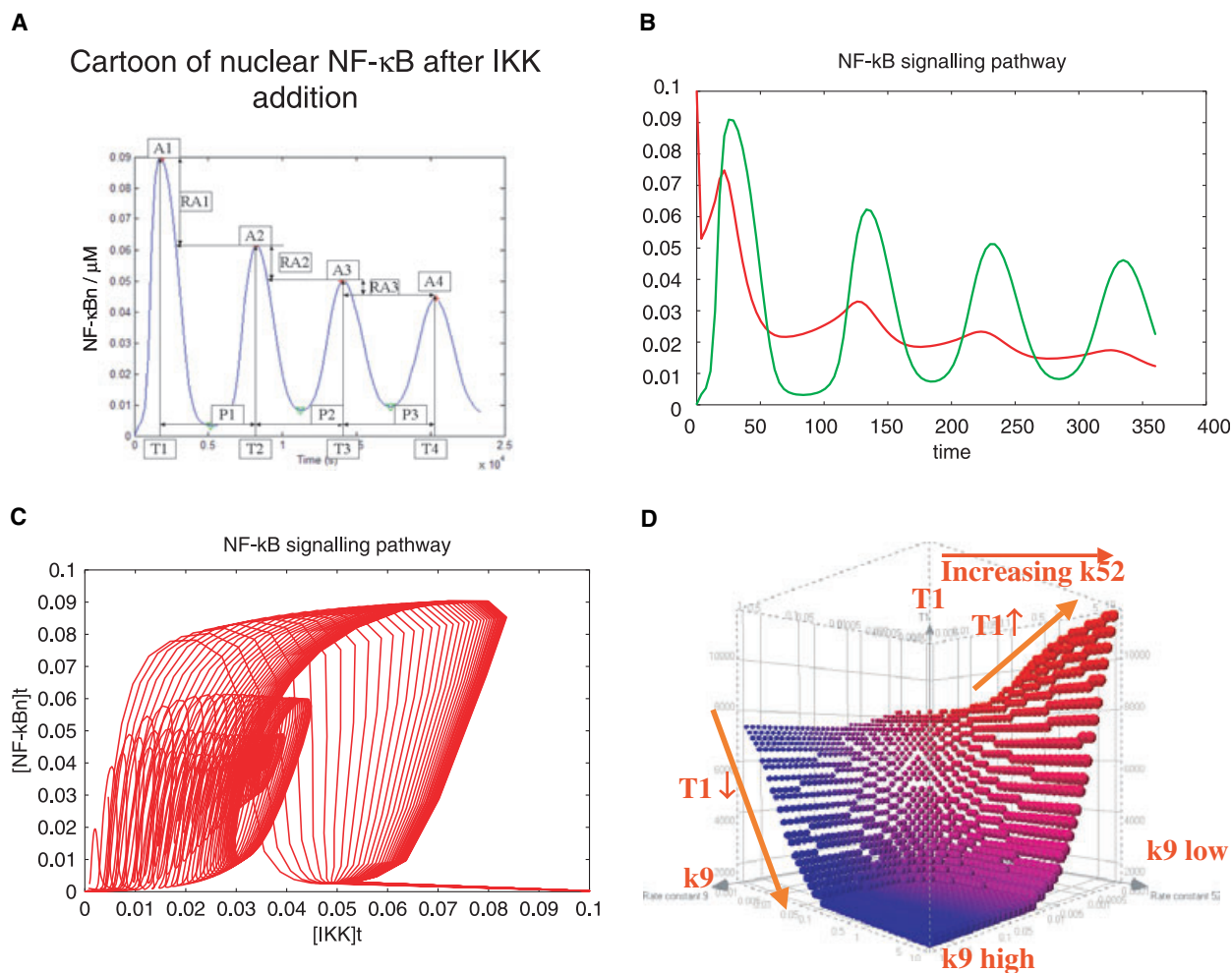


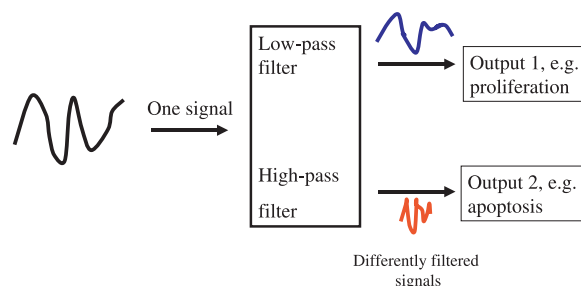
Fig. 4. (A) A cartoon illustrating the characterization of oscillations in the nuclear NF- κ B concentrations, in terms of features such as amplitude (A1, etc.), time (T1, etc.), Period (P1, etc.) and relative amplitude (RA1, etc.). (B) Time series output of a model [18,19] of the NF- κ B pathway showing oscillations in the concentration of NF- κ B in the nucleus (green) and of IKK (red). The model is pre-equilibrated then ‘started’ by adding IKK at 0.1 μ M. As with many such systems, the mechanism underpinning the oscillations is a coupled transcription-translation system with delays. (C) Effect on IKK and of nuclear NF- κ B of varying one rate constant (for reaction 28 in [18]) by two orders of magnitude either side of its basal value. Trajectories start from the right and follow fairly similar pathways for the first oscillation but then diverge considerably. (D) Synergistic effects of individual rate constants in the model [20]. The colour from red to blue shows increasing rate constant 9, while increasing symbol size reflects the increase in rate constant 52. For some values of the rate constants k_9 and k_{52} there is no influence of either on the time to the first oscillation (T1). However, when k_9 is low increasing k_{52} increases T1 while when k_9 is high the same increase in k_{52} decreases T1. Thus the effect of inhibiting a particular step can have qualitatively (directionally) different effects depending on the value of another step. This makes designing safe drugs aimed at targets in such pathways without understanding the system fully a challenging activity. This type of systemic nonlinearity can also account for the unexpected synergism often observed when different metabolic steps or drug targets are affected together, both in theory [349–352] and in practice [294,353,354].

control over the timings and amplitudes of the oscillations in the nuclear NF- κ B concentration [18], that the nonlinearity of the model implied: (a) both a differential control of the frequency and amplitude [18,19] of the first and subsequent oscillations; (b) that interactions between different elements of the model were synergistic [20] (Fig. 4C); and (c) most importantly that it was not so much the concentration of nuclear NF- κ B but its dynamics that were responsible for controlling downstream activities [19]. This leads to a profound emphasis on the role of ‘network motifs’ [21,146,147] as ‘downstream’ signal processing elements that can discriminate the dynamical properties of inputs that otherwise use the same components. Biological signalling is then best seen or understood as signal processing, a major field (mainly developed in areas such as data communications, image processing [148] and so on), in which we recognize that the structure, dynamics and performance of the receiver entirely determine which properties of the upstream signal are actually transduced into downstream (and here biological—see also [149]) events. The crucial point is that in the signal processing world these signals are separated and discriminated by their dynamical, time- and frequency-dependent properties. Normally we model enzyme kinetics on the basis of the effects of a static concentration of substrate or effector [81–84]. Thus, the irreversible Michaelis–Menten reaction ($v = \frac{V_{\max}S}{S+K_m}$) includes only the ‘instantaneous’ concentration but not the dynamics of S. However, if detectors have frequency-sensitive properties, this allows one in principle to solve the ‘crosstalk problem’ (how do cells distinguish identical changes in the ‘static’ NF- κ B concentration that might lead either to apoptosis or to proliferation, when these are in fact entirely orthogonal processes?). Although other factors can always contribute usefully (e.g. spatial segregation in micro-compartments or ‘channelling’ [150–153], and/or further transcription factors that act as a logical AND, OR or NOT [154]), encoding effective signals in the frequency domain allow one to separate signals independently of their amplitudes (i.e. concentrations) while still using the same components.

In the most simplistic way, one could imagine a structure (Fig. 5A) in which there was an input signal that could be filtered via a low-pass or high-pass filter before being passed downstream—a low-frequency signal would ‘go one way’ (i.e. be detected by only one ‘detector’ structure) and a high-frequency signal the other way. In this manner the same components can change their concentrations such that they may be at the same instantaneous levels while nevertheless having entirely different outcomes, solely because of the signal

A

The same signal can lead to two different outputs depending on the filtering/detector



B

Multiple LP filters in series act as a delay line

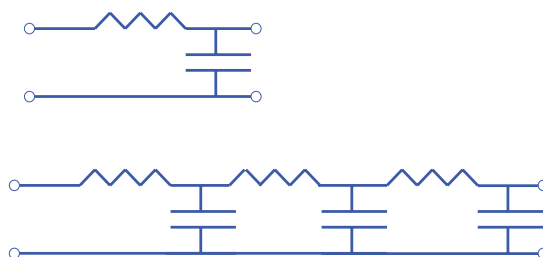


Fig. 5. The importance of signal dynamics and of downstream signal processing in affecting biological responses. (A) A simple system illustrating how two different frequency-selective filters can transduce *different features of the identical signal* into two different downstream signals and hence two different biological responses or events. Such downstream responses might be processes as different as apoptosis and cell proliferation. (B) Simple resistor-capacitor (RC) electrical filters (above) can act as a delay line when they are concatenated in series (below), and every biological reaction can act as an RC element, and this may account in part for the use of such serial devices in biology.

processing, frequency response characteristics of the detectors. Of course the real system and its signal-processing elements will be much more complex than this. We note that there is also precedent for the nonlinear and frequency-selective (bandpass) responses of individual multistate enzymes to exciting alternating electrical fields [155–159].

While the recognition that electrical circuit (signal processing) elements and biological networks are fundamentally similar representations is not especially new [22,47,146,160–167], Alon [21,147,168,169], Arkin [146], Tyson [22] and Sauro and colleagues [167], among others [170] have made these ideas particularly explicit. Any element (Fig. 5B) in a metabolic or signal transduction pathway acts as a resistor–capacitor

element [160] (as indeed do any ‘relaxing’ elements responding to an input, such as an alternating electrical signal [171]). A series of them acts as a delay line (Fig. 5B [17] and see [172] or any other textbook of electrical filters, and in a biological context [173]). This ability to act as a delay element provides another possible ‘reason’, besides signal amplification, for the serial arrangements of kinases and kinase kinases (etc.) in signalling cascades, since amplification alone could (have evolved to) be effected simply by increasing the rate constants of a single kinase. Similarly, a suitably configured (‘coherent’) feedforward network serves to provide resistance to temporally small input perturbations (noise—or at least an amount of fluctuating/diffusing nutrient not worth chasing) whilst transducing longer-lasting ones of the same amplitude into output (biological effects) [174,175]. Other network structures – which like all such network structures effectively act as ‘computational’ or ‘signal processing’ elements – can exhibit robustness of their output(s) to sometimes extreme variations in parameters [22,165,176–187]. Indeed, the evolution of *robustness* is probably an inevitable consequence of the evolution of life in an environment that changes far more rapidly than does the genotype [179].

Thus the recognition that we need to concentrate more on the dynamics of signalling pathways rather than instantaneous concentrations of their components, means that we need to sample very frequently – preferably effectively in real time – and using single cell measurements to avoid oscillations and other more complex and functionally important dynamics being hidden via the combination of signals from individual, out-of-phase cells. It also means that assays for signalling activity, for instance in drug development, should not focus just on the signalling molecules themselves but on the structures that the cell uses to detect them.

A forward look

By concentrating on a restricted subset of issues within the confines of a single lecture, many topics had to be treated only superficially or implicitly, and it is appropriate to set down in slightly more detail some of the directions in which I think progress is required, important or likely.

Data standards and integration

The first is the need to integrate SBML (and other [188]) biochemical models and model representations into postgenomic databases with schemas such as those

for genomics (e.g. GIMS [189]), transcriptomics (e.g. MAGE-ML [190]), protein interactions [191], proteomics (e.g. PEDRo [192] and PSI [193,194]) and metabolomics (e.g. ArMet [195] and SMRS [196]). Progress is being made (e.g. [197]), but significant problems remain before the considerable benefits [198] of extensible markup languages can be fully realized [199], and before well-structured ontologies (<http://suo.ieee.org/>) become the norm [200].

In a related manner, there are many things one might wish to do with an SBML or other biochemical model, including creating it, storing it, editing it, comparing it with other stored models, finding it again in a principled way, visualizing it, sharing it, running it, analysing the results of the run, comparing them with experimental data, finding models that can create a given set of data, and so on. No individual piece of software allows one to do all of these things well or even at all (for a starting point see <http://dbk.ch.umist.ac.uk/sysbio.htm#links>). However, plan A (start from scratch and write the software that one wished existed) would require an enormous and coherent effort involving many person-years. Consequently we are attracted by plan B. This is to create a software environment in which individual software elements appear to – and indeed do – work together transparently [201], such that ‘only’ the software ‘glue’ needs to be written, somewhat in the spirit of the Systems Biology Workbench [202] or of software Application Programming Interfaces more generally. Distributed environments using systems such as Taverna [203] or others [204–206] to enact the necessary bioinformatic workflows may well provide the best way forward, and since the difficulties of interoperability seem in fact to be much more about data structures (syntax) than about their meaning (semantics) [207], this task may turn out to be considerably easier than might have been anticipated.

Synthetic biology

Another emerging and important area is becoming known as ‘synthetic biology’ [208–213] (a portal for this can be found at <http://www.syntheticbiology.org/>). Although this has a variety of subthreads [213], an ‘engineering’-based motivation [214–216] is the one which I regard as paramount. Here one seeks, somewhat in the manner of the ‘network motifs’ mentioned above, to develop principled strategies for determining the kind of networks and computational structures in biology that can effect specific metabolic or signal processing acts or behaviours, and to combine them effectively. Ultimately, as a refined and improved

strategy for metabolic engineering [30,78,217–223] one may hope that this will give sufficient understanding to allow one to design these and more complex bioprocesses (and the organisms that perform them). Similar comments apply to the *de novo* design, synthesis and engineering of proteins [224–234] (where there is already progress with building blocks or elements such as foldamers [235–238]), initially as a complement to effective but more empirical strategies based on the directed evolution and selection of both proteins (e.g. [239–252]), and nucleic acid aptamers (e.g. [253–274]).

Chemical genetics and chemical genomics

The modulation by small molecules of biological activities has proven to be of immense value historically in the dissection of biological pathways (e.g. in oxidative phosphorylation [275,276]). Chemical genetics or chemical genomics (e.g. [277–292]) 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 systematize the knowledge and from which quantitative structure–activity relationships may be discerned [293]). This chemical manipulation is considered to be more discriminating than strategies based on knocking out genes or gene products using the methods of molecular biology since they can be selective towards individual activities that may be among several catalysed by specific gene products. Also, chemical genetics can be used to study multiple effects when the small molecules are added both singly and in combination [294], and such studies – involving only the addition of small molecules – can be performed with far more facility than those requiring complex and serial molecular biological manipulations. As with ‘biological’ genetics, it is usual to discriminate ‘forward’ and ‘reverse’ chemical genetics. In ‘forward’ chemical genetics, the logic goes: screen a library → find cellular or physiological activity → discover molecular target [295], this being somewhat akin to the ‘traditional’ (pregenomic) drug discovery process in the pharmaceutical industry. In ‘reverse’ chemical genetics we start with a purified target, then with the chemical library look for binding activity and then test *in vivo* to see the physiological effects, much as is done (with decreasing success) in the more recent approaches preferred by Pharma. While these strategies should best be seen as iterative (Fig. 6), we would have some preference for the ‘forward’ chemical genetic approach as the hypothesis-generating arm.

Forward and reverse chemical genetics

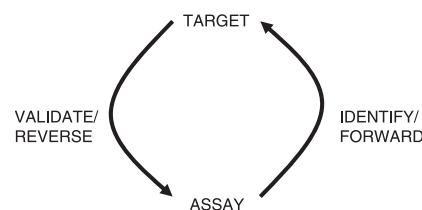


Fig. 6. Chemical genomics as an iterative process in which molecules are screened for effects and their targets identified, thereby allowing the development of mechanistic links between individual targets and (patho-)physiological processes.

Text mining

With the scientific literature expanding by several thousand papers per week, it is obvious that no individual can read them, and there is in addition a large historical database of facts that could be useful to systems biology. Text mining is an emerging field concerned with the process of discovering and extracting knowledge from unstructured textual data, contrasting it with data mining (e.g. [296,297]), which discovers knowledge from structured data. Text mining comprises three major activities: information retrieval, to gather relevant texts; information extraction, to identify and extract a range of specific types of information from texts of interest; and data mining, to find associations among the pieces of information extracted from many different texts [298]. As phrased therein ‘...hypothesis generation relies on background knowledge, and is crucial in scientific discovery’, the pioneering work by Swanson on hypothesis generation [299] is mainly credited with sparking interest in text mining techniques in biology. Text mining aids in the construction of hypotheses from associations derived from vast amounts of text that are then subjected to experimental validation by experts.

Some portals are at <http://www.ccs.neu.edu/home/futrelle/bionlp/> and <http://www.cs.technion.ac.il/~gabr/resources/resources.html>, and a national (UK) centre devoted to the subject is described at <http://www.nactem.ac.uk>. Although these are early days (e.g. [300–308]), we may one day dream of a system that will read the literature for us and produce and parameterize (with linkages, equations and parameters like rate constants) candidate models of chosen parts of biological systems.

Single cell and single molecule biology

Given the heterogeneity of almost all biological systems, and thus for reasons given above the importance

of single cell studies, it is evident that we need to develop improved methods for measuring omics in individual cells, preferably noninvasively and *in vivo*. Buoyed by experience with the fluorescent proteins [309], and indeed with the more recent antibody-based proteomics [310] (<http://www.proteinatlas.org/>), it is evident that optical methods are among the most promising here, with detectors for specific metabolites [311] and transcripts (<http://www.nanostring.com/>) (see also [312]) that can be used in individual cells coming forward as part of the development of Bionanotechnology [313].

What is true about the heterogeneity of single cells [141,142] is also true for that of single molecules [314,315], and many assays capable of detecting the presence or behaviour of single molecules are coming forward. Thus, high-throughput screening for ligand binding [316,317] and nucleic acid sequences [318–320] are now being performed using assays based on miniaturization and single-molecule measurements, bringing the \$1000 human genome well within sight (although amplification techniques can of course also be used to advantage in nucleic acid sequencing [321,322]).

The Manchester Interdisciplinary Biocentre (MIB)

Many of the kinds of problems described above, and certainly the solutions being developed to attack them, require the input of ideas and techniques, and scientific cultures, from the physical sciences, engineering, mathematics and computer science. One solution, that we are adopting in the Manchester Interdisciplinary Biocentre (MIB: <http://www.mib.ac.uk/>, Fig. 7) and the Manchester Centre for Integrative Systems Biology (MCISB: <http://www.mcisb.org/>), is to colocate individuals with the necessary combinations of skills. Within MCISB we are seeking to develop the suite of techniques for the largely ‘bottom up’ systems biology strategies set down in Fig. 2B.

Emergence and a true systems biology

The grand problem of biology, as well as the ‘inverse problem’ (Fig. 2D) of determining parametric causes from measured effects (variables), to which it is related, is understanding at a lower level the time-dependent [323,324] changes of state that are commonly described at a higher level of organization, an issue often referred to using terms such as ‘self-organization’ [325], ‘emergence’ [326–328], networks [329,330] and complexity [161,165,331–333]. Modelling and sensitivity analysis (see above) can begin to decon-



Fig. 7. The Manchester Interdisciplinary Biocentre, a physical building and intellectual environment that brings together workers from a variety of Schools at the University of Manchester focussing on Engineering and Physical Sciences, including mathematics and computing ($\approx 60\%$), with those from biology and medicine (40%).

struct such relations, but it is in areas such as ‘causal inference’ [334–337] that we shall probably see the most focussed development of principled explanations of such causal linkages.

Coda

Having begun with a couple of quotations, and having stressed the role of technology development in science in general and in systems biology in particular, I shall end with another quotation, from the Nobelist Robert Laughlin [338]:

In physics, correct perceptions differ from mistaken ones in that they get clearer when the experimental accuracy is improved. This simple idea captures the essence of the physicist’s mind and explains why they are always so obsessed with mathematics and numbers: through precision one exposes falsehood...A subtle but inevitable consequence of this attitude is that truth and measurement technology are inextricably linked.

Acknowledgements

In addition to the huge contributions of the past and present members of my research group I have enjoyed many friendships and scientific collaborations with numerous colleagues, who are listed as coauthors in the references, but I would especially like to mention Steve Oliver, Hans Westerhoff and Mike White. I also thank the BBSRC, BHF, EPSRC, MRC, NERC and the RSC for financial support.

References

- 1 von Bertalanffy L (1969) *General System Theory*. George Braziller, New York.
- 2 Iberall AS (1972) *Toward a General Science of Viable Systems*. McGraw-Hill, New York.
- 3 Kell DB (1979) On the functional proton current pathway of electron transport phosphorylation: an electrodic view. *Biochim Biophys Acta* **549**, 55–99.
- 4 Brenner S (1997) *Loose Ends*. Current Biology, London.
- 5 Hood L (2003) Systems biology: integrating technology, biology, and computation. *Mech Ageing Dev* **124**, 9–16.
- 6 Ideker T, Galitski T & Hood L (2001) A new approach to decoding life: systems biology. *Annu Rev Genomics Hum Genet* **2**, 343–372.
- 7 Kitano H (2002) Systems biology: a brief overview. *Science* **295**, 1662–1664.
- 8 Kitano H (2002) Computational systems biology. *Nature* **420**, 206–210.
- 9 Davidov E, Holland J, Marple E & Naylor S (2003) Advancing drug discovery through systems biology. *Drug Discov Today* **8**, 175–183.
- 10 Henry CM (2003) Systems biology. *Chem Eng News* **81**, 45–55.
- 11 Westerhoff HV & Palsson BO (2004) The evolution of molecular biology into systems biology. *Nat Biotechnol* **22**, 1249–1252.
- 12 Klipp E, Herwig R, Kowald A, Wierling C & Lehrach H (2005) *Systems Biology in Practice: Concepts, Implementation and Clinical Application*. Wiley/VCH, Berlin.
- 13 Kriete A & Eils R (2005) *Computational Systems Biology*. Academic Press, New York.
- 14 Palsson BØ (2006) *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press, Cambridge.
- 15 Kell DB (2002) Genotype: phenotype mapping: genes as computer programs. *Trends Genet* **18**, 555–559.
- 16 Kell DB & Oliver SG (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.
- 17 Kell DB (2005) Metabolomics, machine learning and modelling: towards an understanding of the language of cells. *Biochem Soc Trans* **33**, 520–524.
- 18 Ihekwa AEC, Broomhead DS, Grimley R, Benson N & Kell DB (2004) Sensitivity analysis of parameters controlling oscillatory signalling in the NF- κ B pathway: the roles of IKK and I κ B α . *Systems Biol* **1**, 93–103.
- 19 Nelson DE, Ihekwa AEC, Elliott M, Gibney CA, Foreman BE, Nelson G, See V, Horton CA, Spiller DG, Edwards SW, McDowell HP, Unitt JF, Sullivan E, Grimley R, Benson N, Broomhead DS, Kell DB & White MRH (2004) Oscillations in NF- κ B signalling control the dynamics of target gene expression. *Science* **306**, 704–708.
- 20 Ihekwa AEC, Broomhead DS, Grimley R, Benson N, White MRH & Kell DB (2005) Synergistic control of oscillations in the NF- κ B signalling pathway. *IEE Systems Biol* **152**, 153–160.
- 21 Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D & Alon U (2002) Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827.
- 22 Tyson JJ, Chen KC & Novak B (2003) Sniffers, buzzers, toggles and blinkers. dynamics of regulatory and signaling pathways in the cell. *Curr Opin Cell Biol* **15**, 221–231.
- 23 Bhalla U.S. (2003) Understanding complex signaling networks through models and metaphors. *Prog Biophys Mol Biol* **81**, 45–65.
- 24 Wall ME, Hlavacek WS & Savageau MA (2004) Design of gene circuits: Lessons from bacteria. *Nat Rev Genet* **5**, 34–42.
- 25 Kacser H (1986) On parts and wholes in metabolism. *The Organization of Cell Metabolism* (Welch, G R & Clegg, J S, eds), pp. 327–337. Plenum Press, New York.
- 26 Kell DB & Mendes P (2000) Snapshots of systems. metabolic control analysis and biotechnology in the post-genomic era. *Technological and Medical Implications of Metabolic Control Analysis* (Cornish-Bowden, A & Cárdenas, M L, eds), pp. 3–25 (and see <http://dbk.ch.umist.ac.uk/WhitePapers/mcabio.htm>). Kluwer Academic Publishers, Dordrecht.
- 27 Kell DB (2004) Metabolomics and systems biology: making sense of the soup. *Curr Op Microbiol* **7**, 296–307.
- 28 Kacser H & Burns JA (1973) The control of flux. *Rate Control of Biological Processes. Symposium of the Society for Experimental Biology*, Vol. 27 (Davies, D D, ed.), pp. 65–104. Cambridge University Press, Cambridge.
- 29 Heinrich R & Rapoport TA (1974) A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. *Eur J Biochem* **42**, 89–95.
- 30 Kell DB & Westerhoff HV (1986) Metabolic control theory: its role in microbiology and biotechnology. *FEMS Microbiol Rev* **39**, 305–320.
- 31 Heinrich R & Schuster S (1996) *The Regulation of Cellular Systems*. Chapman & Hall, New York.
- 32 Fell DA (1996) *Understanding the Control of Metabolism*. Portland Press, London.
- 33 Savageau M (1976) *Biochemical Systems Analysis: a Study of Function and Design in Molecular Biology*. Addison-Wesley, Reading, MA.
- 34 Voit EO (2000) *Computational Analysis of Biochemical Systems*. Cambridge University Press, Cambridge.

- 35 Lazebnik Y (2002) Can a biologist fix a radio? – or, what I learned while studying apoptosis. *Cancer Cell* **2**, 179–182.
- 36 Kell DB & Knowles JD (2005) The role of modeling in systems biology. *System Modeling in Cellular Biology: from Concepts to Nuts and Bolts* (Szallasi, Z Periwal, V & Stelling, J, eds), pp. 3–18. MIT Press, Cambridge.
- 37 Bower JM & Bolouri H (2004) *Computational Modeling of Genetic and Biochemical Networks*. Bradford Books, New York.
- 38 Mendes P & Kell DB (2001) MEG (Model Extender for Gepasi): a program for the modelling of complex, heterogeneous cellular systems. *Bioinformatics* **17**, 288–289.
- 39 Andrews SS & Bray D (2004) Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Phys Biol* **1**, 137–151.
- 40 Salis H & Kaznessis Y (2005) Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. *J Chem Phys* **122**, 54103.
- 41 Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, *et al.* (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531.
- 42 Finney A & Hucka M (2003) Systems biology markup language: Level 2 and beyond. *Biochem Soc Trans* **31**, 1472–1473.
- 43 Mendes P (1997) Biochemistry by numbers: simulation of biochemical pathways with Gepasi 3. *Trends Biochem Sci* **22**, 361–363.
- 44 Mendes P & Kell DB (1998) Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics* **14**, 869–883.
- 45 Kauffman S, Lobo J & Macready WG (2000) Optimal search on a technology landscape. *J Econ Behav Organ* **43**, 141–166.
- 46 Goldberg DE (2002) *The Design of Innovation: Lessons from and for Competent Genetic Algorithms*. Kluwer, Boston.
- 47 Koza JR, Keane MA, Streeter MJ & Mydlowec W, Yu, J & Lanza G (2003) *Genetic Programming: Routine Human-Competitive Machine Intelligence*. Kluwer, New York.
- 48 Raju GK & Cooney CL (1998) Active learning from process data. *AIChE J* **44**, 2199–2211.
- 49 Bryant CH, Muggleton SH, Oliver SG, Kell DB, Reiser P & King RD (2001) Combining inductive logic programming, active learning and robotics to discover the function of genes. *Electronic Transactions on Artificial Intelligence* **5**, 1–36 (<http://www.ep.liu.se/ej/etai/2001/001/>).
- 50 Cohn DA, Ghabhrmani Z & Jordan MI (1996) Active learning with statistical models. *J Artif Intell Res* **4**, 129–145.
- 51 Hasenjäger M & Ritter H (1998) Active learning with local models. *Neural Proc Lett* **7**, 110–117.
- 52 Cohn DA, Atlas L & Ladner R (1994) Improving generalisation with active learning. *Machine Learning* **15**, 201–221.
- 53 Mackay D (1992) Information-based objective functions for active data selection. *Neural Comput* **4**, 590–604.
- 54 Milano M, Schmidhuber J & Koumoutsakos P (2001) (2001) Active learning with adaptive grids. *Artificial Neural Networks-ICANN Proc* **2130**, 436–442.
- 55 King RD, Whelan KE, Jones FM, Reiser PGK, Bryant CH, Muggleton SH, Kell DB & Oliver SG (2004) Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature* **427**, 247–252.
- 56 Whelan KE & King RD (2004) Intelligent software for laboratory automation. *Trends Biotechnol* **22**, 440–445.
- 57 Olansky AS, Parker LR Jr, Morgan SL & Deming SN (1977) Automated development of analytical chemical methods: the determination of serum calcium by the cresolphthalein complexone method. *Anal Chim Acta* **95**, 107–133.
- 58 Olansky AS & Deming SN (1978) Automated development of a kinetic method for the continuous-flow determination of creatinine. *Clin Chem* **24**, 2115–2124.
- 59 O'Hagan S, Dunn WB, Brown M, Knowles JD & Kell DB (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.
- 60 Daniel C, Full J, Gonzalez L, Lupulescu C, Manz J, Merli A, Vajda S & Woste L (2003) Deciphering the reaction dynamics underlying optimal control laser fields. *Science* **299**, 536–539.
- 61 Schlesselman JJ (1982) *Case-Control Studies – Design, Conduct, Analysis*. Oxford University Press, Oxford.
- 62 Logothetis N & Wynn HP (1989) *Quality Through Design: Experimental Design, Off-Line Quality Control, and Taguchi's Contribution*. Clarendon Press, Oxford.
- 63 Hicks CR & Turner KV (1999) Jr. *Fundamental Concepts in the Design of Experiments*, 5th edn. Oxford University Press, Oxford.
- 64 Montgomery DC (2001) *Design and Analysis of Experiments*, 5th edn. Wiley, Chichester.
- 65 Myers RH & Montgomery DC (1995) *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. Wiley, New York.
- 66 Brent R (1999) Functional genomics: Learning to think about gene expression data. *Curr Biol* **9**, R338–R341.
- 67 Kell DB, Darby RM & Draper J (2001) Genomic computing: explanatory analysis of plant expression

- profiling data using machine learning. *Plant Physiol* **126**, 943–951.
- 68 Kell DB (2002) Metabolomics and machine learning: explanatory analysis of complex metabolome data using genetic programming to produce simple, robust rules. *Mol Biol Report* **29**, 237–241.
- 69 Brent R & Lok L (2005) A fishing buddy for hypothesis generators. *Science* **308**, 504–506.
- 70 Förster J, Famili I, Fu P, Palsson BØ & Nielsen J (2003) Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res* **13**, 244–253.
- 71 Reed JL & Palsson BØ (2003) Thirteen years of building constraint-based *in silico* models of *Escherichia coli*. *J Bacteriol* **185**, 2692–2699.
- 72 Borodina I, Krabben P & Nielsen J (2005) Genome-scale analysis of *Streptomyces coelicolor* A3 (2) metabolism. *Genome Res* **15**, 820–829.
- 73 Edwards JS, Ibarra RU & Palsson BØ (2001) *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat Biotechnol* **19**, 125–130.
- 74 Segrè D, Vitkup D & Church GM (2002) Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci USA* **99**, 15112–15117.
- 75 Segrè D, Zucker J, Katz J, Lin X, D'Haeseleer P, Rindone WP, Kharchenko P, Nguyen DH, Wright MA & Church GM (2003) From annotated genomes to metabolic flux models and kinetic parameter fitting. *Omic* **7**, 301–316.
- 76 Covert MW & Palsson BØ (2003) Constraints-based models: regulation of gene expression reduces the steady-state solution space. *J Theor Biol* **221**, 309–325.
- 77 Papin JA, Stelling J, Price ND, Klamt S, Schuster S & Palsson BO (2004) Comparison of network-based pathway analysis methods. *Trends Biotechnol* **22**, 400–405.
- 78 Patil KR, Akesson M & Nielsen J (2004) Use of genome-scale microbial models for metabolic engineering. *Curr Opin Biotechnol* **15**, 64–69.
- 79 Famili I, Mahadevan R & Palsson BO (2005) k-Cone analysis: determining all candidate values for kinetic parameters on a network scale. *Biophys J* **88**, 1616–1625.
- 80 Patil KR, Rocha I, Forster J & Nielsen J (2005) Evolutionary programming as a platform for *in silico* metabolic engineering. *BMC Bioinformatics* **6**, 308.
- 81 Fersht A (1977) *Enzyme Structure and Mechanism*, 2nd edn. W.H. Freeman, San Francisco.
- 82 Keleti T (1986) *Basic Enzyme Kinetics*, Akadémiai Kiadó, Budapest.
- 83 Segel IH (1993) *Enzyme Kinetics*. Wiley, New York.
- 84 Cornish-Bowden A (1995) *Fundamentals of Enzyme Kinetics*, 2nd edn. Portland Press, London.
- 85 Wu L, Wang W, van Winden WA, van Gulik WM & Heijnen JJ (2004) A new framework for the estimation of control parameters in metabolic pathways using lin-log kinetics. *Eur J Biochem* **271**, 3348–3359.
- 86 Ljung L (1987) *System Identification: Theory for the User*. Prentice Hall, Englewood Cliffs, NJ.
- 87 Mendes P & Kell DB (1996) On the analysis of the inverse problem of metabolic pathways using artificial neural networks. *Biosystems* **38**, 15–28.
- 88 Koza JR, Mydlowec W & Lanza G, Yu J & Keane MA (2001) Reverse engineering of metabolic pathways from observed data using genetic programming. *Pac Symp Biocomput* 434–445.
- 89 Moles CG, Mendes P & Banga JR (2003) Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res* **13**, 2467–2474.
- 90 Styczynski MP & Stephanopoulos G (2005) Overview of computational methods for the inference of gene regulatory networks. *Comput Chem Eng* **29**, 519–534.
- 91 Patil KR & Nielsen J (2005) Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci USA* **102**, 2685–2689.
- 92 Oliver SG, Winson MK, Kell DB & Baganz F (1998) Systematic functional analysis of the yeast genome. *Trends Biotechnol* **16**, 373–378.
- 93 Raamsdonk LM, Teusink B, Broadhurst D, Zhang N, Hayes A, Walsh M, Berden JA, Brindle KM, Kell DB, Rowland JJ, *et al.* (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat Biotechnol* **19**, 45–50.
- 94 Fiehn O (2002) Metabolomics: the link between genotypes and phenotypes. *Plant Mol Biol* **48**, 155–171.
- 95 Harrigan GG & Goodacre R (2003) *Metabolic Profiling: its Role in Biomarker Discovery and Gene Function Analysis*, Kluwer Academic Publishers, Boston.
- 96 Sumner LW, Mendes P & Dixon RA (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* **62**, 817–836.
- 97 Weckwerth W (2003) Metabolomics in systems biology. *Annu Rev Plant Biol* **54**, 669–689.
- 98 German JB, Roberts MA & Watkins SM (2003) Personal metabolomics as a next generation nutritional assessment. *J Nutr* **133**, 4260–4266.
- 99 Nicholson JK & Wilson ID (2003) Understanding 'global' systems biology: Metabonomics and the continuum of metabolism. *Nat Rev Drug Disc* **2**, 668–676.
- 100 Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, Draper J, Nikolau BJ, Mendes P, Roessner-Tunali U, Beale MH, *et al.* (2004) Potential of metabolomics as a functional genomics tool. *Trends Plant Sci* **9**, 418–425.
- 101 Nicholson JK, Holmes E, Lindon JC & Wilson ID (2004) The challenges of modeling mammalian biocomplexity. *Nat Biotechnol* **22**, 1268–1274.
- 102 Whitfield PD, German AJ & Noble PJ (2004) Metabolomics: an emerging post-genomic tool for nutrition. *Br J Nutr* **92**, 549–555.

- 103 Gibney MJ, Walsh M, Brennan L, Roche HM, German B & van Ommen B (2005) Metabolomics in human nutrition: opportunities and challenges. *Am J Clin Nutr* **82**, 497–503.
- 104 Vaidyanathan S, Harrigan GG & Goodacre R (2005) *Metabolome Analyses: Strategies for Systems Biology*. Springer, New York.
- 105 Wilson ID, Nicholson JK, Castro-Perez J, Granger JH, Johnson KA, Smith BW & Plumb RS (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.
- 106 Wilson ID & Brinkman UA (2003) Hyphenation and hypernation: the practice and prospects of multiple hyphenation. *J Chromatogr A* **1000**, 325–356.
- 107 Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG & Kell DB (2004) Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol* **22**, 245–252.
- 108 Dunn WB & Ellis DI (2005) Metabolomics: current analytical platforms and methodologies. *Trends Anal Chem* **24**, 285–294.
- 109 Dunn WB, Bailey NJC & Johnson HE (2005) Measuring the metabolome: current analytical technologies. *Analyst* **130**, 606–625.
- 110 Brown M, Dunn WB, Ellis DI, Goodacre R, Handl J, Knowles JD, O'Hagan S, Spasic I & Kell DB (2005) A metabolome pipeline: from concept to data to knowledge. *Metabolomics* **1**, 35–46.
- 111 Vaidyanathan S, Broadhurst DI, Kell DB & Goodacre R (2003) Explanatory optimisation of protein mass spectrometry via genetic search. *Anal Chem* **75**, 6679–6686.
- 112 Barrow JD & Silk J (1995) *The Left Hand of Creation: the Origin and Evolution of the Expanding Universe*. Penguin, London.
- 113 Reeves CR (1995) *Modern Heuristic Techniques for Combinatorial Problems*. McGraw-Hill, London.
- 114 Rayward-Smith VJ, Osman IH, Reeves CR & Smith GD (1996) *Modern Heuristic Search Methods*. Wiley, Chichester.
- 115 Corne D, Dorigo M & Glover F (1999) *New Ideas in Optimization*. McGraw-Hill, London.
- 116 Dasgupta P, Chakrabarti PP & DeSarkar SC (1999) *Multiobjective Heuristic Search*. Vieweg, Braunschweig.
- 117 Michalewicz Z & Fogel DB (2000) *How to Solve It: Modern Heuristics*. Springer-Verlag, Heidelberg.
- 118 Vaidyanathan S, Kell DB & Goodacre R (2004) Selective detection of proteins in mixtures using electrospray ionization mass spectrometry: influence of instrumental settings and implications for proteomics. *Anal Chem* **76**, 5024–5032.
- 119 Bäck T, Fogel DB & Michalewicz Z (1997) *Handbook of Evolutionary Computation*. IOP Publishing/Oxford University Press, Oxford.
- 120 Kell DB & King RD (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.
- 121 Langley P, Simon HA, Bradshaw GL & Zytkow JM (1987) *Scientific Discovery: Computational Exploration of the Creative Processes*. MIT Press, Cambridge, MA.
- 122 Allen JK, Davey HM, Broadhurst D, Heald JK, Rowland JJ, Oliver SG & Kell DB (2003) High-throughput characterisation of yeast mutants for functional genomics using metabolic footprinting. *Nat Biotechnol* **21**, 692–696.
- 123 Allen J, Davey HM, Broadhurst D, Rowland JJ, Oliver SG & Kell DB (2004) Discrimination of the modes of action of antifungal substances by use of metabolic footprinting. *Appl Env Micro* **70**, 6157–6165.
- 124 Kell DB, Brown M, Davey HM, Dunn WB, Spasic I & Oliver SG (2005) Metabolic footprinting and Systems Biology: the medium is the message. *Nat Rev Microbiol* **3**, 557–565.
- 125 Kenny LC, Dunn WB, Ellis DI, Myers J & Baker PN, The GOPEC Consortium & Kell DB (2005) Novel biomarkers for pre-eclampsia detected using metabolomics and machine learning. *Metabolomics* **1** (in press). Online 10.1007/s11306-005-0003-1.
- 126 Catchpole GS, Beckmann M, Enot DP, Mondhe M, Zywicki B, Taylor J, Hardy N, Smith A, King RD, Kell DB, Fiehn O & Draper J (2005) Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops. *Proc Natl Acad Sci USA* **102**, 14458–14462.
- 127 Kaderbhai NN, Broadhurst DI, Ellis DI, Goodacre R & Kell DB (2003) Functional genomics via metabolic footprinting: Monitoring metabolite secretion by *Escherichia coli* tryptophan metabolism mutants using FT-IR and direct injection electrospray mass spectrometry. *Comp Func Genomics* **4**, 376–391.
- 128 Marriott P & Shellie R (2002) Principles and applications of comprehensive two-dimensional gas chromatography. *Trends Anal Chem* **21**, 573–583.
- 129 Ong RC & Marriott PJ (2002) A review of basic concepts in comprehensive two-dimensional gas chromatography. *J Chromatogr Sci* **40**, 276–291.
- 130 Blumberg LM (2003) Comprehensive two-dimensional gas chromatography: metrics, potentials, limits. *J Chromatogr A* **985**, 29–38.
- 131 Plumb R, Castro-Perez J, Granger J, Beattie I, Joncour K & Wright A (2004) Ultra-performance liquid chromatography coupled to quadrupole-orthogonal time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* **18**, 2331–2337.
- 132 Wilson ID, Plumb R, Granger J, Major H, Williams R & Lenz EM (2005) HPLC-MS-based methods for the

- study of metabolomics. *J Chromatogr B Analyt Technol Biomed Life Sci* **817**, 67–76.
- 133 Sauro HM & Kholodenko BN (2004) Quantitative analysis of signaling networks. *Prog Biophys Mol Biol* **86**, 5–43.
- 134 Hoffmann A, Levchenko A, Scott ML & Baltimore D (2002) The I κ B-NF- κ B signaling module: temporal control and selective gene activation. *Science* **298**, 1241–1245.
- 135 Ghosh S & Karin M (2002) Missing pieces in the NF-kappaB puzzle. *Cell* **109** (Suppl.), S81–S96.
- 136 Richmond A (2002) NF- κ B, chemokine gene transcription and tumour growth. *Nat Rev Immunol* **2**, 664–674.
- 137 Tian B & Brasier AR (2003) Identification of a nuclear factor kappa B-dependent gene network. *Recent Prog Horm Res* **58**, 95–130.
- 138 Tian B, Nowak DE, Jamaluddin M, Wang S & Brasier AR (2005) Identification of direct genomic targets downstream of the nuclear factor-kappaB transcription factor mediating tumor necrosis factor signalling. *J Biol Chem* **280**, 17435–17448.
- 139 Nelson G, Paraoan L, Spiller DG, Wilde GJ, Browne MA, Djali PK, Unitt JF, Sullivan E, Floettmann E & White MR (2002) Multi-parameter analysis of the kinetics of NF- κ B signalling and transcription in single living cells. *J Cell Sci* **115**, 1137–1148.
- 140 Mantzaris NV (2005) Single-cell gene-switching networks and heterogeneous cell population phenotypes. *Comput Chem Eng* **29**, 631–643.
- 141 Kell DB, Ryder HM, Kaprelyants AS & Westerhoff HV (1991) Quantifying heterogeneity: Flow cytometry of bacterial cultures. *Antonie van Leeuwenhoek* **60**, 145–158.
- 142 Davey HM & Kell DB (1996) Flow cytometry and cell sorting of heterogeneous microbial populations: the importance of single-cell analysis. *Microbiol Rev* **60**, 641–696.
- 143 Werner SL, Barken D & Hoffmann A (2005) Stimulus specificity of gene expression programs determined by temporal control of IKK activity. *Science* **309**, 1857–1861.
- 144 Covert MW, Leung TH, Gaston JE & Baltimore D (2005) Achieving stability of lipopolysaccharide-induced NF-kappaB activation. *Science* **309**, 1854–1857.
- 145 White TA & Kell DB (2004) Comparative genomic assessment of novel broad-spectrum targets for anti-bacterial drugs. *Comp Func Genomics* **5**, 304–327.
- 146 Wolf DM & Arkin AP (2003) Motifs, modules and games in bacteria. *Curr Opin Microbiol* **6**, 125–134.
- 147 Yeger-Lotem E, Sattath S, Kashan N, Itzkovitz S, Milo R, Pinter RY, Alon U & Margalit H (2004) Network motifs in integrated cellular networks of transcription-regulation and protein–protein interaction. *Proc Natl Acad Sci USA* **101**, 5934–5939.
- 148 Woodward AM, Rowland JJ & Kell DB (2004) Fast automatic registration of images using the phase of a complex wavelet transform: application to proteome gels. *Analyst* **129**, 542–552.
- 149 Isaacs FJ, Blake WJ & Collins JJ (2005) Molecular biology. Signal processing in single cells. *Science* **307**, 1886–1888.
- 150 Mendes P, Kell DB & Welch GR (1995) Metabolic channeling in organized enzyme systems: experiments and models. *Enzymology in Vivo* (Brindle, K M, ed.), pp. 1–19. JAI Press, London.
- 151 Ovádi J (1995) *Cell Architecture and Metabolic Channeling*. Springer-Verlag, New York.
- 152 Agius L & Sherratt HSA (1997) *Channelling in Intermediary Metabolism*. Portland Press, London.
- 153 Ovádi J & Srere PA (2000) Macromolecular compartmentation and channeling. *Int Rev Cytol* **192**, 255–280.
- 154 Buchler NE, Gerland U & Hwa T (2003) On schemes of combinatorial transcription logic. *Proc Natl Acad Sci USA* **100**, 5136–5141.
- 155 Westerhoff HV, Tsong TY, Chock PB, Chen Y & Astumian RD (1986) How enzymes can capture and transmit free energy from an oscillating electric field. *Proc Natl Acad Sci USA* **83**, 4734–4738.
- 156 Westerhoff HV, Astumian RD & Kell DB (1988) Mechanisms for the interaction between nonstationary electric fields and biological systems.2. Nonlinear dielectric theory and free-energy transduction. *Ferroelectrics* **86**, 79–101.
- 157 Woodward AM & Kell DB (1990) On the nonlinear dielectric properties of biological systems. *Saccharomyces cerevisiae. Bioelectrochem Bioenerg* **24**, 83–100.
- 158 Woodward AM, Jones A, Zhang X, Rowland J & Kell DB (1996) Rapid and non-invasive quantification of metabolic substrates in biological cell suspensions using nonlinear dielectric spectroscopy with multivariate calibration and artificial neural networks. Principles and applications. *Bioelectrochem Bioenerg* **40**, 99–132.
- 159 Kell DB, Woodward AM, Davies E, Todd RW, Evans MF & Rowland JJ (2004) Nonlinear dielectric spectroscopy of biological systems: principles and applications. *Nonlinear Dielectric Phenomena in Complex Liquids* (Rzoska SJ & Zhelezny VP, eds), pp. 335–344. Kluwer, Dordrecht.
- 160 Mikulecky DC (1983) Network thermodynamics: a candidate for a common language for theoretical and experimental biology. *Am J Physiol* **245**, R1–R9.
- 161 Mikulecky DC (2001) Network thermodynamics and complexity: a transition to relational systems theory. *Comput Chem* **25**, 369–391.
- 162 Westerhoff HV & van Dam K (1987) *Thermodynamics and Control of Biological Free Energy Transduction*. Elsevier, Amsterdam.
- 163 Koza JR, Mydlowec W, Lanza G, Yu J & Keane MA (2001) Automatic synthesis of both the topology and

- sizing of metabolic pathways using genetic programming. Proceedings of the GECCO-2001. (Spector L, Goodman ED, Wu A, Langdon WB, General M, Sen S, Dorigo M, Pezeshk S, Garzon MH & Burke E, eds), pp. 57–65. Morgan Kaufmann, San Francisco.
- 164 Tyson JJ, Chen K & Novak B (2001) Network dynamics and cell physiology. *Nat Rev Mol Cell Biol* **2**, 908–916.
- 165 Csete ME & Doyle JC (2002) Reverse engineering of biological complexity. *Science* **295**, 1664–1669.
- 166 Kramer BP, Fischer C & Fussenegger M (2004) Biologic gates enable logical transcription control in mammalian cells. *Biotechnol Bioeng* **87**, 478–484.
- 167 Deckard A & Sauro HM (2004) Preliminary studies on the in silico evolution of biochemical networks. *Chem-biochem* **5**, 1423–1431.
- 168 Milo R, Itzkovitz S, Kashtan N, Levitt R, Shen-Orr S, Ayzenshtat I, Sheffer M & Alon U (2004) Superfamilies of evolved and designed networks. *Science* **303**, 1538–1542.
- 169 Kashtan N & Alon U (2005) Spontaneous evolution of modularity and network motifs. *Proc Natl Acad Sci USA* **102**, 13773–13778.
- 170 Endy D & Brent R (2001) Modelling cellular behaviour. *Nat* **409**, 391–395.
- 171 Pethig R & Kell DB (1987) The passive electrical properties of biological systems: their significance in physiology, biophysics and biotechnology. *Phys Med Biol* **32**, 933–970.
- 172 Chen W-K (1986) *Passive and Active Filters: Theory and Implementations*. Wiley, New York.
- 173 Rosenfeld N & Alon U (2003) Response delays and the structure of transcription networks. *J Mol Biol* **329**, 645–654.
- 174 Shen-Orr SS, Milo R, Mangan S & Alon U (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet* **31**, 64–68.
- 175 Mangan S & Alon U (2003) Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci USA* **100**, 11980–11985.
- 176 Barkai N & Leibler S (1997) Robustness in simple biochemical networks. *Nature* **387**, 913–917.
- 177 von Dassow G, Meir E, Munro EM & Odell GM (2000) The segment polarity network is a robust development module. *Nature* **406**, 188–192.
- 178 Ma L & Iglesias PA (2002) Quantifying robustness of biochemical network models. *BMC Bioinformatics* **3**. <http://www.biomedcentral.com/1471-2105/3/38>.
- 179 Morohashi M, Winn AE, Borisuk MT, Bolouri H, Doyle J & Kitano H (2002) Robustness as a measure of plausibility in models of biochemical networks. *J Theor Biol* **216**, 19–30.
- 180 Ebenhoh O & Heinrich R (2003) Stoichiometric design of metabolic networks: multifunctionality, clusters, optimization, weak and strong robustness. *Bull Math Biol* **65**, 323–357.
- 181 Aldana M & Cluzel P (2003) A natural class of robust networks. *Proc Natl Acad Sci USA* **100**, 8710–8714.
- 182 Kitano H (2004) Biological robustness. *Nat Rev Genet* **5**, 826–837.
- 183 Schmitt BM (2004) The concept of ‘buffering’ in systems and control theory: from metaphor to math. *Chembiochem* **5**, 1384–1392.
- 184 Stelling J, Sauer U, Szallasi Z & Doyle FJ (2004) 3rd & Doyle, J. Robustness of cellular functions. *Cell* **118**, 675–685.
- 185 Chaves M, Albert R & Sontag ED (2005) Robustness and fragility of Boolean models for genetic regulatory networks. *J Theor Biol* **235**, 431–449.
- 186 Chen BS, Wang YC, Wu WS & Li WH (2005) A new measure of the robustness of biochemical networks. *Bioinformatics* **21**, 2698–2705.
- 187 Wagner A (2005) Circuit topology and the evolution of robustness in two-gene circadian oscillators. *Proc Natl Acad Sci USA* **102**, 11775–11780.
- 188 Strömbäck L & Lambrix P (2005) Representations of molecular pathways: an evaluation of SBML, PSI MI and BioPAX. *Bioinformatics* **21**, 4401–4407.
- 189 Cornell M, Paton NW, Hedeler C, Kirby P, Delneri D, Hayes A & Oliver SG (2003) GIMS: an integrated data storage and analysis environment for genomic and functional data. *Yeast* **20**, 1291–1306.
- 190 Spellman P, Miller M, Stewart J, Troup C, Sarkans U, Chervitz S, Bernhart D, Sherlock G, Ball C, Lepage M, *et al.* (2002) Design and implementation of microarray gene expression markup language (MAGE-ML). *Genome Biol* **3**, research0046.1-0046.9.
- 191 Hermjakob H, Montecchi-Palazzi L, Bader G, Wojcik J, Salwinski L, Ceol A, Moore S, Orchard S, Sarkans U, von Mering C, *et al.* (2004) The HUPO PSI’s molecular interaction format – a community standard for the representation of protein interaction data. *Nat Biotechnol* **22**, 177–183.
- 192 Garwood KL, McLaughlin T, Garwood C, Joens S, Morrison N, Taylor CF, Carroll K, Evans C, Whetton AD, Hart S, *et al.* (2004) PEDRo: a database for storing, searching and disseminating experimental proteomics data. *BMC Genomics*, doi:10.1186/1471-2164-5-68.
- 193 Orchard S, Hermjakob H & Apweiler R (2003) The proteomics standards initiative. *Proteomics* **3**, 1374–1376.
- 194 Orchard S, Hermjakob H, Julian RK Jr, Runte K, Sherman D, Wojcik J, Zhu W & Apweiler R (2004) Common interchange standards for proteomics data: public availability of tools and schema. *Proteomics* **4**, 490–491.
- 195 Jenkins H, Hardy N, Beckmann M, Draper J, Smith AR, Taylor J, Fiehn O, Goodacre R, Bino R, Hall R,

- et al.* (2004) A proposed framework for the description of plant metabolomics experiments and their results. *Nat Biotechnol* **22**, 1601–1606.
- 196 Lindon JC, Nicholson JK, Holmes E, Keun HC, Craig A, Pearce JT, Bruce SJ, Hardy N, Sansone SA, Antti H, *et al.* (2005) Summary recommendations for standardization and reporting of metabolic analyses. *Nat Biotechnol* **23**, 833–838.
- 197 Xirasagar S, Gustafson S, Merrick BA, Tomer KB, Stasiewicz S, Chan DD, Yost KJ 3rd, Yates JR, 3rd, Sumner S, Xiao N, & Waters MD (2004) CEBS object model for systems biology data, SysBio-OM. *Bioinformatics* **20**, 15.
- 198 Achard F, Vaysseix G & Barillot E (2001) XML, bioinformatics and data integration. *Bioinformatics* **17**, 115–125.
- 199 Jones AR & Paton NW (2005) An analysis of extensible modelling for functional genomics data. *BMC Bioinformatics* **6**, 235 ff.
- 200 Soldatova LN & King RD (2005) Are the current ontologies in biology good ontologies? *Nat Biotechnol* **23**, 1095–1098.
- 201 Goble CA, Stevens R, Ng G, Bechhofer S, Paton NW, Baker PG, Peim M & Brass A (2001) Transparent access to multiple bioinformatics information sources. *IBM Syst J* **40**, 532–551.
- 202 Sauro HM, Hucka M, Finney A, Wellock C, Bolouri H, Doyle J & Kitano H (2003) Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration. *Omic* **7**, 355–372.
- 203 Oinn T, Addis M, Ferris J, Marvin D, Senger M, Greenwood M, Carver T, Glover K, Pocock MR, Wipat A & Li P (2004) Taverna. a tool for the composition and enactment of bioinformatics workflows. *Bioinformatics* **20**, 3045–3054.
- 204 Lu Q, Hao P, Curcin V, He W, Li YY, Luo QM, Guo YK & Li YX (2005) KDE Bioscience: Platform for bioinformatics analysis workflows. *J Biomed Inform*, doi:10.1016/j.jbi.2005.09.001.
- 205 Wilkinson MD & Links M (2002) BioMOBY: an open source biological web services proposal. *Brief Bioinform* **3**, 331–341.
- 206 Curcin V, Ghanem M & Guo Y (2005) Web services in the life sciences. *Drug Discov Today* **10**, 865–871.
- 207 Wilkinson M, Schoof H, Ernst R & Haase D (2005) BioMOBY successfully integrates distributed heterogeneous bioinformatics Web Services. The PlaNet exemplar case. *Plant Physiol* **138**, 5–17.
- 208 Arkin AP (2001) Synthetic cell biology. *Curr Opin Biotechnol* **12**, 638–644.
- 209 Blake WJ & Isaacs FJ (2004) Synthetic biology evolves. *Trends Biotechnol* **22**, 321–324.
- 210 Ferber D (2004) Synthetic biology. Microbes made to order. *Science* **303**, 158–161.
- 211 Gibbs WW (2004) Synthetic life. *Sci Am* **290**, 74–81.
- 212 Sismour AM & Benner SA (2005) Synthetic biology. *Expert Opin Biol Ther* **5**, 1409–1414.
- 213 Benner SA & Sismour AM (2005) Synthetic biology. *Nat Rev Genet* **6**, 533–543.
- 214 Hasty J, Isaacs F, Dolnik M, McMillen D & Collins JJ (2001) Designer gene networks: Towards fundamental cellular control. *Chaos* **11**, 207–220.
- 215 Hasty J, McMillen D & Collins JJ (2002) Engineered gene circuits. *Nature* **420**, 224–230.
- 216 Kærn M, Blake WJ & Collins JJ (2003) The engineering of gene regulatory networks. *Annu Rev Biomed Eng* **5**, 179–206.
- 217 Bailey JE (1991) Toward a science of metabolic engineering. *Science* **252**, 1668–1675.
- 218 Stephanopoulos G & Vallino JJ (1991) Network rigidity and metabolic engineering in metabolite overproduction. *Science* **252**, 1675–1681.
- 219 Stephanopoulos G & Sinskey AJ (1993) Metabolic engineering – methodologies and future prospects. *Trends Biotechnol* **11**, 392–396.
- 220 Keasling JD (1999) Gene-expression tools for the metabolic engineering of bacteria. *Trends Biotechnol* **17**, 452–460.
- 221 Stafford DE, Yanagimachi KS, Lessard PA, Rijhwani SK, Sinskey AJ & Stephanopoulos G (2002) Optimizing bioconversion pathways through systems analysis and metabolic engineering. *Proc Natl Acad Sci USA* **99**, 1801–1806.
- 222 Khosla C & Keasling JD (2003) Metabolic engineering for drug discovery and development. *Nat Rev Drug Discov* **2**, 1019–1025.
- 223 Sweetlove LJ, Last RL & Fernie AR (2003) Predictive metabolic engineering: a goal for systems biology. *Plant Physiol* **132**, 420–425.
- 224 Ulmer KM (1983) Protein engineering. *Science* **219**, 666–671.
- 225 Richardson JS & Richardson DC (1989) The de novo design of protein structures. *Trends Biochem Sci* **14**, 304–309.
- 226 Jones DT (1994) De novo protein design using pairwise potentials and a genetic algorithm. *Protein Sci* **3**, 567–574.
- 227 Tuschcherer G & Mutter M (1995) Templates in protein de novo design. *J Biotechnol* **41**, 197–210.
- 228 Dahiyat BI & Mayo SL (1997) De novo protein design: fully automated sequence selection. *Science* **278**, 82–87.
- 229 Liu LP & Deber CM (1998) Guidelines for membrane protein engineering derived from de novo designed model peptides. *Biopolymers* **47**, 41–62.
- 230 Hill RB, Raleigh DP, Lombardi A & DeGrado WF (2000) De novo design of helical bundles as models for understanding protein folding and function. *Acc Chem Res* **33**, 745–754.

- 231 Park S, Xi Y & Saven JG (2004) Advances in computational protein design. *Curr Opin Struct Biol* **14**, 487–494.
- 232 Park S, Kono H, Wang W, Boder ET & Saven JG (2005) Progress in the development and application of computational methods for probabilistic protein design. *Comput Chem Eng* **29**, 407–421.
- 233 Schueler-Furman O, Wang C, Bradley P, Misura K & Baker D (2005) Progress in modeling of protein structures and interactions. *Science* **310**, 638–642.
- 234 Bradley P, Misura KM & Baker D (2005) Toward high-resolution de novo structure prediction for small proteins. *Science* **309**, 1868–1871.
- 235 Gellman SH (1998) Foldamers: a manifesto. *Acc Chem Res* **31**, 173–180.
- 236 Cubberley MS & Iverson BL (2001) Models of higher-order structure: foldamers and beyond. *Curr Opin Chem Biol* **5**, 650–653.
- 237 Hill DJ, Mio MJ, Prince RB, Hughes TS & Moore JS (2001) A field guide to foldamers. *Chem Rev* **101**, 3893–4012.
- 238 Cheng RP (2004) Beyond de novo protein design – de novo design of non-natural folded oligomers. *Curr Opin Struct Biol* **14**, 512–520.
- 239 Stemmer WPC (1994) Rapid evolution of a protein in vivo by DNA shuffling. *Nature* **370**, 389–391.
- 240 Stemmer WPC (1994) DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc Natl Acad Sci USA* **91**, 10747–10751.
- 241 Colas P, Cohen B, Jessen T, Grishina I, McCoy J & Brent R (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* **380**, 548–550.
- 242 Boder ET, Midelfort KS & Wittrup KD (2000) Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proc Natl Acad Sci USA* **97**, 10701–10705.
- 243 Reetz MT & Jaeger K-E (2000) Enantioselective enzymes for organic synthesis created by directed evolution. *Chemistry – A Eur J* **6**, 407–412.
- 244 Arnold FH, Wintrode PL, Miyazaki K & Gershenson A (2001) How enzymes adapt: lessons from directed evolution. *Trends Biochem Sci* **26**, 100–106.
- 245 Arnold FH (2001) Combinatorial and computational challenges for biocatalyst design. *Nature* **409**, 253–257.
- 246 Alexeeva M, Carr R & Turner NJ (2003) Directed evolution of enzymes: new biocatalysts for asymmetric synthesis. *Org Biomol Chem* **1**, 4133–4137.
- 247 Oates MJ, Corne DW & Kell DB (2003) The bimodal feature at large population sizes and high selection pressure: implications for directed evolution. *Recent Advances in Simulated Evolution and Learning* (Tan, K C Lim, M H Yao, X & Wang, L, eds), pp. 215–240. World Scientific, Singapore.
- 248 Joyce GF (2004) Directed evolution of nucleic acid enzymes. *Annu Rev Biochem* **73**, 791–836.
- 249 Lutz S & Patrick WM (2004) Novel methods for directed evolution of enzymes: quality, not quantity. *Curr Opin Biotechnol* **15**, 291–297.
- 250 Williams GJ, Nelson AS & Berry A (2004) Directed evolution of enzymes for biocatalysis and the life sciences. *Cell Mol Life Sci* **61**, 3034–3046.
- 251 Otten LG & Quax WJ (2005) Directed evolution: selecting today's biocatalysts. *Biomol Eng* **22**, 1–9.
- 252 Reetz MT, Boccola M, Carballeira JD, Zha D & Vogel A (2005) Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test. *Angew Chem Int Ed Engl* **44**, 4192–4196.
- 253 Conrad RC, Giver L, Tian Y & Ellington AD (1996) In vitro selection of nucleic acid aptamers that bind proteins. *Methods Enzymol* **267**, 336–367.
- 254 Brody EN, Willis MC, Smith JD, Jayasena S, Zichi D & Gold L (1999) The use of aptamers in large arrays for molecular diagnostics. *Mol Diagn* **4**, 381–388.
- 255 Jayasena SD (1999) Aptamers: an emerging class of molecules that rival antibodies in diagnostics. *Clin Chem* **45**, 1628–1650.
- 256 Famulok M, Mayer G & Blind M (2000) Nucleic acid aptamers—from selection in vitro to applications in vivo. *Acc Chem Res* **33**, 591–599.
- 257 Hermann T & Patel DJ (2000) Adaptive recognition by nucleic acid aptamers. *Science* **287**, 820–825.
- 258 Jhaveri SD, Kirby R, Conrad R, Maglott EJ, Bowser M, Kennedy RT, Glick G & Ellington AD (2000) Designed signaling aptamers that transduce molecular recognition to changes in fluorescence intensity. *JACS* **122**, 2469–2473.
- 259 Stojanovic MN, de Prada P & Landry DW (2001) Aptamer-based folding fluorescent sensor for cocaine. *J Am Chem Soc* **123**, 4928–4931.
- 260 Gold L, Brody E, Heilig J & Singer B (2002) One, two, infinity: genomes filled with aptamers. *Chem Biol* **9**, 1259–1264.
- 261 Cox JC, Hayhurst A, Hesselberth J, Bayer TS, Georgiou G & Ellington AD (2002) Automated selection of aptamers against protein targets translated in vitro: from gene to aptamer. *Nucl Acids Res* **30**, e108.
- 262 Clark SL & Remcho VT (2002) Aptamers as analytical reagents. *Electrophoresis* **23**, 1335–1340.
- 263 Luzi E, Minunni M, Tombelli S & Mascini M (2003) New trends in affinity sensing: aptamers for ligand binding. *Trac* **22**, 810–818.
- 264 Rimmele M (2003) Nucleic acid aptamers as tools and drugs: recent developments. *ChemBiochem* **4**, 963–971.
- 265 Nutiu R, Yu, JM & Li Y (2004) Signaling aptamers for monitoring enzymatic activity and for inhibitor screening. *ChemBiochem* **5**, 1139–1144.
- 266 Stojanovic MN & Kolpashchikov DM (2004) Modular aptameric sensors. *J Am Chem Soc* **126**, 9266–9270.

- 267 Blank M & Blind M (2005) Aptamers as tools for target validation. *Curr Opin Chem Biol* **9**, 336–342.
- 268 Famulok M & Mayer G (2005) Intramers and aptamers. applications in protein-function analyses and potential for drug screening. *Chembiochem* **6**, 19–26.
- 269 Famulok M (2005) Allosteric aptamers and aptazymes as probes for screening approaches. *Curr Opin Mol Ther* **7**, 137–143.
- 270 Nutiu R & Li Y (2005) In vitro selection of structure-switching signaling aptamers. *Angew Chem Int Ed Engl* **44**, 1061–1065.
- 271 Nutiu R & Li Y (2005) Aptamers with fluorescence-signaling properties. *Methods* **37**, 16–25.
- 272 Stojanovic MN, Semova S, Kolpashchikov D, Macdonald J, Morgan C & Stefanovic D (2005) Deoxyribozyme-based ligase logic gates and their initial circuits. *J Am Chem Soc* **127**, 6914–6915.
- 273 Tombelli S, Minunni M & Mascini M (2005) Analytical applications of aptamers. *Biosens Bioelectron* **20**, 2424–2434.
- 274 Proske D, Blank M, Buhmann R & Resch A (2005) Aptamers-basic research, drug development, and clinical applications. *Appl Microbiol Biotechnol* **69**, 367–374.
- 275 Hitchens GD & Kell DB (1983) Uncouplers can shuttle rapidly between localised energy coupling sites during photophosphorylation by chromatophores of *Rhodospseudomonas capsulata* N22. *Biochem J* **212**, 25–30.
- 276 Westerhoff HV & Kell DB (1988) A control theoretical analysis of inhibitor titrations of metabolic channelling. *Comments Mol Cell Biophys* **5**, 57–107.
- 277 Schreiber SL (1998) Chemical genetics resulting from a passion for synthetic organic chemistry. *Bioorg Medical Chem* **6**, 1127–1152.
- 278 Crews CM & Splittgerber U (1999) Chemical genetics: exploring and controlling cellular processes with chemical probes. *Trends Biochem Sci* **24**, 317–320.
- 279 Stockwell BR (2000) Chemical genetics: ligand-based discovery of gene function. *Nat Rev Genet* **1**, 116–125.
- 280 Stockwell BR (2000) Frontiers in chemical genetics. *Trends Biotechnol* **18**, 449–455.
- 281 Zheng XF & Chan TF (2002) Chemical genomics: a systematic approach in biological research and drug discovery. *Curr Issues Mol Biol* **4**, 33–43.
- 282 Carroll PM, Dougherty B, Ross-Macdonald P, Browman K & FitzGerald K (2003) Model systems in drug discovery: chemical genetics meets genomics. *Pharmacol Ther* **99**, 183–220.
- 283 Zanders ED, Bailey DS & Dean PM (2002) Probes for chemical genomics by design. *Drug Discovery Today* **7**, 711–718.
- 284 Giaever G (2003) A chemical genomics approach to understanding drug action. *Trends Pharmacol Sci* **24**, 444–446.
- 285 Salemme FR (2003) Chemical genomics as an emerging paradigm for postgenomic drug discovery. *Pharmacogenomics* **4**, 257–267.
- 286 Brenner C (2004) Chemical genomics in yeast. *Genome Biol* **5**, 240.
- 287 Darvas F, Dorman G, Krajcsi P, Puskas LG, Kovari Z, Lorincz Z & Urge L (2004) Recent advances in chemical genomics. *Curr Medical Chem* **11**, 3119–3145.
- 288 Meisner NC, Hintersteiner M, Uhl V, Weidemann T, Schmied M, Gstach H & Auer M (2004) The chemical hunt for the identification of drugable targets. *Curr Opin Chem Biol* **8**, 424–431.
- 289 Shim JS & Kwon HJ (2004) Chemical genetics for therapeutic target mining. *Expert Opin Ther Targets* **8**, 653–661.
- 290 Wagner BK, Haggarty SJ & Clemons PA (2004) Chemical genomics: probing protein function using small molecules. *Am J Pharmacogenomics* **4**, 313–320.
- 291 Spring DR (2005) Chemical genetics to chemical genomics: small molecules offer big insights. *Chem Soc Rev* **34**, 472–482.
- 292 Smukste I & Stockwell BR (2005) Advances in chemical genetics. *Annu Rev Genomics Hum Genet* **6**, 261–286.
- 293 Haggarty SJ, Clemons PA, Wong JC & Schreiber SL (2004) Mapping chemical space using molecular descriptors and chemical genetics: deacetylase inhibitors. *Comb Chem High Throughput Screen* **7**, 669–676.
- 294 Fan QW, Specht KM, Zhang C, Goldenberg DD, Shokat KM & Weiss WA (2003) Combinatorial efficacy achieved through two-point blockade within a signaling pathway—a chemical genetic approach. *Cancer Res* **63**, 8930–8938.
- 295 Tochtrop GP & King RW (2004) Target identification strategies in chemical genetics. *Comb Chem High Throughput Screen* **7**, 677–688.
- 296 Hastie T, Tibshirani R & Friedman J (2001) *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer-Verlag, Berlin.
- 297 Han J & Kamber M (2001) *Data Mining: Concepts and Techniques*. Morgan Kaufmann, San Francisco.
- 298 Ananiadou S & McNaught J (2006) *Text Mining in Biology and Biomedicine*. Artech House, London.
- 299 Swanson DR (1990) Medical literature as a potential source of new knowledge. *Bull Medical Libr Assoc* **78**, 29–37.
- 300 Hirschman L, Park JC, Tsujii J, Wong L & Wu CH (2002) Accomplishments and challenges in literature data mining for biology. *Bioinformatics* **18**, 1553–1561.
- 301 Nenadic G, Spasic I & Ananiadou S (2003) Terminology-driven mining of biomedical literature. *Bioinformatics* **19**, 938–943.
- 302 Corney DP, Buxton BF, Langdon WB & Jones DT (2004) BioRAT: extracting biological information from full-length papers. *Bioinformatics* **20**, 3206–3213.

- 303 Daraselia N, Yuryev A, Egorov S, Novichkova S, Nikitin A & Mazo I (2004) Extracting human protein interactions from MEDLINE using a full-sentence parser. *Bioinformatics* **20**, 604–611.
- 304 Hakenberg J, Schmeier S, Kowald A, Klipp E & Leser U (2004) Finding kinetic parameters using text mining. *Omics* **8**, 131–152.
- 305 Rzhetsky A, Iossifov I, Koike T, Krauthammer M, Kra P & Morris M., YuH, Duboue PA, Weng W, Wilbur WJ, Hatzivassiloglou V & Friedman C (2004) GeneWays: a system for extracting, analyzing, visualizing, and integrating molecular pathway data. *J Biomed Inform* **37**, 43–53.
- 306 Hoffmann R, Krallinger M, Andres E, Tamames J, Blaschke C & Valencia A (2005) Text mining for metabolic pathways, signaling cascades, and protein networks. *Sci STKE* pe21.
- 307 Vailaya A, Bluvas P, Kincaid R, Kuchinsky A, Creech M & Adler A (2005) An architecture for biological information extraction and representation. *Bioinformatics* **21**, 430–438.
- 308 Spasic I, Ananiadou S, McNaught J & Kumar A (2005) Text mining and ontologies in biomedicine: Making sense of raw text. *Briefings in Bioinformatics* **6**, 239–251.
- 309 Chalfie M & Kain S (1998) *Green Fluorescent Protein: Properties, Applications, and Protocols*. Wiley-Liss, New York.
- 310 Uhlen M & Ponten F (2005) Antibody-based proteomics for human tissue profiling. *Mol Cell Proteomics* **4**, 384–393.
- 311 Fehr M, Lalonde S, Lager I, Wolff MW & Frommer WB (2003) *In vivo* imaging of the dynamics of glucose uptake in the cytosol of COS-7 cells by fluorescent nanosensors. *J Biol Chem* **278**, 19127–19133.
- 312 Famulok M (2004) Green fluorescent RNA. *Nature* **430**, 976–977.
- 313 Rosi NL & Mirkin CA (2005) Nanostructures in bi-diagnostics. *Chem Rev* **105**, 1547–1562.
- 314 Rotman B (1961) Measurement of activity of single molecules of β -D-galactosidase. *Proc Natl Acad Sci* **47**, 1981–1991.
- 315 Xie XS & Lu HP (1999) Single-molecule enzymology. *J Biol Chem* **274**, 15967–15970.
- 316 Moore KJ, Turconi S, Ashman S, Ruediger M, Haupts U, Emerick V & Pope AJ (1999) Single molecule detection technologies in miniaturized high throughput screening: fluorescence correlation spectroscopy. *J Biomolecular Screening* **4**, 335–353.
- 317 Haupts U, Rüdiger M, Ashman S, Turconi S, Bingham R, Wharton C, Hutchinson J, Carey C, Moore KJ & Pope AJ (2003) Single-molecule detection technologies in miniaturized high-throughput screening: fluorescence intensity distribution analysis. *J Biomol Screening* **8**, 19–33.
- 318 Bannai M, Higuchi K, Akesaka T, Furukawa M, Yamaoka M, Sato K & Tokunaga K (2004) Single-nucleotide-polymorphism genotyping for whole-genome-amplified samples using automated fluorescence correlation spectroscopy. *Anal Biochem* **327**, 215–221.
- 319 Twist CR, Winson MK, Rowland JJ & Kell DB (2004) SNP detection using nanomolar nucleotides and single molecule fluorescence. *Anal Biochem* **327**, 35–44.
- 320 Bennett ST, Barnes C, Cox A, Davies L & Brown C (2005) Toward the \$1000 human genome. *Pharmacogenomics* **6**, 373–382.
- 321 Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380.
- 322 Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD & Church GM (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728–1732.
- 323 Prigogine I (1980) *From Being to Becoming: Time and Complexity in the Physical Sciences*. W.H. Freeman, San Francisco.
- 324 Coveney P & Highfield R (1990) *The Arrow of Time*. W.H. Allen, London.
- 325 Nicolis G & Prigogine I (1977) *Self-organization in Nonequilibrium Systems: From Dissipative Structures to Order Through Fluctuations*. Wiley, New York.
- 326 Kauffman SA (1993) *The Origins of Order*. Oxford University Press, Oxford.
- 327 Holland JH (1998) *Emergence*. Helix, Reading, MA.
- 328 Johnson S (2001) *Emergence: the Connected Lives of Ants, Brains, Cities and Software*. Scribner, New York.
- 329 Barabási A-L (2002) *Linked: the New Science of Networks*. Perseus Publishing, Cambridge, MA.
- 330 Buchanan M (2002) *Nexus: Small Worlds and the Groundbreaking Science of Networks*. W.W. Norton, New York.
- 331 Coveney PV & Highfield RR (1995) *Frontiers of Complexity*. Faber & Faber, London.
- 332 Kauffman SA (1995) *At Home in the Universe: the Search for Laws of Self-Organization and Complexity*. Oxford University Press, Oxford.
- 333 Solé R & Goodwin B (2000) *Signs of Life: How Complexity Pervades Biology*. Basic Books, New York.
- 334 Lipton P (1991) *Inference to the Best Explanation*. Routledge, London.
- 335 Pearl J (2000) *Causality: Models, Reasoning and Inference*. Cambridge University Press, Cambridge.
- 336 Shipley B (2001) *Cause and Correlation in Biology. A User's Guide to Path Analysis, Structural Equations and Causal Inference*. Cambridge University Press, Cambridge.

- 337 Mackay DJC (2003) *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, Cambridge.
- 338 Laughlin RB (2005) *A Different Universe: Reinventing Physics from the Bottom Down*. Basic Books, New York.
- 339 Kell DB & Welch GR (1991) No turning back, Reductionism and Biological Complexity. *Times Higher Educational* (Suppl.) 9th August, p. 15.
- 340 Westerhoff HV (2001) The silicon cell, not dead but live!. *Metab Eng* **3**, 207–210.
- 341 Pe'er D, Regev A, Elidan G & Friedman N (2001) Inferring subnetworks from perturbed expression profiles. *Bioinformatics* **17** (Suppl. 1), S215–S224.
- 342 de la Fuente A, Brazhnik P & Mendes P (2002) Linking the genes: inferring quantitative gene networks from microarray data. *Trends Genet* **18**, 395–398.
- 343 Kholodenko BN, Kiyatkin A, Bruggeman FJ, Sontag E, Westerhoff HV & Hoek JB (2002) Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. *Proc Natl Acad Sci USA* **99**, 12841–12846.
- 344 Stark J, Callard R & Hubank M (2003) From the top down: towards a predictive biology of signalling networks. *Trends Biotechnol* **21**, 290–293.
- 345 Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D & Friedman N (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet* **34**, 166–176.
- 346 King RD, Garrett SM & Coghill GM (2005) On the use of qualitative reasoning to simulate and identify metabolic pathways. *Bioinformatics* **21**, 2017–2026.
- 347 Sachs K & Perez O, Pe'er D, Lauffenburger DA & Nolan GP (2005) Causal protein-signaling networks derived from multiparameter single-cell data. *Science* **308**, 523–529.
- 348 Corne D, Jerram NR, Knowles J & Oates M (2001) PESA-II: Region-based selection in evolutionary multi-objective optimization. Paper presented at the GECCO – Proceedings of the Genetic and Evolutionary Computation Conference, San Francisco. CA.
- 349 Cornish-Bowden A, Hofmeyr J-HS & Cárdenas ML (1995) Strategies for manipulating metabolic fluxes in biotechnology. *Bioorg Chem* **23**, 439–449.
- 350 Fell DA & Thomas S (1995) Physiological control of metabolic flux: the requirement for multisite modulation. *Biochem J* **311**, 35–39.
- 351 Fell DA (1998) Increasing the flux in metabolic pathways: a metabolic control analysis perspective. *Biotechnol Bioeng* **58**, 121–124.
- 352 Cascante M, Boros LG, Comin-Anduix B, de Atauri P, Centelles JJ & Lee PW (2002) Metabolic control analysis in drug discovery and disease. *Nat Biotechnol* **20**, 243–249.
- 353 McCafferty DG, Cudic P, Yu MK, Behenna DC & Kruger R (1999) Synergy and duality in peptide antibiotic mechanisms. *Curr Opin Chem Biol* **3**, 672–680.
- 354 Borisy AA, Elliott PJ, Hurst NW, Lee MS, Lehar J, Price ER, Serbedzija G, Zimmermann GR, Foley MA, Stockwell BR & Keith CT (2003) Systematic discovery of multicomponent therapeutics. *Proc Natl Acad Sci USA* **100**, 7977–7982.