BME Tutorial – Part 1

A conceptual framework for BioModel Engineering (Systems Biology, Synthetic Biology)

Rainer Breitling, Groningen, NL

David Gilbert, Brunel, UK

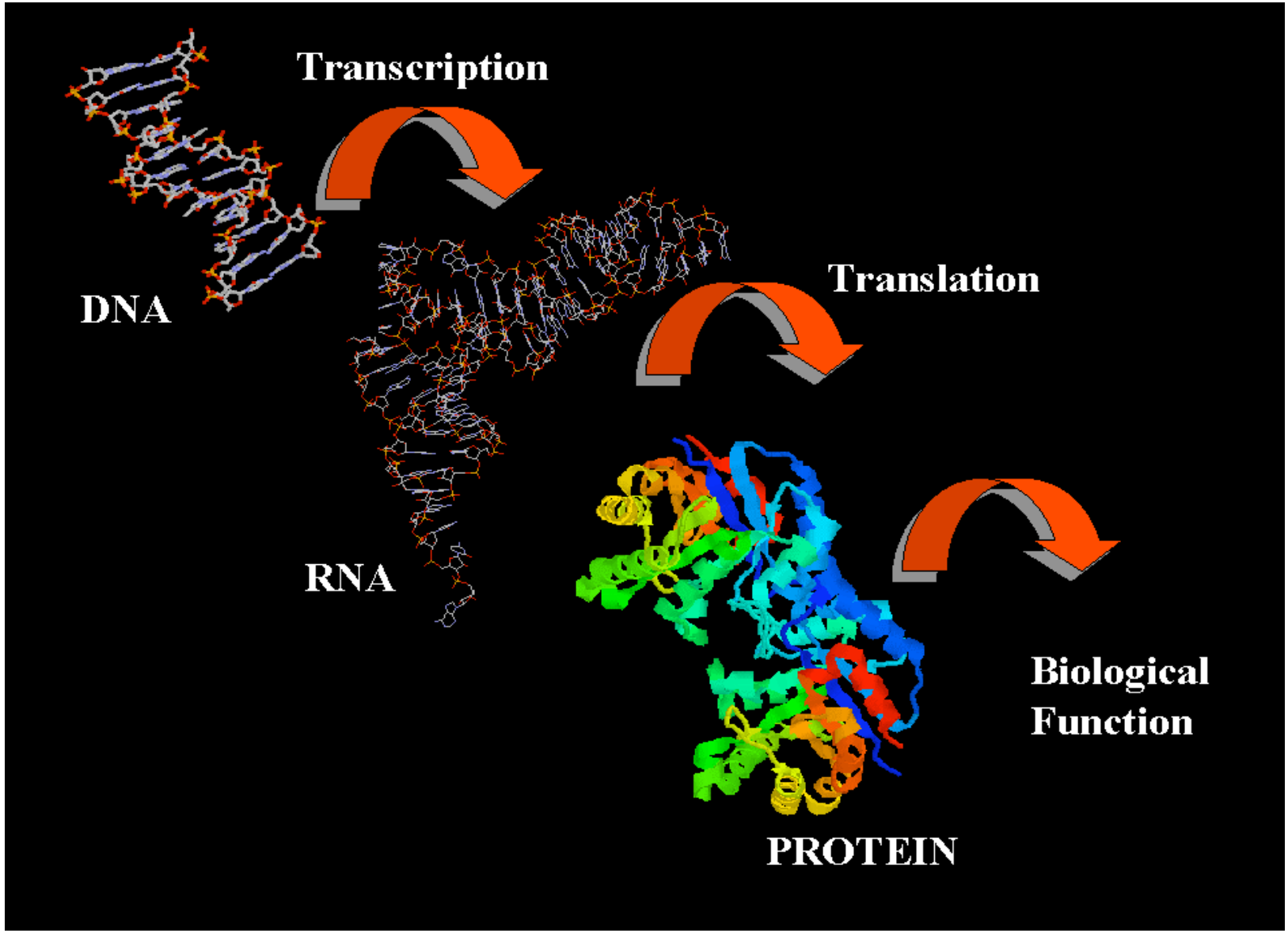
Monika Heiner, Cottbus, DE

This session

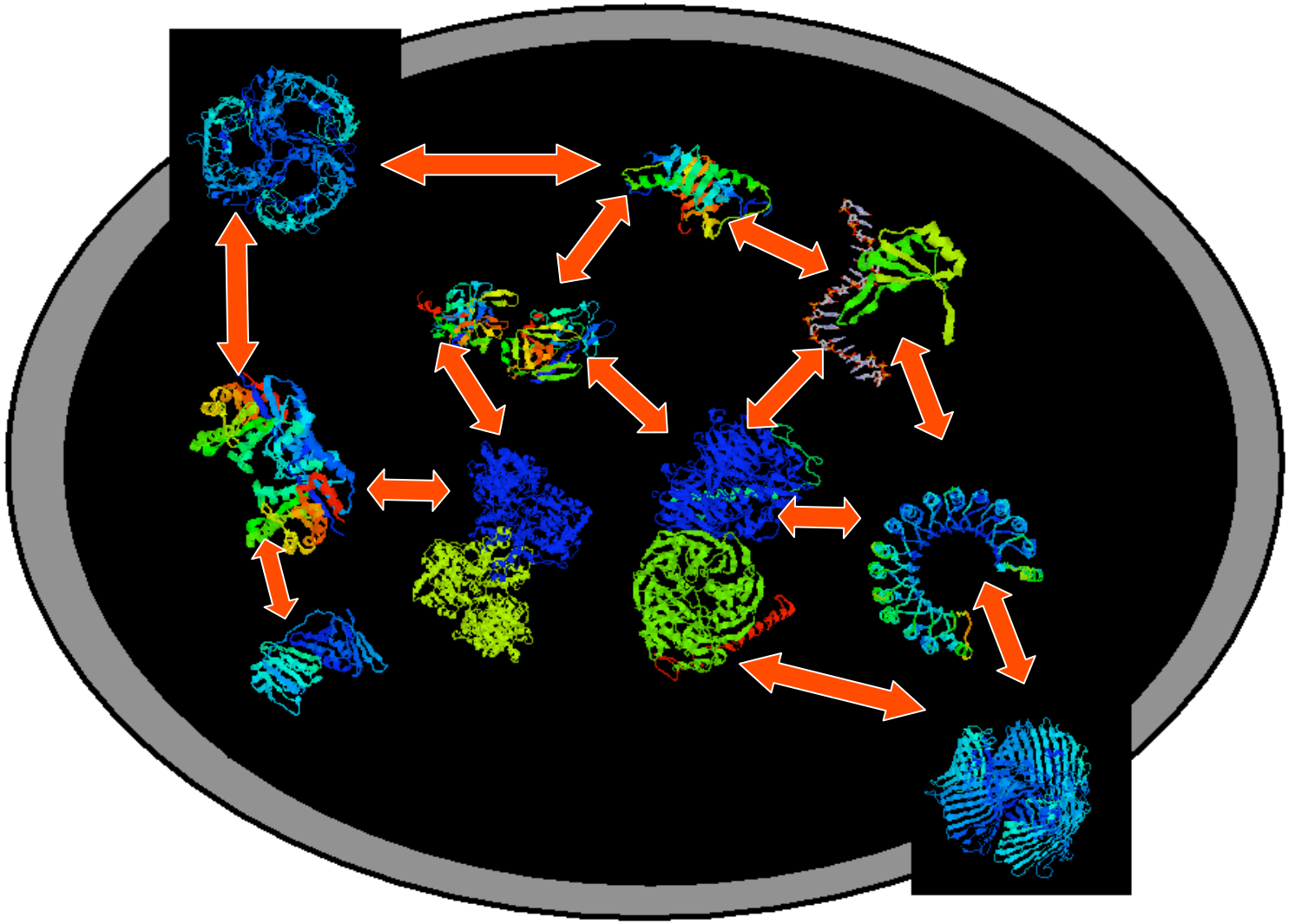
- Introduction: Systems biology, synthetic biology
- Conceptual modelling framework
- BioModel Engineering

Funding from the EU

Bioinformatics Basics



Systems Biology: Interaction in *Networks*



Chemical Reaction

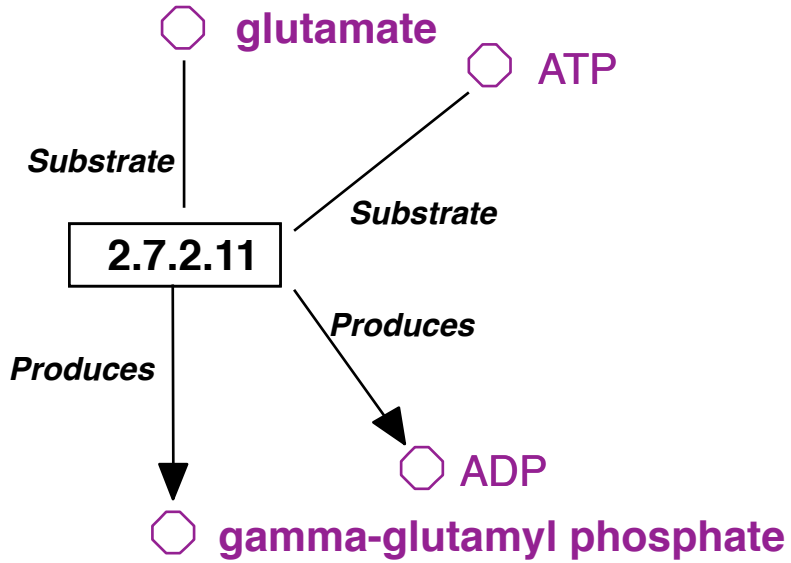
Set of Biochemical Entities

(substrates)

-o [Reaction] ->

Set of Biochemical Entities

(products)

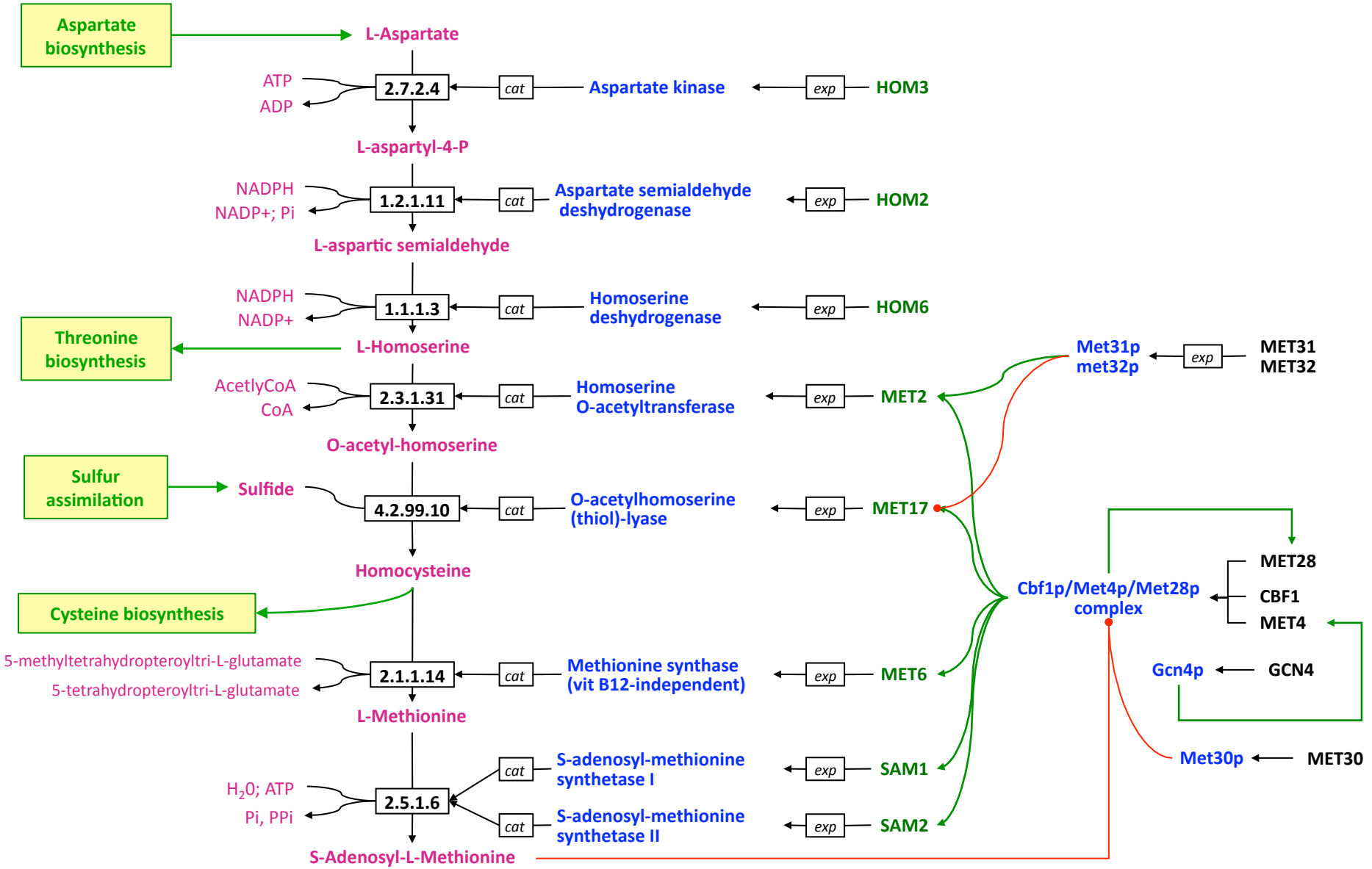


1.5.1.2 **EC (reaction) number**

compound

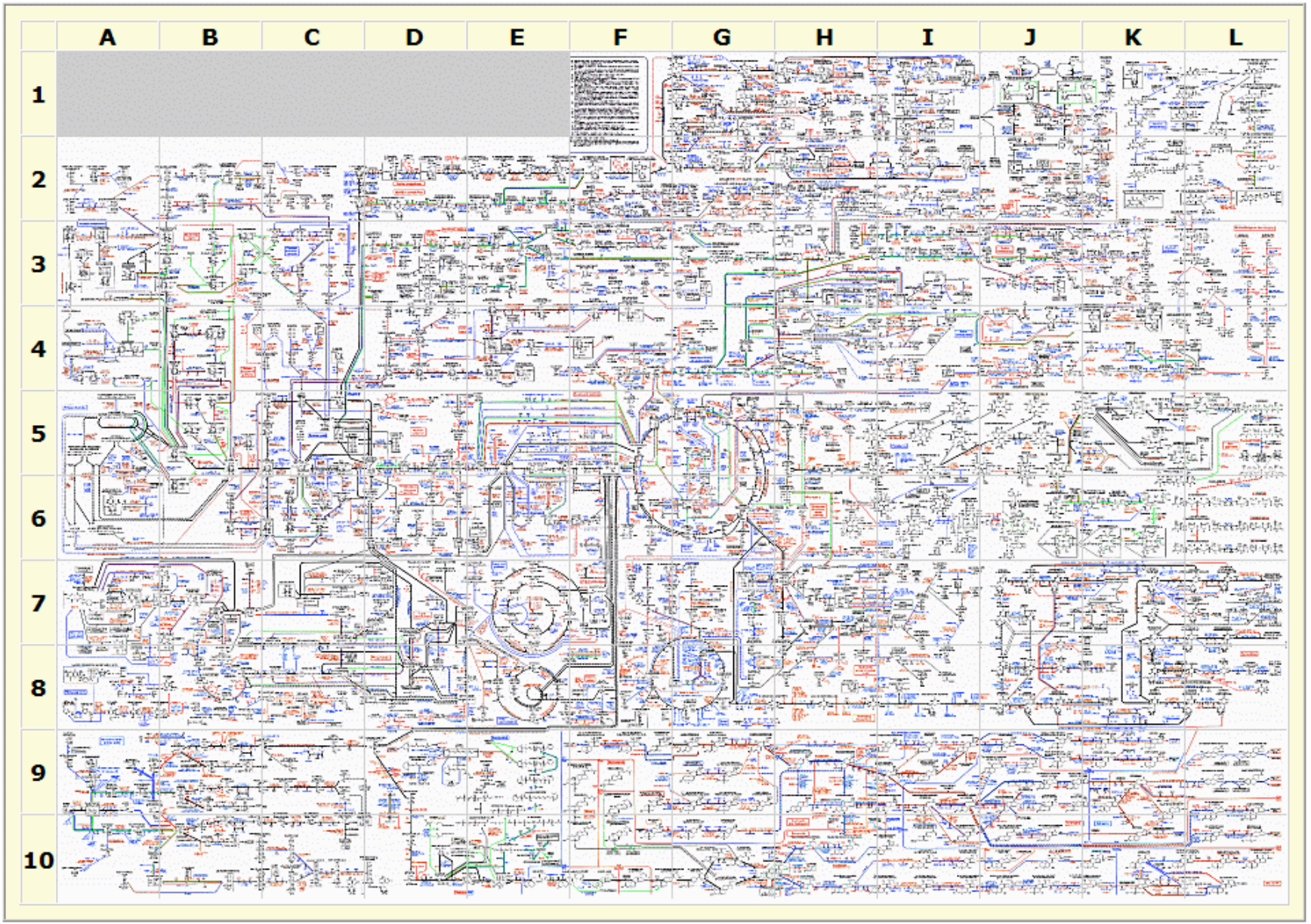
Slide from Jacques van Helden

Methionine Biosynthesis in S.cerevisiae

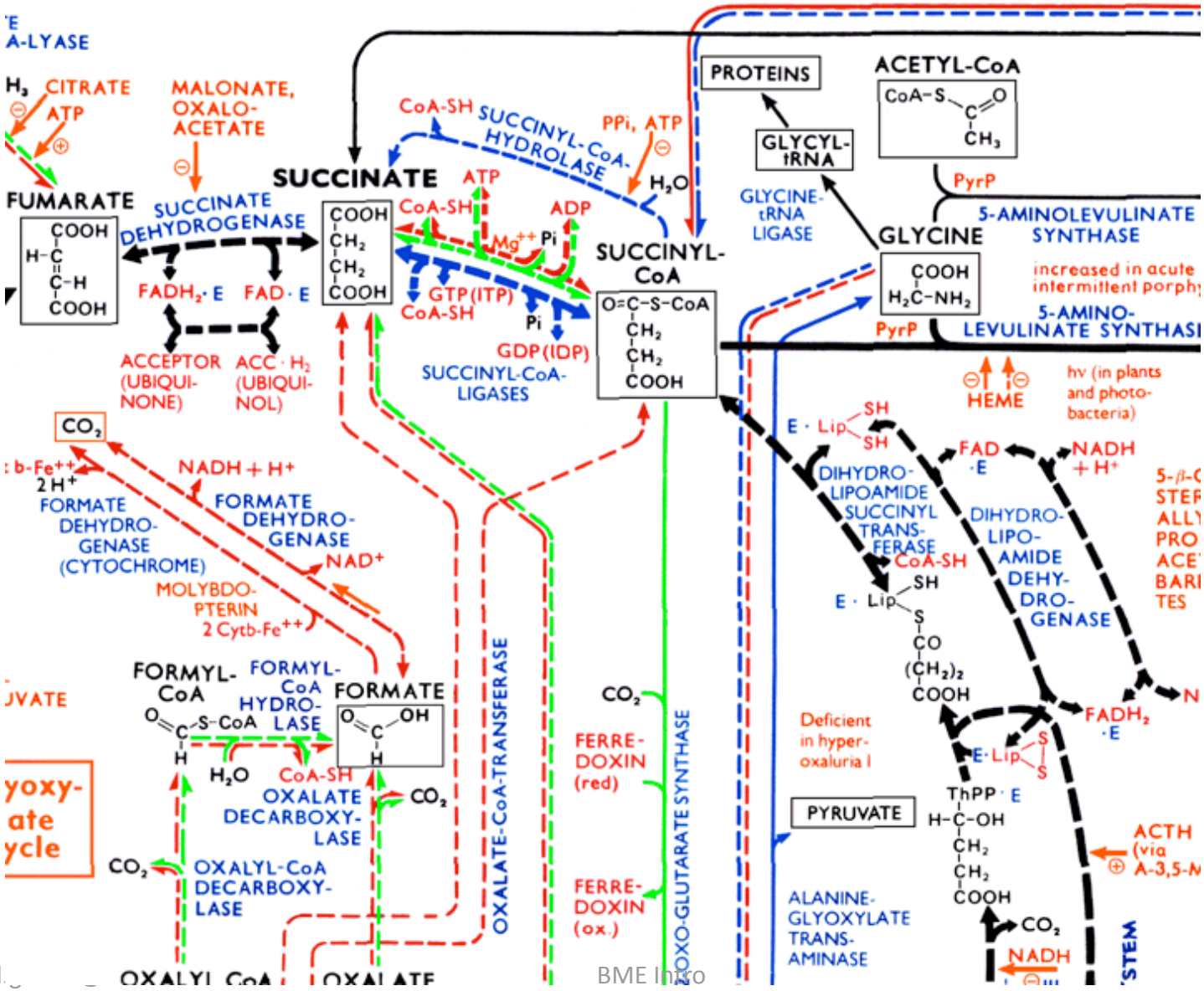


Slide from Jacques van Helden

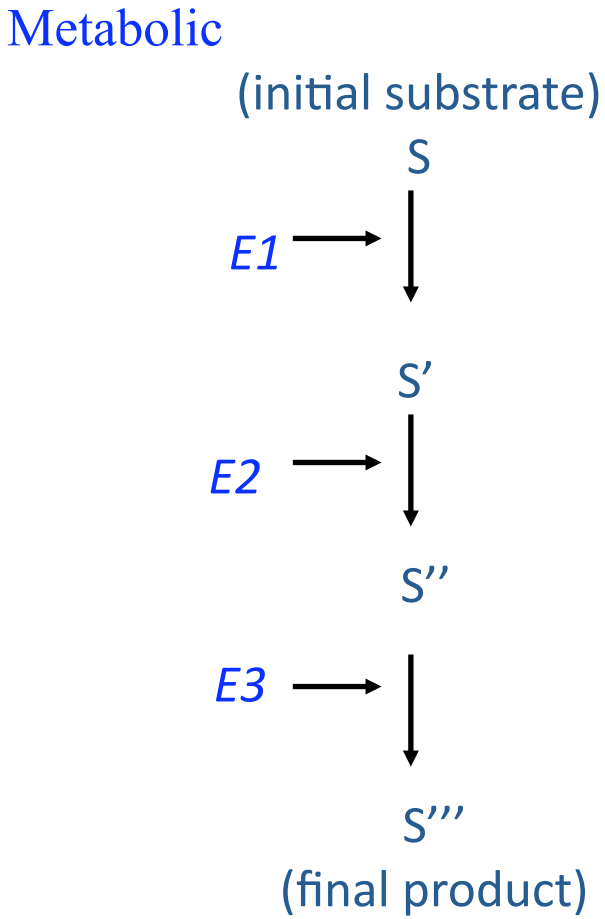
Metabolic Pathways



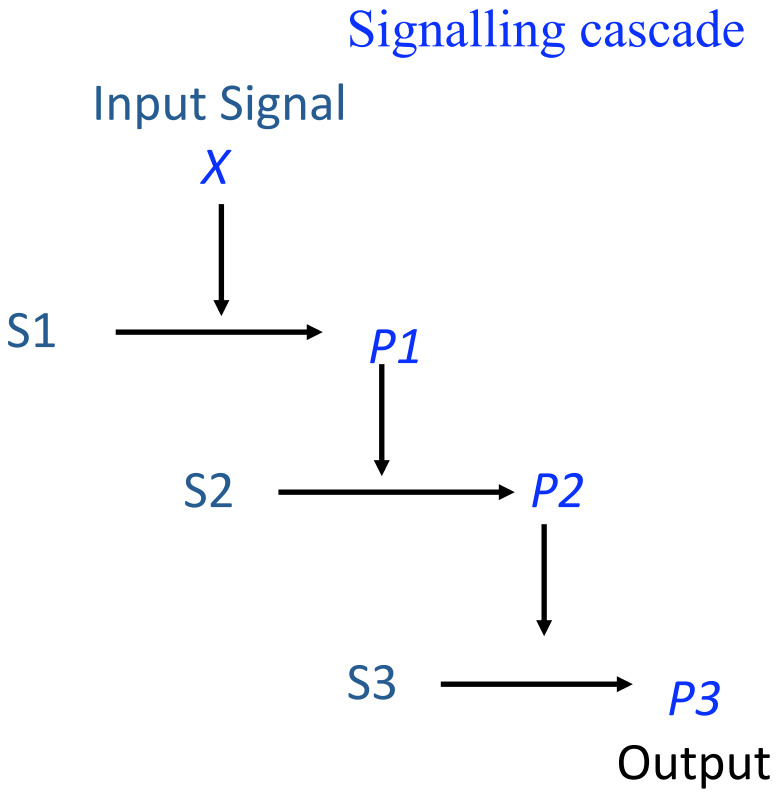
→ general biochemical pathways, → animals,
→ higher plants, → unicellular organisms



Metabolic Pathways vs Signalling Pathways



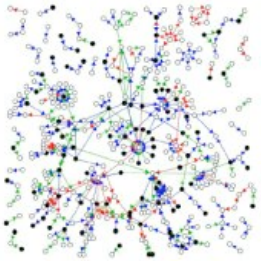
Classical enzyme-product pathway



Product become enzyme at next stage

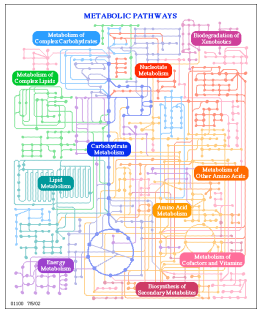
Networks

- Gene regulation



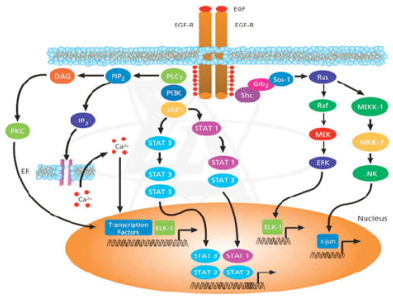
- Protein-protein interaction

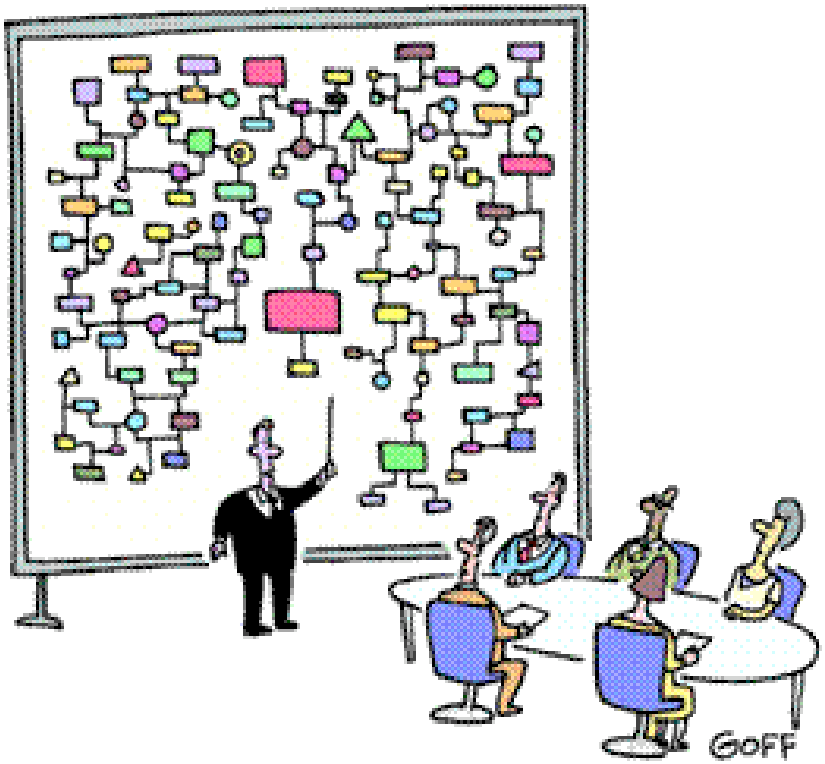
- Metabolic



- Developmental

- Signalling

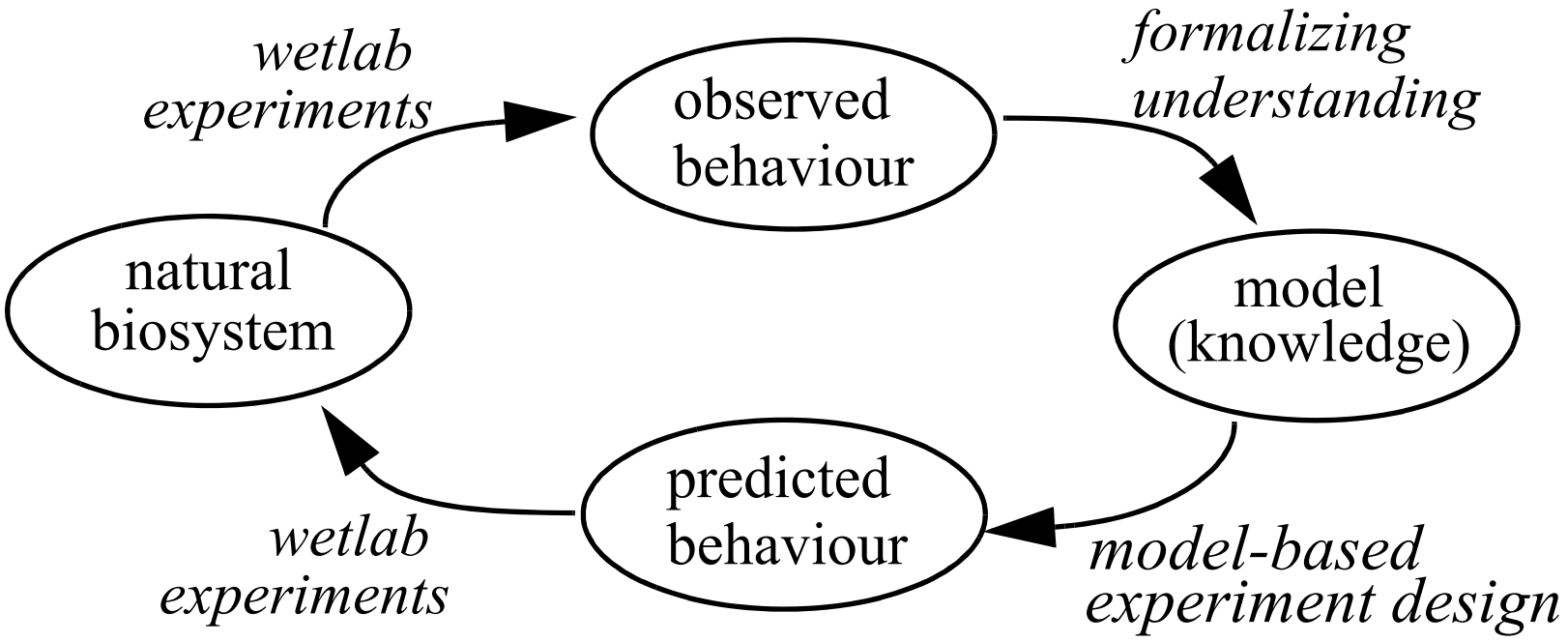




"And that's why we need a computer."

Systems Biology

- Modelling: design and construction of models of existing biological systems, which explain observed properties and predict the response to experimental interventions



Systems Biology – some definitions

- Systems biology is the study of all the elements in a biological system (all genes, mRNAs, proteins, etc) and their relationships one to another in response to perturbations.
- Systems approaches attempt to study the behaviour of all of the elements in a system and relate these behaviours to the systems or emergent properties

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

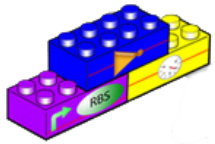
Synthetic Biology - Design & Build it!

Genetic Engineering

- Single gene manipulation
- Gene transfer
- Ad-hoc research

Synthetic Biology

- Genetic circuit design
- Standard parts – BioBricks™
- Modelling and simulation
- Open

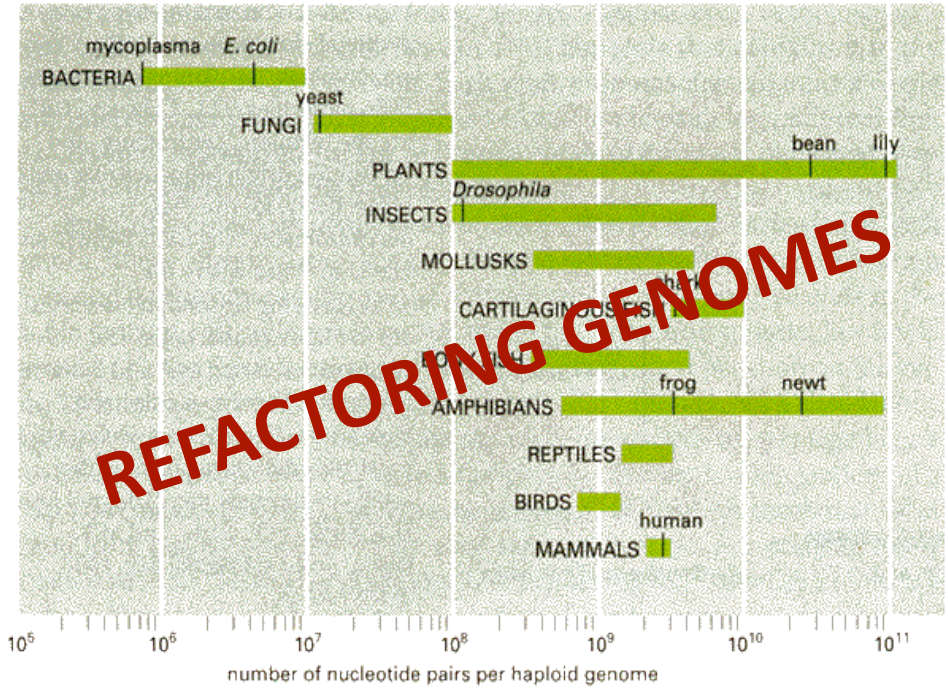


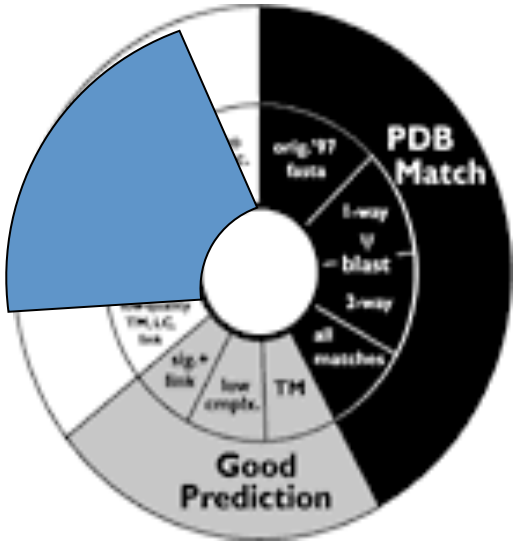
*

Supporting technology: DNA synthesis

* Photographic bacteria - UCSF iGEM 2004 Team

Top-Down- The “North American” model

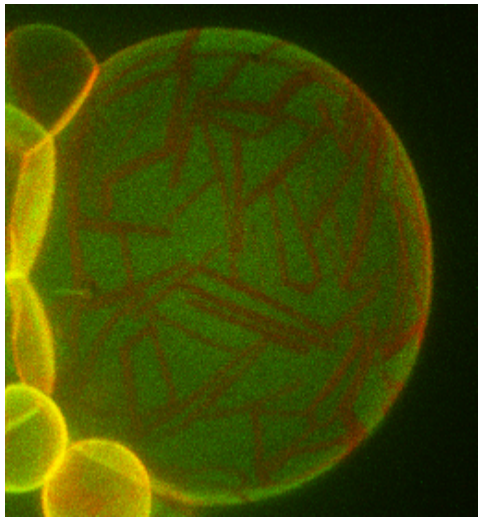




Mycoplamsa genitalium
Genome 580kbp

PNAS 103, 425-430, 2006

Bottom-Up – The "Far-Eastern" Model



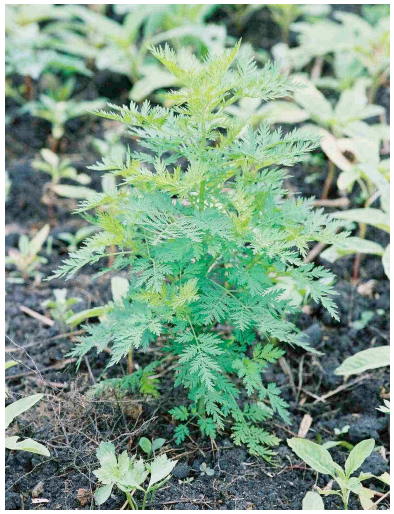
Introduce to the vesicle only what is needed for your uses

- Synthetic DNA
- Refactored organelles
- Membrane pores

Creation of artificial life!

A drug manufacturing plant

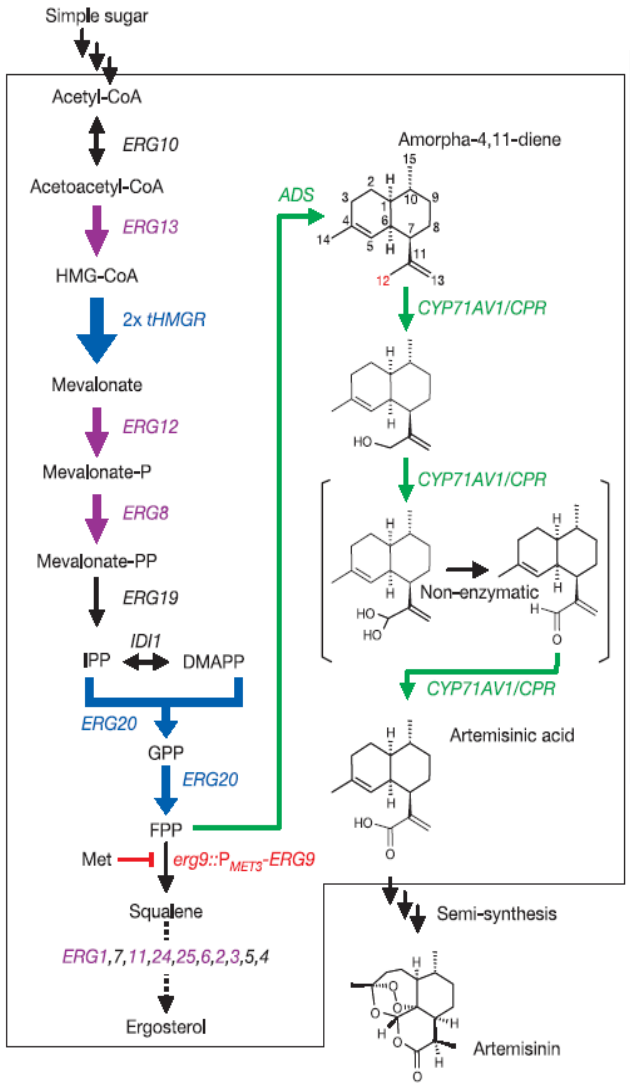
- "Audacious plan" New Scientist, May 2006
- Engineer e.coli / yeast to synthesise the anti-malarial artemisinin
- $42.6 million, Bill & Melinda Gates Foundation

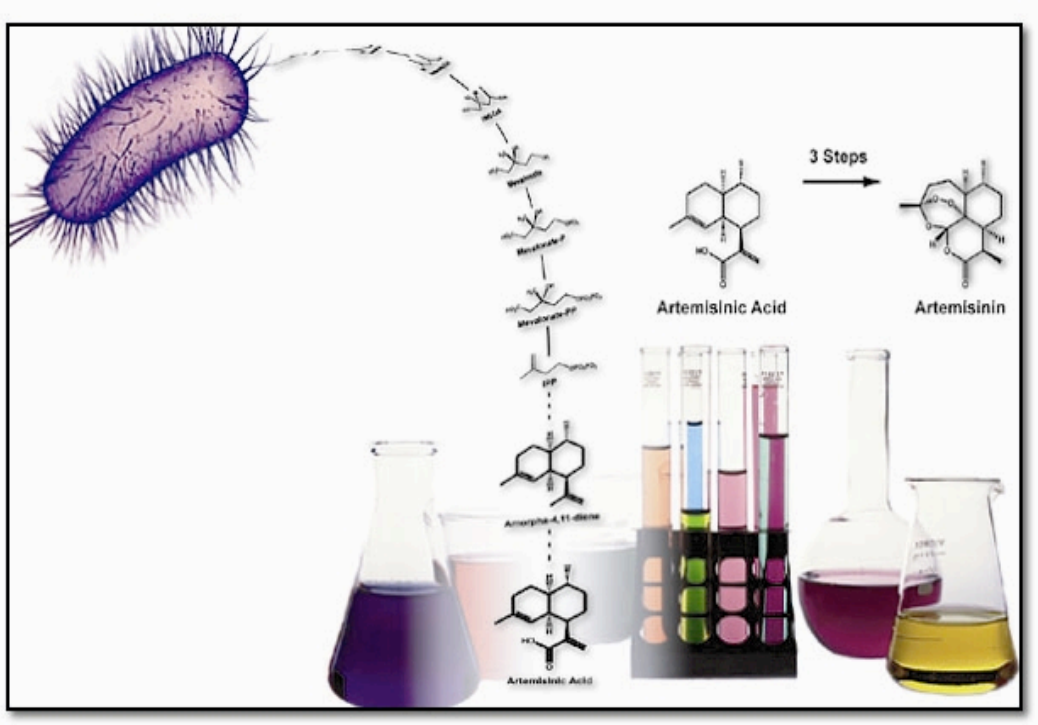


Artemisia annua

- Plant difficult to grow and only yield minute quantities of drug per kilo
- Artemisinin is expensive

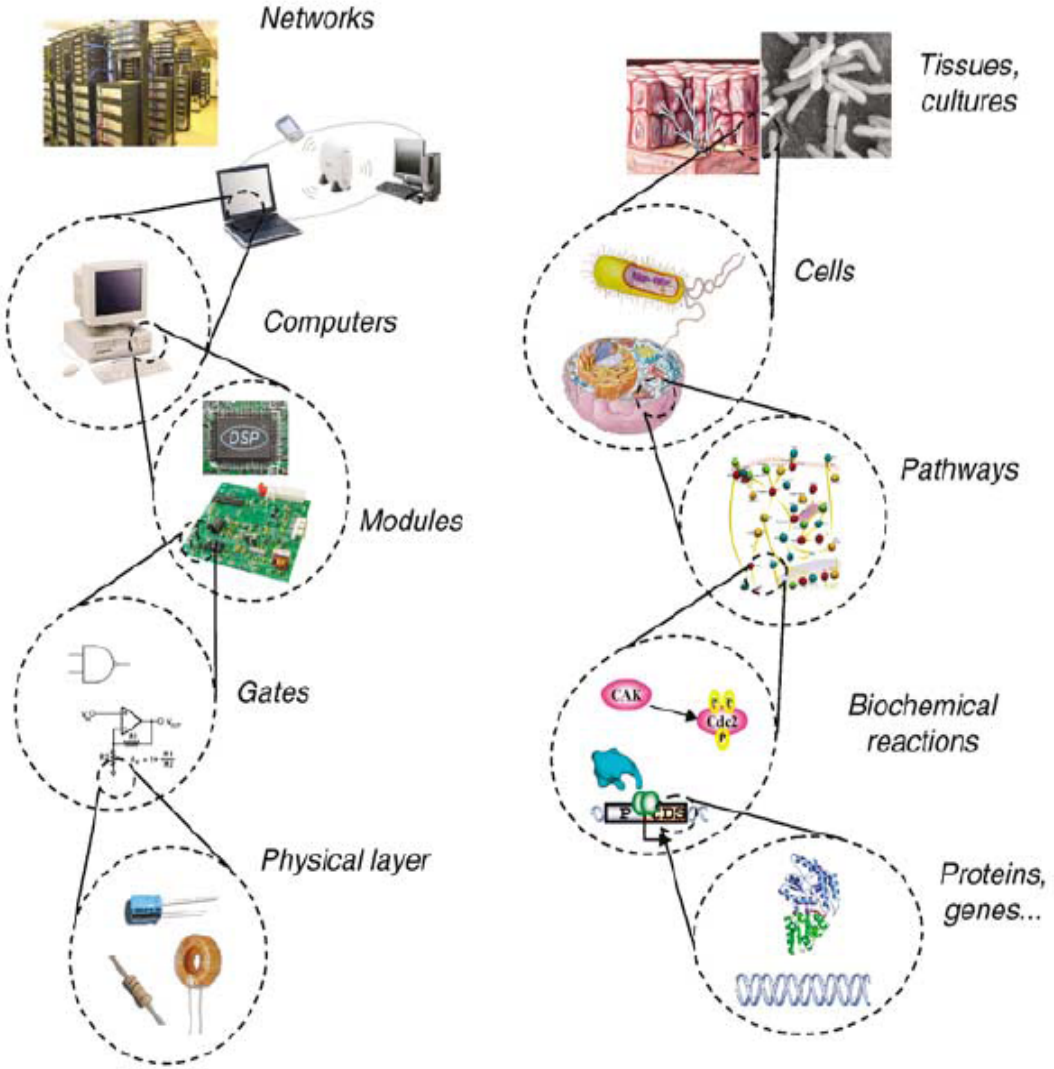
➢ Engineer cheaper alternative and save the world!





Saccharomyces cerevisiae

Levels of Abstraction



Systems

Devices

Parts

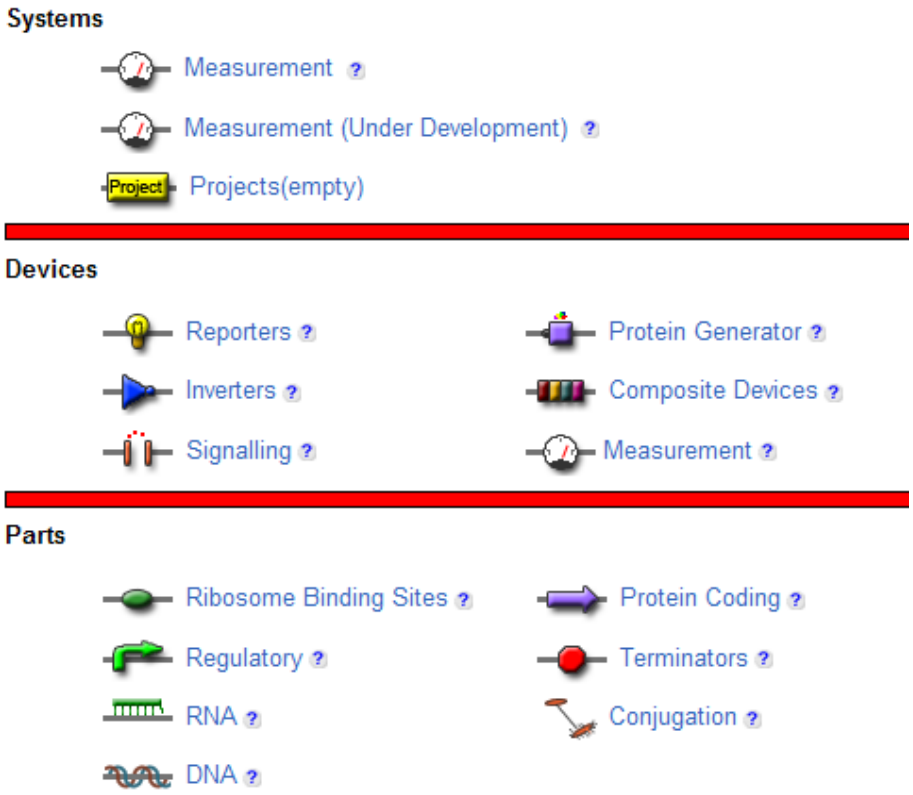
DNA

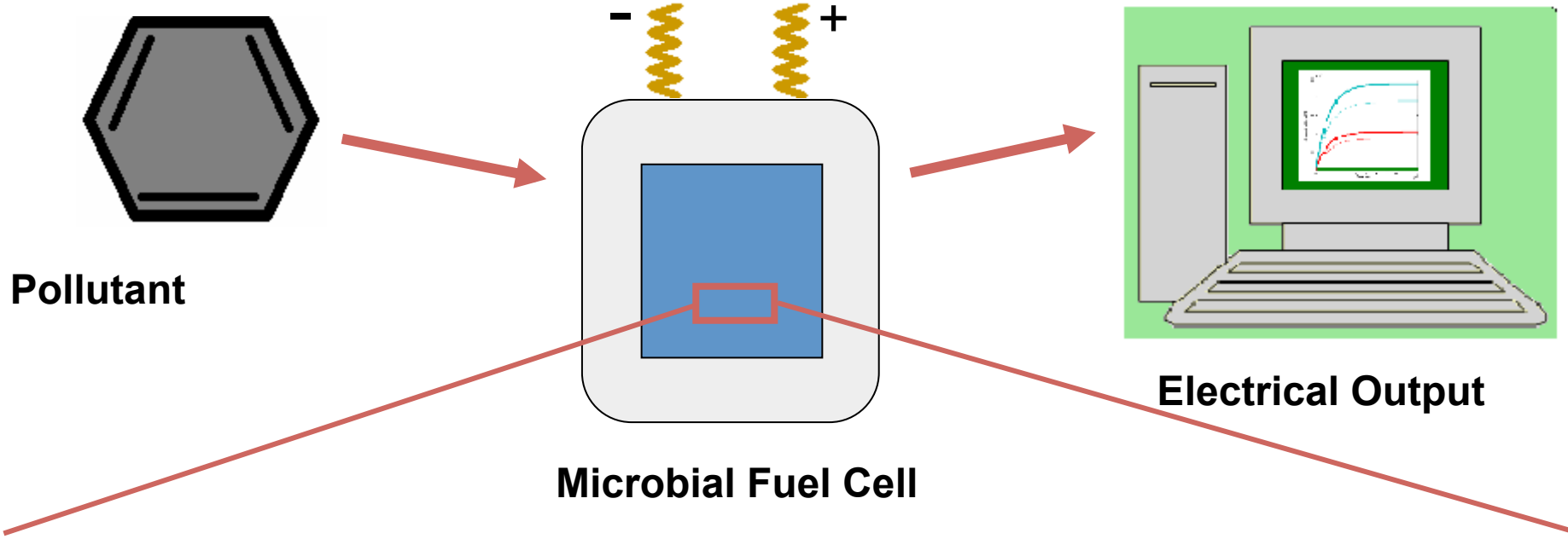
Adrianantoandro et al. Mol Sys Bio 2006

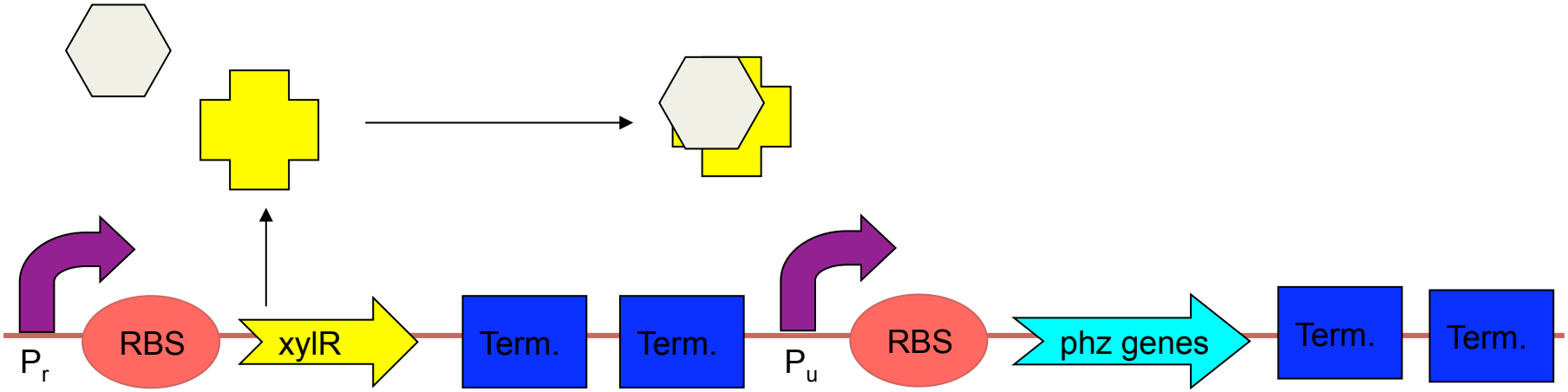
Registry of Parts

http://parts.mit.edu/

BBF continues work on the standard



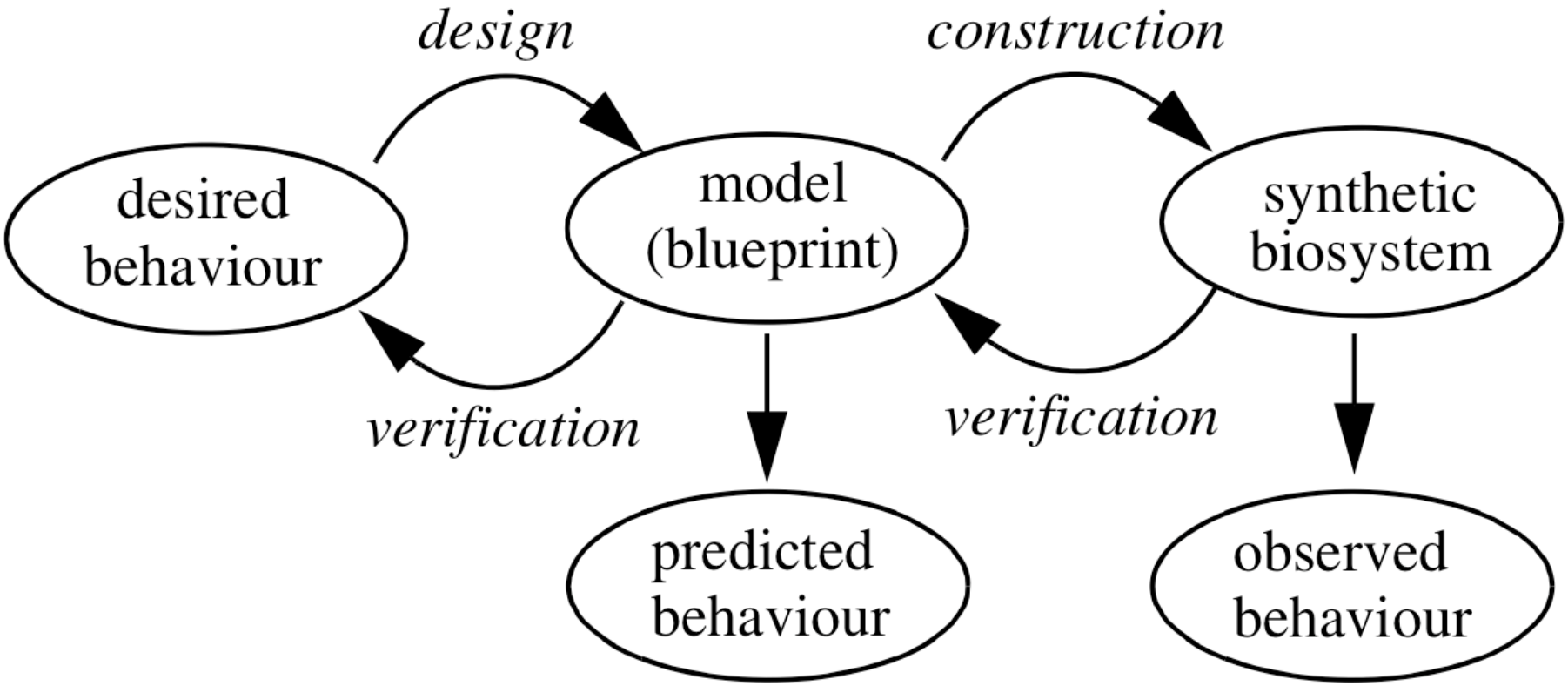




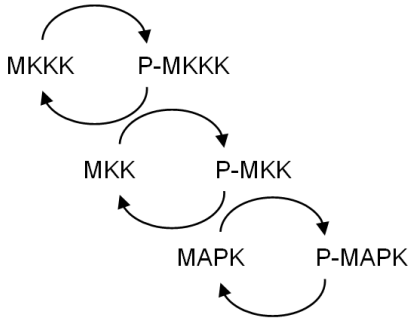
- From pathogenic *Pseudomonas aeruginosa*
- Can act as electron mediator

PYOCYANIN

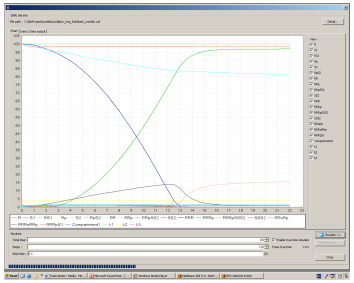
Synthetic Biology



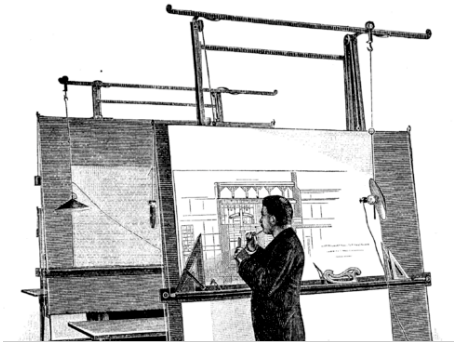
Model



- A model
  - formal representation of the real world
  - simplified abstract view of the complex reality.



- A simulation: implementation of a model over time.
- To design: the process of originating and developing a plan for a product



- A design: (Final) plan, e.g. model, description, for the product

Analysis & Reasoning

- A model may be used to permit (automated) reasoning about the object / system modelled.

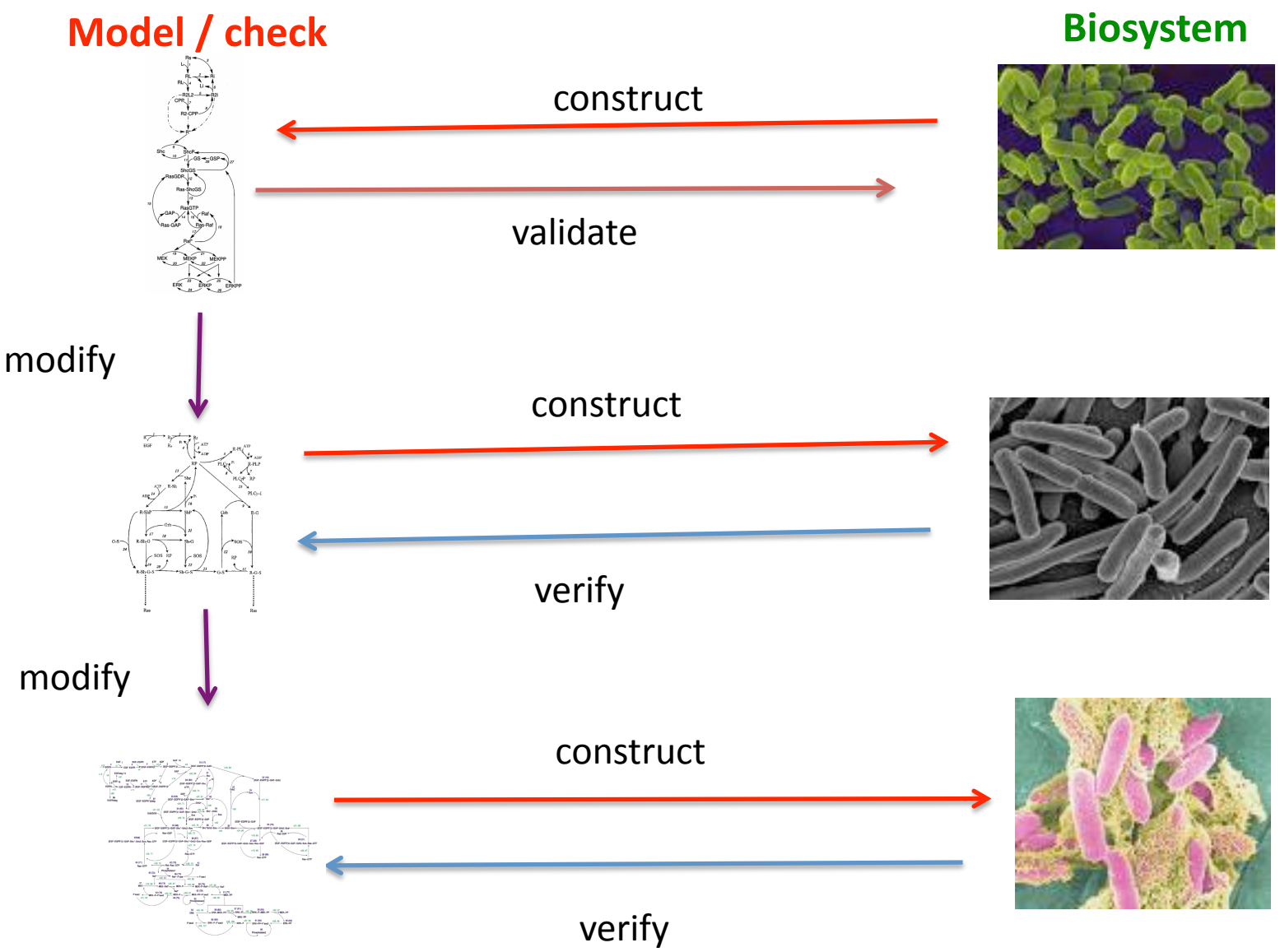
- *Predictive* modelling: the use of a model to predict the behaviour of a system.
  - E.g. predict the effect of drugs on an organism
  - E.g. predict the effect of an inhibitor on a pathway

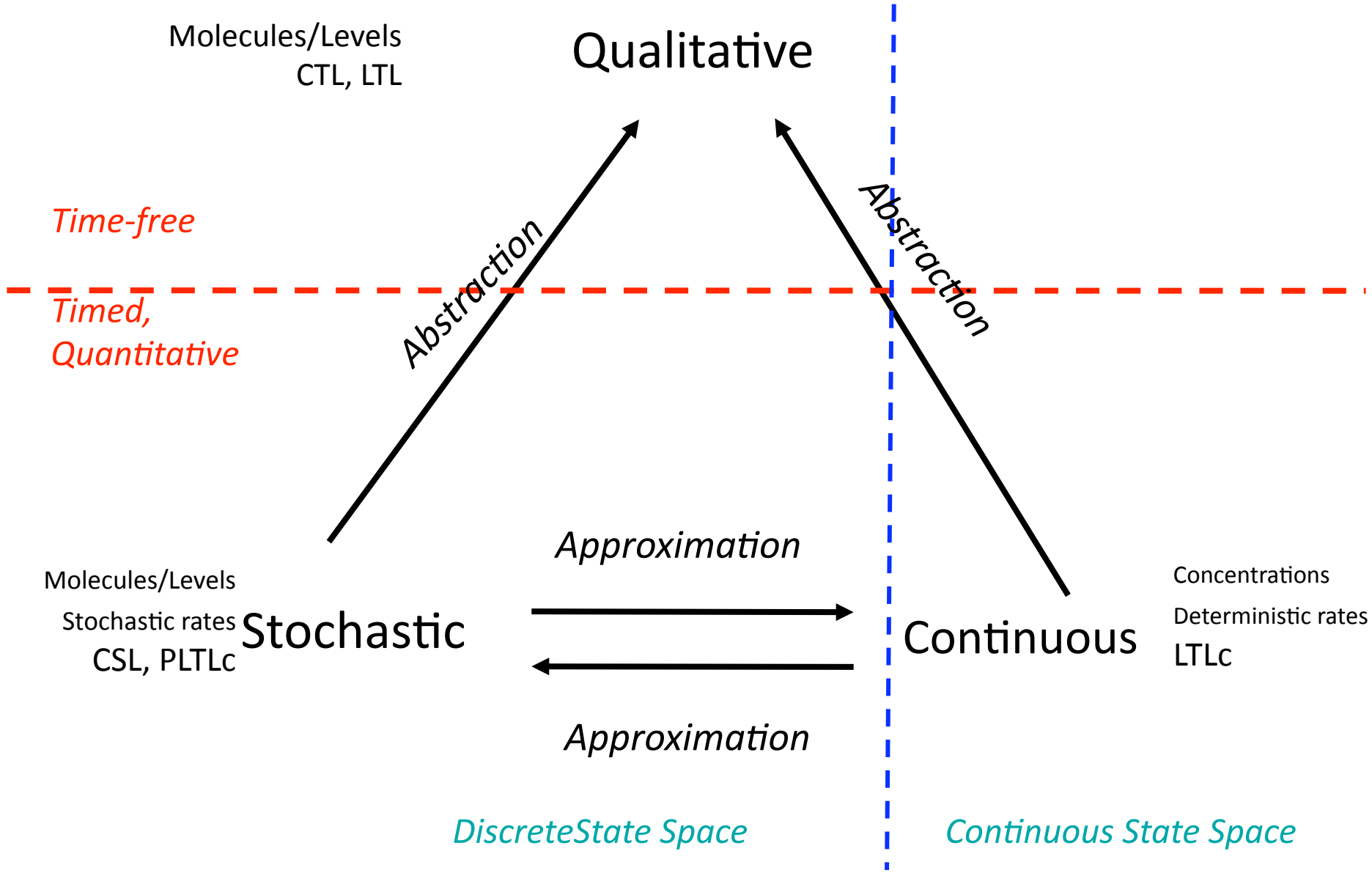
Validation & Verification

- **Validation** – '***You built the right product?***'.
  - Product / system accomplishes its intended requirements.
  - Model / simulation are accurate representations of the real world

- **Verification** - '***You built the product right?***'.
  - System complies with its specification
  - Model / simulation accurately represent the specifications

Synthetic Biology development cycle

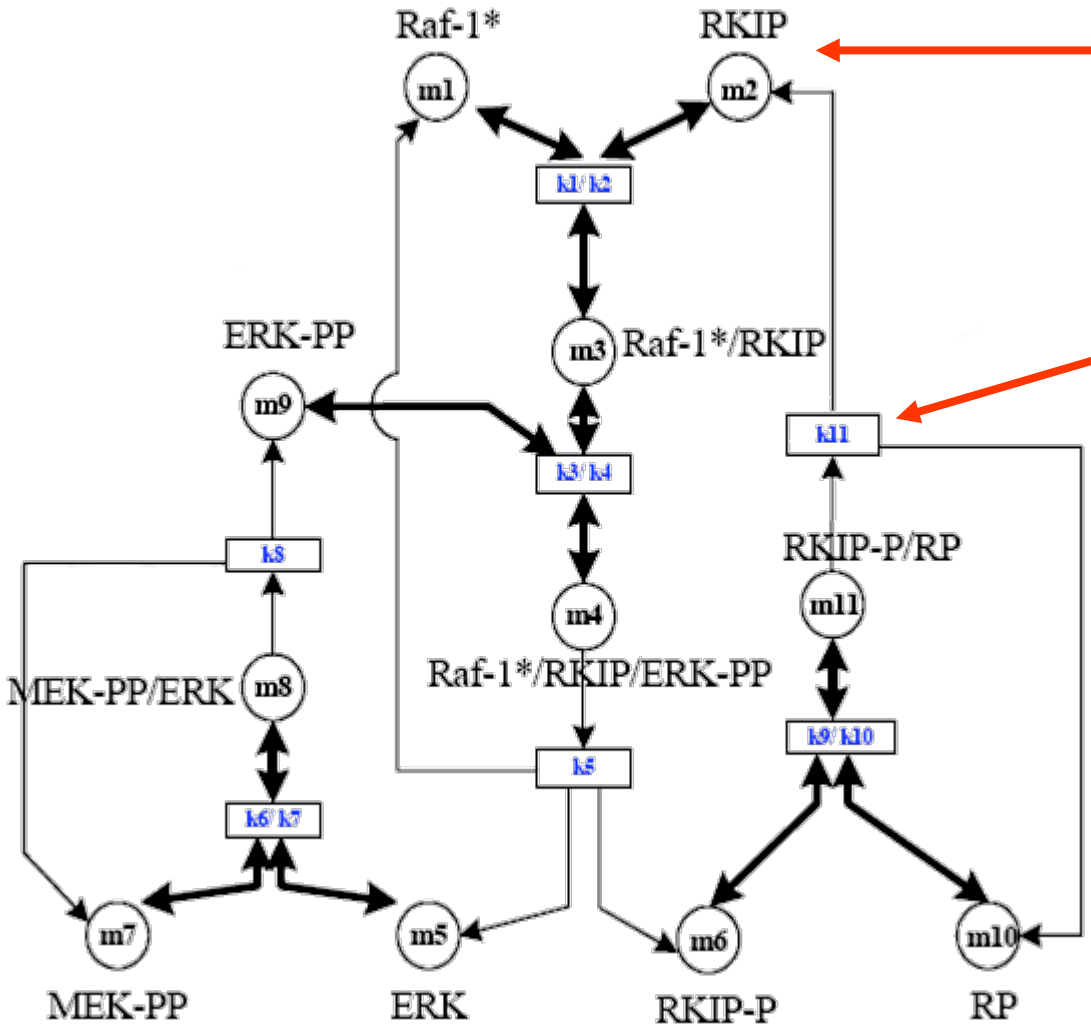




Gilbert, Heiner and Lehrack. ``A Unifying Framework for Modelling and Analysing Biochemical Pathways Using Petri Nets.” Proc CMSB 2007

Case study: small model network
RKIP inhibited ERK pathway

Cho et al, CMSB03

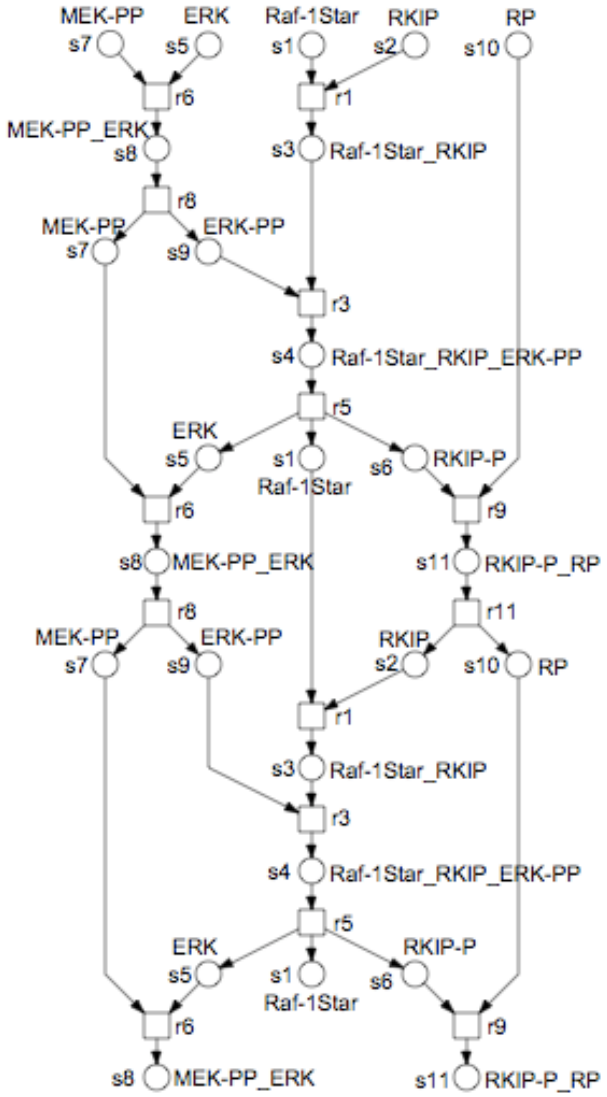


Protein
(concentration)

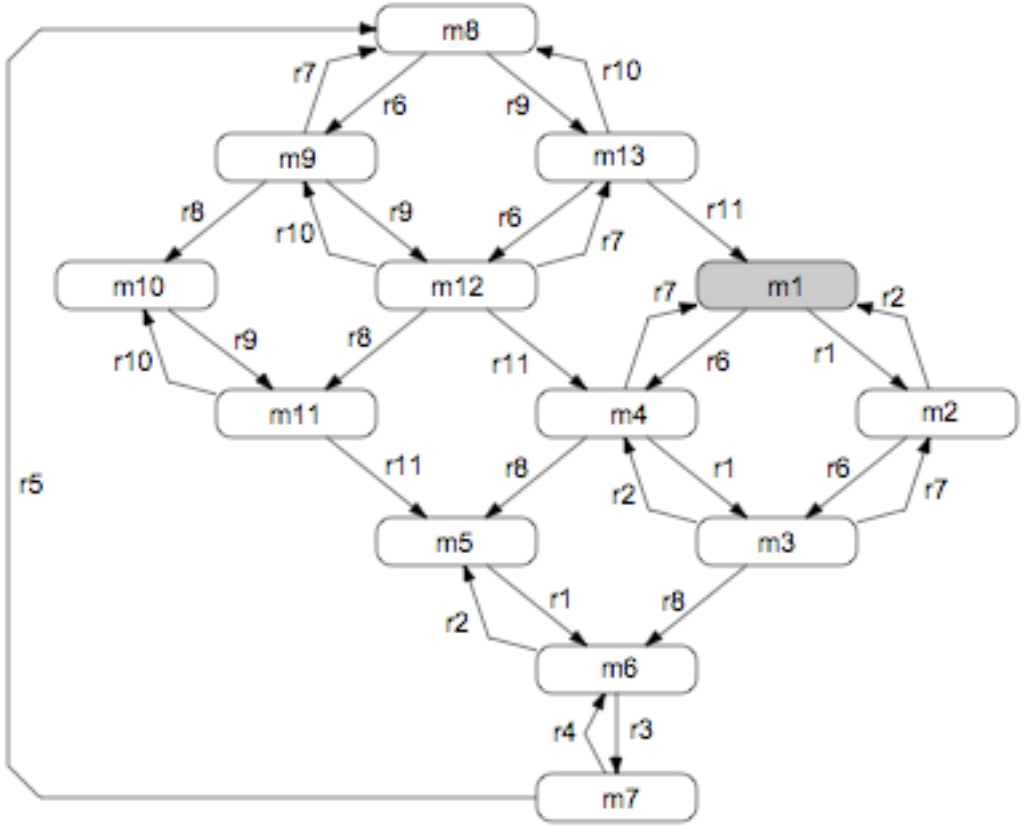
Reaction
+ rate

Qualitative Analysis

Partial order run



Reachability graph



Stochastic Definition

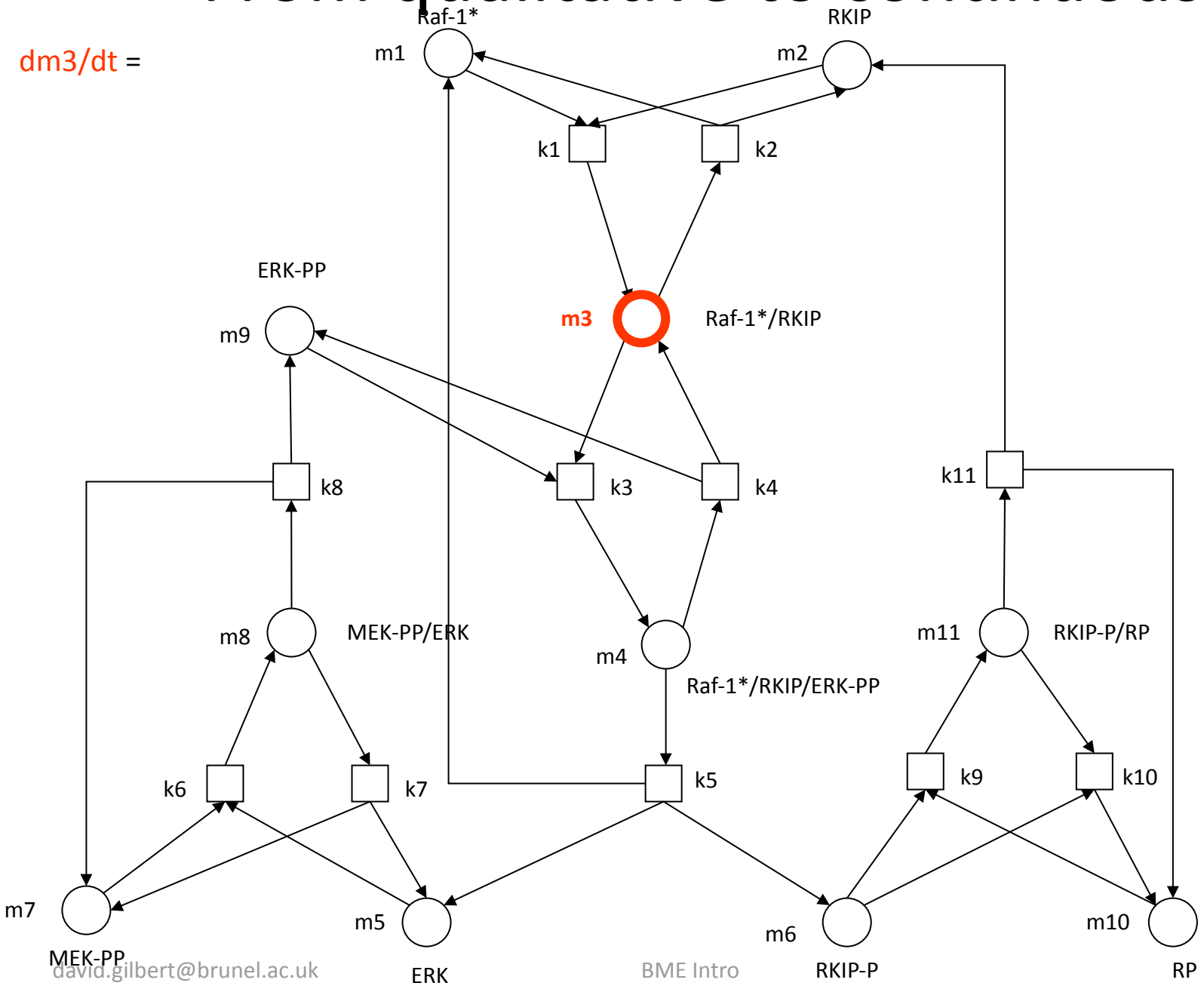
Table 1 The reaction equation, rate function, and rate constants for each reaction (transition). For better readability we use the abbreviations s1 ... s11 for the involved species. All reactions employ mass action kinetics. Backward reactions constants are by two orders of magnitude smaller than for the forward reactions.

#	reaction equation			rate function v_i	rate constant [a] stochastic	deterministic
r1	s1 + s2	$\rightarrow$	s3	$c_1 \cdot$ s1 $\cdot$ s2	$c'_1 = c_1 \cdot f_s$	$c_1 = 0.53$
r2	s3	$\rightarrow$	s1 + s2	$c_2 \cdot$ s3	$c'_2 = c_2$	$c_2 = 0.0072$
r3	s3 + s9	$\rightarrow$	s4	$c_3 \cdot$ s3 $\cdot$ s9	$c'_3 = c_3 \cdot f_s$	$c_3 = 0.625$
r4	s4	$\rightarrow$	s3 + s9	$c_4 \cdot$ s4	$c'_4 = c_4$	$c_4 = 0.00245$
r5	s4	$\rightarrow$	s1 + s5 + s6	$c_5 \cdot$ s4	$c'_5 = c_5$	$c_5 = 0.0315$
r6	s5 + s7	$\rightarrow$	s8	$c_6 \cdot$ s5 $\cdot$ s7	$c'_6 = c_6 \cdot f_s$	$c_6 = 0.8$
r7	s8	$\rightarrow$	s5 + s7	$c_7 \cdot$ s8	$c'_7 = c_7$	$c_7 = 0.0075$
r8	s8	$\rightarrow$	s7 + s9	$c_8 \cdot$ s8	$c'_8 = c_8$	$c_8 = 0.071$
r9	s6 + s10	$\rightarrow$	s11	$c_9 \cdot$ s6 $\cdot$ s10	$c'_9 = c_9 \cdot f_s$	$c_9 = 0.92$
r10	s11	$\rightarrow$	s6 + s10	$c_{10} \cdot$ s11	$c'_{10} = c_{10}$	$c_{10} = 0.00122$
r11	s11	$\rightarrow$	s2 + s10	$c_{11} \cdot$ s11	$c'_{11} = c_{11}$	$c_{11} = 0.87$

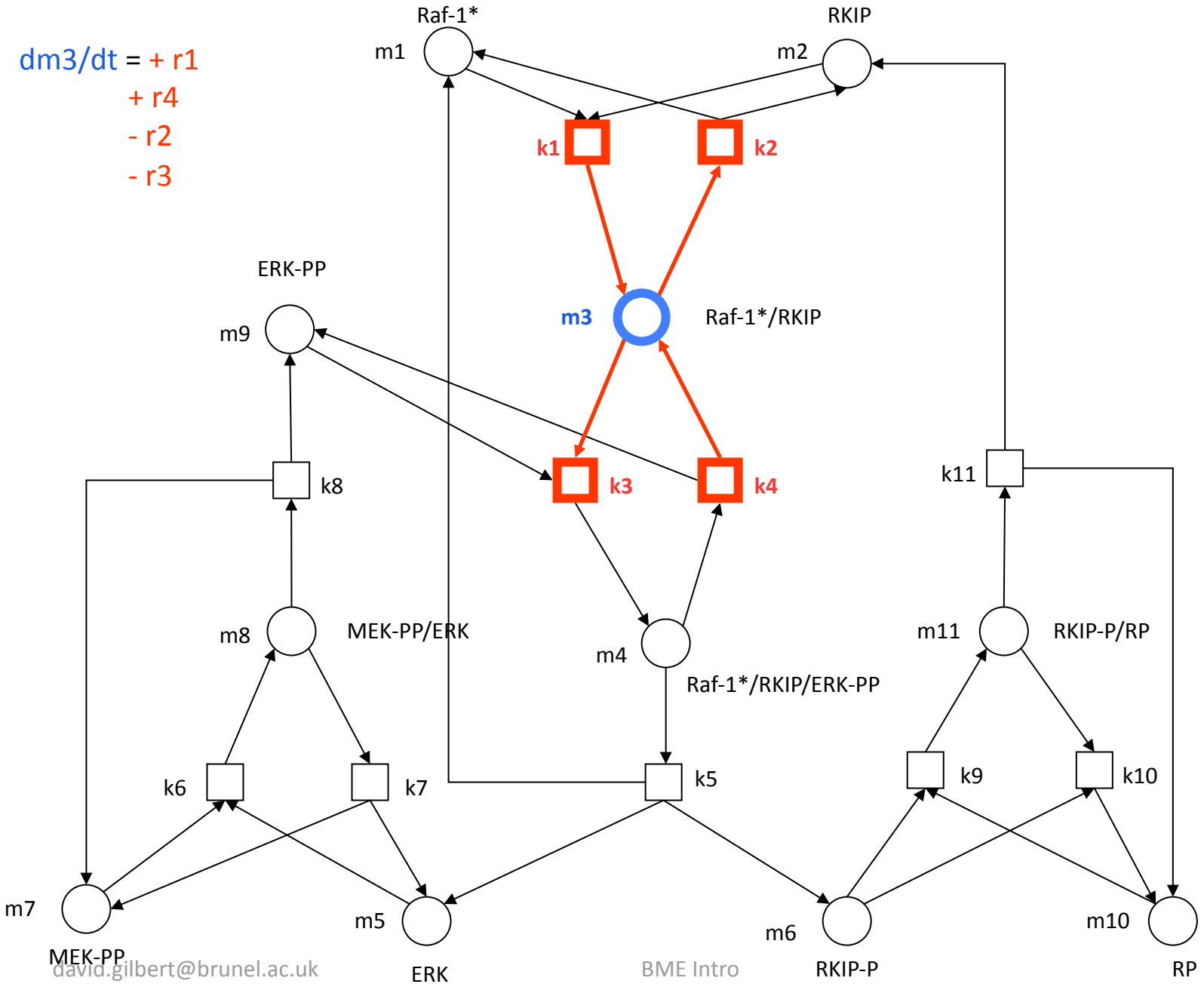
[a] The stochastic and deterministic rate constants are equivalent for first-order reactions. f_s is a scaling factor to map the given *mass* in the continuous concentration onto a finite number of levels (i.e tokens), with N being the highest level number, i.e. $f_s = mass/N$.

From qualitative to continuous

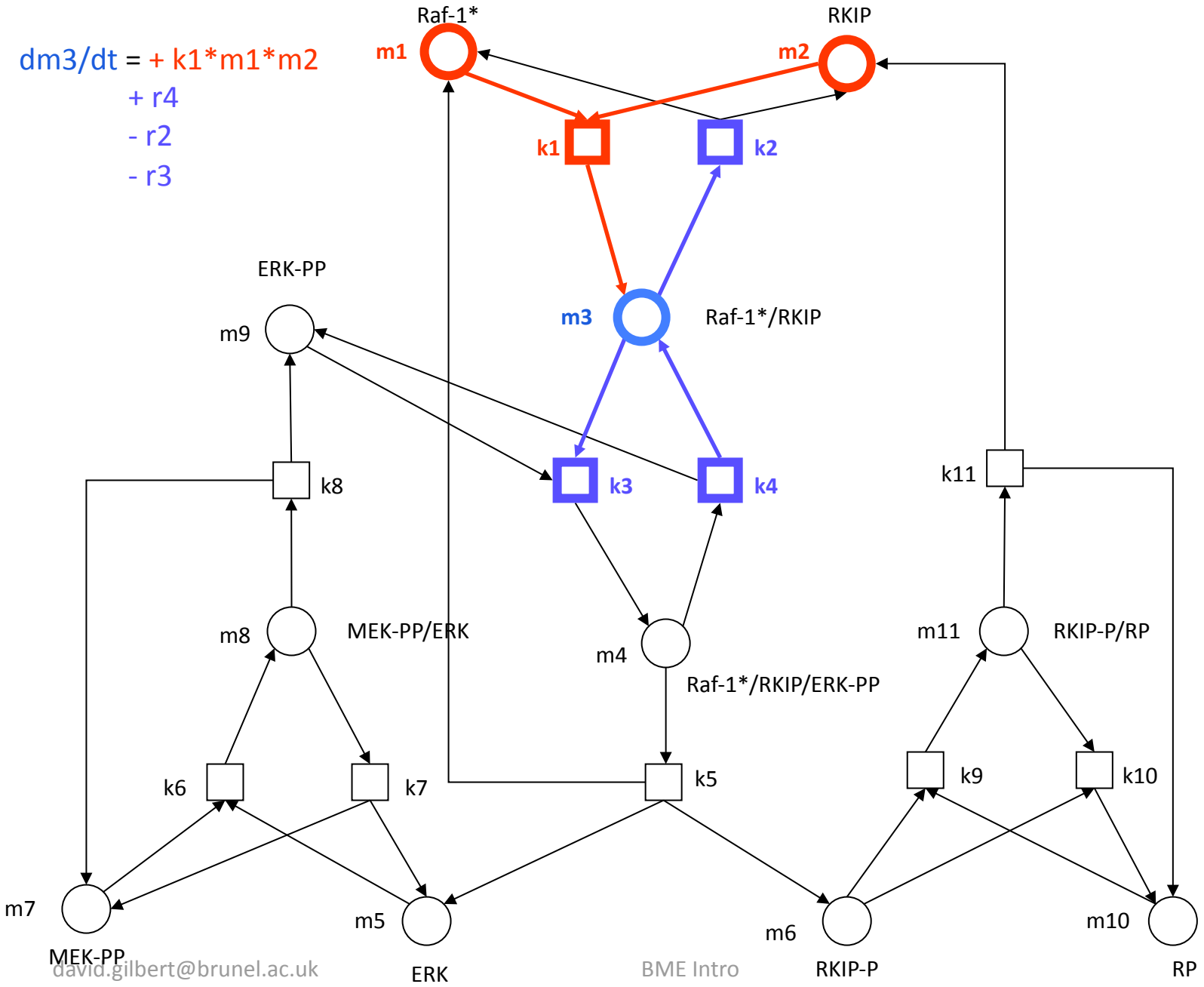
dm3/dt =



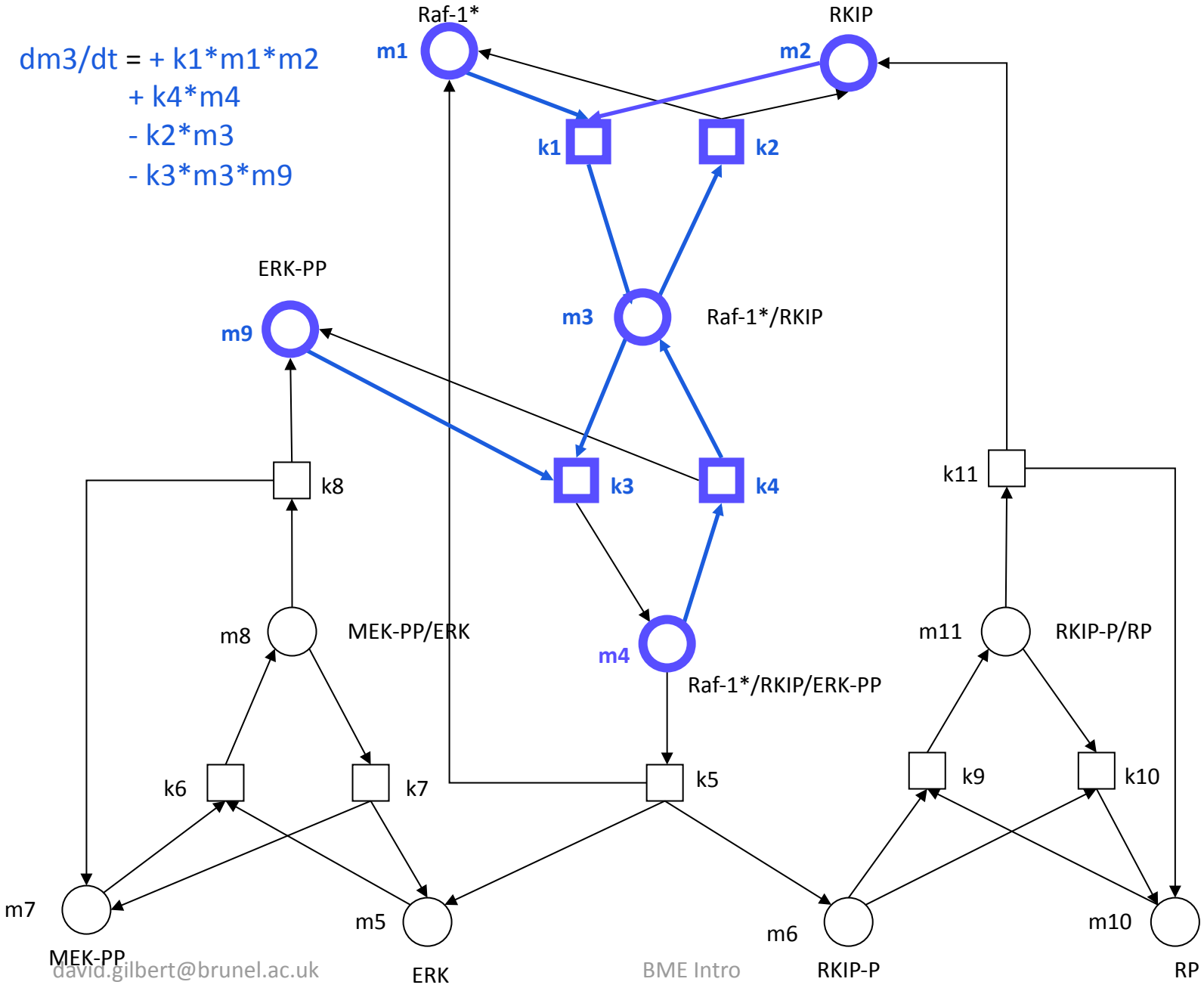
dm3/dt = + r1
+ r4
- r2
- r3

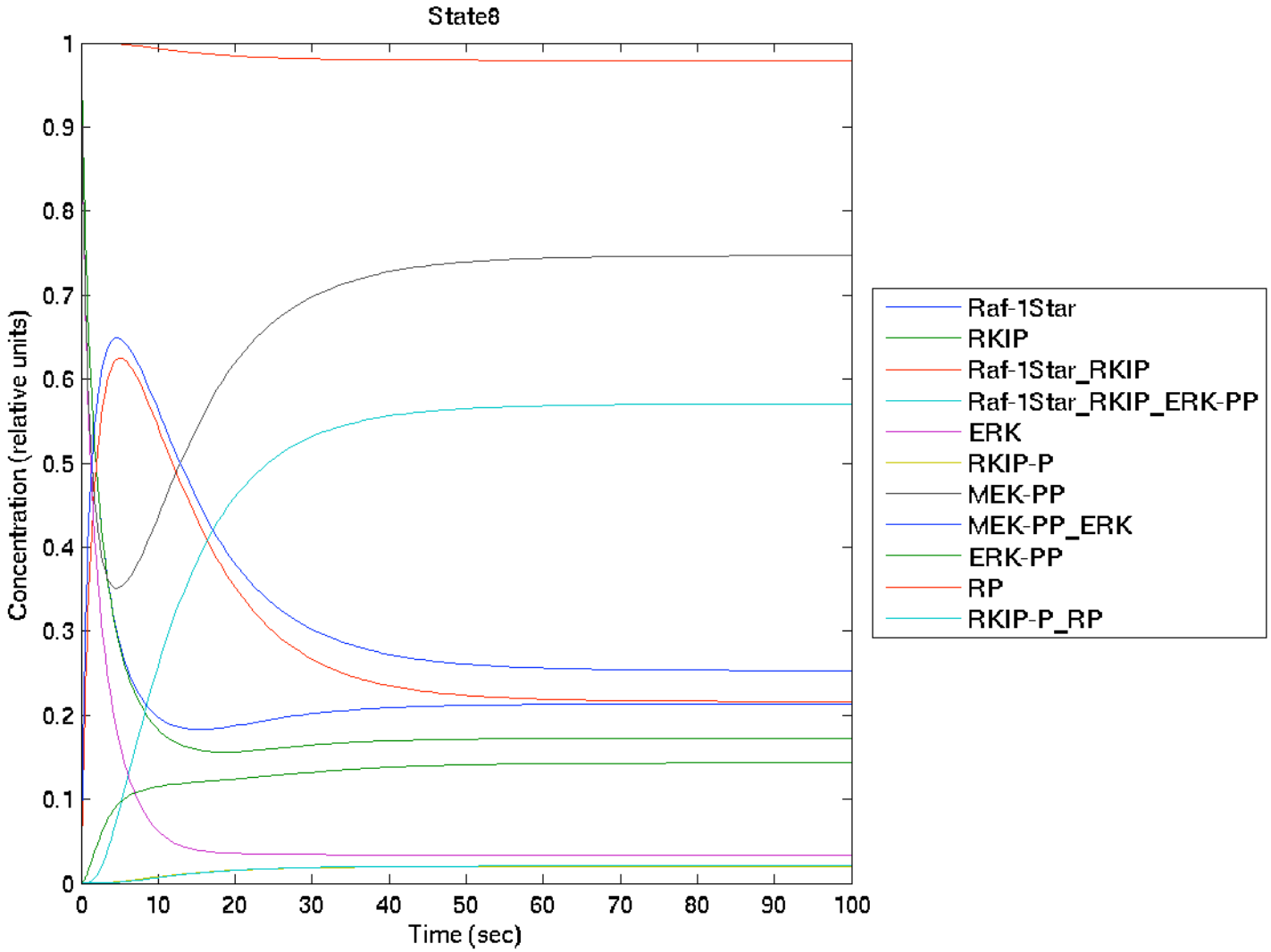


dm3/dt = + k1*m1*m2
+ r4
- r2
- r3

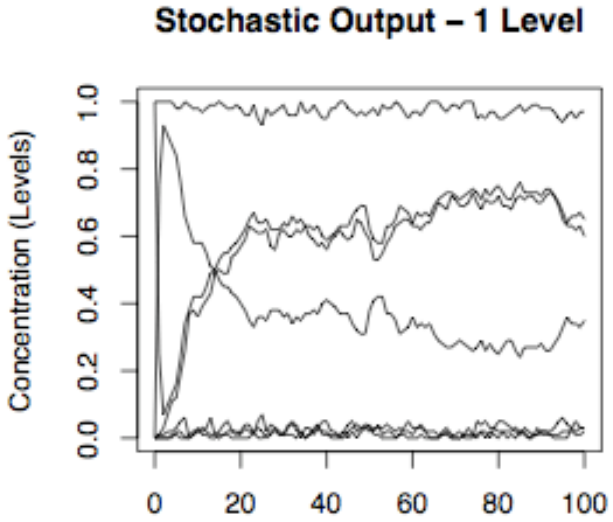


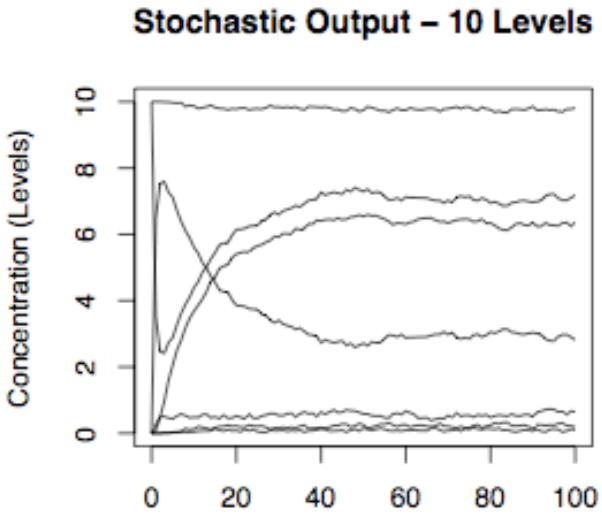
dm3/dt = + k1*m1*m2
+ k4*m4
- k2*m3
- k3*m3*m9

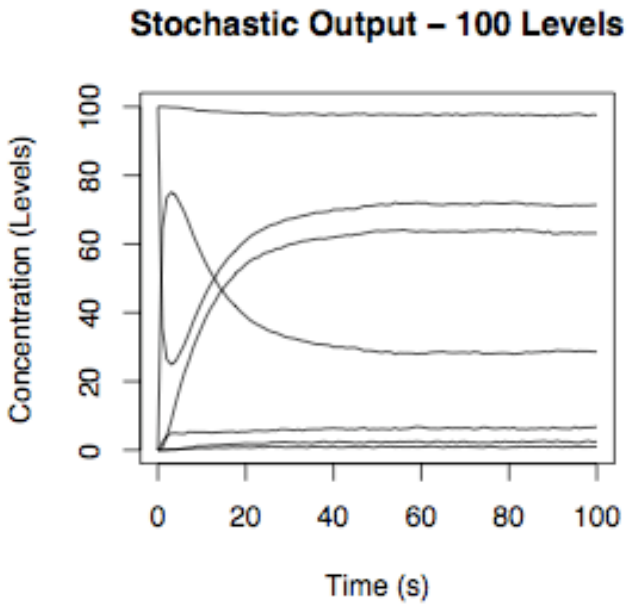


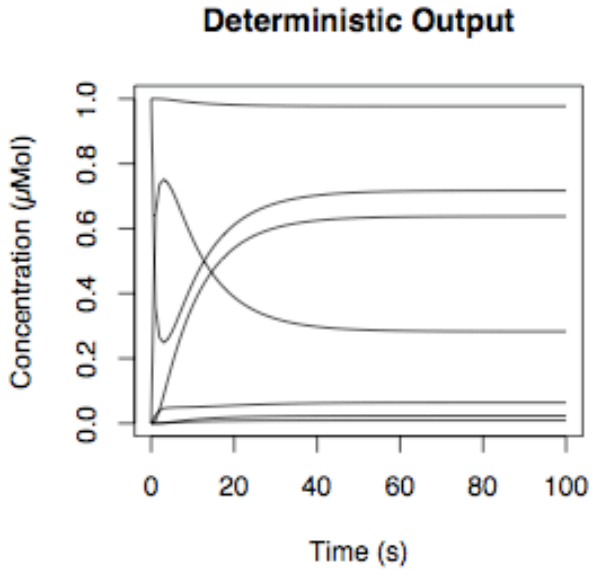

State8
Concentration (relative units)
Time (sec)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
10
20
30
40
50
60
70
80
90
100
Raf-1Star
RKIP
Raf-1Star_RKIP
Raf-1Star_RKIP_ERK-PP
ERK
RKIP-P
MEK-PP
MEK-PP_ERK
ERK-PP
RP
RKIP-P_RP

Stochastic & Deterministic Behaviour









BioModel Engineering

- The science of designing, constructing and analyzing computational models of biological systems
- A systematic and powerful extension of earlier mathematical modeling approaches
- Applied in systems biology and synthetic biology.
- 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 behavior,
  - 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, Robin Donaldosn (2008)
An introduction to BioModel Engineering, illustrated for signal transduction pathways, Proc WMC9, LNCS

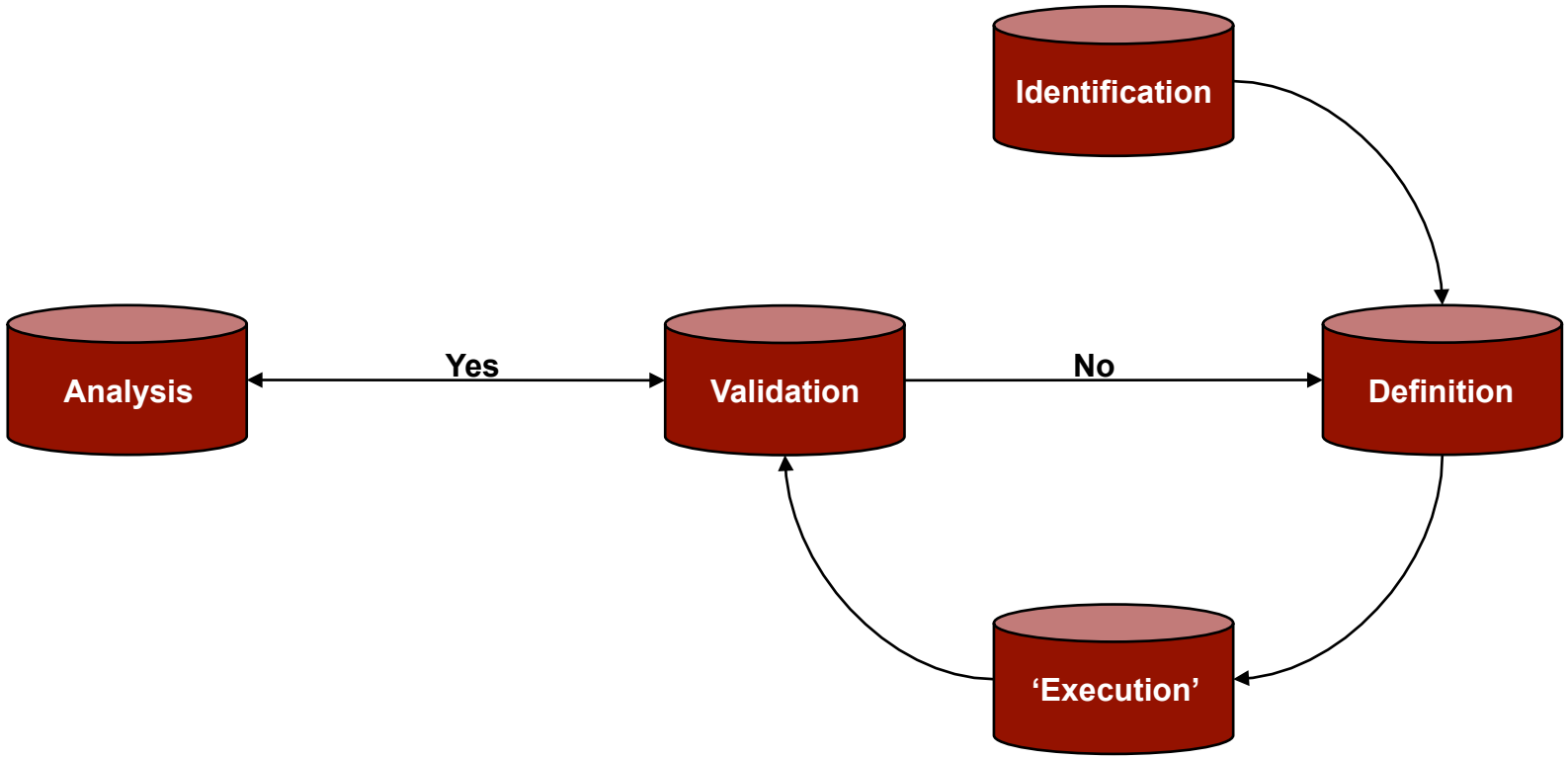
Why model?

- 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)

How to model…Overview



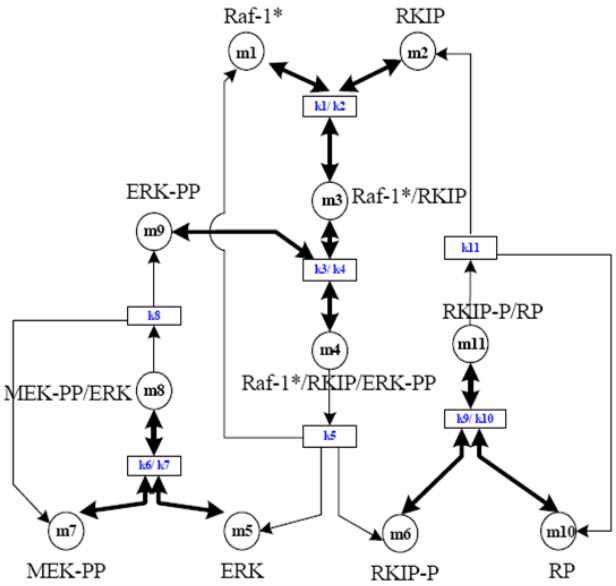
Building computational models

1. Identification
2. Construction
3. Execution, Animation, Simulation
4. Analysis & interpretation
5. Management & development

1: Identification

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

2: Construction



- **Construct topology**
  - Define all the proteins/molecules involved
  - Define the reactions they are involved in
  - Where do you draw the model boundary line?

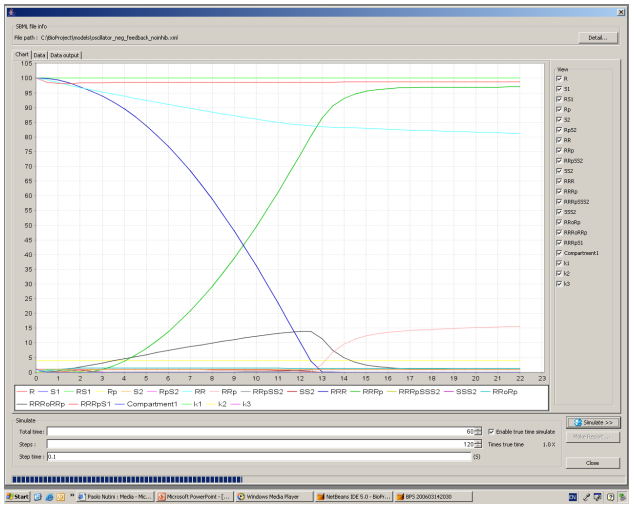
- **Check literature**
  - What is known about the pathway and proteins?
  - What evidence is there that protein A binds directly to protein B?
  - Protein C also binds directly to protein B: does it compete with protein A or do they bind to protein B at different sites?
  - Trust & Conflicts: it is important to recognize which evidence to trust and which to discard (talk to the people in the wet lab)

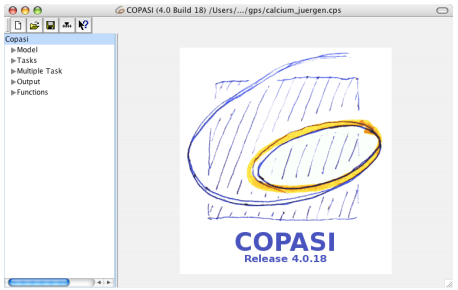
- **Simplifying assumptions**
  - Many biological processes are very complex and not fully understood
  - Therefore, developing a model often involves making simplifying assumptions
  - For example, the activation of Raf by Ras is very complicated and not fully understood but it is often modelled as:
    - Raf + Ras-GTP = Raf/Ras-GTP -> Raf-x + Ras-GTP
  - Although this is a simplification, it is able to explain the observed data

3: Execution, Animation, Simulation

- Relatively straightforward step - many software tools available for token game (animation), to simulate differential equation based models, stochastic models, …

- For example:
  - BioNessie
  - MatLab
  - Copsai / Gepasi
  - CellDesigner
  - Jarnac
  - WinScamp
  - SPiM
  - gillespie2
  - Snoopy
  - Many many more



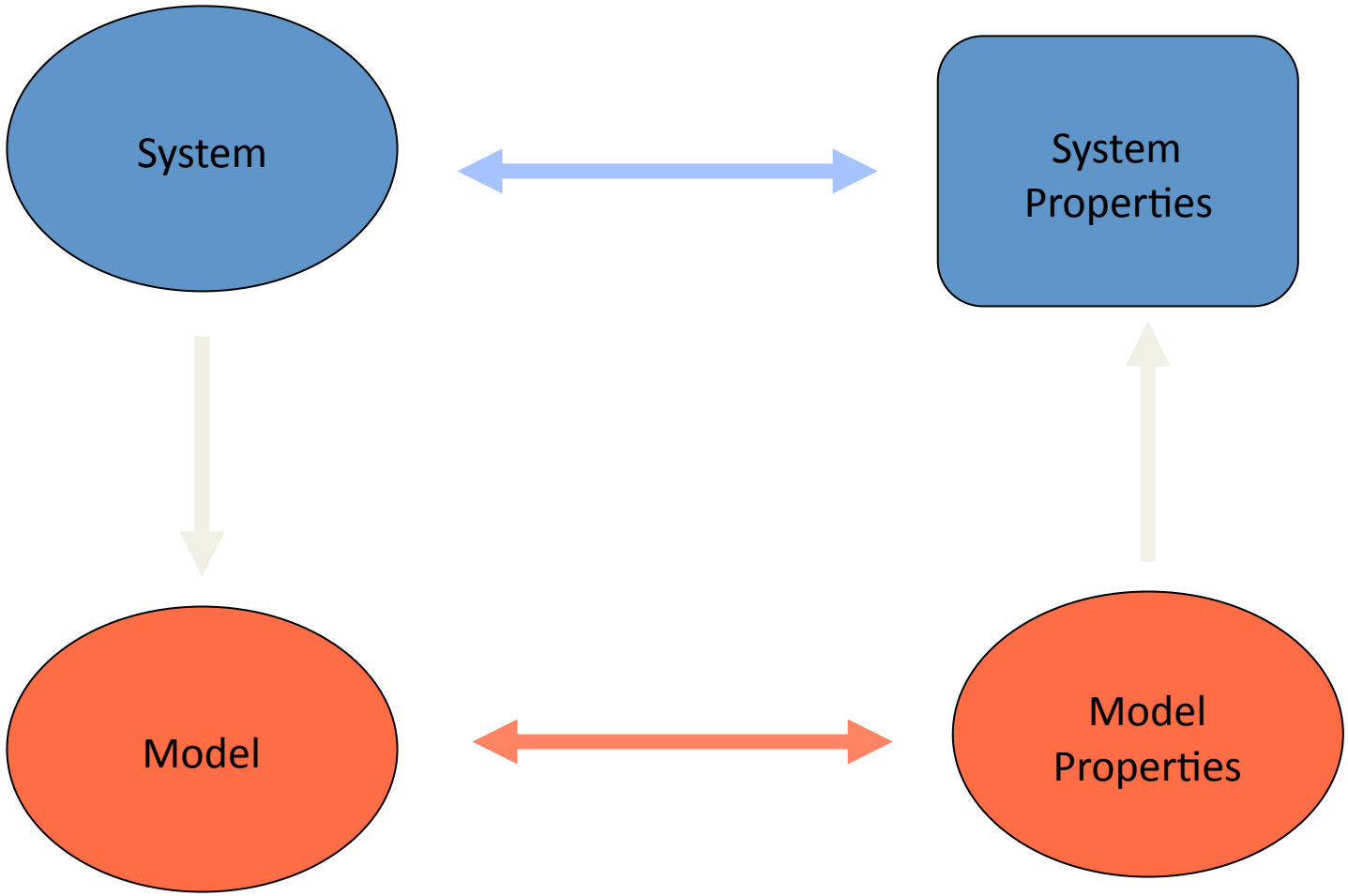


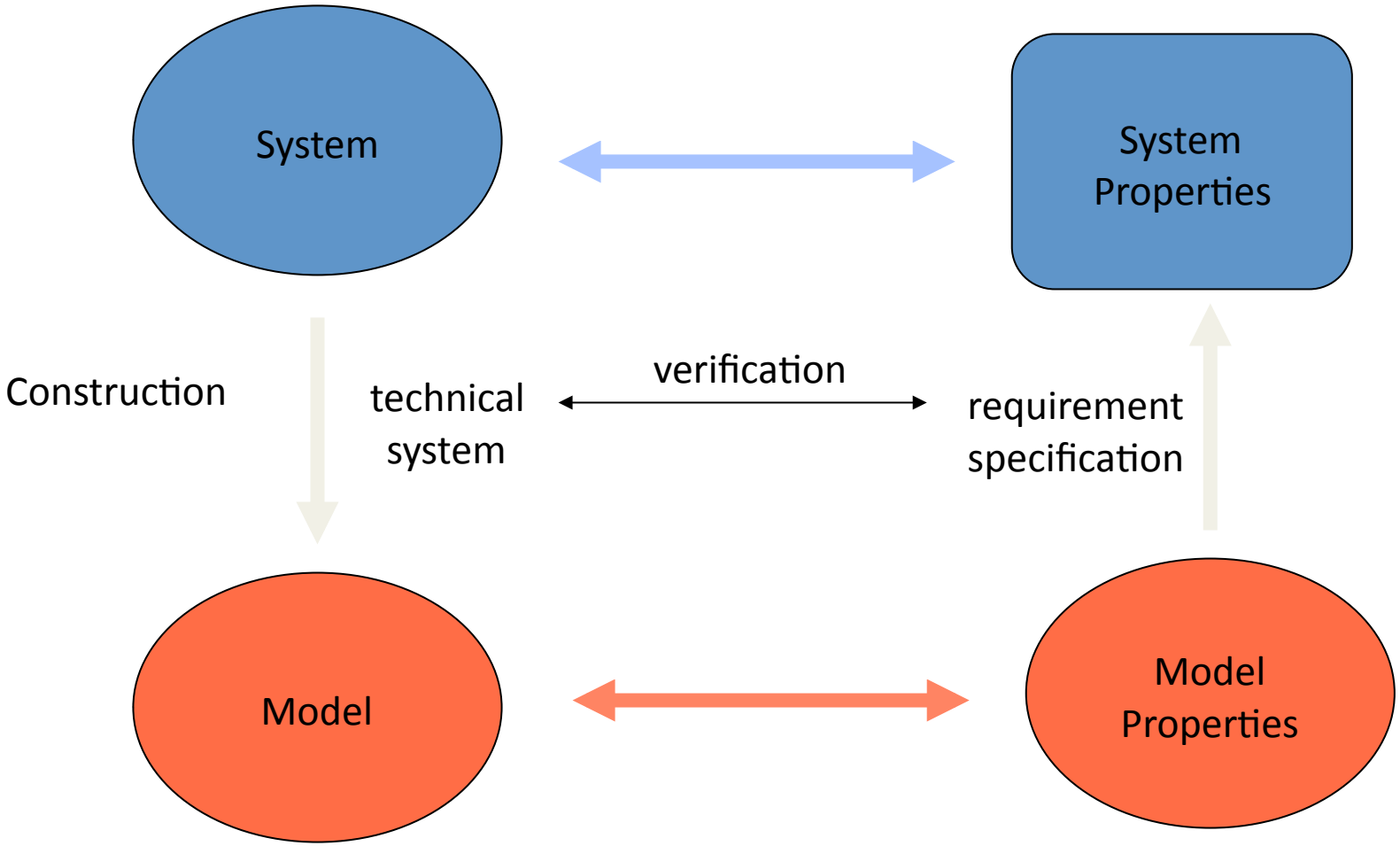
4: Analysis & Interpretation (1)

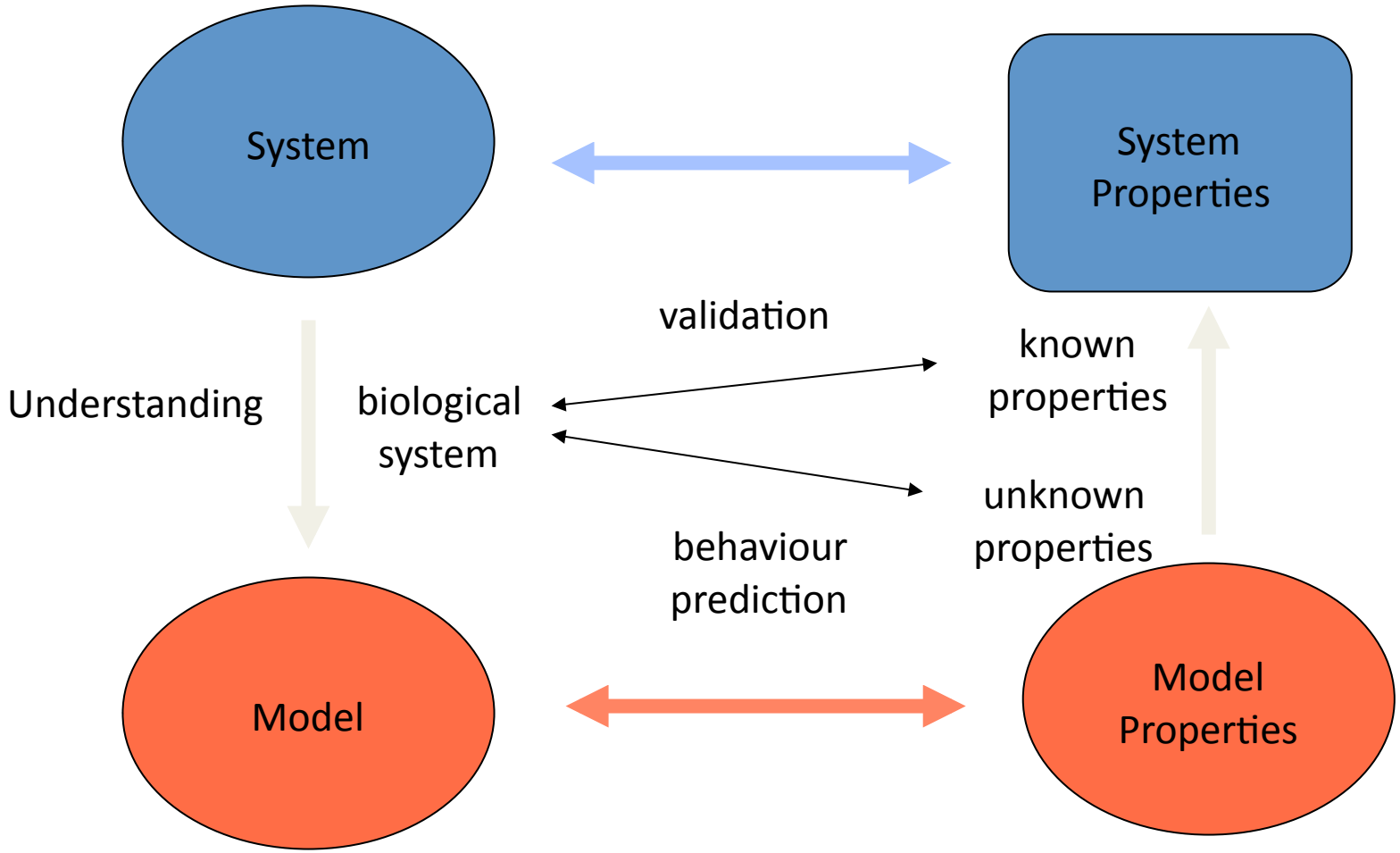
- **Validation**: Do the model results match the experimental data?
  - Yes: validation
  - No: back to definition and check for errors
    - Simple typos
    - Wrong kinetics
    - Over simplifications of processes
    - Missing components from the model
    - Incorrect parameter data

4: Analysis & Interpretation (2)

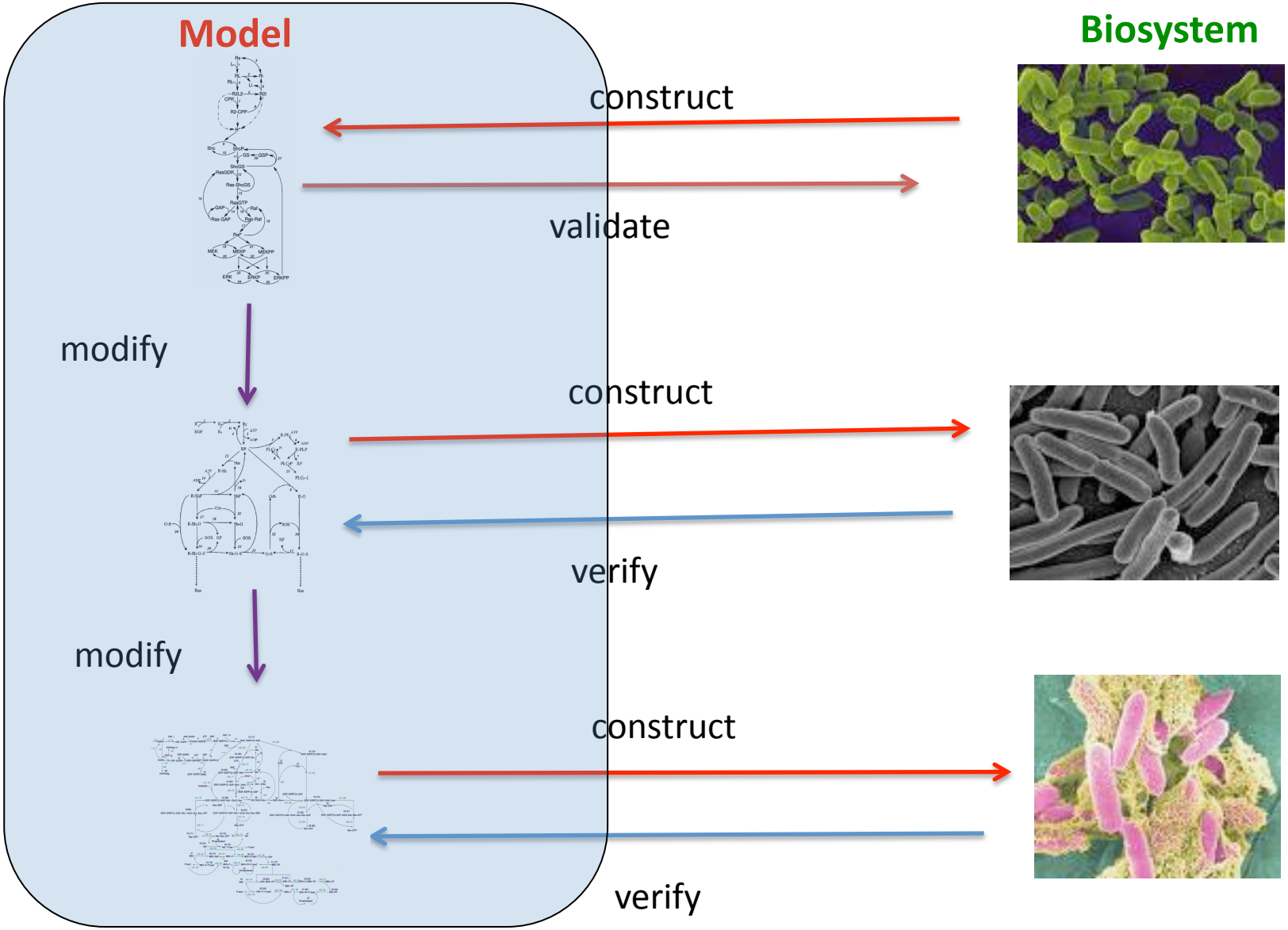
- **Prediction**:
  - What do the results imply or suggest? What do they tell us that is new and that we did not know/understand before?
  - What predictions can we make? (when modifying by drugs, conditions, knockouts, knockdowns,…)
- Sensitivity analysis - identify the key steps and components in the pathway as well as monitoring how robust the system is:
  - Vary an initial concentration or rate by a small amount and see what affect it has on the system as a whole: small changes in a key value are likely to have a large affect
  - How robust is the system to changes?
- Knockout/knockdown experiments
  - k/o components – (initial concentration := 0,…) to identify which components are essential and which are redundant
  - k/o reactions (rate := 0,…) to identify essential and redundant reactions in the system


System
System Properties
Model
Model Properties


System
System Properties
Construction
technical system
verification
requirement specification
Model
Model Properties


System
System Properties
validation
Understanding
biological system
known properties
unknown properties
behaviour prediction
Model
Model Properties

Synthetic Biology development cycle



5: Management & Development

- **Identifying building blocks / submodels (modules)**
- **Database**
  - models, model components
  - behaviours,
  - properties, …
- **Model Version control system**
- **Component reuse**
- **Model checking**:
  Maintaining (temporal logic) properties

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)

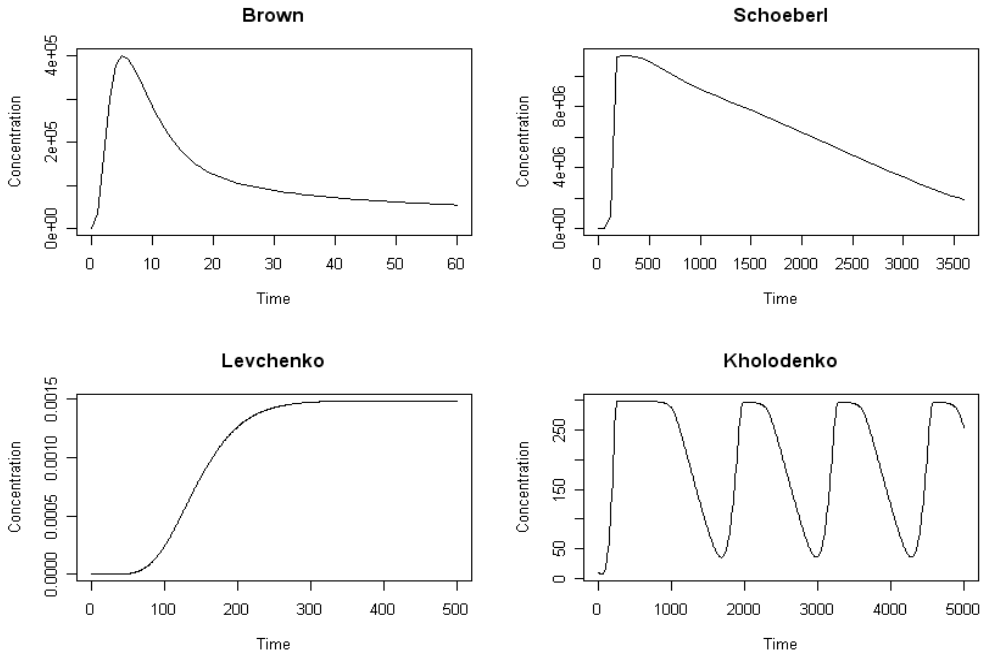
$P_{>=1}[ErkPP <= 0.50*max(ErkPP) \wedge (\ d(ErkPP) > 0\ )\ U$
$(\ G(ErkPP >= 0.99*max(ErkPP))\ )\ ]$

- ***Levchenko***

Oscillates at least 4 times

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

- **Kholodenko**



BioNessie –BioModel Engineering environment

Xuan Liu

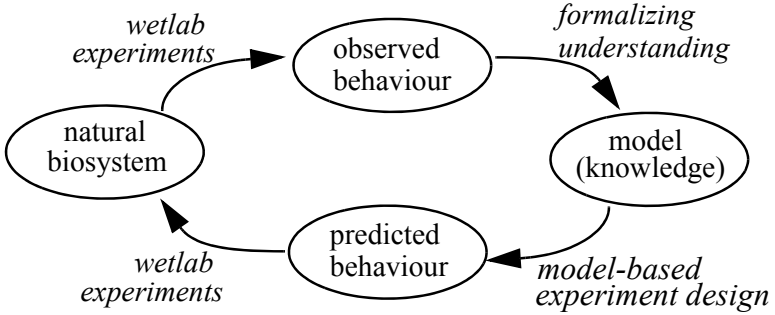
- SBML (Systems Biology Markup Language) enabled.
- Intuitive easy-to-use interface for biochemists & modellers. Input biochemical equations.
- File storage in XML, SBML, text & graphics
- Platform Independent – Java
- Parallel processing - Efficient exploitation of available compute resources – multiple core and multiple CPUs, as well as Grid computing
- Editor, simulator, and analyser
- Model version control
- Kinetic law library creation & management
- Fast efficient ODE solver (stiff & non-stiff)
- (Stochastic solver)
- Parameter scanning
- Sensitivity analysis
- Parameter estimation using a genetic algorithm
- Advanced model checking (MC2 using PLTL)
- Module composition
- Relational database connectivity

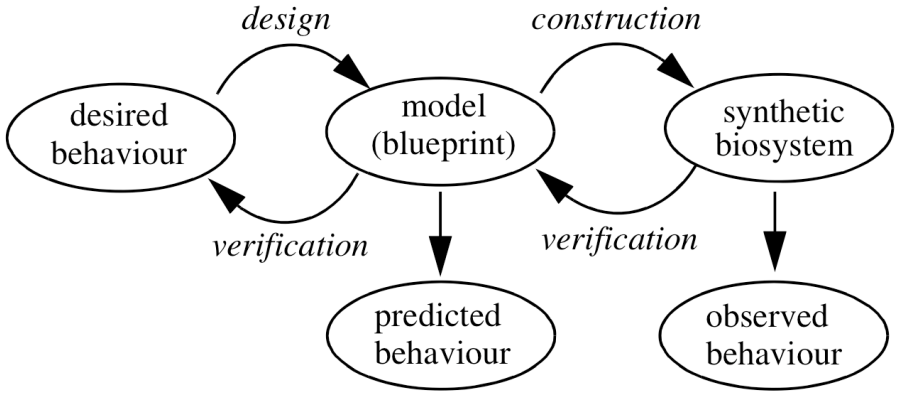
www.BioNessie.org

BME: Systems & Synthetic Biology

Systems Biology: design and construction of models of existing biological systems, which explain observed properties and predict the response to experimental interventions

Synthetic Biology: used as part of a general strategy for designing and constructing synthetic biological systems with novel functionalities.





Some references

David Gilbert, Rainer Breitling, Monika Heiner, Robin Donaldson (2008), **An introduction to BioModel Engineering, illustrated for signal transduction pathways**, Proc WMC9, LNCS Volume 5391, pp13-28

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, 2008 9(5):404-42

Xuan Liu, Jipu Jiang, Oluwafemi Ajayi, Xu Gu, David Gilbert, Richard Sinnott (2008), '**BioNessie(G) - A Grid Enabled Biochemical Networks Simulation Environment**'. Stud Health Technol Inform. IOS Press, 2008, 138: 147-157

David Gilbert, Monika Heiner, Susan Rosser, Rachael Fulton, Xu Gu and Maciej Trybilo (2008), **A Case Study in Model-driven Synthetic Biology. In Biologically Inspired Cooperative Computing**: BICC 2008. IFIP

David Gilbert, Monika Heiner and Sebastian Lehrack (2007). **A Unifying Framework for Modelling and Analysing Biochemical Pathways Using Petri Nets** CMSB 2007, LNCS/LNBI 4695, pp. 200-216.

David Gilbert, Hendrik Fuß, Xu Gu, Richard Orton, Steve Robinson, Vladislav Vyshemirsky, Mary Jo Kurth, C. Stephen Downes and Werner Dubitzky. (2006**) Computational methodologies for modelling, analysis and simulation of signalling networks**, Briefings in Bioinformatics 2006 7(4): 339-353; doi: 10.1093/bib/bbl043 Special Issue: Computational Methodologies for Systems Biology.