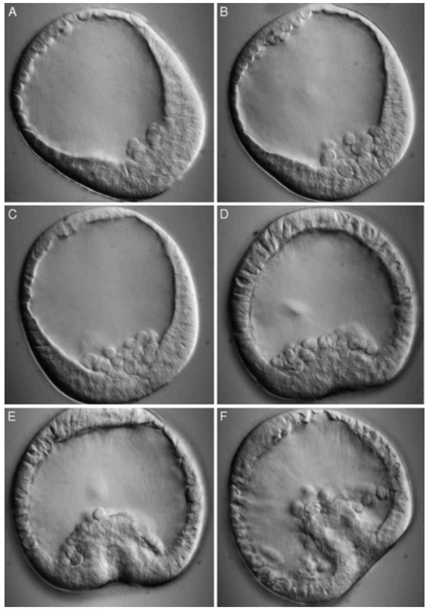
Inaugural Lecture Series 2011

Professor David Gilbert

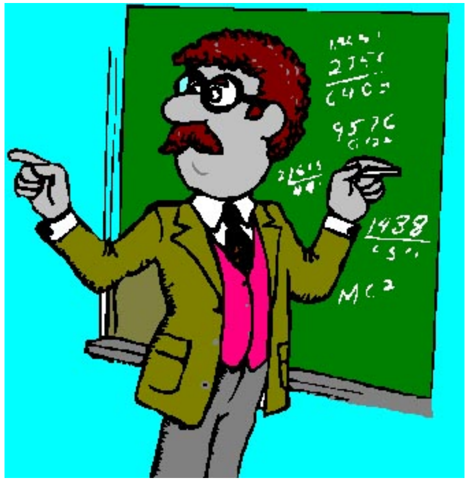
"Computing for biology – can it really help?"

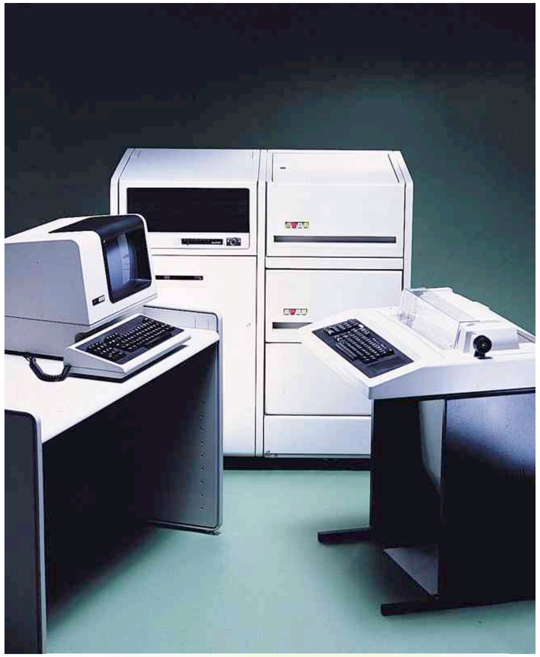
School of Information Systems, Computing & Mathematics
Centre for Systems & Synthetic Biology

Brunel
UNIVERSITY
WEST LONDON

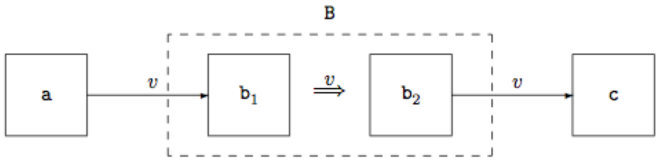

A
B
C
D
E
F

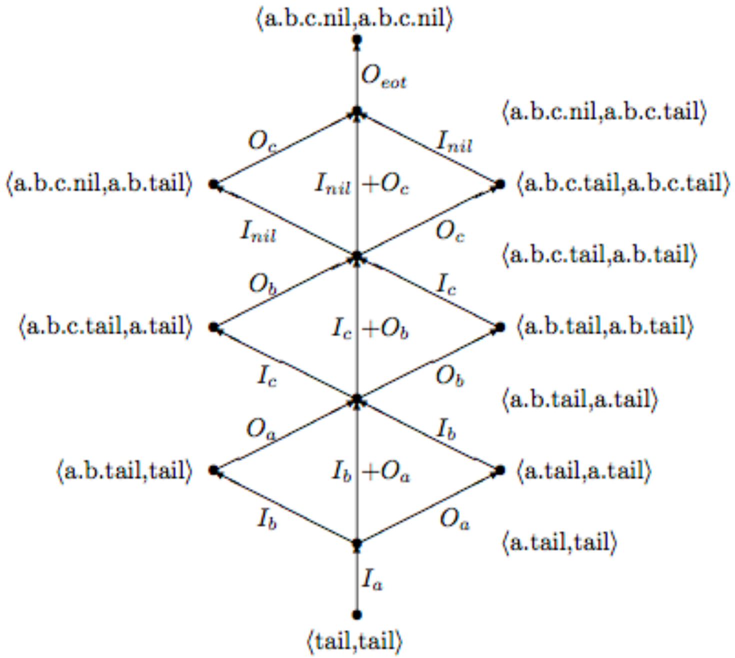
School teacher…

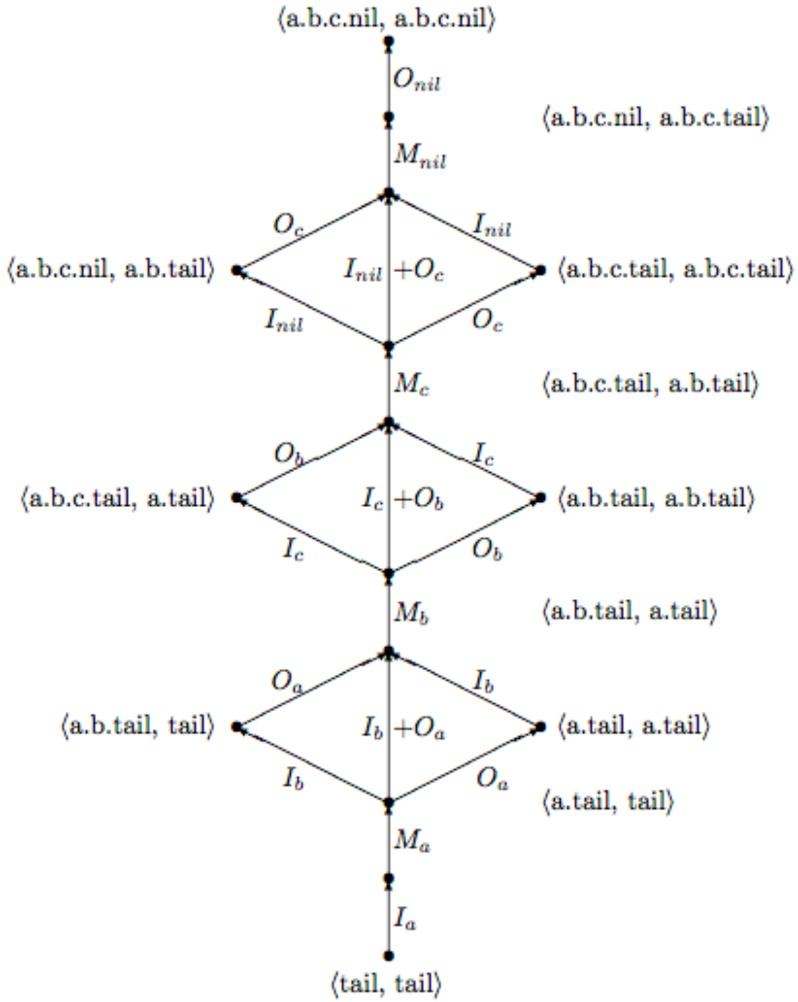




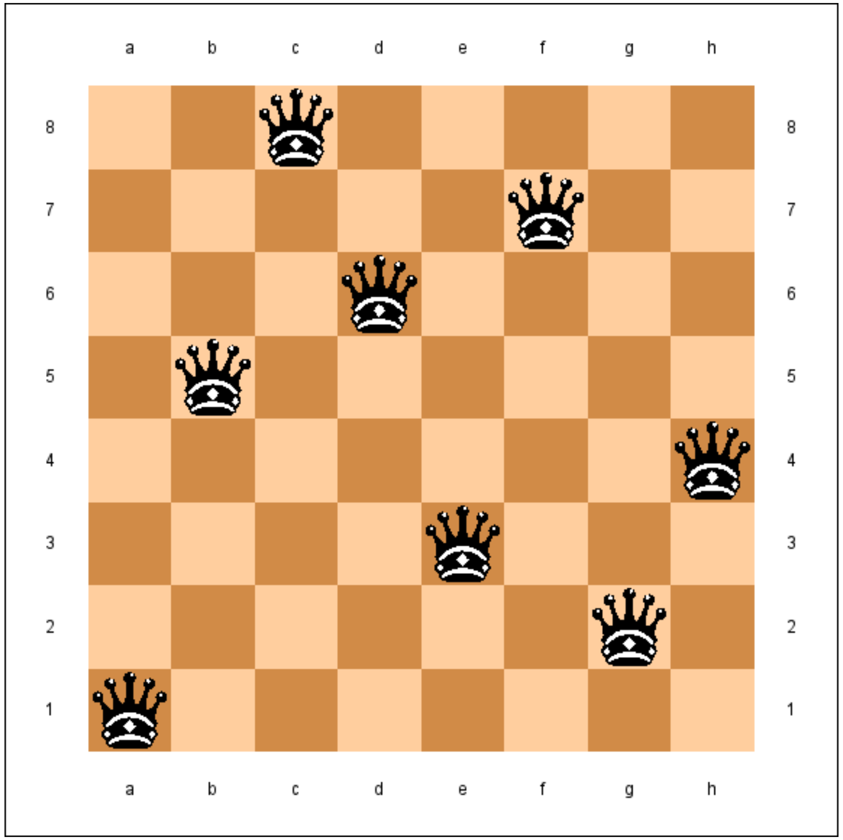
PhD: Specifying and Reasoning about Concurrent Systems in Logic

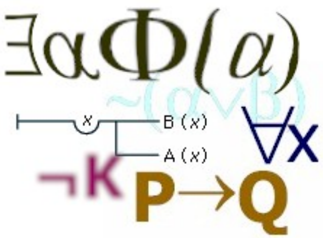






… solving really useful problems
‘8 queens’

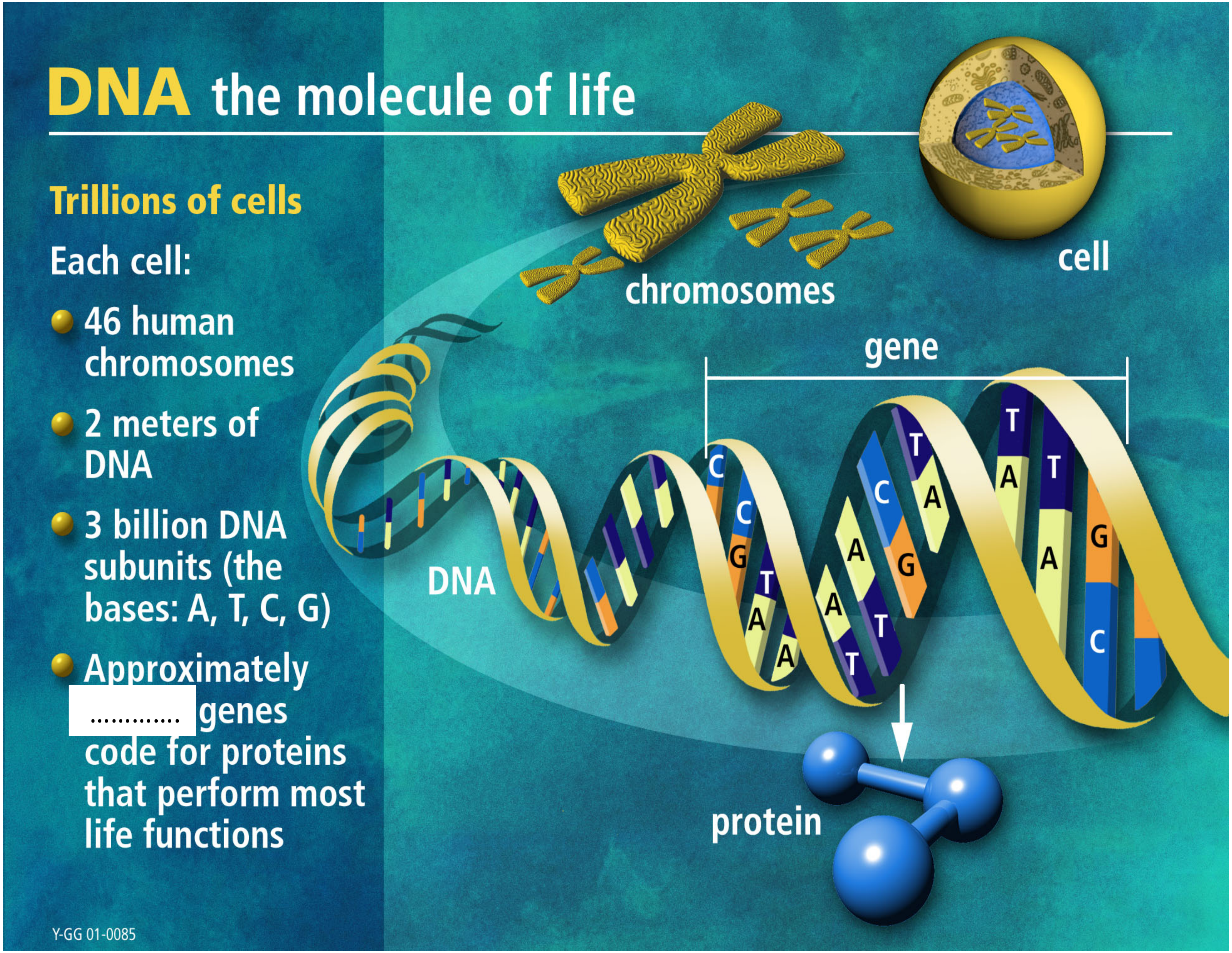




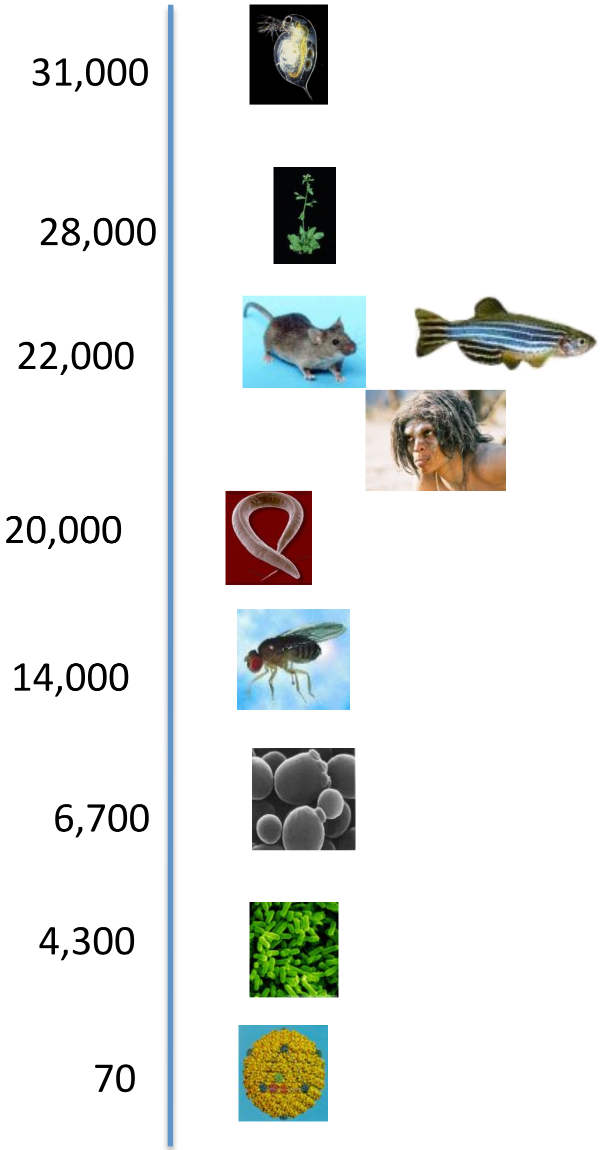
Bob
Kowalski

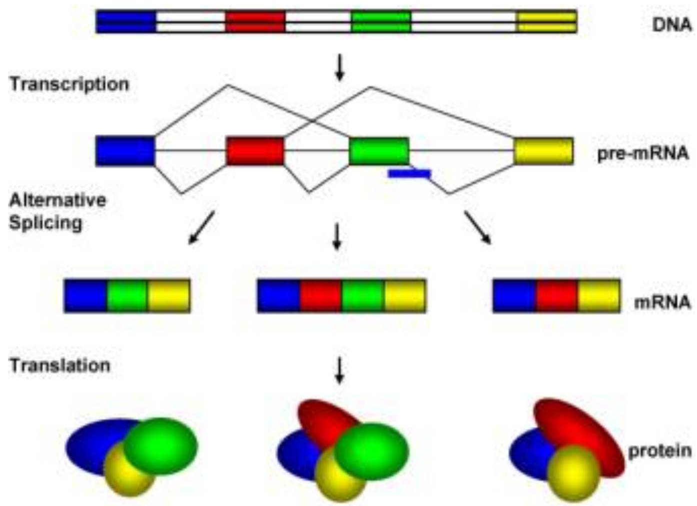
… 'dining philosphers'

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

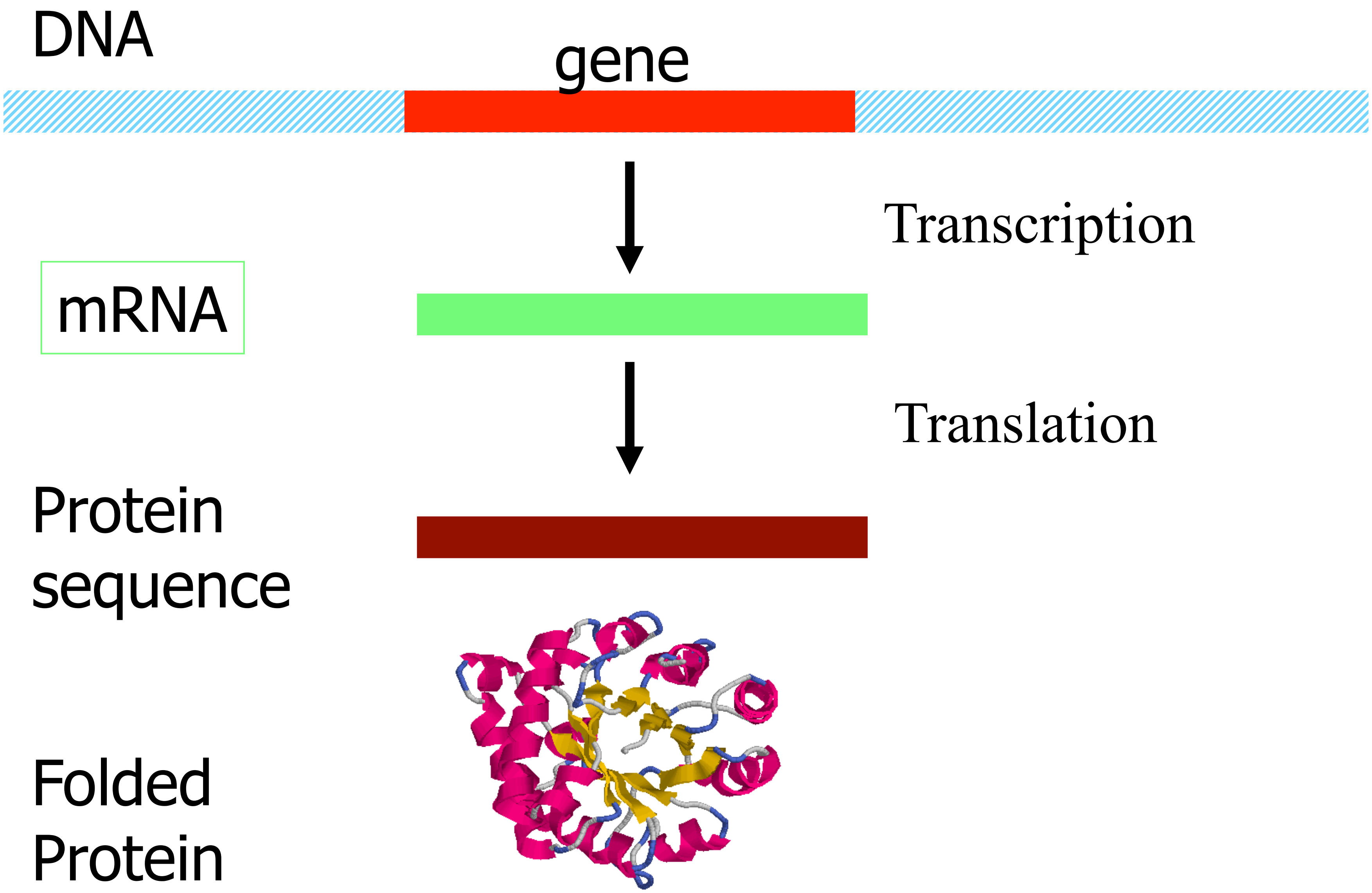


More genes – better?



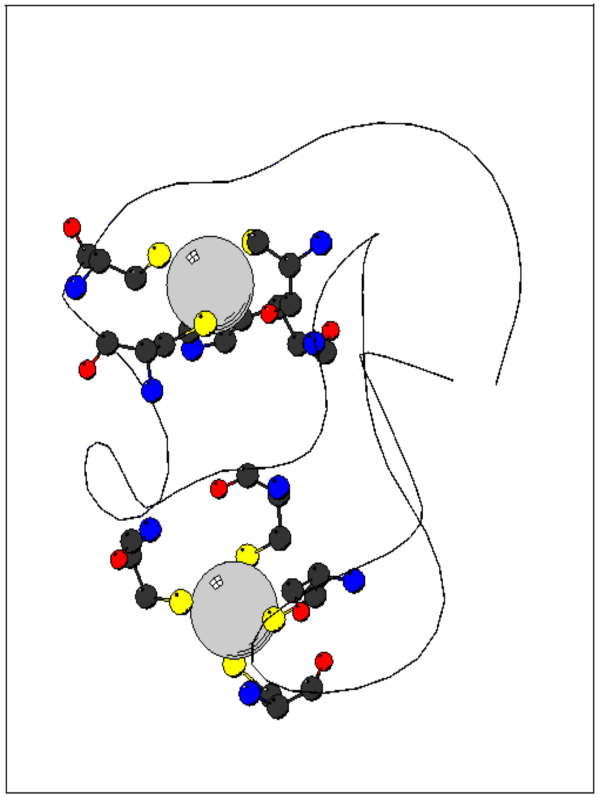


Gene structure



… Patterns …

```
      xxx
   V     x
  x        x
  x        x
  x        x
   C     H
  x \   / x
 x    Zn    x
  x /    \  x
   C       H
xxxx         xxxxxx
```



C-x(2,4)-C-x(3)-[ILVMFYWC]-x(8)-H-x(3,5)-H

Brazma, I. Jonassen, I. Eidhammer and D.R. Gilbert. Approaches to the automatic discovery of patterns in biosequences, Journal of Computational Biology 1998

Stem loops

(1)

```
  c
 a-u
 g-c
 u-a
 c-g
augg   ggcau
```

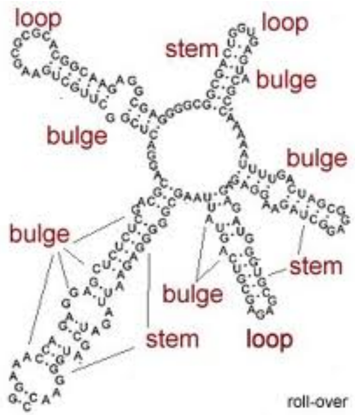
(2)

```
  g
 u-a
 a-u
 g-c
 c-g
aggc   ccgu
```

(1) augg<u>cuga</u>c<u>ucag</u>ggcau

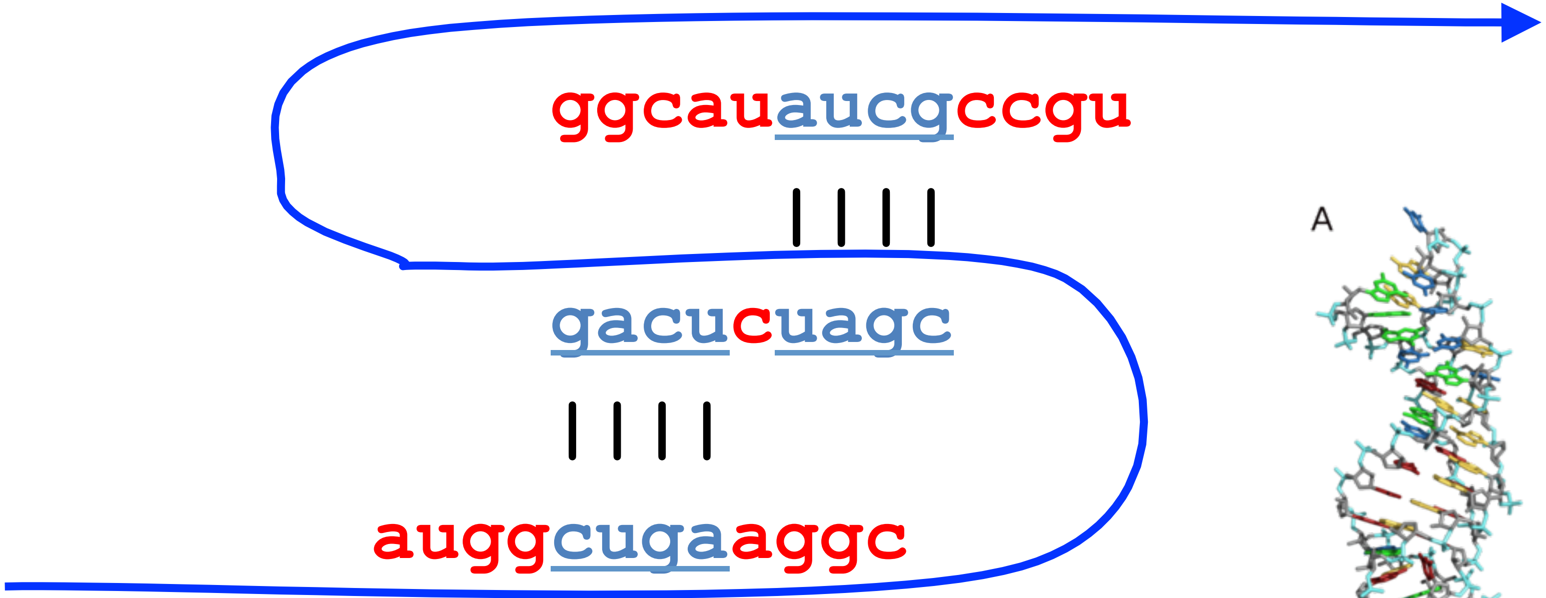
(2) aggc<u>cgau</u>g<u>aucg</u>ccgu

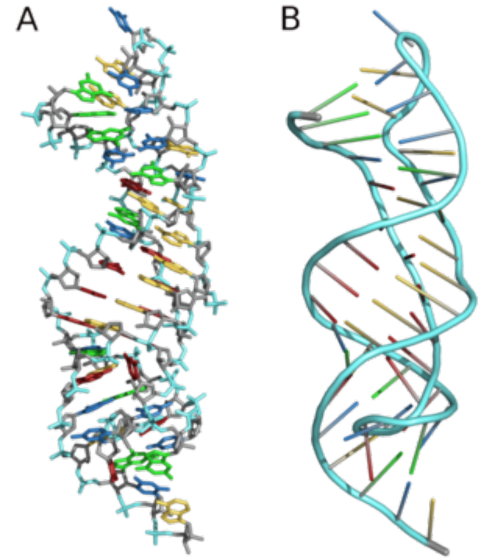
$\alpha \quad \beta \quad \alpha^{rc}$



Eidhammer, Jonassen, Grinhang, Gilbert & Ratnayake, A contraint-based structure description language for biosequences, Journal of Constraints 6:2/3, 2001

Pseudo-knot

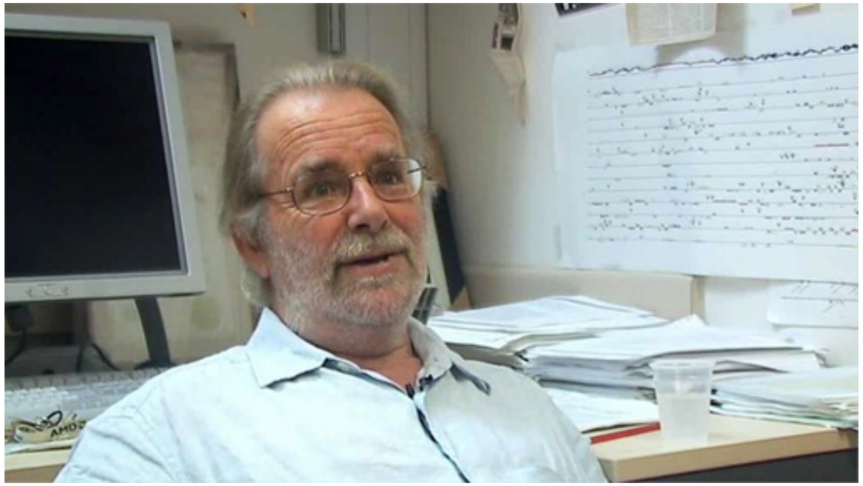




String:

auggcugaaggccgaucucagggcauaucgccgu

α γ_1 β γ_2 α^{rc} γ_3 β^{rc}



Mike Ashburner

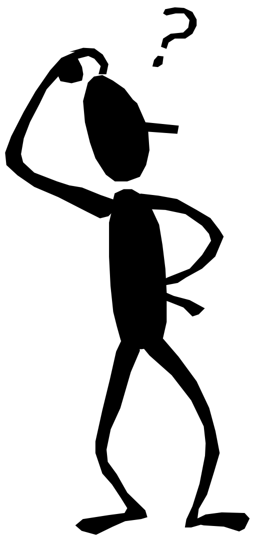
The European Bioinformatics Institute

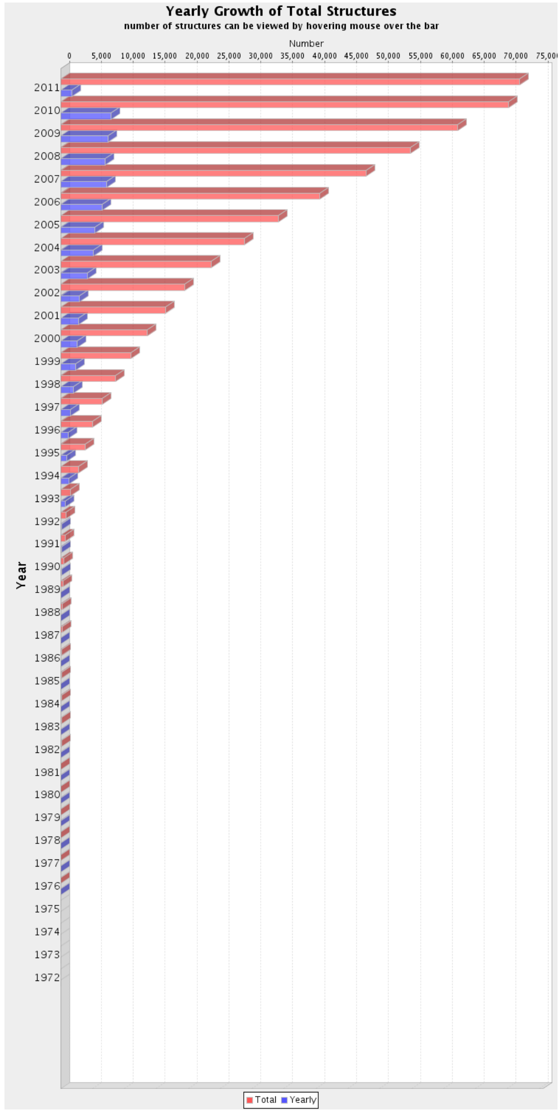
Janet Thornton

EPSRC Visiting Research Fellow at the European Bioinformatics Institute: "Formalising TOPS cartoons, and the development of an interactive similarity search mechanism over TOPS databases” 1998

David Westhead

PDB: Protein Data Bank



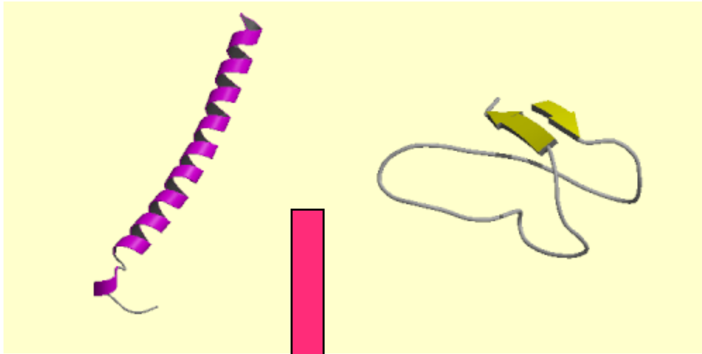


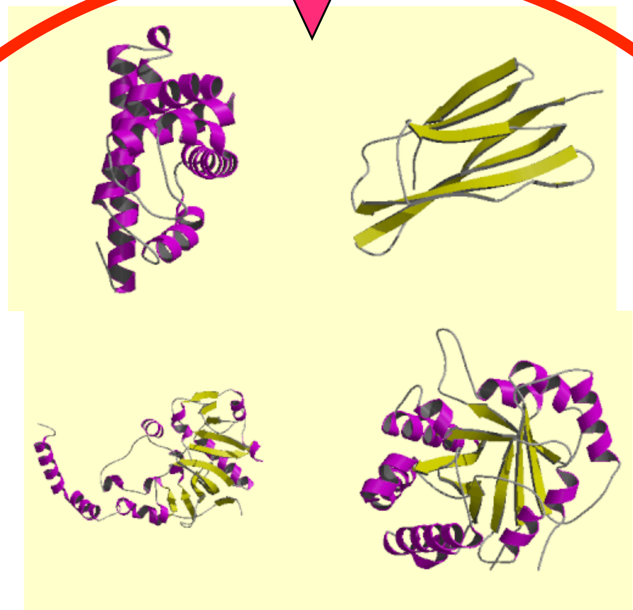
Protein structure - levels

~~PRIMARY STRUCTURE (amino acid sequence)~~

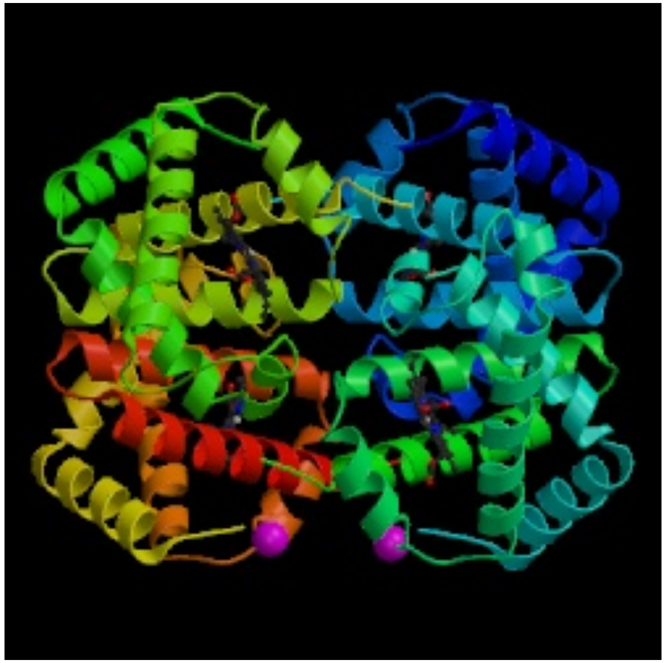
~~VHLTPEEKSAVTALWGKVNVDE~~
~~VGGEALGRLLVVYPWTQRFFE~~
~~SFGDLSTPDAVMGNPKVKAHG~~
~~KKVLGAFSDGLAHLDNLKGTFA~~
~~TLSELHCDKLHVDPENFRLLGN~~
~~VLVCVLAHHFGKEFTPPVQAAY~~
~~QKVVAGVANALAHKYH~~

SECONDARY STRUCTURE (helices, strands)



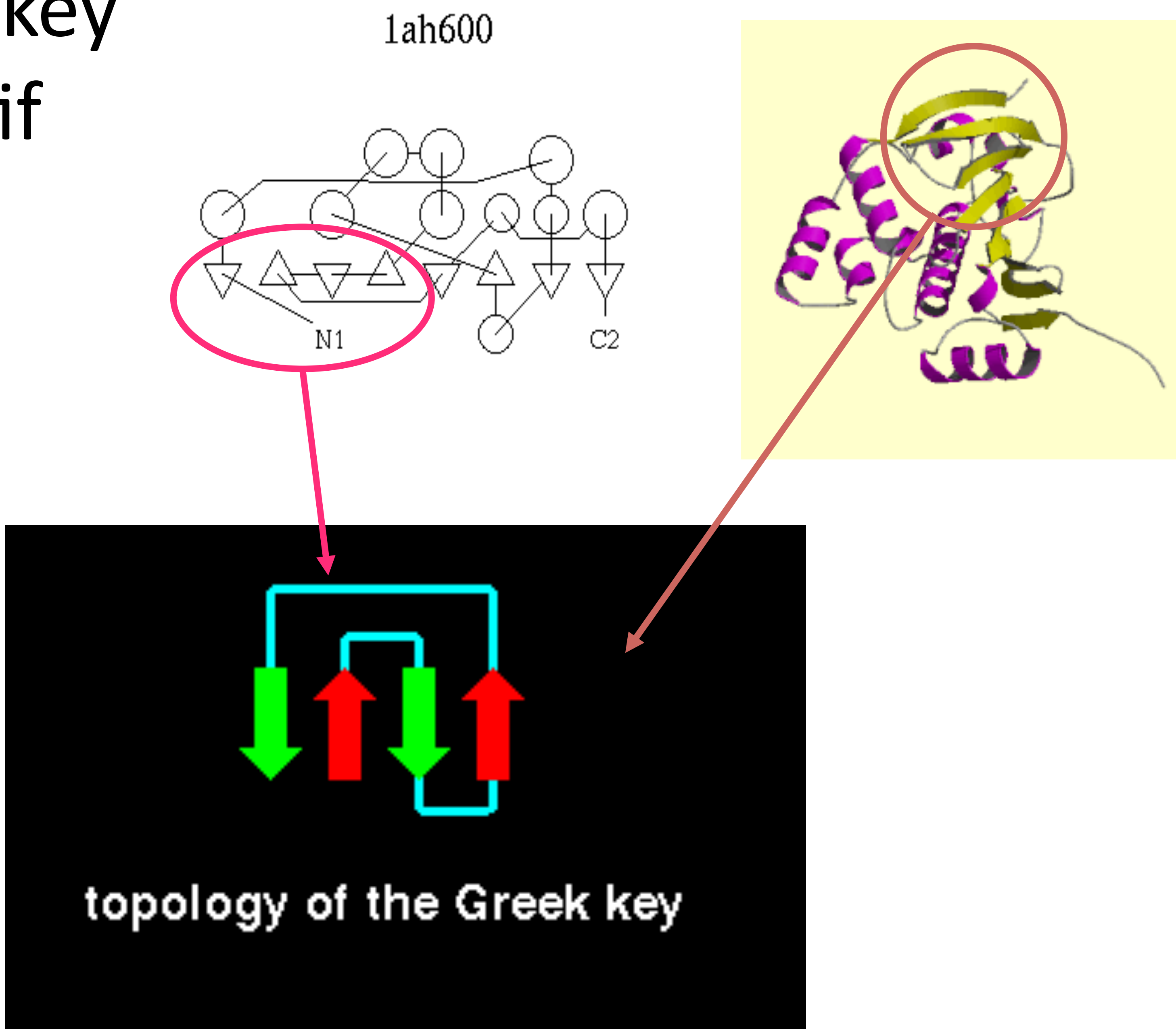


TERTIARY STRUCTURE (fold)

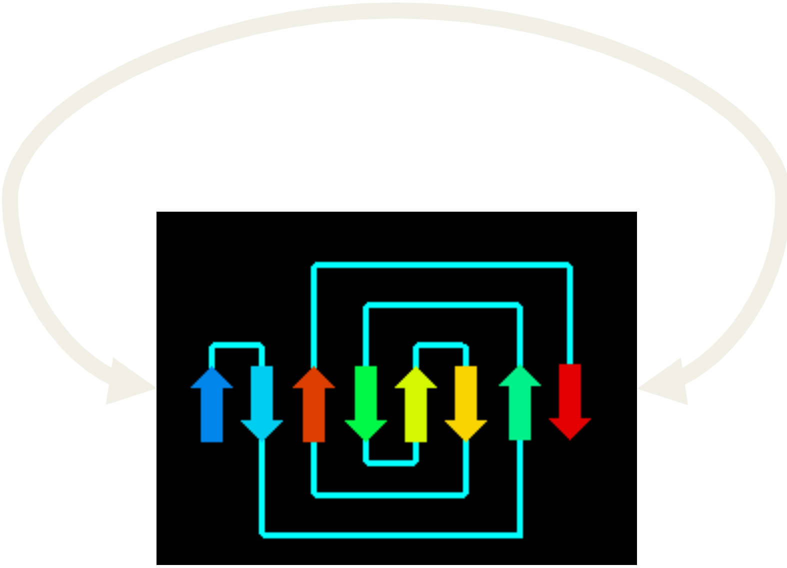


QUATERNARY STRUCTURE

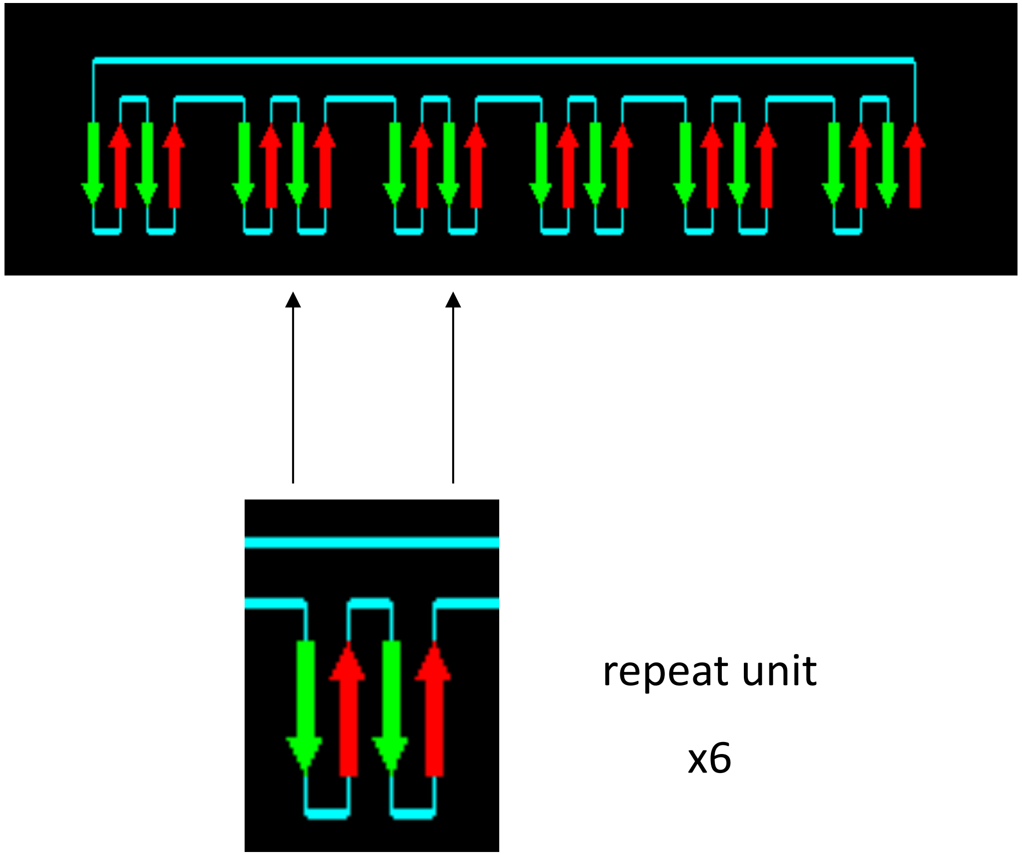
“Greek key” motif



Jelly roll



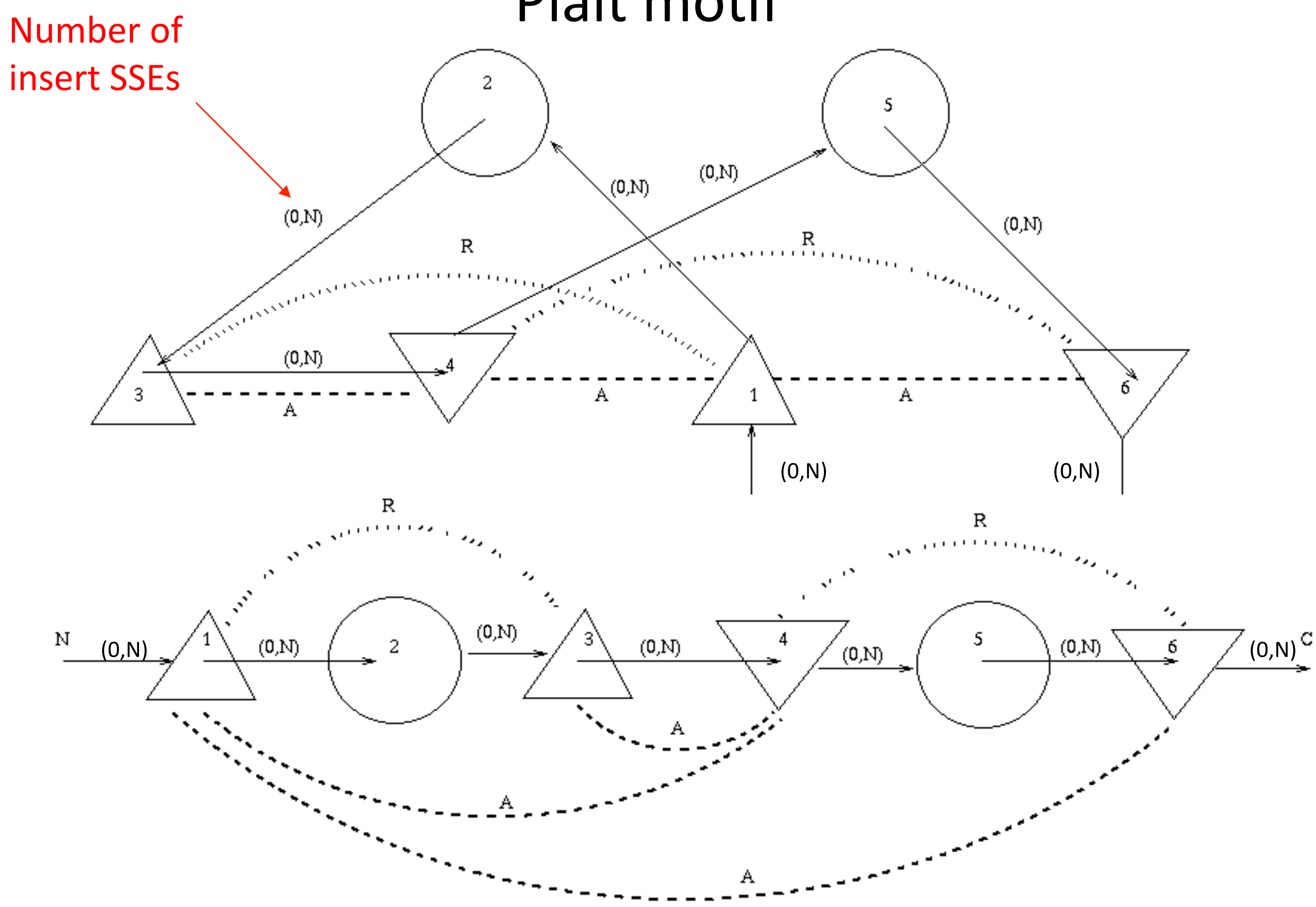
Beta propellor



repeat unit

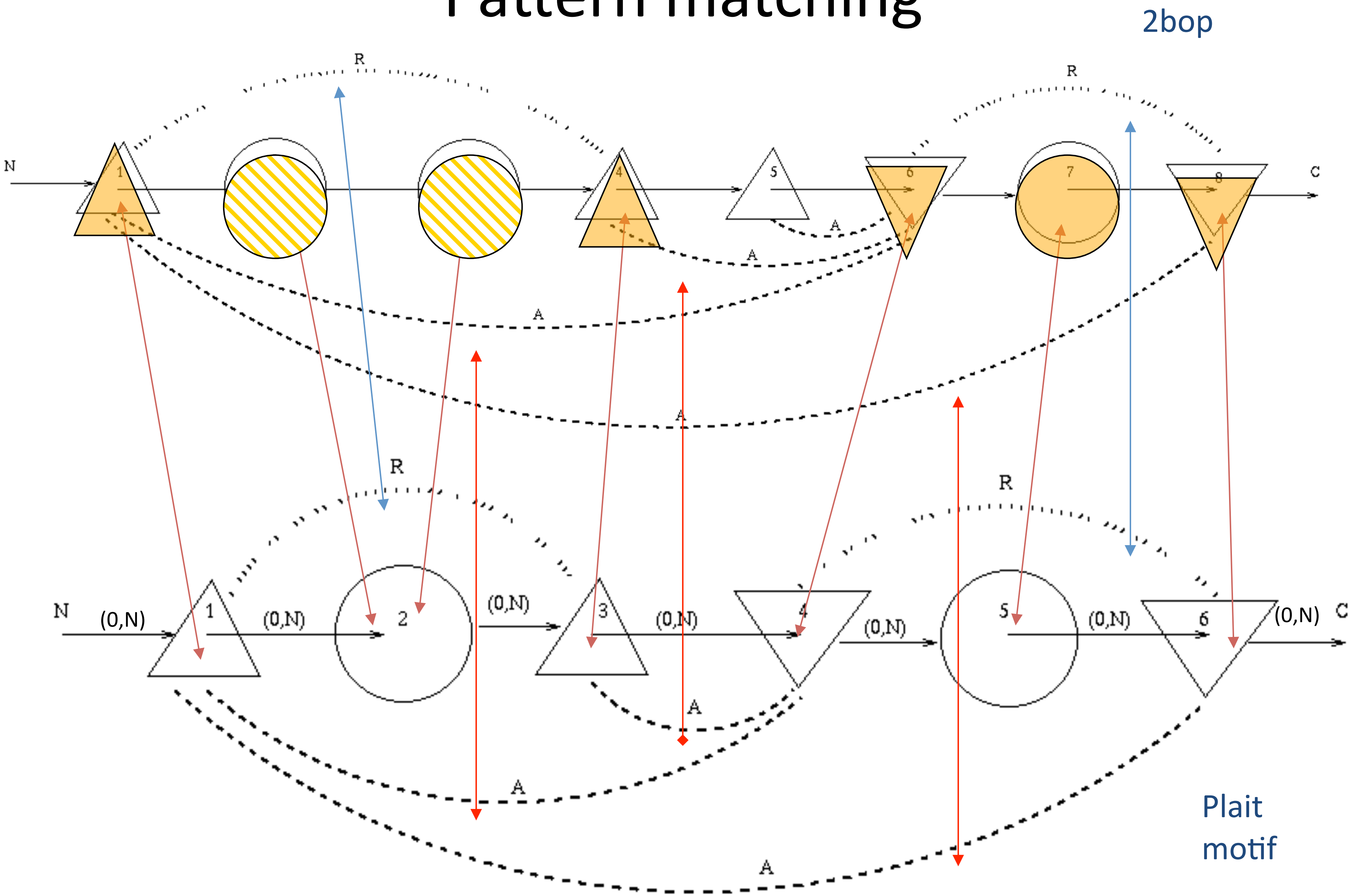
x6

Plait motif




```
Plait E*h*E*e*H*e 1:3R 1:4A 1:6A 3:4A 4:6R
```

Pattern matching



Motif matching: TIM Barrel

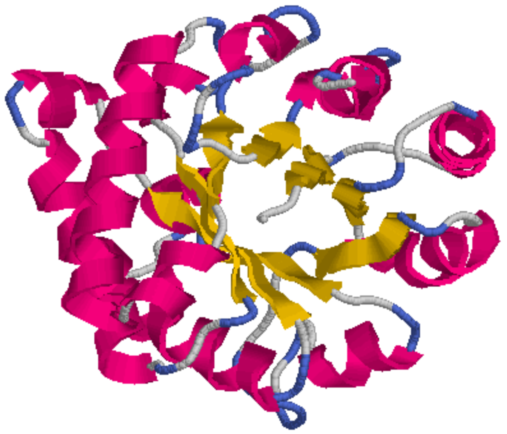
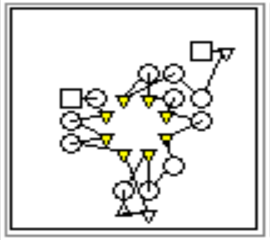
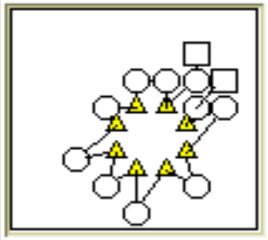
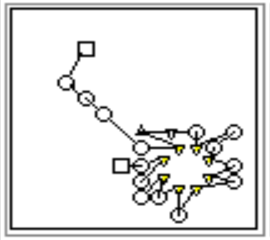
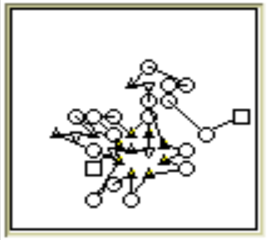
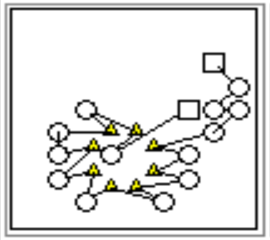
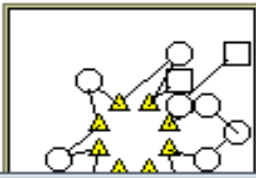
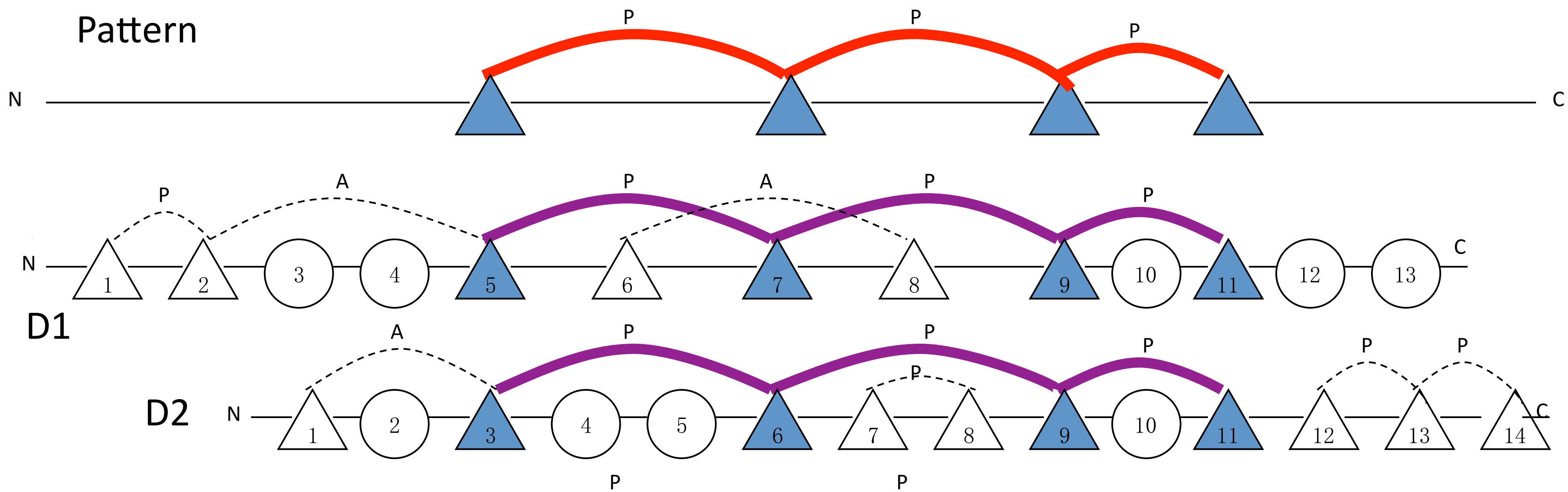


Image	Name	Classification
	1qr7A0	3.20.20.280.1.1.1
	1bd0A1	3.20.20.10.1.1.1
	1aw500	3.20.20.230.1.1.1
	1wkf00	3.20.20.105.1.1.1
	1pymA0	3.20.20.270.1.1.1
	1b5tA0	3.20.20.220.1.1.1

Compression

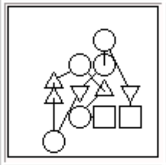
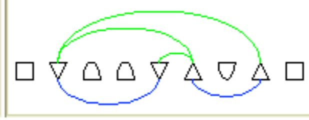
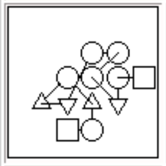
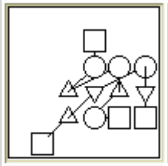
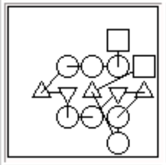
- Send the pattern once, and
- Send the **uncovered parts** of each structure
- Range: 0 (no similarity) to 1 (identical)



Special case: When 2 examples, compression gives *comparison*

Fast protein structure comparison

Pairwise comparison of cath with target: pdb2bop_A0 

Cartoon	Diagram	Compression	Name	Classification
		0.9166667	1vhiA0	3.30.70.390.1.1.1
		0.8833333	1qaxA1	3.30.70.420.1.1.1
		0.8055556	5rubA1	3.30.70.150.1.1.1
		0.78846157	1f0xA1	3.30.70.610.1.1.1
		0.77380955	1f3vA0	3.30.70.680.1.1.1

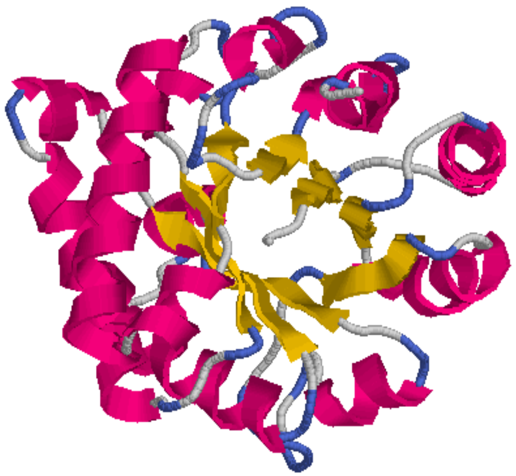
Genes to systems

DNA

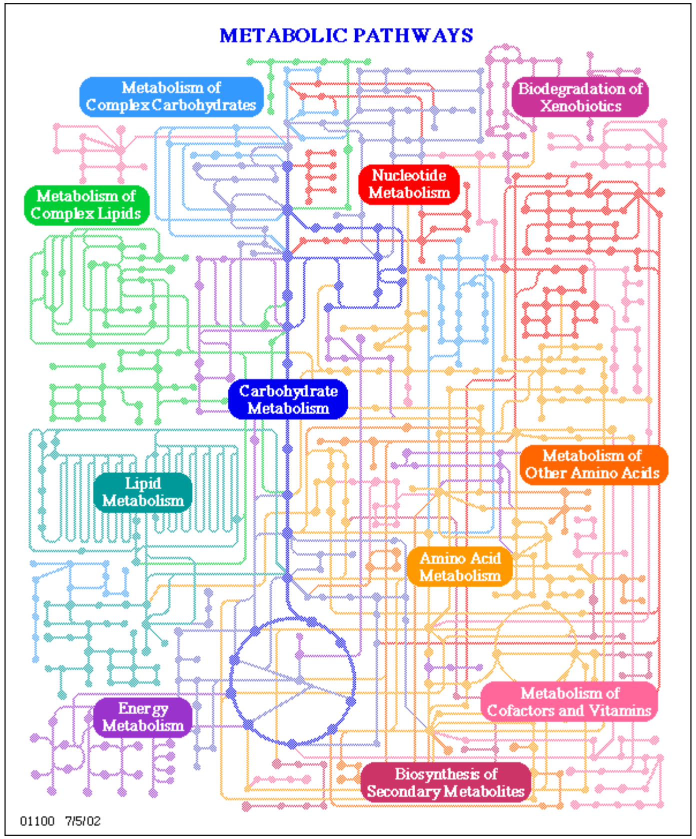
"gene"

mRNA

Protein sequence

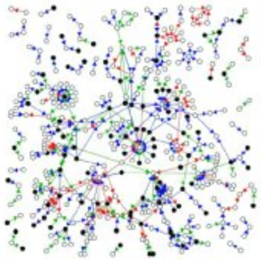


Folded Protein



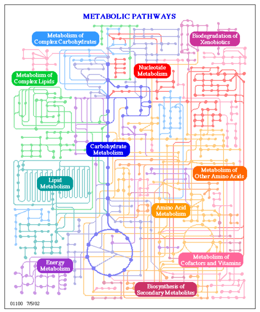
Networks

- Gene regulation



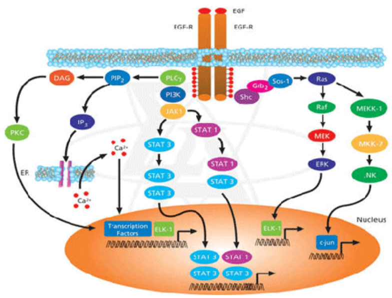
- Protein-protein interaction

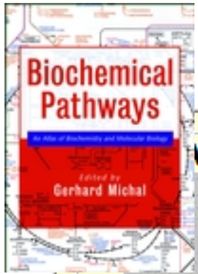
- Metabolic



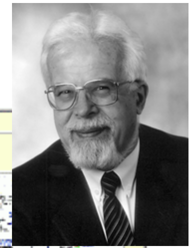
- Developmental

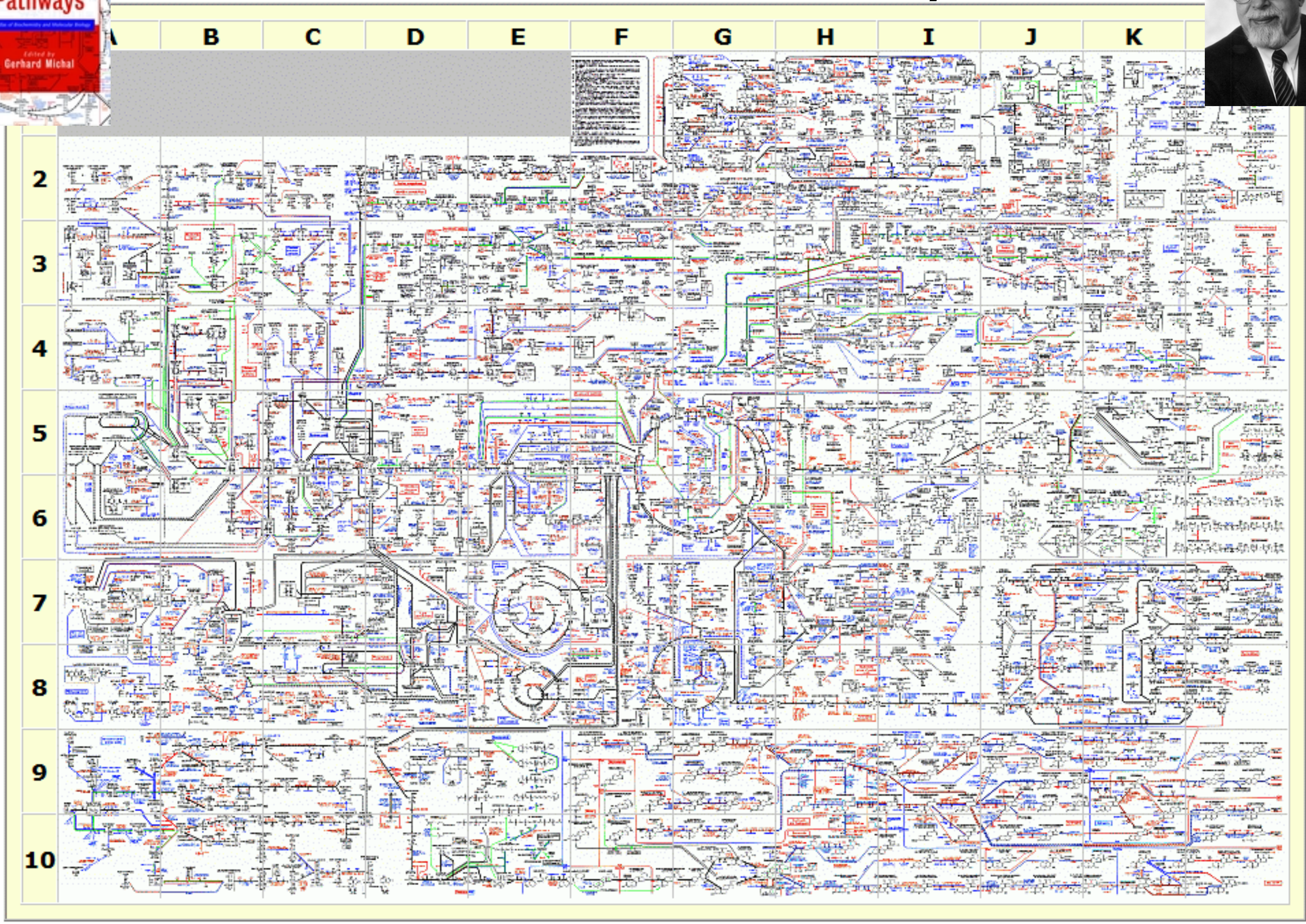
- Signalling





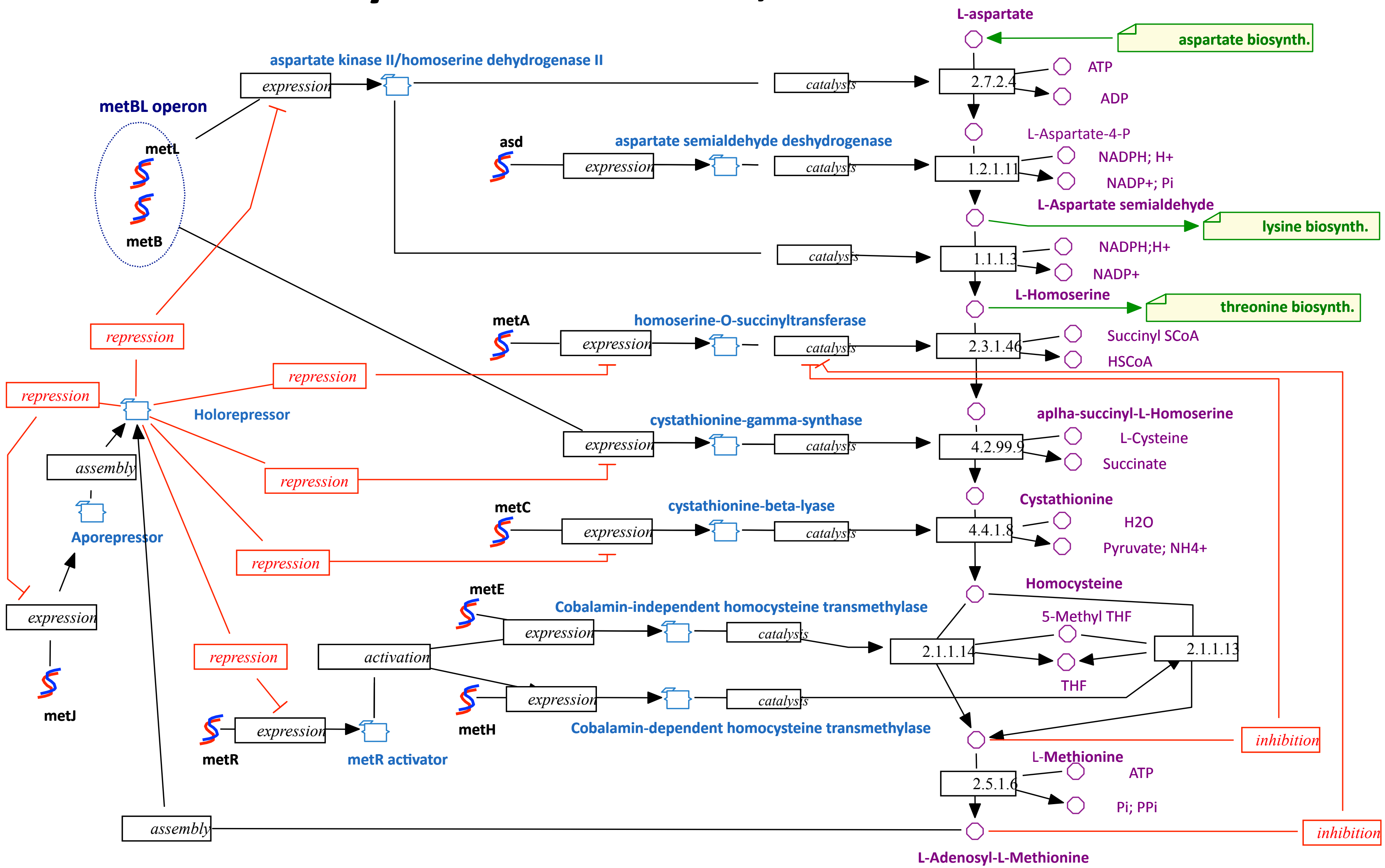
Metabolic Pathways





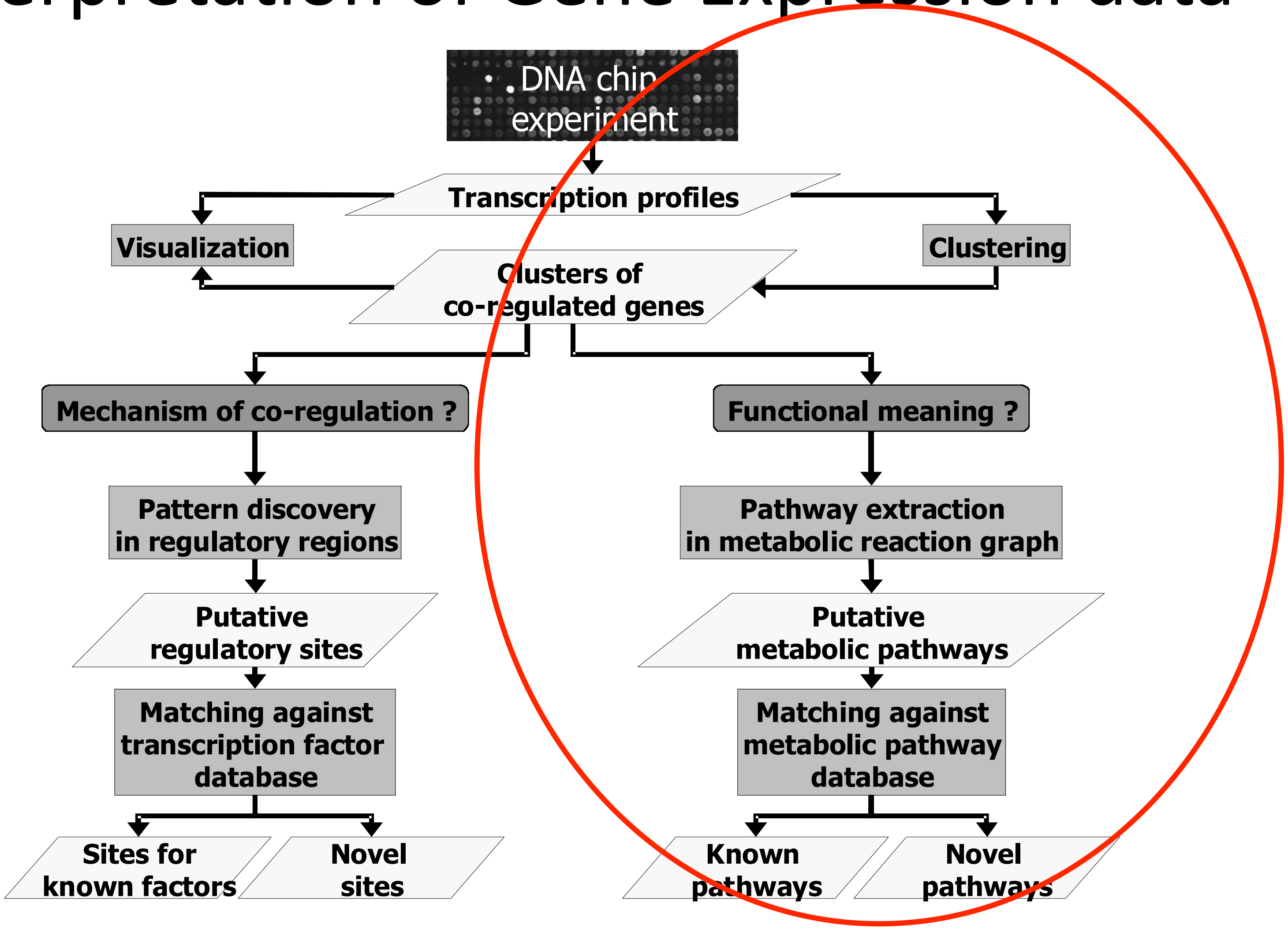
http://ca.expasy.org/tools/pathways/

Methionine Biosynthesis in E.coli, aMAZE data model



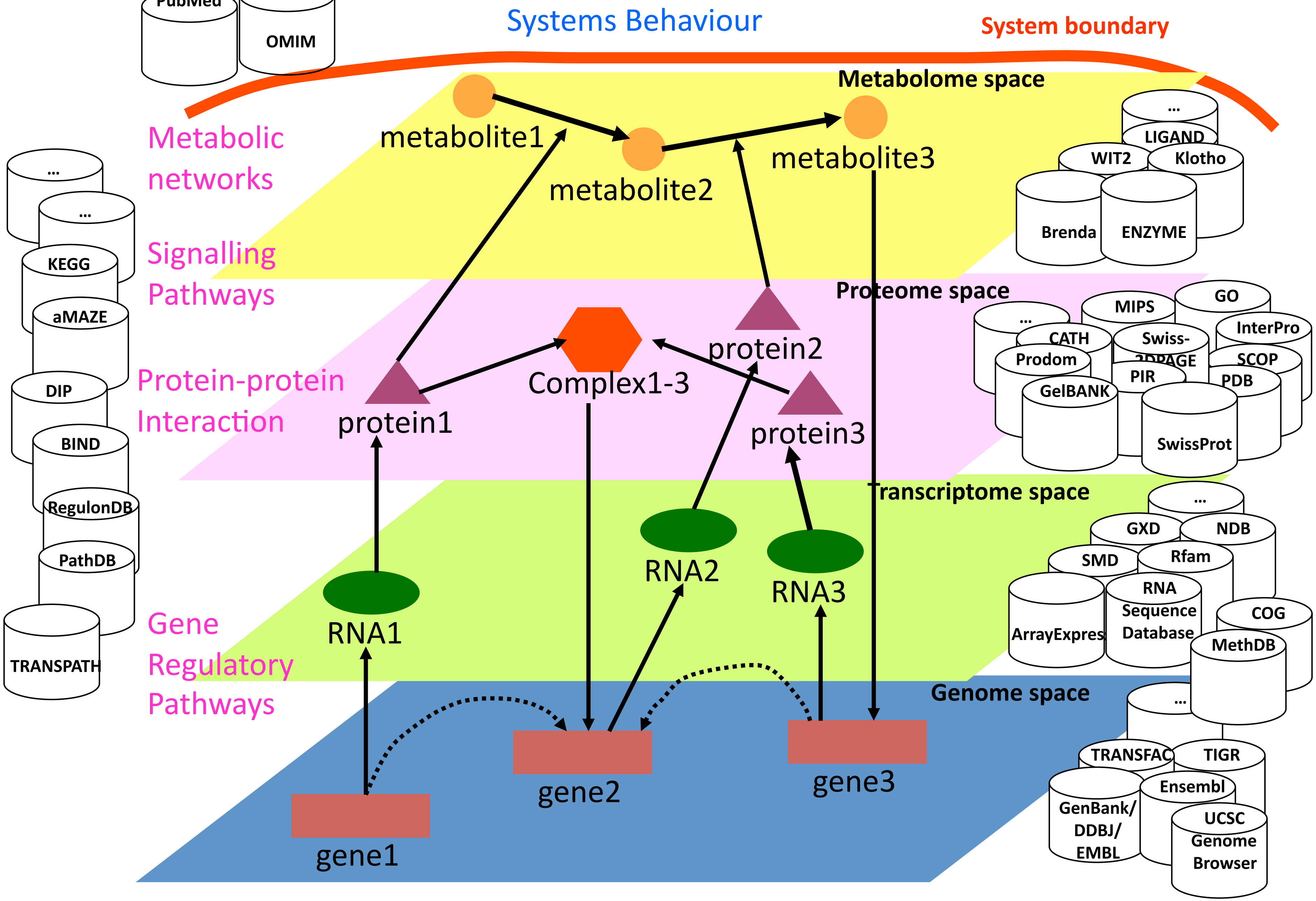
J van Helden, A Naim, R Mancuso, M Eldridge, L Wernisch, D Gilbert, and S Wodak (2000), Representing and analysing molecular and cellular function in the computer, Journal of Biological Chemistry, 381 (9-10):921-35.

Interpretation of Gene Expression data



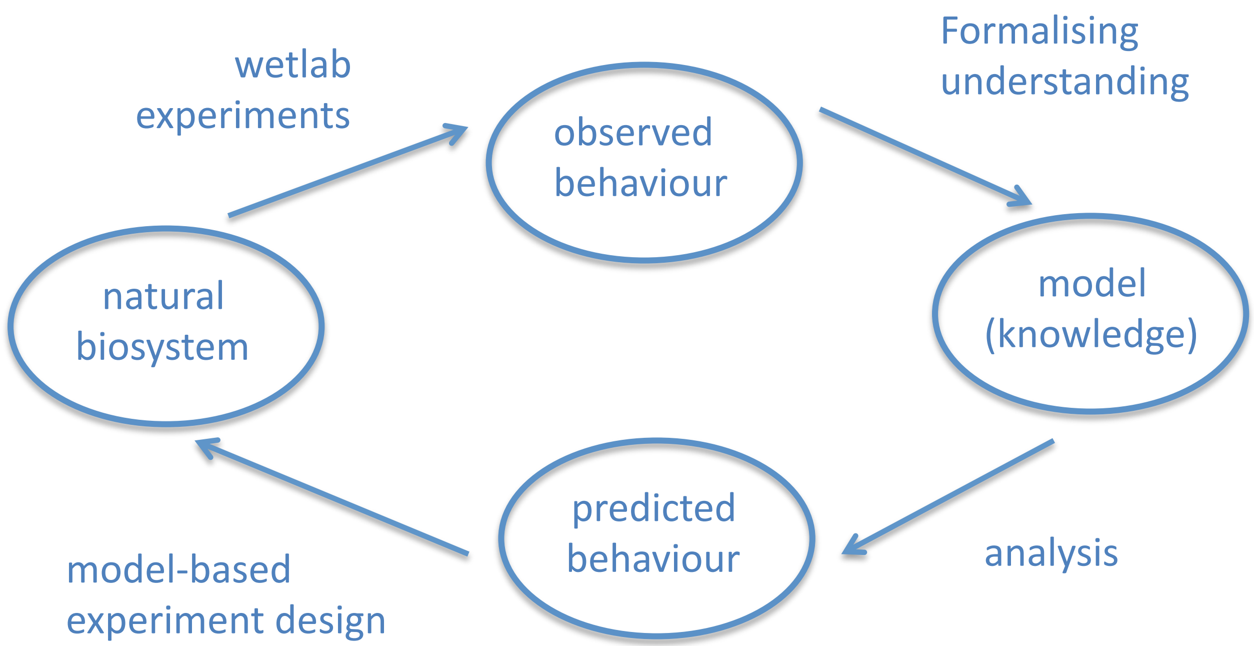
J van Helden, D Gilbert, L Wernisch, M Schroeder, and S Wodak (2001), Application of Regulatory Sequence Analysis and Metabolic Network Analysis to the Interpretation of Gene Expression Data, in Computational Biology,LNCS 2006, pp147-163, ISBN 3-540-42242-0.

Cell levels



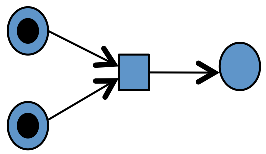
Systems Behaviour
System boundary
Literature
PubMed
OMIM
Metabolic networks
Signalling Pathways
Protein-protein Interaction
Gene Regulatory Pathways
Metabolome space
metabolite1
metabolite2
metabolite3
Proteome space
Complex1-3
protein1
protein2
protein3
Transcriptome space
RNA1
RNA2
RNA3
Genome space
gene1
gene2
gene3
...
...
KEGG
aMAZE
DIP
BIND
RegulonDB
PathDB
TRANSPATH
...
LIGAND
WIT2
Klotho
Brenda
ENZYME
...
MIPS
GO
CATH
Swiss-
InterPro
Prodom
SCOP
PIR
PDB
GelBANK
SwissProt
...
GXD
NDB
SMD
Rfam
RNA Sequence Database
ArrayExpres
COG
MethDB
...
TRANSFAC
TIGR
Ensembl
GenBank/ DDBJ/ EMBL
UCSC Genome Browser

Systems biology



Molecules/Levels
CTL, LTL

Qualitative



Time-free

Timed,
Quantitative

Abstraction

Abstraction

Monika Heiner

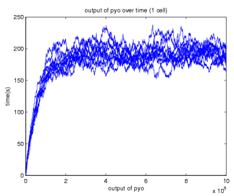
Approximation

Molecules/Levels
Stochastic rates
CSL

Stochastic

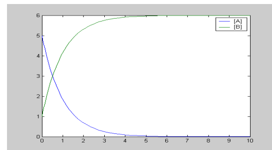
Continuous

Concentrations
Deterministic rates
LTLc



Approximation

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



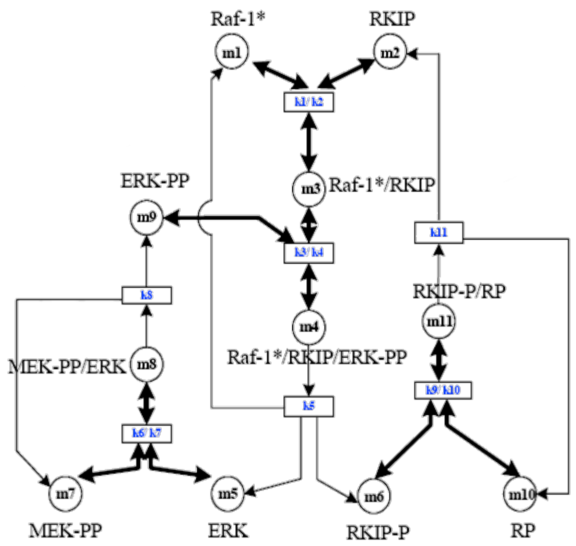
DiscreteState Space

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

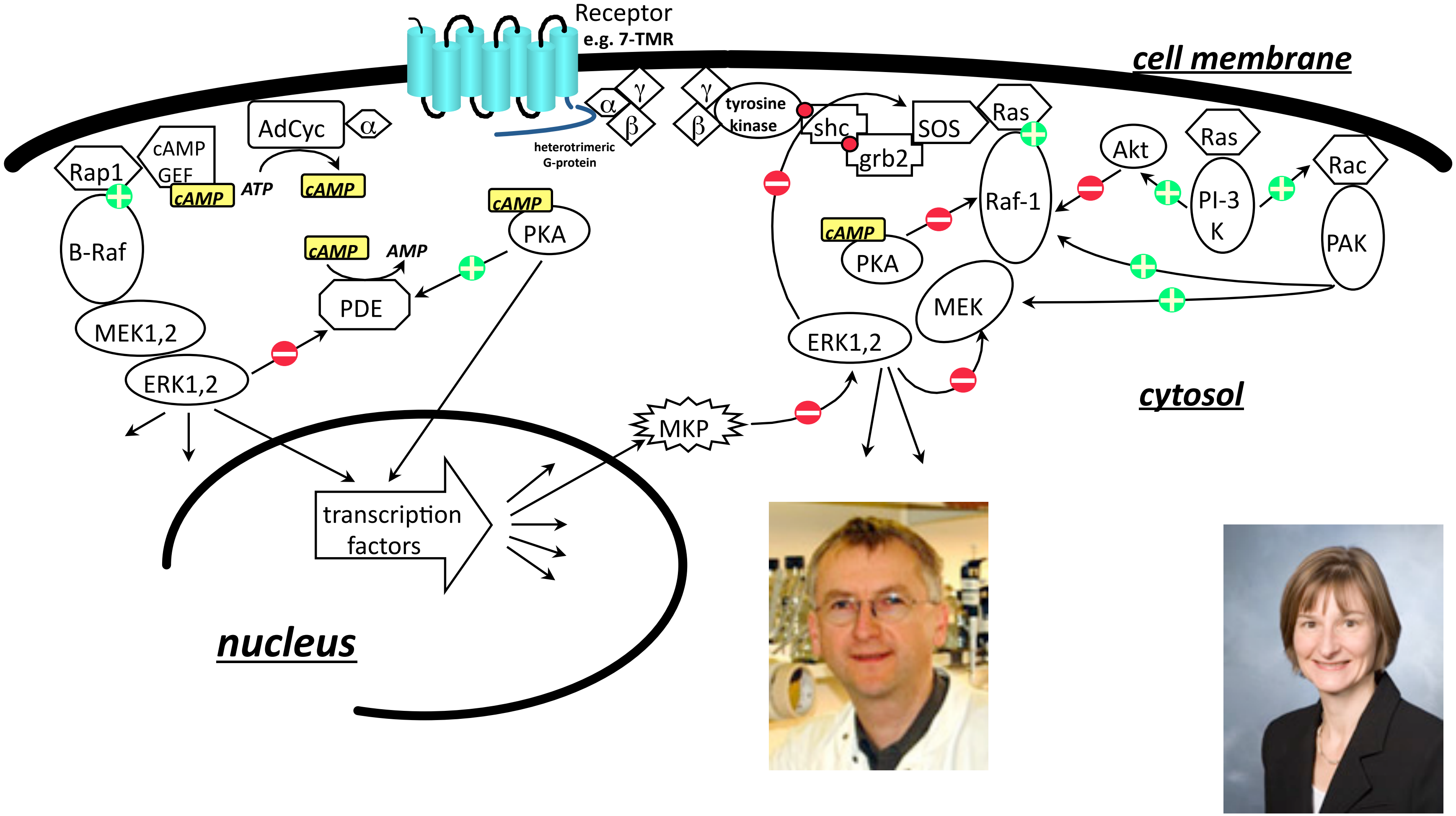
marking , concentrations

$[Raf1^*]_{t=0} = 2\ \mu$Molar

QUANTITATIVE

Biochemical networks

We can describe the general topology and single biochemical steps. However, we do not understand the network function as a whole.

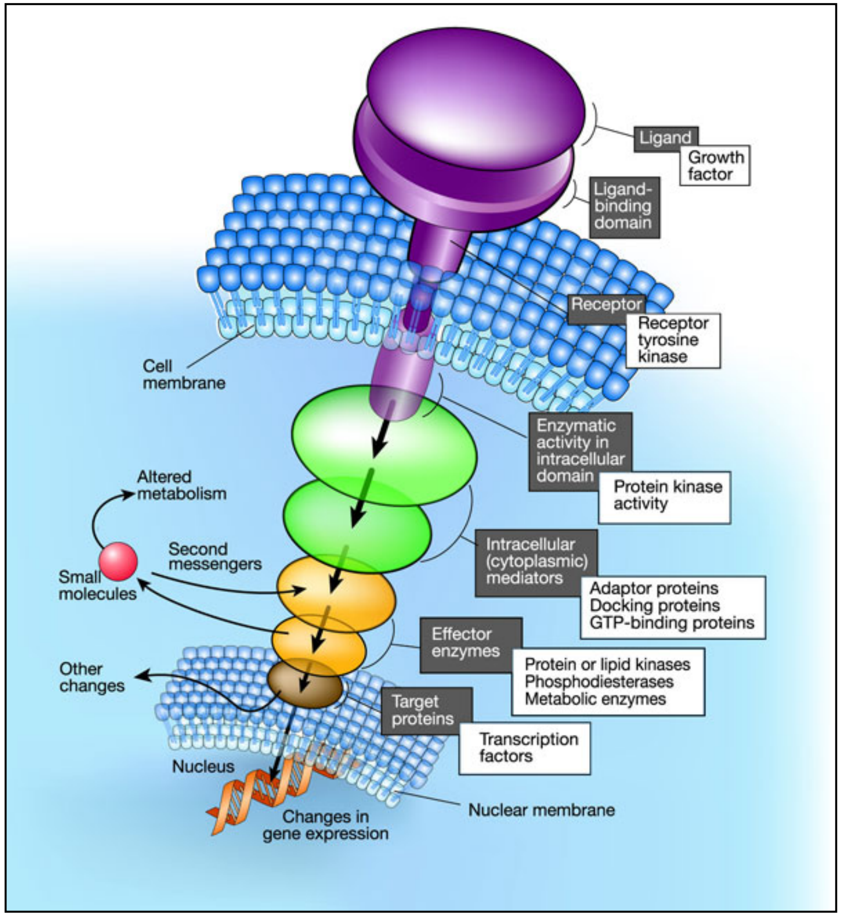


Walter Kolch

Muffy Calder

MAPK Pathway

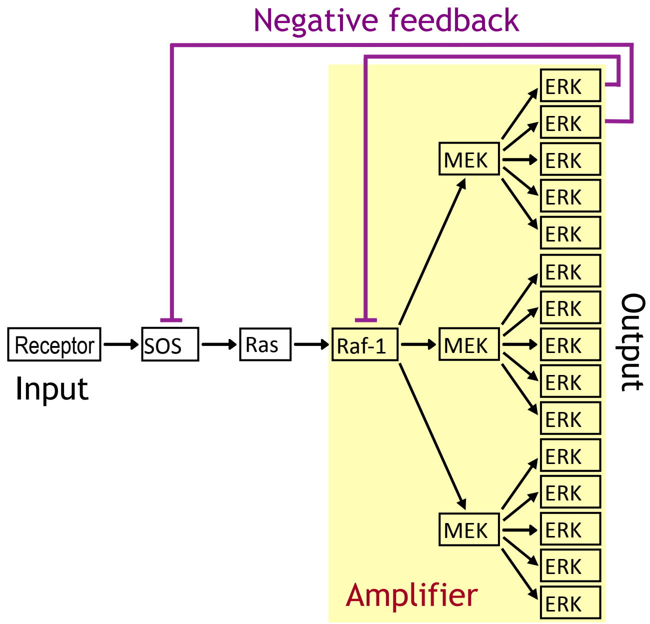
- Responds to wide range of stimuli: cytokines, growth factors, neurotransmitters, cellular stress and cell adherence,…
- Pivotal role in many key cellular processes:
  - growth control in all its variations,
  - cell differentiation and survival
  - cellular adaptation to chemical and physical stress.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



david.gilbert@brunel.ac.uk

ERK cascade well known biological amplifier

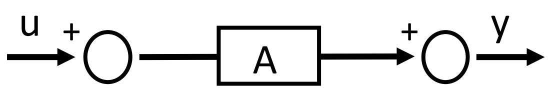
- Amplifies the original signal to create effective cellular responses.
- 1:3:5 are the approximate ratios of Raf-1, MEK and ERK in fibroblasts.
- Well known negative feedback loop: phosphorylation of SOS by ERK-PP (via MAPKAP1) resulting in the dissociation of the Grb2/SOS complex.
- New negative feedback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

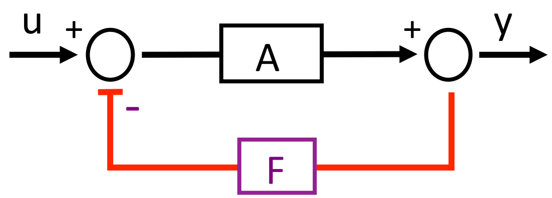
Negative Feedback Amplifier

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

How to test if the ERK pathway is a NFA?

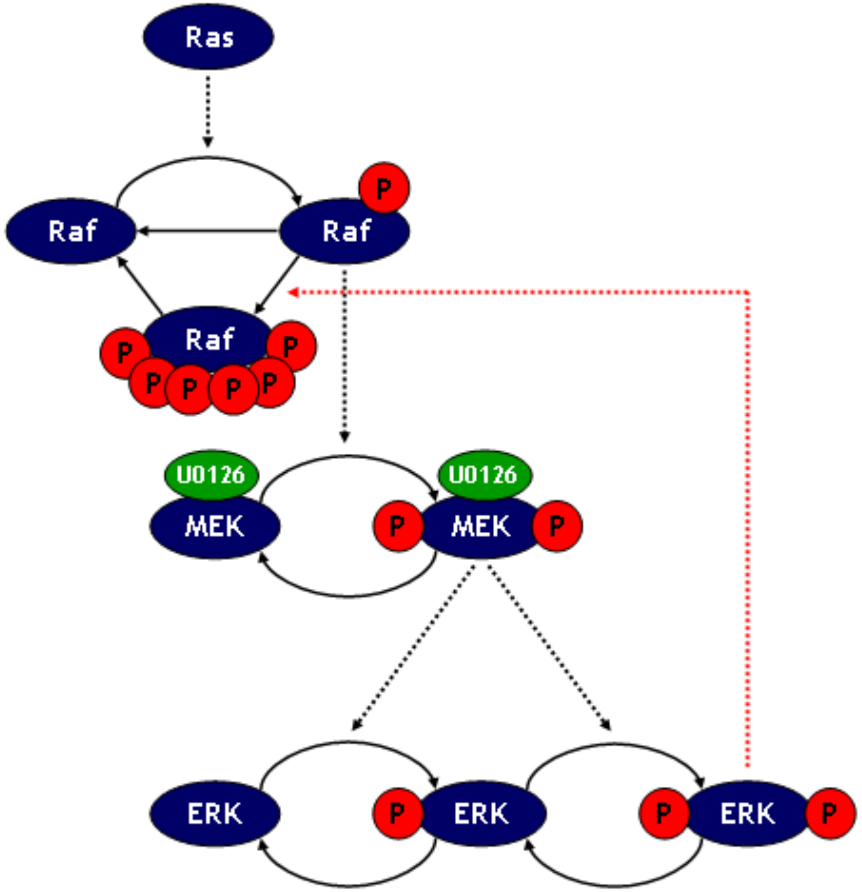
Strategy

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)

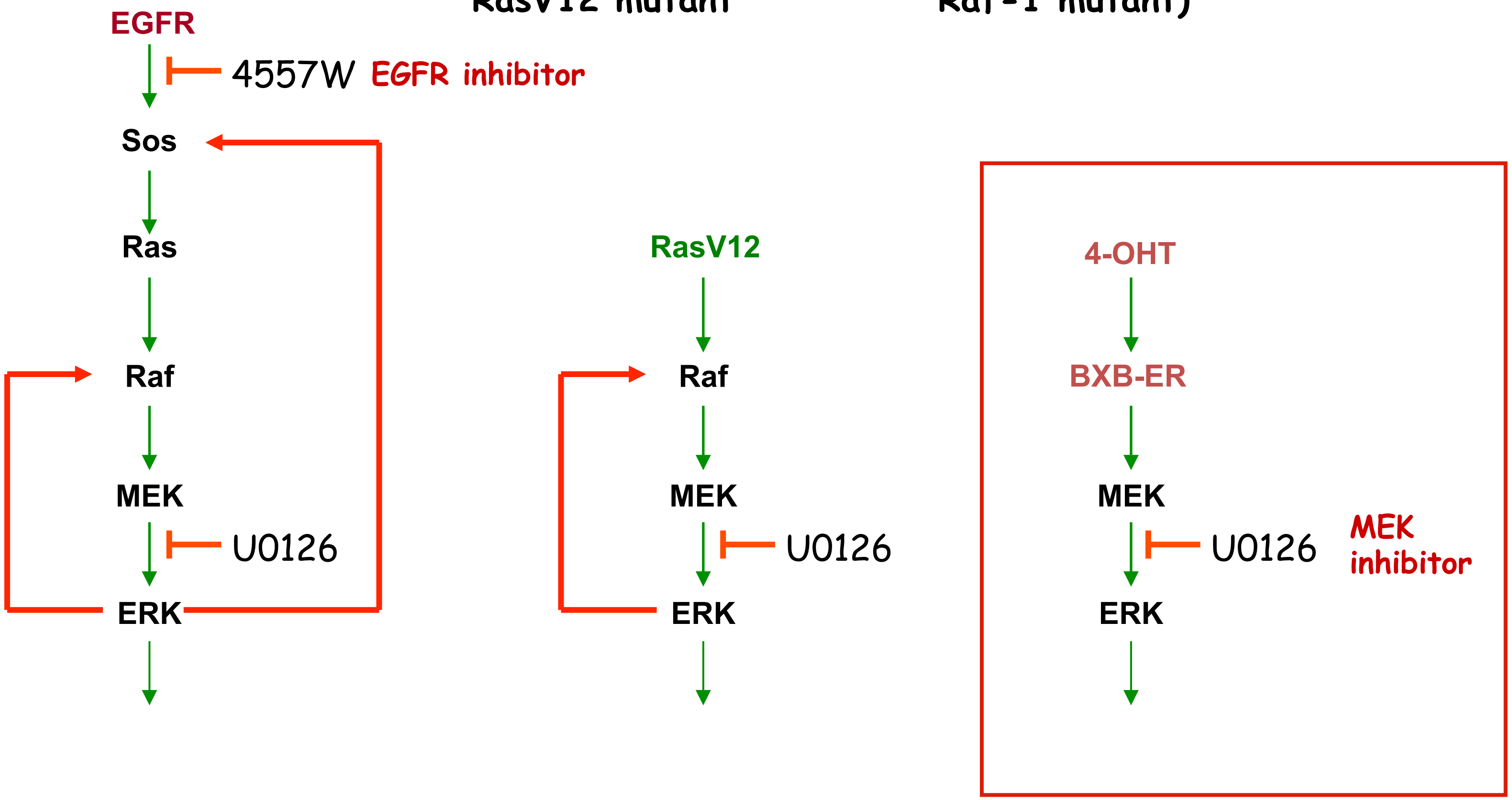
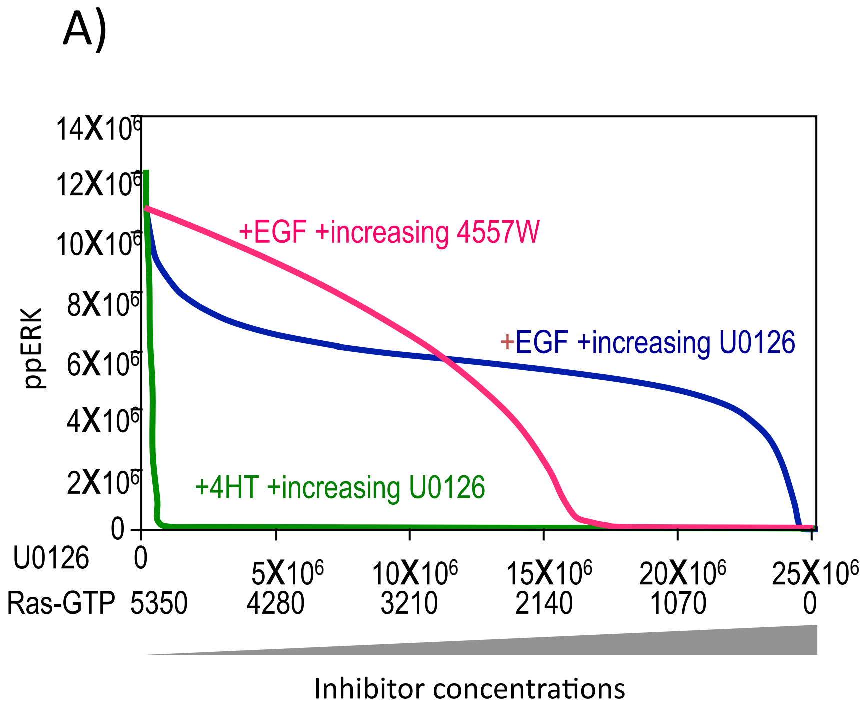
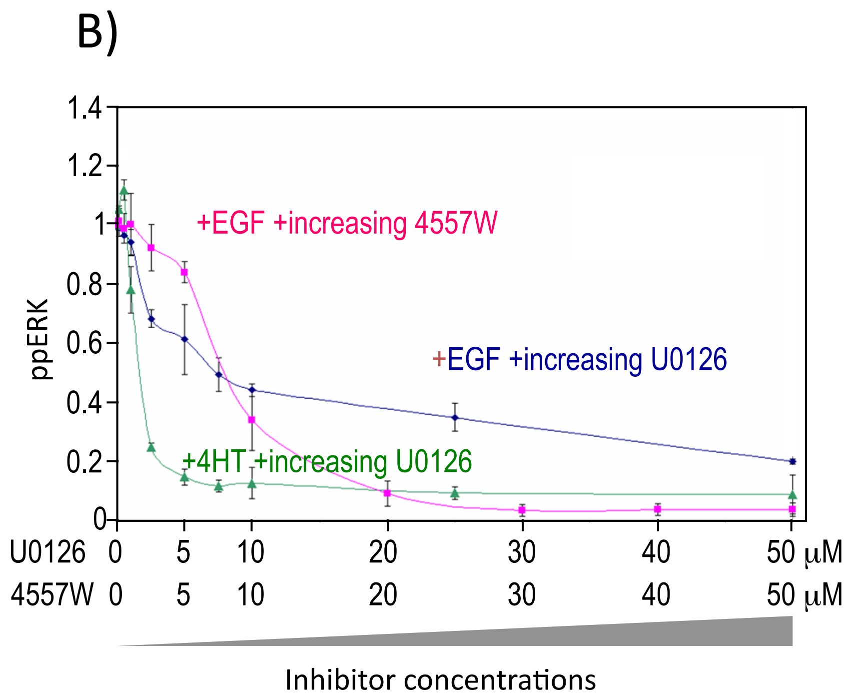


Figure 3

(A) Model prediction (B) Biochemical validation





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

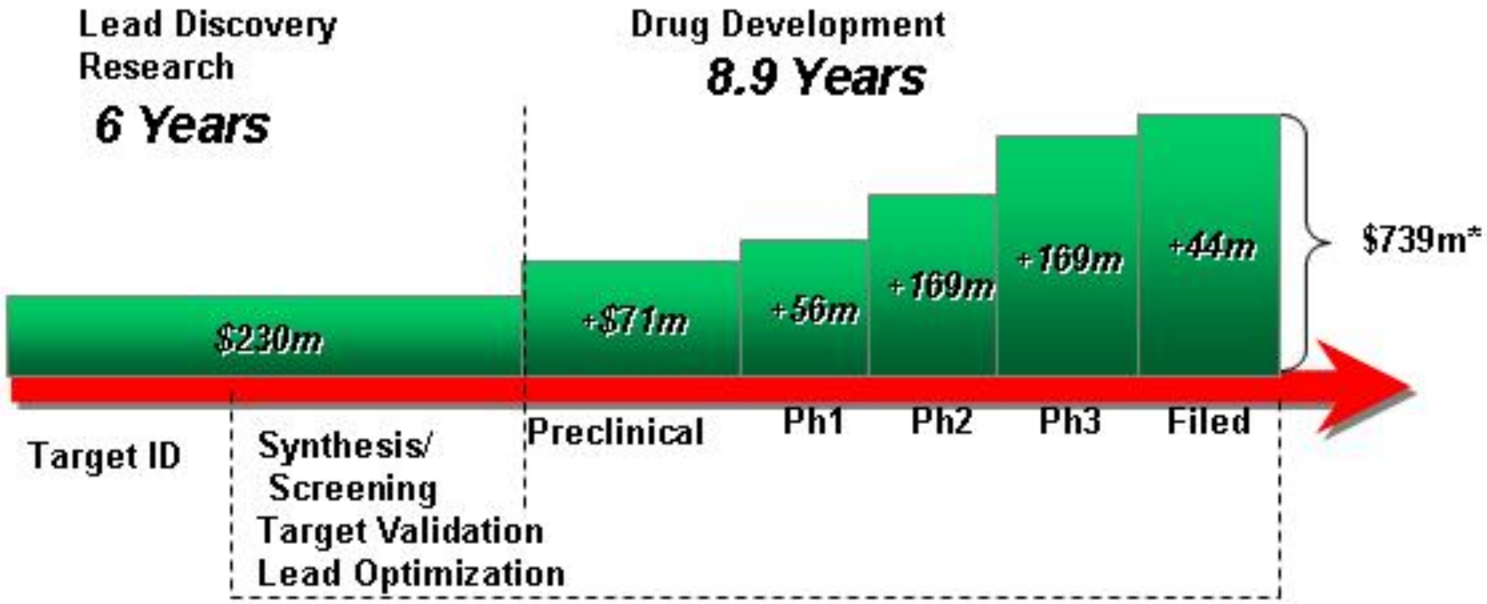
- Three-tiered kinase module, signal amplifier.
- Negative feedback loops - system like negative feedback amplifier
- Smoothens the output to changes in input - system robust to change.
- No feedback loops: cells sensitive to inhibition of MEK
- Feedback intact: cells are resistant to inhibition there.

- **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

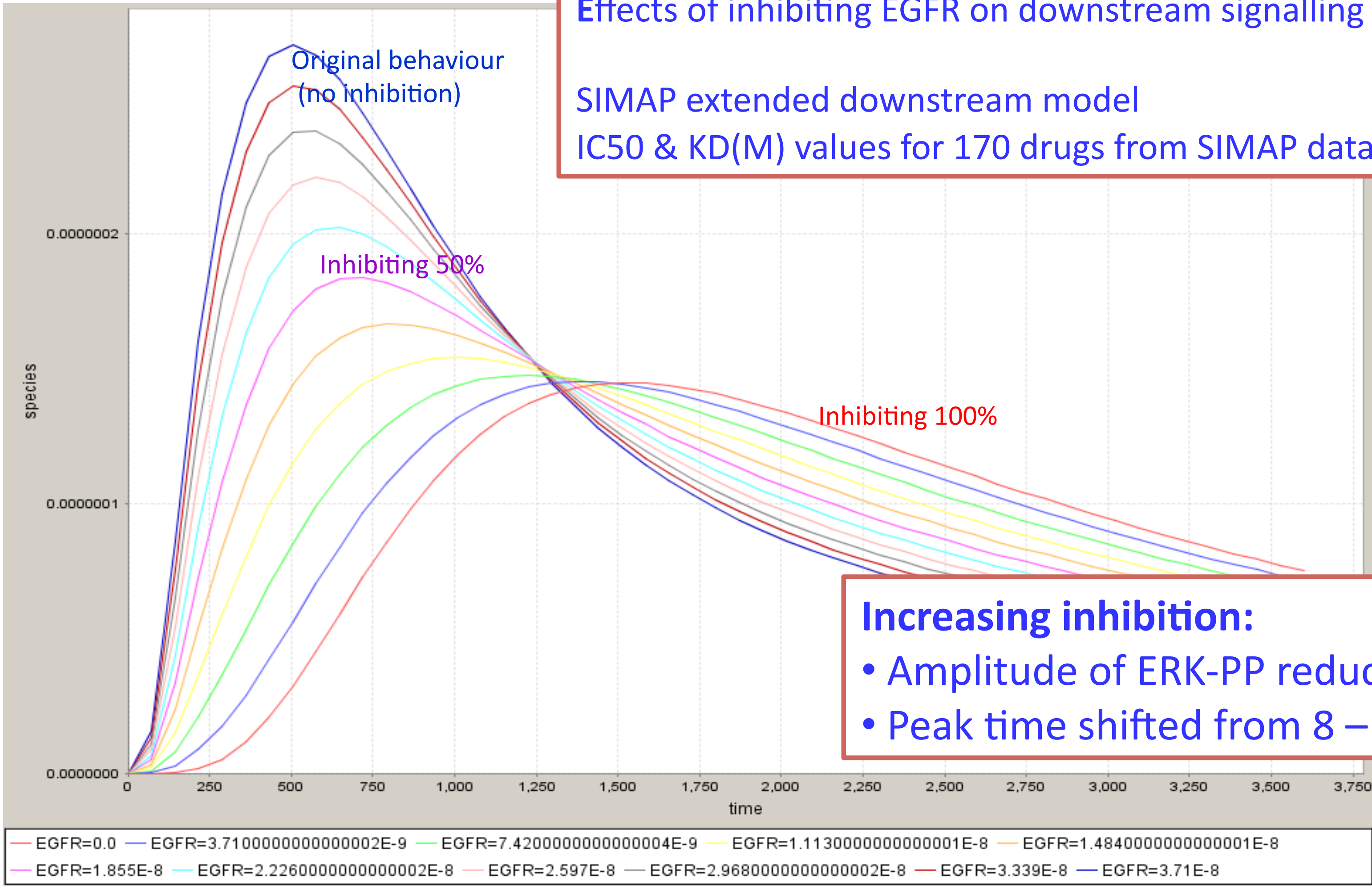
Traditional Pharmaceutical R&D

Costly and Time Consuming***



*Lehman Brothers, 1997; ** Tufts CSDD

Using the Utility to model the effect of drug inhibition

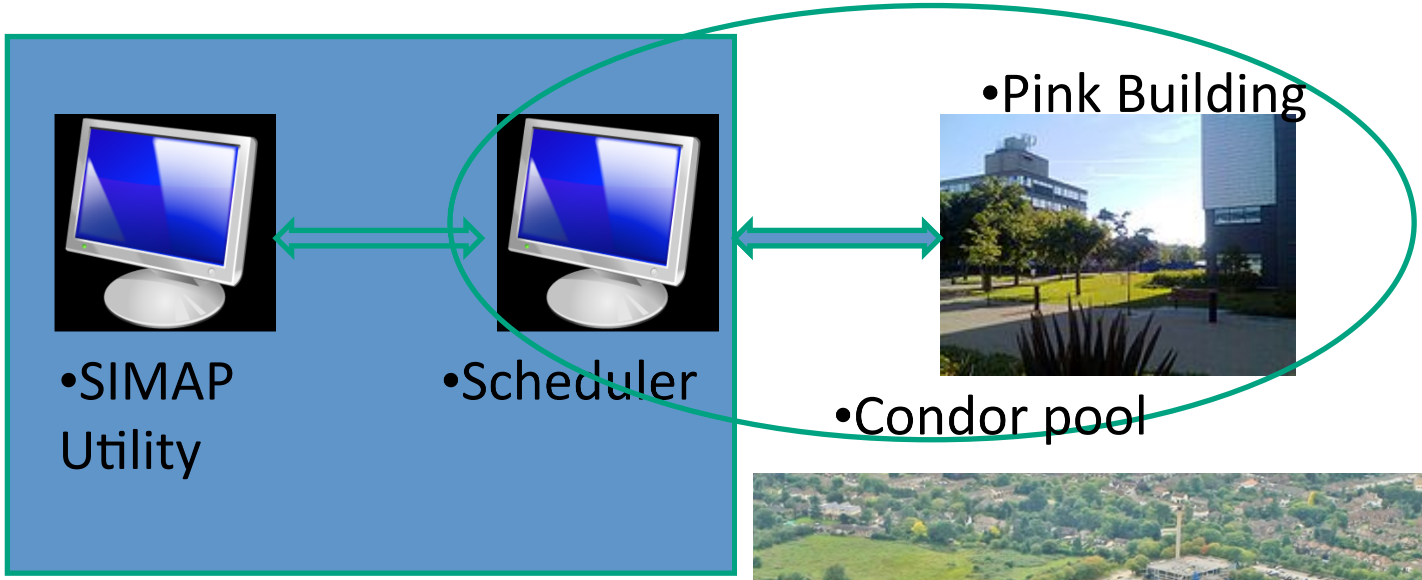


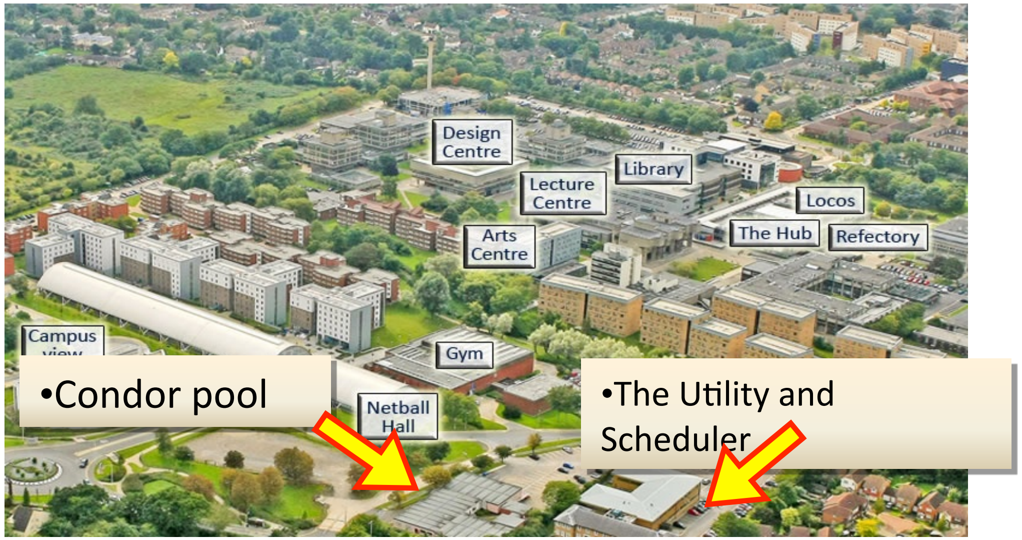
Effects of inhibiting EGFR on downstream signalling (ERK-PP).

SIMAP extended downstream model
IC50 & KD(M) values for 170 drugs from SIMAP database

Increasing inhibition:
- Amplitude of ERK-PP reduced
- Peak time shifted from 8 – 25 mins

Condor Based Grid Component for SIMAP Utility

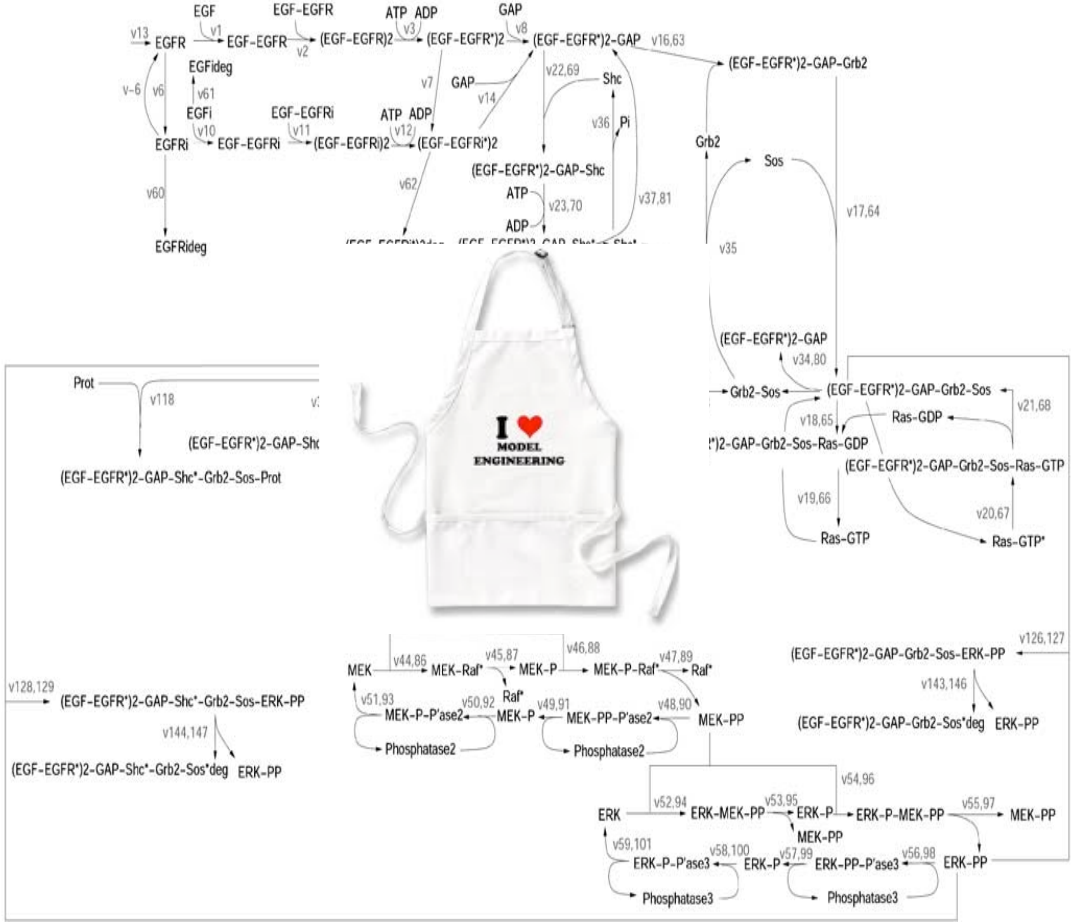




Grid-enabled SIMAP Utility: Motivation, Integration Technology and Performance Results

Jun Wang, Xuan Liu, Navonil Mustafee, Qian Gao, Simon J E Taylor, David Gilbert

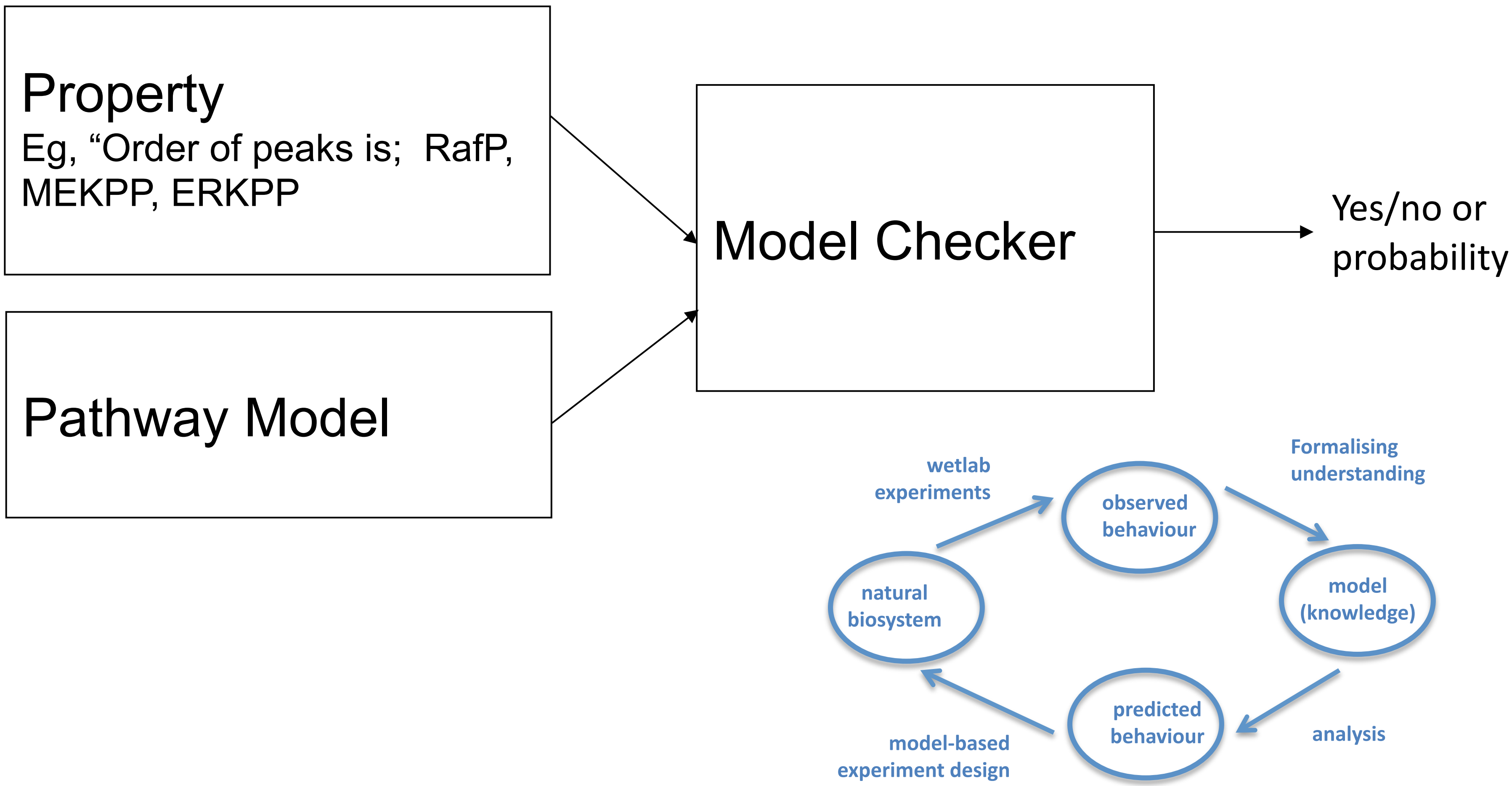
BioModel Engineering



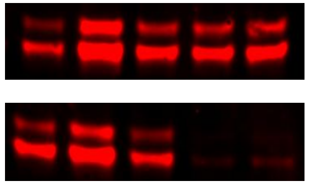
Rainer Breitling, David Gilbert, Monika Heiner, Richard Orton (2008). A structured approach for the engineering of biochemical network models, illustrated for signalling pathways. Briefings in Bioinformatics

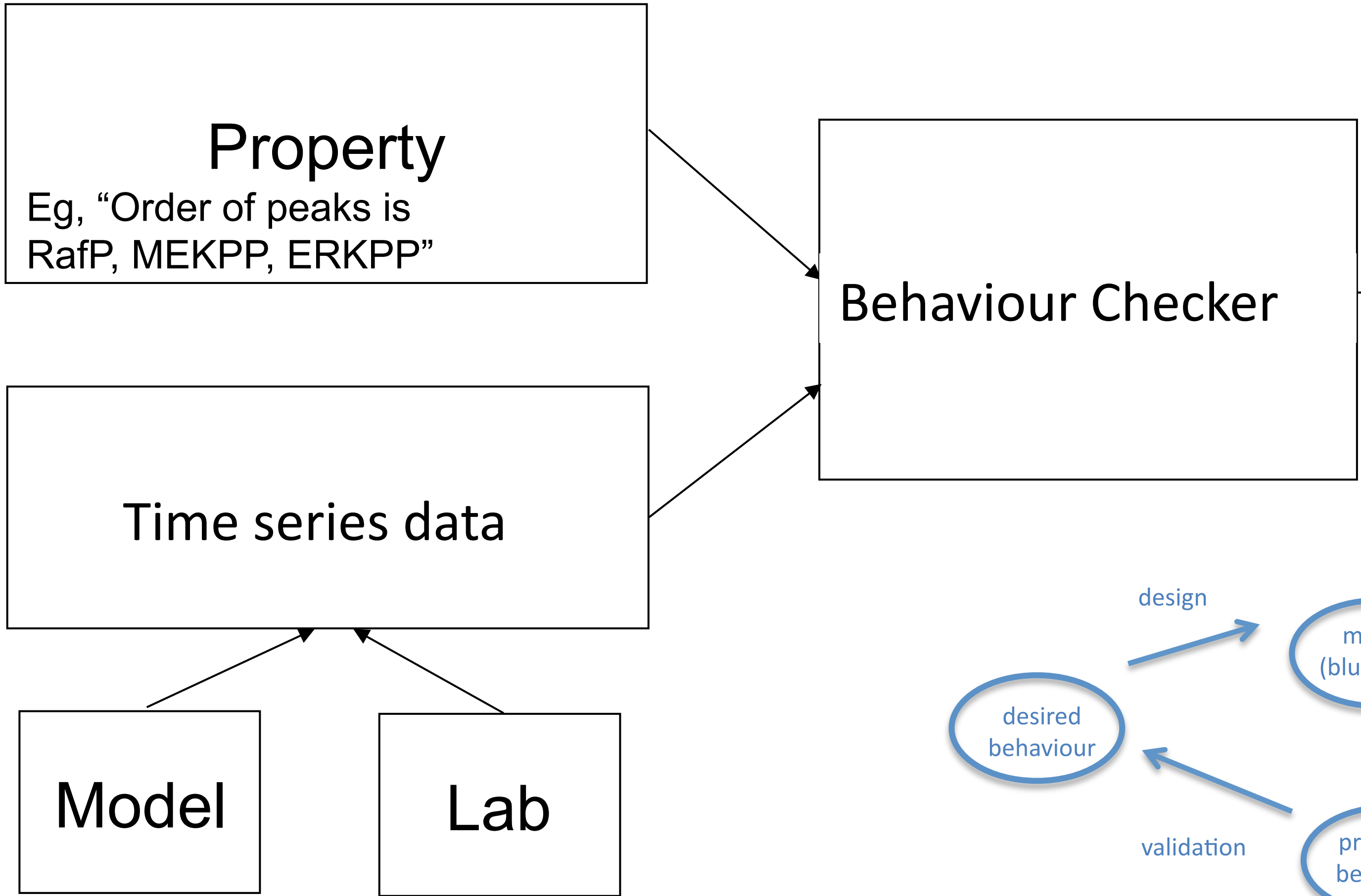
David Gilbert, Rainer Breitling, Monika Heiner, and Robin Donaldson (2009). An introduction to BioModel Engineering, illustrated for signal transduction pathways, 9th International Workshop, WMC 2008, Edinburgh, UK LNCS Volume 539, pp13-28
Rainer Breitling, Robin Donaldson, David Gilbert, Monika Heiner (2010): Biomodel Engineering - From Structure to Behavior; : Trans. Comp Systems Biology XII, Springer LNBI 5945, pp. 1-12

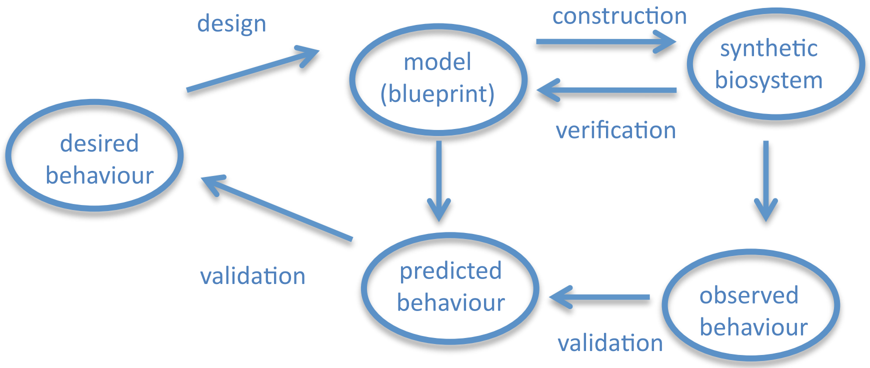
Model Checking
Biochemical Pathways



Simulation-based Model Checking Biochemical Pathways







Qualitative to quantitative descriptions in PLTL

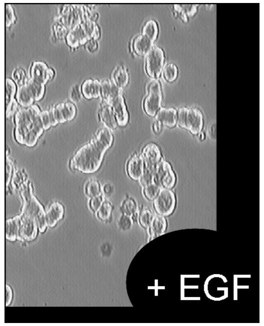
- **Qualitative**:
  Protein rises then falls
  P=? [(d(Protein) > 0) U (G(d(Protein) < 0))]

- **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)))]

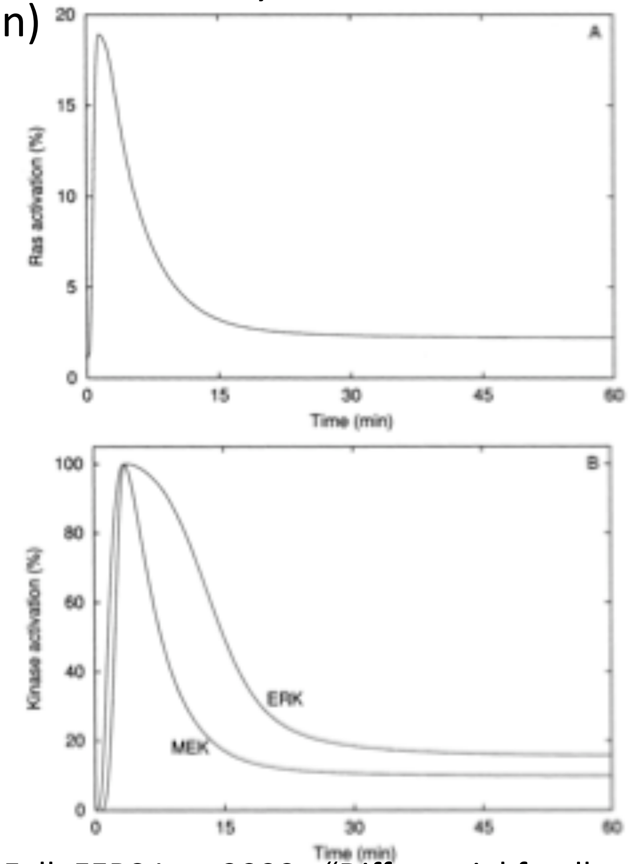
- **Quantitative**:
  Protein rises then falls to less than <u>100μMol</u> by 60 minutes
  P=? [(d(Protein) > 0) U (G(d(Protein) < 0) ∧ F (time = 60 ∧ Protein < 100))]

Parameter estimation
Response with EGF vs. NGF signal

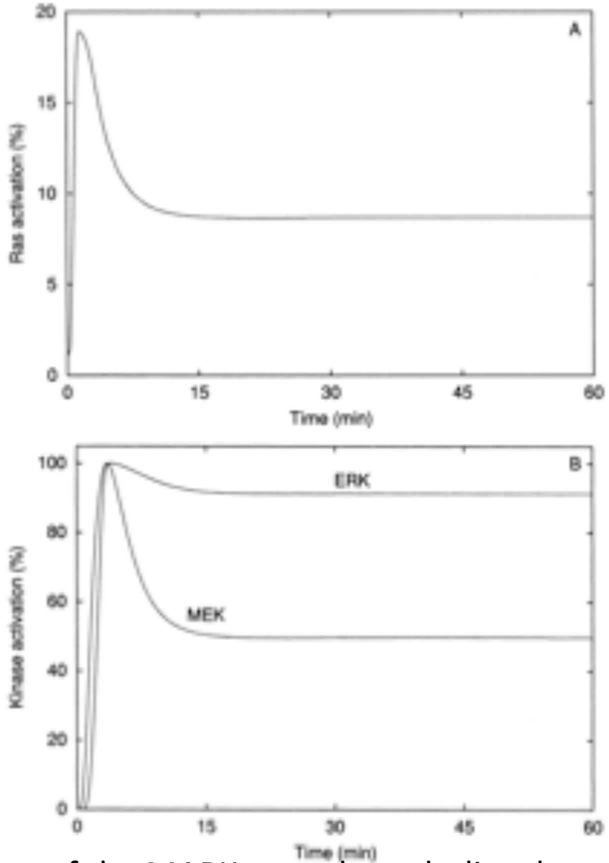


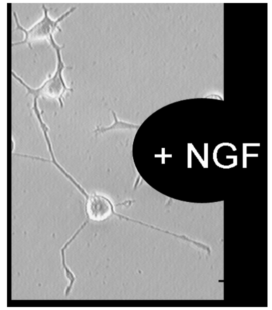
Proliferation
(cell division)
PC12 cells

EGF
Transient activation of Ras, MEK and ERK

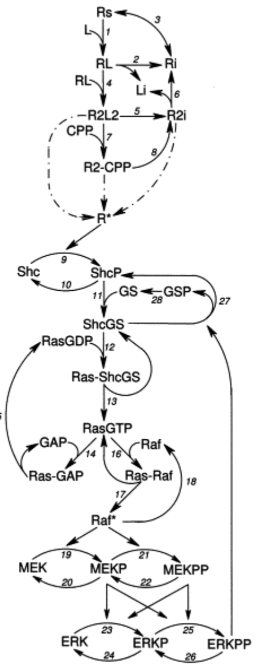


NGF
Sustained activation of Ras, MEK and ERK





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))))]

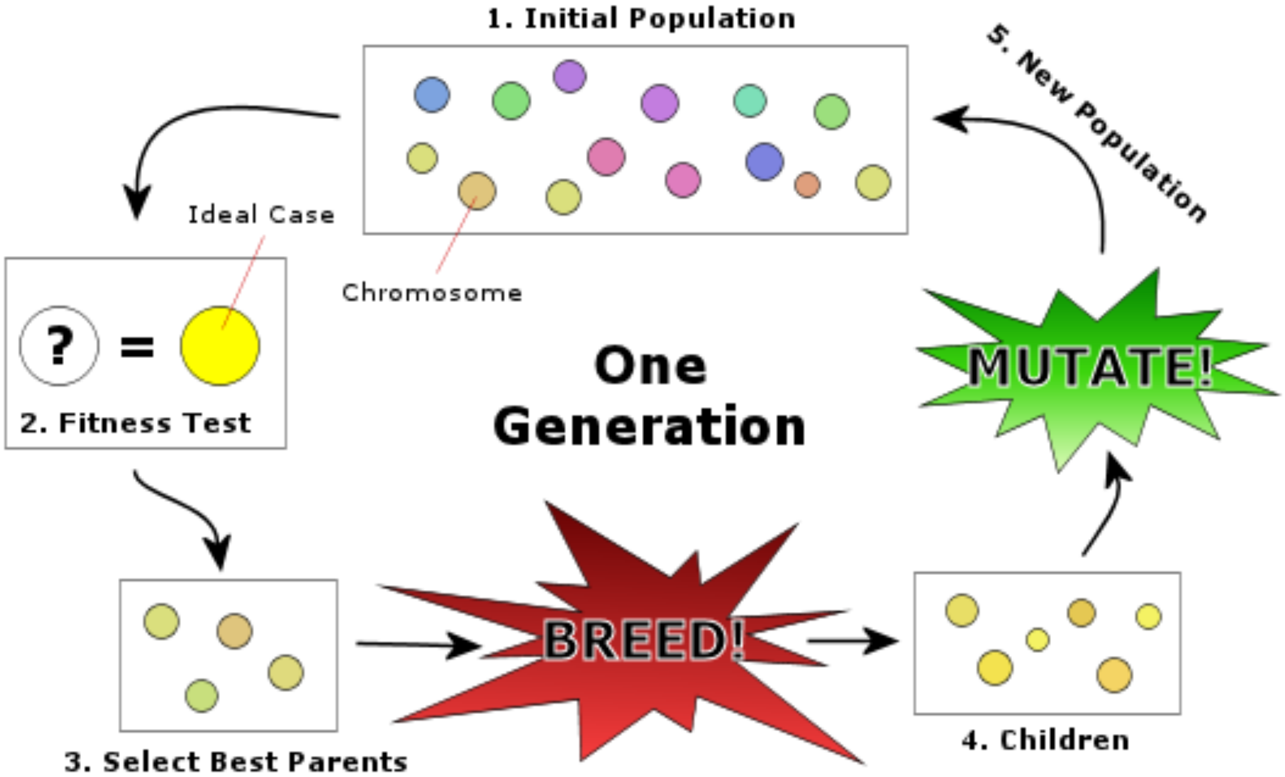
Sustained ERK: Active ERK peaks within 2 to 5 minutes and is stable between 85% and 100% of peak value

$P_{=?}$ [(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

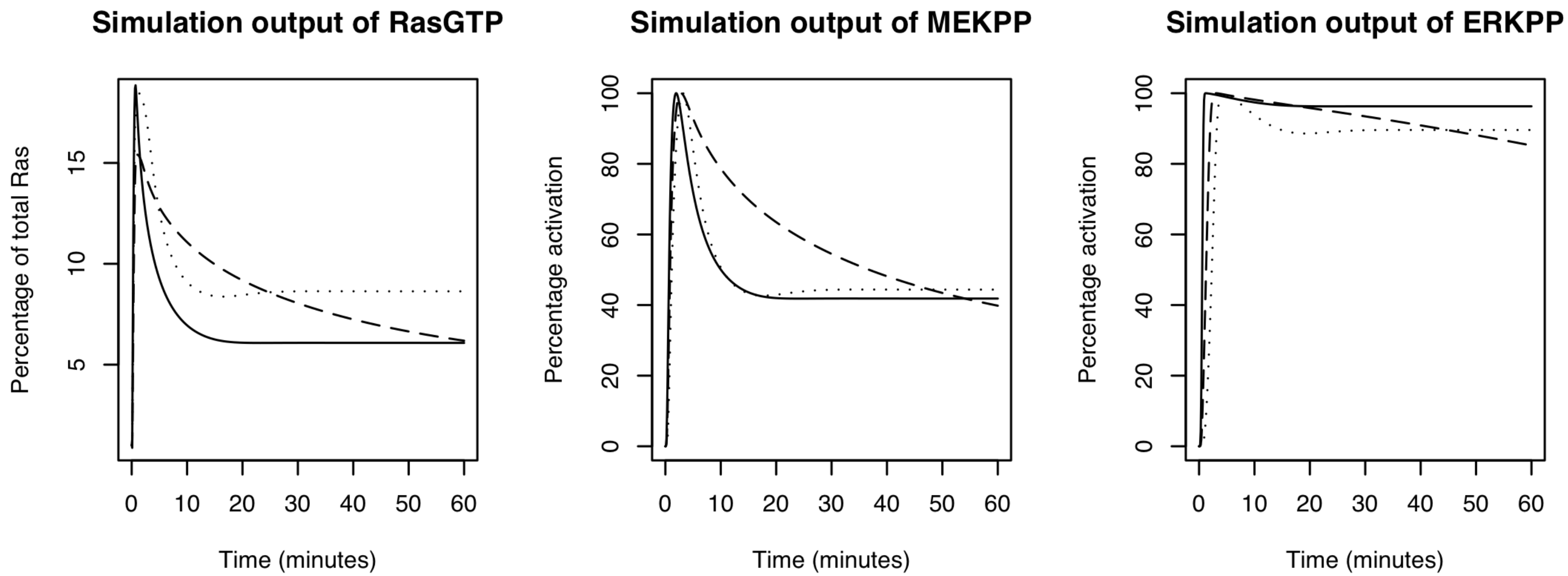
Sustained ERK: Active ERK peaks within 2 to 5 minutes and is stable between 85% and 100% of peak value



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:

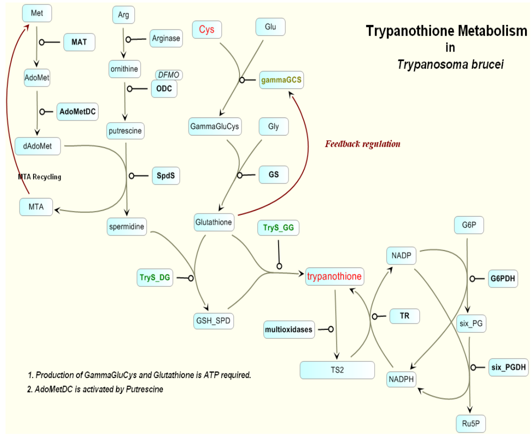


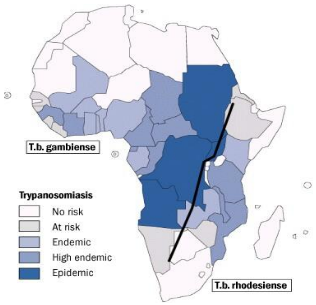
- Original model of the NGF signalling pathway varying V28 (dotted)
- Best model returned when varying the critical parameters (solid)
- Critical parameters without V28 (dashed).

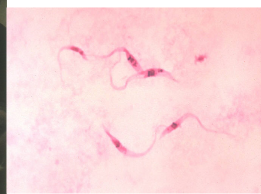
The best model returned when varying the critical parameters only required a ***16-fold*** increase in V28 (compared with 40-fold in original paper)

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

System Biology of Trypanosoma metabolism



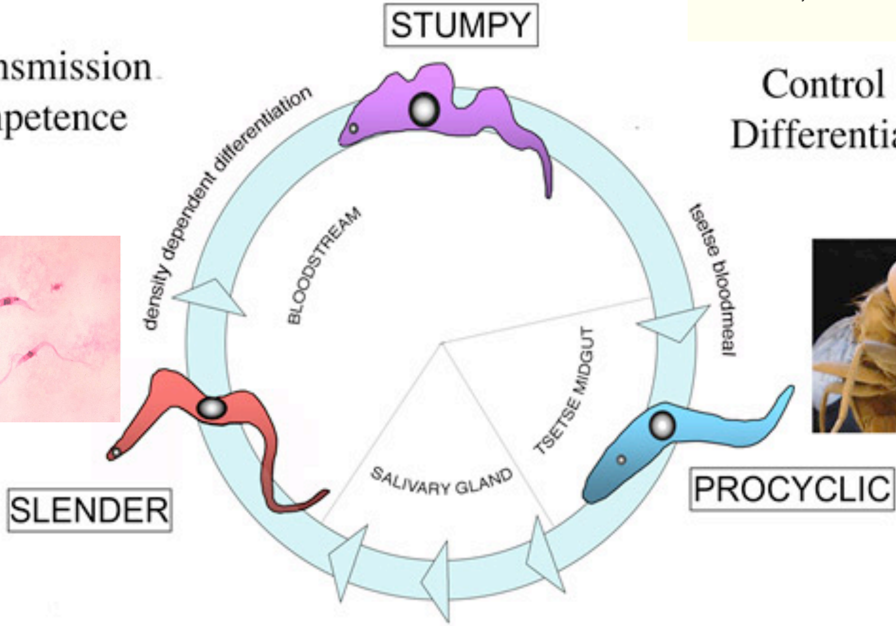




Transmission competence

Initiation of Differentiation

Control of Differentiation



Metabolic systems analysis – differentiation in Trypanosoma

Huma Lodhi & David Gilbert

Aim: Construction of accurate dynamic models of biochemical networks

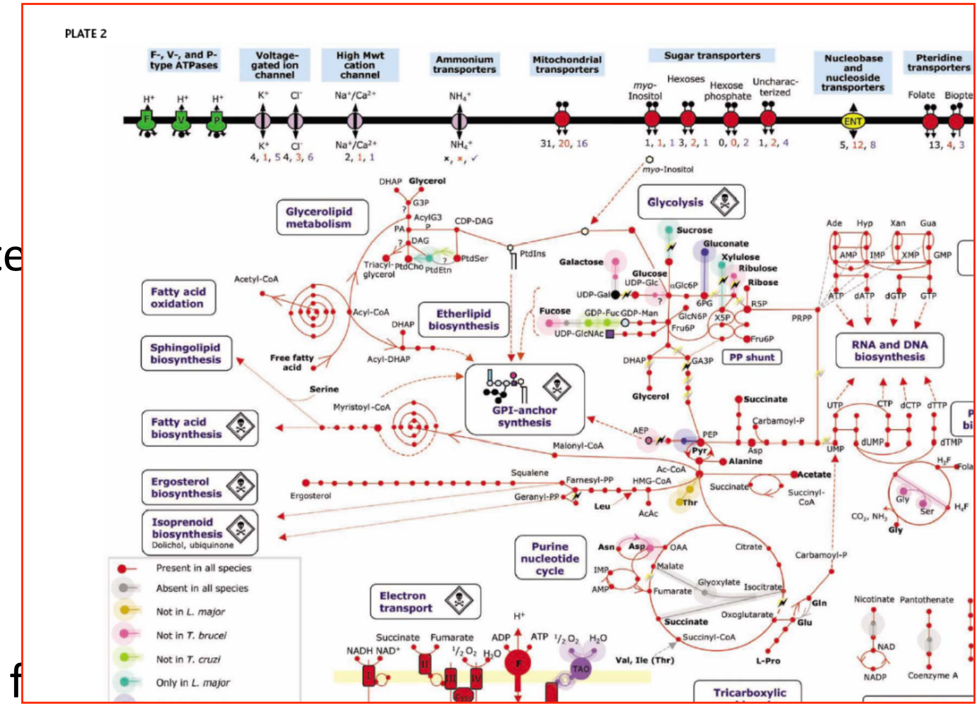
Challenges:

- Unknown / uncertain kinetic paramete
- Short time course data
- Noisy measurements

Solution:

Development of effective and efficient techniques

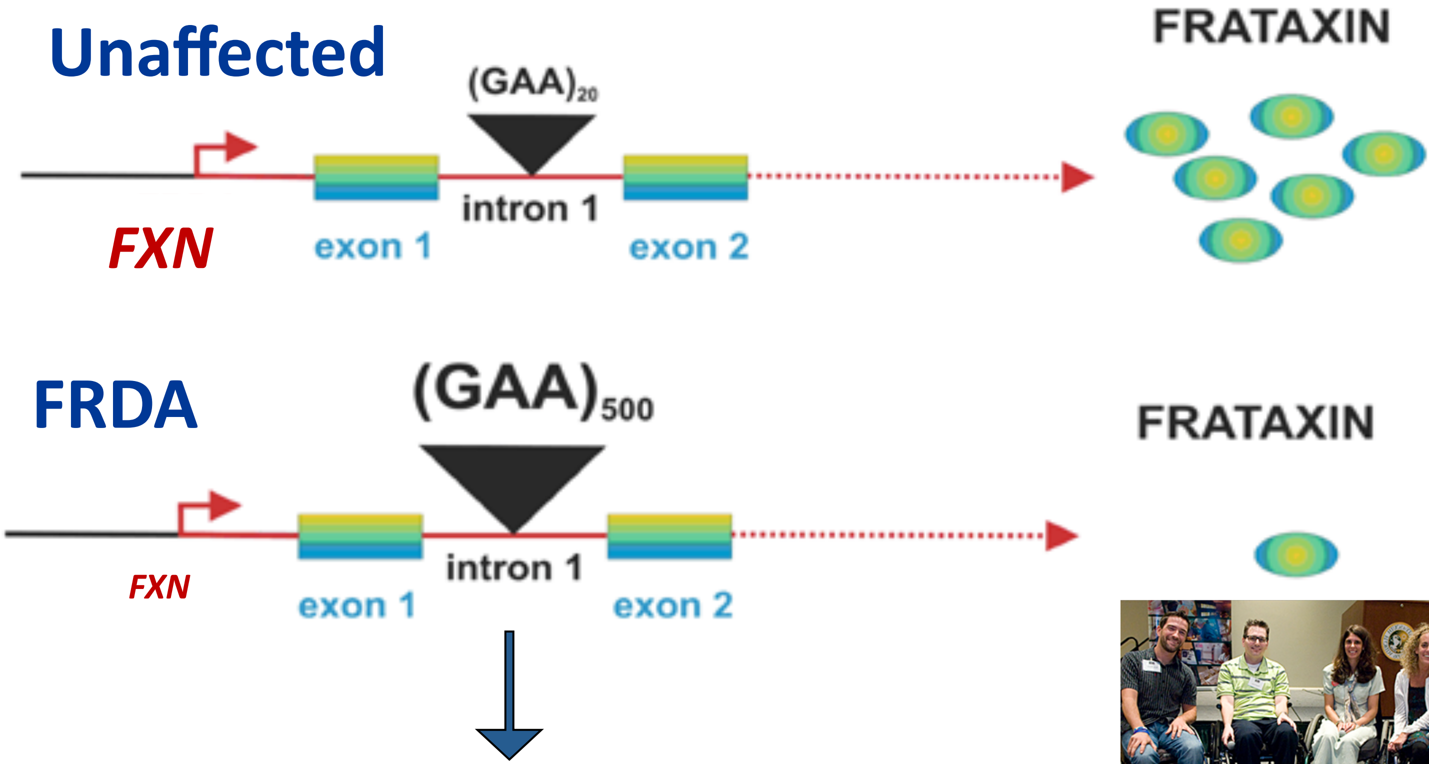
Example: A bootstrapping based method f estimating kinetic parameters.

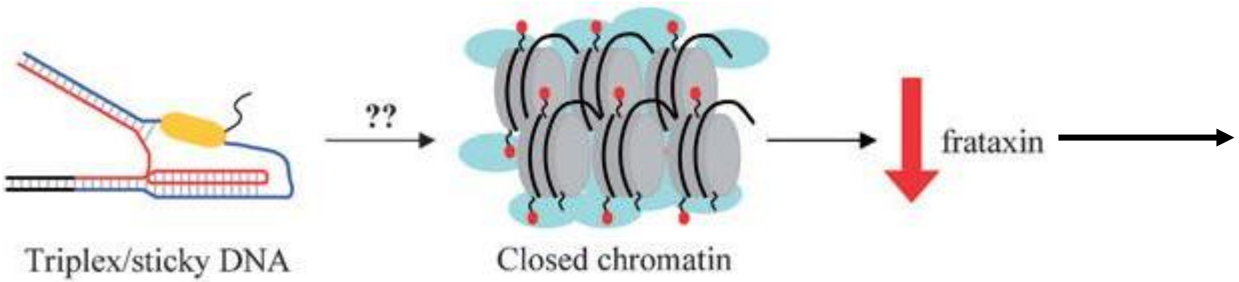


SysTryp

BBSRC (Glasgow) – ANR

Friedreich's Ataxia



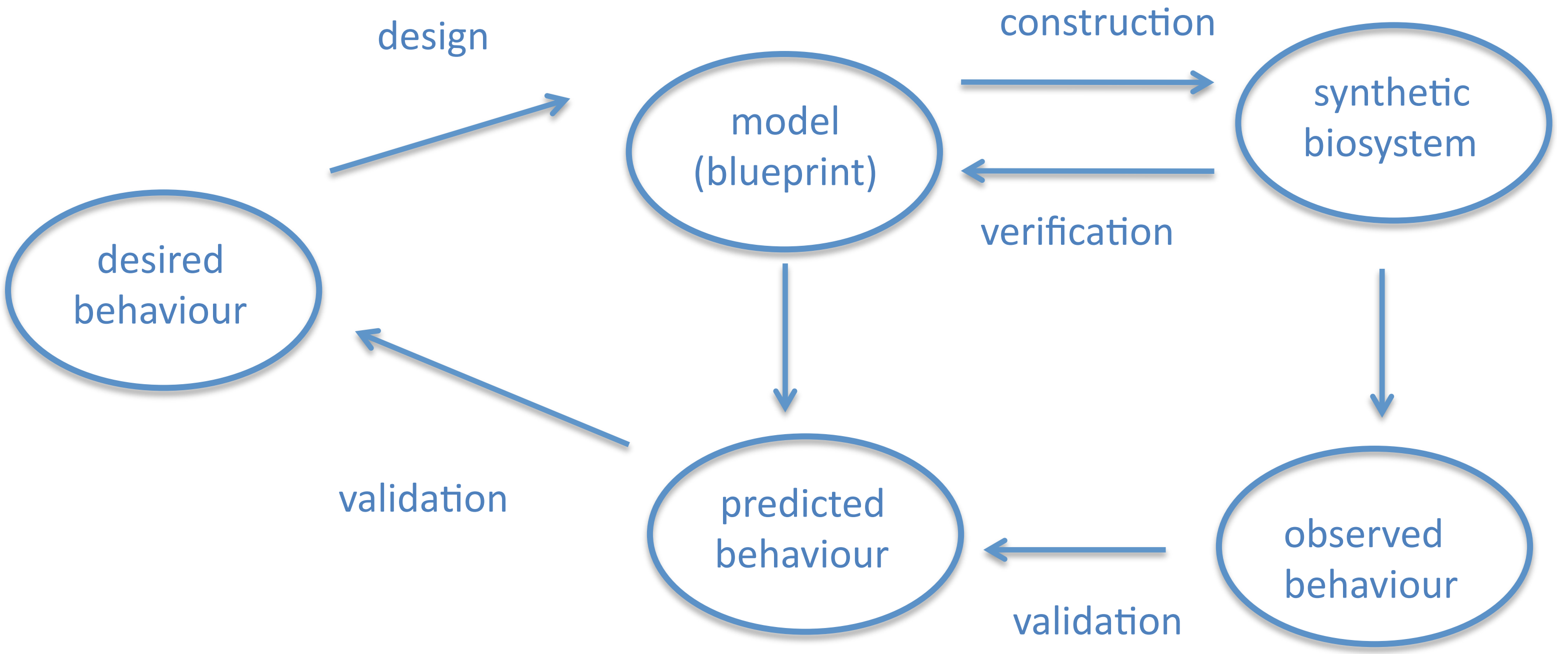


Disease effects

- Europea
  Translational Studies
  - EU funded (€6 million for 4 years)
  - 13 clinical, basic and computational research groups (5 UK) and Pharma (Repligen)

- There is activity in all “Research” and “Pre-clinical” aspects of the “Treatment Pipeline” that aims to take new FRDA treatments through to clinical trials.

Synthetic Biology

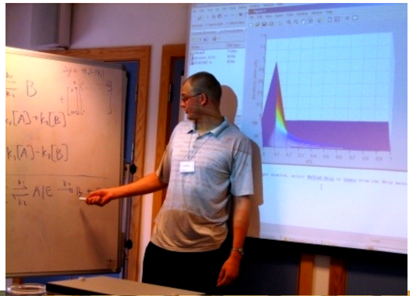


A drug manufacturing plant

- "Audacious plan" New Scientist, May 2006
- Engineer e.coli / yeast to synthesise the anti-malarial artemisinin
- $42.6 million, Bill & Melinda Gates Foundation

Artemisia annua

- Plant difficult to grow and only yield minute quantities of drug per kilo
- Artemisinin is expensive
  - ➢ Engineer cheaper alternative and save the world!

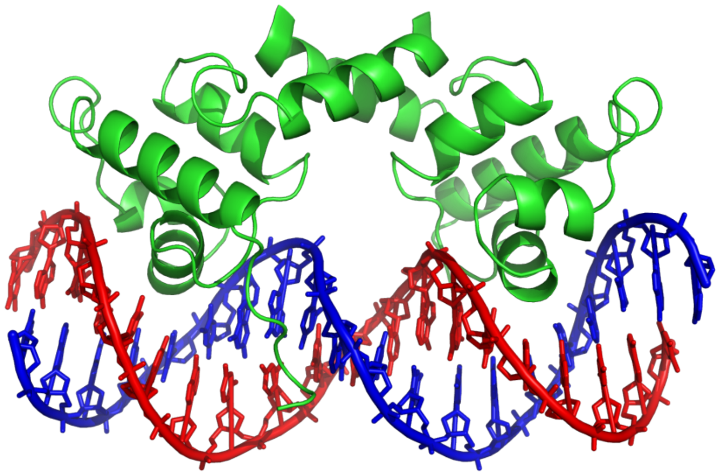



Best Enviro & Sensing
Glasgow

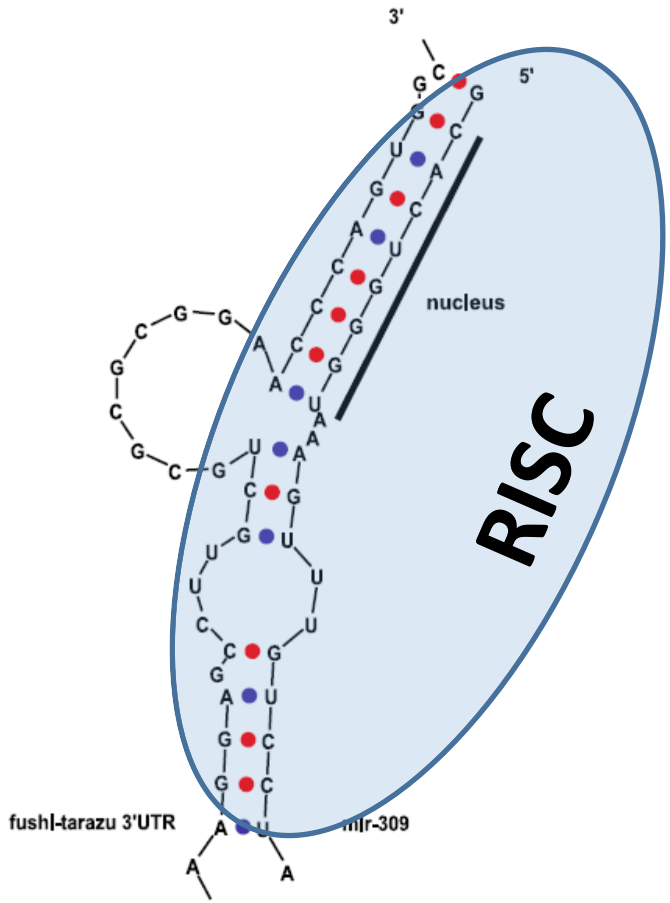

iGEM
International Genetically
Engineered Machines
Competition

Modular design – simple rules

- Enough pairs?
- Crosstalk?



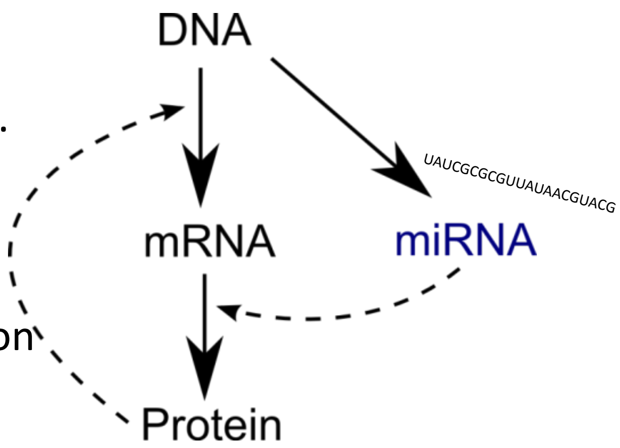
VS.

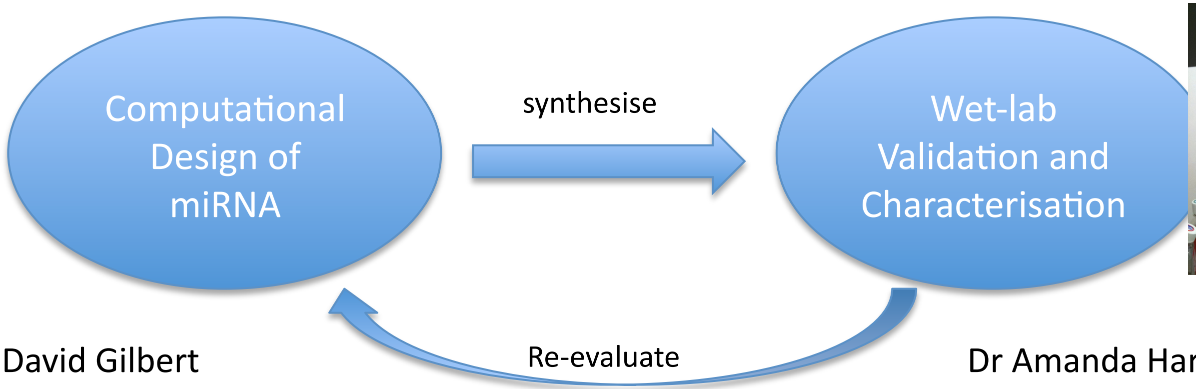


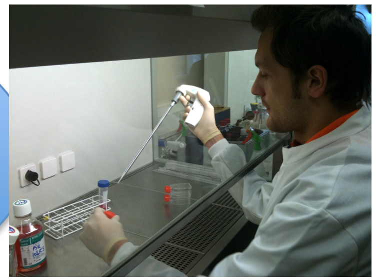
Design of novel miRNA:target site pairs

Maciej Trybiło (PhD student), David Gilbert, Amanda Harvey
Brunel University

- Another major mechanism for gene expression regulation.
- Allow regulatory mechanisms to be more robust, fine tuned.
- Cheap for the cell and faster.
- Rigorous computational design of new miRNAs for use in Synthetic Biology
- Major computational challenge: Avoid off-target hybridisation
- Dry & Wet project



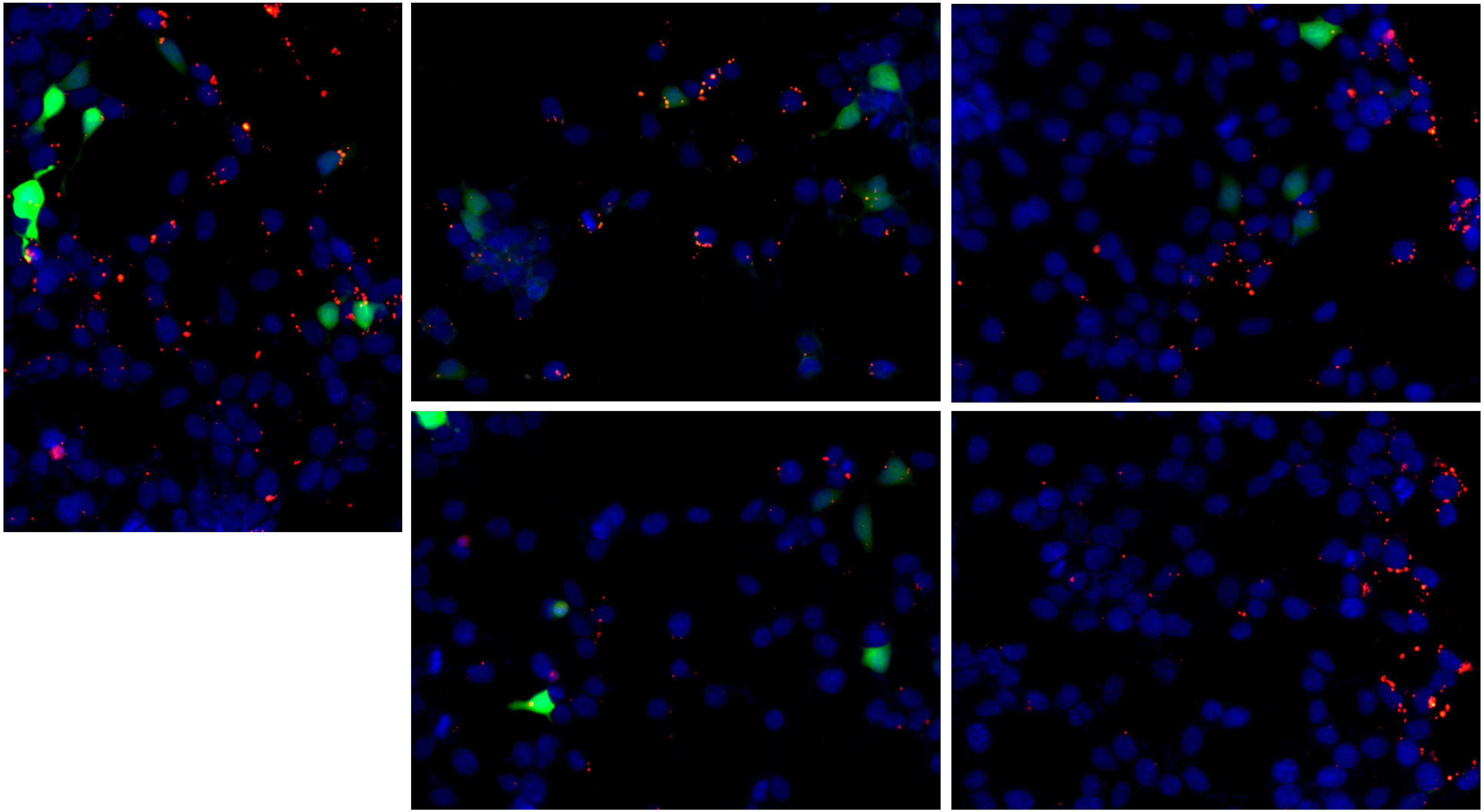




Prof David Gilbert

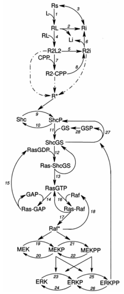
Dr Amanda Harvey

-ve siRNA, EGFP, sample 2, 5.4% green



Top-down Synthetic Biology development cycle

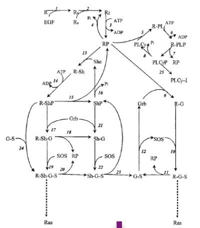
Model / check **Biosystem**



construct

validate

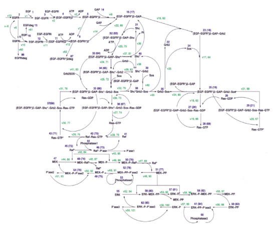
modify



construct

verify

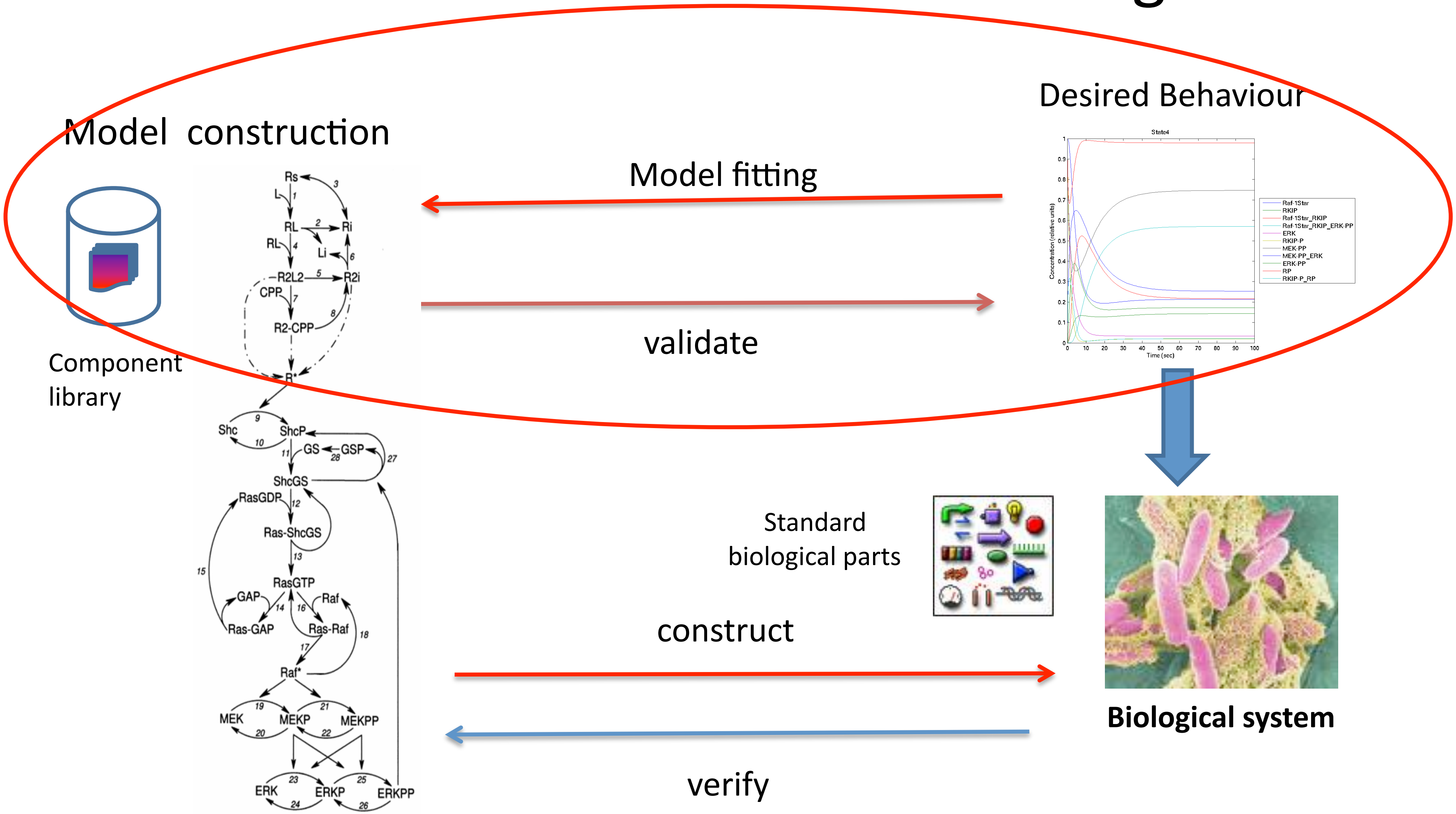
modify



construct

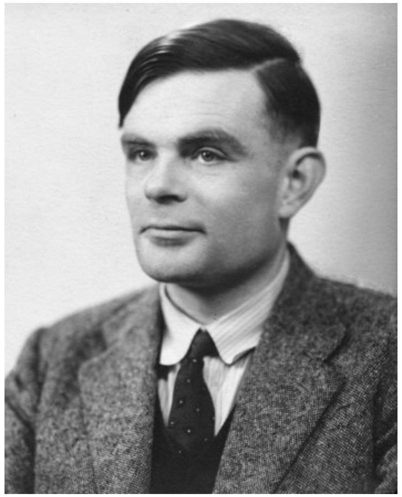
verify

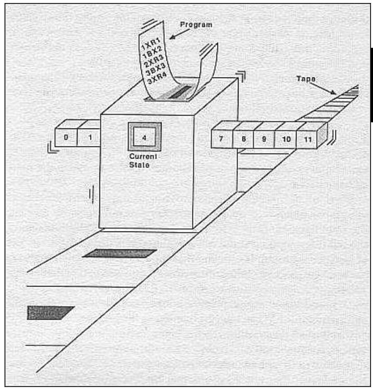
Vision – behaviour driven design

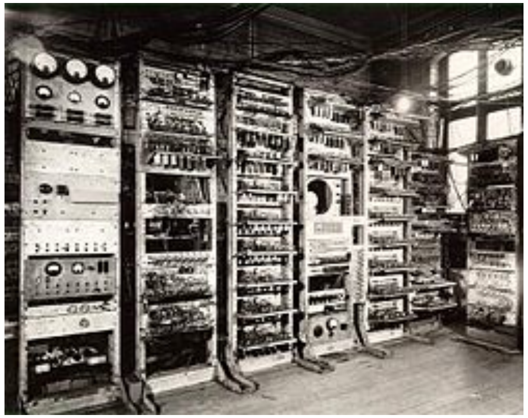


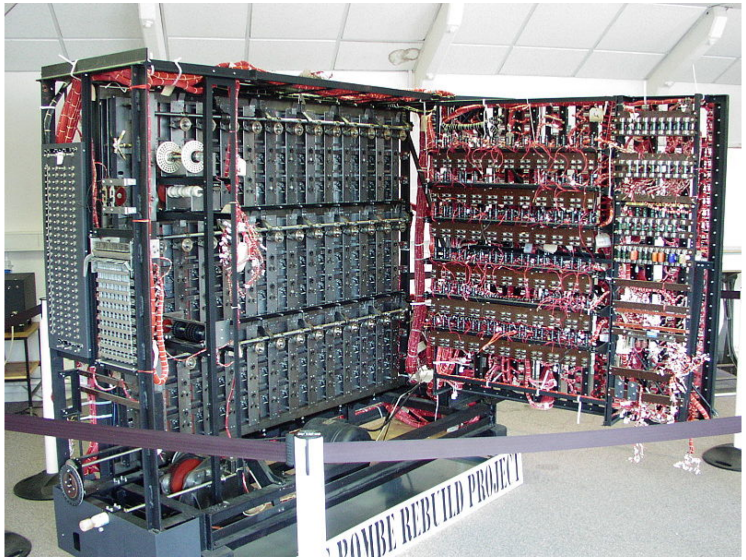
Wu, Gao & Gilbert (CMSB 2010)
Target Driven Biochemical Network Reconstruction Based on Petri nets and Simulated Annealing

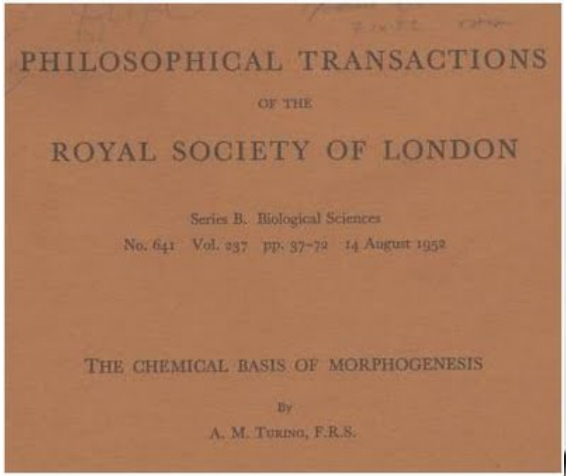
Turing



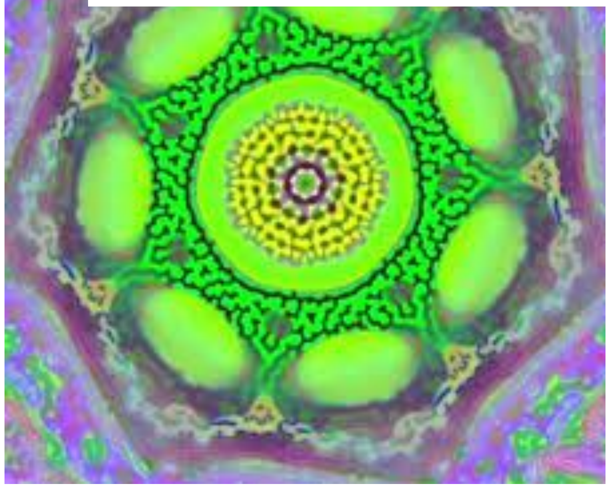




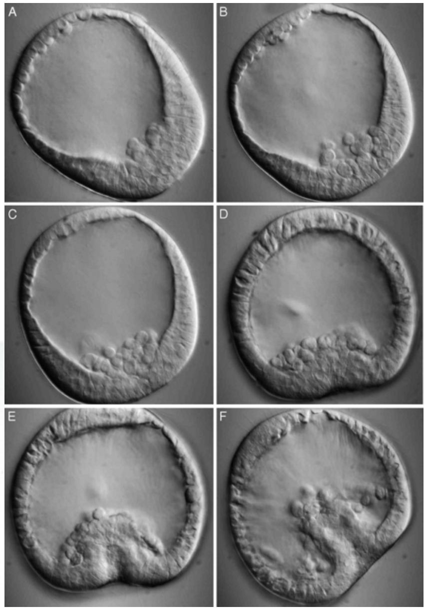




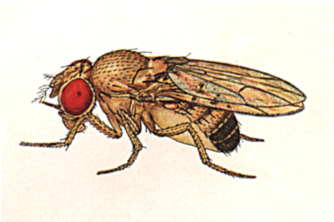
It is suggested that a system of chemical substances, called morphogens, reacting together and diffusing through a tissue, is adequate to account for the main phenomena of morphogenesis. Such a system, although it may originally be quite homogeneous, may later develop a pattern or structure due to an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

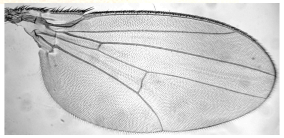


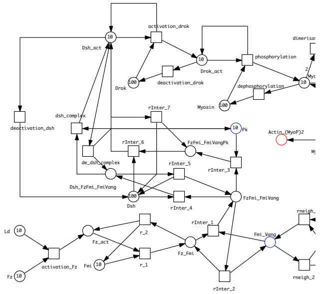
... this might account, for instance, for the tentacle patterns on Hydra and for whorled leaves… account for gastrulation…[and] phyllotaxis. **The purpose of this paper is to discuss a possible mechanism by which the genes of a zygote may determine the anatomical structure of the resulting organism.**



Multiscale from signalling to organs







Petri nets (*coloured, hierarchical*)

Monika Heiner

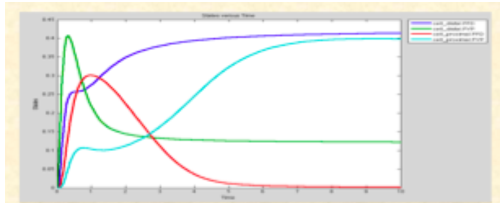
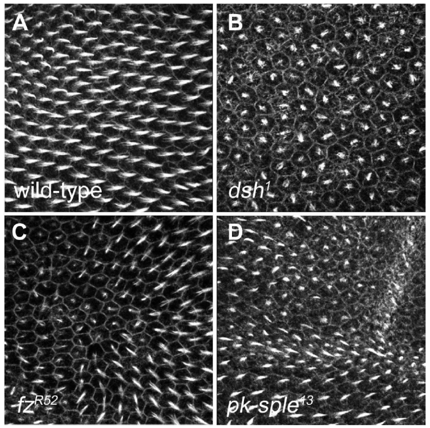


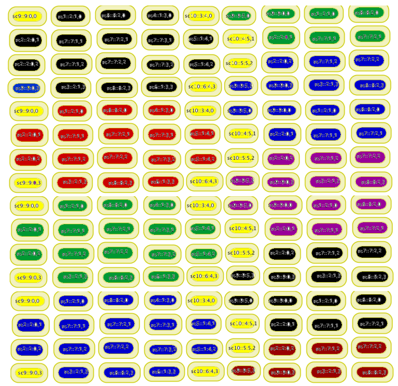
Figure 4: Plots of concentrations of key proteins over time obtained by solving ODEs.

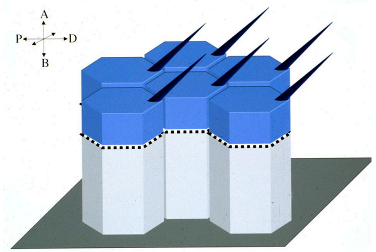
Planar Cell Polarity

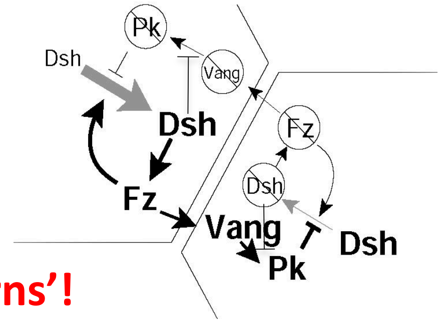


ODEs, stochastics

P-systems (InfoBiotics - Nottingham)







Pam Gao, David Tree

→ Design & genetically engineer 'patterns'!

Acknowledgements

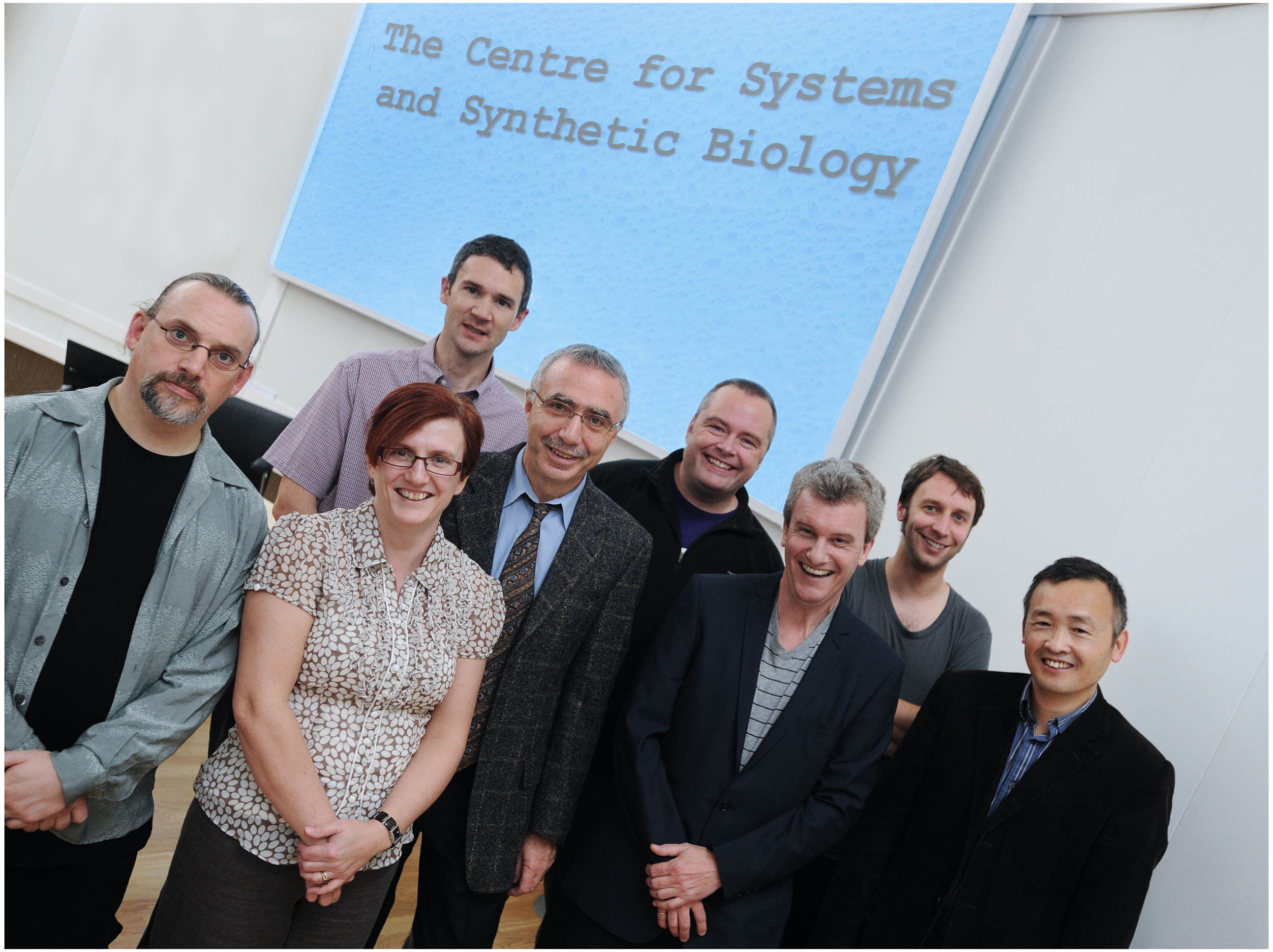
- Pam Gao
- Zujian Wu
- Maciej Trybilo
- Chris Paterson
- Huma Lodhi
- Crina Grosnan
- Hui Liu
- Mark Pook
- Annette Payne
- Amanda Harvey
- David Tree
- Simon Taylor
- Stephen Swift
- Allan Tucker
- Monika Heiner
- Rainer Breitling
- Mike Barrett
- Vlad Vyshemirsky
- Muffy Calder
- Walter Kolch
- Richard 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



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