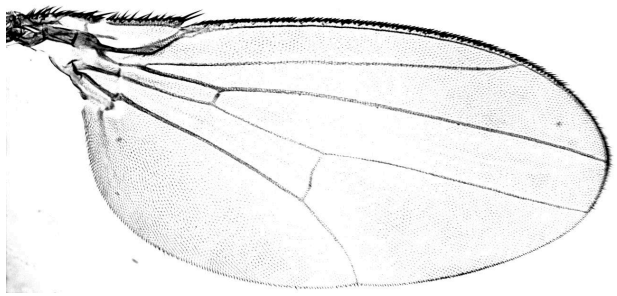
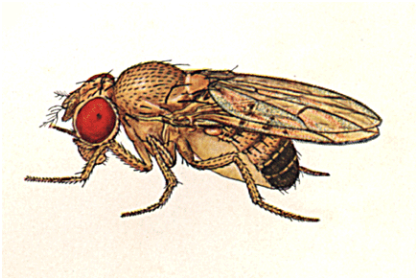




BioModel Engineering: The MultiScale challenge

David Gilbert

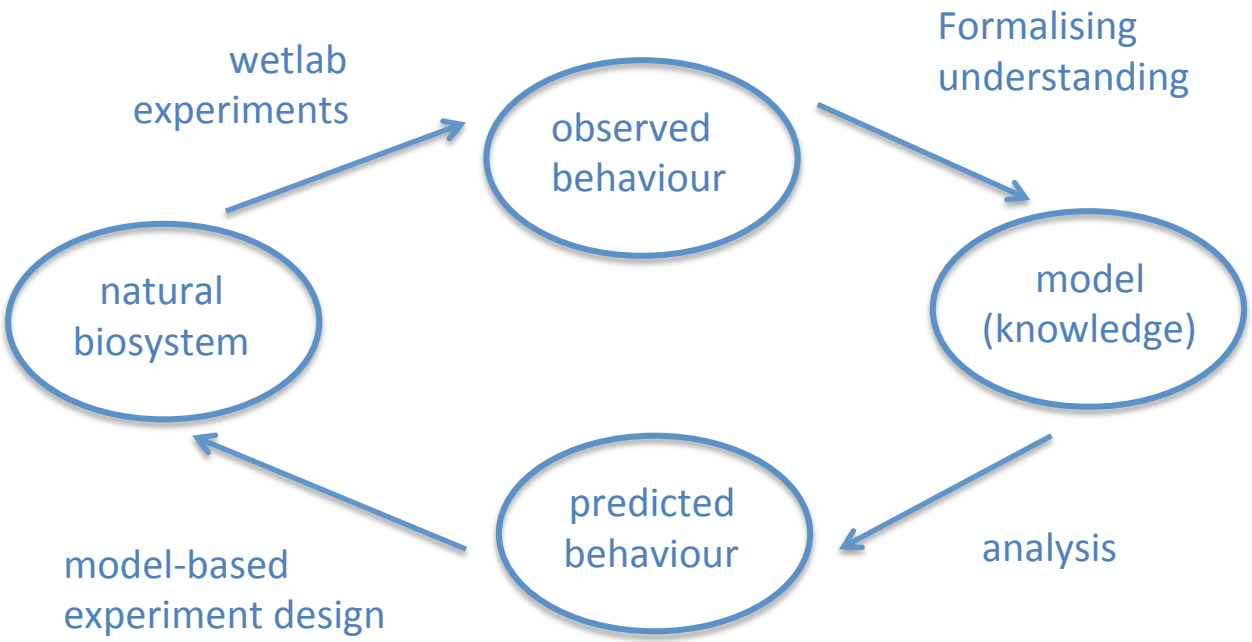
School of Information Systems, Computing & Mathematics

Centre for Systems & Synthetic Biology

Brunel UNIVERSITY LONDON

david.gilbert@brunel.ac.uk

Systems biology



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

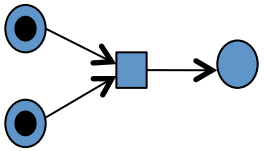
1. Problem identification
2. Construction
3. Simulation
4. Analysis & interpretation
5. Management & development

Biomodel engineering

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

Molecules/Levels
CTL, LTL

Qualitative



Time-free

Timed, Quantitative

Abstraction

Abstraction

Monika Heiner

Markov chain
Molecules/Levels
Stochastic rates
CSL

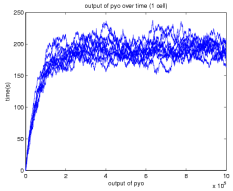
Stochastic

Approximation

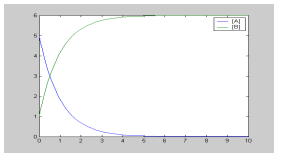
Approximation

Continuous

ODES
Concentrations
Deterministic rates
LTLc



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



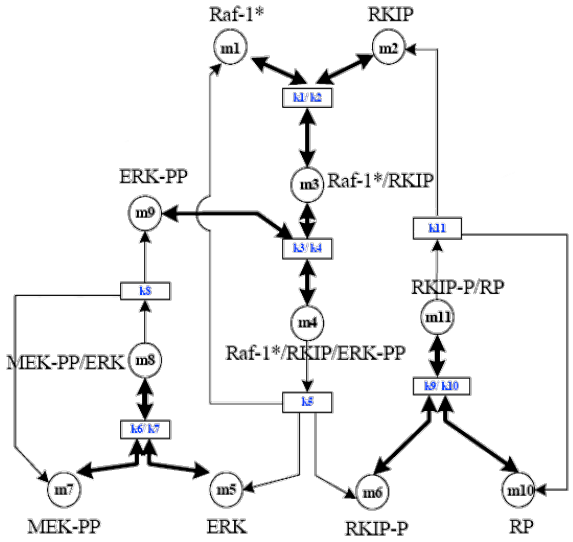
DiscreteState Space

Continuous State Space

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

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

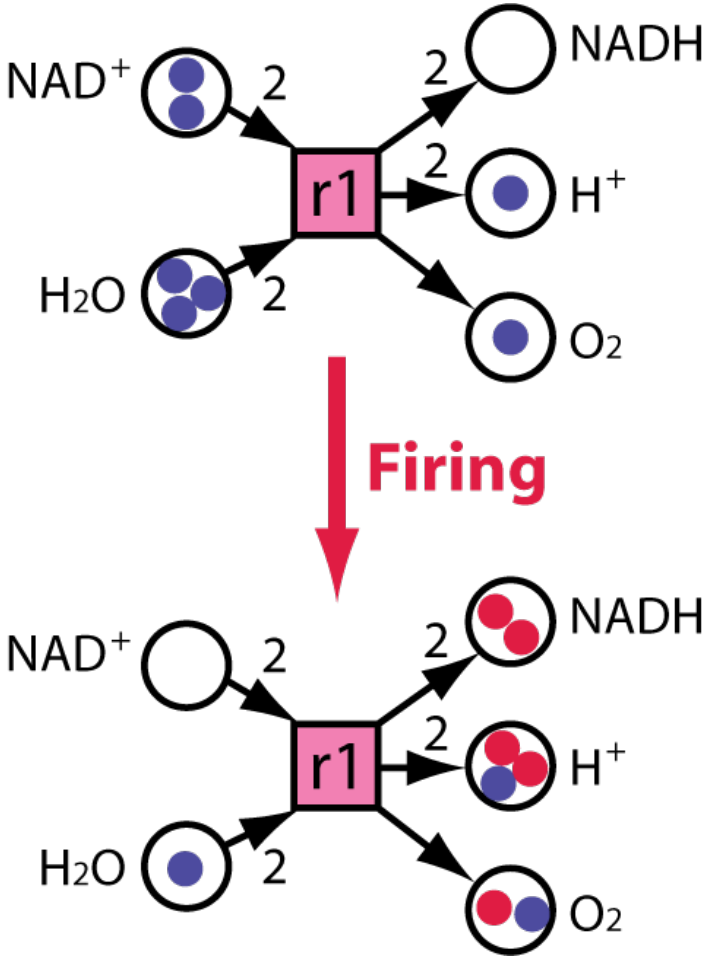
3. Initial conditions

marking , concentrations

$[Raf1^*]_{t=0}$= 2 μMolar

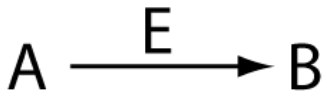
QUANTITATIVE

Petri nets



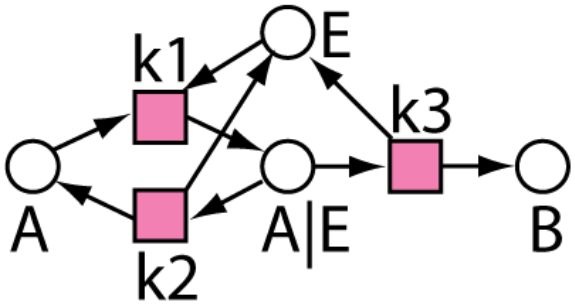
Simple enzymatic reaction

a



b

c



d

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

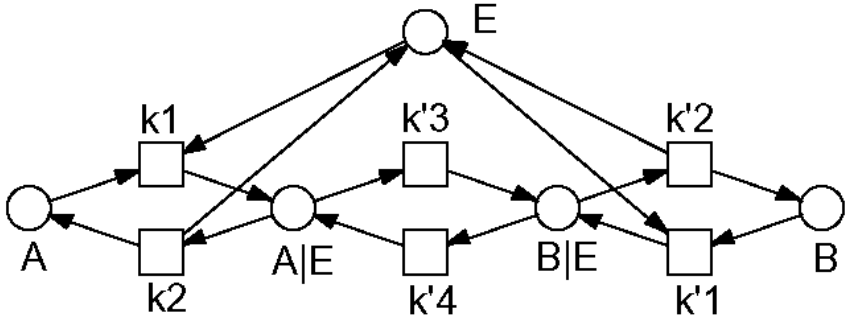
$$\frac{d[E]}{dt} = -k_1[A][E] + k_2[A \mid E] + k_3[A \mid E]$$

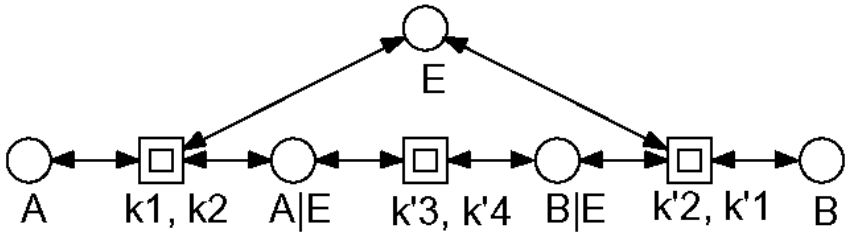
$$\frac{d[A \mid E]}{dt} = k_1[A][E] - k_2[A \mid E] - k_3[A \mid E]$$

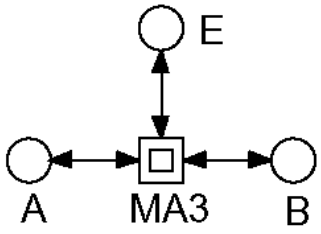
$$\frac{d[B]}{dt} = k_3[A \mid E]$$

MA3 model

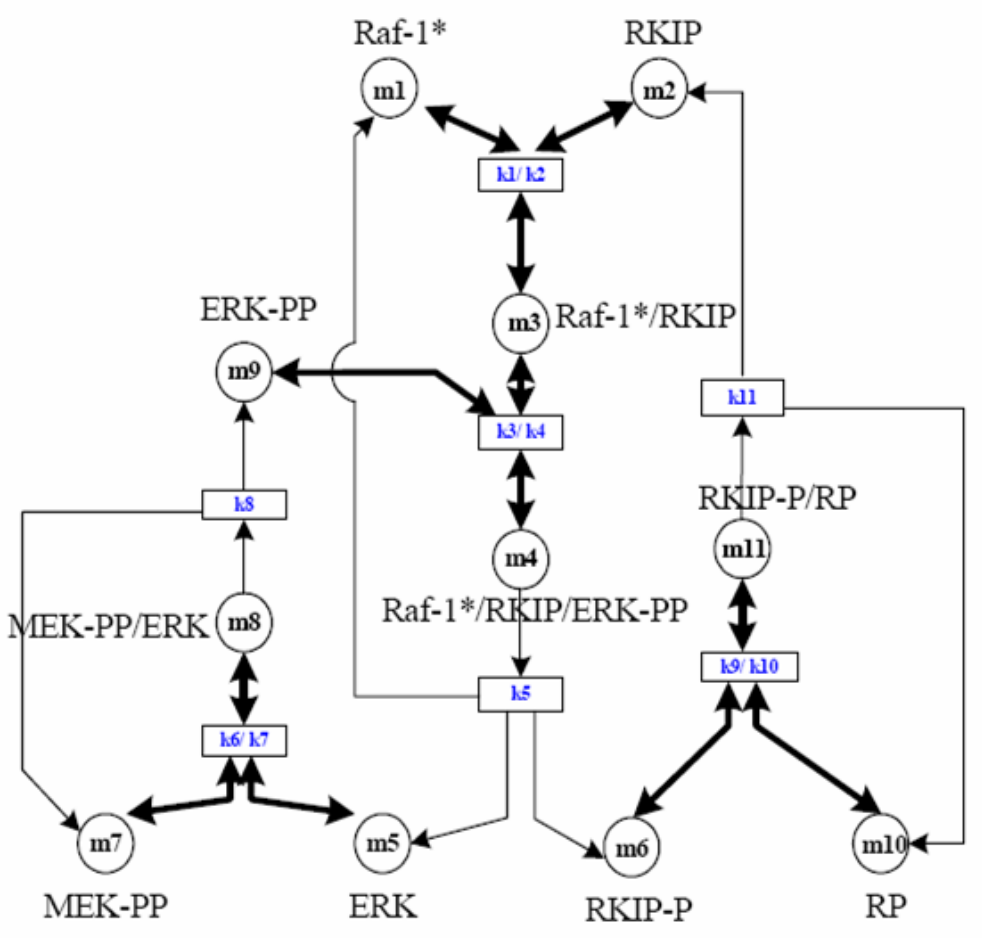
$$A + E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} A|E \underset{k_6}{\overset{k_3}{\rightleftharpoons}} B|E \underset{k_5}{\overset{k_4}{\rightleftharpoons}} B + E$$

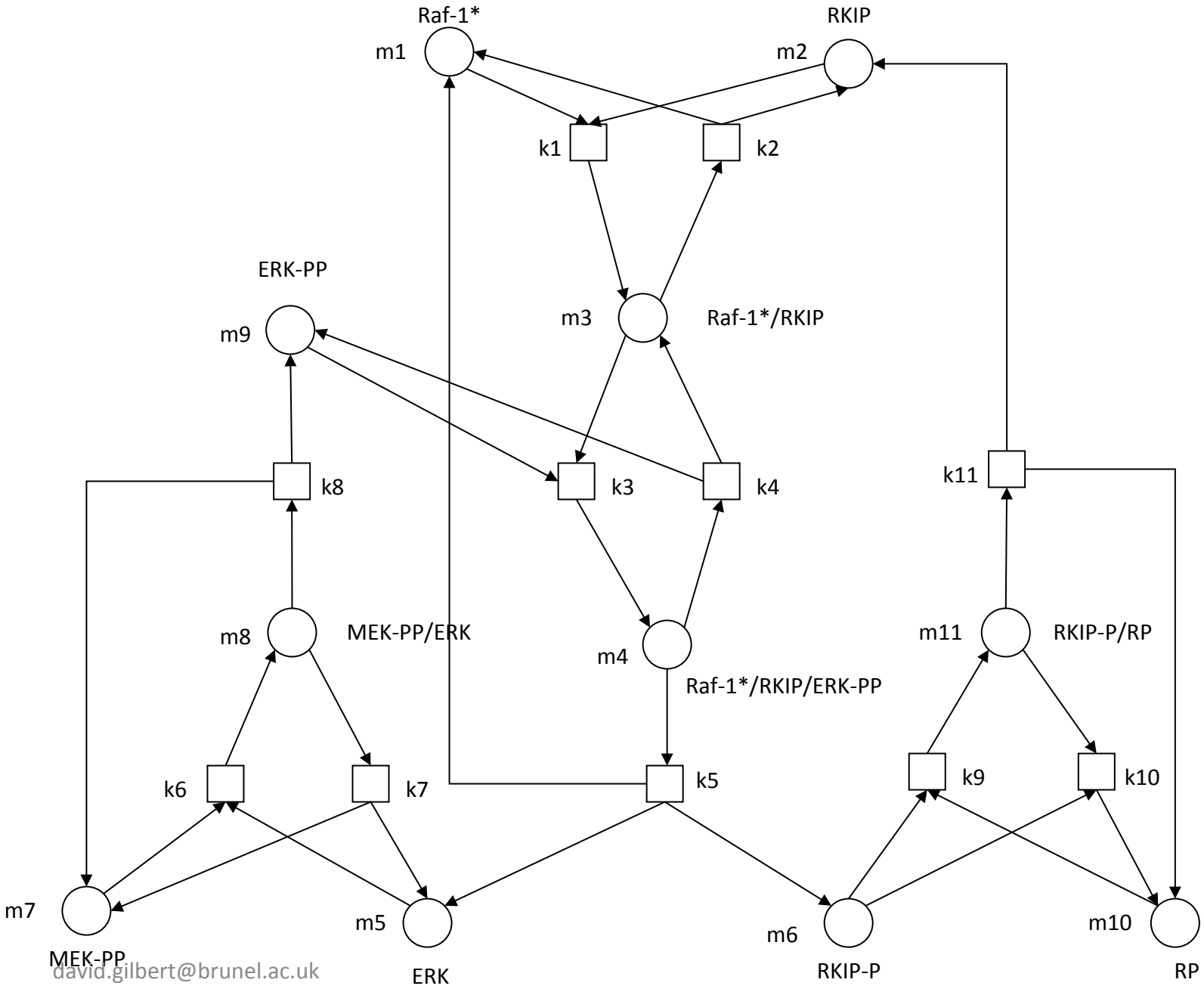




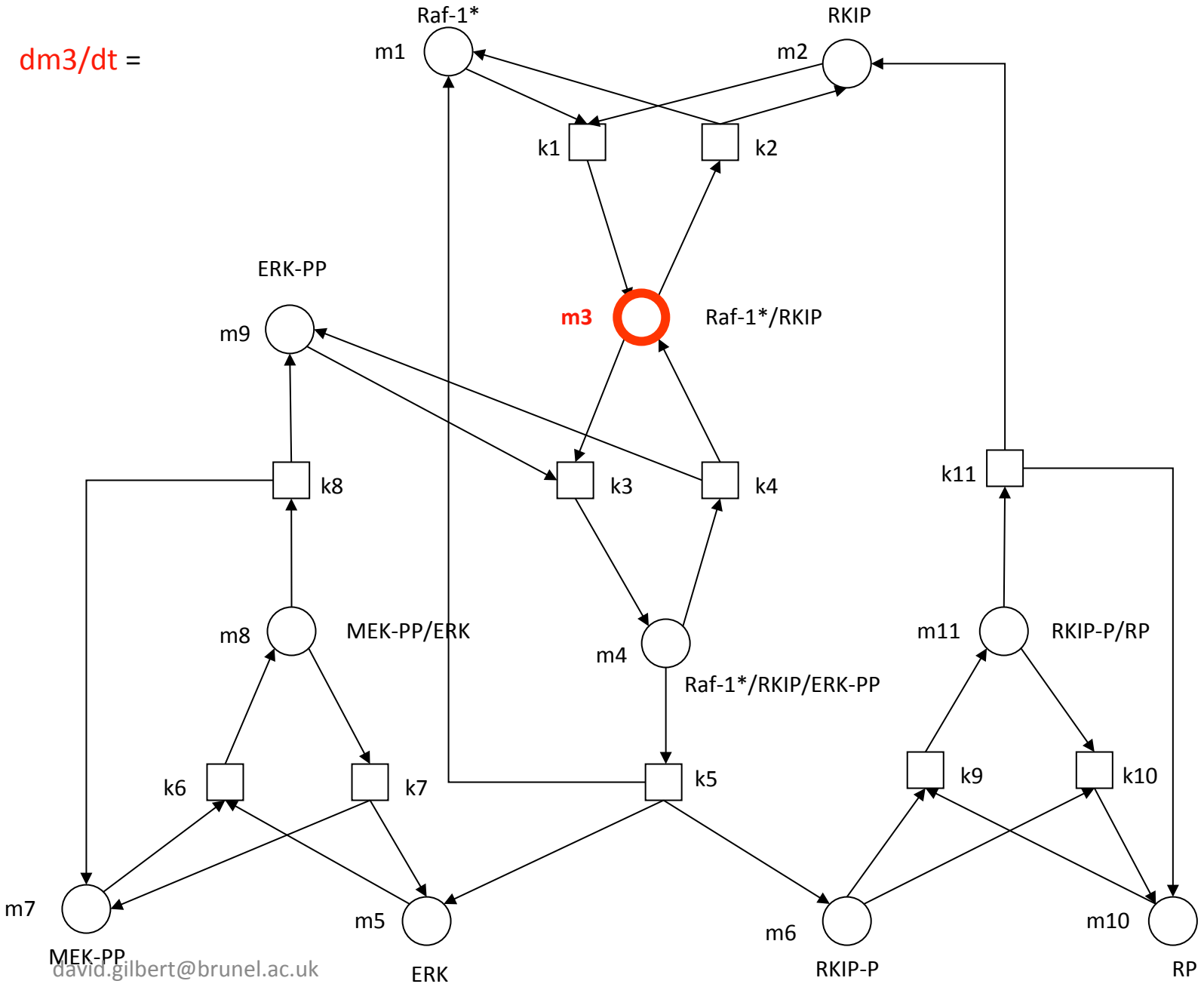


The Raf-1/RKIP/ERK pathway

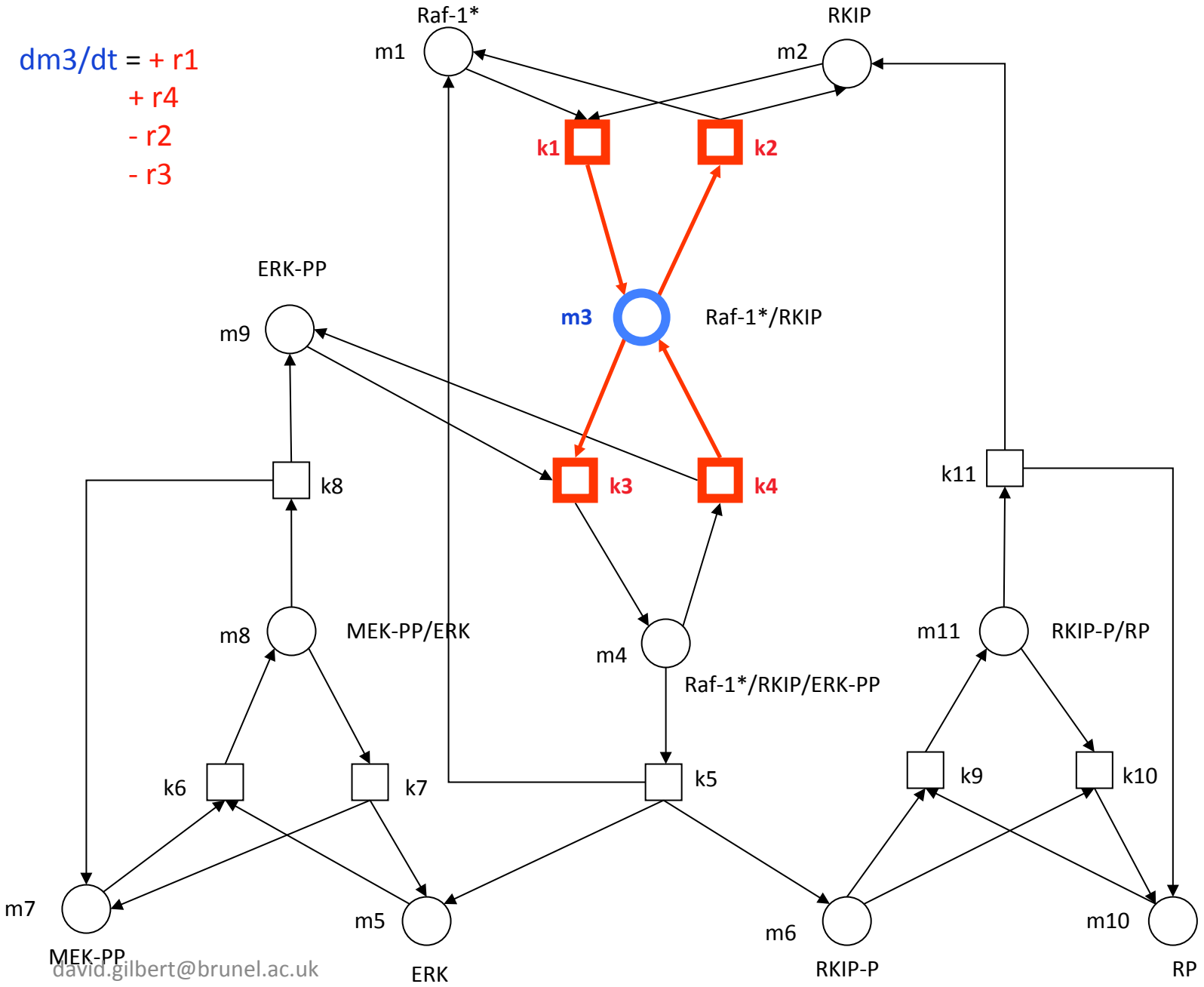



Raf-1*
RKIP
m1
m2
k1
k2
ERK-PP
m3
Raf-1*/RKIP
m9
k8
k3
k4
k11
m8
MEK-PP/ERK
m4
Raf-1*/RKIP/ERK-PP
m11
RKIP-P/RP
k6
k7
k5
k9
k10
m7
MEK-PP
m5
ERK
m6
RKIP-P
m10
RP

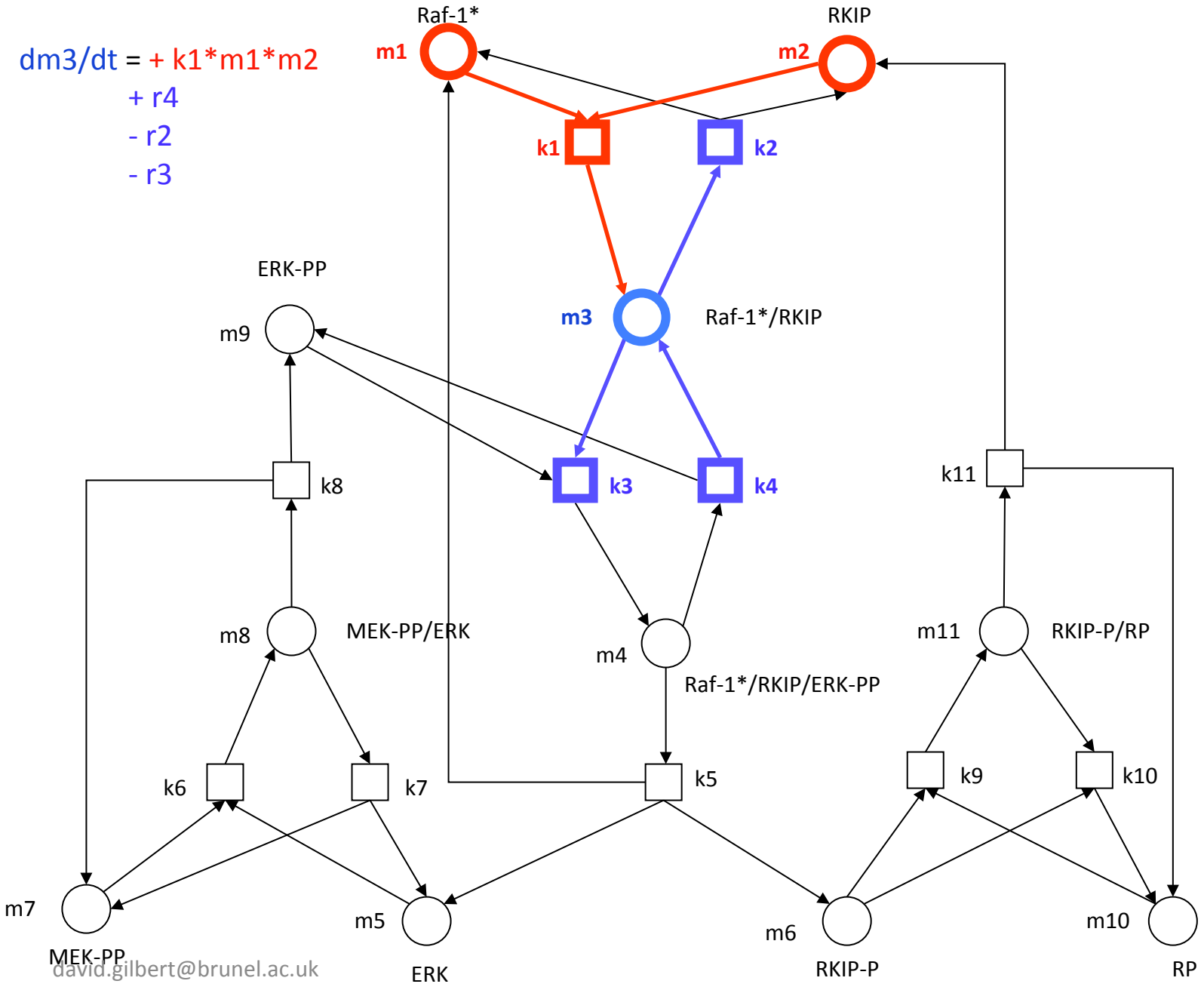
dm3/dt =


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

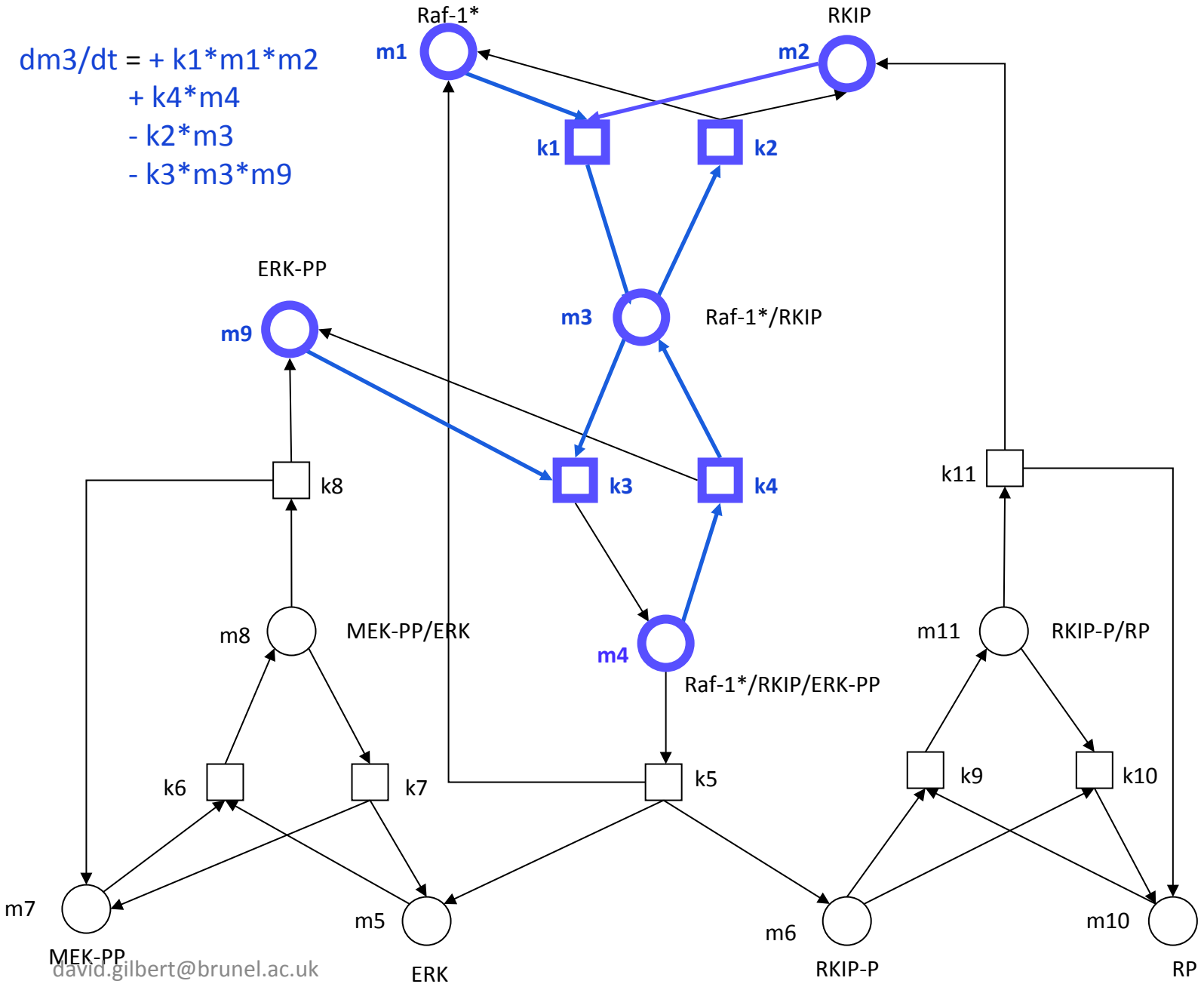
dm3/dt = + r1
+ r4
- r2
- r3



dm3/dt = + k1*m1*m2
+ r4
- r2
- r3

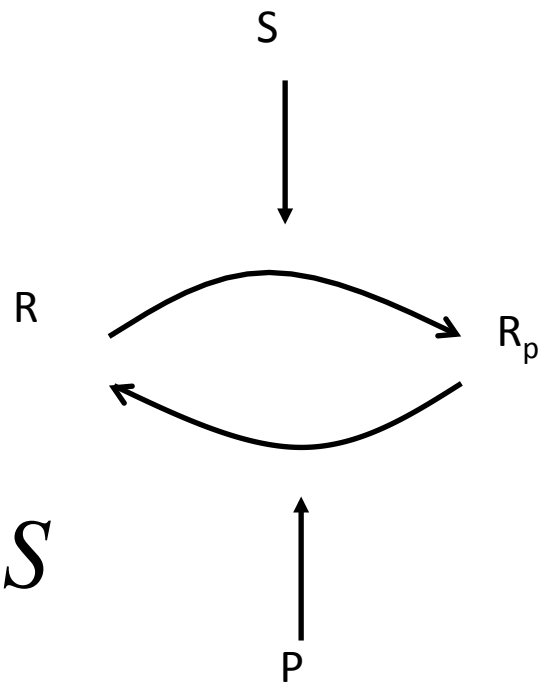


dm3/dt = + k1*m1*m2
+ k4*m4
- k2*m3
- k3*m3*m9



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

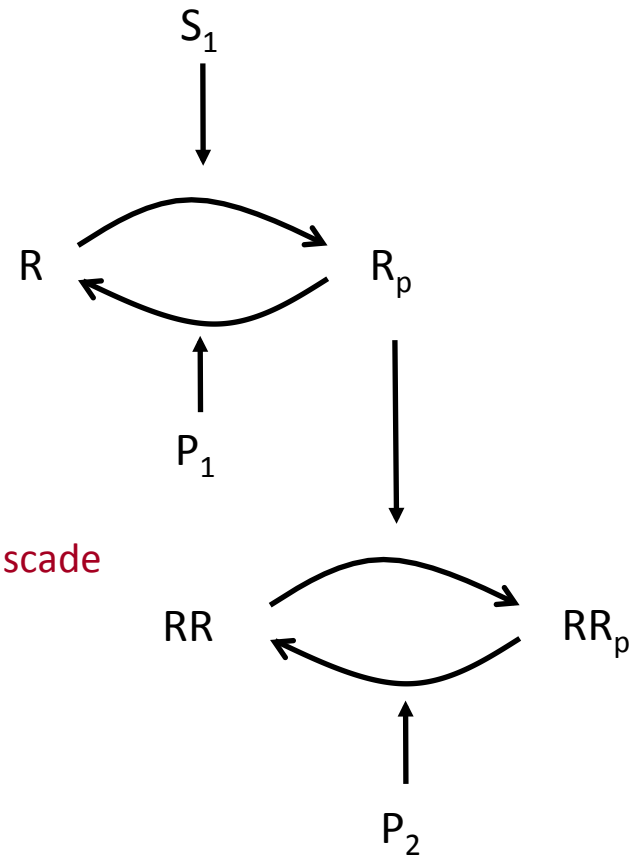


$$R+S \underset{k_2}{\overset{k_1}{\rightleftarrows}} R|S \xrightarrow{k_3} R_p + S$$

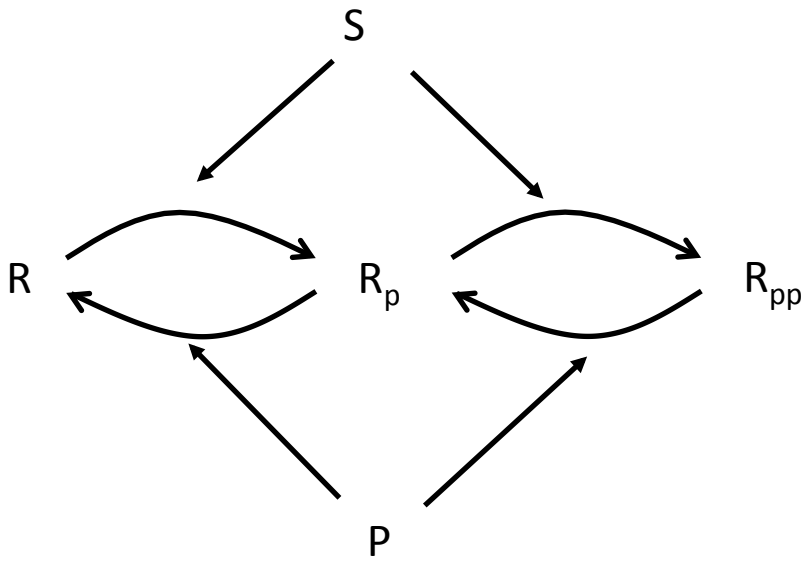
$$R+P \xleftarrow{kr_3} R_p|P \underset{kr_2}{\overset{kr_1}{\rightleftarrows}} R_p + P$$

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

Composition
Vertical & horizontal

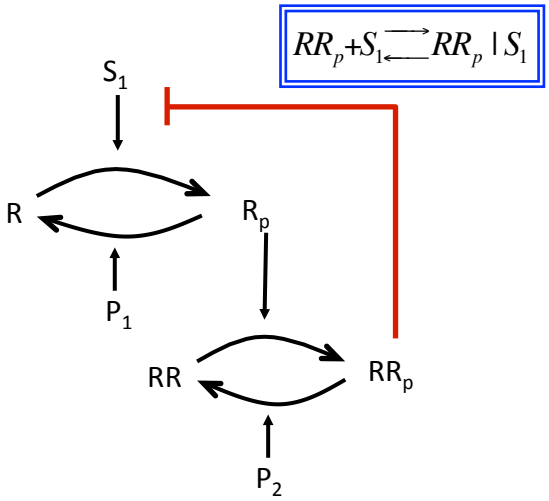


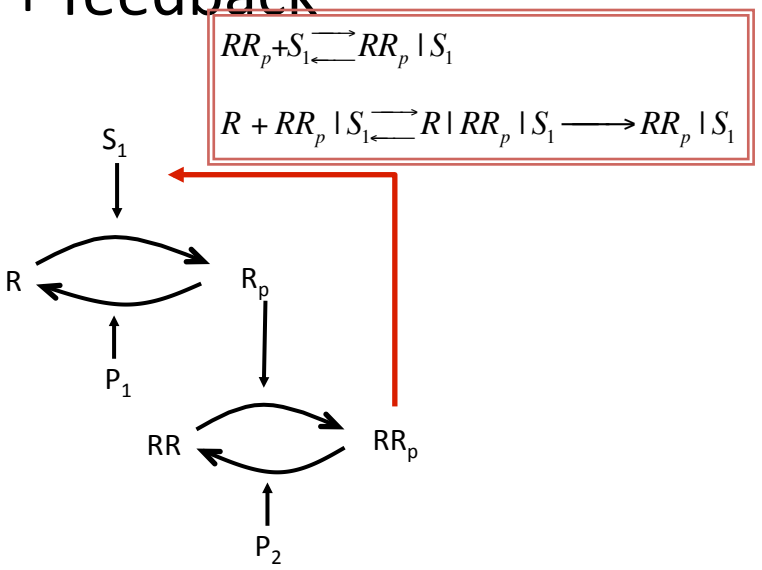
2-stage cascade

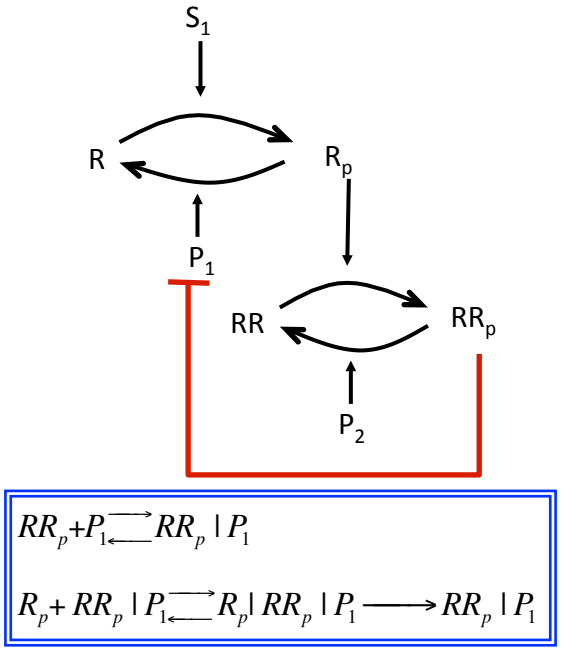


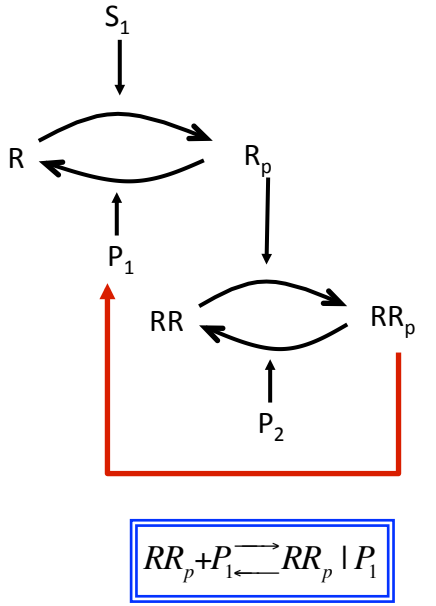
1-stage cascade
double phosphorylation step

Phosphorylation cascade + feedback



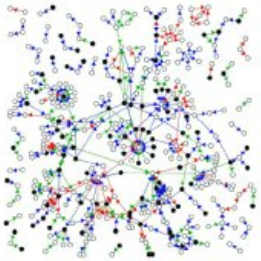




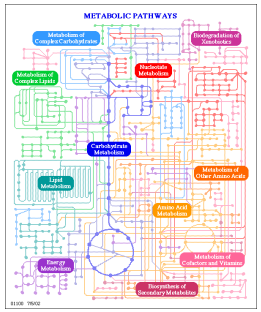


Networks

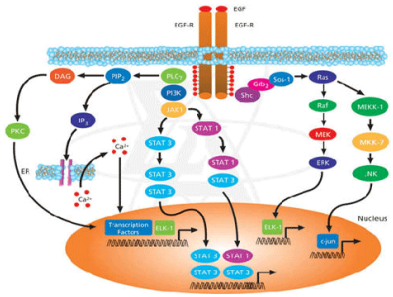
- Gene regulation



- Metabolic



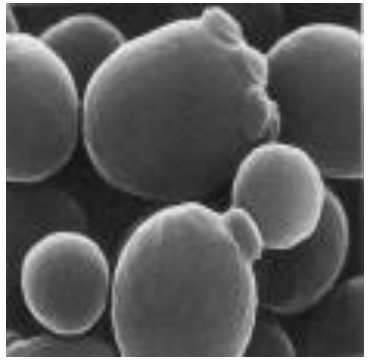
- Signalling



- Protein-protein interaction

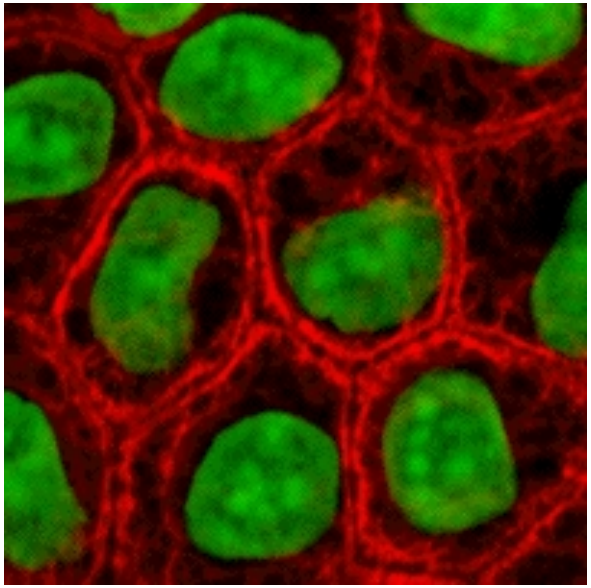
- Developmental

What about scaling up?



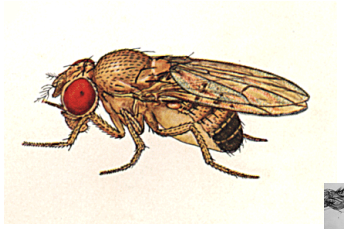
Multiscale modelling challenges

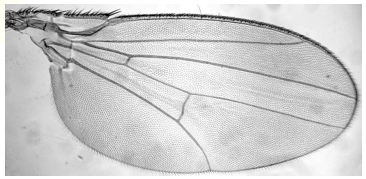
- ***Repetition*** – multiple cells witha similar definitions
- ***Variation*** – mutants.
- ***Organisation -*** regular or irregular patterns over spatial networks in one, two or three dimensions.
- ***Communication*** – between neighbours constrained by neighbour relation, and the position in spatial network.

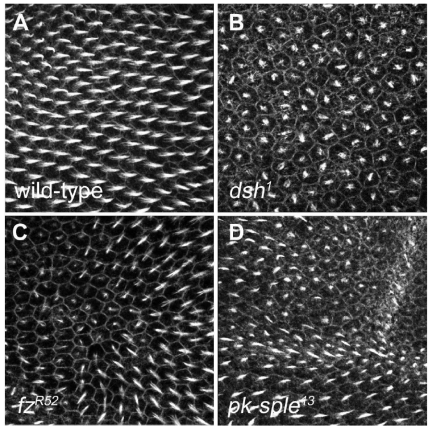


- ***Hierarchical organisation*** –cells containing compartments. Enables abstraction over level of detail of components.

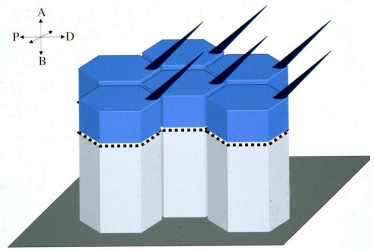
Multiscale from signalling to organs

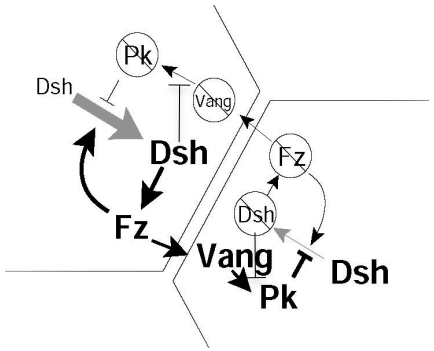






Planar Cell Polarity





Pam Gao, David Tree

Joint work

Brunel:

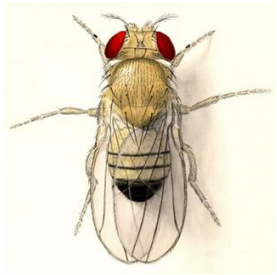
- Pam Gao

- David Tree

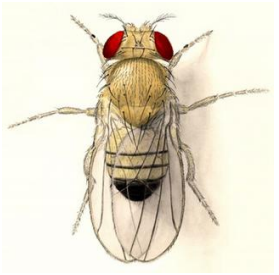
Cottbus:

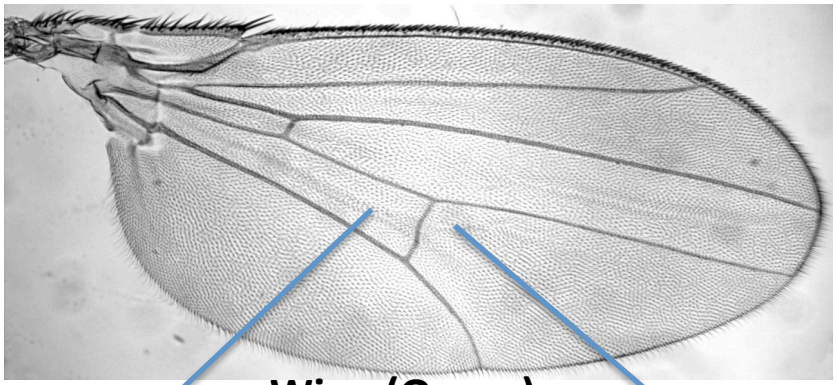
- Monika Heiner

- Fei Liu

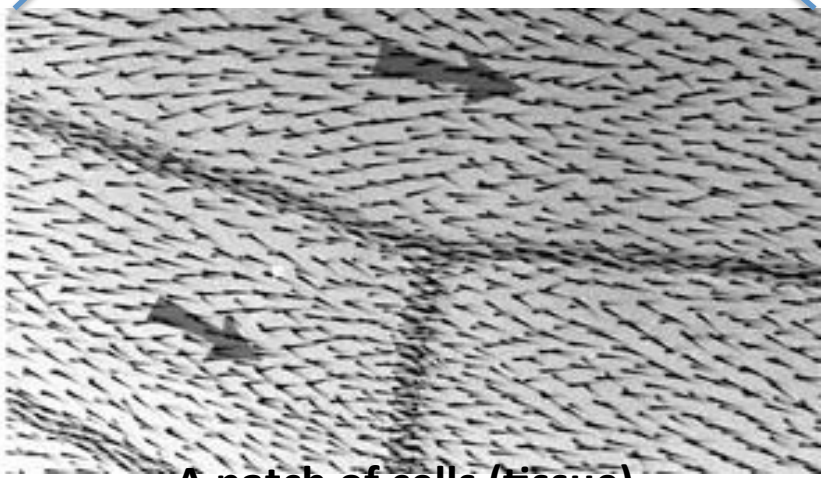


Planar Cell Polarity



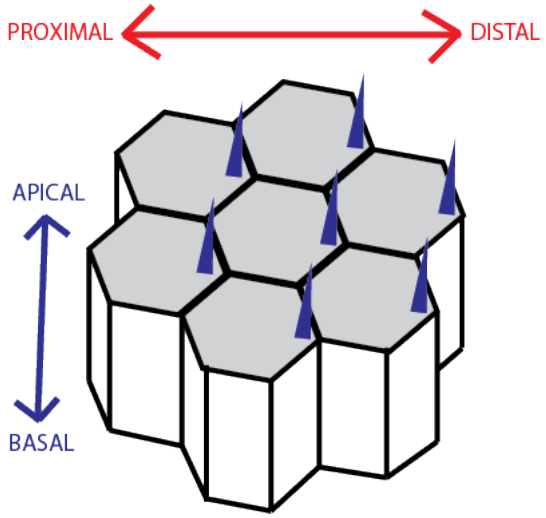


Wing (Organ)



A patch of cells (tissue)

- Drosophila wing hairs point distally virtually error free.
- Hexagonally packed, planar (300,000)
- PCP: the polarization of a field of cells within the plane of a cell sheet.
- Human pathology:
  - ✓Cochlear hair cells
  - ✓Spina bifida
  - ✓Oncogenic Wnt pathway

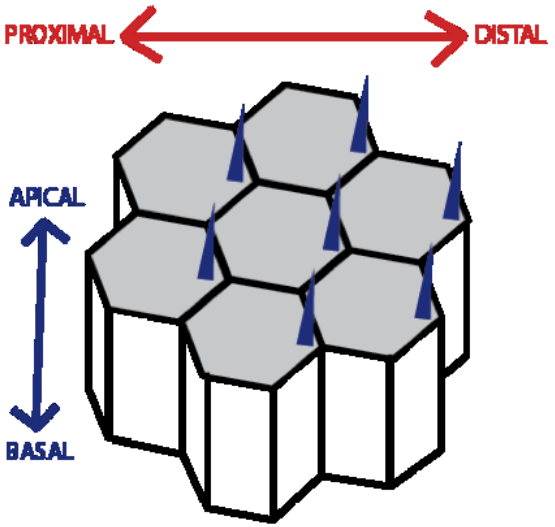


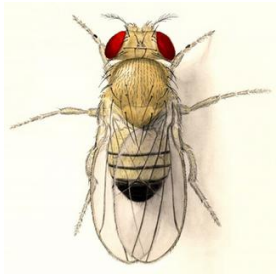
Planar Cell Polarity

- The orientation of cells within the plane of the epithelium, orthogonal to the apical-basal polarity of the cells.

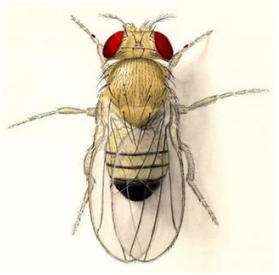
- This polarisation is required for many developmental events in both vertebrates and non- vertebrates.

- Defects in PCP in vertebrates underlie developmental abnormalities in multiple tissues including the neural tube, the kidney and the inner ear

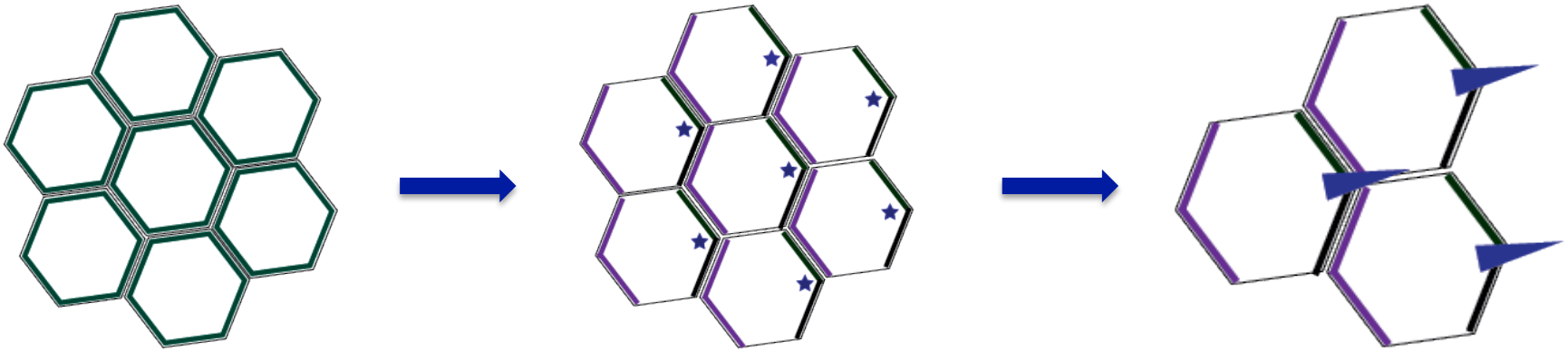




Biological Model



- **A core machinery** mediates a competition between the proximal and distal proteins on *adjacent surfaces of neighboring cells* and amplifies small differences to **result in** a highly **asymmetric distribution** of Frizzled (Fz) and Vang.
- As a result of the above machinery, **Fz** accumulates on the **distal** side of the cell, designating it as the future *site* ***for prehair formation***, while **Vang** accumulates on the **proximal** side of the neighbouring cell.
- There are feedback loops functioning as well. A consequence of these **feedback loops** is that cells tent to align their polarity such that each cell accumulates high levels of **Fz on the same side of the cell and high levels of Vang on the opposite side**.

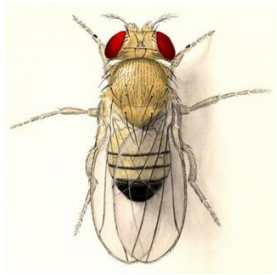


Symmetric distribution of protein complexes

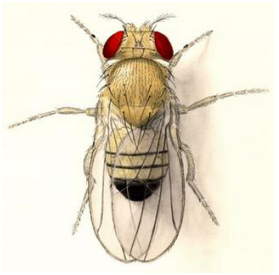
Asymmetric distribution of protein complexes

Prehair formation

david.gilbert@brunel.ac.uk

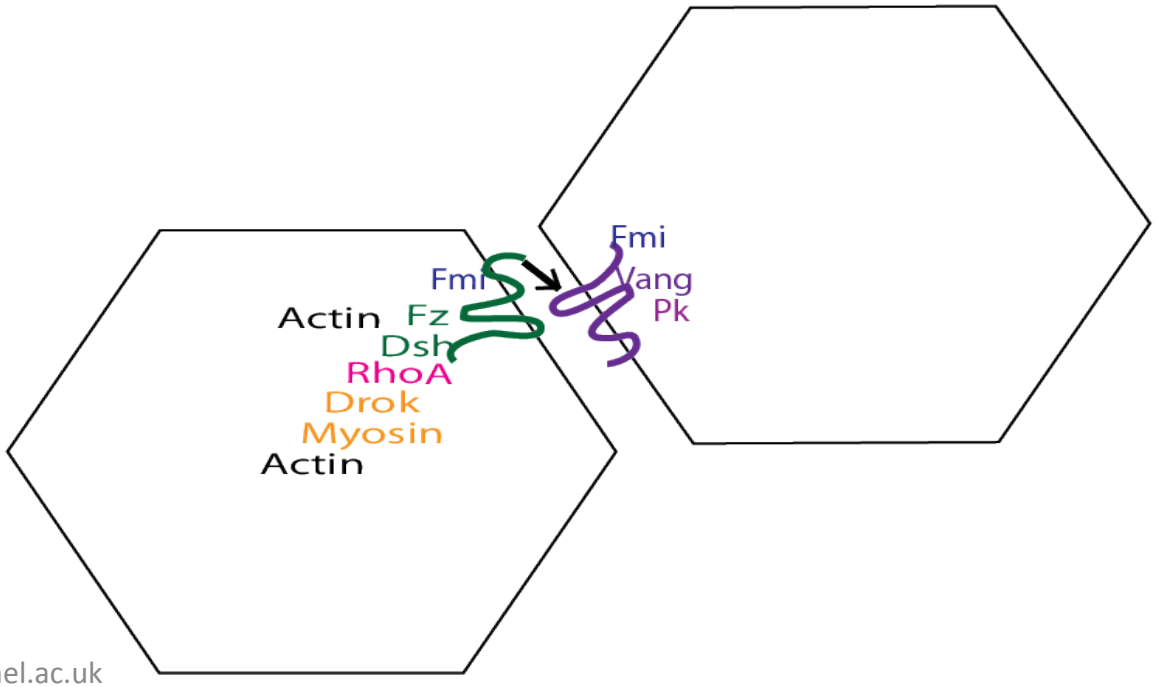


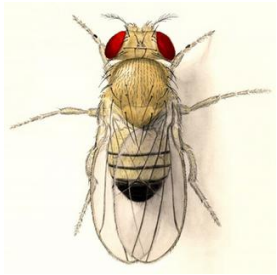
Biological Model



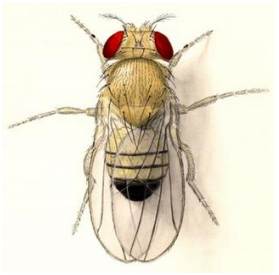
• A**n intracellular signalling cascade** functions downstream of the core machinery, coupling signaling from the core proteins to the cell-type specific responses required to generate PCP in the individual tissues.

• PCP proteins Frizzled (Fz), Dishevelled (Dsh), Prickle (Pk), Flamingo (Fmi) and Van-Gogh (Vang)

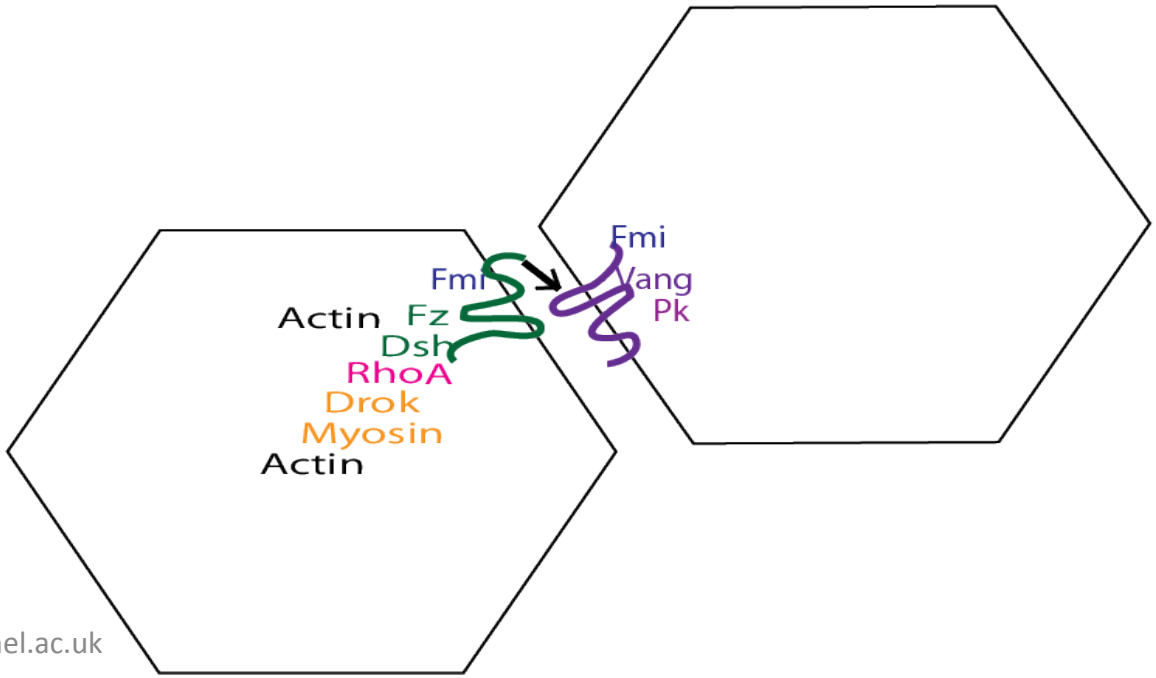




Biological Model

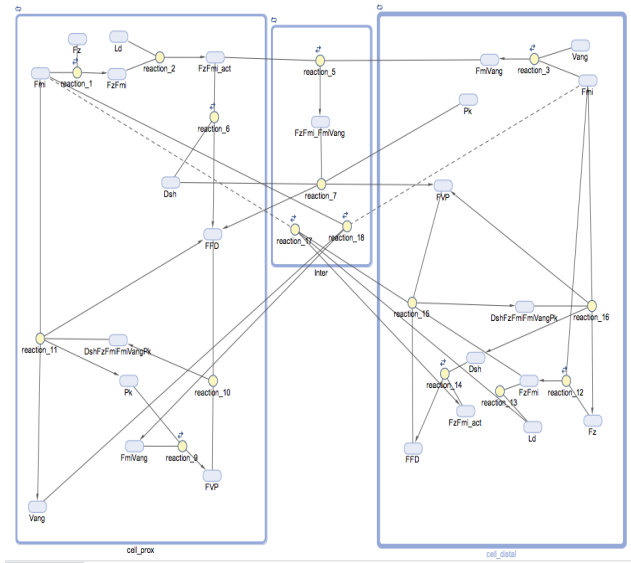


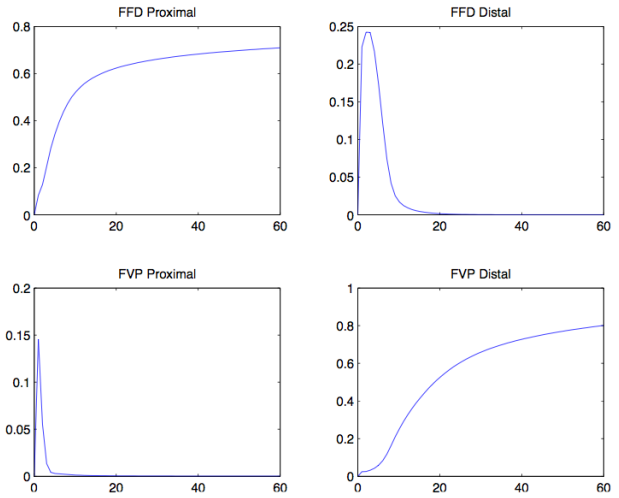
- At initiation of PCP signalling: Fmi, Fz, Dsh, Vang and Pk are all present symmetrically at the cell membrane.
- At the conclusion of PCP signalling: Fmi is found at both the proximal and distal cell membrane, Fz and Dsh are found exclusively at the distal cell membrane and Vang and Pk are found exclusively at the proximal cell membrane.
- Communication between these proteins at cell boundaries. Distally localised Fmi, Fz and Dsh recruit Fmi, Vang and Pk to the proximal cell boundary and vice versa.



ODEs

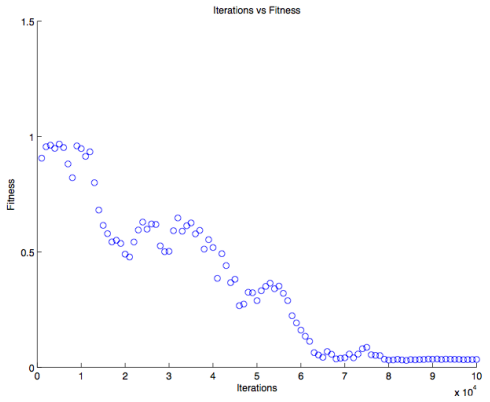
- Current model
  - Inter-cellular signalling:
    25 molecular species and
    23 reactions
  - Kinetic laws: mass action

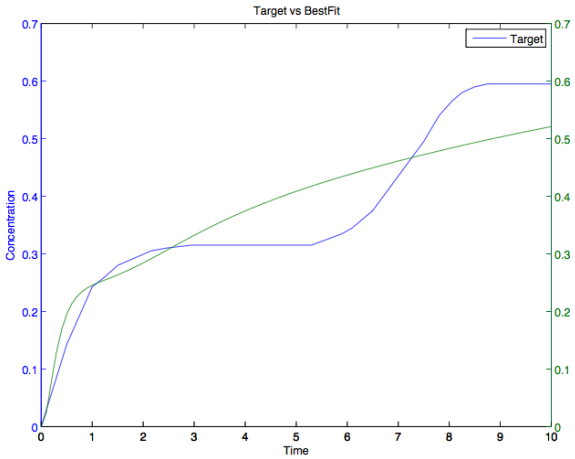


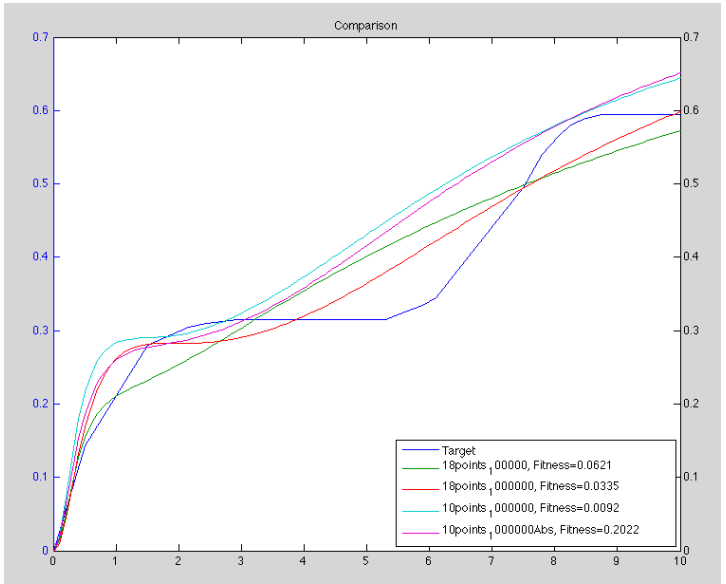


Parameter Estimation

Simulated Annealing







Coloured Petri nets

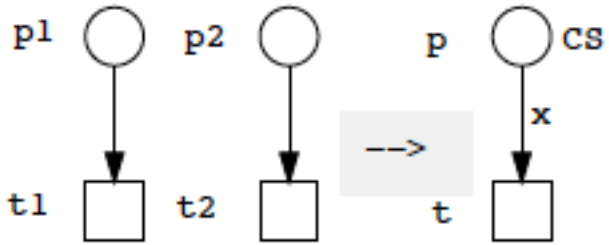
- Tokens distinguished via their colours.
- Each place gets a colour set, specifying the kind of tokens which can reside on the place.
- Each transition gets a guard, specifying which coloured tokens are required for firing.
- Each arc gets an arc inscription specifying the kind of tokens flowing through it

- Allows for the discrimination of species (molecules, metabolites, proteins, secondary substances, genes, etc.).
- Colours can be used to distinguish between sub-populations of a species in different locations (cytosol, nucleus and so on).

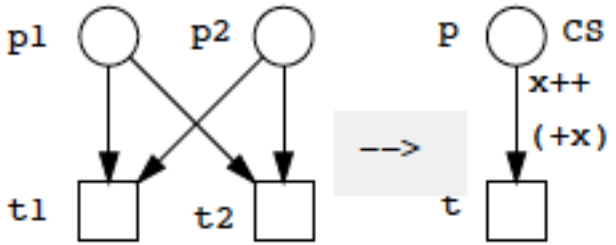
Coloured Petri net

- A coloured Petri net is a tuple N = [P,T,F,Σ,c,g,f,m0], where:
- P is a finite, non-empty set of places.
- T is a finite, non-empty set of transitions.
- F is a finite, non-empty set of directed arcs.
- Σ is a finite, non-empty set of colour sets.
- c : P → Σ is a colour function that assigns to each place p∈P a colourset c(p)∈Σ.
- g : T → EXP is a guard function that assigns to each transition t ∈ T a guard expression of Boolean type.
- f : F → EXP is an arc function that assigns to each arc a ∈ F an arc expression of a multiset type $c(p)_{MS}$, where p is the place connected to the arc a.
- m0 : P → EXP is an initialisation function that assigns to each place p ∈ P an initialisation expression of a multiset type $c(p)_{MS}$.

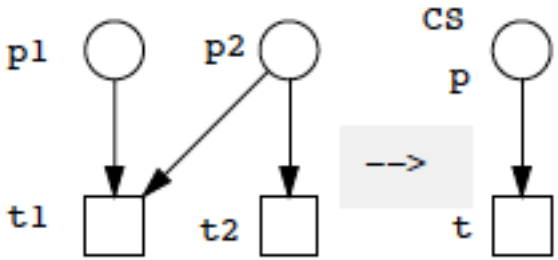
Coloured Petri net folding



(a)



(b)



(c)

[x=1](x++(+x))++
[x=2]x

```
Declarations:
colorset CS = int with 1,2;
variable x : CS ;
```

(d)

++ multiset addition
(+x) successor
[x=2] guard

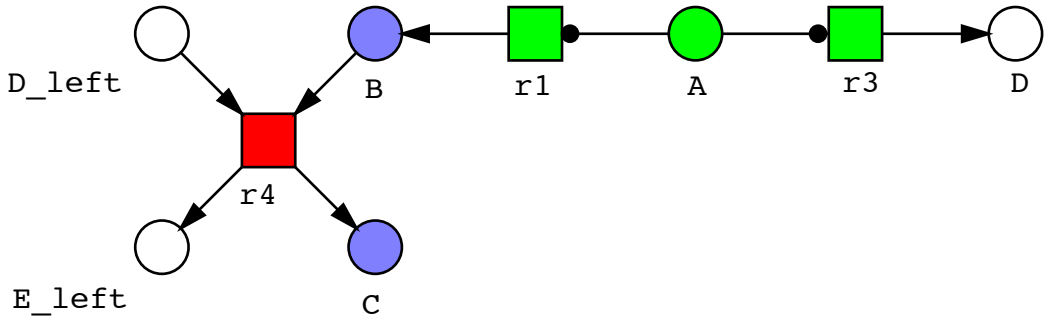
Grid constraints (rectangular)

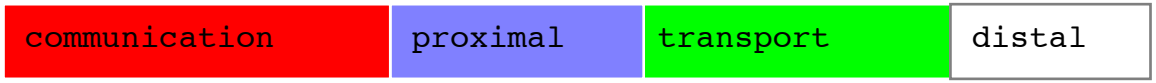
```
neighbour1D(X, A, D1):-
    (A=X; A = X+1; A = X-1),
    not(A=X),
    A <= D1,
    A >= 1.

neighbour2D((X,Y), (A,B), (D1,D2)):-
    (A=X; A = X+1; A = X-1),
    (B=Y; B = Y+1; B = Y-1),
    not((A=X,B=Y)),
    A <= D1, B <= D2,
    A >= 1, B >= 1.
```

```
neighbour3D((X,Y,Z), (A,B,C), (D1,D2,D3)):-
    (A=X; A = X+1; A = X-1),
    (B=Y; B = Y+1; B = Y-1),
    (C=Z; C = Z+1; C = Z-1),
    not((A=X,B=Y,C=Z)),
    A <= D1, B <= D2, C <= D3,
    A >= 1, B >= 1, C >= 1.
```

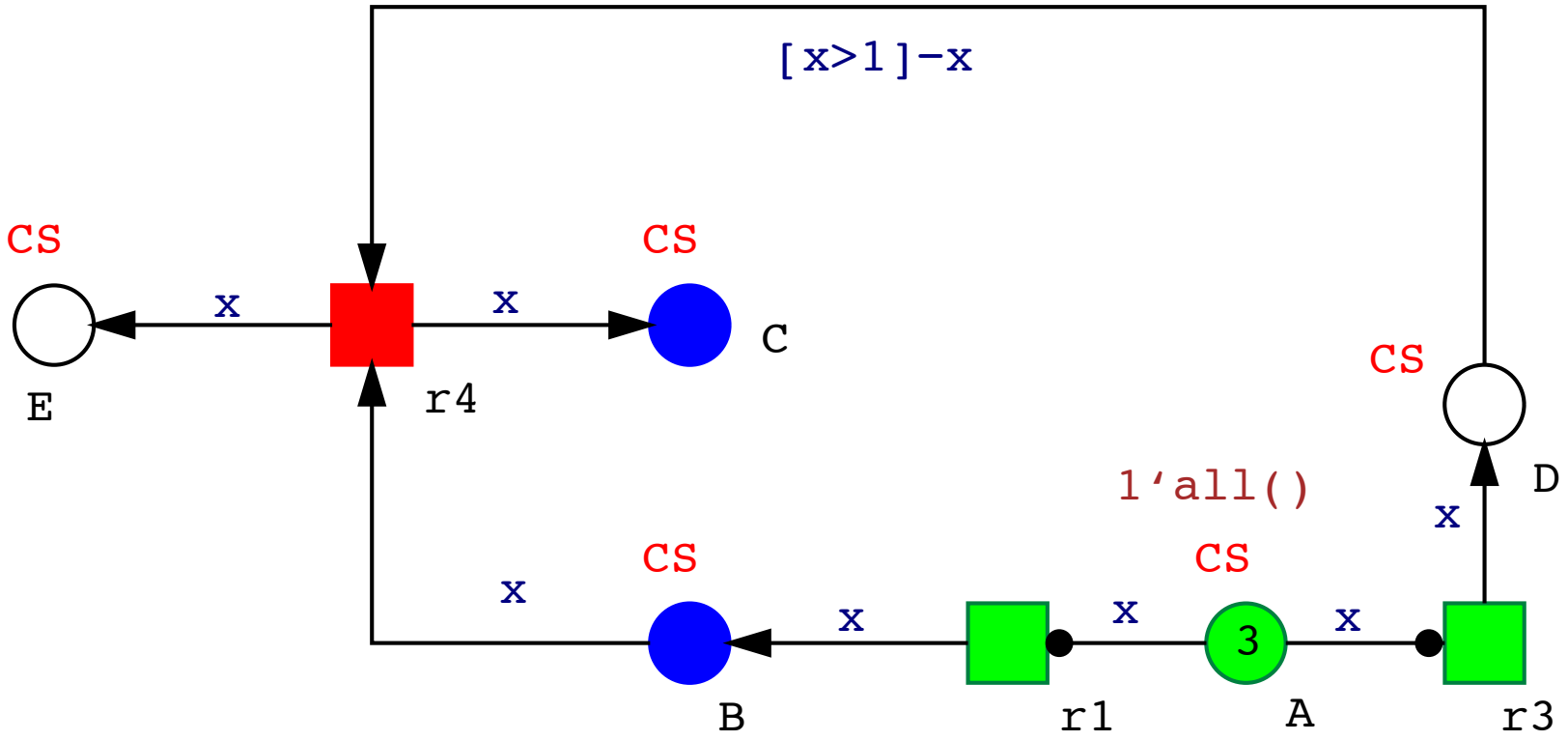
Single cell
Abstract level





(Labelled colours **not** about CPN colour sets)

CPN model for cells linked in a pipeline.



colourset CS = int with 1–N, variable x: CS.

The arc expression [x > 1]–x indicates that the first cell is not linked to the last.

Spatial organisation & colours

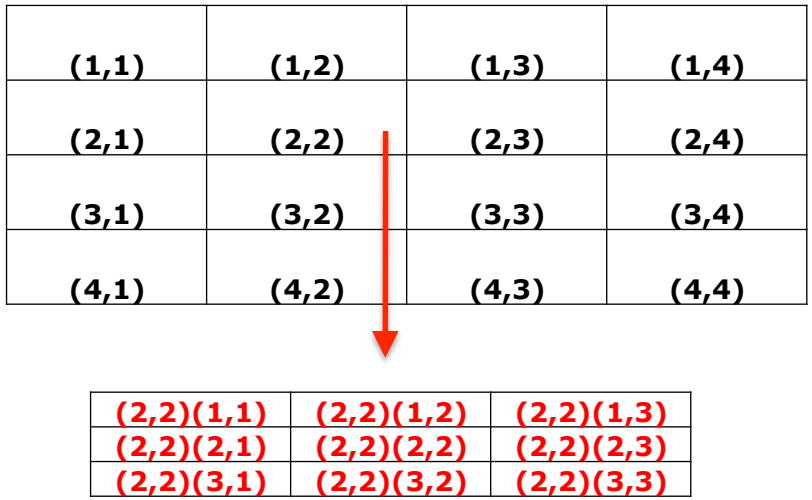
- Reflect organisation by colour structure

(1,1)	(1,2)	(1,3)	(1,4)
(2,1)	(2,2)	(2,3)	(2,4)
(3,1)	(3,2)	(3,3)	(3,4)
(4,1)	(4,2)	(4,3)	(4,4)

Colourset = {(1,1),(1,2),(1,3),(1,4),(2,1),(2,2),(2,3),(2,4), (3,1), (3,2), (3,3), (3,4), (4,1), (4,2), (4,3), (4,4) }

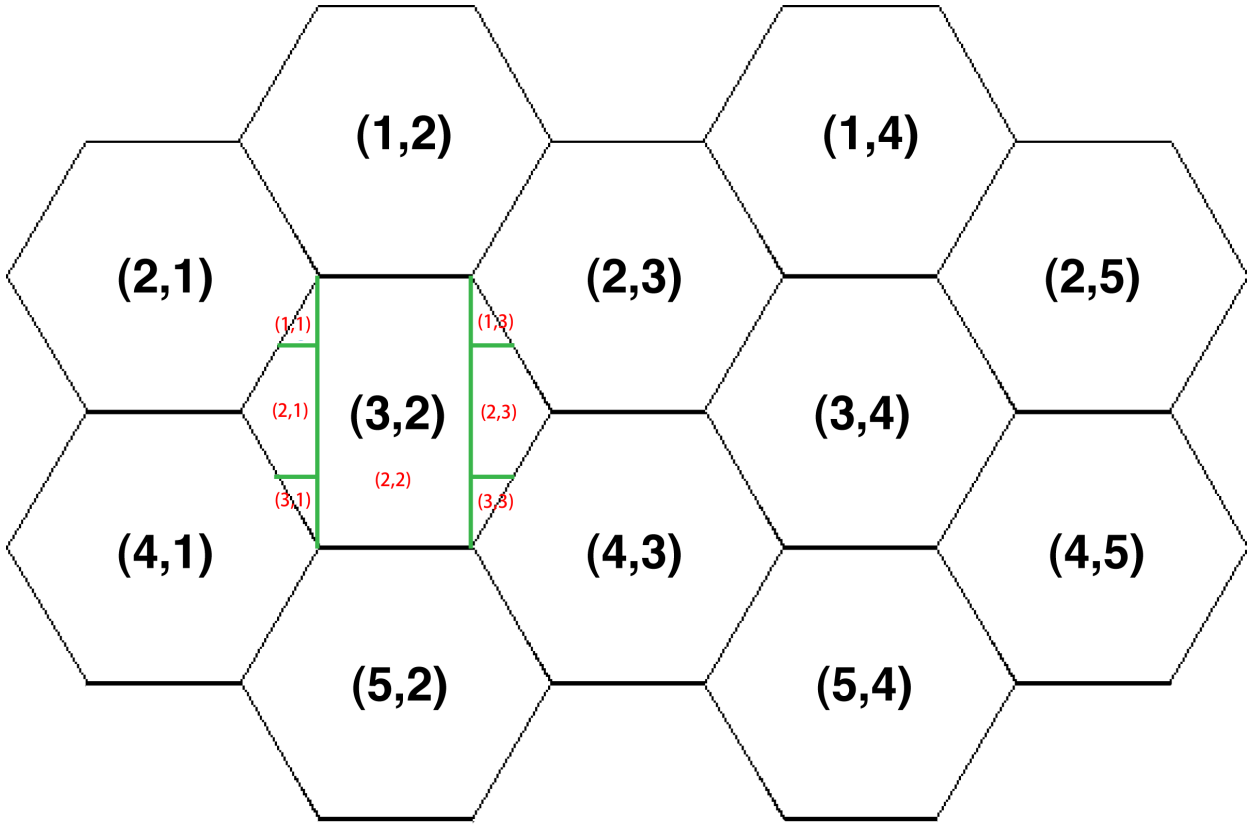
Hierarchical organisation

- Hierarchically coloured

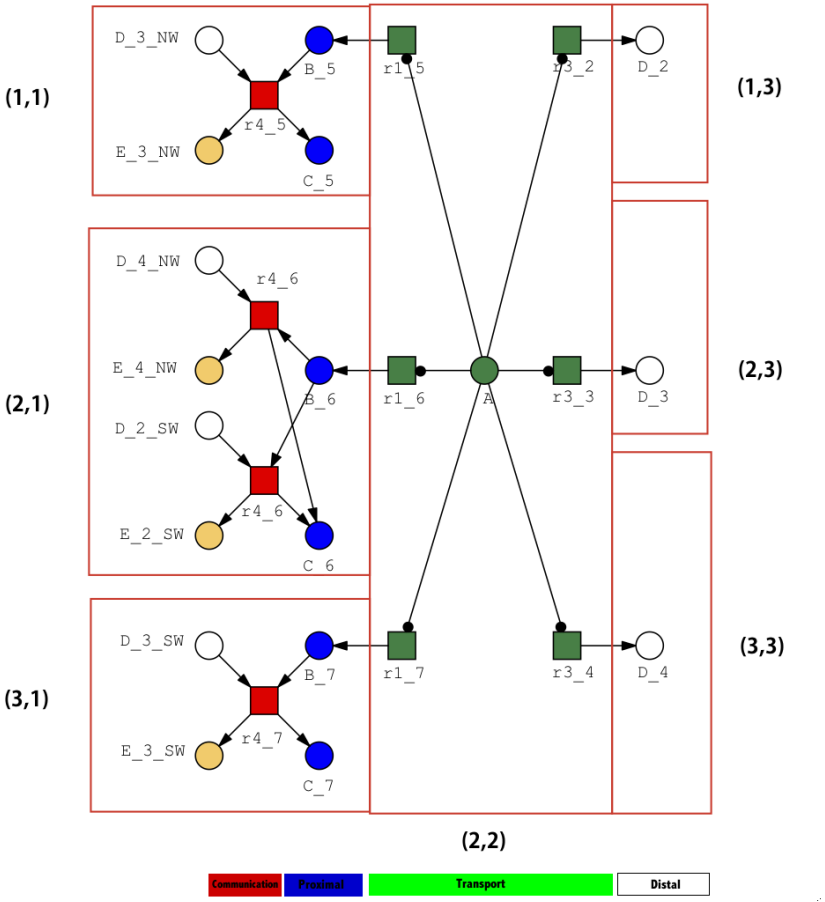


Colourset = {…, {((2,2)(1,1)), ((2,2)(1,2)), ((2,2)(1,3)),……((2,2)(3,3))}, …

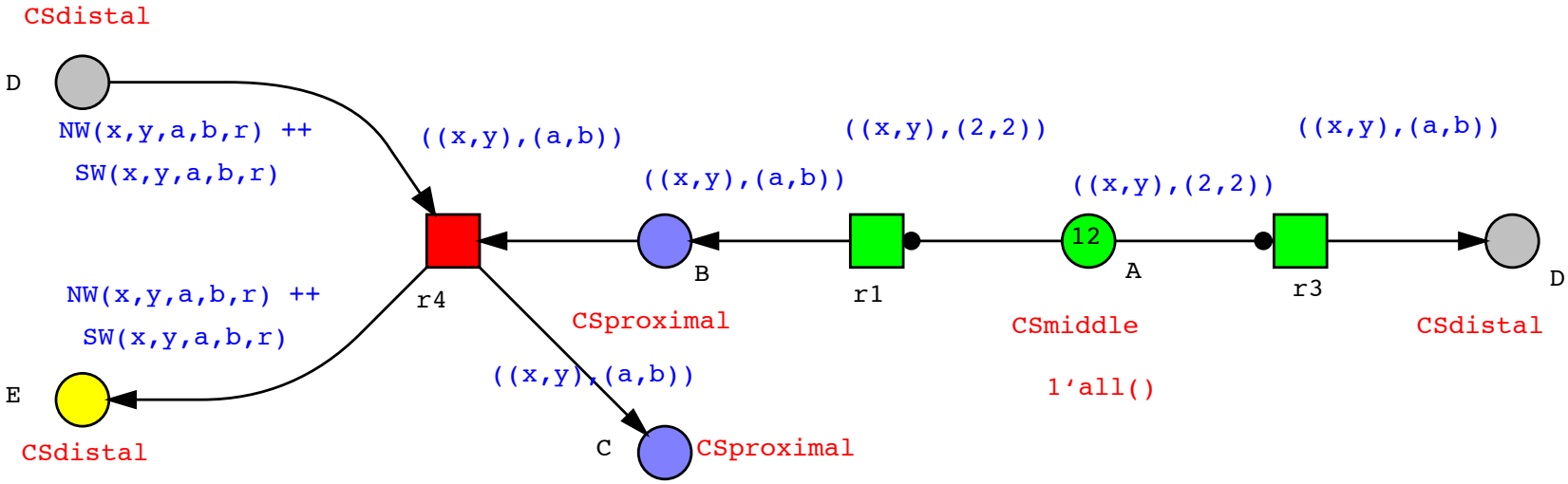
Wing tissue tissue:
Cells with logical compartments



Petri net model for single cell



CPN model of cells with seven compartments in a 2-D matrix.

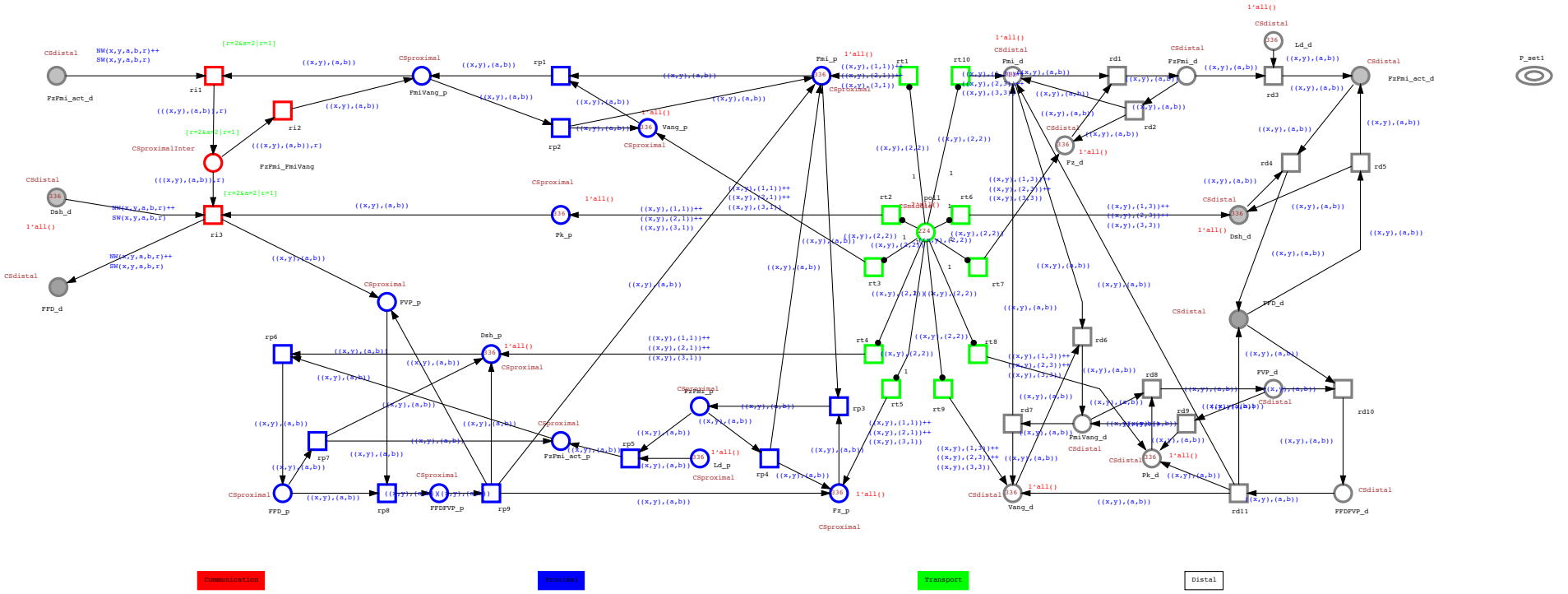


- 4 spatial regions: communication, proximal, transport and distal
- Seven virtual compartments ((1, 1), (2, 1),..., (3,3)).
- Each place or transition belongs to a specific compartment.
- NW and SW denote two left neighbours of the current cell.

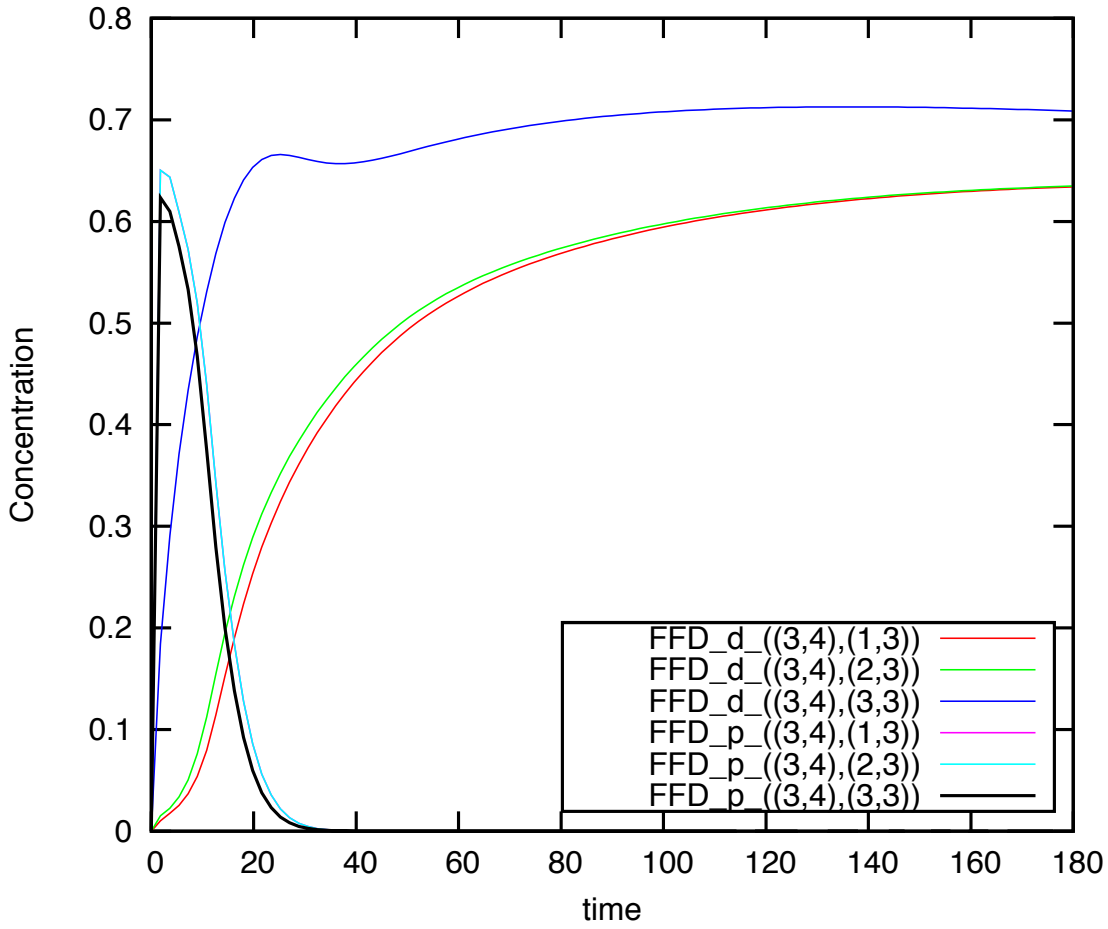
Declarations for the coloured Petri net model

Type	Declaration
con	M = int with 15;
con	N = int with 15;
con	R = int with 3;
con	C = int with 3;
cs	Row = int with $1 - M$;
cs	$Column$ = int with $1 - N$;
cs	$CS1$ = product with $Row \times Column$;
cs	CS_Cell = $CS1$ with $x\%2 = 1\&y\%2 = 0 \mid x\%2 = 0\&y\%2 = 1$;
cs	$ComR$ = int with $1 - R$;
cs	$ComC$ = int with $1 - C$;
cs	CS_ComP = product with $ComR \times ComC$;
cs	$CS2$ = product with $CS\_Cell \times CS\_ComP$;
cs	$CSdistal$ = $CS2$ with $b = 3$;
cs	$CSproximal$ = $CS2$ with $b = 1$;
cs	$CSmiddle$ = $CS2$ with $a = 2\&b = 2$;
cs	$CSInter$ = int with $1 - 2$;
cs	$CS3$ = product with $CSproximal \times CSInter$;
cs	$CSproximalInter$ = $CS3$ with $r = 2\&a = 2 \mid r = 1$;
var	x : Row;
var	y : $Column$;
var	a : $ComR$;
var	b : $ComC$;
var	r : $CSInter$;
fun	$CSproximal\ NW$ ($Row\ x$,$Column\ y$,$ComR\ a$,$ComC\ b$,$CSInter\ r$) {[(!(x=1|y=1))&(r=1&a=1&b=1|r=2&a=2&b=1)] ((x-1,y-1),(a+1,b+2));}
fun	$CSproximal\ SW$ ($Row\ x$,$Column\ y$,$ComR\ a$,$ComC\ b$,$CSInter\ r$) {[(!(x=M|y=1))&(r=2&a=2&b=1|r=1&a=3&b=1)] ((x+1,y-1),(a-1,b+2));}
fun	$bool\ MutReg(Row\ x$,$Column\ y)$ {x>=4&x<=8&y>=3&y<=7;}

CPN model of PCP signalling

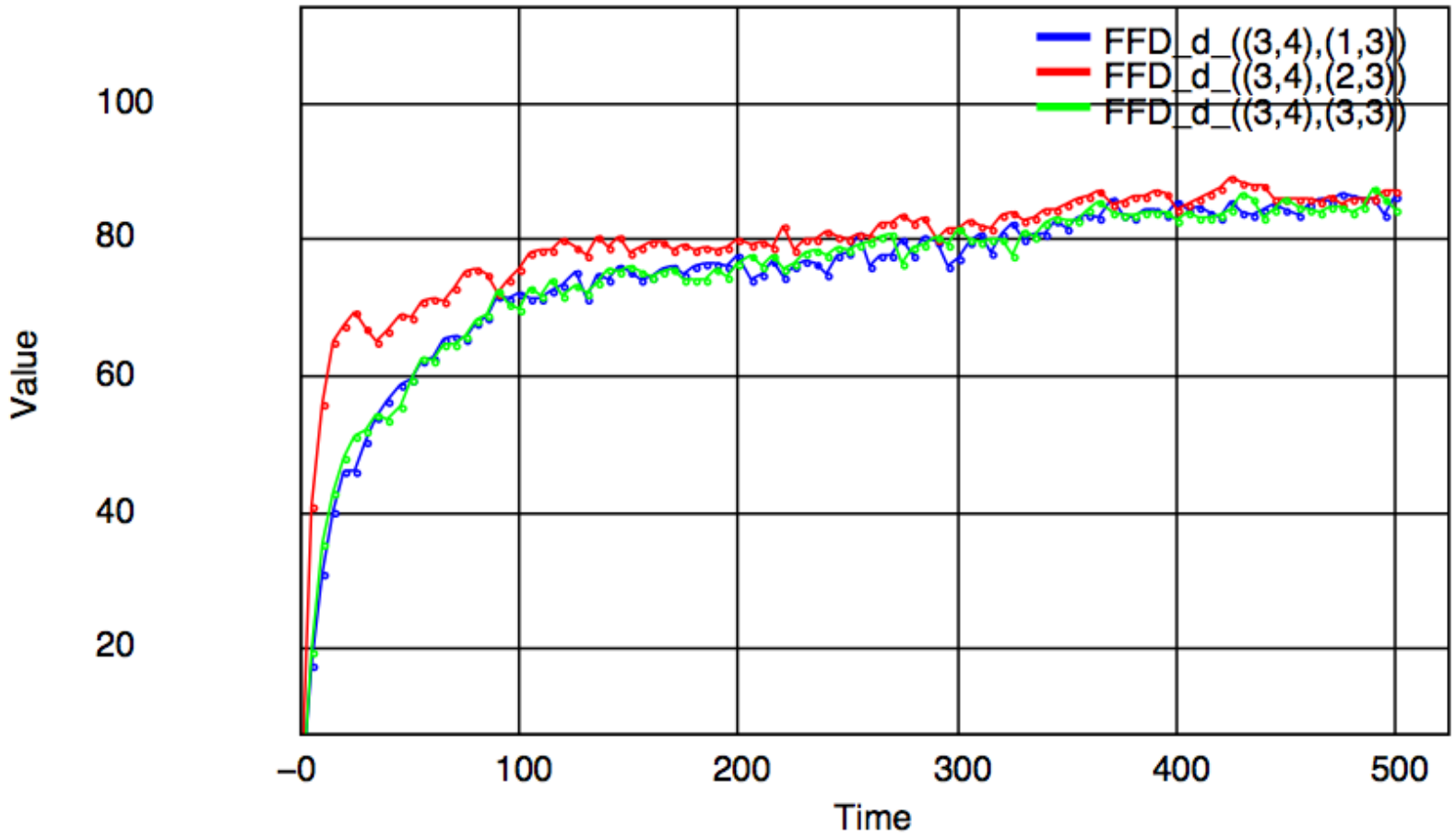


FFD in one cell (3,4)
Continuous simulation



FFD accumulates at the distal edge of the cell rather than the proximal edge at the end of signalling.

Stochastic simulation (average of 10 runs)

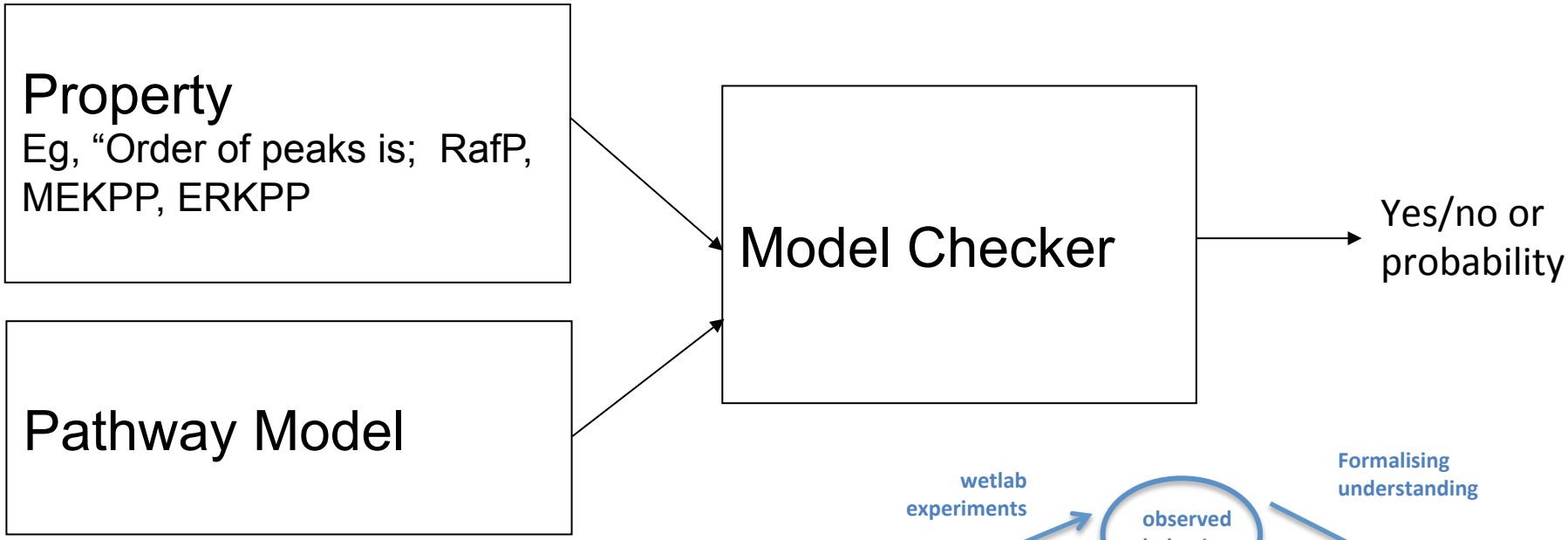


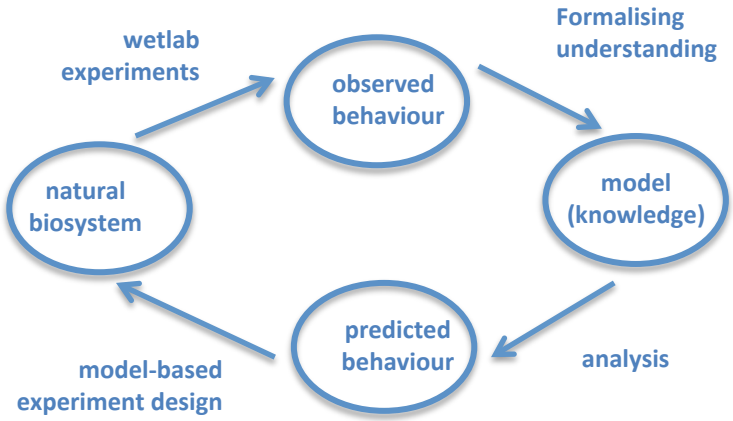
Some statistics

Size			
Grid(M× N)	Cells	Places	Transitions
5×5	12	924	984
10×10	50	3,850	4,100
15×15	112	8,624	9,184
20×20	200	15,400	16,400
50×50	1,250	96,250	102,500

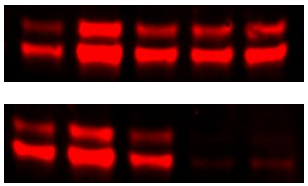
Time (seconds)				
Grid(M× N)	Unfolding	Unfolding/Cells	Simulation	Simulation/Cells
5×5	0.99	0.0825	13.34	1.1117
10×10	3.46	0.0692	235.81	4.7162
15×15	8.04	0.0718	1,366.24	12.1986
20×20	15.52	0.0776	-	-
50×50	161.48	0.1292	-	-

Model Checking Biochemical Pathways





Simulation-based Model Checking Biochemical Pathways



Property

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

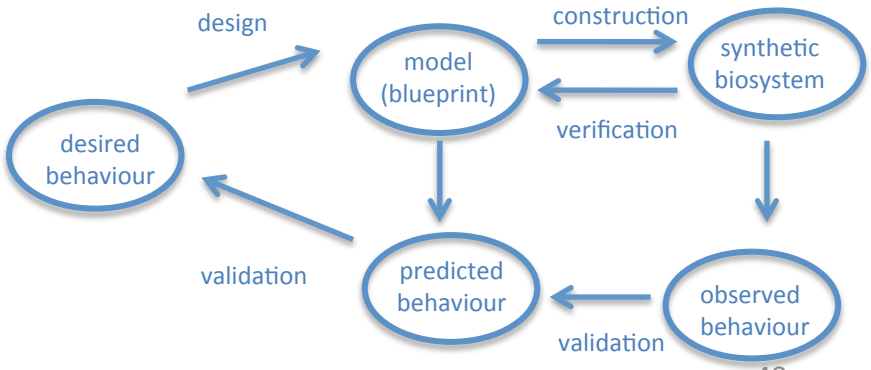
Time series data

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))]

Model searching

Peaks at least once

(rises then falls below 50% max concentration)

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

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

Rises and remains constant

(99% max concentration)

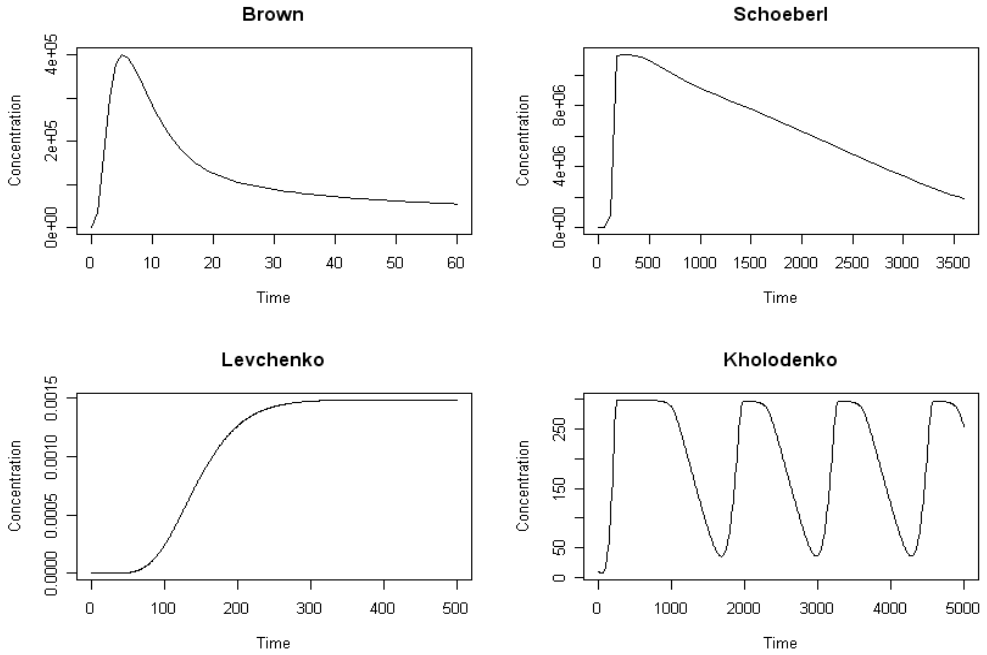
$P_{>=1}[ErkPP <= 0.50*max(ErkPP) \wedge (\ d(ErkPP) > 0\ )\ U\ (\ G(ErkPP >= 0.99*max(ErkPP))\ )\ ]$

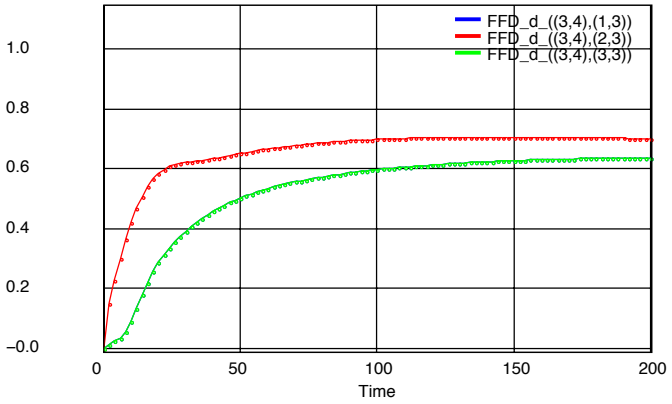
- ***Levchenko***

Oscillates at least 4 times

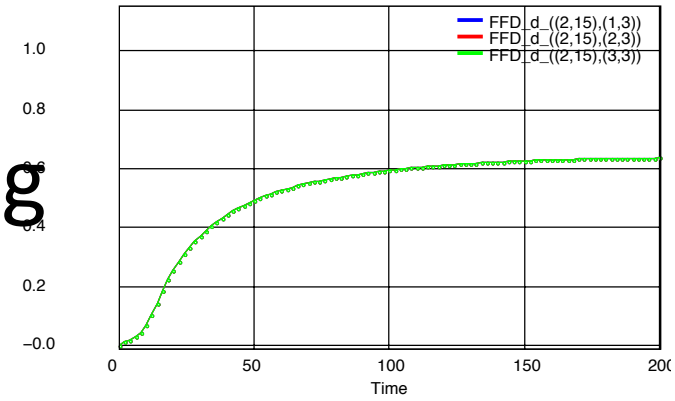
$P_{>=1}[\ F(\ d(ErkPP) > 0 \wedge F(\ d(ErkPP) < 0 \wedge \ldots\ )\ )\ ]$

- **Kholodenko**





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)]&[(3,3)] > [C]))]

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).

Model checking

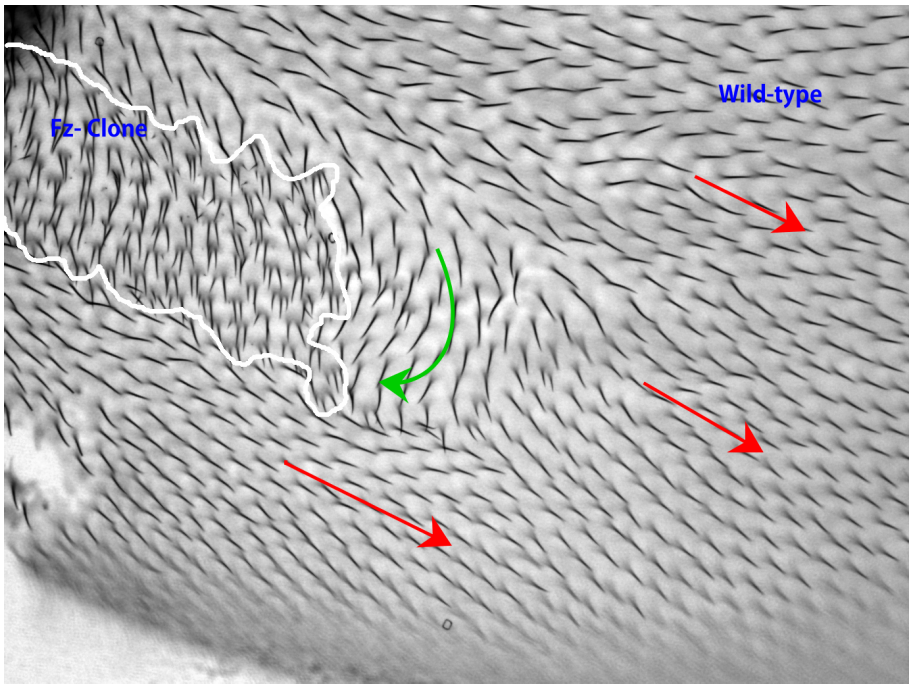
For FFD in each distal compartment of each cell, how many peaks exist in their traces?

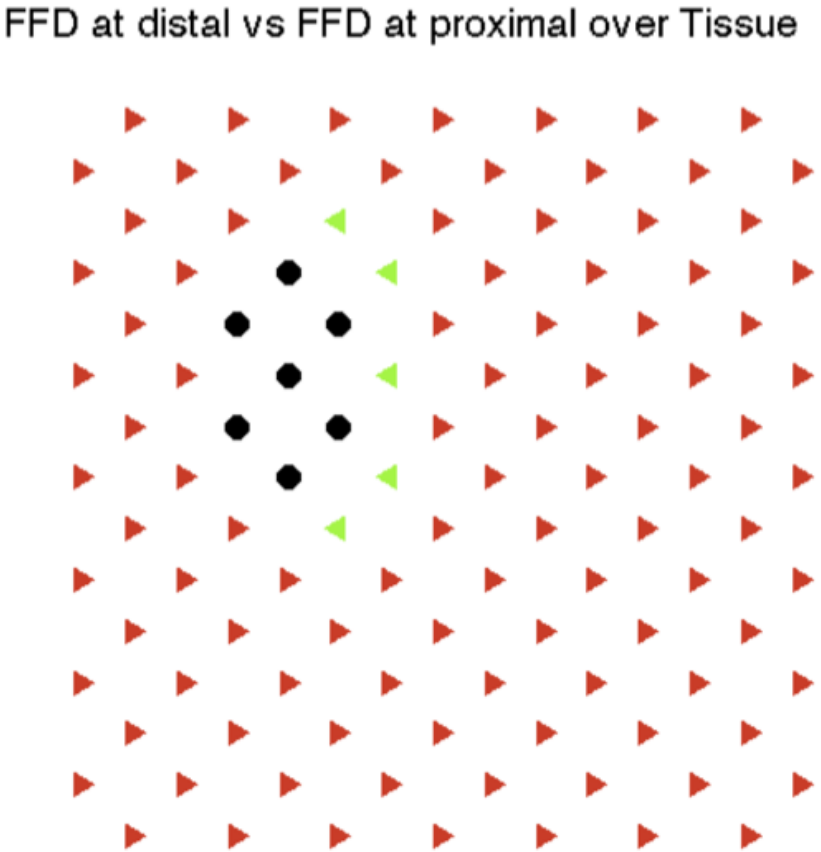
P =?[F ((d[x(2,3)] > 0)&F ((d[x(2,3)] < 0)&F ((d[x(2,3)] > 0)))]
P =?[F ((d[x(2,3)] > 0)&F ((d[x(2,3)] < 0)))]
P =?[F ((d[x(2,3)] > 0))]

P=?[F((d[FFD:(3,4),(1,3)] > 0) & F((d[FFD:(3,4),(1,3)] < 0)))]; Probability: 0.0
P=?[F((d[FFD:(3,4),(2,3)] > 0) & F((d[FFD:(3,4),(2,3)] < 0)))]; Probability: 1.0
P=?[F((d[FFD:(3,4),(3,3)] > 0) & F((d[FFD:(3,4),(3,3)] < 0)))]; Probability: 0.0

- No (2,3) peak in cells in Row 1, Row 15 and Column 15
- Except these boundary cells, other cells have only one peak
- For FFD in other distal compartments, there are no peaks.

Fz clone in WT background





Acknowledgements

Brunel:

Pam Gao

David Tree

Cottbus:

Monika Heiner

Fei Liu

PhD positions available!

david.gilbert@brunel.ac.uk