ELSEVIER

Logic models of pathway biology

Steven Watterson[1,2], Stephen Marshall[3] and Peter Ghazal[1,2]

[1] Division of Pathway Medicine, University of Edinburgh Medical School, Chancellor's Building, 49 Little France Crescent, Edinburgh EH16 4SB, Scotland, United Kingdom
[2] Centre for Systems Biology at Edinburgh, University of Edinburgh, Darwin Building, The King's Buildings, Mayfield Road, Edinburgh EH9 3JZ, Scotland, United Kingdom
[3] Department of Electronic and Electrical Engineering, University of Strathclyde, Royal College Building, 204 George Street, Glasgow G1 1XW, Scotland, United Kingdom

Living systems seamlessly perform complex information processing and control tasks using combinatorially complex sets of biochemical reactions. Drugs that therapeutically modulate the biological processes of disease are developed using single protein target strategies, often with limited knowledge of the complex underlying role of the targets. Approaches that attempt to consider the combinatorial complexity from the outset might help identify any causal relationships that could lead to undesirable or adverse side effects earlier in the development pipeline. Such approaches, in particular logic methodologies, might also aid pathway selection and multiple target strategies during the drug discovery phase. Here, we describe the use of logic as a tractable and informative approach to modelling biological pathways that can allow us to improve our understanding of the dependencies in complex biological processes.

Introduction

Cell biology has historically focussed on the identification of individual interactions between genes, proteins, RNAs and metabolites and on determining how each interaction affects the phenotype. Chains of causal interactions, known as *pathways*, mediate the signals that travel around and between cells, essentially controlling the phenotypic behaviour. However, little is generally known about the structure of the pathways because they comprise large numbers of interactions and, until recently, experiments to study their individual constituents incurred great time and cost.

All this has changed with the arrival of high-dimensional data sets, generated using large-scale genomic, proteomic and metabolomic techniques. Laboratories can now extract an abundance of data from each experiment with relative ease and this has opened the door to both hypothesis driven and data driven studies. This wealth of information has initiated a move away from focussed, single gene or protein investigations, towards system level studies of multiple interaction networks [1]. The annotation and understanding of the cause–effect relationships between multiple genes and gene products can aptly be described as *pathway biology*. It is hoped that such system level studies will integrate high-dimensional data with the wealth of information that exists in previously published investigations to produce consensus descriptions of pathways [2,3].

The current challenge to systems level studies is to model the pathways usefully and this is not an insignificant problem. In general terms, modelling requires us to identify the key features of a complex process and to build a representation that captures those features while simplifying everything else. This allows us to focus on the important properties, without distraction. For pathway modelling, we need a representation that captures the cause–effect relationships associated with each interaction, but that is conceptually and computationally simple.

Computational simplicity is critical to drug target identification. By keeping the representation simple, we minimize the time needed to simulate the activity of each pathway and this, in turn,

Corresponding authors: Watterson, S. (s.watterson@ed.ac.uk), Ghazal, P. (p.ghazal@ed.ac.uk)

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2008.03.019

allows us to consider the activity across a broad range of conditions in a manageable amount of time. With a simple representation, we can comb the pathways for the upstream interactions that control downstream behaviour, allowing us to identify lead candidates for therapeutic treatments. Ultimately, a comprehensive understanding of the pathway structure will also allow us to anticipate potential complications or side effects that a potential treatment could incur.

Several modelling schemes have been touted as suitable representations for pathways. Ordinary and partial differential equations have proven to be very successful in metabolomics [4,5], including the modelling of human metabolism pathways [6], such as cholesterol [7], but they require concentration and reaction rate data, which is largely unknown for pathways and not necessarily relevant to understanding their function. Stochastic schemes capture very well the random nature of individual molecular events [8,9], but inefficiently describe large numbers of proteins. Petri nets have also been considered and, while the formulation is computationally simple, it is conceptually complex and can only approximate large flows [10–12]. Stochastic petri nets [13] and hybrid petri nets [14] have been proposed as extensions. Pi calculus has been used to describe the pathways as a symbolic construction to which quantitative models can be fitted [15,16].

Several authors have considered using logic to discuss pathways [17–21]. In some cases, this has been as part of network inference processes [22,23], but here, we introduce Boolean logic as a means to directly describe a pathway. We provide an overview of the variables and the simplest operations, showing how a logic framework captures the dependency properties of the pathway interactions. We compare and contrast pathway biology with digital electronics and review some of the ways in which the use of logic has been developed.

Defining pathway biology in terms of logic

A biological pathway

Biological pathways can comprise one or more of the following: metabolic pathways, molecular interactions, gene regulatory networks and signalling pathways. The boundaries between these seemingly distinct categories are entirely artificial and exist because historically they bounded different areas of expertise. By definition, a pathway must have a starting state (the input) and an end state (the output).

The principal objective of pathway biology is to determine the relationship between the starting and end states. Similar problems are routinely faced by the electronics industry, in which circuits (which are, in effect, electrical pathways) are studied.

In digital electronics signals can take only one of two states. These states may be referred to as ON or OFF, TRUE or FALSE or more commonly 1 or 0. The two-state signals, suggest themselves as a potential representation for gene activity, which, broadly speaking, also operates on a two-state, ON or OFF basis.

In order to transform expression level data into two-state signals, it is necessary to discretize the data. This can be done by applying thresholds as shown in Figure 1.

Boolean logic

By discretizing expression data, the input and output states of a biological pathway may be recorded, but to unravel the causal relationship between them, we must introduce Boolean logic, which is the branch of mathematics widely used in the design and analysis of digital electronic systems. The origins of Boolean logic are described in Box 1.

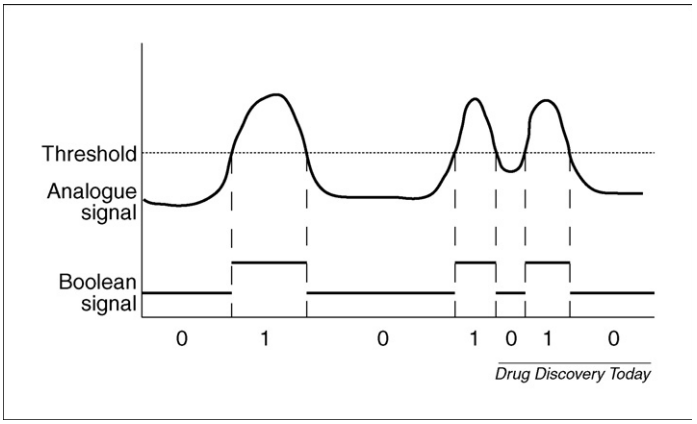

FIGURE 1

Discretizing a continuous, analogue signal. Regions above the threshold have the Boolean value 1 and regions below the threshold have the Boolean value 0.

Boolean logic consists of a small set of basic operations on two-state variables. The simplest is the NOT operation (inversion). If the variable F is related to the variable A by F = NOT A, then when A = 0, F = 1 and vice versa. This relationship can be drawn as the *logic symbol* shown in Figure 2a.

Together with NOT, the most common operations are AND, OR, XOR. The latter three relate two input variables to the output. The AND operation requires that both inputs be 1 in order for the output to be 1; otherwise the output is 0. The logic symbol for AND is shown in Figure 2b. The OR operation requires that only one of its inputs be 1 in order for the output to be 1; otherwise the output

BOX 1

Boolean logic

Boolean logic owes its name to George Boole (1815–1864), the first professor of mathematics at University College, Cork, Ireland. In two publications that straddled his appointment [48,49], he devised a mathematical framework intended for the logical reasoning that was commonly found in philosophy.

Outside of academia, his framework found no natural application and it languished in academic obscurity, until Claude Shannon, a Master's student at MIT some seventy years later, realised that it could be used to simplify the computing systems that were responsible for routing telephone signals [50]. This idea had a great impact on electrical engineering and Shannon saw potential for Boolean logic in other areas of science. Following his Master's success, he took to Mendelian genetics, using Boolean logic to derive a range of laws governing the dynamics and proportions of genetic populations. In 1940, MIT awarded him his PhD for his thesis, an algebra for theoretical genetics (MIT-THESES//1940-3). However, it was his contribution to electronics that left the greatest legacy.

Having used logic to study routing systems, his next idea was revolutionary. He reasoned that this relationship could be turned on its head so that routing systems could be used to analyse complex logic problems. In conceiving this idea, he invented the notion of digital electronics and sowed the seed from which the current multi-billion dollar electronics industry has grown.

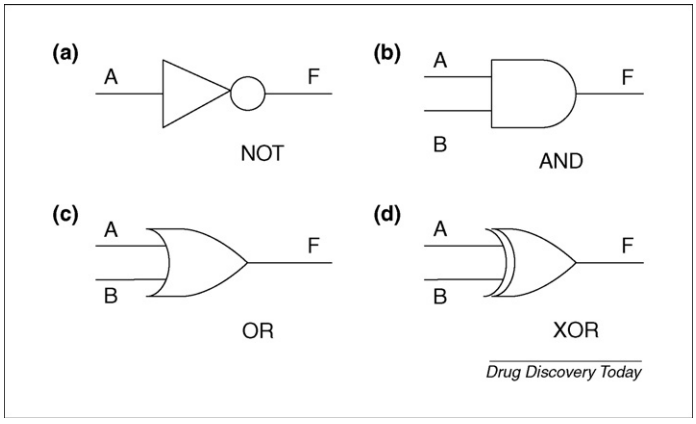

FIGURE 2

Logic gates drawn using the symbols from digital electronics.

is 0. The logic symbol for OR is shown in Figure 2c. The XOR (Exclusive-OR) gives a 1 output when the inputs are different and a 0 output when they are the same. The logic symbols for XOR is shown in Figure 2d.

A truth table showing the outputs of these basic operations for all combinations of inputs is shown in Table 1.

The dependencies represented by the Boolean operations lend themselves to descriptions of physical systems and this was first recognized by Kauffman and co-workers [21,24]. For example, if two different components combine to make a third component, as in the chemical reaction NaCl = Na + Cl, we can state that the presence of NaCl requires the presence of Na and the presence of Cl. If we describe presence as state 1 and absence as state 0, then we can write NaCl = Na AND Cl. Similarly, we can state that the presence of Salts depends on the presence of Cl and either Na or K. We can write this as Salts = (Na OR K) AND Cl. The NOT operation can be used to describe inhibition. The presence of NaCl requires the presence of Na, the presence of Cl and the absence of K, which is more reactive than Na and we can write this as NaCl = (Na AND NOT K) AND Cl.

This type of description omits the molecular details of each interaction and instead describes just the dependencies between components. In doing so, it enjoys great computational efficiency, but is limited to systems where our understanding is reasonably complete and we can assume that the pathway behaves predictably and consistently.

Pathway modelling

The interactions between proteins, genes and complexes can usually be classified into one, or a combination, of complex formation, phosphorylation, gene activation, inhibition, equivalent binding and disassociation. By translating each of these interaction types into a logic description, it is possible to build a general framework that can be used to model pathways.

TABLE 1

The truth table for the basic logic operations

A	B	AND	OR	XOR
0	0	0	0	0
0	1	0	1	1
1	0	0	1	1
1	1	1	1	0

Biological pathways can be readily depicted in diagrammatic form and several diagrammatic schemes have been developed for this purpose [25–28]. Figure 3 shows an example of each type of the listed interactions, drawn using the current proposal of the Systems Biology Graphical Notation (SBGN) [29].

Figure 3a shows complex formation, where the proteins A and B combine to form the complex A:B. A and B are the inputs to the interaction and A:B is the output. Both inputs must be present (in the 1 state) in order for the output to be present (in the 1 state) and this is described by the AND dependency: (complex A:B) = (protein A) AND (protein B).

Figure 3b shows the phosphorylation of protein A. The inputs to this interaction are protein B and unphosphorylated protein A. The output is phosphorylated protein A. Both inputs must be present in order for phosphorylated A to be present: (phosphorylated protein A) = (unphosphorylated protein A) AND (protein B).

Gene activation is shown in Figure 3c. Here, inactive gene B and protein A are inputs to the interaction and active gene B is the output. This also demonstrates an AND dependency: (active gene B) = (inactive gene B) AND (protein A).

Figure 3d shows equivalent binding, where inactive gene B can be activated by either protein A or protein C. We have three inputs: inactive gene B and proteins A and C. The output is active gene B. Inactive gene B must be present for activation to occur, but only one of protein A or protein C need be present. We can describe this in two stages; the first determines whether the proteins can cause activation; the second potentially activates the gene. In the first stage, activation requires either A or C to be present and this can be described as (protein A) OR (protein C). In the second, activation will occur if inactive gene B is present and the result of the first stage is 1. Thus, we have (active gene B) = (inactive gene B) AND ((protein A) OR (protein C)).

Dissociation is shown in Figure 3e. Here, we have one input and two outputs. The complex A:B is the sole input and the constituent proteins A and B are the outputs. Neither protein A nor protein B can exist if the complex A:B does not initially exist, so we can write this as two logic statements: (protein A) = (complex A:B) and (protein B) = (complex A:B).

Inhibition is shown in Figure 3f. The dimer A:A activates gene B and this is inhibited by protein C. The inputs are dimer A:A, inactive gene B and protein C. The output is active gene B. As before, we can describe this in two stages. Activation occurs if dimer A:A is present and protein C is absent. We can describe this as: (dimer A:A) AND (NOT protein C). The second stage integrates this with the activation of the gene: (active gene B) = (inactive gene B) AND ((dimer A:A) AND (NOT protein C)).

By understanding the logical dependencies of the interactions, we can describe entire pathways as logical constructs.

Biological circuitry

Pathways exist to perform information processing within the cell and by translating these functions into a form that captures the functionality without the biochemical detail, it is possible to consider classifying pathways according to their behaviour, rather than by their location or the proteins and genes involved. This opens the door to systems-level analysis, potentially allowing us to

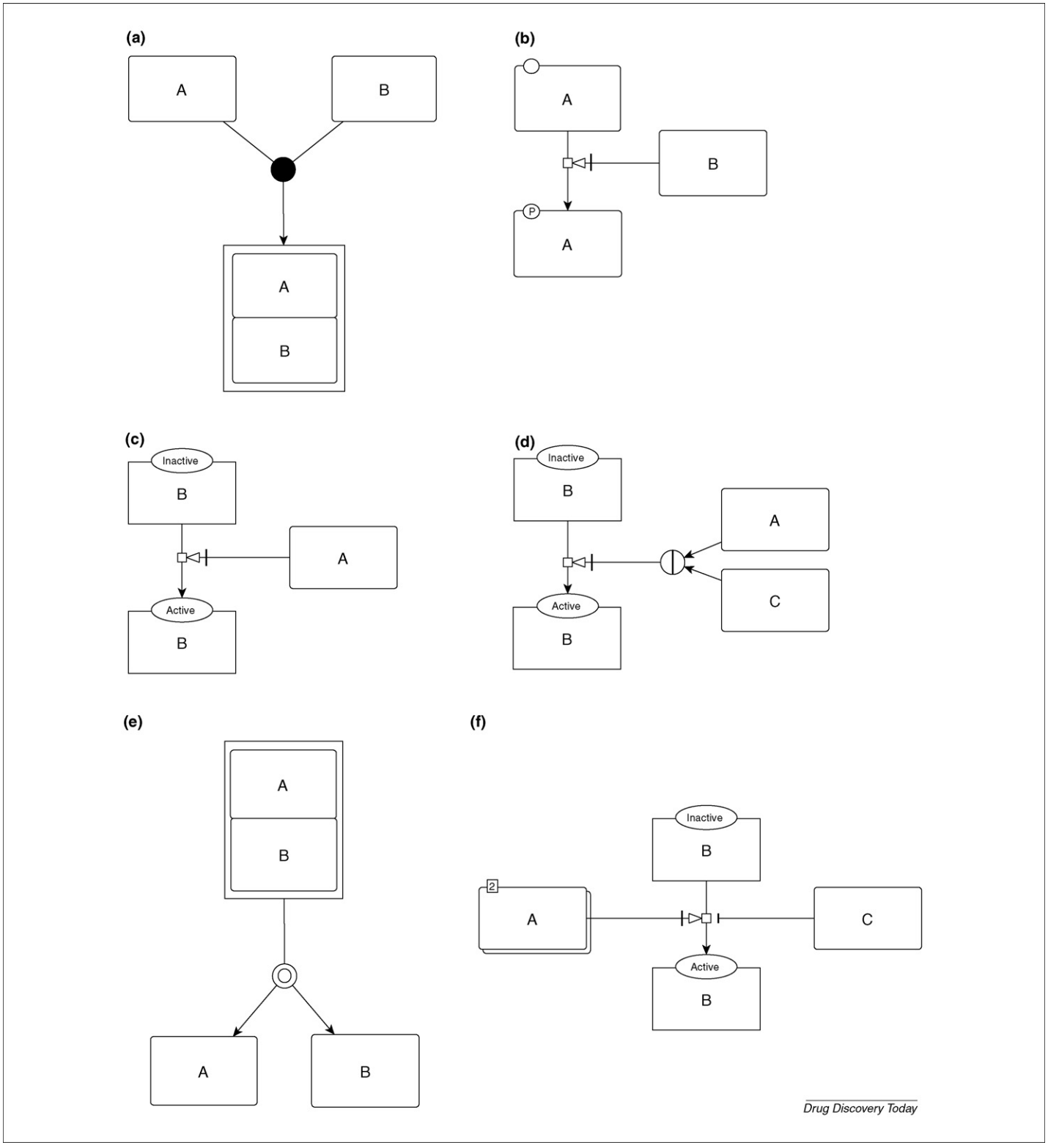

FIGURE 3

Six typical pathway interactions, drawn using the latest available version of Systems Biology Graphical Notation (SBGN). **(a)** Complex formation, **(b)** phosphorylation **(c)** activation **(d)** equivalent binding **(e)** dissociation **(f)** inhibition.

learn how nature builds complex systems and what sub-systems are used as common building blocks.

The digital electronics industry has a long tradition of thinking about electrical pathways in logic terms and, by considering pathway biology in a similar vein, it gives us access to the multitude of mathematical tools that the industry has at its disposal.

In digital electronics, it is routine to consider whole circuits using circuit diagrams. The diagrams allow the circuit designers to

explore the system-wide dependencies. We can build comparable diagrams for biological pathways, although it is important to note that biological pathways represent an evolved system in which our knowledge is incomplete. This is in marked contrast to electronic circuit diagrams which are systems designed so as to fulfil a known function.

Circuits comprise inputs, outputs and a series of gates. Biological pathways comprise inputs, outputs and interactions whose dependencies are operationally comparable to gates. It is important to retain some information of where, within the cell, the interactions take place, in order to explore potential therapies. Therefore, a natural strategy is to divide the cell into compartments and to define the logical pathway that connects the inputs to and outputs from each compartment.

Some signals pass between compartments and some may not. This gives two distinct categories of input and output signal: *migrant* inputs and outputs, which relate to another compartment; and *latent* inputs and outputs, which do not. Placing the inputs to a compartment at the top of the diagram and the outputs at the bottom, we can construct a circuit diagram for a pathway. Figures 4 and 5 illustrate this. In Figure 4 a section of the Jak-Stat signalling pathway is drawn using SBGN notation and in Figure 5 the same pathway is drawn as a circuit describing the logical dependencies of the interactions.

Describing a biological pathway in these terms allows us to test the accuracy of our understanding. By discretizing expression data for the latent inputs to the system, we can use the dependencies to calculate the state of the latent outputs. By comparing this to the discretized expression data for the proteins that correspond to the latent outputs, we have an indicator of the quality of our understanding which can also be used to test hypothetical pathways.

This framework also addresses the combinatorial complexity that accompanies large pathways. By describing signals as two state variables, a pathway with n inputs has the number of possible combination of input signals reduced to 2^n. This can be further reduced by exploiting the critical nature of the AND dependencies. For example, if a pathway has 10 inputs and requires all 10 to be in the 1 state in order for the output to be in the 1 state, the output is related to the inputs through just AND dependencies. If one of the input signals is in state 0, the output cannot take state 1, irrespective of the value of the other input signals. We can use this to represent the 1024 ($=2^{10}$) combinations of input signals with just 11 combinations: one where all the inputs are in state 1 and 10 in which each input in turn is set to zero and the values of the other inputs are irrelevant for that pathway. These combinatorial reductions illustrate that the framework can scale favourably for large pathways.

Capturing dependencies in biological pathways from the literature

The published literature provides a valuable source of information concerning the cause–effect relationships in biological pathways. A summary of the steps involved is provided in Box 2.

Although this information is not always explicit, it can be inferred from various statements in the text, which are justified by experimental evidence. For instance a statement might read: 'An adaptor protein (termed A) in the receptor signalling pathway is shown to interact with factor (named B) resulting in its phosphorylation by kinase (C) upon recruitment by A for activation'. Here, the adaptor protein A, the B factor and the kinase C must all be present in order for B to be activated by phosphorylation. Table 2 shows the truth table for the eight possible combinations, where P denotes the activated protein B.

Such an approach can increase our understanding, but caution should be applied in order to avoid inaccurate annotation. To help minimise false assertions, the dependencies should be corroborated by more than one report or publication, ideally from different research groups. A second solution is to perform the experiments to verify and extend the existing knowledge.

Developments

Because logic provides a tractable framework for modelling schemes, the notion of pathway logic has become a growing area of interest. Consequently, several groups have begun to further extend the concept.

Probabilistic Boolean networks

Large pathway networks can be combinatorially complex to study and can push the validity of the logic approximation to its limit. One strategy to address this is to replace the model of a whole, large pathway with models describing several smaller, simpler pathways, each describing an important feature of the large pathway, and to switch between the pathways according to statistical criteria. The probabilistic Boolean network uses the logical representation of the pathways to achieve this [30–33] and the formulation has been extended to include additional features such as noise and external stimuli.

Network inference

Because many pathways are relatively poorly understood, several groups have developed algorithms to infer pathways from expression data. Depending on the level of detail sought, there are many possible strategies. Liang and co-workers [34,35] have used tools from information theory to construct transition tables, describing the behaviour of a cell. Akutsu *et al.* [36] have calculated the minimum number of experiments required to infer a pathway from expression data and have extended their modelling and inferral to include noise [37]. Husmeier [38,39] has used Bayesian methods to grow best fit pathways. This has been extended to incorporate prior knowledge of the network [40]. Laubenbacher and Sigler [41] derive polynomials to fit time course data.

These algorithms are effective in what they set out to achieve and they mostly succeed in predicting the observed data. However, little progress has been made in developing these methods so that their predictions can be integrated into the known biology.

Attractor homeostasis

Pathways often contain feedback, and as such there is the possibility that the system can enter into a repeated cycle of a particular expression pattern. In some cases, such as in tissue-specific expression patterns, the system may enter into one, self-sustaining pattern. These patterns and cycles are known as *attractors* and it is important to note that a single pathway with feedback can potentially exhibit many different attractors. This flexibility has been proposed as a way to explain the range of phenotypic behaviour displayed by a cell [42]. Huang and Wuensche [43,44] have considered the selection of attractors as a mechanism

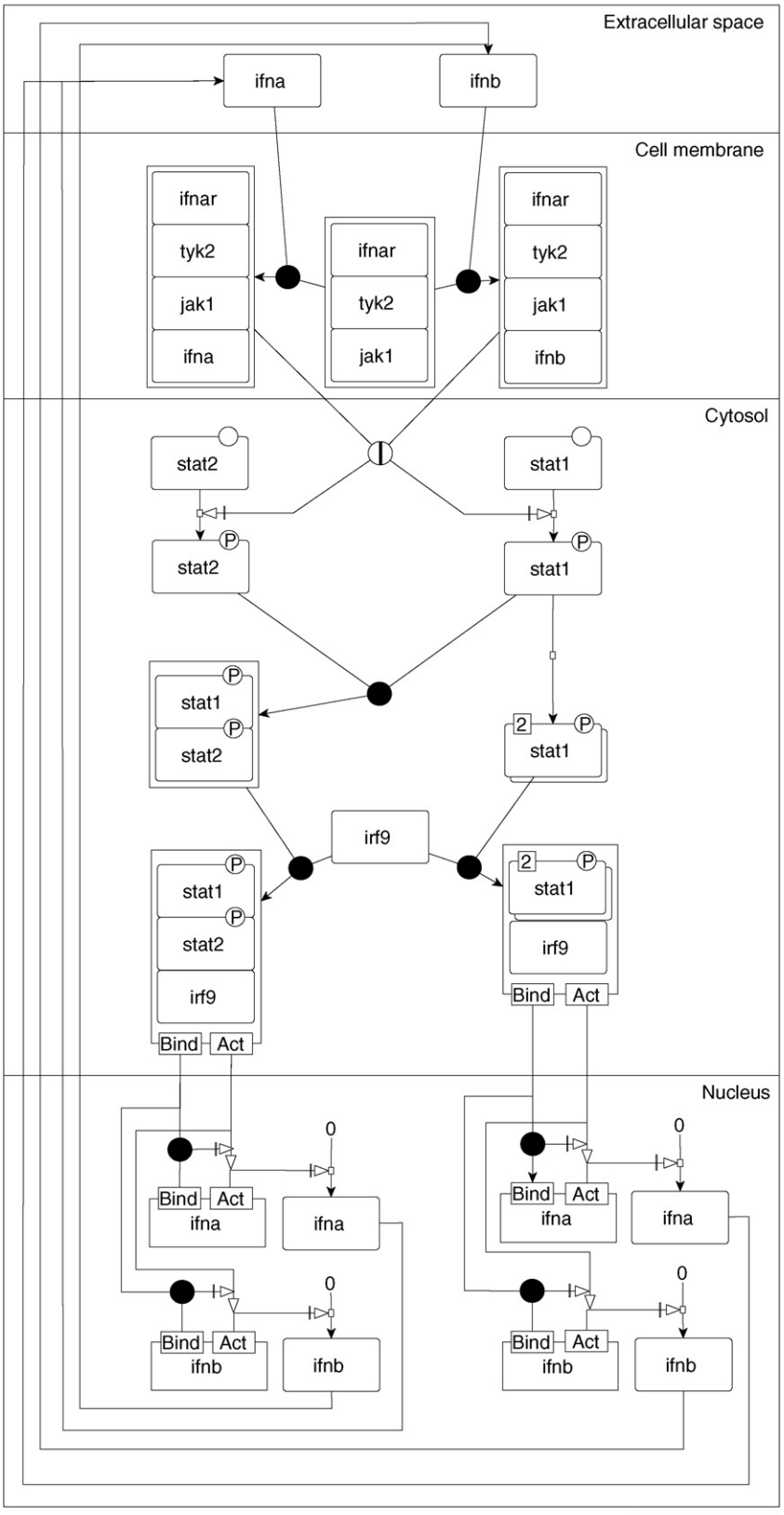

FIGURE 4

A section of the jak-stat signalling pathway drawn using SBGN notation.

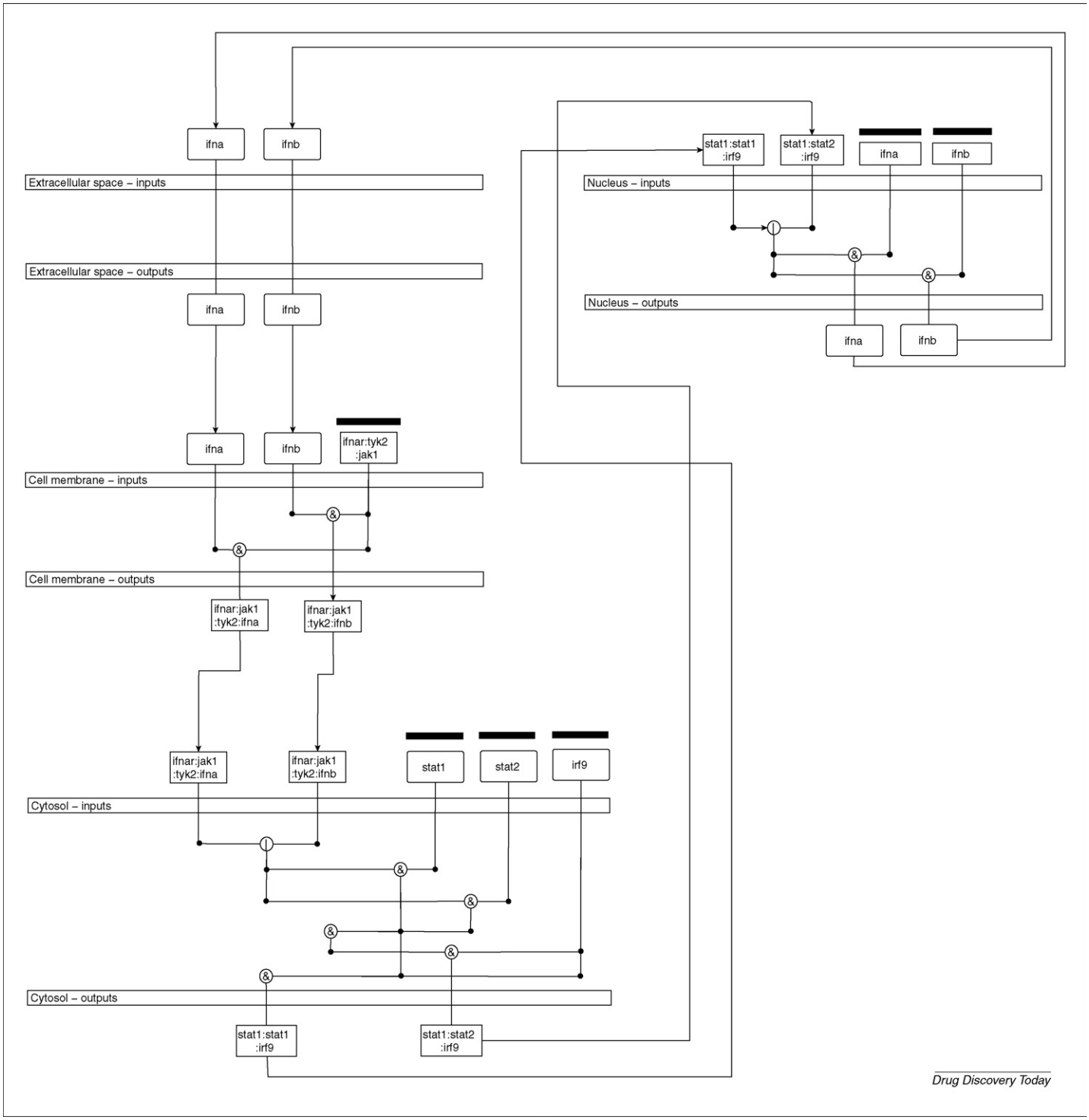

FIGURE 5

The same jak-stat signalling pathway drawn as a form of logic circuit. Each line between the inputs and outputs represents a separate interaction. Bars above (or below) a protein or complex indicate that it is latent and does not translocate to a different compartment of the cell.

to describe three patterns of cellular behaviour: apoptosis, cell proliferation and differentiation. Attractors are defined by the structure of the pathway and, by manipulating the pathway, the activity of the attractors can be controlled.

Application to drug discovery

The key advantage afforded by logic modelling is the potential for accelerating hypothesis driven research and on this basis it could change the success rate of new drugs in the market and the development of new therapeutic regimes. Indeed, industry has learnt over the years that the identification of a target for drug intervention does not necessarily equate with understanding what the target does. The lack of a more complete appreciation of the target biology has often been the underlying reason for complications resulting from adverse drug reactions.

BOX 2

Acquisition of pathway information from the literature

To systematically assemble information relating to the components and interactions of a biological pathway, we recommend a three stage process:

i Literature review. This can use standard Entrez PubMed queries involving keywords, author searches, etc. It is also possible to use more advanced search tools [51–53]. A manual review of the articles is essential to ensure relevance and accuracy. The literature selection can be classified according to species (human or mouse), cell type (macrophage or dendritic cell) and the techniques used.

ii Data-mining of public and proprietary databases. For example Pathway Analysis [54], Ingenuity Systems (http://www.ingenuity.-com), can be used to produce networks of molecular interactions that can be rated with statistical relevance scores. There are also a range of public repositories such has the Human Protein Reference Database [55] and the Biomolecular Interaction Network Database [56] which is an online resource of curated and text-mined information on protein–protein and protein–gene interactions.

iii Integration of the data captured in stages I and II into a relational database that can store and retrieve logical dependencies using their inputs and outputs.

Operationally, a drug target exists as part of one or more pathways that have evolved to process signals in multiple contexts. The architecture of these pathways can be critical to disease and drug discovery. A topical architecture is the 'bow-tie' structure in which pathways with many input and output signals pass through a bottleneck involving a smaller number of interactions [45]. In this case, the centre of the bow-tie would provide a choice pathway target for disruption by pathogens for instance, but also the area of least specificity for drug therapies.

TABLE 2

The truth table leading to activation of factor B

A	B	C	P
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	0
1	0	0	0
1	0	1	0
1	1	0	0
1	1	1	1

In light of this, the approaches taken in industry will be more considered. Instead of associating a singular target with a disease phenotype the aim would be to identify and map the pathway leading to the disease outcome. The key question will be how to inhibit a pathway rather than a target and this will mean identifying the most drug sensitive targets in that pathway. Because a specific target may play a part in many pathways, issues of sensitivity and specificity of the therapy will be based on drug strategies that intervene at multiple points on the disease pathway. Hence, if we consider a hypothetical case in which there are three druggable targets in a disease pathway, each of which also belong to other unrelated pathways, a therapy that inhibits all three targets would be far more effective in curtailing the disease pathway without disrupting other pathways than a therapy that heavily inhibits a single target.

On a disease pathway, sensitivity and specificity are related to the logical dependencies associated with its interactions. If a target was uniquely responsible for activating a protein in another unrelated pathway, the dependency between the gene and the

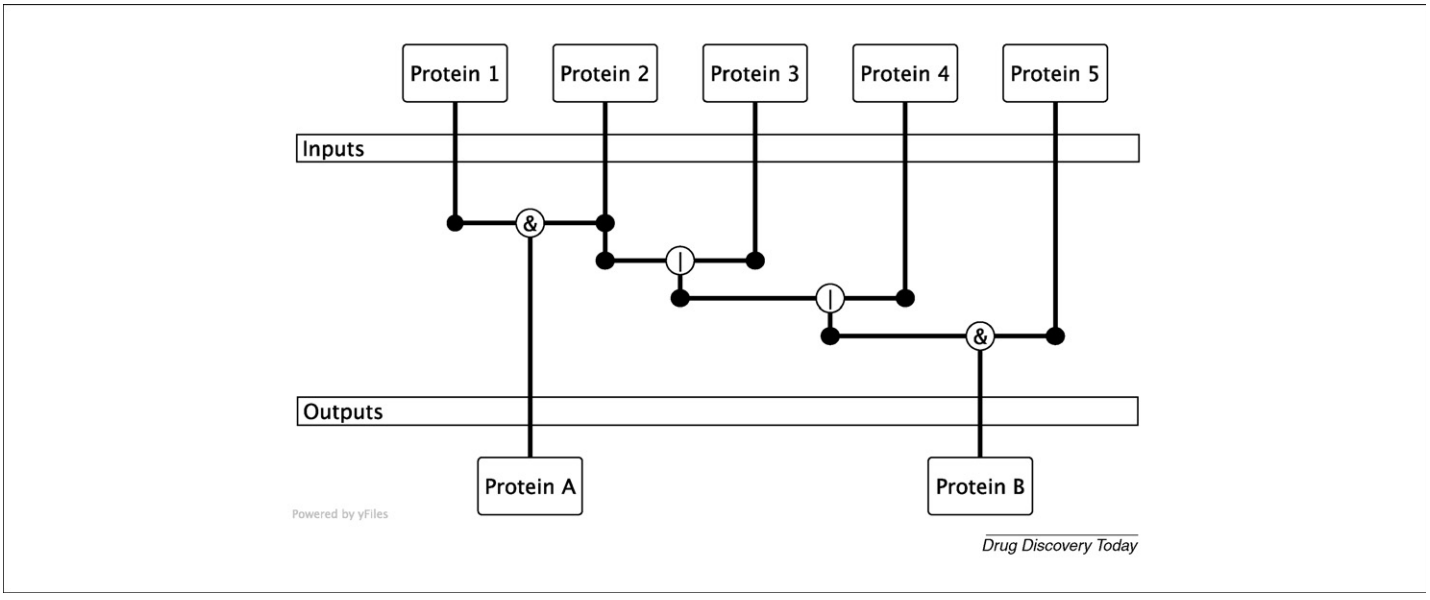

FIGURE 6

Target protein 2 is required for the activation of protein A, but protein 2 is also one of three proteins responsible for the activation of protein B. In this case, protein 2 has a high specificity for the pathway that produces protein A, but a low specificity to the pathway that produces protein B owing to the number of OR dependencies.

target would be AND and the target would have a low specificity to the disease pathway. However, if the target was one of many activators for the protein in the other pathway, the interaction would have the dependencies of equivalent binding (which would be described using OR) and the target would have a higher specificity to the disease pathway. In fact, the more activators that compete with the target (acting as inputs to the OR dependencies), the higher the specificity we would expect to find in a therapy that inhibits that target. This is illustrated in Figure 6.

Future perspectives

Logic modelling can be a powerful tool to help expand our understanding of cellular pathway behaviour. However, it also has great potential outside of the laboratory. A better understanding of pathway structure will position us to translate pathways into representative algorithms, allowing much *in vivo* and *in vitro* work to be replaced with computational studies. It may also be possible to turn the relationship between pathway signalling and logic on its head to allow us to use cell signalling to study logic problems. Claude Shannon did exactly this when he realized that analogue phone systems could be used to study logic problems. Could we use the signalling pathways available across a range of cell types to perform logic calculations?

Pathways have already been used to perform mechanical and computational tasks. The iGem project (http://openwetware.org/wiki/IGEM) is a competition organised by MIT, in which undergraduate teams use a catalogue of *biobricks* (http://parts.mit.edu) to build molecular machines to solve problems of their own choosing [46,47]. The winning team in 2006 produced bacteria to inhibit sepsis in mammalian cells (http://www.igem2006.com).

Conclusion

With computational resources beginning to be introduced into biology, computationally sound modelling schemes will become increasingly important. The key advantages that logic holds as a modelling scheme is that it is conceptually simple, computationally cheap and causally correct. For studying large pathway networks and doing sweeps through pathway configurations, these advantages are key. As such, it is well suited to the searches necessary to develop drug therapies.

Acknowledgements

This work was supported in part by EU grant InfoBioMed (contract number 507585), the MRC and Scottish Enterprise. The Centre for Systems Biology at Edinburgh is a Centre for Integrative Systems Biology supported by the BBSRC and EPSRC.

References

1 Barabasi, A. and Oltvai, Z. (2004) Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113
2 Kitano, H. (2002) Computational systems biology. *Nature* 420, 206–210
3 Kitano, H. (2002) Systems biology: a brief overview. *Science* 205, 1662–1664
4 Kell, D. (2006) Systems biology, metabolic modelling and metabolomics in drug discovery and development. *Drug Discov. Today* 11, 1085–1092
5 Fell, D. (1992) Metabolic control analysis: a survey of its theoretical and experimental development. *Biochem. J.* 286, 313–330
6 Hongwu, M. *et al.* (2007) The Edinburgh human metabolic network reconstruction and its functional analysis. *Mol. Syst. Biol.* 3, 135
7 Ratushnyi, A. *et al.* (2003) A computer model of the gene network of the cholesterol biosynthesis regulation in the cell: analysis of the effect of mutations. *Doklady Biochem. Biophys.* 389, 90–93
8 McAdams, H. and Arkin, A. (1997) Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 94, 814–819
9 Arkin, A. *et al.* (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-Infected *Escherichia coli* cells. *Genetics* 149, 1633–1648
10 Hofestadt, R. and Thelen, S. (1998) Quantitative modelling of biochemical networks. *In Silico Biol.* 1, 39–53
11 Kuffner, R. *et al.* (2000) Pathway analysis in metabolic databases via differential metabolic display (DMD). *Bioinformatics* 16, 825–836
12 Pinney, J. *et al.* (2003) Petri net representations in systems biology. *Biochem. Soc. T.* 31, 1513–1515
13 Goss, P. and Peccoud, J. (1998) Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. *Proc. Natl. Acad. Sci. U.S.A.* 95, 6750–6755
14 Matsuno, H. *et al.* (2000) Hybrid Petri net representation of gene regulatory network. In *Pacific Symposium on Biocomputing 2000: Honolulu, Hawaii, U.S.A., 4–9 January 2000* (Altman, R. *et al.* eds), pp. 338–349, World Scientific Press
15 Phillips, A. and Cardellie, L. (2006) A correct abstract machine for the stochastic pi-calculus. In *Transactions on Computational Systems Biology VII* (Priami, C., ed.),Springer
16 Regev, A. and Shapiro, E. (2004) The pi-calculus as an abstraction for biomolecular systems. In *Modelling in Molecular Biology* (Ciobanu, G. and Rozenberg, G., eds), pp. 219–266, Springer
17 Kaufman, M. *et al.* (1999) A logical analysis of T cell activation and anergy. *Proc. Natl. Acad. Sci. U.S.A.* 96, 3894–3899
18 Mendoza, L. *et al.* (1999) Genetic control of flower morphogenesis in Arabidopsis thaliana: a logical analysis. *Bioinformatics* 15, 593–606
19 Espinosa-Soto, C. *et al.* (2004) A gene regulatory network model for cell-fate determination during *Arabidopsis thaliana* flower development that is robust and recovers experimental gene expression profiles. *Plant Cell* 16, 2923–2939
20 Albert, R. and Othmer, H. (2003) The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster*. *J. Theor. Biol.* 223, 1–18
21 Kauffman, S. (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.* 22, 437–467
22 King, R. *et al.* (2004) Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature* 427, 247–252
23 Tamaddoni-Nezhad, A. *et al.* (2006) Application of abductive ILP to learning metabolic network inhibition from temporal data. *Mach. Learn.* 64, 209–230
24 Glass, L. and Kauffman, S. (1973) The logical analysis of continuous, non-linear biochemical control networks. *J. Theor. Biol.* 39, 103–129
25 Kitano, H. *et al.* (2005) Using process diagrams for the graphical representation of biological networks. *Nat. Biotechnol.* 23, 961–966
26 Kohn, K. *et al.* (2006) Molecular interaction maps of bioregulatory networks: a general rubric for systems biology. *Mol. Biol. Cell* 17, 1–13
27 Demir, E. *et al.* (2002) Patika: an integrated visual environment for collaborative construction and analysis of cellular pathways. *Bioinformatics* 18, 996–1003
28 Moodie, S. *et al.* (2006) A graphical notation to describe the logical interactions of biological pathways. *J. Integr. Bioinform.* 3, 36
29 Systems Biology Graphical Notation Level 1 Specification (see: http://www.sbgn.org)
30 Shmulevich, *et al.* (2002) Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics* 18, 261–274
31 Shmulevich, *et al.* (2002) From Boolean to probabilistic Boolean networks as models of genetic regulatory networks. *Proc. IEEE* 90, 1778–1792
32 Yu, L. and Marshall, S. (2007) Inferring context-sensitive probabilistic Boolean networks from gene expression data under multi-biological conditions. *BMC Syst. Biol.* 1, 63
33 Ching, W. *et al.* (2007) An approximation method for solving the steady-state probability distribution of probabilistic Boolean networks. *Bioinformatics* 23, 1511–1518
34 Liang, S. *et al.* (1998) REVEAL: a general reverse engineering algorithm for inference of genetic network architectures. In *Biocomputing' 98: Proceedings of the Pacific Symposium* (Altman, R., ed.), pp. 18–29, World Scientific Press
35 Akutsu, T. *et al.* (1999) Identification of genetic networks from a small number of gene expression patterns under the Boolean network model. In *Proceedings of the 4th Pacific Symposium on Biocomputing* (Altman, R., ed.), pp. 17–28, World Scientific Press

36 Akutsu, T. *et al.* (1998) Identification of gene regulatory networks by strategic gene disruptions and gene overexpressions. In *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms* (Karloff, H., ed.), pp. 695–702, Society for Industrial and Applied Mathematics

37 Akutsu, T. (2000) Inferring qualitative relations in genetic networks and metabolic pathways. *Bioinformatics* 16, 727–734

38 Husmeier, D. (2003) Reverse engineering of genetic networks with Bayesian networks. *Biochem. Soc. Trans.* 31, 1516–1518

39 Husmeier, D. (2003) Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. *Bioinformatics* 19, 2271–2282

40 Werhli, A. and Husmeier, D. (2007) Reconstructing gene regulatory networks with Bayesian networks by combining expression data with multiple sources of prior knowledge. *Stat. Appl. Genet. Mol. Biol.* 6, 15

41 Laubenbacher, R. and Sigler, B. (2004) A computational algebra approach to the reverse engineering of gene regulatory networks. *J. Theor. Biol.* 229, 523–537

42 Huang, S. *et al.* (2005) Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.* 94, 128701

43 Huang, S. (1999) Gene expression profiling, genetic networks and cellular states: an integrating concept for tumorigenesis and drug discovery. *J. Mol. Med. – JMM* 77, 469–480

44 Wuensche, A. (1998) Genomic regulation modeled as a network with basins of attraction. In *Proceedings of the 3rd Pacific Symposium on Biocomputing* (Altman, R., ed.), pp. 89–102, World Scientific Press

45 Csete, M. and Doyle, J. (2004) Bow ties, metabolism and disease. *Trends Biotechnol.* 22, 446–450

46 Check, E. (2005) Synthetic biology: designs on life. *Nature* 438, 417–418

47 Sander, *et al.* (2007) Engineering a synthetic molecular oscillator based on the Lotka-Volterra dynamic. *BMC Syst. Biol.* 1, 59

48 Boole, G. (1847) Mathematical Analysis of Logic. Being an Essay Towards a Calculus of Deductive Reasoning. Reprinted: Blackwell, Oxford, 1948

49 Boole, G. (1854) An Investigation of the Laws of Thought, on Which are Founded the Mathematical Theories of Logic and Probabilities. Reprinted: Dover Publications, New York, 1958

50 Shannon, C. (1938) A symbolic analysis of relay and switching circuits. *Trans. A.I.E.E.* 57, 713–723

51 Grimes, G. *et al.* (2006) PDQ wizard: automated prioritization and characterization of gene and protein lists using biomedichal literature. *Bioinformatics* 22, 2055–2057

52 Chen, H. and Sharp, B. (2004) Content-rich biological networks constructed by mining PubMed abstracts. *BMC Bioinform.* 5, 147

53 Becker, K. *et al.* (2003) PubMatrix: a tool for multiplex literature mining. *BMC Bioinform.* 4, 61

54 Nikitin, A. *et al.* (2003) Pathway Studio – the analysis and navigation of molecular networks. *Bioinformatics* 19, 2155–2157

55 Peri, S. *et al.* (2003) Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res.* 13, 2363–2371

56 Bader, G. *et al.* (2001) BIND – The biomolecular interaction network database. *Nucleic Acids Res.* 29, 242–245