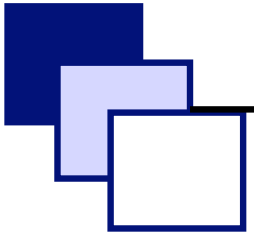


BPS: Biochemical Pathway Simulator

A Software Tool for Simulation & Analysis of Biochemical Networks

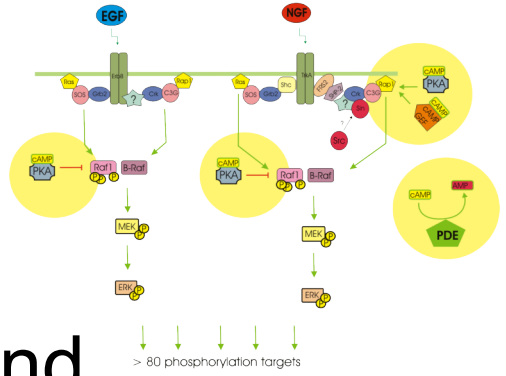
www.brc.dcs.gla.ac.uk/projects/bps

University of Glasgow
Bioinformatics Research Centre
Beatson Laboratories

DTI 'Beacon' project

David Gilbert
Muffy Calder
Walter Kolch
Richard Orton
Vladislav Vyshemirsky
Oliver Sturm
Amelie Gormand

The Problem



- The behaviour of cells is governed and coordinated by *biochemical signalling networks* that translate external cues (hormones, growth factors, stress etc) into adequate biological responses such as cell proliferation, specialisation or death, and metabolic control.

- Since regulatory malfunction underlies many diseases such as cancer, a deep understanding is crucial for drug development and other therapies.

The Challenge in Systems Biology…

- One of the great challenges in Systems Biology is to model and analyse biochemical networks when there is a ***lack of exact quantitative data***.
- Traditional approaches to modelling biochemical networks are based on ***differential equations, which require exact rate constants*** in order model realistically the reactions.
- *The aim of this interdisciplinary project is to model diverse biochemical networks and develop an associated computational system to facilitate their simulation and analysis which does not require exact quantitative data for all reactions*.
- The biochemical behaviour of the pathways in which we are interested is being studied in the context of the biological effects of differentiation; these data have to be generated in cells, and are not amenable to analysis using differential equations without making unwarranted assumptions about concentrations.
- Our approach is based on modelling quantitative and qualitative aspects using concurrency theories and tools.

Aims

- *Model* dynamic behaviour of biochemical networks
- Develop a computational system to *analyse* their behaviour
- Guide *experimental* design

Approach
Continuous cross check between modelling & real experimental data.

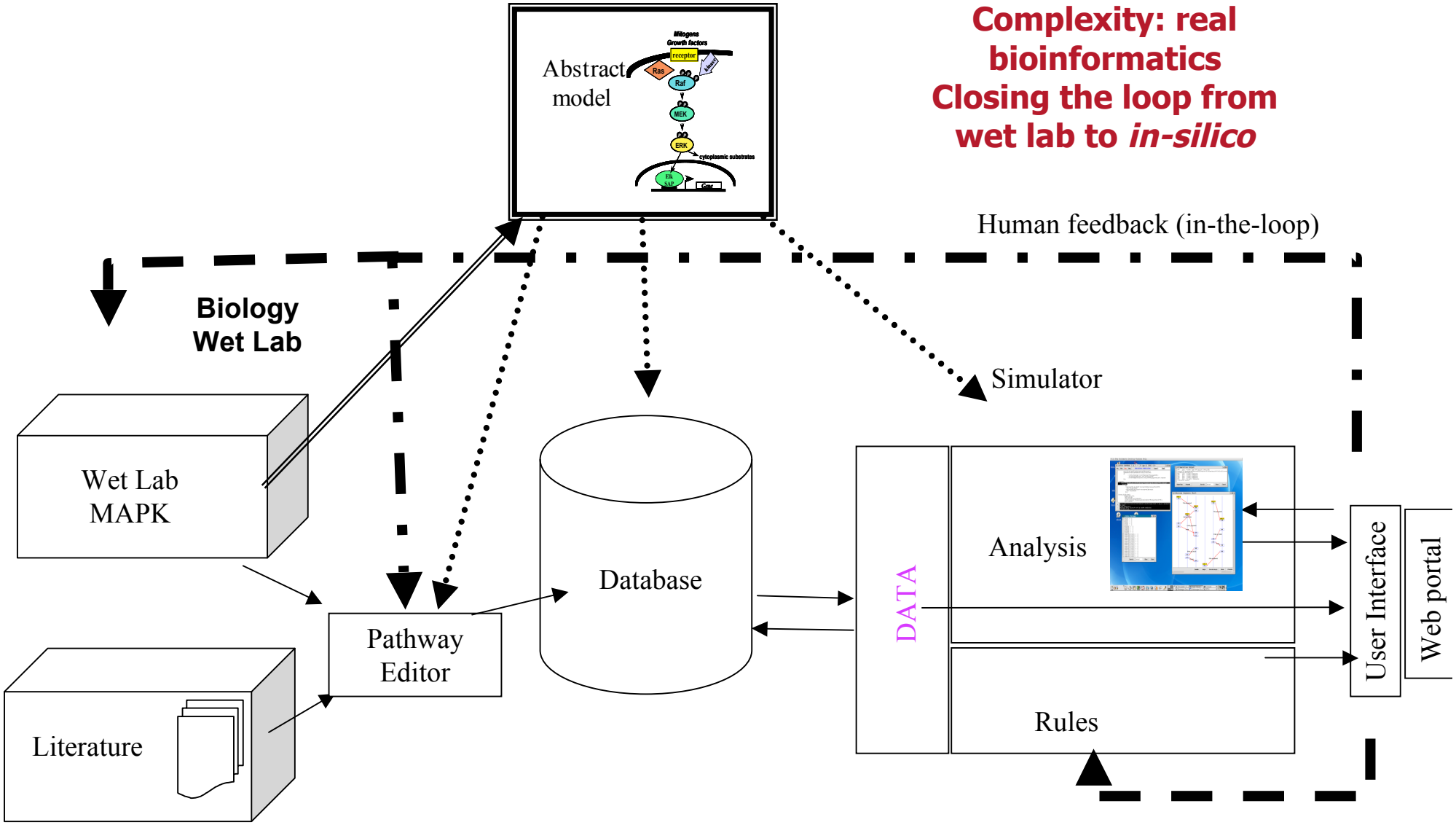
Model system: MAPK signalling network

A target of current drug discovery efforts in important disease areas e.g. cancer, arteriosclerosis, stroke, heart disease, chronic inflammatory and degenerative diseases.

BPS Project Team

- **Collaborative**: Dept Computing Science, Inst Biomedical and Life Sciences, Beatson Cancer Research Institute.
- David Gilbert (PI) [BRC/DCS]
- Muffy Calder [DCS]
- Walter Kolch, [IBLS & Beatson Institute]
- 3 researchers:
  - bioinformatician – Richard Orton
  - concurrency specialist – Vladislav Vyshemirsky
  - biologist – Oliver Sturm
- 2 PhD studentships (Scottish Enterprise)
  - Amelie Gormand (wet lab)
  - Text mining (vacant)

Complexity: real bioinformatics Closing the loop from wet lab to *in-silico*



Wet Lab
Walter, Oliver & Amelie

Bioinformatics
Tools, database, interface
David & Richard

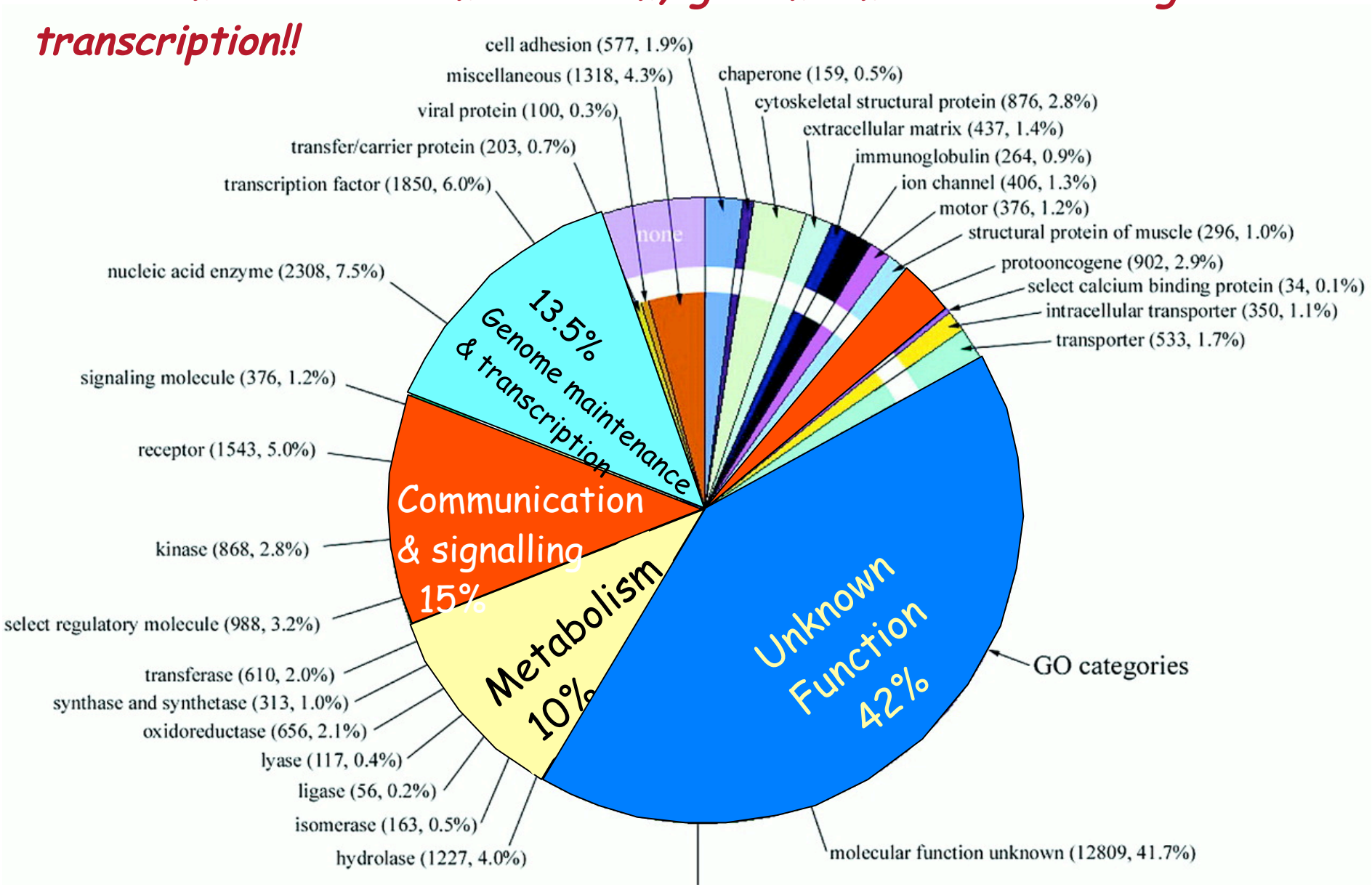
Simulator & analyser
Concurrency theory
Muffy & Vlad

Wet lab

Walter Kolch, Oliver Sturm, Amelie Gormand

15% of our known genes are involved in signal transduction.

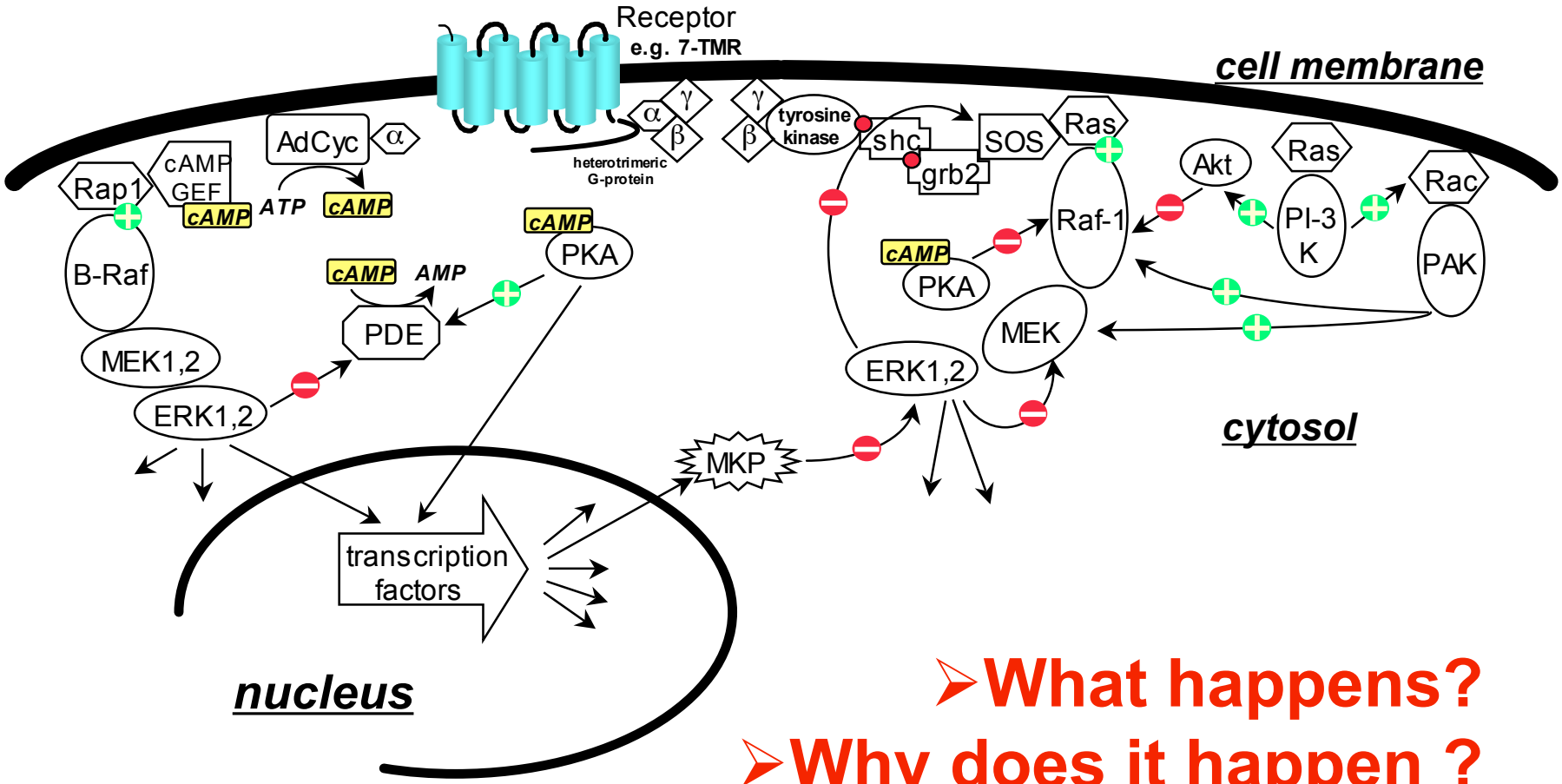
This is more than in metabolism, genome maintenance or gene transcription!!



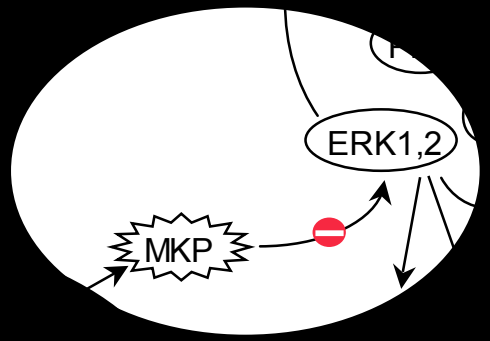
Signal transduction pathways are organised as networks.

We can describe the general topology and single biochemical steps.

However, we do not understand how the network functions as a whole.



- What happens?
- Why does it happen ?
- How is specificity achieved?

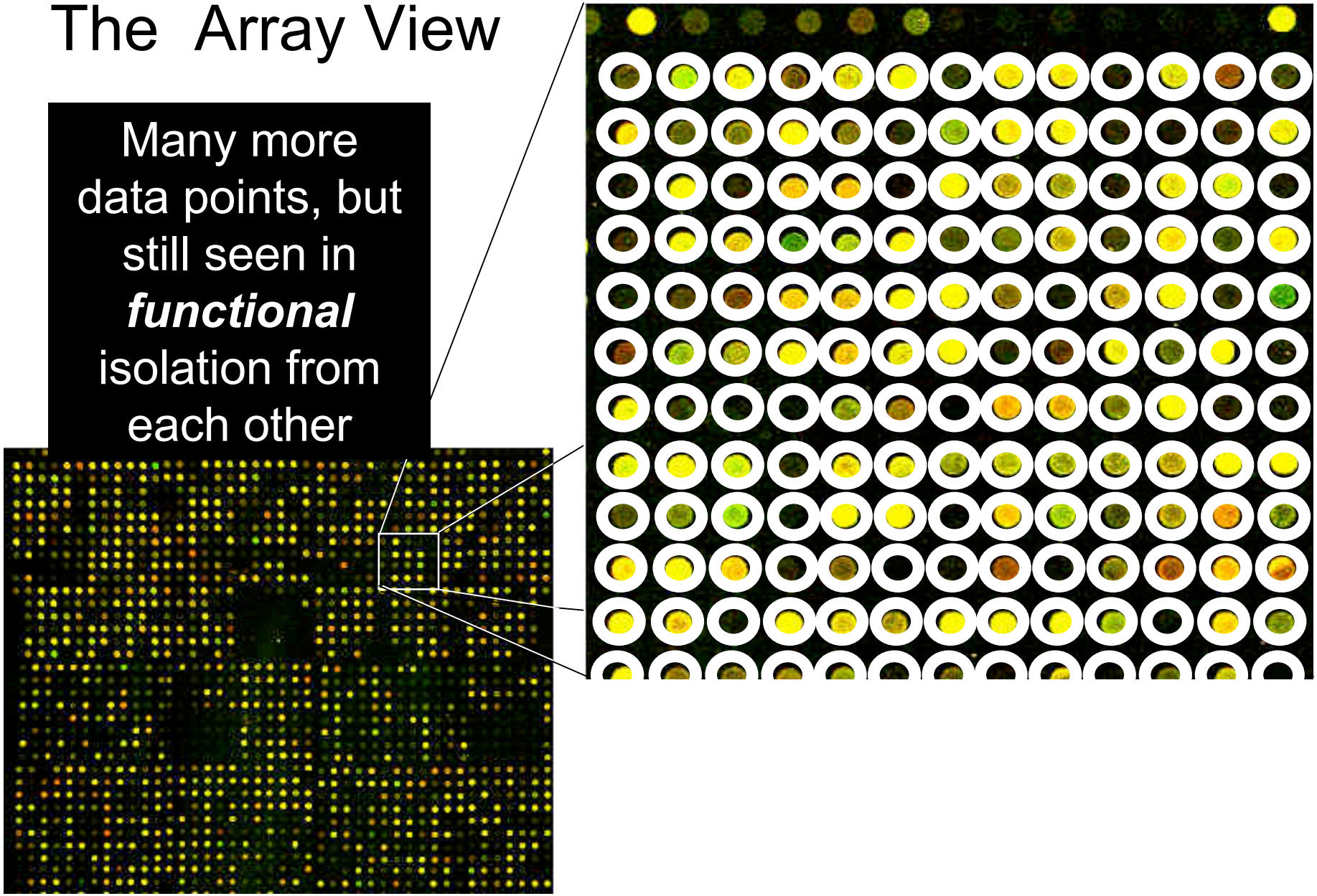


The biochemical view

activ

The Array View

Many more data points, but still seen in ***functional*** isolation from each other



Pathway modelling: challenges & opportunities

Problem: The complexity of signalling networks makes an exhaustive analysis of the kinetics that connect various inputs with outputs impractical.

Hypothesis: Pathway modelling reduces the experimental load by identifying the network nodes that are important for biological decision making.

Problem: How do biochemical networks make biological decisions?

Hypothesis: Pathway modelling can reveal thresholds that translate quantitative biochemical reactions into qualitative biological responses.

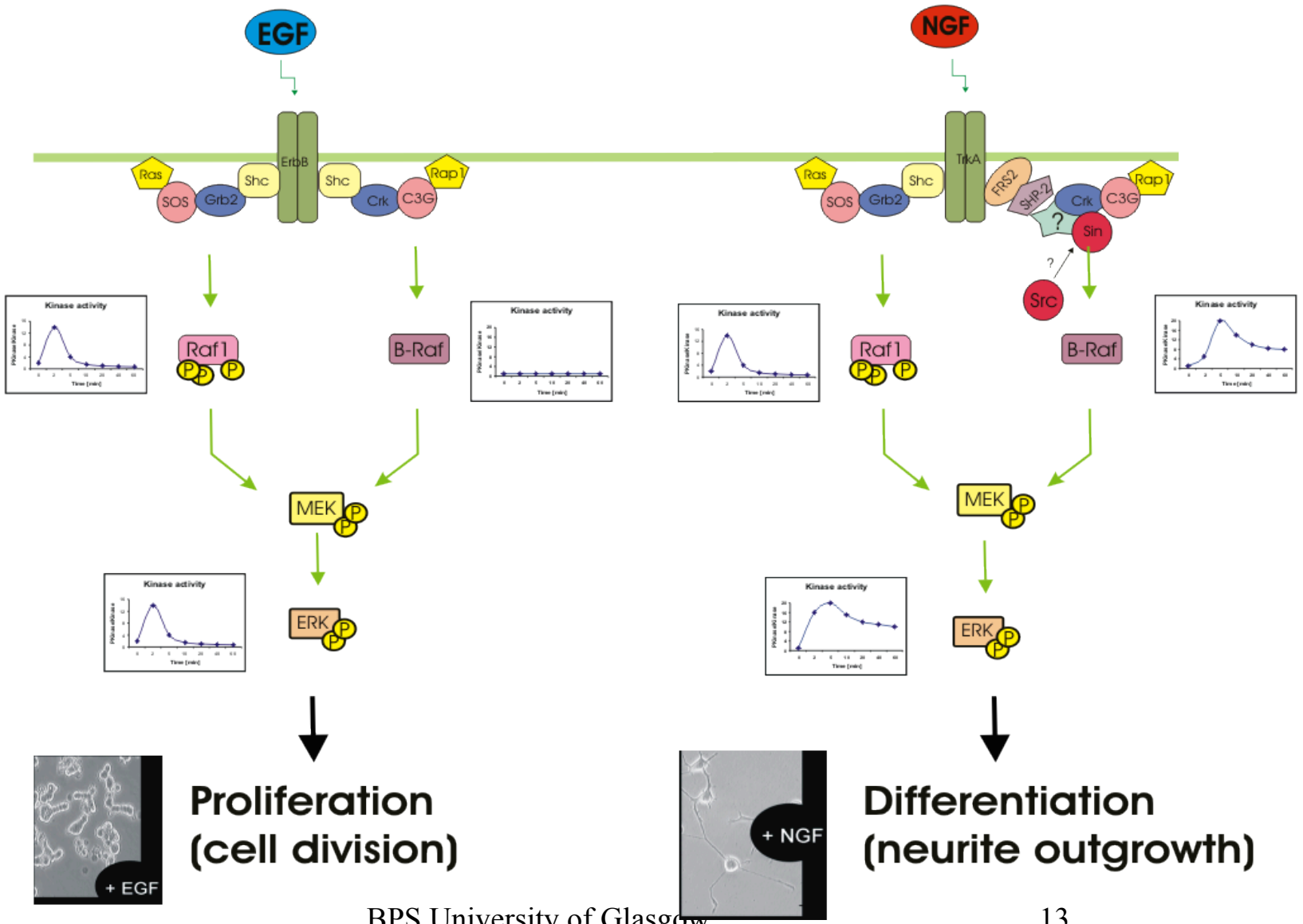
Problem: How do cells distinguish signals from noise?

Hypothesis: Pathway modelling can reveal relevant mechanisms.

Problem: Probably there are a few more ….

Hypothesis: We can faithfully model biological processes, if we can match the unexpected “few more”…

MAPK pathway in action: PC12 cells switch between neuronal differentiation and proliferation



Acquisition of quantitative data

- How to measure kinetics in vivo?

  → Quantitative Immunoblotting (ERK)

  → Kinase assays (C-Raf, B-Raf)

  → RBD – pulldown assays (Ras/Rap1)

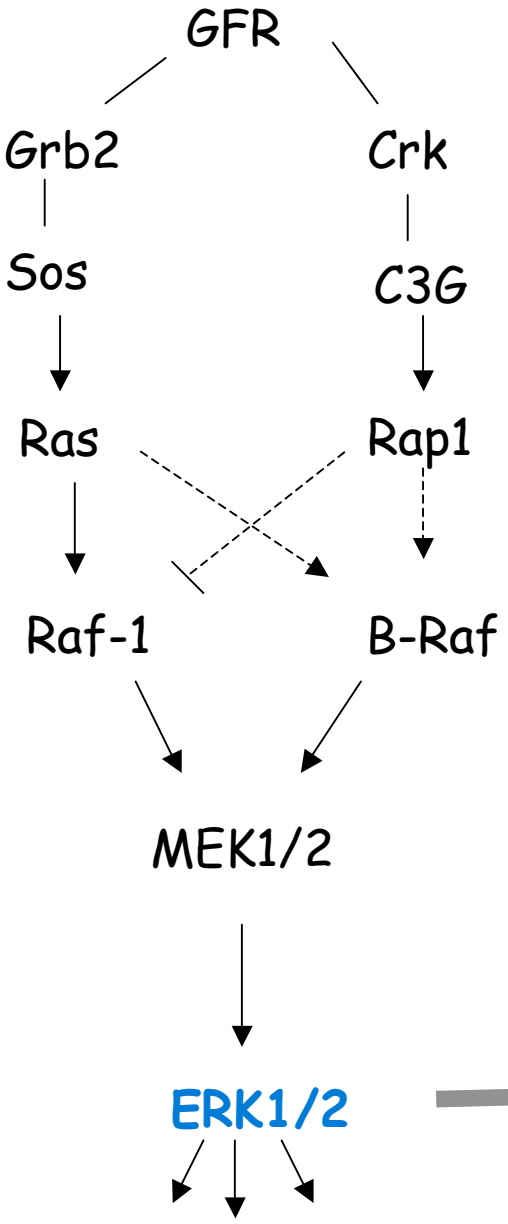
- Prediction of feedback loops/signaling hubs?

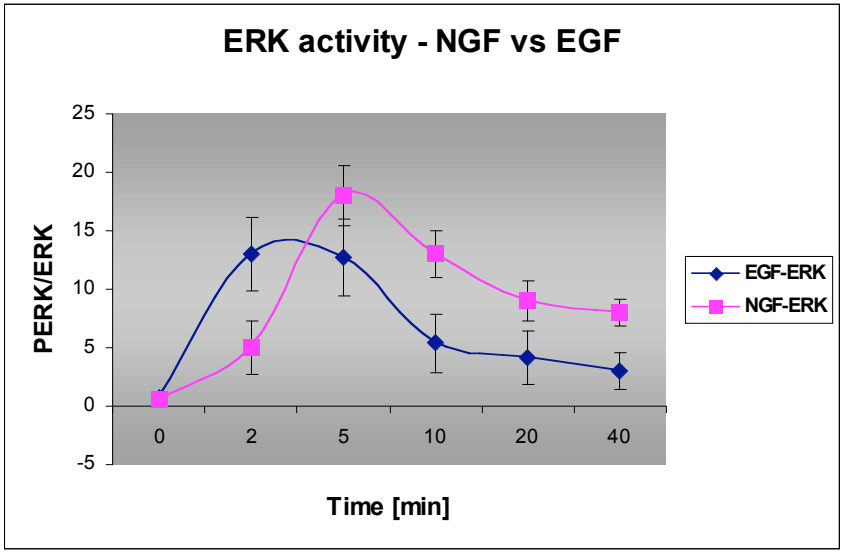
  → antagonists, RNAi interference

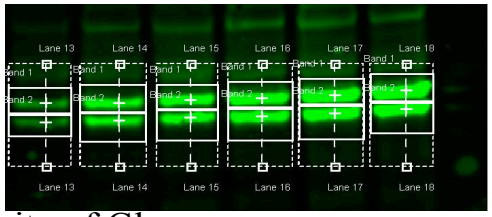
- Quantify protein concentrations

  → Fluorophore labelling of proteins

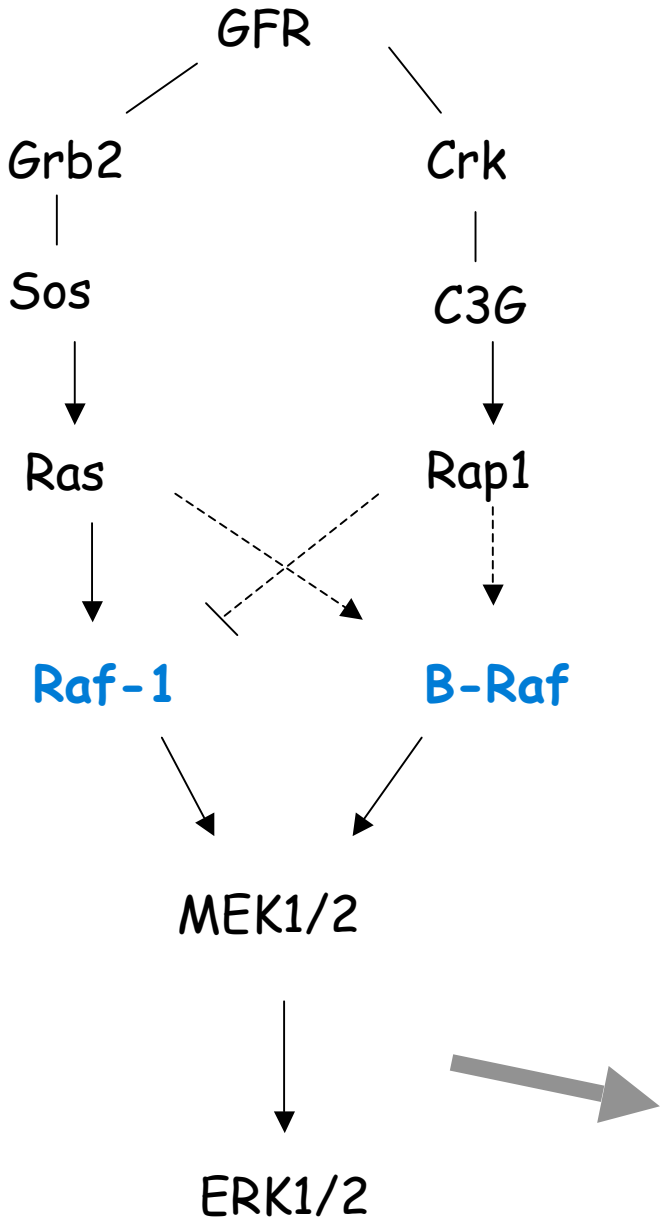
Quantitative Immunoblotting (ERK activation)

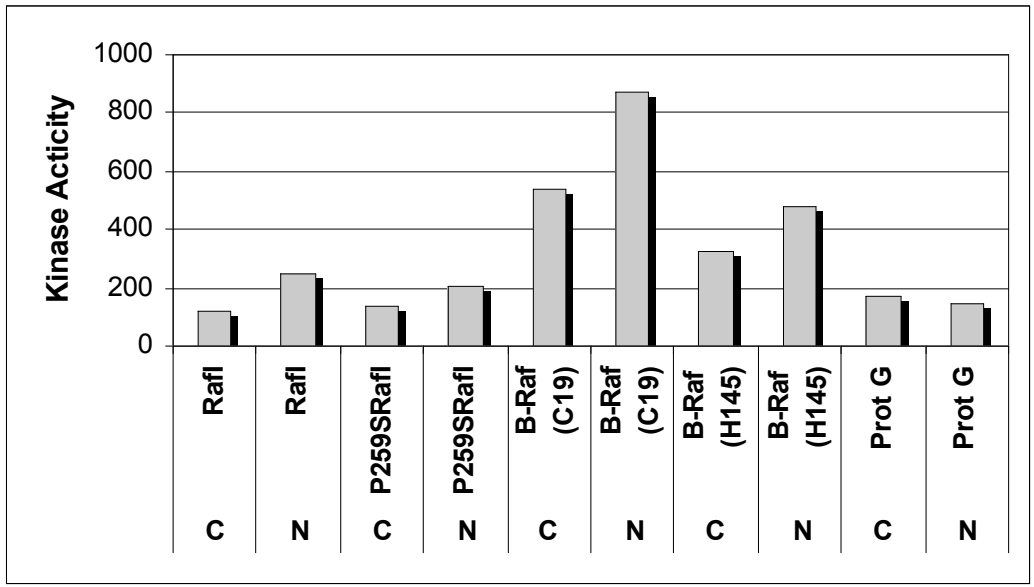


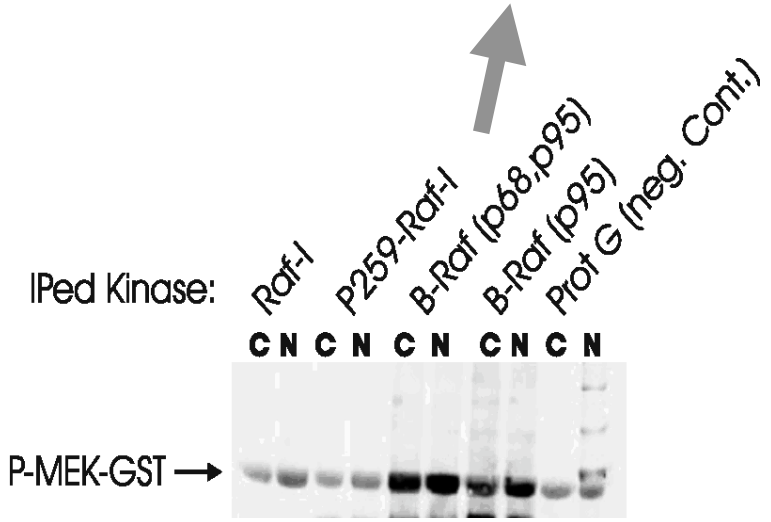




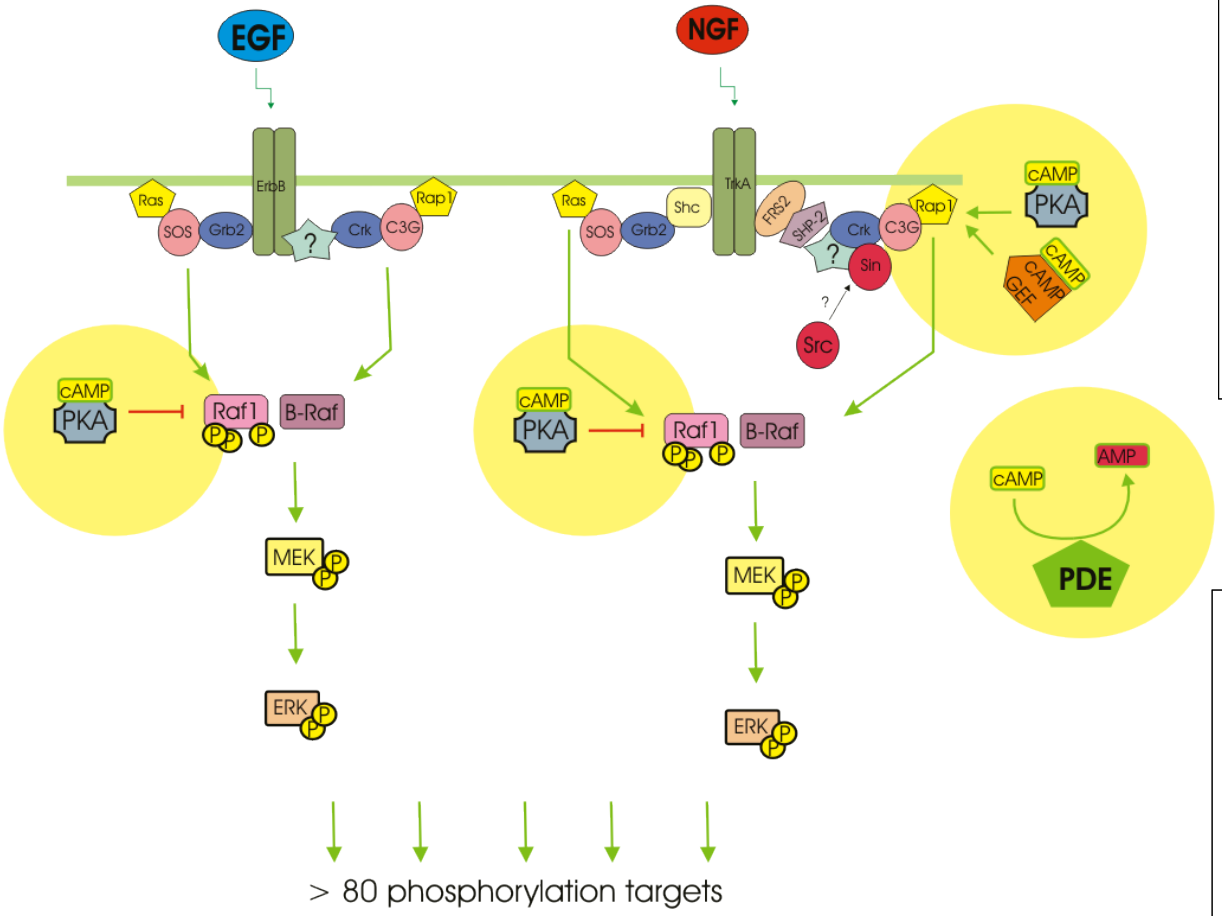
Kinase assays - Raf activation

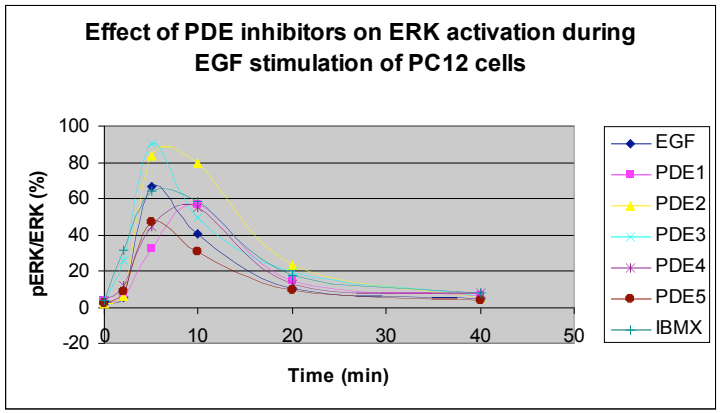


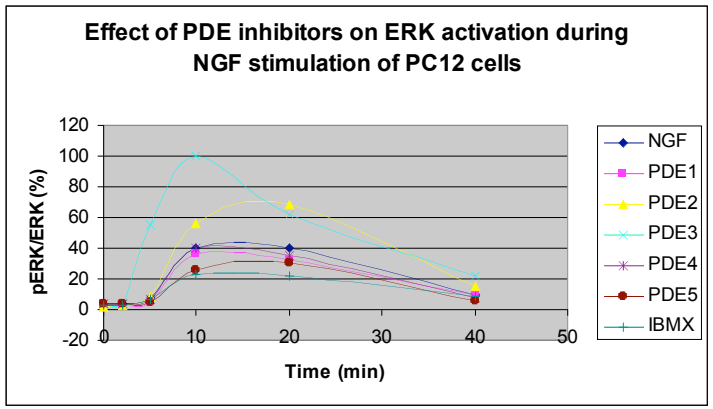


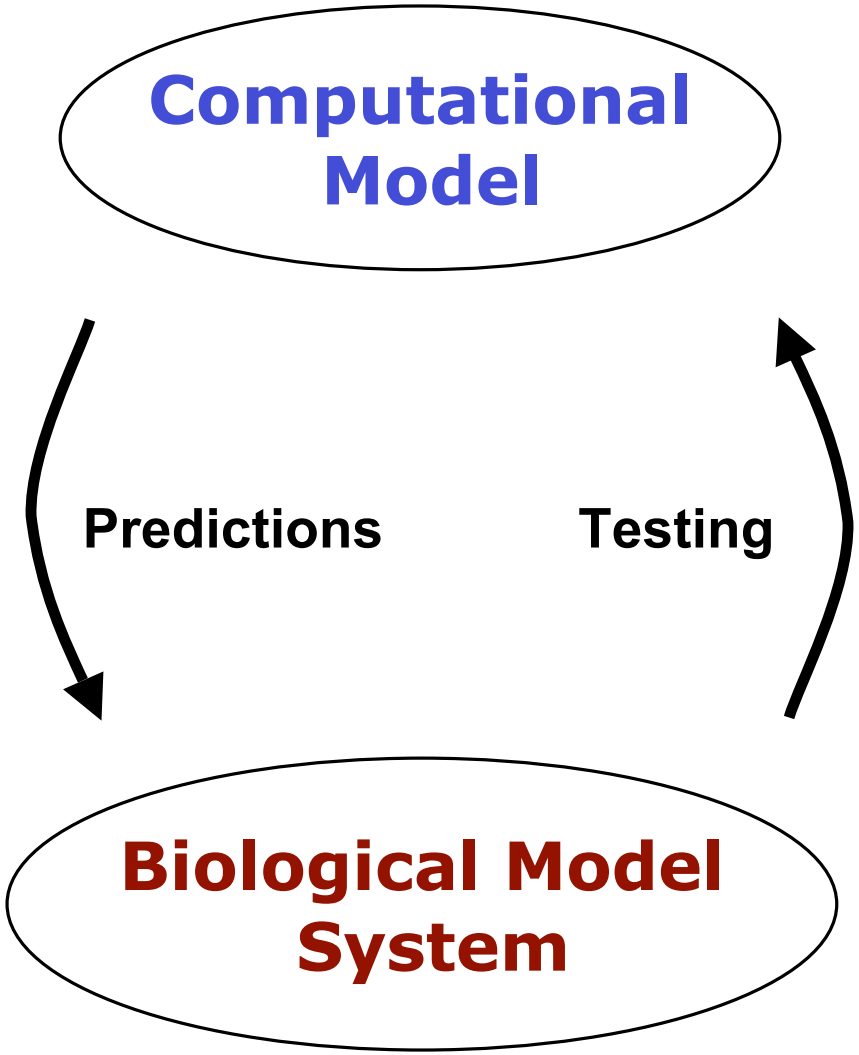


cAMP crosstalk with the MAPK pathway in PC12 cells (Amelie Gormand)









– What is the best formalism for modelling signal transduction?

– Is it valid to model pathways as isolated systems?

– How useful is a computational model for biological questions ?

– What experiments can we do to supply meaningful kinetic data?

– What biological questions can be solved by modelling?

Bioinformatics

Richard Orton & David Gilbert

Overview

- **Modelling with differential equations**
  - Preliminary RKIP model
  - Published MAPK models
  - Development of new MAPK model
- **Model Database**
  - Biochemical models and associated parameter data
  - Modelling data generated from wet lab
  - Information on the pathways and proteins involved
- **Model Control Tool**
  - Construction
  - Analysis
  - Visualisation
  - Conversion

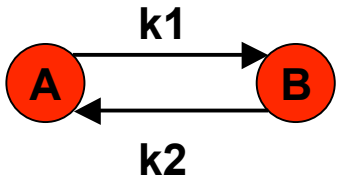
Modelling

- **What is modelling?**
  - Translating a biological pathway (e.g. MAPK) into mathematics for subsequent analysis
- **Why model?**
  - A computer model can generate new insights: in a complex pathway, knowing all the proteins involved and what they do, may still not tell you how the pathway works
  - A computer model can make testable predictions
  - A computer model can test conditions that may be difficult to study in the laboratory
  - A computer model can rule out particular explanations for an experimental observation
  - A computer model can help you identify what's right and wrong with your hypotheses

Differential Equations

- Differential equations are an established technique for modelling biological pathways

- They can be used to model the changes in concentration of all the species in a pathway over time (requires both the rate constants and initial concentrations)



$$\frac{d[A]}{dt} = -k_1[A] + k_2[B]$$

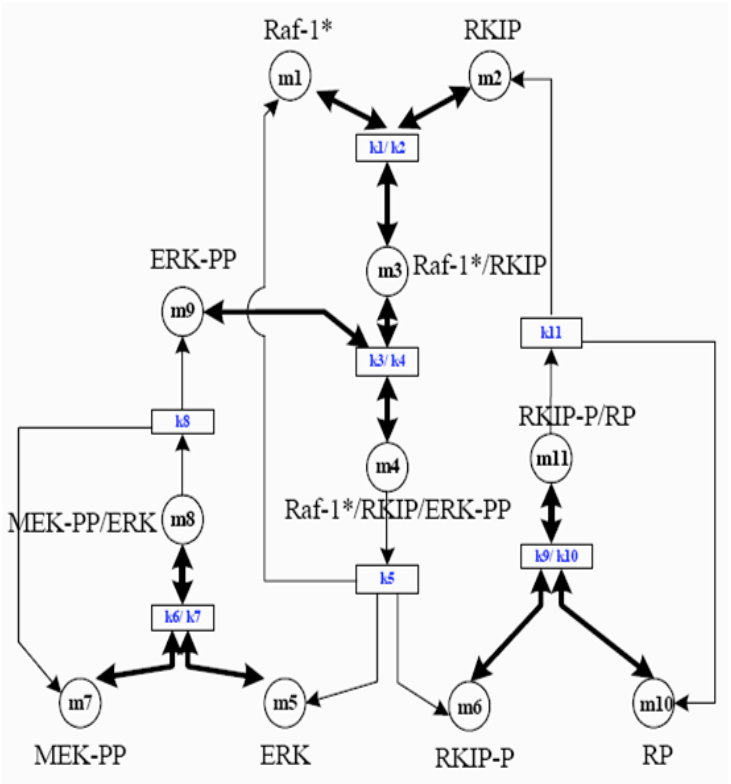
$$\frac{d[B]}{dt} = k_1[A] - k_2[B]$$

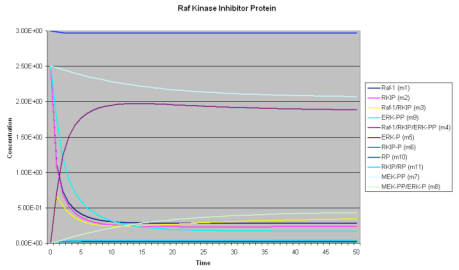
- The simple example above can easily be expanded to include more species and reactions

- Differential equations can be used to model various kinetic types including mass action and michaelis-menten (constants must be known)

Preliminary RKIP Model

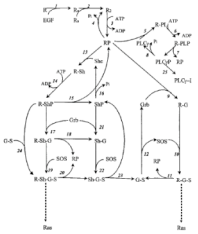
- RKIP (Raf Kinase Inhibitor Protein)
- This differential equation based model was originally published in:
  - Mathematical modelling of the influence of RKIP on the ERK signalling pathway (Cho *et al*., 2003)
- This is one of the models currently being used in the development of our concurrency based approaches
- Therefore, this model was recreated to enable direct comparisons between the different modelling approaches to be made:
  - Compare simulation results
  - Compare performance
  - What types of analysis can be performed
- The RKIP model is a small and relatively simple model that was chosen as a preliminary test model.
- It is not a 'complete' model of the MAPK pathway as the receptors, adaptor proteins and Ras are not considered



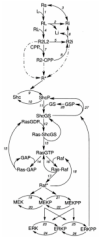


Published MAPK Models

- Currently, there are a variety of published differential equation based models of the MAPK pathway (activated by EGF)
- These models all differ in the way they represent the MAPK pathway
  - Differ in the way some processes are modelled
  - Differ in what proteins are involved in the pathway
  - Differ in the reactions particular proteins are involved in
- A number of these models were recreated and analysed to assess their relative strengths and weaknesses, compare their kinetic data and to see how they performed

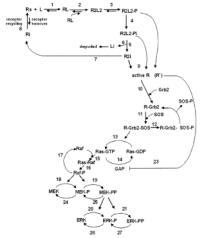


Kholodenko *et al*., 1999

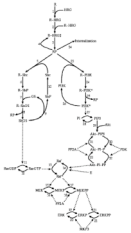


Brightman & Fell, 2000

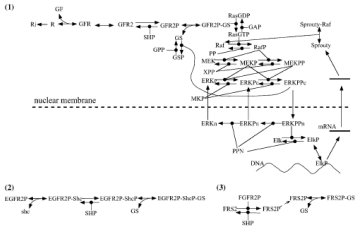
Schoeberl *et al*., 2002



Aksan & Kurnaz 2003



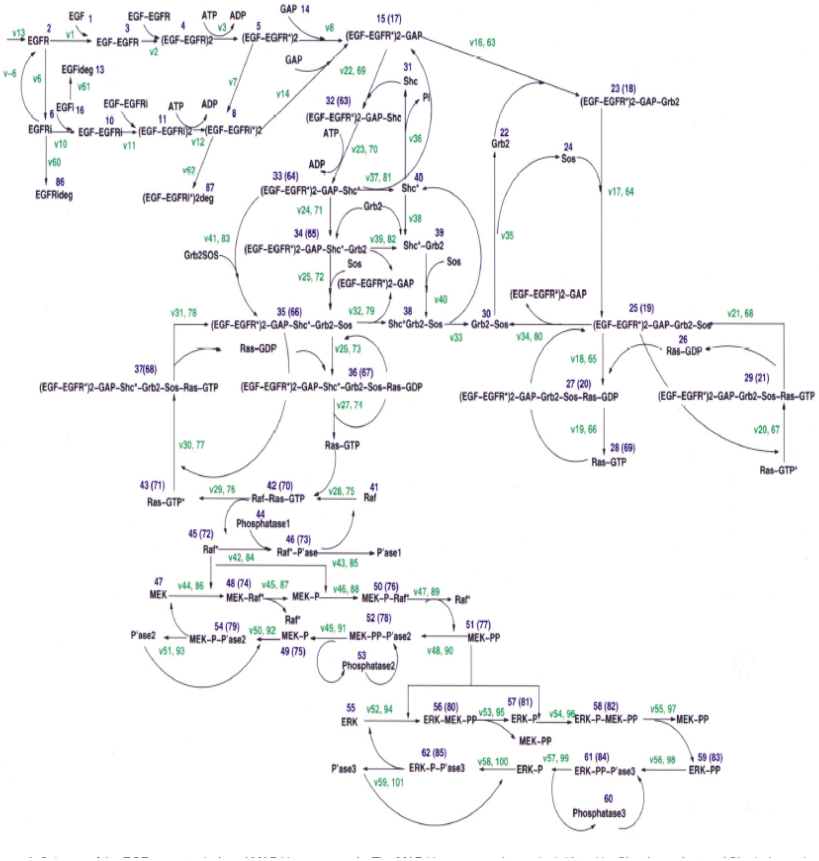
Hatakeyama *et al*., 2003



Yamada *et al*., 2004

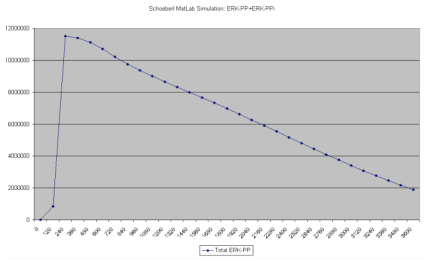
The Schoeberl MAPK Model

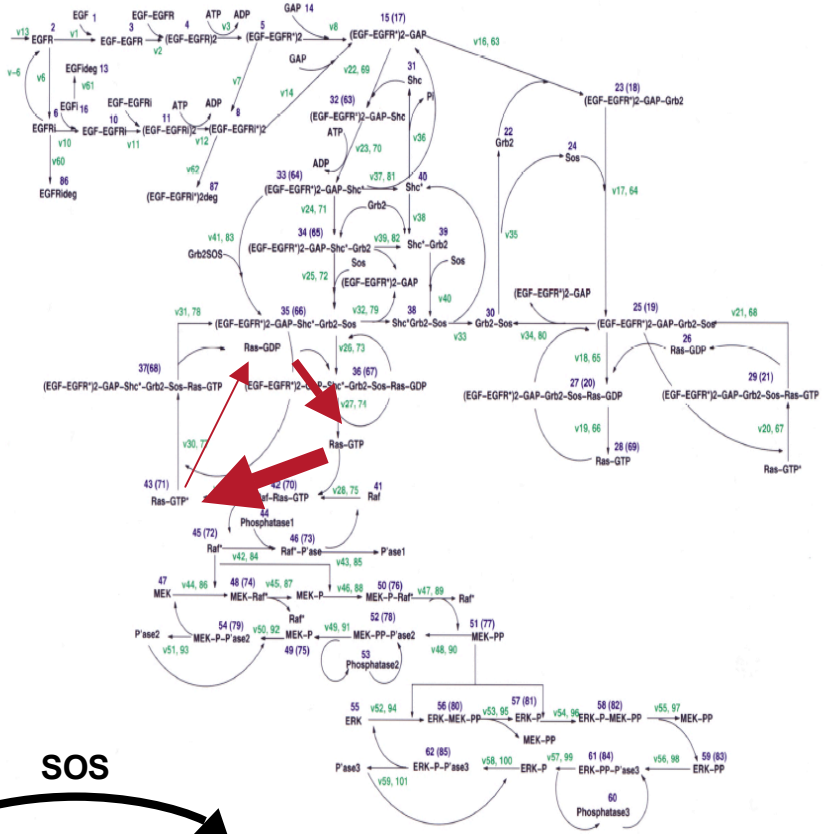
- The Schoeberl model is one of the most comprehensive models of the MAPK pathway available:
  - Computational modelling of the dynamics of the MAP kinase cascade activated by surface and internalised EGF receptors (Schoeberl *et al*. 2002)
  - 125 reactions
  - 94 species
  - Receptor complex strategy
  - Receptor internalisation
  - Shc dependent & independent pathway
- However, it does have a number of errors..



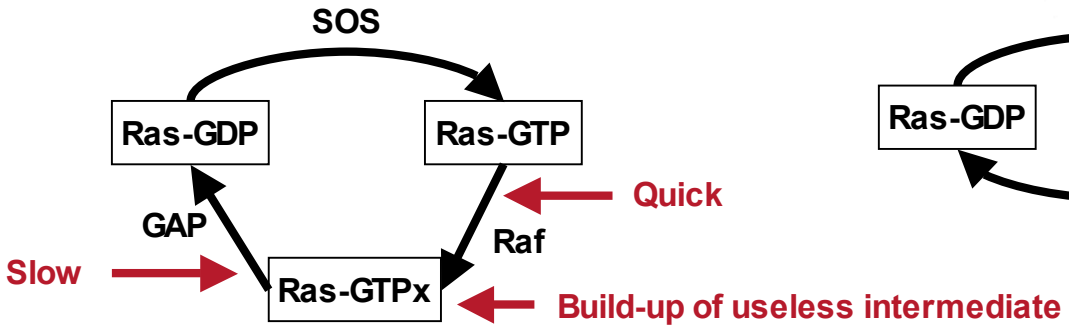
Model Problems

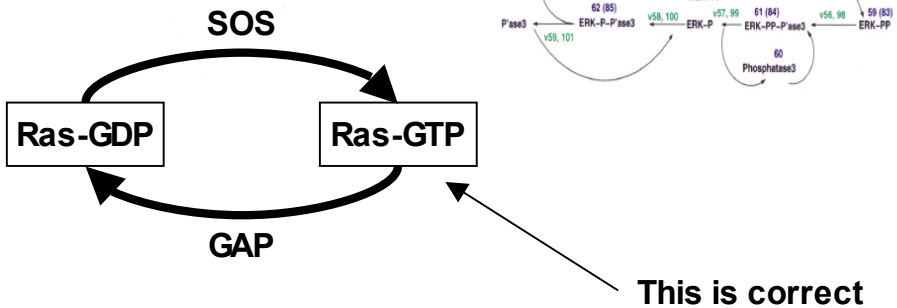
- The major error in this model is that there is no negative feedback loop:
  - ERK-PP (via MAPKAP1) should phosphorylate SOS causing it to dissociate from the receptor complex forming a negative feedback loop
- So why is ERK activation transient and not sustained?





- The deactivation of Ras is incorrectly modelled



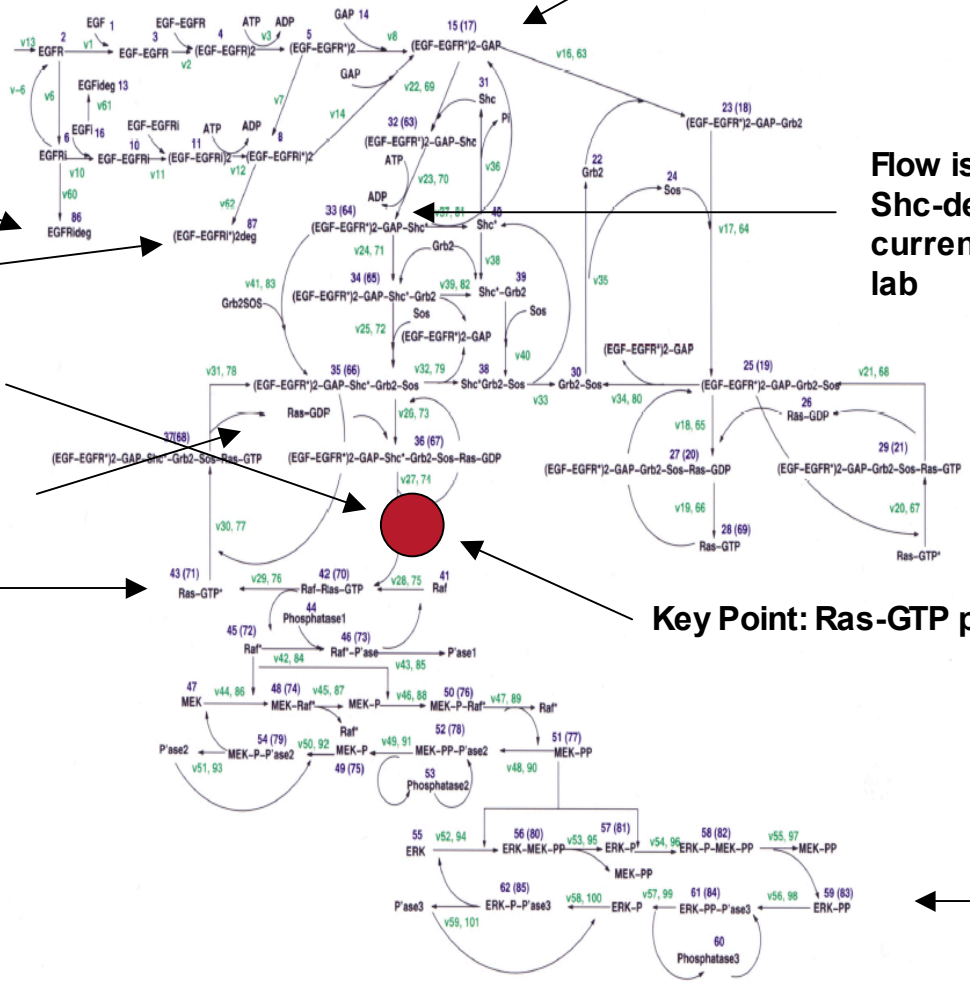


Model Flow

Molecular Flow

Start

Branch Point



Does not restart (no oscillations) as by this time too many receptors have been internalised and degraded

Flow is predominately down the Shc-dependent pathway: currently confirming in the wet lab

Lack of Ras-GTP to keep Raf and therefore MEK and ERK activated

Converted slowly back to Ras-GDP

Build up of useless intermediate

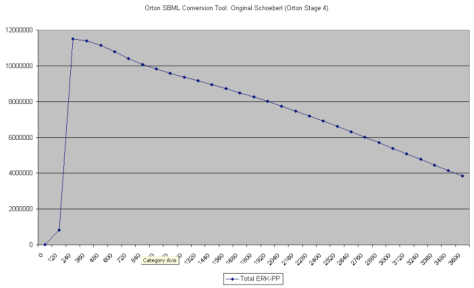
Key Point: Ras-GTP produced

ERK Activated:

But only transiently

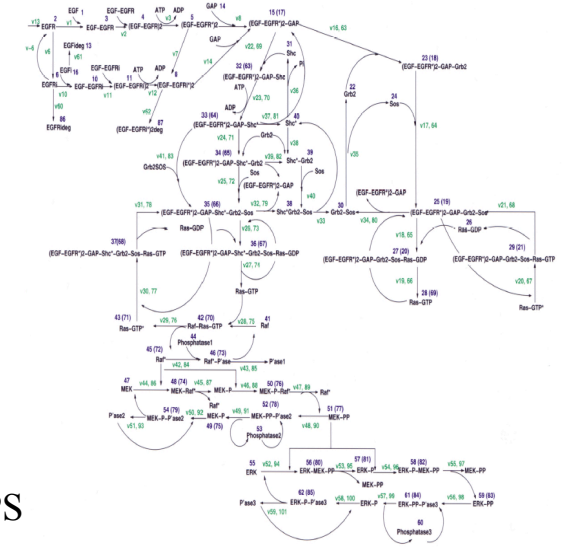
Model Development

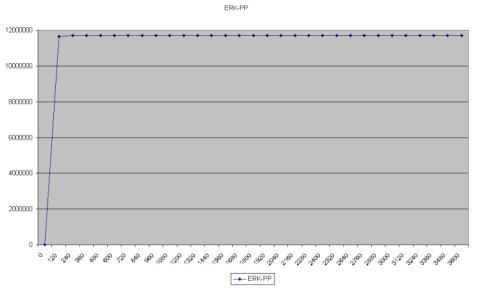
- We are now in the process of developing and validating our own model of the MAPK pathway
- We are using the Schoeberl model as a base to develop from: it is one of the most comprehensive models of the MAPK pathway available, it agrees well with experimental data and it has been used in further analyses by various groups
- Most importantly we were able to identify and fix all of the errors (real errors and significant simplifications)
- Fixing the major (Ras-GTPx) error caused a switch in ERK behaviour from transient to sustained
- This is because there is no now no negative feedback loop to deactivate the signal



ERK-PP before fixes

(transient)





ERK-PP after fixes

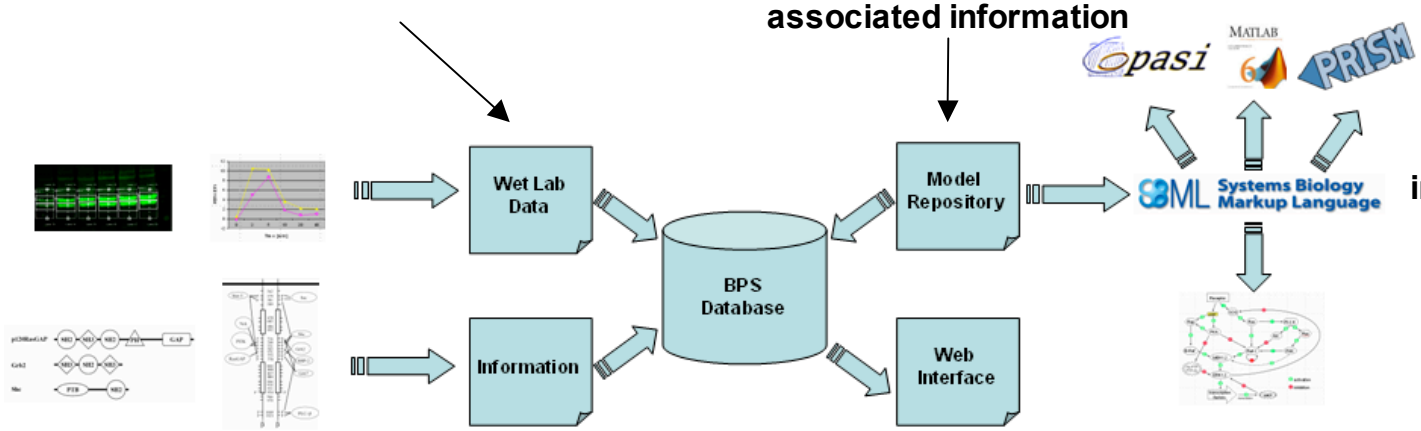
(sustained)

BPS Database and Tools

- We are currently developing a database to store a variety of the data generated from the project and a number of software tools to aid in the modelling process

Wet Lab Section to store modelling data generated from the wet lab

Model Repository Section to store the final and development versions of the models and associated information



Use friendly Model Construction Tool including version handling and tracking, model conversion and visualisation.

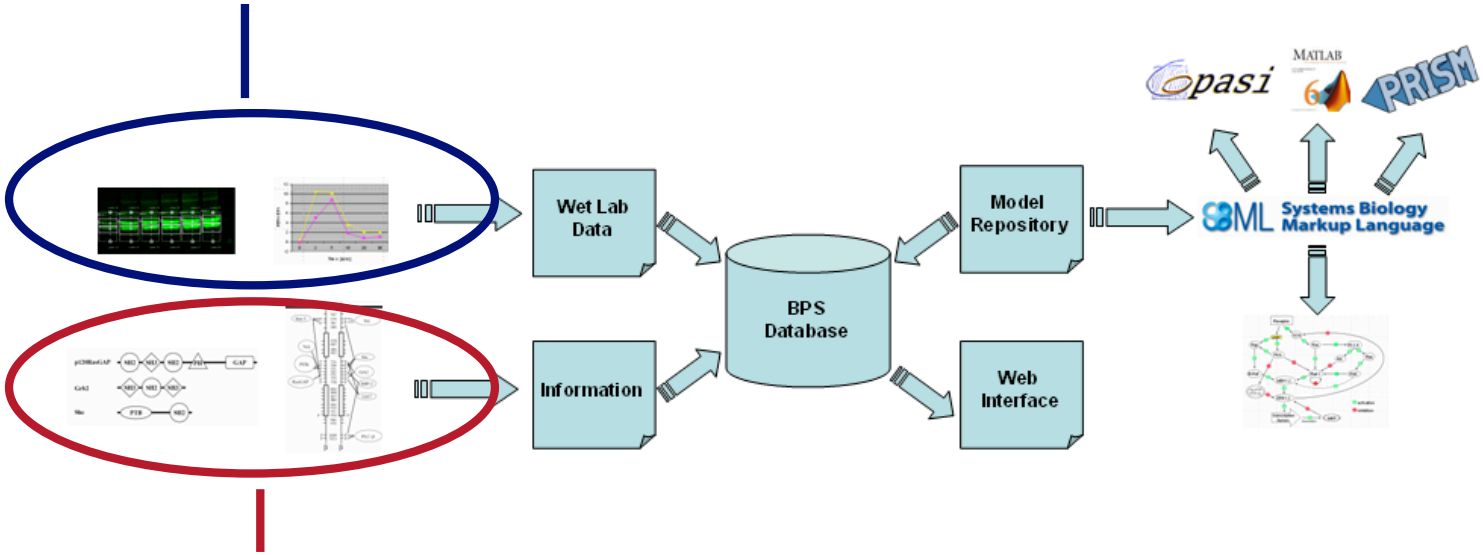
Information on the proteins involved in signal transduction: sequence, domains, structure, annotations, descriptions and references

Web interface to the database and tools

In this system we will be able to link parameter data in the models to the experiments in the wet lab that determined them. Furthermore, biochemical knowledge will be linked to all of the proteins used in the models.

BPS Database and Tools

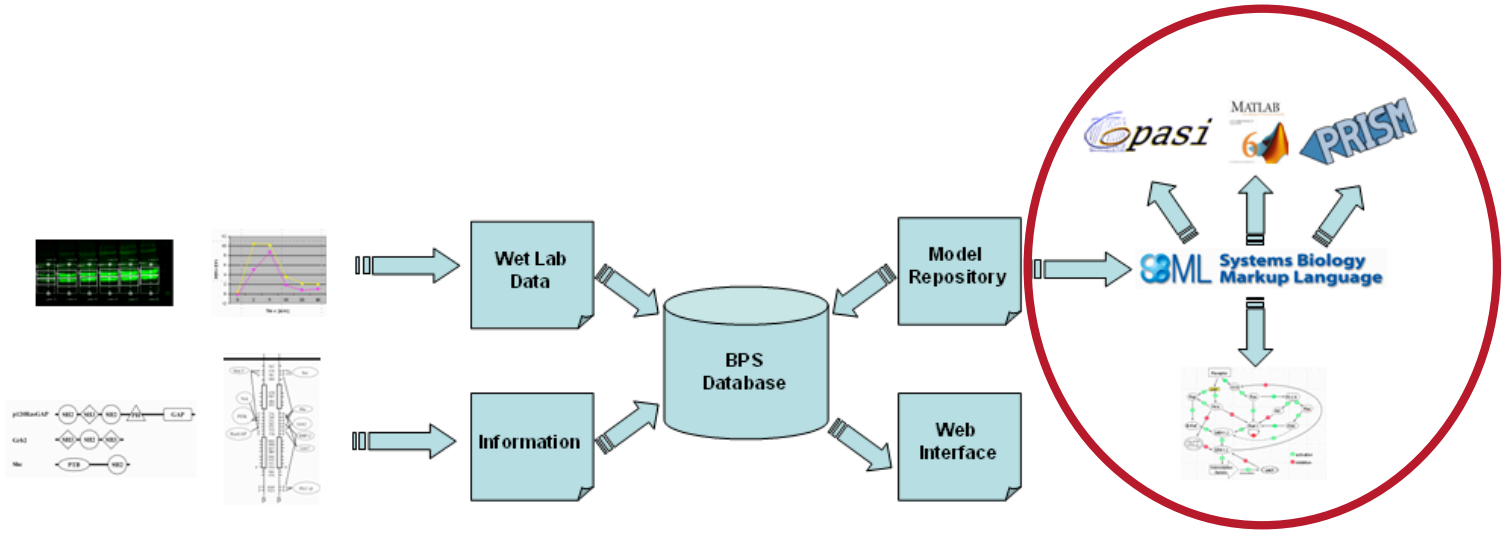
- We are currently developing a database to store a variety of the data generated from the project and a number of software tools to aid in the modelling process
- Modelling data generated from the wet lab will be stored in the database



- Biochemical information on all the proteins in the models, as well as other signal transduction proteins, will also be stored in the database e.g.:
  - Sequence, Domains, Structure, Annotations, Descriptions, Reactions, References

Model Construction Tool

- User friendly Model Construction Tool to construct biochemical models, define reaction kinetics and assign parameter data.
- This tool will also be able to graphically display models
- This tool will be SBML compatible and have a number of built-in functions to convert models to other formats e.g. MatLab and PRISM
- The tool will include a model version tracking system that tracks all changes to a model from its creation and also hierarchically links it to related models.
- The tool will run off the BPS database where the models will be stored and integrated with other data
  - Parameter data can be linked to the wet lab experiments that determined them
  - Biochemical information on any of the proteins in a model can be readily accessed
- The entire system will be web based
- Currently working with Vlad to develop database schema and tool overview

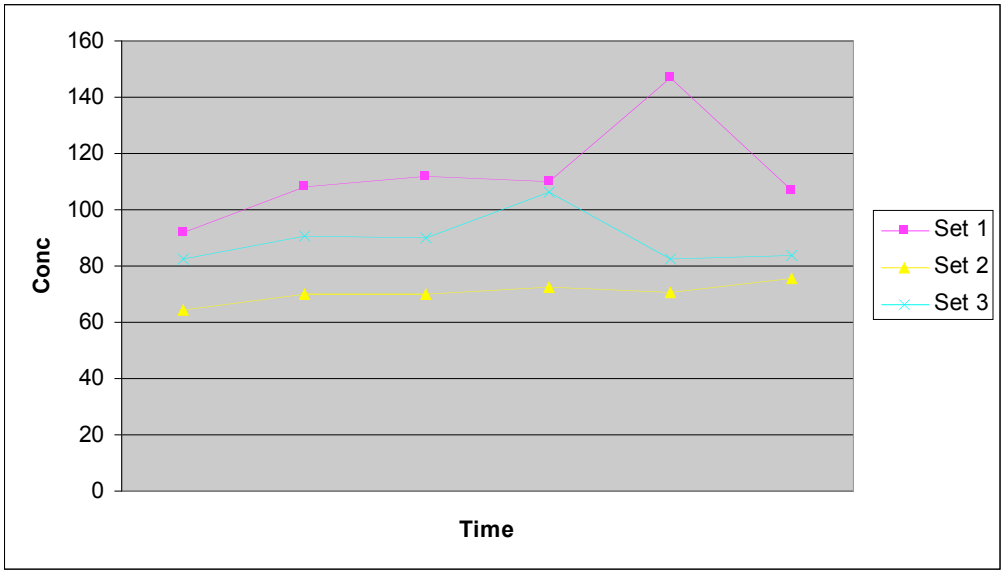


The Qualitative vs Quantitative Challenge

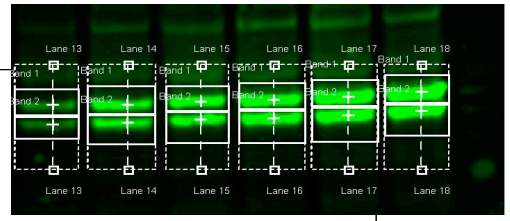
- How to model & analyse biochemical networks when there is a ***lack of exact quantitative data***.
- Traditional approaches based on ***differential equations, which require exact rate constants*** **& [initial] concentrations**
- A pathway modelled by constructing a system of differential equations to describe the changes in concentration of each of the species in the pathway, then solved numerically (‘system simulation').
- A severe limitation of this approach :
  - the rate constants must be known exactly for each reaction, and
  - the concentrations must also be known exactly.
- Obtaining data in sufficient quality and density requires dedicated experimental efforts.
- In addition, for many of the most interesting molecular systems, such as signalling pathways, intracellular concentrations are almost impossible to determine exactly, and associated rate constants are also inexact.

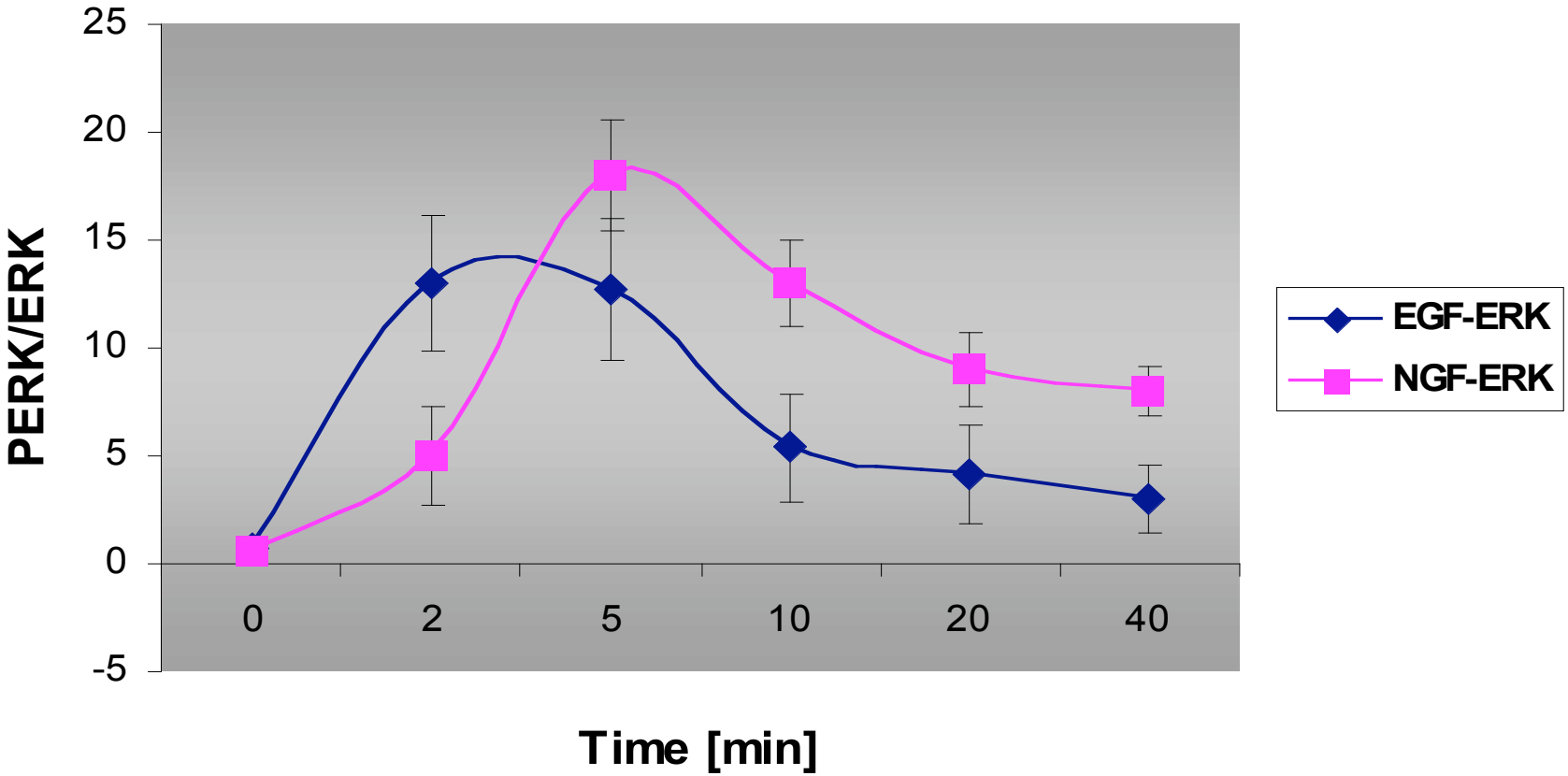
Experimental observations…

Time	Set 1	Set 2	Set 3
0	92.06901	64.25505	82.57557
2	108.3974	69.72303	90.80861
5	111.9357	70.16582	90.28703
10	110.2452	72.76562	106.1241
20	146.861	70.43287	82.39227
40	106.9155	75.45979	83.9535
Average	112.7373	70.46703	89.35684



Multiple observations – error bars?





Parameter fitting

- Popular method to overcome inexact data.

- Variety of techniques:
  - deterministic optimisation
  - random search
  - clustering
  - evolutionary computation
  - Simulated annealing
  - taboo search
  - multiple shooting
  - …

Parameter fitting:

√ Obtain (one) curve

√ Can apply standard simulation & analysis techniques

X Do not preserve the variations possible in the observed values - are only a 'best representative' of a set of possible values.

X Need to be combined with methods like sensitivity analysis to understand the relevance of variability for each parameter.

X Can be misleading in published results

Further limitations of DE's

- Lack of techniques for a systematic analysis of the computed behaviour of the pathways.

- The normal validation of such models is to check by eye that the shape of the curves for the predicted behaviour accords with experimental observations.

- Once a model is 'trusted' it can be used as the basis for predicting the behaviour of the pathway when it is modifed, e.g. by gene knock-out, gene knock-down (RNAi), overexpression, various drug treatments.

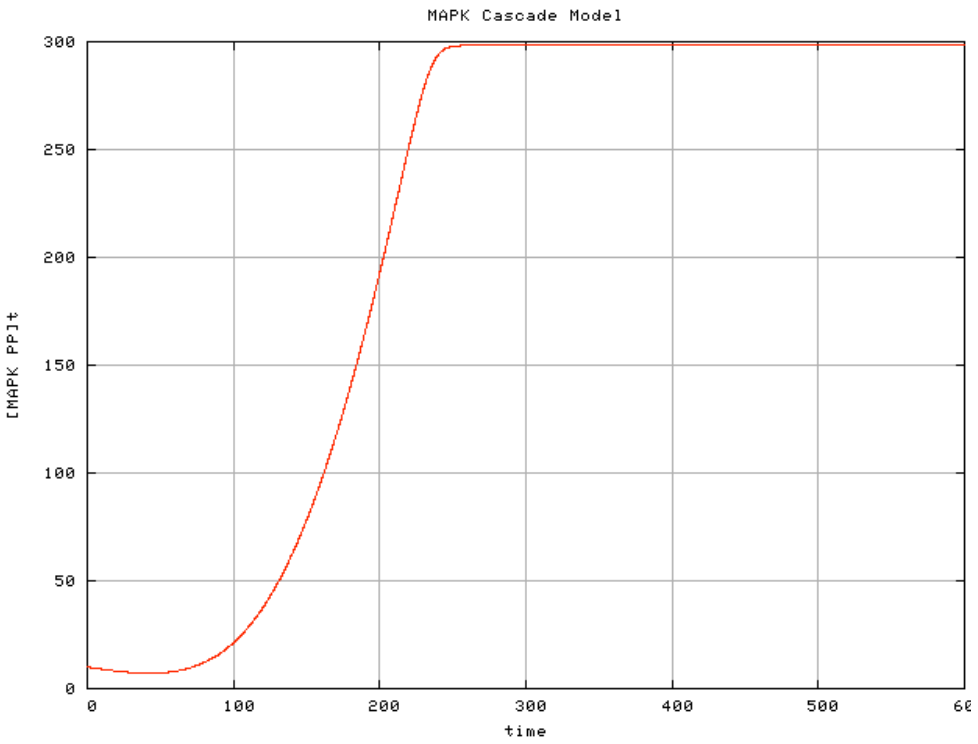
UNIVERSITY
of
GLASGOW

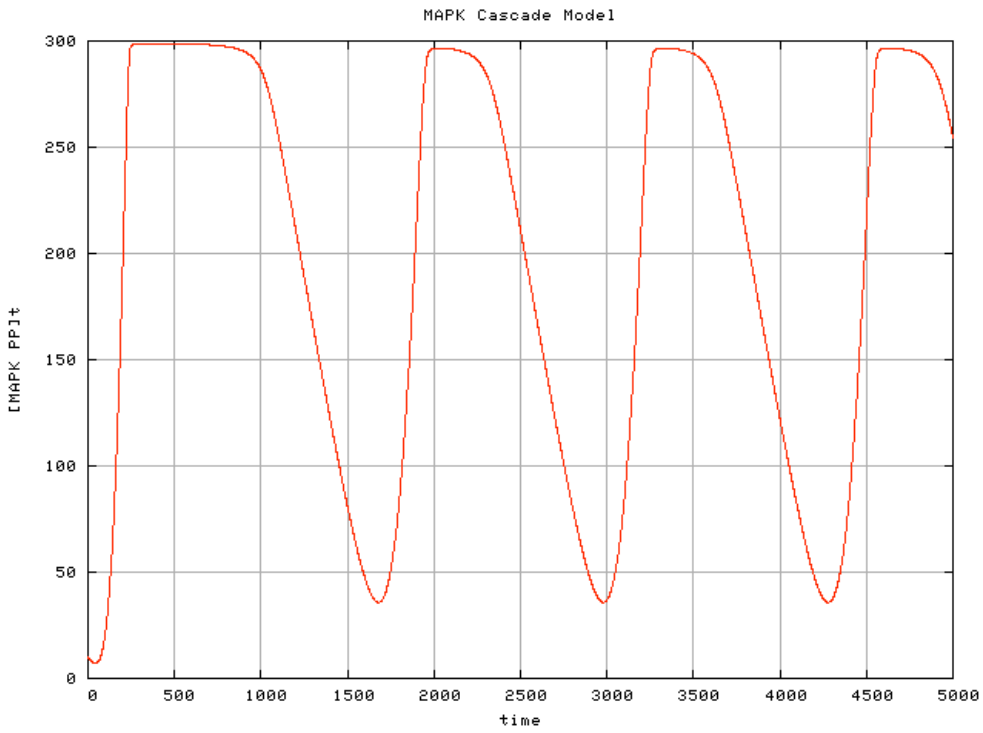
Analysis of Biochemical Networks with the PRISM Model Checker

Vladislav Vyshemirsky
Muffy Calder

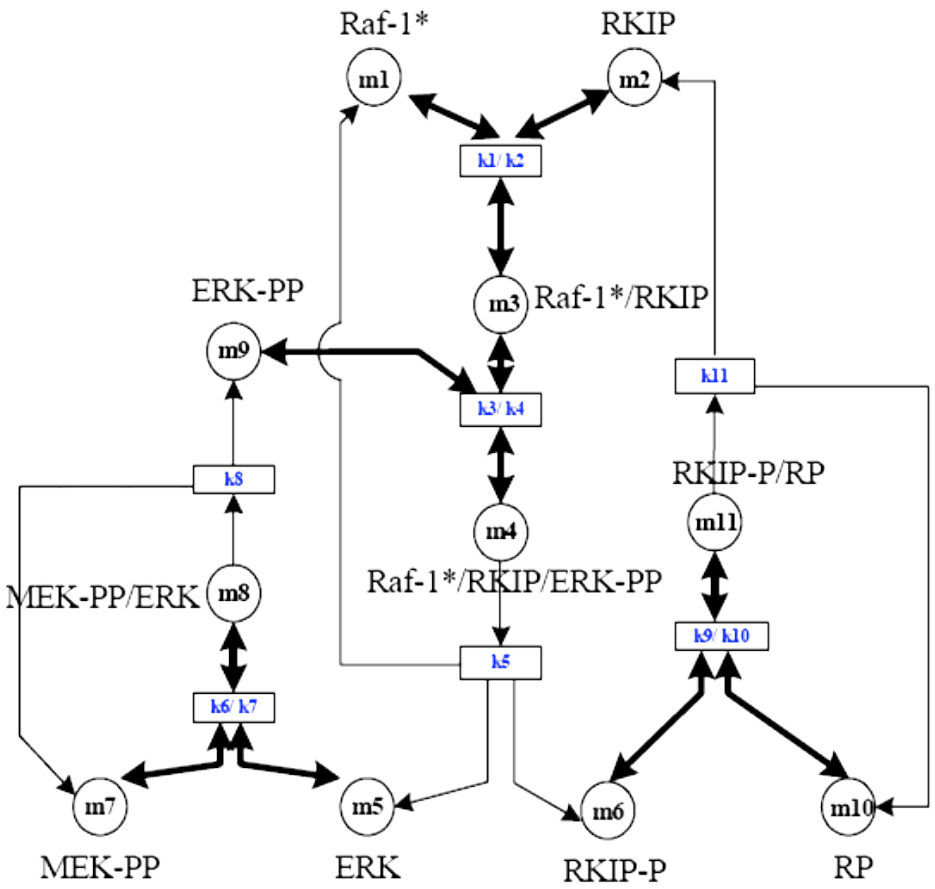
Logical Modelling Motivation

- Reasoning about system, not about observed behaviour





Model



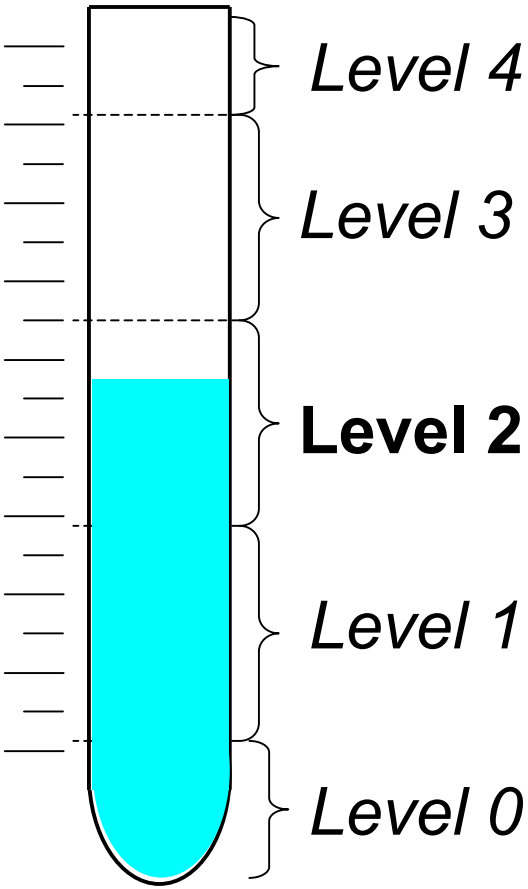
We developed an approach to modelling networks using stochastic process algebras and stochastic model checking and implemented a computational model of the RKIP inhibited ERK pathway.

The model allows evaluation of semiquantitative properties, such as the probabilities of an activation precedence and stable state analysis.

Model Parameters

- Number of Concentration Levels
  - Discrete levels are used to model uncertain data
  - These levels provide an approximation of “master behaviour” defined by ordinary differential equations
  - The concept of levels allows discrete models, and consequently, analysis with model checking tools

Modelling Technique

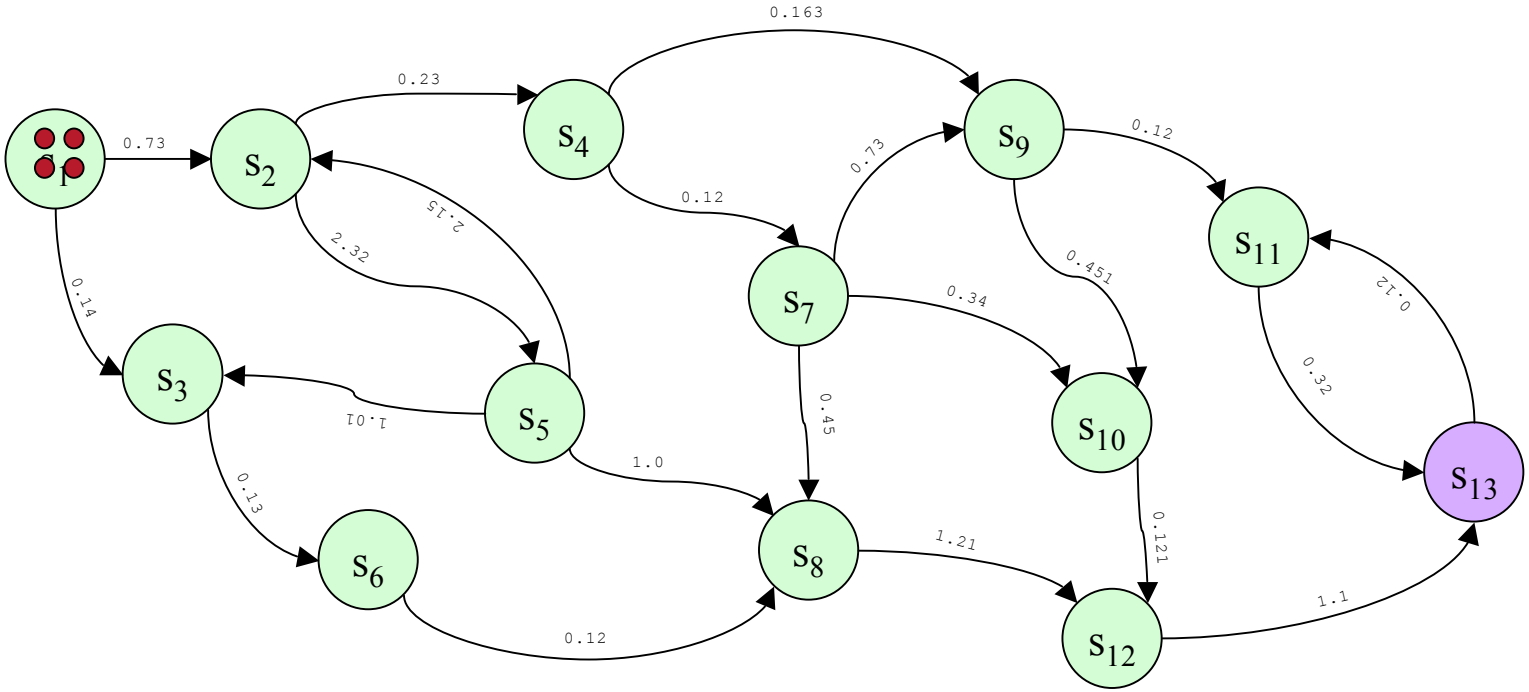


Time **is continuous**

other variables are discrete.

Model Structure

- Model is represented with Continuous Time Markov Chain

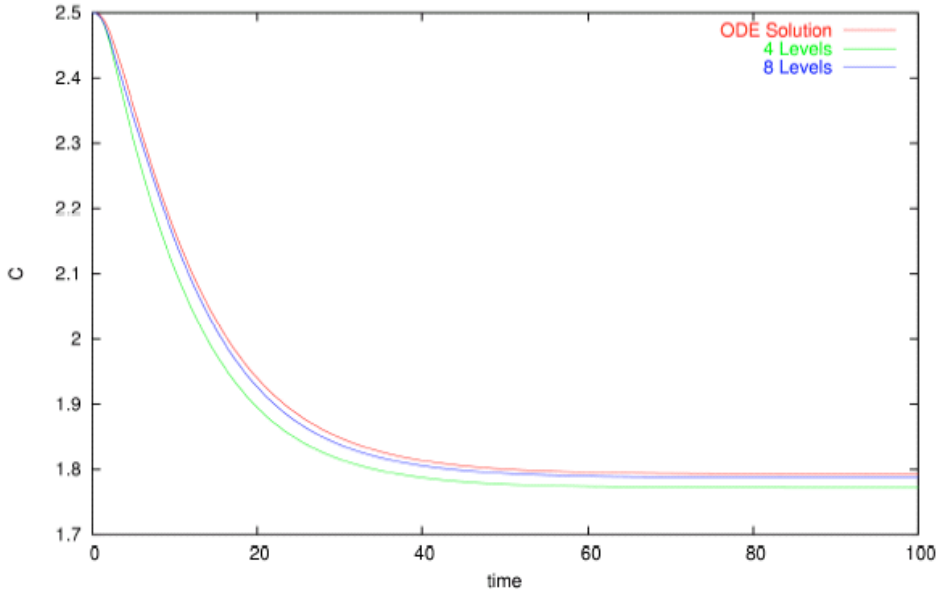


Model Structure

- The internal structure of the model can be found in the submitted papers
  - For CMSB'05: Analysis of Biochemical Pathways with the PRISM model checker
  - For IEEE Transactions in Computational Biology and Bioinformatics: Analysis of Biochemical Pathways with Continuous Stochastic Logic

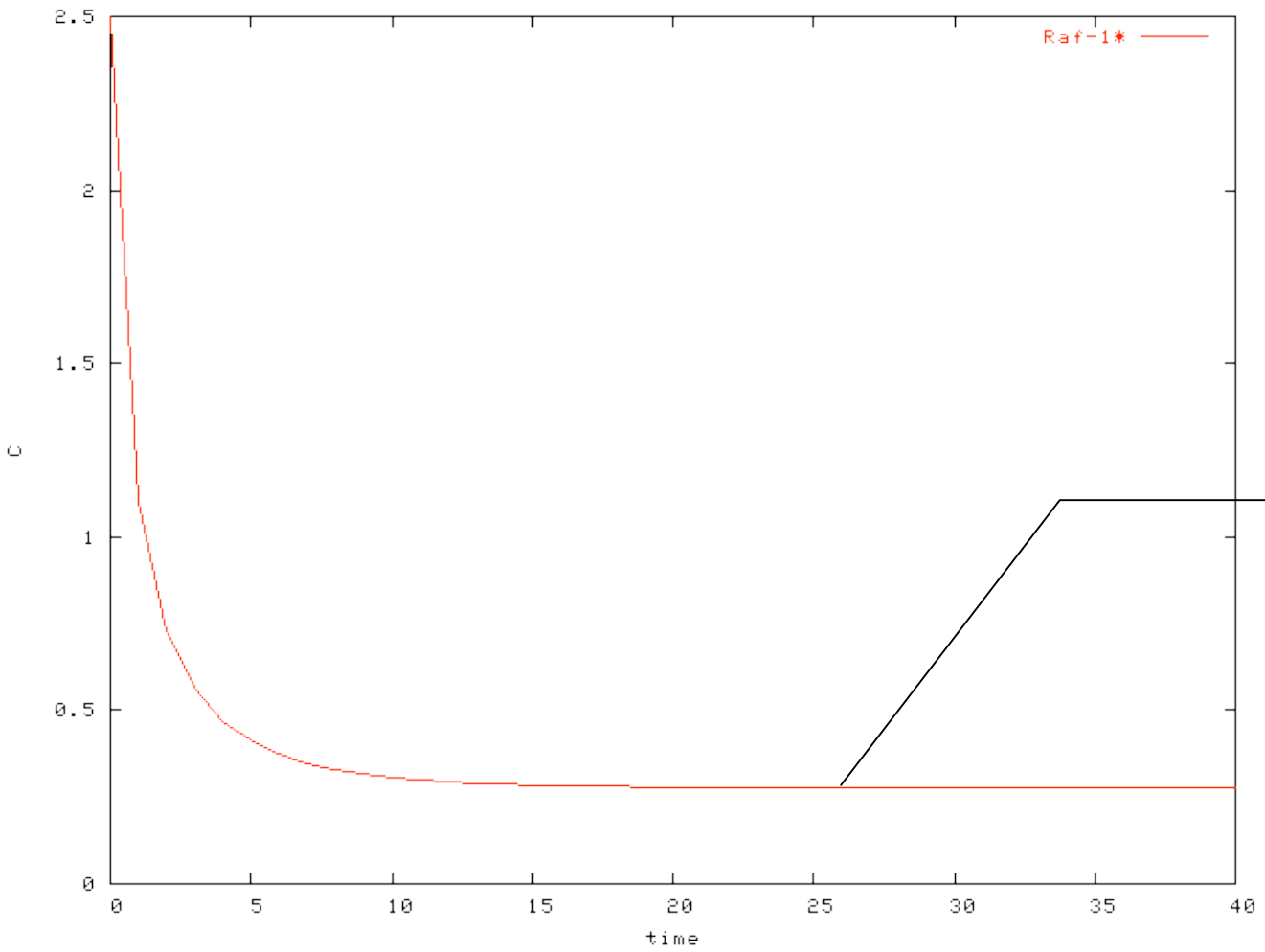
Comparison to ODEs

- Similarity of results to classical ODE model
  - We developed a technique to build model simulation trace
  - Now we can compare the results



Levels	ϵ_a	ϵ_r	$C\epsilon_a$	$C\epsilon_a^2$
4	0.126 mM	0.280	21.557 mM	2.58
5	0.103 mM	0.217	17.569 mM	1.727
6	0.086 mM	0.176	14.582 mM	1.191
8	0.061 mM	0.122	10.402 mM	0.605
12	0.036 mM	0.071	6.042 mM	0.204

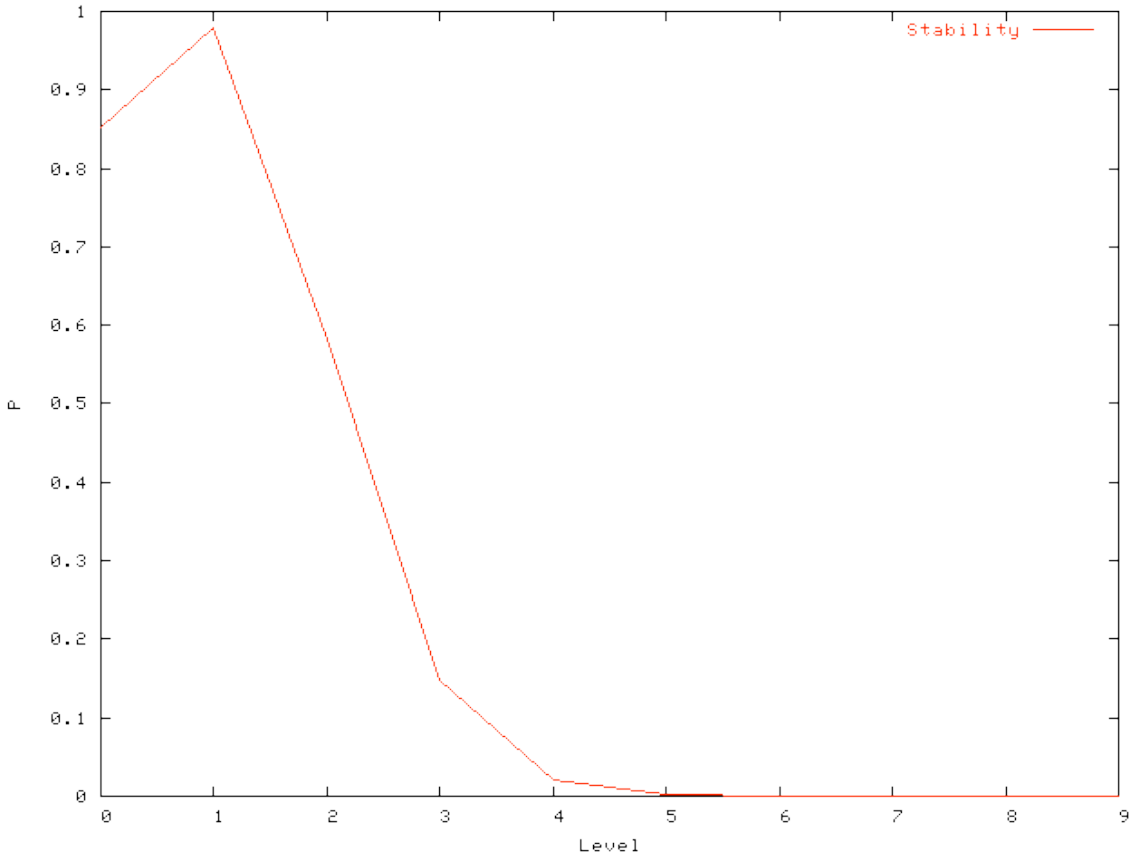
Steady State Analysis of Stability



The query is: "What is the probability that the protein will stabilise on some level of concentration?"

Protein becomes stable from some point of time

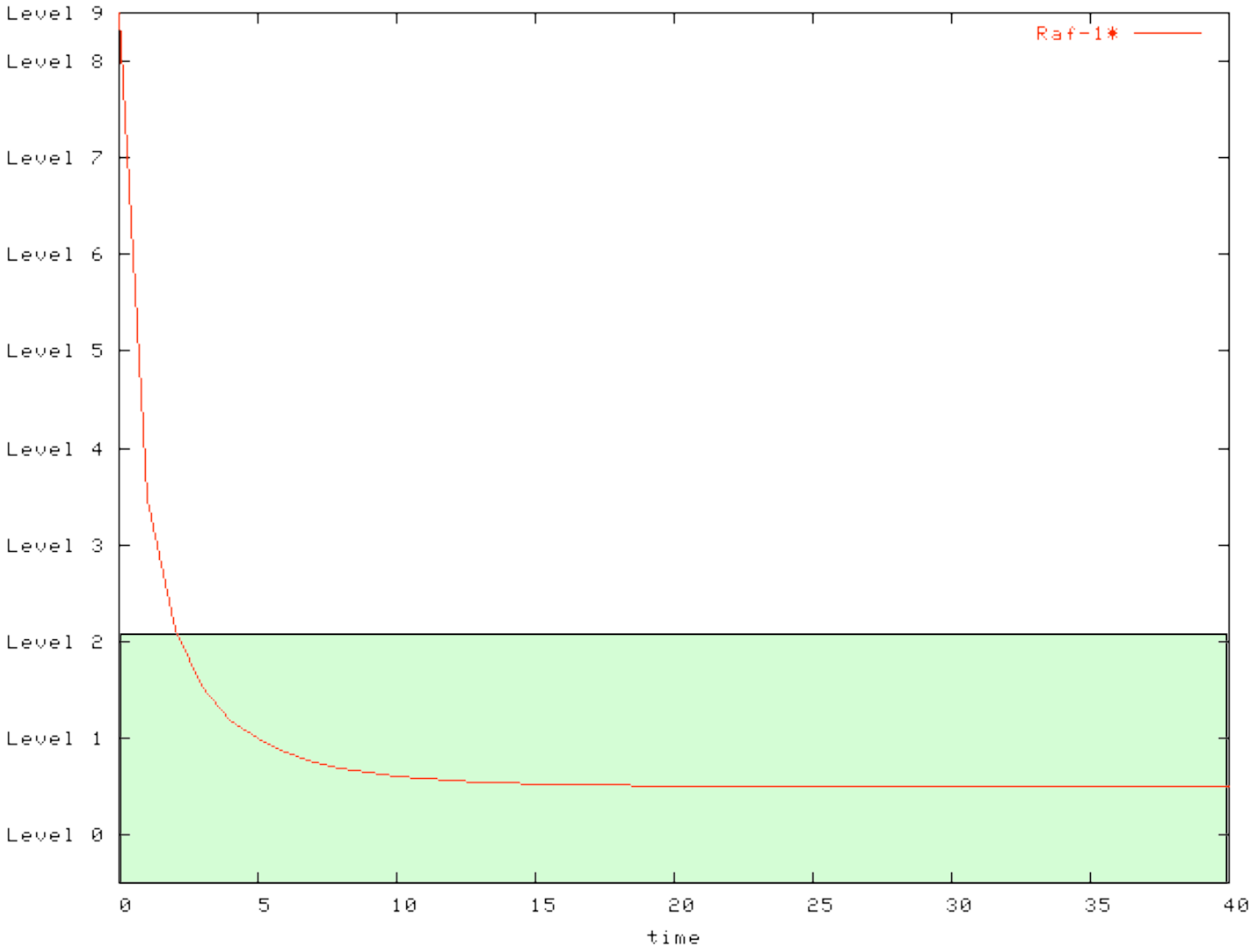
Steady State Analysis of Stability



The evaluation of the steady state probabilities in the model with 10 discrete levels.

The probabilities to be stable at levels between 0 and 2 are quite high, when the other levels are less likely to be stable

Steady State Analysis of Stability



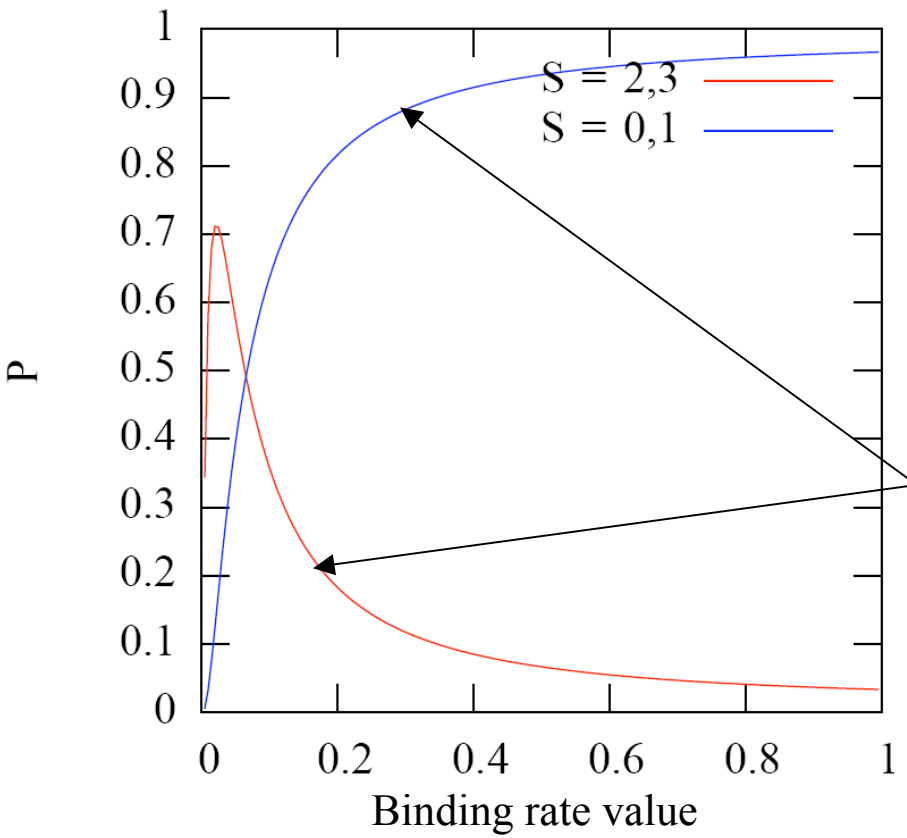
The high probable stability area is shaded green.

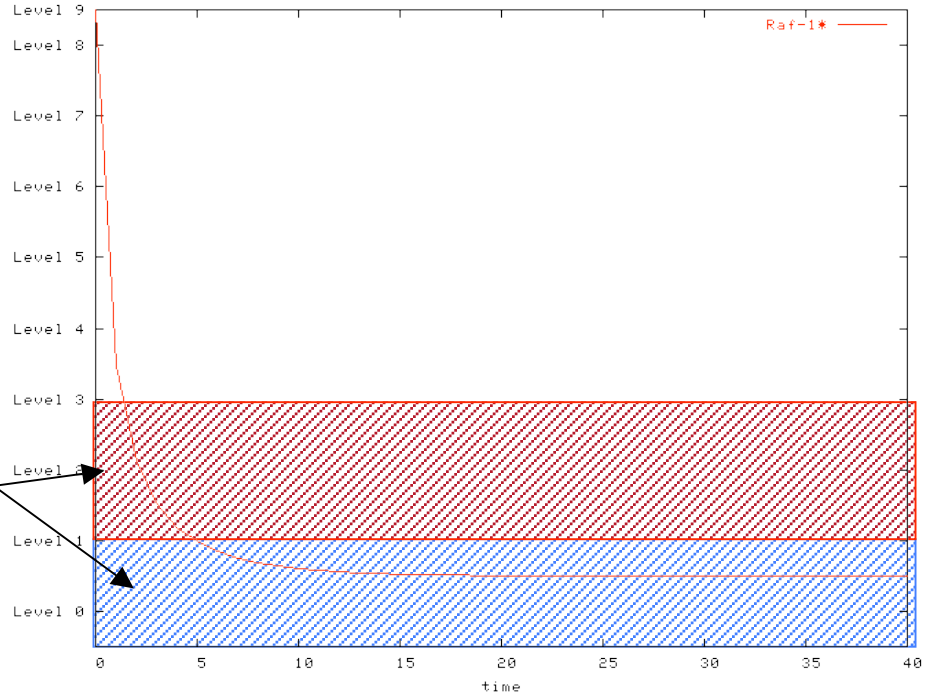
More levels of discrete concentration will help to make this area smaller.

Steady State while varying Parameters

- The value of the stable state can change if the parameters of the model are changed.
- This experiment considers the change of the steady state while the rate of protein binding is changed in some interval.

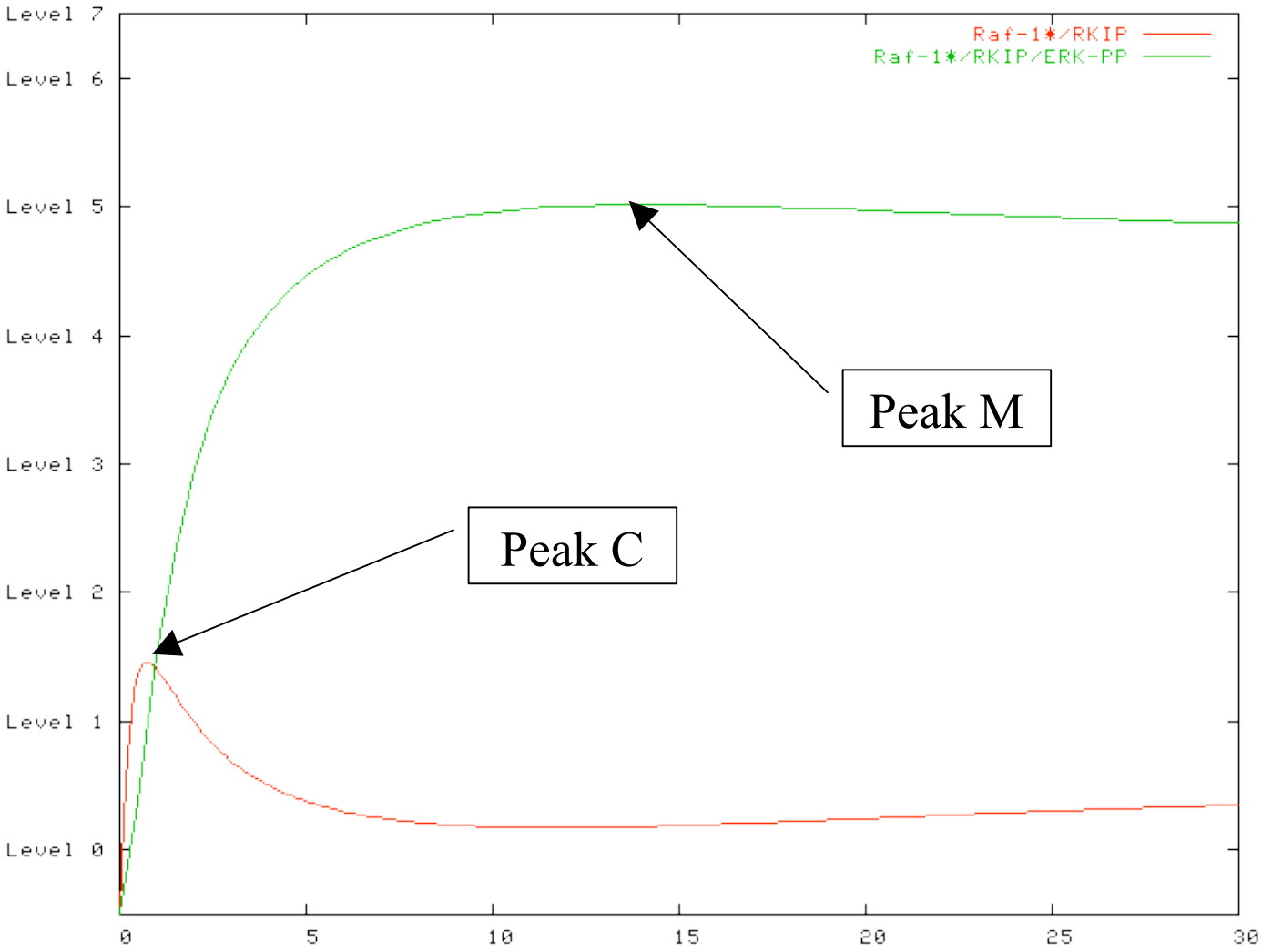
Steady State while varying Parameters





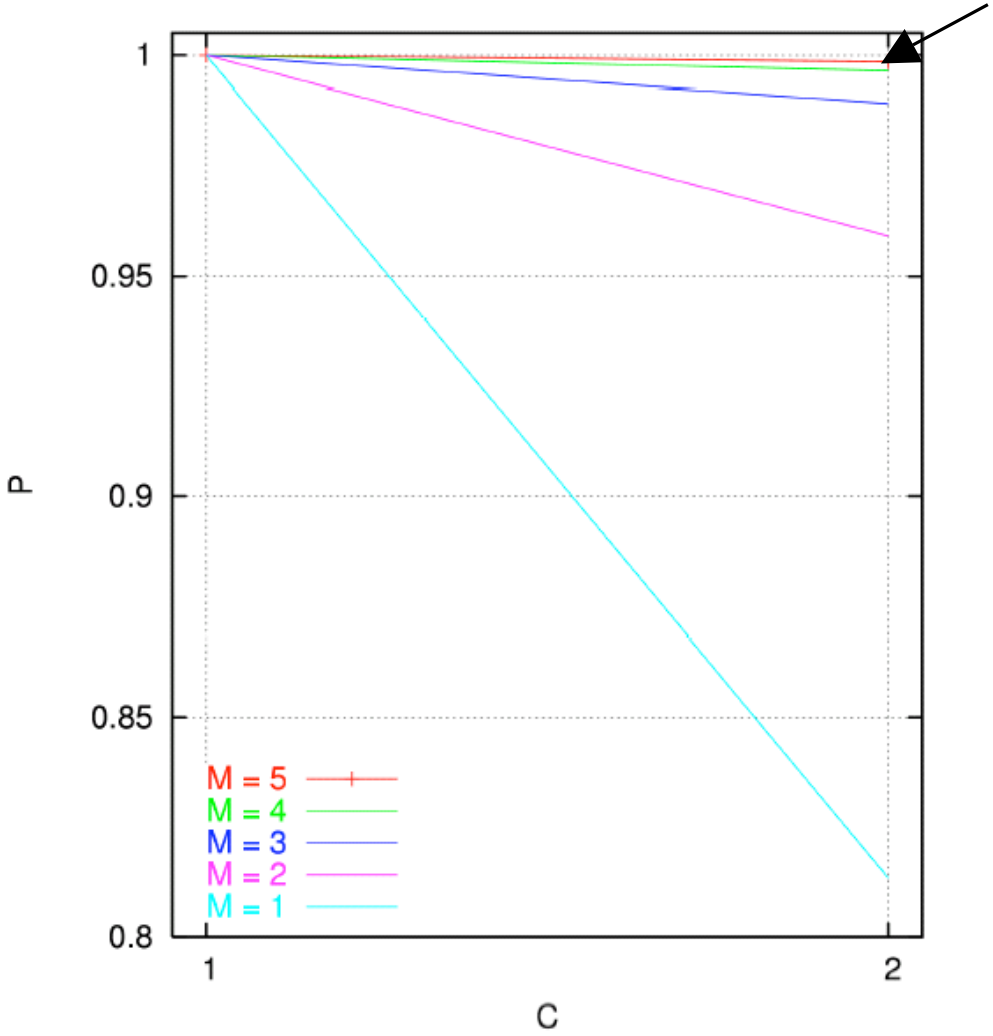
When the binding rate is increased, the probability to stabilise on levels 2 or 3 (red square) falls down, and the probability to stabilise on levels 0 or 1 (blue square) rises.

Activation Sequence Analysis



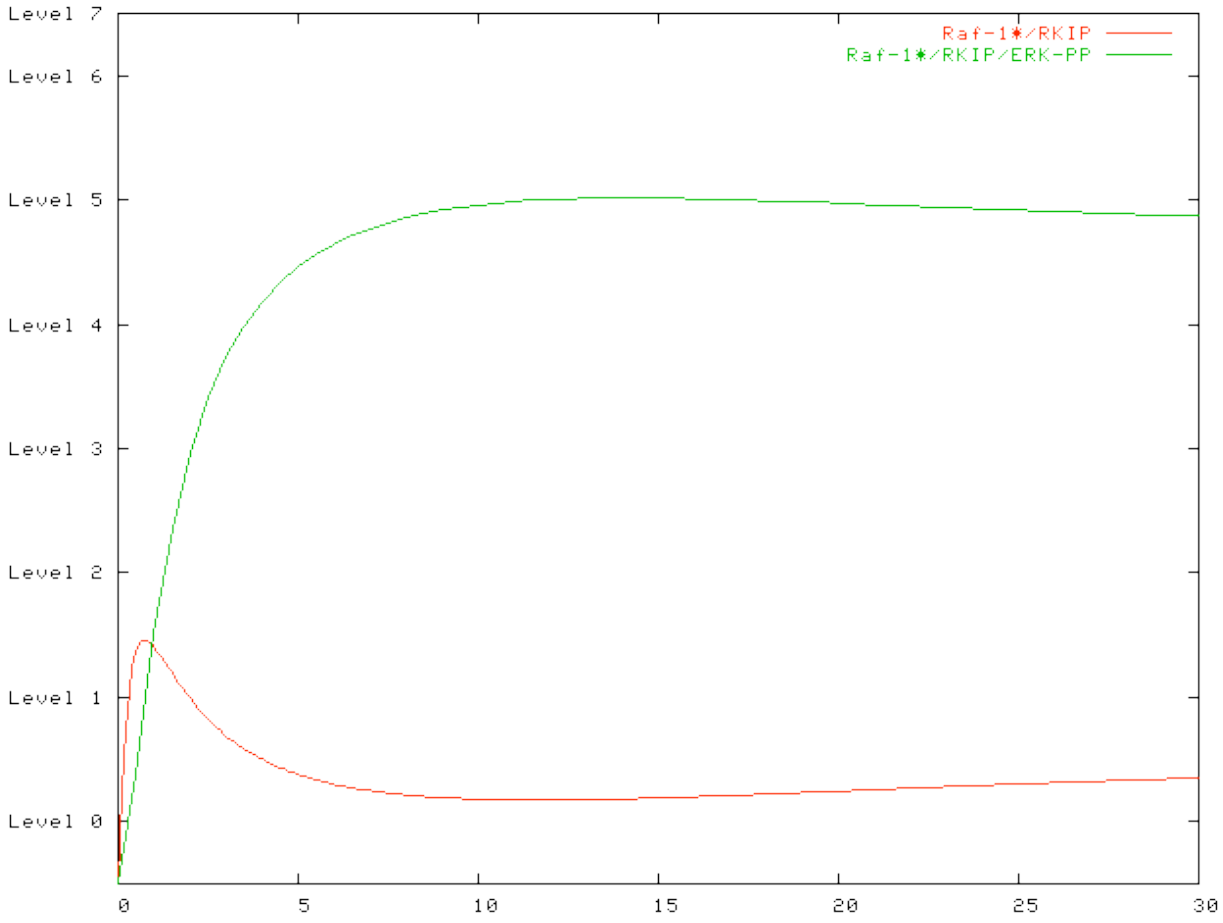
This kind of analysis helps to decide whether the sequence of activation is stable (high probable) or not.

Activation Sequence Analysis



This point shows, that peak C of “red” protein will be before peak M of “green” protein with the probability over 98%

Activation Sequence Analysis



The probability that the "red" protein will reach level 2 before the "green" one reaches level 5 is more than 98%.

The probability that the "red" protein will reach level 2 before the "green" one reaches level 2 is almost 96%.

Conclusion: This activation sequence is very stable.

Problems

- Scalability
  - We already work with practical networks
  - Of the small size
- Properties Library
  - We need to develop a library of meaningful properties

Take home messages

- Zoom in / zoom out modelling…
  (fine/coarse grain)
- Building block approach
- Model validation in-house (wet-dry team)
- More than simulation – (logical) analysis
- Future – software platform
- Interdisciplinary research is exciting!

BPS: Biochemical Pathway Simulator
A Software Tool for Simulation & Analysis of Biochemical Networks

www.brc.dcs.gla.ac.uk/projects/bps

University of Glasgow
Bioinformatics Research Centre
Beatson Laboratories

David Gilbert (PI)	drg@brc.dcs.gla.ac.uk
Muffy Calder	muffy@dcs.gla.ac.uk
Walter Kolch	wkolch@beatson.gla.ac.uk
Richard Orton	rorton@dcs.gla.ac.uk
Vlad Vyshemirsky	vvv@dcs.gla.ac.uk
Oliver Sturm	sturmo@dcs.gla.ac.uk
Amelie Gormand	agormand@dcs.gla.ac.uk

Contact Details:
Prof. David Gilbert, Director
Bioinformatics Research Centre
A416, Fourth Floor, Davidson Building
University of Glasgow
Glasgow G12 8QQ
Scotland UK

0141 330 2563