

# Modelling the MAPK pathway

## Richard Orton

## Overview

- Modelling the MAPK pathway with differential equations
- Modelling
  - What is modelling?
  - Why model?
  - How to model
  - RKIP Model
  - Gepasi
  - SBML
- Example: MAPK Pathway
  - What is it?
  - Why is it important?
  - Why model it?
  - How to model it?

2

## What is modelling?

- In this context:
  - Translating a biological pathway into mathematics for subsequent analysis

Translating a biological pathway

Into mathematics

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

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

[A] = 10; [B] = 0; k1 = 2; k2 = 1; Time = 10

For subsequent analysis

3

## Why model?

- Simplistic answers:
  - Because it's there...
  - Why not?
- Technical answer:
  - "The benefit of formal mathematical models is that they can show whether proposed causal mechanisms are at least theoretically feasible and can help to suggest experiments that might further discriminate between alternatives." (Franks & Tofts, 1994)
- Realistic answers:
  - A computer model can generate new insights
  - 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 (could/is the proposed mechanism correct)

4

## Why model?

- In a complex pathway, knowing all the proteins involved and what they do, may still not tell you how the pathway works
- Furthermore, if all the initial concentrations and rate constants are known in the pathway, a computer simulation will probably still be needed to show how the system behaves over time

5

## How to model

6

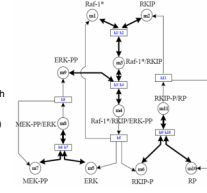
## How to model...1: Identification

- Identify the biological pathway to model (what)
  - RKIP
  - EGF and NGF activated MAPK
- Or, more importantly, identify the biological question to answer (why)
  - What influence does the Raf Kinase Inhibitor Protein (RKIP) have on the Extracellular signal Regulated Kinase (ERK) signalling pathway?
  - How do EGF and NGF cause differing responses in ERK activation, transient and sustained, respectively?

7

## How to model...2: Definition

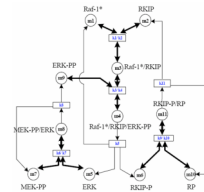
- This is the key step and is not trivial
- Draw a detailed picture of the pathway to model
  - Define all the proteins/molecules involved
  - Define the reactions they are involved in
  - Where do you draw the model boundary line?
- Check the literature
  - What is known about the pathway and proteins?
  - What evidence is there that protein A binds directly to protein B?
  - Protein C also binds directly to protein B: does it compete with protein A or do they bind to protein B at different sites?
  - Trust & Conflicts: it is important to recognize which evidence to trust and which to discard (talk to the people in the wet lab)
- Simplifying assumptions
  - Many biological processes are very complex and not fully understood
  - Therefore, developing a model often involves making simplifying assumptions
  - For example, the activation of Raf by Ras is very complicated and not fully understood but it is often modelled as:
    - $Raf + Ras-GTP \rightarrow RafRas-GTP \rightarrow Raf^* + Ras-GTP$
  - Although this is a simplification, it is able to explain the observed data



8

## How to model...2: Definition

- Define the kinetic types
  - Each reaction has a specific kinetic type
  - All the reactions in the RKIP model are mass action (plain, uncatalysed kinetic type):
    - $V = k_1[m][n_2] + k_2m^3$
  - Another common kinetic type is Michaelis Menten (enzyme catalysis):
    - $V = V_{max}[S] / (K_m + [S])$
- Define the rate constants (k's, km's, Vmax's etc)
- Define the initial concentrations
- Check the literature
  - What values have been previously reported?
  - What values are used in similar models?
  - Do you trust them? Are there any conflicts?
  - Measure them yourself in the wet lab
  - Parameter estimation techniques: estimate some parameters based on others and observed data



9

## How to model...3: Simulation

- Once the model has been constructed and parameter data has been assigned you can simulate (run) the model
- This is a relatively straightforward step as there are many software tools available to simulate differential equation based models
- For example:
  - MatLab
  - Gepasi
  - CellDesigner
  - Janac
  - WinScamp
  - Many many more
- Runtime options include setting the time to run the model for and the number of data points to take



10

## How to model...4: Validation

- Simulating the model typically returns a table of data which shows how each specie's concentration varies over time
- This table can then be used to generate graphs of specie concentrations
- Do the model results match the experimental data?
  - Yes: validation
  - No: back to definition and check for errors
    - Simple types
      - Wrong kinetics
      - Over simplifications of processes
      - Missing components from the model
      - Incorrect parameter data
- The model can then be validated further by checking the system behaves correctly when things are varied:
  - It might be known how the system behaves when you over-express or knockout a component
  - The model should be able to recreate this behaviour
- If the model's results do not match known biology, we cannot rely on predictions about unknown biology

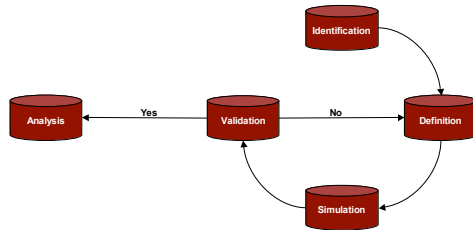


## How to model...5: Analysis

- After the model has been validated we can then analyse and interpret the results
  - What do the results imply or suggest?
  - What do they tell us that is new and that we did not know/understand before?
  - What predictions can we make?
- Sensitivity analysis can be used to identify the key steps and components in the pathway as well as monitoring how robust the system is:
  - Vary an initial concentration or rate by a small amount and see what affect it has on the system as a whole: small changes in a key value are likely to have a large affect
  - How robust is the system to changes?
- Knockout experiments are easy to do in a model: for example, simply set the initial concentration of the desired component to 0
  - Knockout experiments can be used to identify which components are essential and which are redundant
  - Can also knockout reactions (set rate to 0) to identify essential and redundant reactions in the system

12

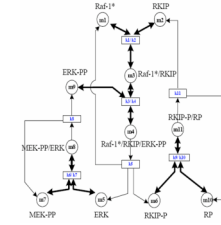
## How to model... Overview



13

## The RKIP Model

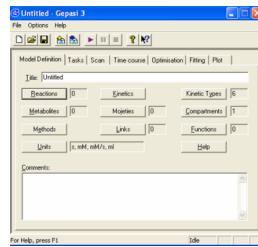
- RKIP (Raf Kinase Inhibitor Protein)
- This differential equation based model was originally published in:
  - Mathematical modelling of the influence of RKIP on the ERK signalling pathway (Cho et al., 2003)
- It is a relatively small and simple model:
  - 11 species
  - 11 reactions
  - Mass action
- This was introduced last week: MatLab



14

## Gepasi

- GEPASI: General Pathway Simulator
- GEPASI is a software package for modeling biochemical systems
- GEPASI integrates the systems of differential equations using LSODA (Livermore Solver of Ordinary Differential Equations)
- Performs well and has a lot of features but has limited graphical displays and is only available on the Windows platform



15

## Gepasi Demo

- Gepasi demo of the RKIP model

16

## SBML: <http://www.sbml.org>

- The Systems Biology Markup Language (SBML) is a computer-readable format for representing models of biochemical reaction networks. SBML is applicable to metabolic networks, cell-signaling pathways, regulatory networks, and many others.
- SBML has been evolving since mid-2000 through the efforts of an international group of software developers and users. Today, SBML is supported by over 75 software systems including Gepasi. Also an SBML->MatLab converter
- Advances in biotechnology are leading to larger, more complex quantitative models. The systems biology community needs information standards if models are to be shared, evaluated and developed cooperatively. SBML's widespread adoption offers many benefits, including:
  - enabling the use of multiple tools without rewriting models for each tool
  - enabling models to be shared and published in a form other researchers can use even in a different software environment
  - ensuring the survival of models (and the intellectual effort put into them) beyond the lifetime of the software used to create them.



17

## SBML Example

- XML Based Language
- Species representation: m1 in RKIP model:
  - <specie name="m1" compartment="compartment" initialAmount="2.5" boundaryCondition="false" />
- Reaction representation: k1 in RKIP model: m1 + m2 -> m3 (rate = k1 = 0.53)
  - <reaction name="k1" reversible="false">
  - <listOfReactants>
  - <specieReference specie="m1" stoichiometry="1" />
  - <specieReference specie="m2" stoichiometry="1" />
  - </listOfReactants>
  - <listOfProducts>
  - <specieReference specie="m3" stoichiometry="1" />
  - </listOfProducts>
  - <kineticLaw formula="k\_1\*m1\*m2">
  - <listOfParameters>
  - <parameter name="k\_1" value="0.53" />
  - </listOfParameters>
  - </kineticLaw>
- </reaction>

18

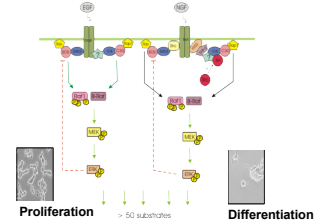
## SBML Demo

- SBML demo of RKIP Model

19

## The MAPK pathway

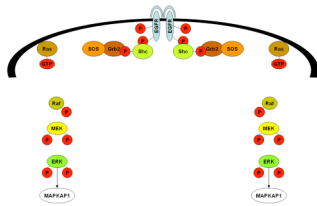
- The MAPK pathway is one of the most important and intensively studied signalling pathways.
- It is at the heart of a molecular signalling network that governs the growth, proliferation, differentiation and survival of many, if not all cell types.
- It is deregulated in various diseases ranging from cancer to immunological, inflammatory and degenerative syndromes, and thus represents an important drug target.



20

## The EGF Activated MAPK Cascade

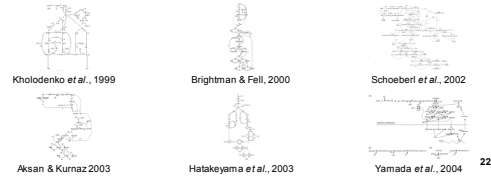
- Demo of EGF activated MAPK cascade



21

## Published MAPK Models

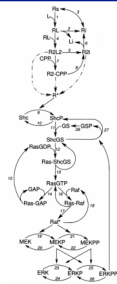
- Currently, there are a variety of published differential equation based models of the MAPK pathway (activated by EGF)
- These models all differ in the way they represent the MAPK pathway
  - Differ in the way some processes are modelled
  - Differ in what proteins are involved in the pathway
  - Differ 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



22

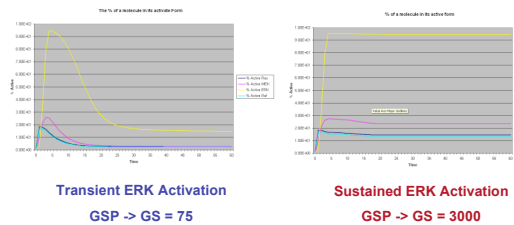
## The Brightman & Fell MAPK Model

- Brightman & Fell (2000) FEBS Letters 482 169-174
  - Differential feedback regulation of the MAPK cascade underlies the quantitative differences in EGF and NGF signalling in PC12 cells
- Differential equation based model of the EGF activated MAPK cascade (used Gepasi)
  - 29 species and 30 reactions
- Used sensitivity analysis to investigate how ERK activation could be switched from transient to sustained
  - Take each reaction in the pathway, change its kinetic rate (up and down) and monitor the level of ERK-PP
- MAPK Cascade
  - EGF: Transient ERK-PP activation: Proliferation
  - NGF: Sustained ERK-PP activation: Differentiation
- Part of the negative feedback loop,  $GSP \rightarrow GS$  (dephosphorylation of the Grb2-SOS complex), was found to be the most important reaction



23

## Brightman & Fell Simulation



Transient ERK Activation  
GSP > GS = 75

Sustained ERK Activation  
GSP > GS = 3000

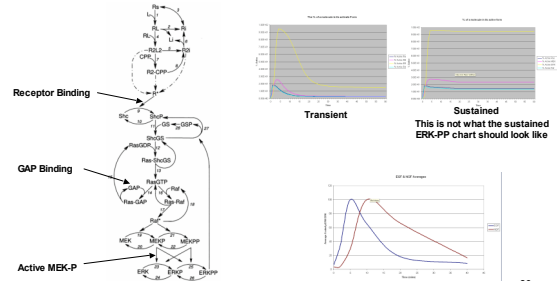
24

## Brightman & Fell Demo

- Gepasi demo of the Brightman & Fell model

25

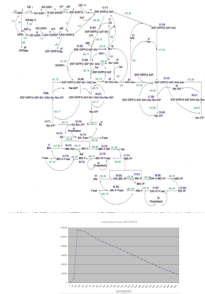
## Brightman & Fell Errors



26

## The Schoeberl MAPK Model

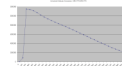
- The Schoeberl model is one of the most comprehensive models of the MAPK pathway available:
  - Computational modelling of the dynamics of the MAP kinase cascade activated by surface and internalised EGF receptors (Schoeberl et al., 2002)
  - Showed that the critical parameter is the initial velocity of receptor activation
  - 125 reactions
  - 94 species
  - Receptor complex strategy
  - Receptor internalisation
  - Shc dependent & independent pathway
- However, it does have a number of errors.
- The major one being there is no negative feedback loop from ERK-PP to SOS
- The transient activation of ERK is caused by an incorrect modelling of Ras and associated reactions



27

## Schoeberl Errors

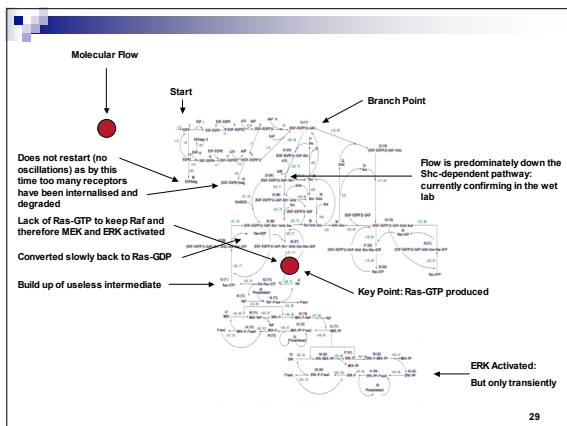
- The major error in this model is that there is no negative feedback loop:
  - ERK-PP (via MAPKAP1) should phosphorylate SOS causing it to dissociate from the receptor complex forming a negative feedback loop
- So why is ERK activation transient and not sustained?



- The deactivation of Ras is incorrectly modelled



28



29

## Schoeberl Demo

- Gepasi demo of Schoeberl model

30

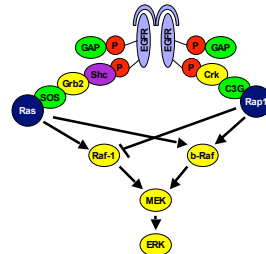
## BPS Project: What we are doing

- 1) Identification
  - We want to build realistic and comprehensive models of the MAPK cascade activated by EGF and NGF receptors
  - We want to know how EGF causes a transient ERK response and how NGF causes a sustained ERK response: what are the key steps involved?
- 2) Definition
  - We are currently in the process of defining 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)
- 3) Simulation
  - We currently use MatLab (Gepasi started to get very slow)
- 4) Validation
  - Oliver and Amelie are generating wet lab data to validate the model with
- 5) Analysis
  - Will hope to analyse the models to find the key steps in determining the transient/sustained behaviour

31

## Model Expansion: Rap1

- We are now also in the process of expanding the model as so far, only the Ras/Raf-1 pathway has been discussed.
- But, there is another route to ERK activation: the Rap1/b-Raf pathway

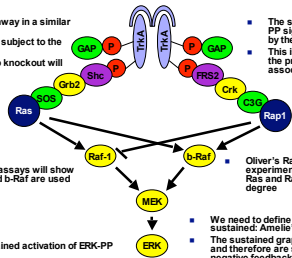


- However, there is not as much known about this pathway
- We firstly need to confirm if it is indeed used and then to what extent
- The model can then be expanded to included this pathway if it is used
- Oliver's Raf kinase assays will show us if Raf1 and b-Raf are used and to what degree. However, they will not tell us the true story because...
- Therefore, Oliver will also measure the levels of active Ras and Rap1 (Ras binding domains) as these will tell us the true story
- However, the overall ERK-PP signal is transient
- Therefore, there must also be some sort of negative feedback loop on this side (if it is used)
- Currently investigating this: Crk phosphorylation, GAP binding

32

## Model Expansion: NGF

- So far, the model has been concerned with the EGF receptor and we are now applying it to the NGF receptor (TrkA)
- NGF uses the Ras pathway in a similar transient way as EGF
- Therefore it should be subject to the same feedback loop
- Oliver's feedback loop knockout will investigate this
- Oliver's Raf kinase assays will show us whether Raf1 and b-Raf are used and to what degree
- NGF causes a sustained activation of ERK-PP
- The sustained nature of the ERK-PP signal is believed to be caused by the Rap1/b-Raf pathway
- This is believed to be caused by the protein FRS2 which can stably associate with Crk
- Oliver's Raf binding domain experiments will show us whether Ras and Rap1 are used and to what degree
- We need to define what we mean by sustained: Amelie's ERK-PP data
- The sustained graphs are not pure sustained and therefore are subject to some sort of negative feedback



33

## Overview

- Modelling with differential equations is an established technique that has been and is widely used
- Requires the correct parameter data (initial concentrations, kinetic constants)
- Currently, there are many models of the MAPK pathway activated by EGF receptors
- However, all these models differ, some have errors and some have large simplifications
- We are creating realistic and comprehensive models of the MAPK pathway which we believe better reflect real live which will be thoroughly validated with data generated from our wet lab
- The models will be analysed and any predictions made will be tested in the wet lab

34