Multi-pathway network analysis of mammalian epithelial cell responses in inflammatory environments

David C. Clarke and Douglas A. Lauffenburger[1]
Department of Biological Engineering and the Center for Cellular Decision Processes, Massachusetts Institute of Technology, 77 Mass Ave., Cambridge, MA 02139, U.S.A.

Abstract

Inflammation is a key physiological response to infection and injury and, although usually beneficial, it can also be damaging to the host. The liver is a prototypical example in this regard because inflammation helps to resolve liver injury, but it also underlies the aetiology of pathologies such as fibrosis and hepatocellular carcinoma. Liver cells sense their environment, including the inflammatory environment, through the activities of receptor-mediated signal transduction pathways. These pathways are organized in a complex interconnected network, and it is becoming increasingly recognized that cellular adaptations result from the quantitative integration of multi-pathway network activities, rather than isolated pathways causing particular phenotypes. Therefore comprehending liver cell signalling in inflammation requires a scientific approach that is appropriate for studying complex networks. In the present paper, we review our application of systems analyses of liver cell signalling in response to inflammatory environments. Our studies feature broad measurements of cell signalling and phenotypes in response to numerous experimental perturbations reflective of inflammatory environments, the data from which are analysed using Boolean and fuzzy logic models and regression-based methods in order to quantitatively relate the phenotypic responses to cell signalling network states. Our principal biological insight from these studies is that hepatocellular carcinoma cells feature uncoupled inflammatory and growth factor signalling, which may underlie their immune evasion and hyperproliferative properties.

Introduction

Inflammation is a mechanism for preserving homoeostasis in response to noxious stimuli such as infection and injury. Cells of the innate immune system drive the inflammatory response, which typically begins when pro-inflammatory stimuli activate macrophages residing in the affected tissue to produce chemoattractants that recruit neutrophils to the site of inflammation. Neutrophils are white blood cells that sense and eliminate pathogens. They carry out their functions in part by release of oxygen radicals and degradative enzymes whose leakage can cause collateral tissue damage [1]. Once the inflammatory stimulus is cleared, the inflammatory response is damped by a process called resolution, which is marked by the transition from neutrophil to monocyte recruitment. Monocytes are blood-borne macrophages that differentiate into macrophages once inside tissue. They serve to clear debris and promote tissue repair. The inflammatory response is therefore characterized by processes that cause both damage and repair. The damaging processes must be sufficiently strong to eliminate the inflammatory stimulus, but not too strong so as to cause excess tissue damage. A maladapted inflammatory response can lead to chronic inflammation, which is a hallmark of many complex diseases such as cancer, fibrosis, cardiovascular disease and diabetes. A better understanding of inflammation could therefore improve therapeutic approaches to acute and chronic diseases.

The liver plays a central role in homoeostasis through its functions in metabolism, detoxification and inflammation. The liver is a key participant in the initial systemic response to inflammation, called the acute-phase response, because it synthesizes acute-phase proteins such as C-reactive protein, serum amyloid A and fibrinogen [2]. Conversely, components of the inflammatory response are important in liver physiology and pathophysiology. The cytokine IL (interleukin)-6, for example, serves to protect the liver when it is injured and promotes its regeneration [3]. Inflammation can contribute to pathological states of the liver, perhaps best exemplified by chronic inflammation due to viral infection (e.g. hepatitis B and C), toxic substance exposure (e.g. aflatoxin-B1) or steatosis serving as a precursor to HCC (hepatocellular carcinoma) [4]. HCC is the fifth most prevalent cancer worldwide and is notoriously difficult to treat, which underlies its status as the third most lethal type of cancer [5]. The considerable burdens of liver disease and diseases linked to chronic inflammation emphasize the

Key words: cell signalling, hepatocellular carcinoma, inflammation, liver, logic, mathematical model.

Abbreviations used: cFL, constrained fuzzy logic; ERK, extracellular-signal-regulated kinase; HCC, hepatocellular carcinoma; IL, interleukin; IL6R, IL-6 receptor; MAPK, mitogen-activated protein kinase; MEK, MAPK/ERK kinase; ODE, ordinary differential equation; PI3K, phosphoinositide 3-kinase; PKN, prior knowledge network; PLSR, partial least-squares regression; PSN, protein signalling network; STAT3, signal transducer and activator of transcription 3; TNFα, tumour necrosis factor α.

[1]To whom correspondence should be addressed (email lauffen@mit.edu).

need for investigating the interplay between the liver and inflammation.

Cells adapt to their environments by the activities of receptor-mediated signalling pathways. The biochemical activities of these signalling pathways regulate gene expression, metabolism and/or cell structure in order to modify cell physiology. For example, in the case of hepatocytes during the acute-phase response, IL-6 released by macrophages and stromal fibroblasts at the site of inflammation acts hormonally on hepatocytes by binding and activating a receptor complex leading to phosphorylation and dimerization of the STAT3 (signal transducer and activator of transcription 3) [6,7]. STAT3 is a transcription factor that regulates the transcription of a number of genes including many involved in the acute-phase response. The communication between cells acting at the site of inflammation and those of the liver exemplify how the cells of multicellular organisms communicate in homoeostasis.

The activities of individual pathways such as IL-6/STAT3 have been well studied in liver cells. However, studies of isolated pathways have limited applicability to the *in vivo* situation in which cells are continually exposed to multiple extracellular molecules leading to the simultaneous activity of many signalling pathways. In inflammation, for example, cells are exposed to multiple pro- and anti-inflammatory cytokines [e.g. TNFα (tumour necrosis factor α), IL-1, IL-4 or IL-10], growth factors, hormones and other molecules. Intracellular signalling pathways cross-talk with each other, effectively forming a network, but it is poorly understood how intracellular signalling networks processes the combinatorial action of multiple environmental cues. Progress in this area will improve our understanding of complex physiological responses such as inflammation and help to guide the development of better therapeutics. In the present paper, we review our studies of the intracellular signalling systems of hepatocytes and HCC cells in response to inflammatory environments using a systems approach. We begin by explaining what a systems approach to biology means followed by reviewing three studies in which we used this approach to investigate liver cell signalling and physiology in inflammatory contexts.

Scientific approach: 'cue–signal–response' experiments and mathematical modelling

A hallmark of engineering practice is applying a systems approach to the design process. In this context, a 'systems approach' denotes studying a system by applying diverse inputs to the system and measuring the outputs. Mathematical models are then used to quantify the relationship between input and output. The parameters of the mathematical model are tuned such that the desired outputs are obtained from the inputs expected under operating conditions. The engineer then modifies the design to reflect these optimal parameter values.

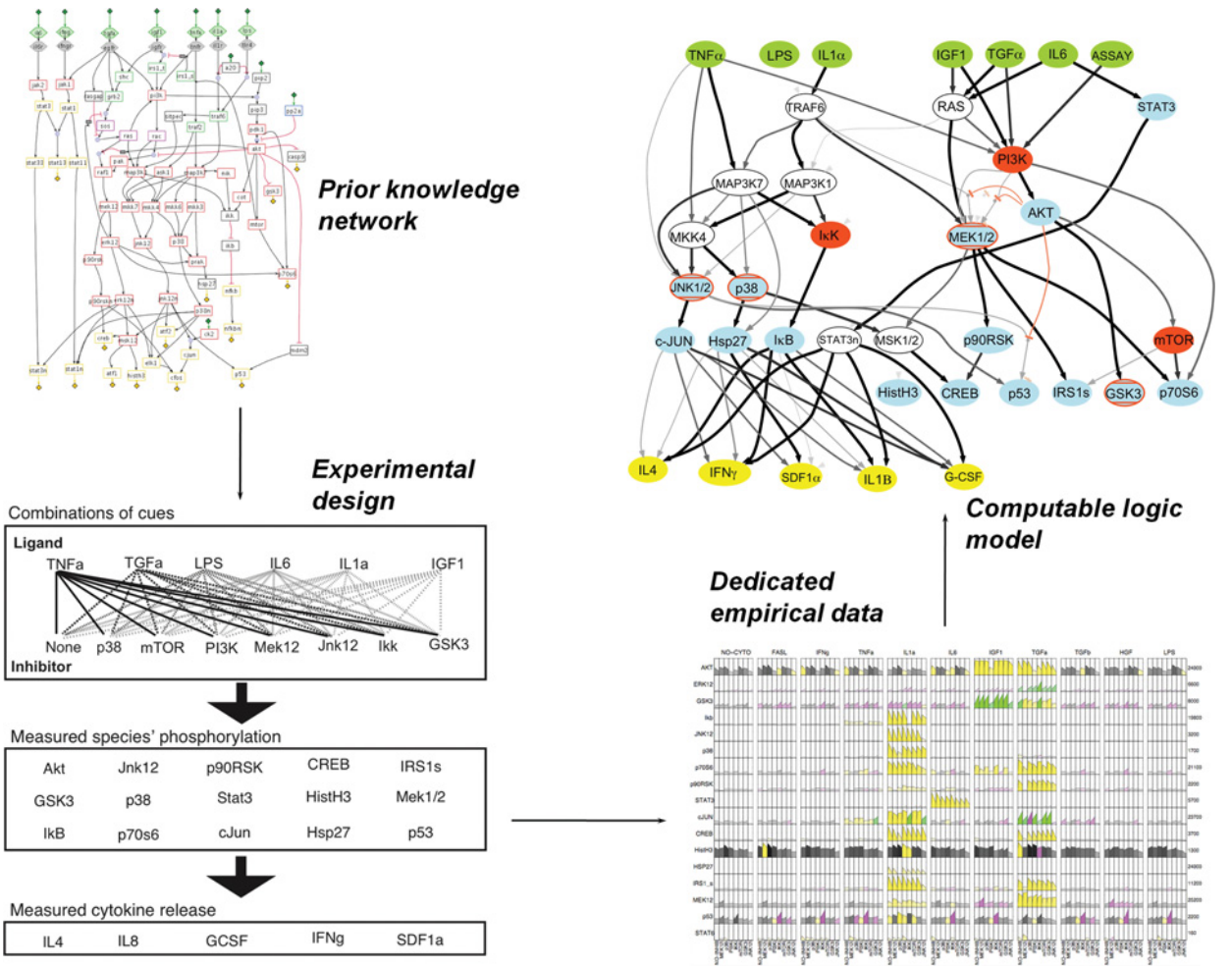
We adopt a similar engineering approach in investigating biological systems, with a difference being that we seek to 'reverse engineer' the system by using input–output relationships and selected measurements of the system to constrain a model of the intracellular signalling network. We implement an experimental paradigm called cue–signal–response that reflects this input–system–output relationship [8]. The input consists of molecules in the cell's environment, the system is the cell signalling network and the output is the behaviour that the cell executes to adapt to the input. In practice, a cue–signal–response paradigm applied to hepatocyte physiology during inflammation involves applying inflammatory cytokines and growth factors (cues) to cultured hepatocytes or HCC cells, perturbing intracellular signalling by inhibiting kinases with small-molecule inhibitors, assessing intracellular signalling by multiplexed measurement of phospho-protein levels (signals) and measuring the secretion of cytokines (responses). Specific cues include inflammatory cytokines such as IL-6 and IL-1 and inhibitors target kinases such as MEK [MAPK (mitogen-activated protein kinase)/ERK (extracellular-signal-regulated kinase) kinase], PI3K (phosphoinositide 3-kinase) and p38 MAPK. The cells are lysed at specific time points and multiplexed bead-based flow cytometric immunoassays based on Luminex xMAP technology are used to measure the levels of ~15 phosphorylated proteins including Akt, ERK and STAT3. Similar assays are also used to measure the levels of ~50 secreted proteins in the cell culture medium.

The resulting dataset features thousands of data points, which makes it challenging to interpret. Mathematical tools are therefore used for downstream analyses. Classically, ODEs (ordinary differential equations) have been the method of choice for analysing cell signalling systems because they make direct use of biochemical rate equations that describe the kinetics of enzyme-catalysed reactions, protein–protein interactions and transport processes. Disadvantages of ODE models include their critical requirement for firm specification of network topological interactions and their reliance on adjustable parameters that must be robustly estimated in order to effectively represent the system. As the size of the model grows, so does the uncertainty in the topology and the corresponding number of parameters, which in turn increases the demand for more comprehensive biological knowledge and intensive experimental data (as well as computational power, although this is a lesser challenge at this point). Although studying signalling from one or two pathways with ODEs is feasible, the networks that we study are too uncertain and large for ODEs to be practically useful. We therefore use modelling techniques that represent the system in a coarser-grained fashion. In so doing, we require less data to obtain a quantitative insight into the system, albeit less insight than could be obtained with ODEs.

Our studies feature two types of mathematical frameworks, regression-based methods and logic-based methods. Our implementation of these techniques has been reviewed in detail elsewhere [9–11]. Briefly, regression models, such as multiple linear regression and PLSR (partial least-squares regression), are useful for quantifying the correlation between variables in the context of one another. Regression models do not incorporate information beyond the data itself, except

Figure 1 | **Logic modelling workflow**

A PKN is derived from literature, databases and/or existing data. Experiments are conducted to systematically perturb and/or measure nodes distributed throughout the network. The data are then used by a model-optimization (i.e. data-fitting) algorithm to tune the model topology and, if applicable, the model parameters, to minimize the discrepancy between the model output and data. The resulting model is then analysed to derive insight into the biology of the system. Adapted from [14] with permission under a Creative Commons Attribution Licence.



that the variables included in the model are specified by the modeller. This prior specification makes the models supervised, but they are also strictly empirical. Logic modelling, in contrast, involves translating prior knowledge or hypotheses about the system structure or function into computable language. In this way, logic models are capable of bringing a network diagram to operational function. We investigate cell signalling networks using logic modelling by first constructing a diagram of the network based on published data [which we call a PKN (prior knowledge network)], then collecting a cue–signal–response dataset devoted to perturbing and measuring aspects of the network, followed by using optimization algorithms to identify and quantify the connections in the hypothetical network that are most important for explaining the data (Figure 1). The resulting fitted models can then be used for simulation or analysis purposes. Irrespective of the modelling approach, model predictions are experimentally validated. We have used this workflow to obtain considerable insights into the systems-level operation of cell signalling networks in effecting phenotypic responses in diverse contexts, which we discuss in the following section.

Liver cell information processing during inflammation

We applied our interdisciplinary approach to the issue of epithelial cell signalling in inflammation by devising logical and statistical modelling methods and applying them to data from cultured liver cells exposed to inflammatory conditions. In the first paper from these studies, Saez-Rodriguez et al. [12] extended Boolean logic methods used previously to study biological networks from a theoretical standpoint to allow the model to directly interface

with experimental data. Specifically, their algorithm translates a PKN, a database-derived PSN (protein signalling network) map in this specific example, into a Boolean logic model and optimizes the model topology to best fit experimental data. In this example, the data comprised phospho-protein levels of intracellular signalling intermediates in HepG2 HCC cells exposed to inflammatory cytokines and inhibitors of several kinases. Interestingly, they found that the resulting optimal models consisted of substantially fewer connections than found in the PKN. Remarkably, an empty model that contained nodes but no edges connecting them, fitted the data better than the PKN [12]. This result implies that the comprehensive protein–protein interaction network maps and PSNs commonly used to depict biological networks do not necessarily reflect networks operating in a specific cell type under specific conditions. This lack of fit stems from two apparently paradoxical sources: (i) the networks include too many interactions, presumably because they are typically curated from multiple sources, cell types, time points and experimental conditions; and (ii) the networks lack interactions that are present and functional in the network under study due to imperfect databases or incomplete understanding of the biology. Indeed, by examining the data points that the original optimal model failed to adequately fit, Saez-Rodriguez et al. [12] tested new interactions to see which ones best improved the fit. The existence of two candidate interactions, one linking TRAF6 (TNF-receptor associated factor 6) and MEK and another linking ERK and IRS-1 (insulin receptor substrate 1), was supported by published evidence.

An alternative modelling approach was used by Alexopoulos et al. [13], who performed a comparative analysis of the intracellular signalling networks of healthy and cancerous liver cells. A cue–signal–response dataset was generated in which inflammatory cytokines and growth factors were applied to primary human hepatocytes and HepG2 HCC cells in concert with small-molecule inhibitors targeting seven kinases from different signalling pathways. Multiple linear regression analysis was used to estimate the strength of relationships between the cytokines and signals, the inhibitors and signals, and the signals and secreted cytokines. The networks were defined by the relationships featuring the highest regression coefficient magnitudes. The networks for the primary hepatocytes and the HepG2 cells were then compared, revealing that HepG2 cells displayed reduced responsiveness to inflammatory stimuli, but increased responsiveness to pro-growth stimuli, relative to the primary hepatocytes. In particular, alterations of NF-κB (nuclear factor κB) signalling in HCC cells had profound phenotypic consequences because primary hepatocytes secreted a number of cytokines that the HCC cells did not. These cytokines are thought to be involved in recruiting cells of the innate immune system, which implies that HCC cells modify the secretion patterns in order to avoid detection and elimination by the immune system.

The above studies clearly demonstrate the utility of Boolean logic and regression methods in exploring how normal and healthy liver cells differentially process environmental information. However, both modelling methods have important limitations. Specifically, Boolean logic describes the activity of each node as either 'on' or 'off', which ignores potentially important graded activity, and regression models do not incorporate potentially valuable prior knowledge. Morris et al. [14] addressed these limitations by developing a novel logic-based modelling method called cFL (constrained fuzzy logic), which incorporates prior knowledge in the same manner as Boolean logic, but also models quantitative behaviour. Specifically, this method converts an input value from an upstream node into a continuous value between 0 and 1 for the downstream node through a sigmoidal transfer function [14]. This capability facilitates the ability of the resulting trained models to fit weak responses. The quantitative relationships between proteins are also estimated, allowing for modelling of dose–response data, which could prove valuable for pharmacological applications.

cFL was applied to the same dataset to which the Boolean logic approach had been applied previously [12]. cFL was able to capture interactions that were missed by Boolean logic [14]. This included the moderate levels of phosphorylation of JNK (c-Jun N-terminal kinase) and c-Jun by TGFα stimulation. This interaction was the only instance of growth factor pathway cross-talk with inflammatory pathways observed in measurements of HepG2 cells, which the previous studies failed to detect [12,13]. Furthermore, stimulating HepG2 cells with IL-6 led to moderately increased phosphorylation of several species, including Akt, MEK and p70 S6 kinase, in addition to strong phosphorylation of its canonical downstream STAT3 pathway [14]. The PKN did not include links from the IL-6 receptor that allowed for the observed moderate phosphorylation levels [14]. In the case of Boolean logic, the resulting lack of fit of these data points did not adversely affect the overall fit, presumably because a similar absolute deviation resulted between the intermediate levels of the measured phosphorylations and the model outputs of 0 or 1. In contrast, the cFL model was sufficiently sensitive to this error that the model was deemed to inadequately fit this data [14]. Morris et al. [14] followed up this result by seeking to distinguish the pathway that most likely caused the phosphorylations. To do so, they tested PKNs with new interactions either between the IL6R (IL-6 receptor) and PI3K or IL6R and Ras. Most of the resulting fitted models contained the IL6R–Ras link, thus predicting that the Ras/Raf/MEK pathway and not a PI3K-downstream pathway mediated the phosphorylations [14]. This result was validated with dedicated experiments.

Conclusions and future directions

We have performed a series of studies in which mathematical models of proteomic data revealed important insights into the signal transduction networks of healthy and cancerous liver cells in inflammatory environments. Our principal biological insight is that HCC cells feature both decreased responsiveness to inflammatory stimuli and increased responsiveness to growth factors relative to normal

hepatocytes, which could promote immune evasion and increased proliferation. Our principal mathematical advances include devising methods for formally fitting Boolean logic models to data and creating a fuzzy logic method useful for making quantitative models. For relatively small networks such as the one studied here (i.e. downstream of approximately five to seven receptors), cFL is a powerful approach. Given the higher computational burden of cFL, Boolean logic will still be needed to model larger networks until more efficient algorithms are developed. We note, however, that larger networks do not necessarily provide additional predictive power because maximal predictivity was observed with models featuring substantially reduced numbers of edges from the initial PKN. This finding supports the use of our functional biochemistry approach because it provides data on components of the system that actually carry out the cellular response. A distinctive feature of our approach is that it requires broad sampling of network states, which is achieved by applying diverse experimental conditions (in our case, cytokines and inhibitors, but could also include other treatments such as small interfering RNAs). Attempting to process samples from hundreds of independent experiments with other proteomics techniques such as MS is currently unfeasible owing to technical limitations and time and fiscal costs. Our approach therefore represents a rational, efficient and informative means to elucidating epithelial cellular signalling and physiology in inflammatory contexts.

We emphasize that systems-level approaches can be effectively used *in vivo*. A recent study from our laboratory successfully extended previous systems-level analyses of data collected from colon cancer cells *in vitro* [15–17] by applying PLSR modelling to signalling measurements taken from the intestines of mice treated systemically with TNFα [18]. Determining the biological effects of TNFα is not straightforward because stimulation of TNFα receptors increases the activity of multiple downstream signalling pathways, the quantitative integration of which determines the ultimate biological outcome. In the case of mouse intestinal epithelial cells *in vivo*, for example, TNFα was found to promote apoptosis in cells of the proximal part of the small intestine, but not in the distal part, with the timing of apoptosis being dose-dependent [18]. TNFα administration also affected cell proliferation in a region-specific manner. PLSR modelling of signalling and phenotype data revealed that the differential sensitivity of apoptosis was due to quantitative differences in MAPK signalling kinetics between the two intestinal regions and that growth arrest was related to c-Jun and ATF (activating transcription factor) activation as well as MAPK signalling kinetics [18]. Subsequent experiments validated the hypotheses generated from the original dataset and model. This study demonstrates that our systems-level approach can be applied successfully to *in vivo* contexts, despite their added complexity compared with *in vitro* cell-culture-based experiments.

We contend that our approach works irrespective of the experimental system because cells integrate complex contextual information into biochemical activities of signalling pathways that form the basis for phenotypic decisions. The cell signalling network is complex, but manageable, such that by measuring selected nodes across this network and using mathematical models to infer the network output, we are able to predict the ultimate biological outcome. We therefore anticipate systems-level approaches becoming broadly applicable to the study of cellular signalling.

Acknowledgements

We thank Melody Morris, Julio Saez-Rodriguez and Scott Dixon for critically reading the paper before submission.

Funding

Our work is supported by the National Institutes of Health (NIH) [grant number R24-DK090963], the Center for Cellular Decision Processes [grant number NIH P50-GM68762] and the Institute for Collaborative Biotechnologies [through contract number W911NF-09-D-0001 from the U.S. Army Research Office].

References

1 Sipe, J.D. (1985) Cellular and humoral components of the early inflammatory reaction. In the Acute-Phase Response to Injury and Infection: the Roles of Interleukin 1 and Other Mediators (Gordon, A.H. and Koj, A., eds), p. 339, Elsevier, Amsterdam
2 Baumann, H. and Gauldie, J. (1994) The acute phase response. Immunol. Today **15**, 74–80
3 Taub, R. (2004) Liver regeneration: from myth to mechanism. Nat. Rev. Mol. Cell Biol. **5**, 836–847
4 Farazi, P.A. and DePinho, R.A. (2006) Hepatocellular carcinoma pathogenesis: from genes to environment. Nat. Rev. Cancer **6**, 674–687
5 El-Serag, H.B. and Rudolph, K.L. (2007) Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. Gastroenterology **132**, 2557–2576
6 Singh, A., Jayaraman, A. and Hahn, J. (2006) Modeling regulatory mechanisms in IL-6 signal transduction in hepatocytes. Biotechnol. Bioeng. **95**, 850–862
7 Heinrich, P.C., Behrmann, I., Muller-Newen, G., Schaper, F. and Graeve, L. (1998) Interleukin-6-type cytokine signalling through the gp130/Jak/STAT pathway. Biochem. J. **334**, 297–314
8 Gaudet, S., Janes, K.A., Albeck, J.G., Pace, E.A., Lauffenburger, D.A. and Sorger, P.K. (2005) A compendium of signals and responses triggered by prodeath and prosurvival cytokines. Mol. Cell. Proteomics **4**, 1569–1590
9 Kumar, N., Hendriks, B.S., Janes, K.A., de Graaf, D. and Lauffenburger, D.A. (2006) Applying computational modeling to drug discovery and development. Drug Discovery Today **11**, 806–811
10 Lazzara, M.J. and Lauffenburger, D.A. (2009) Quantitative modeling perspectives on the ErbB system of cell regulatory processes. Exp. Cell Res. **315**, 717–725
11 Morris, M.K., Saez-Rodriguez, J., Sorger, P.K. and Lauffenburger, D.A. (2010) Logic-based models for the analysis of cell signaling networks. Biochemistry **49**, 3216–3224
12 Saez-Rodriguez, J., Alexopoulos, L.G., Epperlein, J., Samaga, R., Lauffenburger, D.A., Klamt, S. and Sorger, P.K. (2009) Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. Mol. Syst. Biol. **5**, 331
13 Alexopoulos, L.G., Saez-Rodriguez, J., Cosgrove, B.D., Lauffenburger, D.A. and Sorger, P.K. (2010) Networks inferred from biochemical data reveal profound differences in Toll-like receptor and inflammatory signaling between normal and transformed hepatocytes. Mol. Cell. Proteomics **9**, 1849–1865

14 Morris, M.K., Saez-Rodriguez, J., Clarke, D.C., Sorger, P.K. and Lauffenburger, D.A. (2011) Training signaling pathway maps to biochemical data with constrained fuzzy logic: quantitative analysis of liver cell responses to inflammatory stimuli. PLoS Comput. Biol. **7**, e1001099
15 Janes, K.A., Albeck, J.G., Gaudet, S., Sorger, P.K., Lauffenburger, D.A. and Yaffe, M.B. (2005) A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. Science **310**, 1646–1653
16 Miller-Jensen, K., Janes, K.A., Brugge, J.S. and Lauffenburger, D.A. (2007) Common effector processing mediates cell-specific responses to stimuli. Nature **448**, 604–608
17 Kreeger, P.K., Mandhana, R., Alford, S.K., Haigis, K.M. and Lauffenburger, D.A. (2009) RAS mutations affect tumor necrosis factor-induced apoptosis in colon carcinoma cells via ERK-modulatory negative and positive feedback circuits along with non-ERK pathway effects. Cancer Res. **69**, 8191–8199
18 Lau, K.S., Juchheim, A.M., Cavaliere, K.R., Philips, S.R., Lauffenburger, D.A. and Haigis, K.M. (2011) *In vivo* systems analysis identifies spatial and temporal aspects of the modulation of TNF-α-induced apoptosis and proliferation by MAPKs. Sci. Signaling **4**, ra16

Received 14 July 2011
doi:10.1042/BST20110633