THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Use of logic theory in understanding regulatory pathway signaling in response to infection

OPEN ACCESS

Use of logic theory in understanding regulatory pathway signaling in response to infection

Steven Watterson & Peter Ghazal[†]

[†]Author for correspondence: Division of Pathway Medicine, University of Edinburgh Medical School, Chancellor's Building, 49 Little France Crescent, Edinburgh, EH16 4SB, Scotland, UK ■ Tel.: +44 131 242 6242 ■ Fax: +44 131 242 6244 ■ p.ghazal@ed.ac.uk

Biological pathways link the molecular and cellular levels of biological activity and perform complex information processing seamlessly. Systems biology aims to combine an understanding of the cause–effect relationships of each individual interaction to build an understanding of the function of whole pathways. Therapies that target the 'host' biological processes in infectious diseases are often limited to the use of vaccines and biologics rather than small molecules. The development of host drug targets for small molecules is constrained by a limited knowledge of the underlying role of each target, particularly its potential to cause harmful side effects after targeting. By considering the combinatorial complexity of pathways from the outset, we can develop modeling tools that are better suited to analyzing large pathways, enabling us to identify new causal relationships. This could lead to new drug target strategies that beneficially disrupt host–pathogen interactions, minimizing the number of side effects. We introduce logic theory as part of a pathway modeling approach that can provide a new framework for understanding pathways and refine 'host-based' drug target identification strategies.

With the increasing availability of high through-put technologies has come a richness of data regarding the inner workings of the cell. This data has brought us closer than ever to quantifiable analyses of cell behavior and teases us with the prospect that we may one day be able to understand, comprehensively and exactly, certain cell functions or even a cell's whole function. Presently, we are only beginning to learn how cells operate and the key to the control of their behavior is the network of signaling pathways that mediate a cells response to internal and external stimuli.

In most areas, our knowledge of pathway composition and structure is provisional and our understanding is continually undergoing revision and refinement. However, composition and structure are critical to pathway function and so by using our understanding to make quantitative predictions about pathway behavior and comparing these to observations *in vitro*, we can validate our understanding and infer improvements. Naturally, the improvements lead to new predictions and so our understanding is iteratively refined.

A detailed understanding of pathways presents great opportunities for the development of therapeutic strategies. A detailed understanding will allow us to devise more sophisticated and subtle controls over cell function that maximize efficacy and minimize side effects. This will enable us to disrupt pro-pathogen pathways and augment protective pathways in new ways. However, this presupposes that we can identify the appropriate methods for making predictions of pathway behavior using our current state of knowledge.

Accurate pathway models are required in order to facilitate predictions and building these models is an ongoing challenge. Accurate pathway models need appropriate modeling methodologies and it is not yet clear how best to apply these methodologies. Several modeling methodologies have been proposed. Ordinary differential equations (ODEs) have proven to be very successful in modeling metabolic pathways [1–4]. However, they generally require a relatively large set of parameters and it is often not clear *a priori* which parameters require precision and which can be obtained more coarsely with little impact. Obtaining realistic *in vitro* or *in vivo* values for parameters can often be difficult owing to the specialist experiments required.

Alternatives include the stochastic schemes that accurately capture the microscopic fluctuations that accompany small populations of events. In the context of signaling pathways, these schemes are well suited to the behavior of small populations of interacting proteins and molecules [5,6]. There are comparable issues surrounding parameter values for stochastic schemes, but their most significant limitation

future medicine
part of fsg

comes with large populations. As larger population sizes are considered, the computational demands of stochastic methods become inhibitively large. However, in these cases, the behavior will often start to describe results similar to those of ODEs and so computationally less demanding ODE methods can often be employed.

Petri nets [7–9], process algebras [10,11] and a range of grammatical languages [12,13] have been proposed as vehicles in which to formally describe the range of interactions that can occur in pathway systems. Their strength lies in their ability to describe pathway systems modularly and to prove what behavior is and isn't possible. However, these methods still require a quantitative methodology of the type described above in order to make quantitative predictions. As such, they remain a tool of the future.

The value of any model is limited by the availability of accurate, high-confidence parameters. The fewer the number of parameters a model requires, the more valuable a model is likely to become. However, this comes with a trade-off. Generally speaking, more parameters allow a model to be more flexible and to predict behavior in greater detail. A model with fewer parameters can only make coarser predictions.

Nonetheless, a model and a modeling methodology that is simpler and requires fewer parameters will be experimentally more tractable and computationally less demanding, so if we can accept the associated coarse-grained predictions, it can offer significant advantages.

The simplest modeling scheme imaginable is one in which the activity level of the components on each pathway (e.g., the genes, proteins and complexes) are described in only one of two states: active or inactive. Each component is then represented by a two state variable describing its activity. We can then write the interactions between components by introducing logical dependencies between them. Such a scheme is best suited to pathways in which the response by each component is clear, unambiguous and absolute, rather than by degree. The low level of detail means that where a small change in activity level indicates a response, this will be missed by the scheme, so the signaling response must be dramatic and dominant over other signaling influences.

A modeling scheme that is well suited to this application is Boolean logic [14–16]. This has been discussed previously as a modeling methodology in a variety of forms and is of increasing interest [17–25]. Here we review it's application to signaling pathways and discuss what it can tell us about the host–pathogen immune response, as an example of a critical, combinatorially complex signaling system.

This paper is organized as follows. In the next section we discuss how best to assemble a model from established resources, such as the published literature. In the following section we discuss modeling pathway systems. This section does not explore any particular modeling scheme in depth, but reviews some of the general considerations in modeling pathways. Next, we introduce logic and demonstrate how it can be applied to a pathway system, then we apply these approaches to a section of the Jak–Stat signaling pathway. The final section is our conclusion.

Building pathway models

Before we can begin to model, we must first assemble a description of the pathway from previous research findings. This in itself is nontrivial. We must compile, integrate and visualize the components and interactions along pathway using a standardized synthesis methodology [26]. This generally follows a four-stage process:

- A literature review identifies the relevant pathway components and interactions from peer reviewed publications. This can be performed using standard Entrez PubMed queries involving keywords, author searches and so on. A variety of tools can also be used to facilitate this process, such as PDQ wizard [27]. A manual review of the resultant articles is essential to ensure the relevance and accuracy of the results. To build confidence, the obtained literature set should include at least two, ideally three, independent reports corroborating the functional interaction.

- Appending to the literature review using a data mining approach. Available resources include KEGG [101], HPRD [102], Chilibot [103] and Ingenuity Pathway Analysis [104]. These resources can be used to consolidate the literature-derived interactions and identify new components or interactions. They utilize databases of curated data obtained from further text mining and experiments.

- Graphical presentation of the resulting interactions. This provides an intuitive, if not always unambiguous presentation of the results. This can be achieved with a variety of packages, including yEd [105], EPE [28], Cell Designer [29], Copasi [30] and SimBiology [106] and can be done using the newly introduced, community-driven systems biology graphical notation [31]

or one of its forerunners [32–34]. Due to the fashion in which they are compiled, the resulting diagrams represent a consensus view derived from the community's published results and not necessarily a canonical pathway that has been thoroughly validated.

■ A database of the resulting pathways. This facilitates the interrogation and sharing of pathway data. Several approaches to the storage and review of network interaction data have been published [35–37], but this area is still in its infancy and will likely see further development in the future. One of the most recent developments in this area has been the emergence of the community-driven Wikipathways wiki, which has been established for open curation by the biology community [38].

Several recent studies have applied this research synthesis approach to signaling pathways in macrophage biology [34,39,40].

There is a degree of *caveat emptor* to pathway information assembled from the established literature. Such pathways assume that the current state of the literature reflects an adequate level of understanding of the biology for the purpose in hand. However, it is not always the case that the published data truthfully reflects the underlying biology and it is prone to change. Accordingly, it is important to consider how best to independently validate the pathway constructed for the system under investigation. By combining pathway modeling and hypothesis generation with experimental tools such as RNAi, we can use the correlation between experimental and computational results as a means to validate the obtained pathway information.

Quantitative modeling

Graphical representations of pathways can be translated into the system of equations needed to simulate the pathway. Since this process is laborious, the graphing packages mentioned above include features that automate this process to produce ODE systems. A standardized file format has been established in a community-driven effort for exchanging such ODE models. The systems biology markup language [41] uses the XML file formatting system and has been adopted widely in new software tools, online repositories and the supplementary online material of publications.

The validation of the pathway is achieved by comparing the predictions from the systems of equations to the equivalent data. However, the system of equations can only relate the output of the pathway to the inputs to the

pathway. Consequently, the equations require the input data in order to make a prediction of the output.

Comparisons between the prediction and the equivalent data are rarely exact as both the equivalent data, and the data used in the inputs to the pathway contain some variation due to noise. As a result, the comparison must take the form of a statistical correlation.

The true advantage of building a quantitative model (often referred to as an *in silico* model) comes once confidence has been established. A high confidence quantitative model allows us to explore 'what if' scenarios without being limited by the financial and ethical constraints of experiment. It also allows us to explore considerably more scenarios than an experiment would permit, for the same investment of time and money.

The workflow describing how pathway models can be used to refine our understanding is shown in Figure 1.

Complexity

Complexity is something of an umbrella term that groups together the many factors that make analysis of quantitative models difficult and time consuming. More sophisticated analyses require more computing time than simpler analyses and so the level of sophistication of the analysis is often limited by the computing resources available.

Generally, there are two factors that contribute to the computing time required for analysis. The first is the efficiency with which a pathway can be simulated. ODE representations are simulated
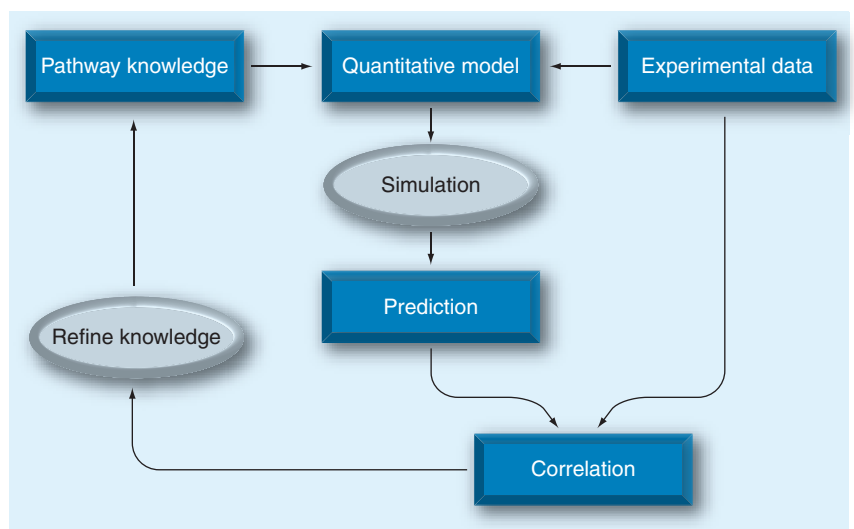

Figure 1. A workflow describing how quantitative models can contribute to refining our understanding of pathways.

by numerical integration [42], stochastic systems are simulated using a variant of the Gillespie algorithm [43] and logic systems are simulated using sequences of update steps that can be either synchronized or unsynchronized [18]. By choosing the most efficient methodology and simulation method, the time taken by each simulation of the pathway can be kept to a minimum.

The second factor arises when the simulation must be run multiple times. This is the case when there are uncertain parameters in the quantitative model. A typical strategy to identify the value of an unknown parameter is to experimentally measure an initial state of the pathway and a final state some time later. Several representative values for the unknown parameter are then taken from across a likely range and the pathway is simulated multiple times, once for each value, using the same initial state. The simulation whose final state best matches the experimentally observed final state provides an indicator of the likely value for the parameter. This is a fundamental strategy in parameter estimation. However, because it requires multiple simulations, it can be very time consuming. When there are multiple unknown parameters, all permutations of the representative values of each parameter must be considered and the number of permutations grows exponentially in the number of parameters. Thus for N parameters each with R representative values, the number of simulations required is of the order of R^N.

Another common challenge is to determine which initial states of the pathway lead to a known final state. Here, the levels of the inputs to the pathway become the unknowns and the strategy is to choose representative values for each input, generate the necessary permutations of parameter values and simulate the pathway, taking each permutation as an initial state. By comparing the known final state to the states generated by each simulation, we can determine which initial state generated the known final state. As in the previous example, the number of simulations required grows exponentially with the number of unknown inputs. Thus for M inputs each with V representative values, the number of simulations required is of the order of V^M.

Both of these scenarios are examples of combinatorial complexity, which generally indicates an exponential growth in the number of permutations to be considered, when there are multiple unknowns.

Any modeling methodology that can describe the behavior of interest, but that reduces the number of parameters and the number of representative values to be considered will facilitate a quicker analysis of the pathway. This, in turn, allows us to use the same computing time to explore larger pathways and to perform deeper analyses.

Modularity

By exploiting modularity in the pathway, we can reduce the number of simulations to be considered. A module is a section of pathway that takes one or more input signals and produces one or more output signals through a self-contained process that can be arbitrarily complex. In a modular pathway, it is often enough to consider only the signals that are passed between modules to study the pathway's behavior, without necessarily needing to analyze the internal functioning of each module. This strategy can reduce the number of parameters or the number of initial states to be considered.

Modularity is a feature that has been highly sought in pathway biology, partly because of the analytical simplicity it brings and partly because it may hint at underlying organizational principles in pathway biology. At the level of cell signaling, it is plausible that pathways may be arranged in a modular fashion to avoid unnecessary cross-talk and to maximize their independence of function. For both functional and evolutionary reasons, this may, in turn, lead to improved levels of flexibility (adaptability) and robustness.

The innate immune response demonstrates a hierarchy of connected responses that hints at a modular control mechanism. The primary level response is mediated by innate and adaptive immunity and addresses infection. The secondary level response includes cell-proliferation modulation and apoptosis triggering along with inter-cell signaling [44]. The second level response involves significant interplay between the host and the pathogen as each try to control the other's function.

At this level, we see another example of modular signaling in the interplay between the host and pathogen. Both have complex internal networks of signaling pathways that regulate their function. Both also have pathways that manipulate and exploit the others function. We can regard each as a module and the pathways between as inter-modular signaling. As such, the pathogen represents a modular plug-in to the host's pathway network.

For the purposes of developing logic as a modeling methodology, we shall focus on the host-cell regulatory pathways that dominate the second level of host regulatory control. The pathogen suppresses the immune response

using these pathways and exploits the cellular and metabolic processes of the host for its own survival [45]. They therefore serve as potential targets for drug intervention.

Homeostasis

A further feature that gives us insight into the function of pathways is homeostasis, in which a pathway functions at a steady, equilibrium level. The principal of homeostasis is key to determining the pathway function in chronic or persistent infections [44]. System–wide, this is achieved through the constant and dynamic adjustment of regulatory interaction pathways between the host and pathogen [46]. However, at a cellular level this is achieved by ensuring that key regulatory pathways sustain a fixed level of infection activity.

Exploiting homeostasis can be a valuable approach to studying pathways with feedback. When simulated for an adequate length of time, pathways with feedback invariably enter into either a single state or a repeated cycle of states, neither of which they can escape without a change in external signals. Both are known as attractors (or limit cycles). One pathway may have multiple attractors and many initial states are likely to lead to each attractor.

It has been speculated that the range of attractors that can be sustained by cellular pathways might explain the process through which cells differentiate into different cell phenotypes [47]. Individual pathways or pathway systems may also demonstrate a range of behaviors due to their attractor structure.

Logic modeling

Signals can be propagated along pathways by both subtle and coarse variations in the activity of pathway components. Whether subtle variations represent significant signaling is dependent, amongst other factors, on the sensitivity of the subsequent downstream interactions. In systems in which the sensitivity is low and the signaling behavior is dominated by coarse, dramatic variations in activity, we can approximate the signaling activity of the pathway components using one of two values: active or inactive. This dramatically reduces the number of permutations that must be considered when we wish to determine which initial states of a pathway lead to which final states and therefore achieves greater computational efficiency.

As part of this scenario, we can also reduce the number of parameters required by assuming that all interactions take place reliably and completely. This eliminates the parameters that describe the strength of an interaction. Besides reducing the computational demands of searching through parameter values, this simplification makes validation more experimentally tractable.

This scenario lends itself very well to the form of propositional logic known as Boolean logic that is widely used in computing [16]. By using Boolean variables (which are two state variables with values ON or OFF) to describe the activity of components and the logical dependencies, AND, OR and NOT to describe the interactions between pathway components, we can study pathway biology in a manner consistent the computational sciences.

Amongst the many signaling interactions that appear on pathways, six common interaction types are binding, inhibition, complex formation, equivalent binding, dissociation and phosphorylation. Here, we review the logic description of each, using A, B and C to denote proteins, D to denote a complex, G to denote a gene and A^P to denote a phosphorylated state of A. In a Boolean logic description, A, B, C, D and G are all two-state variables.

- Binding: suppose transcription factor A binds to gene G which leads to the transcription and translation of protein C. We require both A and G to be in an active state for C to be produced. Thus, C = A AND G;

- Inhibition: suppose transcription factor A binds to gene G which leads to the transcription and translation of protein C and that this is inhibited by protein B. We require A and G to be active for C to be produced and B to be absent. Thus, C = (A AND NOT B) AND G;

- Complex formation: suppose protein A and protein B, bind to form complex D. Both A and B must be active for D to be produced. Thus, D = A AND B;

- Equivalent binding: suppose both A and B bind to G and that only one need bind in order for transcription and translation of protein C to occur. We require at least one of A and B to be active together with G. Thus, C = (A OR B) AND G;

- Dissociation: suppose the complex D dissociates into the proteins A and B. We require D to be active for A and B to be produced. Thus, we have A = D and B = D;

- Phosphorylation: suppose comple D phosphorylates protein A. We require D and A to be active for A^P to be produced. Thus, we have A^P = A AND D.

We can combine the dependencies of individual interactions to determine the dependencies of larger pathways. For example, suppose transcription factor A binds to gene *G1*, producing protein C, and that protein A itself is the result of binding between transcription factor B and gene *G2*. This gives us two interactions: C = A AND *G1* and A = B AND *G2*. We can substitute for A in the former to give C = (B AND *G2*) AND *G1*.

Logic implementation of pathways

The key to using logic to analyze pathway behavior is the mechanism by which we relate experimental data to the two-state variables of Boolean logic. In order to convert experimental data to a form suitable for use in a logic description, we must discretize the data.

Discretization requires the introduction of a threshold and we generate two state activity levels by setting all experimental values above the threshold to the active state and all experimental values below the threshold to the inactive state. The proportion of experimental values that fall exactly on threshold will be small and any errors that arise from misclassifying these points are likely to be small for sufficiently long time courses.

In Figure 2, we demonstrate how continuous experimental data can be discretized and that this retains the coarse trends of the data. The continuous line represents expression data obtained from microarray experiment (in this case the TNF transcript level obtained when bone marrow derived macrophages were infected with murine cytomegalovirus). The dotted line represents the discretization threshold and the dashed line the discrete expression levels obtained from the discretization process.
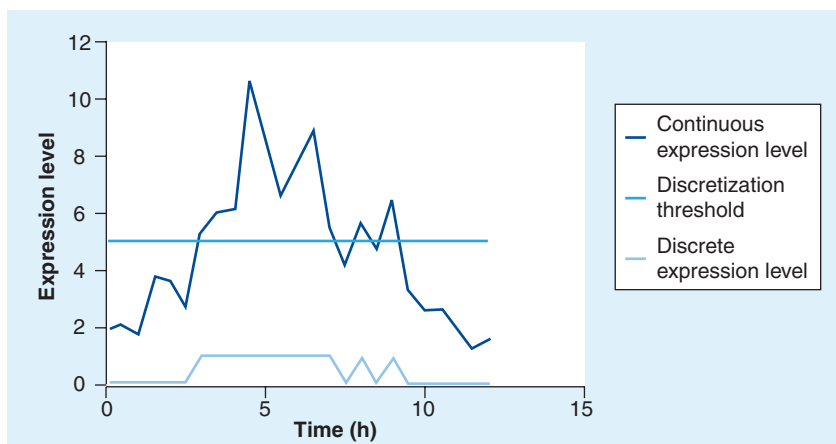

Figure 2. The discretization of continuous data. Here we show TNF expression levels recorded by microarray during infection of bone marrow derived macrophage cells by cytomegalovirus.

Logic diagrams

In order to communicate the structure and dependencies of pathways, a graphical notation system that captures the logic dependencies of the interactions without retaining the biochemical details must be introduced, analogous to that used in electrical engineering. Here, we expand on a notational system that we have developed for pathway systems [21].

In order to describe large networks of signaling pathways, it is necessary to consider the network in sections. The choice of sections is largely arbitrary, but a good choice will allow us to study how the signals propagate through the system and identify whether there is feedback. We can segment many pathways using the compartments of the cell, for example. Each section has input signals and output signals and some of these signals are communicated to other sections (which we denote migrant signals) and others are not (latent signals). Latent outputs are important to capture because signals may be passed on to pathways that are not of interest in the current study and that we would want to truncate. Latent inputs are likely to arise because their signals derive either from pathways that are not of interest in the current study or from pathways that are unknown. Migrant inputs receive a signal from another section and migrant outputs pass a signal onto a different section. To distinguish between migrant and latent inputs, we place a bar above the latent input indicating that it should not receive a signal. To distinguish between latent and migrant outputs, we place a bar below the latent output indicating that no output signal can leave.

Figure 3 shows a hypothetical cell compartment in which there are no interactions. Proteins A and B pass through this compartment without interacting. They have migrant inputs and outputs. The origin of protein C has not been deemed relevant, but is involved in downstream interactions and so has a latent input. Protein D propagates into the compartment, but triggers subsequent pathways that are not of interest and so are truncated. Protein D has a latent output.

In order to distinguish between proteins, complexes and genes, we use rectangles with rounded corners to describe proteins and rectangles with square corners to describe complexes.

In order to capture the logic of the pathways, we use the logic operators described above. Here we denote AND, OR and NOT using the symbols '&', '|' and '!', respectively and we enclose these symbols in circular nodes. AND and OR both take two inputs and produce one output. NOT takes one input and produces one output.
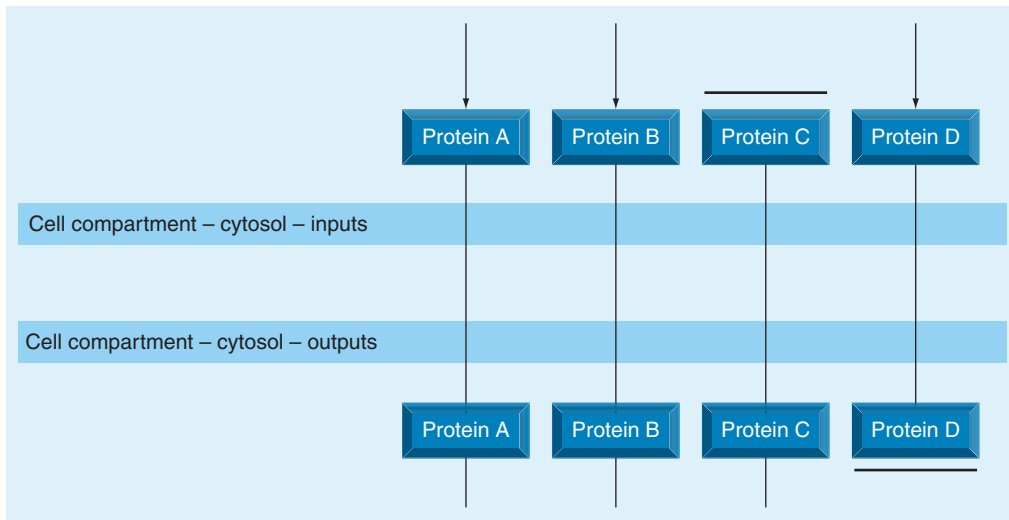
Figure 3. Proteins translocating into and out of the cytosol. Protein C is a latent input. Protein D is a latent output.

The inputs are generally related to the outputs by the logic operators. For example, suppose we have a hypothetical pathway in which, within one compartment, proteins A and B bind to form the complex A:B and proteins C and D bind to form the complex C:D. Suppose also that a fifth protein, E, binds to either complex before translocating to a different compartment. We would describe this as shown in Figure 4.

One input can relate to several outputs and, for large networks of signaling pathways, this can lead to overlapping lines. A small black circle is used to distinguish between meaningful junctions between lines, where a signal is propagated, and overlapping lines where a signal is not propagated (see Figure 5 for an example).

In pathways where transcriptional capacity is itself variable, for example in systems in which genes may be knocked down, we want to include genes as two state signals that may or may not be active. We include genes as grey rounded rectangles, similar in shape to the proteins. A transcriptional event, in which transcription factor A activates gene *B* to produce protein B, is shown in Figure 6. Genes do not carry signals from other cell compartments and so, by definition, are latent inputs to a section.
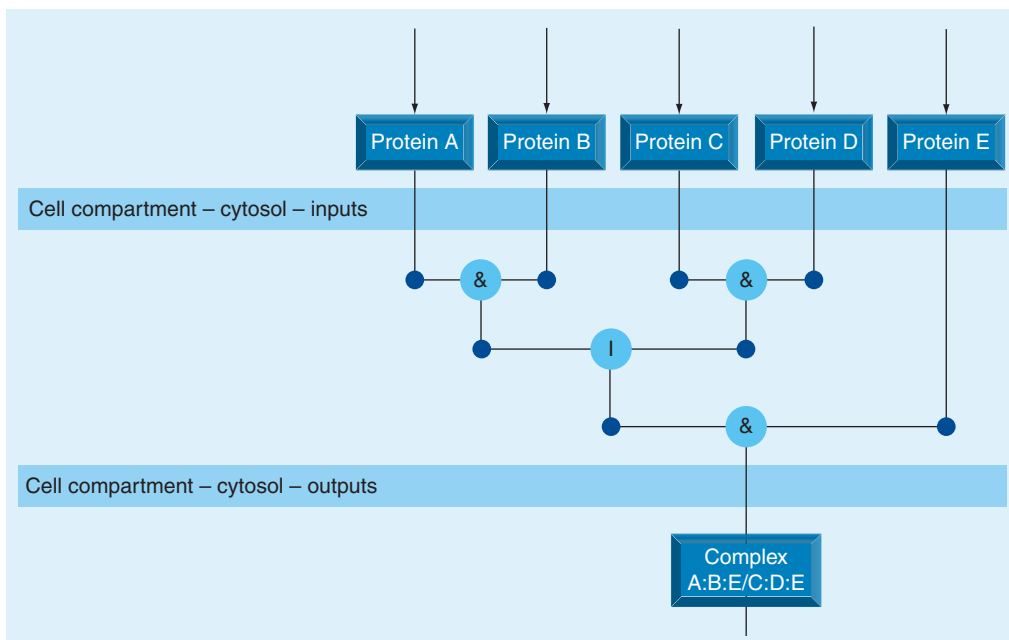

Figure 4. A hypothetical pathway through the cytosol in which either complex [A:B] or complex [C:D] can bind with protein E before translocating out of the cytosol.
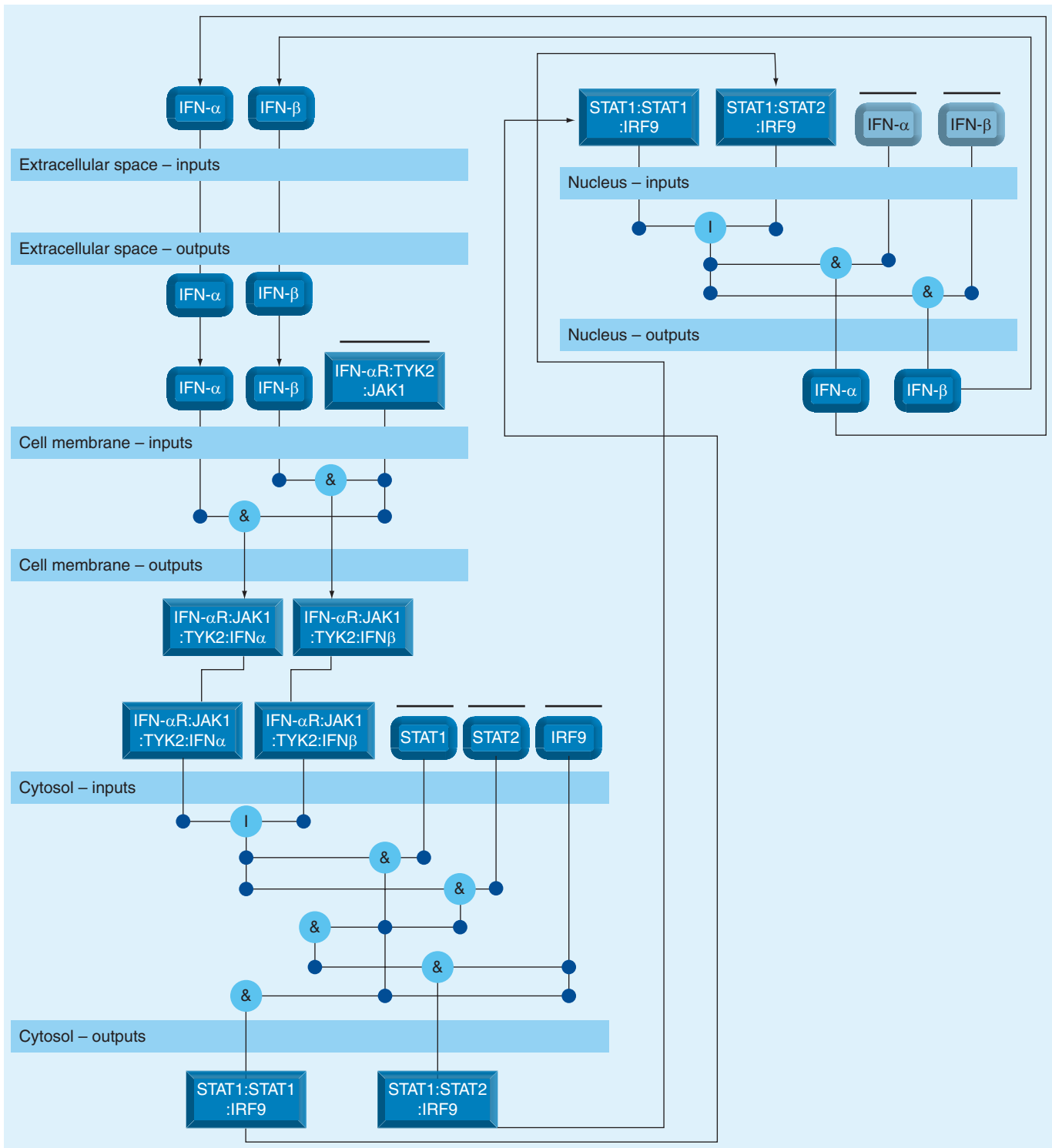
Figure 5. The logic representation of a section of the Jak–Stat signaling pathway.

A key feature of this type of notation is that new pathways can be added to existing models in an orderly and neat fashion by adding new lines between the inputs and outputs.

A modification of this notation allows us to present the pathways in a modular and more compact fashion. Each output from a compartment draws upon multiple input signals and so we can represent a pathway at a level that describes just the inputs and the output. Each of these relations would only require one line and so the interaction described in Figure 4 would be simplified to the form shown in Figure 7. Such a representation would not

allow us to make predictions of the output, but may be useful for succinctly summarizing a pathway.

Several freely available software platforms facilitate the construction and analysis of logic pathways. The notation we have described here can be built in free-form schematic software, such as yEd. Electrical engineering software can also be used, using the iconography of the field [107–109].

Logic statements

The logic also allows us to write the pathway dependencies as a logic statement. In order to do this in a fashion that is consistent with the diagrams, we must introduce a text-based notation.

Each entity on the pathway is given a name and we use parenthesis to indicate the type of the entity. Square brackets denote complexes, curly brackets denote genes and angled brackets denote proteins. Thus protein A, gene *B* and complex C:D are denoted <A>, {B} and [C:D], respectively. Activation and phosphorylation are denoted with the superscripts * and P. For example, phosphorylated protein A and active gene *B* are denoted with <A>^{P} and {B}*. Latent inputs and outputs are denoted with an underline, for example <u>A</u>.

Using this notation, we can write logic statements of the dependencies of each individual interaction and assemble, from these statements, descriptions of whole pathways. For example, referring to Figure 4, we can see the following individual interactions. The existence of complex [A:B] depends on proteins <A> and <B> and so we have [A:B] = <A> AND <B>. Similarly we can say that the existence of complex [C:D] requires
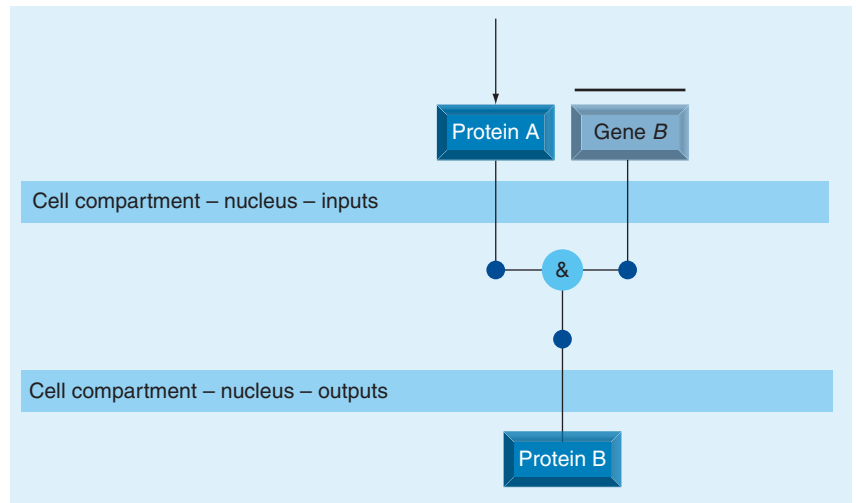

Figure 6. Transcription factor A binds to gene *B* causing transcription and translation of protein B.

the proteins <C> and <D>. Thus [C:D] = <C> AND <D>. The state that binds with the protein <E> is either [A:B] or [C:D] and we describe this state with [A:B/C:D] = [A:B] OR [C:D]. Finally, we know that complex [A:B:E/C:D:E] is the bound state of <E> with [A:B/C:D] and so we can say [A:B:E/C:D:E] = [A:B/C:D] AND <E>.

We can now use substitution to develop a description of the whole pathway. Starting from the pathway output, we have [A:B:E/C:D:E] = [A:B/C:D] AND <E> and here we can substitute for [A:B/C:D] using the interaction [A:B/C:D] = [A:B] OR [C:D]. This gives us [A:B:E/C:D:E] = ([A:B] OR [C:D]) AND <E>. We can now substitute for the complex [A:B] to give [A:B:E/C:D:E] = ((<A> AND <B>) OR [C:D]) AND <E>. We can also substitute for [C:D] to give [A:B:E/C:D:E] = ((<A>
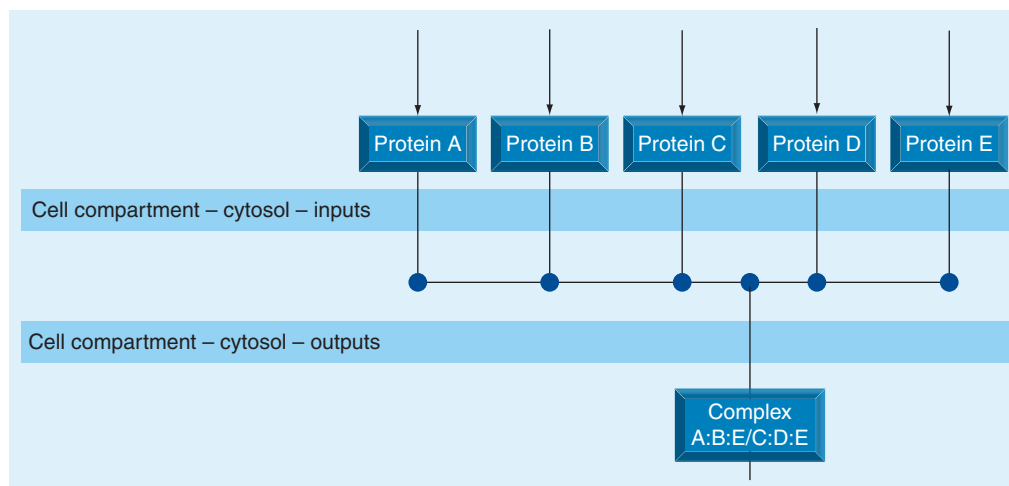

Figure 7. A concise, modular representation of the pathway shown in Figure 4. A single line can describe all the inputs to and the output from a single pathway.

AND <B>) OR (<C> AND <D>)) AND <E>. This now describes the output of the pathway entirely in terms of the inputs.

There is no loss of information in going from a diagram of the form of FIGURE 4 to a textual notation of this form and so it provides and efficient representation with which to analyse the behavior of whole pathways. It also gives us an efficient way of storing and communicating the dependencies of whole pathways.

Attractors of a logic system

When the latent inputs remain fixed, the pathway will enter into either a fixed state (one that is not changed by the pathway logic) or a repeated cycle of states, if we simulate for a sufficient length of time. These fixed states and limit cycles of states are the attractors of the system [48,49]. This is a property of all Turing machines [50,51], of which a logic representation of a pathway is an example.

A pathway can have multiple attractors and the set of pathway states that will go on to reach each attractor are the known as the basin of that attractor [20,52].

The state of the whole pathway system, at a given moment, can be described by the values of all the variables representing the components. Thus, if there are N proteins, genes and complexes in their various states passed between the sections in a pathway, we can describe the state of the whole pathway as a binary number N digits long. We can then describe the attractors and accompanying basin structure in terms of these binary numbers. As described earlier, it has been speculated that the set of attractors belonging to a pathway may correspond to different phenotypes [43].

In a logic description of a pathway with N components, there are 2^N possible initial states. The logic of the pathway will yield multiple attractors and determining which initial states belong to which attractor basins can be computationally complex for large pathway models. In the following section, we demonstrate this for a section of the Jak–Stat signaling pathway.

Jak–Stat signaling pathway

In FIGURE 5, we see a section of the Jak–Stat signaling pathway drawn as a logic system. The Jak–Stat signaling pathway is a relatively well understood mammalian host immune signaling system that contains feedback via extracellular signaling. It is small enough to be tractable to the analysis we propose and large enough to demonstrate, through the

extracellular feedback, a range of attractors. Here, we use the Jak–Stat pathway to provide an example of how the functional behavior of a pathway may be analyzed and a control strategy hypothesized.

The pathway itself crosses four cell compartments (the extracellular space, cell membrane, cytosol and nucleus) and comprises eight migrant components (two proteins, two genes and four complexes) and six latent components (three proteins, two genes and one complex).

If we distinguish between components in different compartments by labeling them using their location, we have the following logic to describe the pathway.

<IFNα_cell_membrane> = <IFNα_extracellular_space>

<IFNβ_cell_membrane> = <IFNβ_extracellular_space>

[IFNαR:Jak1:Tyk2:IFNα_cytosol] = <IFNα_cell_membrane> AND [IFNαR:Tyk2:Jak1_cell_membrane]

[IFNαR:Jak1:Tyk2:IFNβ_cytosol] = <IFNβ_extracellular_space> AND [IFNαR:Tyk2:Jak1_cell_membrane]

[Stat1:Stat1:IRF9_nucleus] = ((([IFNαR:Jak1:Tyk2:IFNα_cytosol] OR [IFNαR:Jak1:Tyk2:IFNβ_cytosol]) AND <Stat1_cytosol>) AND <IRF9_cytosol>)

[Stat1:Stat2:IRF9_nucleus] = ((([IFNαR:Jak1:Tyk2:IFNα_cytosol] OR [IFNαR:Jak1:Tyk2:IFNβ_cytosol]) AND <Stat1_cytosol>) AND ((([IFNαR:Jak1:Tyk2:IFNα_cytosol] OR [IFNαR:Jak1:Tyk2:IFNβ_cytosol]) AND <Stat2_cytosol>)) AND <IRF9_cytosol>

<IFNα_extracellular_space> = ([Stat1:Stat1:IRF9_nuclear] OR [Stat1:Stat2:IRF9_nuclear]) AND {IFNα_nuclear}

<IFNβ_extracellular_space> = ([Stat1:Stat1:IRF9_nuclear] OR [Stat1:Stat2:IRF9_nuclear]) AND {IFNβ_nuclear}.

In order to explore the complete attractor basin structure of this pathway system, we would need to consider $2^{(8+6)} = 2^{14}$ (~16,384) initial states of the pathway. However, we can reduce this to a manageable level if we assume that all the latent inputs are present in an active state. This reduces the number of initial states to be considered to $2^8 = 256$. We describe the states of the pathway in the form of an eight digit binary number in which the digits use the following order to describe the state of the pathway components.

<IFNα_extracellular_space><IFNβ_extracellular_space><IFNα_cell_membrane><IFNβ_cell_membrane>[IFNαr:Jak1:Tyk2:IFNα_

cytosol][IFNαR:Jak1:Tyk2:IFNβ_cytosol][Stat1:Stat1:IRF9_nucleus][Stat1:Stat2:IRF9_nucleus]

To determine to which basin an initial state belongs, we apply the pathway logic repeatedly until repetition appears. For example, taking 01001100 as the initial state, repeated application of the pathway logic gives:

01001100→00010011→11000100→00110011 → 11001100 → 00110011

The last state is the same as the fourth state so this sequence will repeat indefinitely. From this, we can conclude that state 01001100 belongs to the attractor 00110011 → 11001100 → 00110011.

The full set of basins of attractors for this pathway are shown in SUPPLEMENTARY TABLE 1 (see online www.futuremedicine.com/toc/fmb/5/2).

The attractor states describe the stable states of the Jak–Stat signaling pathway as it has been presented and we could speculate on its modes of operation. By understanding the basins of these states, we can learn how best to intervene to switch the pathway from one phenotype to another. On pathways of this scale, this may be possible by pharmacologically intervening with a single component of the pathway. For example, if the pathway were operating in the attractor 00110011 → 11001100 → 00110011 and we changed the state of one protein by introducing <IFNα> to the extracellular space, we could transform the state 00110011 to 10110011. The state 10110011 belongs to the basin of the attractor 00111111 → 11001111 → 11110011 → 11111100 → 00111111 and so, provided that <IFNα> dissipates quickly, the pathway will adopt the new mode of operation as a result of the intervention.

On larger pathways, there are likely to be more attractor states and more complex means of switching the pathway between attractors. It may be possible to switch the pathway between different modes of operations by intervening at a single component, but optimal changes in the mode of operations could also require intervention at several points in the pathway.

Conclusion

Understanding how pathogen subsystems exploit our own immune pathways will, by necessity, require a blend of computational and experimental biology. Here, we put forward a framework based on logic theory for reducing the impact of combinatorial complexity on pathway analysis that is amenable to computational and experimental testing. Such methods promise to shed much needed light on the emergent properties of pathway signaling response and regulation. They place us at the start of an exciting new era in microbiology, holding the promise of fundamentally new understandings of pathway biology and host–pathogen interactions.

We are hopeful, although it has not yet been demonstrated, that a pathway biology approach may lead to new and predictive insights in the targeting of host–pathogen interaction pathways. In this endeavor, new drugs that target the responding host's pathways rather than the pathogen in isolation will become a first-line anti-infective strategy. To quote Ernest Rutherford, "it is impossible until you understand it and then it becomes trivial". Hopefully, in the near future, this challenge will become trivial.

Future perspective

Modeling methodologies that recognize, from the outset, the combinatorial difficulties of dealing with large pathways have enormous potential to improve our predictive understanding of pathway biology. Once models are developed that have been sufficiently validated for us to have a high confidence in their predictions, these methods will allow us to explore pathway behavior in ways that are cheaper, quicker and more ethical than through laboratory research. An *in silico* experiment is considerably simpler and quicker to perform than its *in vitro* equivalent and, with the development of high confidence models, one would expect the *in silico* models to become the first port of call in future studies. *In vitro* and *in vivo* studies will be critical for enlarging and consolidating the body of mapped pathways and this would be by developing boundary knowledge to a level sufficient to be incorporated into the existing high-confidence models. In other sciences, this blend of theory and experiment is well established and so the movement of cell biology in a direction that formalizes the position of a theoretical understanding will serve to bring cell biology to a level that allows it to better integrate with the other sciences.

However, before such high confidence models can be developed, progress must be made on several fronts. Much of our understanding of the basic signaling biology has derived from *in vitro* experiments that poorly recreate *in vivo* conditions and so our current understanding is relatively poor of which interactions are dominant and, therefore, key and which are not. This information is encoded in the parameterization of more detailed and computationally

less tractable modeling methodologies and this problem lies at the heart of the problems with introducing *in silico* research into cell biology. For *in silico* methods to produce high confidence models, high confidence parameters need to be first obtained, comprehensively, for all types of interaction and in all conditions. For a single pathway, this is a colossal experimental undertaking and is very hard to justify when the publishable output will be significantly less than if the same financial resources were allocated to new *in vitro* studies. On a larger, more comprehensive scale, this barrier is more acute.

The likelihood is that in the next 5 years several small high confidence pathway models will appear with possible pharmacological value. However, the difficulty of obtaining broad and comprehensive parameter values across all interactions is likely to be a longer-term issue. In other sciences, this level of detail has tended to be addressed once the boundaries of the science have been reached, meaning that parameter studies will be evidence of a level of maturity in the field of systems biology.

Executive summary

Pathway modeling

- By combining our understanding of the interactions within the cell, we can gain insight into the pathways that propagate signals around the cell.
- Pathways can demonstrate emergent behavior, meaning behavior that cannot be determined purely by considering the individual interactions.
- By learning more about the pathway structure, we will be able to devise new therapeutic strategies for manipulating cellular behavior, particularly in response to infection.

Pathway assembly

- Consensus pathways can be assembled by searching the published literature.
- This can be augmented with data-mining approaches.
- The resulting pathway information can be presented graphically using systems biology graphical notation or one of a range of other notations.

Quantitative modeling

- Quantitative modeling is part of a process of hypothesis generation, experimental testing and knowledge refinement.
- The simplest starting point from which to build a model is from a consensus pathway diagram.
- Combinatorial complexity ensures that model analysis can be extremely computationally demanding.
- Strategies such as modularity and homeostasis can have a limited impact in reducing the computational demands.
- Logic can be used as a modeling methodology that minimizes the combinatorial impact of pathway simulation and analysis as far as possible.

Logic representations

- Drawing on methods analogous to electrical engineering, diagrams and logic statements can be assembled to describe the dependencies of pathways.
- Logic models of pathways require fewer simulations in order to determine unknown parameters of unknown starting states than other models.
- Discretization of data is required to fit data to a logic model.

Attractor states & phenotypes

- All logic models progress to reach either a steady state or a repeated cycle of states (collectively, known as attractors).
- The dominant behaviors of pathways correspond to attractors and the phenotypic behavior of cells has been speculated to correspond to the attractors of cellular signaling networks.

Conclusion

- Logic provides a possible approach to understanding the behavior of larger networks of signaling pathways than is currently possible.
- We propose the use of logic models for devising future therapeutic strategies that maximize efficacy and minimize side effects.

Bibliography

Papers of special note have been highlighted as:
- ▪ of interest
- ▪▪ of considerable interest

1. Fell D: Metabolic control analysis: a survey of its theoretical and experimental development. *Biochem. J.* 286, 313–330 (1992).

2. Kell D: Systems biology, metabolic modelling and metabolomics in drug discovery and development. *Drug Discov. Today* 11, 1085–1092 (2006).

3. Ratushnyi A, Likhoshvai V, Ignat'eva E *et al.*: 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 (2003).

4. Ma A, Sorokin A, Mazein A *et al.*: The Edinburgh human metabolic network reconstruction and its functional analysis. *Mol. Sys. Biol.* 3, 135 (2007).

5. McAdams H, Arkin A: Stochastic mechanisms in gene expression. *Proc. Natl Acad. Sci. USA* 94, 814–819 (1997).

6. Arkin A, Ross J, McAdams H: Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *Escherichia coli* cells. *Genetics* 149, 1633–1648 (1998).

7. Hofestadt R, Thelen S: Quantitative modelling of biochemical networks. *In Silico Biol.* 1, 39–53 (1998).

8. Küffner R, Zimmer R, Lengauer T: Pathway analysis in metabolic databases via differential metabolic display (DMD). *Bioinformatics* 16, 825–836 (2000).

9. Goss P, Peccoud J: Quantitative modelling of stochastic systems in molecular biology by using stochastic Petri Nets. *Proc. Natl Acad. Sci. USA* 95, 6750-6755 (1998).

10. Calder M, Gilmore S, Hillston J: Modelling the influence of RKIP on the ERK signalling pathway using the stochastic process algebra PEPA. In: *Transactions on Computational Systems Biology VII*. Istrail S, Pevzner P, Waterman M (Eds). Springer Berlin/ Heidelberg, Germany, 1–23 (2006).

11. Regev A, Silverman W, Shapiro E: Representation and simulation of biochemical processes using the π-calculus process algebra. *Pac. Symp. Biocomput.* 459–470 (2001).

12. Danos V, Feret J, Fontana W, Harmer R, Krivine J: Rule-based modelling of cellular signalling. Presented at: *Proceedings of the 18th International Conference on Concurrency Theory (CONCUR'07)*. Caires L, Vasconcelos, V (Eds), Lecture Notes in Computer Science (2007).

13. Faeder J, Blinov M, Hlavacek W: Graphical rule-based representation of signal-transduction networks. *Proc. ACM Symp. Appl. Comp.* 133–140 (2005).

14. Boole G: Mathematical analysis of logic. Being an essay towards a calculus of deductive reasoning. (1847), reprinted: Blackwell, Oxford, UK (1948).

15. Boole G: An investigation of the laws of thought, on which are founded the mathematical theories of logic and probabilities. (1854) Reprinted: Dover Publications, New York, USA (1958).

16. Shannon C: A symbolic analysis of relay and switching circuits. *Trans. AIEE* 57, 713–722 (1938).

17. Kauffman S: Homeostasis and differentiation in random genetic control networks. *Nature* 224(5215), 177–178 (1969).

▪▪ **Original work in which logic was posited to the biological community as a description of gene activity.**

18. Kaufman M, Andris F, Leo O: A logical analysis of T cell activation and anergy. *Proc. Natl Acad. Sci. USA* 96(7), 3894–3899 (1999).

▪▪ **Clear example of the application of logic to a signaling system.**

19. Laubenbacher R, Stigler B: A computational algebra approach to the reverse engineering of gene regulatory networks. *J. Theor. Biol.* 229(4), 523–537 (2004).

20. Shmulevich I, Dougherty E, Kim S, Zhang W: Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics* 18(2), 261–274 (2002).

▪ **Clear description of a scheme that allows statistical switching between pathway logics.**

21. Watterson S, Marshall S, Ghazal P: Logic models of pathway biology. *Drug Discov. Today* 13(9/10), 447–456 (2008).

▪ **Introduction to the use of logic in modeling signaling.**

22. Samaga R, Saez-Rodriguez J, Alexopoulos L, Sorger P, Klamt S: The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data. *PLoS Comp. Biol.* 5(8), E1000438 (2009).

23. Schlatter R, Schmich K, Vizcarra I *et al.*: ON/OFF and beyond – a Boolean model of apoptosis. *PLoS Comp. Biol.* 5(12), E1000595 (2009).

24. Zhang B, Shah M, Yang J *et al.*: Network model of survival signaling in large granular lymphocyte leukemia. *Proc. Natl Acad. Sci. USA* 105(42), 16308–16313 (2008).

25. Shmulevich I, Dougherty E, Zhang W: From Boolean to probabilistic Boolean networks as models of genetic regulatory networks. *Proc. IEEE* 90(11), 1778–1792 (2002).

26. Cooper H, Hedges L: *The Handbook of Research Synthesis*. Russell Sage Foundation, NY, USA (1994).

27. Grimes G, Wen T, Mewissen M *et al.*: PDQ Wizard: automated prioritization and characterization of gene and protein lists using biomedical literature. *Bioinformatics* 22, 2055–2057 (2006).

28. Sorokin A, Paliy K, Selkov A *et al.*: The pathway editor: a tool for managing complex biological networks. *IBM J. Res. Dev.* 50, 561–573 (2006).

29. Funahashi A, Tanimura N, Morohashi M, Kitano H: Cell designer: a process diagram editor for gene-regulatory and biochemical networks. *BIOSILICO* 1, 159–162 (2003).

30. Hoops S, Sahle S, Gauges R *et al.*: COPASI – a complex pathway simulator. *Bioinformatics* 22, 3067–3074 (2006).

31. Le Novere N, Hucka M, Mi H: The systems biology graphical notation. *Nat. Biotechnol.* 27, 735–741 (2009).

▪ **Recent, community-driven effort to establish a common graphical notation for drawing pathways.**

32. Kitano H, Funahashi A, Matsuoka Y, Oda K: Using process diagrams for the graphical representation of biological networks. *Nat. Biotechnol.* 23(8), 961–966 (2005).

33. Kohn K: Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Mol. Biol. Cell* 10(8), 2703–2734 (1999).

34. Raza S, Robertson K, Lacaze P *et al.*: A logic-based diagram of signalling pathways central to macrophage activation. *BMC Syst. Biol.* 2, 36 (2008).

35. Bader G, Donaldson I, Wolting C: BIND – the biomolecular interaction network database. *Nucleic Acids Res.* 29(1), 242–245 (2001).

36. Hermjakob H, Montecchi-Palazzi L, Bader G *et al.*: The HUPO PSI's molecular interaction format – a community standard for the representation of protein interaction data. *Nat. Biotechnol.* 22, 177–183 (2004).

37. Joshi-Tope G, Gillespie M, Vastrik I *et al.*: Reactome: a knowledgebase of biological pathways, *Nucleic Acids Res.* 33, D428–D432 (2005).

38. Pico A, Kelder T, van Iersel M: WikiPathways: pathway editing for the people. *PLoS Biol.* 6(7), E184 (2008).

39. Oda K, Kimura T, Matsuoka Y, Muramatsu M, Kitano H: Molecular interaction map of macrophage. *AfCS Res. Rep.* 2(14), DA (2004).

40. Moodie S, Sorokin A, Ghazal P: A graphical notation to describe the logical interactions of biological pathways. *J. Integr. Bioinform.* 3(2), 36 (2006).

41. Hucka M, Finney A, Bornstein B *et al.*: Evolving a lingua franca and associated software infrastructure for computational systems biology: the systems biology markup language (SBML) project. *IEE Syst. Biol.* 1(1), 41–53 (2004).

■ **Recent, community-driven effort to establish a file format in which to exchange pathway models.**

42. Davis P, Rabinowitz P: *Methods of Numerical Integration (2nd edition)*. Academic Press, NY, USA (1984).

43. Gillespie D: Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81(25), 2340–2361 (1977).

44. Ghazal P, Garcia-Ramirez J, Gonzales-Armas JC, Kurz S, Angulo A: Principles of homeostasis in governing virus activation and latency. *Immunol. Res.* 21(2–3), 219–223 (2000).

45. Ghazal P, Gonzalez-Armas JC, Garcia-Ramirez JJ, Kurz S, Angulo A: Viruses: hostages to the cell. *Virology* 275(2), 233–237 (2000).

46. Virgin HW, Wherry EJ, Ahmed R: Redefining chronic viral infection. *Cell* 138(1), 30–50 (2009).

47. Huang S, Eichler G, Bar-Yam Y, Ingber D: Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.* 94, 128701 (2005).

48. Glass L, Pasternack J: Predictions of limit cycles in mathematical models of biological oscillations. *Bull. Math. Biol.* 40(1), 27–44 (1978).

49. Yan-Qian Y, Chi Y: *Theory of Limit Cycles*. Gould SH, Hale JK (Eds). American Mathematical Society, RI, USA (2009).

50. Turing A: On computable numbers with an application to the Entscheidungs problem. *Proc. Lond. Math. Soc.* 2(42), 230–265 (1937).

51. Goldin D, Wegner P: The Church–Turing thesis: breaking the myth. Lecture notes in computer science. Springer Berlin/Hamberg, Germany, 3526 (2005).

52. Yu L, Watterson S, Marshall S, Ghazal P: Inferring Boolean networks with perturbation from sparse gene expression data: a general model applied to the interferon regulatory network. *Mol. Biosyst.* 4, 1024–1030 (2008).

■■ **The authors apply attractor and basin analysis to the cell differentiation of human promyelocytic HL60 cells as the differentiate to neutrophil cells.**

Websites

101. Kyoto Encyclopedia of Genes and Genomes
www.genome.jp/kegg/

102. Human Protein Reference Database
www.hprd.org

103. Chilibot
www.chilibot.net

104. Ingenuity Pathway Analysis
www.ingenuity.com

105. yEd
www.yworks.com

106. SimBiology
www.mathworks.com/products/simbiology/

107. Logic Simulator
http://www.tetzl.de/java_logic_simulator.html

108. Logic Circuit Designer
http://download.cnet.com/Logic-Circuit-Designer/3000–2054_4–10840569.html

109. Logic Circuit Simulator
www.itlocation.com/en/software/prd58983,,.htm

Affiliations

▪ Steven Watterson
Division of Pathway Medicine, University of Edinburgh Medical School, Chancellor's Building, 49 Little France Crescent, Edinburgh, EH16 4SB, Scotland, UK
Tel.: +44 131 242 6242
Fax: +44 131 242 6244
and
Centre for Systems Biology at Edinburgh, CH Waddington Building, King's Buildings, Mayfield Road, Edinburgh, EH9 3JY, Scotland, UK
Tel.: +44 131 651 9065
Fax: +44 131 651 9068
s.watterson@ed.ac.uk

▪ Peter Ghazal
Division of Pathway Medicine, University of Edinburgh Medical School, Chancellor's Building, 49 Little France Crescent, Edinburgh, EH16 4SB, Scotland, UK
Tel.: +44 131 242 6242
Fax: +44 131 242 6244
and
Centre for Systems Biology at Edinburgh, CH Waddington Building, King's Buildings, Mayfield Road, Edinburgh, EH9 3JY, Scotland, UK
Tel.: +44 131 651 9065
Fax: +44 131 651 9068
p.ghazal@ed.ac.uk