

Full Title

Computational Modelling of the Receptor Tyrosine Kinase Activated MAPK Pathway

Authors Names

Richard J. Orton^{*}, Oliver E. Sturm^{*}, Vladislav Vyshemirsky^{*}, Muffy Calder[†], David R. Gilbert^{*},
and Walter Kolch^{‡,§}

Establishments

* Bioinformatics Research Centre, Department of Computing Science, University of Glasgow,
Glasgow, G12 8QQ, UK

† Department of Computing Science, University of Glasgow, Glasgow, G12 8QQ, UK

‡ Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK

§ Beatson Institute for Cancer Research, Garscube Estate, Glasgow, G61 1BD, UK

Corresponding Author

Walter Kolch

Beatson Institute for Cancer Research

Garscube Estate

Glasgow G61 1BD, UK

Tel: ++44-141-330 3983

Fax: ++44-141-942 6521

Email: wkolch@beatson.gla.ac.uk

Abstract

The Mitogen Activated Protein Kinase (MAPK) pathway is one of the most important and intensively studied signalling pathways. It is at the heart of a molecular signalling network that governs the growth, proliferation, differentiation and survival of many, if not all cell types. It is deregulated in various diseases ranging from cancer to immunological, inflammatory and degenerative syndromes, and thus represents an important drug target. Over recent years, the computational or mathematical modelling of biological systems has become increasingly valuable and there is now a wide variety of mathematical models of the MAPK pathway which have led to some novel insights and predictions as to how this system functions. In this review, we give an overview of the processes involved in modelling a biological system using the popular approach of ordinary differential equations. Focussing on the MAPK pathway, we introduce the features and functions of the pathway itself before comparing the available models and describing what new biological insights they have led to.

Keywords

Systems Biology; Computational Modelling; Signalling Networks; Receptor Tyrosine Kinase; Mitogen Activated Protein Kinase; Extracellular Signal Regulated Kinase

The general structure of MAPK pathways

In molecular biology the mitogen activated protein kinase (MAPK) pathway is considered to be a paradigm for signal transduction as it occupies a central role in key cellular processes and is evolutionary conserved. Various manifestations of the MAPK pathway are found in all eukaryotic cells so far examined and have been studied extensively in a multitude of organisms, ranging from yeast to man. On the basis of the substantial body of data available in the literature, this pathway has frequently been the system of choice for computational modelling of biological signal transduction over the last decade.

The term “MAPK pathway” refers to a module of three kinases which are activated by sequentially phosphorylating each other in response to a diverse range of stimuli, such as cytokines, growth factors, neurotransmitters, cellular stress and cell adherence. Accordingly, the pathway plays a pivotal role in many key cellular processes, ranging from growth control in all its variations, cell differentiation and survival to cellular adaptation to chemical and physical stress. For reviews see [1-3].

The MAPK pathway employs one of the most generic signalling designs found in biological signal transduction, namely that of a cycle formed by a kinase phosphorylating a target protein and an opposing phosphatase, which is in charge of dephosphorylating the target (Figure 1). This type of protein phosphorylation presents a fundamental mechanism by which the activities of numerous enzymes, receptors, transporters, docking and scaffolding proteins are regulated.

The general layout of the MAPK pathway consists of three kinase/phosphatase cycles built into a three-tiered cascade, consisting of a mitogen activated protein kinase (MAPK), which is activated via phosphorylation by a MAPK-kinase (MAPKK, MKK), which in turn is phosphorylated by a MAPKK-kinase (MAPKKK, MKKK) (Figure 2). MAP-kinases (MAPKs) are deactivated by a family of phosphatases termed mitogen activated protein kinase phosphatases (MKPs) [4].

Most activators of the MAPK pathway initiate signalling by activating receptors in the cell membrane, which assemble into receptor signalling complexes and activate a MAPKKK typically through a small GTPase. Signal transduction within the MAPK pathway module appears to be fairly specific and often is perceived as a pathway with a linear structure. The number of known MAPK effectors, however, is very large and diverse, including mainly transcription factors, protein kinases and cytoskeletal proteins. Upon activation, MAPK can translocate from the cytoplasm to the nucleus, where it regulates gene transcription through affecting chromatin structure and modifying the activity of transcription factors [5]. Eukaryotic cells contain at least 12 different MAPKKKs, 7 MAPKKs, and 8 MAPKs, which can be

attributed to at least four functionally distinct MAPK modules. MAPK modules have evolved by gene duplication and several closely related modules, delivering specific biological responses, are coexpressed in a particular cell [2]. Currently, it is not entirely clear exactly to what extent kinases of different MAPK modules are specific within their module or whether they are able to crosstalk within other MAPK modules or with other targets [3, 6]. However, the fact that only relatively few MAPKs receive and integrate a plethora of extracellular stimuli to control a widely diverse range of cellular processes, implies that considerable versatility and specificity is built into MAPK signalling. Some of the mechanisms which explain how specificity is achieved in MAPK signalling have recently come to light:

- 1) The duration and amplitude of MAPK phosphorylation is critical for the control of distinctive processes, such as proliferation or differentiation in PC12 cells [7] or the activation of c-Fos [8]
- 2) Specific binding interactions, mediated by the common docking domain (CD domain) on MAPKs and scaffolds such as JIP, KSR or MP1, tether several components of a MAPK module into a protein complex [9, 10].
- 3) Crosstalk with other signalling pathways could modulate MAPK signalling, examples include regulation of Raf by cAMP activated protein kinase (PKA) and PDE4 by ERK [11].

The ERK/MAPK pathway

In the following we will focus on the Raf/MEK/ERK pathway, because the majority of computational MAPK models address this particular MAPK module. Signal transduction along the Raf/MEK/ERK pathway (Figure 3) begins with the activation of the small GTPases Ras (and possibly Rap) by receptor tyrosine kinases, G-protein coupled receptors and/or integrins [12]. These membrane proteins assemble large signalling complexes upon activation, which recruit and activate Ras proteins by inducing the exchange of Ras-bound GDP with GTP, converting Ras to its activated conformation. This process is mediated by the interaction of Ras with GDP/GTP-exchange factors, such as SOS (son of sevenless). Deactivation of Ras, on the other hand, is controlled by GTPase activating proteins (GAPs), which significantly enhance the otherwise very low GTPase activity of Ras and thus effectively enhance the hydrolysis of GTP to GDP [13]. Upon activation, the small G-proteins recruit the MKKKs c-Raf (and, if present, A and B-Raf) to the plasma membrane, where Raf is activated in a complex, only partially understood process that involves binding to Ras, phosphorylation and changes in conformation and protein interactions [12]. Raf then activates the MKK/extracellular signal regulated kinases MEK1/2 by phosphorylation of two serine residues. MEKs generally recognize only specific MAPKs as substrates. MEK1/2 phosphorylates ERK1/2 at threonine and tyrosine residues in a “TEY” motif within its activation loop. This dual phosphorylation is correlated with ERK1/2 activity and can be used

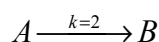
as a measurement for the indirect quantification of ERK activation [14]. ERK is a serine/threonine kinase. Activated ERK can phosphorylate over 80 substrates in the cytoplasm and the nucleus. It can regulate gene expression by directly phosphorylating transcription factors such as Ets, Elk, and Myc, or indirectly by targeting substrates such as p90-RSK (ribosomal S6 kinase) family kinases, which can modify transcription factors and histones [15]. The term MAPK originally was synonymous with ERK, but increasingly is used to indicate the superfamily of different MAPKs. The latter usage is adopted in the text below.

Computational Modelling

Systems biology is concerned with the study of biological systems in terms of their underlying network structure rather than simply their individual molecular components. A biological system can be anything from a simple biochemical reaction cycle, to a gene regulatory network or signal transduction pathway to a cell, tissue or an entire organism. At the heart of systems biology is mathematical modelling - the process of translating a biological system into a mathematical model for subsequent computer simulation and analysis. In order to be useful and applicable to biological questions, models have to faithfully describe the biological system and be able to make predictions about its behaviour. Thus, while the basis of a model is the topological representation of its components and their links, it is the description of the biological system's dynamic behaviour which equips the model with predictive power [16]. However, it is important to emphasise that a model is not a real or exact representation of the biological system; rather, it is a simplified description to assist in analysis and to help us to better understand the real world. A mathematical model can be used to generate new insights, make testable predictions, identify gaps in our biochemical knowledge, test conditions that may be difficult to study in the laboratory and help identify what is right and wrong with our hypotheses, for example, could the proposed mechanism give the observed results? Ultimately, the purpose of a model is to increase our understanding of the real biological system, to identify the key components and processes and to predict biological behaviour such as what effect will a particular drug have on the system, or which components need to be disturbed for the system to behave in a specific way.

Ordinary Differential Equations

One of the most commonly used approaches to modelling biological systems is that of ordinary differential equations (ODEs). The technical definition of a differential equation is an equation involving one or more unknown functions and their derivatives. Essentially, a differential equation describes how a property of interest, such as $[A]$ -- the concentration of A, changes over time; this is usually expressed by describing how the rate of change of the concentration is related to the concentration at that moment. For example, consider the simple reaction below which depicts the decay (or conversion) of A to B:



Equation 1

This reaction is a plain uncatalysed reaction that can be modelled using mass action kinetics; k represents the rate constant of the reaction which in this example is equal to 2 mM/s. Therefore, the reaction proceeds at the following rate:

$$v = k[A] = 2[A]$$

Equation 2

As can be seen, the rate of the reaction (v) is dependent on the concentration of A, the higher the concentration of A the faster the rate will be and therefore the faster A will be consumed and the faster B will be produced. From the above equations it is relatively straightforward to construct differential equations representing the rate of change of the concentration of A and B over time:

$$\frac{d[A]}{dt} = -k[A] \quad \frac{d[B]}{dt} = k[A] \quad \text{Equation 3}$$

In order to simulate the above reaction we also need to know the initial (time zero) concentrations of A and B, which in this example are set to be equal to 5 and 0 mM, respectively. Simulation uses numerical methods to solve the differential equations and approximate the change in concentration of all the species in a system over time. One of the simplest methods of numerical integration is Euler's method, the basic idea of which is to approximate a curve with a series of straight lines tangential to the curve. The rate of change is the same as the slope of a line drawn tangent to the curve at that time; therefore, a series of these tangent lines can be used to follow the function over time. Euler's method is a point-slope type method as it uses an initial concentration (the point) and a differential equation (the slope) to monitor a species concentration over time. In practice this method involves using this initial slope for a very small finite step-size (Δt) in the time direction. The method is then repeated for another step-size using the result from the previous step as the new starting point. For example, the simple reaction above (Equation 1) was simulated using Euler's method over two time steps (Δt) of 0.01 seconds; the calculations used to compute the concentration of A at these two time points is shown below (the extension of this example to longer periods of times is simple):

$$\begin{aligned}
[A(t + \Delta t)] &= [A(t)] - k[A(t)]\Delta t \\
\Delta t &= 0.01 \\
k &= 2 \\
[A(0)] &= 5 \\
[A(0.01)] &= 5 - 2 \times 5 \times 0.01 = 4.9 \\
[A(0.02)] &= 4.9 - 2 \times 4.9 \times 0.01 = 4.802
\end{aligned}$$

Equation 4

Euler's method described above is a very basic method of numerical integration. However, there are many more advanced methods such as the Runge-Kutta, Rosenbrock and Richardson extrapolation; for more information on the different methods of numerical integration methods see [17]. Furthermore, although the above example reaction (Equation 1) is both small and simple, differential equations can easily be used to model large biological systems involving many more species and reactions as well as more complex reaction mechanisms such as Michaelis-Menten kinetics. It is also important to note that when modelling a system comprising several biochemical species, each represented by a differential equation, ODE methods require simultaneous solving of all the differential equations that represent the system.

The Modelling Process

One way to view the actual process of modelling the behaviour of a biological system is to consider it as five distinct steps (Figure 4):

Step 1) System Delimitation: this step involves selecting the biological system to be modelled, e.g. the EGF or NGF activated ERK cascade. More importantly it also includes identifying the biological question the model aims to answer, e.g. how do EGF and NGF produce differing responses in ERK activation?

Step 2) Definition: this is the key step in the modelling process and is by no means trivial. It involves defining the model to represent the biological system of interest. Essentially, this involves drawing a detailed topological chart of the system that shows all the species (e.g. proteins) involved, what reactions they can participate in and where. The kinetic types for each reaction then need to be defined and the parameter data (e.g. rate constants and initial concentrations) assigned to give a set of detailed kinetic reactions. This definition step can involve exhaustive searches of the scientific literature to see what is already known about the system and what parameter data are already available as well as performing laboratory experiments to provide data or the use of computational techniques to estimate missing parameters (for example: [18]). It is important to note that many biological processes are very complex and not fully understood. Therefore, defining a model often involves making simplifying assumptions that reduce complex and poorly understood processes into simpler

ones that can still represent the biological processes well enough to explain the observed data.

Step 3) Simulation: after the model has been defined, the next step is to translate the kinetic reactions into a set of differential equations that describe how the concentration of each species in the system changes over time; this set of differential equations is then simulated (or solved) over a desired period of time. This is a relatively straightforward step as there are many software tools available for the generation and simulation of models of biological systems that are based on differential equations (see below for more details on simulation tools).

Step 4) Validation: simulating a model typically returns a table of data or curve showing how each species' concentration varies over time. This data must then be validated against available experimental data. If the model behaves as the experimental data suggests then the model can be analysed further. If it does not, then the definition step must be revisited where the model is checked for errors such as incorrect kinetics or parameter data, over-simplifications of processes and perhaps missing components. In essence, this is a 'debugging' loop involving model definition, simulation and validation where the model is refined in order to obtain behaviour which conforms with experimental observations. Sometimes the refinement already reveals useful information about the system. For instance, refining the mathematical model of STAT5 phosphorylation in response to erythropoietin receptor stimulation uncovered that the model can capture the experimental kinetics of STAT5 phosphorylation when a time delay was introduced to account for the shuttling of STAT5 between the nucleus and cytosol. This indicated that STAT5 cycles between the nucleus and cytosol, which subsequently was proven experimentally [19]. The validation step is crucial if a reliable model is to be generated; if the model's results do not match known biology, we cannot trust predictions about unknown biology.

Step 5) Analysis: after the model has been validated it can be analysed and the simulation results interpreted. Analysis can come in various forms from the simple examination of the species' concentration graphs to more complex statistical analyses. Sensitivity analysis is a commonly used approach that studies the response of system variables to changes in parameter values and can therefore be used to identify the key reactions and species as well as monitoring how robust a model is. Essentially, sensitivity analysis works by varying the data for a parameter by a small amount and analysing what effect it has on a specific system variable, such as the peak height or duration of the phosphorylated ERK signal; a small change to a key parameter data is likely to have a large effect on the system variable. Different parameters can have widely different sensitivities and the sensitivity of a specific parameter can also vary depending on which system variable is considered. On the other hand, sensitivity analysis can also be used to estimate the impact of uncertainty in

parameters on system variables. Robustness is a very important factor in biological systems as it allows a system to absorb fairly large perturbations and still function reasonably well; this is because the functionally important behaviour of a system has a certain degree of resilience to damage. If a system variable has a low sensitivity with respect to a parameter, it is robust to alterations in that parameter, however caused. The structural robustness of a model can also be analysed by monitoring how it performs when parts of it are removed; for example, how does the system behave if a specific specie, reaction or entire pathway is removed. This can be useful because there is often redundancy in biological systems where multiple pathways are available for the production or activation of a certain protein. Overall, the analysis step aims to generate new predictions and hypotheses about biological processes that were not known or unproven before thus increasing our overall understanding of the system itself.

Computational Tools

Currently, there a number of software tools available for the simulation and analysis of differential equation based models of biological systems such as Gepasi [20], E-CELL [21], Virtual Cell [22], GENESIS [23] combined with Kinetikit [24], Jarnac combined with JDesigner [25], Mathematica (Wolfram Research, <http://www.wolfram.com/>) and Matlab (Mathworks, www.mathworks.com); for a recent review of simulation tools see [26]. The majority of these tools have a graphical interface that permits the user to enter the biochemical reactions and kinetic constants which the tool then uses to automatically generate the corresponding ODEs and simulate the model. Some have a number of advanced features to visualise and analyse models, display simulation results and also estimate missing parameter data. Furthermore, the majority of tools now support the Systems Biology Markup Language (SBML; www.sbml.org) [27] which is concerned with introducing a standard representation of models of biological systems. This enables models to be shared, evaluated and developed cooperatively as well as enabling the use of multiple tools without having to rewrite models for each tool; a full list of SBML compatible modelling tools is available from the SBML website. A recommendable website for the novice is the pathway model repository of the Silicon Cell project (<http://www.jjj.bio.vu.nl/index.html>; [28]). Via an interactive webpage, pathway models can be viewed and the effects of changes in kinetic parameters can be simulated.

ERK Models

The ERK cascade is one of the most important cell signalling pathways and has been the subject of intensive study in the laboratory and more recently through mathematical modelling techniques. Early mathematical models of the ERK cascade focused on investigating the properties and behaviour of the core cascade itself. The first model was published in 1996 by Huang & Ferrell [29] and showed that the ERK cascade exhibited ultrasensitivity, i.e. a non

linear sigmoid activation curve, with the degree of ultrasensitivity increasing as one moves down the cascade. This was quickly followed by two models in 1997 [30, 31] that showed that the activating dual phosphorylation of ERK itself was accomplished via a two-collision, distributive mechanism whereby MEK phosphorylates one site, dissociates and then has to rebind to phosphorylate the second site. This generates a pool of largely single phosphorylated, i.e. inactive ERK molecules, which appears as a gentle response curve. When enough single phosphorylated ERK molecules have accumulated, most further phosphorylation events produce double phosphorylated, i.e. active ERK, causing the slope of the activation curve to increase sharply and steeply. This provided a mechanistic basis for the ultrasensitivity of ERK activation and explained how ERK can convert graded inputs into switch-like outputs [31]. Then in 1998, Ferrell and Machleder [32] using *Xenopus* oocytes showed that because of ultrasensitivity, ERK is activated essentially in an all-or-none fashion in individual cells when they are treated with increasing concentrations of progesterone. Thus, the apparently graded concentration dependent response curve observed when a whole cell population was analysed, actually was composed of increasing numbers of responders versus non-responders on the level of the individual cells. Over the past decade, an ever increasing number of models of the ERK cascade have been developed, growing in both size and complexity through the years. Models now routinely incorporate growth factor receptors and the plethora of adaptor proteins that can bind to them and subsequently activate the ERK cascade. Currently, there are over thirty mathematical models that in some way incorporate the ERK cascade (Figure 5). These models have been used to investigate various aspects of the biological behaviour of this system such as bistable feedback loops [24, 33], oscillations [34], feedback inhibition [35], autocrine loops [36, 37], scaffold proteins [38, 39], feedback effects [40], temperature dependence [41], receptor internalisation [42], signal specificity [43], receptor expression [44], robustness [45], crosstalk [46], receptor trafficking [47, 48], memory [49], bistability and hysteresis [50], Ras activation [51], receptor regeneration [52], receptor comparison [53] and temporal dynamics [54]. For a recent review of the relationships between some of these ERK models see [55] or for more general reviews of models of cell signalling pathways see [56-58].

The most common growth factor receptor that is currently incorporated into models of the ERK cascade is the Epidermal Growth Factor Receptor (EGFR); for a recent review of models of the EGFR system itself see [59]. This is because the EGFR system has been well studied, is present at substantial levels in various cell types, and good antibodies and molecular reagents are widely available enabling a range of quantitative studies to be performed. We have selected three popular models of the ERK cascade encompassing the EGFR system for discussion in detail below; we review what each model considers and more importantly what biological insights and predictions they have led to. Our selection of models is a good representation of the existing models; they are spread over the timeline, are ODE based and represent the same biological system and are therefore directly comparable.

Additional information on the models including links to simulation files is available at: <http://www.brc.dcs.gla.ac.uk/~rorton/mapk/>.

Model 1: Kholodenko *et al.* 1999

Kholodenko *et al.* (1999) [60] developed an ODE based mathematical model of the EGFR signalling network to investigate the short term pattern of cellular responses to EGF in isolated rat hepatocytes. The model consists of 25 reactions involving 23 different species (Figure 6) and includes three adaptor proteins that can directly interact with phosphotyrosine residues on EGFR (namely Shc, Grb2 and phospholipase C gamma [PLC γ]). The kinetic parameters in the model were based on the scientific literature and/or derived from basic physical-chemical quantities. In order to effectively validate the model before analysis, a number of wet laboratory experiments were performed such as time courses of EGFR phosphorylation and EGF-induced tyrosine phosphorylation of adaptor proteins. The simulation was then compared to these data to show that the model gives a good fit to the experimentally observed time courses.

Analysis of the model showed that the rapid, and short-lived pattern of EGFR phosphorylation can be explained by the fact that bound adaptor proteins protect receptor phosphotyrosine residues against dephosphorylation by tonically active phosphatases, rather than having to invoke the activation of tyrosine phosphates by the receptor [61]. The protection of phosphotyrosines from dephosphorylation is transient as bound adaptors turn over and allow dephosphorylation of receptors. This kinetic model also explains why the levels of phosphorylated adaptors such as Shc, which becomes phosphorylated when recruited to the receptor, stays high even when receptor phosphorylation already has declined. Sensitivity analysis of the model showed that the dynamics of the EGFR signalling pathway appeared to be robust to significant changes in many of the rate constants of the protein interactions involved. Interestingly, however, the time course of phosphorylation/activation responses to EGF appeared to be more sensitive to variations in the relative concentrations of adaptor proteins than to most variations of the kinetic constants. In particular, the Shc:Grb2 ratio was suggested to be an important controlling factor of the kinetics of the EGFR signalling response. This finding can be explained by the competition between adaptor molecules during the formation of multiprotein signalling complexes. It also highlights that changes in the expression levels of adaptor proteins are potent modulators of EGF receptor signalling providing a rational basis for the observation that many adaptor proteins when overexpressed, or stabilised through mutations, can usurp the EGFR mitogenic signalling pathways efficiently and act as oncogenes. Obviously, this finding also has bearings for normal physiology because various cell types and also cells in different functional states can show considerable variation in the abundance of signalling proteins.

Technically, this model does not include the core ERK cascade of Raf, MEK and ERK in its description as it only goes down as far as the Ras guanine nucleotide exchange factor SOS. However, this model was a pacemaker for the field in several ways. It included the feedback between theoretical prediction and experimental validation, which now is deemed essential for modern systems biology. It also was one of the first models to incorporate the EGF receptor with its associated adaptor proteins and it does predict a transient recruitment of SOS to EGFR at the plasma membrane where Ras is located. This transient recruitment of SOS is therefore predicted to give rise to a transient activation of Ras and the ERK cascade as expected for an EGF response. No wonder, this model has been used as a basis for many other models of the EGFR system which do include the core ERK cascade (for example: [42, 46]).

Model 2: Brightman & Fell (2000)

Brightman & Fell (2000) [35] developed an ODE based mathematical model of the EGF signal transduction pathway in PC12 cells to investigate the factors influencing the kinetics of ERK cascade activation. Their model consists of 30 reactions involving 29 species (Figure 7) and includes a self contained module of the activation and internalization of EGF receptors induced by EGF. Through the phosphorylation of Shc, activated receptors can then initiate an intracellular signal transduction pathway that results in the activation of Ras and ultimately a cytosolic ERK cascade comprising Raf, MEK and ERK. Feedback regulation of the pathway is mediated by the inhibitory phosphorylation of SOS which causes the dissociation of the Shc/Grb2/SOS complex. The kinetic constants of reactions and initial concentrations of species were largely based on a range of measured or estimated values published in the existing scientific literature.

The analysis of this model indicated that negative feedback inhibition of the ERK cascade was the most important factor in determining whether the cascade activation was transient or sustained, and that differences in feedback regulation were likely to underlie the characteristic patterns of EGF and NGF induced ERK activation in PC12 cells. In the model, EGF initially activates Ras and the ERK cascade via formation of the Shc/Grb2/SOS signalling complex. This signal is rapidly terminated through the negative feedback phosphorylation of SOS, resulting in the dissociation of the Shc/Grb2/SOS complex. However, the analysis showed that the dephosphorylation of SOS (reaction number 28 in Figure 5) was one of the most important steps in determining the duration of the signal and that only a 40 fold increase in the rate of this reaction generated a time course of ERK cascade activation similar to that observed when PC12 cells are stimulated with NGF rather than EGF. This suggests that NGF, but not EGF, could enhance phosphatase activity towards phosphorylated SOS, resulting in sustained signalling through the Shc/Grb2/SOS complex and therefore a sustained activation of MEK and ERK. This model drew attention to the importance of the

regulation of the Ras family proteins as critical regulators of the kinetics of ERK signalling. An elegant and elaborate analysis confirming and significantly extending these results was recently presented [54], showing that the activation kinetics of Ras and its relative Rap1 can account for the differences in ERK activation in response to EGF versus NGF.

Model 3: Schoeberl et al. (2002)

Schoeberl *et al.* (2002) [42] developed an ODE based mathematical model describing the dynamics of the EGF signal transduction pathway to investigate the effects of receptor internalization on the ERK cascade, and also the signal-response relationship between the binding of EGF to its receptor at the cell surface and the activation of downstream proteins in the signalling cascade. Their model consists of 125 reactions involving 94 species (Figure 8) advancing model building by including two principal pathways of Ras activation, Shc-dependent and Shc-independent, as well as EGFR internalisation by endocytosis. Receptor internalisation is comprehensively represented in the model with receptor species able to become internalised via two distinct routes. The majority of kinetic parameters were based on values published in the scientific literature and initial concentrations were either compiled from the literature or based on laboratory experiments.

This model is one of the most comprehensive available as it includes a large range of dynamic processes. A main conclusion of the modelling is that it is the initial rate of change of receptor activation that determines the cellular response to EGF. Varying the concentration of EGF, analysis of the model suggested that the cell maintains a high sensitivity over a relatively broad EGF concentration range and that the initial velocities of EGFR activation rather than the peak maxima are important for signal propagation. The model was also used to investigate the roles of internalized and cell-surface receptors in generating a cellular response leading to the conclusion that EGF receptor internalization has a dual role: (1) signal attenuation by protection from prolonged external EGF stimulation at high EGF concentrations; and (2) signal amplification after internalization at low EGF concentrations. The analysis of sensitivity showed that the model was robust to variation in the parameters and initial conditions.

Model Comparison

It is interesting to note that although the three models discussed above are describing the same biochemical system, namely the EGF regulated ERK pathway, they are all different and yet seem all to be able to explain the observed data, and also make interesting predictions about the system's behaviour. This leads to the inevitable question of how is it possible for different models of the same system to all be both correct and reliable? This is a very important question, as the implications are enormous. Does it condemn modelling into the

realm of reproducible artefacts, or does it demonstrate that modelling can give valid answers even in the light of different approaches, incompleteness and imperfect data? We do not know the definite answer yet, but all indications point to the latter possibility. Although the three models do differ, they do not do so to a great extent and many of the differences can be accounted for by simplifications and where the model boundaries lie; a direct comparison of the size and features of the three models is presented in Table 1.

One interesting aspect is that the Brightman & Fell (2000) [35] model utilises a different strategy of dealing with activated receptors when compared to the Kholodenko *et al.* (1999) [60] and Schoeberl *et al.* (2002) [42] models. In the Brightman & Fell (2000) model, activated receptors simply catalyse the phosphorylation of Shc. However, the other two models utilise a more realistic receptor complex strategy where adaptor proteins such as Shc have to bind to activated receptors and stay bound in order for the signal to propagate. These two different strategies have led to different predictions as to the basis of the transient signal response to EGF. In the Brightman & Fell (2000) model, Shc is phosphorylated by activated receptors enabling it to bind Grb2 and SOS before activating Ras and the core ERK cascade; the Shc/Grb2/SOS complex is a functional complex and does not need to bind to the receptor. The transient nature of the EGF response is in part caused by the negative feedback phosphorylation of SOS by activated ERK (ERKPP) resulting in the dissociation of the Shc/Grb2/SOS complex and therefore stopping the activation of Ras and the core ERK cascade. Whereas in the Kholodenko *et al.* (1999) model, Shc followed by Grb2 and SOS rapidly bind to activated receptors, which would enable the activation of the membrane bound Ras and the core ERK cascade. However, the receptor bound Shc/Grb2/SOS complex then dissociates and there is a build up of this unbound complex. As unbound receptors are exposed to phosphatases and are rapidly deactivated, a transient response can result without the need for a negative feedback loop.

The boundary of a model is essentially the point or points at which the model stops; the model does not consider and therefore define the biological processes beyond these points. The decision of where the boundary lies is typically determined by what biological processes and questions the model is intended to investigate and answer. A good illustration of model boundaries is given by the comparison of the Kholodenko *et al.* (1999) model to the Schoeberl *et al.* (2002) model. Both models have a similar general structure as they essentially have the same Shc-dependent and Shc-independent pathways, and utilise a receptor complex strategy. However, the Kholodenko *et al.* (1999) model includes the PLC γ pathway whereas the Schoeberl *et al.* (2002) model includes the ERK cascade and receptor internalisation. This is because the Kholodenko *et al.* (1999) model was designed to investigate the short term pattern of cellular responses to EGF through EGFR and its adaptor proteins; therefore only the receptor proximal events were included. In contrast, the Schoeberl *et al.* (2002) model was designed to investigate the downstream effects of EGFR activation as

well as the effects of receptor internalization on the ERK cascade. This illustrates an experimental design strategy for reductionism that is familiar to the biologist: depending on the question that is the focus of the investigation, the experimental system is simplified in the way most appropriate to address the question at hand.

Simplifications in a model can come in many forms and typically involve the simplification of biological processes or events to reduce the number of reactions needed in the model whilst still being able to represent the biological processes well enough to explain the observed data. The classic example of a simplification is the activation of Raf by Ras-GTP. Ras-GTP recruits Raf from the cytosol to the plasma membrane, where it is activated through a still not completely known process that involves interaction with adaptor proteins, lipids and changes in phosphorylation. However, this complex process has been effectively represented in many models of the ERK pathway as a simple two step process: $\text{Ras-GTP} + \text{Raf} \leftrightarrow \text{Ras-GTP/Raf} \rightarrow \text{Ras-GTP} + \text{Rafx}$. Other simplifications can be found when examining, for example, the Schoeberl *et al.* (2002) model: (1) Only one member of the EGFR family is considered; (2) EGFR dimers are considered as single molecules; (3) The binding of the adaptor proteins Shc and Grb2 to EGFR is assumed to be competitive; and (4) GAP must be bound to the EGFR before any other adaptor proteins can bind.

There may also be some biological inaccuracies in the models discussed. In the Brightman & Fell (2000) model, single phosphorylated MEK is able to phosphorylate ERK and ERKP; this is probably irrelevant as single phosphorylated MEK species have not been observed. Moreover, mutating one of the two phosphorylation sites in MEK abrogates activation suggesting that only double phosphorylated MEK is active [62]. In the Schoeberl *et al.* (2002) model there is an apparent inactive active form of Ras called Ras-GTP* (species 43/71 in Figure 8). The role of this Ras-GTP* is to limit the number of Raf molecules Ras-GTP can activate to one. However, this is probably an artificial and incorrect assumption as Ras is either active in its GTP bound state or inactive in its GDP bound state. There is no in-between, and there is no evidence that the number of Raf molecules Ras-GTP can activate is limited (for a review of Raf activation, see [63]). Although, these are not major errors it could be viewed as a concern that models with apparent errors can still explain the observed data and be used to make valid predictions about the biological system. However, as robustness seems to be an inherent and ubiquitous property of biological systems [64] the robustness of the models to tolerate small errors may simply reflect the biological property. From a pragmatic point of view, it could be viewed as an advantage as models, despite not being perfect, can be used to suggest interesting new hypotheses and explanations for the observed data that challenge our current understanding, which is by no means complete. Importantly, these predictions need to be verified experimentally in the laboratory.

After a model has been completed it can easily be modified or expanded upon and used in the development of new models of the same or related biological systems. For example, the Kholodenko *et al.* (1999) model has been used as a basis for many other models of the EGFR system [41, 46, 48, 51]. The Brightman & Fell (2000) model has since been analysed further by [65] who investigated the effects of a number of MEK inhibitors in this system, and it was also used in the development of other models of the ERK cascade [44, 52]. The Schoeberl *et al.* (2002) model has since been analysed further in a number of studies [66-69], for example, [66] investigated the relevance of the Shc-dependent and Shc-independent pathways; it has also been combined with data on metalloprotease activation to build a model of autocrine signal transduction by cancer cells exposed to ionizing radiation [70].

Alternative Receptor Systems

Although the EGFR is the most common growth factor receptor system modelled together with the ERK cascade, other growth factor receptors have also been successfully modelled. Bhalla *et al.* (2002) [33] developed an ODE based model of the Platelet Derived Growth Factor (PDGF) activated ERK cascade revealing novel and interesting design principles of this network, in particular that ERK activation can operate in different states depending on the history of the cell. The first application of PDGF to a naive cell induces a bistable, switchlike ERK activation response, where even a brief stimulus results in sustained ERK activity. As part of this response the expression of MKPs is induced. Thus, when the cells are restimulated with PDGF after the initial ERK activation has returned to baseline, there is a higher level of MKPs in the cell, causing ERK activation to proceed in a monostable fashion with the ERK activity increasing proportionally with the dose of PDGF. Qiu *et al.* (2004) [52] developed an ODE based model of the NGF activated ERK cascade and suggested that the sustained behaviour of the response was mainly due to a continual regeneration of NGF receptors. Yamada *et al.* (2004) [53] developed ODE based models of the EGF and Fibroblast Growth Factor (FGF) activated ERK cascade and proposed that the protein FRS2 plays a key role in generating the sustained behaviour of the FGF response by recruiting more SOS to the plasma membrane; furthermore, they found that the negative feedback system in the model did not profoundly affect the time course of ERK activation. In one of the most recent models, Sasagawa *et al.* (2005) [54] developed an extensive ODE based model of the EGF and NGF activated ERK cascade in PC12 cells. This model is amongst the most comprehensive to date as it includes both the Ras and Rap1 pathways to ERK activation and combines theoretical predictions with experimental validation. The model was used to investigate how EGF and NGF encode transient and sustained dynamics of ERK activation, respectively. A salient finding was that the transient activation of ERK depended on the rate of receptor activation rather than the concentration of the growth factors. In contrast, sustained ERK activation by NGF depended on the final concentration of NGF, but not the temporal rate of increase in NGF concentration. These diverse response modes are due to differences in

Ras and Rap1 inactivation. Ras, which mediates transient ERK activation, is inactivated by the receptor induced recruitment of Ras-GAP to the membrane resulting in a stringent temporal regulation of Ras activity. In contrast, Rap1, which in this model is responsible for sustained ERK activation via stimulation of B-Raf, is inactivated by Rap-GAP, which is not regulated by growth factors, but functions constitutively. Thus, the activation of Rap1 becomes a function of the NGF concentration. In this model the Ras and Rap1 pathways capture the temporal rate of increase and concentration of growth factors and encode these distinct properties into transient and sustained ERK activation, respectively.

Alternative Modelling Methods

Although ODEs are commonly used to model biological systems such as the ERK cascade they have one major drawback and that is they are reliant on high frequency sampling and absolute parameter data being available, such as detailed kinetic rates and absolute initial concentrations. However, a lot of the data generated by biologists, including data generated from high-throughput techniques, are not directly amenable to modelling as they often contains sparse time series, are qualitative rather than quantitative and show relative changes rather than changes in absolute concentrations. Furthermore, as there is only very little standardisation of measurements, data from different laboratories usually can only be compared in a semiquantitative or qualitative fashion [16]. However, although high frequency and absolute data is required if one wants to produce a near exact replica of the experimental system, it is often not necessary to know every single parameter with high accuracy due to issues such as variable sensitivity. Therefore, ODE based models can readily be used to assess whether the system is capable of showing specific qualitative features such as oscillations. In addition, there are also a number of techniques to estimate missing parameter data in a model (for example: [18]) which typically work by varying the missing parameters values until the expected behaviour is obtained. An alternative method developed by Brown *et al.* (2004) [71] used an ensemble approach to model complex signalling networks, namely the NGF and EGF activated ERK cascade; these models also included both the Ras and Rap1 pathways. Instead of using kinetic parameters, which are commonly not available, the ensemble method was used to match the model to experimental time courses of the activities of signalling molecules which then generates an ensemble of weighted parameters which can then be used to analyse the model. This model was used to evaluate the importance of different regulatory loops in generating a sustained activation of ERK. Furthermore, this approach suggested that only a small fraction of parameter combinations are likely to be well constrained and that the few well-constrained parameters reveal critical focal points in the signalling network. Two additional approaches are Flux Balance Analysis (FBA; [72]) and Metabolic control analysis (MCA; [73]). FBA is an approach to constrain a metabolic network based on the stoichiometry of the metabolic reactions and does not require kinetic information. MCA is a quantitative sensitivity analysis of fluxes and concentrations; the

relative control exerted by each step on a system variable is measured by applying a perturbation to the step and measuring the effect on the variable of interest after the system has settled to a new steady state. It is important to note that FBA and MCA are essentially analyses of the steady state, and are therefore less suited to the dynamic aspects of signal transduction; however, MCA has recently been extended to the dynamics of signal transduction [74].

There are also a number of alternative approaches to ODEs that can be used to model and analyse biological systems. Resat *et al.* (2003) [48] developed one of the largest models of the EGFR system to date using a probability weighted-dynamic Monte Carlo stochastic simulation. They developed an integrated model of both the trafficking and signalling components of the EGFR system that consisted of hundreds of distinct endocytic compartments and ~13,000 reactions that occurred over a broad spatio-temporal range. The signalling component of this model consisted of the ODE model developed by Kholodenko *et al.* (1999) [60], highlighting that ODE based models can be utilised by other modelling techniques. Pi-calculus [75] has successfully been used to model biological systems, with molecules and their domains represented by computational processes, and reactions by communication and rearranging the communication channels. Further developments in this area include stochastic Pi-calculus [76] and the hybrid system BioSPi (<http://www.wisdom.weizmann.ac.il/~biospi/>).

Petri nets are another class of innovative modelling notations used to analyse biological systems [77]. In this approach biological networks are represented by intuitively readable but strictly formalized graphical diagrams (of molecules and connecting reactions) that are directly subjected to algebraic analysis. For example, Oliveira *et al.* (2004) [78] constructed an algebraic-combinatorial model of the SOS compartment of the EGFR system by using a Petri net approach. Extended petri net approaches that incorporate both discrete and continuous state transitions (hybrid function Petri nets) combine the powerful simulation capabilities of differential equations with the formal logical analysis available in the Petri net computations [79-83]. The Biochemical Abstract Machine BIOCHAM [84, 85] is a programming environment for modelling biochemical systems, making simulations and querying the model in temporal logic, which allows the behaviour of a model to be checked against biological predictions expressed in logical statements. The interface is based on a simple language for representing biochemical networks. BIOCHAM provides mechanisms to reason about the reachability of certain states, the existence of stable states, and some types of temporal behaviour (e.g. oscillations). An alternative algebraic modelling and analysis approach is proposed in [86] where the stochastic process algebra PEPA was used to model the ERK signalling pathway. The main advantage of algebraic modelling techniques such as these lies in their ability to systematically reason over structural properties such as the interaction of sub-networks (for example crosstalk) and the behavioural equivalence of different networks.

Future Directions

Signalling pathways have traditionally been drawn as separate linear entities reflecting the history of how they were discovered rather than their functional context. However, signalling pathways are extensively interconnected and embedded in networks with common protein components and a multitude of links and crosstalk between pathways. Due to the complexity of these networks, computational methods are required in order to explain in detail how they function and predict possible behaviours. Over recent years, the computational or mathematical modelling of biological systems has become increasingly valuable and can provide useful information to understand their behaviour. The ERK cascade is currently one of the most popular systems to be modelled as it is one of the most intensely studied signalling pathways and has also been implicated in various diseases. Currently, there are many differential equation based models of the ERK cascade that have been used to investigate various aspects of its biological behaviour and have led to some novel insights and predictions as to how this system functions. However, no two models appear to be the same topologically or dynamically as they consider different biological processes and components, use different representations and have different kinetic properties. Over the years, models have increased in both size and complexity. The first model of the ERK cascade [29] only considered the core cascade itself, whereas the most recent model [54] considered both the EGF and NGF receptors, numerous adaptor proteins and both the Ras and Rap1 pathways leading to the activation of the ERK cascade, as well as numerous feedback loops.

Coupled with the increase in the number, size and complexity of mathematical models is an increase in the number of techniques and software tools available to simulate and analyse them. Although differential equation methods are currently the most widely used, there are a number of good alternatives available and a number of promising alternatives in development. The adoption of some sort of unifying standard, such as SBML, for all published models of biological systems would now be a welcome development enabling models to be shared and evaluated with ease and thus eliminating the need to painstakingly recreate models based on static tables of supplementary data and compared to simulation graphs in papers. To this end, the BioModels database (www.ebi.ac.uk/biomodels; [87]) was recently launched and aims to be a curated database for the deposition of models of biological systems in an SBML format. Similarly, the recently founded Receptor Tyrosine Kinase (RTK) Consortium (<http://www.rtkconsort.org>) is an international research effort to advance the understanding of RTK signalling systems including the MAPK pathways through a combination of mathematical modelling and quantitative biological experimentation. Thus, with the first steps accomplished we now can expect quantitative biology to rapidly gain momentum.

Acknowledgements

The authors acknowledge the support through the Department of Trade and Industry (DTI) Beacons Project "BPS: Biochemical Pathway Simulator".

References

- 1 Cobb, M. H. (1999) MAP kinase pathways. *Prog Biophys Mol Biol* **71**, 479-500
- 2 Widmann, C., Gibson, S., Jarpe, M. B. and Johnson, G. L. (1999) Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human. *Physiol Rev* **79**, 143-180
- 3 Chang, L. and Karin, M. (2001) Mammalian MAP kinase signalling cascades. *Nature* **410**, 37-40
- 4 Farooq, A. and Zhou, M. M. (2004) Structure and regulation of MAPK phosphatases. *Cell Signal* **16**, 769-779
- 5 Murphy, L. O., MacKeigan, J. P. and Blenis, J. (2004) A network of immediate early gene products propagates subtle differences in mitogen-activated protein kinase signal amplitude and duration. *Mol Cell Biol* **24**, 144-153
- 6 Hindley, A. and Kolch, W. (2002) Extracellular signal regulated kinase (ERK)/mitogen activated protein kinase (MAPK)-independent functions of Raf kinases. *J Cell Sci* **115**, 1575-1581
- 7 Stork, P. J. (2002) ERK signaling: duration, duration, duration. *Cell Cycle* **1**, 315-317
- 8 Murphy, L. O., Smith, S., Chen, R. H., Fingar, D. C. and Blenis, J. (2002) Molecular interpretation of ERK signal duration by immediate early gene products. *Nat Cell Biol* **4**, 556-564
- 9 Tanoue, T. and Nishida, E. (2002) Docking interactions in the mitogen-activated protein kinase cascades. *Pharmacol Ther* **93**, 193-202
- 10 Pouyssegur, J. and Lenormand, P. (2003) Fidelity and spatio-temporal control in MAP kinase (ERKs) signalling. *Eur J Biochem* **270**, 3291-3299
- 11 Houslay, M. D. and Kolch, W. (2000) Cell-type specific integration of cross-talk between extracellular signal-regulated kinase and cAMP signaling. *Mol Pharmacol* **58**, 659-668
- 12 Kolch, W. (2000) Meaningful relationships: the regulation of the Ras/Raf/MEK/ERK pathway by protein interactions. *Biochem J* **351 Pt 2**, 289-305
- 13 Bernards, A. (2003) GAPs galore! A survey of putative Ras superfamily GTPase activating proteins in man and *Drosophila*. *Biochim Biophys Acta* **1603**, 47-82
- 14 Yung, Y., Dolginov, Y., Yao, Z., Rubinfeld, H., Michael, D., Hanoch, T., Roubini, E., Lando, Z., Zharhary, D. and Seger, R. (1997) Detection of ERK activation by a novel monoclonal antibody. *FEBS Lett* **408**, 292-296
- 15 Davie, J. R. and Spencer, V. A. (2001) Signal transduction pathways and the modification of chromatin structure. *Prog Nucleic Acid Res Mol Biol* **65**, 299-340
- 16 Kolch, W., Calder, M. and Gilbert, D. (2005) When kinases meet mathematics: the systems biology of MAPK signalling. *FEBS Lett* **579**, 1891-1895
- 17 Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (2002) Numerical recipes in C++: the art of scientific computing. Cambridge University Press
- 18 Mendes, P. and Kell, D. (1998) Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics* **14**, 869-883
- 19 Swameye, I., Muller, T. G., Timmer, J., Sandra, O. and Klingmuller, U. (2003) Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. *Proc Natl Acad Sci U S A* **100**, 1028-1033
- 20 Mendes, P. (1993) GEPASI: a software package for modelling the dynamics, steady states and control of biochemical and other systems. *Comput Appl Biosci* **9**, 563-571
- 21 Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T. S., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J. C. and Hutchison, C. A., 3rd (1999) E-CELL: software environment for whole-cell simulation. *Bioinformatics* **15**, 72-84

- 22 Schaff, J., Fink, C. C., Slepchenko, B., Carson, J. H. and Loew, L. M. (1997) A general computational framework for modeling cellular structure and function. *Biophys J* **73**, 1135-1146
- 23 Bower, J. M. and Beeman, D. (1998) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System*, 2nd edn.
- 24 Bhalla, U. S. and Iyengar, R. (1999) Emergent properties of networks of biological signaling pathways. *Science* **283**, 381-387
- 25 Sauro, H. M., Hucka, M., Finney, A., Wellock, C., Bolouri, H., Doyle, J. and Kitano, H. (2003) Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration. *Omics* **7**, 355-372
- 26 Pettinen, A., Aho, T., Smolander, O. P., Manninen, T., Saarinen, A., Taattola, K. L., Yli-Harja, O. and Linne, M. L. (2005) Simulation tools for biochemical networks: evaluation of performance and usability. *Bioinformatics* **21**, 357-363
- 27 Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., Arkin, A. P., Bornstein, B. J., Bray, D., Cornish-Bowden, A., Cuellar, A. A., Dronov, S., Gilles, E. D., Ginkel, M., Gor, V., Goryanin, I., Hedley, W. J., Hodgman, T. C., Hofmeyr, J. H., Hunter, P. J., Juty, N. S., Kasberger, J. L., Kremling, A., Kummer, U., Le Novere, N., Loew, L. M., Lucio, D., Mendes, P., Minch, E., Mjolsness, E. D., Nakayama, Y., Nelson, M. R., Nielsen, P. F., Sakurada, T., Schaff, J. C., Shapiro, B. E., Shimizu, T. S., Spence, H. D., Stelling, J., Takahashi, K., Tomita, M., Wagner, J. and Wang, J. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524-531
- 28 Olivier, B. G. and Snoep, J. L. (2004) Web-based kinetic modelling using JWS Online. *Bioinformatics* **20**, 2143-2144
- 29 Huang, C. Y. and Ferrell, J. E., Jr. (1996) Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc Natl Acad Sci U S A* **93**, 10078-10083
- 30 Burack, W. R. and Sturgill, T. W. (1997) The activating dual phosphorylation of MAPK by MEK is nonprocessive. *Biochemistry* **36**, 5929-5933
- 31 Ferrell, J. E., Jr. and Bhatt, R. R. (1997) Mechanistic studies of the dual phosphorylation of mitogen-activated protein kinase. *J Biol Chem* **272**, 19008-19016
- 32 Ferrell, J. E., Jr. and Machleder, E. M. (1998) The biochemical basis of an all-or-none cell fate switch in *Xenopus* oocytes. *Science* **280**, 895-898
- 33 Bhalla, U. S., Ram, P. T. and Iyengar, R. (2002) MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. *Science* **297**, 1018-1023
- 34 Kholodenko, B. N. (2000) Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur J Biochem* **267**, 1583-1588
- 35 Brightman, F. A. and Fell, D. A. (2000) Differential feedback regulation of the MAPK cascade underlies the quantitative differences in EGF and NGF signalling in PC12 cells. *FEBS Lett* **482**, 169-174
- 36 Shvartsman, S. Y., Hagan, M. P., Yacoub, A., Dent, P., Wiley, H. S. and Lauffenburger, D. A. (2002) Autocrine loops with positive feedback enable context-dependent cell signaling. *Am J Physiol Cell Physiol* **282**, C545-559
- 37 Maly, I. V., Wiley, H. S. and Lauffenburger, D. A. (2004) Self-organization of polarized cell signaling via autocrine circuits: computational model analysis. *Biophys J* **86**, 10-22
- 38 Levchenko, A., Bruck, J. and Sternberg, P. W. (2000) Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. *Proc Natl Acad Sci U S A* **97**, 5818-5823
- 39 Somsen, O. J., Siderius, M., Bauer, F. F., Snoep, J. L. and Westerhoff, H. V. (2002) Selectivity in overlapping MAP kinase cascades. *J Theor Biol* **218**, 343-354
- 40 Asthagiri, A. R. and Lauffenburger, D. A. (2001) A computational study of feedback effects on signal dynamics in a mitogen-activated protein kinase (MAPK) pathway model. *Biotechnol Prog* **17**, 227-239
- 41 Moehren, G., Markevich, N., Demin, O., Kiyatkin, A., Goryanin, I., Hoek, J. B. and Kholodenko, B. N. (2002) Temperature dependence of the epidermal growth factor receptor signaling network can be accounted for by a kinetic model. *Biochemistry* **41**, 306-320

- 42 Schoeberl, B., Eichler-Jonsson, C., Gilles, E. D. and Muller, G. (2002) Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat Biotechnol* **20**, 370-375
- 43 Swain, P. S. and Siggia, E. D. (2002) The role of proofreading in signal transduction specificity. *Biophys J* **82**, 2928-2933
- 44 Aksan, I. and Kurnaz, M. L. (2003) A computer-based model for the regulation of mitogen activated protein kinase (MAPK) activation. *J Recept Signal Transduct Res* **23**, 197-209
- 45 Bluthgen, N. and Herzog, H. (2003) How robust are switches in intracellular signaling cascades? *J Theor Biol* **225**, 293-300
- 46 Hatakeyama, M., Kimura, S., Naka, T., Kawasaki, T., Yumoto, N., Ichikawa, M., Kim, J. H., Saito, K., Saeki, M., Shirouzu, M., Yokoyama, S. and Konagaya, A. (2003) A computational model on the modulation of mitogen-activated protein kinase (MAPK) and Akt pathways in heregulin-induced ErbB signalling. *Biochem J* **373**, 451-463
- 47 Hendriks, B. S., Opreko, L. K., Wiley, H. S. and Lauffenburger, D. (2003) Quantitative analysis of HER2-mediated effects on HER2 and epidermal growth factor receptor endocytosis: distribution of homo- and heterodimers depends on relative HER2 levels. *J Biol Chem* **278**, 23343-23351
- 48 Resat, H., Ewald, J. A., Dixon, D. A. and Wiley, H. S. (2003) An integrated model of epidermal growth factor receptor trafficking and signal transduction. *Biophys J* **85**, 730-743
- 49 Xiong, W. and Ferrell, J. E., Jr. (2003) A positive-feedback-based bistable 'memory module' that governs a cell fate decision. *Nature* **426**, 460-465
- 50 Markevich, N. I., Hoek, J. B. and Kholodenko, B. N. (2004) Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. *J Cell Biol* **164**, 353-359
- 51 Markevich, N. I., Moehren, G., Demin, O. V., Kiyatkin, A., Hoek, J. B. and Kholodenko, B. N. (2004) Signal processing at the Ras circuit: what shapes Ras activation patterns? *Systems Biology*, 104-113
- 52 Qiu, D., Mao, L., Kikuchi, S. and Tomita, M. (2004) Sustained MAPK activation is dependent on continual NGF receptor regeneration. *Dev Growth Differ* **46**, 393-403
- 53 Yamada, S., Taketomi, T. and Yoshimura, A. (2004) Model analysis of difference between EGF pathway and FGF pathway. *Biochem Biophys Res Commun* **314**, 1113-1120
- 54 Sasagawa, S., Ozaki, Y., Fujita, K. and Kuroda, S. (2005) Prediction and validation of the distinct dynamics of transient and sustained ERK activation. *Nat Cell Biol* **7**, 365-373
- 55 Vayttaden, S. J., Ajay, S. M. and Bhalla, U. S. (2004) A spectrum of models of signaling pathways. *Chembiochem* **5**, 1365-1374
- 56 Bhalla, U. S. (2004) Models of cell signaling pathways. *Curr Opin Genet Dev* **14**, 375-381
- 57 Sauro, H. M. and Kholodenko, B. N. (2004) Quantitative analysis of signaling networks. *Prog Biophys Mol Biol* **86**, 5-43
- 58 Eungdamrong, N. J. and Iyengar, R. (2004) Modeling cell signaling networks. *Biol Cell* **96**, 355-362
- 59 Wiley, H. S., Shvartsman, S. Y. and Lauffenburger, D. A. (2003) Computational modeling of the EGF-receptor system: a paradigm for systems biology. *Trends Cell Biol* **13**, 43-50
- 60 Kholodenko, B. N., Demin, O. V., Moehren, G. and Hoek, J. B. (1999) Quantification of short term signaling by the epidermal growth factor receptor. *J Biol Chem* **274**, 30169-30181
- 61 Saso, K., Moehren, G., Higashi, K. and Hoek, J. B. (1997) Differential inhibition of epidermal growth factor signaling pathways in rat hepatocytes by long-term ethanol treatment. *Gastroenterology* **112**, 2073-2088
- 62 Alessi, D. R., Saito, Y., Campbell, D. G., Cohen, P., Sithanandam, G., Rapp, U., Ashworth, A., Marshall, C. J. and Cowley, S. (1994) Identification of the sites in MAP kinase kinase-1 phosphorylated by p74raf-1. *Embo J* **13**, 1610-1619
- 63 Dhillon, A. S. and Kolch, W. (2002) Untying the regulation of the Raf-1 kinase. *Arch Biochem Biophys* **404**, 3-9
- 64 Kitano, H. (2004) Biological robustness. *Nat Rev Genet* **5**, 826-837

- 65 Babu, C. V., Yoon, S., Nam, H. S. and Yoo, Y. S. (2004) Simulation and sensitivity analysis of phosphorylation of EGFR signal transduction pathway in PC12 cell model. *Systems Biology* **2**
- 66 Gong, Y. and Zhao, X. (2003) Shc-dependent pathway is redundant but dominant in MAPK cascade activation by EGF receptors: a modeling inference. *FEBS Lett* **554**, 467-472
- 67 Conzelmann, H., Saez-Rodriguez, J., Sauter, T., Bullinger, E., Allgower, F. and Gilles, E. D. (2004) Reduction of mathematical models of signal transduction networks: simulation-based approach applied to EGF receptor signalling. *Systems Biology*, 159-169
- 68 Liu, G., Swihart, M. T. and Neelamegham, S. (2004) Sensitivity, principal component and flux analysis applied to signal transduction: the case of epidermal growth factor mediated signaling. *Bioinformatics*
- 69 Mayawala, K., Gelmi, C. A. and Edwards, J. S. (2004) MAPK cascade possesses decoupled controllability of signal amplification and duration. *Biophys J* **87**, L01-02
- 70 Miller, J. H. and Zheng, F. (2004) Large-scale simulations of cellular signaling processes. *Parallel Computing* **30**, 1137-1149
- 71 Brown, K. S., Hill, C. C., Calero, G. A., Myers, C. R., Lee, K. H., Sethna, J. P. and Cerione, R. A. (2004) The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Physical Biology*
- 72 Bonarius, H. P. J., Schmid, G. and Tramper, J. (1997) Flux analysis of underdetermined metabolic networks: The quest for the missing constraints. *Trends in Biotechnology* **15**, 308-314
- 73 Kacser, H. and Burns, J. A. (1973) The control of flux. *Symp. Soc. Exp. Biol.* **27**, 65-104
- 74 Hornberg, J. J., Bruggeman, F. J., Binder, B., Geest, C. R., de Vaate, A. J., Lankelma, J., Heinrich, R. and Westerhoff, H. V. (2005) Principles behind the multifarious control of signal transduction. ERK phosphorylation and kinase/phosphatase control. *Febs J* **272**, 244-258
- 75 Regev, A., Silverman, W. and Shapiro, E. (2001) Representation and simulation of biochemical processes using Pi-calculus process algebra. *Pacific Symposium on Biocomputing*, 459-470
- 76 Priami, C., Regev, A., Silverman, W. and Shapiro, E. (2001) Application of a stochastic name passing calculus to representation and simulation of molecular processes. *Information Processing Letters* **80**, 25-31
- 77 Pinney, J., Westhead, D. and McConkey, G. (2003) Petri Net representations in systems biology. *Biochemical Society Transactions* **31**, 1513-1515
- 78 Oliveira, J. S., Jones-Oliveira, J. B., Dixon, D. A., Bailey, C. G. and Gull, D. W. (2004) Hyperdigraph-theoretic analysis of the EGFR signaling network: initial steps leading to GTP:Ras complex formation. *J Comput Biol* **11**, 812-842
- 79 Koch, I. and Heiner, M. (2004) Qualitative modelling and analysis of biochemical pathways with petri nets. *Tutorial Notes, 5th International Conference on Systems Biology*
- 80 Heiner, M., Koch, I. and Will, J. (2004) Model validation of biological pathways using Petri nets, demonstrated for apoptosis. *Biosystems* **75**, 15-28
- 81 Matsuno, H., Doi, A., Nagasaki, M. and Miyano, S. (2000) Hybrid petri net representation of gene regulatory network. *Pacific Symposium on Biocomputing* 5, 341-352
- 82 Popova-Zeugmann, L., Heiner, M. and Koch, I. (2004) Modelling and analysis of biochemical networks with time petri nets. *Informatik-Berichte der HUB* **170**, 136-143
- 83 Goss, P. J. and Peccoud, J. (1998) Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. *Proceedings of the National Academy of Sciences* **95**, 6750-6755
- 84 Chabrier, N. and Fages, F. (2003) Symbolic model checking of biochemical networks. *Lecture Notes in Computer Science* **2602**, 149-162
- 85 Chabrier-Rivier, N., Chiaverini, M., Danos, V., Fages, F. and Schachter, V. (2004) Modeling and querying biomolecular interaction networks. *Theoretical Computer Science* **325**

- 86 Calder, M., Gilmore, S. and Hillston, J. (2004) Modelling the influence of RKIP on the ERK signalling pathway using the stochastic process algebra PEPA. Proceedings of Bio-Concur
- 87 (2005) In pursuit of systems. *Nature* **435**, 1
- 88 González, P. P., Cárdenas, M., Gershenson, C. and Lagunez, J. (2001) Integration of Computational Techniques for the Modelling of Signal Transduction. *Advances in Systems Science: Measurement, Circuits and Control*. WSES Press.
- 89 Cho, K. H., Shin, S. Y., Kim, H. W., Wolkenhauer, O., McFerran, B. and Kolch, W. (2003) Mathematical modeling of the influence of RKIP on the ERK signaling pathway. CMSB'03
- 90 Chapman, S. and Asthagiri, A. R. (2004) Resistance to signal activation governs design features of the MAP kinase signaling module. *Biotechnol Bioeng* **85**, 311-322
- 91 Aksan, I. (2005) Kinetic Analysis of RSK2 and Elk-1 Interaction on the Serum Response Element and Implications for Cellular Engineering. *Biotechnology and Bioengineering* **88**, 890-900
- 92 Perez-Jimenez, M. J. and Romero-Campero, F. J. (2005) Modelling EGFR signalling cascade using continuous membrane systems. *Computational Methods in Systems Biology*, 118-129
- 93 Oney, I., Kurnaz, I. A. and Kurnaz, M. L. (2005) Cytoplasmic-to-nuclear volume ratio affects AP-1 complex formation as an indicator of cell cycle responsiveness. *FEBS Lett* **579**, 433-440

Tables

	Kholodenko <i>et al.</i>, (1999)	Brightman & Fell (2000)	Schoeberl <i>et al.</i>, (2002)
Cell Line	Rat Hepatocytes	PC12	HeLa
Reactions	25	30	125
Species	23	29	94
Receptor Complex Strategy	YES	NO	YES
Receptor Internalisation	NO	YES	YES
Receptor Degradation	NO	NO	YES
Shc-dependent Pathway	YES	YES	YES
Shc-independent Pathway	YES	NO	YES
PLC γ Pathway	YES	NO	NO
Core ERK Cascade	NO	YES	YES
Negative Feedback Loop	NO	YES	NO

Table 1: Comparison of the size and features of the selected models

This table contains a direct comparison between the three selected models: Kholodenko *et al.* (1999) [60], Brightman & Fell (2000) [35] and Schoeberl *et al.* (2002) [42]. The comparison criteria are defined as followed (more details can be found in the text below): Cell Line: the cell line the model is based on; Reactions: the number of reactions in the model; Species: the number of biochemical species in the model; Receptor Complex Strategy: whether adaptor proteins have to bind to the EGF receptor for the signal to propagate; Receptor Internalisation: the process of internalisation of receptors from the plasma membrane; Receptor Degradation: the process of degradation of internalised receptor; Shc-dependent pathway: the pathway comprising Shc, Grb2 and SOS leading to the activation of Ras; Shc-independent pathway: the pathway comprising Grb2 and SOS leading to the activation of Ras (i.e. Shc not required); PLC γ Pathway: the pathway comprising the reactions of the adaptor protein PLC γ ; Core ERK Cascade: the Raf, MEK and ERK cascade; Negative Feedback Loop: the negative feedback phosphorylation of SOS by ERKPP.

Figure Legends

Figure 1: Kinase / Phosphatase signalling reaction

A phospho – group is transferred from a nucleotide to a serine, threonine or tyrosine residue of a target protein (T). The reaction is catalysed by a kinase. The phosphorylated target protein (T-P) is in turn dephosphorylated by a phosphatase to complete the cycle. Phosphorylation of a protein often entails conformational changes that modulate the function of the protein. Phosphorylation of MEK, e.g. activates the kinase domain of the enzyme. MEK then catalyses the phosphorylation and, as a consequence thereof, activation of MAPK.

Figure 2: General structure of the three tiered cascade of the MAPK pathway

A signal is propagated along this cascade by sequential phosphorylation as described in the text.

Figure 3: Structure of the ERK pathway

Upon ligand binding receptor tyrosine kinases (RTK) autophosphorylate (phosphates are shown as red circles) on tyrosine residues, which serve as docking sites for adaptor and signalling molecules. Ras and Rap1 are activated by the recruitment of Guanosine nucleotide exchange factors (SOS, C3G) via adaptor proteins (Shc and Grb2; Crk). Ras can activate Raf-1 and B-Raf, Rap1 presumably can activate B-Raf. Raf proteins phosphorylate and activate MEK-1/2, which in turn activate ERK-1/2 (indicated by black arrows). Negative feedback loops (indicated by red lines) include the induction of MAPK phosphatases (MKPs) by ERK as well as the inhibitory phosphorylation of Raf-1 and SOS.

Figure 4: The Five Steps of Modelling

This diagram depicts the five steps involved in modelling a biological system. The first step is identifying the biological system to model, followed by actually defining the model to represent the system, simulating the model and validating the simulation results. If the model is valid it can be analysed further, if it is not the definition step is revisited where the model is checked for various types of errors. For more information on each of the steps see the accompanying text.

Figure 5: Timeline of ERK Models

This diagram is a timeline of mathematical models that in some way incorporate the ERK cascade. Models are represented as ovals labelled with the name of the first author and located above the year they were published. White ovals represent models of the core ERK cascade whereas grey ovals represent larger models generally including growth factor receptors, adaptor proteins as well as the ERK cascade itself. Models highlighted in red are the models we have selected for discussion in detail below. 1996: Huang [29]; 1997: Burack [30], Ferrell [31]; 1998: Ferrell [32]; 1999: Bhalla [24], Kholodenko [60]; 2000: Brightman [35],

Kholodenko [34], Levchenko: [38]; 2001: Asthagiri [40], Gonzalez [88]; 2002: Bhalla [33], Moehren [41], Schoeberl [42], Shvartsman [36], Somsen [39], Swain [43]; 2003: Aksan [44], Hatakeyama [46], Hendriks [47], Resat [48], Bluthgen [45], Cho [89], Xiong [49]; 2004: Maly [37], Markevich [51], Oliveira [78], Qiu [52], Yamada [53], Chapman [90], Markevich [50]; 2005: Aksan [91], Perez-Jimenez [92], Oney [93], Sasagawa [54].

Figure 6: Kholodenko *et al.* (1999) Model Diagram

This schematic representation of the Kholodenko *et al.* (1999) model of EGFR signalling mediated by adapter and target proteins is taken from the original publication.

Figure 7: Brightman & Fell (2000) Model Diagram

This schematic representation of the Brightman & Fell (2000) model of EGF signal transduction is taken from the original publication. In this schema, GS represents the Grb2/SOS complex.

Figure 8: Schoeberl *et al.* (2002) Model Diagram

This schematic representation of the Schoeberl *et al.* (2002) model of the EGF receptor induced ERK kinase cascade is taken from the original publication.

Figures

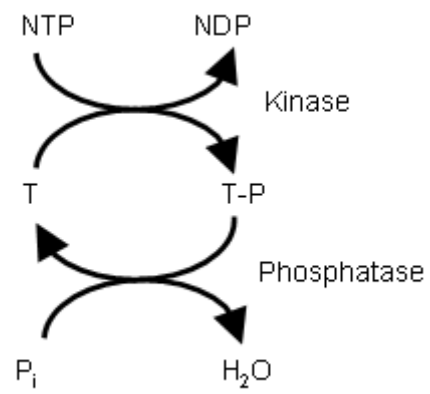


Figure 1

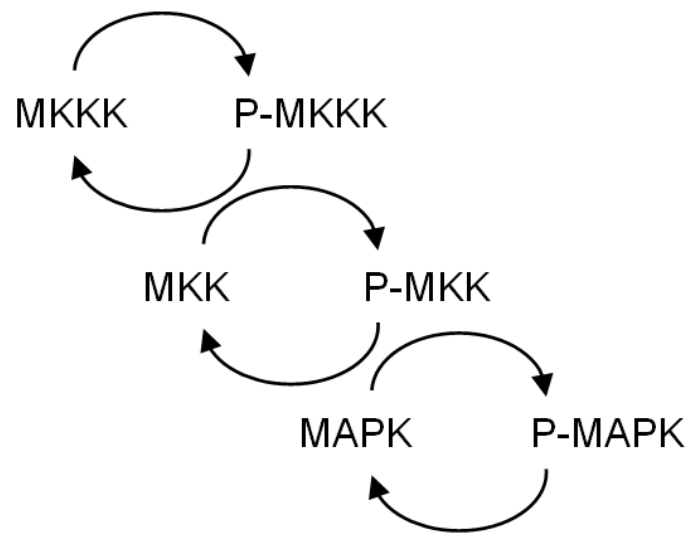


Figure 2

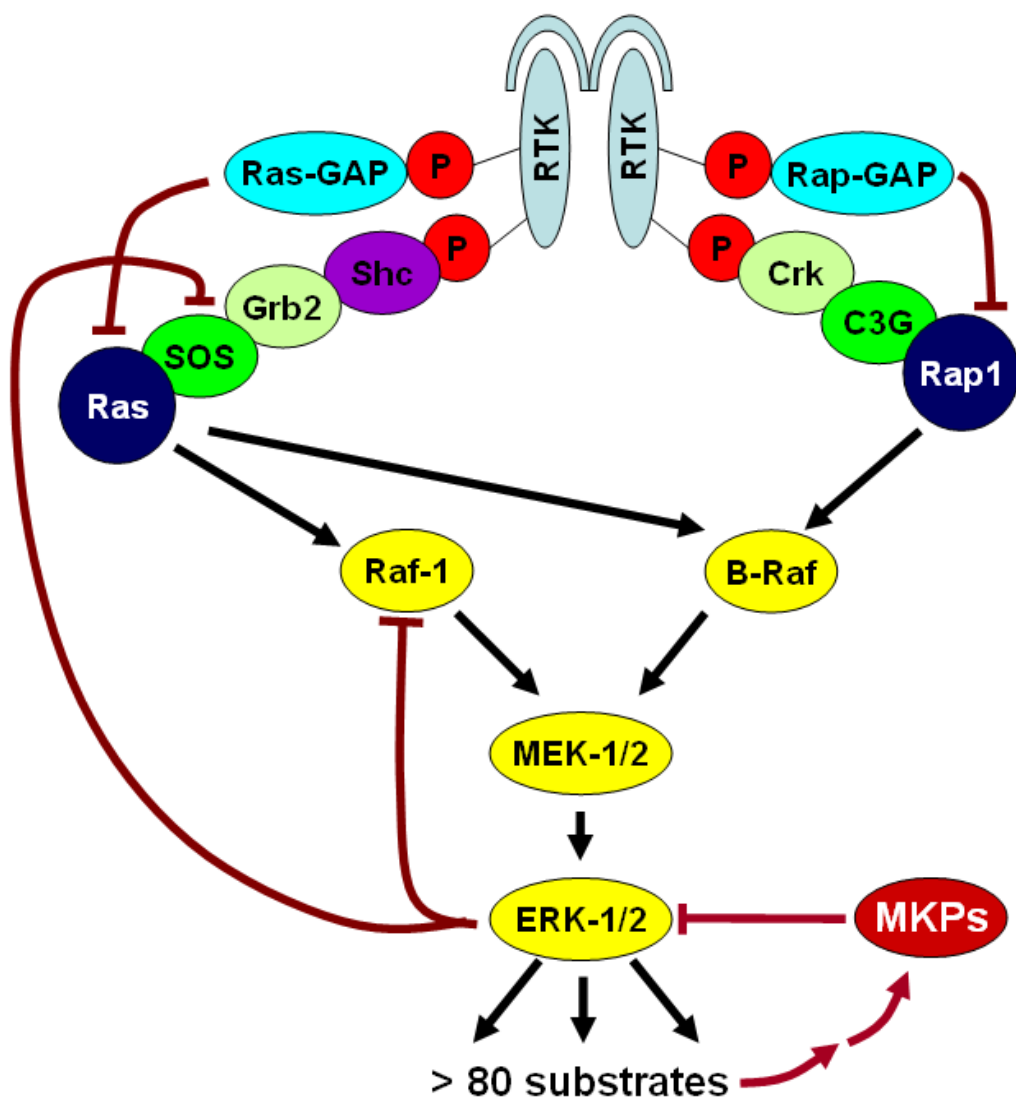


Figure 3

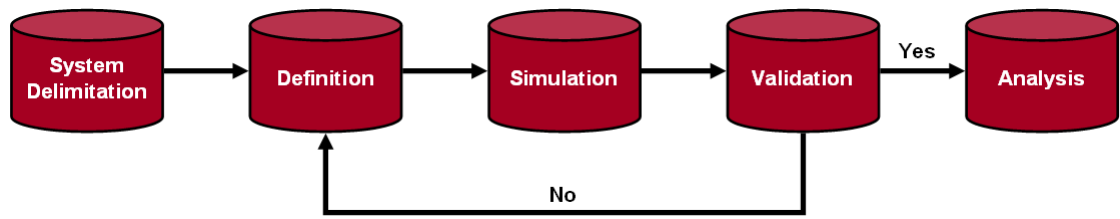


Figure 4

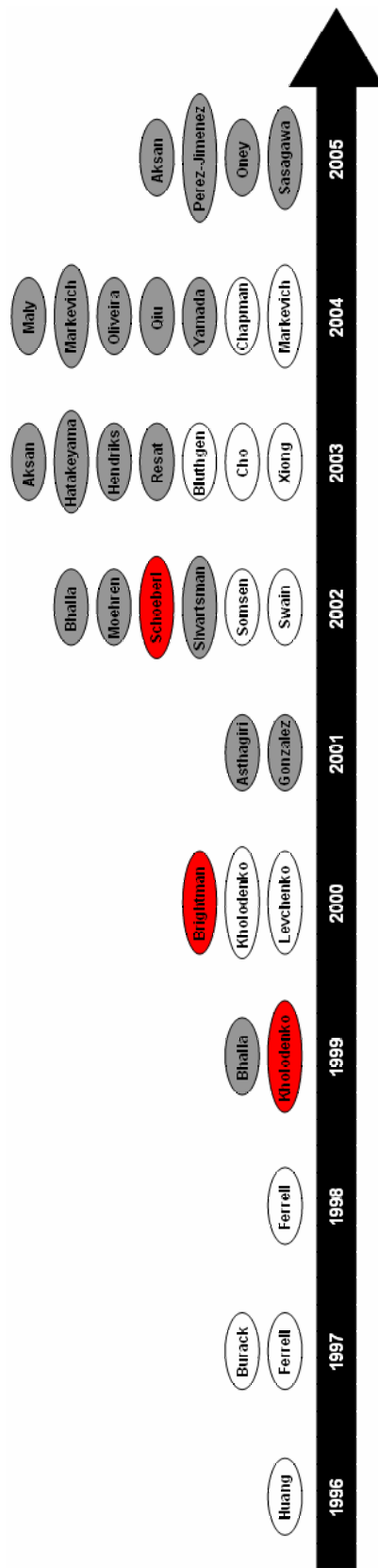


Figure 5

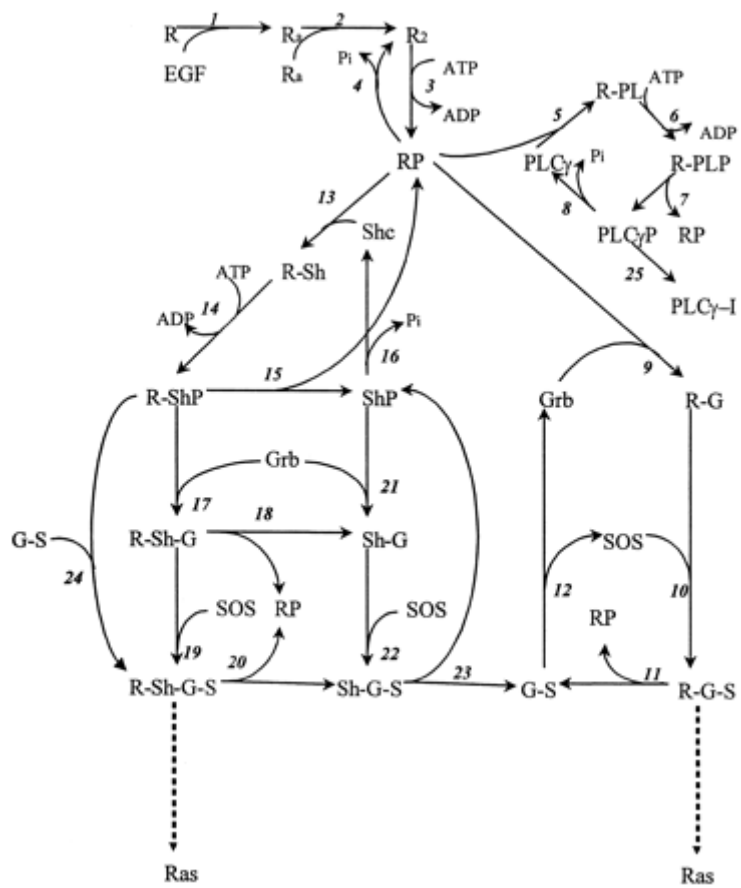


Figure 6

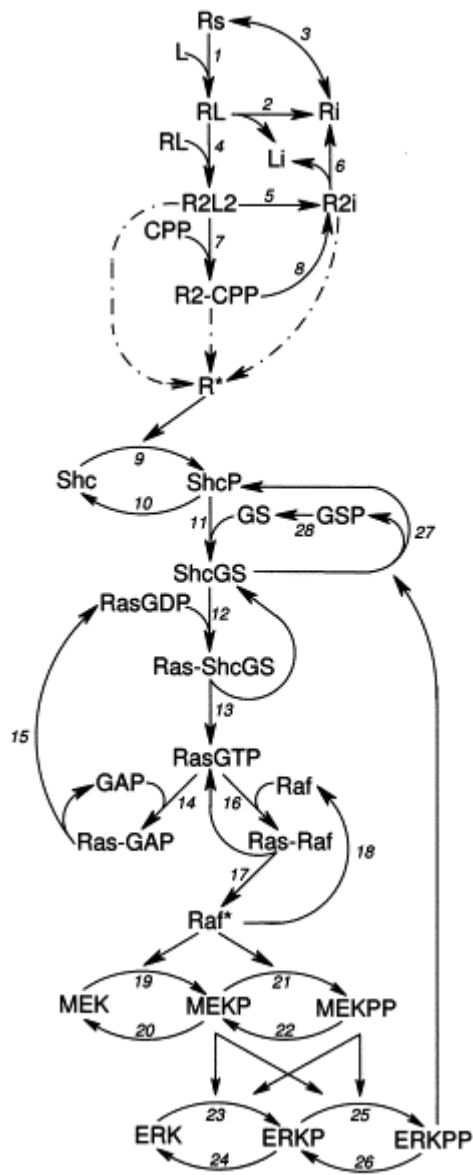


Figure 7

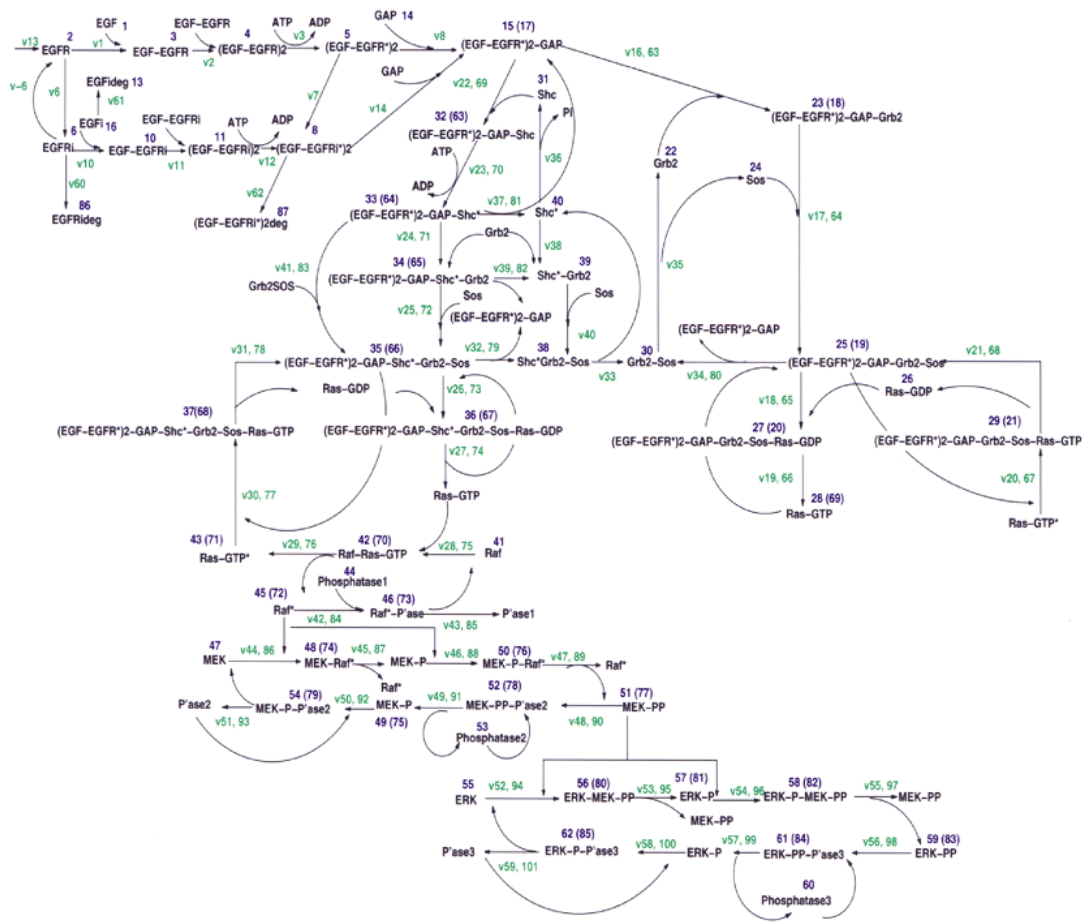


Figure 8