Computational Methods for Systems Biology

Computing for Biology – what is the payoff?

David Gilbert
School of Information Systems, Computing & Mathematics
Centre for Systems & Synthetic Biology

people.brunel.ac.uk/~csstdrg/courses

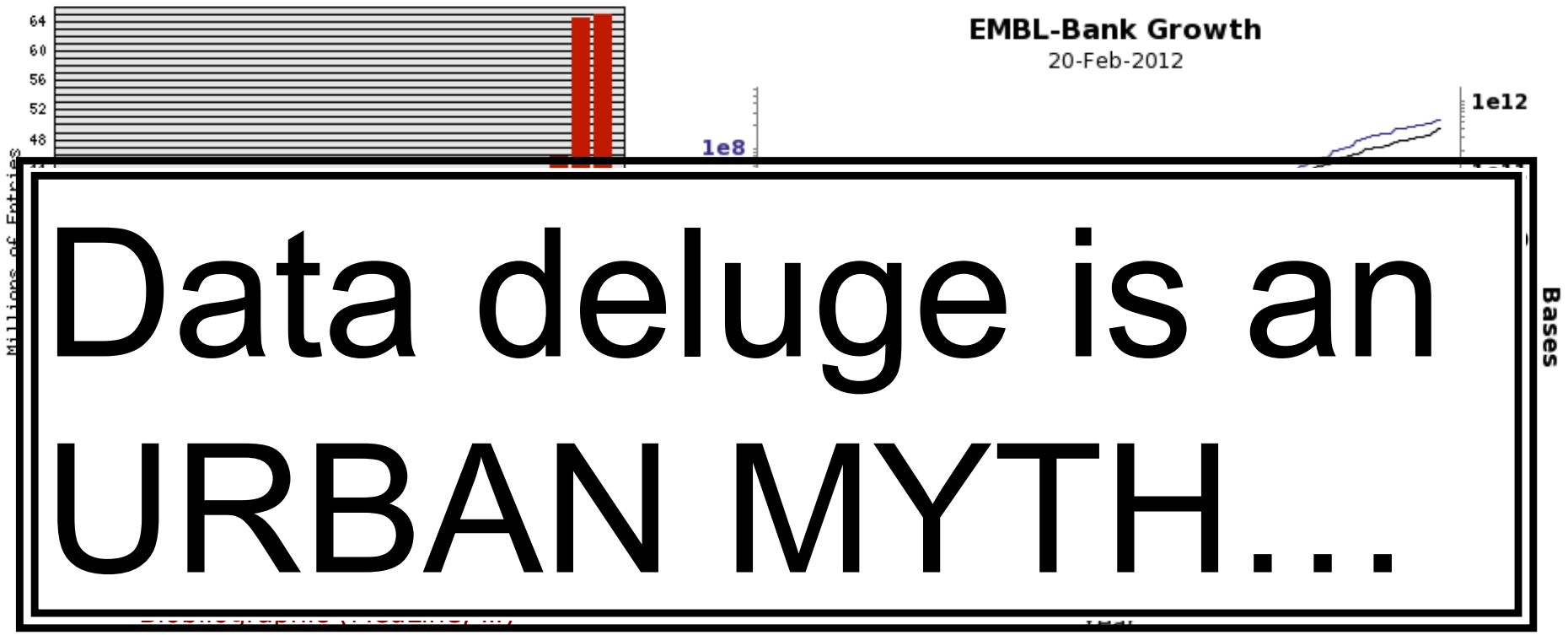
Outline

- Motivation
- Basic biochemistry & cell biology
- Biological data
- Biomodel engineering
- Modelling – qualitative; quantitative (continuous, stochastic)
- Simulation
- Analysis
- Model checking

DNA sequencing
what challenge?

Database Growth

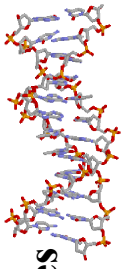
EMBL - sequences

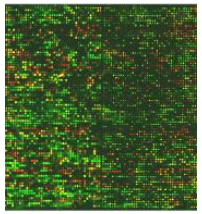


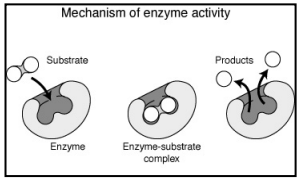
Data deluge is an URBAN MYTH…

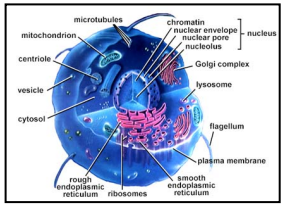
- Amino Acid Seq (SWISS-PROT, …)
- 3D Molecular Structure (PDB, …)
- Nucleotide Seq (GenBank, EMBL, …)
- Biochemical Pathways (KEGG, WIT…)
- Molecular Classifications (SCOP, CATH,…)
- Motif Libraries (PROSITE, Blocks, …)

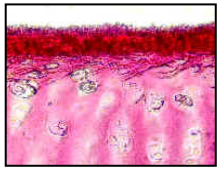
The Complexity of Biological Data

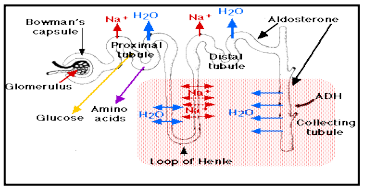












Nucleotide sequences

Nucleotide structures

Gene expressions

Protein Structures

Protein functions

Protein-protein interaction (pathways)

Cell

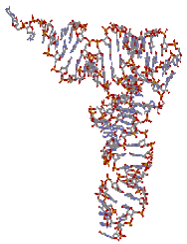
Cell signalling

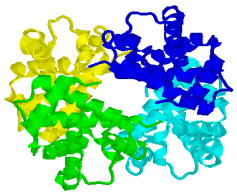
Tissues

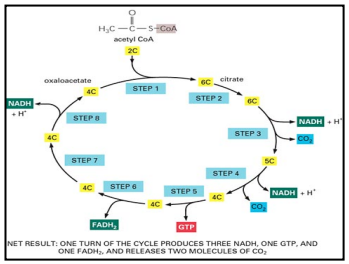
Organs

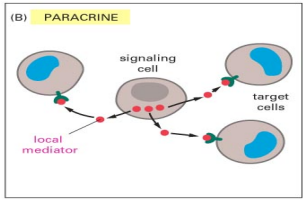
Physiology

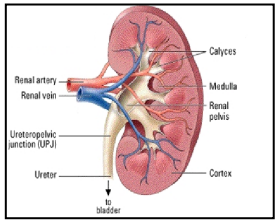
Organisms











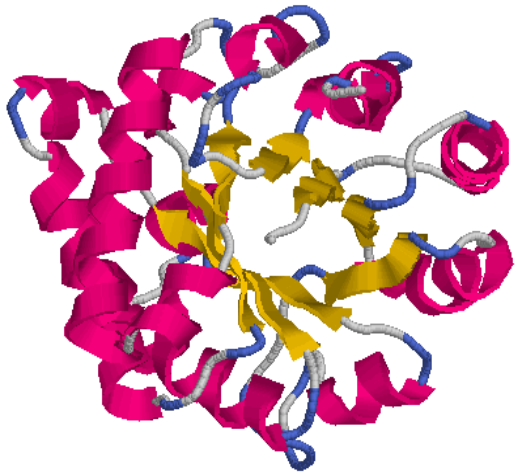
Genes to systems

DNA

"gene"

mRNA

Protein sequence



Folded Protein

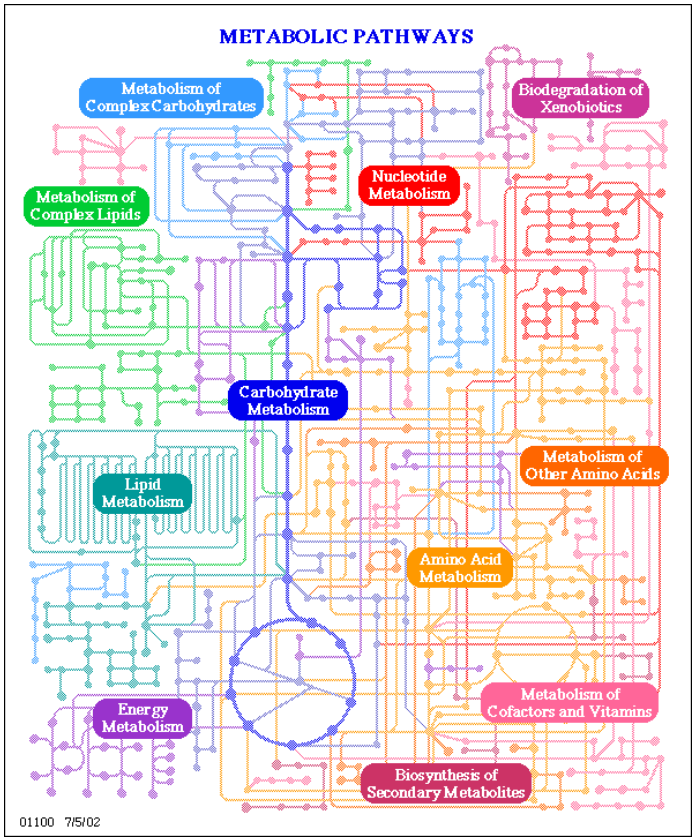
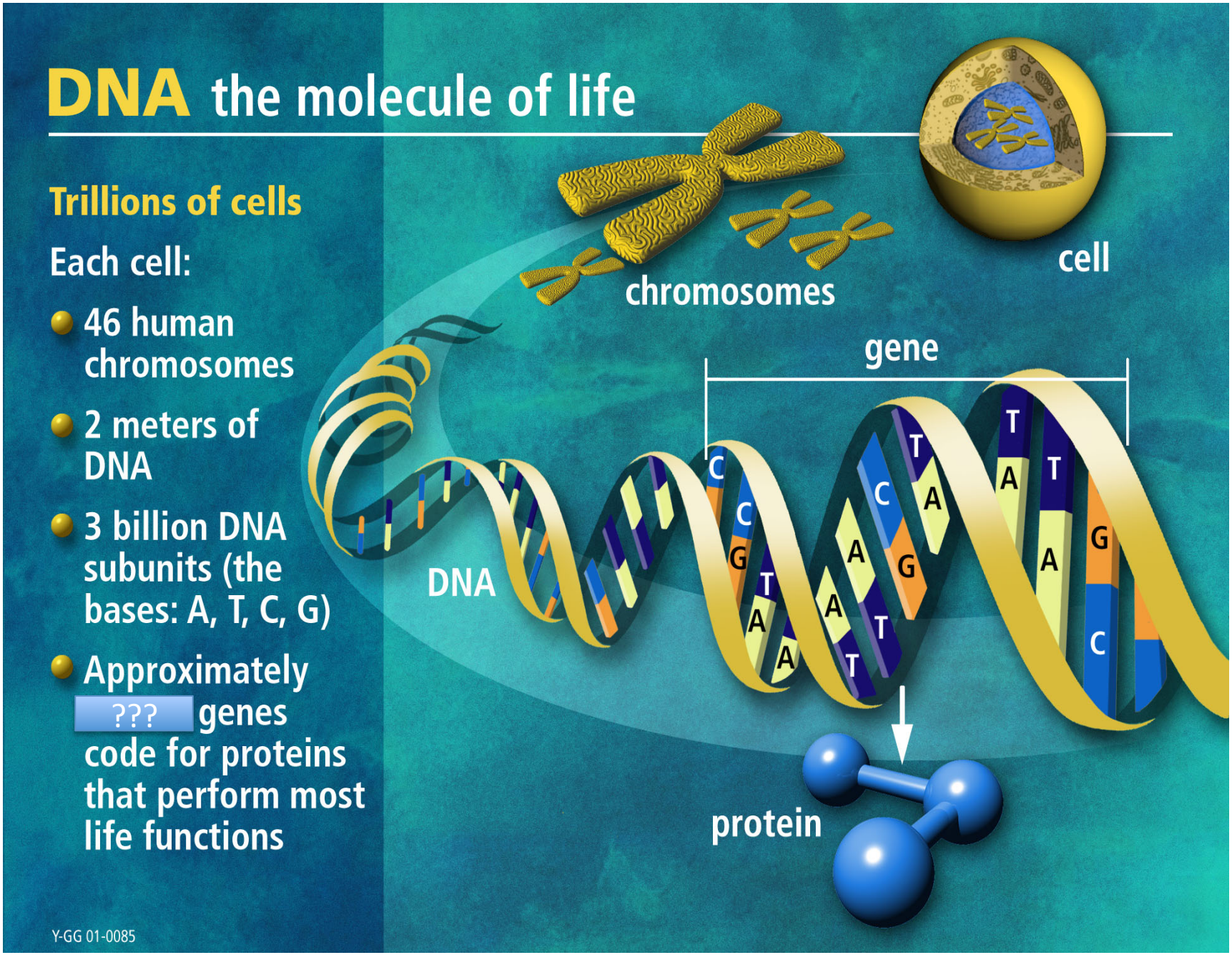


Image from Human Genome Project http://www.ornl.gov/hgmis



More genes – better?

31,000

28,000

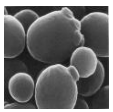
22,000

huh?

20,000

14,000

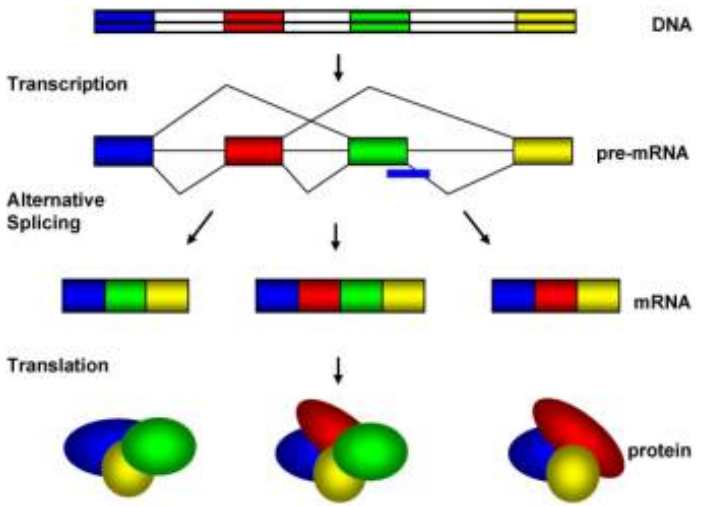
6,700



4,300

70

Alternative splicing



Bioinformatics
(Computational Biology - USA)

- **Bio** - Molecular Biology
- **Informatics** - Computer Science

- Bioinformatics - the study of the application of
  - *molecular biology, computer science, artificial intelligence, statistics* and *mathematics*
  - to **model, organise, understand** and **discover** interesting information associated with the large scale molecular biology databases
  - to **guide assays** for biological experiments

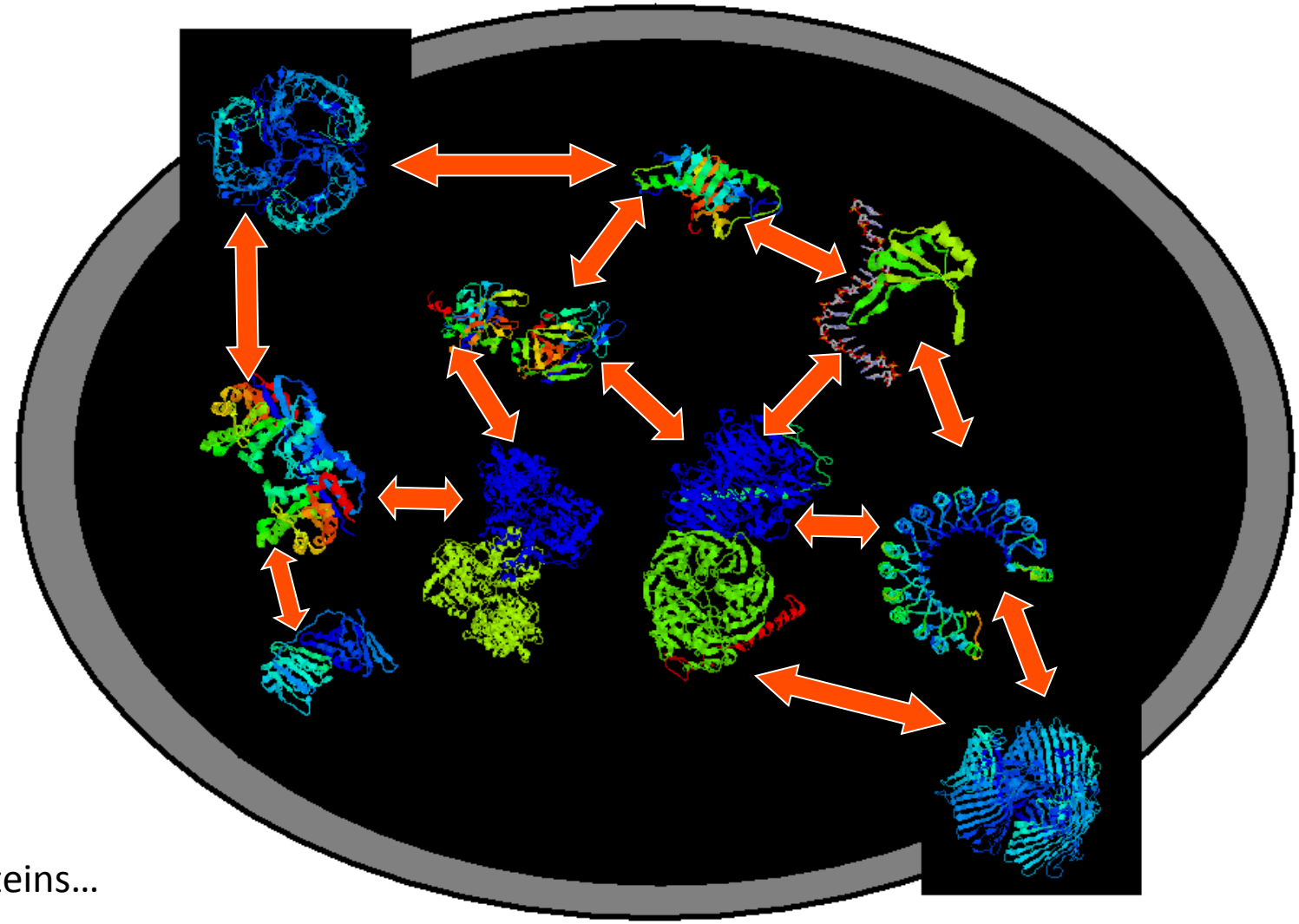
- '***One gene at a time***'
  - even if all genes in a genome

BTW, related but different...

Apply principles from biology to derive novel approaches in computer science:

- biocomputing
- neural computing
- genetic algorithms
- evolutionary computing

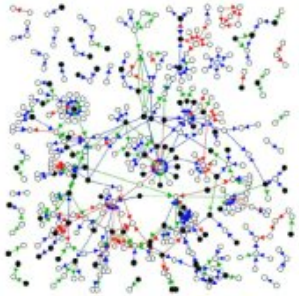
But what about interactions?



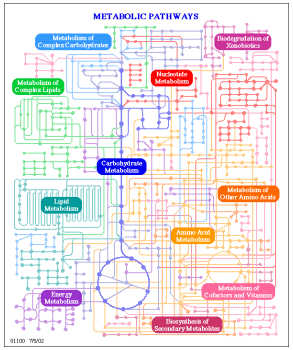
Proteins…

Networks

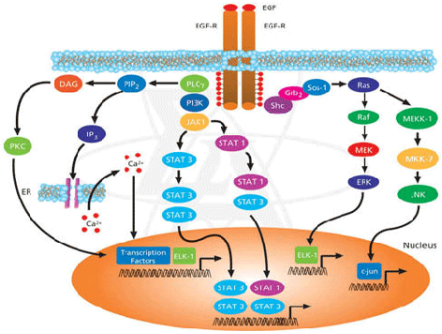
- Gene regulation



- Metabolic



- Signalling

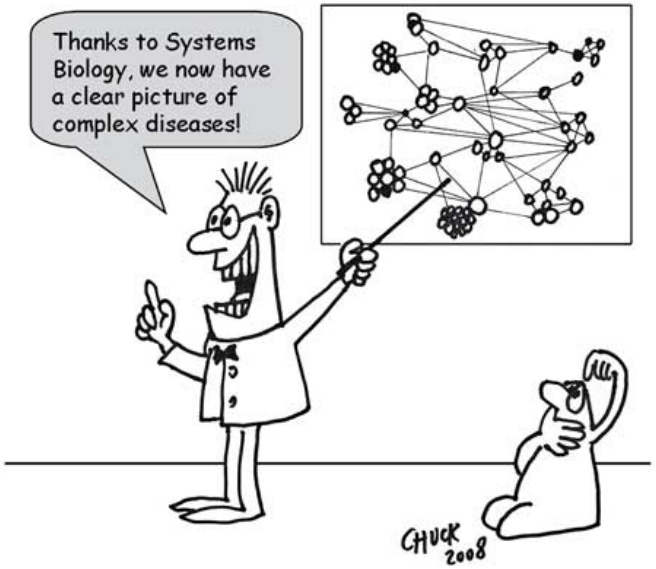


- Protein-protein interaction

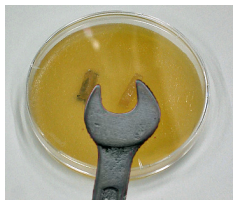
- Developmental

And so…

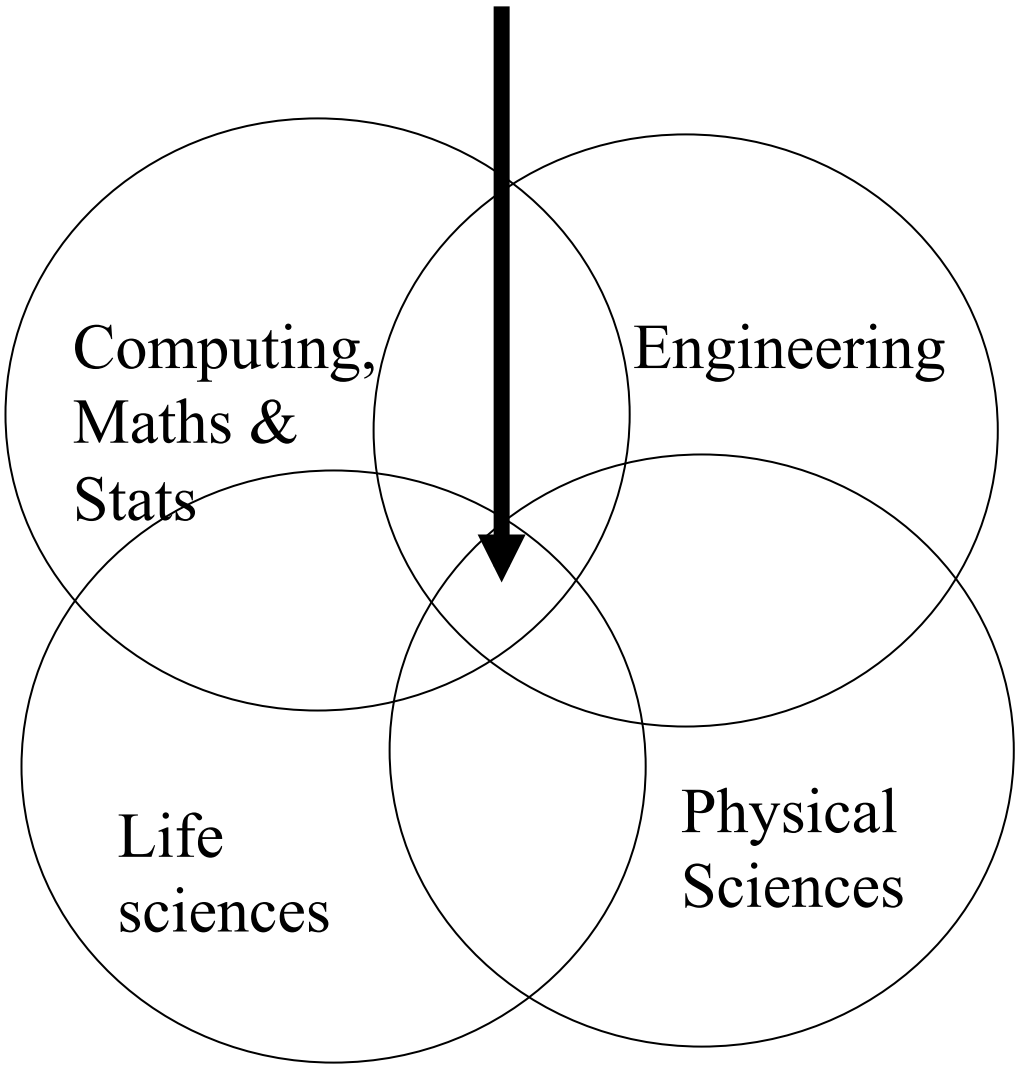
- **Systems Biology** studies the relationships and interactions between various parts of a biological system in order to understand how the whole system functions



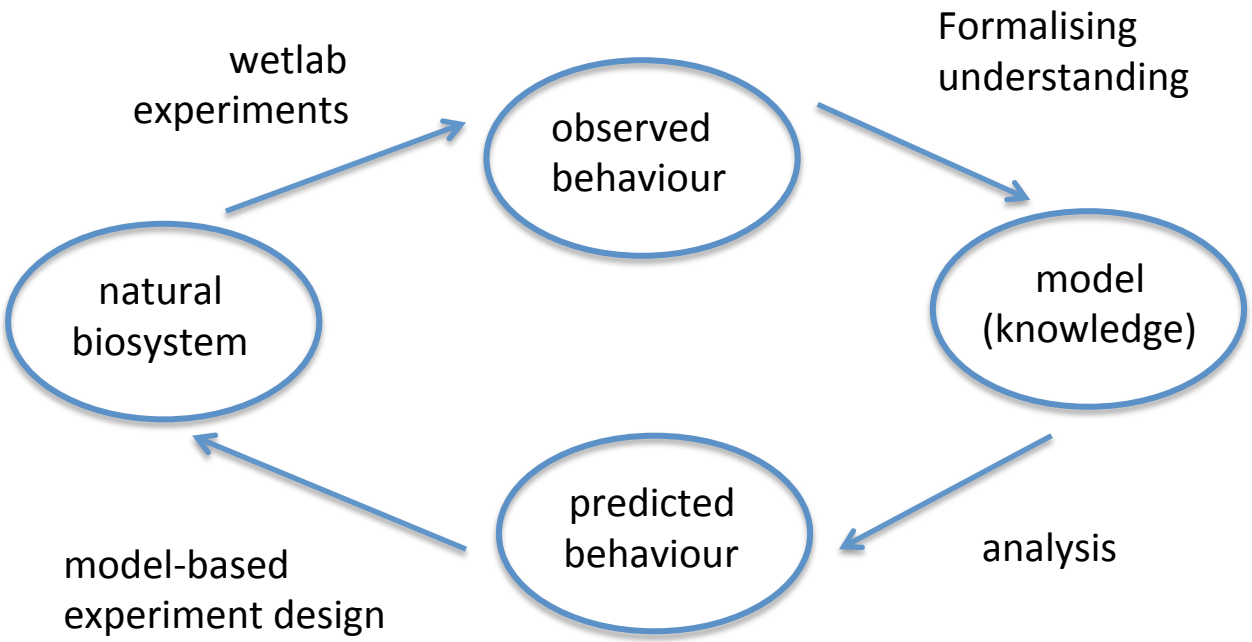
- **Synthetic Biology** - the structured engineering of biological systems for useful purposes.



Systems Biology in context - a new discipline?



Systems biology



A Framework for Systems Biology

(Ideker, Galitski & Hood, 2001)

- Define all of the components of the system
- Systematically perturb and monitor components of the system
- Reconcile the experimentally observed responses with those predicted by the model
- Design and perform new perturbation experiments to distinguish between multiple or competing model hypotheses

BioModel Engineering

- Takes place at the interface of computing science, mathematics, engineering and biology.
- A systematic approach for **designing**, **constructing** and **analyzing** computational models of biological systems.
- Some inspiration from efficient software engineering strategies.
- Not engineering biological systems *per se,* but
  - describes their structure and behaviour,
  - in particular at the level of intracellular molecular processes,
  - using computational tools and techniques in a principled way.

Rainer Breitling, David Gilbert, Monika Heiner, Richard Orton (2008). A structured approach for the engineering of biochemical network models, illustrated for signalling pathways. Briefings in Bioinformatics

David Gilbert, Rainer Breitling, Monika Heiner, and Robin Donaldson (2009). An introduction to BioModel Engineering, illustrated for signal transduction pathways, 9th International Workshop, WMC 2008, Edinburgh, UK LNCS Volume 539, pp13-28

Rainer Breitling, Robin Donaldson, David Gilbert, Monika Heiner (2010): Biomodel Engineering - From Structure to Behavior; : Trans. Comp Systems Biology XII, Springer LNBI 5945, pp. 1-12

Biomodel engineering

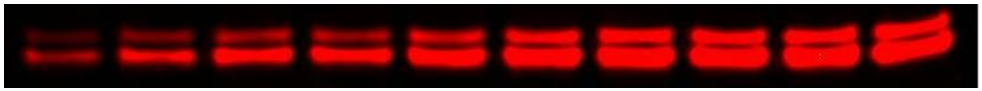
1. Problem identification
2. Construction
3. Simulation
4. Analysis & interpretation
5. Management & development

Biomodel engineering

1. Problem identification
2. **Construction**
3. **Simulation**
4. **Analysis & interpretation**
5. Management & development

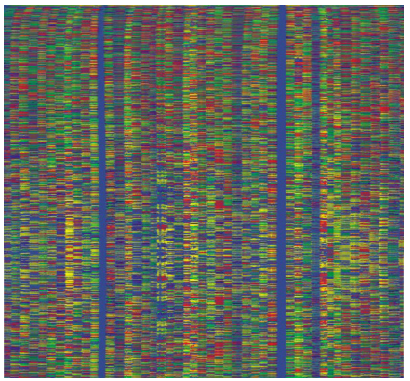
Where do the data come from?

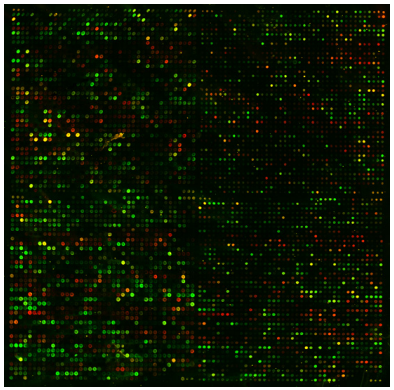
- 'Traditional' biochemistry and genetics
  - Reductionist
  - Descriptive
  - "One gene = one career"
  - Published in scientific journals
  - Text mining
- Low throughput: immuno-precipitation,…

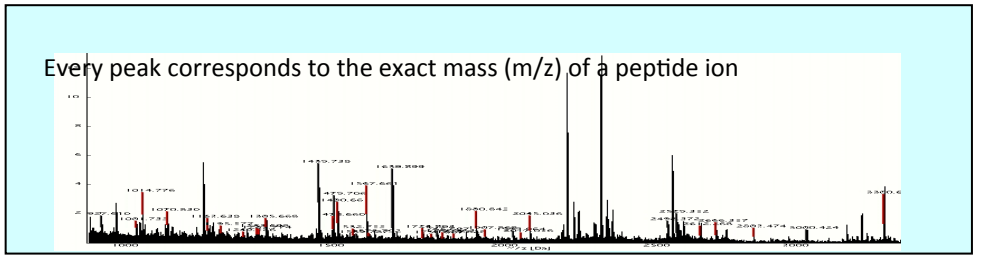


Drivers - technology

- High throughput:
  - Genome sequencing
  - Gene expression array
  - Protein array
  - Mass spectrometry
  - Metabolomics







NCBI

National Library of Medicine PubMed

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM

Search PubMed for Go Clear

Limits Preview/Index History Clipboard

About Entrez

Entrez PubMed
Overview
Help | FAQ
New/Noteworthy

PubMed Services
Journal Browser
MeSH Browser
Single Citation Matcher
Batch Citation Matcher
Clinical Queries

Related Resources
Order Documents
Grateful Med
Consumer Health
Clinical Alerts
ClinicalTrials.gov

Privacy Policy

Display Citation Save Text Order Add to Clipboard

1: *Science* 1998 Sep 4;281(5382):1509-12 Related Articles, Books, OMIM, LinkOut

Full text article at www.sciencemag.org

Identification of c-MYC as a target of the APC pathway.

He TC, Sparks AB, Rago C, Hermeking H, Zawel L, da Costa LT, Morin PJ, Vogelstein B, Kinzler KW

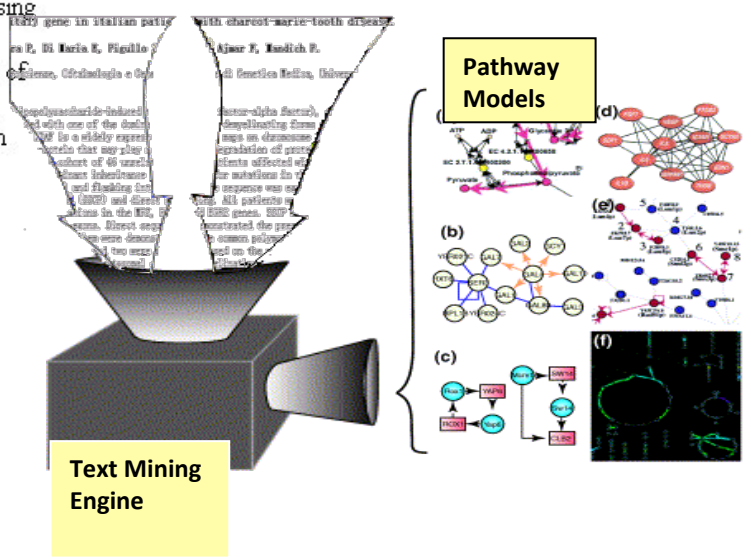
Howard Hughes Medical Institute and Johns Hopkins Oncology Center, 424 North Bond Street, Baltimore, MD 21231, USA.

The adenomatous polyposis coli gene (APC) is a tumor suppressor gene that is inactivated in most colorectal cancers. Mutations of APC cause aberrant accumulation of beta-catenin, which then binds T cell factor-4 (Tcf-4), causing increased transcriptional activation of unknown genes. Here, the c-MYC oncogene is identified as a target gene in this signaling pathway. Expression of c-MYC was shown to be repressed by wild-type APC and activated by beta-catenin, and these effects were mediated through Tcf-4 binding sites in the c-MYC promoter. These results provide a molecular framework for understanding the previously enigmatic overexpression of c-MYC in colorectal cancers.

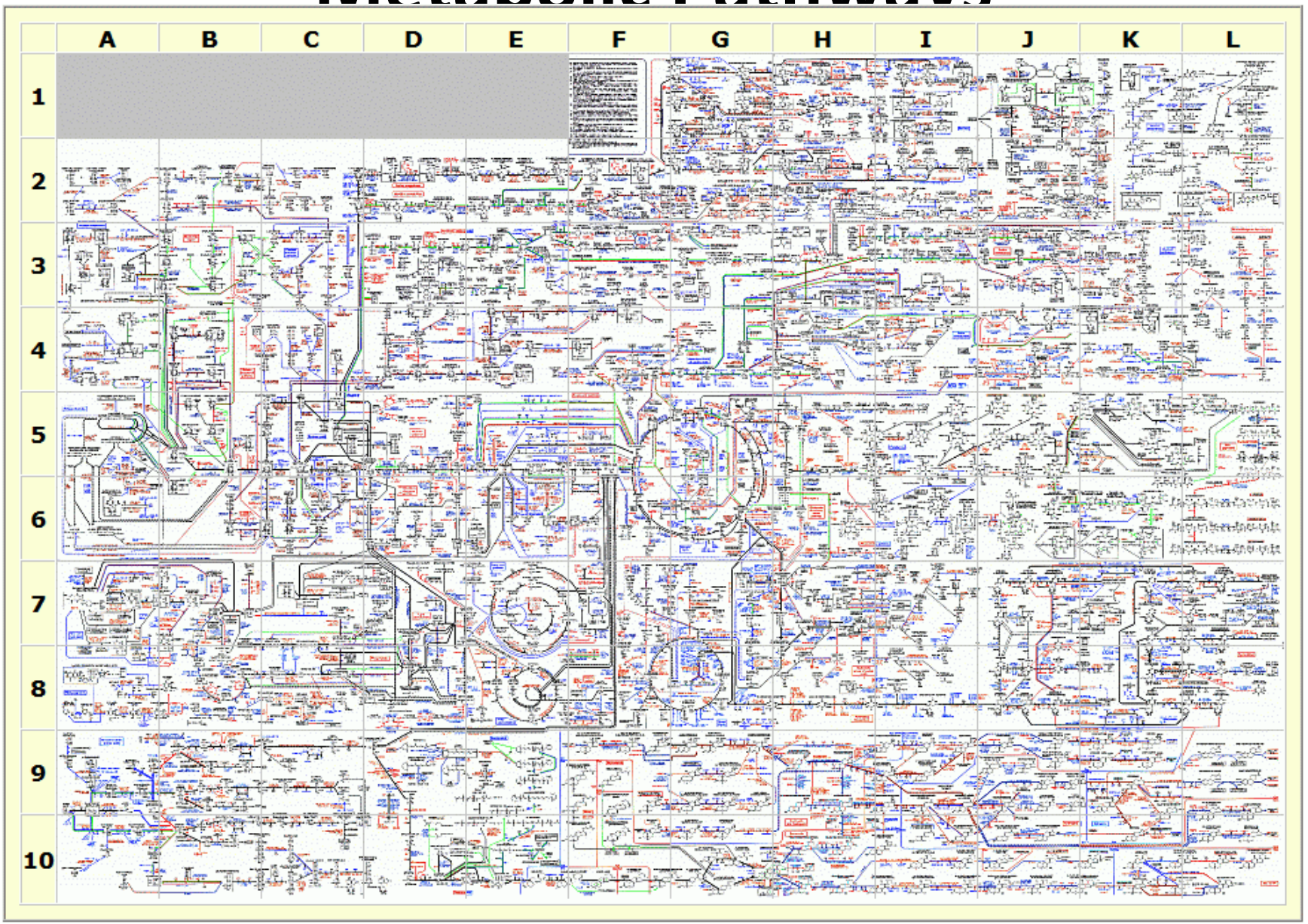
Comments:
- Comment in: Science 1998 Sep 4;281(5382):1438-41

MeSH Terms:
- Binding Sites
- Cell Line
- Colorectal Neoplasms/genetics*
- Cytoskeletal Proteins/metabolism
- Cytoskeletal Proteins/genetics
- Gene Expression Regulation, Neoplastic*

- Maintained by National Library of Medicine
- Free of charge, since 1997
- > 14 million references since 1971
- > 4000 biomedical journals
- > 80% in English
- > 80% have an abstract

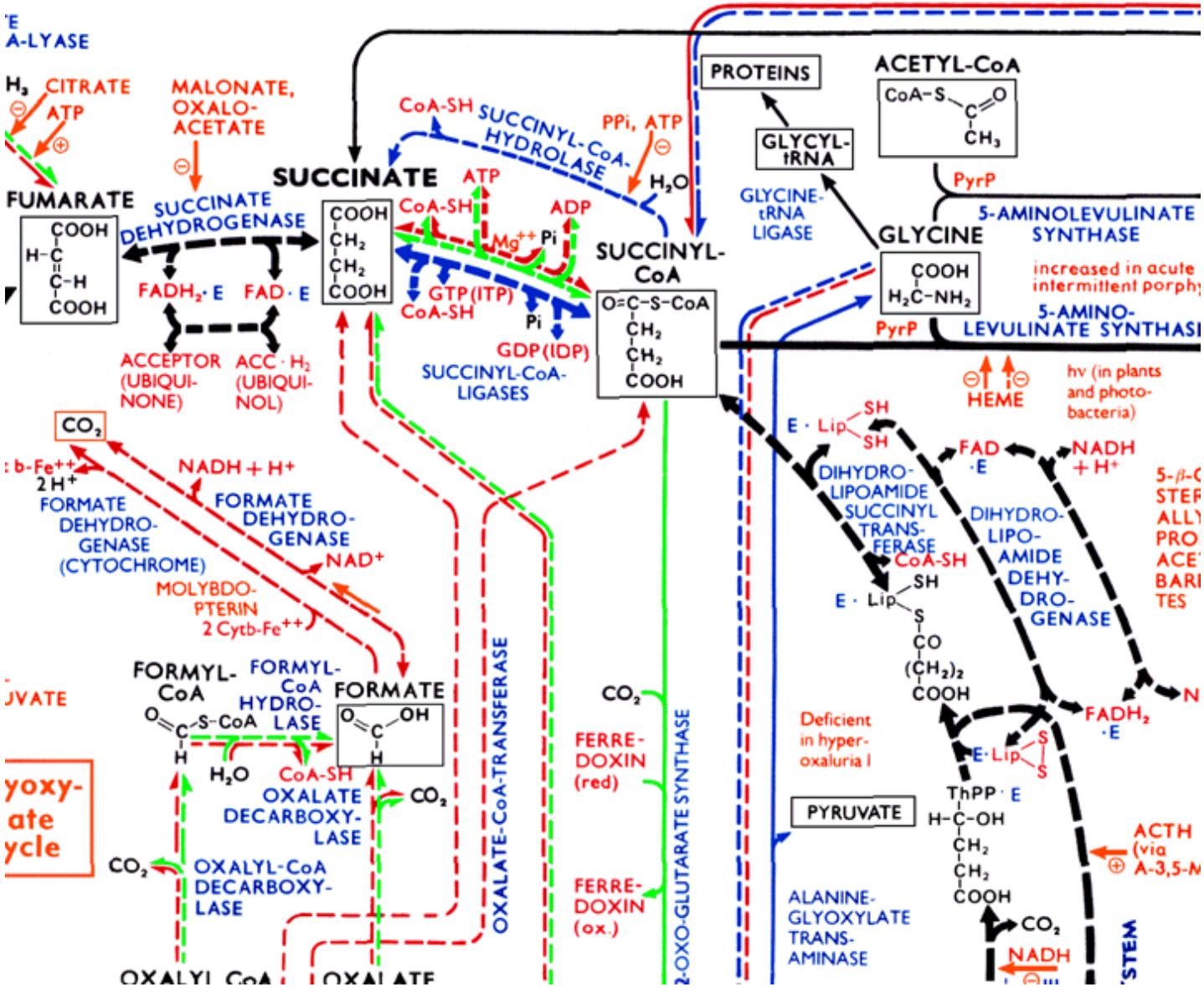


Metabolic Pathways



http://ca.expasy.org/tools/pathways/

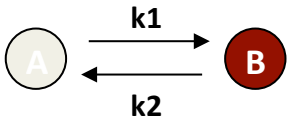
→ general biochemical pathways, → animals,
→ higher plants, → unicellular organisms



What is modelling?

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

Translating a biological pathway



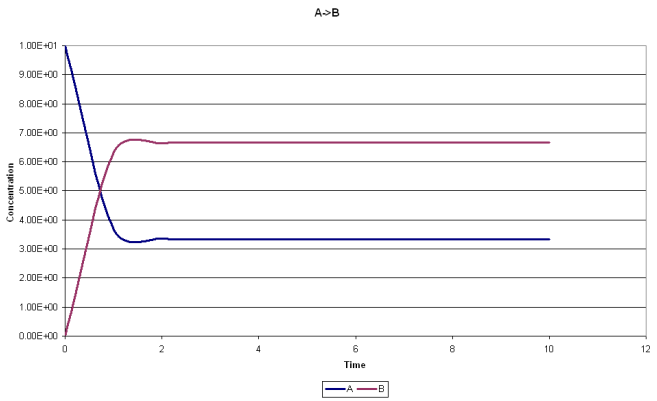
Into mathematics

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

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

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

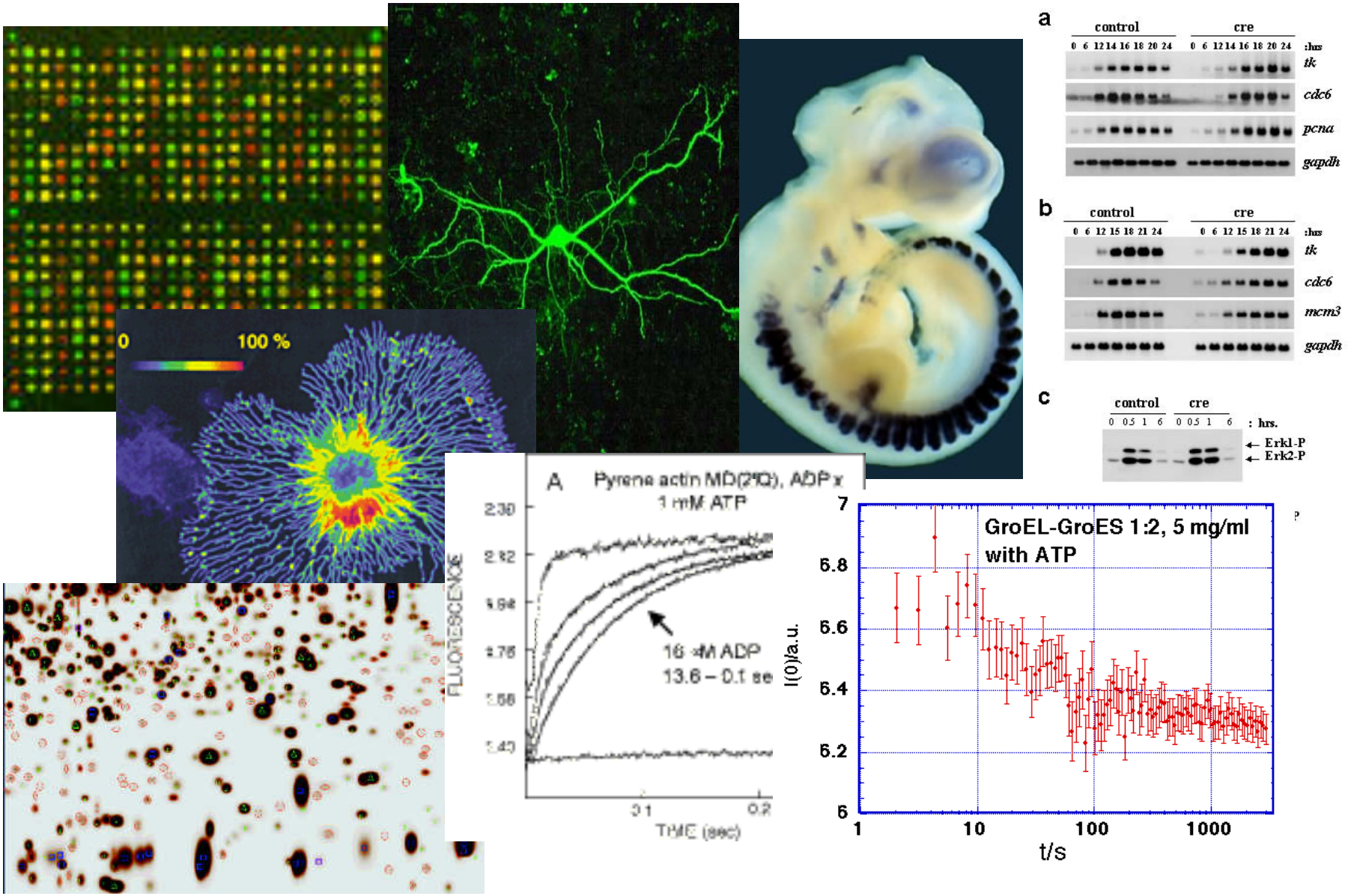
For subsequent analysis



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)

Biology = Concentrations



The simplest chemical reaction

$$A \rightarrow B$$

- irreversible, one-molecule reaction
- examples: all sorts of decay processes, e.g. radioactive, fluorescence, activated receptor returning to inactive state
- any metabolic pathway can be described by a combination of processes of this type (including reversible reactions and, in some respects, multi-molecule reactions)

The simplest chemical reaction

$$A \rightarrow B$$

various levels of description:

- homogeneous system, large numbers of molecules = ordinary differential equations, **kinetics**
- small numbers of molecules = probabilistic equations, **stochastics**
- spatial heterogeneity = partial differential equations, **diffusion**
- small number of heterogeneously distributed molecules = single-molecule tracking (e.g. cytoskeleton modelling)

Kinetics Description

Main idea: Molecules don't talk

- Imagine a box containing N molecules.
  How many will decay during time t? **k*N**
- Imagine two boxes containing N/2 molecules each.
  How many decay? **k*N**
- Imagine two boxes containing N molecules each.
  How many decay? **2k*N**
- In general:

$$-\frac{dn(t)}{dt} = \lambda * n(t) \quad \Leftrightarrow \quad n(t) = N_0 e^{-\lambda t}$$

differential equation (ordinary, linear, first-order)

exact solution (in more complex cases replaced by a numerical approximation)

Some (Bio)Chemical Conventions

Concentration of Molecule A = [A], usually in units mol/litre (molar)

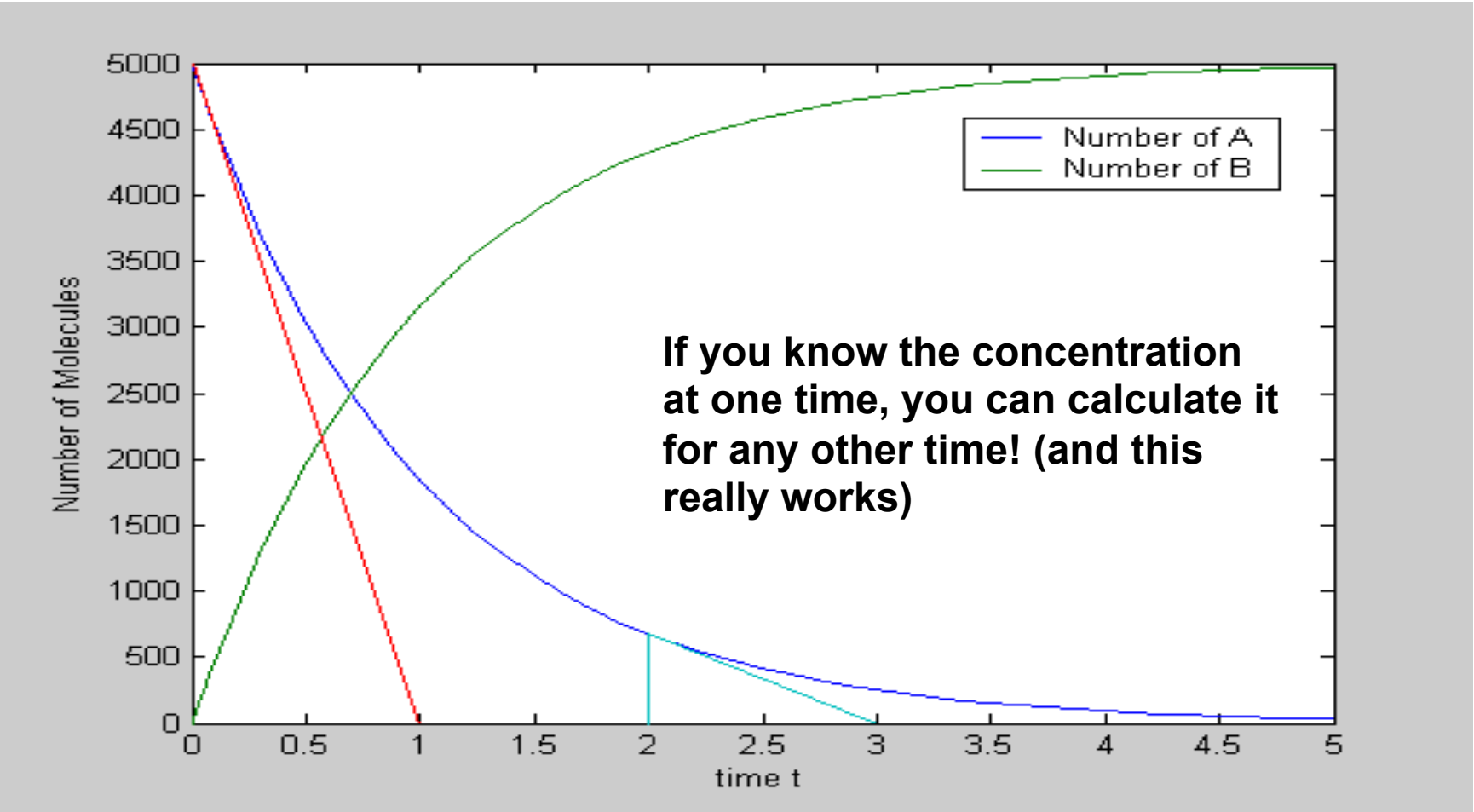
Rate constant = k, with indices indicating constants for various reactions (k_1, k_2...)

Therefore:

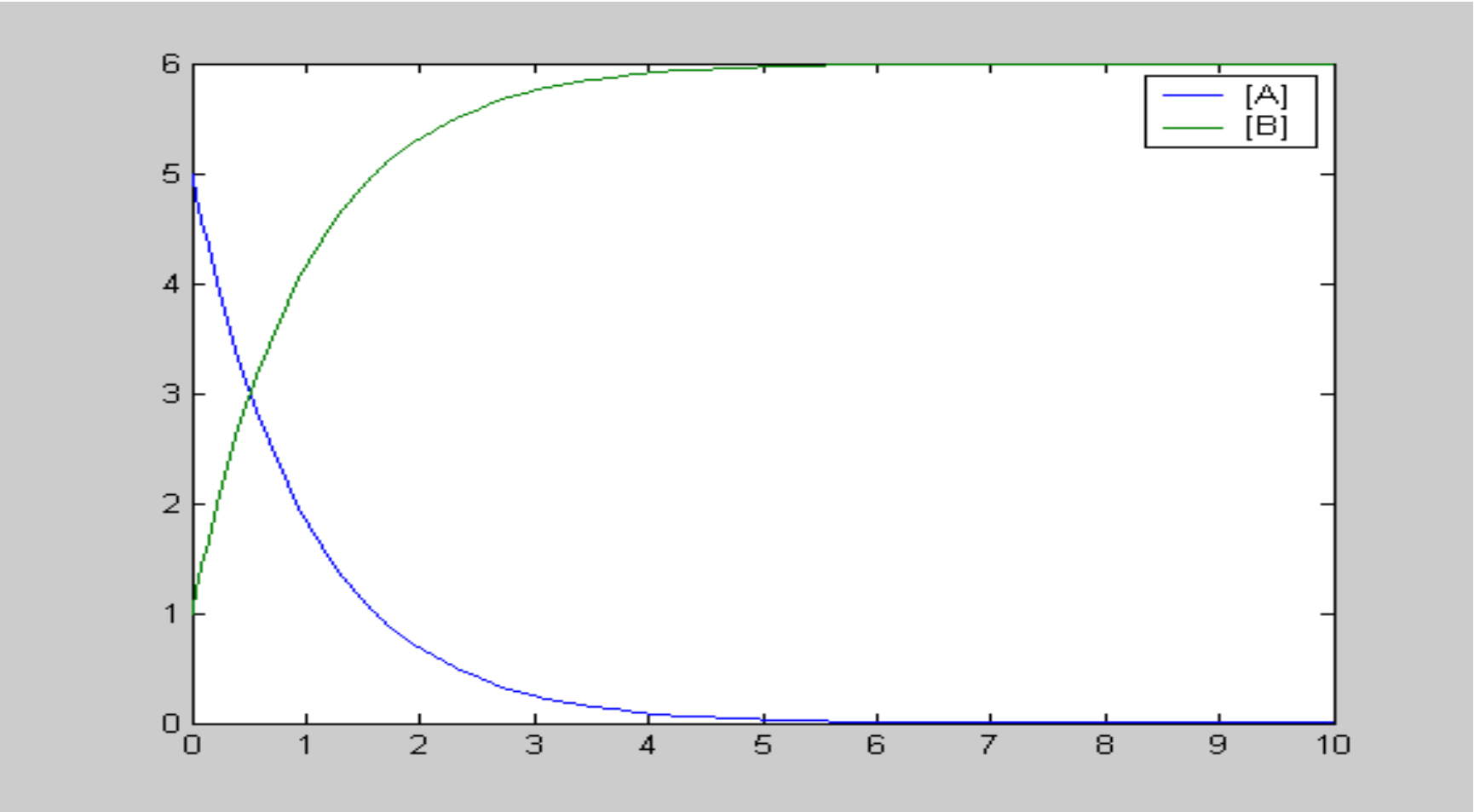
A→B

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

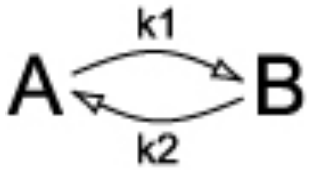
Kinetics Description



Decay Reaction in MATLAB



Reversible, Single-Molecule Reaction

A ⇔ B, or 

Differential equations:

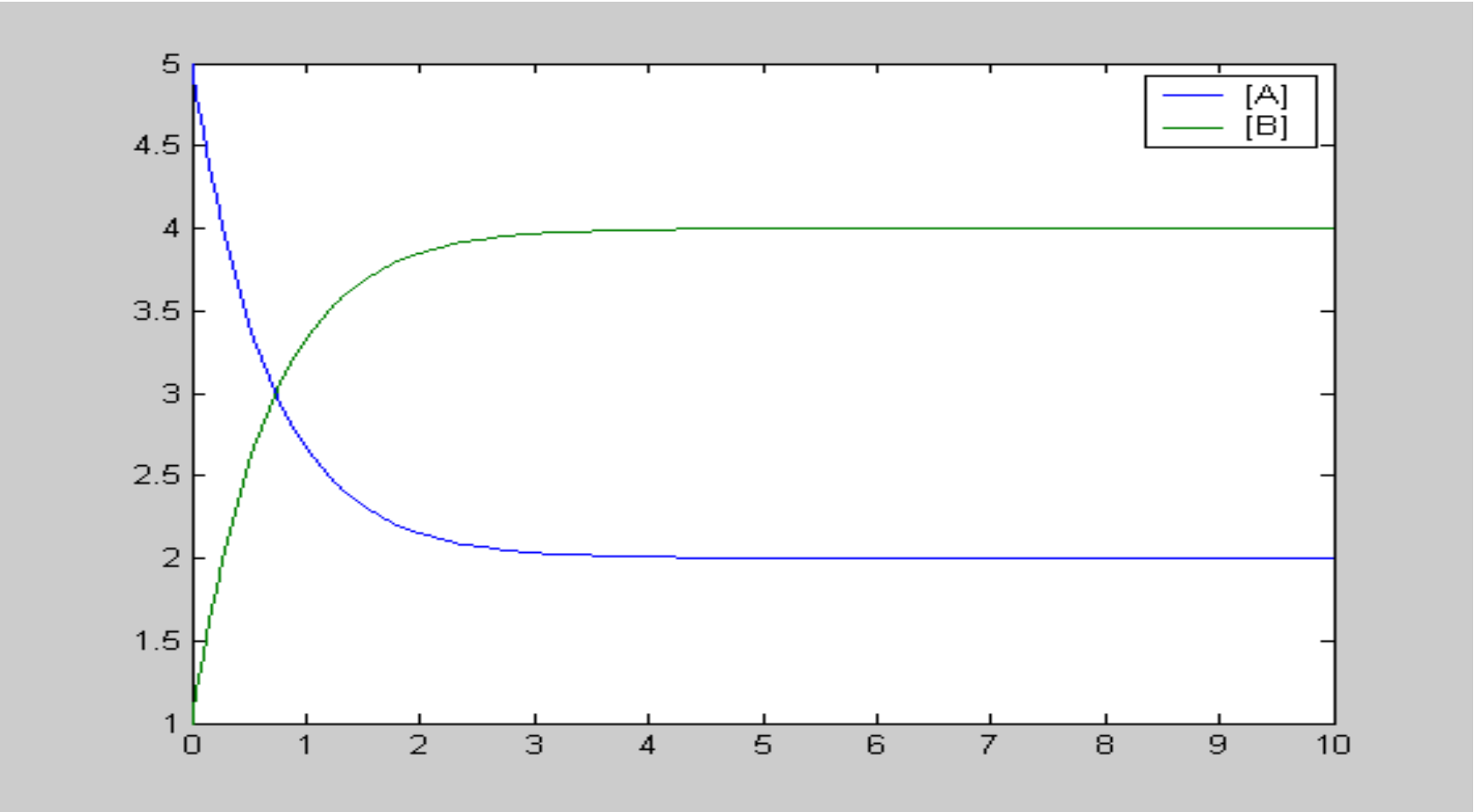
forward reverse

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

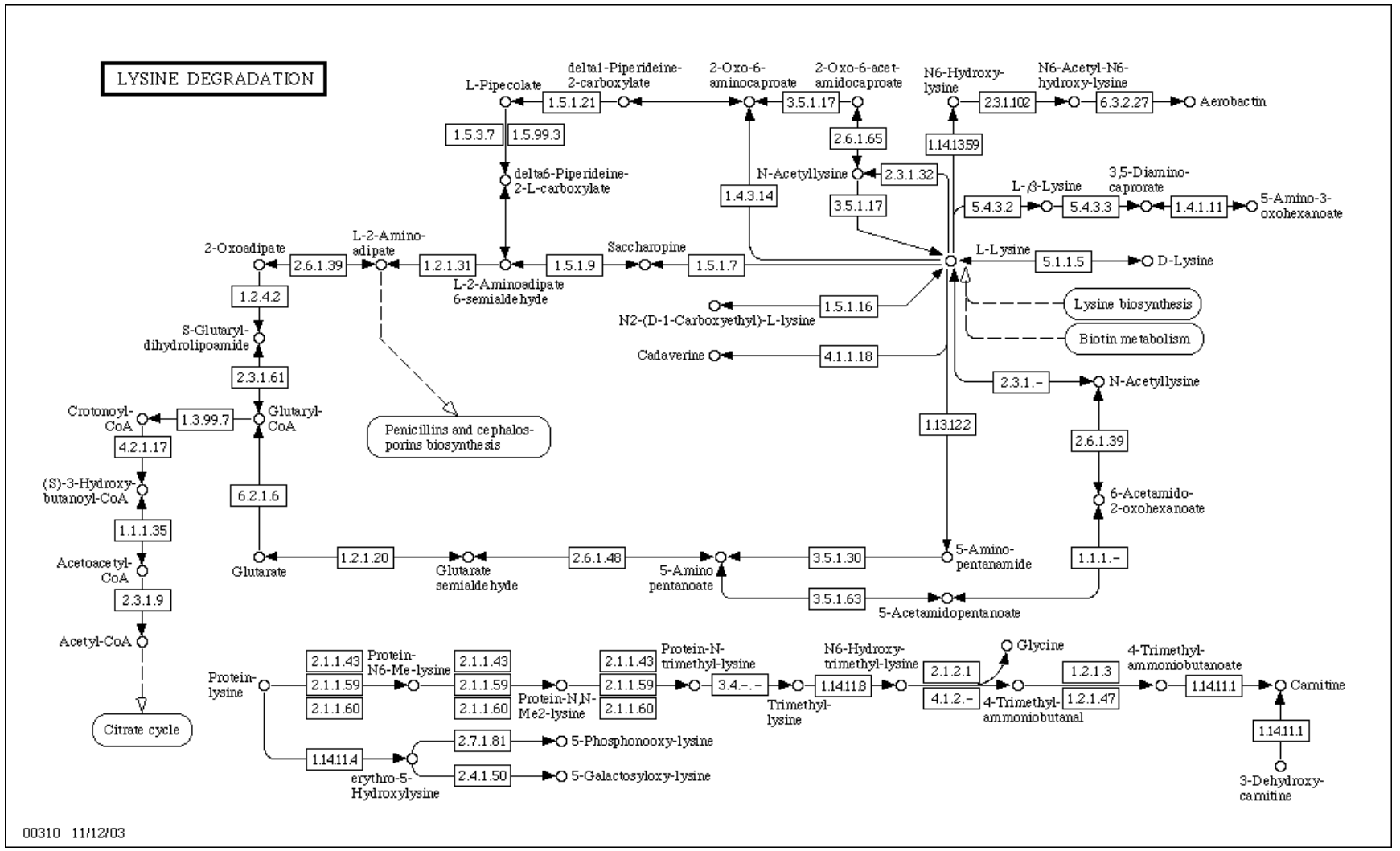
$$\frac{d[B]}{dt} = k_1[A] - k_2[B]$$

Main principle: Partial reactions are **independent**!

Isomerization Reaction

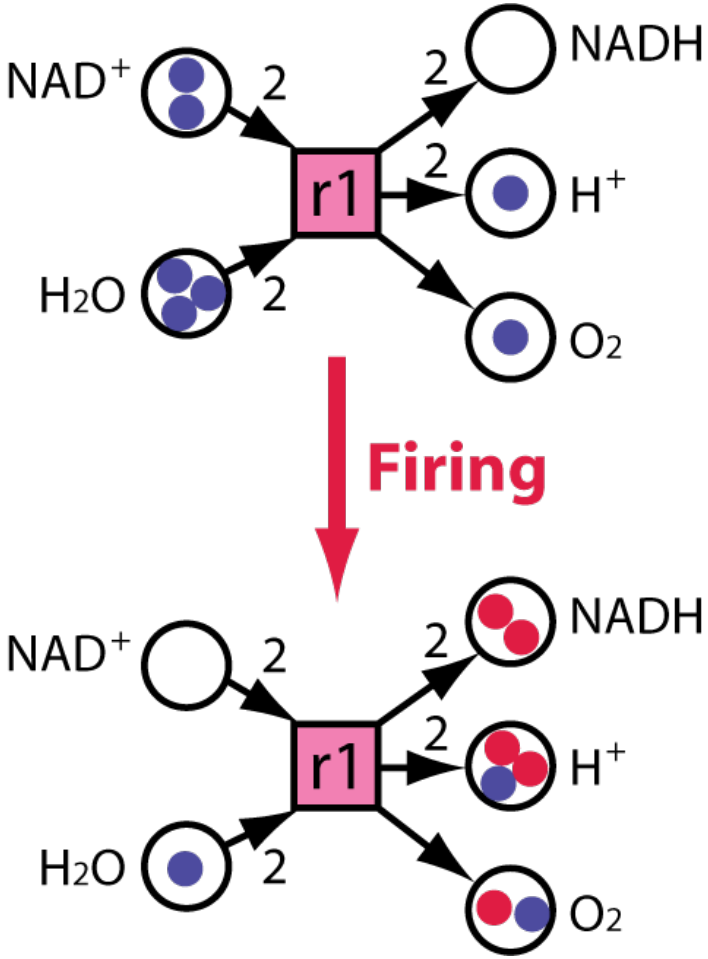


Biological description → bigraph → differential equations



KEGG

Petri nets



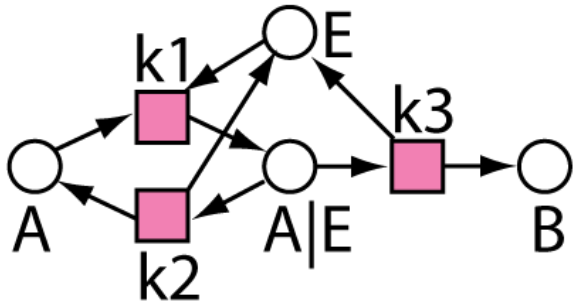
Simple enzymatic reaction

Mass-action kinetics

A: substrate, B: product, E: enzyme

E|A substrate-enzyme complex

$$A + E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} A \mid E \xrightarrow{k_3} B + E$$



$$A + E \xrightarrow{k_1} A \mid E$$

$$A \mid E \xrightarrow{k_2} A + E$$

$$A \mid E \xrightarrow{k_3} B + E$$

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

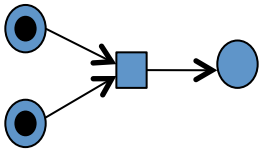
$$\frac{d[A \mid E]}{dt} = +k_1[A][E] - k_2[A \mid E] - k_3[A \mid E]$$

$$\frac{d[B]}{dt} = +k_3[A \mid E]$$

$$\frac{d[E]}{dt} = -k_1[A] \times [E] + k_2[A \mid E] + k_3[A \mid E]$$

Molecules/Levels
CTL, LTL

Qualitative



Time-free

Timed, Quantitative

Abstraction

Abstraction

Monika Heiner

Markov chain
Molecules/Levels
Stochastic rates
CSL

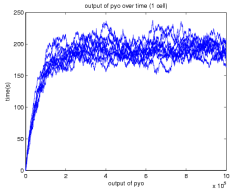
Stochastic

Approximation

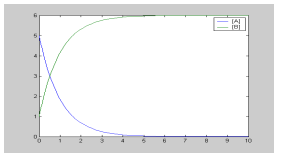
Approximation

Continuous

ODES
Concentrations
Deterministic rates
LTLc



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



DiscreteState Space

Continuous State Space

Gilbert, Heiner and Lehrack. ``A Unifying Framework for Modelling and Analysing Biochemical Pathways Using Petri Nets.'' Proc CMSB 2007

Demo time!
Enzymatic reactions

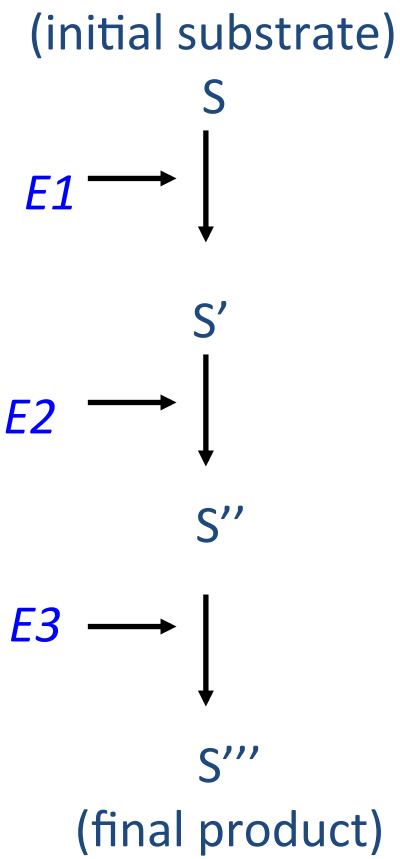
Snoopy (Petri net system)

- Qualitative (token game)

- Quantitative
  - Continuous (ODEs)
  - Stochastic

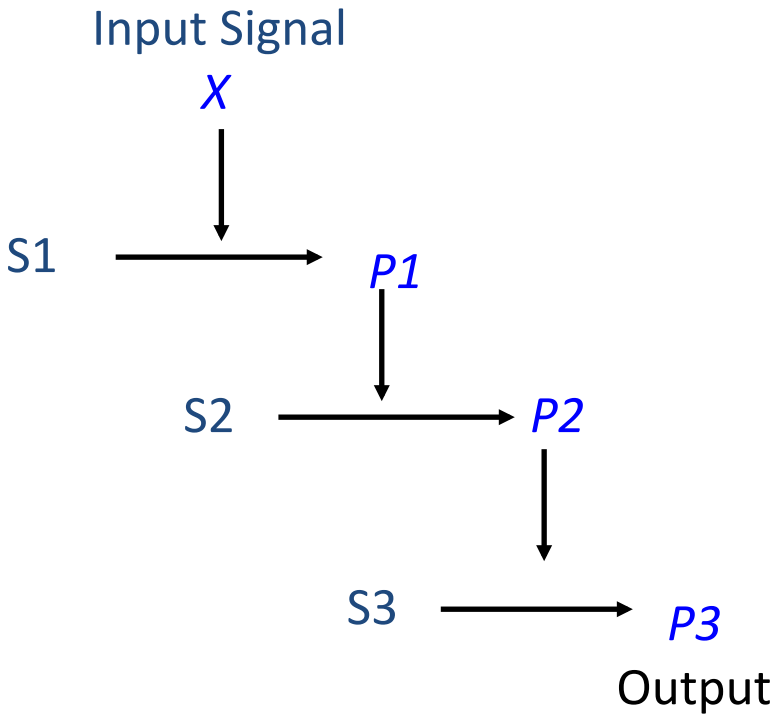
Metabolic pathways vs Signalling Pathways

Metabolic



Classical enzyme-product pathway

Signalling cascade



Product become enzyme at next stage

Biological description → bigraph → differential equations

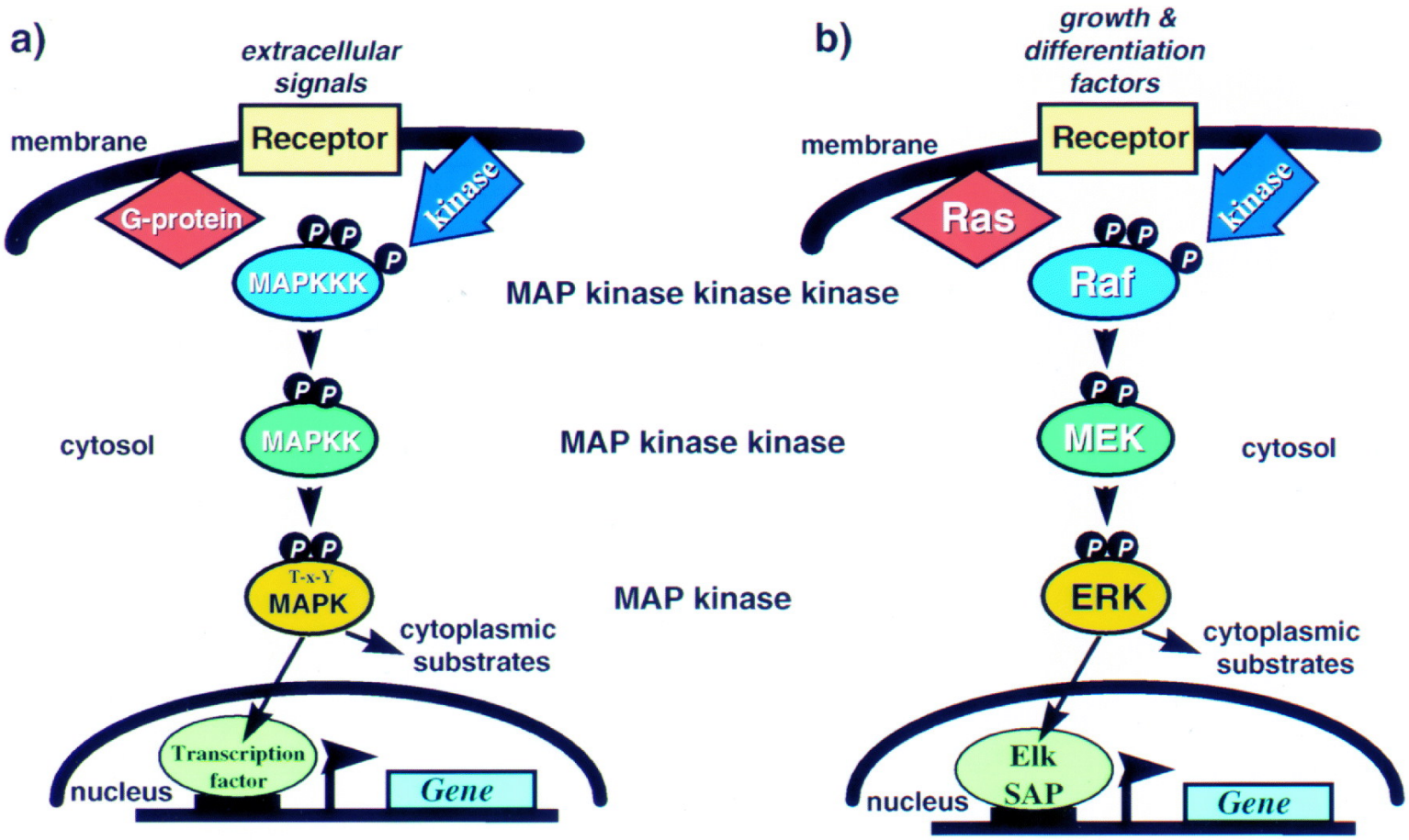


Fig. courtesy of W. Kolch

Biological description → bigraph → differential equations

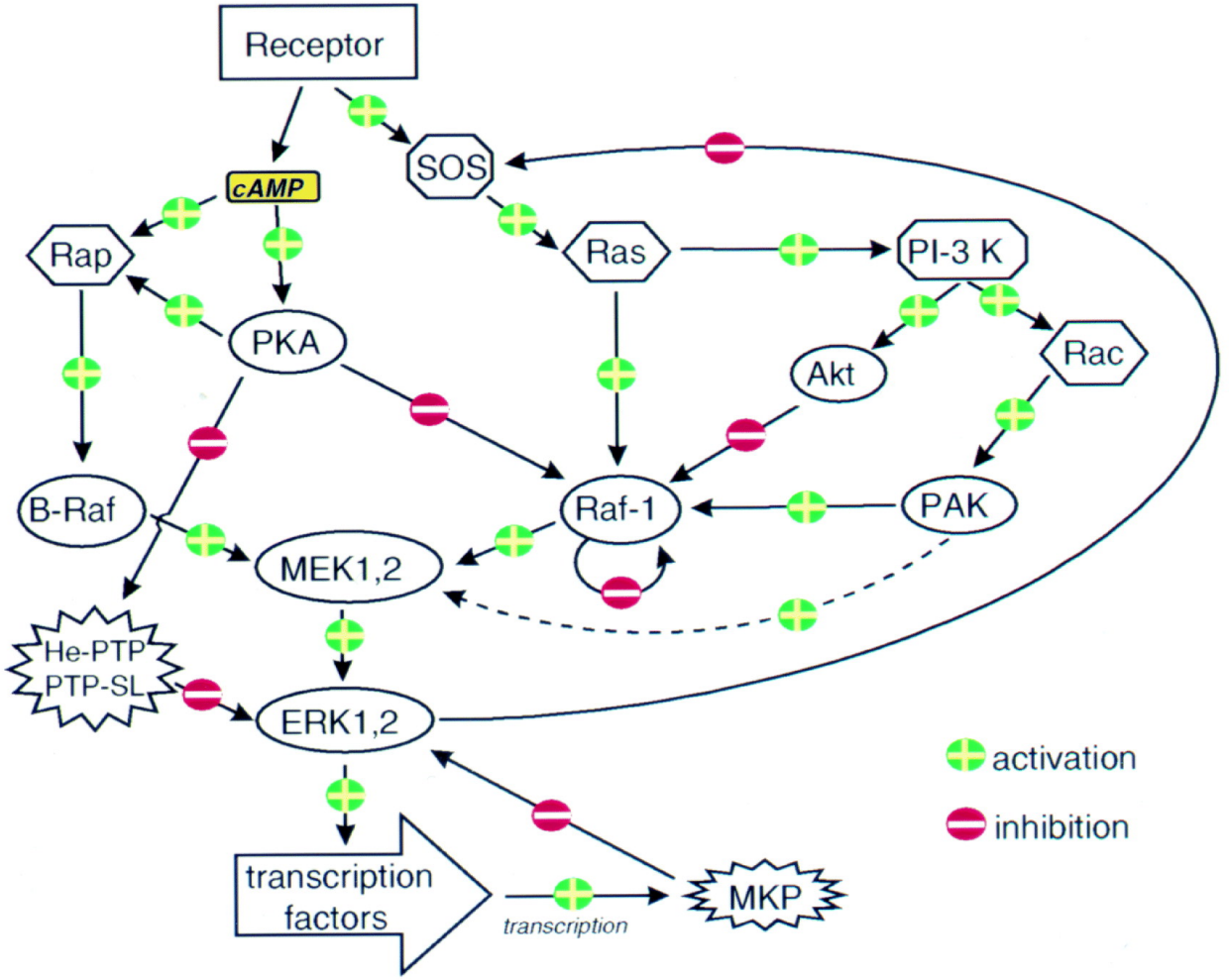
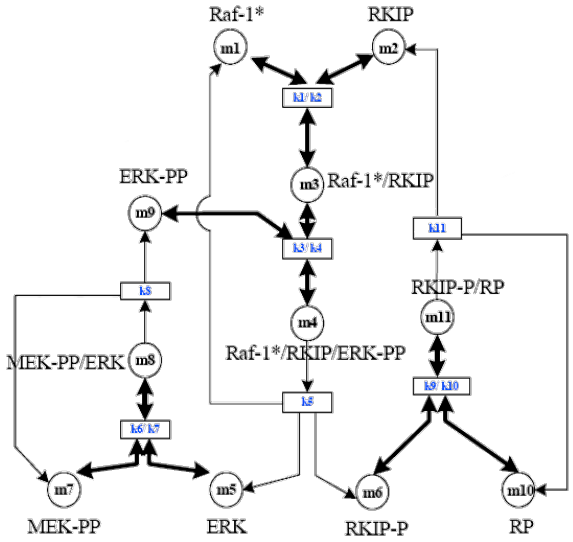


Fig. courtesy of W. Kolch

What is a biochemical network model?

1. Structure



graph
QUALITATIVE

2. Kinetics (if you can)

reaction rates

$$d[Raf1^*]/dt = k1*m1*m2 + k2*m3 + k5*m4$$

$$k1 = 0.53;\ k2 = 0.0072;\ k5 = 0.0315$$

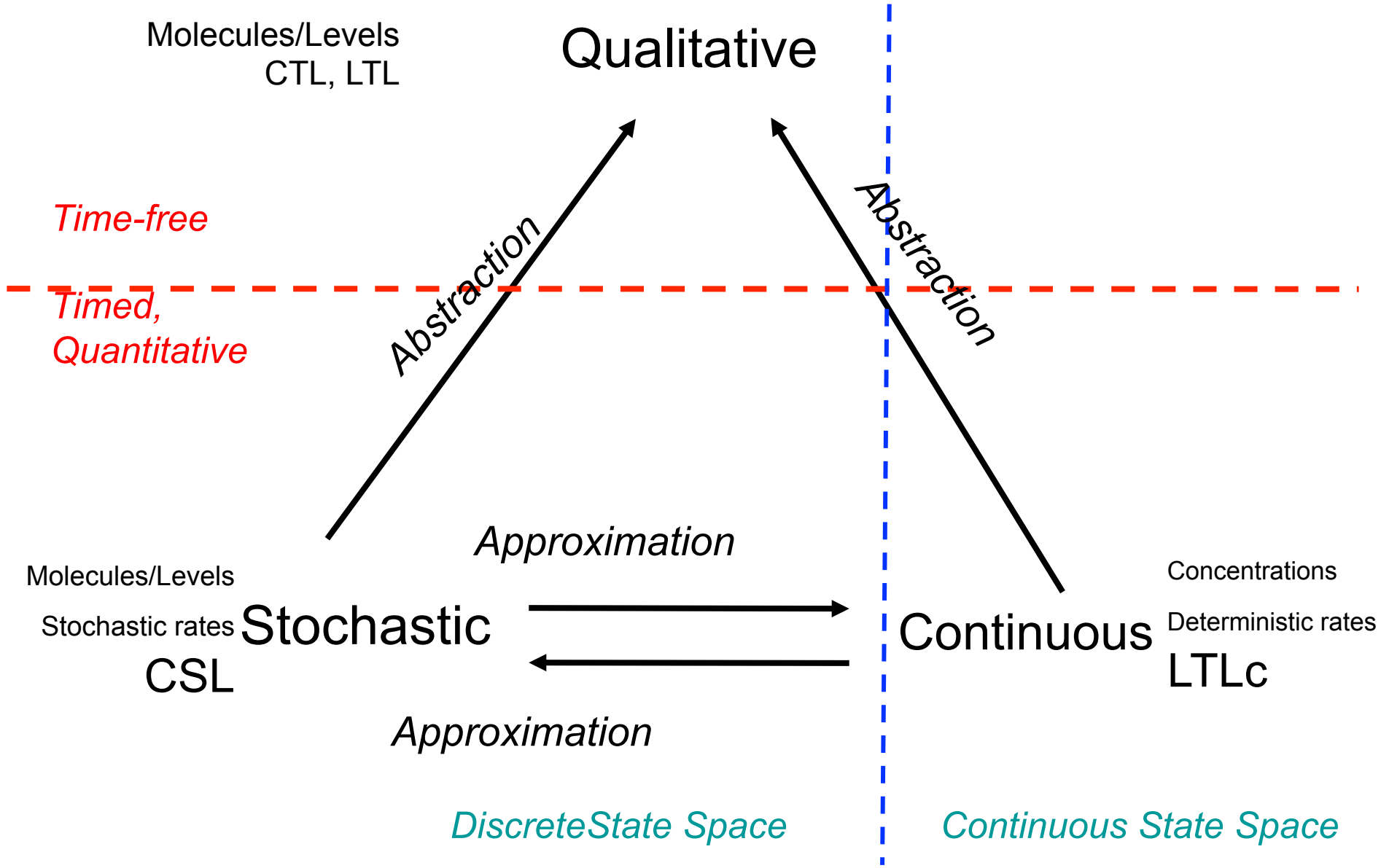
QUANTITATIVE

3. Initial conditions

marking , concentrations

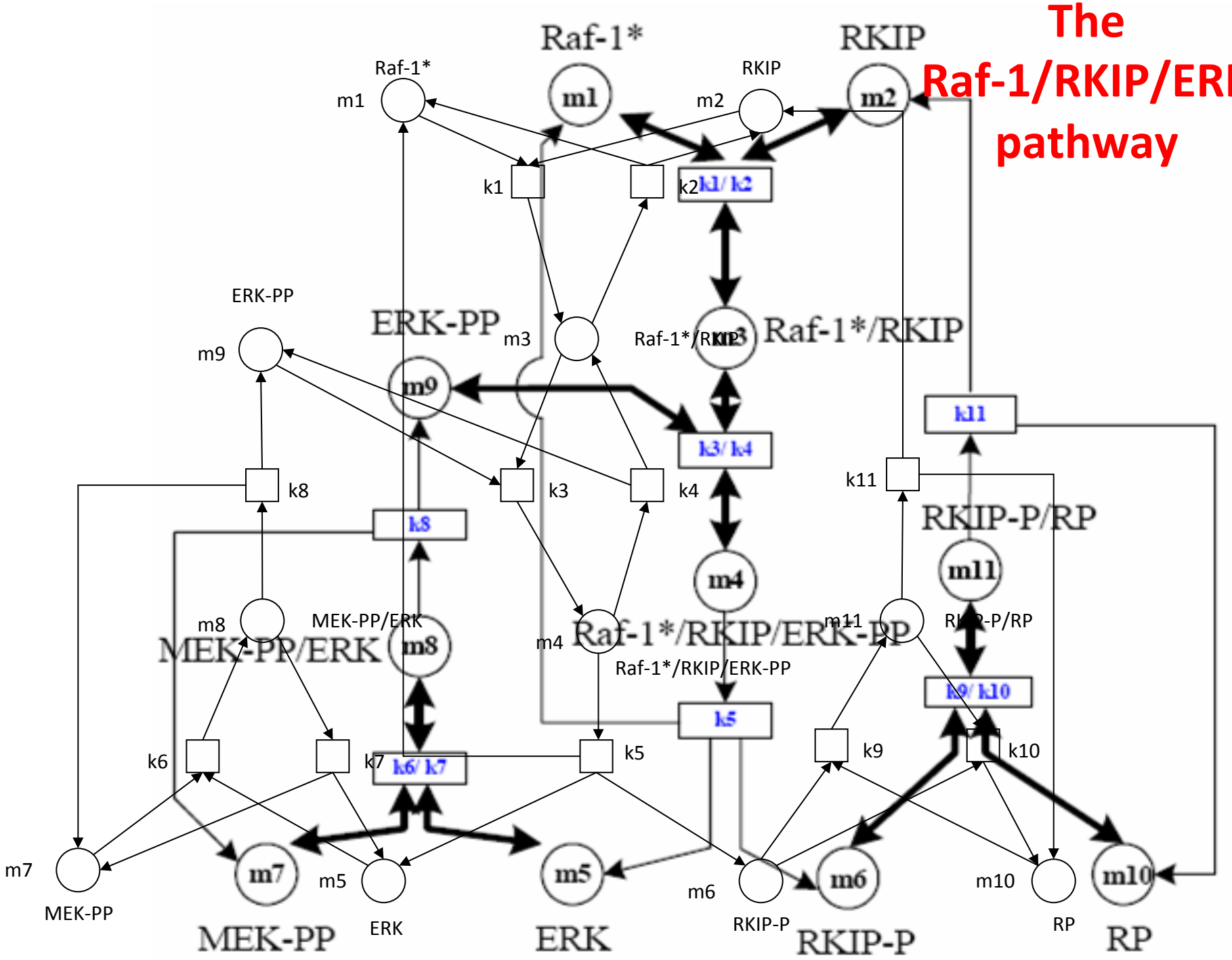
$[Raf1^*]_{t=0}$= 2 μMolar

QUANTITATIVE

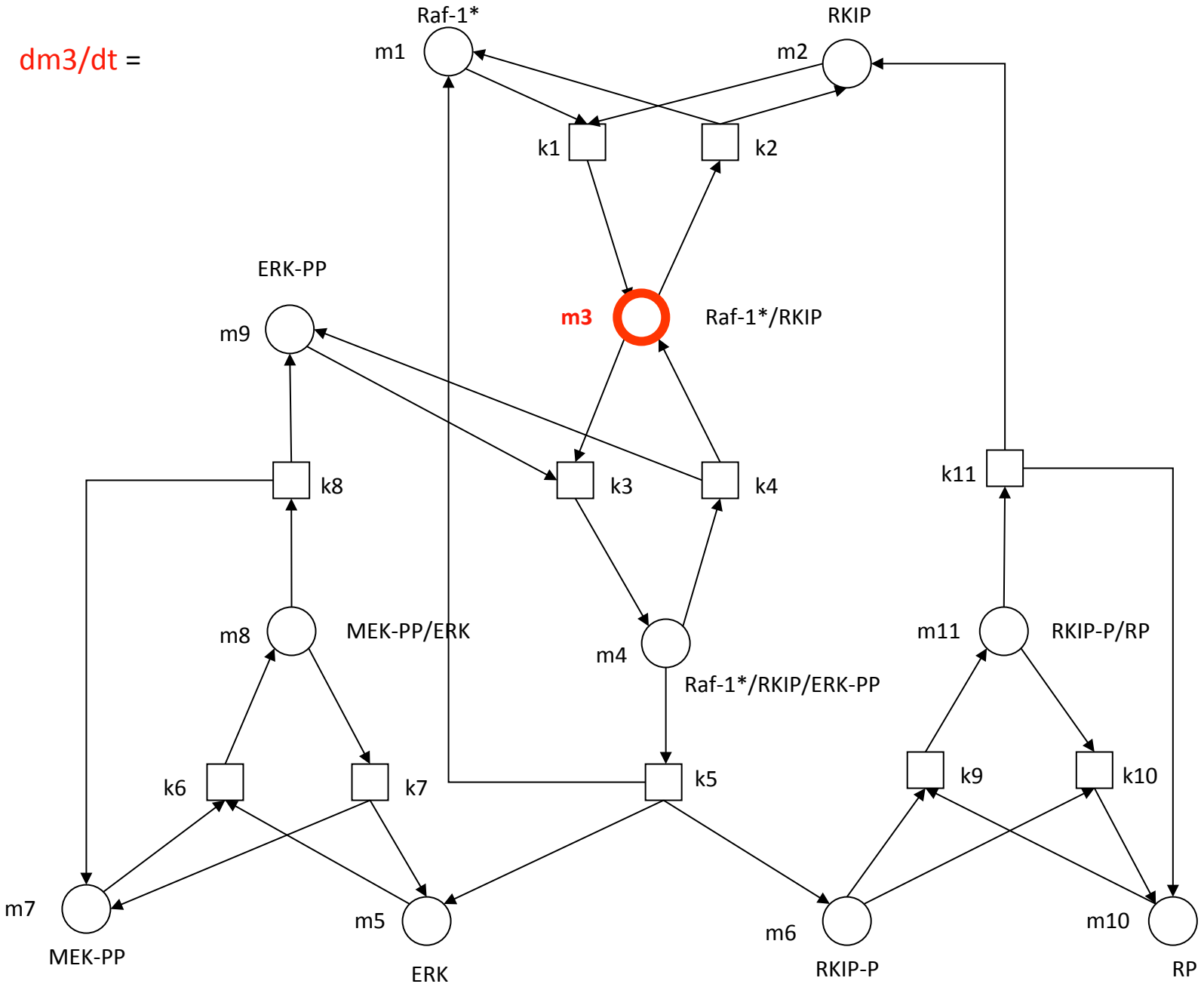


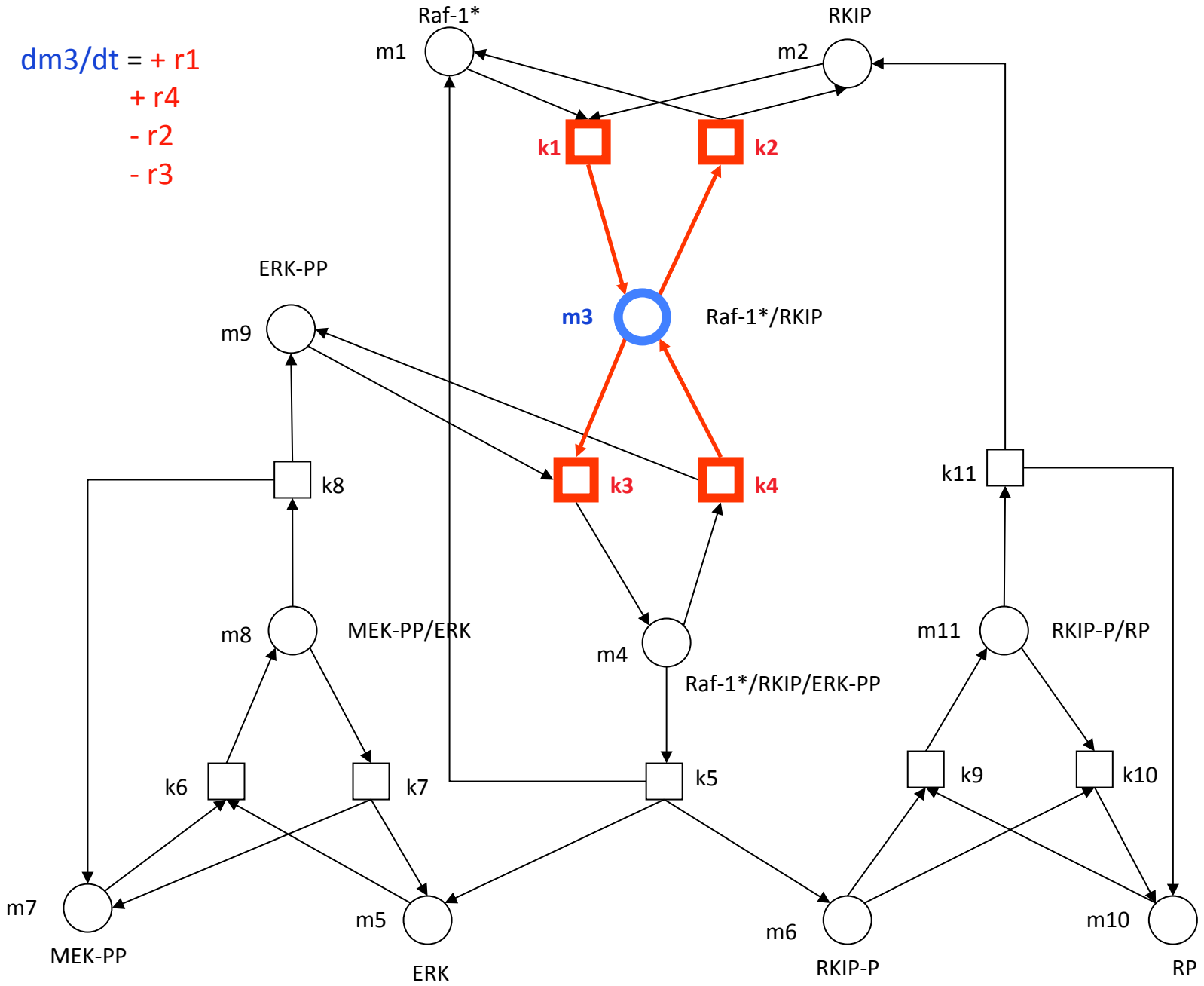
Gilbert, Heiner and Lehrack. "A Unifying Framework for Modelling and Analysing Biochemical Pathways Using Petri Nets." Proc CMSB 2007

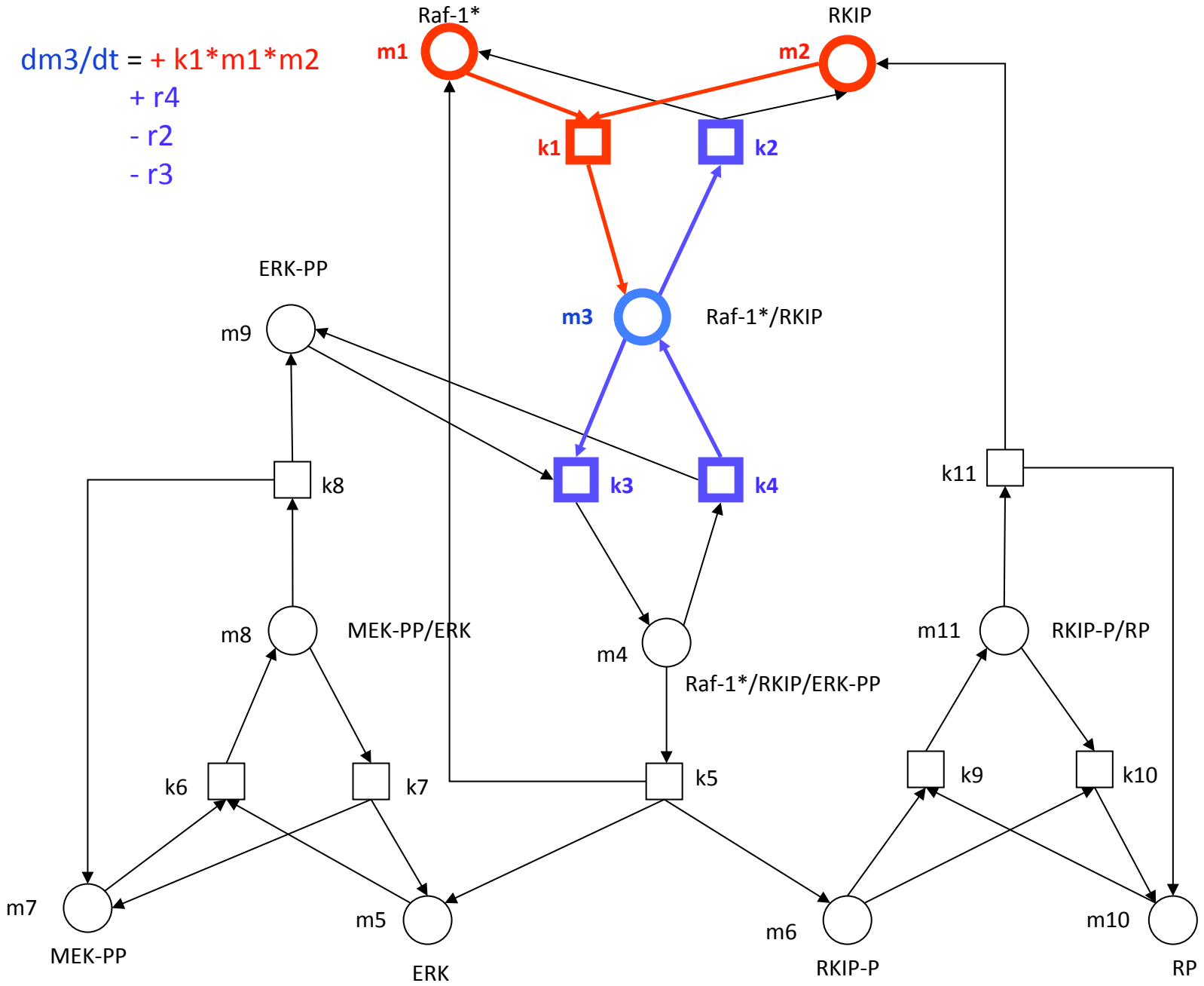
The Raf-1/RKIP/ERK pathway

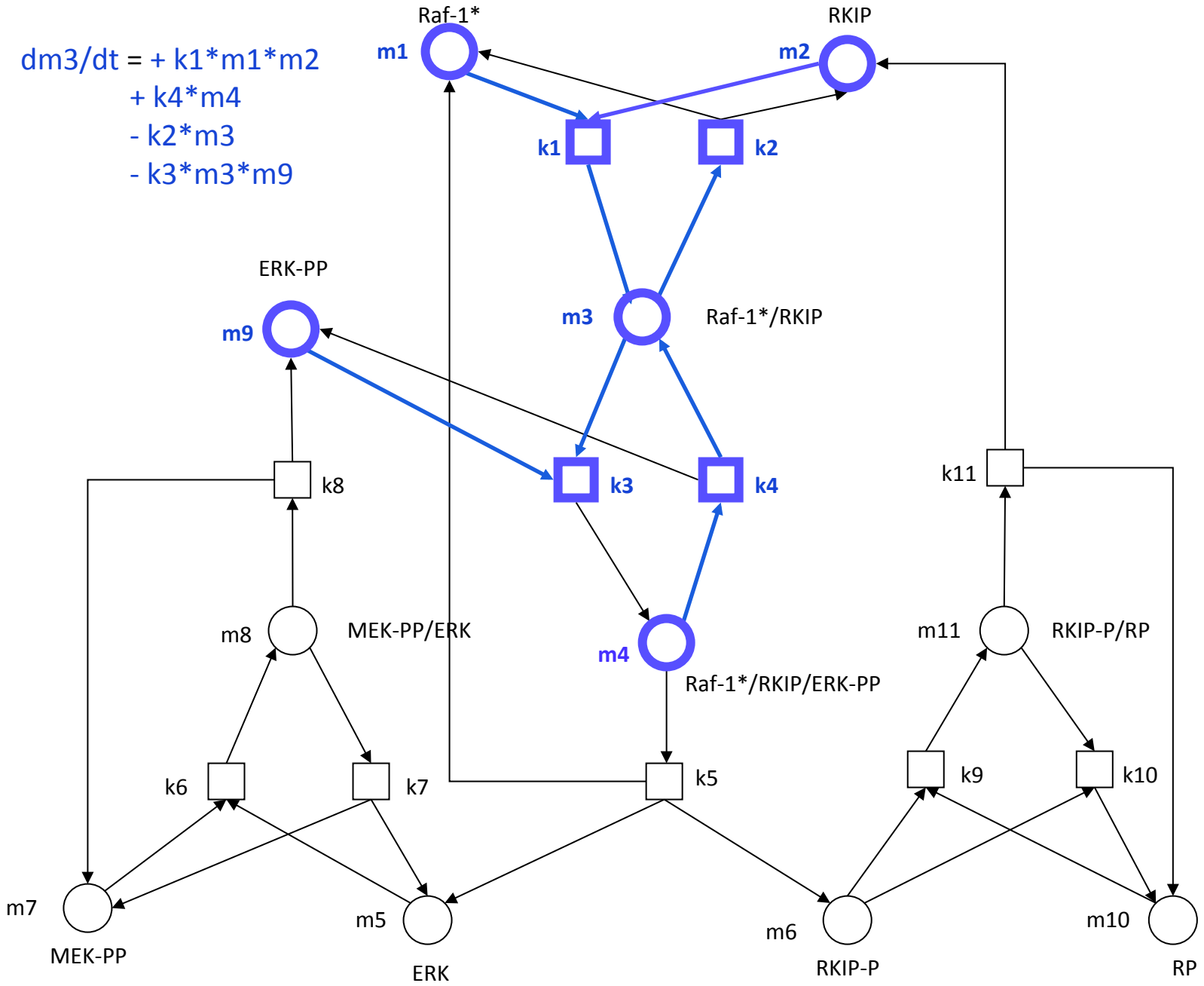


dm3/dt =

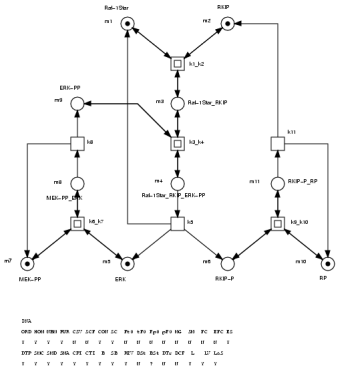

Raf-1*
m1
RKIP
m2
k1
k2
ERK-PP
m9
m3
Raf-1*/RKIP
k8
k3
k4
k11
m8
MEK-PP/ERK
m4
Raf-1*/RKIP/ERK-PP
m11
RKIP-P/RP
k6
k7
k5
k9
k10
m7
MEK-PP
m5
ERK
m6
RKIP-P
m10
RP

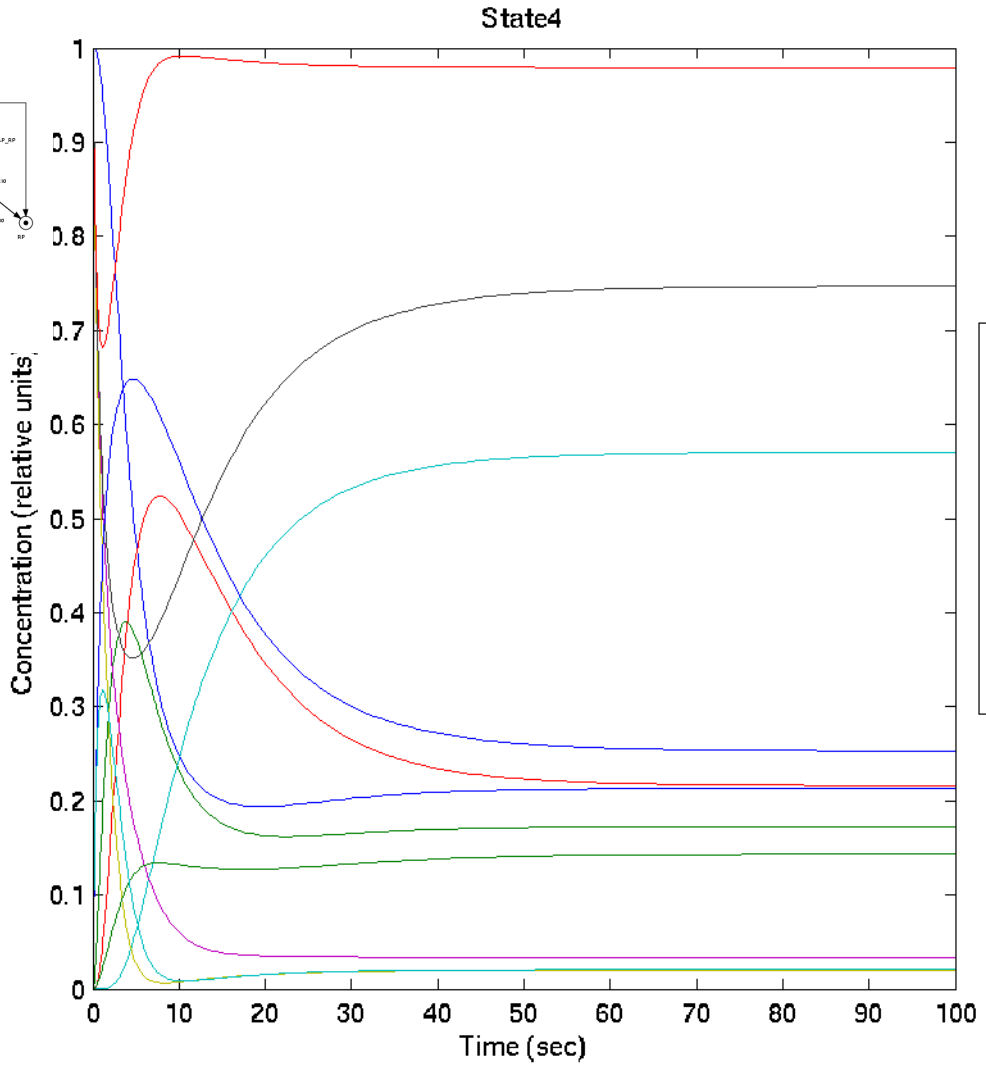

dm3/dt = + r1
+ r4
- r2
- r3
Raf-1*
m1
RKIP
m2
k1
k2
ERK-PP
m9
m3
Raf-1*/RKIP
k3
k4
k8
k11
m8
MEK-PP/ERK
m4
Raf-1*/RKIP/ERK-PP
m11
RKIP-P/RP
k6
k7
k5
k9
k10
m7
MEK-PP
m5
ERK
m6
RKIP-P
m10
RP

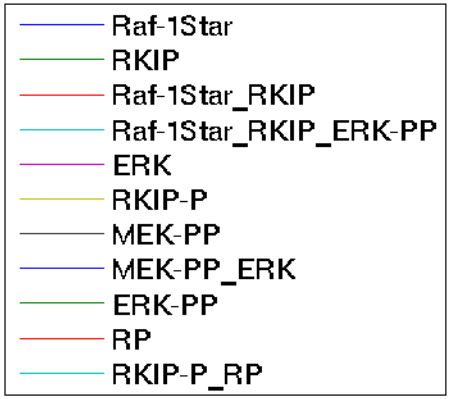

dm3/dt = + k1*m1*m2
+ r4
- r2
- r3
Raf-1*
m1
RKIP
m2
k1
k2
ERK-PP
m9
m3
Raf-1*/RKIP
k3
k4
k8
k11
m8
MEK-PP/ERK
m4
Raf-1*/RKIP/ERK-PP
m11
RKIP-P/RP
k6
k7
k5
k9
k10
m7
MEK-PP
m5
ERK
m6
RKIP-P
m10
RP

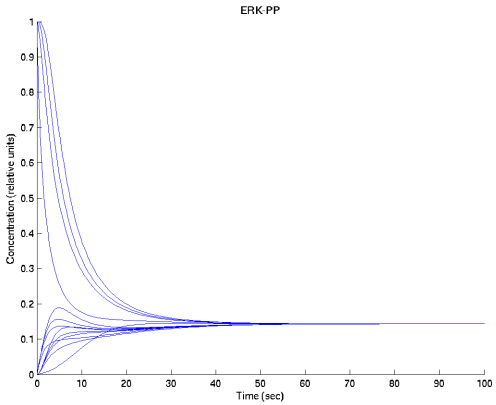

dm3/dt = + k1*m1*m2
+ k4*m4
- k2*m3
- k3*m3*m9
Raf-1*
m1
RKIP
m2
k1
k2
ERK-PP
m9
m3
Raf-1*/RKIP
k3
k4
k8
k11
m8
MEK-PP/ERK
m4
Raf-1*/RKIP/ERK-PP
m11
RKIP-P/RP
k6
k7
k5
k9
k10
m7
MEK-PP
m5
ERK
m6
RKIP-P
m10
RP

From Petri Nets to Differential Equations - an Integrative Approach
David Gilbert & Monika Heiner, Proc ATPN06





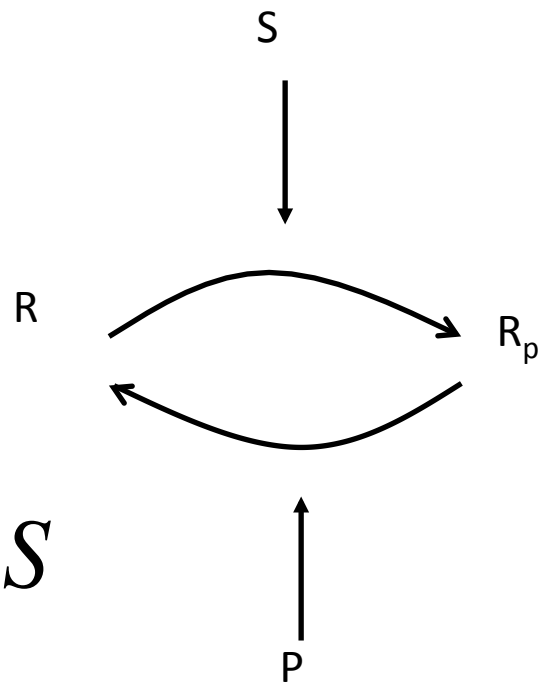




Demo time!
RKIP

Phosphorylation - dephosphorylation step
Mass action

- R: unphosphorylated form
- R_p: phosphorylated form
- S: kinase
- P: phosphotase
- R|S unphosphorylated+kinase complex
- R|P unphosphorylated+phosphotase complex

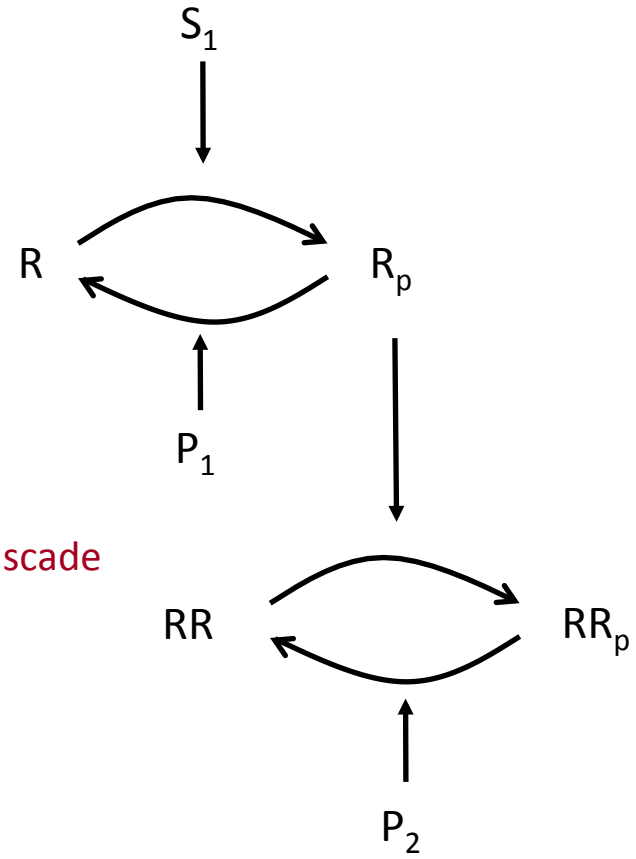


$$R+S \underset{k_2}{\overset{k_1}{\rightleftarrows}} R|S \xrightarrow{k_3} R_p + S$$

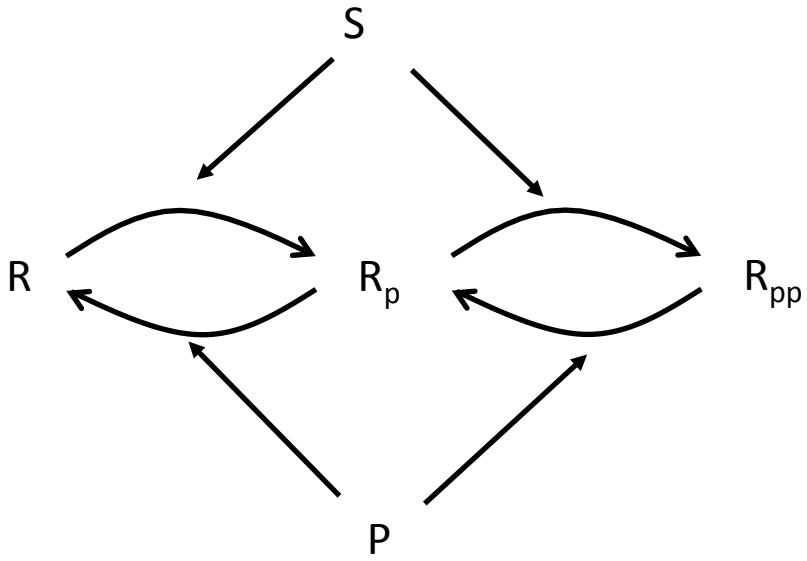
$$R+P \xleftarrow{kr_3} R_p|P \underset{kr_2}{\overset{kr_1}{\rightleftarrows}} R_p + P$$

Breitling, Gilbert, Heiner & Orton "A structured approach for the engineering of biochemical models, illustrated for signalling pathways". Briefings in Bioinformatics, 2008

Composition
Vertical & horizontal



2-stage cascade



1-stage cascade
double phosphorylation step

Cell signaling pathways – feedback loops

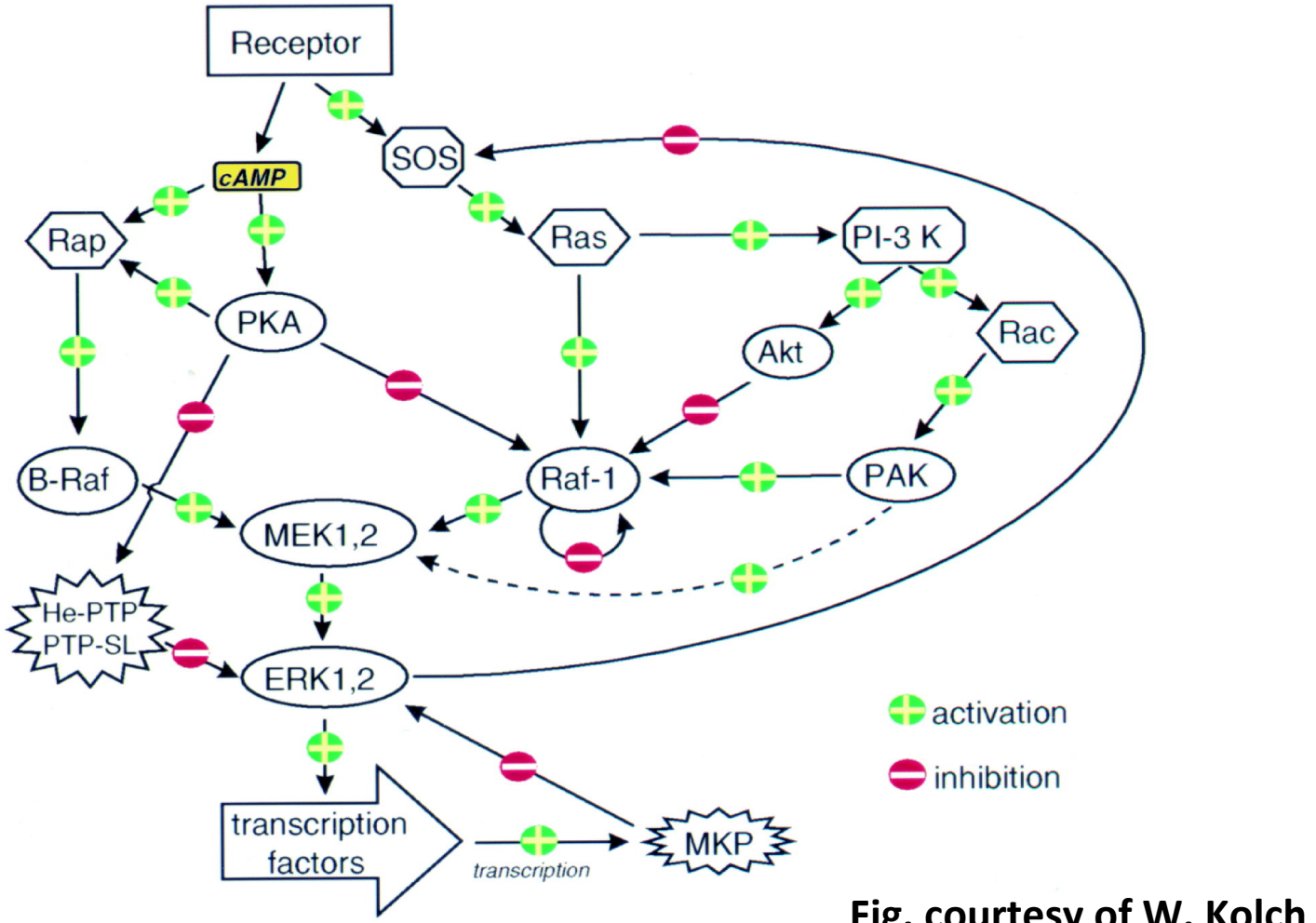
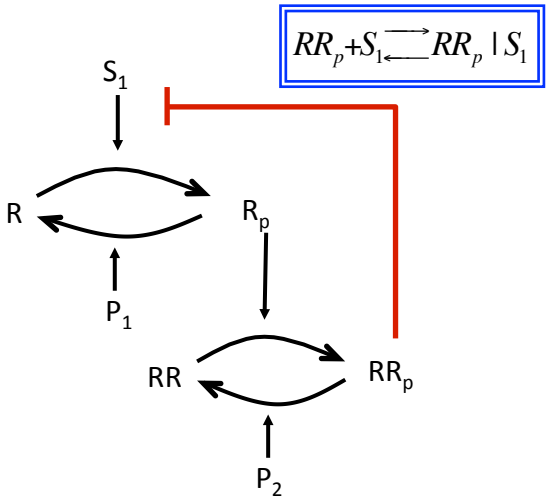
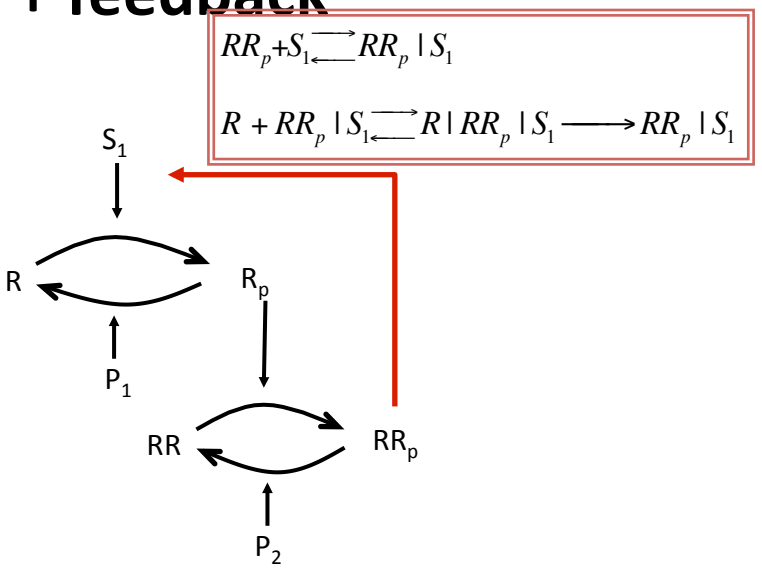
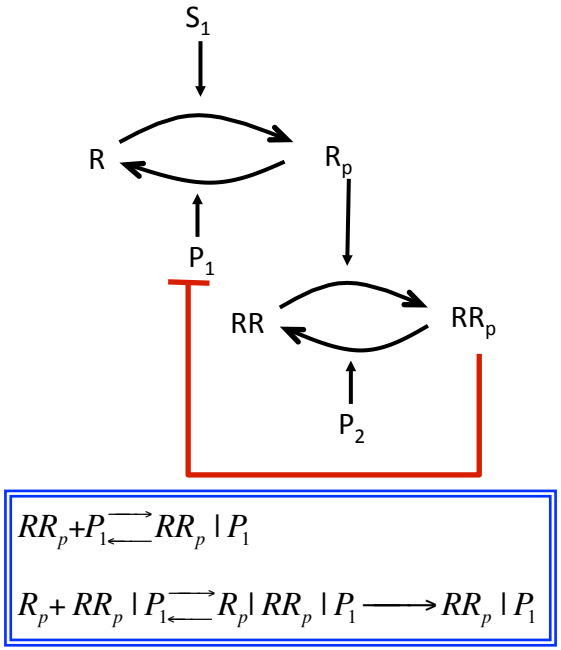


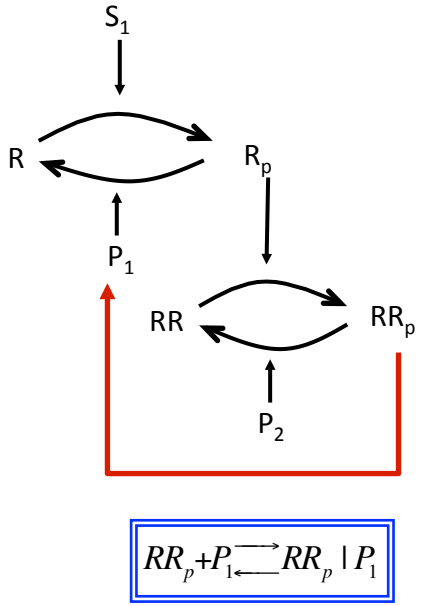
Fig. courtesy of W. Kolch

Phosphorylation cascade + feedback

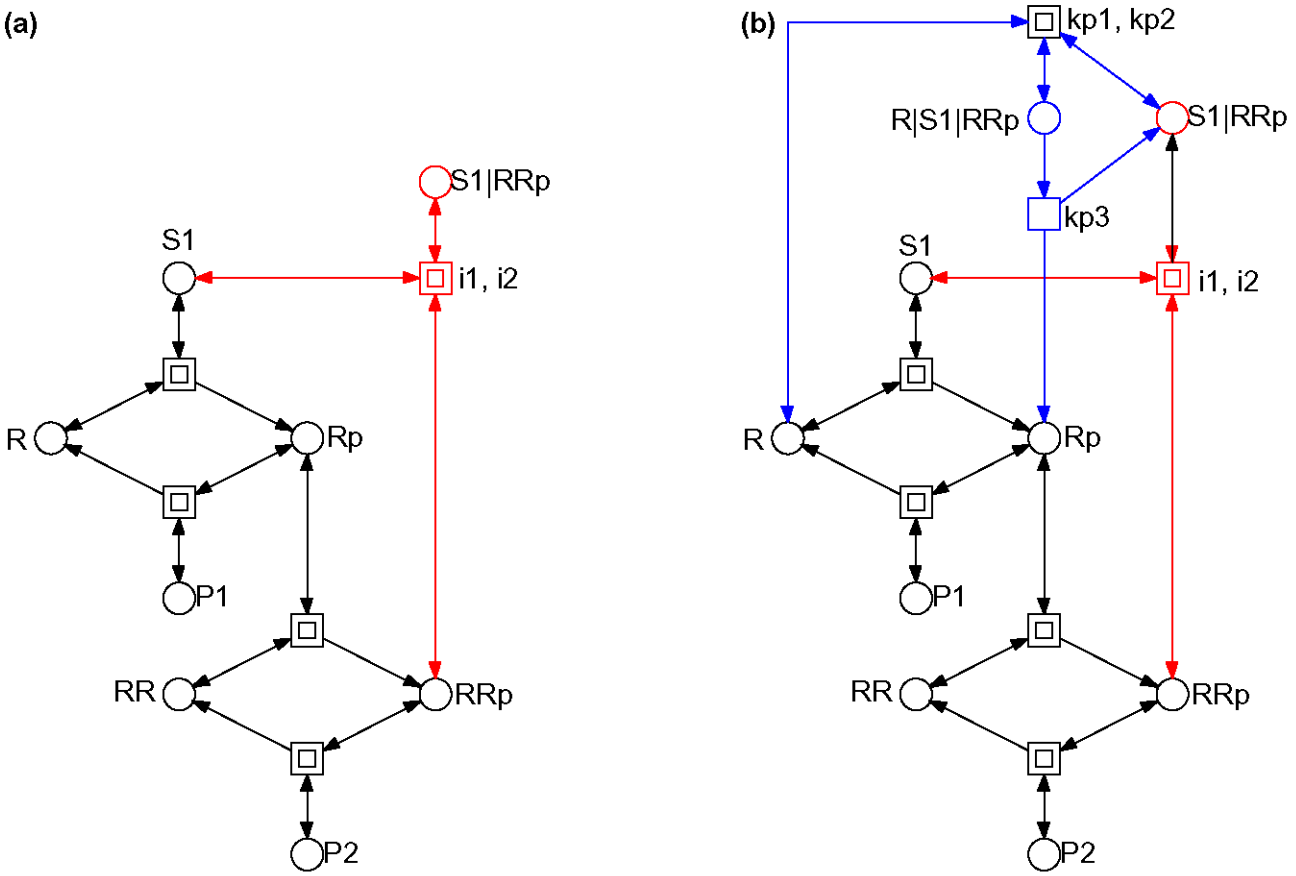




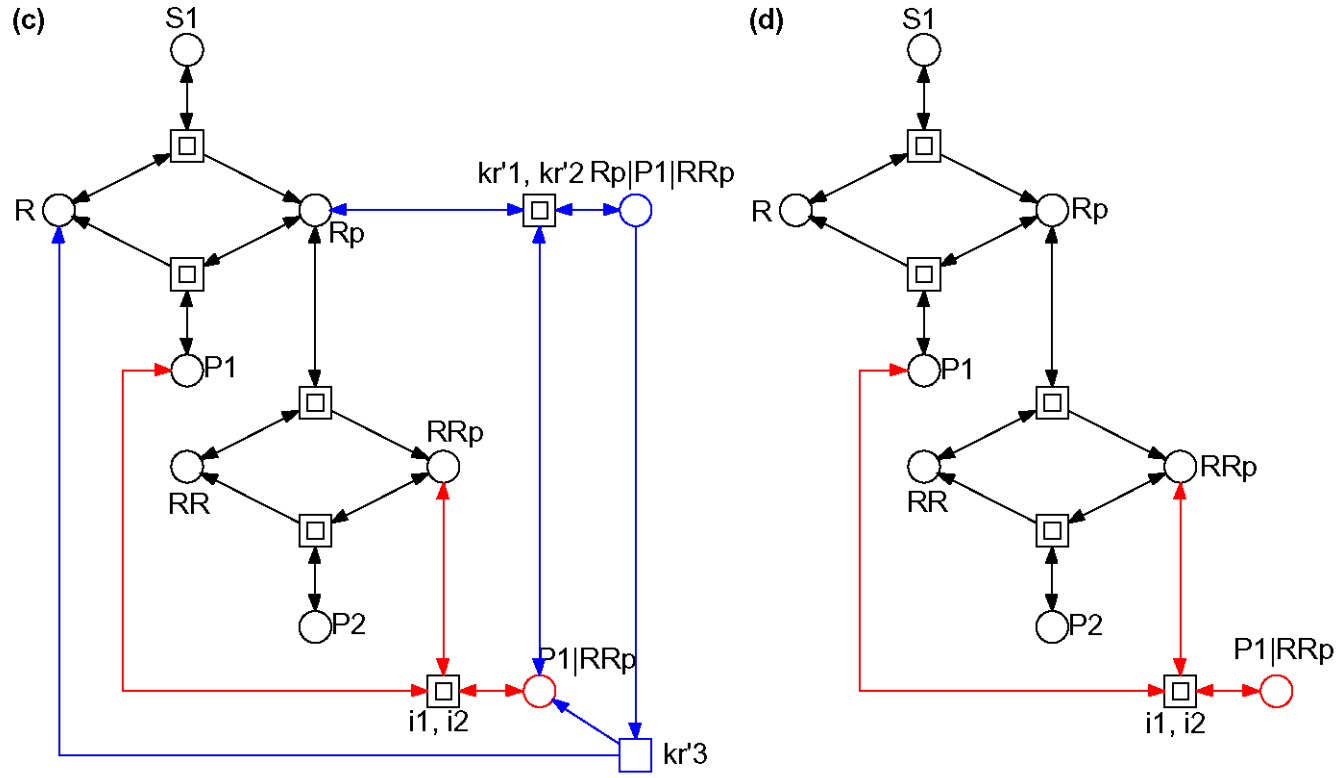




Feedback loops in Petri Nets



Feedback loops in Petri Nets

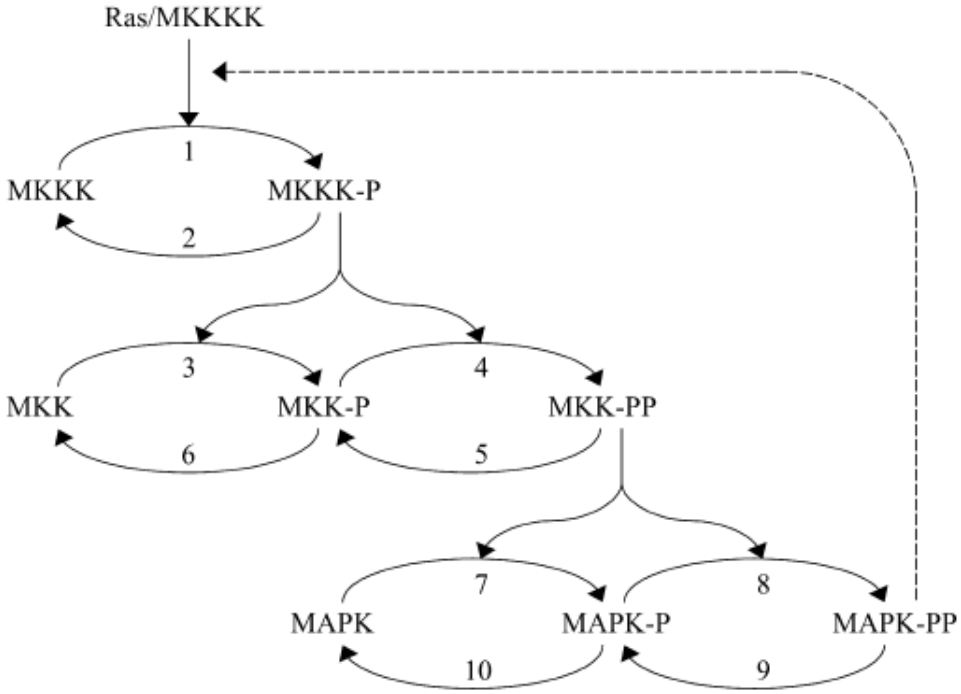


MAPK Pathway - Kholodenko

- Continuous model of the Mitogen Activated Protein Kinase (MAPK) Pathway

- Kholodenko (2000): The oscillatory behaviour is because of the negative feedback loop.

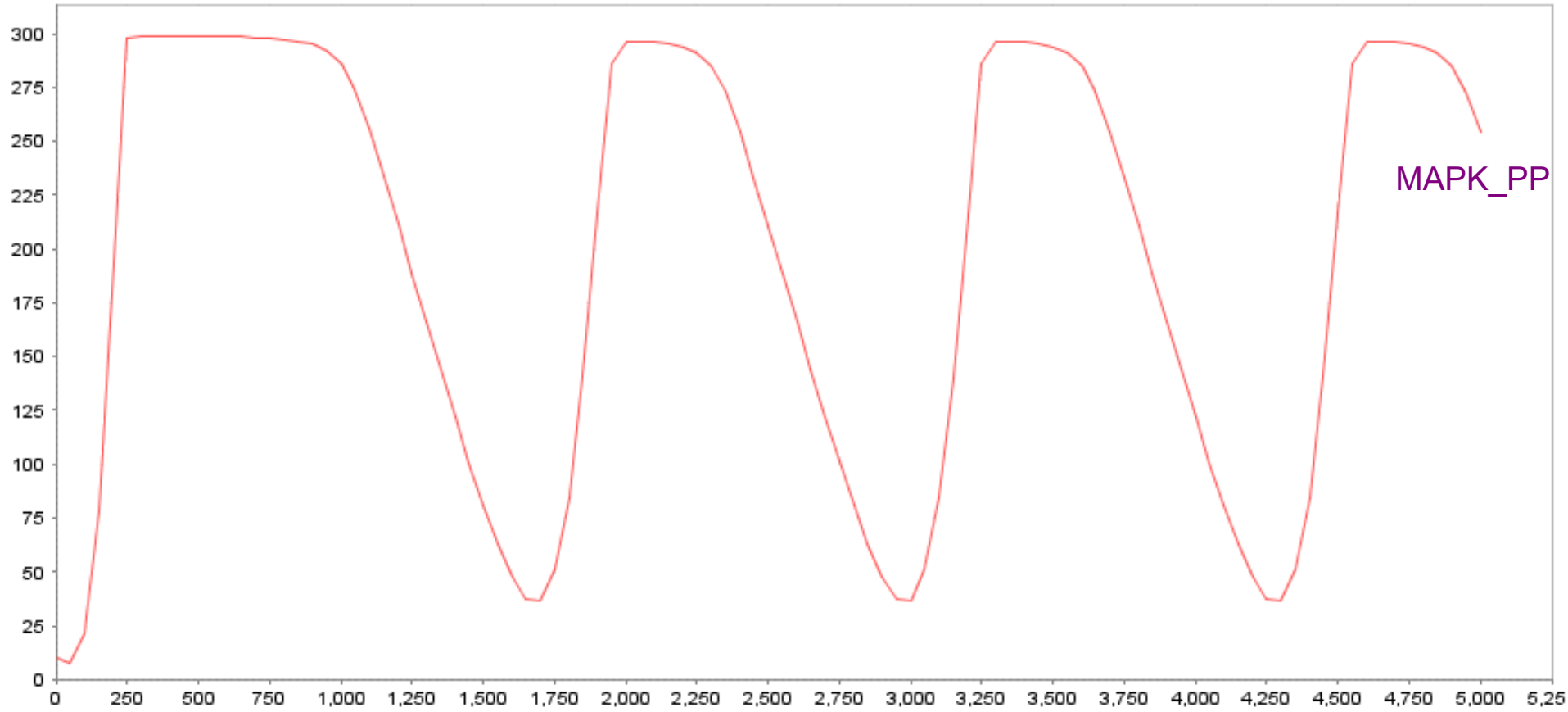
- Using parameter scanning, we vary the negative feedback effect (parameter Ki) and assess the effect on #oscillations detected:



Kholodenko (2000), "Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades". Eur. J. Biochem. 2000, 267(6): 1583–1588.

Kholodenko Model

- Oscillatory output:



Demo time!
Kholodenko

Adding a drug (inhibitor on 2nd stage)

$$RRR_p + S_1 \underset{ki'}{\overset{ki}{\rightleftarrows}} RRR_p \mid S_1$$

$$R + S_1 \underset{k_2}{\overset{k_1}{\rightleftarrows}} R \mid S_1 \xrightarrow{k_3} R_p + S_1$$

$$R + P_1 \xleftarrow{k_3'} R_p \mid P_1 \underset{k_2'}{\overset{k_1'}{\rightleftarrows}} R_p + P_1$$

$$RR + R_p \underset{kk_2}{\overset{kk_1}{\rightleftarrows}} RR \mid R_p \xrightarrow{kk_3} RR_p + R_p$$

$$RR + P_2 \xleftarrow{kk_3'} RR_p \mid P_2 \underset{kk_2'}{\overset{kk_1'}{\rightleftarrows}} RR_p + P_2$$

$$RRR + RR_p \underset{kkk_2}{\overset{kkk_1}{\rightleftarrows}} RRR \mid RR_p \xrightarrow{kkk_3} RRR_p + RR_p$$

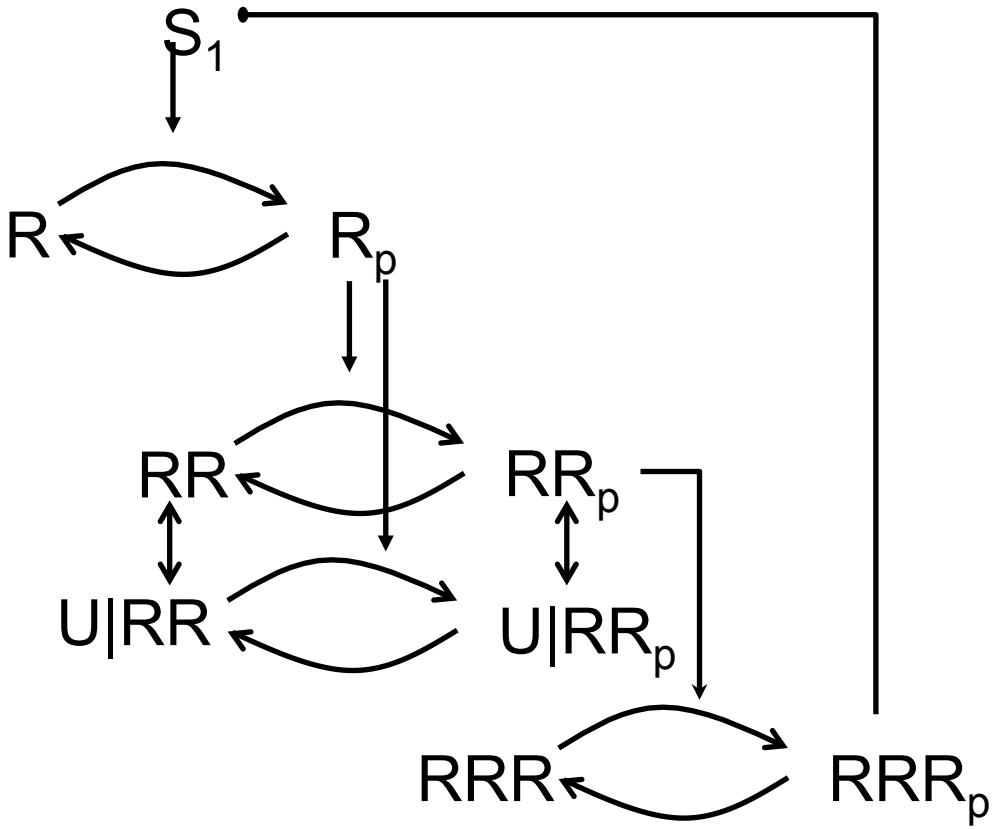
$$RRR + P_3 \xleftarrow{kkk_3'} RRR_p \mid P_3 \underset{kkk_2'}{\overset{kkk_1'}{\rightleftarrows}} RRR_p + P_3$$

$$U + RR \underset{ku_2}{\overset{ku_1}{\rightleftarrows}} U \mid RR$$

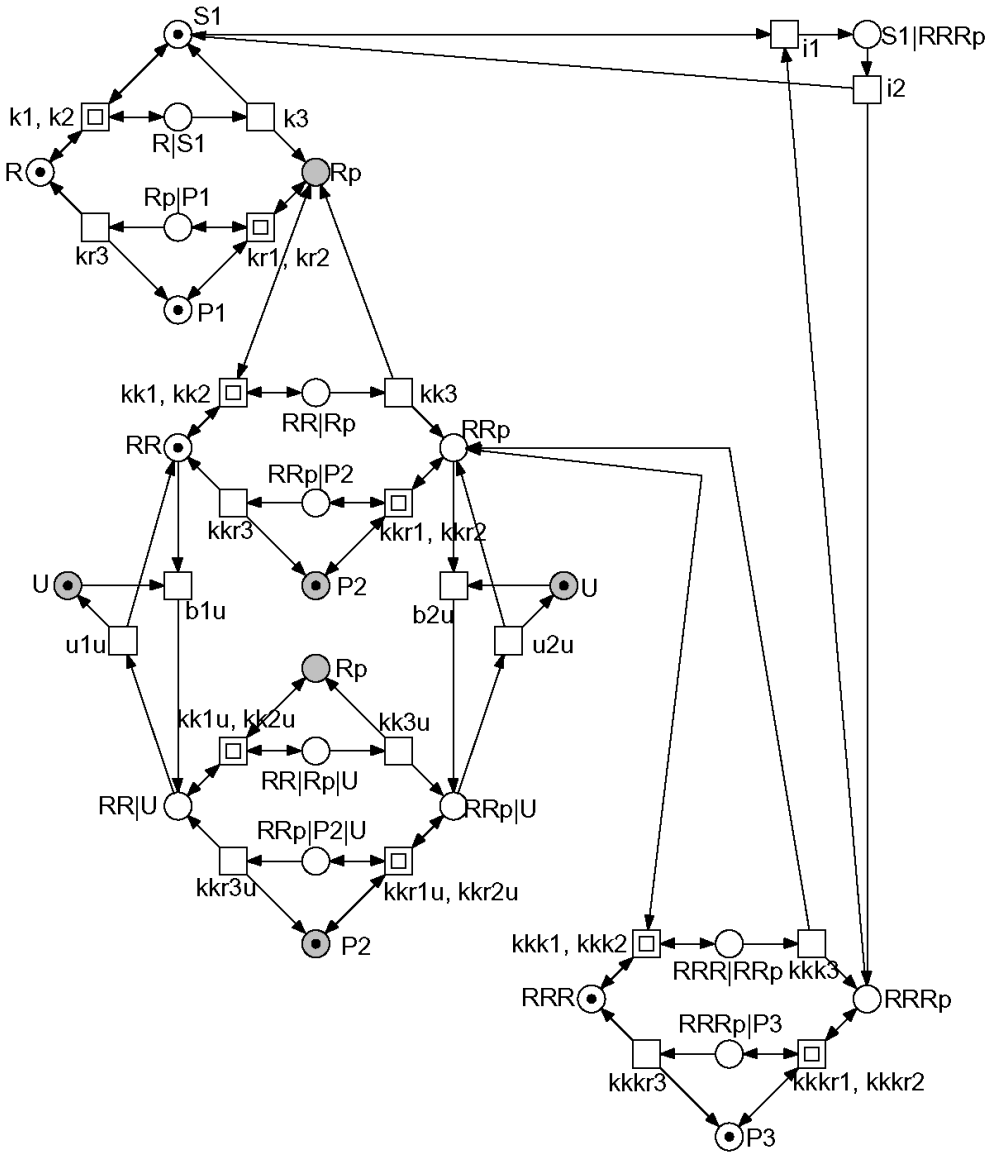
$$U + RR_p \underset{ku_2}{\overset{ku_1}{\rightleftarrows}} U \mid RR_p$$

$$U \mid RR + R_p \underset{kk_2}{\overset{kk_1}{\rightleftarrows}} U \mid RR \mid R_p \xrightarrow{kk_3} U \mid RR_p + R_p$$

$$U \mid RR + P_2 \xleftarrow{kk_3'} U \mid RR_p \mid P_2 \underset{kk_2'}{\overset{kk_1'}{\rightleftarrows}} U \mid RR_p + P_2$$

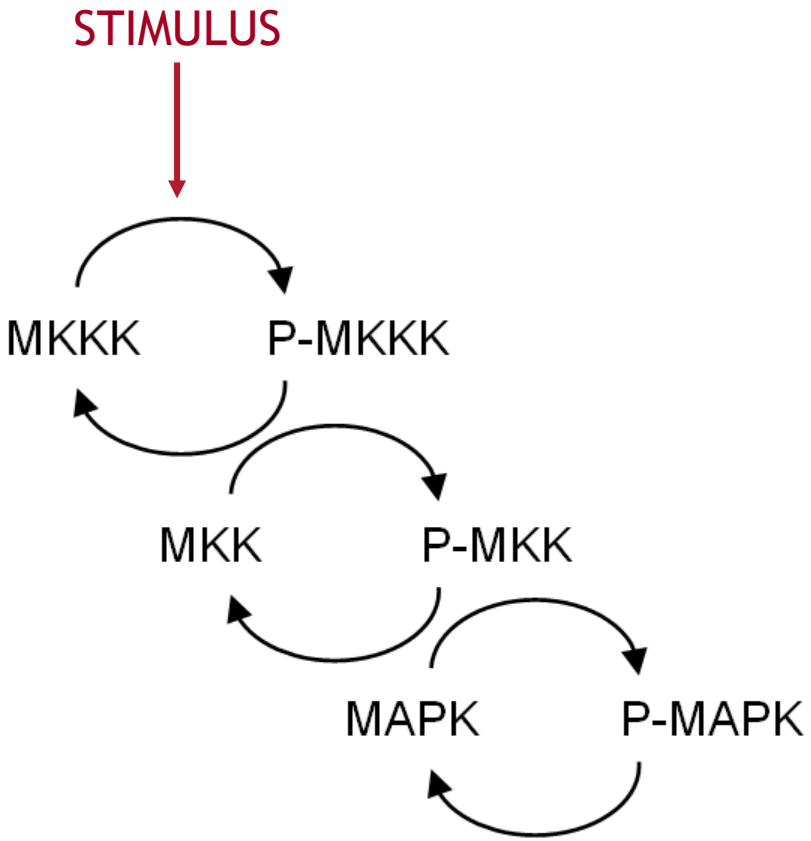


…and added inhibitor



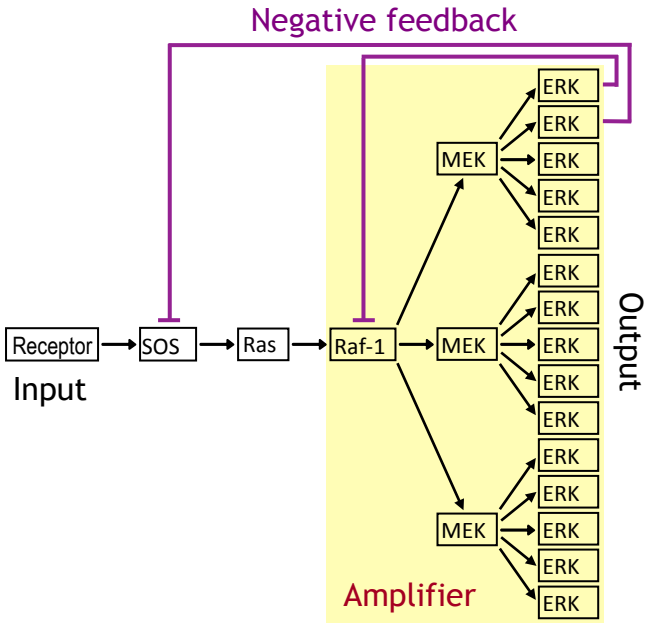
MAPK Pathway

- Responds to wide range of stimuli: cytokines, growth factors, neurotransmitters, cellular stress and cell adherence,…
- Pivotal role in many key cellular processes:
  - growth control in all its variations,
  - cell differentiation and survival
  - cellular adaptation to chemical and physical stress.
- Deregulated in various diseases: cancer; immunological, inflammatory and degenerative syndromes,
- Represents an important drug target.



ERK cascade well known biological amplifier

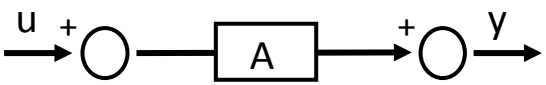
- Amplifies the original signal to create effective cellular responses.
- 1:3:5 are the approximate ratios of Raf-1, MEK and ERK in fibroblasts.
- Well known negative feedback loop: phosphorylation of SOS by ERK-PP (via MAPKAP1) resulting in the dissociation of the Grb2/SOS complex.
- New negative feedback loop: ERK-PP phosphorylates Raf-1 resulting in a hyper-phosphorylated inactive form of Raf (Dougherty *et al*. 2005)



Dougherty *et al*. (2005), Regulation of Raf-1 by Direct Feedback Phosphorylation, Molecular Cell 17 215-224

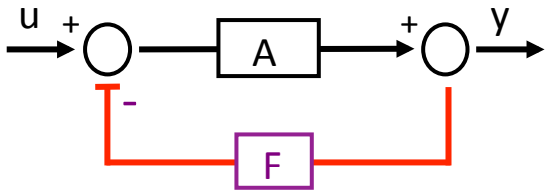
Negative Feedback Amplifier

Standard Amplifier



$y = u*A$

Negative Feedback Amplifier



$$y = u \frac{A}{1+A*F}$$

- Negative feedback amplifier from electronics
- Amplifier with a negative feedback loop from the output of the amplifier to its input.

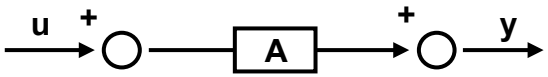
- NF loop → a system much more robust to disturbances in the amplifier.

- NFA was invented in 1927 by Harold Black of Western Electric.
- Originally used for reducing distortion in long distance telephone lines.

- NFA a key electrical component used in a wide variety of applications

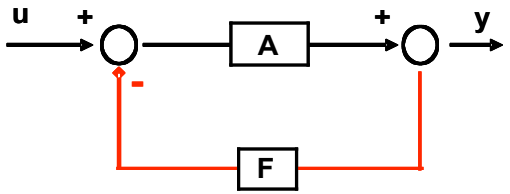
The negative feedback imparts signalling robustness

Standard Amplifier



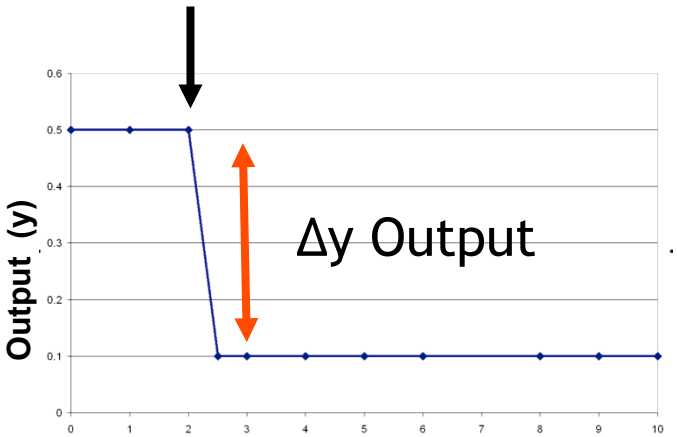
y=A*u

Negative Feedback Amplifier

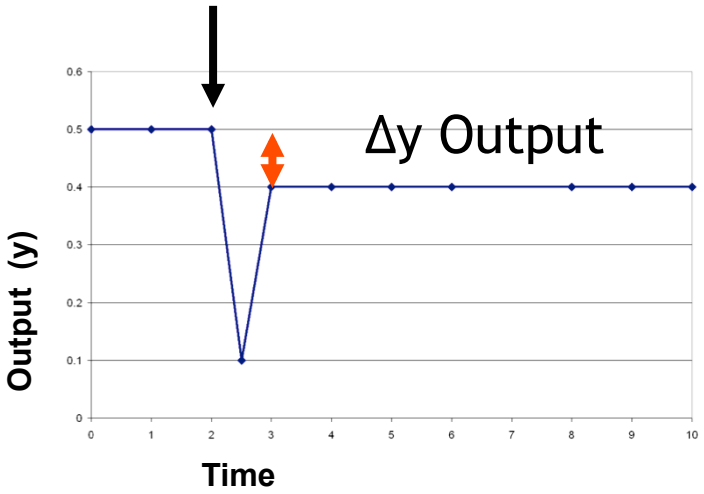


y=A*u/(1+A*F)

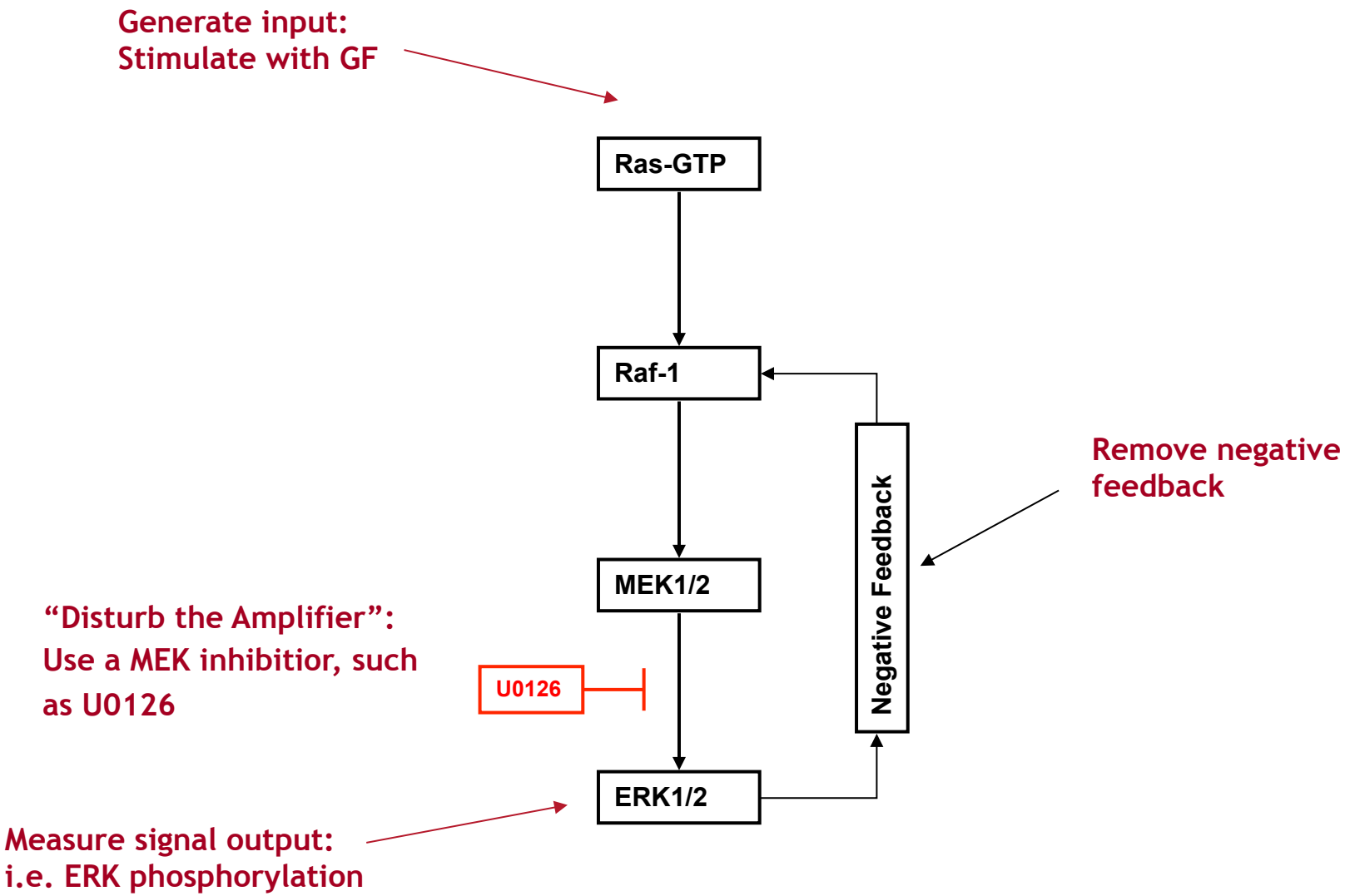
Sudden drop in Amplifier (A) gain



Sudden drop in Amplifier (A) gain

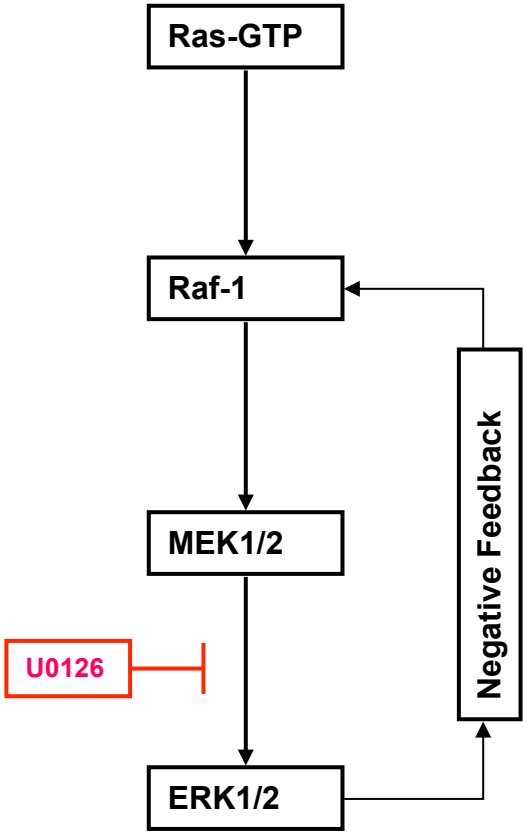


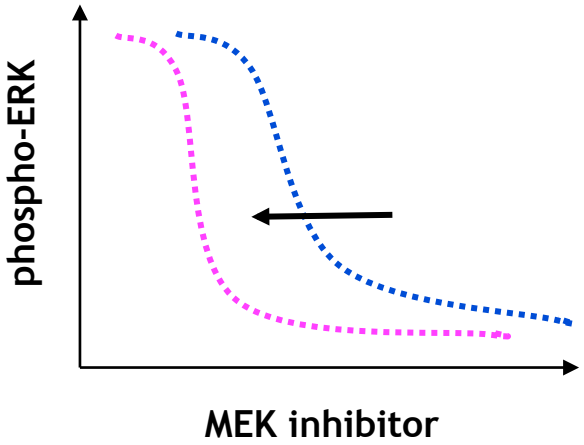
How to test if the ERK pathway is a NFA?



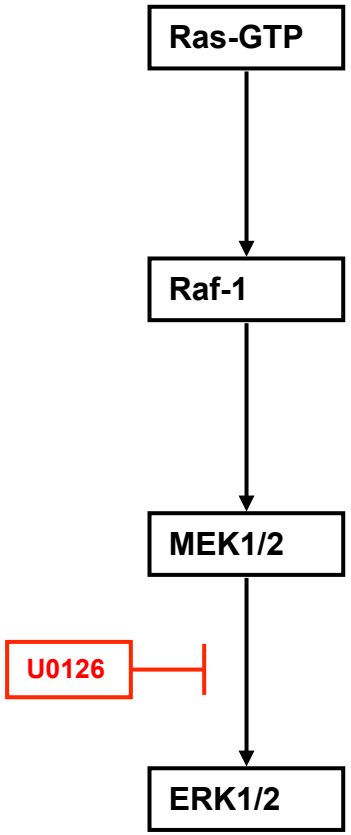
Hypothesis: Breaking the feedback should sensitise the ERK pathway to MEK-inhibitor

Feedback intact





Feedback removed



How to test if the ERK pathway is a NFA?

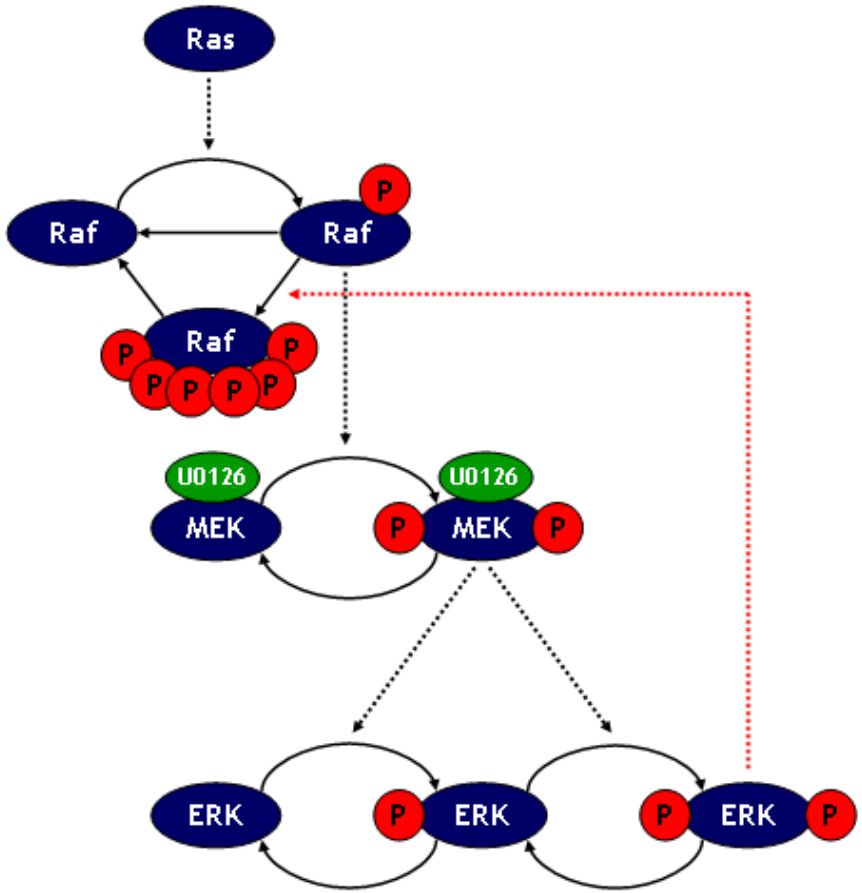
Strategy

In vivo system that allows us to compare feedback broken to feedback intact model.

Computational Model of ERK pathway with/without feedback

Computational Modeling 1: Build the model

- Non-linear ordinary differential equations (ODE's).
- ODE's were solved using Math Lab and Gepasi.
- Models are based on the Schoeberl et al. (2002) model
- Mass Action Kinetics instead of Michaelis Menten
- Kinetic parameters are from literature, previous models and "guesstimates"



Schoeberl *et al*. (2002), Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors, Nature Biotechnology 20, 370-375

The experimental systems

Negative feedback loops intact

One feedback loop eliminated by constitutively active RasV12 mutant

Both feedback loops eliminated by BXB-ER (4-OHT regulatable Raf-1 mutant)

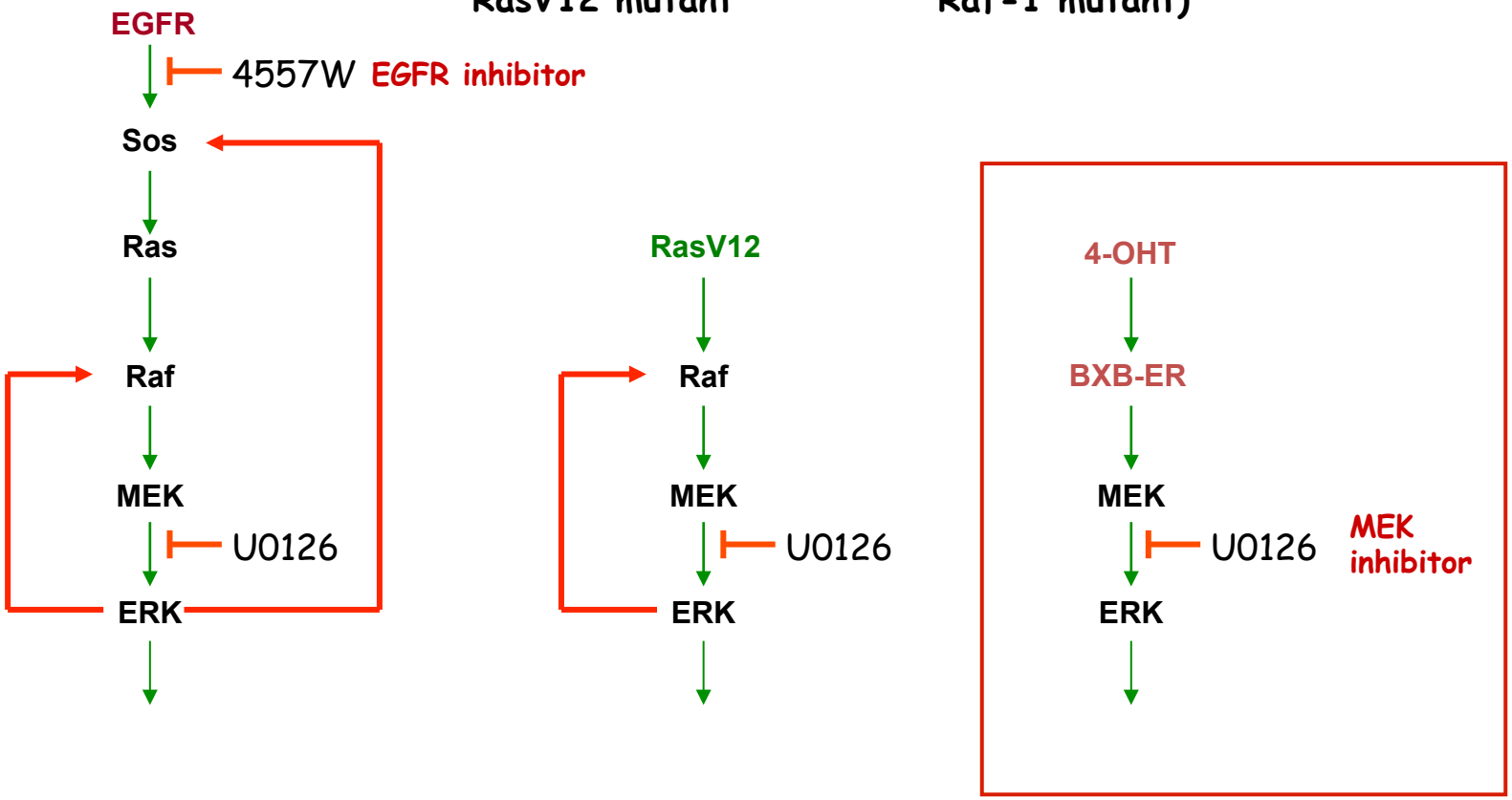
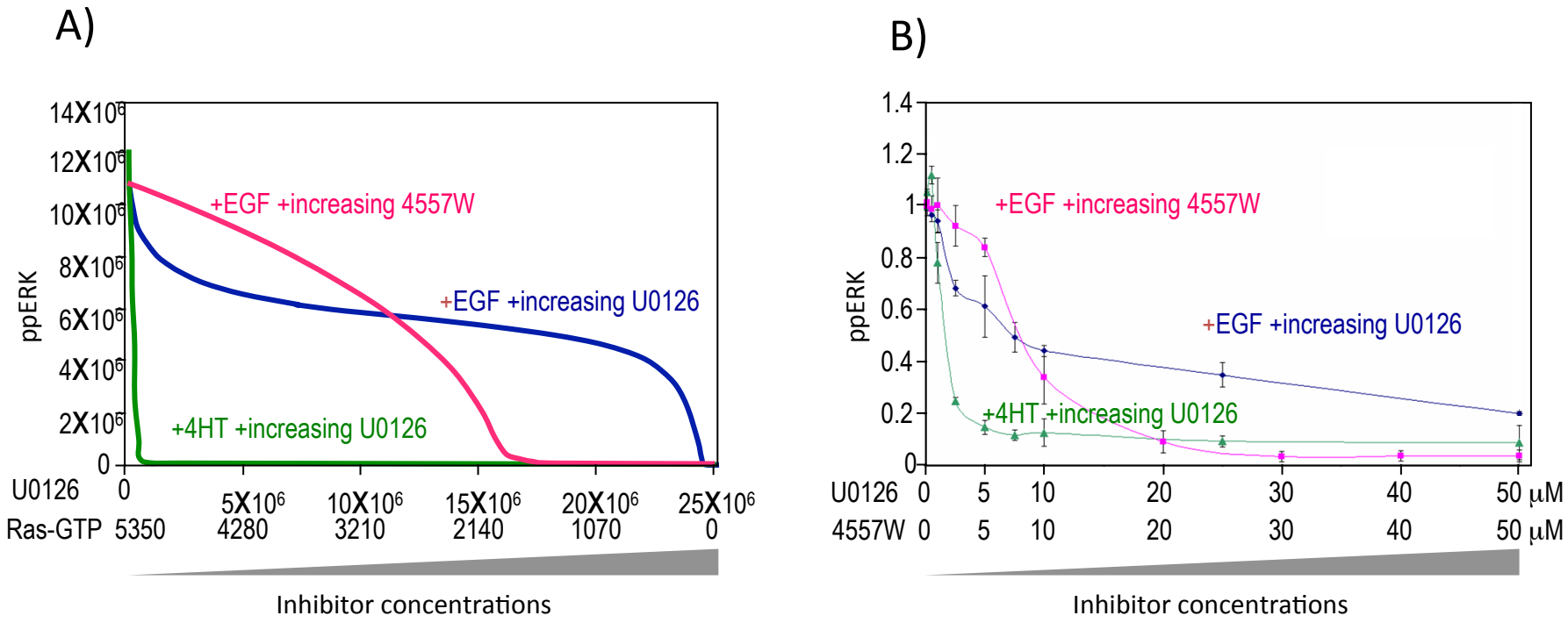


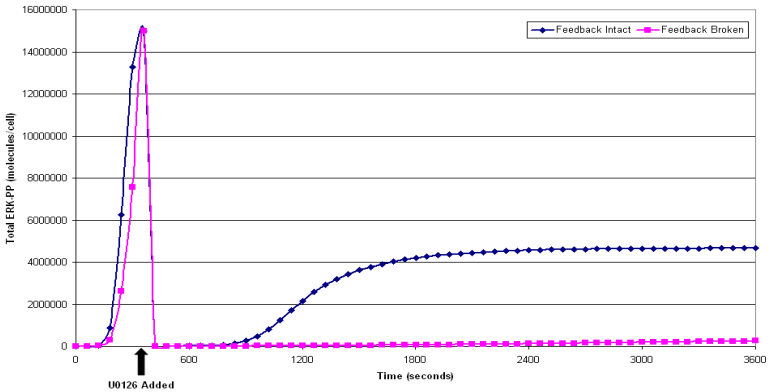
Figure 3

(A) Model prediction (B) Biochemical validation

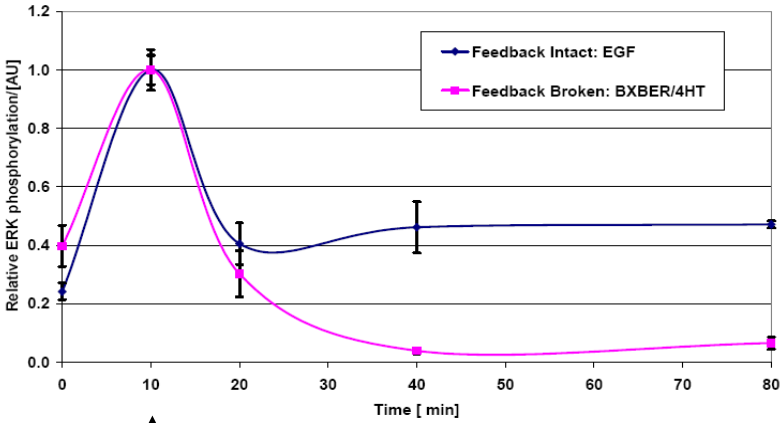


Signal recovery after MEK inhibition

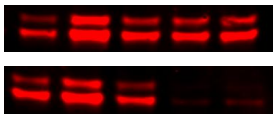
Simulation



Experiment



U0126 added



pERK1/2, +EGF

pERK1/2, + BXBER/4HT

0 10 20 40 80 min stimulation

The Mammalian MAPK/ERK Pathway Exhibits Properties of a Negative Feedback Amplifier

- Three-tiered kinase module, signal amplifier.
- Negative feedback loops - system like negative feedback amplifier
- Smoothens the output to changes in input - system robust to change.
- No feedback loops: cells sensitive to inhibition of MEK
- Feedback intact: cells are resistant to inhibition there. D

- **Drug development: inhibitors targetting components outside NFA are more effective at inhibiting the pathway.**

Sturm, Orton, Vyshemirsky, Grindlay, Birtwistle, Gilbert, Calder, Pitt, Kholodenko and Kolch., Science Signalling Dec 21;3

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.

SBML - XML Based Language

```
<sbml>
<model>
        <listOfCompartments> <compartment/> </listOfCompartments>
        <listOfSpecies> <specie/>          < /listOfSpecies>
        <listOfReactions>
                <reaction>
                        <listOfReactants>
                                <specieReference/>
                        </listOfReactants>
                        <listOfProducts>
                                <specieReference/>
                        </listOfProducts>
                        <kineticLaw>
                                <listOfParameters>
                                <parameter/>
                                </listOfParameters>
                        </kineticLaw>
                </reaction>
        </listOfReactions>
</model>
</sbml>
```

SBML Example

Specie 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>

Model Checking

In a sentence:

- “Formally check whether a model of a biochemical system does what we want”

Components:

- A model
  - the current description of a biochemical system of interest
- A property
  - a property which we think the system should have
- A model checker
  - a program to test whether the model has the property

Why model check in Systems Biology

- Biologists will often talk in qualitative or semi-quantitative language (trends).
  - "this protein peaks after 5 minutes, then falls to half concentration"
  - Often quite certain about time.

- Systems biology; Part of model design process, validate the model conforms to the **observed** data.

- Synthetic biology; Make sure the model and constructed bio system conform to the **desired** behaviour.

Analytical vs Simulative Model Checking

- Analytical:
  - Exact probabilities & prove properties
  - A model state is an association of #molecules to the each species
    - **Protein1** has 10 molecules & **Protein2** has 20 molecules
  - Analytical assesses every state that the model can be in (reachable states)
  - State space can grow even worse than exponentially with increasing molecules
  - Stochastic model checking with even as little as 12 molecules can be impossible with today's technology

- Solve this problem using…
- Simulative:
  - Instead of analysing the state space, analyses simulation outputs
  - Simulate the model X times and check these simulations
  - Simulation run = finite path through the state space
  - Can't prove probabilities

MC2 model checker

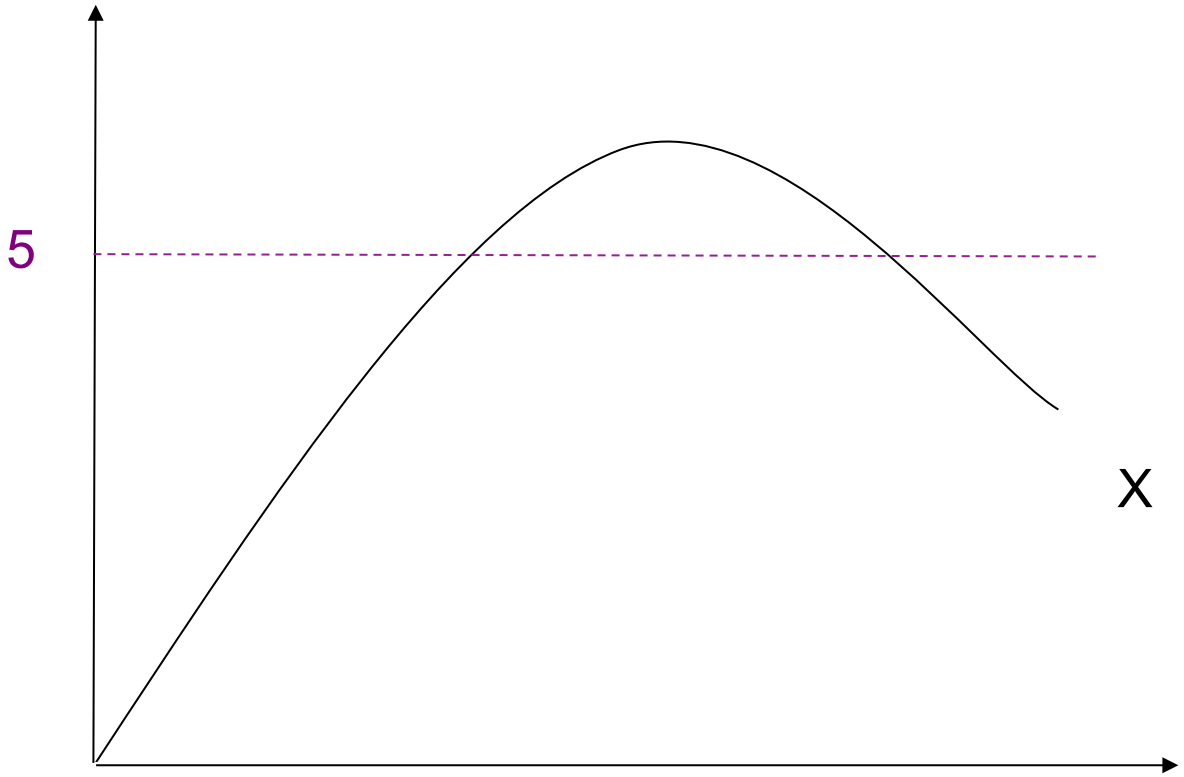
- Our approach
- Developed an ***offline*** Monte Carlo Model Checker for PLTLc properties, MC2(PLTLc) for short.
- Operates on a finite set of simulations – simulative approach
- Typically, many stochastic simulations (1,000 in this presentation)

  Approximate probability = fraction of simulations which satisfy the property over the #simulations
- Monte Carlo approximation – 2 approximations made:
  - finite number of simulations
  - Simulations of finite length

MC2 with ODE Output

$P_{=?}[\ F(\ X > 5\ )\ ]$

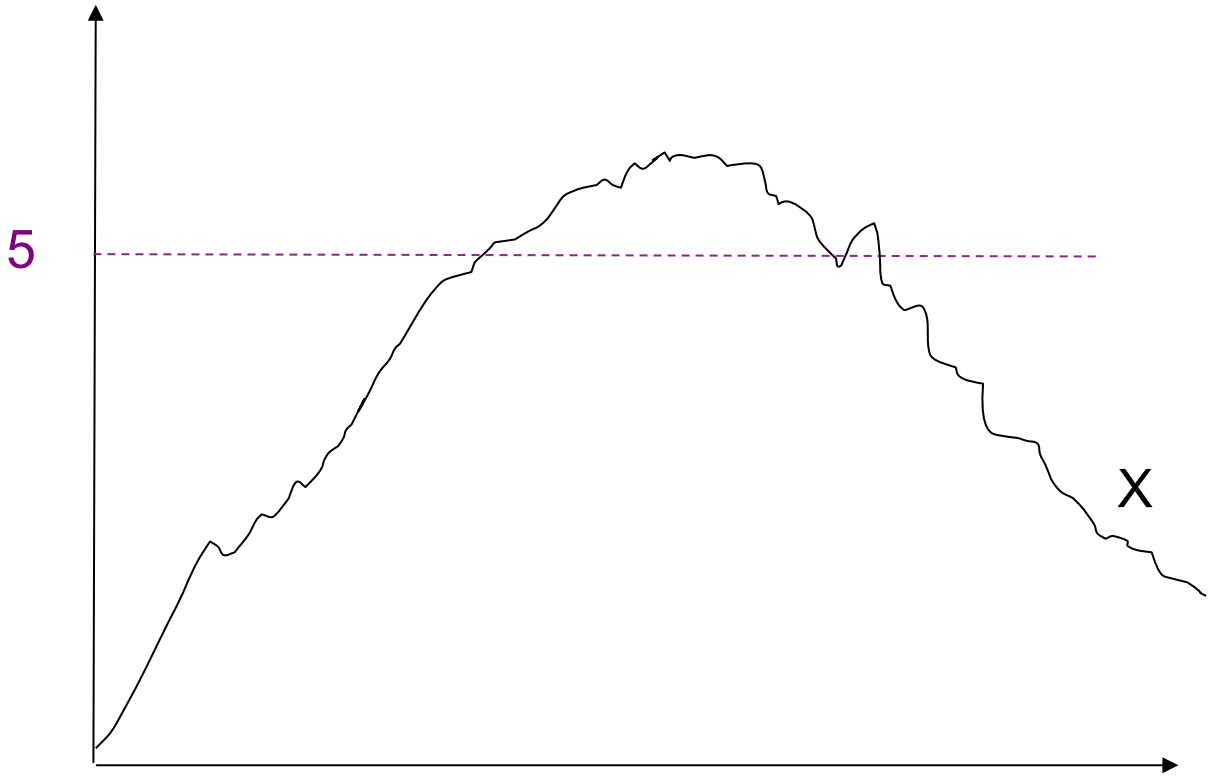
=> P = 1



MC2 with Gillespie Output

$P_{=?}[\ F(\ X > 5\ )\ ]$

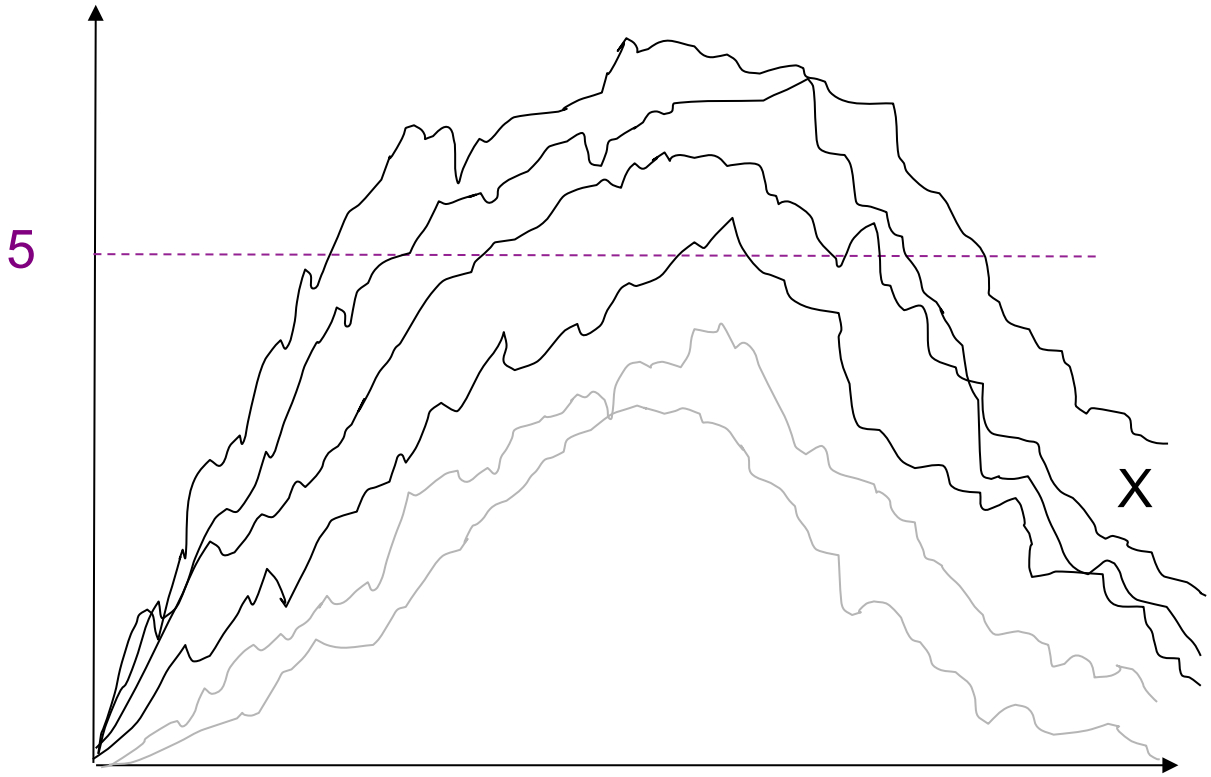
=> P = 1



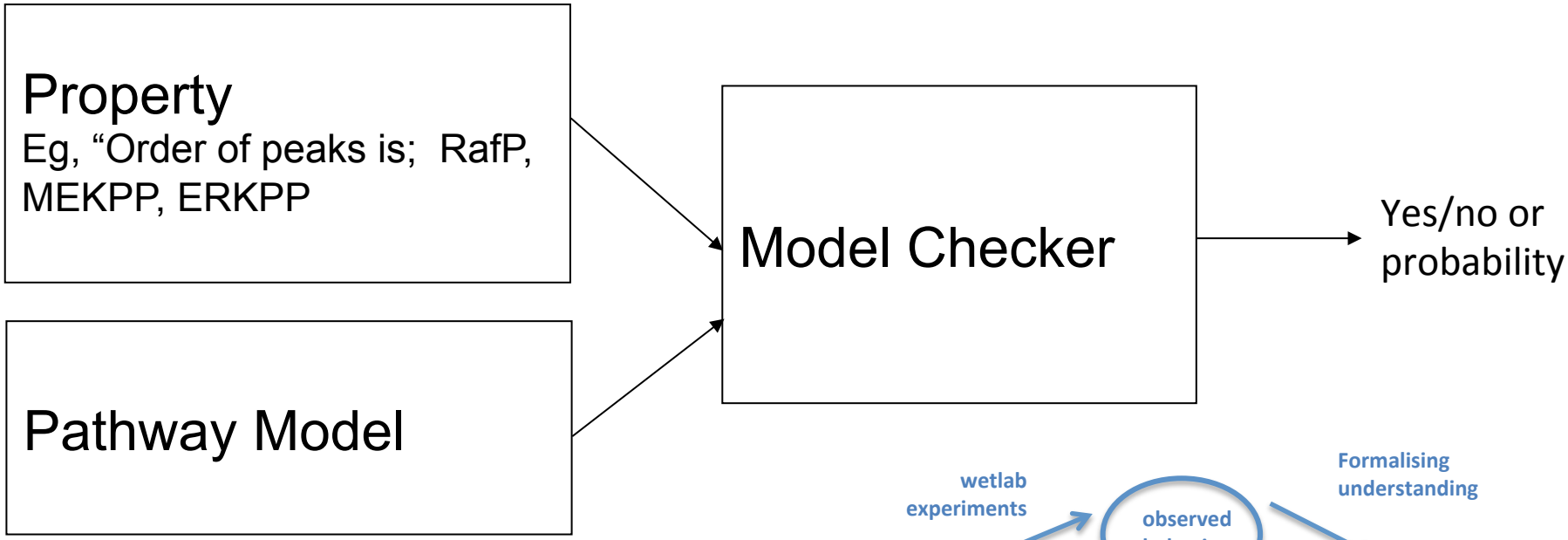
MC2 with Gillespie Output

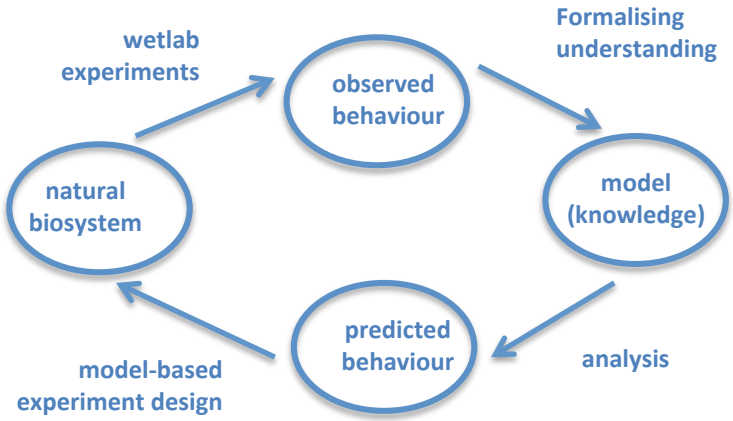
$$P_{=?}[\ F(\ X > 5\ )\ ]$$

=> P = 4/6

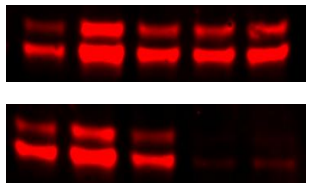


Model Checking Biochemical Pathways





Simulation-based Model Checking Biochemical Pathways



Property

Eg, “Order of peaks is RafP, MEKPP, ERKPP”

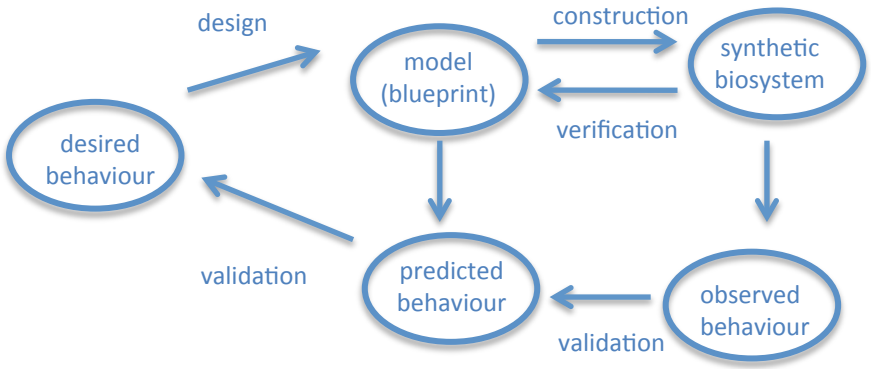
Time series data

Model

Lab

Behaviour Checker

Yes/no or probability



PLTL language

- Behaviours to be checked against a model is expressed in temporal logic

- We chose:
  Probabilistic logic called Probabilistic Linear-time Temporal Logic (PLTL)

- Main PLTL operators:

  G (P) – P always happens

  F (P) – P happens at some time

  X (P) – P happens in the next time point

  (P1) U (P2) – P1 happens until P2 happens

  P1 { P2 } – P1 happens from the first time P2 happens

Range of expressivity in PLTL

- **Qualitative**:
  Protein rises then falls
  P=? [(d(Protein) > 0) U (G(d(Protein) < 0))]

- **Semi-qualitative**:
  Protein rises then falls to less than 50% of peak concentration
  P=? [(d(Protein) > 0) U (G(d(Protein) < 0) ∧ F ([Protein] < 0.5 ∗ max[Protein]))]

- **Semi-quantitative**:
  Protein rises then falls to less than 50% of peak concentration by 60 minutes
  P=? [(d(Protein) > 0) U (G(d(Protein) < 0) ∧ F (time = 60 ∧ Protein < 0.5 ∗ max(Protein)))]

- **Quantitative**:
  Protein rises then falls to less than 100µMol by 60 minutes
  P=? [(d(Protein) > 0) U (G(d(Protein) < 0) ∧ F (time = 60 ∧ Protein < 100))]

What can we do with model checking in sys/syn bio?

- **Model validation**:
  - Show that your model of the pathway matches the lab data
  - Show that the (constructed) biosystem conforms to the specification
  - May not be obvious behaviours, so not easy to see by eye!
  - Might have a high probability of doing what you want, but doesn't always do it!

- **Model building**:
  - If the model doesn't do what we want, we can change the model (automatically?) until it does!
  - Change the parameters of a model (reaction rates/initial concentrations) until the pathway behaves as you want

- **Model finding**:
  - Many models in a database, can use PLTL as a query language like SQL.
  - "Give me all the models in the database which oscillate"

Model searching

Peaks at least once

(rises then falls below 50% max concentration)

$P_{>=1}[$ *ErkPP <= 0.50*max(ErkPP) ∧ d(ErkPP) > 0 U (ErkPP = max(ErkPP) ∧ F(ErkPP <= 0.50*max(ErkPP)))]*

- ***Brown***
- ***Kholodenko***
- ***Schoeberl***

Rises and remains constant

(99% max concentration)

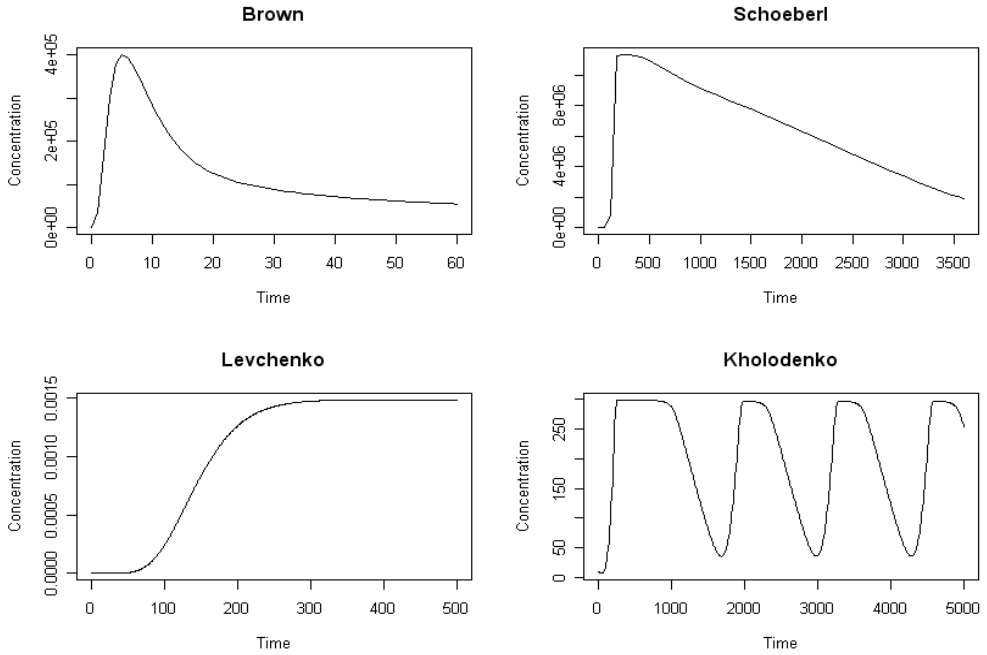
$P_{>=1}$*[ErkPP <= 0.50*max(ErkPP) ∧ (d(ErkPP) > 0) U (G(ErkPP >= 0.99*max(ErkPP)))]*

- ***Levchenko***

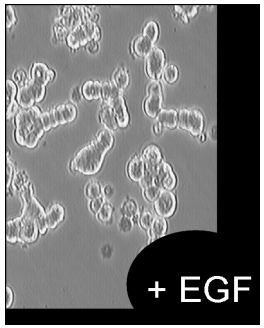
Oscillates at least 4 times

$P_{>=1}$*[F(d(ErkPP) > 0 ∧ F(d(ErkPP) < 0 ∧ …))]*

- **Kholodenko**



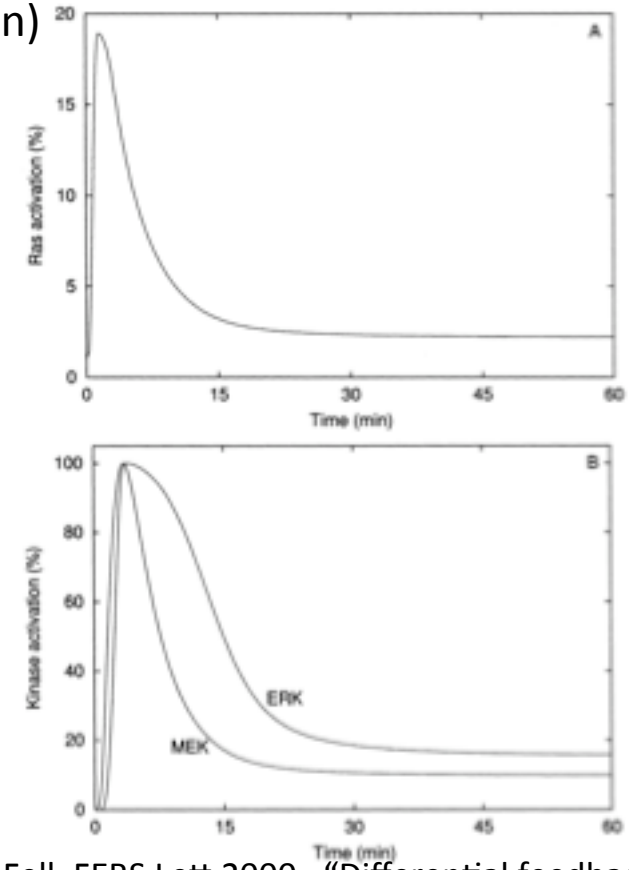
Parameter estimation
Response with EGF vs. NGF signal



Proliferation
(cell division)
PC12 cells

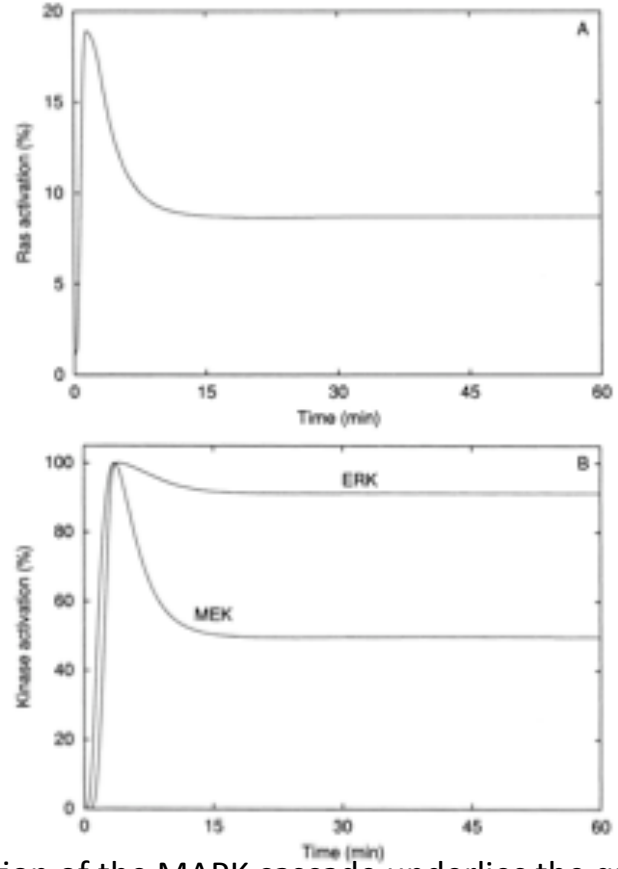
EGF

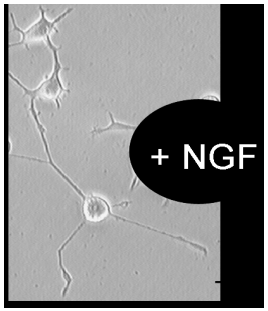
Transient activation of Ras, MEK and ERK



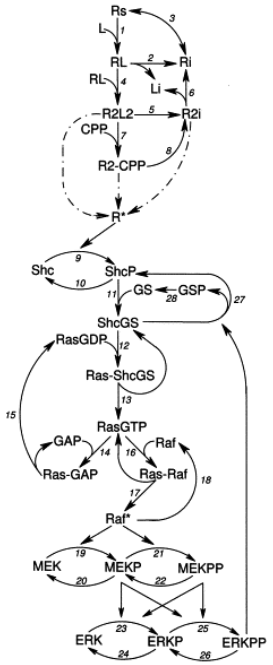
NGF

Sustained activation of Ras, MEK and ERK





Differentiation
(neurite
outgrowth)



Brightman & Fell, FEBS Lett 2000. "Differential feedback regulation of the MAPK cascade underlies the quantitative differences in EGF and NGF signalling in PC12 cells"

Desired Behaviour in PLTLc

The desired (sustained) NGF behaviour of the pathway was written in the original model paper.
Can be written in PLTLc as:

Sustained Ras: Active Ras peaks within 2 minutes to a maximum of 20% of total Ras and is stable between 5% and 10%

$P_{=?}$ [d(active Ras) > 0 U (time ≤ 2 ⋀ active Ras ≥ 0.15∗total Ras
⋀ active Ras ≤ 0.2∗total Ras ⋀ (d(active Ras) < 0)
U (G(active Ras ≥ 0.05∗total Ras ⋀ active Ras ≤ 0.10∗total Ras)))]

Sustained MEK: Active MEK peaks within 2 to 5 minutes and is stable between 40% and 50% of peak value

$P_{=?}$ [d(MEKPP) > 0 U (time ≥ 2 ⋀ time ≤ 5 ⋀ d(MEKPP) < 0
U (G(MEKPP ≥ 0.40∗max(MEKPP) ⋀ MEKPP ≤ 0.50∗max(MEKPP))))]

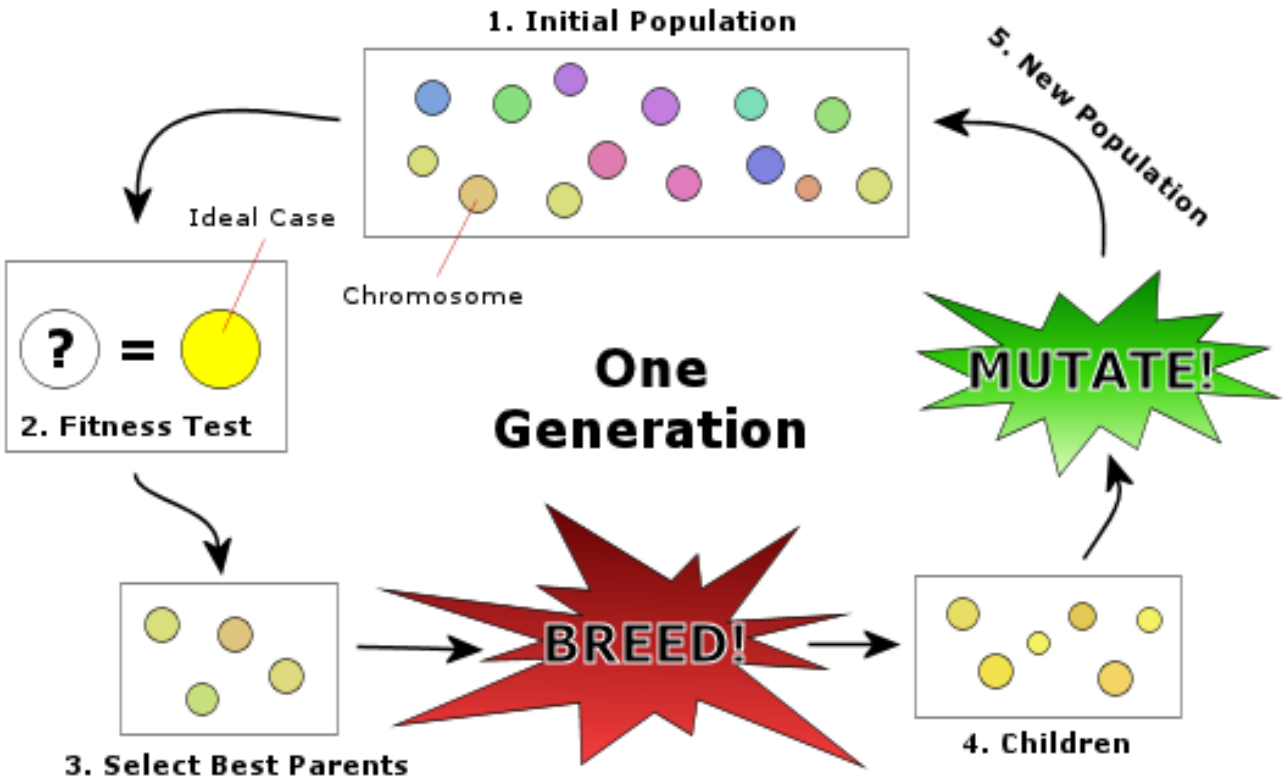
Sustained ERK: Active ERK peaks within 2 to 5 minutes and is stable between 85% and 100% of peak value

$P_{=?}$ [(d(ERKPP) > 0) U (time ≥ 2 ⋀ time ≤ 5 ⋀ d(ERKPP) < 0
U (G(ERKPP ≥ 0.85 ∗ max(ERKPP))))]

Robin Donaldson and David Gilbert (2008). A Model Checking Approach to the Parameter Estimation of Biochemical Pathways In proceedings CMSB 2008 (Computational Methods in Systems Biology). To Appear.

Model construction using a genetic algorithm

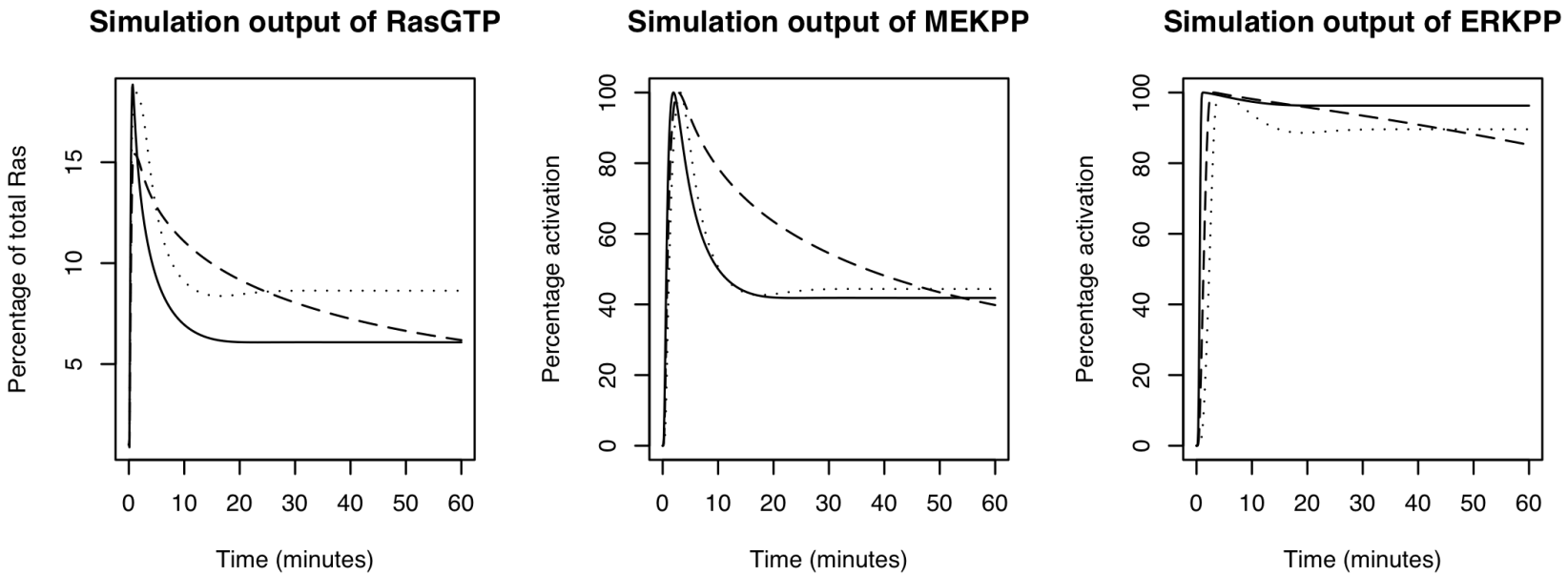
Sustained ERK: Active ERK peaks within 2 to 5 minutes and is stable between 85% and 100% of peak value



2000 models, 100 generations: 200,000 simulations/checks

Parameter fitting results

- Built a fitness function for sustained Ras, MEK and ERK
- Ran the genetic algorithm with 100 generations and obtained results:



- Original model of the NGF signalling pathway varying V28 (dotted)
- Best model returned when varying the critical parameters (solid)
- Critical parameters without V28 (dashed).

The best model returned when varying the critical parameters only required a ***16-fold*** increase in V28 (compared with 40-fold in original paper)

Even possible to get similar behaviour **without** varying V28

Biochemical Models
Construction Based on Reuse of Components

Zujian Wu

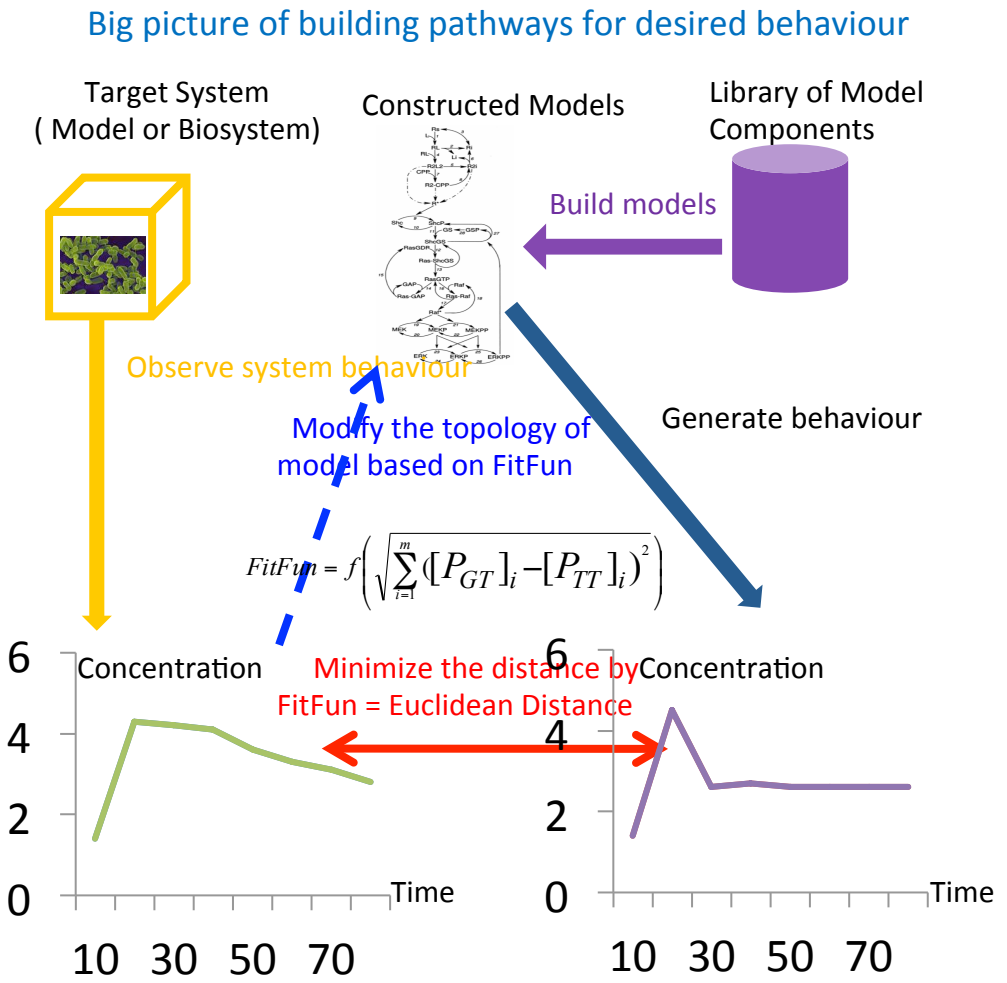
Aims:

- To achieve the target-driven construction of biochemical models by reference to their desired behaviours

- To address the construction problem by
  - building a library for storing reliable biochemical functional submodels (as components)
  - intelligently selecting, combining and mutating these submodels in order to generate complex

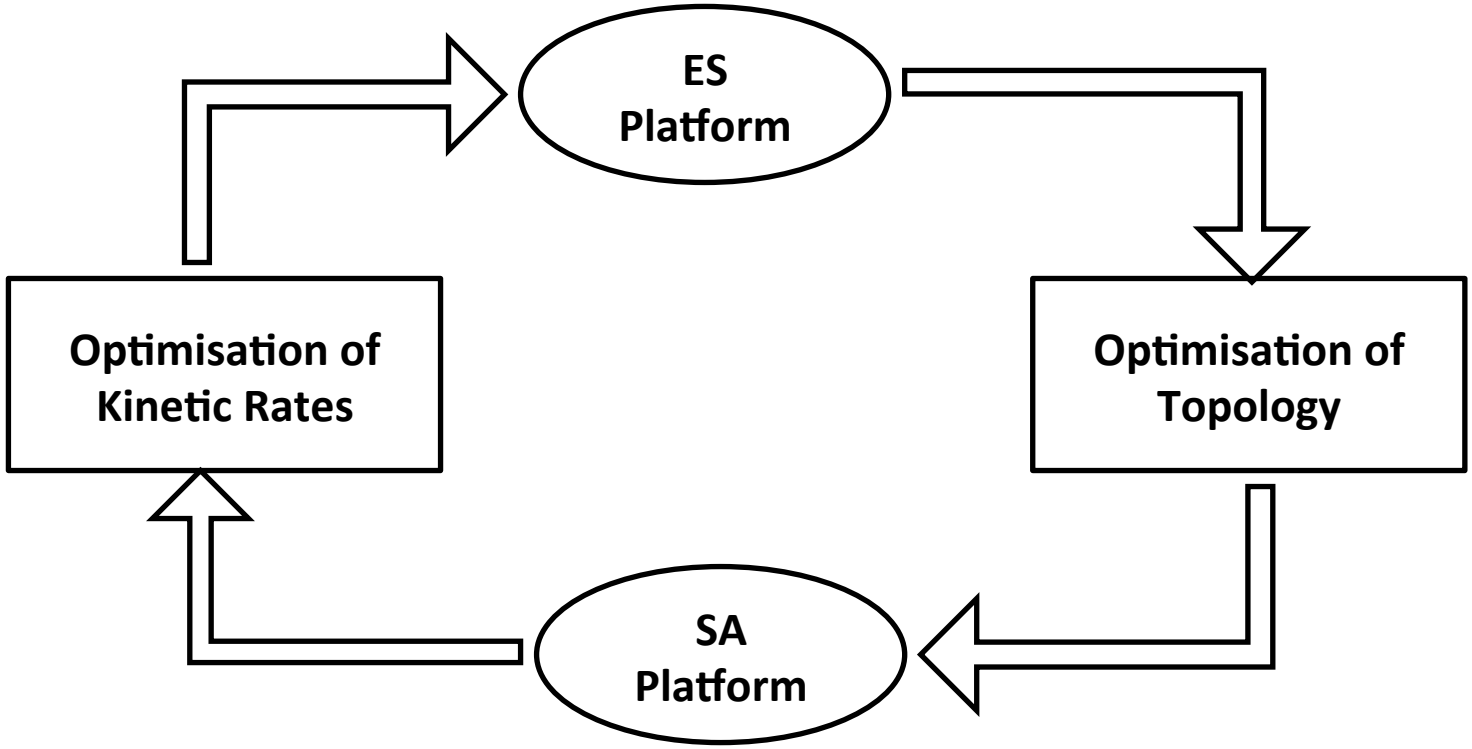
Methods:

- Petri Nets
- Simulated Annealing
- Evolutionary Algorithms

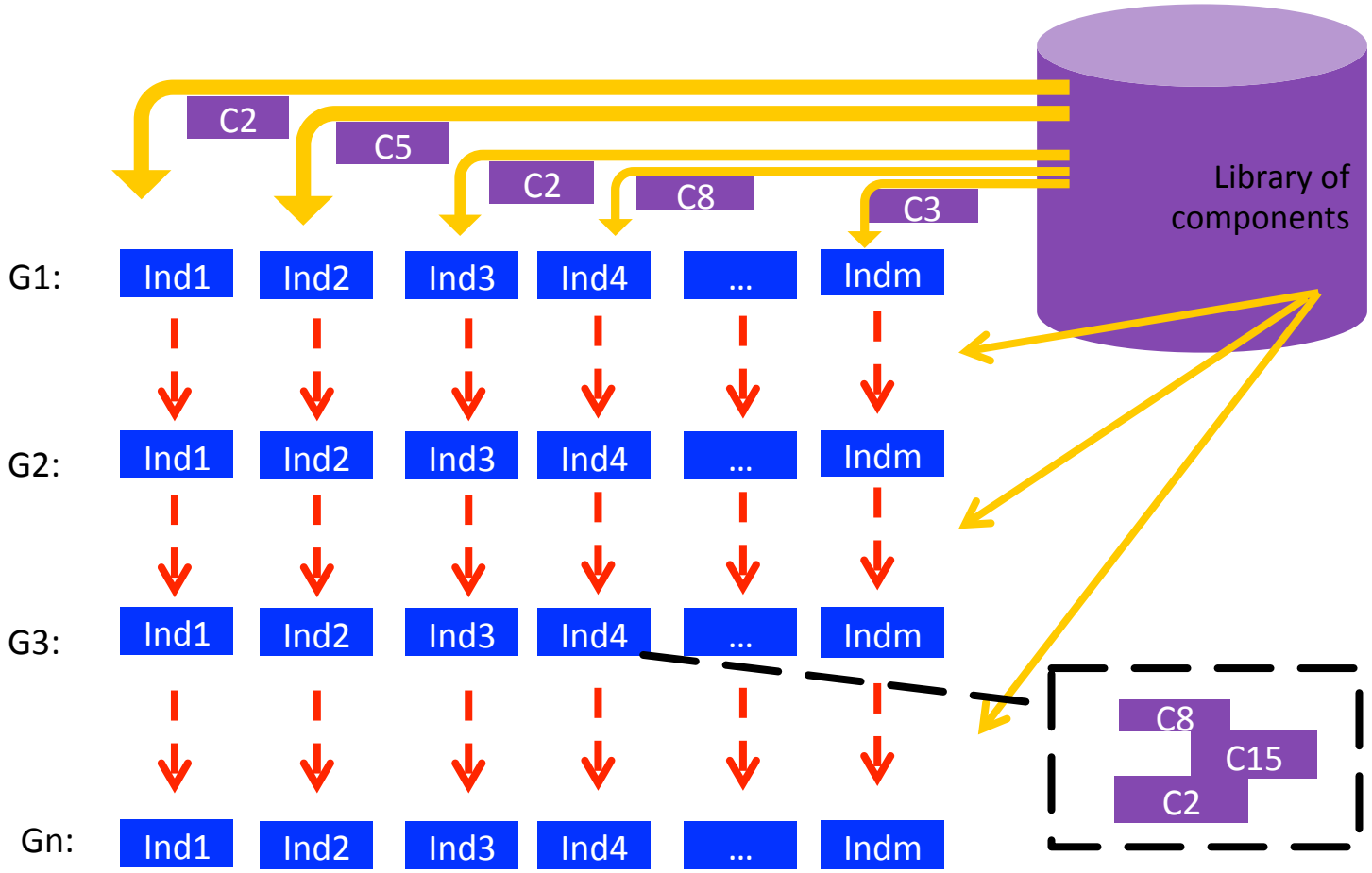


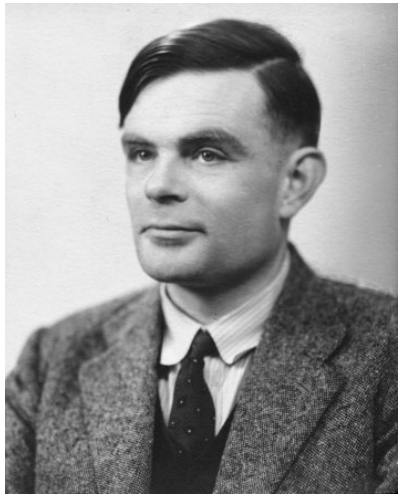
Current Research

Hybrid Optimisation of Topology and Kinetic Rates based on Evolution Strategy(ES) and Simulated Annealing(SA)

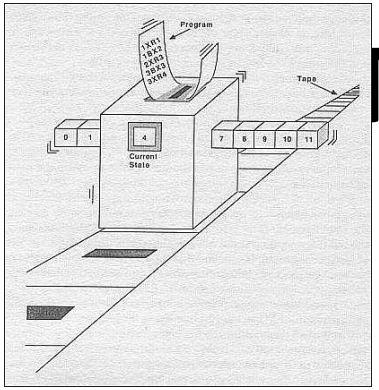


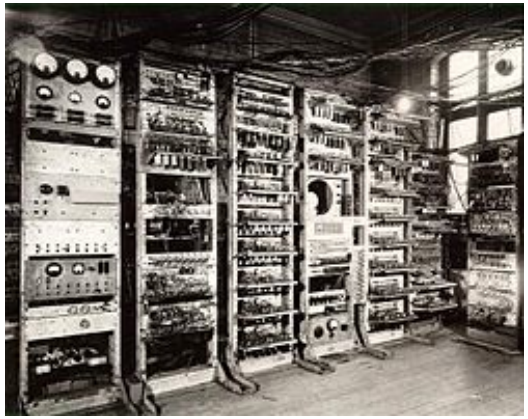
Evolution Strategy – Topology Optimisation

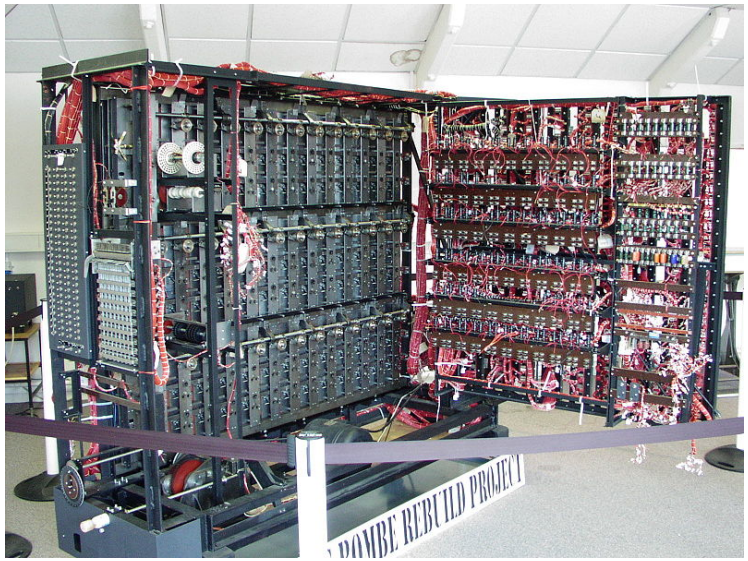


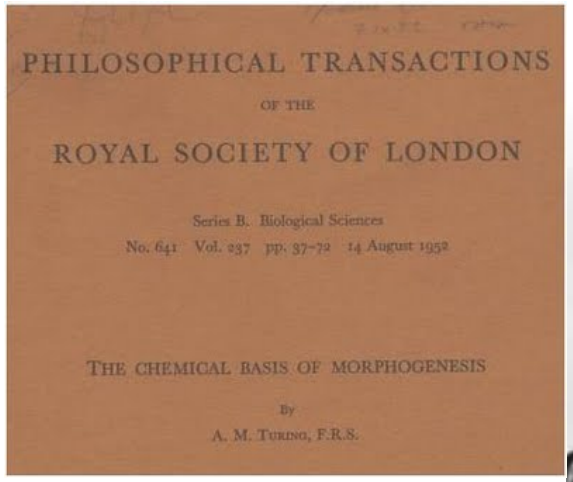


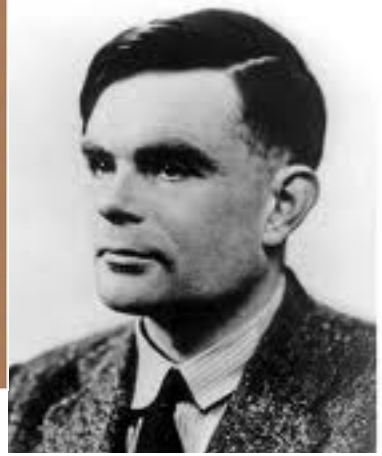
Turing



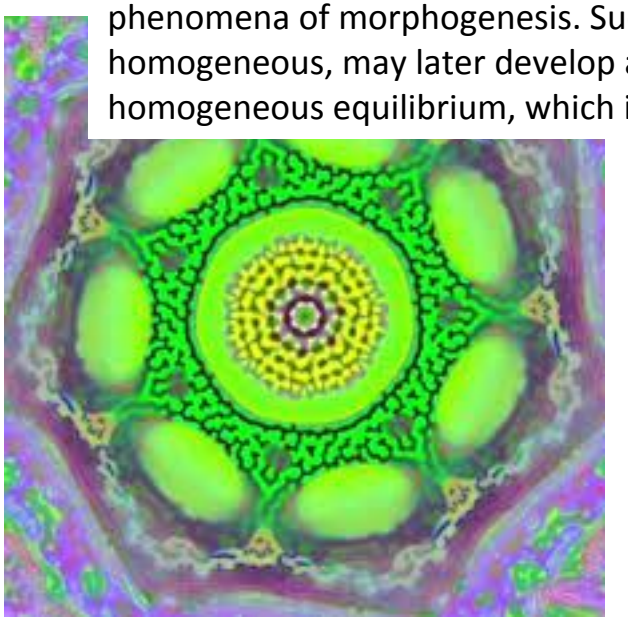






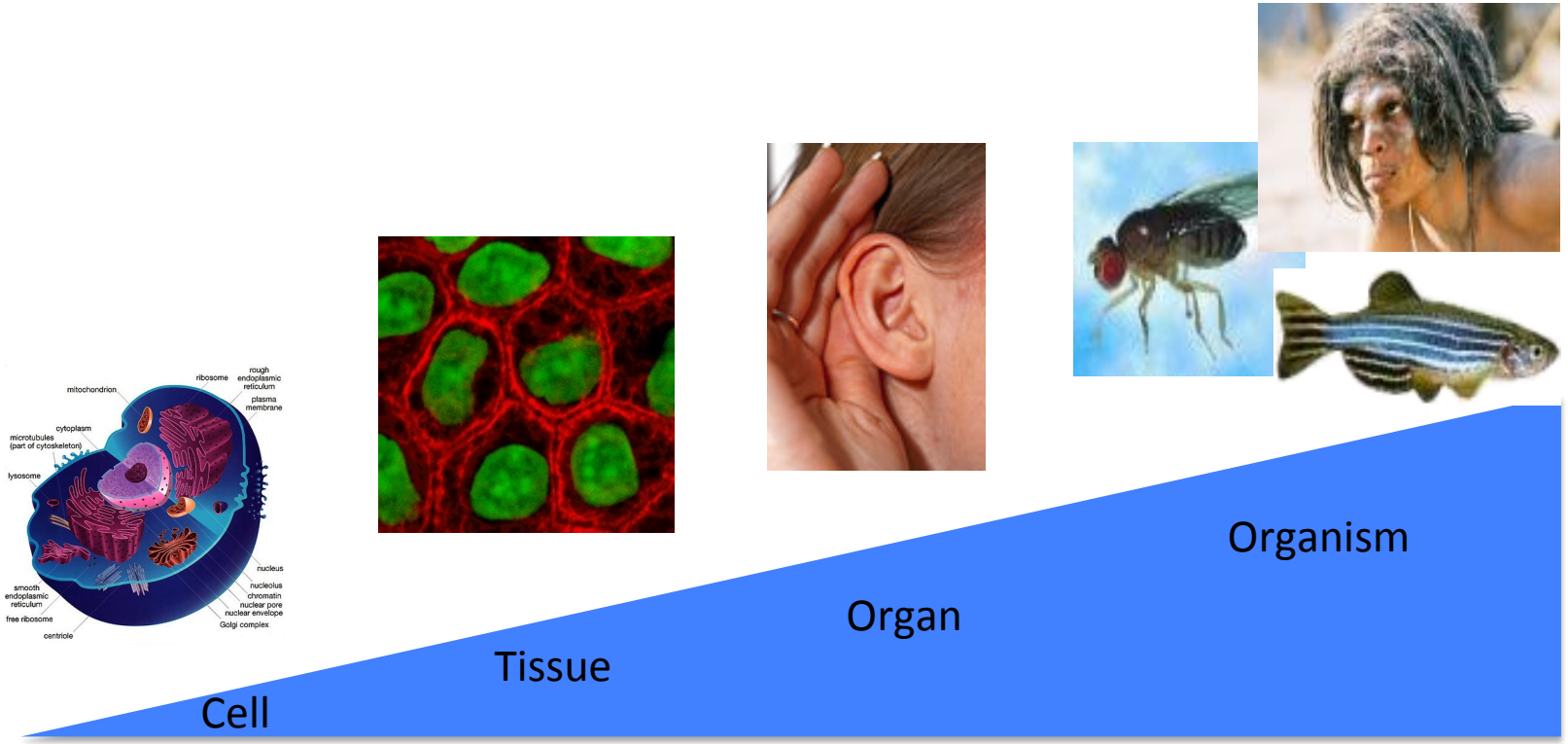


It is suggested that a system of chemical substances, called morphogens, reacting together and diffusing through a tissue, is adequate to account for the main phenomena of morphogenesis. Such a system, although it may originally be quite homogeneous, may later develop a pattern or structure due to an instability of the homogeneous equilibrium, which is triggered off by random disturbances..



Multiscale Modelling

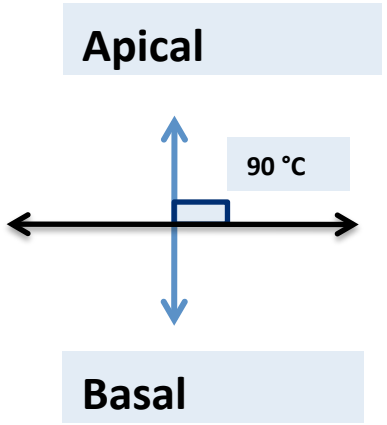
Solving physical problems which have important features at multiple scales, particularly multiple spatial and/or temporal

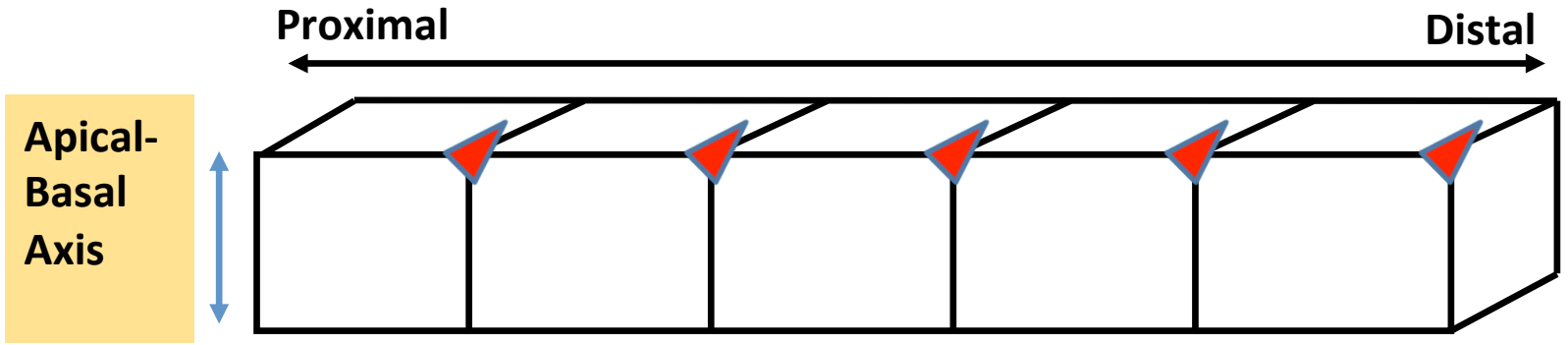


Planar Cell Polarity

- **Mechanism through which a number of tissue types determine the polarity of their cells perpendicular to their cellular apical-basal axis.**

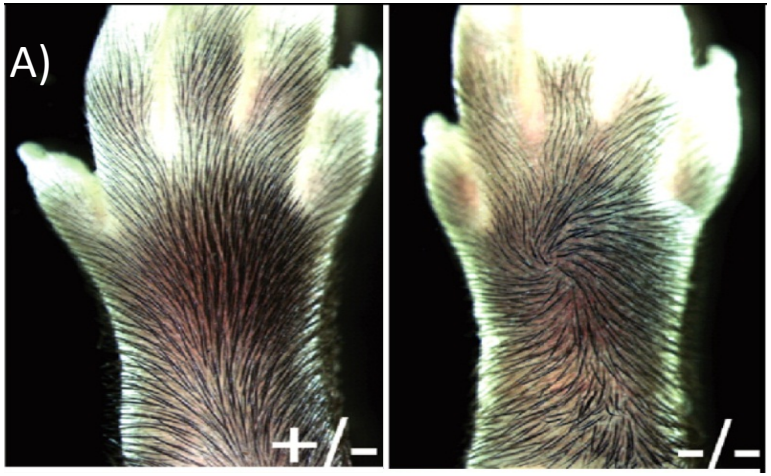
  In mammals: ear (sensory hair cell), eye (Equipotent R3/R4 cells) epithelia



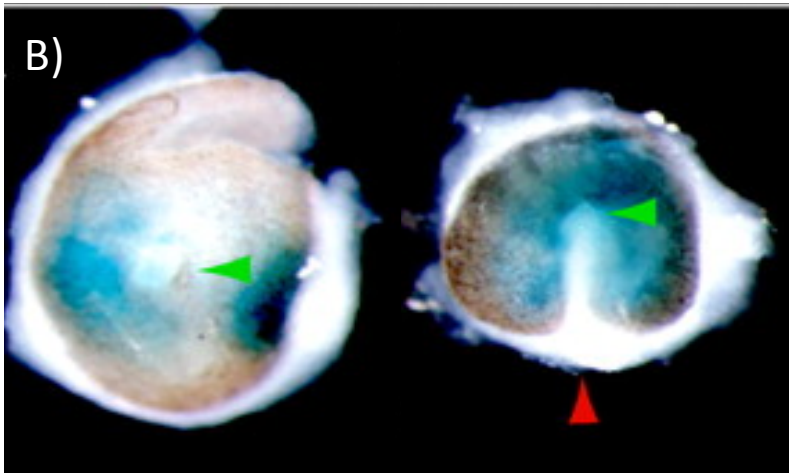


Errors in PCP

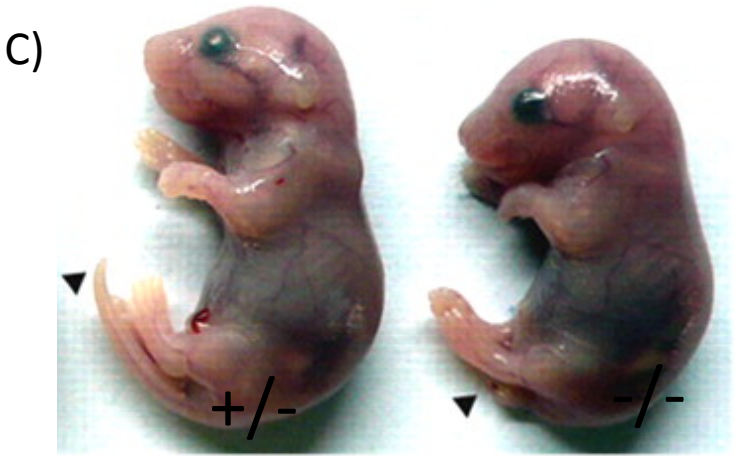
Fz6 Knockout: Hair Polarity



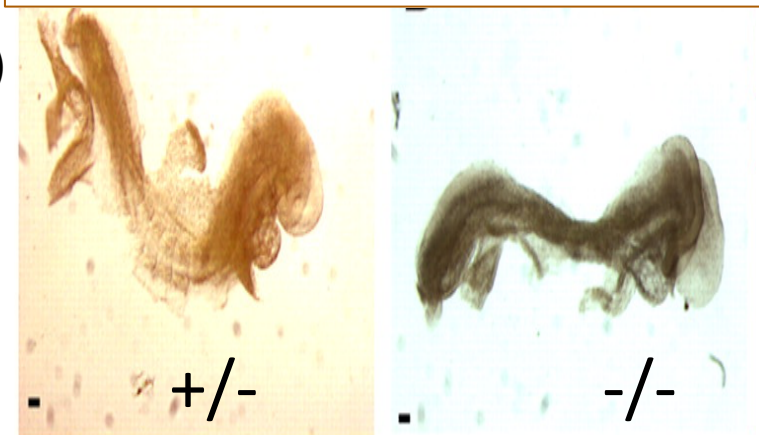
Fz5 Knockout: Coloboma



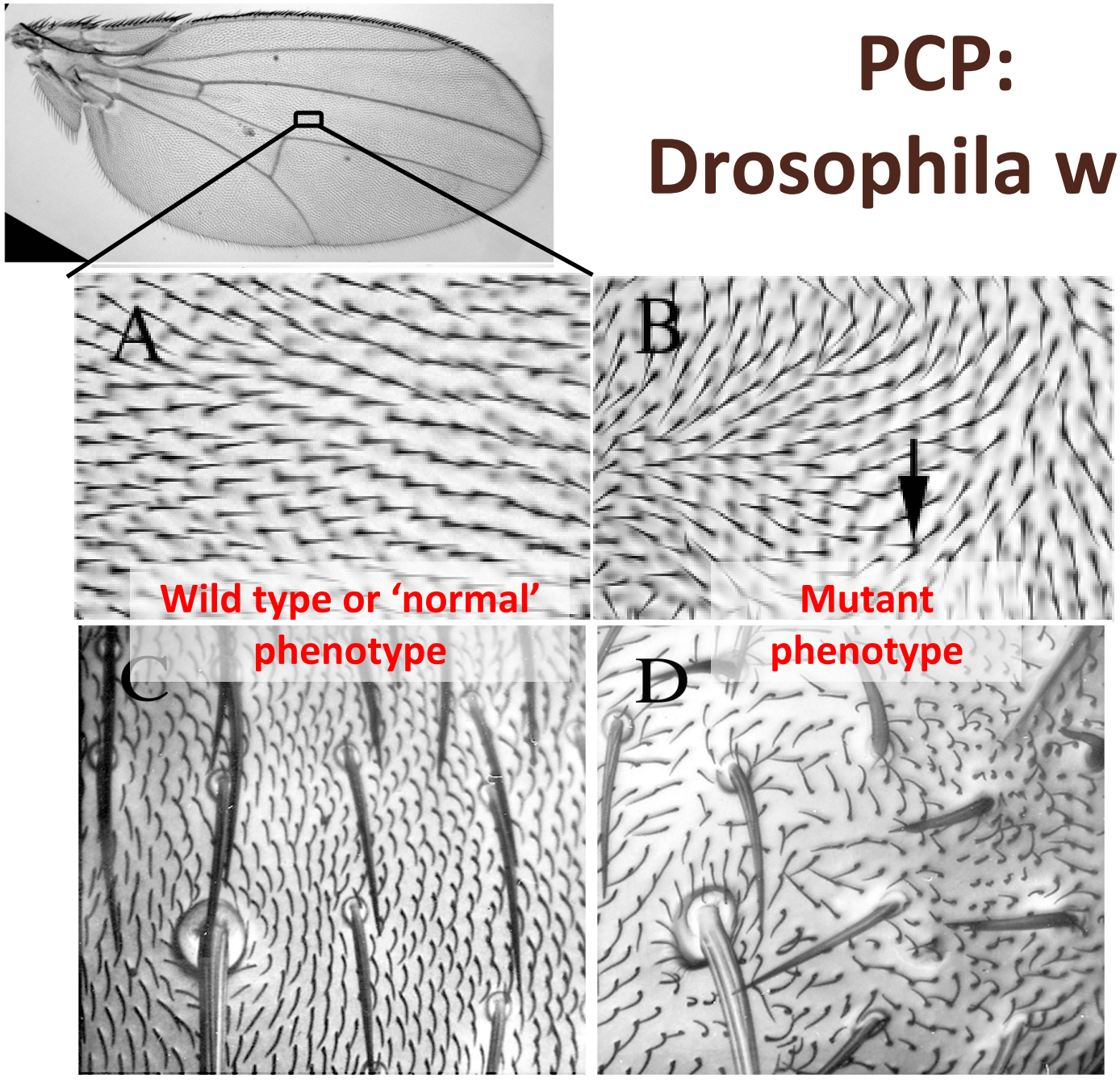
Dact1 Knockout: Urogenital Defects



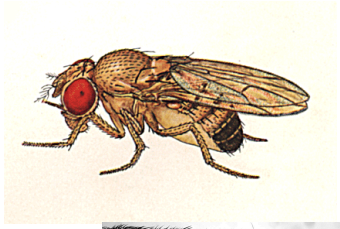
Vangl2 Knockout: Impairs Cranial Neurulation

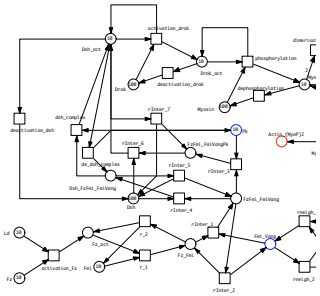


PCP:
Drosophila wing



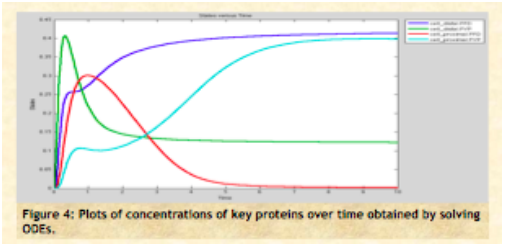
Multiscale from signalling to organs



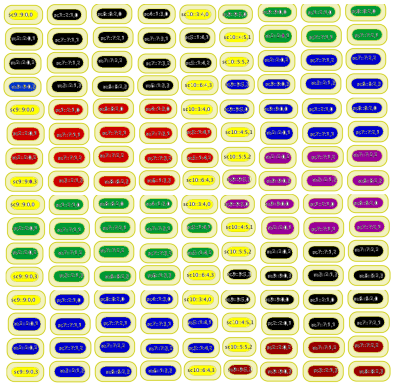


Petri nets (*coloured, hierarchical*)

Monika Heiner



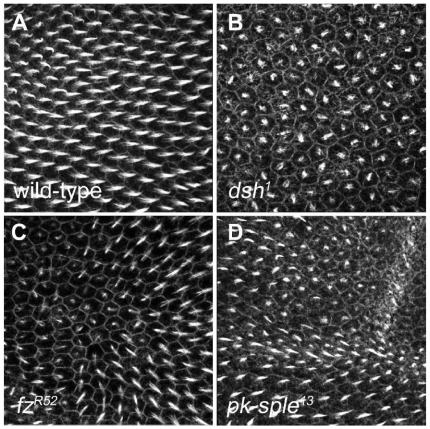
ODEs, stochastics

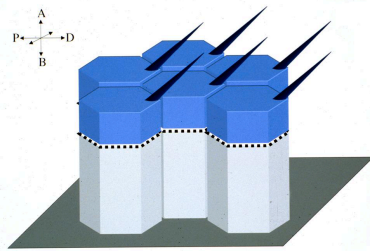


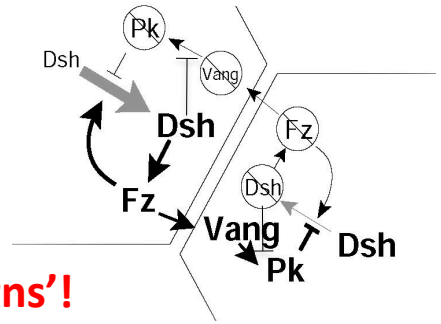
P-systems

(InfoBiotics - Nottingham)

Planar Cell Polarity





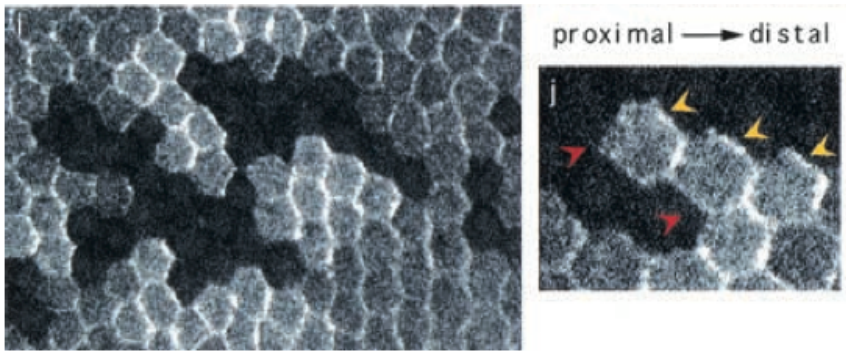


Pam Gao, David Tree

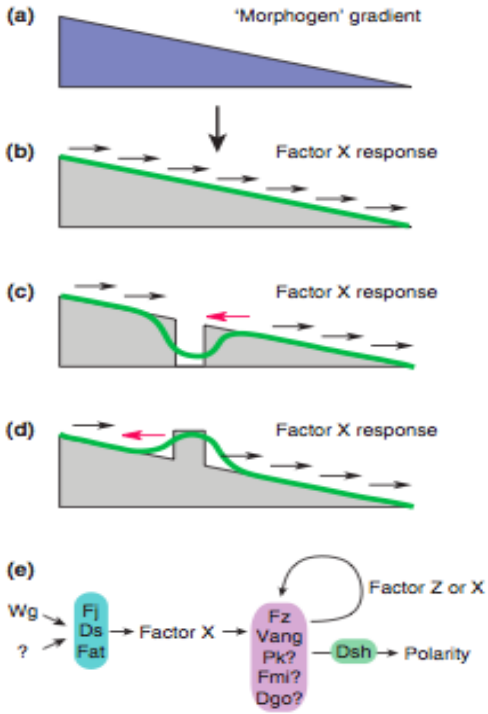
→ **Design & genetically engineer 'patterns'!**

Hypothesis in PCP

- Morphogen (Factor X) gradient
  - Between cells
  - Inside cells



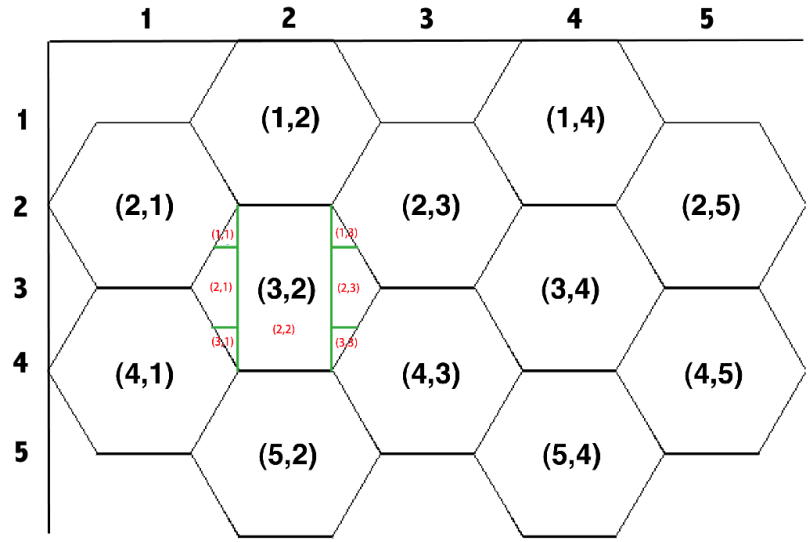
• J.D. Axelrod, C.J. Tomlin (2011). Modeling the control of planar cell polarity. WIREs Systems Biology and Medicine, p865

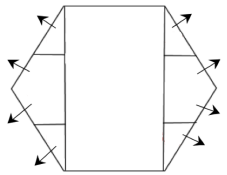


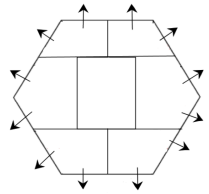
- Microtubules (biased transport of proteins)

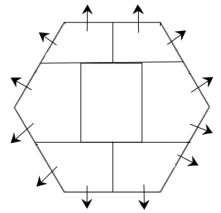
Hierarchical Organisation

- Hierarchically coloured



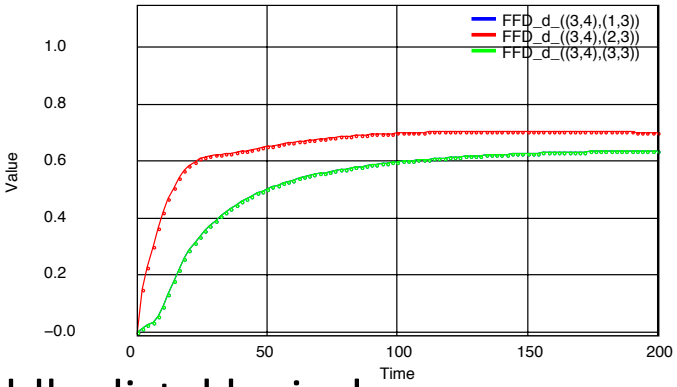






Colourset = {…, {((3,2)(1,1)), ((3,2)(1,2)), ((3,2)(1,3)),……((3,2)(3,3))}, …

Clustering & model checking



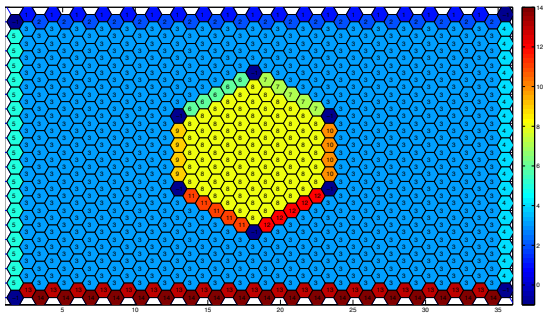
For each cell (x,y) in the honeycomb:

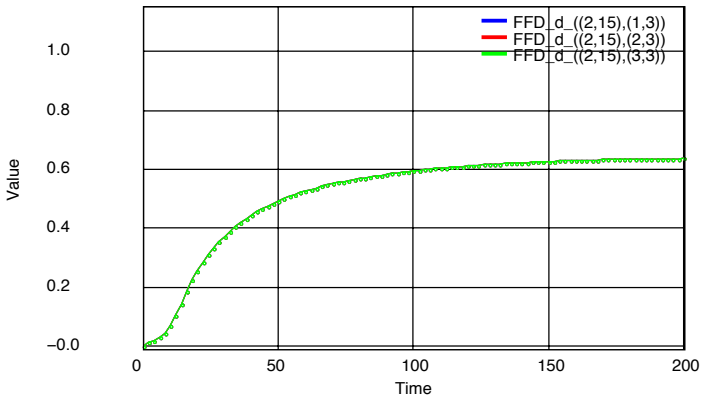
After some initialisation phase, FFD in the middle distal logical compartment (2,3) is always greater than in the other distal compartments (1,3) and (3,3), and will remain so:

P =? [G(time > init → ([(2,3)] > [(1,3)] & [(2,3)] > [(3,3)]))]

15*15 honeycomb grid: 112 cells in total

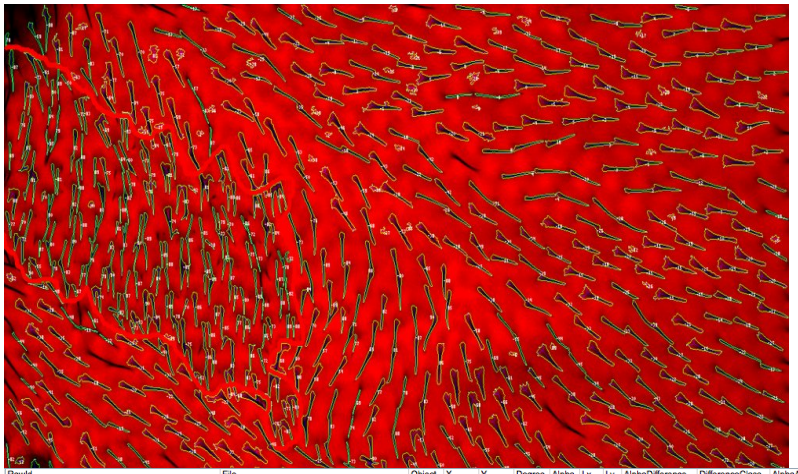
Query holds for all these cells except the cells in the last column, cells (2,15) to (14,15).





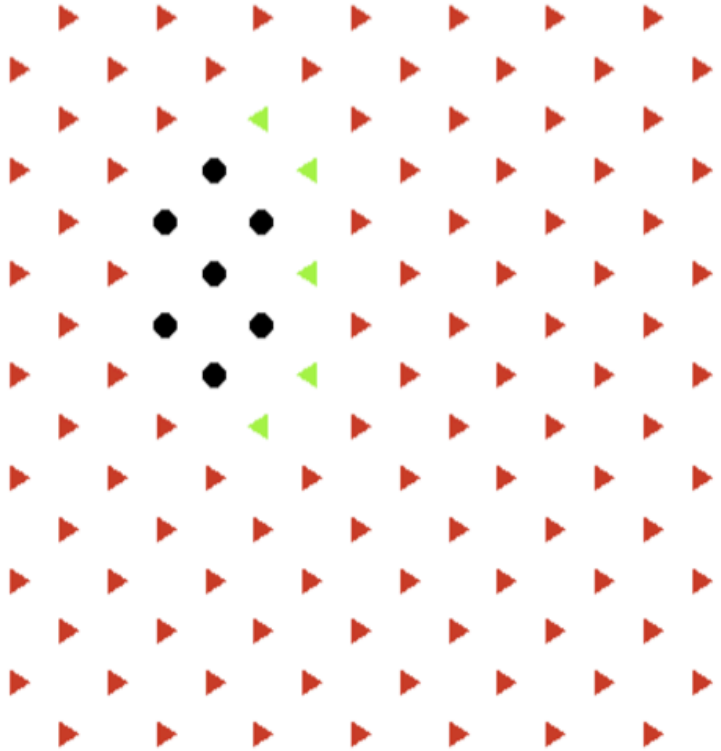
- Q. Gao, F. Liu, D. Gilbert, M. Heiner & D. Tree. **CMSB 2011**, Paris, France.

Mutated tissue
Experimental vs In-silico

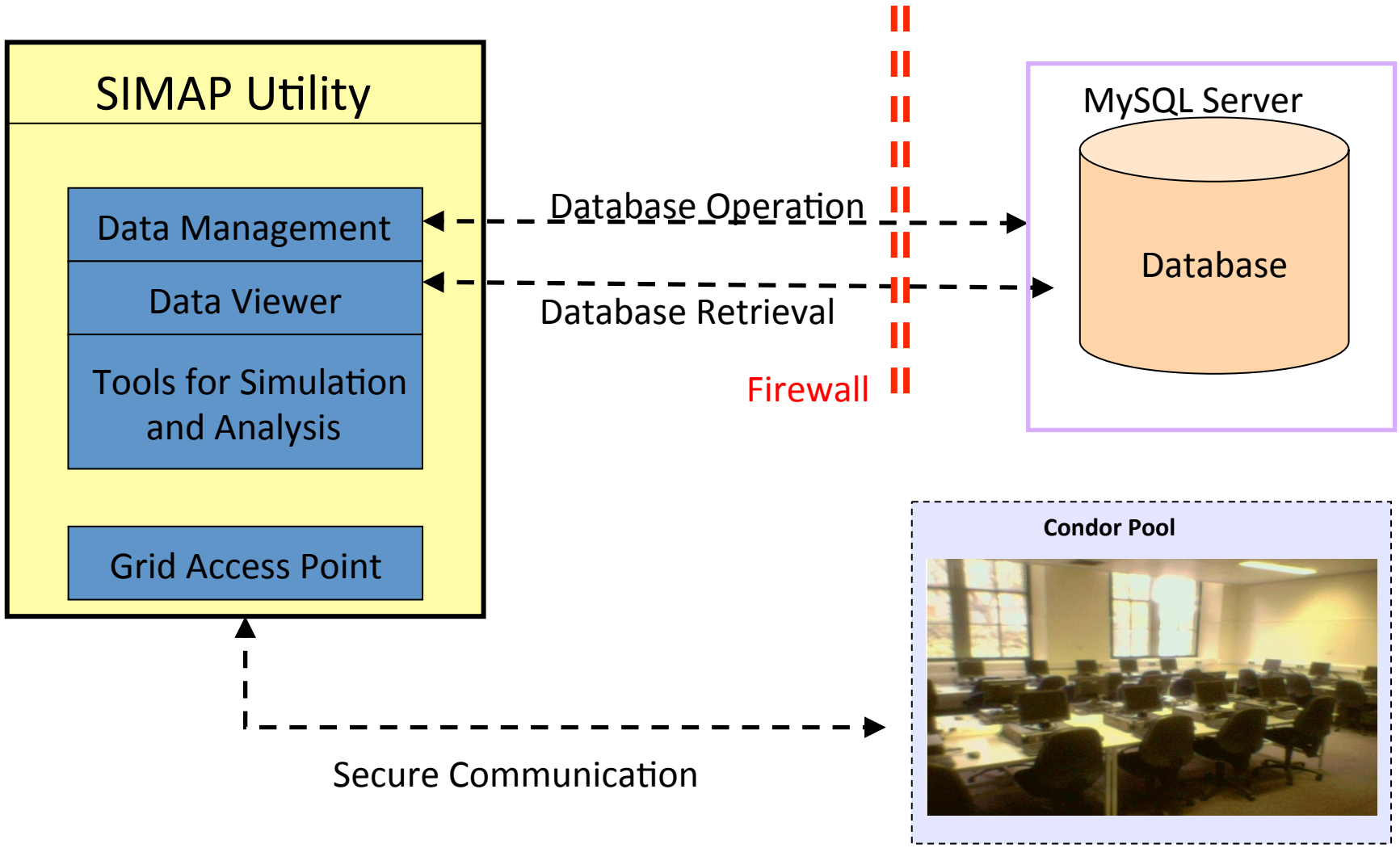


RowId	File	Object	X	Y	Degree	Alpha	Lx	Ly	AlphaDifference	DifferenceClass	AlphaAverage
dsh1-1001005_012.TIF_1	dsh1-1001005_012.TIF	1	23	1104	105	3.40	15	1	0.27	5	3.12
dsh1-1001005_012.TIF_2	dsh1-1001005_012.TIF	2	23	1469	28	2.06	15	1	1.06	2	3.12
dsh1-1001005_012.TIF_3	dsh1-1001005_012.TIF	3	24	822	113	3.55	15	1	0.43	5	3.12
dsh1-1001005_012.TIF_4	dsh1-1001005_012.TIF	4	25	537	158	4.33	15	1	1.21	2	3.12
dsh1-1001005_012.TIF_5	dsh1-1001005_012.TIF	5	25	1602	46	2.37	15	1	0.75	3	3.12
dsh1-1001005_012.TIF_6	dsh1-1001005_012.TIF	6	26	1479	29	2.08	15	1	1.04	2	3.12
dsh1-1001005_012.TIF_7	dsh1-1001005_012.TIF	7	28	1489	20	1.92	15	1	1.21	2	3.12
dsh1-1001005_012.TIF_8	dsh1-1001005_012.TIF	8	28	1545	177	4.65	15	1	1.53	0	3.12
dsh1-1001005_012.TIF_9	dsh1-1001005_012.TIF	9	30	1517	67	2.73	15	1	0.39	5	3.12
dsh1-1001005_012.TIF_10	dsh1-1001005_012.TIF	10	31	1635	11	1.76	15	1	1.36	1	3.12
dsh1-1001005_012.TIF_11	dsh1-1001005_012.TIF	11	32	1619	38	2.19	15	1	0.93	3	3.12
dsh1-1001005_012.TIF_12	dsh1-1001005_012.TIF	12	33	1426	59	2.60	15	1	0.52	4	3.12
dsh1-1001005_012.TIF_13	dsh1-1001005_012.TIF	13	35	1389	40	2.27	15	1	0.85	3	3.12
dsh1-1001005_012.TIF_14	dsh1-1001005_012.TIF	14	35	1563	82	3.01	15	1	0.12	6	3.12
dsh1-1001005_012.TIF_15	dsh1-1001005_012.TIF	15	35	1777	15	1.84	15	1	1.28	1	3.12
dsh1-1001005_012.TIF_16	dsh1-1001005_012.TIF	16	37	835	149	4.17	15	1	1.05	2	3.12
dsh1-1001005_012.TIF_17	dsh1-1001005_012.TIF	17	37	1446	68	2.75	15	1	0.37	5	3.12
dsh1-1001005_012.TIF_18	dsh1-1001005_012.TIF	18	37	1535	15	1.84	15	1	1.28	1	3.12
dsh1-1001005_012.TIF_19	dsh1-1001005_012.TIF	19	39	805	153	4.25	15	1	1.12	2	3.12
dsh1-1001005_012.TIF_20	dsh1-1001005_012.TIF	20	41	1464	50	2.45	15	1	0.68	4	3.12
dsh1-1001005_012.TIF_21	dsh1-1001005_012.TIF	21	44	992	123	3.72	15	1	0.60	4	3.12
dsh1-1001005_012.TIF_22	dsh1-1001005_012.TIF	22	44	1554	16	1.86	15	1	1.27	1	3.12
dsh1-1001005_012.TIF_23	dsh1-1001005_012.TIF	23	44	1608	19	1.90	15	1	1.23	1	3.12
dsh1-1001005_012.TIF_24	dsh1-1001005_012.TIF	24	46	1376	50	2.45	15	1	0.68	4	3.12
dsh1-1001005_012.TIF_25	dsh1-1001005_012.TIF	25	46	1484	53	2.50	15	1	0.62	4	3.12
dsh1-1001005_012.TIF_26	dsh1-1001005_012.TIF	26	47	1087	87	3.08	15	1	0.04	6	3.12
dsh1-1001005_012.TIF_27	dsh1-1001005_012.TIF	27	47	1626	178	4.67	15	1	1.55	0	3.12
dsh1-1001005_012.TIF_28	dsh1-1001005_012.TIF	28	49	846	111	3.51	15	1	0.39	5	3.12
dsh1-1001005_012.TIF_29	dsh1-1001005_012.TIF	29	49	1524	18	1.88	15	1	1.24	1	3.12
dsh1-1001005_012.TIF_30	dsh1-1001005_012.TIF	30	49	1645	40	2.27	15	1	0.86	3	3.12
dsh1-1001005_012.TIF_31	dsh1-1001005_012.TIF	31	50	1502	36	2.20	15	1	0.93	3	3.12
dsh1-1001005_012.TIF_32	dsh1-1001005_012.TIF	32	51	818	144	4.09	15	1	0.97	2	3.12
dsh1-1001005_012.TIF_33	dsh1-1001005_012.TIF	33	51	1164	26	2.02	15	1	1.11	2	3.12
dsh1-1001005_012.TIF_34	dsh1-1001005_012.TIF	34	51	1787	16	1.86	15	1	1.27	1	3.12
dsh1-1001005_012.TIF_35	dsh1-1001005_012.TIF	35	53	1009	59	2.60	15	1	0.52	4	3.12
dsh1-1001005_012.TIF_36	dsh1-1001005_012.TIF	36	53	1145	51	2.47	15	1	0.66	4	3.12
dsh1-1001005_012.TIF_37	dsh1-1001005_012.TIF	37	53	1431	75	2.89	15	1	0.24	6	3.12
dsh1-1001005_012.TIF_38	dsh1-1001005_012.TIF	38	53	1667	10	1.74	15	1	1.38	1	3.12
dsh1-1001005_012.TIF_39	dsh1-1001005_012.TIF	39	53	1683	10	1.74	15	1	1.38	1	3.12
dsh1-1001005_012.TIF_40	dsh1-1001005_012.TIF	40	54	1573	22	1.96	15	1	1.17	2	3.12
dsh1-1001005_012.TIF_41	dsh1-1001005_012.TIF	41	55	922	108	3.42	15	1	0.29	5	3.12
dsh1-1001005_012.TIF_42	dsh1-1001005_012.TIF	42	55	1472	60	2.62	15	1	0.51	4	3.12
dsh1-1001005_012.TIF_43	dsh1-1001005_012.TIF	43	57	1393	39	2.25	15	1	0.87	3	3.12
dsh1-1001005_012.TIF_44	dsh1-1001005_012.TIF	44	57	1451	69	2.77	15	1	0.35	5	3.12

FFD at distal vs FFD at proximal over Tissue



SIMAP Utility



The Silicon Cell - ultimate goal?

- http://www.bio.vu.nl/hwconf/Silicon/
- The long-term goal of the Silicon Cell (SiC) Consortium is the computation of Life at the cellular level on the basis of the complete genomic, transcriptomic, proteomic, metabolomic and cell-physiomic information that will become available in the forthcoming years.
- 3 major challenges, i.e. networks, space and time; systematic handling of data and results.

- Key objectives

(i) Computational models of catabolism, signal transduction, gene-expression regulation, coupling between supramolecular structures and fluxes, and biochemical cycling.

(ii) Model integration to calculate system properties for two real cells (E. coli and S. cerevisiae).

(iii) Demonstration of the cellular bioinformatics approach: calculating without fitting.

(iv) Methodology for modularisation to accurate mesoscopic descriptions.

(v) Visualisation, systematic data access and a www resource for two real living cells.

- Approach: focus on three different, but interconnected dimensions of cell functioning,

(i) the 'chemical and information dimension': networks of biochemical reactions and their regulation,

(ii) space: gradients and dynamic structures in signal transduction and gene expression (chromatin), and

(iii) biological time: coherent glycolytic and cell-cycle oscillations.

-

Software

Snoopy (Petri nets)
www-dssz.informatik.tu-cottbus.de/DSSZ/Software/Snoopy

Copasi (Simulator, SBML based)
www.copasi.org

Cell Designer (Graphical environment, simulator)
celldesigner.org/

Bionessie (BioModel engineering environment)
disc.brunel.ac.uk/bionessie
www.brc.dcs.gla.ac.uk/software/bionessielite

MC2 (model checker)
www.brc.dcs.gla.ac.uk/software/mc2

Sancho (general workstation cluster system)
people.brunel.ac.uk/~cspgmmt/sancho

Materials

- D Gilbert, M Heiner, R Breitling and Robin DonaldsoA
  A
  structured approach for the engineering of biochemical network models, illustrated for signalling pathways;
  Sixteenth International Conference on Intelligent Systems for Molecular Biology (ISMB 2008), Toronto, July 2008.
  - http://www.brc.dcs.gla.ac.uk/~drg/workshops/ismb08/

- Snoopy examples for ODEs
  - http://www-dssz.informatik.tu-cottbus.de/examples/ode_tutorial/

Acknowledgements

- Pam Gao
- Zujian Wu
- Maciej Trybilo
- Chris Paterson
- Huma Lodhi
- Crina Grosnan
- Hui Liu
- Mark Pook
- Annette Payne
- Amanda Harvey
- David Tree
- Simon Taylor
- Stephen Swift
- Allan Tucker
- Monika Heiner
- Rainer Breitling
- Mike Barrett
- Vlad Vyshemirsky
- Muffy Calder
- Walter Kolch
- Richard Orton
- Oliver Sturm
- Robin Donaldson
- Xuan Liu
- Jun Wang
- Aik Choon Tan
- Mallika Veeramalai
- Xu Gu
- Julie Morrison
- Jose Reyes
- Ali Al-Shahib
- Des Higham
- Susan Rosser
- Fabien Jourdan
- J-C Nebel
- Juris Viksna
- Gabriel Valiente
- Sung Hee Park

… and apologies for anyone whom I may have inadvertently omitted.

Centre for Systems and Synthetic Biology

Principles and applications of systems and synthetic biology:

- To conduct interdisciplinary research (computer science, biology, engineering, mathematics, and statistics) concerned with the principles behind systems and synthetic biology
- To apply this research to challenging real-world problems, especially in biology, environment, engineering and health science.
- www.brunel.ac.uk/research/centres/cssb

Join us!

www.brunel.ac.uk/research/centres/cssb